

IBM SPSS Text Analytics for Surveys
4.0.1 – Benutzerhandbuch



Anmerkung: Lesen Sie vor der Verwendung dieser Informationen und des zugehörigen Produkts die allgemeinen Informationen unter Hinweise auf S. 273.

Diese Ausgabe gilt für IBM® SPSS® Text Analytics for Surveys 4 und alle nachfolgenden Versionen und Abwandlungen, und zwar bis es in neuen Ausgaben anderweitig angegeben wird.

Screenshots von Adobe-Produkten nachgedruckt mit Genehmigung durch Adobe Systems Incorporated.

Screenshots von Microsoft-Produkten nachgedruckt mit Genehmigung durch Microsoft Corporation.

Lizenziertes Material – Eigentum von IBM

© **Copyright IBM Corporation 2004, 2011.**

Eingeschränkte Rechte für Mitarbeiter der US-Regierung – Benutzung, Duplizierung und Veröffentlichung beschränkt durch GSA ADP Schedule-Vertrag mit der IBM Corp.

Vorwort

Willkommen bei IBM® SPSS® Text Analytics for Surveys Version 4.0.1, einer Anwendung zur Umfragetextkodierung für eine sinnvolle Analyse von Freitext-Antworten. Mit diesem Produkt kann jeder, der Umfragen durchführt, unstrukturierte Antworten rasch in quantitative Daten umwandeln. Das “Erschließen” dieser offenen Textdaten kann die Qualität der Analyse und die Entscheidungsfindung erheblich verbessern. Mit dieser Anwendung können Sie Umfragedaten importieren, Schlüsselkonzepte extrahieren, die Ergebnisse verfeinern und Antworten kategorisieren. Sobald Ihre Daten kategorisiert sind, können Sie Ihre Kategorien zur weiteren Analyse und Diagrammerstellung zum Import in Tools zur quantitativen Analyse wie z. B. das IBM® SPSS® Statistics-System exportieren.

SPSS Text Analytics for Surveys kombiniert fortschrittliche linguistische Technologien zur zuverlässigen Extrahierung und Klassifizierung von Schlüsselkonzepten in Freitext-Antworten mit manuellen Techniken. Mithilfe robuster Algorithmen zur Kategorieerstellung und einfacher Drag&Drop-Funktionalität können Sie Kategorien oder “Codes” erstellen, in denen Ihre Umfrageantworten kategorisiert werden. Die erzeugten Kategorien können auch wiederverwendet werden, um konsistente Ergebnisse über dieselbe oder mehrere Studien hinweg zu erhalten. Da offene Antwortdaten von einer Umfrage zur nächsten stark variieren können, werden keine zwei Projekte exakt gleich sein. Jedoch können Sie davon ausgehen, dass Sie zur Fertigstellung Ihrer Analyse demselben grundlegenden Prozess folgen können. [Für weitere Informationen siehe Thema Der typische Prozess in Kapitel 2 auf S. 11.](#)

Informationen zu IBM Business Analytics

Die Software IBM Business Analytics liefert umfassende, einheitliche und korrekte Informationen, mit denen Entscheidungsträger die Unternehmensleistung verbessern können. Ein umfassendes Portfolio aus [Business Intelligence](#), [Vorhersageanalyse](#), [Finanz- und Strategiemangement](#) sowie [Analyseanwendungen](#) bietet Ihnen sofort klare und umsetzbare Einblicke in die aktuelle Leistung und gibt Ihnen die Möglichkeit, zukünftige Ergebnisse vorherzusagen. Durch umfassende Branchenlösungen, bewährte Vorgehensweisen und professionellen Service können Unternehmen jeder Größe die Produktivität maximieren, Entscheidungen automatisieren und bessere Ergebnisse erzielen.

Als Teil dieses Portfolios unterstützt IBM SPSS Predictive Analytics-Software Unternehmen dabei, zukünftige Ereignisse vorherzusagen und proaktiv Maßnahmen zu ergreifen, um bessere Geschäftsergebnisse zu erzielen. Kunden aus Wirtschaft, öffentlichem Dienst und dem Bildungsbereich weltweit nutzen IBM SPSS-Technologie als Wettbewerbsvorteil für Kundengewinnung, Kundenbindung und Erhöhung der Kundenumsätze bei gleichzeitiger Eindämmung der Betrugsmöglichkeiten und Minderung von Risiken. Durch die Einbindung von IBM SPSS-Software in ihre täglichen Operationen wandeln sich Organisationen zu “Predictive Enterprises” – die Entscheidungen auf Geschäftsziele ausrichten und automatisieren und einen messbaren Wettbewerbsvorteil erzielen können. Wenn Sie weitere Informationen wünschen oder Kontakt zu einem Mitarbeiter aufnehmen möchten, besuchen Sie die Seite <http://www.ibm.com/spss>.

Technischer Support

Kunden mit Wartungsvertrag können den technischen Support in Anspruch nehmen. Kunden können sich an den technischen Support wenden, wenn sie Hilfe bei der Arbeit mit IBM Corp.-Produkten oder bei der Installation in einer der unterstützten Hardware-Umgebungen benötigen. Zur Kontaktaufnahme mit dem technischen Support besuchen Sie die Website von IBM Corp. unter <http://www.ibm.com/support>. Beachten Sie, dass Sie nach Ihrem Namen, dem Namen Ihrer Organisation und Ihrer Supportvereinbarung gefragt werden, wenn Sie Unterstützung anfordern.

Teil I: Erste Schritte

1 Erläuterung der Textanalyse 1

Neuheiten	1
Offene Umfragedaten	2
Informationen zum Text-Mining	3
Wie Extrahierung funktioniert	4
Wie die Kategorisierung funktioniert	6
Vorbereitung für die Textanalyse	8
Zuverlässigkeit und Feineinstellung	8
Verfeinern von linguistischen Ressourcen	9
Verfeinern von Kategoriedefinitionen	9

2 Erste Schritte 11

Der typische Prozess	11
Das Textanalysefenster	12
Ansicht "Frage"	13
Die Ansicht "Gesamtes Projekt"	15
Das Fenster "Ressourceneditor"	16
Festlegen von Optionen	19
Optionen: Registerkarte "System"	19
Optionen: Registerkarte "Anzeige"	21
Optionen: Registerkarte "Klänge"	23
Optionen: Registerkarte "Übersetzung"	24
Microsoft Internet Explorer-Einstellungen für die Hilfe	25

Teil II: Textanalyse

3 Erstellen neuer Projekte und Pakete 27

Erstellen von Projekten	27
Ihre Daten vorbereiten	28
Starten neuer Projekte	29
Auswählen von Datenquellen	30

Auswählen von Variablen	35
Übersetzen ins Englische.	37
Auswählen von Kategorien und Ressourcen	39
Verwendung von Text Analysis Packages	43
Erstellung von Text Analysis Packages.	43
Aktualisierung von Text Analysis Packages	45

4 Arbeiten mit Projekten 49

Öffnen von Projekten	49
Bearbeiten von Projekteigenschaften	50
Anzeigen von Projektdaten	51
Sortieren von Variablen	52
Bearbeiten von Variableneigenschaften	52
Speichern von Projekten.	53
Exportieren von Kategorisierungsergebnissen	55
Exportieren in IBM SPSS Statistics oder IBM SPSS Data Collection.	56
Exportieren in Microsoft Excel.	58
Exportieren von Zusammenfassungsdiagrammen	60
Ändern von Datenquellen	63
Auswählen von Datenquellen	64
Auswählen von Variablen	69
Abgleichen von Variablen	71
Übersetzen ins Englische.	72
Aktualisieren von Daten	74
Übersetzen ins Englische	74
Gemeinsame Benutzung von Projekten	76
Kennzeichnen von Antworten	76
Projektstatusleiste	77

5 Daten extrahieren 80

Extrahierte Ergebnisse: Konzepte, Typen und Muster	80
Daten extrahieren.	85
Extrahierungsergebnisse speichern	88
Extrahierungsergebnisse verfeinern	88
Synonyme hinzufügen	89
Konzepte zu Typen hinzufügen	91

Konzepte von der Extrahierung ausschließen	94
Extrahierung von Wörtern erzwingen	95

6 Kategorisieren von Textdaten 96

Der Fensterbereich "Kategorien"	97
Der Fensterbereich "Daten"	100
Kategorierelevanz.	102
Methoden und Strategien zur Erstellung von Kategorien	104
Methoden für die Kategorieerstellung	104
Strategien für die Kategorieerstellung	104
Tipps zur Erstellung von Kategorien	105
Auswahl der besten Deskriptoren	107
Erläuterung von Kategorien	110
Kategorieeigenschaften.	111
Erstellen von Kategorien.	112
Erweiterte linguistische Einstellungen	116
Über linguistische Verfahren	121
Erweiterte Einstellungen für Häufigkeit	127
Erweitern von Kategorien	129
Manuelle Erstellung von Kategorien	134
Erstellen neuer Kategorien bzw. Umbenennen von Kategorien.	134
Erstellen von Kategorien durch Ziehen und Ablegen (Drag&Drop)	135
Importieren und Exportieren von vordefinierten Kategorien.	136
Importieren vordefinierter Kategorien	136
Exportieren von Kategorien	145
Verwenden von Kategorieregeln.	148
Kategorieregelsyntax.	149
Verwenden von TLA-Mustern in Kategorieregeln.	151
Platzhalter in Kategorieregeln	153
Beispiele für Kategorieregeln	155
Erstellen von Kategorienregeln	157
Bearbeiten und Löschen von Regeln	159
Bearbeiten und Verfeinern von Kategorien	159
Bearbeiten von Kategorieeigenschaften	160
Hinzufügen von Deskriptoren zu Kategorien.	161
Bearbeiten von Kategoriedeskriptoren.	162
Verschieben von Kategorien	163
Zusammenführen bzw. Kombinieren von Kategorien	164
Einschließen von Antworten in Kategorien erzwingen	165
Textübereinstimmungen in Kategorien	167

Kopieren von Kategorien	168
Drucken von Kategorien	169
Löschen von Kategorien	170

7 Visualisierung von Diagrammen 172

Kategoriebalkendiagramm	173
Kategorienetzdiagramm	174
Tabelle für Kategorienetzdiagramm.	175
Verwenden von Diagrammsymboleisten und Paletten	175
Bearbeiten von Visualisierungen	176
Allgemeine Regeln zur Bearbeitung von Visualisierungen	177
Bearbeiten und Formatieren von Text.	179
Ändern von Farben, Mustern, Strichmustern und Transparenz	179
Drehen und Ändern der Form und des Verhältnisses von Punktelementen	181
Ändern der Größe von Grafikelementen	182
Festlegen von Rändern und Abständen	182
Formatieren von Zahlen	183
Ändern der Einstellungen für Achsen und Skalen.	184
Bearbeiten von Kategorien	186
Ändern der Orientierung von Feldern	188
Transformieren des Koordinatensystems	188
Ändern von Statistiken und Grafikelementen	190
Ändern der Position der Legende.	193
Kopieren von Visualisierungen und Visualisierungsdaten.	193
Tastaturkombinationen	194

Teil III: Ressourceneditor

8 Vorlagen und Ressourcen 196

Die Editoroberfläche	197
Erstellen und Aktualisieren von Vorlagen	200
Wechseln von Ressourcenvorlagen	201
Vorlagen verwalten	202
Importieren und Exportieren von Vorlagen	203
Ressourcen sichern	205
Ressourcendateien importieren	206

9 Mit Bibliotheken arbeiten **209**

Mitgelieferte Bibliotheken	209
Bibliotheken erstellen	211
Öffentliche Bibliotheken hinzufügen	211
Fachausdrücke und Typen suchen	212
Bibliotheken anzeigen	213
Local Libraries verwalten	213
Local Libraries umbenennen	214
Local Libraries deaktivieren	214
Local Libraries löschen	215
Public Libraries verwalten	215
Bibliotheken gemeinsam nutzen	217
Bibliotheken veröffentlichen	219
Bibliotheken aktualisieren	219
Konflikte auflösen	220

10 Informationen zu Bibliothekwörterbüchern **223**

Typ-Wörterbücher	223
Integrierte Typen	224
Erstellen von Typen	225
Hinzufügen von Fachausdrücken	227
Erzwingen von Fachausdrücken	231
Umbenennen von Typen	232
Verschieben von Typen	233
Deaktivieren und Löschen von Typen	234
Substitutions-/Synonymwörterbücher	234
Definieren von Synonymen	236
Definieren optionaler Elemente	238
Deaktivieren und Löschen von Substitutionen	239
Ausschlusswörterbücher	240

11 Informationen zu erweiterten Ressourcen **242**

Suchen	243
Ersetzen	244
Unschärfe Gruppierung	245

Nicht linguistische Elemente	246
Definitonen regulärer Ausdrücke	247
Normalisierung	250
Konfiguration.	250
Sprachbehandlung	252
Extraktionsmuster	252
Erzwungene Definitionen	252
Abkürzungen	253

Anhänge

A Ausnahmen bei japanischem Text 255

Extrahieren und Kategorisieren von japanischem Text	255
Wie Extrahierung funktioniert	255
Wie sekundäre Extrahierung funktioniert.	257
Wie die Kategorisierung funktioniert	260
Bearbeitungsressourcen für japanischen Text	260
Japanischer Bibliotheksbaum, Typen und Fachausdrucksbereich	262
Verfügbare Typen für japanischen Text	265
Bearbeiten von japanischen Typeigenschaften	269
Verwenden des Synonymwörterbuchs für japanischen Text.	269
Validieren und Kompilieren von japanischen Ressourcen.	271
Weitere Ausnahmen für Japanisch.	271

B Hinweise 273

Index 276

Teil I: Erste Schritte

Erläuterung der Textanalyse

Willkommen bei IBM® SPSS® Text Analytics for Surveys Version 4.0.1, einer Anwendung zur Umfragetextkodierung für eine sinnvolle Analyse von Freitext-Antworten. Mit diesem Produkt kann jeder, der Umfragen durchführt, unstrukturierte Antworten rasch in quantitative Daten umwandeln. Das “Erschließen” dieser offenen Textdaten kann die Qualität der Analyse und die Entscheidungsfindung erheblich verbessern. Mit dieser Anwendung können Sie Umfragedaten importieren, Schlüsselkonzepte extrahieren, die Ergebnisse verfeinern und Antworten kategorisieren. Sobald Ihre Daten kategorisiert sind, können Sie Ihre Kategorien zur weiteren Analyse und Diagrammerstellung zum Import in Tools zur quantitativen Analyse wie z. B. das IBM® SPSS® Statistics-System exportieren.

SPSS Text Analytics for Surveys kombiniert fortschrittliche linguistische Technologien zur zuverlässigen Extrahierung und Klassifizierung von Schlüsselkonzepten in Freitext-Antworten mit manuellen Techniken. Mithilfe robuster Algorithmen zur Kategorieerstellung und einfacher Drag&Drop-Funktionalität können Sie Kategorien oder “Codes” erstellen, in denen Ihre Umfrageantworten kategorisiert werden. Die erzeugten Kategorien können auch wiederverwendet werden, um konsistente Ergebnisse über dieselbe oder mehrere Studien hinweg zu erhalten. Da offene Antwortdaten von einer Umfrage zur nächsten stark variieren können, werden keine zwei Projekte exakt gleich sein. Jedoch können Sie davon ausgehen, dass Sie zur Fertigstellung Ihrer Analyse demselben grundlegenden Prozess folgen können. [Für weitere Informationen siehe Thema Der typische Prozess in Kapitel 2 auf S. 11.](#)

Neuheiten

Sie finden die folgenden neuen Funktionen in IBM® SPSS® Text Analytics for Surveys 4.0.1:

Hierarchische Kategorien. Kategorien können auch eine hierarchische Struktur besitzen, d. h., sie können Unterkategorien enthalten, die wiederum eigene Unterkategorien enthalten können usw. Sie können vordefinierte Kategorien, früher Coderahmen genannt, mit hierarchischen Kategorien importieren und diese hierarchischen Kategorien auch im Produkt aufbauen.

Hierarchische Kategorien ermöglichen Ihnen den Aufbau einer Baumstruktur mit einer oder mehreren Unterkategorien, die eine genauere Gruppierung von Elementen, beispielsweise verschiedenen Konzept- oder Themenbereichen, gestatten. Ein einfaches Beispiel könnte sich auf Freizeitaktivitäten beziehen: Bei der Beantwortung einer Frage wie *Welcher Aktivität würden Sie sich gerne widmen, wenn Sie mehr Zeit hätten?* könnten die Kategorien der ersten Ebene *Sport, Kunst und Handarbeiten, Angeln* usw. lauten. Auf der Ebene unter *Sport* könnten Sie Unterkategorien einrichten, um zu sehen, ob es sich um *Ballsportarten, Wassersportarten* usw. handelt.

Zugriff auf Language Weaver. Der Zugriff auf die Übersetzungsoberfläche von Language Weaver wurde vereinfacht. Es sind nur noch eine einzige URL und die zugehörigen Sicherheitsdetails erforderlich.

Offene Umfragedaten

Umfragen enthalten in der Regel unterschiedliche Arten von Fragen einschließlich offener und geschlossener Fragen. Die erste, eine **geschlossene Frage**, präsentiert einen eingeschränkten Satz an Antworten, der verschiedene Arten von quantitativen Analysen gestattet. Die zweite, eine **offene Frage**, gestattet einem Antwortenden die Angabe einer unstrukturierten Antwort in variierender Länge und Ausführlichkeit.

Die Wörter, mit denen Menschen eine Frage beantworten, teilen viel über ihre Gedanken und Gefühle mit. Deshalb sind offene Fragen häufig in Umfragen enthalten: Sie bieten vielfältigere und strukturiertere Informationen als geschlossene Fragen und können häufig Einblicke bieten, die vom Gestalter der Umfrage gar nicht vorgesehen waren.

Jedoch verbot sich traditionell aus Kostengründen die Verwendung einer großen Anzahl an offenen Fragen, da mit der Interpretation der Antworten ein großer analytischer Aufwand einherging. Zudem müssen diese langen Antworten anhand von detaillierten Kodierungsanweisungen auf standardisierte Weise kodiert werden. Diese Aufgabe kann schwierig sein, da Kodierer häufig uneinig darin sind, wie bestimmte Antworten kategorisiert werden sollen. Wenn Kodierer unterschiedlicher Ansicht sind, verringert sich die Zuverlässigkeit der Ergebnisse. Aus all diesen Gründen wurde die Kodierung von offenen Fragen lange als eine Art der Befragung betrachtet, die zeitraubend und kostspielig ist und in keinem Verhältnis zum Nutzen der gesammelten Daten steht.

IBM® SPSS® Text Analytics for Surveys bietet eine Alternative zu diesem kostspieligen Verfahren, da es die Kodierung von offenen Fragen in einem Bruchteil der Zeit erledigen kann, die für die manuelle Bearbeitung erforderlich wäre. Mithilfe von fortschrittlicher linguistischer Theorie und Technologie analysiert SPSS Text Analytics for Surveys offenen Antworttext als Menge an Phrasen und Sätzen, deren grammatikalische Struktur einen Kontext für die Bedeutung einer Antwort liefert. Nach der Analyse dieses Textes werden die Schlüsselkonzepte und Muster extrahiert und in Kategorien klassifiziert.

Sie können integrierte Kategorieerstellungstechniken nutzen, um automatisch Kategorien zu erstellen, und manuelle Techniken, um die Ergebnisse feiner einzustellen. Die Zuverlässigkeit der Ergebnisse steigt dynamisch, da Extrahierung und Kategorisierung stets konsistent und wiederholbar ausgeführt werden: Dieselbe Antwort wird jedesmal denselben Kategorien zugeordnet, es sei denn, Sie wählen eine Feineinstellung Ihrer Kategoriedefinitionen oder linguistischen Bibliotheken.

Eine erfolgreiche Umfrageanalyse hängt nicht nur von einer einzigen Methode ab. Die subjektive Natur der Interpretation einer offenen Antwort verlangt den Einsatz mehrerer Techniken. Neben seinen integrierten Extrahierungs- und Kategorieerstellungstechniken verlässt sich SPSS Text Analytics for Surveys auch auf das Verständnis des Benutzers für die Ziele der spezifischen Textanalyse für jede Umfrage. Textanalyse ist am leistungsstärksten, wenn sie regelmäßig ausgeführt wird (Extrahieren, Überprüfen, Verfeinern, Erneut extrahieren), und ihre Nützlichkeit hängt häufig von der Menge an Zeit und Aufwand für die manuelle Überprüfung und die Verfeinerung der Extrahierungsergebnisse und Kategoriedefinitionen ab. [Für weitere Informationen siehe Thema Zuverlässigkeit und Feineinstellung auf S. 8.](#)

Wenn Sie in wiederkehrenden Umfragen mit identischen und ähnlichen Fragen arbeiten, können Sie Kategorien in anderen Fragen oder Projekten wiederholt verwenden. Die Wiederverwendung von Kategorien ermöglicht eine größere Konsistenz bei der Kodierung und bietet zudem immense Einsparungen an Zeit und Aufwand.

Zusätzlich sollten Sie auch andere Analysen ausführen. Die Kategorien, die Sie erzeugen, können in zahlreichen Typen von statistischen Analysen mit den übrigen Fragen im Fragebogen oder anderen demographischen Daten benutzt werden, um weitere Einblicke in die Umfrageteilnehmer und ihre Meinungen und Verhaltensweisen zu vermitteln. Nach der Verwendung von SPSS Text Analytics for Surveys zur Ermittlung der Kategorien, die einer Gruppe von Antworten zugrunde liegen, können Sie die Kategorien auch zur weiteren quantitativen Analyse in ein anderes Programm exportieren, z. B. in IBM® SPSS® Statistics Base.

Informationen zum Text-Mining

Bei der Textanalyse als Form einer qualitativen Analyse werden nützliche Daten aus Texten extrahiert (wie z. B. offene Antworten), so dass die Schlüsselbegriffe und Konzepte, die im betreffenden Text enthalten sind, unter einer angemessenen Zahl von Kategorien zusammengefasst werden können. Textanalysen können auf Texte aller Arten und Längen angewendet werden. Allerdings unterscheiden sich die jeweiligen Analyseansätze dabei ein wenig.

Kürzere Datensätze lassen sich am leichtesten kategorisieren, da sie eine geringere Komplexität aufweisen und für gewöhnlich weniger mehrdeutige Wörter und Antworten enthalten. Wenn Personen beispielsweise im Rahmen einer Umfrage mit offenen Antworten nach ihren drei Lieblingsaktivitäten im Urlaub gefragt werden, sind möglicherweise viele kurze Antworten zu erwarten, etwa: *an den Strand gehen*, *Nationalparks besuchen* oder *Nichtstun*. Längere offene Antworten können dagegen ziemlich komplex und weitschweifig ausfallen, besonders bei Befragten, die gebildet und motiviert sind und genug Zeit für das Ausfüllen eines Fragebogens zur Verfügung haben. Bei Umfragen zu den politischen Überzeugungen von Personen oder bei einem langen Blog-Feed zum Thema Politik sind möglicherweise längere Kommentare zu allerlei Fragen und Positionen zu erwarten.

Umfrageforscher analysieren für gewöhnlich keine sehr langen Antworten. Antworten im Rahmen der meisten Umfragen sind normalerweise kurz bis mittellang (zwischen einem Satz und einem kurzen Absatz). IBM® SPSS® Text Analytics for Surveys wurde für diese Textlänge konzipiert, ist aber auch in der Lage dazu, Antworten zu analysieren, die deutlich länger sind.

Es gibt verschiedene Methoden der Textanalyse. Zunächst einmal gibt es den manuellen Ansatz: Durchlesen der Umfrage-Antworten, Festhalten des Inhalts, Feststellung der wichtigsten Konzepte, die darin enthalten sind, und Zuweisung von Codes für diese Konzepte. Da Menschen über ein gutes Textverständnis verfügen, ist diese Methode ziemlich akkurat. Andererseits ist sie zeitaufwändig, arbeitsintensiv und, angesichts des immensen Textumfangs, der heute vorliegt, zunehmend unpraktisch. Sie ist außerdem in hohem Maße von der Interpretation der einzelnen Personen abhängig.

Eine andere Methode ist die Verwendung automatisierter Lösungen. Viele verschiedene automatisierte Lösungen einschließlich statistischer und linguistischer Lösungen stehen zur Auswahl. SPSS Text Analytics for Surveys bietet eine Kombination aus automatisierten linguistischen und statistischen Methoden, mit denen sich für jede Phase des Projekts die zuverlässigsten Ergebnisse erzielen lassen. In diesem Produkt werden linguistisch basierte Methoden zur automatischen Extrahierung der wichtigsten Konzepte aus den Antworten eingesetzt; zur Erstellung der Kategoriedefinitionen (Codes), die Antworten zugewiesen werden, können sowohl linguistische als auch statistische Methoden eingesetzt werden.

Wie Extrahierung funktioniert

Während der Extrahierung von Schlüsselkonzepten und -begriffen aus Ihren Antworten wird bei IBM® SPSS® Text Analytics for Surveys die linguistisch basierte Textanalyse angewendet. Diese Methode bietet die Geschwindigkeit und Kosteneffektivität von statistisch basierten Systemen. Sie ermöglicht jedoch ein weitaus höheres Maß an Genauigkeit, während ein deutlich geringeres Maß an Eingriffen seitens des Benutzers erforderlich ist. Linguistisch basierte Textanalyse baut auf einem Forschungsgebiet namens Verarbeitung natürlicher Sprache auf, das auch als Computerlinguistik bekannt ist.

Zur Veranschaulichung des Unterschieds zwischen statistikbasierten und linguistisch basierten Ansätzen beim Extrahierungsprozess dient die Überlegung, wie der jeweilige Ansatz auf eine Abfrage zum Begriff `Dokumentreproduktion` reagieren würde. Sowohl bei den statistikbasierten als auch bei den linguistisch basierten Lösungen müsste eine Begriffserweiterung für das Wort `Reproduktion` erfolgen, damit auch Synonyme wie `Kopie` und `Vervielfältigung` berücksichtigt werden. Andernfalls werden relevante Informationen übersehen. Wenn jedoch bei einer statistikbasierten Lösung eine derartige Synonymik—die Suche nach anderen Termen mit derselben Bedeutung—angewendet werden soll, wird wahrscheinlich auch der Term `Geburt` berücksichtigt, was zur Erzeugung einer Reihe von irrelevanten Ergebnissen führt. Das Verstehen von Sprache beseitigt die Mehrdeutigkeit von Texten, weshalb das linguistisch basierte Text-Mining definitionsgemäß den verlässlicheren Ansatz darstellt.

Wenn Sie verstehen, wie der Extrahierungsprozess funktioniert, fällt es Ihnen leichter, bei der Feinabstimmung Ihrer linguistischen Ressourcen (Bibliotheken, Typen, Synonyme und anderer) zentrale Entscheidungen zu treffen. Der Extrahierungsprozess umfasst folgende Schritte:

- Konvertieren von Quelldaten in ein Standardformat
- Ermittlung von Kandidaten
- Ermittlung von Äquivalenzklassen und Integration von Synonymen
- Zuweisung eines Typs
- Indexerstellung
- Musterabgleich und Ereignisextrahierung

Schritt 1. Konvertieren von Quelldaten in ein Standardformat

Im ersten Schritt werden die importierten Daten in ein einheitliches Format konvertiert, das für weitergehende Analysen genutzt werden kann. Diese Konvertierung erfolgt intern. Ihre Ausgangsdaten werden dabei nicht geändert.

Schritt 2. Ermittlung von Kandidaten

Es ist wichtig zu verstehen, welche Rolle die linguistischen Ressourcen während der linguistischen Extrahierung bei der Ermittlung von Kandidaten spielen. Linguistische Ressourcen kommen jedes Mal zum Einsatz, wenn ein Extrahierungsvorgang ausgeführt wird. Sie liegen in Form von Vorlagen, Bibliotheken und zusammengestellten Ressourcen vor. Bibliotheken bestehen aus Wortlisten, Beziehungen und weiteren Informationen, die eingesetzt werden, um die Extrahierung abzustimmen oder zu spezifizieren. Die zusammengestellten Ressourcen können nicht angezeigt oder bearbeitet werden. Die übrigen Ressourcen (Vorlagen) können jedoch im Resource Editor bearbeitet werden.

Zusammengestellte Ressourcen sind interne Kernkomponenten der Extraktor-Engine in IBM® SPSS® Text Analytics for Surveys. Diese Ressourcen umfassen ein allgemeines Wörterbuch, in dem eine Liste von Grundformen mit einem Code für die Wortart (Part of Speech) enthalten ist (Nomen, Verb, Adjektiv, Adverb, Partizip, Koordinator, Determinator oder Präposition). Die Ressourcen beinhalten auch belegte integrierte Typen, die verwendet werden, um den folgenden Typen eine Vielzahl von extrahierten Termen zuzuweisen: <Location>, <Organization> oder <Person>. [Für weitere Informationen siehe Thema Integrierte Typen in Kapitel 10 auf S. 224.](#)

Zusätzlich zu diesen zusammengestellten Ressourcen sind auch mehrere Bibliotheken im Lieferumfang enthalten. Diese können verwendet werden, um die Typen und Konzeptdefinitionen der zusammengestellten Ressourcen zu ergänzen und weitere Typen und Synonyme zu liefern. Diese Bibliotheken—sowie sämtliche benutzerdefinierte Bibliotheken, die Sie erstellen—bestehen aus mehreren Wörterbüchern. Diese umfassen Typ-Wörterbücher, Substitutionswörterbücher (Synonyme und optionale Elemente) sowie Ausschlusswörterbücher. [Für weitere Informationen siehe Thema Mit Bibliotheken arbeiten in Kapitel 9 auf S. 209.](#)

Sobald die Daten importiert und konvertiert wurden, beginnt die Extrahierungsengine, Kandidaten für die Extrahierung zu identifizieren. Kandidaten sind Wörter oder Wortgruppen, die verwendet werden, um Konzepte im Text zu ermitteln. Während der Verarbeitung des Texts werden einzelne Wörter (**Uniterms**), die nicht in den zusammengestellten Ressourcen enthalten sind, als Kandidaten für die Extrahierung betrachtet. Kandidaten, die aus zusammengesetzten Wörtern bestehen (**Multiterms**), werden über Extraktoren auf der Grundlage von Wortklasse-Mustern (Part of Speech) ermittelt. Der Multiterm `Sportwagen`, der dem Wortklasse-Muster “Adjektiv-Nomen” entspricht, besteht beispielsweise aus zwei Komponenten. Der Multiterm `schneller Sportwagen`, der dem Wortklasse-Muster “Adjektiv-Adjektiv-Nomen” entspricht, besteht aus drei Komponenten.

Anmerkung: Die Terme aus dem oben genannten zusammengestellten allgemeinen Wörterbuch stellen eine Liste aller Wörter dar, die als Uniterms wahrscheinlich uninteressant sind oder sprachliche Mehrdeutigkeiten aufweisen. Diese Wörter werden von der Extrahierung ausgeschlossen, wenn die Uniterms ermittelt werden. Sie werden jedoch erneut ausgewertet, wenn Wortarten bestimmt oder längere zusammengesetzte Wörter (Multiterms) als Kandidaten geprüft werden.

Abschließend wird ein bestimmter Algorithmus für die Verarbeitung von Zeichenfolgen verwendet, die aus Großbuchstaben bestehen (z. B. bei Berufsbezeichnungen), so dass diese speziellen Muster extrahiert werden können.

Schritt 3. Ermittlung von Äquivalenzklassen und Integration von Synonymen

Im Anschluss an die Ermittlung von Uniterms und Multiterms, die als Kandidaten infrage kommen, werden diese über eine Reihe von Algorithmen der Software miteinander verglichen und Äquivalenzklassen ermittelt. Bei einer Äquivalenzklasse handelt es sich um eine Grundform eines Ausdrucks oder einer einzelnen Form von zwei Varianten desselben Ausdrucks. Ausdrücke werden Äquivalenzklassen zugeordnet, damit beispielsweise die Begriffe `Unternehmensleiter` und `Leiter des Unternehmens` nicht als unterschiedliche Konzepte betrachtet werden. Um festzustellen, welches Konzept für die betreffende Äquivalenzklasse als Hauptterm verwendet wird, —`Leiter des Unternehmens` oder `Unternehmensleiter`, werden die folgenden Regeln in der aufgeführten Reihenfolge durch die Extrahierungsengine angewendet:

- Die vom Benutzer festgelegte Form in einer Bibliothek.

- Die Form, die im gesamten Textkörper am häufigsten vorkommt.
- Die kürzeste Form im gesamten Textkörper (die normalerweise der Grundform entspricht).

Schritt 4. Zuweisen eines Typs

Anschließend werden den extrahierten Konzepten Typen zugewiesen. Bei einem Typ handelt es sich um Konzepte, die nach semantischen Gesichtspunkten gruppiert werden. Für diesen Schritt werden sowohl zusammengestellte Ressourcen als auch die Bibliotheken verwendet. Zu den Typen gehören beispielsweise übergeordnete Konzepte, positive und negative Wörter, Vornamen, Orte, Organisationen und anderes. Der Benutzer kann zusätzliche Typen definieren. [Für weitere Informationen siehe Thema Typ-Wörterbücher in Kapitel 10 auf S. 223.](#)

Schritt 5. Indexerstellung

Die gesamte Reihe der Datensätze wird indiziert. Dazu wird ein Zeiger zwischen einer Textstelle und dem bezeichnenden Term für jede Äquivalenzklasse erstellt. Das setzt voraus, dass sämtliche gebeugten Formen, die als Kandidaten für ein Konzept vorkommen, als Grundform für den Kandidaten indiziert werden. Die globale Häufigkeit wird für jede Grundform berechnet.

Schritt 6. Musterabgleich und Ereignisextrahierung.

Mit IBM SPSS Text Analytics for Surveys können nicht nur Typen und Konzepte sondern auch Beziehungen ermittelt werden, die zwischen diesen bestehen. Mit diesem Produkt stehen mehrere Algorithmen und Bibliotheken zur Verfügung, über die Beziehungsmuster zwischen Typen und Konzepten extrahiert werden können. Besonders hilfreich ist dies für die Erfassung von bestimmten Meinungen (z. B. Reaktionen auf ein Produkt) oder von Beziehungsgefügen, die zwischen Personen oder Objekten (etwa zwischen politischen Gruppen oder Genomen) bestehen.

Wie die Kategorisierung funktioniert

Sie haben die Wahl zwischen verschiedenen Methoden, um Kategorien zu erstellen. Da jeder Datensatz seine besonderen Eigenheiten aufweist, kann die Zahl der Methoden und die Reihenfolge, in der sie angewendet werden, gegebenenfalls variieren. Da sich Ihre Interpretation der Ergebnisse möglicherweise von der einer anderen Person unterscheidet, müssen Sie gegebenenfalls ein wenig mit den unterschiedlichen Methoden experimentieren, um zu erkennen, mit welcher Sie die besten Ergebnisse für Ihre Textdaten erzielen.

In diesem Handbuch bezieht sich **Kategorieerstellung** auf die Erstellung von Kategoriedefinitionen und Klassifikation über eine oder mehrere integrierte Methoden und **Kategorisierung** auf das Scoring oder Labeling, einen Prozess, bei dem den Kategoriedefinitionen für jeden Datensatz unverwechselbare IDs (Name/ID/Wert) zugewiesen werden.

Während der Kategorieerstellung werden die extrahierten Konzepte und Typen als Bausteine für Ihre Kategorien verwendet. Bei der Erstellung von Kategorien werden den Kategorien automatisch die Datensätze zugewiesen, die Text enthalten, der einem Element der jeweiligen Kategoriedefinition entspricht.

IBM® SPSS® Text Analytics for Surveys bietet Ihnen mehrere automatisierte Kategorieerstellungsmethoden, mit denen Sie Ihre Datensätze schnell kategorisieren können.

Gruppierverfahren

Die einzelnen verfügbaren Verfahren sind für bestimmte Datentypen und Situationen jeweils sehr gut geeignet, doch ist es oftmals nützlich, bei einer Analyse mehrere Verfahren miteinander zu verbinden, um die Datensätze vollständig zu erfassen. Sie können ein Konzept in mehreren Kategorien erkennen oder redundante Kategorien vorfinden.

Konzeptwurzelaufleitung. Mit diesem Verfahren werden Kategorien erstellt, indem ausgehend von einem Konzept andere verwandte Konzepte ermittelt werden (durch Analyse, ob bestimmte Konzeptkomponenten morphologisch verwandt sind oder gemeinsame Wurzeln haben). Dieses Verfahren ist sehr nützlich bei der Identifizierung von bedeutungsgleichen Konzepten aus zusammengesetzten Wörtern, da die Konzepte in jeder generierten Kategorie die gleiche oder ähnliche Bedeutung haben. Das Verfahren funktioniert mit Daten unterschiedlicher Länge und erzeugt eine geringere Anzahl an kompakten Kategorien. So wird beispielsweise das Konzept `Möglichkeiten zum Aufstieg` mit den Konzepten `Möglichkeit des Aufstiegs` und `Aufstiegsmöglichkeit` zu einer Kategorie zusammengefasst. [Für weitere Informationen siehe Thema Konzeptwurzelaufleitung in Kapitel 6 auf S. 122.](#)

Semantisches Netz. Bei diesem Verfahren wird zunächst auf der Grundlage eines umfassenden Index von Wortbeziehungen jedes Konzept auf seine möglichen Bedeutungen untersucht. Anschließend werden Kategorien durch Gruppieren zusammenhängender Konzepte erstellt. Diese Technik empfiehlt sich, wenn die Konzepte dem semantischen Netz bekannt und nicht zu zweideutig sind. Sie ist weniger hilfreich, wenn der Text spezielle Terminologie oder Sprache enthält, die dem Netz unbekannt ist. Das Konzept `Granny Smith Apfel` würde zum Beispiel mit `Gala Apfel` und `Winesap Apfel` gruppiert, da es sich um Geschwister von `Granny Smith` handelt. In einem anderen Beispiel würde das Konzept `Tier` mit `Katze` und `Känguru` gruppiert, da dies Hyponyme von `Tier` sind. Dieses Verfahren ist in dieser Version nur für englischen Text verfügbar. [Für weitere Informationen siehe Thema Semantische Netze in Kapitel 6 auf S. 124.](#)

Konzepteinbeziehung. Diese Technik erstellt Kategorien durch die Gruppierung von Multiterm-Konzepten (zusammengesetzte Wörter) basierend darauf, ob sie Wörter enthalten, die Unter- oder Übermengen eines Worts in dem anderen sind. So wird beispielsweise `Sitz` mit `Kindersitz`, `Sitzheizung` und `Kindersitzgurt` zu einer Gruppe zusammengefasst. [Für weitere Informationen siehe Thema Konzepteinbeziehung in Kapitel 6 auf S. 123.](#)

Kookkurrenz. Diese Technik erstellt Kategorien aus Kookkurrenz im Text. Dahinter steht folgende Idee: Wenn Konzepte oder Konzeptmuster häufig in Dokumenten bzw. Datensätzen gefunden werden, ist diese Kookkurrenz Ausdruck einer zugrunde liegenden Beziehung, die wahrscheinlich in Ihren Kategoriedefinitionen von Nutzen ist. Wenn Wörter eine signifikante Kookkurrenz aufweisen, wird eine Kookkurrenzregel erstellt, die als Kategoriedeskriptor für eine neue Unterkategorie verwendet werden kann. Wenn beispielsweise viele Datensätze die Wörter `Preis` und `Verfügbarkeit` enthalten (wenige jedoch nur eines von beiden), könnten diese Konzepte in eine Kookkurrenzregel zusammengefasst (`Preis & verfügbar`) und beispielsweise einer Unterkategorie der Kategorie `Preis` zugewiesen werden. [Für weitere Informationen siehe Thema Kookkurrenzregeln in Kapitel 6 auf S. 126.](#)

- **Minimale Anzahl an Datensätzen.** Um festzustellen, wie interessant Kookkurrenzen sind, definieren Sie die minimale Anzahl an Datensätzen, die eine bestimmte Kookkurrenz enthalten muss, um als Deskriptor in einer Kategorie verwendet zu werden.

Vorbereitung für die Textanalyse

Die Textanalyse umfasst mehr als nur Extrahierung und Kategorisierung. Beachten Sie für eine erfolgreiche Textanalyse die folgenden Punkte:

- Wie beim Umfragedesign beeinflusst die Qualität der Antworten, die Sie in IBM® SPSS® Text Analytics for Surveys importieren, direkt die Qualität der resultierenden Kategorisierungen. Im Allgemeinen führen vage oder unklare Fragen zu weitschweifigen und undeutlichen Antworten, die schwer zu analysieren sind.
- Wie die statistische Analyse muss auch die Textanalyse mit klaren Zielen ausgeführt werden. Bevor Sie eine Analyse beginnen, sollten Sie über Ihre Studie nachdenken und bestimmen, was Sie erfahren möchten.

Nehmen wir zum Beispiel an, eine Umfrage in einer örtlichen Schule sollte die Haltungen von Eltern zur Qualität der Bildung messen, die ihre Kinder erhielten. Bei der Analyse könnten wir uns auf Themen wie Namen der Lehrer, Schulprogramme usw. konzentrieren oder auf das Identifizieren und Gruppieren von positiven Rückmeldungen und negativen Rückmeldungen. Ebenso muss die erforderliche Granularität für die Analyse definiert werden, z. B. das Gruppieren aller Bemerkungen über die generelle Mittelbeschaffung oder die weitere Untergliederung dieser Kategorie in Mittelbeschaffung pro Programm. Die Codes oder Kategorien, die wir erstellen, sollten den Fokus und die Ziele unserer Analyse reflektieren.

- Weit weniger als bei einer statistischen Analyse handelt es sich bei einer Textanalyse um eine exakte Wissenschaft, da es nicht nur ein "richtiges" Ergebnis gibt. Textanalyse wird zwar mit bestimmten Zielen ausgeführt, aber sie ist auch subjektiv, da sie von der Interpretation einer Aussage durch den Analytiker abhängt. Ein Beispiel wäre etwa, wie sarkastische Antworten von Teilnehmern identifiziert oder klassifiziert werden. Abhängig von ihren Zielen und ihrem Fokus können zwei kompetente Personen dieselben Daten analysieren und unterschiedliche Schlüsse daraus ziehen.
- Textanalyse ist vor allem ein *iterativer Prozess*. Bei Ihrer Arbeit mit Umfrageantworten werden Sie sicher die Antworten immer wieder anhand verschiedener Kategoriedefinitionen (d. h. Kodierungsschemata), unterschiedlicher Konzept- oder Synonymdefinitionen und unterschiedlicher Antwortgruppierungen extrahieren und neu kategorisieren.

Nachdem Sie Konzepte aus Ihrem Text extrahiert und Ihre Kategorien angelegt haben, sollten Sie Ihre Ergebnisse sorgfältig prüfen. Wenn Sie Elemente finden, die Sie manipulieren möchten, passen Sie einfach Ihre Analyse an, indem Sie Ihre Kategoriedefinitionen und die Definitionen der linguistischen Bibliothek feineinstellen. Die Antworten werden dann beim erneuten Extrahieren automatisch neu kategorisiert. Sie können diesen Prozess einmal oder mehrmals durchlaufen, bis Sie mit dem Ergebnis Ihrer Analyse zufrieden sind.

Anmerkung: Weitere Informationen zu Überlegungen vor dem Import von Daten finden Sie unter [Ihre Daten vorbereiten auf S. 28](#)

Zuverlässigkeit und Feineinstellung

Beim Kodieren von Daten sollen die resultierenden Kategorien immer zuverlässig sein. Im Zusammenhang mit dem Kodieren offener Antworten bedeutet dies, dass zwei unabhängige Kodierer mit denselben Regeln (demselben Coderahmen) dieselbe Antwort identisch kodieren

werden. Bei der manuellen Textanalyse ist dies ein kritischer Punkt. Eine wertvolle Gruppe an Kategorien kann angelegt werden, aber wenn sie nicht zuverlässig auf die Antworten angewendet werden, kann sich ihr Wert erheblich verringern. Wenn IBM® SPSS® Text Analytics for Surveys auf dieselben Daten mit denselben linguistischen Ressourcen angewendet wird, reproduziert dies eine vorherige Analyse stets perfekt. Die Zuverlässigkeit beträgt 100 %.

Dies bedeutet jedoch nicht, dass die Analyse keine Fehler enthält, aber die Konzentration auf die Kodierung kann sich nun auf eine andere Aufgabe verlagern — *die Feineinstellung*. Beim manuellen Kodieren lesen die Kodierer die Antwort und können alle Nuancen einer Aussage erfassen (selbst wenn sie Probleme beim Anwenden der Kodierungskategorien haben). SPSS Text Analytics for Surveys kann die Kodierungskategorien anwenden, aber die Kategorien müssen so definiert sein, dass Nuancen und Unterschiede erfasst werden. Es gibt zwei Möglichkeiten zur Feineinstellung:

- Verfeinern der linguistischen Ressourcen
- Verfeinern der Kategoriedefinitionen

Verfeinern von linguistischen Ressourcen

IBM® SPSS® Text Analytics for Surveys erstellt problemlos Kategorien ohne Eingriffe von Ihrer Seite, aber diese werden ausnahmslos nicht alle Informationen in den Antworten erfassen. Sie müssen die linguistische Basis verbessern, die das Programm verwendet, damit seine Kategoriererstellung immer mehr die Eigenheiten des Textes berücksichtigt. Um diese Basis zu verbessern, können Sie eine Anpassung und Feinabstimmung der für die Extrahierung aus dem Text verwendeten linguistischen Ressourcen vornehmen.

Eine Feineinstellung umfasst in diesem Fall das Hinzufügen von Wörtern und Phrasen zu zahlreichen linguistischen Bibliotheken und Wörterbüchern, die Angabe von Wörtern, die aus der Analyse auszuschließen sind, das Definieren von Synonymen oder das Erstellen von benutzerdefinierten Bibliotheken – alles in Hinblick auf ein bestimmtes Ziel. Dieses Ziel ist die exakte Erfassung der Gedanken der Umfrageteilnehmer im Text und das Entfernen von Mehrdeutigkeiten aus dem Ergebnis.

Verfeinern von Kategoriedefinitionen

Neben der Verfeinerung der linguistischen Ressourcen sollten Sie Ihre Kategorien überprüfen, indem Sie nach Möglichkeiten suchen, die zugehörigen Definitionen zu kombinieren oder zu bereinigen, und einige der kategorisierten Antworten prüfen. Sie können die automatisierten Kategorieerstellungsverfahren zum Erstellen Ihrer Kategorien verwenden; Sie werden jedoch sicherlich noch einige Optimierungen an diesen Definitionen vornehmen wollen. Nach der Verwendung eines Verfahrens wird eine Reihe neuer Kategorien im Fenster angezeigt. Sie können unter den Kategorien die Konzepte einblenden, die jede Kategorie definieren. Anschließend können Sie die Antworten in einer Kategorie überprüfen und Anpassungen vornehmen, bis Sie mit den Kategoriedefinitionen zufrieden sind.

Keines der automatischen Verfahren ergibt eine perfekte Kategorisierung Ihrer Daten, weshalb wir empfehlen, ein oder mehrere automatische Verfahren anzuwenden, die gut mit Ihren Daten funktionieren. Anschließend können Sie mithilfe von manuellen Verfahren kleinere Anpassungen

vornehmen, etwaige Fehlklassifizierungen beheben oder Datensätze oder Konzepte hinzufügen, die nicht erfasst wurden.

Erste Schritte

Die vorliegende Dokumentation präsentiert die Aufgaben, die Sie mit IBM® SPSS® Text Analytics for Surveys ausführen können, und die Techniken, mit deren Hilfe Sie Ihre Antworten kategorisieren können. Diese Informationen führen Sie durch Ihre erste Analyse. Hier werden alle Prozesse zur vollständigen Analyse Ihrer Daten behandelt. Da aber jede Datenmenge anders ist, müssen Sie letztendlich entscheiden, wann Ihre Analyse abgeschlossen ist.

In diesem Kapitel behandeln wir den typischen Prozess, den Benutzer während einer Textanalyse durchlaufen. Zusätzlich erhalten Sie einen allgemeinen Überblick über die Oberfläche und die wesentlichen Aufgaben und Elemente Ihrer Arbeit.

Der typische Prozess

Im Folgenden finden Sie eine Übersicht über den typischen Workflow-Prozess für die Verwendung von IBM® SPSS® Text Analytics for Surveys.

- **Erstellen Sie ein Projekt**, indem Sie Ihre Umfragedaten einschließlich offener Fragen, einer ID-Variablen und anderer Referenzvariablen in SPSS Text Analytics for Surveys importieren. Daten können aus IBM® SPSS® Statistics-Datendateien, Microsoft Excel, jedem ODBC-kompatiblen Datenbankprogramm oder einer IBM® SPSS® Data Collection-Datenquelle eingelesen werden. Sie können ein Textanalysepaket wählen, um einige vordefinierte Kategorien und spezielle Ressourcen zu nutzen und schnell produktiv zu werden.
- **Extrahieren** Sie Konzepte und Muster für jede offene Frage, die Sie importiert haben. Die interne Extrahierungsengine identifiziert und erfasst automatisch die im Text ausgedrückten Schlüsselbegriffe. Diese Begriffe werden unter einem Hauptkonzept gruppiert. Konzepte werden dann zu Typen zusammengefasst, bei denen es sich um Sammlungen ähnlicher Begriffe handelt, wie beispielsweise Unternehmen, Produkte oder positive Meinungen. Außerdem werden Muster extrahiert. Diese stellen Kombinationen von Begriffen und Typen dar, die für Meinungen und Beziehungen stehen, wie beispielsweise positive Kommentare zu einem Unternehmen.
- **Verfeinern** Sie die extrahierten Konzepte und optimieren Sie Ihre Extrahierungen durch die direkte Bearbeitung einer oder mehrerer Bibliotheken, die Worttypen, Begriffe, Synonyme, Ausschlusslisten und andere linguistische Konstrukte enthalten. Wie bereits erwähnt, handelt es sich bei der Textanalyse um einen iterativen Prozess, bei dem die Verfeinerung von Bibliotheken und Wörterbüchern zu Ergebnissen führt, die durch eine Feinabstimmung für Ihre Daten optimiert werden.
- **Kategorisieren** Sie Ihre Antworten, indem Sie Kategorien manuell mithilfe von Kategorieregeln und Coderahmen erstellen und bearbeiten und/oder automatisch Kategorieerstellungstechniken verwenden. Die Kategorien stehen für Konzepte einer höheren

Ebene, die die wichtigsten Ideen und Haltungen erfassen, die die Befragten zum Ausdruck bringen.

- **Exportieren** Sie Ihre Kategorien zusammen mit der ID-Variablen in übliche Dateiformate zur weiteren Analyse bzw. grafischen Darstellung in anderen Programmen. Bei der Ausgabe kann es sich um ein Set mit Mehrfachantwortvariablen handeln, entweder als SPSS Statistics- oder als Microsoft Excel-Datei.

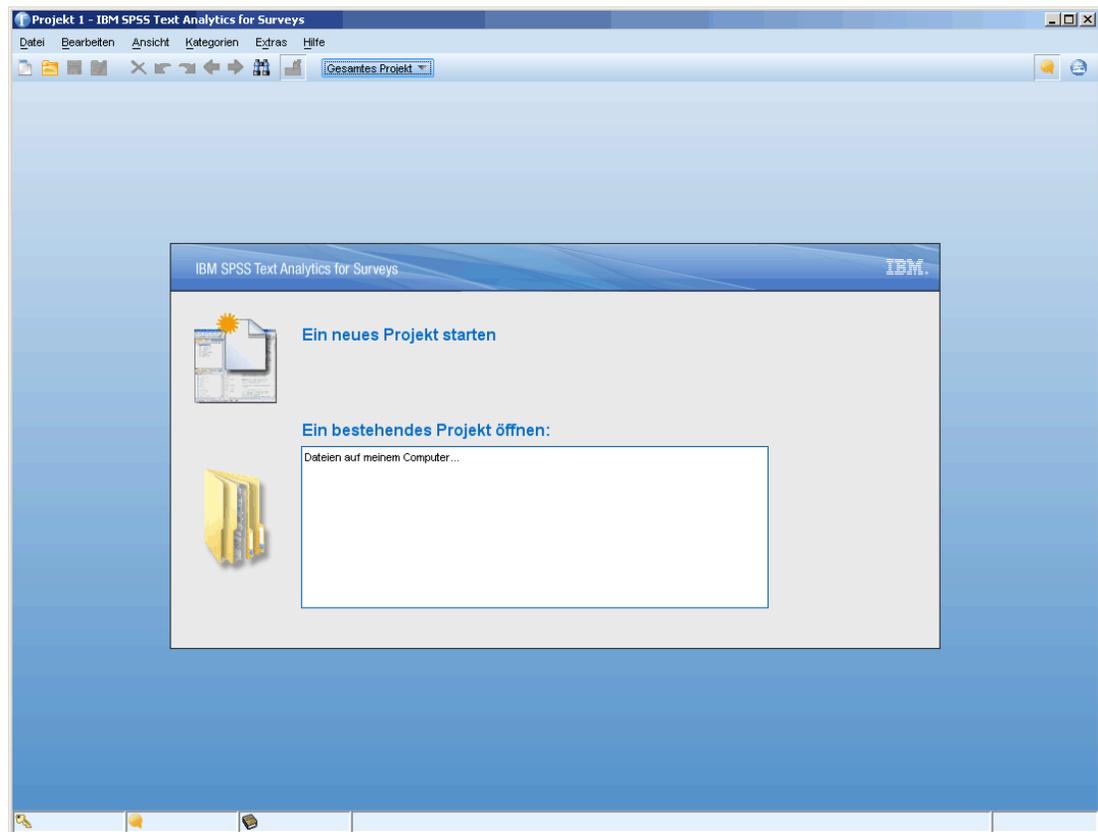
Das Textanalysefenster

Die Oberfläche der Anwendung besteht aus mehreren Fenstern. Das erste ist das Textanalysefenster, in dem Sie den größten Teil Ihrer Arbeit erledigen. In diesem Fenster können Sie jede Frage in Ihren Daten analysieren. Für jede Frage können Sie Konzepte, Typen und Muster extrahieren und anschließend Ihre Antworten kategorisieren.

Beim Start der Anwendung sehen Sie einen Bildschirm, in dem Sie ein bestehendes Projekt öffnen oder ein neues Projekt anlegen können. Wenn Sie sich für das Erstellen eines neuen Projekts entscheiden, begleitet Sie ein Assistent durch den Vorgang der Projekterstellung. [Für weitere Informationen siehe Thema Erstellen von Projekten in Kapitel 3 auf S. 27.](#)

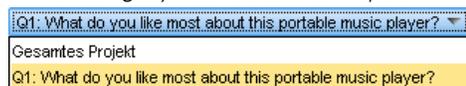
Abbildung 2-1

Textanalysefenster beim Produktstart



Sobald Sie Daten importiert haben, können Sie die Ansicht(en) “Frage” oder “Gesamtes Projekt” ansehen. Sie können die Ansicht wechseln, indem Sie eine Ansicht aus der Dropdown-Liste in der Symbolleiste des Textanalysefensters oder aus dem Menü “Ansicht” wählen. Der im Listenfeld angezeigte Text wird aus der Variablenbezeichnung für jede Frage übernommen.

Abbildung 2-2
Anwendungssymbolleiste mit Dropdown-Liste “Ansichtauswahl”



Ansicht “Frage”

Die Ansicht “Frage” bietet Ihnen einen Bereich, in dem Sie die Antworten auf eine bestimmte Frage analysieren und kategorisieren können. Nach dem Anlegen eines neuen Projekts wird die Ansicht “Frage” mit den Antworten für die erste importierte offene Textvariable angezeigt. Sie können diese Ansicht aus der Dropdown-Liste wählen oder einen Fragenamen aus dem Menü Ansicht > Frage > wählen.

Abbildung 2-3
Ansicht “Frage”

Kategorie	Anzahl D...	Anzahl Tr...
Alle Datensätze	-	405
Nicht kategorisiert	-	262
Keine Konzepte extrahiert	-	4
songs++Positive>	3	9
fx [listening to songs + like]	-	1
fx [songs + like] [songs + meets needs] [songs	-	7
drag & dropped songs from their site++Positive>	1	1
fx [drag & dropped songs from their site + eat	-	1
memory	2	2
fx [memory + easy to use]	-	1
fx [storage + able]	-	1
battery++Contextual>	2	3
design++Positive>	2	7
store++Positive>	2	3
skipping++Contextual>	2	2

Kategorie	Balken	Auswahl %	Befragte	Gesamt %
volume++Contextual>		11,1	1	0,5
songs++Positive>		88,9	8	2,2

ID	Antwort	Kategorien
1	69 Stores 5000 songs in a compact portable player.	songs++Positive>
2	117 Listening to any songs like any time like.	songs++Positive>
3	146 It has a lot of storage capacity. I can fit a lot of songs on it. Also, it's very lightweight.	songs++Positive>
4	161 Product A is the best. Drag & drop songs from their site is really easy and I love all the random play options. Like the long battery life.	...ongs from their site++Positive>
5	168 It like has all of my songs. I mean all of them, and it still has room for like a gazillion more.	songs++Positive>
6	274 Well being a long haul truck driver I really like the fact that there is always music to listen to and that I can turn up the volume real loud when my favorite song comes on.	songs++Positive> volume++Contextual>
7	299 It holds loads of songs and is easy to use.	songs++Positive>
8	340 I can listen to only the songs that like.	songs++Positive>

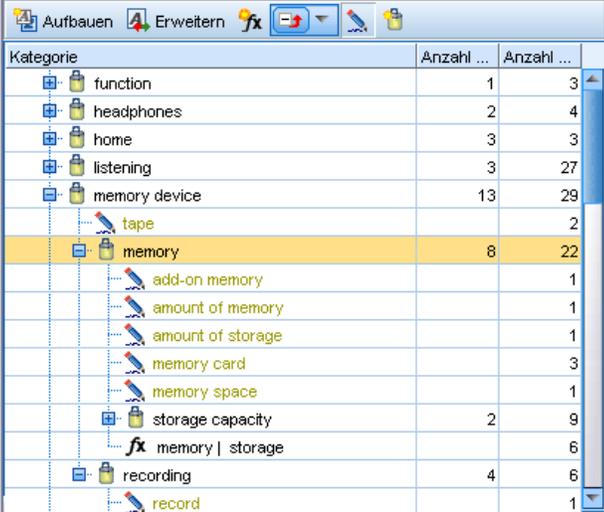
Die Operationen, die Sie in der Ansicht “Frage” ausführen, drehen sich um drei Elemente: Extrahierungsergebnisse, Kategorien und Antwortdaten. Zur Unterstützung Ihrer Arbeit und der unabhängigen Analyse jedes der Elemente ist dieses Fenster in vier Bereiche gegliedert:

Fensterbereich “Kategorien”

Dieser Bereich befindet sich links oben und zeigt eine Arbeitsfläche, in der Sie alle von Ihnen erstellten Kategorien verwalten und neue Kategorien erstellen können. Nachdem Sie die Konzepte, Typen und Muster aus Ihren Textdaten extrahiert haben, können Sie Kategorien mithilfe von automatischen Verfahren (z. B. semantische Netze und Konzeptbeziehung) oder manuell erstellen. Sie können auf eine Kategorie klicken und sie erweitern, um alle Deskriptoren zu sehen, aus denen ihre Definition besteht, wie beispielsweise Konzepte, Typen und Regeln. Wenn Sie eine Kategorie oder einen Deskriptor auswählen, können Sie Informationen über entsprechende Datensätze in den Bereichen “Daten” und “Visualisierung” anzeigen. [Für weitere Informationen siehe Thema Der Fensterbereich “Kategorien” in Kapitel 6 auf S. 97.](#)

Abbildung 2-4

Fensterbereich “Kategorien”: Erweiterte Kategoriedefinition



Kategorie	Anzahl ...	Anzahl ...
function	1	3
headphones	2	4
home	3	3
listening	3	27
memory device	13	29
tape		2
memory	8	22
add-on memory		1
amount of memory		1
amount of storage		1
memory card		3
memory space		1
storage capacity	2	9
memory storage		6
recording	4	6
record		1

Fensterbereich “Extrahierte Ergebnisse”

Dieser Bereich befindet sich links unten und zeigt die Ergebnisse der Extrahierung an. Wenn Sie eine Extrahierung ausführen, liest die Extrahierungseingabe die Textdaten, identifiziert die relevanten Konzepte und weist jedem davon einen Typ zu. **Konzepte** sind Wörter bzw. Ausdrücke, die aus Ihren Textdaten extrahiert wurden. **Typen** sind semantische Gruppierungen von Konzepten, die in Form von Typ-Wörterbüchern gespeichert sind. Wenn die Extrahierung abgeschlossen ist, werden die Konzepte, Typen und Muster im Fensterbereich “Extrahierungsergebnisse” angezeigt. Konzepte und Typen sind farbkodiert, damit Sie leichter erkennen können, zu welchem Typ sie gehören. Anhand dieser Konzepte, Typen und Muster, die Sie hier sammeln, können Sie Ihre Kategorien erstellen. [Für weitere Informationen siehe Thema Extrahierte Ergebnisse: Konzepte, Typen und Muster in Kapitel 5 auf S. 80.](#)

Text-Mining ist ein iterativer Prozess, in dem Extrahierungsergebnisse dem Kontext der Textdaten gemäß überprüft, für die Gewinnung neuer Ergebnisse verfeinert und dann neu bewertet werden. Die Extrahierungsergebnisse können durch Bearbeiten der linguistischen Ressourcen verfeinert werden. Diese Feinabstimmung kann teilweise direkt im Fensterbereich “Extrahierte Ergebnisse” oder “Daten” vorgenommen werden, aber auch direkt in der Ansicht “Ressourceneditor”. [Für weitere Informationen siehe Thema Das Fenster “Ressourceneditor” auf S. 16.](#)

Datenbereich

In der unteren rechten Ecke dieser Ansicht werden die Antwortdaten für die ausgewählte offene Frage in tabellarischem Format präsentiert. Standardmäßig zeigt der Bereich “Daten” drei Spalten (Datensatz-IDs, Textantworten und zugewiesene Kategorien) an. Die Anzahl der in diesem Bereich angezeigten Antworten wird entsprechend Ihrer Auswahl in einem anderen Fensterbereich gefiltert. Sie können zwar die in diesen Bereich importierten Daten anzeigen, jedoch können Sie die Datensätze nicht bearbeiten, löschen oder erweitern. [Für weitere Informationen siehe Thema Der Fensterbereich “Daten” in Kapitel 6 auf S. 100.](#)

Fensterbereich “Visualisierung”

In der oberen rechten Ecke dieser Ansicht, standardmäßig ausgeblendet. Sie können diesen Fensterbereich einblenden. (Wählen Sie Ansicht > Visualisierung.) Dieser Bereich bietet auf der Grundlage Ihrer Auswahl in den anderen Fensterbereichen drei unterschiedliche Ansichten dazu, wie Antworten in Kategorien passen oder wie Kategorien Antworten gemeinsam enthalten können (Netzdiagramm, Balkendiagramm und Tabelle). [Für weitere Informationen siehe Thema Visualisierung von Diagrammen in Kapitel 7 auf S. 172.](#)

Abhängig davon, ob Sie die Extrahierungsoption im Assistenten “Neues Projekt” wählen, können im Fensterbereich unten links Extrahierungsergebnisse angezeigt sein. Klicken Sie im Bereich “Extrahierungsergebnisse” auf Extrahieren, um die Extrahierung zu beginnen. Nach der Extrahierung können Sie die Ergebnisse anzeigen, um zu sehen, ob Feineinstellungen wie die Gruppierung von Synonymen unter einem Begriff oder das Ausschließen allgemeiner, uninteressanter Begriffe aus der Liste erforderlich sind.

Sobald die Extrahierungsergebnisse Ihren Wünschen entsprechen, können Sie die Kategorisierung Ihrer Antworten beginnen, indem Sie Konzepte verschieben und als Kategorien ablegen oder Techniken der Kategorieerstellung wie die Einbeziehung von Konzepten oder ein semantisches Netzwerk verwenden.

Die Ansicht “Gesamtes Projekt”

Die Ansicht “Gesamtes Projekt” bietet einen Überblick über alle Variablen, die Sie in das Projekt importiert haben. Sie können diese Ansicht aus der Dropdown-Liste oder dem Menü “Ansicht” (Ansicht > Gesamtes Projekt) wählen. In dieser Ansicht können Sie die importierten Daten prüfen, die Rolle einer Variablen ändern (z. B. von einer Frage- in eine Referenzvariable) und Variablen Bezeichnungen zuweisen. [Für weitere Informationen siehe Thema Anzeigen von Projektdaten in Kapitel 4 auf S. 51.](#) Sie können zwar die in diese Ansicht importierten Daten anzeigen, jedoch können Sie die Datensätze nicht bearbeiten, korrigieren, löschen oder erweitern.

Anmerkung: Den vollständigen Inhalt einer Zelle in dieser Ansicht sehen Sie, wenn Sie die Maus auf der Zelle platzieren. Eine QuickInfo zeigt den Zelleninhalt an.

Abbildung 2-5
Ansicht "Gesamtes Projekt"

	Serial	3. Was hat Ihnen während Ihres Aufenthaltes am besten gefallen?	4. Womit waren Sie überhaupt nicht zufrieden?	1. Was ist der Anlass Ihrer Reise:	Einige Angaben zu Ihrer Person:6. Wie alt sind Sie?	Einige / Person
1	1	Zimmer sauber	Ziemlich teuer	Geschäftsreise	51 bis 60 Jahre	Weiblich
2	2	nichts	Schlechtes Personal	Urlaub	31 bis 40 Jahre	Weiblich
3	3	Preis-leistungsverhältnis sehr ok	n/a	Geschäftsreise	21 bis 30 Jahre	Männlich
4	4	Sauberes Zimmer	Parkplatz zu klein	Geschäftsreise	41 bis 50 Jahre	Weiblich
5	5	die komfortablen Zimmer, reichhaltiges Fr...	mit der Sauberkeit der Badezimmer und des Schi...	Urlaub	21 bis 30 Jahre	Männlich
6	6	Die ruhige Lage direkt am Strand.	Das Essen war zu wenig abwechslungsreich, di...	Urlaub	41 bis 50 Jahre	Weiblich
7	7	Angenehmer Service	Frühstück	Geschäftsreise	21 bis 30 Jahre	Weiblich
8	8	Der Ausblick auf die schöne Landschaft ...	Das Personal war sehr unfreundlich.	Geschäftsreise	31 bis 40 Jahre	Männlich
9	9	Die besonders schöne Landschaft	nichts	Urlaub	31 bis 40 Jahre	Weiblich
10	10	Der Service, die Sauberheit und Gepflegt...	Die technischen Anlagen, Fahrstuhl, elektrische ja...	Urlaub	41 bis 50 Jahre	Weiblich
11	11	Die Bedienung.	Mit dem Frühstück.	Urlaub	41 bis 50 Jahre	Männlich
12	12	gute Strandlage, vielfältige Ausflugs-Ang...	keine Moskitonetze, Angebote überteuert, zu wen...	Urlaub	21 bis 30 Jahre	Männlich
13	13	Freundlicher Room-Service, aufmerksame...	keine Verlängerung der Check out Zeit möglich	Geschäftsreise	41 bis 50 Jahre	Weiblich
14	14	Das Hotelpersonal war stets sehr zuvork...	Das Frühstücksbuffet war sehr eintönig, d.h. es ...	Geschäftsreise	21 bis 30 Jahre	Weiblich
15	15	Sauberes Hotel	kleines Zimmer	Urlaub	Über 60 Jahre	Männlich
16	16	die Sonnenterrasse war wunderbar, auch...	die Öffnungszeiten des Fitness-Centers waren z...	Urlaub	21 bis 30 Jahre	Weiblich
17	17	Der Late Check-in und die Internetverbind...	Die Kosten für eine mehrstündige LAN-Verbindun...	Geschäftsreise	41 bis 50 Jahre	Männlich
18	18	Sehr ruhig	sehr teuer	Geschäftsreise	31 bis 40 Jahre	Männlich
19	19	Frühstücksbuffet, Luxuriöse Bäder	Freundlichkeit des Personals, Reiseleitung auf Au...	Urlaub	31 bis 40 Jahre	Weiblich
20	20	Preis/Leistung, technische Ausstattung d...	Unruhe in der Nacht	Geschäftsreise	51 bis 60 Jahre	Männlich
21	21	Der freundliche Service sowie die Ruhe u...	Die Betten waren etwas zu klein, man konnte sic...	Geschäftsreise	41 bis 50 Jahre	Männlich
22	22	Service, ruhige Lage, Ausstattung der Zi...	Parkplatzangebot	Geschäftsreise	31 bis 40 Jahre	Männlich
23	23	Auswahl beim Frühstücksbuffet, Größe d...	keine Angabe	Geschäftsreise	31 bis 40 Jahre	Weiblich
24	24	schöne Atmosphäre, tolle Einrichtung, her...	zu kurze Frühstückszeiten	Urlaub	21 bis 30 Jahre	Weiblich
25	25	Frühstücksbuffet, Lounge, Guter Service	Handtücher waren zu hart	Geschäftsreise	21 bis 30 Jahre	Weiblich
26	26	Das gute Essen, das freundliche Personal...	die Sauberkeit des Zimmers (Ungeziefer), Flecken...	Urlaub	21 bis 30 Jahre	Männlich
27	27	Das Wetter, die Aussicht, der Internetans...	Preise für Internetzugang viel zu hoch, das Perso...	Geschäftsreise	21 bis 30 Jahre	Männlich
28	28	Die Zimmer und der Service	Mit dem Abendessen	Urlaub	31 bis 40 Jahre	Männlich
29	29	der gute Service	Frühstücksbuffet	Geschäftsreise	31 bis 40 Jahre	Weiblich

Das Fenster "Ressourceneditor"

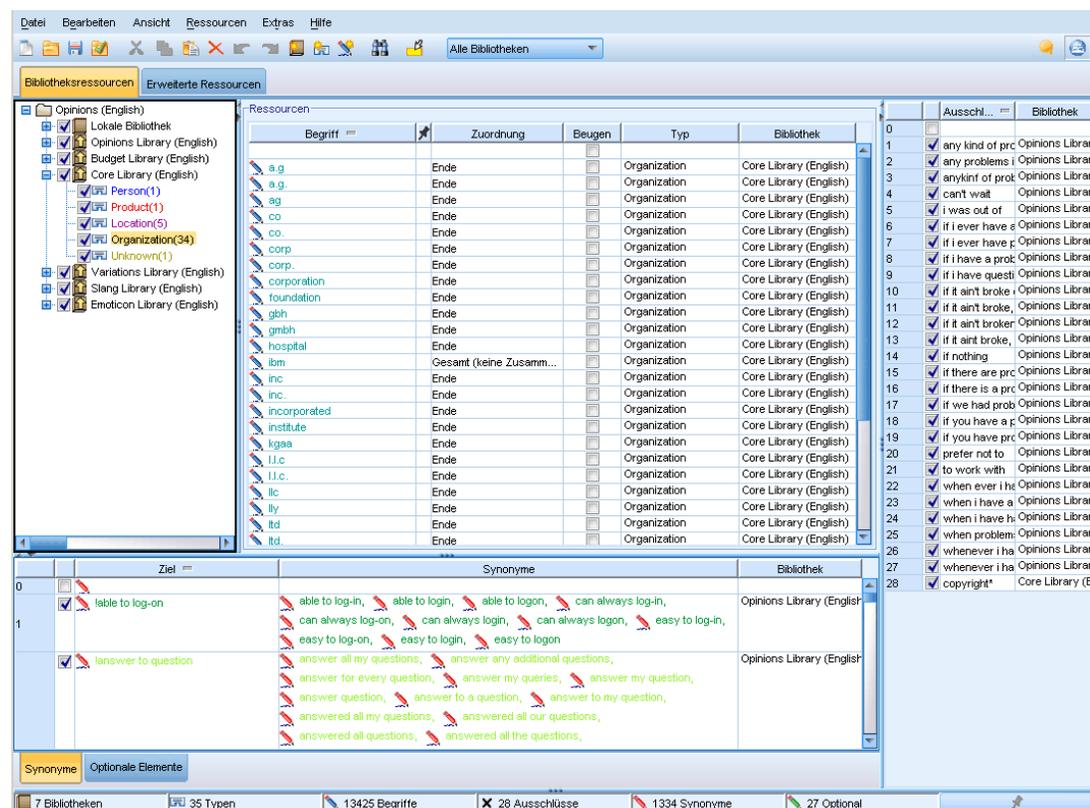
IBM® SPSS® Text Analytics for Surveys erfasst mithilfe einer robusten Extrahierungs-Engine schnell und genau die Schlüsselkonzepte aus Textdaten. Diese Engine ist stark auf linguistische Ressourcen angewiesen, die vorgeben, wie große Mengen an unstrukturierten Textdaten zu analysieren und interpretieren sind.

In der Ansicht Resource Editor können Sie die zum Extrahieren von Konzepten verwendeten linguistischen Ressourcen anzeigen und durch Feinabstimmung optimieren, zu Typen zusammenfassen, Muster in den Textdaten ermitteln und vieles mehr. IBM® SPSS® Text Analytics for Surveys bietet verschiedene vorkonfigurierte Ressourcenvorlagen. In einigen Sprachen können Sie auch die Ressourcen in einem Textanalysepaket verwenden. [Für weitere Informationen siehe Thema Verwendung von Text Analysis Packages in Kapitel 3 auf S. 43.](#)

Da diese Ressourcen möglicherweise nicht immer perfekt an den Kontext Ihrer Daten angepasst sind, können Sie im Resource Editor Ihre eigenen Ressourcen für einen bestimmten Kontext oder eine bestimmte Domäne erstellen, bearbeiten und verwalten. [Für weitere Informationen siehe Thema Mit Bibliotheken arbeiten in Kapitel 9 auf S. 209.](#) Einige Benutzer werden dieses Fenster nur selten verwenden, da die im Lieferumfang des Produkts enthaltenen Ressourcen oft ausreichen. Außerdem können Sie einen Großteil der einfachen Arbeiten an Bibliotheken direkt vom Bereich "Extrahierungsergebnisse" im Fenster "Textanalyse" aus durchführen.

Um die Feinabstimmung Ihrer linguistischen Ressourcen zu vereinfachen, können Sie über Kontextmenüs in den Fensterbereichen “Extrahierungsergebnisse” und “Daten” häufig verwendete Wörterbuchaufgaben direkt aus der Ansicht Text Analytics heraus durchführen. [Für weitere Informationen siehe Thema Extrahierungsergebnisse verfeinern in Kapitel 5 auf S. 88.](#)

Abbildung 2-6
Ressourceneditoransicht



Bei den in der Ansicht Resource Editor durchgeführten Operationen geht es um die Verwaltung und Feinabstimmung der linguistischen Ressourcen. Diese Ressourcen werden in Form von Vorlagen und Bibliotheken gespeichert. Die Ansicht “Resource Editor” gliedert sich in vier Bereiche: den Fensterbereich “Bibliotheksbaum”, den Fensterbereich “Typ-Wörterbuch”, den Fensterbereich “Austauschwörterbuch” und den Fensterbereich “Ausschlusswörterbuch”.

Die Oberfläche gliedert sich in die vier folgenden Bereiche:

1. Fensterbereich “Bibliotheksbaum”. Dieser Plan befindet sich links oben und zeigt eine Baumstruktur mit den Bibliotheken. Sie können die Bibliotheken in diesem Baum aktivieren bzw. deaktivieren sowie die Ansichten in den anderen Fensterbereichen filtern, indem Sie eine Bibliothek im Baum auswählen. In diesem Baum können Sie mithilfe der Kontextmenüs eine Vielzahl von Operationen durchführen. Wenn Sie eine Bibliothek im Baum erweitern, wird das darin enthaltene Typen-Set angezeigt. Sie können diese Liste auch über das Menü Ansicht filtern, wenn Sie sich nur auf eine bestimmte Bibliothek konzentrieren möchten.

2. Bereich für Fachausdruckslisten von Typen-Wörterbüchern. Dieser Fensterbereich befindet sich rechts neben dem Bibliotheksbaum und zeigt die Fachausdruckslisten der Typ-Wörterbücher für die im Baum ausgewählten Bibliotheken an. Ein **Typ-Wörterbuch** ist eine Zusammenstellung von Fachausdrücken, die unter einer Bezeichnung (bzw. einem Typ oder einem Namen) gruppiert werden sollen. Beim Lesen der Textdaten vergleicht die Extrahierungsengine die im Text gefundenen Wörter mit den Fachausdrücken in den Typ-Wörterbüchern. Wenn ein extrahiertes Konzept als Fachausdruck in einem Typ-Wörterbuch vorkommt, wird der entsprechende Typname zugewiesen. Sie können das Typ-Wörterbuch als eigenes Wörterbuch mit Fachausdrücken betrachten, die etwas gemeinsam haben. Der Typ <Location> (Ort) in der Core Library beispielsweise enthält Konzepte wie *new orleans*, *great britain* und *new york*. Diese Fachausdrücke stehen jeweils für geografische Orte. Eine Bibliothek kann ein oder mehrere Typ-Wörterbücher enthalten. [Für weitere Informationen siehe Thema Typ-Wörterbücher in Kapitel 10 auf S. 223.](#)

3. Fensterbereich "Ausschlusswörterbuch". Dieser Fensterbereich befindet sich auf der rechten Seite und zeigt die Sammlung der Fachausdrücke, die aus den endgültigen Extrahierungsergebnissen ausgeschlossen werden. Die Fachausdrücke in diesem Ausschlusswörterbuch werden nicht im Fensterbereich "Extrahierungsergebnisse" angezeigt. Ausgeschlossene Fachausdrücke werden in der Bibliothek Ihrer Wahl gespeichert. Jedoch zeigt der Bereich "Ausschlussbibliothek" alle ausgeschlossenen Fachausdrücke für alle Bibliotheken an, die im Bibliotheksbaum sichtbar sind. [Für weitere Informationen siehe Thema Ausschlusswörterbücher in Kapitel 10 auf S. 240.](#)

4. Fensterbereich "Substitutionswörterbuch". Dieser Fensterbereich befindet sich links unten und zeigt Synonyme und optionale Elemente in einer jeweils eigenen Registerkarte an. Synonyme und optionale Elemente helfen bei der Gruppierung ähnlicher Fachbegriffe unter einem Haupt- oder Zielkonzept in den endgültigen Extrahierungsergebnissen. Das Wörterbuch kann bekannte Synonyme, benutzerdefinierte Synonyme und Elemente sowie häufig vorkommende Rechtschreibfehler zusammen mit der korrekten Schreibung enthalten. Synonyme und optionale Elemente können in einer Bibliothek Ihrer Wahl gespeichert werden. Jedoch zeigt der Bereich "Substitutionswörterbuch" den Inhalt für alle Bibliotheken an, die im Bibliotheksbaum sichtbar sind. Während dieser Bereich alle Synonyme und optionale Elemente aus allen Bibliotheken anzeigt, werden die Substitutionen für alle Bibliotheken in der Baumstruktur zusammen in diesem Fensterbereich angezeigt. Eine Bibliothek kann nur ein einziges Substitutionswörterbuch enthalten. [Für weitere Informationen siehe Thema Substitutions-/Synonymwörterbücher in Kapitel 10 auf S. 234.](#)

Anmerkung:

- Wenn Sie einen Filtervorgang durchführen möchten, sodass nur die Informationen angezeigt werden, die zu einer einzigen Bibliothek gehören, können Sie die Bibliotheksansicht mithilfe der Dropdown-Liste in der Symbolleiste ändern. Sie enthält auf der obersten Ebene den Eintrag Alle Bibliotheken sowie einen zusätzlichen Eintrag für jede einzelne Bibliothek. [Für weitere Informationen siehe Thema Bibliotheken anzeigen in Kapitel 9 auf S. 213.](#)

Festlegen von Optionen

Im Dialogfeld “Optionen” können Sie allgemeine Optionen für IBM® SPSS® Text Analytics for Surveys festlegen. Dieses Dialogfeld enthält folgende Registerkarten:

Die Registerkarte **System** enthält Optionen für Standard-Bibliothekslisten, automatisches Speichern, Speichern von Extrahierungsergebnissen, Trennzeichen sowie zur Sprache der Benutzeroberfläche.

Die Registerkarte **Anzeige** enthält Optionen für die auf der Benutzeroberfläche verwendeten Farben.

Die Registerkarte **Klänge** enthält Optionen für Tonsignale.

Die Registerkarte **Übersetzung** enthält Optionen für Übersetzungsverbindungen.

So bearbeiten Sie Optionen:

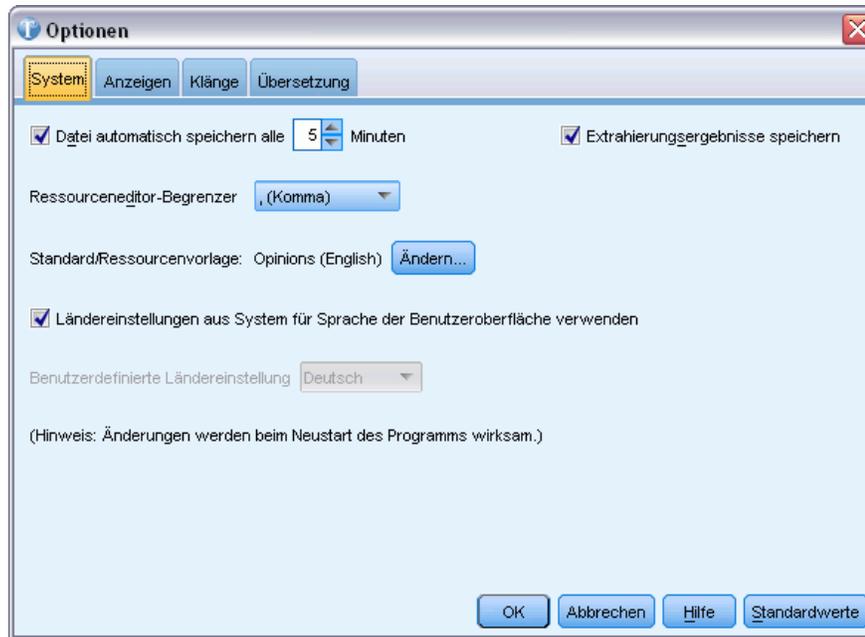
- ▶ Wählen Sie in den Menüs die Optionsfolge Extras > Optionen aus. Das Dialogfeld “Optionen” wird geöffnet.
- ▶ Wählen Sie die Registerkarte mit den Informationen aus, die geändert werden sollen.
- ▶ Ändern Sie die Optionen nach Bedarf.
- ▶ Klicken Sie auf OK, um die Änderungen zu speichern.

Optionen: Registerkarte “System”

In dieser Registerkarte können Sie unter anderem folgende Projekteinstellungen definieren:

- Hinzufügen oder Entfernen von Bibliotheken, die standardmäßig in allen neuen Projekten erscheinen sollen
- Aktivieren oder Deaktivieren der Funktion für automatisches Speichern und Wiederherstellung
- Aktivieren oder Deaktivieren der Speicherung von Extrahierungsergebnissen
- Definieren des globalen Trennzeichens, das im Resource Editor zur Trennung von Elementen benutzt wird

Abbildung 2-7
Dialogfeld "Optionen": Registerkarte "System"



Datei automatisch speichern alle n Minuten. Wählen Sie diese Option aus, damit das Produkt für den Fall eines Computerausfalls automatisch eine temporär gespeicherte Version der geöffneten Projektdatei erstellt. Stellen Sie auch die Anzahl der Minuten zwischen den einzelnen Speichervorgängen ein. Wenn Sie diese Funktion aktivieren und das Produkt unerwartet beendet wird oder ein Computerproblem auftritt, erhalten Sie beim nächsten Start des Produkts die Gelegenheit, eine Wiederherstellungsversion Ihrer Datei zu öffnen und zu bearbeiten.

Extrahierungsergebnisse speichern. Wählen Sie diese Option aus, um die Ergebnisse Ihrer Extrahierungen in Ihrem Projekt zu speichern. Damit können Sie Zeit sparen, wenn Sie noch an Ihren Kategorien arbeiten. Jedoch kann es zusätzliche Zeit beim Laden beanspruchen und Ihr Projekt vergrößern. Zur Sicherheit werden diese Extrahierungsergebnisse während des Speicherns verschlüsselt und in der Datenbank abgelegt. Dieses Verfahren erschwert es sogar einem fortgeschrittenen Benutzer, Daten in der Datenbank ausfindig zu machen. Außerdem werden Extrahierungsergebnisse nie in IBM® SPSS® Text Analytics for Surveys dargestellt, bis die Anwendung die Datenquelle gefunden hat. Wenn die Daten passwortgeschützt sind, müssen daher der Benutzername und das Passwort für diese Datenquelle eingegeben werden, um die Extrahierungsergebnisse auf dem Bildschirm anzuzeigen.

- *Speichern* ist für Zeiteffizienz am vorteilhaftesten. Da der Extrahierungsprozess bei großen Datensätzen eine gewisse Zeit in Anspruch nehmen kann, gewährt das Speichern Ihnen direkten Zugriff auf die Ergebnisse, wenn Sie das Projekt erneut öffnen. Die Wartezeit beim Öffnen des Projekts kann jedoch etwas länger sein.
- *Nicht speichern* wird verwendet, wenn Sie nicht möchten, dass der Antworttext nur in den Originaldatendateien vorkommen soll, selbst wenn Sicherheitsmaßnahmen aktiv sind.

Resource Editor Trennzeichen. Dient zur Auswahl des Zeichens, das bei der Eingabe von Elementen wie beispielsweise Begriffen, Synonymen und optionalen Elementen im Resource Editor als Trennzeichen verwendet werden soll.

Ressourcenvorlage. Wenn Sie kein Textanalysepaket gewählt haben, wird ein Satz an Standardressourcen verwendet. Diese Ressourcen sind in einer Vorlage gespeichert. Klicken Sie auf “Ändern”, um eine andere Standard-Ressourcenvorlage zu wählen. Wählen Sie dann im Dialogfeld “Vorlage ändern” die Tabellenzeile für die gewünschte Vorlage aus und klicken Sie auf OK.

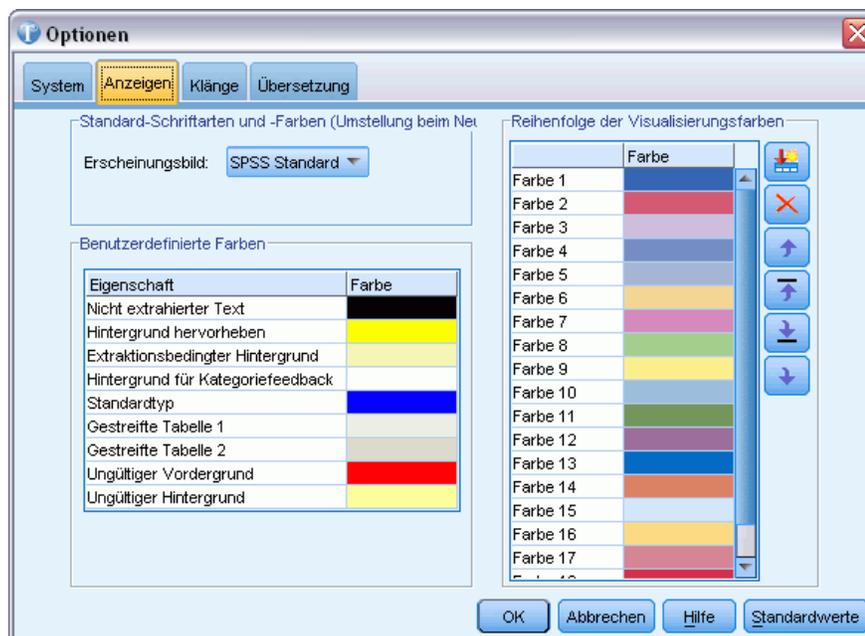
Ländereinstellungen aus System für Sprache der Benutzeroberfläche verwenden. Wählen Sie diese Option, damit SPSS Text Analytics for Surveys die Angaben aus den Ländereinstellungen Ihres Systems für die Sprache der Benutzeroberfläche verwendet. Alternativ können Sie die Auswahl dieser Option aufheben und eine andere Sprache für die Benutzeroberfläche auswählen. Dies kann beispielsweise sinnvoll sein, wenn Sie Informationen analysieren möchten, die in einer anderen Sprache aufgezeichnet wurden als der in der Ländereinstellung Ihres Systems festgelegten und SPSS Text Analytics for Surveys in dieser Sprache ausführen möchten.

Beachten Sie, dass Änderungen an dieser Option erst beim Neustart von SPSS Text Analytics for Surveys wirksam werden.

Optionen: Registerkarte “Anzeige”

Auf dieser zweiten Registerkarte können Sie Optionen bearbeiten, die das allgemeine Erscheinungsbild der Anwendung und die zur Unterscheidung der verschiedenen Elemente verwendeten Farben betreffen.

Abbildung 2-8
Dialogfeld “Optionen”: Registerkarte “Anzeige”



Standard-Schriftarten und -Farben (Umstellung beim Neustart). Die Optionen in diesem Kontrollfeld werden verwendet, um das Farbschema und das Erscheinungsbild anzugeben. Die hier ausgewählten Optionen werden erst nach einem Neustart der Anwendung wirksam. Zur Auswahl stehen folgende Optionen:

- **SPSS Standard (Standardvorgabe),** ein Design, das in bei allen Produkten der Marke SPSS (Teil von IBM Corp.) verbreitet ist.
- **SPSS Classic,** ein Design, mit dem Benutzer früherer Versionen dieser Anwendung vertraut sind.
- **Microsoft Windows,** ein Microsoft Windows-Design das nützlich sein kann, um einen höheren Kontrast bei Stream-Zeichenbereich und Paletten zu erreichen.

Benutzerdefinierte Farben. Dient zum Bearbeiten der Farben für Elemente, die auf dem Bildschirm angezeigt werden. Sie können die Farbe für jedes Element in der Tabelle ändern. Um eine benutzerdefinierte Farbe anzugeben, klicken Sie auf den Farbbereich rechts neben dem zu ändernden Element und wählen Sie in der Dropdown-Farbliste die gewünschte Farbe aus.

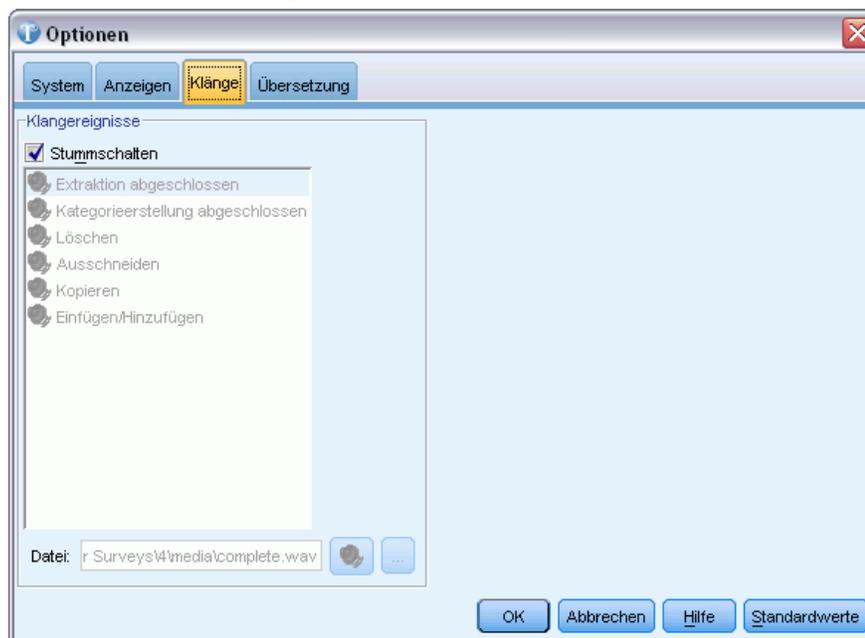
- **Nicht extrahierter Text.** Antworttext, der nicht extrahiert wurde, aber dennoch im Bereich “Daten” sichtbar ist.
- **Hintergrund hervorheben.** Hintergrundfarbe für die Textauswahl (bei der Auswahl von Elementen in den Fensterbereichen oder von Text im Bereich “Daten”).
- **Für Extrahierung erforderlicher Hintergrund.** Hintergrundfarbe des Fensterbereichs “Extrahierungsergebnisse”, die anzeigt, dass Änderungen an den Bibliotheken vorgenommen wurden und eine Extrahierung erforderlich ist.
- **Hintergrund für Kategorie-Feedback.** Farbe des Kategorienhintergrunds, die nach einer Operation wie dem Verschieben und Ablegen von Antworten und dem Erzwingen von Antworten aus dem Bereich “Daten” in den Bereich “Kategorien” angezeigt wird.
- **Standardtyp.** Standardfarbe für Typen und Begriffe, die in den Fensterbereichen “Daten” und “Extrahierungsergebnisse” angezeigt werden. Diese Farbe wird immer dann an der Oberfläche gezeigt, wenn der Typ `Unknown` oder einer der verknüpften Begriffe auftritt. Diese Farbe gilt auch für alle benutzerdefinierten Typen, die Sie im Resource Editor erstellen. Sie können diese Standardfarbe für Ihre benutzerdefinierten Typ-Wörterbücher außer Kraft setzen, indem Sie die Eigenschaften für diese Typ-Wörterbücher in Resource Editor bearbeiten. [Für weitere Informationen siehe Thema Erstellen von Typen in Kapitel 10 auf S. 225.](#)
- **Gestreifte Tabelle 1.** Die erste der beiden Farben, die sich in der Tabelle im Dialogfeld “Erzwungene Fachausdrücke bearbeiten” abwechseln, damit die einzelnen Zeilen besser auseinandergehalten werden können.
- **Gestreifte Tabelle 2.** Die zweite der beiden Farben, die sich in der Tabelle im Dialogfeld “Erzwungene Fachausdrücke bearbeiten” abwechseln, damit die einzelnen Zeilen besser auseinandergehalten werden können.
- **Ungültiger Vordergrund.** Farbe für den Text in doppelten Einträgen im Coderahmenmanager, die einen Fehler anzeigt.
- **Ungültiger Hintergrund.** Farbe für die Hintergrundmarkierung in doppelten Einträgen im Coderahmenmanager, die einen Fehler anzeigt.

Reihenfolge der Visualisierungsfarben. Wenn Sie das Kategorien-Balkendiagramm im Bereich “Visualisierung” und zusätzlich eine Referenzvariable auswählen, sehen Sie die möglichen Werte für die Referenzvariable in einer Legende am unteren Rand des Fensterbereichs. Jeder Wert ist auch mit einer Farbe markiert, damit Sie ihn im Balkendiagramm besser erkennen können. Sie können diese Standardfarben hier ändern. [Für weitere Informationen siehe Thema Visualisierung von Diagrammen in Kapitel 7 auf S. 172.](#)

Optionen: Registerkarte “Klänge”

Auf dieser Registerkarte können Sie die im Produkt verwendeten Klänge angeben. Unter “Klänge” können Sie einen Klang angeben, der als Signal verwendet werden soll, wenn ein Ereignis eintritt. Indem Sie Klänge ein- oder ausschalten oder bestimmte Klänge zuweisen, können Sie steuern, wie Sie auf bestimmte Operationen in der Software hingewiesen werden. Sie können beispielsweise Klänge für Ereignisse wie das Ende des Extrahierungsvorgangs, das Ende einer automatischen Kategorisierungstechnik oder für gängigere Aufgaben wie Ausschneiden, Einfügen, Kopieren und Löschen aktivieren.

Abbildung 2-9
Dialogfeld “Optionen”: Registerkarte “Klänge”



Es stehen mehrere Klänge zur Auswahl zur Verfügung. Mit der Auslassungsschaltfläche (...) wählen Sie einen Klang aus. Die .wav-Dateien, mit denen Klänge für IBM® SPSS® Text Analytics for Surveys erstellt werden, sind im Unterverzeichnis /media des Installationsverzeichnis gespeichert. Wenn keine Klänge abgespielt werden sollen, wählen Sie die Option Alle Klänge stummschalten. Die Klänge sind standardmäßig stummgeschaltet.

Optionen: Registerkarte "Übersetzung"

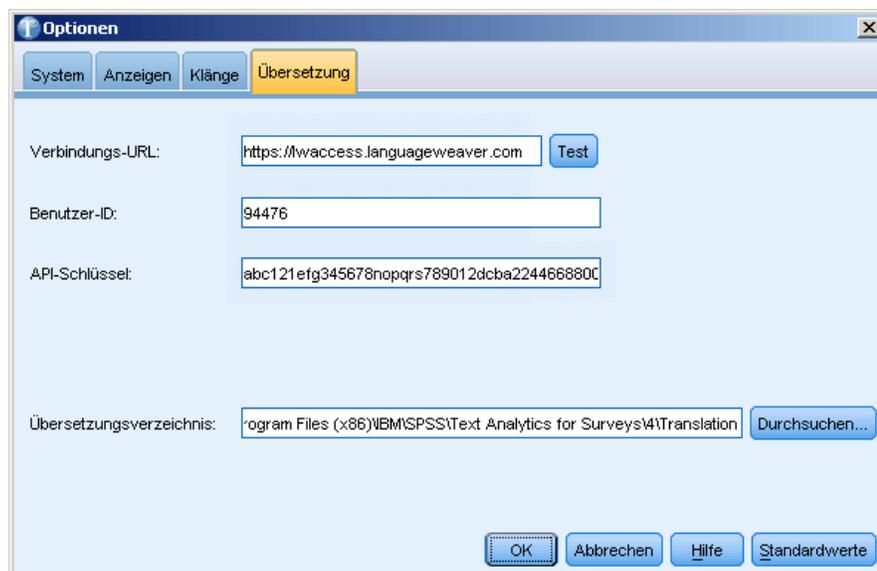
Wichtig: Übersetzungen stehen nur ins Englische zur Verfügung.

Auf dieser Registerkarte können Sie die Verbindung des Language Weaver-Übersetzungsservers definieren und verwalten, die Sie dann bei jeder Übersetzung wieder verwenden können. Sobald Sie hier eine Verbindung definiert haben, können Sie beim Übersetzen rasch eine Sprachpaar-Verbindung wählen, ohne erneut alle Verbindungseinstellungen eingeben zu müssen.

Eine Sprachpaar-Verbindung identifiziert die Quell- und Zielsprache sowie die Verbindungsdetails für den Server. Beispiel: *Chinesisch – Englisch* bedeutet, dass der Quelltext in Chinesisch ist und die resultierende Übersetzung in Englisch sein wird. Sie müssen die Verbindung für den Language Weaver-Server, auf den Sie über die Language Weaver-Online-Services zugreifen, manuell definieren.

Die Übersetzungsergebnisse werden in dem in diesem Dialogfeld definierten Verzeichnis gespeichert. Sie können Ihre Übersetzungsdateien direkt in diesem Verzeichnis verwalten bzw. hier ein anderes Verzeichnis angeben.

Abbildung 2-10
Dialogfeld "Optionen": Registerkarte "Übersetzung"



Verbindungs-URL. Geben Sie den Server-URL oder die Web-Adresse für den Language Weaver-Online-Server ein.

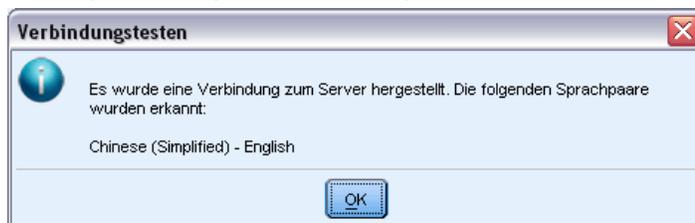
Benutzer-ID. Geben Sie die eindeutige ID ein, die Ihnen von Language Weaver bereitgestellt wurde.

API-Schlüssel. Geben Sie den Schlüssel ein, der Ihnen von Language Weaver bereitgestellt wurde.

Test. Klicken Sie auf Test, um sicherzustellen, dass die Verbindung korrekt konfiguriert ist, und um das bzw. die gefundenen Sprachpaare auf dieser Verbindung anzuzeigen.

Übersetzungsverzeichnis. Klicken Sie auf Durchsuchen, um in ein anderes Verzeichnis zu wechseln, oder geben Sie den Ordnerpfad direkt in das Feld ein.

Abbildung 2-11
Meldung über erfolgreiche Verbindung



Microsoft Internet Explorer-Einstellungen für die Hilfe

Einstellungen für Microsoft Internet Explorer

Die meisten Hilfsfunktionen in dieser Anwendung verwenden Technologie, die auf Microsoft Internet Explorer beruht. Einige Versionen von Internet Explorer (insbesondere die in Microsoft Windows XP, Service Pack 2, bereitgestellte Version) blockieren standardmäßig als "aktiver Inhalt" betrachtete Elemente in Internet Explorer-Fenstern auf dem lokalen Computer. Diese Standardeinstellung kann dazu führen, dass bestimmte Inhalte in Hilfsfunktionen blockiert werden. Um alle Hilfe-Inhalte anzuzeigen, können Sie das Standardverhalten von Internet Explorer ändern.

- ▶ Wählen Sie im Menü des Internet Explorer folgende Optionen:
Extras > Internetoptionen...
- ▶ Klicken Sie auf die Registerkarte Erweitert.
- ▶ Führen Sie einen Bildlauf nach unten zum Abschnitt Sicherheit durch.
- ▶ Aktivieren Sie Ausführung aktiver Inhalte in Dateien auf dem lokalen Computer zulassen.

Teil II:

Textanalyse

Erstellen neuer Projekte und Pakete

Erstellen von Projekten

In IBM® SPSS® Text Analytics for Surveys arbeiten Sie mit Umfragedaten und kategorisieren diese. Dazu erstellen Sie Projekte, in denen Sie die Daten aus Ihrer Datenquelle importieren, einige Variablen auswählen und Kategorien sowie Ressourcen auswählen. Nachdem ein Projekt erstellt ist, können Sie Ihre Ressourcen und Kategorien genauer einstellen, bis Sie annehmen, dass Sie über den endgültigen Satz an Kategorien verfügen. Ein Projekt kann die folgenden Elemente enthalten: Umfragedaten, linguistische Ressourcen, Extrahierungsergebnisse und Kategorien.

Umfragedaten

Auf die importierte Umfragedatenquelle wird innerhalb des Projekts verwiesen, die Umfragedaten selbst werden nicht im Projekt gespeichert. Stattdessen werden beim Öffnen eines Projekts die Umfragedaten erneut aus der originalen Datenquelle eingelesen. Während es sich bei den wichtigsten Variablen in diesem Kontext um offene Textvariablen handelt, wird die Variable für die eindeutige ID beibehalten, ebenso wie etwaige Referenzvariablen (z. B. demografische Variablen), die beim Importieren der Daten angegeben wurden. Werte für alle diese Variablen können im Fensterbereich "Daten" in der Ansicht "Frage" im Textanalysefenster oder in der Ansicht "Gesamtes Projekt" im Textanalysefenster angezeigt werden.

Linguistische Ressourcen

Herstellerspezifische und benutzerdefinierte Bibliotheken mit Listen von Begriffen, Synonymen, ausgeschlossenen Wörtern und Typdeklarationen werden in einem Projekt gespeichert und können geändert werden. Zusätzlich werden bestimmte kompilierte Ressourcen zur Verarbeitung von Text verwendet und ebenfalls im Projekt gespeichert, lassen sich allerdings nicht bearbeiten. Bibliotheken können veröffentlicht werden, wodurch sie innerhalb der Datenbank zur Verwendung in anderen Projekten verfügbar werden. Eine veröffentlichte Bibliothek kann anderen Projekten hinzugefügt werden.

Extrahierte Ergebnisse

Die Extrahierungsergebnisse sind Schlüsselwörter und -phrasen (Konzepte), ihre semantischen Gruppierungen (Typen) und ihre Beziehungen (Muster), die identifiziert und aus den Textantworten extrahiert werden. Diese Extrahierungsergebnisse sind Teil des Projekts und dienen als Grundlage für die Kategorieerstellung. Standardmäßig werden Extrahierungsergebnisse im Projekt gespeichert. Wenn Sie aber befürchten, dass sie die Projektdatei zu sehr vergrößern, können Sie diese Speicherfunktion deaktivieren und beim nächsten Öffnen des Projekts die Daten erneut extrahieren. [Für weitere Informationen siehe Thema Extrahierungsergebnisse speichern in Kapitel 5 auf S. 88.](#)

Kategorien

Textantworten werden in Kategorien verteilt, die automatisch mithilfe von Kategorieerstellungstechniken, manuell durch Ziehen und Ablegen, durch Importieren von Kategoriedefinitionsdateien oder durch Verwendung des Code Frame Manager erstellt werden können. Wenn Sie die Option wählen, die Extrahierungsergebnisse nicht zu speichern, bleiben die Kategoriedefinitionen bei jedem erneuten Öffnen eines Projekts erhalten, jedoch werden Antwortzahlen für sämtliche Teile der Definitionen, die aus den Extrahierungsergebnissen stammten, solange mit einem Fragezeichen(?) angezeigt, bis Sie die Daten erneut extrahieren.

Ihre Daten vorbereiten

Bevor Sie Ihre Daten in IBM® SPSS® Text Analytics for Surveys importieren, bedenken Sie folgende Punkte:

- **Eingabedaten.** Für den Import in SPSS Text Analytics for Surveys muss Ihre Datenquelle einige grundlegende Elemente enthalten, z. B. eine ID-Variable und mindestens eine offene Frage. Die ID-Variable darf nur eindeutige Werte enthalten. Jedes Duplikat verhindert den Import. Sie können mehrere offene Fragen sowie Referenzvariablen importieren. [Für weitere Informationen siehe Thema Auswählen von Variablen auf S. 35.](#)
- **Rechtschreibfehler.** Das Programm erlaubt zwar einige Rechtschreibfehler, aber es wird dennoch empfohlen, dass Sie vor dem Import Ihrer Daten in das Programm die Fehler korrigieren. Rechtschreibfehler können sowohl Mitarbeitern als auch Softwareprogrammen bei der Textanalyse Probleme verursachen. Je mehr Rechtschreibfehler Sie vorab korrigieren können, umso zuverlässiger sind die resultierenden Kategorien. Sie können auch Synonyme mit der korrekten Schreibung eines Worts und den häufig falsch geschriebenen Variationen im Programm erstellen. Tatsächlich sind etliche häufige Rechtschreibfehler in der Core library vordefiniert. Wenn Sie nicht sicher sind, wie viel Aufwand Sie der Rechtschreibprüfung widmen sollen, können Sie Experimente mit einem kleineren Auszug aus den Antworten durchführen, um zu erkennen, wie sehr die Analyse durch Rechtschreibfehler beeinflusst wird.
- **Leere Antworten.** Nicht selten befinden sich leere Antworten in offenen Umfragedaten. Zwar bieten leere Antworten keine Information, aber sie können dennoch nützlich sein. Beispielsweise kann es interessant sein, wie viele Personen eine Frage nicht beantwortet haben oder welche Art von Personen nicht geantwortet hat. Da SPSS Text Analytics for Surveys jedoch Text zur Extrahierung von Begriffen und zum Kategorisieren von Antworten verwendet, können diese leeren Antworten nicht einfach kategorisiert werden.

Eine Methode ist, vor dem Import alle leeren Antworten in Ihren Daten durch das Wort *leer* oder einen anderen passenden Begriff zu ersetzen. Nach dem Datenimport können Sie dann einen neuen Typ anlegen, der eine leere Antwort darstellt, wobei das Wort *leer* (bzw. ein anderes Wort, das Sie eingefügt haben) der Begriff ist, der diesen Typ repräsentiert.

Eine weitere Option ist es, leere Antworten in eine Kategorie zu zwingen. Nach dem Kategorisieren der Antworten sind leere Antworten anfangs nicht kategorisiert. Sie können manuell die Kategorie "Leer" anlegen, indem Sie mit der rechten Maustaste in den Bereich

“Kategorien” klicken. Anschließend können Sie alle leere Antworten auswählen und in die neue Kategorie “Leer” zwingen.

- **Fragen mit Mehrfachantworten.** Gewöhnlich erhalten offene Fragen eine Antwort, jedoch ist dies nicht immer der Fall. Gelegentlich werden offene Fragen als ein Satz an Mehrfachantworten verwendet. Wenn Sie beispielsweise einen Umfrageteilnehmer bitten, auf “Sagen Sie uns drei Dinge, die wir in diesem Hotel verbessern können” zu antworten, und drei separate Bereiche (Variablen) für die Antwort bieten, stellt dies eine Mehrfachantwortenfrage dar.

Da SPSS Text Analytics for Surveys jede Fragevariable separat analysiert, können Sie erneut die Kategorien und linguistischen Ressourcen verwenden, die zur Analyse der ersten Antwort erstellt wurden, um die zweite und dritte Antwort zu kategorisieren. Jedoch ist dies nicht unbedingt die effizienteste Methode. Sie sollten erwägen, alle drei Antwortvariablen in einer Variablen zu kombinieren, bevor Sie die Daten in das Programm importieren. Wenn Sie sie kombinieren, stellen Sie sicher, dass sich mindestens ein Leerschritt oder vorzugsweise ein Punkt zwischen dem letzten Wort in einer Antwort und dem ersten Wort der nächsten Antwort befindet. Da dies bei größeren Datenmengen eine zeitraubende Aufgabe darstellen kann, ist es unter Umständen sinnvoll, die Antworten beim Erstellen der Datendatei zu kombinieren, nicht erst hinterher.

- **Antwortstichproben.** Je größer die Anzahl der Antworten und je länger jede Antwort durchschnittlich ist, umso mehr Zeit kann eine Extrahierung oder Kategorisierung erfordern. Für effizienteres Arbeiten können Sie bei einem großen Dateiumfang (etwa 1.500 Fälle oder mehr) zuerst eine zufällige Stichprobe nehmen und anhand dieser kleineren Untermenge an Antworten einen ersten Analysedurchlauf ausführen.

Eine kleinere Stichprobe ist häufig völlig ausreichend, um zu entscheiden, wie die linguistischen Ressourcen bearbeitet werden sollten. Und sobald die kleinere Datendatei kategorisiert ist, können Sie die komplette Datei einlesen und erneut extrahieren, wodurch automatisch viele der Antworten kategorisiert werden. Im Anschluss können Sie nach Antworten suchen, die nicht in die von Ihnen definierten Kategorien fallen, und nach Bedarf Änderungen vornehmen. Die Größe der zufälligen Stichprobe kann variieren, aber etwa 300 Fälle sind in der Regel ausreichend.

Wichtig: Es gibt noch weitere Überlegungen zum Textanalyseprozess als Ganzes. [Für weitere Informationen siehe Thema Vorbereitung für die Textanalyse in Kapitel 1 auf S. 8.](#)

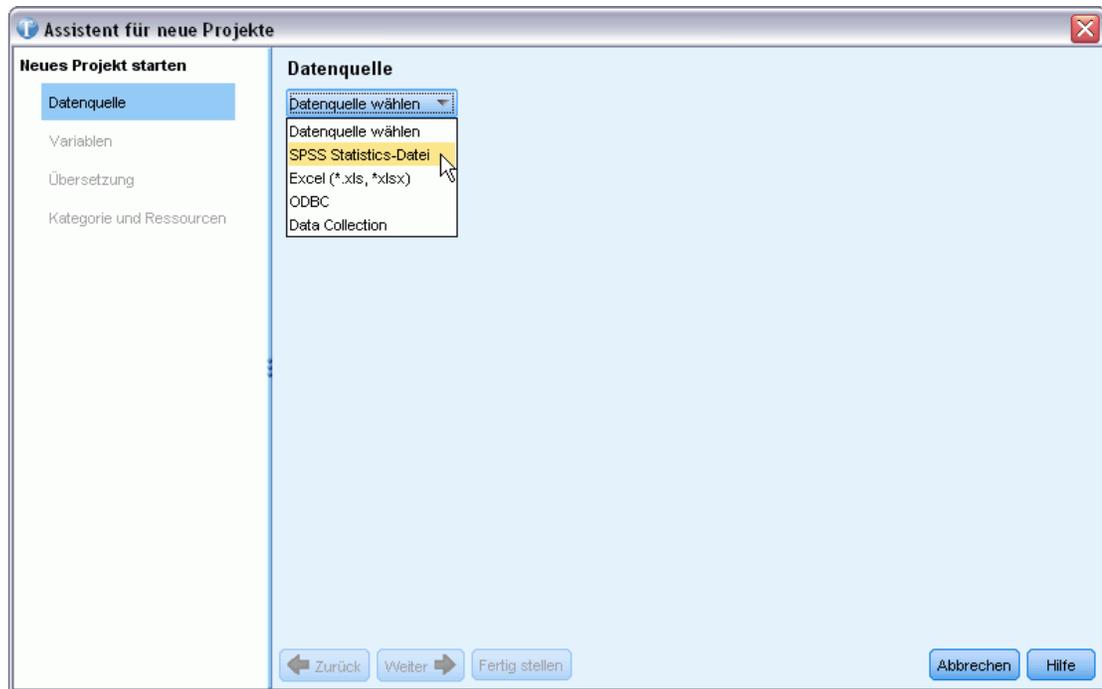
Starten neuer Projekte

Um die Kategorisierung Ihrer Antwortdaten zu beginnen, müssen Sie zunächst ein Projekt anlegen. Ein Assistent begleitet Sie durch die Auswahl der Datenquelle und der Variablen, die Angabe von Kategorien und Ressourcen usw. Bevor Sie mit der Erstellung Ihres Projekts beginnen, sollten Sie Ihre Daten vorbereiten.

So starten Sie ein neues Projekt

- ▶ Wählen Sie in den Menüs die Optionsfolge Datei > Neues Projekt. Klicken Sie alternativ im Startbildschirm auf Neues Projekt starten, wenn keine Projekte geöffnet sind. Der Assistent für neue Projekte wird gestartet.

Abbildung 3-1
Assistent für neue Projekte



- Beginnen Sie durch Auswahl des Typs der Datenquelle aus der Dropdown-Liste “Datenquelle wählen”. Für weitere Informationen siehe Thema [Auswählen von Datenquellen](#) auf S. 30.

Auswählen von Datenquellen

Wenn der Assistent geöffnet wird, beginnen Sie mit der Auswahl einer Datenquelle. IBM® SPSS® Text Analytics for Surveys wurde für die Verarbeitung von Datenmengen von bis zu 10.000 Datensätzen optimiert. Je nachdem, wie viel Text in diesen Datensätzen enthalten ist, kann die Leistung jedoch variieren. Weitere Informationen zu Leistungsstatistiken und Empfehlungen finden Sie in den Installationsanweisungen.

Wichtig: Es muss eine ID-Variable mit einem eindeutigen Wert für jeden Datensatz vorhanden sein, um die Daten zu importieren.

Sie können eine der folgenden Datenquellen wählen:

- **SPSS Statistics-Dateien** (*.sav).
- **Microsoft Excel-Dateien** (*.xls/*.xlsx).
- **ODBC-Datenbank** (Open Database Connectivity-Protokoll von Microsoft).
- **Data Collection-Datenmodell**. Diese Option ist nur verfügbar, wenn das Datenmodell installiert ist.

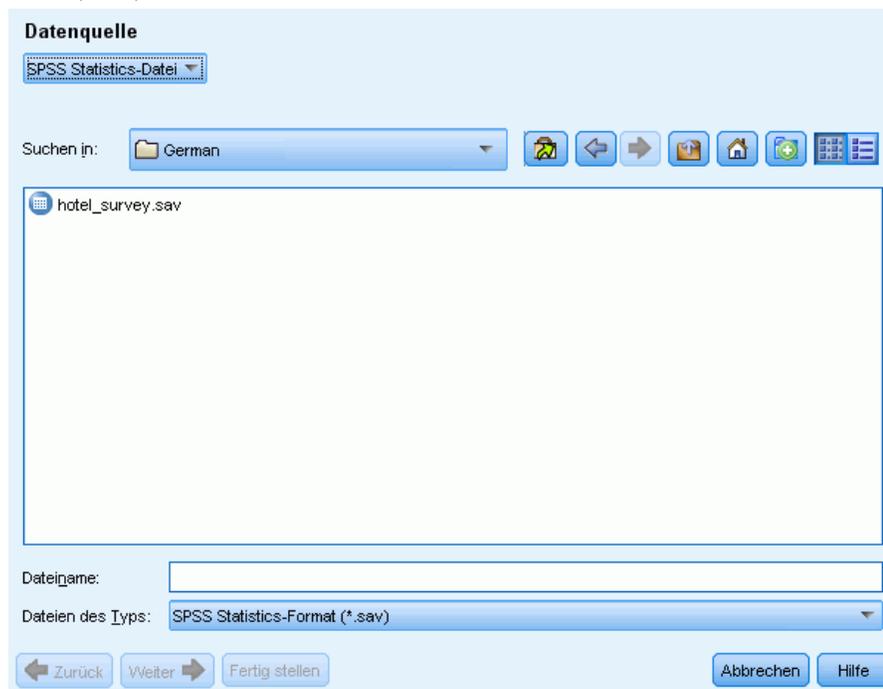
Verwenden von IBM SPSS Statistics-Dateien

Sie können eine IBM® SPSS® Statistics-Datei (.sav) in IBM® SPSS® Text Analytics for Surveys importieren. Eine ID-Variable mit einem eindeutigen Wert für jeden Datensatz muss vorhanden sein, um die Daten zu importieren.

Wichtig: Es können keine SPSS Statistics-(.sav-)Dateien mit Datensätzen importiert werden, die mehr als 4.000 Zeichen umfassen.

Anmerkung: SPSS Text Analytics for Surveys wurde für die Verarbeitung von Datenmengen von bis zu 10.000 Datensätzen optimiert. Je nachdem, wie viel Text in diesen Datensätzen enthalten ist, kann die Leistung jedoch variieren. Weitere Informationen zu Leistungsstatistiken und Empfehlungen finden Sie in den Installationsanweisungen.

Abbildung 3-2
Datenquelloptionen für IBM SPSS Statistics-Dateien



So rufen Sie Daten von IBM SPSS Statistics ab

- ▶ Wählen Sie im ersten Bildschirm des Assistenten die SPSS Statistics-Datei aus der Dropdown-Liste aus. Der Assistent zeigt die Optionen für SPSS Statistics-Dateien an.
- ▶ Wählen Sie in der Dropdown-Liste "Suchen in" das Laufwerk und den Ordner aus, in dem sich die Datei befindet.
- ▶ Wählen Sie die Datei aus der Liste aus. Sie wird im Textfeld "Dateiname" angezeigt.
- ▶ Klicken Sie auf Weiter, um Variablen auszuwählen. [Für weitere Informationen siehe Thema Auswählen von Variablen auf S. 35.](#)

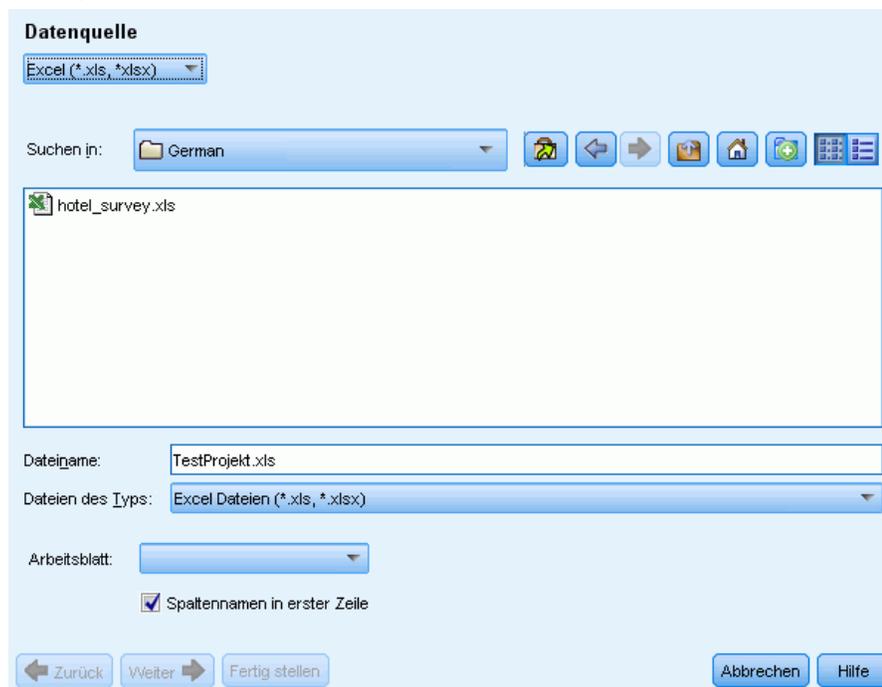
Verwenden von Microsoft Excel-Dateien

Sie können eine Microsoft Excel-Datei (.xls/.xlsx) in IBM® SPSS® Text Analytics for Surveys importieren. Eine ID-Variable mit einem eindeutigen Wert für jeden Datensatz muss vorhanden sein, um die Daten zu importieren.

Wichtig: Beim Importieren der Microsoft Excel-Datei können Sie die Option Spaltennamen in erster Zeile wählen. Um diese Option verwenden zu können, muss die erste Zeile des Arbeitsblatts — nicht die Zeile direkt darüber, wo die Daten beginnen — die Spaltennamen enthalten. Wenn Ihre Daten und Spaltennamen beispielsweise in Zeile 7 beginnen, müssen Sie die Zeilen 1–6 löschen, bevor Sie den Import starten, damit die Datei korrekt importiert wird.

Anmerkung: SPSS Text Analytics for Surveys wurde für die Verarbeitung von Datenmengen von bis zu 10.000 Datensätzen optimiert. Je nachdem, wie viel Text in diesen Datensätzen enthalten ist, kann die Leistung jedoch variieren. Weitere Informationen zu Leistungsstatistiken und Empfehlungen finden Sie in den Installationsanweisungen.

Abbildung 3-3
Datenquelloptionen für Microsoft Excel-Dateien



So rufen Sie Daten von Microsoft Excel ab

- ▶ Wählen Sie im ersten Bildschirm des Assistenten die Option Excel aus der Dropdown-Liste aus. Der Assistent zeigt die Optionen für Microsoft Excel-Dateien an.
- ▶ Wählen Sie in der Dropdown-Liste "Suchen in" das Laufwerk und den Ordner aus, in dem sich die Datei befindet.
- ▶ Wählen Sie die Datei aus der Liste aus. Sie wird im Textfeld "Dateiname" angezeigt.

- ▶ Wählen Sie aus der Dropdown-Liste das Arbeitsblatt aus. Sie können nur Daten aus einem einzelnen Arbeitsblatt importieren. Um mit Daten auf mehreren Arbeitsblättern zu arbeiten, müssen Sie mehrere Projekte erstellen.
- ▶ Wenn die erste Zeile des Arbeitsblatts die Spaltenüberschriften enthält, wählen Sie die Option Spaltennamen in erster Zeile. Um diese Option verwenden zu können, muss die erste Zeile des Arbeitsblatts — nicht die Zeile direkt darüber, wo die Daten beginnen — die Spaltennamen enthalten. Wenn Ihre Daten und Spaltennamen beispielsweise in Zeile 7 beginnen, müssen Sie die Zeilen 1–6 löschen, bevor Sie den Import starten, damit die Datei korrekt importiert wird. Die Anwendung kann diese (oder eine konvertierte Version, falls die Spaltenüberschriften nicht den IBM® SPSS® Statistics-Namensgebungskonventionen für Variablen entsprechen) als Variablennamen verwenden. Anderenfalls verwendet die Anwendung die Spaltenbuchstaben des Arbeitsblatts als ID.
- ▶ Klicken Sie auf Weiter, um Variablen auszuwählen. [Für weitere Informationen siehe Thema Auswählen von Variablen auf S. 35.](#)

Verwenden von Daten über ODBC

Daten aus Datenbankquellen, allgemein Datenbanken, können problemlos in IBM® SPSS® Text Analytics for Surveys importiert werden. Jede Datenbank, die Open Database Connectivity-(ODBC) Treiber verwendet, kann direkt vom Produkt gelesen werden, wenn die entsprechenden Treiber auf dem Rechner installiert sind, auf dem auch SPSS Text Analytics for Surveys installiert ist. Eine ID-Variablen mit einem eindeutigen Wert für jeden Datensatz muss vorhanden sein, um die Daten zu importieren.

Anmerkung: SPSS Text Analytics for Surveys wurde für die Verarbeitung von Datenmengen von bis zu 10.000 Datensätzen optimiert. Je nachdem, wie viel Text in diesen Datensätzen enthalten ist, kann die Leistung jedoch variieren. Weitere Informationen zu Leistungsstatistiken und Empfehlungen finden Sie in den Installationsanweisungen.

Abbildung 3-4
Datenquelloptionen für ODBC

Datenquelle

ODBC

Quelle (DSN)

Name	Beschreibung
MS Access Database	Microsoft Access Driver (*.mdb)
Excel Files	Microsoft Excel Driver (*.xls)
dBASE Files	Microsoft dBase Driver (*.dbf)
Visio Database Samples	Microsoft Access Driver (*.mdb)

ODBC

Nutzer:

Passwort:

Tabelle

SQL

Zurück Weiter Fertig stellen Abbrechen Hilfe

Verwendung über ODBC

- ▶ Wählen Sie im ersten Bildschirm des Assistenten die Option ODBC aus der Dropdown-Liste aus. Der Assistent zeigt die Optionen für ODBC-Dateien an.
- ▶ Legen Sie die Datenquelle fest, indem Sie sie aus der Liste der registrierten ODBC-Quellen auswählen oder den Namen in das Textfeld "Quelle (DSN)" eingeben. Wenn Sie neue Datenquellen registrieren müssen, die nicht in der Liste angezeigt werden, klicken Sie auf ODBC. Damit wird der ODBC-Datenquellen-Administrator geöffnet, der auf den meisten Microsoft Windows-Computern installiert ist. Wenn das nicht der Fall ist, können Sie den ODBC-Import nicht verwenden. Weitere Informationen finden Sie in der Microsoft Windows-Hilfe.
- ▶ Wenn die Datenquelle kennwortgeschützt ist, müssen Sie einen Benutzernamen und ein Kennwort eingeben. Sie werden bei jedem Öffnen des Projekts dazu aufgefordert, da der Benutzername und das Kennwort aus Sicherheitsgründen nicht im Projekt gespeichert werden.
- ▶ Wählen Sie Ihre Daten auf eine der folgenden beiden Arten aus: Tabelle oder SQL. Sie können direkt eine Tabelle auswählen oder SQL-Befehle für die Auswahl der Daten verwenden.
- ▶ Klicken Sie auf Weiter, um Variablen auszuwählen. [Für weitere Informationen siehe Thema Auswählen von Variablen auf S. 35.](#)

Verwenden von IBM SPSS Data Collection-Daten

So importieren Sie über IBM SPSS Data Collection

- ▶ Wählen Sie im ersten Bildschirm des Assistenten die Option Data Collection aus der Dropdown-Liste aus. Die IBM® SPSS® Data Collection-Datenmodelloption ist nur verfügbar, wenn das Datenmodell mit einem anderen Produkt installiert ist.

Auswählen von Variablen

Nach der Auswahl der Datenquelle besteht der nächste Schritt in der Angabe der Variablen, die importiert werden sollen. Drei Variablentypen können in ein Projekt importiert werden.

Eindeutige ID-Variable (erforderlich)

Die ID-Variable ist eine eindeutige numerische Erläuterung oder Texterläuterung, die jeden Teilnehmer identifiziert. Die Datendatei muss nicht nach der eindeutigen ID-Variable sortiert sein, um erfolgreich eingelesen werden zu können. Nach dem Einlesen in das Programm können die Datensätze nach unterschiedlichen Kriterien sortiert werden. [Für weitere Informationen siehe Thema Sortieren von Variablen in Kapitel 4 auf S. 52.](#) Diese ID-Variable ist erforderlich, um Daten zu importieren. Jeder importierte Datensatz (oder Fall) muss einen eindeutigen ID-Wert besitzen.

In folgenden zwei Situationen schlägt der Import fehl:

- Doppelte ID-Werte erkannt,
- Datensätze mit leeren ID-Werten.

Anmerkung: Wenn eine doppelte ID erkannt wird und IBM® SPSS® Statistics auf Ihrem Rechner installiert ist, können Sie die Prozedur "Doppelte Fälle ermitteln" verwenden, um die doppelten IDs zu ermitteln, und anschließend die Optionen verwenden, um zu ermitteln, welche Datensätze beibehalten werden sollen (primäre Fälle).

Variable(n) für offenen Text (erforderlich)

Die Variablen für offenen Text stellen die Textantworten auf die Frage(n) in der Umfrage dar. Mindestens eine dieser Variablen ist erforderlich, um Daten zu importieren. Bei diesen Variablen kann es sich um String- oder lange String-Variablen in SPSS Statistics, Spalten mit allgemeinen Zellen oder Textzellen in Microsoft Excel oder Text- oder Notizenfelder aus Datenbanken handeln. Jede Variable für offenen Text wird separat analysiert. Für jede aus einer .SAV-Datei importierte Textvariable besteht eine Begrenzung von 4.000 Zeichen.

Referenzvariable(n) (optional)

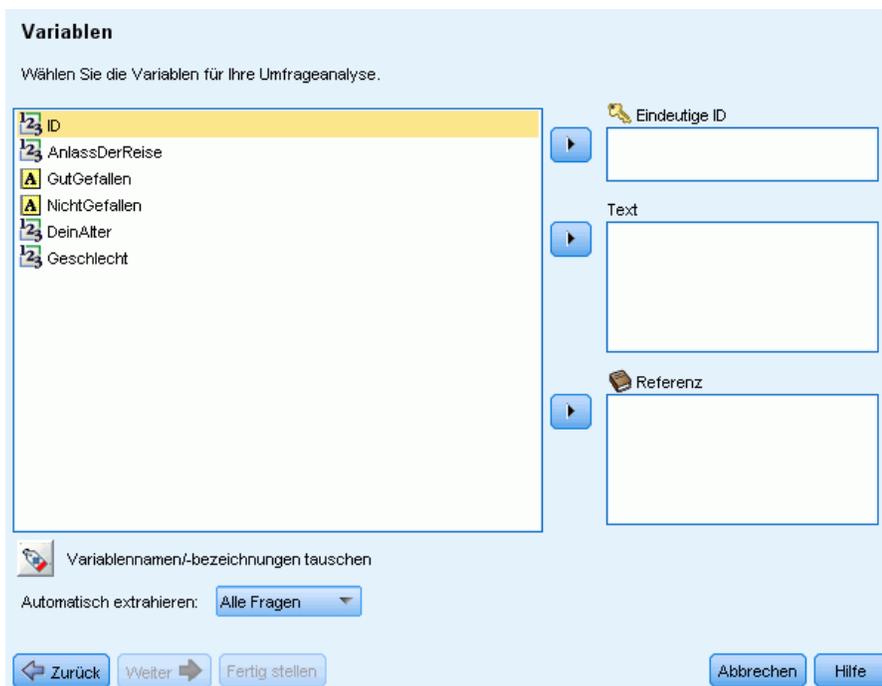
Die Referenzvariablen sind zusätzliche, optionale, in der Regel kategoriale Variablen, die zu Referenzzwecken importiert werden können. Referenzvariablen werden nicht für die Textanalyse verwendet, sondern liefern zusätzliche Informationen, die den Teilnehmer beschreiben und zum Verständnis oder zur Interpretation beitragen können. Demografische Variablen werden oft als Referenzvariablen einbezogen, da Sie zum Verständnis beitragen können, welche Terme oder

Kategorien von welchen Gruppen von Einzelfällen verwendet werden. Beispiele sind Geschlecht, Abteilung, Beruf und Lehrpläne (für Schüler- und Schulungsbeurteilungen). Sie können alle Referenzvariablen anzeigen, wenn sie in die Ansicht “Gesamtes Projekt” importiert wurden. Referenzvariablen können zudem im Bereich “Daten” der Ansicht “Frage” angezeigt werden. Zudem können Sie Referenzvariablen im Balkendiagramm im Visualisierungsbereich auswählen, um eine Untergruppe von Teilnehmern herauszufiltern.

Anmerkung: Für Referenzvariablen, die aus einer SPSS Statistics-Datendatei ausgelesen wurden, werden Variablenlabels (falls vorhanden) als Spaltenüberschriften und die Wertelabels (falls vorhanden) in den Zellen des Bereichs “Daten” angezeigt.

Abbildung 3-5

Auswählen von Variablen



So wählen Sie Variablen und Extrahierungsoptionen aus

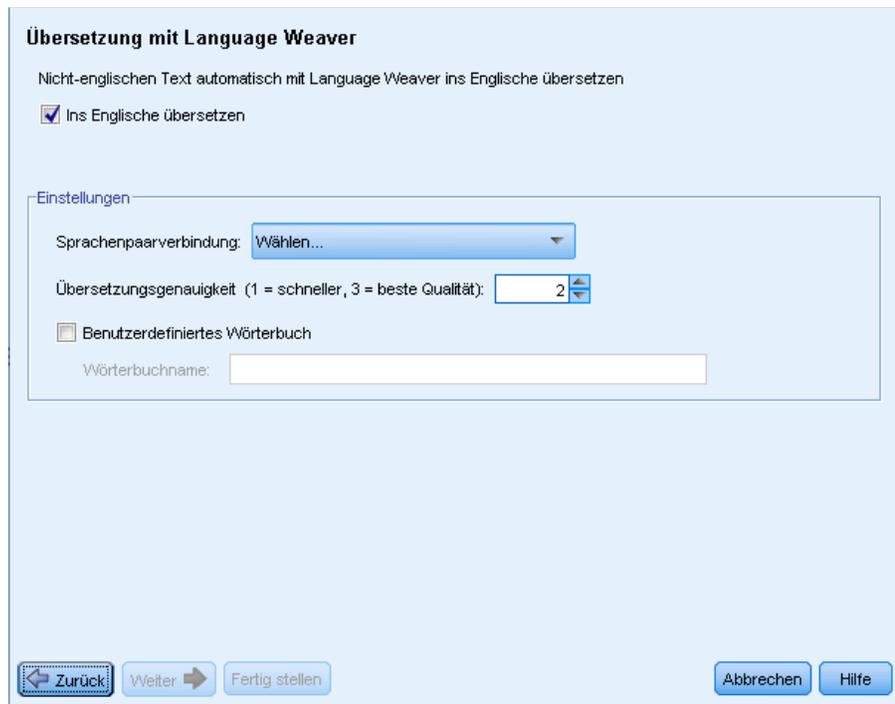
- ▶ Wählen Sie aus der Liste der verfügbaren Variablen die Variable aus, die der ID-Variablen in Ihrer Datenmenge entspricht, und klicken Sie auf die Pfeilschaltfläche, um sie in das Feld “Eindeutige ID” zu verschieben. Die ID muss eine eindeutige Zahl oder ein eindeutiger alphanumerischer String sein, der einen Datensatz von einem anderen unterscheidet. Wenn Ihre Datenmenge doppelte IDs enthält, wird eine Fehlermeldung angezeigt. In diesem Fall müssen Sie Ihre Daten bereinigen, bevor Sie den Versuch wiederholen.
- ▶ Wählen Sie aus der Liste der verfügbaren Variablen eine oder mehrere Variablen aus, die den Variablen für offene Antworten entsprechen, und klicken Sie auf die Pfeilschaltfläche, um die Variable(n) in die Liste “Text” zu verschieben. Die Variablen werden als separate Fragen importiert, deren Antworten Sie analysieren und kategorisieren werden.

- ▶ Wählen Sie aus der Liste der verfügbaren Variablen eine oder mehrere Variablen aus, die den Referenzvariablen entsprechen, und klicken Sie auf die Pfeilschaltfläche, um die Variable(n) in die Liste “Referenz” zu verschieben. Referenzvariablen werden nicht von den automatischen Verfahren zum Aufbau von Kategorien verwendet. Sie können jedoch ihre Inhalte anzeigen und sie verwenden, um bei der Kategorisierung Ihrer Antworten fundiertere Entscheidungen treffen zu können.
- ▶ Um die Variablenlabels anstelle der Variablennamen anzuzeigen, klicken Sie auf die Schaltfläche unter der links angezeigten Variablenliste.
- ▶ Um die Extrahierungseinstellungen zu ändern, wählen Sie eine Option aus der Dropdown-Liste aus. Standardmäßig ist die Option Nur erste Frage ausgewählt, bei der, wenn Sie mehr als eine Variable für offenen Text ausgewählt haben, der Extrahierungsprozess automatisch für die erste Frage gestartet wird, wenn der Assistent beendet ist. Die Extrahierung kann bei größeren Datenmengen einige Zeit in Anspruch nehmen. Daher können Sie je nach verfügbarer Zeit Keine oder Alle Fragen extrahieren.
- ▶ Klicken Sie auf Weiter>, wenn Sie alle gewünschten Variablen ausgewählt haben.

Übersetzen ins Englische

Wenn Sie mit nicht-englischsprachigem Quelltext arbeiten, können Sie eine Verbindung zu Language Weaver herstellen und eine Übersetzung ins Englische vornehmen. Übersetzungen stehen nur ins Englische zur Verfügung. Language Weaver muss ordnungsgemäß konfiguriert und mit Verbindungen definiert sein, damit eine Übersetzung durchgeführt werden kann. [Für weitere Informationen siehe Thema Optionen: Registerkarte “Übersetzung” in Kapitel 2 auf S. 24.](#)

Abbildung 3-6
Übersetzungsoptionen



So nehmen Sie eine Übersetzung ins Englische vor:

- ▶ Um die Textdaten aus einer lizenzierten Sprache ins Englische zu übersetzen, aktivieren Sie das Kontrollkästchen **Ins Englische übersetzen**.
- ▶ Wählen Sie in der Liste **“Sprachpaar-Verbindung** die Verbindung für das zu verwendene Language Weaver-Sprachpaar aus. Wenn Language Weaver auf Ihrem lokalen Computer konfiguriert ist, erscheinen diese Sprachpaare automatisch in dieser Liste. Sie können die Verbindung zu Online-Diensten im Dialogfeld **“Optionen”** auf der Registerkarte **“Übersetzung”** hinzufügen, ändern bzw. testen. [Für weitere Informationen siehe Thema Optionen: Registerkarte “Übersetzung” in Kapitel 2 auf S. 24.](#)
- ▶ Geben Sie die gewünschte Übersetzungsgenauigkeit an. Wählen Sie einen Wert von 1 bis 3, der angibt, welche Geschwindigkeit bzw. Genauigkeit Sie wünschen. Ein niedrigerer Wert führt zu einer schnelleren Übersetzung, jedoch auch zu einer geringeren Genauigkeit. Ein höherer Wert führt zu Ergebnissen mit größerer Genauigkeit, jedoch höherer Verarbeitungszeit. Zur Zeitoptimierung wird empfohlen, mit einer niedrigen Stufe zu beginnen und diese nur zu erhöhen, wenn Sie nach einer Überprüfung der Ergebnisse das Gefühl haben, dass größere Genauigkeit erforderlich ist.
- ▶ Wenn Sie zuvor benutzerdefinierte Wörterbücher erstellt haben, die unter Language Weaver gespeichert sind, können Sie sie zusammen mit der Übersetzung verwenden. Aktivieren Sie zur Auswahl eines benutzerdefinierten Wörterbuchs die Option **Benutzerdefiniertes Wörterbuch**

und geben Sie den Namen unter Wörterbuchname ein. Zur Verwendung mehrerer Wörterbücher, müssen Sie die Namen mit Kommas trennen.

- ▶ Klicken Sie im Assistenten für neue Projekte auf Weiter>, um mit der Auswahl von Kategorien und Ressourcen zu beginnen. [Für weitere Informationen siehe Thema Auswählen von Kategorien und Ressourcen auf S. 39.](#)
- ▶ Klicken Sie im Assistenten zum Ändern des Daten-Sets auf Fertig stellen um die Änderung im Daten-Set abzuschließen und den Übersetzungsvorgang zu starten.

So überspringen Sie die Übersetzung:

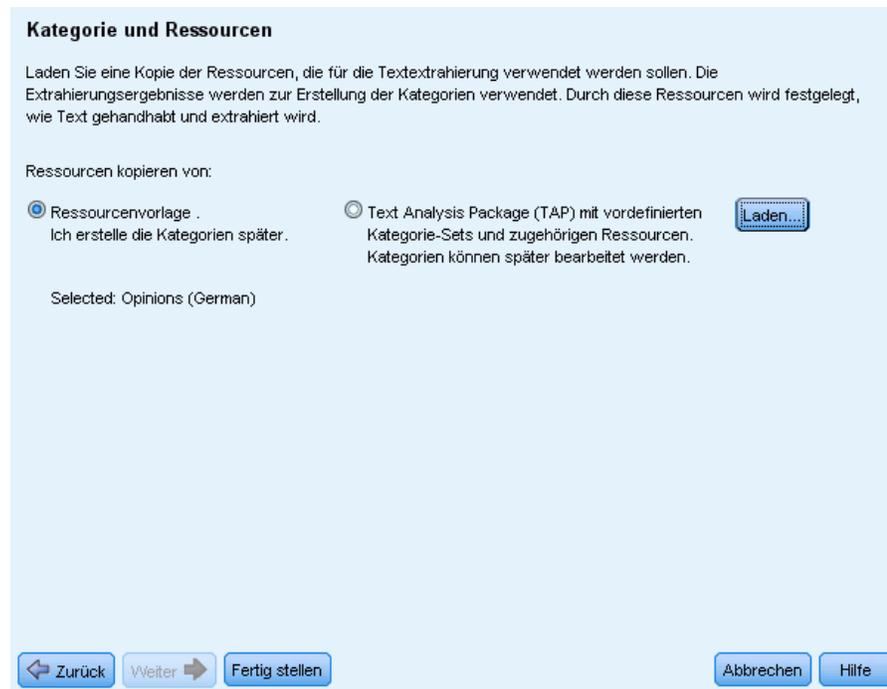
- ▶ Deaktivieren Sie die Option Ins Englische übersetzen.
- ▶ Klicken Sie im Assistenten für neue Projekte auf Weiter>, um mit der Auswahl von Kategorien und Ressourcen zu beginnen. [Für weitere Informationen siehe Thema Auswählen von Kategorien und Ressourcen auf S. 39.](#)
- ▶ Klicken Sie im Assistenten zum Ändern des Daten-Sets auf Fertig stellen um die Änderung im Daten-Set abzuschließen.

Auswählen von Kategorien und Ressourcen

In diesem abschließenden Schritt können Sie die linguistischen Ressourcen auswählen, die verwendet werden sollen, um hervorstechende Konzepte und Muster aus Ihrem Text zu extrahieren. Alternativ können Sie ein Text Analysis Package (TAP) laden, das nicht nur die linguistischen Ressourcen enthält, sondern auch ein oder mehr vordefinierte Kategoriesets, die erweiterte Coderahmen darstellen. [Für weitere Informationen siehe Thema Verwendung von Text Analysis Packages auf S. 43.](#) Mehrere vorgefertigte TAP-Dateien für englischsprachigen Text werden von IBM® SPSS® Text Analytics for Surveys angeboten. Jede der im Lieferumfang dieses Produkt enthaltenen TAP-Dateien ist auf einen bestimmten Umfragetyp zugeschnitten, wie beispielsweise Mitarbeiterumfragen, Produktumfragen oder Umfragen zur Kundenzufriedenheit. Sie können auch Ihre eigenen TAPs für jede beliebige Textsprache erstellen, die vom Produkt unterstützt wird.

Standardmäßig ist eine Ressourcenvorlage bereits geladen. Sie können die vorgeschlagene Standardressourcenvorlage auf der ersten Registerkarte im Dialogfeld "Optionen" ändern. [Für weitere Informationen siehe Thema Festlegen von Optionen in Kapitel 2 auf S. 19.](#) Sie können eine andere Ressourcenvorlage laden oder stattdessen ein TAP auswählen.

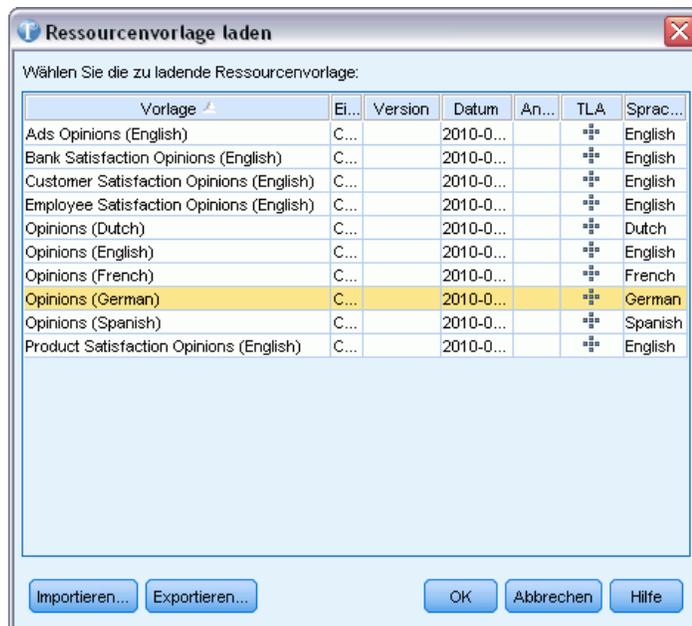
Abbildung 3-7
Auswahl von Ressourcen.



So wählen Sie eine andere Ressourcenvorlage aus:

- ▶ Um eine andere Ressourcenvorlage zu laden, müssen Sie sicherstellen, dass die Option Ressourcenvorlagen ausgewählt ist, und auf Laden klicken. Das Dialogfeld “Ressourcenvorlage laden” wird angezeigt.

Abbildung 3-8
Ressourcenvorlage laden



- ▶ Wählen Sie im Dialogfeld “Ressourcenvorlage laden” die zu verwendende Vorlage aus und klicken Sie auf OK. Das Dialogfeld wird geschlossen und der Assistent zeigt nun die neue Vorlage an, die Sie ausgewählt haben.

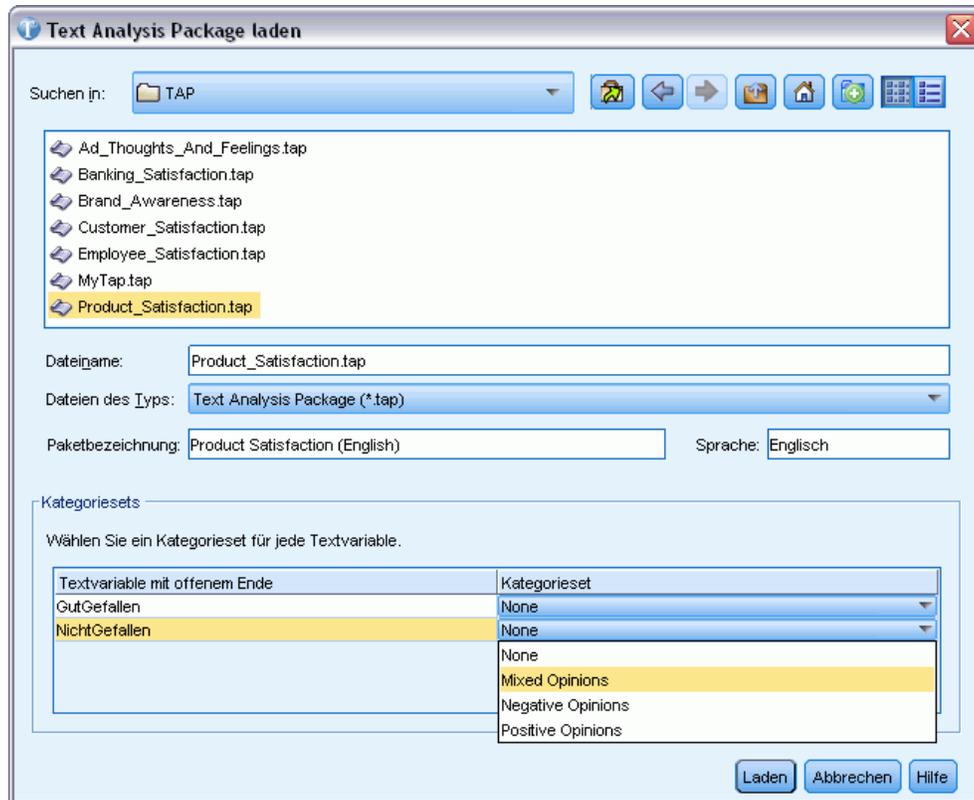
Beachten Sie: Wenn Sie Vorlagen in Sprachen besitzen, für die Sie keine Lizenz haben, wird unten im Dialogfeld ein Kontrollkästchen angezeigt, mit dem Sie die Vorlagen in nicht lizenzierten Sprachen von der Anzeige ausschließen können.

- ▶ Klicken Sie auf Fertig stellen, um mit dem Datenimport zu beginnen. Wenn Sie diese Option wählen, verwendet das resultierende Projekt linguistische Standardressourcen und Sie können nach der Extrahierung Ihre eigenen Kategorien manuell oder mithilfe einer automatisierten Technik erstellen. [Für weitere Informationen siehe Thema Kategorisieren von Textdaten in Kapitel 6 auf S. 96.](#)

So wählen Sie ein Textanalysepaket aus:

- ▶ Um ein Textanalysepaket zu laden, müssen Sie sicherstellen, dass die Option Text Analysis Package ausgewählt ist, und auf Laden klicken. Das Dialogfeld “Text Analysis Package laden” wird angezeigt.

Abbildung 3-9
Text Analysis Package laden



- ▶ Wählen Sie im Dialogfeld das zu verwendende TAP aus. Nur die Pakete, die im Standardverzeichnis <installation_directory>\TAP installiert sind, werden direkt in der Liste angezeigt. Die unten stehenden Felder werden mit den konkreten Details für das ausgewählte TAP aktualisiert.
- ▶ In der Tabelle Kategoriesets können Sie jeder Textvariablen ein Kategorienset zuweisen. Klicken Sie in der Spalte “Kategorieset” auf die Dropdown-Liste in jeder Zelle, um ein verfügbares Kategorienset zu wählen. Wenn Sie Keine wählen, verfügen Sie so lange über keine Kategorien für diese Textvariable, bis Sie später welche erstellen.
- ▶ Klicken Sie auf OK. Das Dialogfeld wird geschlossen und der Assistent zeigt nun das neue TAP an, das Sie ausgewählt haben. Nach Auswahl von TAP und etwaigen Kategoriensets wird der Assistent beendet und nach wenigen Augenblicken sehen Sie Ihre Datensätze in die vordefinierten Kategorien kodiert. Von dort können Sie die Ergebnisse exportieren oder die Kategorien als Ausgangspunkt für Ihre Analyse verwenden.
- ▶ Klicken Sie auf Fertigstellen, um das Dialogfeld zu schließen und das Projekt zu erstellen. Nach der Fertigstellung öffnet die Anwendung automatisch die Ansicht “Frage” für die erste offene Textfrage in Ihrem Projekt. Wenn Sie die Extrahierung wählen, wird ein entsprechendes Dialogfeld geöffnet und es kann einige Zeit dauern, bis der Extrahierungsvorgang abgeschlossen ist. Sie können nun mit der Analyse Ihrer Fragen beginnen. Wählen Sie für den Wechsel zu einer anderen Frage aus den Menüs Ansicht > Frage.

Verwendung von Text Analysis Packages

Ein Text Analysis Package, auch TAP genannt, dient als Vorlage für die Kategorisierung von Textdaten. Der Einsatz eines TAP ist eine einfache Methode zur Kategorisierung Ihrer Textdaten, die minimalen Benutzereingriff erfordert, da das TAP den Coderahmen und die linguistischen Ressourcen enthält, die zur schnellen und automatischen Kodierung einer großen Anzahl von Datensätzen nötig sind. Mit Hilfe der linguistischen Ressourcen werden Textdaten analysiert und dem Mining-Verfahren unterzogen, um die wichtigsten Konzepte zu extrahieren. Auf Basis von wichtigen Konzepten und Mustern, die im Text gefunden werden, können die Datensätze dem Kategorienset zugeordnet werden, das Sie im TAP ausgewählt haben. Sie können Ihr eigenes TAP erstellen oder ein TAP aktualisieren.

Ein TAP besteht aus folgenden Elementen:

- **Kategorienset(s).** Ein Kategorie-Set besteht im Wesentlichen aus vordefinierten Kategorien, Kategoriecodes, Deskriptoren für jede Kategorie und schließlich aus einem Namen für das gesamte Kategorie-Set. Deskriptoren sind linguistische Elemente (Konzepte, Typen, Muster und Regeln), so z. B. der Begriff *billig* oder das Muster *guter Preis*. Deskriptoren werden verwendet, um eine Kategorie zu definieren, sodass der Datensatz dieser Kategorie zugeordnet wird, wenn der Text mit einem Kategoriedeskriptor übereinstimmt.
- **Linguistische Ressourcen.** Linguistische Ressourcen sind eine Reihe von Bibliotheken und erweiterten Ressourcen, die darauf abgestimmt sind, wichtige Konzepte und Muster zu extrahieren. Diese Extrahierungskonzepte und -muster wiederum werden als Deskriptoren verwendet, die es ermöglichen, Datensätze einer Kategorie im Kategorienset zuzuordnen.

Sie können Textanalysepakete erstellen und aktualisieren.

Nachdem das TAP ausgewählt und ein Kategorie-Set für jede Textvariable im Assistenten für neue Projekte gewählt wurde, kann IBM® SPSS® Text Analytics for Surveys Ihre Datensätze extrahieren und kategorisieren. Danach können Sie entweder die Ergebnisse exportieren oder mit der Feinabstimmung der Kategorisierung fortfahren, bis Sie die gewünschten Ergebnisse erhalten.

Anmerkung: TAPs können austauschbar zwischen SPSS Text Analytics for Surveys und IBM® SPSS® Text Analytics verwendet werden.

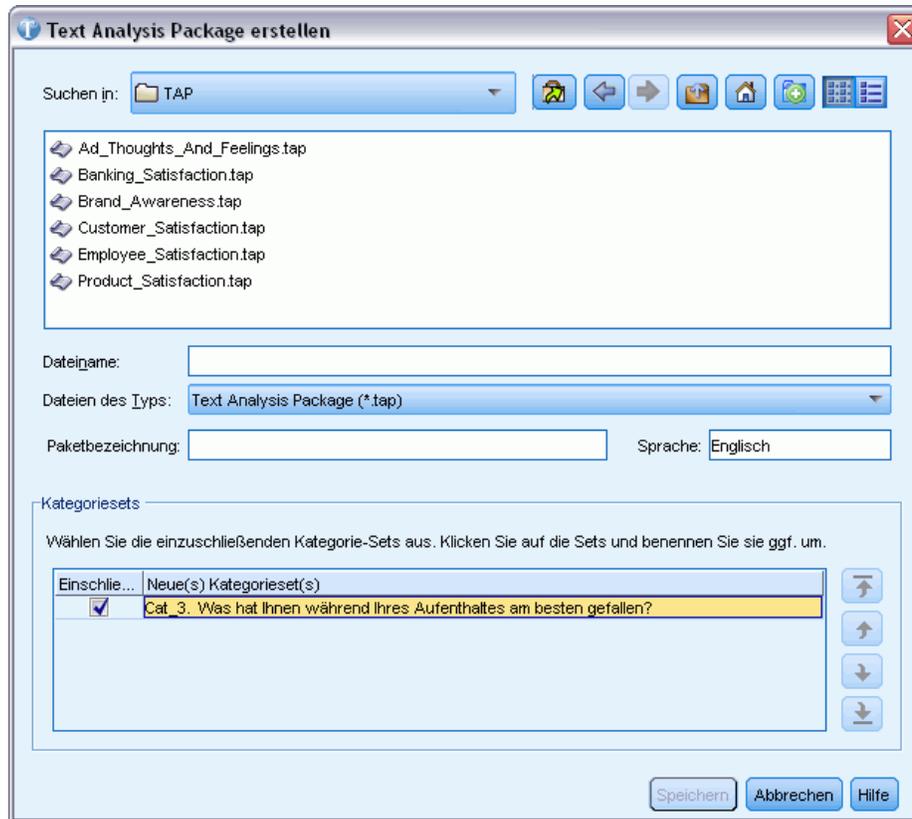
Erstellung von Text Analysis Packages

Wenn Sie ein Projekt mit mindestens einer Kategorie und einigen Ressourcen haben, können Sie aus dem Inhalt der offenen Projektsitzung ein Text Analysis Package (TAP) erstellen. Die Menge der Kategorien und Deskriptoren (Konzepte, Typen, Regeln oder TLA-Musterausgaben) in jeder Frage kann zusammen mit allen linguistischen Ressourcen, die im Ressourceneditor geöffnet sind, zur Erstellung eines TAP verwendet werden.

Sie können die Sprache sehen, für die die Ressourcen erstellt wurden. Die Sprache wird auf der Registerkarte "Erweiterte Ressourcen" von Resource Editor eingestellt.

Wichtig: Falls Ihre Kategorien Textübereinstimmungen, erzwungene Datensätze oder Flags enthalten, werden diese nicht in Kategoriensets gespeichert, da sie datenspezifisch und in beinahe allen Fällen für andere Daten unbrauchbar sind. Bezeichnungen und Kategoriecodes werden jedoch gespeichert.

Abbildung 3-10
Dialogfeld "Text Analysis Package erstellen"

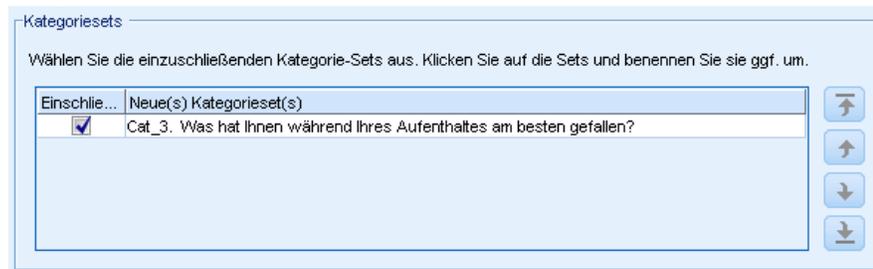


So erstellen Sie ein Text Analysis Package

- ▶ Wählen Sie in den Menüs die Optionsfolge Datei > Text Analysis Packages > Paket erstellen. Das Dialogfeld "Paket erstellen" erscheint.
- ▶ Wechseln Sie zum Verzeichnis, in dem Sie das TAP speichern möchten. Standardmäßig werden TAPs im Unterverzeichnis \TAP des Produktinstallationsverzeichnisses gespeichert.
- ▶ Geben Sie im Feld Dateiname einen Namen für das TAP ein.
- ▶ Geben Sie im Feld Paketbezeichnung eine Bezeichnung ein. Wenn Sie einen Dateinamen eingeben, erscheint dieser Name automatisch als Bezeichnung, die Sie aber ändern können. Wenn Sie ein TAP in diesem Standardverzeichnis speichern, wird die Paketbezeichnung als eine Option im Dropdown-Menü des Assistenten für neue Projekte angezeigt.
- ▶ Um ein Kategorienset aus dem TAP auszuschließen, deaktivieren Sie das Kontrollkästchen Einschließen. Dadurch wird sichergestellt, dass es Ihrem Paket nicht hinzugefügt wird. Standardmäßig ist ein Kategorienset pro Frage im TAP enthalten. Im TAP muss sich immer mindestens ein Kategorienset befinden.
- ▶ Umbenennen von Kategoriensets. Die Spalte Neues Kategorienset enthält standardmäßig generische Namen, die durch Hinzufügen des Präfix Cat_ zum Textvariablennamen generiert

werden. Durch einmaliges Klicken in die Zelle können Sie den Namen bearbeiten. Durch Drücken der Eingabetaste oder Klicken an einer anderen Stelle wird die Umbenennung angewendet. Wenn Sie ein Kategorienset umbenennen, ändert sich der Name nur im TAP; der Variablenname in der offenen Projekt wird nicht geändert.

Abbildung 3-11
Kategoriensets umbenennen

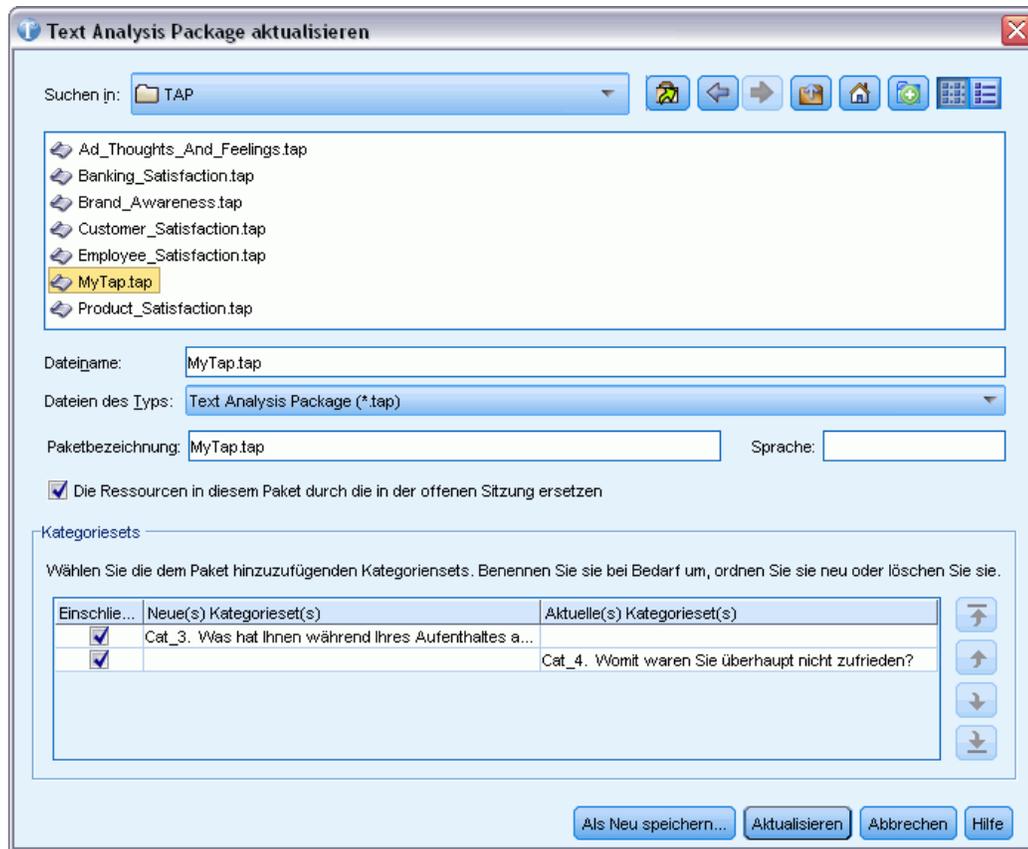


- ▶ Falls gewünscht, ändern Sie die Reihenfolge der Kategoriensets mit Hilfe der Pfeiltasten rechts von der Tabelle "Kategorienset".
- ▶ Klicken Sie auf Speichern, um das Text Analysis Package zu erstellen. Das Dialogfeld wird geschlossen.

Aktualisierung von Text Analysis Packages

Um Verbesserungen an einem Kategorie-Set oder an linguistischen Ressourcen vorzunehmen oder um ein völlig neues Kategorienset zu erstellen, können Sie ein Text Analysis Package (TAP) aktualisieren, um die spätere Nutzung dieser Verbesserungen zu erleichtern. Dafür müssen Sie die Projekt geöffnet haben, die die Informationen enthält, die Sie in das TAP aufnehmen möchten. Beim Aktualisieren können Sie wählen, ob Sie Kategoriensets anfügen, Ressourcen ersetzen, die Paketbezeichnung ändern oder Kategoriensets umbenennen/neu ordnen möchten.

Abbildung 3-12
Dialogfeld "Text Analysis Package aktualisieren"



So aktualisieren Sie ein Text Analysis Package

- ▶ Wählen Sie in den Menüs die Optionsfolge Datei > Text Analysis Packages > Paket aktualisieren. Das Dialogfeld "Text Analysis Package aktualisieren" wird angezeigt.
- ▶ Wechseln Sie in das Verzeichnis, das das Text Analysis Package enthält, das Sie aktualisieren möchten.
- ▶ Geben Sie im Feld Dateiname einen Namen für das TAP ein.
- ▶ Um die linguistischen Ressourcen im TAP durch die in der aktuellen Projekt zu ersetzen, wählen Sie die Option Ressourcen in diesem Paket durch die in der offenen Sitzung ersetzen. Es macht für gewöhnlich Sinn, die linguistischen Ressourcen zu aktualisieren, da sie dazu verwendet wurden, die wichtigen Konzepte und Muster zu extrahieren, mit denen die Kategoriedefinitionen erstellt wurden. Durch die Verwendung der aktuellen linguistischen Ressourcen können Sie sicherstellen, dass Sie bei der Kategorisierung Ihrer Datensätze die besten Ergebnisse erzielen. Wenn Sie diese Option nicht auswählen, bleiben die linguistischen Ressourcen, die bereits im Paket vorhanden waren, unverändert.

- ▶ Um nur die linguistischen Ressourcen zu aktualisieren, stellen Sie sicher, dass die Option Ressourcen in diesem Paket durch die in der offenen Sitzung ersetzen aktiviert ist und Sie nur die aktuellen Kategoriensets auswählen, die bereits im TAP vorhanden waren.
- ▶ Um das (die) neue(n) Kategorie-Set(s) aus der offenen Projekt in das TAP zu übernehmen, aktivieren Sie das Kontrollkästchen für jede Kategorie, die hinzugefügt werden soll. Sie können ein Kategorienset, mehrere oder keine Kategoriensets hinzufügen.
- ▶ Um Kategoriensets aus dem TAP zu entfernen, deaktivieren Sie die jeweiligen Einschließen-Kontrollkästchen. Unter Umständen möchten Sie ein Kategorienset, das bereits im TAP vorhanden war, entfernen, da Sie ein verbessertes Set hinzufügen. Deaktivieren Sie zu diesem Zweck das Kontrollkästchen Einschließen für das jeweilige Kategorienset in der Spalte "Aktuelles Kategorienset". Im TAP muss sich immer mindestens ein Kategorienset befinden.
- ▶ Falls nötig, benennen Sie die Kategoriensets um. Durch einmaliges Klicken in die Zelle können Sie den Namen bearbeiten. Durch Drücken der Eingabetaste oder Klicken an einer anderen Stelle wird die Umbenennung angewendet. Wenn Sie ein Kategorienset umbenennen, ändert sich der Name nur im TAP; der Variablenname in der offenen Projekt wird nicht geändert. Wenn zwei Kategoriensets denselben Namen haben, werden ihre Namen rot angezeigt, bis Sie das Duplikat korrigieren.

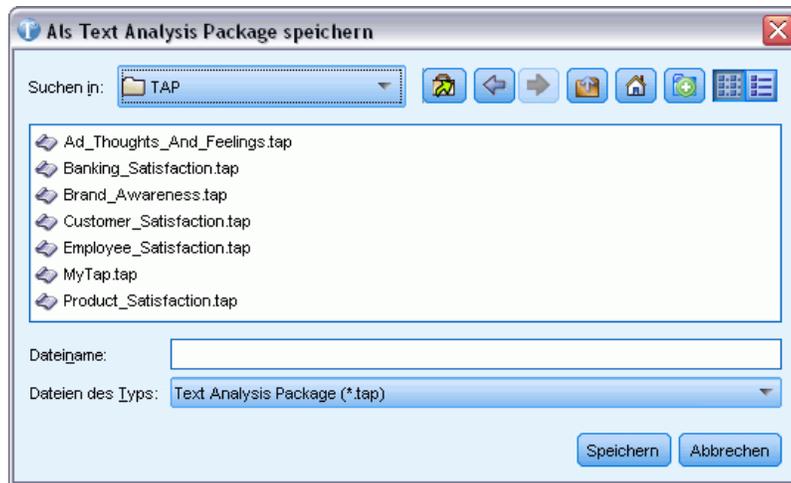
Abbildung 3-13

Doppelt vorhandene Namen

Einschlie...	Neue(s) Kategorienset(s)	Aktuelle(s) Kategorienset(s)
<input checked="" type="checkbox"/>	Cat_3. Was hat Ihnen während Ihres Aufenthaltes a...	
<input checked="" type="checkbox"/>	Cat_4. Womit waren Sie überhaupt nicht zufrieden?	
<input checked="" type="checkbox"/>		Cat_3. Was hat Ihnen während Ihres Aufenthaltes a...
<input checked="" type="checkbox"/>		Cat_4. Womit waren Sie überhaupt nicht zufrieden?

- ▶ Um ein neues Paket mit den Inhalten der Sitzung und den Inhalten des ausgewählten TAP zu erstellen, klicken Sie auf Als neu speichern. Das Dialogfeld "Als Text Analysis Package speichern" wird angezeigt. Beachten Sie die folgenden Anweisungen.
- ▶ Klicken Sie auf Aktualisieren, um die Änderungen, die Sie am ausgewählten TAP vorgenommen haben, zu speichern.

Abbildung 3-14
Dialogfeld "Als Text Analysis Package speichern"



So speichern Sie ein Text Analysis Package

- ▶ Wechseln Sie zum Verzeichnis, in dem Sie die TAP-Datei speichern möchten. Standardmäßig werden TAP-Dateien im Unterverzeichnis "TAP" des Installationsverzeichnisses gespeichert.
- ▶ Geben Sie im Feld "Dateiname" einen Namen für die TAP-Datei ein.
- ▶ Geben Sie im Feld "Paketbezeichnung" eine Bezeichnung ein. Wenn Sie einen Dateinamen eingeben, wird dieser Name automatisch als Bezeichnung verwendet. Sie können diese Bezeichnung jedoch umbenennen. Sie müssen eine Bezeichnung haben. Wenn Sie ein TAP in diesem Standardverzeichnis speichern, wird die Paketbezeichnung als eine Option im Dropdown-Menü des Assistenten für neue Projekte angezeigt.
- ▶ Klicken Sie auf Speichern, um das neue Paket zu erstellen.

Arbeiten mit Projekten

In IBM® SPSS® Text Analytics for Surveys arbeiten Sie mit Umfragedaten und kategorisieren diese. Dafür erstellen Sie Projekte, in denen Sie Kategoriendefinitionen und die entsprechenden Antworten anlegen und speichern. Ein Projekt kann die folgenden Elemente enthalten:

- **Umfragedaten.** Textantwortvariable für offene Fragen, eine eindeutige ID-Variable und ggf. optionale Referenzvariablen. Die Umfragedaten werden nicht im Projekt gespeichert, sondern beim Öffnen des Projekts von der originalen Datenquelle eingelesen.
- **Linguistische Ressourcen.** Herstellerspezifische und benutzerdefinierte Vorlagen und Bibliotheken (Synonyme, Ausschlüsse und Typenwörterbücher), die beim Extrahieren von Konzepten und Mustern aus dem Text verwendet werden.
- **Extrahierte Ergebnisse.** Nach einer Extrahierung vorhanden. Dabei handelt es sich um die Schlüsselwörter und -phrasen, die in Ihren Antwortdaten identifiziert und extrahiert wurden. Sie werden diese Konzepte zum Erstellen Ihrer Kategorien verwenden.
- **Kategorien.** Stammen aus TAP-Kategoriensets, manueller Erstellung und/oder automatisierter Kategorieerstellung. Umfrageantworten werden diesen Kategorien zugewiesen.

Öffnen von Projekten

Sie können zu einem bestehenden Projekt zurückkehren, indem Sie es öffnen. Es kann nur jeweils ein Projekt gleichzeitig geöffnet sein. Wenn Sie versuchen, ein Projekt zu öffnen, während bereits eines geöffnet ist, werden Sie aufgefordert, das andere Projekt falls erforderlich zuvor zu speichern.

Beim Öffnen eines Projekts prüft IBM® SPSS® Text Analytics for Surveys Ihre linguistischen Ressourcen, um festzustellen, ob neuere öffentliche Bibliotheken als die im Projekt vorhanden sind. Wenn dies der Fall ist, müssen Sie angeben, ob die Bibliotheken aktualisiert werden sollen. Sie können dann wählen, Ihre Version ohne Aktualisierung zu behalten oder die Aktualisierung in Ihr Projekt einzulesen. [Für weitere Informationen siehe Thema Bibliotheken aktualisieren in Kapitel 9 auf S. 219.](#)

Wichtig! SPSS Text Analytics for Surveys speichert die Quelldaten nicht physisch in seinen Projekten. Stattdessen wird ein Verweis auf diese Daten auf Ihrem Computer im Projekt gespeichert. Wenn jemand die ursprünglich importierten Variablen in der Datenquelle ändert, wird eine Warnung angezeigt, dass die Daten nicht gefunden werden. In diesem Fall müssen Sie die Daten neu importieren und die “neuen” Fragen mit den ursprünglichen Variablennamen abgleichen, um an diesem Projekt weiterzuarbeiten. Im Allgemeinen wird davon abgeraten, Variablen oder Spaltenüberschriften in Ihrer Datenquelle umzubenennen.

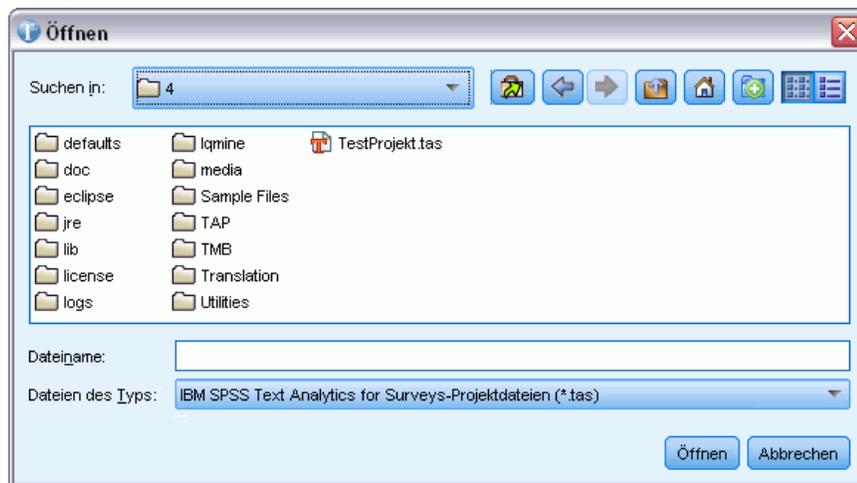
Anmerkung: Die Extrahierungsergebnisse werden in Ihren Projekten gespeichert, es sei denn, Sie wählen eine andere Option (Extras > Optionen). Beim Schließen eines Projekts werden Ihre Kategoriendefinitionen gespeichert, aber der Bereich “Extrahierungsergebnisse” wird gelöscht. Beim erneuten Öffnen dieses Projekts müssen Sie eine Extrahierung durchführen, wenn Sie mit der Kategorisierung Ihrer Antworten fortfahren möchten. Die bestehenden Kategoriendefinitionen

zeigen ein Fragezeichen (?) anstelle einer Antwortzahl. Nach dem erneuten Extrahieren werden die Antwortzahlen wieder angezeigt.

So öffnen Sie ein Projekt

- ▶ Wählen Sie in den Menüs die Optionsfolge Datei > Projekt öffnen. Das Dialogfeld “Öffnen” wird geöffnet.

Abbildung 4-1
Dialogfeld “Öffnen”



- ▶ Wählen Sie aus der Liste das Verzeichnis und den Namen des zu öffnenden Projekts. Sie können die Details wie Dateigröße und Datum in diesem Dialogfeld nicht sortieren.
- ▶ Klicken Sie auf OK, um das Projekt im Hauptfenster zu öffnen.
- ▶ Wenn Ihr Projekt Daten aus einer passwortgeschützten Datenbank enthält, werden Sie bei jedem Öffnen dieses Projekts zur Eingabe des Passworts aufgefordert.
- ▶ Wenn Ihr Projektinhalt aus einer früheren Version des Produkts stammt, werden Sie aufgefordert, Ihre Ressourcen in das neue Format umzuwandeln. Das bedeutet, dass Sie Ihr Projekt nach dem Speichern nicht mehr in einer früheren Version des Produkts öffnen können.
- ▶ Wenn sich seit dem letzten Öffnen des Projekts öffentliche Bibliotheken verändert haben, werden Sie in einer Warnmeldung darüber informiert.

Wichtig: Bei jedem Öffnen eines Projekts wird die entsprechende Datenmenge geöffnet. Wenn die Daten nicht gefunden werden, wird eine Fehlermeldung angezeigt. Damit Sie weiterhin mit Ihren Daten arbeiten können, müssen Sie die Daten erneut importieren. [Für weitere Informationen siehe Thema Ändern von Datenquellen auf S. 63.](#)

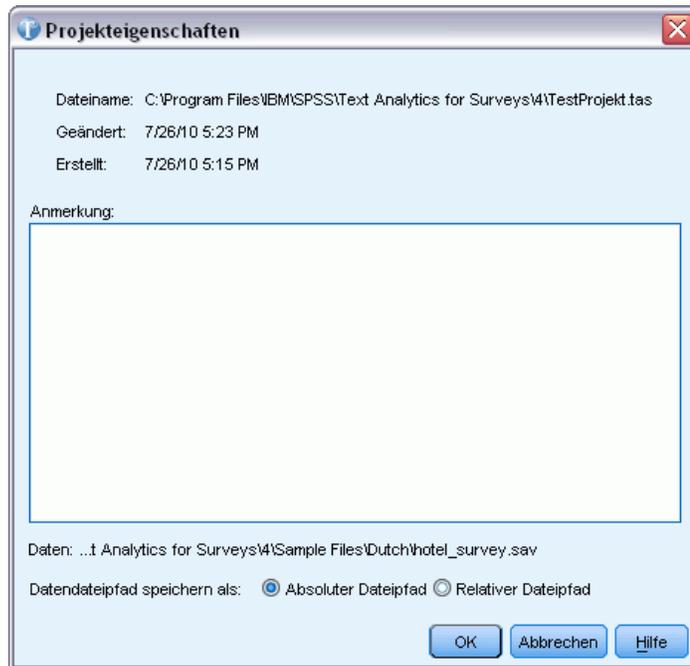
Bearbeiten von Projekteigenschaften

Standardmäßig erhält jedes neue Projekt den Namen *Projekt 1*. Sie können die grundlegenden Eigenschaften Ihres Projekts prüfen und Ihrem Projekt eine Anmerkung hinzufügen oder die Anmerkung ändern.

So bearbeiten Sie Projekteigenschaften

- ▶ Wählen Sie in den Menüs die Optionsfolge Datei > Projekteigenschaften. Das Dialogfeld “Projekteigenschaften” wird geöffnet.

Abbildung 4-2
Dialogfeld “Projekteigenschaften”



- ▶ Geben Sie nach Wunsch einen Kommentar oder eine Beschreibung für das Projekt im Textfeld “Anmerkungen” ein.

Anmerkung: Der Name der Datendatei wird in diesem Dialogfeld angezeigt. Da Sie ein Projekt erstellen und noch keine Daten importiert haben, ist der Dateiname unbekannt. Nach dem Import werden hier nur die letzten 60 Zeichen des Namens der Datendatei angezeigt. Wenn der Dateiname länger ist, können Sie die Maus auf dem Namen platzieren, um den vollständigen Namen zu sehen.

- ▶ Klicken Sie auf OK, um die neuen Eigenschaften zu übernehmen. Das Dialogfeld wird geschlossen und die Projekteigenschaften werden angewendet.

Anzeigen von Projektdaten

Nach dem Import der Datendatei wird automatisch die Ansicht “Frage” für die erste offene Textfrage in Ihrem Projekt angezeigt. Sie sollten jedoch alle importierten Daten ansehen. Dies ist in der Ansicht “Gesamtes Projekt” möglich, in der Sie eine umfassende Darstellung Ihrer Daten erhalten. Wählen Sie zur Anzeige dieser Ansicht Ansicht > Gesamtes Projekt. In dieser Ansicht können Sie folgende Aufgaben ausführen:

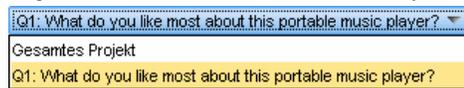
- Inhalt aller importierten Variablen überprüfen.
- Den Variablen Werte und Beschriftungen zuweisen.

- Die Variablentypen ändern.
- Die Variablen sortieren.
- Daten aus benachbarten Zellen kopieren und in andere Anwendungen einfügen.
- Die Breite von Variablenspalten ändern.

Wichtig: Die in das Projekt importierten Daten sind schreibgeschützt. Sie können diese Daten innerhalb von IBM® SPSS® Text Analytics for Surveys nicht bearbeiten.

Sie können dann beginnen, Konzepte aus diesen Antworten zu extrahieren, mit denen Sie Ihre Kategorien erstellen werden. [Für weitere Informationen siehe Thema Kategorisieren von Textdaten in Kapitel 6 auf S. 96.](#)

Abbildung 4-3
Zugriff auf die Ansicht "Gesamtes Projekt"



Sortieren von Variablen

Sie können Ihre Daten in der Ansicht "Gesamtes Projekt" alphabetisch oder nach Datenlänge sortieren.

So sortieren Sie die Daten in der Ansicht "Gesamtes Projekt"

- ▶ Wählen Sie die Spalte aus, die Sie sortieren möchten, und klicken Sie mit der rechten Maustaste auf die Spaltenüberschrift, um ein Kontextmenü zu öffnen.
- ▶ Wählen Sie die gewünschte Sortieroption:
 - **Natürliche Sortierung.** Ergebnisse werden in die Reihenfolge gebracht, in der sie beim Import eingelesen wurden.
 - **Aufsteigend sortieren: A–Z.** Ergebnisse werden alphabetisch sortiert, beginnend mit leeren Zellen, Zahlen und dann *A* bis *Z*.
 - **Absteigend sortieren: A–Z.** Ergebnisse werden alphabetisch sortiert, beginnend mit *Z* bis *A*, anschließend Zahlen und dann leere Zellen.
 - **Aufsteigend sortieren: Länge.** Ergebnisse werden nach Länge sortiert mit den kürzesten Antworten am Beginn.
 - **Absteigend sortieren: Länge.** Ergebnisse werden nach Länge sortiert mit den längsten Antworten am Beginn.

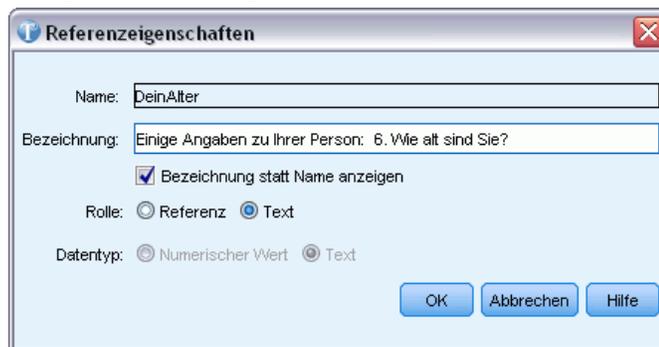
Bearbeiten von Variableneigenschaften

Während Sie Ihre Daten beim Importvorgang definieren, werden Sie aufgefordert, die Variablen für die eindeutigen IDs, die Variablen für die zu analysierenden Fragen und ggf. alle Referenzvariablen anzugeben, die Sie einschließen möchten. Nach dem Datenimport können Sie den Eigenschaften Informationen für diese Variablen hinzufügen oder ihre Rolle im Projekt

ändern. Sie könnten beispielsweise eine Variable analysieren, die Sie als eine Referenzvariable importiert haben. Sie können die folgenden Variableneigenschaften ändern:

- Variablenname oder -bezeichnung hinzufügen oder ändern.
- Referenzvariable in offene Textvariable ändern.
- Offene Textvariable in Referenzvariable ändern.
- Datentyp einer ID oder Referenzvariablen ändern.

Abbildung 4-4
Dialogfeld "Referenzeigenschaften"



So bearbeiten Sie Variableneigenschaften

- ▶ Wählen Sie in der Ansicht "Gesamtes Projekt" die Spalte für die Variable aus, deren Eigenschaften Sie ändern möchten, und klicken Sie mit der rechten Maustaste auf die Spaltenüberschrift, um ein Kontextmenü zu öffnen.
- ▶ Wählen Sie Eigenschaften aus dem Menü. Das Dialogfeld "Eigenschaften" wird geöffnet.
- ▶ Fügen Sie nach Wunsch den Variablennamen oder die Variablenbezeichnung hinzu oder ändern Sie sie.
- ▶ Um die Variablenbezeichnungen anstelle des Variablennamens im Produkt zu verwenden, aktivieren Sie die Option Bezeichnung statt Name anzeigen.
- ▶ Ändern Sie nach Wunsch die Rolle der Variablen in der Analyse zu Referenz oder Offener Text. Die Rolle der ID-Variablen kann nicht geändert werden. Falls Sie die Arbeit an einer offenen Textvariablen (oder Frage) begonnen haben und deren Rolle in eine Referenzvariable ändern, geht sämtliche bis zu diesem Zeitpunkt geleistete Kategorisierungsarbeit verloren.
- ▶ Ändern Sie den Datentyp der Variablen zu Text oder Numerisch.

Speichern von Projekten

Bei jedem Schließen eines Projekts oder Beenden der Sitzung werden Sie aufgefordert, vorgenommene Änderungen zu speichern. Projekte werden mit der Dateinamenerweiterung *.tas in Dateien gespeichert.

So speichern Sie Projekte

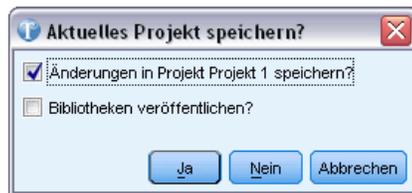
- ▶ Wählen Sie in den Menüs die Optionsfolge Datei > Projekt speichern. Das Projekt wird gespeichert.

So speichern Sie ein Projekt beim Beenden

- ▶ Wenn Sie ein Projekt schließen, werden Sie in einem Dialogfeld gefragt, ob Sie die am Projekt vorgenommenen Änderungen speichern und die Bibliotheken (erneut) veröffentlichen möchten.

Abbildung 4-5

Dialogfeld "Aktuelles Projekt speichern"



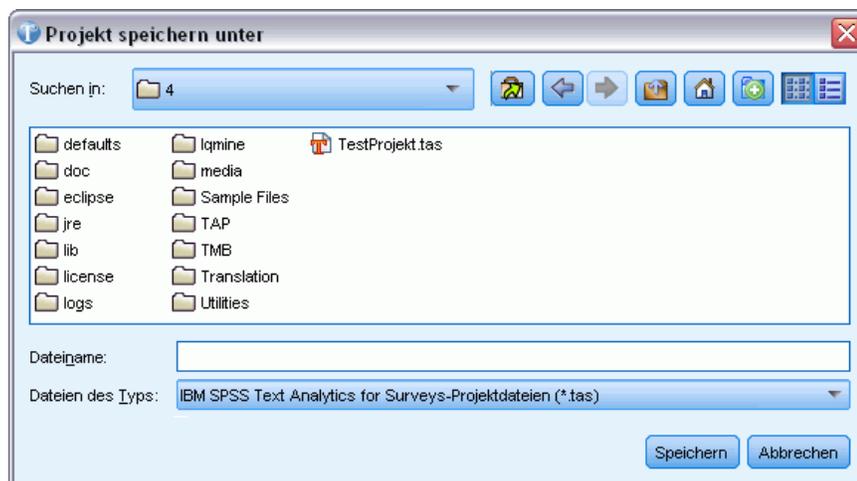
- ▶ Wählen Sie Änderungen an Projekt speichern.
- ▶ Wenn Sie Bibliotheken zur späteren erneuten Verwendung veröffentlichen oder öffentliche Versionen aktualisieren möchten, aktivieren Sie auch Bibliotheken veröffentlichen. Wenn keine Bibliotheken veröffentlicht werden müssen, ist diese Option inaktiv. [Für weitere Informationen siehe Thema Bibliotheken veröffentlichen in Kapitel 9 auf S. 219.](#)
- ▶ Klicken Sie auf Ja, um zu speichern. Wenn Sie die Veröffentlichung von Bibliotheken gewählt haben, wird ein weiteres Dialogfeld geöffnet. [Für weitere Informationen siehe Thema Bibliotheken gemeinsam nutzen in Kapitel 9 auf S. 217.](#)

So speichern Sie das Projekt unter einem anderen Namen

Wenn Sie eine Warnmeldung über einen doppelten Namen erhalten oder das Projekt unter einem anderen Namen speichern möchten, wird das Dialogfeld "Projekt speichern unter" geöffnet.

Abbildung 4-6

Dialogfeld "Projekt speichern unter"



- ▶ Geben Sie den neuen eindeutigen Namen für das Projekt in das Textfeld "Dateiname" ein.
- ▶ Klicken Sie auf Speichern, um den neuen Namen zu speichern.

Exportieren von Kategorisierungsergebnissen

In einigen Fällen kann Ihre Analyse mit der Erstellung von Kategorien enden. Für Textanalysezwecke kann es ausreichen, die wichtigsten Themen zu kennen, die von den Befragten zum Ausdruck gebracht wurden, und zu wissen, wie viele Befragte jeweils ein Thema erwähnten. Häufig sollen jedoch Berichte erstellt werden oder eine weitere Analyse der Kategorien durchgeführt werden, wie beispielsweise das Erstellen von Tabellen und Diagrammen zur Anzeige der Ergebnisse. Möglicherweise möchten Sie auch andere Variablen aus dem Fragebogen verwenden, um die Befragten in den einzelnen Kategorien weiter zu charakterisieren, oder die Kategorien gar zur Untersuchung anderer Antworten verwenden.

Wenn Sie in der Lage sein möchten, die Arbeit mit den neuen Kategorisierungsergebnissen fortzusetzen, können Sie Ihre Kategorien in einem Textformat exportieren, um sie anschließend in eine Anwendung zur quantitativen Analyse (z. B. IBM® SPSS® Statistics Base) zu importieren. Die so entstehende Datei enthält IDs für die Antworten sowie Kategorienamen und -bezeichnungen, jedoch nicht die Werte für Referenzvariablen oder die Freitextantworten.

Anmerkung: Sie können auch Auswertungsdiagramme wie ein Balkendiagramm der Top-5-Kategorien erzeugen. Diese Diagramme, die in HTML exportiert werden, können dann in Präsentationen verwendet werden. [Für weitere Informationen siehe Thema Exportieren von Zusammenfassungsdiagrammen auf S. 60.](#)

Exportierte Dateitypen

Beim Exportieren stehen mehrere Dateitypen zum Erstellen zur Auswahl:

- SPSS Statistics-Dateien (*.sav). [Für weitere Informationen siehe Thema Exportieren in IBM SPSS Statistics oder IBM SPSS Data Collection auf S. 56.](#)
- Microsoft Excel-Dateien (*.xls/*.xlsx). [Für weitere Informationen siehe Thema Exportieren in Microsoft Excel auf S. 58.](#)
- IBM® SPSS® Data Collection. [Für weitere Informationen siehe Thema Exportieren in IBM SPSS Statistics oder IBM SPSS Data Collection auf S. 56.](#) Informationen finden Sie auch in der Data Collection Development Library unter *Data Collection Data Model*.

Die Ausgaben "Dichotomien" und "Kategorien" im Vergleich

Mit IBM® SPSS® Text Analytics for Surveys kodierte Textdaten bilden ein Mehrfachantworten-Set, da die einzelnen Befragten mehrere Antworten geben und für eine Frage mehreren Kategorien zugewiesen werden können. Daher müssen die Daten beim Export in einem speziellen Format kodiert werden. Beim Export stehen zwei verschiedene Ausgabeformate zur Verfügung: **Dichotomien** und **Kategorien**.

Dichotomien. Im Zentrum der Ergebnisse stehen Flags für die Kategoriezugehörigkeit für die einzelnen Antwort-IDs. Für jede Kategorie in den Daten wird mithilfe eines binären Flags, das entweder als "wahr" oder als "falsch" kodiert ist, angegeben, ob die einzelnen Befragten (nach ID)

zu der Kategorie gehören oder nicht. Die Daten sind in einem Tabellenformat strukturiert, mit der ID in der linken Spalte und einer Spalte für jede Kategorie. Bei diesem Datentyp sind beliebig viele Kategorien pro Antwort zulässig. Bei 10 Kategorien gibt es 10 neue Variablen.

Kategorien. Die Ergebnisse konzentrieren sich auf das Kategorien-Set, zu dem eine Antwort gehört. Für jede Antwort in den Daten wird jede Kategorie, der sie zugewiesen ist, entweder als Wert (für SPSS Statistics) oder als die Kategorie selbst (für Microsoft Excel) angezeigt. Die Kategorieexportdaten sind in einem Tabellenformat strukturiert, mit der ID in der linken Spalte, gefolgt von einer Spalte pro Kategorie, zu der mindestens eine Antwort gehört. Diese Spalten stehen nicht für eine bestimmte Kategorie, sondern einen Slot zur Aufzeichnung eines zugewiesenen Kategoriecodes. Für jede Antwort werden die einzelnen Kategoriecodes, zu denen die Antwort gehört, in einem separaten Slot gespeichert. Die Anzahl der zu erstellenden Variablen richtet sich nach der Antwort, der die maximale Anzahl an Kategorien zugewiesen ist. Wenn es 10 Kategorien gibt, aber kein Befragter mit mehr als 4 Kategorien kodiert ist, werden 4 Variablen für die Darstellung der Kategorien benötigt.

- **Für SPSS Statistics/Data Collection.** Für jede Antwort-ID in den Daten werden die einzelnen Kategorien, denen die Antwort-ID zugewiesen ist, als separater Wert von 1 bis N angezeigt, wobei N der höchste Wert für den Kategoriecode ist. Wenn Sie im Coderrahmenmanager keine Codes zugewiesen haben, wurden die Codes automatisch beim Erstellen der Kategorie zugewiesen. Wenn ein Befragter weniger als der maximalen Anzahl an Kategorien zugewiesen wurde, werden die verbleibenden nicht verwendeten Variablen mit dem SPSS Statistics-Wert für “systemdefiniert fehlend” (einem Punkt) kodiert.
- **Für Microsoft Excel.** Bei jeder Antwort-ID in den Daten wird jede Kategorie, der sie zugewiesen ist, entweder als Kategorienname oder Kategoriebezeichnung angezeigt, je nachdem, was Sie in der Benutzeroberfläche des Produkts verwenden. Wenn ein Befragter weniger als der maximalen Anzahl an Kategorien zugewiesen wurde, werden die verbleibenden nicht verwendeten Variablen in Microsoft Excel mit einer Leerzelle kodiert.

Exportieren in IBM SPSS Statistics oder IBM SPSS Data Collection

Nach der Kategorisierung der Antworten sollen sie zumeist mithilfe von statistischen Verfahren analysiert werden. Mit IBM® SPSS® Text Analytics for Surveys können Sie eine Datendatei erstellen, die für die Verwendung in anderen Produkten formatiert ist — die folgenden Anweisungen beziehen sich auf den Export für die Verwendung in IBM® SPSS® Statistics (Programm für statistische Analysen) und verschiedene IBM® SPSS® Data Collection-Produkte. SPSS Text Analytics for Surveys erstellt automatisch die Mehrfachantwortvariable in Ihrer exportierten Datei. Das genaue Format der Datei hängt vom ausgewählten Datentyp ab: Dichotomien oder Kategorien.

Anmerkung: Die so entstehende Datei enthält IDs für die Antworten sowie Kategorienamen und -bezeichnungen, jedoch nicht die Werte für Referenzvariablen oder die Freitextantworten.

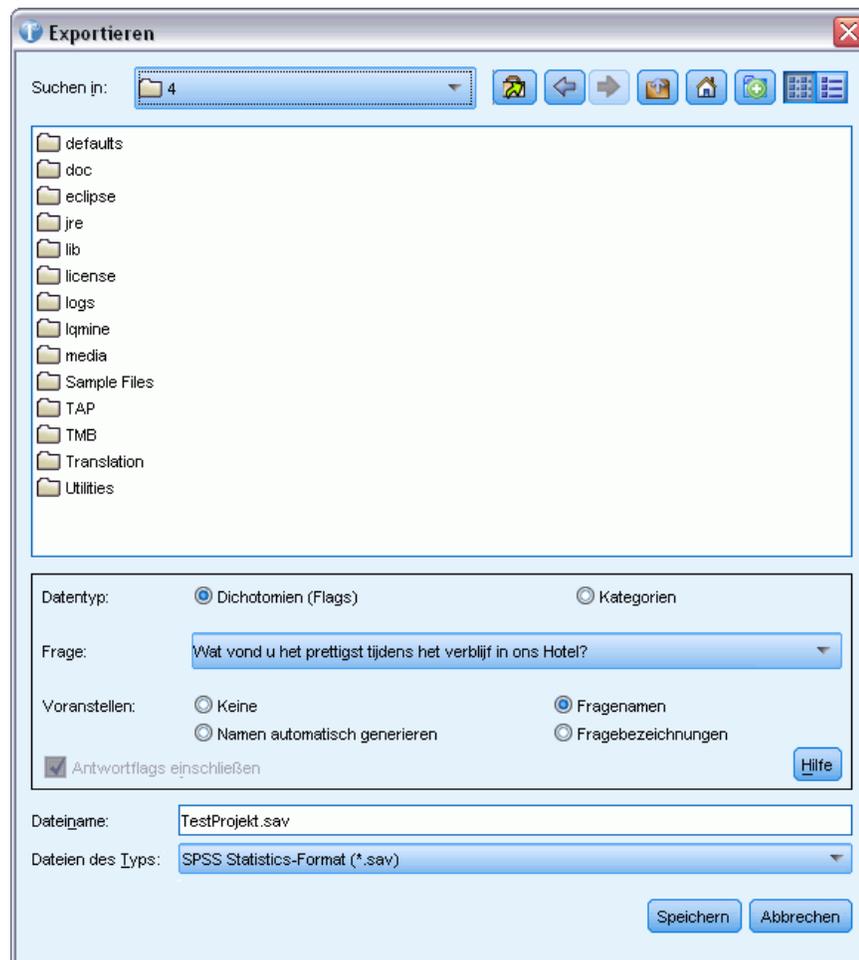
SPSS Statistics nur. Wenn die Datenmenge bei der Ausgabe fehlende Daten oder Fälle enthält, bei denen ein Befragter auf eine bestimmte Frage nicht geantwortet hat, weist die Anwendung diesen Fällen den Wert “systemdefiniert fehlend” zu. SPSS Statistics-Dateien, die aus SPSS Text Analytics for Surveys exportiert wurden, werden von SPSS Statistics-Versionen vor 7.5 nicht unterstützt.

So exportieren Sie Daten:

- ▶ Wählen Sie im Menü Datei > Ergebnisse exportieren eine der folgenden Optionen, um das Dialogfeld “Exportieren” zu öffnen:
 - SPSS Statistics Datei
 - Data Collection Datei

Abbildung 4-7

Dialogfeld “Exportieren” für IBM SPSS Statistics-Dateiformate



- ▶ Wählen Sie in der Dropdown-Liste “Speichern in” das Laufwerk und den Ordner aus, in dem die Datei gespeichert werden soll.
- ▶ Wählen Sie eine Option für den Datentyp. [Für weitere Informationen siehe Thema Exportieren von Kategorisierungsergebnissen auf S. 55.](#)
 - Dichotomien.
 - Kategorien. Diese Option steht für die Data Collection-Datendatei nicht zu Verfügung. Standardmäßig ist “Dichotomien” ausgewählt.

- ▶ Wählen Sie in der Dropdown-Liste “Frage” die zu exportierende Frage aus. Sie können auswählen, ob die Kategorisierungsergebnisse für eine einzelne Frage oder für das gesamte Projekt exportiert werden sollen. Wenn Sie die einzelnen Fragen gesondert exportieren möchten, müssen Sie jeweils immer eine Frage auswählen und exportieren. Alternativ können Sie mit **Gesamtes Projekt** die Ergebnisse für alle offenen Fragen exportieren.
- ▶ Wählen Sie eine Voranstellungsoption aus, um ein Präfix anzugeben, das beim Export der Kategorienamen für das *gesamte Projekt* verwendet werden soll. Diese Option ist besonders nützlich, wenn die Daten für mehrere Fragen exportiert werden sollen. Beim Voranstellen wird dem ursprünglichen Variablenlabel bzw. Kategorienamen ein Präfix beigefügt und sichergestellt, dass keine Dopplungen vorkommen, wenn beim Export des gesamten Projekts die Ergebnisse für mehrere Fragen kombiniert werden. Folgende Optionen stehen zur Auswahl:
 - Keine. Es wird kein Präfix hinzugefügt.
 - Fragenamen. Stellt den Namen der ausgegebenen Kategorievariablen (entweder Kategorienamen oder Kategoriebezeichnung, je nachdem, was Sie im Projekt verwendet haben) dem Variablennamen für offenen Text (Frage) voran. Der Name der Fragevariablen stammt aus der ursprünglichen Datenquelle. Wenn der Name der ausgegebenen Kategorievariablen nicht den Namensgebungskonventionen für Variablen entspricht oder länger ist als 40 Zeichen, werden Standardnamen erstellt (mithilfe der Option für die automatische Generierung).
 - Namen automatisch generieren. Stellt den Kategorienamen automatisch *Q1*, *Q2*, *Q3* usw. voran. *Q1* bezieht sich auf die erste Frage, die Sie exportieren usw.
 - Fragebezeichnungen. Stellt den Namen der ausgegebenen Kategorievariablen (entweder Kategorienamen oder Kategoriebezeichnung, je nachdem, was Sie im Projekt verwendet haben) das Variablenlabel für offenen Text (Frage) voran. Der Name der Fragevariablen stammt aus der ursprünglichen Datenquelle. Wenn der Name der ausgegebenen Kategorievariablen nicht den Namensgebungskonventionen für Variablen entspricht oder länger ist als 40 Zeichen, werden Standardnamen erstellt (mithilfe der Option für die automatische Generierung).
- ▶ Wenn in Ihren Daten Antwortflags vorliegen, können Sie auswählen, ob auch diese exportiert werden sollen. Um Antwortflags zu exportieren, wählen Sie die betreffende Option. [Für weitere Informationen siehe Thema Kennzeichen von Antworten auf S. 76.](#)
- ▶ Wählen Sie im Textfeld “Dateiname” den angezeigten Standardprojektnamen oder geben Sie einen anderen Namen für diese Datei ein.
- ▶ Klicken Sie auf Speichern, um die Ergebnisse zu exportieren.

Exportieren in Microsoft Excel

Nach der Kategorisierung der Antworten sollen sie zumeist mithilfe von statistischen Verfahren analysiert werden. Mit IBM® SPSS® Text Analytics for Surveys können Sie eine Datendatei erstellen, die für die Verwendung in anderen Produkten formatiert ist. Folgende Anweisungen beziehen sich auf den Export in ein Microsoft Excel-Format. SPSS Text Analytics for Surveys erstellt automatisch die Mehrfachantwortvariable in Ihrer exportierten Datei. Das genaue Format der Datei hängt vom ausgewählten Datentyp ab: Dichotomien oder Kategorien. Die so

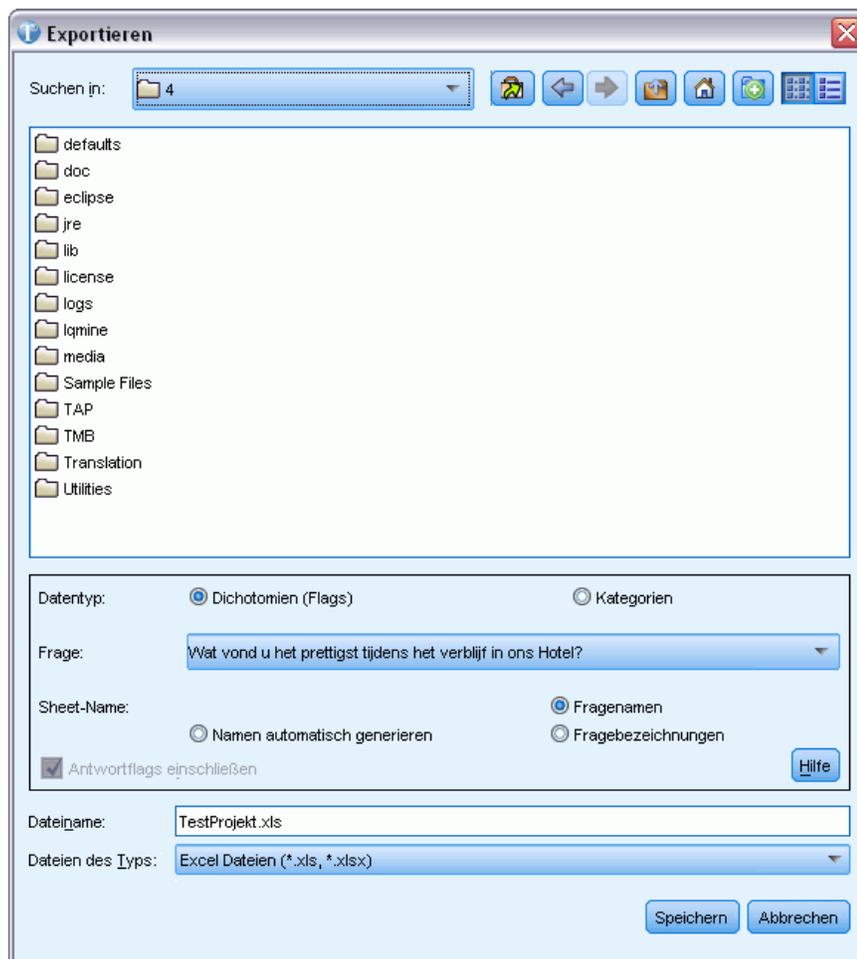
entstehende Datei enthält IDs für die Antworten sowie Kategorienamen und -bezeichnungen, jedoch nicht die Werte für Referenzvariablen oder die Freitextantworten.

So exportieren Sie Daten:

- ▶ Wählen Sie in den Menüs die Optionsfolge Datei > Ergebnisse exportieren > Microsoft Excel-Datei. Das Dialogfeld “Exportieren” wird geöffnet.

Abbildung 4-8

Dialogfeld “Exportieren” für Microsoft Excel-Dateien



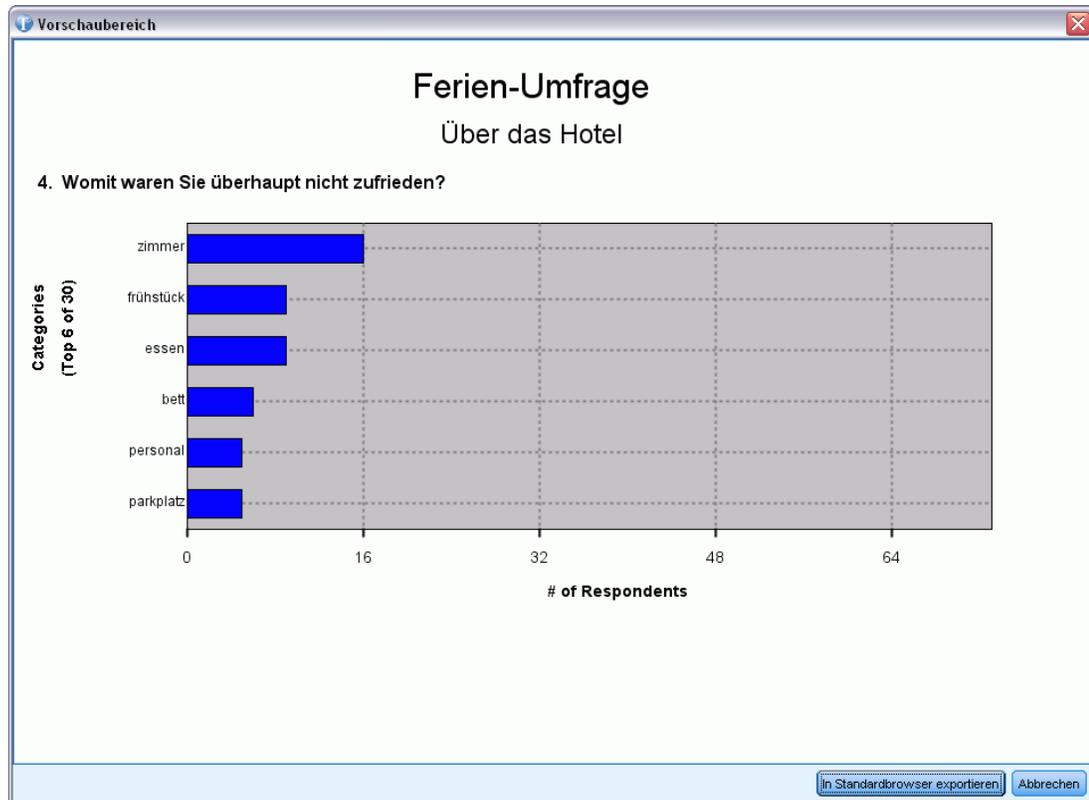
- ▶ Wählen Sie in der Dropdown-Liste “Speichern in” das Laufwerk und den Ordner aus, in dem die Datei gespeichert werden soll.
- ▶ Wählen Sie eine Option für den Datentyp. [Für weitere Informationen siehe Thema Exportieren von Kategorisierungsergebnissen auf S. 55.](#)
 - Dichotomien.
 - Kategorien. Diese Option steht für die IBM® SPSS® Data Collection-Datendatei nicht zu Verfügung. Standardmäßig ist “Dichotomien” ausgewählt.

- ▶ Wählen Sie in der Dropdown-Liste “Frage” die zu exportierende Frage aus. Sie können auswählen, ob die Kategorisierungsergebnisse für eine einzelne Frage oder für das gesamte Projekt exportiert werden sollen. Wenn Sie die einzelnen Fragen gesondert exportieren möchten, müssen Sie jeweils immer eine Frage auswählen und exportieren. Alternativ können Sie mit **Gesamtes Projekt** die Ergebnisse für alle offenen Fragen exportieren.
- ▶ Wählen Sie eine Benennungsoption für Arbeitsblätter, um anzugeben, wie die einzelnen, in der exportierten Microsoft Excel-Datei generierten Arbeitsblätter benannt werden sollen. Folgende Optionen stehen zur Auswahl:
 - **Fragennamen.** Verwendet den Variablennamen für offenen Text (Frage) als Name des Arbeitsblatts. Der Fragename stammt aus der ursprünglichen Datenquelle. Wenn der Name der ausgegebenen Kategorievariablen nicht den Namensgebungskonventionen für Variablen entspricht oder länger ist als 40 Zeichen, werden Standardnamen erstellt (mithilfe der Option für die automatische Generierung).
 - **Namen automatisch generieren.** Benennt die Arbeitsblätter automatisch mit *Q1*, *Q2*, *Q3* usw. *Q1* bezieht sich auf die erste Frage, die Sie exportieren usw.
 - **Fragebezeichnungen.** Verwendet das Variablenlabel für offenen Text (Frage), sofern vorhanden, als Name des Arbeitsblatts. Wenn der Name der ausgegebenen Kategorievariablen nicht den Namensgebungskonventionen für Variablen entspricht oder länger ist als 40 Zeichen, werden Standardnamen erstellt (mithilfe der Option für die automatische Generierung).
- ▶ Wenn in Ihren Daten Antwortflags vorliegen, können Sie auswählen, ob auch diese exportiert werden sollen. Um Antwortflags zu exportieren, wählen Sie die betreffende Option. [Für weitere Informationen siehe Thema Kennzeichen von Antworten auf S. 76.](#)
- ▶ Wählen Sie im Textfeld “Dateiname” den angezeigten Standardprojektname oder geben Sie einen anderen Namen für diese Datei ein.

Exportieren von Zusammenfassungsdiagrammen

Wenn Sie mit der Arbeit an Ihren Kategorien und Daten fertig sind, können Sie grafische Auswertungsberichte exportieren, um Ihre Analyseergebnisse und Erkenntnisse mit anderen zu teilen. Die Ausgabe erzeugt ein Balkendiagramm je Frage. Sie können die Anzahl der Top-Kategorien wählen, die in jedem Diagramm verwendet werden sollen, sodass Sie die Top-5- oder Top-10-Kategorien für eine bestimmte Frage optisch präsentieren können. Das Diagramm kann an Ihren Standardbrowser exportiert werden, von dem aus Sie das Bild zur Weiterverwendung in anderen Produkten oder Präsentationen speichern können.

Abbildung 4-9
Beispielzusammenfassungsdiagramme in einem Browserfenster



So exportieren Sie Zusammenfassungsdiagramme

- ▶ Wählen Sie in den Menüs die Optionsfolge Kategorien > Zusammenfassungsdiagramme exportieren. Das Dialogfeld "Zusammenfassungsdiagramme exportieren" wird geöffnet.
- ▶ Konfigurieren Sie Ihr Zusammenfassungsdiagramm entsprechend den in diesem Thema beschriebenen Optionen.
- ▶ Klicken Sie auf Aufbauen, um das Diagramm zu erzeugen und es im Vorschaubereich anzuzeigen.
- ▶ Klicken Sie auf In Standardbrowser exportieren, um das Diagramm in einem Browserfenster anzuzeigen.

Abbildung 4-10
Dialogfeld "Zusammenfassingsdiagramme exportieren"

Berichtsoptionen

Diagramm(e) exportieren für. Wählen Sie, ob Sie ein Zusammenfassingsdiagramm für alle Fragen in Ihrem Projekt oder nur die aktuell ausgewählte Frage erzeugen möchten.

Anzahl an einzuschließenden Top-Kategorien. Wählen Sie die maximale Anzahl an im Diagramm anzuzeigenden Kategorien. Die Kategorien mit der höchsten Anzahl an Datensätzen werden zuerst verwendet.

Neu extrahieren, wenn veraltete Extrahierungsergebnisse festgestellt werden. Wählen Sie diese Option, um eine neue Extrahierung zu erzwingen, bevor das Diagramm erzeugt wird, wenn die Extrahierungsergebnisse nicht aktuell sind.

Haupttitel. Geben Sie einen Haupttitel für Ihre Diagramme ein. Das könnte zum Beispiel der Name Ihrer Umfrage sein.

Untertitel. Geben Sie einen Untertitel für Ihre Diagramme ein. Das könnte zum Beispiel der Name der Firma oder das Jahr der Umfrage sein.

Frage-titel. Um bei der Identifizierung jedes Diagramms zu helfen, wird der Titel aus der Frage abgeleitet. Sie können den Variablennamen der Frage, die Bezeichnung oder keinen Namen verwenden.

X-Achsen-Bezeichnung. Definieren Sie eine Bezeichnung für die X-Achse der Diagramme. Standardmäßig wird eine Bezeichnung vorgeschlagen.

Y-Achsen-Bezeichnung: Definieren Sie eine Bezeichnung für die Y-Achse der Diagramme. Standardmäßig wird eine Bezeichnung vorgeschlagen.

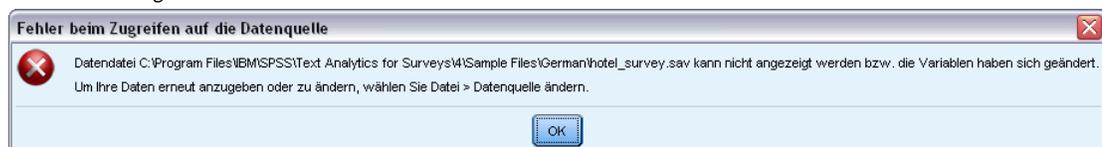
Balkenfarbe. Wählen Sie eine Farbe für die Balken in den Zusammenfassungsdiagrammen. Diese Farbe gilt für alle Fragen.

Schriftart. Wählen Sie eine Schriftart für die Titel und Bezeichnungen im Diagramm.

Ändern von Datenquellen

Bei jedem Öffnen eines Projekts wird die entsprechende Datenmenge geöffnet. Wenn die Daten nicht gefunden werden, wird eine Fehlermeldung angezeigt. Gelegentlich werden Daten nicht gefunden, weil sie an einen anderen Speicherort verschoben, versehentlich gelöscht oder umbenannt wurden. Möglicherweise möchten Sie auch einfach die Datenquellen ändern.

Abbildung 4-11
Fehlermeldung über fehlende Daten

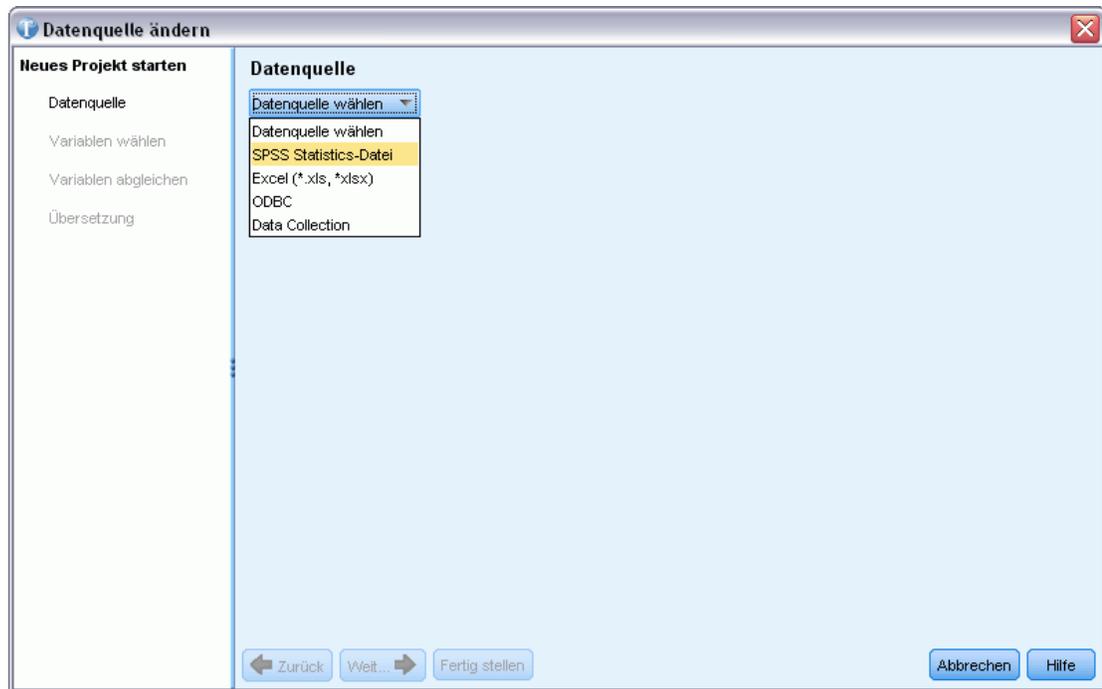


Um mit Ihren Daten weiterzuarbeiten, müssen Sie den Speicherort zur korrekten Datenquelle ändern. Wenn in den Daten Variablenänderungen festgestellt werden, etwa neue Variablen, umbenannte Variablen oder fehlende Variablen, werden Sie aufgefordert, die zuvor importierten Variablen mit den neuen Variablen abzugleichen.

So ändern Sie Ihre Datenquelle:

- ▶ Wenn Sie diese Fehlermeldung erhalten, klicken Sie auf OK.
- ▶ Wählen Sie in den Menüs die Optionsfolge Datei > Datenquelle ändern. Das Dialogfeld "Datenquelle ändern" wird angezeigt.

Abbildung 4-12
Assistent zum Ändern der Datenquelle



Auswählen von Datenquellen

Wenn der Assistent geöffnet wird, beginnen Sie mit der Auswahl einer Datenquelle. IBM® SPSS® Text Analytics for Surveys wurde für die Verarbeitung von Datenmengen von bis zu 10.000 Datensätzen optimiert. Je nachdem, wie viel Text in diesen Datensätzen enthalten ist, kann die Leistung jedoch variieren. Weitere Informationen zu Leistungsstatistiken und Empfehlungen finden Sie in den Installationsanweisungen.

Wichtig: Es muss eine ID-Variable mit einem eindeutigen Wert für jeden Datensatz vorhanden sein, um die Daten zu importieren.

Sie können eine der folgenden Datenquellen wählen:

- **SPSS Statistics-Dateien** (*.sav).
- **Microsoft Excel-Dateien** (*.xls/*.xlsx).
- **ODBC-Datenbank** (Open Database Connectivity-Protokoll von Microsoft).
- **Data Collection-Datenmodell**. Diese Option ist nur verfügbar, wenn das Datenmodell installiert ist.

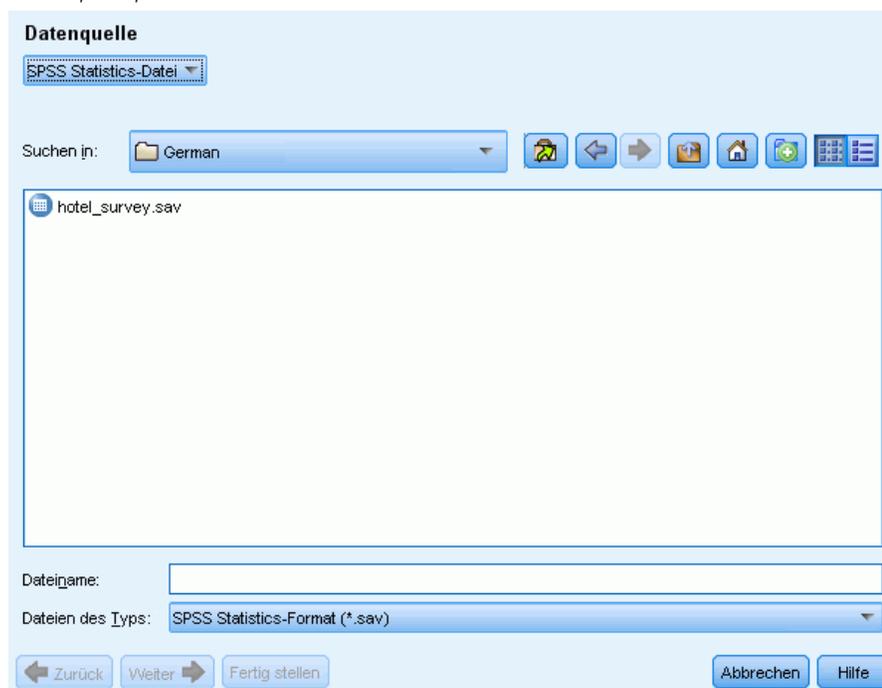
Verwenden von IBM SPSS Statistics-Dateien

Sie können eine IBM® SPSS® Statistics-Datei (.sav) in IBM® SPSS® Text Analytics for Surveys importieren. Eine ID-Variable mit einem eindeutigen Wert für jeden Datensatz muss vorhanden sein, um die Daten zu importieren.

Wichtig: Es können keine SPSS Statistics-(.sav-)Dateien mit Datensätzen importiert werden, die mehr als 4.000 Zeichen umfassen.

Anmerkung: SPSS Text Analytics for Surveys wurde für die Verarbeitung von Datenmengen von bis zu 10.000 Datensätzen optimiert. Je nachdem, wie viel Text in diesen Datensätzen enthalten ist, kann die Leistung jedoch variieren. Weitere Informationen zu Leistungsstatistiken und Empfehlungen finden Sie in den Installationsanweisungen.

Abbildung 4-13
Datenquelloptionen für IBM SPSS Statistics-Dateien



So rufen Sie Daten von IBM SPSS Statistics ab

- ▶ Wählen Sie im ersten Bildschirm des Assistenten die SPSS Statistics-Datei aus der Dropdown-Liste aus. Der Assistent zeigt die Optionen für SPSS Statistics-Dateien an.
- ▶ Wählen Sie in der Dropdown-Liste "Suchen in" das Laufwerk und den Ordner aus, in dem sich die Datei befindet.
- ▶ Wählen Sie die Datei aus der Liste aus. Sie wird im Textfeld "Dateiname" angezeigt.
- ▶ Klicken Sie auf Weiter, um Variablen auszuwählen. [Für weitere Informationen siehe Thema Auswählen von Variablen auf S. 35.](#)

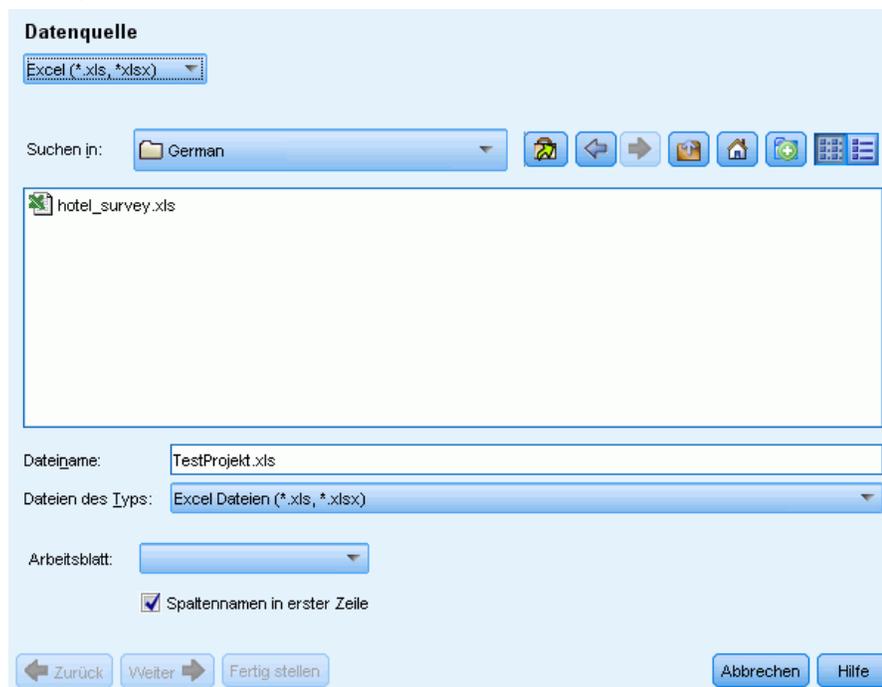
Verwenden von Microsoft Excel-Dateien

Sie können eine Microsoft Excel-Datei (.xls/.xlsx) in IBM® SPSS® Text Analytics for Surveys importieren. Eine ID-Variable mit einem eindeutigen Wert für jeden Datensatz muss vorhanden sein, um die Daten zu importieren.

Wichtig: Beim Importieren der Microsoft Excel-Datei können Sie die Option Spaltennamen in erster Zeile wählen. Um diese Option verwenden zu können, muss die erste Zeile des Arbeitsblatts — nicht die Zeile direkt darüber, wo die Daten beginnen — die Spaltennamen enthalten. Wenn Ihre Daten und Spaltennamen beispielsweise in Zeile 7 beginnen, müssen Sie die Zeilen 1–6 löschen, bevor Sie den Import starten, damit die Datei korrekt importiert wird.

Anmerkung: SPSS Text Analytics for Surveys wurde für die Verarbeitung von Datenmengen von bis zu 10.000 Datensätzen optimiert. Je nachdem, wie viel Text in diesen Datensätzen enthalten ist, kann die Leistung jedoch variieren. Weitere Informationen zu Leistungsstatistiken und Empfehlungen finden Sie in den Installationsanweisungen.

Abbildung 4-14
Datenquelloptionen für Microsoft Excel-Dateien



So rufen Sie Daten von Microsoft Excel ab

- ▶ Wählen Sie im ersten Bildschirm des Assistenten die Option Excel aus der Dropdown-Liste aus. Der Assistent zeigt die Optionen für Microsoft Excel-Dateien an.
- ▶ Wählen Sie in der Dropdown-Liste “Suchen in” das Laufwerk und den Ordner aus, in dem sich die Datei befindet.
- ▶ Wählen Sie die Datei aus der Liste aus. Sie wird im Textfeld “Dateiname” angezeigt.

- ▶ Wählen Sie aus der Dropdown-Liste das Arbeitsblatt aus. Sie können nur Daten aus einem einzelnen Arbeitsblatt importieren. Um mit Daten auf mehreren Arbeitsblättern zu arbeiten, müssen Sie mehrere Projekte erstellen.
- ▶ Wenn die erste Zeile des Arbeitsblatts die Spaltenüberschriften enthält, wählen Sie die Option Spaltennamen in erster Zeile. Um diese Option verwenden zu können, muss die erste Zeile des Arbeitsblatts — nicht die Zeile direkt darüber, wo die Daten beginnen — die Spaltennamen enthalten. Wenn Ihre Daten und Spaltennamen beispielsweise in Zeile 7 beginnen, müssen Sie die Zeilen 1–6 löschen, bevor Sie den Import starten, damit die Datei korrekt importiert wird. Die Anwendung kann diese (oder eine konvertierte Version, falls die Spaltenüberschriften nicht den IBM® SPSS® Statistics-Namensgebungskonventionen für Variablen entsprechen) als Variablennamen verwenden. Anderenfalls verwendet die Anwendung die Spaltenbuchstaben des Arbeitsblatts als ID.
- ▶ Klicken Sie auf Weiter, um Variablen auszuwählen. [Für weitere Informationen siehe Thema Auswählen von Variablen auf S. 35.](#)

Verwenden von Daten über ODBC

Daten aus Datenbankquellen, allgemein Datenbanken, können problemlos in IBM® SPSS® Text Analytics for Surveys importiert werden. Jede Datenbank, die Open Database Connectivity-(ODBC) Treiber verwendet, kann direkt vom Produkt gelesen werden, wenn die entsprechenden Treiber auf dem Rechner installiert sind, auf dem auch SPSS Text Analytics for Surveys installiert ist. Eine ID-Variablen mit einem eindeutigen Wert für jeden Datensatz muss vorhanden sein, um die Daten zu importieren.

Anmerkung: SPSS Text Analytics for Surveys wurde für die Verarbeitung von Datenmengen von bis zu 10.000 Datensätzen optimiert. Je nachdem, wie viel Text in diesen Datensätzen enthalten ist, kann die Leistung jedoch variieren. Weitere Informationen zu Leistungsstatistiken und Empfehlungen finden Sie in den Installationsanweisungen.

Abbildung 4-15
Datenquelloptionen für ODBC

Datenquelle

ODBC

Quelle (DSN)

Name	Beschreibung
MS Access Database	Microsoft Access Driver (*.mdb)
Excel Files	Microsoft Excel Driver (*.xls)
dBASE Files	Microsoft dBase Driver (*.dbf)
Visio Database Samples	Microsoft Access Driver (*.mdb)

ODBC

Nutzer:

Passwort:

Tabelle

SQL

Zurück Weiter Fertig stellen Abbrechen Hilfe

Verwendung über ODBC

- ▶ Wählen Sie im ersten Bildschirm des Assistenten die Option ODBC aus der Dropdown-Liste aus. Der Assistent zeigt die Optionen für ODBC-Dateien an.
- ▶ Legen Sie die Datenquelle fest, indem Sie sie aus der Liste der registrierten ODBC-Quellen auswählen oder den Namen in das Textfeld "Quelle (DSN)" eingeben. Wenn Sie neue Datenquellen registrieren müssen, die nicht in der Liste angezeigt werden, klicken Sie auf ODBC. Damit wird der ODBC-Datenquellen-Administrator geöffnet, der auf den meisten Microsoft Windows-Computern installiert ist. Wenn das nicht der Fall ist, können Sie den ODBC-Import nicht verwenden. Weitere Informationen finden Sie in der Microsoft Windows-Hilfe.
- ▶ Wenn die Datenquelle kennwortgeschützt ist, müssen Sie einen Benutzernamen und ein Kennwort eingeben. Sie werden bei jedem Öffnen des Projekts dazu aufgefordert, da der Benutzername und das Kennwort aus Sicherheitsgründen nicht im Projekt gespeichert werden.
- ▶ Wählen Sie Ihre Daten auf eine der folgenden beiden Arten aus: Tabelle oder SQL. Sie können direkt eine Tabelle auswählen oder SQL-Befehle für die Auswahl der Daten verwenden.
- ▶ Klicken Sie auf Weiter, um Variablen auszuwählen. [Für weitere Informationen siehe Thema Auswählen von Variablen auf S. 35.](#)

Verwenden von IBM SPSS Data Collection-Daten

So importieren Sie über IBM SPSS Data Collection

- ▶ Wählen Sie im ersten Bildschirm des Assistenten die Option Data Collection aus der Dropdown-Liste aus. Die IBM® SPSS® Data Collection-Datenmodelloption ist nur verfügbar, wenn das Datenmodell mit einem anderen Produkt installiert ist.

Auswählen von Variablen

Nach der Auswahl der Datenquelle besteht der nächste Schritt in der Angabe der Variablen, die importiert werden sollen. Drei Variablentypen können in ein Projekt importiert werden.

Eindeutige ID-Variable (erforderlich)

Die ID-Variable ist eine eindeutige numerische Erläuterung oder Texterläuterung, die jeden Teilnehmer identifiziert. Die Datendatei muss nicht nach der eindeutigen ID-Variable sortiert sein, um erfolgreich eingelesen werden zu können. Nach dem Einlesen in das Programm können die Datensätze nach unterschiedlichen Kriterien sortiert werden. [Für weitere Informationen siehe Thema Sortieren von Variablen auf S. 52.](#) Diese ID-Variable ist erforderlich, um Daten zu importieren. Jeder importierte Datensatz (oder Fall) muss einen eindeutigen ID-Wert besitzen.

In folgenden zwei Situationen schlägt der Import fehl:

- Doppelte ID-Werte erkannt,
- Datensätze mit leeren ID-Werten.

Anmerkung: Wenn eine doppelte ID erkannt wird und IBM® SPSS® Statistics auf Ihrem Rechner installiert ist, können Sie die Prozedur "Doppelte Fälle ermitteln" verwenden, um die doppelten IDs zu ermitteln, und anschließend die Optionen verwenden, um zu ermitteln, welche Datensätze beibehalten werden sollen (primäre Fälle).

Variable(n) für offenen Text (erforderlich)

Die Variablen für offenen Text stellen die Textantworten auf die Frage(n) in der Umfrage dar. Mindestens eine dieser Variablen ist erforderlich, um Daten zu importieren. Bei diesen Variablen kann es sich um String- oder lange String-Variablen in SPSS Statistics, Spalten mit allgemeinen Zellen oder Textzellen in Microsoft Excel oder Text- oder Notizenfelder aus Datenbanken handeln. Jede Variable für offenen Text wird separat analysiert. Für jede aus einer .SAV-Datei importierte Textvariable besteht eine Begrenzung von 4.000 Zeichen.

Referenzvariable(n) (optional)

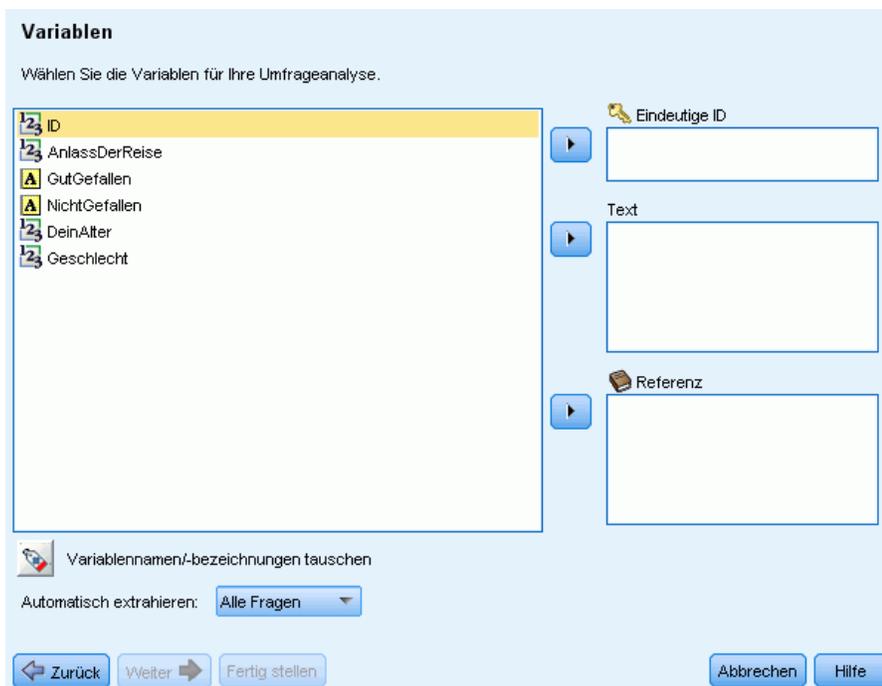
Die Referenzvariablen sind zusätzliche, optionale, in der Regel kategoriale Variablen, die zu Referenzzwecken importiert werden können. Referenzvariablen werden nicht für die Textanalyse verwendet, sondern liefern zusätzliche Informationen, die den Teilnehmer beschreiben und zum Verständnis oder zur Interpretation beitragen können. Demografische Variablen werden oft als Referenzvariablen einbezogen, da Sie zum Verständnis beitragen können, welche Terme oder

Kategorien von welchen Gruppen von Einzelfällen verwendet werden. Beispiele sind Geschlecht, Abteilung, Beruf und Lehrpläne (für Schüler- und Schulungsbeurteilungen). Sie können alle Referenzvariablen anzeigen, wenn sie in die Ansicht "Gesamtes Projekt" importiert wurden. Referenzvariablen können zudem im Bereich "Daten" der Ansicht "Frage" angezeigt werden. Zudem können Sie Referenzvariablen im Balkendiagramm im Visualisierungsbereich auswählen, um eine Untergruppe von Teilnehmern herauszufiltern.

Anmerkung: Für Referenzvariablen, die aus einer SPSS Statistics-Datendatei ausgelesen wurden, werden Variablenlabels (falls vorhanden) als Spaltenüberschriften und die Wertelabels (falls vorhanden) in den Zellen des Bereichs "Daten" angezeigt.

Abbildung 4-16

Auswählen von Variablen



So wählen Sie Variablen und Extrahierungsoptionen aus

- ▶ Wählen Sie aus der Liste der verfügbaren Variablen die Variable aus, die der ID-Variablen in Ihrer Datenmenge entspricht, und klicken Sie auf die Pfeilschaltfläche, um sie in das Feld "Eindeutige ID" zu verschieben. Die ID muss eine eindeutige Zahl oder ein eindeutiger alphanumerischer String sein, der einen Datensatz von einem anderen unterscheidet. Wenn Ihre Datenmenge doppelte IDs enthält, wird eine Fehlermeldung angezeigt. In diesem Fall müssen Sie Ihre Daten bereinigen, bevor Sie den Versuch wiederholen.
- ▶ Wählen Sie aus der Liste der verfügbaren Variablen eine oder mehrere Variablen aus, die den Variablen für offene Antworten entsprechen, und klicken Sie auf die Pfeilschaltfläche, um die Variable(n) in die Liste "Text" zu verschieben. Die Variablen werden als separate Fragen importiert, deren Antworten Sie analysieren und kategorisieren werden.

- ▶ Wählen Sie aus der Liste der verfügbaren Variablen eine oder mehrere Variablen aus, die den Referenzvariablen entsprechen, und klicken Sie auf die Pfeilschaltfläche, um die Variable(n) in die Liste “Referenz” zu verschieben. Referenzvariablen werden nicht von den automatischen Verfahren zum Aufbau von Kategorien verwendet. Sie können jedoch ihre Inhalte anzeigen und sie verwenden, um bei der Kategorisierung Ihrer Antworten fundiertere Entscheidungen treffen zu können.
- ▶ Um die Variablenlabels anstelle der Variablennamen anzuzeigen, klicken Sie auf die Schaltfläche unter der links angezeigten Variablenliste.
- ▶ Um die Extrahierungseinstellungen zu ändern, wählen Sie eine Option aus der Dropdown-Liste aus. Standardmäßig ist die Option Nur erste Frage ausgewählt, bei der, wenn Sie mehr als eine Variable für offenen Text ausgewählt haben, der Extrahierungsprozess automatisch für die erste Frage gestartet wird, wenn der Assistent beendet ist. Die Extrahierung kann bei größeren Datenmengen einige Zeit in Anspruch nehmen. Daher können Sie je nach verfügbarer Zeit Keine oder Alle Fragen extrahieren.
- ▶ Klicken Sie auf Weiter>, wenn Sie alle gewünschten Variablen ausgewählt haben.

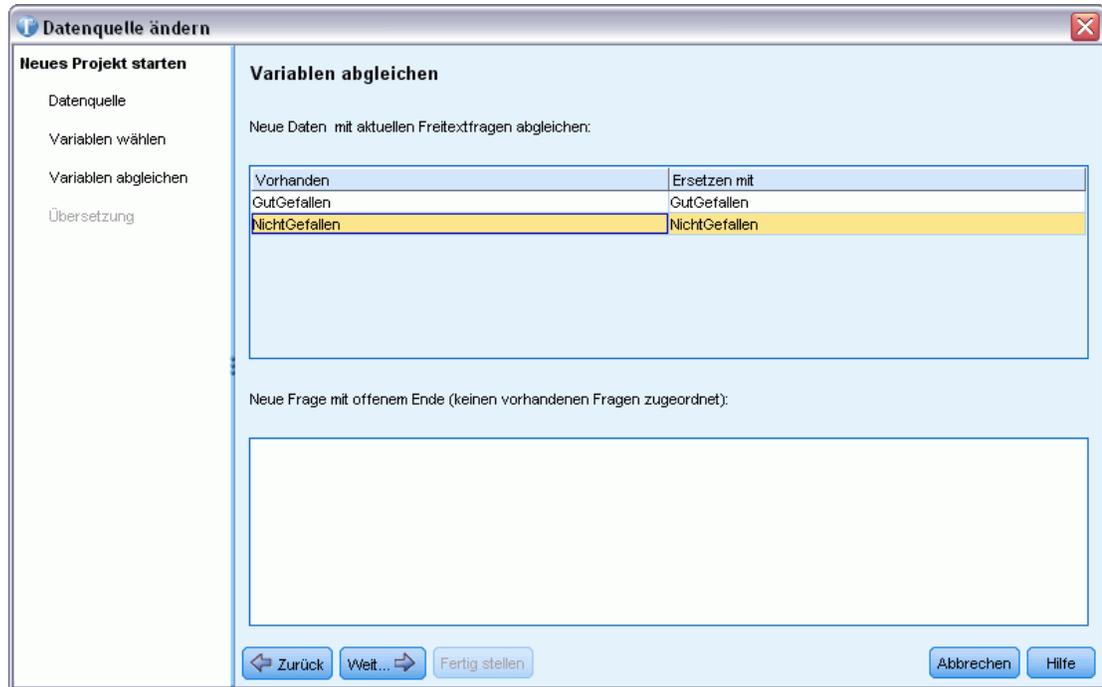
Abgleichen von Variablen

Nach der Auswahl von Variablen im vorherigen Schritt versucht IBM® SPSS® Text Analytics for Surveys, die zuvor importierten Variablen mit den soeben ausgewählten Variablen abzugleichen. Zuordnungen werden automatisch vorgeschlagen, aber Sie können die Variablen auch anders zuordnen, indem Sie in die Spalte Ersetzen durch klicken und eine andere Variable wählen.

Wenn die neue Datendatei keine offene Textvariable enthält, die zuvor im Projekt vorhanden war, können Sie KEINE aus der Liste wählen und die mit der alten Frage verbundenen Daten werden aus dem Projekt entfernt.

Alle Variablen in der neuen Datendatei, die nicht dem vorhandenen Projekt zugeordnet wurden, erscheinen in der Liste “Neue Frage mit offenem Ende” am Ende des Dialogfelds. Nach dem Ändern von Datenmengen erscheinen die verbleibenden Variablen als neue Fragen in Ihrem Projekt.

Abbildung 4-17
Dialogfeld "Inhalt mit vorhandenem Projekt abgleichen"



So gleichen Sie Variablen mit bestehenden Variablen ab:

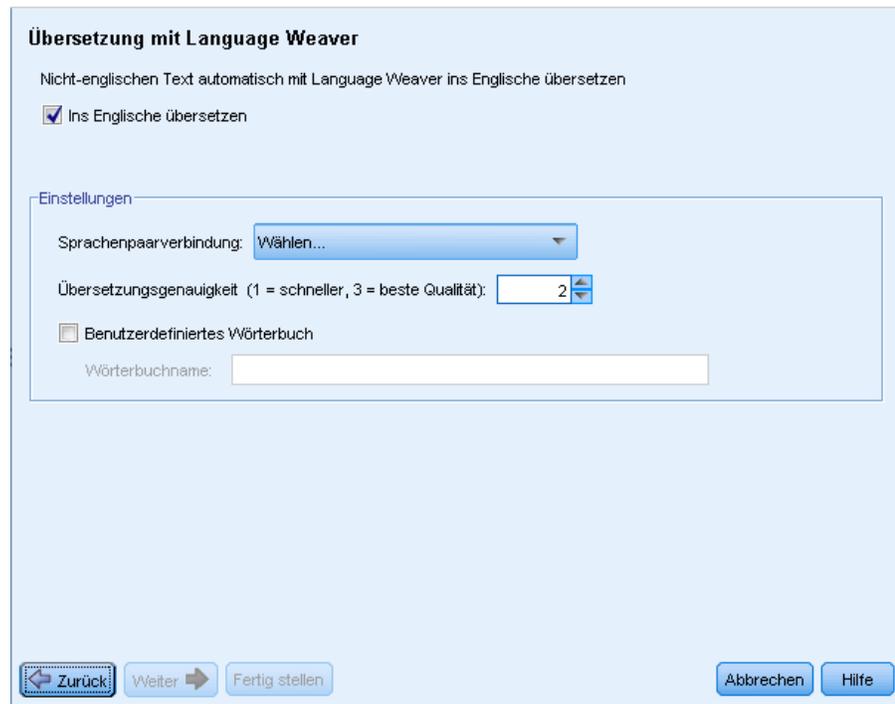
- ▶ Während das Produkt versucht, Ihre neuen Variablen mit denjenigen abzugleichen, die zuvor in der Datendatei gefunden wurden, können Sie ändern, wie die Variablen zugeordnet werden, indem Sie in die Spalte Ersetzen durch klicken und die Variablenentsprechung auswählen. Wenn Sie im vorherigen Schritt mehr Variablen ausgewählt haben als in der Spalte Ersetzen durch enthalten sind, erscheinen sie in Neue Fragen mit offenem Ende (keinen vorhandenen Fragen zugeordnet). Nach dem Ändern von Datenmengen erscheinen die verbleibenden Variablen als neue Fragen in Ihrem Projekt.
- ▶ Wenn für die vorhandene Variable keine Übereinstimmung in der neuen Datenmenge vorhanden ist, wählen Sie KEINE. Die Daten aus dieser vorhandenen Frage werden dann verworfen.
- ▶ Wenn Ihr Projekt Antworten, die in oder aus Kategorien gezwungen wurden, oder Flags enthält, werden Sie aufgefordert, diese Antwort-ID-spezifischen Ergebnisse zu behalten oder zu verworfen. In der Regel werden Sie diese Informationen beim Import anderer Daten (z. B. neue Fragen oder andere Befragte) verworfen, um keine falschen Ergebnisse zu erzeugen. Wenn Sie dieselbe Datendatei importieren, werden Sie diese Information in der Regel behalten, da die Antwort-IDs den alten Daten entsprechen würden.

Übersetzen ins Englische

Wenn Sie mit nicht-englischsprachigem Quelltext arbeiten, können Sie eine Verbindung zu Language Weaver herstellen und eine Übersetzung ins Englische vornehmen. Übersetzungen stehen nur ins Englische zur Verfügung. Language Weaver muss ordnungsgemäß konfiguriert und

mit Verbindungen definiert sein, damit eine Übersetzung durchgeführt werden kann. [Für weitere Informationen siehe Thema Optionen: Registerkarte “Übersetzung” in Kapitel 2 auf S. 24.](#)

Abbildung 4-18
Übersetzungsoptionen



So nehmen Sie eine Übersetzung ins Englische vor:

- ▶ Um die Textdaten aus einer lizenzierten Sprache ins Englische zu übersetzen, aktivieren Sie das Kontrollkästchen **Ins Englische übersetzen**.
- ▶ Wählen Sie in der Liste **“Sprachpaar-Verbindung** die Verbindung für das zu verwendene Language Weaver-Sprachpaar aus. Wenn Language Weaver auf Ihrem lokalen Computer konfiguriert ist, erscheinen diese Sprachpaare automatisch in dieser Liste. Sie können die Verbindung zu Online-Diensten im Dialogfeld **“Optionen”** auf der Registerkarte **“Übersetzung”** hinzufügen, ändern bzw. testen. [Für weitere Informationen siehe Thema Optionen: Registerkarte “Übersetzung” in Kapitel 2 auf S. 24.](#)
- ▶ Geben Sie die gewünschte Übersetzungsgenauigkeit an. Wählen Sie einen Wert von 1 bis 3, der angibt, welche Geschwindigkeit bzw. Genauigkeit Sie wünschen. Ein niedrigerer Wert führt zu einer schnelleren Übersetzung, jedoch auch zu einer geringeren Genauigkeit. Ein höherer Wert führt zu Ergebnissen mit größerer Genauigkeit, jedoch höherer Verarbeitungszeit. Zur Zeitoptimierung wird empfohlen, mit einer niedrigen Stufe zu beginnen und diese nur zu erhöhen, wenn Sie nach einer Überprüfung der Ergebnisse das Gefühl haben, dass größere Genauigkeit erforderlich ist.
- ▶ Wenn Sie zuvor benutzerdefinierte Wörterbücher erstellt haben, die unter Language Weaver gespeichert sind, können Sie sie zusammen mit der Übersetzung verwenden. Aktivieren Sie

zur Auswahl eines benutzerdefinierten Wörterbuchs die Option Benutzerdefiniertes Wörterbuch und geben Sie den Namen unter Wörterbuchname ein. Zur Verwendung mehrerer Wörterbücher, müssen Sie die Namen mit Kommas trennen.

- ▶ Klicken Sie im Assistenten für neue Projekte auf Weiter>, um mit der Auswahl von Kategorien und Ressourcen zu beginnen. [Für weitere Informationen siehe Thema Auswählen von Kategorien und Ressourcen in Kapitel 3 auf S. 39.](#)
- ▶ Klicken Sie im Assistenten zum Ändern des Daten-Sets auf Fertig stellen um die Änderung im Daten-Set abzuschließen und den Übersetzungsvorgang zu starten.

So überspringen Sie die Übersetzung:

- ▶ Deaktivieren Sie die Option Ins Englische übersetzen.
- ▶ Klicken Sie im Assistenten für neue Projekte auf Weiter>, um mit der Auswahl von Kategorien und Ressourcen zu beginnen. [Für weitere Informationen siehe Thema Auswählen von Kategorien und Ressourcen in Kapitel 3 auf S. 39.](#)
- ▶ Klicken Sie im Assistenten zum Ändern des Daten-Sets auf Fertig stellen um die Änderung im Daten-Set abzuschließen.

Aktualisieren von Daten

Bei der Arbeit mit Ihren Projektdaten ändern Sie möglicherweise die ursprüngliche Datenquelle. Sie könnten beispielsweise Datensätze hinzufügen oder entfernen. Sie können die Daten mithilfe der Funktion "Daten aktualisieren" aktualisieren. Wenn Sie jedoch z. B. die Variablennamen oder den Dateinamen geändert haben, müssen Sie Ihre Daten komplett neu importieren. [Für weitere Informationen siehe Thema Ändern von Datenquellen auf S. 63.](#)

So aktualisieren Sie Ihre Daten:

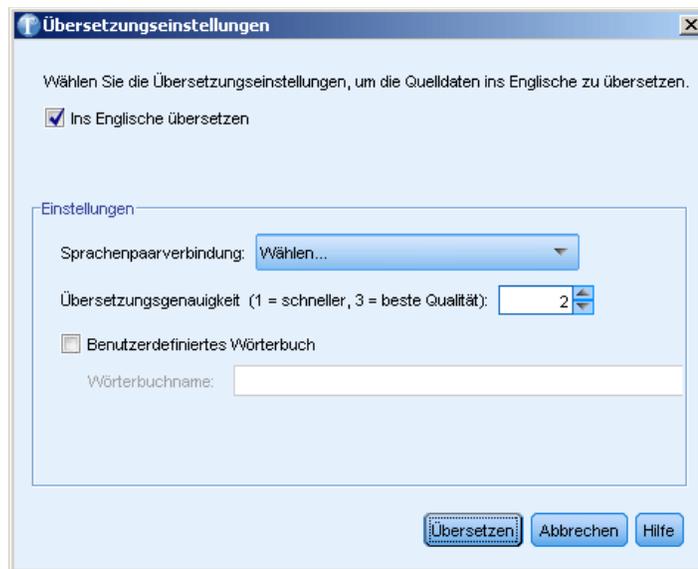
- ▶ Wählen Sie in den Menüs die Optionsfolge Datei > Daten aktualisieren. Die Daten werden erneut gelesen, um Ihre neuen Änderungen zu berücksichtigen.
- ▶ Wenn zuvor eine Übersetzung ins Englische vorgenommen wurde, wird das Dialogfeld "Übersetzungseinstellungen" angezeigt, damit Sie das Sprachpaar auswählen und die Übersetzung erneut durchführen können. [Für weitere Informationen siehe Thema Übersetzen ins Englische auf S. 74.](#)

Übersetzen ins Englische

Sie können Übersetzungen jederzeit aktualisieren. Nach der Übersetzung müssen Sie erneut eine Extrahierung vornehmen, da die Übersetzungsergebnisse nicht mit der neuen Übersetzung synchron sind.

Anmerkung: Wenn Sie neue Daten übersetzen möchten, können Sie dies direkt beim Erstellen eines neuen Projekts im Assistenten für neue Projekte tun. [Für weitere Informationen siehe Thema Übersetzen ins Englische auf S. 37.](#)

Abbildung 4-19
Übersetzungseinstellungen



So nehmen Sie eine Übersetzung ins Englische vor:

- ▶ Wählen Sie in den Menüs die Optionsfolge Extras > Übersetzungseinstellungen. Das Dialogfeld “Übersetzungseinstellungen” wird angezeigt.
- ▶ Um die Textdaten aus einer lizenzierten Sprache ins Englische zu übersetzen, aktivieren Sie das Kontrollkästchen Ins Englische übersetzen.
- ▶ Wählen Sie in der Liste “Sprachpaar-Verbindung die Verbindung für das zu verwendenden Language Weaver-Sprachpaar aus. Wenn Language Weaver auf Ihrem lokalen Computer konfiguriert ist, erscheinen diese Sprachpaare automatisch in dieser Liste. Sie können Verbindungen von Netzwerken (WAN) oder Online-Diensten (HTTP) im Dialogfeld “Optionen” auf der Registerkarte “Übersetzung” hinzufügen bzw. testen. [Für weitere Informationen siehe Thema Optionen: Registerkarte “Übersetzung” in Kapitel 2 auf S. 24.](#)
- ▶ Geben Sie die gewünschte Übersetzungsgenauigkeit an. Wählen Sie einen Wert von 1 bis 3, der angibt, welche Geschwindigkeit bzw. Genauigkeit Sie wünschen. Ein niedrigerer Wert führt zu einer schnelleren Übersetzung, jedoch auch zu einer geringeren Genauigkeit. Ein höherer Wert führt zu Ergebnissen mit größerer Genauigkeit, jedoch höherer Verarbeitungszeit. Zur Zeitoptimierung wird empfohlen, mit einer niedrigen Stufe zu beginnen und diese nur zu erhöhen, wenn Sie nach einer Überprüfung der Ergebnisse das Gefühl haben, dass größere Genauigkeit erforderlich ist.
- ▶ Wenn Sie zuvor benutzerdefinierte Wörterbücher erstellt haben, die unter Language Weaver gespeichert sind, können Sie sie zusammen mit der Übersetzung verwenden. Aktivieren Sie zur Auswahl eines benutzerdefinierten Wörterbuchs die Option Benutzerdefiniertes Wörterbuch und geben Sie den Namen unter Wörterbuchname ein. Zur Verwendung mehrerer Wörterbücher, müssen Sie die Namen mit Kommas trennen.

- ▶ Klicken Sie auf Übersetzen, um den Übersetzungsvorgang zu starten. Das Dialogfeld mit dem Übersetzungsverlauf wird angezeigt.

Gemeinsame Benutzung von Projekten

Sie können Ihre Projekte gemeinsam mit anderen Benutzern verwenden oder nach Wunsch an einem anderen Computer an einem Projekt arbeiten.

So benutzen Sie ein Projekt gemeinsam:

- ▶ Wählen Sie in den Menüs die Optionsfolge Datei > Projekt speichern. Das Projekt wird gespeichert.
- ▶ Senden Sie die Projektdatei an einen anderen Computer oder eine andere Person. Diese Projektdatei enthält einen Verweis auf die Datendatei, die Sie ursprünglich importiert haben. Wenn der andere Benutzer dieselben Quelldaten für dieses Projekt verwenden soll, müssen Sie ihm auch die originale Datendatei bereitstellen und ihm den Pfad mitteilen, unter den er diese Datendatei kopieren muss, damit IBM® SPSS® Text Analytics for Surveys die Datendatei beim Öffnen des Projekts findet.
- ▶ Wenn der andere Benutzer das Projekt in SPSS Text Analytics for Surveys öffnet, kann er wählen, ob er die Local libraries in der Projektdatei oder öffentliche Versionen dieser Bibliotheken verwenden möchte, die ihm bereits zur Verfügung standen. Im Allgemeinen sollten die lokalen Versionen verwendet werden, um dieselben Ergebnisse sicherzustellen.
- ▶ Wenn SPSS Text Analytics for Surveys die Datendatei nicht findet, teilt eine Meldung dem Benutzer mit, dass die Daten erneut importiert werden müssen. [Für weitere Informationen siehe Thema Ändern von Datenquellen auf S. 63.](#)

Kennzeichen von Antworten

Zur besseren Überwachung Ihres Fortschritts bei der Umfragenanalyse können Sie Antworten mithilfe von Flags im Bereich "Daten" markieren. Es gibt zahlreiche Gründe, eine Antwort zu markieren, zum Beispiel:

- Abhaken der Antworten, die Sie bereits manuell überprüft haben, damit Sie wissen, wo Sie später fortfahren müssen
- Markieren einer Antwort, bei der Sie nicht genau wissen, wie sie behandelt werden soll
- Kennzeichnen und Exportieren der Flags in ein anderes Programm

Sobald Sie eine Antwort mit einem Flag kennzeichnen, können Sie mit diesen Antworten weiterarbeiten. Sie dienen nur für Sie selbst als Dokumentation. Sie können aus den folgenden Flags wählen:

Tabelle 4-1
Flag-Beschreibungen

Flag	Beschreibung
	Abgeschlossen-Flag zur Markierung von Antworten, die Sie als fertig bearbeitet betrachten.
	Wichtig-Flag zur Markierung von Antworten, die Sie als wichtig betrachten.

Abbildung 4-20
Antwort-Flags im Fensterbereich "Daten"

	ID	Antwort	Kategorien
1	1	Zimmer sauber	zimmer
2	4	Sauberes Zimmer	zimmer
3	5	die komfortablen Zimmer, reichhaltiges Frühstück, netter Service	frühstück service zimmer
4	6	Die ruhige Lage direkt am Strand.	lage
5	7	Angenehmer Service	service
6	8	Der Ausblick auf die schöne Landschaft und auf das Meer.	landschaft
7	9	Die besonders schöne Landschaft	landschaft
8	10	Der Service, die Sauberheit und Gepflegtheit des Zimmers und des Bades. Das reichhaltige und phantasievolle Frühstücksbuffet	büffet bad service
9	11	Die Bedienung.	bedienung
10	12	gute Strandlage, vielfältige Ausflugs-Angebote, abwechslungsreiches Büffet, freundliches Service-Personal	büffet lage service

So kennzeichnen Sie eine Antwort mit einem Flag:

- ▶ Wählen Sie im Bereich "Daten" die Antwort aus, die Sie kennzeichnen möchten.
- ▶ Wählen Sie in den Menüs die Optionsfolge Bearbeiten > Treffer hervorheben mit und wählen Sie dann den gewünschten Flag-Typ (Wichtig-Flag oder Abgeschlossen-Flag). Das ausgewählte Flag wird zugewiesen. Wenn die Spalte "Flag" im Bereich "Daten" nicht sichtbar ist, wird sie nun eingeblendet. Die Statusleiste wird mit der Anzahl an gekennzeichneten Antworten aktualisiert.

So löschen Sie Flags:

- ▶ Klicken Sie im Bereich "Daten" mit der rechten Maustaste auf die Antworten, von denen Sie ein Flag entfernen möchten.
- ▶ Wählen Sie im Kontextmenü die Optionsfolge Treffer hervorheben mit > Flags löschen. Die ausgewählten Flags werden entfernt.

Projektstatusleiste

Abhängig vom Fenster oder der Ansicht, in der Sie arbeiten, sind unterschiedliche Statusleisten vorhanden. Standardmäßig wird immer eine Statusleiste angezeigt, wenn ein Projekt geöffnet ist. Die Statusleiste bietet zusammenfassende Informationen über das Projekt und die Elemente, die es enthält. Sie können die Statusleiste nach Wunsch auch ein- und ausblenden.

So blenden Sie die Statusleiste in einem Fenster ein oder aus:

- ▶ Wählen Sie in den Menüs die Optionsfolge Ansicht > Statusleiste.

Textanalysefenster

Die Statusleiste bietet zusammenfassende Informationen über Fragen und Antworten im Projekt. Abhängig von der aktuellen Position im Textanalysefenster ändert sich die Information in der Statusleiste. Sie können auch die Anzahl der Antworten sehen, die als wichtig oder abgeschlossen gekennzeichnet sind.

Abbildung 4-21
Statusleiste im Textanalysefenster: Ansicht "Frage"



In der Ansicht "Frage" sehen Sie die Anzahl der Kategorien für diese Frage und die prozentuale Kategorisierung der Antworten. In der Ansicht "Gesamtes Projekt" sehen Sie Informationen für das ganze Projekt.

Abbildung 4-22
Statusleiste im Textanalysefenster: Ansicht "Gesamtes Projekt"



Die folgende Tabelle beschreibt jedes Element in der Statusleiste.

Tabelle 4-2
Textanalysefenster: Beschreibung der Statusleiste

Element	Beschreibung
Datensätze	Anzahl der Datensätze in Ihren Daten.
Fragen	Anzahl der Fragen in Ihrem Projekt.
Referenz	Anzahl der importierten Referenzvariablen. Referenzvariablen sind zusätzliche Variablen, die zu Referenzzwecken importiert, aber nicht analysiert werden.
Categories	Anzahl der Kategorien für eine bestimmte Frage. Wenn leere Kategorien vorhanden sind, wird die Anzahl der leeren Kategorien in Klammern angezeigt.
Kategorisiert	Anzahl der Antworten für die Frage, gefolgt vom Kategorisierungsprozentsatz in Klammern.

Resource Editor-Fenster

Die Statusleiste bietet Informationen über die linguistischen Ressourcen für das Projekt. Der Bereich "Erzwungene Begriffe" in der Leiste enthält Aktionen, auf die Sie klicken können, um sie auszuführen. Beim Arbeiten in Bibliotheken können Sie Elemente in Bibliotheken deaktivieren, um sie aus der Verarbeitung auszuschließen. [Für weitere Informationen siehe Thema Local Libraries deaktivieren in Kapitel 9 auf S. 214.](#) Wenn das Projekt deaktivierte Elemente enthält, erscheinen zwei Zahlenangaben in der Statusleiste: die erste ist die Anzahl der vorhandenen Elemente, die zweite die Anzahl der aktivierten Elemente. Wenn Ihre Statusleiste beispielsweise 5(2) Bibliotheken anzeigt, bedeutet dies, dass Ihr Projekt fünf Bibliotheken enthält, aber nur zwei aktiviert sind.

Abbildung 4-23
Statusleiste im Fenster "Wörterbucheditor"



Die folgende Tabelle beschreibt jedes Element in der Statusleiste.

Tabelle 4-3
Resource Editor-Fenster: Beschreibung der Statusleiste

Element	Beschreibung
Bibliothek	Anzahl der Bibliotheken in Ihrem Projekt.
Typ	Anzahl der Typen im gesamten Projekt.
Begriff	Anzahl der Begriffe in allen Bibliotheken. Wenn sich ein Begriff in einer Ausschlussliste befindet, wird er dennoch in der Anzahl berücksichtigt. Wenn ein Typ deaktiviert ist, beachten Sie, dass dann auch alle Begriffe dieses Typs deaktiviert sind.
Ausschließen	Die Anzahl der ausgeschlossenen Einträge in allen Bibliotheken im Projekt.
Synonym	Die Anzahl der Synonymziele in allen Bibliotheken im Projekt.
Optional	Die Anzahl der definierten optionalen Einträge in allen Bibliotheken im Projekt. Beachten Sie, dass die Anzahl jeweils jeden mit Trennzeichen versehenen Eintrag in einer Zelle individuell umfasst.
Erzwungene Begriffe	Eine Schaltfläche, die aktiviert ist, sofern in den Bibliotheken Ihres Projekts erzwungene Begriffe vorhanden sind. Wenn Sie auf diese Schaltfläche klicken, wird das Dialogfeld "Erzwungene Begriffe bearbeiten" geöffnet. Für weitere Informationen siehe Thema Erzwingen von Fachausdrücken in Kapitel 10 auf S. 231.

Daten extrahieren

Wenn Sie ein Projekt über den Assistenten “Neues Projekt” anlegen, ist die Standardauswahl eine automatische Extrahierung für die erste Frage. Wenn Sie eine Extraktion aktualisieren oder für eine neue Frage extrahieren möchten, können Sie dies manuell tun (Extras > Extrahieren) oder Sie extrahieren, wenn Sie beginnen, Kategorien aufzubauen. Das Endergebnis dieser Extrahierung ist eine Reihe von Konzepten, Typen und Mustern. Sie können diese Ergebnisse im Bereich “Extrahierungsergebnisse” anzeigen und dort mit ihnen arbeiten.

Wenn keine Extraktion durchgeführt wurde, als Sie Ihr Projekt erstellt haben, oder wenn Sie die Extrahierungsergebnisse nicht speichern, können Sie zu der Frage navigieren, an der Sie arbeiten möchten (Ansicht > Frage > “Frage”) und dann extrahieren.

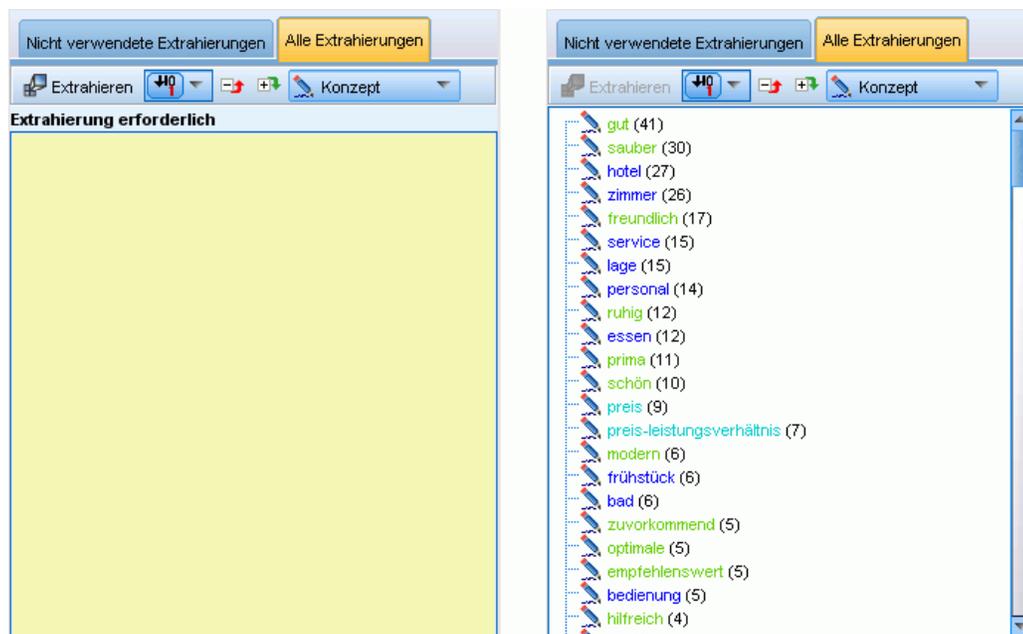
Überprüfen Sie nach dem Extrahieren die Ergebnisse und nehmen Sie Änderungen vor, die Sie für erforderlich halten. [Für weitere Informationen siehe Thema Extrahierungsergebnisse verfeinern auf S. 88.](#) Sie können dann erneut extrahieren, um die neuen Ergebnisse anzuzeigen. Beim manuellen Kodieren von Antworten könnten zwei Personen Antworten etwas unterschiedlich gruppieren. Jedoch sind Präzision und Kontinuität von größter Wichtigkeit für die Kategorisierung von Antworten in Umfragen.

Die Stärke von IBM® SPSS® Text Analytics for Surveys liegt in der Fähigkeit, Ihre Kategoriedefinitionen konsistent immer wieder anzuwenden. Wenn Sie Ihre Extrahierungsergebnisse von Anfang an verfeinern, gehen Sie sicher, dass Sie bei jeder neuen Extrahierung dieselben gut auf den Kontext der Daten angepassten Ergebnisse in Ihren Kategoriedefinitionen erhalten. So wird die Zuweisung von Antworten zu Ihren Kategoriedefinitionen genauer und wiederholbar.

Extrahierte Ergebnisse: Konzepte, Typen und Muster

Nachdem Sie ein Projekt erstellt haben, wird die erste offene Frage, die Sie importiert haben, im Fenster automatisch angezeigt. Der Bereich “Extrahierungsergebnisse” befindet sich in der unteren linken Ecke der Frageansicht. Auf diese Ansicht kann im Menü “Ansicht” zugegriffen werden (Ansicht > Frage > “Frage”). Falls es keine Extrahierungsergebnisse gibt, müssen Sie mit einer Extrahierung beginnen. [Für weitere Informationen siehe Thema Daten extrahieren auf S. 85.](#)

Abbildung 5-1
Bereich "Extrahierungsergebnisse" vor und nach der Extrahierung



Wenn der Bereich "Extrahierungsergebnisse" leer oder nicht mehr aktuell ist, wird er in gelb dargestellt. Klicken Sie auf die Schaltfläche Extrahieren, um den Extrahierungsprozess zu starten. Nach dem Extrahieren können Sie die Ergebnisse überprüfen, indem Sie in der Dropdown-Liste auswählen, was Sie anzeigen möchten.

Abbildung 5-2
Die Dropdown-Liste im Bereich "Extrahierungsergebnisse"



Die extrahierten Konzepte, Typen und TLA-Muster werden als **Extrahierungsergebnisse** bezeichnet und dienen als Deskriptoren oder Bausteine für Ihre Kategorien. Sie können auch Konzepte, Typen und Muster in Ihren Kategorieregeln verwenden. Außerdem verwenden die automatischen Verfahren Konzepte und Typen zum Aufbau der Kategorien.

Text-Analysis ist ein iterativer Prozess, in dem Extrahierungsergebnisse dem Kontext der Textdaten gemäß überprüft, für die Gewinnung neuer Ergebnisse verfeinert und dann neu bewertet werden. Überprüfen Sie nach dem Extrahieren die Ergebnisse und nehmen Sie Änderungen, die Sie für erforderlich halten, durch Bearbeiten der linguistischen Ressourcen vor. Sie können die Ressourcen zum Teil direkt vom Bereich "Extrahierungsergebnisse" oder vom Datenbereich aus mithilfe von Kontextmenüs verfeinern. [Für weitere Informationen siehe Thema](#)

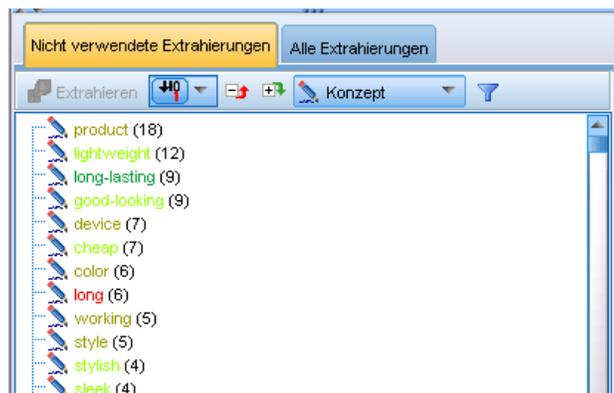
[Extrahierungsergebnisse verfeinern auf S. 88](#). Außerdem können Sie Verfeinerungen direkt in der Resource Editor-Ansicht vornehmen. [Für weitere Informationen siehe Thema Das Fenster “Ressourceneditor” in Kapitel 2 auf S. 16](#).

Nach der Verfeinerung können Sie erneut extrahieren, um die neuen Ergebnisse anzuzeigen. Wenn Sie Ihre Extrahierungsergebnisse von Anfang an verfeinern, gehen Sie sicher, dass Sie bei jeder neuen Extrahierung dieselben perfekt auf den Kontext der Daten angepassten Ergebnisse in Ihren Kategoriedefinitionen erhalten. So wird die Zuweisung von Datensätzen zu Ihren Kategoriedefinitionen genauer und wiederholbar.

Konzepte

Beim Extrahieren werden die Textdaten untersucht und analysiert, um interessante oder relevante einzelne Wörter (z. B. Wahl oder Frieden) und zusammengesetzte Ausdrücke (z. B. Wahl zur Präsidentschaft, Wahl des Präsidenten, oder Verhandlungen um Frieden) im Text zu ermitteln. Diese Wörter und Zusammensetzungen werden auch als *Fachausdrücke* bezeichnet. Die relevanten Fachausdrücke werden mithilfe der linguistischen Ressourcen extrahiert und anschließend werden ähnliche Fachausdrücke unter einem übergeordneten Fachausdruck, einem **Konzept**, gruppiert.

Abbildung 5-3
Fensterbereich “Extrahierungsergebnisse”: Konzeptansicht



Standardmäßig werden die Konzepte in Kleinbuchstaben angezeigt und absteigend entsprechend der Anzahl der Datensätze, in denen das Konzept vorkommt, sortiert. Beim Extrahieren wird den Konzepten ein Typ zugewiesen, um das Gruppieren ähnlicher Konzepte zu erleichtern. Sie sind gemäß diesem Typ farblich gekennzeichnet. Die Farben sind in den Typ-Eigenschaften im Resource Editor definiert. [Für weitere Informationen siehe Thema Typ-Wörterbücher in Kapitel 10 auf S. 223](#).

Wenn ein Konzept, Typ oder Muster in einer Kategoriedefinition verwendet wird, angezeigt, es erscheint in Kursivschrift in der Tabelle.

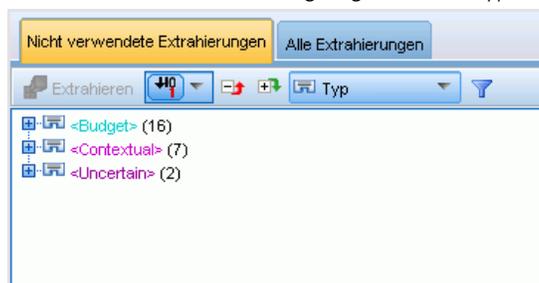
Typen

Typen sind semantische Gruppierungen von Konzepten. Beim Extrahieren wird den Konzepten ein Typ zugewiesen, um das Gruppieren ähnlicher Konzepte zu erleichtern. Es sind mehrere integrierte Typen im Lieferumfang von IBM® SPSS® Text Analytics for Surveys enthalten, z.

B. <Location>, <Organization>, <Person>, <Positive>, <Negative> usw. Der Typ <Location> gruppiert z. B. geografische Stichwörter und Orte. Dieser Typ würde Konzepten wie z. B. Chicago, Paris und Tokio zugewiesen. Konzepte, die in keinem Typen-Wörterbuch gefunden, aber aus dem Text extrahiert werden, erhalten automatisch den Typ <Unbekannt>. [Für weitere Informationen siehe Thema Integrierte Typen in Kapitel 10 auf S. 224.](#)

Wählen Sie die Ansicht "Typ", um die extrahierten Typen standardmäßig in absteigender Reihenfolge nach Häufigkeit anzuzeigen. Wenn der Baum erweitert wird, sehen Sie die Konzepte, die für diesen Typ extrahiert wurden. Sie sehen außerdem, dass die Typen farblich gekennzeichnet sind, um ihre Unterscheidung zu erleichtern. Farben sind Teil der Typeigenschaften. [Für weitere Informationen siehe Thema Erstellen von Typen in Kapitel 10 auf S. 225.](#) Sie können auch eigene Typen erstellen.

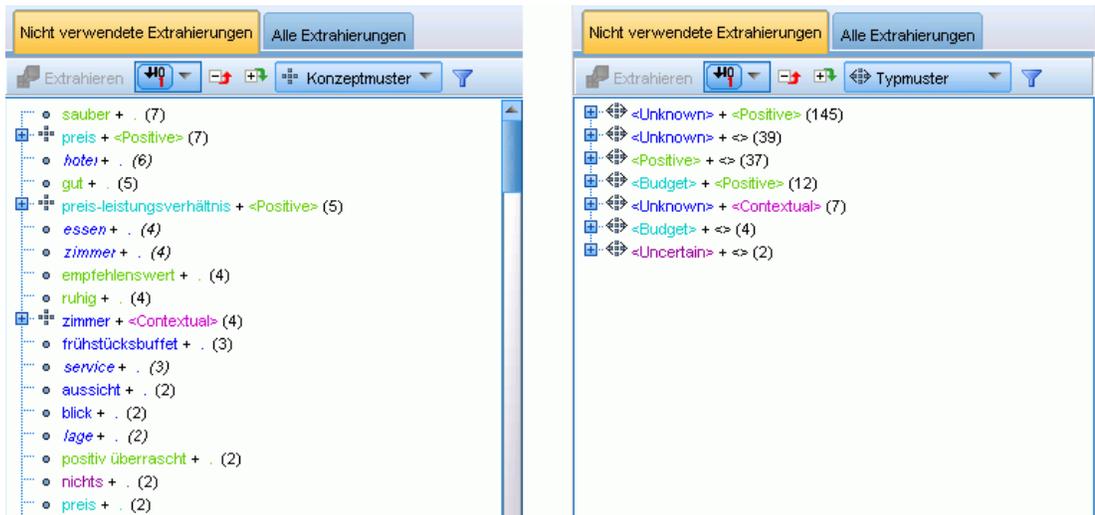
Abbildung 5-4
Fensterbereich "Extrahierungsergebnisse": Typansicht



Muster

Muster bestehen aus zwei Teilen: Sie sind eine Kombination aus Konzepten und integrierten Typen, die für Vermerke und Adjektive stehen. Muster sind besonders nützlich, wenn Sie versuchen, Meinungen zu einem bestimmten Thema zu ermitteln. Es könnte sein, dass es Ihnen nicht reicht, den Produktnamen Ihres Mitbewerbers zu extrahieren. In diesem Fall können Sie die extrahierten Muster dahingehend überprüfen, ob Sie Beispiele finden, bei denen Befragte das Produkt für gut, schlecht oder teuer befanden. Es gibt zwei verschiedene Musteransichten: Konzeptmuster und Typmuster.

Abbildung 5-5
 Fensterbereich "Extrahierungsergebnisse": Konzeptmusteransicht und Typmusteransicht



Konzeptmuster. In dieser Ansicht wird die höchste Ebene des Baums im Fensterbereich "Extrahierungsergebnisse" mit folgender Struktur angezeigt: Konzept1 + <Typ1> für Konzeptmuster, z. B. Textanalyse + <Positive> oder Kosten + <Negative>. Wird der Baum erweitert, können Sie die exakten Muster sehen, z. B. Textanalyse + leistungsstark oder Kosten + teuer.

Muster können auch von Bedeutung sein, wenn sie keinen zweiten Teil aufweisen. Sie könnten zum Beispiel daran interessiert sein, Fälle zu finden, in denen der Befragte keine negative oder positive Meinung zum Thema geäußert hat. In diesem Fall wird dies durch Konzept1 + . ausgedrückt, wobei . ein Null-Qualifizierungszeichen darstellt. Falls ein Befragter zum Beispiel "Kosten. Lage." auf die Frage "Welche Faktoren beeinflussen Ihre Entscheidung, einen MP3-Player zu kaufen?" antwortet, könnte die Extrahierung Kosten + . und Lage_Geschäft + . als Nullmuster hervorbringen. Wenn Muster angezeigt werden, werden allen Elementen im Muster abhängig von ihrem Typ Farben zugeordnet.

Typmuster. In dieser Ansicht wird die höchste Ebene des Baums im Fensterbereich "Extrahierungsergebnisse" mit folgender Struktur angezeigt: <Typ> + <Typ>, z. B. <Budget> + <Positive>. Wenn Sie den Baum abermals erweitern, werden Sie Beziehungen sehen, die der Beschreibung und Darstellung in der Konzeptmusteransicht entsprechen. Wenn Muster angezeigt werden, werden allen Elementen im Muster abhängig von ihrem Typ Farben zugeordnet.

Registerkarten "Nicht verwendete Extrahierungen" und "Alle Extrahierungen"

Der Bereich "Extrahierungsergebnisse" präsentiert die Ausgabe aus dem Extrahierungsprozess. Während Sie beginnen, Kategorien zu definieren, werden einige der Extrahierungsergebnisse (Konzepte, Typen und Muster) Teil der Kategoriedeskriptoren. Daher präsentiert SPSS Text Analytics for Surveys diese Informationen mithilfe von Registerkarten auf zwei Arten. Sie können zwischen den Ansichten derjenigen Elemente, die bereits in Kategoriedefinitionen verwendet werden, und dem vollständigen Satz extrahierter Konzepte hin- und herwechseln. Klicken Sie dazu einfach auf die Registerkarten "Nicht verwendete Extrahierungen" und "Alle Extrahierungen". Die Registerkarte "Nicht verwendete Extrahierungen" zeigt alle Elemente, die

derzeit noch keinem Kategoriedeskriptor angehören. Die Registerkarte “Alle Extrahierungen” zeigt alle extrahierten Elemente an, wobei die verwendeten Elemente kursiv dargestellt werden.

Daten extrahieren

Wenn eine Extrahierung nötig ist, wird der Bereich “Extrahierungsergebnisse” in gelb dargestellt und die Nachricht Schaltfläche “Extrahieren” drücken, um Konzepte zu extrahieren wird unter der Symbolleiste in diesem Bereich angezeigt.

Eventuell müssen Sie extrahieren, wenn Sie noch keine Extrahierungsergebnisse haben, linguistische Ressourcen geändert haben und die Extrahierungsergebnisse aktualisieren müssen oder ein Projekt neu geöffnet haben, wo Sie die Extrahierungsergebnisse nicht gespeichert haben (Extras > Optionen).

Anmerkung: Wenn Sie den Quellenknoten für Ihren Stream ändern, nachdem Extrahierungsergebnisse mit der Option Arbeit der Sitzung verwenden... im Cache abgelegt wurden, müssen Sie eine neue Extrahierung ausführen, sobald die interaktive Workbench-Sitzung gestartet wird, damit Sie aktualisierte Extrahierungsergebnisse erhalten.

Wenn Sie eine Extrahierung durchführen, erscheint ein Fortschrittsbalken, der den Status der Extrahierung anzeigt. Währenddessen liest die Extrahierungsengine alle Textdaten, identifiziert die relevanten Ausdrücke und Muster, extrahiert sie und weist sie einem Typ zu. Dann versucht das Modul, synonyme Ausdrücke unter einem Leitausdruck, einem Konzept, zu gruppieren. Wenn der Vorgang abgeschlossen ist, werden die resultierenden Konzepte, Typen und Muster im Bereich “Extrahierungsergebnisse” angezeigt. Sie können beginnen, mit den Ergebnissen zu arbeiten und sie zu prüfen.

Anmerkung: Die für den Extrahierungsprozess benötigte Zeit steht in direkter Beziehung zur Größe Ihrer Datenmenge. In den Installationsanweisungen finden Sie Leistungsstatistiken und Empfehlungen.

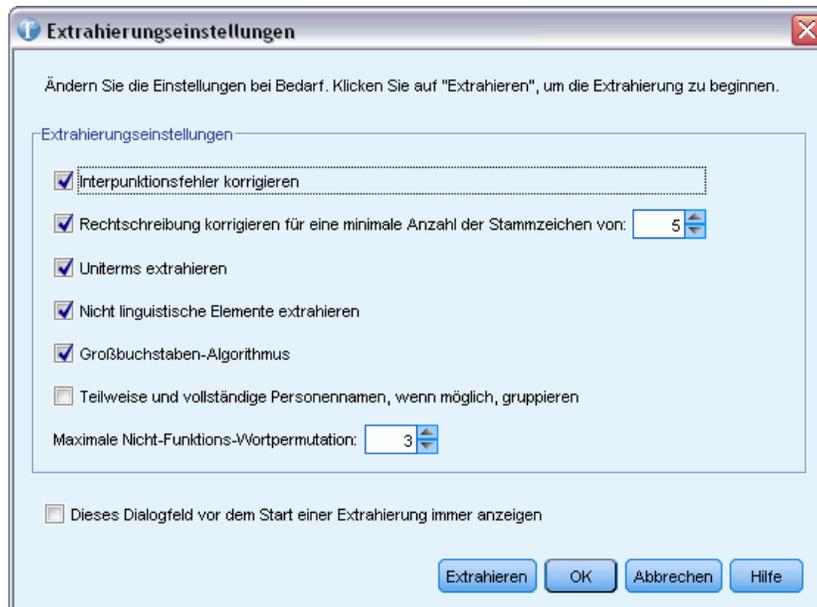
Daten extrahieren

- ▶ Wählen Sie in den Menüs die Optionsfolge Extras > Extrahieren aus. Alternativ können Sie auf die Symbolleistenschaltfläche Extrahieren klicken.
- ▶ Wenn Sie das Dialogfeld “Extrahierungseinstellungen” immer anzeigen lassen, wird es angezeigt, damit Sie Änderungen vornehmen können. Weitere Informationen zu Deskriptoren für jede Einstellung finden Sie in diesem Thema.
- ▶ Klicken Sie auf Extrahieren, um die Extrahierung zu starten. Sobald die Extrahierung beginnt, öffnet sich die Statusanzeige. Nach der Extrahierung werden die Ergebnisse im Bereich “Extrahierungsergebnisse” dargestellt. Standardmäßig werden die Konzepte in Kleinbuchstaben angezeigt und absteigend entsprechend der Anzahl der Datensätze, in denen das Konzept vorkommt, sortiert.

Sie können die Ergebnisse überprüfen, indem Sie sie mithilfe der Optionen in der Symbolleiste unterschiedlich sortieren und filtern oder die Ansicht wechseln (Konzepte, Muster oder Typen). Sie können die Ergebnisse auch verfeinern, indem Sie mit den linguistischen Ressourcen arbeiten. [Für weitere Informationen siehe Thema Extrahierungsergebnisse verfeinern auf S. 88.](#)

Das Dialogfeld “Extrahierungseinstellungen” enthält einige grundlegende Extrahierungsoptionen.

Abbildung 5-6
Dialogfeld “Extrahierungseinstellungen”



Interpunktionsfehler korrigieren. Diese Option normalisiert Text mit Punktationsfehlern (zum Beispiel ungeeignete Verwendung) während der Extrahierung, um die Extrahierbarkeit von Konzepten zu verbessern. Diese Option ist besonders nützlich bei kurzem Text und niedriger Textqualität (wie dies beispielsweise bei offenen Antworten bei Umfragen, E-Mails und CRM-Daten der Fall ist) oder wenn der Text viele Abkürzungen enthält.

Rechtschreibung korrigieren für eine minimale Anzahl an Stammzeichen von [n]. Diese Option wendet ein unscharfes Gruppierungsverfahren an, das hilft, häufig falsch geschriebene Wörter oder genau geschriebene Wörter unter einem Konzept zu gruppieren. Der Algorithmus für die unscharfe Gruppierung entfernt alle Vokale (außer den ersten) und doppelte/dreifache Konsonanten temporär aus extrahierten Wörtern und vergleicht sie, um festzustellen, ob sie gleich sind, so dass *Modellierung* und *Modelierung* zusammen gruppiert werden würden. Wenn jedoch jeder Fachausdruck einem anderen Typ (ausschließlich des Typs <Unknown>) zugewiesen ist, wird das unscharfe Gruppierungsverfahren nicht angewendet.

Sie können auch die minimal erforderliche Zahl von *Stammzeichen* definieren, bevor unscharfe Gruppierung eingesetzt wird. Die Anzahl der Stammzeichen in einem Ausdruck berechnet sich aus der Summe aller Zeichen abzüglich aller Zeichen, die Beugungsendungen und – bei zusammengesetzten Ausdrücken – Determinatoren und Präpositionen bilden. So würde beispielsweise der Ausdruck *Aufgaben* durch die Form “Aufgabe” mit 7 Stammzeichen gezählt werden, da der Buchstabe *n* am Ende des Worts eine Beugung darstellt (Pluralform). Gleichmaßen werden für *Apfelmus* 8 Stammzeichen (“Apfelmus”) gezählt und *Hersteller von Autos* zählt als 14 Stammzeichen (“Hersteller Auto”). Diese Zählmethode dient nur zur Überprüfung, ob die Fuzzy-Gruppierung angewendet werden soll, hat jedoch keinen Einfluss auf den Abgleich der Wörter.

Anmerkung: Wenn Sie feststellen, dass bestimmte Wörter später falsch gruppiert werden, können Sie Wortpaare aus diesem Verfahren ausschließen, indem Sie sie explizit im Abschnitt Unschärfe Gruppierung: Ausnahmen im erweiterten Ressourceneditor deklarieren. [Für weitere Informationen siehe Thema Unschärfe Gruppierung in Kapitel 11 auf S. 245.](#)

uniterms extrahieren. Diese Option extrahiert einzelne Wörter (Uniterms), solange das Wort nicht bereits Teil eines zusammengesetzten Worts ist und es entweder ein Nomen oder eine nicht erkannte Wortart ist.

Nicht linguistische Elemente extrahieren. Diese Option extrahiert nicht linguistische Elemente wie beispielsweise Telefonnummern, Personalausweisnummern, Uhrzeiten, Datumsangaben, Währungen, Ziffern, Prozentsätze, E-Mail-Adressen und HTTP-Adressen. Sie können bestimmte Typen von nicht linguistischen Elementen im Abschnitt Nicht linguistische Elemente: Konfiguration des erweiterten Ressourceneditors ein- bzw. ausschließen. Durch Deaktivierung unnötiger Elemente vergeudet die Extrahierungsengine keine Verarbeitungszeit. [Für weitere Informationen siehe Thema Konfiguration in Kapitel 11 auf S. 250.](#)

Großbuchstaben-Algorithmus. Diese Option extrahiert einfache und zusammengesetzte Ausdrücke, die sich nicht in den integrierten Wörterbüchern befinden, solange der erste Buchstabe des Begriffs in Großbuchstaben geschrieben ist. Diese Option ist eine gute Möglichkeit, die geeignetsten Substantive zu extrahieren.

Teilweise und vollständige Personennamen, wenn möglich, gruppieren. Diese Option gruppiert Namen, die zusammen im Text unterschiedlich erscheinen. Diese Funktion ist nützlich, da Namen zu Beginn des Textes oft in voller Länge angegeben werden und später nur noch mit einer Kurzform auf sie verwiesen wird. Diese Option versucht, jeden Uniterm mit dem Typ <Unknown> mit dem letzten Wort aller zusammengesetzten Ausdrücke abzugleichen, die dem Typ <Person> zugeordnet sind. Wird beispielsweise *doe* gefunden und anfänglich dem Typ <Unknown> zugeordnet, überprüft die Extrahierungsengine, ob ein zusammengesetzter Ausdruck vom Typ <Person> als letztes Wort *doe* enthält, z. B. *john doe*. Diese Option wird nicht auf Vornamen angewendet, da sie in den meisten Fällen nicht als Uniterms extrahiert werden.

Maximale Nicht-Funktions-Wortpermutation. Diese Option gibt die maximale Anzahl von Füllwörtern an, die für die Anwendung des Permutationsverfahrens vorhanden sein müssen. Dieses Permutationsverfahren gruppiert ähnliche Phrasen, die sich nur durch die enthaltenen Füllwörter (zum Beispiel von und der) unabhängig von der Beugung unterscheiden. Nehmen wir zum Beispiel an, dass Sie diesen Wert auf höchstens zwei Wörter eingestellt haben und sowohl Unternehmen des Vertreters und Vertreter des Unternehmens extrahiert wurden. In diesem Fall würden beide extrahierte Ausdrücke in der endgültigen Konzeptliste zusammen gruppiert, da beide Ausdrücke als gleich betrachtet werden, wenn des ignoriert wird.

Dieses Dialogfeld vor der Extrahierung immer anzeigen. Legen Sie fest, ob Sie das Dialogfeld "Extrahierungseinstellungen" bei jeder Extrahierung anzeigen möchten, ob Sie es nie anzeigen möchten (außer beim Aufruf über das Menü "Extras") oder ob Sie bei jeder Extrahierung gefragt werden möchten, ob Sie Änderungen an den Extrahierungseinstellungen vornehmen wollen.

Extrahierungsergebnisse speichern

Beim Extrahieren werden die Ergebnisse im Bereich “Extrahierungsergebnisse” angezeigt und können zur Kategorisierung der Treffer verwendet werden. Diese Extrahierungsergebnisse werden während einer IBM® SPSS® Text Analytics for Surveys-Sitzung gespeichert, sodass Sie mit ihnen arbeiten können. Standardmäßig werden die Extrahierungsergebnisse in den Projekten gespeichert. Ob die Ergebnisse beim Schließen des Projekts gespeichert werden, ist eine globale Einstellung, die Sie jederzeit im Dialogfeld “Optionen” (Extras > Optionen) ändern können. [Für weitere Informationen siehe Thema Optionen: Registerkarte “System” in Kapitel 2 auf S. 19.](#)

Zur Sicherheit werden diese Extrahierungsergebnisse während des Speicherns verschlüsselt und in der Datenbank abgelegt. Dieses Verfahren erschwert es, Daten in der Datenbank ausfindig zu machen. Außerdem werden Extrahierungsergebnisse nie in SPSS Text Analytics for Surveys dargestellt, bis die Anwendung die Datenquelle gefunden hat. Wenn die Daten passwortgeschützt sind, müssen daher der Benutzername und das Passwort für diese Datenquelle eingegeben werden, um die Extrahierungsergebnisse auf dem Bildschirm anzuzeigen.

Extrahierungsergebnisse verfeinern

Die Extrahierung ist ein iterativer Prozess, in dem Sie extrahieren, die Ergebnisse überprüfen und ändern und dann erneut extrahieren können, um die Ergebnisse zu aktualisieren. Da Genauigkeit und Kontinuität für die erfolgreiche Durchführung von Text-Mining und Kategorisierung unverzichtbar sind, gewährleistet das Verfeinern Ihrer Extrahierungsergebnisse von Anfang an, dass Sie bei jeder erneuten Extrahierung genau dieselben Ergebnisse in Ihren Kategoriedefinitionen erhalten. So wird die Zuweisung von Datensätzen zu Ihren Kategorien genauer und wiederholbar.

Die Extrahierungsergebnisse dienen als Bausteine für Ihre Kategorien. Beim Erstellen von Kategorien mithilfe der Extrahierungsergebnisse werden Datensätze automatisch Kategorien zugewiesen, wenn sie Text enthalten, der mit einem oder mehreren Kategoriedeskriptoren übereinstimmt. Sie können zwar vor dem Verfeinern der linguistischen Ressourcen mit der Kategorisierung beginnen, aber es ist nützlich, die Extrahierungsergebnisse vorher mindestens einmal zu überprüfen.

Beim Überprüfen Ihrer Ergebnisse stoßen Sie möglicherweise auf Elemente, die die Extrahierungsengine anders verarbeiten soll. Betrachten Sie die folgende Beispiele:

- **Nicht erkannte Synonyme.** Angenommen, Sie entdecken mehrere Konzepte, die Sie als Synonyme betrachten, z. B. *schlau*, *intelligent*, *gescheit* und *klug*, und alle sind als einzelne Konzepte in den Extrahierungsergebnissen enthalten. Dann können Sie eine Synonymdefinition erstellen, nach der *intelligent*, *gescheit* und *klug* unter dem Zielkonzept *schlau* gruppiert werden. So werden diese Konzepte mit *schlau* zusammengefasst und die globale Häufigkeitsanzahl steigt ebenfalls. [Für weitere Informationen siehe Thema Synonyme hinzufügen auf S. 89.](#)
- **Den falschen Typen zugeordnete Konzepte.** Angenommen, die Konzepte in Ihren Extrahierungsergebnissen werden bei einem Typ aufgeführt und Sie möchten Sie einem anderen Typen zuweisen. Ein weiteres Beispiel: Sie finden in den Extrahierungsergebnissen 15 Konzepte zu Pflanzen und Sie möchten alle einem neuen Typ mit der Bezeichnung `<Pflanze>` hinzufügen. Konzepte, die in keinem Typen-Wörterbuch gefunden, aber aus dem Text extrahiert werden, erhalten automatisch den Typ `<Unbekannt>`. Sie können

Konzepte zu Typen hinzufügen. [Für weitere Informationen siehe Thema Konzepte zu Typen hinzufügen auf S. 91.](#)

- **Bedeutungslose Konzepte.** Angenommen, Sie stoßen auf ein extrahiertes Konzept mit einer sehr hohen Häufigkeitsanzahl—, d. h., es kommt in vielen Datensätzen vor. Sie halten dieses Konzept aber für bedeutungslos für Ihre Analyse. Dann können Sie es von der Extrahierung ausschließen. [Für weitere Informationen siehe Thema Konzepte von der Extrahierung ausschließen auf S. 94.](#)
- **Falsch erkannte Übereinstimmungen.** Angenommen, Sie stellen bei der Überprüfung der Datensätze mit einem bestimmten Konzept fest, dass zwei Wörter fälschlicherweise gruppiert worden sind, z. B. `Fakultät` und `Faktura`. Diese Übereinstimmung kann durch einen als Fuzzy-Gruppierung bezeichneten internen Algorithmus entstehen, mit dem vorübergehend doppelt/dreifach auftretende Konsonanten und Vokale ignoriert werden, um häufige Schreibfehler zu gruppieren. Nehmen Sie diese Wörter in eine Liste von Wortpaaren auf, die nicht gruppiert werden sollen. [Für weitere Informationen siehe Thema Unschärfe Gruppierung in Kapitel 11 auf S. 245.](#)
- **Nicht extrahierte Konzepte.** Angenommen, Sie erwarten die Extrahierung bestimmter Konzepte, stellen jedoch bei der Überprüfung des Datensatztexts fest, dass einige Wörter oder Ausdrücke nicht extrahiert worden sind. Häufig handelt es sich bei diesen Wörtern um Verben oder Adjektive, die für Sie uninteressant sind. Manchmal möchten Sie vielleicht trotzdem nicht extrahierte Wörter oder Ausdrücke als Teil einer Kategoriedefinition verwenden. Um das Konzept zu extrahieren, können Sie die Aufnahme eines Fachausdrucks in ein Typ-Wörterbuch erzwingen. [Für weitere Informationen siehe Thema Extrahierung von Wörtern erzwingen auf S. 95.](#)

Viele dieser Änderungen können Sie direkt im Bereich “Extrahierungsergebnisse” oder im Datenbereich vornehmen, indem Sie mit einem Klick auf die rechte Maustaste auf die Kontextmenüs zugreifen.

Wenn Sie die Änderungen vorgenommen haben, wechselt die Hintergrundfarbe des Bereichs und zeigt so, dass zum Anzeigen der Änderungen eine erneute Extrahierung erforderlich ist. [Für weitere Informationen siehe Thema Daten extrahieren auf S. 85.](#) Beim Arbeiten mit größeren Datensätzen kann es effizienter sein, erst nach mehreren Änderungen statt nach jeder einzelnen erneut zu extrahieren.

Anmerkung: In der Ansicht Resource Editor (Ansicht > Resource Editor) können Sie die Gesamtmenge der für die Gewinnung der Extrahierungsergebnisse verwendeten editierbaren linguistischen Ressourcen anzeigen. Diese Ressourcen werden in dieser Ansicht als Bibliotheken und Wörterbücher angezeigt. Sie können die Konzepte und Typen direkt in den Bibliotheken und Wörterbüchern anpassen. [Für weitere Informationen siehe Thema Mit Bibliotheken arbeiten in Kapitel 9 auf S. 209.](#)

Synonyme hinzufügen

Synonyme verknüpfen zwei oder mehr Wörter mit derselben Bedeutung. Synonyme werden häufig auch verwendet, um Fachausdrücke mit ihren Abkürzungen oder häufig falsch geschriebene Wörter mit der richtigen Schreibweise zu gruppieren. Durch die Verwendung von Synonymen nimmt die Häufigkeit des Zielkonzepts zu, sodass ähnliche Informationen, die in unterschiedlicher Form in den Textdaten vorhanden sind, viel leichter zu erkennen sind.

Die im Lieferumfang enthaltenen Vorlagen für linguistische Ressourcen und Bibliotheken beinhalten bereits viele vordefinierte Synonyme. Sie können jedoch nicht erkannte Synonyme definieren, sodass sie bei der nächsten Extrahierung erkannt werden.

Im ersten Schritt legen Sie das Zielkonzept oder Hauptkonzept fest. Das **Zielkonzept** ist das Wort oder der Ausdruck, unter dem Sie alle synonymen Fachausdrücke in den endgültigen Ergebnissen gruppieren möchten. Während der Extrahierung werden die Synonyme unter diesem Zielkonzept gruppiert. Im zweiten Schritt werden alle Synonyme für dieses Konzept ermittelt. Bei der endgültigen Extrahierung werden alle Synonyme durch das Zielkonzept substituiert. Ein Fachausdruck muss extrahiert werden, um ein Synonym sein zu können. Das Zielkonzept muss jedoch nicht extrahiert werden, damit die Substitution stattfinden kann. Wenn Sie z. B. *intelligent* durch *schlau* ersetzen möchten, dann ist *intelligent* das Synonym und *schlau* das Zielkonzept.

Beim Erstellen einer neuen Synonymdefinition wird ein neues Zielkonzept in das Wörterbuch aufgenommen. Anschließend nehmen Sie Synonyme in das Zielkonzept auf. Wenn Sie Synonyme erstellen oder bearbeiten, werden die Änderungen im Resource Editor in Synonymwörterbücher aufgenommen. Um den gesamten Inhalt der Synonymwörterbücher anzuzeigen oder eine beträchtliche Anzahl von Änderungen vorzunehmen, sollten Sie direkt im Resource Editor arbeiten. [Für weitere Informationen siehe Thema Substitutions-/Synonymwörterbücher in Kapitel 10 auf S. 234.](#)

Neue Synonyme werden automatisch in der ersten Bibliothek im Bibliotheksbaum in der Ansicht Resource Editor gespeichert – standardmäßig ist dies die *Local Library*.

Anmerkung: Wenn Sie vergeblich nach einer Synonymdefinition in den Kontextmenüs oder direkt im Resource Editor suchen, könnte eine Übereinstimmung möglicherweise aus einem internen unscharfen Gruppierungsverfahren resultieren. [Für weitere Informationen siehe Thema Unschärfe Gruppierung in Kapitel 11 auf S. 245.](#)

So erstellen Sie ein neues Synonym:

- ▶ Wählen Sie die Konzepte, für die Sie ein neues Synonym erstellen möchten, im Bereich “Extrahierungsergebnisse” oder im Datenbereich aus.
- ▶ Wählen Sie in den Menüs Bearbeiten > Zu Synonym hinzufügen> Neu. Das Dialogfeld “Synonym erstellen” wird geöffnet.

Abbildung 5-7
Dialogfeld “Synonym erstellen”



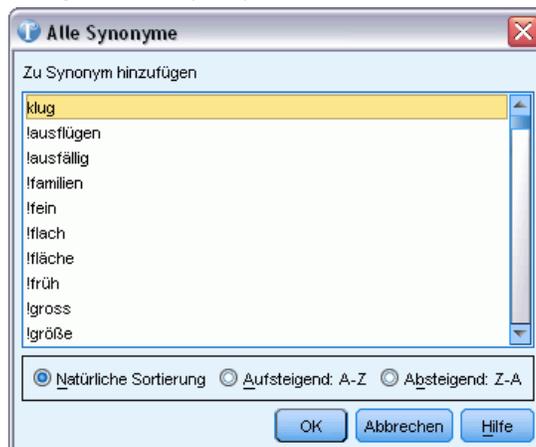
- ▶ Geben Sie im Textfeld “Ziel” ein Zielkonzept ein. Unter diesem Konzept werden alle Synonyme gruppiert.

- ▶ Um weitere Synonyme aufzunehmen, geben Sie sie in das Listenfeld “Synonyme” ein. Trennen Sie die synonymen Fachausdrücke mit dem globalen Trennzeichen. [Für weitere Informationen siehe Thema Optionen: Registerkarte “System” in Kapitel 2 auf S. 19.](#)
- ▶ Klicken Sie auf OK, um die Änderungen anzuwenden. Das Dialogfeld wird geschlossen und die Hintergrundfarbe im Bereich “Extrahierungsergebnisse” wechselt und zeigt so, dass zum Anzeigen der Änderungen eine erneute Extrahierung erforderlich ist. Nehmen Sie alle gewünschten Änderungen vor der erneuten Extrahierung vor.

So fügen Sie Konzepte zu einem Synonym hinzu:

- ▶ Wählen Sie die Konzepte, die Sie zu einer vorhandenen Synonymdefinition hinzufügen möchten, im Bereich “Extrahierungsergebnisse” oder im Datenbereich aus.
- ▶ Wählen Sie in den Menüs Bearbeiten > Zu Synonym hinzufügen>. Im Menü werden die Synonyme angezeigt, wobei das zuletzt erstellte am Anfang der Liste steht. Wählen Sie den Namen des Synonyms aus, zu dem Sie die ausgewählten Konzepte hinzufügen möchten. Wenn Sie das gesuchte Synonym entdecken, dann wählen Sie es aus und die ausgewählten Konzepte werden zu dieser Synonymdefinition hinzugefügt. Wenn Sie es nicht entdecken, wählen Sie Weitere aus, um das Dialogfeld “Alle Synonyme” anzuzeigen.

Abbildung 5-8
Dialogfeld “Alle Synonyme”



- ▶ Im Dialogfeld “Alle Synonyme” können Sie die Liste in der natürlichen Reihenfolge (Erstellungsreihenfolge) oder in aufsteigender bzw. absteigender Reihenfolge sortieren. Wählen Sie den Namen des Synonyms aus, zu dem Sie die ausgewählten Konzepte hinzufügen möchten, und klicken Sie auf OK. Das Dialogfeld wird geschlossen und die Konzepte werden zu den Synonymdefinitionen hinzugefügt.

Konzepte zu Typen hinzufügen

Bei der Extrahierung werden die extrahierten Konzepte Typen zugewiesen, damit Fachausdrücke mit Gemeinsamkeiten gruppiert werden. Es sind bereits zahlreiche Typen im Lieferumfang von IBM® SPSS® Text Analytics for Surveys enthalten. [Für weitere Informationen siehe Thema](#)

[Integrierte Typen in Kapitel 10 auf S. 224](#). Konzepte, die in keinem Typen-Wörterbuch gefunden, aber aus dem Text extrahiert werden, erhalten automatisch den Typ <Unbekannt>.

Wenn Sie Ihre Ergebnisse prüfen, finden Sie eventuell einige Konzepte, die in einem Typ erscheinen, die Sie einem anderen zuweisen möchten, oder Sie stellen fest, dass eine Gruppe von Wörtern eigentlich in einen neuen Typ gehört. In diesen Fällen sollten Sie die Konzepte einem anderen Typ neu zuweisen oder einen neuen Typ erstellen.

Angenommen, Sie arbeiten z. B. mit Umfragedaten zu Kraftfahrzeugen und möchten mit Fokussierung auf verschiedene Fahrzeugbereiche kategorisieren. Sie können einen Typ <Armaturenbrett> erstellen, um alle Konzepte zu Anzeigeinstrumenten und Schaltern von Armaturenbrettern bei Fahrzeugen zu gruppieren. Dann können Sie Konzepte wie Tankanzeige, Heizung, Radio und Kilometerzähler diesem neuen Typ zuweisen.

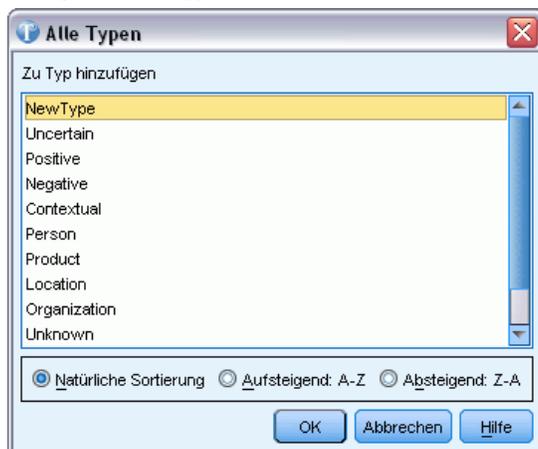
Oder angenommen, Sie arbeiten mit Umfragedaten zu Universitäten und Fachhochschulen und beim Extrahieren hat Johann Wolfgang Goethe (die Universität) den Typ <Person> statt des Typs <Organization> erhalten. In diesem Fall fügen Sie dieses Konzept dem Typ <Organization> hinzu.

Wenn Sie einen Typ erstellen oder Konzepte als Fachausdrücke zu einer Ausdruckliste eines Typs hinzufügen, werden die Änderungen in Typ-Wörterbüchern in den Bibliotheken der linguistischen Ressourcen im Resource Editor aufgezeichnet. Um den Inhalt dieser Bibliotheken anzuzeigen oder eine beträchtliche Anzahl von Änderungen vorzunehmen, sollten Sie direkt im Resource Editor arbeiten. [Für weitere Informationen siehe Thema Hinzufügen von Fachausdrücken in Kapitel 10 auf S. 227](#).

So fügen Sie ein Konzept zu einem Typ hinzu:

- ▶ Wählen Sie die Konzepte, die Sie zu einem vorhandenen Typ hinzufügen möchten, im Bereich “Extrahierungsergebnisse” oder im Datenbereich aus.
- ▶ Klicken Sie mit der rechten Maustaste, um das Kontextmenü aufzurufen.
- ▶ Wählen Sie in den Menüs Bearbeiten > Zu Typ hinzufügen>. Im Menü werden die Typen angezeigt, wobei der zuletzt erstellte am Anfang der Liste steht. Wählen Sie den Namen des Typs aus, zu dem Sie die ausgewählten Konzepte hinzufügen möchten. Wenn Sie den gesuchten Typnamen sehen, dann wählen Sie ihn aus und die ausgewählten Konzepte werden zu diesem Typ hinzugefügt. Wenn Sie ihn nicht sehen, wählen Sie Weitere aus, um das Dialogfeld “Alle Typen” anzuzeigen.

Abbildung 5-9
Dialogfeld "Alle Typen"



- ▶ Im Dialogfeld "Alle Typen" können Sie die Liste in der natürlichen Sortierung (Erstellungsreihenfolge) oder in aufsteigender bzw. absteigender Reihenfolge sortieren. Wählen Sie den Namen des Typs aus, dem Sie die ausgewählten Konzepte hinzufügen möchten, und klicken Sie auf OK. Das Dialogfeld wird geschlossen und die Konzepte werden als Fachausdrücke zu den Typen hinzugefügt.

So erstellen Sie einen "Neuen Typ":

- ▶ Wählen Sie die Konzepte, für die Sie einen neuen Typ erstellen möchten, im Bereich "Extrahierungsergebnisse" oder im Datenbereich aus.
- ▶ Wählen Sie in den Menüs Bearbeiten > Zu Typ hinzufügen> Neu. Das Dialogfeld "Typ-Eigenschaften" wird geöffnet.

Abbildung 5-10
Dialogfeld "Typ-Eigenschaften"



- ▶ Geben Sie einen neuen Namen für diesen Typ im Textfeld “Name” ein und nehmen Sie die gewünschten Änderungen in den anderen Feldern vor. [Für weitere Informationen siehe Thema Erstellen von Typen in Kapitel 10 auf S. 225.](#)
- ▶ Klicken Sie auf OK, um die Änderungen anzuwenden. Das Dialogfeld wird geschlossen und die Hintergrundfarbe im Bereich “Extrahierungsergebnisse” wechselt und zeigt so, dass zum Anzeigen der Änderungen eine erneute Extrahierung erforderlich ist. Nehmen Sie alle gewünschten Änderungen vor der erneuten Extrahierung vor.

Konzepte von der Extrahierung ausschließen

Bei der Überprüfung der Ergebnisse entdecken Sie ggf. gelegentlich unerwünschte Konzepte oder Konzepte, die von automatisierten Kategorieaufbauverfahren verwendet werden. In einigen Fällen haben diese Konzepte eine hohe globale Häufigkeitsanzahl, sind aber für Ihre Analyse völlig bedeutungslos. In diesem Fall markieren Sie ein Konzept, um ihn aus der endgültigen Extrahierung auszuschließen. In der Regel sind Konzepte in dieser Liste Füllwörter oder -ausdrücke zur Verbesserung des Textflusses, die keine Bedeutung tragen und die Extrahierungsergebnisse unnötig anfüllen. Wenn Sie diese Konzepte in das Ausschlusswörterbuch aufnehmen, verhindern Sie ihre Extrahierung.

Durch das Ausschließen von Konzepten werden bei der nächsten Extrahierung alle Varianten des ausgeschlossenen Konzepts aus den Extrahierungsergebnissen entfernt. Wenn dieses Konzept bereits als ein Deskriptor in einer Kategorie angezeigt wird, bleibt er nach der erneuten Extrahierung mit der Anzahl null in der Kategorie.

Wenn Sie ausschließen, werden die Änderungen im Resource Editor in einem Ausschlusswörterbuch aufgezeichnet. Um alle ausgeschlossenen Definitionen anzuzeigen und sie direkt zu bearbeiten, sollten Sie direkt im Resource Editor arbeiten. [Für weitere Informationen siehe Thema Ausschlusswörterbücher in Kapitel 10 auf S. 240.](#)

So schließen Sie Konzepte aus

- ▶ Wählen Sie die Konzepte, die Sie von der Extrahierung ausschließen möchten, im Bereich “Extrahierungsergebnisse” oder im Datenbereich aus.
- ▶ Klicken Sie mit der rechten Maustaste, um das Kontextmenü aufzurufen.
- ▶ Wählen Sie Aus Extrahierung ausschließen aus. Das Konzept wird im Resource Editor dem Ausschlusswörterbuch hinzugefügt und die Hintergrundfarbe im Bereich “Extrahierungsergebnisse” wechselt und zeigt so, dass zum Anzeigen der Änderungen eine erneute Extrahierung erforderlich ist. Nehmen Sie alle gewünschten Änderungen vor der erneuten Extrahierung vor.

Anmerkung. Ausgeschlossene Wörter werden automatisch in der ersten Bibliothek im Bibliotheksbaum im Resource Editor gespeichert – standardmäßig ist dies die *Local Library*.

Extrahierung von Wörtern erzwingen

Wenn Sie nach der Extrahierung die Textdaten im Bereich “Daten” überprüfen, stellen Sie möglicherweise fest, dass einige Wörter oder Ausdrücke nicht extrahiert wurden. Häufig handelt es sich bei diesen Wörtern um Verben oder Adjektive, die für Sie uninteressant sind. Manchmal möchten Sie vielleicht trotzdem nicht extrahierte Wörter oder Ausdrücke als Teil einer Kategoriedefinition verwenden.

Wenn diese Wörter und Ausdrücke extrahiert werden sollen, haben Sie zwei Optionen:

- Einen Begriff in eine Typenbibliothek zwingen. [Für weitere Informationen siehe Thema Erzwingen von Fachausdrücken in Kapitel 10 auf S. 231.](#)
- Wörter direkt zu einer vorhandenen Kategoriedefinition hinzufügen. Dies wird im Allgemeinen angewendet, wenn die erste Option nicht zu den erwarteten Ergebnissen geführt hat. [Für weitere Informationen siehe Thema Textübereinstimmungen in Kategorien in Kapitel 6 auf S. 167.](#)

Wichtig: Einen Fachausdruck in einem Wörterbuch als erzwungen zu markieren, ist kein absolut sicheres Verfahren. Das heißt, obwohl ein Fachausdruck explizit einem Wörterbuch hinzugefügt worden ist, ist er möglicherweise nach einer erneuten Extrahierung nicht im Bereich “Extrahierungsergebnisse” vorhanden oder er wird nicht genau so angezeigt, wie Sie ihn deklariert haben. Dies kommt selten vor, ist aber möglich, wenn ein Wort oder Ausdruck bereits als Teil eines längeren Ausdrucks extrahiert wurde. Um das zu verhindern, wenden Sie die Abgleichoption Gesamt (keine Zusammensetzungen) auf diesen Ausdruck im Typwörterbuch an. [Für weitere Informationen siehe Thema Hinzufügen von Fachausdrücken in Kapitel 10 auf S. 227.](#)

Kategorisieren von Textdaten

In der IBM® SPSS® Text Analytics for Surveys können Sie **Kategorien** erstellen, die im Prinzip allgemeinere (auf einer höheren Ebene liegende) Konzepte oder Themen darstellen, die die Grundideen, das Grundwissen bzw. die Grundhaltungen erfassen, die im Text ausgedrückt werden.

Kategorien können auch eine hierarchische Struktur besitzen, d. h., sie können Unterkategorien enthalten, die wiederum eigene Unterkategorien enthalten können usw. Sie können vordefinierte Kategorien, früher Coderahmen genannt, mit hierarchischen Kategorien importieren und diese hierarchischen Kategorien auch im Produkt aufbauen.

Hierarchische Kategorien ermöglichen Ihnen den Aufbau einer Baumstruktur mit einer oder mehreren Unterkategorien, die eine genauere Gruppierung von Elementen, beispielsweise verschiedenen Konzept- oder Themenbereichen, gestatten. Ein einfaches Beispiel könnte sich auf Freizeitaktivitäten beziehen: Bei der Beantwortung einer Frage wie *Welcher Aktivität würden Sie sich gerne widmen, wenn Sie mehr Zeit hätten?* könnten die Kategorien der ersten Ebene *Sport, Kunst und Handarbeiten, Angeln* usw. lauten. Auf der Ebene unter *Sport* könnten Sie Unterkategorien einrichten, um zu sehen, ob es sich um *Ballsportarten, Wassersportarten* usw. handelt.

Kategorien bestehen aus einer Reihe von Deskriptoren wie *Konzepten, Typen, Mustern* und *Kategorieregeln*. Diese Deskriptoren werden zusammen verwendet, um zu bestimmen, ob ein Datensatz zu einer gegebenen Kategorie gehört oder nicht. Der Text in einem Datensatz kann gescannt werden, um zu überprüfen, ob es Text gibt, der mit einem Deskriptor übereinstimmt. Liegt eine Übereinstimmung vor, wird der Datensatz dieser Kategorie zugeordnet. Dieser Prozess wird als **Kategorisierung** bezeichnet.

Mithilfe der in den vier Fensterbereichen angezeigten Daten können Sie Kategorien erstellen, damit arbeiten und sie visuell untersuchen. Jeder der vier Fensterbereiche kann durch Auswahl seines Namens im Menü "Ansicht" ein- bzw. ausgeblendet werden.

- **Fensterbereich "Kategorien"**. In diesem Fensterbereich können Sie Kategorien erstellen und verwalten. [Für weitere Informationen siehe Thema Der Fensterbereich "Kategorien" auf S. 97.](#)
- **Fensterbereich "Extrahierungsergebnisse"**. In diesem Fensterbereich können Sie mit den extrahierten Konzepten und Typen arbeiten. [Für weitere Informationen siehe Thema Extrahierte Ergebnisse: Konzepte, Typen und Muster in Kapitel 5 auf S. 80.](#)
- **Visualisierungsbereich**. In diesem Fensterbereich können Sie die Kategorien und ihre Interaktionen visuell untersuchen. [Für weitere Informationen siehe Thema Visualisierung von Diagrammen in Kapitel 7 auf S. 172.](#)
- **Datenbereich**. In diesem Fensterbereich können Sie den Text untersuchen und überprüfen, der in Datensätzen enthalten ist, die Ihrer Auswahl entsprechen. [Für weitere Informationen siehe Thema Der Fensterbereich "Daten" auf S. 100.](#)

Abbildung 6-1
Ansicht "Frage"

The screenshot shows a software interface for text analysis. The main window is titled "G1: What do you like most about this portable music player?". The interface is divided into several panes:

- Top Left:** A menu bar with "Datei", "Bearbeiten", "Ansicht", "Kategorien", "Extras", and "Hilfe". Below it is a toolbar with icons for "Aufbauen", "Erweitern", and "fx".
- Top Right:** A "Kategoriebaken" (Category Breadcrumbs) pane showing a tree structure of categories. The selected category is "songs++<Positive>".
- Middle Right:** A "Tabelle für Kategorienetzdiagramm" (Table for Category Network Diagram) showing a table with columns: "Kategorie", "Balken", "Auswahl %", "Beträge", and "Gesamt %". The table contains two rows:

Kategorie	Balken	Auswahl %	Beträge	Gesamt %
volume++<Contextual>		11,1	1	0,5
songs++<Positive>		88,9	8	2,2
- Bottom Left:** A "Nicht verwendete Extrahierungen" (Unused Extractions) pane showing a list of concepts like "product (18)", "lightweight (12)", "long-lasting (9)", etc.
- Bottom Right:** A table showing document classification results with columns: "ID", "Antwort", and "Kategorie".

ID	Antwort	Kategorie
1	Stores 5000 songs in a compact portable player.	songs++<Positive>
2	Listening to any songs like any time like.	songs++<Positive>
3	It has a lot of storage capacity. I can fit a lot of songs on it. Also, it's very lightweight.	songs++<Positive>
4	Product A is the best. Drag & drop songs from their site is really easy and I love all the random play options. Like the long battery life.	...ongs from their site++<Positive>
5	It like has all of my songs. I mean all of them, and it still has room for like a gazillion more.	songs++<Positive>
6	Well being a long haul truck driver I really like the fact that there is always music to listen to and that I can turn up the volume real loud when my favorite songs comes on.	songs++<Positive> volume++<Contextual>
7	It holds loads of songs and is easy to use.	songs++<Positive>
8	I can listen to only the songs that like.	songs++<Positive>

Sie können zwar mit einem Kategorien-Set aus einem Text Analysis Package (TAP) beginnen oder einen Import aus einer vordefinierten Kategoriendatei durchführen, aber eventuell müssen Sie auch Ihre eigenen erstellen. Kategorien können mit den leistungsfähigen automatisierten Methoden des Produkts automatisch erstellt werden, wobei Kategorien und deren Deskriptoren anhand von Extrahierungsergebnissen (Konzepte, Typen und Muster) erzeugt werden. Daneben können Sie Kategorien auch manuell erstellen und dabei zusätzliche Erkenntnisse mit einbeziehen, die Sie hinsichtlich der Datengrundlage möglicherweise gewonnen haben. Sie können Kategoriedefinitionen manuell erstellen, indem Sie die Extrahierungsergebnisse in die Kategorien ziehen und ablegen. Sie können diese Kategorien oder jede leere Kategorie anreichern, indem Sie einer Kategorie Kategorieregeln hinzufügen, Ihre eigenen vordefinierten Kategorien verwenden, ein Wort oder einen Ausdruck hinzufügen, der nie extrahiert wurde (als **Textabgleich** bezeichnet), und/oder indem Sie Antworten direkt zwangsweise in eine Kategorie einordnen.

Diese Verfahren und Methoden eignen sich jeweils gut für bestimmte Arten von Daten und Situationen, häufig ist es jedoch sinnvoll, in einer Analyse mehrere Verfahren zu kombinieren, um das gesamte Spektrum an Datensätzen zu erfassen. Außerdem können Ihnen im Verlauf der Kategorisierung andere Änderungen auffallen, die an den linguistischen Ressourcen vorgenommen werden sollten.

Der Fensterbereich "Kategorien"

Im Fensterbereich "Kategorien" können Sie Ihre Kategorien erstellen und verwalten. Dieser Fensterbereich befindet sich in der oberen linken Ecke der der Ansicht "Frage". Sie können über das Menü "Ansicht" darauf zugreifen (Ansicht > Frage > "Ihre_Frage"). Nachdem Sie die Konzepte

und Typen aus Ihren Textdaten extrahiert haben, können Sie Kategorien automatisch mithilfe von Verfahren wie Konzeptbeziehung, Kookkurrenz usw. oder manuell erstellen. [Für weitere Informationen siehe Thema Erstellen von Kategorien auf S. 112.](#)

Abbildung 6-2
Fensterbereich "Kategorien" mit und ohne Kategorien

Kategorie	Anzahl ...	Anzahl ...
Alle Datensätze	-	405
Nicht kategorisiert	-	405
Keine Konzepte extrahiert	-	405

Kategorie	Anzahl ...	Anzahl ...
Alle Datensätze	-	405
Nicht kategorisiert	-	282
Keine Konzepte extrahiert	-	4
songs+<Positive>	3	9
fx [listening to songs + like]		1
fx [songs + like] [songs + meets needs]		7
drag & dropped songs from their site+<P	1	1
memory	2	2
battery+<Contextual>	2	3
design+<Positive>	2	7
store+<Positive>	2	3
skipping+<Contextual>	2	2
look+<Positive>	2	3
feature+<Positive>	2	3
software+<Positive>	1	3

Immer wenn eine Kategorie erstellt oder aktualisiert wird, können die Datensätze automatisch gescannt werden, um zu sehen, ob es Text gibt, der einem Deskriptor in einer bestimmten Kategorie entspricht. Liegt eine Übereinstimmung vor, wird der Datensatz dieser Kategorie zugeordnet. Das Endergebnis besteht darin, dass die meisten, wenn nicht alle Datensätze, anhand der Deskriptoren in den Kategorien bestimmten Kategorien zugewiesen werden.

Kategoriebaumtabelle

Die Baumtabelle in diesem Bereich zeigt das Set von Kategorien, Unterkategorien und Deskriptoren an. Der Baum besitzt auch mehrere Spalten mit Informationen für jedes Bauelement. Folgende Spalten stehen ggf. zur Anzeige zur Verfügung:

- Code.** Listet den Codewert für jede Kategorie auf. Diese Spalte ist standardmäßig ausgeblendet. Sie können diese Spalte anzeigen, indem Sie mit der rechten Maustaste in die Baumtabelle klicken und die Optionsfolge Anzeigen > Kategoriecode auswählen.
- Kategorie.** Enthält den Kategoriebaum mit dem Namen der Kategorie und der Unterkategorien. Außerdem wird durch einen Klick auf das Deskriptor-Symboleistensymbol auch das Set der Deskriptoren angezeigt.
- Deskriptoren.** Gibt die Anzahl der Deskriptoren an, aus der die Kategorie besteht. Diese Zahl enthält nicht die Anzahl an Deskriptoren in den Unterkategorien. Es wird keine Zahl angezeigt, wenn ein Deskriptorname in der Spalte Kategorien angezeigt wird. Sie können diese Spalte anzeigen, indem Sie mit der rechten Maustaste in die Baumtabelle klicken und die Optionsfolge Anzeigen > Alle Deskriptoren auswählen.
- Dokumente.** Nach dem Scoring enthält diese Spalte die Anzahl von Datensätzen, die einer Kategorie und allen ihren Unterkategorien zugeordnet werden. Wenn also 5 Datensätze mit Ihrer obersten Kategorie auf der Basis ihrer Deskriptoren übereinstimmen und 7 unterschiedliche Datensätze mit einer Unterkategorie auf der Basis ihrer Deskriptoren übereinstimmen, ist die Gesamtanzahl an Dokumenten für die oberste Kategorie die Summe

dieser beiden, in diesem Fall also 12. Wenn jedoch der gleiche Datensatz mit der obersten Kategorie und ihrer Unterkategorie übereinstimmt, betrüge die Anzahl 11.

Wenn keine Kategorien vorhanden sind, enthält die Tabelle dennoch zwei Zeilen. Die oberste Zeile, Alle Datensätze, gibt die Gesamtzahl der Datensätze an. Die zweite Zeile Nicht kategorisiert, zeigt die Anzahl der Dokumente/Datensätze an, die noch kategorisiert werden müssen.

Bei jeder Kategorie im Fensterbereich geht dem Kategorienamen ein kleines Symbol in Form eines gelben Eimers voran. Wenn Sie auf eine Kategorie doppelklicken oder mit der rechten Maustaste in die Baumtabelle klicken und Kategoriedefinitionen auswählen, wird das Dialogfeld "Kategoriedefinitionen" geöffnet, in dem alle Elemente (so genannte **Deskriptoren** angezeigt werden, aus denen die Definition besteht, beispielsweise Konzepte, Typen, Muster und Kategorieregeln. [Für weitere Informationen siehe Thema Erläuterung von Kategorien auf S. 110.](#) Standardmäßig werden in der Kategoriebaumtabelle die Deskriptoren in den Kategorien nicht angezeigt. Wenn Sie die Deskriptoren direkt in der Tabelle und nicht im Dialogfeld "Kategoriedefinitionen" sehen möchten, klicken Sie auf die Umschaltfläche mit dem Stiftsymbol in der Symbolleiste. Durch Auswahl dieser Umschaltfläche können Sie Ihren Baum erweitern, um auch die Deskriptoren sehen zu können.

Scoren von Kategorien

Die Spalte Dokumente in der Kategoriebaumtabelle zeigt die Anzahl von Datensätzen, die dieser bestimmten Kategorie zugeordnet werden. Es erscheint ein Symbol in der Spalte, wenn die Zahlen veraltet sind oder nicht berechnet wurden. Denken Sie daran, dass der Scoring-Vorgang bei der Arbeit mit größeren Daten-Sets einige Zeit in Anspruch nehmen kann.

Auswahl von Kategorien im Baum

Wenn Sie eine Auswahl im Baum treffen, können Sie nur Schwesterkategorien auswählen, d. h., wenn Sie Kategorien der obersten Ebene auswählen, können Sie nicht auch eine Unterkategorie auswählen. Oder wenn Sie zwei Unterkategorien einer bestimmten Kategorie auswählen, können Sie nicht gleichzeitig eine Unterkategorie einer anderen Kategorie auswählen. Durch die Auswahl einer nicht zusammenhängenden Kategorie geht die vorherige Auswahl verloren.

Anzeige in den Fensterbereichen "Daten" und "Visualisierung"

Wenn Sie eine Zeile in der Tabelle auswählen, werden die Fensterbereiche "Visualisierung" und "Daten" automatisch entsprechend Ihrer Auswahl mit Informationen aktualisiert.

Verfeinern der Kategorien

Die Kategorisierung liefert möglicherweise nicht beim ersten Versuch ideale Ergebnisse für Ihre Daten und es kann Kategorien geben, die Sie löschen oder mit anderen Kategorien kombinieren möchten. Außerdem können Sie durch eine Überprüfung der Extrahierungsergebnisse herausfinden, dass einige Kategorien, die Sie für sinnvoll halten würden, nicht erstellt wurden. In diesem Fall können Sie manuelle Änderungen an den Ergebnissen vornehmen, um sie für den jeweiligen Kontext zu optimieren. [Für weitere Informationen siehe Thema Bearbeiten und Verfeinern von Kategorien auf S. 159.](#)

- Sie können Kategoriedefinitionen bearbeiten und erweitern und Kategorien verschieben, zusammenführen oder löschen. Weitere Informationen finden Sie unter [Bearbeiten und Verfeinern von Kategorien](#).
- Erzwingen Sie die Einordnung spezifischer Antwort-IDs in Kategorien oder ihre Entfernung daraus. [Für weitere Informationen siehe Thema Einschließen von Antworten in Kategorien erzwingen auf S. 165](#).
- Fügen Sie Textübereinstimmungen zu Kategorien hinzu, um Antworten zu erfassen, die den gleichen Text in der Kategorie enthalten. [Für weitere Informationen siehe Thema Textübereinstimmungen in Kategorien auf S. 167](#).
- Fügen Sie einer Kategorie Kategorieregeln hinzu, um Antworten aufgrund eines logischen Ausdrucks für eine Kategorie zu klassifizieren. [Für weitere Informationen siehe Thema Verwenden von Kategorieregeln auf S. 148](#).
- Visualisieren Sie, wie Ihre Kategorien zusammenarbeiten. [Für weitere Informationen siehe Thema Visualisierung von Diagrammen in Kapitel 7 auf S. 172](#).
- Exportieren Ihrer Kategorisierungsergebnisse. [Für weitere Informationen siehe Thema Exportieren von Kategorisierungsergebnissen in Kapitel 4 auf S. 55](#).

Der Fensterbereich “Daten”

Beim Erstellen von Kategorien kann es vorkommen, dass Sie einen Teil der Textdaten, mit denen Sie gerade arbeiten, überprüfen möchten. Wenn Sie beispielsweise eine Kategorie erstellen, in der 640 Datensätze kategorisiert sind, kann es erforderlich sein, einen Blick auf einige oder alle diese Datensätze zu werfen, um zu sehen, was dort tatsächlich geschrieben wurde. Sie können Datensätze im Datenbereich überprüfen, der sich unten rechts befindet. Wird dieser nicht standardmäßig angezeigt, wählen Sie die Befehlsfolge Ansicht > Bereiche > Daten.

In diesem Bereich werden die Antwortdatensätze für Ihre offene Fragedaten in tabellarischem Format präsentiert. Abhängig von Ihrer Auswahl in den anderen Bereichen dieser Ansicht werden nur die entsprechenden Datensätze in dem Bereich angezeigt. Wenn Sie beispielsweise ein Konzept im Bereich für Extrahierungsergebnisse ausgewählt haben, werden nur die Datensätze, die dieses Konzept (und verknüpfte Begriffe) enthalten, im Bereich “Daten” angezeigt.

Wenn Sie ein Konzept oder eine Kategorie in einem anderen Bereich markieren und die Daten anzeigen, werden die in diesen Dokumenten oder Datensätzen gefundenen Konzepte (und verknüpfte Begriffe) farblich hervorgehoben, damit Sie diese leicht im Text identifizieren können. Die Farbkodierung entspricht den Typen, die den Konzepten zugewiesen sind. Alternativ können Sie die Maus über farbkodierte Elemente bewegen, um das Konzept anzuzeigen, unter dem das betreffende Element extrahiert wurde, und den Typ, dem es zugewiesen wurde. Nicht extrahierter Text wird schwarz angezeigt. Bei diesen nicht extrahierten Wörtern handelt es sich meistens um Verbindungselemente (*und* oder *mit*), Pronomen (*mich* oder *sie*) und Verben (*ist*, *haben* oder *nehmen*).

Abbildung 6-3
Datenbereich

	Id	Antwort	Kategorien
1	14	I like that Product A has a lot of storage. Also, the interface is very easy to use.	memory device/memory
2	45	Everything! Product A rules! I can't wait to get a video one!	memory device/recording/video
3	68	Large storage capacity	...vice/memory/storage capacity
4	75	Small size. It has 512Mb of add-on memory, so it is quick to load and play music. It can also encode directly from external devices from the radio or a CD player.	...sumer electronics/home audio memory device/memory music radio size
5	76	storage capacity	...vice/memory/storage capacity
6	89	Small but lots of space (60 GB). Video is a bit of a toy but cool.	memory device/recording/video
7	98	Big storage capacity, also does video	...vice/memory/storage capacity memory device/recording/video
8	114	Large storage capacity and a good LCD screen for viewing digital photos	...vice/memory/storage capacity photos screen
9	120	The sound quality and the ability to record and mix my own play lists.	...nics/audio/sound/sound quality memory device/recording playlists
10	121	Storage capacity (40GB)	...vice/memory/storage capacity
11	146	It has a lot of storage capacity. I can fit a lot of songs on it. Also, it's very lightweight.	...vice/memory/storage capacity songs
	151	This has 256MB of memory; it holds about 50 songs. I've got another chip in my bag with another 50 songs on it. The cool thing about this is it cost \$200 plus \$60 for a portable	music computer network listening

Anmerkung: Um alle Datensätze für eine bestimmte Frage im Bereich “Daten” anzuzeigen, klicken Sie auf den Knoten Alle Datensätze am oberen Rand des Bereichs “Kategorien”.

Standardmäßig zeigt der Bereich “Daten” drei Spalten (ID, Treffer und Kategorien) an. Sie können diesem Bereich jedoch weitere Spalten hinzufügen. Folgende Spalten sind möglich:

- **ID.** Führt die Datensatz- bzw. Dokument-ID auf, sofern eine importiert wurde.
- **Treffer.** Führt die Textdaten auf, aus denen Konzepte und Typen extrahiert wurden.
- **Kategorien.** Führt jede Kategorie auf, der ein Datensatz angehört. Wenn diese Spalte angezeigt wird, kann die Aktualisierung des Datenbereichs ein wenig länger dauern, da jeweils die aktuellsten Informationen angezeigt werden. Kategorien werden in dieser Spalte nach ihrer Relevanz für den Datensatz aufgeführt. [Für weitere Informationen siehe Thema Kategorierelevanz auf S. 102.](#)
- **Erzwungenermaßen einschließen.** Listet die Kategorien auf, in die Sie eine Antwort erzwungenermaßen eingeschlossen haben. Antworten lassen sich mithilfe der Menüauswahl Bearbeiten > Erzwungenermaßen einschließen in eine Kategorie zwingen. [Für weitere Informationen siehe Thema Einschließen von Antworten in Kategorien erzwingen auf S. 165.](#)
- **Erzwungenermaßen ausschließen.** Listet die Kategorien auf, aus denen Sie eine Antwort erzwungenermaßen ausgeschlossen haben. Antworten lassen sich mithilfe der Menüauswahl Bearbeiten > Erzwungenermaßen ausschließen aus einer Kategorie ausschließen. In der Regel wird dies verwendet, wenn eine Antwort durch den Sarkasmus eines Befragten falsch kategorisiert wurde. [Für weitere Informationen siehe Thema Einschließen von Antworten in Kategorien erzwingen auf S. 165.](#)

- **Textübereinstimmung.** Listet sämtliche gefundenen Textübereinstimmungen für jede Antwort auf. Zeichenfolgen können so definiert werden, dass sie bestimmten Text in eine Kategoriedefinition zwingen, unabhängig davon, ob diese Zeichenfolge extrahiert wurde. [Für weitere Informationen siehe Thema Textübereinstimmungen in Kategorien auf S. 167.](#)
- **Kategorieanzahl.** Liefert die Gesamtanzahl der Kategorien, denen die Antwort auf diese Frage angehört.
- **Relevanzrang.** Führt den Rang jedes Datensatzes einer einzelnen Kategorie auf. Dieser Rang gibt an, wie gut der Datensatz im Verhältnis zu den anderen Datensätzen in der Kategorie zu der Kategorie passt. Wählen Sie eine Kategorie im Fensterbereich “Kategorien” (oben links) aus, um den Rang in dieser Spalte anzuzeigen. [Für weitere Informationen siehe Thema Kategorierelevanz auf S. 102.](#)
- **Antwort-Flags.** Fügt eine Spalte hinzu, die alle Antwort-Flags anzeigt, die Sie eventuell verwenden. Indem Sie in diese Spalte klicken, können Sie den Typ des Flags ändern, den Sie jeder Antwort zuweisen.
- **<alle Referenzvariablenamen>.** Fügt dem Bereich “Daten” eine Spalte für die ausgewählte Referenzvariable hinzu. Wenn Sie keine Referenzvariablen importiert haben, werden hier auch keine angeboten. Für jede Referenzvariable ist eine eigene Spalte verfügbar. [Für weitere Informationen siehe Thema Auswählen von Variablen in Kapitel 3 auf S. 35.](#)

So zeigen Sie andere Datenbereichsspalten an

- ▶ Klicken Sie innerhalb des Bereichs “Daten” mit der rechten Maustaste auf eine Spaltenüberschrift, um das Kontextmenü zu öffnen.
- ▶ Wählen Sie Spalten anzeigen aus dem Menü und wählen Sie dann die Spalte aus, die im Fensterbereich “Daten” angezeigt werden soll. Die neue Spalte wird im Datenbereich angezeigt.

Anmerkung: Das Erzwingen des Einschlusses von Antworten in Kategorien bzw. des Ausschlusses von Aufnahmen aus Kategorien ermöglicht es Ihnen, die durch automatische Kategorieerstellungsmethoden erstellten Kategoriedefinitionen zu überschreiben, ohne die tatsächliche Kategoriedefinition zu ändern. [Für weitere Informationen siehe Thema Einschließen von Antworten in Kategorien erzwingen auf S. 165.](#)

Kategorierelevanz

Um bessere Kategorien aufzubauen, können Sie die Relevanz der Datensätze in jeder Kategorie sowie die Relevanz aller Kategorien überprüfen, zu der ein Datensatz gehört.

Relevanz einer Kategorie für einen Datensatz

Wenn ein Datensatz im Datenbereich angezeigt wird, werden alle zugehörigen Kategorien in der Spalte “Kategorien” aufgeführt. Wenn ein Datensatz zu mehreren Kategorien gehört, werden die Kategorien in dieser Spalte von der relevantesten absteigend bis zur am wenigsten relevanten angezeigt. Die erste Kategorie stimmt also am besten mit dem Datensatz überein. [Für weitere Informationen siehe Thema Der Fensterbereich “Daten” auf S. 100.](#)

Relevanz eines Datensatzes für eine Kategorie

Wenn Sie eine Kategorie auswählen, können Sie die Relevanz der jeweiligen Datensätze im Datenbereich in der Spalte “Relevanzrang” überprüfen. Dieser Relevanzrang gibt an, wie gut der Datensatz im Verhältnis zu anderen Datensätzen in dieser Kategorie in die gewählte Kategorie passt. Um den Rang der Datensätze für eine einzelne Kategorie anzuzeigen, wählen Sie diese Kategorie im Fensterbereich “Kategorien” (oben links) aus und der Rang für den Datensatz wird in der Spalte angezeigt. Diese Spalte wird standardmäßig nicht eingeblendet. Sie können jedoch festlegen, dass sie eingeblendet werden soll. [Für weitere Informationen siehe Thema Der Fensterbereich “Daten” auf S. 100.](#)

Je niedriger die Rangnummer eines Datensatzes, desto besser passt er zur gewählten Kategorie, das heißt, eine 1 steht für die beste Entsprechung. Wenn mehrere Datensätze dieselbe Relevanz aufweisen, werden alle mit demselben Rang gefolgt von einem Gleichheitszeichen (=) angezeigt, um zu kennzeichnen, dass sie dieselbe Relevanz besitzen. Sie könnten beispielsweise folgende Ränge haben: 1=, 1=, 3, 4 usw., d. h., es gibt zwei Datensätze, von denen beide die beste Übereinstimmung für diese Kategorie aufweisen.

Tipp: Sie können den Text des relevantesten Datensatzes in die Anmerkung zur Kategorie einfügen, um eine bessere Beschreibung für diese Kategorie zu erstellen. Fügen Sie den Text direkt über den Datenbereich ein, indem Sie den Text und anschließend die Menüoptionen Kategorien > Zu Anmerkung hinzufügen auswählen.

Abbildung 6-4

Datenbereich mit Kategorien und Relevanzrang

	ID	Antwort	Kategorien	Relevanzrang
1	1	Zimmer sauber	zimmer	10=
2	4	Sauberes Zimmer	zimmer	10=
3	5	die komfortablen Zimmer, reichhaltiges Frühstück, netter Service	frühstück service zimmer	20=
4	26	Das gute Essen, das freundliche Personal, das Zimmer, die Sauberkeit, die Aussicht, die Nähe zur Stadtmitte, die Nähe zum Zentrum, die Nähe zum Strand, die Nähe zum Meer, die Verkehrsanbindung, die Erreichbarkeit	essen personal zimmer	26
5	27	Das Wetter, die Aussicht, der Internetanschluß, das Zimmer, die Fitnessmöglichkeiten, der Wellnessbereich	wellnessbereich zimmer	25
6	28	Die Zimmer und der Service	service zimmer	23
7	31	Kleines aber gemütliches Zimmer	zimmer	7=
8	35	Die Zimmer sind sehr ansprechend gestaltet	zimmer	10=
9	38	Die Zimmer waren neu renoviert und sauber und für diesen Preis recht geräumig.	zimmer	4=
10	47	Die Zimmer sind ordentlich sauber und werden täglich gereinigt	zimmer	10=
11	60	Die Zimmer sind sehr gut eingerichtet und sauber	zimmer	2=
12	74	Zimmer war gut, sauber und sehr gepflegt. Klimaanlage war sogar zimmerweise regelbar.	zimmer	1
13	91	die Zimmer sind sehr modern und komfortabel, das Bad groß	zimmer bad	4=
14	107	Super Service, Essen, Hotel, Zimmer.	service essen hotel zimmer	24
15	109	Sehr geräumige Zimmer	zimmer	10=

Methoden und Strategien zur Erstellung von Kategorien

Wenn Sie noch keine Extrahierung durchgeführt haben oder die Extrahierungsergebnisse nicht auf dem neuesten Stand sind, werden Sie automatisch bei Verwendung einer dieser Verfahren zum Aufbau oder zur Erweiterung einer Kategorie zu einer Extrahierung aufgefordert. Nachdem Sie ein Verfahren angewendet haben, stehen die Konzepte und Typen, die in eine Kategorie gruppiert wurden, weiterhin für die Kategorieerstellung mit anderen Verfahren zur Verfügung. Das bedeutet, dass Sie unter Umständen ein Konzept in mehreren Kategorien sehen, es sei denn, Sie beschließen, diese nicht wiederzuverwenden.

Als Unterstützung zur Ermittlung der besten Kategorien, beachten Sie bitte die folgenden Abschnitte:

- **Methoden für die Kategorieerstellung**
- **Strategien für die Kategorieerstellung**
- **Tipps zur Erstellung von Kategorien**

Methoden für die Kategorieerstellung

Da jedes Daten-Set anders ist, können sich die Anzahl der Kategorieerstellungsmethoden und die Reihenfolge, in der sie angewendet werden, im Laufe der Zeit ändern. Da zusätzlich Ihre Text-Mining-Ziele von Daten-Set zu Daten-Set unterschiedlich sein können, müssen Sie möglicherweise mit den verschiedenen Methoden experimentieren, um zu sehen, welches die besten Ergebnisse für die jeweiligen Textdaten bringt. Keines der automatischen Verfahren ergibt eine perfekte Kategorisierung Ihrer Daten, weshalb wir empfehlen, ein oder mehrere automatische Verfahren anzuwenden, die gut mit Ihren Daten funktionieren.

Neben der Verwendung von Text Analysis Packages (TAPs, *.tap) mit vordefinierten Kategoriensets können Sie Ihre Antworten auch mit jeder Kombination der folgenden Methoden kategorisieren:

- **Automatische Erstellungsverfahren.** Verschiedene linguistisch basierte und häufigkeitsbasierte Kategorieoptionen stehen zur Verfügung, um automatisch Kategorien für Sie zu erstellen. [Für weitere Informationen siehe Thema Erstellen von Kategorien auf S. 112.](#)
- **Automatische Erweiterungsverfahren.** Verschiedene linguistische Verfahren stehen zur Verfügung, um vorhandene Kategorien durch Hinzufügen und Verbessern von Deskriptoren zu erweitern, so dass sie mehr Datensätze erfassen. [Für weitere Informationen siehe Thema Erweitern von Kategorien auf S. 129.](#)
- **Manuelle Verfahren.** Es gibt verschiedene manuelle Methoden wie Drag-and-Drop. [Für weitere Informationen siehe Thema Manuelle Erstellung von Kategorien auf S. 134.](#)
- **Coderahmen.** Importieren Sie Ihren eigenen Coderahmen oder Code in den Coderahmenmanager kopieren/einfügen. [Für weitere Informationen siehe Thema Importieren vordefinierter Kategorien auf S. 136.](#)

Strategien für die Kategorieerstellung

Die folgende Liste der Strategien ist keinesfalls vollständig, kann Ihnen aber einige Ideen geben, sich dem Aufbau Ihrer Kategorien zu nähern.

- Wenn Sie ein Projekt beginnen, wählen Sie ein Kategorienset aus einem Text Analysis Package (TAP), um Ihre Analyse mit einigen vordefinierten Kategorien zu beginnen. Diese Kategorien kategorisieren Ihren Text vielleicht gleich zu Beginn in ausreichender Form. Wenn Sie jedoch weitere Kategorien hinzufügen möchten, können Sie die Kategorieaufbaueinstellungen (Kategorien > Aufbaueinstellungen) bearbeiten. Öffnen Sie das Dialogfeld Erweiterte Einstellungen: Linguistik, wählen Sie die Kategorieeingabeoption Nicht verwendete Extrahierungsergebnisse und bauen Sie die zusätzlichen Kategorien auf.
- Wenn Sie ein Projekt beginnen, wählen Sie ein Kategorienset aus einem TAP. Ziehen Sie dann nicht verwendete Konzepte oder Muster nach Bedarf in die Kategorien. Erweitern Sie dann die bestehenden Kategorien, die Sie eben bearbeitet haben (Kategorien > Kategorien erweitern), um weitere Deskriptoren zu erhalten, die mit den vorhandenen Kategoriedeskriptoren in Bezug stehen.
- Bauen Sie Kategorien automatisch über die erweiterten linguistischen Einstellungen (Kategorien > Kategorien aufbauen) auf. Verfeinern Sie dann die Kategorien manuell durch Löschen von Deskriptoren, Löschen von Kategorien oder Zusammenführen ähnlicher Kategorien, bis Sie mit den resultierenden Kategorien zufrieden sind. Zusätzlich können Sie, wenn die Kategorien ursprünglich **ohne** die Option Mit Platzhaltern, wenn möglich, verallgemeinern erzeugt wurden, auch versuchen, die Kategorien automatisch über “Kategorien erweitern” mit der Option Verallgemeinern zu vereinfachen.
- Importieren Sie eine vordefinierte Kategoriedatei mit sehr beschreibenden Kategoriennamen und/oder Anmerkungen. Zusätzlich können Sie, wenn der Import ursprünglich durchgeführt wurde, **ohne** dass die Option zum Import oder Erzeugen von Deskriptoren aus Kategoriennamen gewählt wurde, später das Dialogfeld “Kategorien erweitern” verwenden und die Option Leere Kategorien mit aus dem Kategoriennamen erzeugten Deskriptoren erweitern auswählen. Erweitern Sie dann diese Kategorien ein zweites Mal, aber verwenden Sie dieses Mal die Gruppierverfahren.
- Erstellen Sie manuell ein erstes Set an Kategorien, indem Sie Konzepte oder Konzeptmuster nach Häufigkeit sortieren und dann die interessantesten in den Kategoriebereich ziehen. Wenn Sie dieses erste Set an Kategorien haben, verwenden Sie die Funktion “Erweitern” (Kategorien > Kategorien erweitern), um alle ausgewählten Kategorien zu erweitern und zu verfeinern, so dass sie andere zugehörige Deskriptoren einschließen und so mehr Datensätzen entsprechen.

Es wird empfohlen, dass Sie nach Anwenden dieser Verfahren die resultierenden Kategorien prüfen und manuelle Verfahren verwenden, um kleinere Anpassungen vorzunehmen, etwaige Fehlklassifizierungen zu beheben oder Datensätze oder Wörter hinzuzufügen, die nicht erfasst wurden. Da die Verwendung verschiedener Verfahren auch zu redundanten Kategorien führen kann, können Sie zusätzlich Kategorien zusammenführen bzw. löschen. [Für weitere Informationen siehe Thema Bearbeiten und Verfeinern von Kategorien auf S. 159.](#)

Tipps zur Erstellung von Kategorien

Um Sie dabei zu unterstützen, bessere Kategorien zu erstellen, finden Sie hier einige Tipps, die Ihnen helfen, Entscheidungen zu Ihrem Ansatz zu treffen.

Tipps zum Verhältnis zwischen Kategorien und Antworten

Bei der Erstellung von Codes für eine geschlossene Frage, wie etwa “*Wann haben Sie das letzte Mal unseren Outlet-Store besucht?*”, sollten die Kategorien für die Antworten sich gegenseitig ausschließen und umfassend sein. Dieses Prinzip gilt aus mindestens zwei Gründen nicht notwendigerweise für eine qualitative Textanalyse:

- Zunächst gilt eine allgemeine Faustregel, die besagt: Je länger Datensatz, desto eindeutiger sind die darin ausgedrückten Ideen und Meinungen. Dadurch wird die Wahrscheinlichkeit, dass ein Datensatz mehreren Kategorien zugewiesen werden kann, stark erhöht.
- Zum anderen gibt es häufig verschiedene Möglichkeiten zur Gruppierung und Interpretation von Datensätzen, die nicht logisch getrennt sind. Bei einer Umfrage mit offenen Fragen zur politischen Haltung des Befragten können wir Kategorien erstellen wie *rechts* und *links* oder *konservativ* und *sozialdemokratisch* sowie speziellere Kategorien wie *sozialliberal*, *fiskalkonservativ* usw. Diese Kategorien müssen sich nicht gegenseitig ausschließen und brauchen nicht erschöpfend zu sein.

Tipps zur Anzahl der zu erstellenden Kategorien

Außer bei extrem einfachen offenen Fragen ist es niemals intuitiv klar, wieviele Kategorien erstellt werden sollten. Die Anzahl der Kategorien ist kein Anlass zur Sorge. Die Kategorieerstellung sollte sich direkt aus den Daten ergeben — Wenn Sie etwas Interessantes hinsichtlich der Ziele dieser Umfrage sehen, können Sie eine Kategorie erstellen, die diese Haltungen und Ideen repräsentiert.

- **Kategoriehäufigkeit.** Damit eine Kategorie sinnvoll eingesetzt werden kann, muss sie eine Mindestanzahl an Datensätzen enthalten. Ein oder zwei Datensätze können sehr interessante Aspekte enthalten, doch wenn es sich dabei um ein oder zwei von 1.000 Datensätzen handelt, sind die darin enthaltenen Informationen mit großer Wahrscheinlichkeit nicht häufig genug in der Grundgesamtheit vorhanden, um von praktischem Nutzen zu sein.
- **Komplexität.** Je mehr Kategorien Sie erstellen, desto mehr Informationen müssen Sie nach Abschluss der Analyse überprüfen und zusammenfassen. Sehr viele Kategorien vergrößern zwar die Komplexität, führen jedoch nicht unbedingt zu zusätzlichen nützlichen Details.

Leider gibt es keine Regeln zur Ermittlung, ab wann zu viele Kategorien vorliegen oder wie viele Datensätze mindestens pro Kategorie vorhanden sein sollen. Sie müssen diese Festlegungen je nach den Anforderungen der jeweiligen Situation treffen.

Wir können jedoch einige Ratschläge für den Anfang geben. Die Anzahl der Kategorien sollte zwar nicht übermäßig hoch sein, in den frühen Phasen einer Analyse ist es jedoch besser, eher zu viele als zu wenige Kategorien zu verwenden. Es ist einfacher, Kategorien, die sich relativ ähnlich sind, zusammenzufassen, als Fälle in neue Kategorien abzuspalten. Daher ist eine Strategie, die von relativ vielen Kategorien ausgeht und zu einer Verringerung ihrer Zahl führt, normalerweise die beste Vorgehensweise. Angesichts des iterativen Charakters des Text-Minings und der Leichtigkeit, mit der es mithilfe dieses Softwareprogramms durchgeführt werden kann, stellt eine hohe Kategorienzahl für den Anfang kein Problem dar.

Auswahl der besten Deskriptoren

Die folgenden Informationen enthalten einige Richtlinien zur Auswahl oder zum Erstellen der besten Deskriptoren (Konzepte, Typen, TLA-Muster und Kategorieregeln) für Ihre Kategorien. Deskriptoren sind die Bausteine von Kategorien. Wenn Text in einem Datensatz ganz oder teilweise mit einem Deskriptor übereinstimmt, wird das Datensatz mit der Kategorie abgeglichen.

Ein Deskriptor wird nur dann mit Datensätzen abgeglichen, wenn er ein extrahiertes Konzept oder Muster enthält oder ihm entspricht. Verwenden Sie daher Konzepte, Typen, Muster und Kategorieregeln wie beschrieben in den folgenden Absätzen.

Da Konzepte nicht nur sich selbst, sondern auch eine Reihe zugrundeliegender Fachausdrücke darstellen, von Plural- und Singularformen über Synonyme bis hin zu Rechtschreibvariationen, sollte nur das Konzept selbst als Deskriptor oder Teil eines Deskriptors verwendet werden. Um mehr über die zugrunde liegenden Fachausdrücke für ein bestimmtes Konzept zu erfahren, klicken Sie auf den Konzeptnamen in den Extrahierungsergebnissen. Wenn Sie mit der Maus über den Konzeptnamen fahren, erscheint eine QuickInfo, in der alle bei der letzten Extrahierung in Ihrem Text gefundenen, zugrunde liegenden Fachausdrücke angezeigt werden. Nicht alle Konzepte haben zugrunde liegende Fachausdrücke. Falls beispielsweise `Auto` und `Fahrzeug` als Synonyme gelten, aber `Auto` als Konzept und `Fahrzeug` als zugrunde liegender Fachausdruck extrahiert wurde, sollten Sie nur `Auto` in einem Deskriptor verwenden, da dieser automatisch mit Datensätzen abgeglichen wird, die `Fahrzeug` enthalten.

Konzepte und Typen als Deskriptoren

Verwenden Sie ein Konzept als Deskriptor, wenn Sie alle Datensätze mit diesem Konzept (oder seine zugrunde liegenden Fachausdrücke) finden möchten. In diesem Fall ist es nicht notwendig, eine komplexere Kategorieregel zu verwenden, da der exakte Konzeptname ausreicht. Denken Sie daran, dass sich Konzepte bei der Verwendung von Ressourcen, die Meinungen extrahieren, manchmal während der Extrahierung von TLA-Mustern verändern können, um den wahren Inhalt des Satzes zu erfassen (siehe dazu das Beispiel im nächsten Abschnitt über TLA).

Beispielsweise könnte die Antwort in einer Umfrage über das Lieblingsobst der Teilnehmer, etwa *“Äpfel und Ananas sind die Besten”*, zur Extrahierung von `Apfel` und `Ananas` führen. Indem Sie das Konzept `Apfel` Ihrer Kategorie als Deskriptor hinzufügen, werden alle Antworten mit dem Konzept `Apfel` (oder seinen zugrunde liegenden Fachausdrücken) mit dieser Kategorie abgeglichen.

Wenn Sie jedoch einfach nur wissen möchten, welche Antworten in irgendeiner Form das Wort *Apfel* enthalten, können Sie eine Kategorieregel mit `* Apfel *` erstellen, wodurch alle Antworten erfasst werden, die Konzepte wie `Apfel`, `Apfelsauce` oder `französischer Apfelkuchen` enthalten.

Sie können auch alle Datensätze mit Konzepten mit der gleichen Schreibweise erfassen, indem Sie einen Typ direkt als Deskriptor verwenden, z. B. `<Obst>`. Hinweis: Sternchen (*) können nicht mit Typen verwendet werden.

Für weitere Informationen siehe [Thema Extrahierte Ergebnisse: Konzepte, Typen und Muster in Kapitel 5 auf S. 80](#).

Textlinkanalyse- (TLA-) Muster als Deskriptoren

Verwenden Sie ein TLA-Musterergebnis als Deskriptor, wenn Sie feinere, nuancierte Ideen erfassen möchten. Wenn während einer TLA-Extrahierung Text analysiert wird, wird der Text satz- oder absatzweise verarbeitet, anstatt den Text als Ganzes zu betrachten (das Datensatz). Indem alle Teile eines einzelnen Satzes zusammen betrachtet werden, kann die TLA beispielsweise Meinungen, Beziehungen zwischen zwei Elementen oder eine Negation identifizieren und so den wahren Sinn des Satzes erfassen. Sie können Konzeptmuster oder Typmuster als Deskriptoren verwenden.

Beispielsweise könnten in dem Satz *“Das Zimmer war nicht sonderlich sauber.”* die folgenden Konzepte extrahiert werden: `zimmer` und `sauber`. Falls jedoch die TLA-Extrahierung in den Extrahierungseinstellungen deaktiviert wurde, könnte TLA erkennen, dass `sauber` negativ verwendet wurde und eigentlich `nicht sauber` entspricht, was als Synonym zu dem Konzept `dreckig` gilt. Hier zeigt sich, dass die Verwendung des Konzepts `sauber` als eigenständiger Deskriptor mit diesem Text übereinstimmen würde, jedoch auch andere oder Datensätze mit dem Inhalt Sauberkeit erfasst werden könnten. Aus diesem Grund ist es besser, das TLA-Konzeptmuster mit `dreckig` als Ausgabe konzept zu verwenden, da dieses mit dem Text übereinstimmen würde und wahrscheinlich ein passenderer Deskriptor wäre.

Kategoriegeschäftsregeln als Deskriptoren

Kategorieregeln sind Anweisungen, mit denen Datensätze auf Basis eines logischen Ausdrucks mithilfe von extrahierten Konzepten, Typen und Mustern sowie von Boole'schen Operatoren automatisch einer Kategorie zugewiesen werden. Sie könnten beispielsweise einen Ausdruck schreiben, der bedeutet *Schließe alle Datensätze, die das extrahierte Konzept Botschaft enthalten, nicht jedoch Argentinien, in diese Kategorie ein.*

Sie können Kategorieregeln in Ihren Kategorien als Deskriptoren erstellen und verwenden, um mit den Boole'schen Operatoren `&`, `|` und `!()` unterschiedliche Ideen auszudrücken. Nähere Informationen über die Syntax dieser Regeln und wie sie geschrieben und bearbeitet wird finden Sie unter [Verwenden von Kategorieregeln auf S. 148](#)

- Verwenden Sie eine Kategorieregel mit dem Boole'schen Operator `&` (UND), um Datensätze zu finden, in denen zwei oder mehrere Konzepte vorkommen. Die zwei oder mehr durch `&`-Operatoren verbundenen Konzepte müssen nicht im selben Satz oder Satzteil auftreten, sondern können an beliebiger Stelle im selben Datensatz auftreten, um als Übereinstimmung mit einer Kategorie zu gelten. Wenn Sie beispielsweise die Kategorieregel `Essen & günstig` als Deskriptor erstellen, würde dadurch der Datensatz *“Das Essen war ziemlich teuer, aber das Zimmer war günstig”* als Übereinstimmung gelten, obwohl das Nomen `Essen` nicht als `günstig` bezeichnet wurde, da der Text sowohl `Essen` als auch `günstig` enthielt.
- Verwenden Sie eine Kategorieregel mit dem Boole'schen Operator `!()` (NICHT) als Deskriptor, um Datensätze zu finden, in denen manche Ausdrücke vorkommen, andere jedoch nicht. So können Sie vermeiden, dass Informationen gruppiert werden, deren Wörter zwar ohne Zusammenhang erscheinen, nicht jedoch deren Kontext. Wenn Sie beispielsweise die Kategorieregel `<Unternehmen> & !(ibm)` als Deskriptor erstellen, würde der Text *SPSS Inc. ist ein 1967 gegründetes Unternehmen.* als Übereinstimmung gefunden werden, der Text *Das Softwareunternehmen wurde von IBM aufgekauft.* jedoch nicht.

- Verwenden Sie eine Kategorieregel mit dem Boole'schen Operator | (ODER) als Deskriptor, um Datensätze mit einem oder mehreren Konzepten oder Typen zu finden. Wenn Sie beispielsweise die Kategorieregel `(Personal|Belegschaft|Team|Kollegen) & schlecht` als Deskriptor erstellen, würden alle Datensätze als Übereinstimmung gelten, in denen mindestens eines dieser Nomen mit dem Konzept `schlecht` gefunden wurde.
- Verwenden Sie Typen in Kategorieregeln, um diese allgemeiner und möglicherweise anwendbarer zu gestalten. Beispielsweise möchten Sie bei der Arbeit mit Hoteldaten erfahren, was Ihre Kunden von dem Hotelpersonal halten. Verwandte Begriffe enthalten unter Umständen Wörter wie `Rezeptionist`, `Kellner`, `Kellnerin`, `Hotelrezeption`, `Empfang` usw. In diesem Fall könnten Sie einen neuen Typ namens `<HotelStaff>` erstellen und diesem Typ alle oben erwähnten Begriffe hinzufügen. Es ist zwar möglich, eine Kategorieregel für jede Personalart zu erstellen, etwa `[* Kellnerin * & nett]`, `[* Empfang * & freundlich]`, `[* Rezeptionist * & entgegenkommend]`, Sie können jedoch auch eine einzelne, allgemeinere Kategorieregel mit dem Typ `<HotelStaff>` erstellen, um alle Antworten zu erfassen, die sich positiv über das Hotelpersonal äußern, und zwar in der Form `[<Hotelpersonal> & <Positive>]`.

Anmerkung: Sie können sowohl `+` als auch `&` in Kategorieregeln verwenden, wenn Sie TLA-Muster in diese Regeln einschließen. Für weitere Informationen siehe [Thema Verwenden von TLA-Mustern in Kategorieregeln](#) auf S. 151.

Beispiele für unterschiedliche Übereinstimmungen bei Konzepten, TLA- oder Kategorieregeln

Das folgende Beispiel zeigt, wie sich die Verwendung eines Konzepts als Deskriptor, einer Kategorieregel als Deskriptor oder eines TLA-Musters als Deskriptor auf die Kategorisierung von Datensätzen auswirkt. Nehmen wir an, Sie haben folgende fünf Datensätze.

- A: *“Hervorragendes Restaurantpersonal, köstliches Essen und die Zimmer bequem und sauber.”*
- B: *“Das Restaurantpersonal war schrecklich, aber die Zimmer waren sauber.”*
- C: *“Die Zimmer waren bequem und sauber.”*
- D: *“Mein Zimmer war nicht sonderlich sauber.”*
- E: *“Saubere.”*

Da die Datensätze das Wort *sauber* enthalten und Sie diese Informationen erfassen möchten, könnten Sie einen der in der folgenden Tabelle gezeigten Deskriptoren erstellen. Auf der Basis des wahren Inhalts, den Sie erfassen möchten, können Sie sehen, wie die Verwendung einer Deskriptorart anstatt einer anderen zu unterschiedlichen Ergebnissen führen kann.

Tabelle 6-1
Übereinstimmungen zwischen Beispieldatensätzen und Deskriptoren

Deskriptor	A	B	C	D	E	Erklärung
sauber	Übereinstimmung	Übereinstimmung	Übereinstimmung	Übereinstimmung	Übereinstimmung	Deskriptor ist ein extrahiertes Konzept. Jeder Datensatz enthielt das Konzept <i>sauber</i> , selbst Datensatz D, denn ohne TLA ist nicht automatisch bekannt, dass " <i>nicht sauber</i> " laut den TLA-Regeln dreckig bedeutet.
sauber + .	-	-	-	-	Übereinstimmung	Deskriptor ist ein TLA-Muster, das selbst für <i>sauber</i> steht. Nur Übereinstimmung mit dem Datensatz, in dem <i>sauber</i> während der TLA-Extrahierung ohne zugehöriges Konzept extrahiert wurde.
[sauber]	Übereinstimmung	Übereinstimmung	Übereinstimmung	-	Übereinstimmung	Deskriptor ist eine Kategorieregel, die nach einer TLA-Regel sucht, die das Wort <i>sauber</i> allein stehend oder in Verbindung mit anderen Wörtern enthält. Übereinstimmung mit dem Datensatz, in dem eine TLA-Ausgabe mit <i>sauber</i> gefunden wurde, unabhängig davon, ob <i>sauber</i> mit einem anderen Konzept wie <i>Zimmer</i> und in einer beliebigen Slot-Position verknüpft wurde.

Erläuterung von Kategorien

Kategorien bezeichnen eine Gruppe von eng verwandten Konzepten, Meinungen oder Haltungen. Um nützlich zu sein, sollte sich eine Kategorie auch leicht durch einen kurzen Ausdruck oder eine kurze Bezeichnung beschreiben lassen, der bzw. die ihre grundlegende Bedeutung erfasst.

Wenn Sie beispielsweise Umfrageantworten von Verbrauchern zu einem neuen Waschmittel analysieren, können Sie eine Kategorie mit dem Label *Duft* erstellen, das alle Antworten enthält, die den Geruch des Produkts beschreiben. Eine solche Kategorie würde jedoch nicht zwischen den Personen unterscheiden, die den Duft angenehm fanden, und den Personen, denen der Duft unangenehm war. Da IBM® SPSS® Text Analytics for Surveys mithilfe der geeigneten Ressourcen Meinungen extrahieren kann, könnten Sie dann zwei andere Kategorien definieren, um Befragte zu identifizieren, die *den Duft mochten*, und Befragte, die *den Duft nicht mochten*.

Sie können Ihre Kategorien im Bereich "Kategorien" im oberen linken Bereich des Textanalysefensters erstellen und mit ihnen arbeiten. Die einzelnen Kategorien sind durch ein oder mehrere Deskriptoren definiert. **Deskriptoren** sind Konzepte, Typen und Muster sowie Kategorieregeln, die zum Definieren einer Kategorie verwendet wurden.

Wenn Sie die Deskriptoren sehen möchten, aus denen eine bestimmte Kategorie besteht, können Sie die Kategorie im Baum erweitern. Im Baum werden Symbole angezeigt, über die Sie die einzelnen Elemente problemlos identifizieren können. Nur die erste Ebene definiert die Kategorie. Wenn Sie die Definitionen zusätzlich erweitern, sehen Sie Beispiele dafür, was in den Daten gefunden wurde.

Wenn Sie Kategorien automatisch mithilfe von Kategorieerstellungsverfahren (z. B. semantische Netze oder Konzeptbeziehungen) erstellen, verwenden die Verfahren Konzepte und Typen als Deskriptoren zum Erstellen der Kategorien. Sie können außerdem diese Muster oder Teile davon als Kategoriedeskriptoren hinzufügen. Und schließlich können Sie manuell Kategorieregeln erstellen, die als Deskriptoren in den Kategorien verwendet werden sollen. [Für weitere Informationen siehe Thema Verwenden von Kategorieregeln auf S. 148.](#)

Wenn Sie beispielsweise einer Kategoriedefinition einen Typ hinzufügen, sind alle diesem Typ zugewiesenen Konzepte automatisch enthalten, auch zu dem Zeitpunkt, da nur eine Handvoll dieser Konzepte in den Daten vorhanden ist. Dies hilft bei der erneuten Verwendung von Kategoriedefinitionen mit neuen Daten. [Für weitere Informationen siehe Thema Kopieren von Kategorien auf S. 168.](#) Sie können auch manuell Kategorieregeln erstellen, die als Deskriptoren in den Kategorien verwendet werden sollen. [Für weitere Informationen siehe Thema Verwenden von Kategorieregeln auf S. 148.](#)

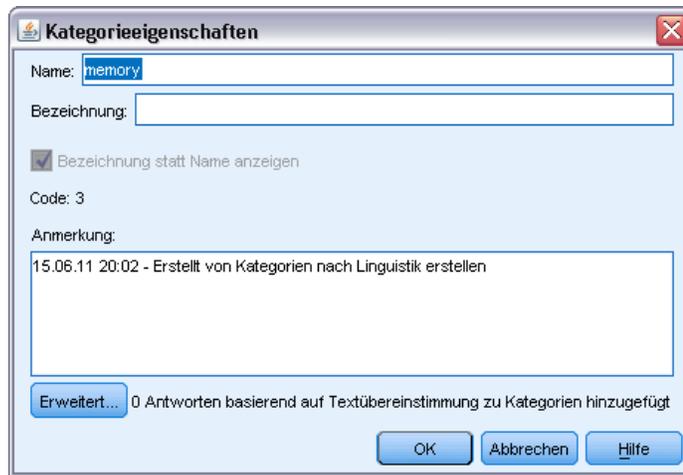
Tabelle 6-2
Symbole, die Elemente in Definitionen identifizieren

Symbol	Beschreibung
	Konzept.
	Typ, der erweitert werden kann, um die enthaltenen Konzepte anzuzeigen
	Konzeptmuster, das erweitert werden kann, um die spezifischen Konzepte in Mustern zu sehen
	Typmuster, das erweitert werden kann, um die Ebene der Konzeptmuster zu sehen
	Kategorieregeln in der Kategorie. Klicken Sie mit der rechten Maustaste auf den Regelnamen, um die Regel zu bearbeiten.

Kategorieeigenschaften

Neben Deskriptoren verfügen Kategorien zudem über Eigenschaften, die Sie bearbeiten können, um Kategorien umzubenennen oder eine Bezeichnung oder Anmerkung hinzuzufügen oder auf das Dialogfeld für Textübereinstimmung zuzugreifen.

Abbildung 6-5
Dialogfeld "Kategorieeigenschaften"



Folgende Eigenschaften sind vorhanden:

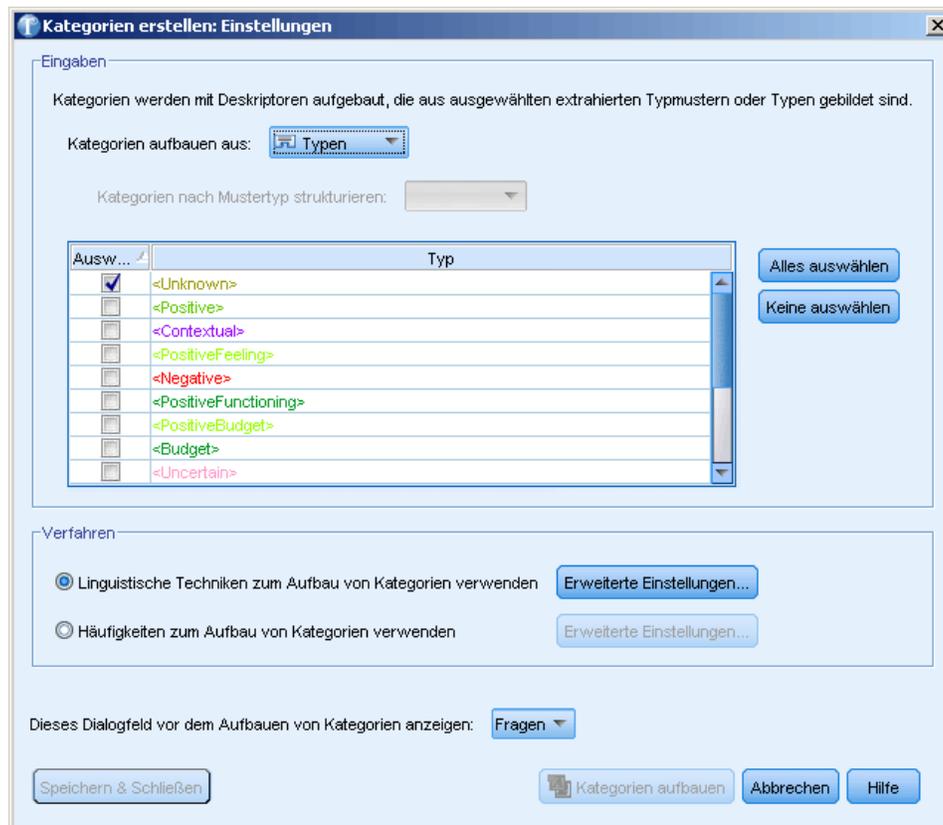
- **Name.** Dieser Name wird standardmäßig im Baum angezeigt. Wenn eine Kategorie mithilfe eines automatisierten Verfahrens erstellt wird, erhält sie automatisch einen Namen.
- **Bezeichnung.** Die Verwendung von Bezeichnungen ist nützlich beim Erstellen aussagekräftigerer Kategoriebeschreibungen zur Verwendung in anderen Produkten oder anderen Tabellen bzw. Grafiken. Wenn Sie die Option zur Anzeige der Bezeichnung auswählen, wird die Bezeichnung auf der Benutzeroberfläche zur Angabe der Kategorie verwendet. In IBM® SPSS® Statistics werden sie als Variablenlabels exportiert. In Microsoft Excel werden sie als gesonderte Zeile exportiert.
- **Code.** Die Codenummer entspricht dem Codewert für diese Kategorie. Sie können diesen Code im Coderahmenmanager bearbeiten. Im Coderahmenmanager können Sie den Namen, die Bezeichnung und den Code für jede Kategorie bearbeiten sowie ganze Coderahmen kopieren und einfügen. .
- **Anmerkung.** In diesem Feld können Sie eine kurze Beschreibung für die einzelnen Kategorien hinzufügen. Wenn eine Kategorie über das Dialogfeld "Kategorien aufbauen" erstellt wird, wird dieser Anmerkung automatisch ein Hinweis hinzugefügt. Sie können auch direkt über den Datenbereich Text in eine Anmerkung einfügen, indem Sie den Text und anschließend aus den Menüs die Optionen Kategorien > Zu Anmerkung hinzufügen auswählen.
- **Erweiterte Textübereinstimmung.** Mithilfe der Schaltfläche "Erweitert" können Sie einer Kategoriedefinition Wörter oder zusammengesetzte Ausdrücke hinzufügen. Dies wird häufig verwendet, um eine falsche Extrahierung zu überschreiben. [Für weitere Informationen siehe Thema Textübereinstimmungen in Kategorien auf S. 167.](#)

Erstellen von Kategorien

Sie können zwar über Kategorien aus einem Textanalysepaket verfügen, aber Sie können auch mit einigen linguistischen und häufigkeitsbasierten Verfahren automatisch Kategorien erstellen. Über das Dialogfeld Kategorien aufbauen: Einstellungen können Sie die automatisierten

linguistischen und häufigkeitsbasierten Verfahren anwenden, um Kategorien aus Konzepten oder Konzeptmustern zu erstellen.

Abbildung 6-6
Dialogfeld "Kategorien aufbauen"



In der Regel können Kategorien aus unterschiedlichen Deskriptoren (Typen, Konzepte, TLA-Muster, Kategorieregeln) aufgebaut werden. Wenn Sie Kategorien mit automatisierten Verfahren zum Aufbauen von Kategorien erstellen, wird jede erstellte Kategorie nach einem Konzept oder Konzeptmuster (je nach Auswahl) benannt und enthält eine Reihe an Deskriptoren. Diese Deskriptoren können aus Kategorieregeln oder Konzepten bestehen und enthalten alle zugehörigen Konzepte, die von den Verfahren erkannt werden.

Nach dem Aufbau der Kategorien können Sie viel über die Kategorien erfahren, indem Sie sie im Fensterbereich "Kategorien" überprüfen oder in den Diagrammen und Tabellen untersuchen. Anschließend können Sie mithilfe von manuellen Verfahren kleinere Anpassungen vornehmen, etwaige Fehlklassifizierungen beheben oder Datensätze oder Wörter hinzufügen, die nicht erfasst wurden. Nachdem Sie ein Verfahren angewendet haben, stehen die Konzepte, Typen und Muster, die in eine Kategorie gruppiert wurden, weiterhin für andere Verfahren zur Verfügung. Da die Verwendung verschiedener Verfahren auch zu redundanten oder ungeeigneten Kategorien führen kann, können Sie außerdem Kategorien zusammenführen bzw. löschen. [Für weitere Informationen siehe Thema Bearbeiten und Verfeinern von Kategorien auf S. 159.](#)

Wichtig: In früheren Versionen wurden Kookkurenz- und Synonymregeln von eckigen Klammern umgeben. In dieser neuen Version zeigen nun eckige Klammern ein Musterergebnis für die an. Stattdessen stehen Kookkurenz- und Synonymregeln in runden Klammern, z. B. (Lautsprechersysteme|Lautsprecher).

So erstellen Sie Kategorien

- ▶ Wählen Sie in den Menüs die Optionsfolge Kategorien > Kategorien aufbauen. Wenn Sie nicht das Kontrollkästchen markiert haben, dass diese Aufforderung nie angezeigt werden soll, wird ein Dialogfeld geöffnet.

Abbildung 6-7

Aufforderung vor dem Aufbau



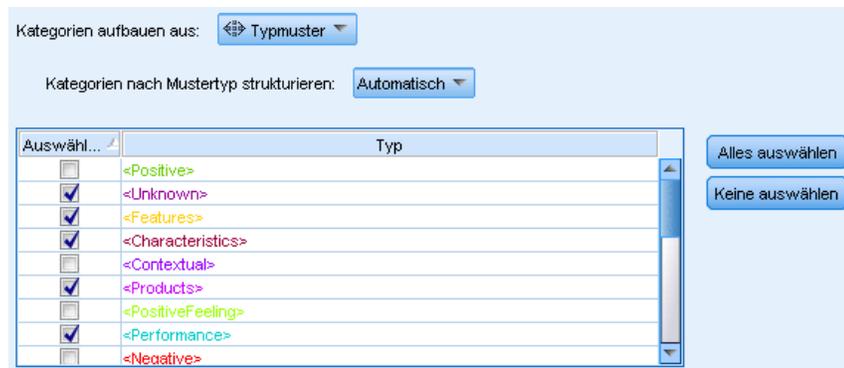
- ▶ Wählen Sie aus, ob Sie die Kategorie jetzt aufbauen oder zuerst die Einstellungen bearbeiten möchten.
 - Klicken Sie auf Jetzt aufbauen, um mit den aktuellen Einstellungen den Aufbau einer Kategorie zu beginnen. Die standardmäßig ausgewählten Einstellungen sind für den Beginn der Kategorisierung oft ausreichend. Der Kategorieaufbauprozess wird gestartet und ein Dialogfeld über den Fortschritt wird angezeigt.
 - Klicken Sie auf Bearbeiten, um die Aufbaueinstellungen zu überprüfen und zu ändern.

Eingaben

Die Kategorien werden aus Deskriptoren aufgebaut, die aus Typmuster oder Typen abgeleitet werden. Standardmäßig sind Typmuster in der Dropdown-Liste ausgewählt. In der Tabelle können Sie die einzelnen Typen oder Muster auswählen, die in den Kategorieaufbauprozess aufgenommen werden.

Typmuster. Wenn Sie Typmuster auswählen, werden die Kategorien aus Mustern anstelle von Typen und aus einzelnen Konzepten aufgebaut. So werden Datensätze kategorisiert, die ein Konzeptmuster enthalten, das zum gewählten Typmuster gehört. Wenn Sie in der Tabelle also das Typmuster <Budget> und <Positiv> auswählen, können Kategorien wie *Kosten & &Positiv* oder *Sätze & Hervorragend* erzeugt werden. Die Tabelle zeigt nur eine Zeile für jede Typkombination wie <Standort> + <Positiv> und <Positiv> + <Standort> und ihre Reihenfolge ist unwichtig dafür, wie die Kategorien erzeugt werden.

Abbildung 6-8
Dialogfeld "Kategorien aufbauen" mit den verfügbaren Typmustern



Wenn Typmuster als Eingabe zum automatisierten Kategorieaufbau verwendet werden, können die Techniken manchmal mehrere Möglichkeiten zur Bildung der Kategoriestructur identifizieren. Technisch gesehen gibt es nicht nur eine richtige Lösung zum Erzeugen von Kategorien, Sie finden jedoch eventuell eine Struktur besser für Ihre Analyse geeignet als andere. Um in diesem Fall die Ausgabe besser anpassen zu können, können Sie einen Typ als bevorzugt kennzeichnen. Wählen Sie diesen Typ im Feld Kategorien nach Mustertyp strukturieren: aus und die Tabelle wird aktualisiert und zeigt nur die jeweiligen Muster mit dem ausgewählten Typ an. In den meisten Fällen ist <Unbekannt> vorausgewählt. Das führt dazu, dass alle Muster, die den Typ <Unbekannt> enthalten, ausgewählt werden. Die Tabelle zeigt die Typen in absteigender Reihenfolge, beginnend mit der größten Anzahl an Datensätzen.

Typen. Wenn Sie Typen auswählen, werden die Kategorien aus Konzepten aufgebaut, die zu den gewählten Typen gehören. Wenn Sie also den Typ <Budget> in der Tabelle auswählen, können Kategorien wie *Kosten* oder *Preis* erzeugt werden, da *Kosten* und *Preis* Konzepte sind, die dem Typ <Budget> zugewiesen wurden.

Standardmäßig werden nur die Typen ausgewählt, die die meisten Datensätze erfassen. Dank dieser Vorauswahl können Sie sich schnell auf die interessantesten Typen konzentrieren und den Aufbau uninteressanter Kategorien vermeiden. Die Tabelle zeigt die Typen in absteigender Reihenfolge, beginnend mit der größten Anzahl an Datensätzen. Typen aus der Bibliothek *Opinions* werden standardmäßig in der Typtabelle abgewählt.

Die sich ergebenden Kategorien hängen von der ausgewählten Eingabe ab. Wenn Sie Typen als Eingabe verwenden, können Sie die eindeutig verwandten Konzepte leichter erkennen. Wenn Sie beispielsweise Kategorien mit Typen als Eingaben erstellen, könnten Sie eine Kategorie *Obst* erhalten, die die Konzepte *Apfel*, *Birne*, *Zitrusfrüchte*, *Orange* usw. enthält. Wenn Sie stattdessen Typmuster als Eingabe wählen und beispielsweise das Muster <Unbekannt> + <Positiv> auswählen, erhalten Sie möglicherweise die Kategorie *Obst + <Positiv>* mit ein oder zwei Arten von Obst, wie beispielsweise *Obst + lecker* und *Apfel + gut*. Bei diesem zweiten Ergebnis werden nur 2 Konzeptmuster angezeigt, da die anderen Vorkommnisse von Obst nicht unbedingt positiv qualifiziert sind. Und während dies möglicherweise für Ihre aktuellen Textdaten brauchbar ist, kann es für Longitudinalstudien, in denen verschiedene Mengen von Dokumenten verwendet werden, sinnvoll sein, von Hand weitere Deskriptoren hinzuzufügen, wie beispielsweise *Zitrusfrucht + positiv*, oder aber Typen zu verwenden. Wenn Sie ausschließlich Typen als Eingabe verwenden, können Sie alle möglichen Arten von Obst finden.

Abbildung 6-9
Dialogfeld "Kategorien aufbauen" mit den verfügbaren Typen



Verfahren

Da jedes Daten-Set anders ist, kann sich die Anzahl der Methoden und die Reihenfolge, in der sie angewendet werden, im Laufe der Zeit ändern. Da Ihre Text-Mining-Ziele von Daten-Set zu Daten-Set unterschiedlich sein können, müssen Sie möglicherweise mit den verschiedenen Verfahren experimentieren, um zu sehen, welches die besten Ergebnisse für die jeweiligen Textdaten bringt.

Sie müssen nicht besonders gut mit diesen Einstellungen vertraut sein, um sie verwenden zu können. Standardmäßig sind die gängigsten Einstellungen bereits ausgewählt. Daher können Sie die Dialogfelder für die erweiterten Einstellungen überspringen und gleich mit dem Aufbau der Kategorien beginnen. Wenn Sie hier Änderungen vornehmen, müssen Sie diese nicht mit jedem Öffnen des Dialogfelds erneut vornehmen, da immer die neuesten Einstellungen beibehalten werden.

Wählen Sie entweder die linguistischen oder häufigkeitsbasierten Verfahren und klicken Sie auf die Schaltfläche "Erweiterte Einstellungen", um die Einstellungen für die gewählten Verfahren anzuzeigen. Keines der automatischen Verfahren ergibt eine perfekte Kategorisierung Ihrer Daten, weshalb wir empfehlen, ein oder mehrere automatische Verfahren anzuwenden, die gut mit Ihren Daten funktionieren. Ein gleichzeitiger Einsatz von linguistischen und häufigkeitsbasierten Verfahren ist beim Aufbau nicht möglich.

- **Erweiterte linguistische Verfahren.** Weitere Informationen finden Sie unter auf S. 116.
- **Erweiterte häufigkeitsbasierte Verfahren.** Weitere Informationen finden Sie unter auf S. 127.

Erweiterte linguistische Einstellungen

Beim Aufbau von Kategorien haben Sie die Auswahl zwischen einigen erweiterten linguistischen Verfahren für den Kategorieaufbau, darunter *Konzeptwurzelableitung*, *Konzepteinbeziehung*, *semantische Netze* (nur für englischen Text) und *Kookkurrenzregeln*. Zum Erstellen von Kategorien können diese Verfahren einzeln oder in Kombination verwendet werden.

Beachten Sie, dass jedes Daten-Set anders ist und sich die Anzahl der Methoden und die Reihenfolge, in der sie angewendet werden, daher im Laufe der Zeit ändern kann. Da Ihre Text-Mining-Ziele von Daten-Set zu Daten-Set unterschiedlich sein können, müssen Sie möglicherweise mit den verschiedenen Verfahren experimentieren, um zu sehen, welches die

besten Ergebnisse für die jeweiligen Textdaten bringt. Keines der automatischen Verfahren ergibt eine perfekte Kategorisierung Ihrer Daten, weshalb wir empfehlen, ein oder mehrere automatische Verfahren anzuwenden, die gut mit Ihren Daten funktionieren.

Abbildung 6-10

Erweiterte Einstellungen: Dialogfeld "Linguistik" für die Erstellung von Kategorien

Erweiterte Einstellungen: Linguistik

Eingabe und Ausgabe

Kategorieeingabe: Nicht verwendete Extrahierungsergebnisse Alle Extrahierungsergebnisse

Kategorieausgabe:

Hierarchisch mit Unterkategorien Maximal erstellte Ebenen: 5

Kategorien glätten (nur Einzelebene)

Gruppierungsmethoden

Konzeptwurzelableitung Konzepteinbeziehung

Semantisches Netz Kookkurrenz

Minimale Anzahl an Datensätzen: 2

Maximaler Suchabstand: 3

Paarung spezifischer Konzepte verhindern

Mit Platzhaltern, wenn möglich, verallgemeinern (z. B. apfel *)

Andere Optionen

Maximale Anzahl an erstellten Kategorien der obersten Ebene: 30

Mindestanzahl von Deskriptoren und/oder Unterkategorien pro Kategorie: 2

Deskriptoren können in mehr als einer Kategorie erscheinen

Doppelte Kategorienamen auflösen durch: Neue zusammenführen, falls vorhanden

Eingabe und Ausgabe

Kategorieeingabe. Auswahl, woraus die Kategorien aufgebaut werden:

- Nicht verwendete Extrahierungsergebnisse. Diese Option aktiviert Kategorien, die aus Extrahierungsergebnissen erstellt werden, die nicht in vorhandenen Kategorien verwendet werden. So wird die Tendenz für Datensätze minimiert, mehrere Kategorien abzugleichen und die Anzahl der erzeugten Kategorien zu begrenzen.
- Alle Extrahierungsergebnisse. Diese Option aktiviert unter Verwendung der Extrahierungsergebnisse aufzubauende Kategorien. Dies ist am sinnvollsten, wenn noch keine oder nur sehr wenige Kategorien vorhanden sind.

Kategorieausgabe. Auswahl der allgemeinen Struktur für die aufzubauenden Kategorien:

- Hierarchisch mit Unterkategorien. Diese Option aktiviert das Erstellen von Unterkategorien und Unter-Unterkategorien. Sie können die Tiefe Ihrer Kategorien einstellen, indem Sie die maximale Anzahl an zu erstellenden Ebenen (Feld *Maximal erstellte Ebenen*) auswählen. Wenn Sie 3 auswählen, können Kategorien Unterkategorien enthalten, welche wiederum ebenfalls Unterkategorien enthalten können.
- Flache Kategorien (nur eine Ebene). Mit dieser Option wird nur eine Kategorieebene aufgebaut, d. h., es werden keine Unterkategorien erzeugt.

Gruppierverfahren

Die einzelnen verfügbaren Verfahren sind für bestimmte Datentypen und Situationen jeweils sehr gut geeignet, doch ist es oftmals nützlich, bei einer Analyse mehrere Verfahren miteinander zu verbinden, um die Datensätze vollständig zu erfassen. Sie können ein Konzept in mehreren Kategorien erkennen oder redundante Kategorien vorfinden.

Konzeptwurzelaufleitung. Mit diesem Verfahren werden Kategorien erstellt, indem ausgehend von einem Konzept andere verwandte Konzepte ermittelt werden (durch Analyse, ob bestimmte Konzeptkomponenten morphologisch verwandt sind oder gemeinsame Wurzeln haben). Dieses Verfahren ist sehr nützlich bei der Identifizierung von bedeutungsgleichen Konzepten aus zusammengesetzten Wörtern, da die Konzepte in jeder generierten Kategorie die gleiche oder ähnliche Bedeutung haben. Das Verfahren funktioniert mit Daten unterschiedlicher Länge und erzeugt eine geringere Anzahl an kompakten Kategorien. So wird beispielsweise das Konzept *Möglichkeiten zum Aufstieg mit den Konzepten Möglichkeit des Aufstiegs und Aufstiegsmöglichkeit* zu einer Kategorie zusammengefasst. [Für weitere Informationen siehe Thema Konzeptwurzelaufleitung auf S. 122.](#)

Semantisches Netz. Bei diesem Verfahren wird zunächst auf der Grundlage eines umfassenden Index von Wortbeziehungen jedes Konzept auf seine möglichen Bedeutungen untersucht. Anschließend werden Kategorien durch Gruppieren zusammenhängender Konzepte erstellt. Diese Technik empfiehlt sich, wenn die Konzepte dem semantischen Netz bekannt und nicht zu zweideutig sind. Sie ist weniger hilfreich, wenn der Text spezielle Terminologie oder Sprache enthält, die dem Netz unbekannt ist. Das Konzept *Granny Smith Apfel* würde zum Beispiel mit *Gala Apfel* und *Winesap Apfel* gruppiert, da es sich um Geschwister von *Granny Smith* handelt. In einem anderen Beispiel würde das Konzept *Tier* mit *Katze* und *Känguru* gruppiert, da dies Hyponyme von *Tier* sind. Dieses Verfahren ist in dieser Version nur für englischen Text verfügbar. [Für weitere Informationen siehe Thema Semantische Netze auf S. 124.](#)

Konzepteinbeziehung. Diese Technik erstellt Kategorien durch die Gruppierung von Multiterm-Konzepten (zusammengesetzte Wörter) basierend darauf, ob sie Wörter enthalten, die Unter- oder Übermengen eines Worts in dem anderen sind. So wird beispielsweise `sitz` mit `kindersitz`, `sitzheizung` und `kindersitzgurt` zu einer Gruppe zusammengefasst. [Für weitere Informationen siehe Thema Konzepteinbeziehung auf S. 123.](#)

Kookkurrenz. Diese Technik erstellt Kategorien aus Kookkurrenz im Text. Dahinter steht folgende Idee: Wenn Konzepte oder Konzeptmuster häufig in Dokumenten bzw. Datensätzen gefunden werden, ist diese Kookkurrenz Ausdruck einer zugrunde liegenden Beziehung, die wahrscheinlich in Ihren Kategoriedefinitionen von Nutzen ist. Wenn Wörter eine signifikante Kookkurrenz aufweisen, wird eine Kookkurrenzregel erstellt, die als Kategoriedeskriptor für eine neue Unterkategorie verwendet werden kann. Wenn beispielsweise viele Datensätze die Wörter `preis` und `verfügbarkeit` enthalten (wenige jedoch nur eines von beiden), könnten diese Konzepte in eine Kookkurrenzregel zusammengefasst (`preis & verfügbar`) und beispielsweise einer Unterkategorie der Kategorie `preis` zugewiesen werden. [Für weitere Informationen siehe Thema Kookkurrenzregeln auf S. 126.](#)

- **Minimale Anzahl an Datensätzen.** Um festzustellen, wie interessant Kookkurrenzen sind, definieren Sie die minimale Anzahl an Datensätzen, die eine bestimmte Kookkurrenz enthalten muss, um als Deskriptor in einer Kategorie verwendet zu werden.

Maximaler Suchabstand. Legen Sie fest, wie weit die Verfahren suchen sollen, bevor Kategorien erstellt werden. Je niedriger der Wert, desto weniger Ergebnisse erhalten Sie. Allerdings sind die Ergebnisse weniger verrauscht und mit größerer Wahrscheinlichkeit auf signifikante Weise miteinander verknüpft oder verbunden. Je höher der Wert, desto mehr Ergebnisse erhalten Sie. Allerdings sind diese Ergebnisse möglicherweise weniger zuverlässig oder relevant. Während diese Option global auf alle Verfahren angewendet wird, hat sie die größte Auswirkung auf Kookkurrenzen und semantische Netze.

Paarung spezifischer Konzepte verhindern. Markieren Sie dieses Kontrollkästchen, um den Vorgang der Gruppierung oder Paarung von zwei Konzepten in der Ausgabe zu verhindern. Klicken Sie zum Erstellen oder Verwalten von Konzeptpaaren auf `Paare verwalten....` [Für weitere Informationen siehe Thema Verwalten von Verknüpfungsausnahmepaaren auf S. 120.](#)

Nach Möglichkeit mit Platzhaltern verallgemeinern. Wählen Sie diese Option, um dem Produkt zu ermöglichen, für Kategorien, in denen ein Sternchen als Platzhalter verwendet wird, allgemeine Regeln aufzustellen. Beispielsweise könnte anstelle der Erzeugung mehrerer Deskriptoren wie `[Apfelkuchen + .]` und `[Apfelsauce + .]` der Einsatz von Platzhaltern `[Apfel * + .]` erzeugen. Wenn Sie mit Platzhaltern verallgemeinern, erhalten Sie oft genau die gleiche Anzahl an Datensätzen wie zuvor. Diese Option hat jedoch den Vorteil, die Zahl zu verringern und die Kategoriedeskriptoren zu vereinfachen. Zusätzlich erhöht diese Option die Möglichkeit, mehr Datensätze unter Verwendung dieser Kategorien zu neuen Textdaten (zum Beispiel bei Langzeit-/Wellenstudien) zu kategorisieren.

Weitere Optionen für den Aufbau von Kategorien

Neben der Auswahl der anzuwendenden Gruppierungstechniken können Sie folgende weitere Optionen bearbeiten:

Maximale Anzahl an erstellten Kategorien . Verwenden Sie diese Option zur Beschränkung der Anzahl an Kategorien, die erstellt werden können, wenn Sie als Nächstes auf die Schaltfläche “Kategorien aufbauen” klicken. In einigen Fällen erzielen Sie bessere Ergebnisse, wenn Sie diesen Wert hoch setzen und dann etwaige uninteressante Kategorien löschen.

Mindestanzahl an Deskriptoren und/oder Unterkategorien pro Kategorie. Verwenden Sie diese Option, um die Mindestanzahl an Deskriptoren und Unterkategorien zu definieren, die eine Kategorie enthalten muss, um erstellt zu werden. Durch diese Option kann das Erstellen von Kategorien eingeschränkt werden, die keine hohe Zahl von Datensätzen erfassen.

Deskriptoren können in mehr als einer Kategorie angezeigt werden. Wenn ausgewählt, ermöglicht diese Option, dass Deskriptoren in mehr als einer der Kategorien verwendet werden, die als nächste erstellt werden. Diese Option ist allgemein ausgewählt, da Elemente häufig oder “natürlich” in zwei oder mehr Kategorien fallen, so dass sie in der Regel zu Kategorien höherer Qualität führen. Wenn Sie diese Option nicht auswählen, verringern Sie die Überschneidung von Datensätzen in mehreren Kategorien und abhängig von dem vorhandenen Datentyp. Dies könnte gewünscht sein. Bei den meisten Datentypen jedoch führt die Einschränkung von Deskriptoren auf eine einzelne Kategorie zu einem Verlust an Qualität oder Kategorieabdeckung. Angenommen, Sie hätten das Konzept *Autositzhersteller*. Mit dieser Option könnte dieses Konzept in einer Kategorie basierend auf dem Text *Autositz* und in einem anderen auf *Hersteller* basieren. Wenn diese Option aber nicht ausgewählt ist, wird das Konzept *Autositzhersteller*, auch wenn Sie noch beide Kategorien erhalten, nur als Deskriptor in der Kategorie angezeigt, in der es basierend auf verschiedenen Faktoren einschließlich der Anzahl an Datensätzen, in denen *Autositz* und *Hersteller* jeweils auftreten, am besten passt.

Konflikte mit doppelten Namen lösen. Wählen Sie, wie mit neuen Kategorien oder Unterkategorien verfahren werden soll, deren Namen mit denen von bestehenden Kategorien identisch wären. Sie können entweder die neuen Kategorien (und ihre Deskriptoren) mit den bestehenden Kategorien desselben Namens zusammenführen. Alternativ können Sie die Erstellung jeglicher Kategorien überspringen, wenn in den bestehenden Kategorien ein Namensduplikat gefunden wird.

Verwalten von Verknüpfungsausnahmepaaren

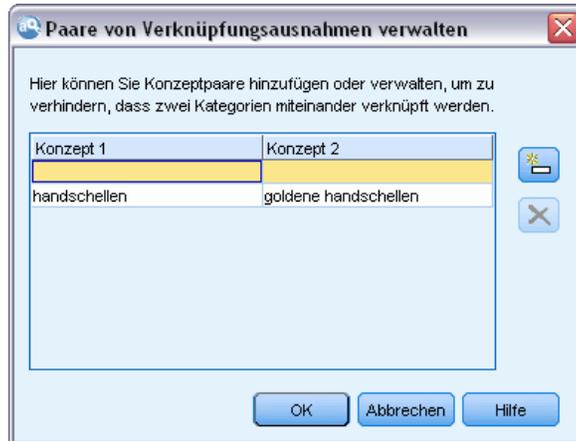
Bei der automatischen Kategorieerstellung und Zuordnen von Konzepten gruppieren die internen Algorithmen Wörter anhand bekannter Zuweisungen. Damit zwei Konzepte nicht gepaart oder miteinander verknüpft werden können, aktivieren Sie diese Funktion im Dialogfeld *Kategorien aufbauen*: *Erweiterte Einstellungen* und im Dialogfeld *Konzeptkartenindex*: *Einstellungen* und klicken auf die Schaltfläche *Paare verwalten*.

Im anschließend gezeigten Dialogfeld *Verknüpfungsausnahmen verwalten* können Sie Konzeptpaare hinzufügen, bearbeiten oder löschen. Geben Sie ein Paar pro Zeile ein. Durch die Eingabe von Paaren an dieser Stelle wird verhindert, dass die Paarbildung beim Erstellen oder Erweitern von Kategorien und beim Zuordnen von Konzepten erfolgt. Geben Sie Wörter exakt wie gewünscht ein, z. B. unterscheidet sich die mit einem Akzent versehene Version eines Wortes von der Wortversion ohne Akzent.

Wenn Sie z. B. sicherstellen möchten, dass `Hot Dog` und `Dog` nicht gruppiert werden, können Sie das Set als separate Zeile in der Tabelle hinzufügen.

Abbildung 6-11

Dialogfeld zum Verwalten von Verknüpfungsausnahmepaaren



Über linguistische Verfahren

Beim Aufbau bzw. der Erweiterung von Kategorien haben Sie die Auswahl zwischen einigen erweiterten linguistischen Verfahren für den Kategorieaufbau, darunter *Konzeptwurzelableitung*, *Konzepteinbeziehung*, *semantische Netze* (nur für Englisch) und *Kookkurrenzregeln*. Zum Erstellen von Kategorien können diese Verfahren einzeln oder in Kombination verwendet werden.

Sie müssen nicht besonders gut mit diesen Einstellungen vertraut sein, um sie verwenden zu können. Standardmäßig sind die gängigsten Einstellungen bereits ausgewählt. Sie können die Dialogfelder für die erweiterten Einstellungen überspringen und gleich mit dem Aufbau oder der Erweiterung der Kategorien beginnen. Wenn Sie hier Änderungen vornehmen, müssen Sie diese nicht mit jedem Öffnen des Dialogfelds erneut vornehmen, da immer die neuesten Einstellungen beibehalten werden.

Beachten Sie jedoch, dass jedes Daten-Set anders ist und sich die Anzahl der Methoden und die Reihenfolge, in der sie angewendet werden, daher im Laufe der Zeit ändern kann. Da Ihre Text-Mining-Ziele von Daten-Set zu Daten-Set unterschiedlich sein können, müssen Sie möglicherweise mit den verschiedenen Verfahren experimentieren, um zu sehen, welches die besten Ergebnisse für die jeweiligen Textdaten bringt. Keines der automatischen Verfahren ergibt eine perfekte Kategorisierung Ihrer Daten, weshalb wir empfehlen, ein oder mehrere automatische Verfahren anzuwenden, die gut mit Ihren Daten funktionieren.

Die wichtigsten automatischen linguistischen Verfahren für den Aufbau von Kategorien:

- **Konzeptwurzelableitung.** Bei diesem Verfahren werden Kategorien erstellt, indem mit einem Konzept verwandte Konzepte durch die Analyse einzelner Konzeptkomponenten hinsichtlich ihrer morphologischen Verwandtschaft gefunden werden. [Für weitere Informationen siehe Thema Konzeptwurzelableitung auf S. 122.](#)
- **Konzepteinbeziehung.** Dieses Verfahren erstellt Kategorien, indem es ausgehend von einem Konzept andere Konzepte ermittelt, die dieses Konzept enthalten. [Für weitere Informationen siehe Thema Konzepteinbeziehung auf S. 123.](#)

- **Semantisches Netz.** Bei diesem Verfahren wird zunächst auf der Grundlage eines umfassenden Index von Wortbeziehungen jedes Konzept auf seine möglichen Bedeutungen untersucht. Anschließend werden Kategorien durch Gruppieren zusammenhängender Konzepte erstellt. [Für weitere Informationen siehe Thema Semantische Netze auf S. 124.](#) Diese Option ist nur für englischen Text verfügbar.
- **Kookkurrenz.** Diese Technik erstellt Kookkurrenzregeln, die verwendet werden können, um eine neue Kategorie zu erstellen oder eine Kategorie zu erweitern, oder als Eingabe einer anderen Kategorietechnik verwendet werden. [Für weitere Informationen siehe Thema Kookkurrenzregeln auf S. 126.](#)

Konzeptwurzelableitung

Beim Konzeptwurzelableitungsverfahren werden Kategorien erstellt, indem mit einem Konzept verwandte Konzepte durch die Analyse einzelner Konzeptkomponenten hinsichtlich ihrer morphologischen Verwandtschaft gefunden werden. Eine Komponente ist ein Wort. Das Verfahren versucht, Konzepte durch Untersuchung der Endungen (Suffixe) der einzelnen Komponenten und durch Ermittlung anderer Konzepte, die daraus abgeleitet werden können, zusammenzufassen. Dahinter steht die Idee, dass Wörter, die voneinander abgeleitet werden, mit hoher Wahrscheinlichkeit auch dieselbe oder eine ähnliche Bedeutung aufweisen. Zur Ermittlung der Endungen werden interne sprachspezifische Regeln angewendet. So wird beispielsweise das Konzept `Möglichkeiten` zum `Aufstieg` mit den Konzepten `Möglichkeit` des `Aufstiegs` und `Aufstiegsmöglichkeit` zu einer Kategorie zusammengefasst.

Sie können die Konzeptwurzelableitung bei allen Textsorten einsetzen. Für sich genommen führt sie zu relativ wenigen Kategorien und die einzelnen Kategorien enthalten in der Regel nur wenige Konzepte. Die Konzepte in den einzelnen Kategorien sind entweder synonym oder situationsbezogen verwandt. Es kann nützlich sein, diesen Algorithmus zu verwenden, selbst bei einer manuellen Erstellung der Kategorien; die mithilfe dieses Algorithmus gefundenen Synonyme sind möglicherweise Synonyme der Konzepte, an denen Sie besonders interessiert sind.

Anmerkung: Sie können verhindern, dass Konzepte miteinander gruppiert werden, indem Sie sie explizit angeben. [Für weitere Informationen siehe Thema Verwalten von Verknüpfungsausnahmepaaren auf S. 120.](#)

Begriffskomponentenbildung und Bildung der Grundform

Bei Anwendung der Verfahren zur Konzeptwurzelableitung bzw. Konzepteinbeziehung werden die Ausdrücke zunächst in Komponenten (Wörter) gegliedert und anschließend wird die Grundform der Komponenten gebildet. Bei der Anwendung eines Verfahrens werden die Konzepte und die einem Konzept zugeordneten Fachausdrücke geladen und auf der Grundlage von Trennzeichen wie Leerzeichen, Bindestriche und Apostrophe in Komponenten aufgespalten. Der Ausdruck `Systemadministrator` beispielsweise wird wie folgt in Komponenten aufgespalten `{Administrator, System}`.

Einige Teile des ursprünglichen Ausdrucks werden jedoch möglicherweise nicht verwendet. Diese werden als "Stoppwörter" bezeichnet. Im Deutschen gehören beispielsweise folgende Wörter zu diesen ignorierbaren Komponenten: `ein`, `und`, `als`, `durch`, `für`, `von`, `in`, `auf`, `oder`, `der`, `bis` und `mit`.

So weist der Ausdruck *examination of the data* das Komponenten-Set {*data*, *examination*} auf; *von* und *der* werden als ignorierbar betrachtet. Die Reihenfolge der Komponenten in einem Komponenten-Set ist nicht von Bedeutung. Daher können folgende drei Ausdrücke äquivalent sein: *Hustenmedikament für Kinder*, *Kindermedikament gegen Husten* und *Medikament gegen Husten bei Kindern*, da alle dasselbe Komponenten-Set {*Kind*, *Husten*, *Medikament*} aufweisen. Die einzelnen Begriffspaare werden als äquivalent identifiziert. Die zugehörigen Konzepte werden zu einem neuen Konzept zusammengeführt, das alle drei Ausdrücke referenziert.

Da die Komponenten eines Terms außerdem flektiert sein können, werden intern sprachspezifische Regeln angewendet, um äquivalente Ausdrücke unabhängig von ihren Flexionsabweichungen zu ermitteln. Somit können die Ausdrücke *Waldschutz* und *Schutz der Wälder* als äquivalent identifiziert werden, da die Grundform im Singular *Wald* lautet.

Funktionsprinzipien der Konzeptwurzelableitung

Nachdem die Ausdrücke in Komponenten zerlegt und ihre Grundform gebildet wurde (siehe vorheriger Abschnitt), analysiert der Konzeptwurzelableitungsalgorithmus die Komponentenendungen (Suffixe), um den Stamm der Komponente zu ermitteln, und fasst anschließend die Konzepte mit anderen Konzepten, die denselben oder einen ähnlichen Stamm besitzen, zusammen. Die Endungen werden mithilfe einer Reihe linguistischer Ableitungsregeln für die jeweilige Textsprache identifiziert. So gibt es beispielsweise eine Ableitungsregel für englischsprachige Texte, die besagt, dass eine Konzeptkomponente, die auf das Suffix *ical* endet, von einem Konzept abgeleitet sein könnte, das denselben Stamm aufweist und auf das Suffix *ic* endet. Mit dieser Regel (und der Bildung der Grundform) kann der Algorithmus die Konzepte *epidemiologic study* und *epidemiological studies* zusammenfassen.

Da die Begriffe bereits in Komponenten zerlegt sind und die ignorierbaren Konzepte (z. B. *in* und *von*) identifiziert wurden, kann der Konzeptwurzelableitungsalgorithmus auch das Konzept *studies in epidemiology* mit *epidemiological studies* in diese Kategorie einordnen.

Der Satz an Komponentenableitungsregeln wurde so gewählt, dass die meisten durch diesen Algorithmus zusammengefassten Konzepte synonym sind: Die Konzepte *epidemiologic studies*, *epidemiological studies*, *studies in epidemiology* sind alle äquivalente Begriffe. Zur größeren Vollständigkeit gibt es auch Ableitungsregeln, die dem Algorithmus die Zusammenfassung von Konzepten gestatten, die situationsbezogen verwandt sind. So kann der Algorithmus beispielsweise Konzepte wie *Gründer des Reichs* und *Gründung des Reichs* zusammenfassen.

Konzepteinbeziehung

Beim Konzepteinbeziehungsverfahren werden Kategorien aufgebaut, indem mit Algorithmen für lexikalische Reihen Konzepte identifiziert werden, die in anderen Konzepten enthalten sind. Dahinter steht folgende Idee: Wenn Wörter in einem Konzept eine Untergruppe eines anderen Konzepts bilden, ist dies Ausdruck einer zugrunde liegenden semantischen Beziehung. Die Einbeziehung ist ein leistungsstarkes Verfahren, das auf Texte aller Art angewendet werden kann.

Dieses Verfahren funktioniert am besten in Verbindung mit semantischen Netzen, kann aber auch getrennt verwendet werden. Die Konzepteinbeziehung kann bessere Ergebnisse erzielen, wenn die Datensätze einen großen Anteil an domänenspezifischer Terminologie oder an

Fachjargon enthalten. Dies gilt insbesondere dann, wenn die Wörterbücher zuvor abgestimmt wurden, sodass die speziellen Fachausdrücke extrahiert und entsprechend gruppiert werden (mit Synonymen).

Funktionsprinzipien der Konzepteinbeziehung

Vor der Anwendung des Konzepteinbeziehungsalgorithmus werden die Terme in Komponenten zerlegt und auf ihre Grundform zurückgeführt. [Für weitere Informationen siehe Thema Konzeptwurzelaufleitung auf S. 122](#). Anschließend analysiert der Einbeziehungsalgorithmus die Komponenten-Sets. Bei jedem Komponenten-Set sucht der Algorithmus nach einem weiteren Komponenten-Set, bei dem es sich um eine Untergruppe des ersten Komponenten-Sets handelt.

Wenn Ihnen beispielsweise das Konzept `kontinentales Frühstück`, das das Komponenten-Set `{Frühstück, kontinental}` aufweist, und das Konzept `Frühstück`, das das Komponenten-Set `{Frühstück}` aufweist, vorliegen, folgert der Algorithmus, dass `kontinentales Frühstück` eine Art `Frühstück` ist, und fasst die beiden Konzepte zusammen.

Oder ein umfangreicheres Beispiel: Wenn Ihnen im Fensterbereich “Extrahierungsergebnisse” das Konzept `Sitz` vorliegt und Sie diesen Algorithmus anwenden, werden Konzepte wie `Kindersitz`, `Ledersitz`, `Sitzheizung`, `Elektro-Sitzheizung`, `Kindersitzgurt` und `Kindersitzvorschriften` ebenfalls in diese Kategorie eingeordnet.

Da die Begriffe bereits in Komponenten zerlegt sind und die ignorierbaren Konzepte (z. B. `in` und `von`) identifiziert wurden, würde der Konzepteinbeziehungsalgorithmus erkennen, dass das Konzept `Fortgeschrittenenkurs in Spanisch` das Konzept `Spanischkurs` beinhaltet.

Anmerkung: Sie können verhindern, dass Konzepte miteinander gruppiert werden, indem Sie sie explizit angeben. [Für weitere Informationen siehe Thema Verwalten von Verknüpfungsausnahmepaaren auf S. 120](#).

Semantische Netze

In dieser Version ist das Verfahren mit semantischen Netzen nur für englischsprachige Texte verfügbar.

Bei diesem Verfahren werden Kategorien mithilfe eines integrierten Netzes von Wortbeziehungen erstellt. Aus diesem Grund können mit diesem Verfahren sehr gute Ergebnisse erzielt werden, wenn die Ausdrücke konkret sind und nur einen geringen Grad an Mehrdeutigkeit aufweisen. Es ist jedoch nicht zu erwarten, dass dieses Verfahren viele Zusammenhänge zwischen sehr technischen/spezialisierten Konzepten findet. Beim Umgang mit solchen Konzepten sind das Konzepteinbeziehungs- und das Konzeptwurzelaufleitungsverfahren zumeist von größerem Nutzen.

Funktionsprinzipien semantischer Netze

Hinter dem Verfahren mit semantischen Netzen steht die Idee, bekannte Wortbeziehungen zu nutzen, um Kategorien von Synonymen bzw. Hyponymen zu erzeugen. Ein **Hyponym** liegt vor, wenn ein Konzept eine Sorte eines zweiten Konzepts ist, dergestalt, dass eine hierarchische Beziehung (auch als ISA-Beziehung bezeichnet) vorliegt. Beispiel: Wenn `animal` ein Konzept

ist, dann sind `cat` und `kangaroo` Hyponyme von `animal`, da es sich dabei jeweils um eine Art Tier (`animal`) handelt.

Neben Synonym- und Hyponymbeziehungen untersucht das semantische Netz auch Teilzusammenhänge und vollständige Zusammenhänge zwischen Konzepten aus dem Typ `<Standort>`. Beispielsweise ordnet das Verfahren die Konzepte `normandy`, `provence` und `france` in dieselbe Kategorie ein, da Normandie (Normandy) und Provence Teile von Frankreich (France) sind.

Bei dem Verfahren mit semantischen Netzen werden zunächst die möglichen Bedeutungen der einzelnen Konzepte im semantischen Netz ermittelt. Wenn Konzepte als Synonyme oder Hyponyme identifiziert werden, werden sie alle in dieselbe Kategorie eingeordnet. Beispielsweise erstellt dieses Verfahren eine einzelne Kategorie, die die folgenden drei Konzepte enthält: `Speiseapfel`, `Dessertapfel` und `Granny Smith`, da das semantische Netz folgende Informationen enthält: 1) `Dessertapfel` ist ein Synonym von `Speiseapfel` und 2) `Granny Smith` ist eine Sorte von `Speiseapfel` (d. h. ist ein Hyponym von `Speiseapfel`).

Isoliert betrachtet sind viele Konzepte, insbesondere Uniterms, mehrdeutig. Das Konzept `buffet` beispielsweise kann eine Art Mahlzeit oder ein Möbelstück bezeichnen. Wenn die Menge der Konzepte `Mahlzeit`, `Möbel` und `Schrank` beinhaltet, ist der Algorithmus gezwungen zu entscheiden, ob `Schrank` in dieselbe Kategorie eingeordnet werden soll wie `Mahlzeit` oder `Möbel`. Beachten Sie, dass es vorkommen kann, dass die vom Algorithmus getroffene Wahl im Kontext eines bestimmten Sets an Datensätzen nicht angemessen ist.

Das Verfahren mit semantischen Netzen führt bei bestimmten Arten von Daten zu besseren Ergebnissen als die Konzeptenbeziehung. Sowohl das semantische Netz als auch die Konzeptenbeziehung erkennen, dass `Apfelkuchen` eine Art von `Kuchen` ist, aber nur das semantische Netz erkennt, dass `Tarte` auch eine Art von `Kuchen` ist.

Semantische Netze können auch zusammen mit anderen Verfahren eingesetzt werden. Beispielsweise angenommen, Sie haben sowohl das Verfahren mit semantischen Netzen als auch das Einbeziehungsverfahren ausgewählt und das semantische Netz hat das Konzept `teacher` in dieselbe Kategorie eingeordnet wie das Konzept `tutor` (da ein Tutor eine Art von Lehrer (`teacher`) ist). Der Einbeziehungsalgorithmus kann das Konzept `graduate tutor` mit `tutor` zusammenfassen und als Ergebnis erstellen die beiden Algorithmen zusammen eine Ausgabekategorie, die alle drei Konzepte enthält: `tutor`, `graduate tutor` und `teacher`.

Optionen für semantische Netze

Es sind einige zusätzliche Einstellungen vorhanden, die für dieses Verfahren interessant sein könnten.

- Ändern der Einstellung `Maximaler Suchabstand`. Legen Sie fest, wie weit die Verfahren suchen sollen, bevor Kategorien erstellt werden. Je niedriger der Wert, desto weniger Ergebnisse erhalten Sie. Allerdings sind die Ergebnisse weniger verrauscht und mit größerer Wahrscheinlichkeit auf signifikante Weise miteinander verknüpft oder verbunden. Je höher der Wert, desto mehr Ergebnisse erhalten Sie. Allerdings sind diese Ergebnisse möglicherweise weniger zuverlässig oder relevant.

Abhängig vom Abstand sucht der Algorithmus zum Beispiel von `Dänischem Gebäck` bis zu `Blätterteiggebäck` (übergeordnet), dann `Teiggebäck` (über-übergeordnet) und aufwärts bis `Gebäck`.

Durch Verringern des Suchabstands führt dieses Verfahren zu kleineren Kategorien, mit denen sich möglicherweise leichter arbeiten lässt, wenn Sie den Eindruck haben, dass die erstellten Kategorien zu groß sind oder zu viele Elemente zu einer Gruppe zusammenfassen.

Wichtig: Außerdem wird empfohlen, bei Verwendung dieses Verfahrens die Option Rechtschreibung korrigieren für eine minimale Anzahl an Stammzeichen von (definiert im Dialogfeld “Extrahieren”) für Fuzzy-Gruppierung nicht anzuwenden, da Fehlgruppierungen große negative Auswirkungen auf die Ergebnisse haben können.

Kookkurrenzregeln

Mit Kookkurrenzregeln können Sie Konzepte ermitteln und zusammenfassen, die im Set an Datensätzen eng miteinander verwandt sind. Dahinter steht folgende Idee: Wenn Konzepte häufig in Datensätzen gefunden werden, ist diese Kookkurrenz Ausdruck einer zugrunde liegenden Beziehung, die wahrscheinlich in Ihren Kategoriedefinitionen von Nutzen ist. Diese Technik erstellt Kookkurrenzregeln, die verwendet werden können, um eine neue Kategorie zu erstellen oder eine Kategorie zu erweitern, oder als Eingabe einer anderen Kategorietechnik verwendet werden. Zwei Konzepte weisen eine starke Kookkurrenz auf, wenn sie häufig zusammen in einer Menge von Datensätzen vorkommen und sehr selten einzeln in den anderen Datensätzen. Dieses Verfahren kann bei größeren Daten-Sets mit mindestens mehreren Hundert Datensätzen zu guten Ergebnissen führen.

Wenn beispielsweise viele Datensätze die Wörter `Preis` und `Verfügbarkeit` enthalten, könnten diese Konzepte in eine Kookkurrenzregel zusammengefasst werden (`Preis & verfügbar`). Wenn die Konzepte `Erdnussbutter`, `Gelee` und `Sandwich` häufiger zusammen als getrennt vorkommen, werden sie gemeinsam in eine Konzeptkookkurrenzregel (`Erdnussbutter&Gelee & Sandwich`) aufgenommen.

Wichtig: In früheren Versionen wurden Kookkurrenz- und Synonymregeln von eckigen Klammern umgeben. In dieser neuen Version zeigen nun eckige Klammern ein Musterergebnis für die an. Stattdessen stehen Kookkurrenz- und Synonymregeln in runden Klammern, z. B. (`Lautsprechersysteme|Lautsprecher`).

Funktionsprinzipien von Kookkurrenzregeln

Bei diesem Verfahren werden die Datensätze durchsucht, um zwei oder mehr Konzepte zu finden, die häufig gemeinsam vorkommen. Zwei oder mehr Konzepte weisen eine starke Kookkurrenz auf, wenn sie häufig zusammen in einer Menge von Datensätzen vorkommen und selten einzeln in den anderen Datensätzen.

Wenn kookkurrierende Konzepte gefunden werden, wird eine Kategorieregel erstellt. Diese Regeln bestehen aus zwei oder mehr Konzepten, die über den Boole’schen Operator `&` verbunden sind. Bei diesen Regeln handelt es sich um logische Aussagen, die einen Datensatz automatisch in eine Kategorie einordnen, wenn die in der Regel enthaltenen Konzepte alle in dem betreffenden Datensatz kookkurrieren.

Optionen für Kookkurrenzregeln

Wenn Sie das Kookkurrenzregelverfahren verwenden, können Sie mehrere Einstellungen optimieren, die Einfluss auf die resultierenden Regeln haben:

- Ändern der Einstellung Maximaler Suchabstand. Legen Sie fest, wie weit die Verfahren suchen sollen, bevor Kategorien erstellt werden. Je niedriger der Wert, desto weniger Ergebnisse erhalten Sie. Allerdings sind die Ergebnisse weniger verrauscht und mit größerer Wahrscheinlichkeit auf signifikante Weise miteinander verknüpft oder verbunden. Je höher der Wert, desto mehr Ergebnisse erhalten Sie. Allerdings sind diese Ergebnisse möglicherweise weniger zuverlässig oder relevant. Bei der Arbeit mit Kookkurrenzen führt der Standardwert für den Suchabstand zu extrem vielen Kookkurrenzen, von denen viele nur eine schwache Verknüpfung aufweisen und daher uninteressant sind. Wenn Sie den Suchabstand verringern, werden die schwächeren Kookkurrenzen herausgefiltert und Sie erhalten nur die signifikanteren Ergebnisse.
- Minimale Anzahl an Datensätzen. Um festzustellen, wie interessant Kookkurrenzen sind, definieren Sie die minimale Anzahl an Datensätzen, die eine bestimmte Kookkurrenz enthalten muss, um als Deskriptor in einer Kategorie verwendet zu werden. Wenn Sie bei kleineren Daten-Sets diese Option niedriger einstellen, wird das Auffinden von Kookkurrenzen leichter.

Anmerkung: Sie können verhindern, dass Konzepte miteinander gruppiert werden, indem Sie sie explizit angeben. [Für weitere Informationen siehe Thema Verwalten von Verknüpfungsausnahmepaaren auf S. 120.](#)

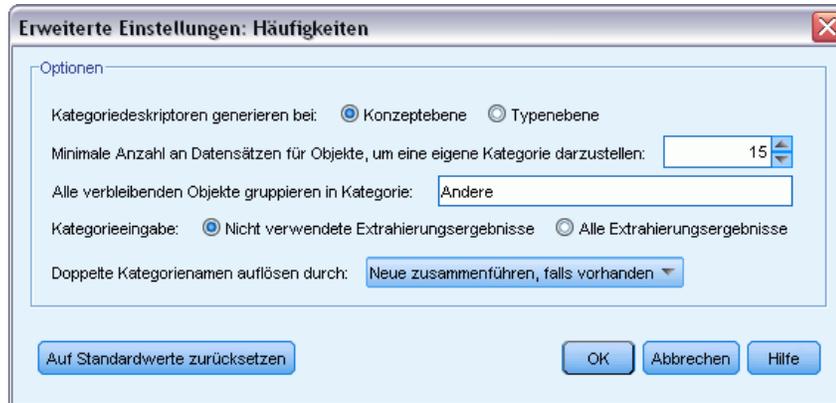
Erweiterte Einstellungen für Häufigkeit

Sie können Kategorien basierend auf einem direkten und mechanischen Häufigkeitsverfahren erstellen. Mit diesem Verfahren können Sie eine Kategorie für jedes Element (Typ, Konzept oder Muster) aufbauen, das über einem gegebenen Datensatz gefunden wurde. Zusätzlich können Sie eine einzelne Kategorie für jedes der weniger häufig auftretenden Elemente erstellen. Häufigkeit bezieht sich auf die Anzahl an Datensätzen, die das extrahierte Konzept (und seine Synonyme), den Typ oder das Muster enthalten. Es geht nicht um die Gesamtanzahl an Vorkommnissen im gesamten Text.

Das Gruppieren häufig auftretender Objekte kann interessante Ergebnisse bringen, da dies eine übliche oder signifikante Antwort angeben kann. Das Verfahren ist sehr nützlich auf die unbenutzten Extrahierungsergebnisse anzuwenden, nachdem andere Verfahren angewendet wurden. Eine andere Anwendung besteht in der direkten Ausführung dieses Verfahrens nach der Extrahierung, wenn keine anderen Kategorien vorhanden sind, der Bearbeitung der Ergebnisse, um nicht interessante Kategorien zu löschen und dann der Erweiterung dieser Kategorien, damit sie noch mehr Datensätze abdecken. [Für weitere Informationen siehe Thema Erweitern von Kategorien auf S. 129.](#)

Anstelle dieser Technik könnten Sie die Konzepte und Konzeptmuster nach absteigenden Nummern von Datensätzen in den Extrahierungsergebnissen sortieren und dann die obersten in den Kategorienbereich ziehen, um die entsprechenden Kategorien zu erstellen.

Abbildung 6-12
Erweiterte Einstellungen: Dialogfeld "Häufigkeiten"



Kategoriedeskriptoren generieren bei. Wählen Sie die Eingabeart für Deskriptoren. [Für weitere Informationen siehe Thema Erstellen von Kategorien auf S. 112.](#)

- **Konzeptebene.** Wenn Sie diese Option auswählen, werden Konzept- oder Konzeptmusterhäufigkeiten verwendet. Konzepte werden verwendet, wenn Typen als Eingabe für den Kategorieaufbau gewählt wurden. Konzeptmuster werden verwendet, wenn Typmuster ausgewählt wurden. Im Allgemeinen ergibt dieses Verfahren für die Konzeptebene spezifischere Ergebnisse, da Konzepte und Konzeptmuster eine geringere Messungsebene repräsentieren.
- **Typenebene.** Wenn Sie diese Option auswählen, werden Typen- oder Typenmusterhäufigkeiten verwendet. Typen werden verwendet, wenn Typen als Eingabe für den Kategorieaufbau gewählt wurden. Typmuster werden verwendet, wenn Typmuster ausgewählt wurden. Wenn Sie dieses Verfahren auf die Typenebene anwenden, erhalten Sie einen raschen Einblick in die Bandbreite der gegebenen Antworten.

Minimale Häufigkeit Datensatz für Elemente für eine eigene Kategorie. Mithilfe dieser Option können Sie Kategorien aus häufig auftretenden Elementen aufbauen. Diese Option beschränkt die Ausgabe auf nur die Kategorien mit einem Deskriptor, die in mindestens der Anzahl X von Datensätzen aufgetreten sind, wobei X den Wert angibt, der für diese Option eingegeben werden soll.

Alle verbleibenden Objekte gruppieren in Kategorie. Mithilfe dieser Option können Sie alle Konzepte oder Typen, die selten auftreten, in einer einzigen Kategorie für alle mit einem Namen Ihrer Wahl zusammenfassen. Standardmäßig heißt diese Kategorie *Andere*.

Kategorieeingabe. Wählen Sie die Gruppe aus, auf die Sie die Verfahren anwenden möchten:

- **Nicht verwendete Extrahierungsergebnisse.** Diese Option aktiviert Kategorien, die aus Extrahierungsergebnissen erstellt werden, die nicht in vorhandenen Kategorien verwendet werden. So wird die Tendenz für Datensätze minimiert, mehrere Kategorien abzugleichen und die Anzahl der erzeugten Kategorien zu begrenzen.
- **Alle Extrahierungsergebnisse.** Diese Option aktiviert unter Verwendung der Extrahierungsergebnisse aufzubauende Kategorien. Dies ist am sinnvollsten, wenn noch keine oder nur sehr wenige Kategorien vorhanden sind.

Konflikte mit doppelten Namen lösen. Wählen Sie, wie mit neuen Kategorien verfahren werden soll, deren Namen mit denen von bestehenden Kategorien identisch wären. Sie können entweder die neuen Kategorien (und ihre Deskriptoren) mit den bestehenden Kategorien desselben Namens zusammenführen. Alternativ können Sie die Erstellung jeglicher Kategorien überspringen, wenn in den bestehenden Kategorien ein Namensduplikat gefunden wird.

Erweitern von Kategorien

Das Erweitern ist ein Prozess, durch den Deskriptoren automatisch hinzugefügt oder verbessert werden, um vorhandene Kategorien auszubauen. Ziel ist es, eine bessere Kategorie zu erzeugen, die verwandte Datensätze erfasst, die der Kategorie ursprünglich nicht zugeordnet waren.

Die automatischen Gruppierungsverfahren, die Sie auswählen, versuchen, Konzepte, TLA-Muster und Kategorieregeln zu den vorhandenen Kategoriedeskriptoren zu identifizieren. Diese neuen Konzepte, Muster und Kategorieregeln werden dann als neue Deskriptoren hinzugefügt oder vorhandenen Deskriptoren hinzugefügt. Zu den Erweiterungsgruppierungsverfahren gehören *Konzeptwurzelableitung*, *Konzepteinbeziehung*, *semantische Netze* (nur Englisch) sowie *Kookkurrenzregeln*. Die Methode Leere Kategorien mit aus dem Kategorienamen erzeugten Deskriptoren erweitert erzeugt Deskriptoren mit den Wörtern in den Kategorienamen. Je beschreibender ein Kategorienamen, um so besser daher die Ergebnisse.

Anmerkung: Die häufigkeitsbasierten Verfahren stehen beim Erweitern von Kategorien nicht zur Verfügung.

Erweitern ist ideal, um Ihre Kategorien interaktiv zu verbessern. Hier einige Beispiele für das Erweitern einer Kategorie:

- Nach dem Ziehen von Konzeptmustern, um Kategorien im Bereich “Kategorien” zu erstellen
- Nach dem Erstellen von Kategorien per Hand und dem Hinzufügen einfacher Kategorieregeln und Deskriptoren
- Nach dem Import eines Coderahmens, bei dem die Kategorien sehr aussagekräftige Namen hatten
- Nach dem Verfeinern der Kategorien aus dem TAP, das Sie während der Projekterstellung gewählt hatten

Sie können eine Kategorie mehrfach erweitern. Wenn Sie zum Beispiel eine vordefinierte Kategoriedatei mit sehr beschreibenden Namen importiert haben, könnten Sie mit der Option Leere Kategorien mit aus dem Kategorienamen erzeugten Deskriptoren erweitern, um eine erste Menge an Deskriptoren zu erhalten, und dann diese Kategorien erneut erweitern. In anderen Fällen könnte das mehrfache Erweitern jedoch zu zu generischen Kategorien führen, wenn die Deskriptoren mehr und mehr erweitert werden. Da den Aufbau- und Erweiterungsgruppierungsverfahren ähnliche Algorithmen zugrunde liegen, führt das Erweitern direkt nach dem Aufbau von Kategorien mit nur geringer Wahrscheinlichkeit zu interessanteren Ergebnissen.

Tipps:

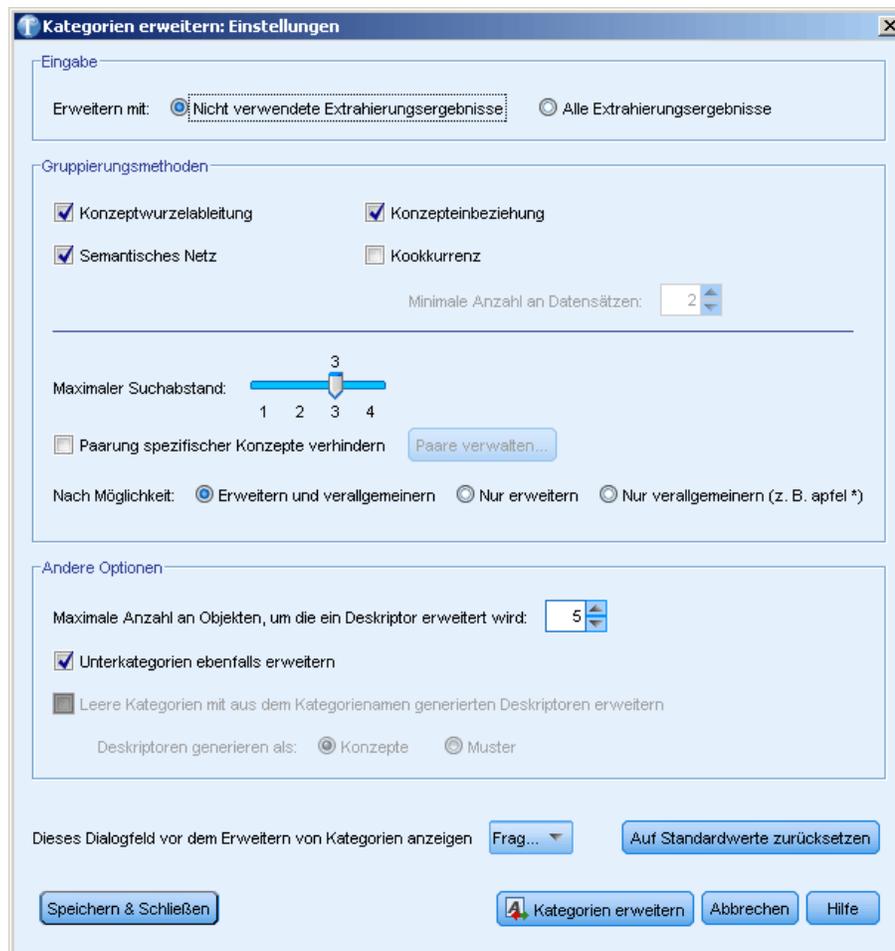
- Wenn Sie versuchen zu erweitern und die Ergebnisse nicht verwenden möchten, können Sie den Vorgang stets direkt nach dem Erweitern widerrufen (Bearbeiten > Rückgängig).
- Das Erweitern kann zwei oder mehr Kategorieregeln in einer Kategorie erzeugen, die mit genau dem gleichen Satz an Dokumenten übereinstimmen, da Regeln während des Prozesses unabhängig erzeugt werden. Wenn gewünscht, können Sie die Kategorien prüfen und Redundanzen durch manuelle Bearbeitung der Kategoriebeschreibung entfernen. [Für weitere Informationen siehe Thema Bearbeiten von Kategoriedeskriptoren auf S. 162.](#)

So erweitern Sie Kategorien

- ▶ Wählen Sie im Fensterbereich “Kategorien” die Kategorien aus, die Sie erweitern möchten.
- ▶ Wählen Sie in den Menüs die Optionsfolge Kategorien > Kategorien erweitern. Wenn Sie nicht die Option markiert haben, dass diese Aufforderung nie angezeigt werden soll, wird ein Dialogfeld geöffnet.
- ▶ Wählen Sie aus, ob Sie die Kategorie jetzt aufbauen oder zuerst die Einstellungen bearbeiten möchten.
 - Klicken Sie auf Jetzt erweitern, um mit den aktuellen Einstellungen mit der Erweiterung von Kategorien zu beginnen. Der Prozess wird gestartet und ein Dialogfeld über den Fortschritt angezeigt.
 - Klicken Sie auf Bearbeiten, um die Einstellungen zu überprüfen und zu ändern.

Nach dem Versuch der Erweiterung werden alle Kategorien, für die neue Deskriptoren gefunden werden, mit dem Wort Erweitert im Bereich “Kategorien” gekennzeichnet, so dass Sie sie schnell erkennen. Der Text “Erweitert” bleibt, bis Sie erneut erweitern oder die Kategorie anderweitig bearbeiten oder über das Kontextmenü löschen.

Abbildung 6-13
Dialogfeld "Kategorien erweitern"



Diese Verfahren, die für den Aufbau und die Erweiterung von Kategorien verfügbar sind, eignen sich jeweils gut für bestimmte Arten von Daten und Situationen, häufig ist es jedoch sinnvoll, in einer Analyse mehrere Verfahren zu kombinieren, um das gesamte Spektrum an Datensätzen zu erfassen. Es gibt Konzepte und Typen, die unter einer Kategorie gruppiert wurden und bei der nächsten Erstellung von Kategorien ebenfalls zur Verfügung stehen. Das bedeutet, dass Sie ein Konzept in mehreren Kategorien erkennen oder redundante Kategorien vorfinden können.

Kategorieeingabe. Auswahl, welche Eingabe zur Erweiterung der Kategorien verwendet wird:

- **Nicht verwendete Extrahierungsergebnisse.** Diese Option aktiviert Kategorien, die aus Extrahierungsergebnissen erstellt werden, die nicht in vorhandenen Kategorien verwendet werden. So wird die Tendenz für Datensätze minimiert, mehrere Kategorien abzugleichen und die Anzahl der erzeugten Kategorien zu begrenzen.
- **Alle Extrahierungsergebnisse.** Diese Option aktiviert unter Verwendung der Extrahierungsergebnisse aufzubauende Kategorien. Dies ist am sinnvollsten, wenn noch keine oder nur sehr wenige Kategorien vorhanden sind.

Gruppierverfahren

Eine kurze Beschreibung dieser Verfahren finden Sie unter [Erweiterte linguistische Einstellungen](#) auf S. 116. Zu diesen Techniken zählen:

- Konzeptwurzelaufleitung (*nicht für Japanisch verfügbar*)
- Semantisches Netz (*nur englischer Text*)
- Konzeptbeziehung
- Kookkurrenz und Unteroption Mindestanzahl an Dokumenten.

Eine Reihe von Typen wird dauerhaft aus dem Verfahren mit semantischen Netzen ausgeschlossen, da diese Typen nicht zu relevanten Ergebnissen führen. Sie umfassen <Positiv>, <Negativ>, <IP>, andere nichtlinguistische Typen usw.

Maximaler Suchabstand. Legen Sie fest, wie weit die Verfahren suchen sollen, bevor Kategorien erstellt werden. Je niedriger der Wert, desto weniger Ergebnisse erhalten Sie. Allerdings sind die Ergebnisse weniger verrauscht und mit größerer Wahrscheinlichkeit auf signifikante Weise miteinander verknüpft oder verbunden. Je höher der Wert, desto mehr Ergebnisse erhalten Sie. Allerdings sind diese Ergebnisse möglicherweise weniger zuverlässig oder relevant. Während diese Option global auf alle Verfahren angewendet wird, hat sie die größte Auswirkung auf Kookkurrenzen und semantische Netze.

Paarung spezifischer Konzepte verhindern. Markieren Sie dieses Kontrollkästchen, um den Vorgang der Gruppierung oder Paarung von zwei Konzepten in der Ausgabe zu verhindern. Klicken Sie zum Erstellen oder Verwalten von Konzeptpaaren auf [Paare verwalten...](#) [Für weitere Informationen siehe Thema Verwalten von Verknüpfungsausnahmepaaren auf S. 120.](#)

Wenn möglich: Wählen Sie, ob Sie die Deskriptoren einfach erweitern, mithilfe von Platzhaltern verallgemeinern oder beides tun möchten.

- **Erweitern und verallgemeinern.** Mit dieser Option werden die ausgewählten Kategorien erweitert und dann die Deskriptoren verallgemeinert. Wenn Sie verallgemeinern möchten, erstellt das Produkt allgemeine Kategorieregeln in Kategorien mithilfe eines Sternchens als Platzhalter. Beispielsweise könnte anstelle der Erzeugung mehrerer Deskriptoren wie [Apfelkuchen + .] und [Apfelsauce + .] der Einsatz von Platzhaltern [Apfel * + .] erzeugen. Wenn Sie mit Platzhaltern verallgemeinern, erhalten Sie oft genau die gleiche Anzahl an Datensätzen wie zuvor. Diese Option hat jedoch den Vorteil, die Zahl zu verringern und die Kategoriedeskriptoren zu vereinfachen. Zusätzlich erhöht diese Option die Möglichkeit, mehr Datensätze unter Verwendung dieser Kategorien zu neuen Textdaten (zum Beispiel bei Langzeit-/Wellenstudien) zu kategorisieren.
- **Nur erweitern.** Mit dieser Option werden Ihre Kategorien erweitert, ohne die Deskriptoren zu verallgemeinern. Es kann hilfreich sein, für manuell erstellte Kategorien zunächst die Option Nur erweitern zu wählen und dann die gleichen Kategorien mit der Option Erweitern und verallgemeinern noch einmal zu erweitern.
- **Nur verallgemeinern.** Mit dieser Option werden die Deskriptoren verallgemeinert, ohne Ihre Kategorien auf andere Weise zu erweitern.

Weitere Optionen für die Erweiterung von Kategorien

Neben der Auswahl der anzuwendenden Verfahren können Sie folgende weitere Optionen bearbeiten:

Maximale Anzahl an Objekten, um die ein Deskriptor erweitert wird. Definieren Sie bei der Erweiterung eines Deskriptors um Objekte (Konzepte, Typen und andere Ausdrücke) die maximale Anzahl an Objekten, die einem einzelnen Deskriptor hinzugefügt werden können. Wenn Sie als Limit 10 wählen, können einem vorhandenen Deskriptor höchstens 10 zusätzliche Objekte hinzugefügt werden. Wenn mehr als 10 Objekte hinzugefügt werden sollen, beendet das Verfahren das Hinzufügen neuer Objekte nach dem zehnten Objekt. Dies kann eine Deskriptorliste verkürzen, garantiert aber nicht, dass die interessantesten Objekte zuerst verwendet wurden. Eventuell möchten Sie die Größe der Erweiterung ohne Abstriche an der Qualität verringern, indem Sie die Option Mit Platzhaltern, wenn möglich, verallgemeinern. Diese Option gilt nur für Deskriptoren, die die Boole'schen Operatoren & (UND) bzw. ! (NICHT) enthalten.

Unterkategorien ebenfalls erweitern. Mit dieser Option werden auch alle Unterkategorien unter den ausgewählten Kategorien erweitert.

Kategorien mit Deskriptoren basierend auf Kategorienamen erweitern. Diese Option versucht, Deskriptoren für jede importierte Kategorie, basierend auf den Wörtern, die den Namen der Kategorie ausmachen, automatisch zu erstellen. Der Kategoriename wird gescannt, um festzustellen, ob Wörter im Namen einem extrahierten Konzept entsprechen. Wenn ein Konzept erkannt wird, wird es verwendet, um nach passenden Konzeptmustern zu suchen. Diese werden dann beide herangezogen, um Deskriptoren für die Kategorie zu bilden. Diese Option erzeugt die besten Ergebnisse, wenn die Kategorienamen lang und beschreibend sind. Dies ist eine schnelle Methode, um Kategoriedeskriptoren zu erzeugen, die es wiederum der Kategorie ermöglichen, Datensätze zu erfassen, die diese Deskriptoren enthalten. Diese Option ist vor allem dann hilfreich, wenn Sie Kategorien importieren oder Kategorien mit langen beschreibenden Namen manuell erstellen. Diese Methode wird nur auf leere Kategorien mit null Deskriptoren angewendet. Wenn eine Kategorie bereits Deskriptoren enthält, kann Sie nicht auf diese Art erweitert werden.

Deskriptoren erzeugen als. Diese Option ist nur verfügbar, wenn die vorherige Option ausgewählt wurde.

- **Konzepte.** Wählen Sie diese Option, um die resultierenden Deskriptoren in der Form von Konzepten zu erzeugen, ungeachtet dessen, ob sie aus dem Quelltext extrahiert wurden.
- **Muster.** Wählen Sie diese Option, um die resultierenden Deskriptoren in der Form von Mustern zu erzeugen, ungeachtet dessen, ob die resultierenden Muster oder ein beliebiges Muster extrahiert wurde.

Manuelle Erstellung von Kategorien

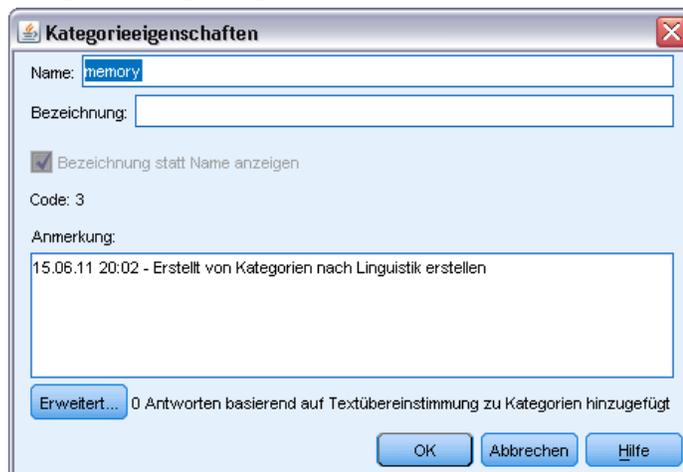
Neben dem Erstellen von Kategorien mithilfe der automatisierten Verfahren, des Coderahmenmanagers und des Regeleditors können Sie Kategorien auch manuell definieren. Es gibt folgende manuelle Methode:

- Erstellen einer leeren Kategorie, der Sie nacheinander Elemente hinzufügen. [Für weitere Informationen siehe Thema Erstellen neuer Kategorien bzw. Umbenennen von Kategorien auf S. 134.](#)
- Ziehen von Begriffen, Typen und Mustern in den Kategorienbereich. [Für weitere Informationen siehe Thema Erstellen von Kategorien durch Ziehen und Ablegen \(Drag&Drop\) auf S. 135.](#)

Erstellen neuer Kategorien bzw. Umbenennen von Kategorien

Sie können leere Kategorien erstellen, um Konzepte und Typen hinzuzufügen. Außerdem können Sie die Kategorien umbenennen.

Abbildung 6-14
Dialogfeld "Kategorieeigenschaften"



So erstellen Sie eine neue, leere Kategorie:

- ▶ Rufen Sie den Fensterbereich "Kategorien" auf.
- ▶ Wählen Sie in den Menüs die Optionsfolge Kategorien > Leere Kategorie erstellen. Das Dialogfeld wird geöffnet.
- ▶ Geben Sie im Namensfeld einen Namen für diese Kategorie ein.

- ▶ Klicken Sie auf OK, um den Namen zu übernehmen und das Dialogfeld zu schließen. Das Dialogfeld wird geschlossen und im Fensterbereich wird ein neuer Kategorienname angezeigt.

Sie können nun weitere Elemente in diese Kategorie aufnehmen. [Für weitere Informationen siehe Thema Hinzufügen von Deskriptoren zu Kategorien auf S. 161.](#)

So benennen Sie eine Kategorie um:

- ▶ Wählen Sie eine Kategorie aus und wählen Sie die Optionsfolge Kategorien > Kategorie umbenennen. Das Dialogfeld wird geöffnet.
- ▶ Geben Sie im Namensfeld einen neuen Namen für diese Kategorie ein.
- ▶ Klicken Sie auf OK, um den Namen zu übernehmen und das Dialogfeld zu schließen. Das Dialogfeld wird geschlossen und im Fensterbereich wird ein neuer Kategorienname angezeigt.

Erstellen von Kategorien durch Ziehen und Ablegen (Drag&Drop)

Das Drag&Drop-Verfahren erfolgt manuell und beruht nicht auf Algorithmen. Sie können Kategorien im Bereich "Kategorien" erstellen, indem Sie folgende Elemente ziehen:

- Extrahierte Konzepte, Typen oder Muster aus dem Bereich "Extrahierungsergebnisse" in den Bereich "Kategorien".
- Extrahierte Konzepte aus dem Bereich "Daten" in den Bereich "Kategorien".
- Ganze Zeilen aus dem Bereich "Daten" in den Bereich "Kategorien". Damit wird eine Kategorie erstellt, die aus allen extrahierten Konzepten und Mustern in dieser Zeile besteht.

Anmerkung: Der Bereich "Extrahierungsergebnisse" unterstützt die Mehrfachauswahl, um das Ziehen und Ablegen mehrerer Elemente zu vereinfachen.

Wichtig: Sie können keine Konzepte aus dem Bereich "Daten" ziehen, wenn diese nicht aus dem Text extrahiert wurden. Wenn Sie die Extrahierung eines Konzepts erzwingen möchten, das Sie in Ihren Daten gefunden haben, müssen Sie das Konzept einem Typ hinzufügen. Führen Sie dann die Extrahierung erneut aus. Die neuen Extrahierungsergebnisse enthalten das soeben hinzugefügte Konzept. Sie können es dann in Ihrer Kategorie verwenden. [Für weitere Informationen siehe Thema Konzepte zu Typen hinzufügen in Kapitel 5 auf S. 91.](#)

So erstellen Sie Kategorien durch Ziehen und Ablegen:

- ▶ Wählen Sie im Bereich "Extrahierungsergebnisse" oder im Datenbereich eines oder mehrere Konzepte, Muster, Typen, Datensätze oder Datensatzteile aus.
- ▶ Ziehen Sie das Element bei gedrückter Maustaste in eine bestehende Kategorie oder in den Fensterbereich, um eine neue Kategorie anzulegen.
- ▶ Wenn Sie den Bereich erreicht haben, in dem Sie das Element ablegen möchten, lassen Sie die Maustaste los. Das Element wird dem Bereich "Kategorien" hinzugefügt. Die geänderten Kategorien werden mit einer besonderen Hintergrundfarbe gekennzeichnet. Diese Farbe ist der Hintergrund für Kategoriefeedback. [Für weitere Informationen siehe Thema Festlegen von Optionen in Kapitel 2 auf S. 19.](#)

Anmerkung: Die resultierende Kategorie wurde automatisch benannt. Wenn Sie einen Namen ändern möchten, können Sie die Kategorie umbenennen. [Für weitere Informationen siehe Thema Bearbeiten von Kategorieeigenschaften auf S. 160.](#)

Um anzuzeigen, welche Datensätze einer Kategorie zugewiesen wurden, wählen Sie die betreffende Kategorie im Fensterbereich “Kategorieen” aus. Der Datenbereich wird automatisch aktualisiert und zeigt alle Datensätze für diese Kategorie an. Wenn Sie das vollständige Antwort-Set für eine Frage sehen möchten, wählen Sie den Knoten Alle Datensätze am Beginn des Kategoriebaums aus.

Importieren und Exportieren von vordefinierten Kategorien

Wenn Sie Ihre eigenen Kategorien in einer Microsoft Excel-Datei (*.xls, *.xlsx) gespeichert haben, können Sie sie in IBM® SPSS® Text Analytics for Surveys importieren.

Sie können die vorhandenen Kategorien in einem geöffneten Projekt in eine Microsoft Excel-Datei (*.xls, *.xlsx) exportieren. Wenn Sie Ihre Kategorien exportieren, können Sie wie erforderlich einige zusätzliche Informationen einschließen oder ausschließen, z. B. Deskriptoren und Werte. [Für weitere Informationen siehe Thema Exportieren von Kategorien auf S. 145.](#)

Wichtig: Ab SPSS Text Analytics for Surveys Version 4.0.1 wurde die Verwendung von Coderahmen größtenteils durch vordefinierte Kategorien ersetzt. So wurde beispielsweise der Assistent für den Coderahmen-Import durch den Assistenten für den Import vordefinierter Kategorien ersetzt. Mit diesem neuen Assistenten können Sie jedoch weiterhin etwaige bestehende Coderahmen importieren. Außerdem wird der Coderahmenmanager nicht mehr unterstützt. Wählen Sie zur Bearbeitung von Codewerten die Menüoptionen Anzeigen > Kategoriecodes aus, um die Spalte “Code” im Fensterbereich “Kategorie” anzuzeigen, in der Sie alle Codes nach Bedarf bearbeiten können.

Wenn Ihre vordefinierten Kategorien keine Codes haben bzw. Sie neue Codes wünschen, können Sie automatisch ein neues Set an Codes für das Kategorienset im Bereich “Kategorien” generieren, indem Sie aus den Menüs die Befehle Kategorien > Kategorien verwalten > Codes automatisch generieren wählen. Damit werden alle bestehenden Codes entfernt und automatisch neu nummeriert.

Importieren vordefinierter Kategorien

Sie können vordefinierte Kategorien in IBM® SPSS® Text Analytics for Surveys importieren. Stellen Sie vor dem Import sicher, dass sich die vordefinierte Kategoriedatei in einer Microsoft Excel-Datei (*.xls, *.xlsx) befindet und in einem der unterstützten Formate strukturiert ist. Sie

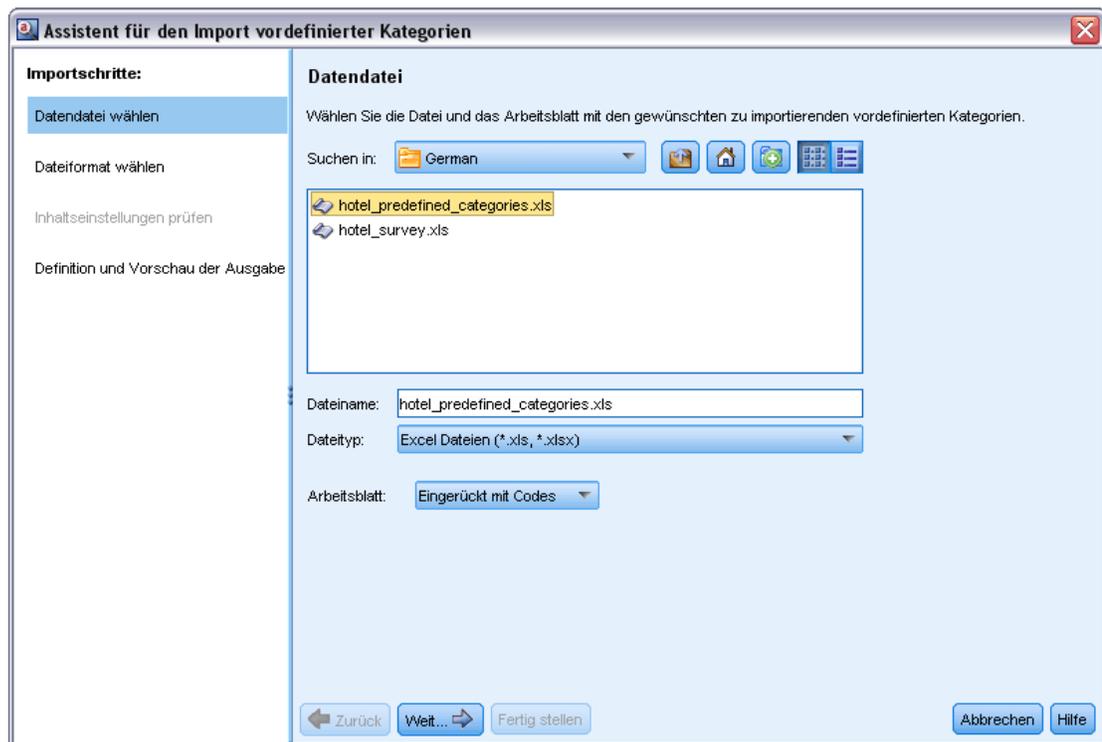
können auch wählen, dass das Produkt das Format automatisch erkennen soll. Die folgenden Formate werden unterstützt:

- Einfaches Listenformat: [Für weitere Informationen siehe Thema Einfaches Listenformat auf S. 141.](#)
- Kompaktes Format: [Für weitere Informationen siehe Thema Kompaktes Format auf S. 142.](#)
- Eingerücktes Format: [Für weitere Informationen siehe Thema Eingerücktes Format auf S. 144.](#)

So importieren Sie vordefinierte Kategorien:

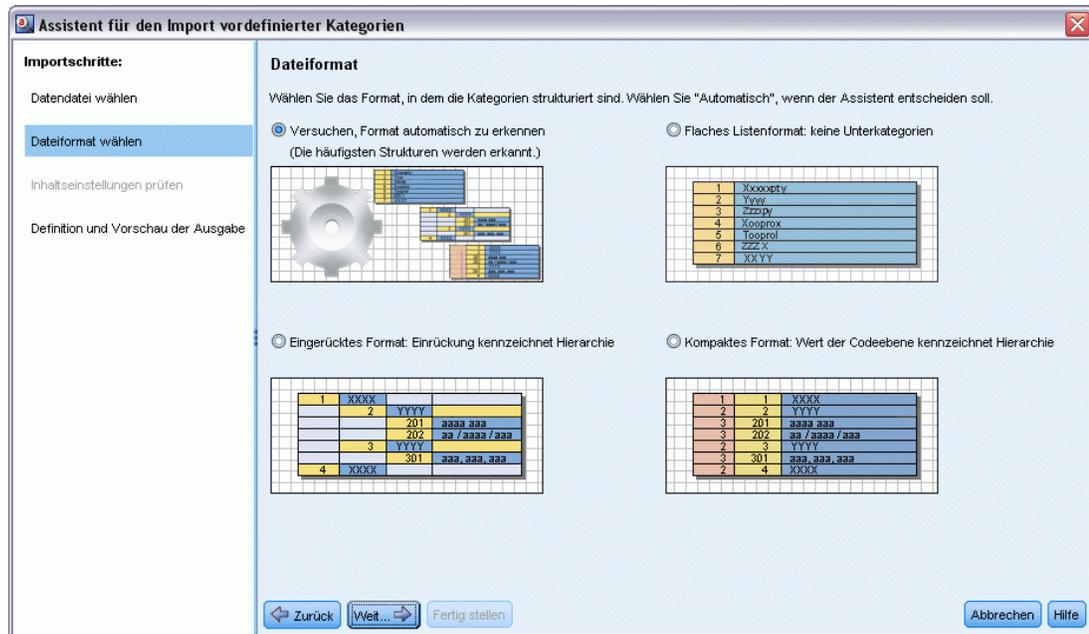
- ▶ Wählen Sie in den Menüs die Optionsfolge Kategorien > Kategorien verwalten > Vordefinierte Kategorien importieren. Ein Assistent für den Import vordefinierter Kategorien wird angezeigt.

Abbildung 6-15
Assistent für den Import vordefinierter Kategorien



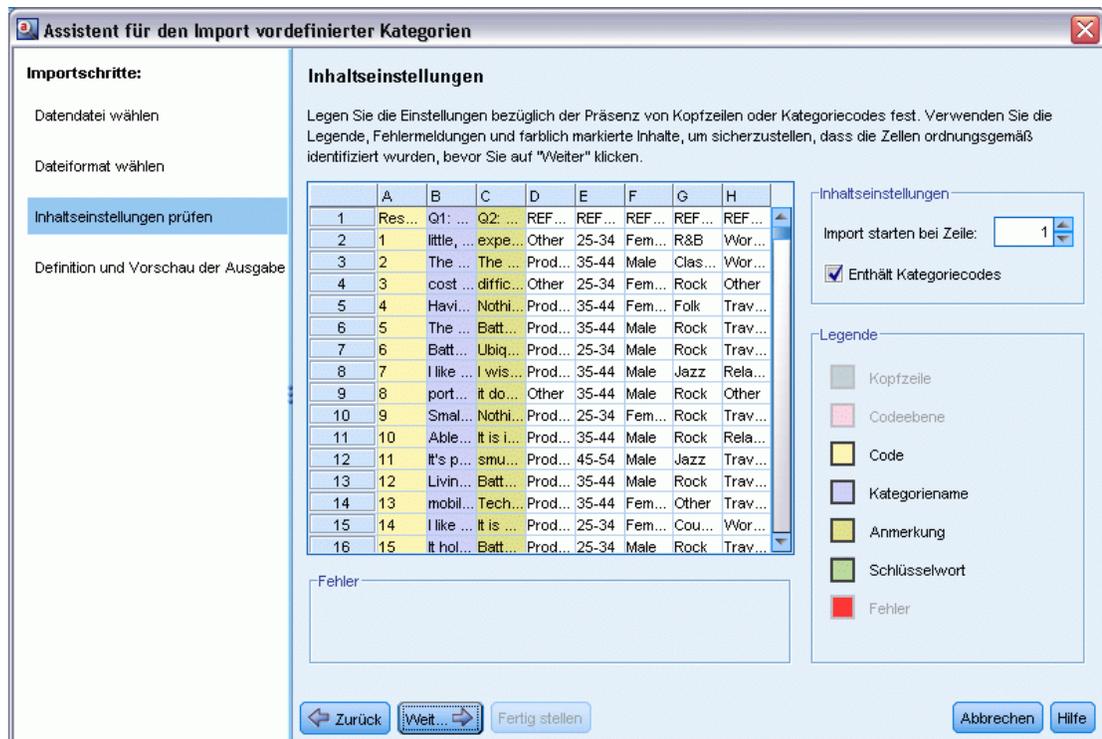
- ▶ Wählen Sie in der Dropdown-Liste “Suchen in” das Laufwerk und den Ordner aus, in dem sich die Datei befindet.
- ▶ Wählen Sie die Datei aus der Liste aus. Der Name der Datei wird im Feld “Dateiname” angezeigt.
- ▶ Wählen Sie aus der Liste das Arbeitsblatt aus, das die vordefinierten Kategorien enthält. Der Name des Arbeitsblatts wird im Feld “Arbeitsblatt” angezeigt.
- ▶ Klicken Sie auf Weiter, um zu beginnen, das Datenformat zu wählen.

Abbildung 6-16
Dialogfeld "Vordefinierte Kategorien importieren," Schritt "Datenformat"



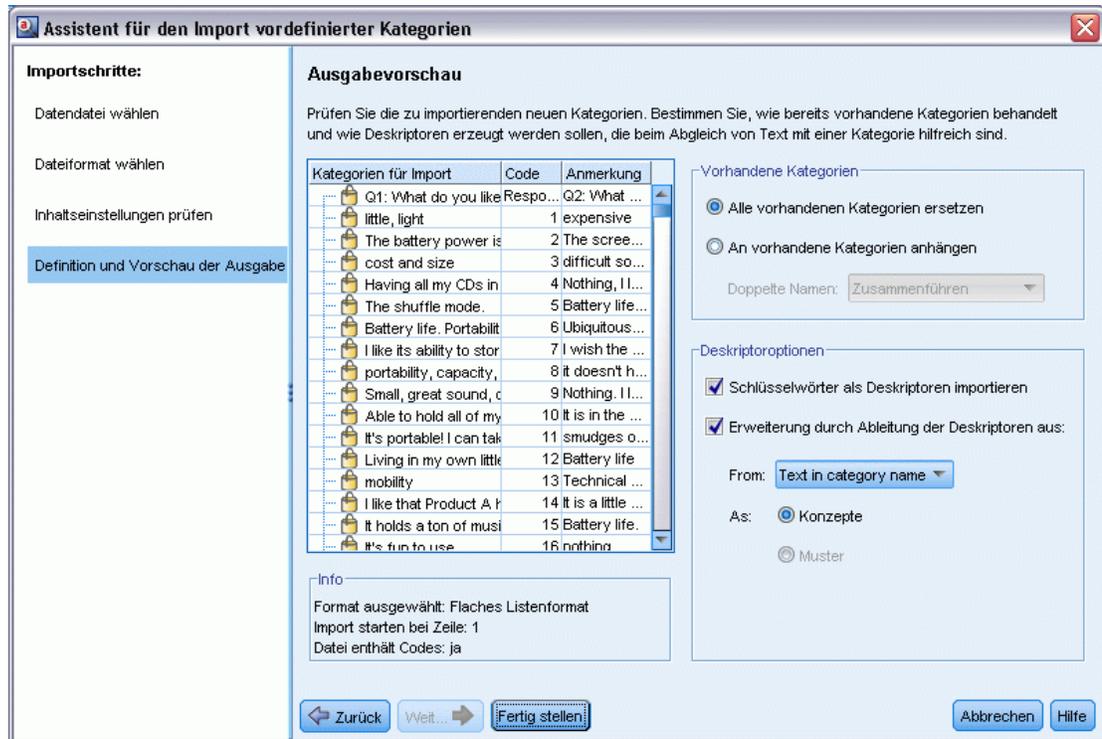
- ▶ Wählen Sie das Format für Ihre Datei oder wählen Sie die Option, damit das Produkt versucht, das Format automatisch zu erkennen. Die automatische Erkennung funktioniert am besten bei den häufigsten Formaten.
 - Einfaches Listenformat: [Für weitere Informationen siehe Thema Einfaches Listenformat auf S. 141.](#)
 - Kompaktes Format: [Für weitere Informationen siehe Thema Kompaktes Format auf S. 142.](#)
 - Eingerücktes Format: [Für weitere Informationen siehe Thema Eingerücktes Format auf S. 144.](#)
- ▶ Klicken Sie auf Weiter, um die zusätzlichen Importoptionen zu definieren. Wenn Sie die automatische Formaterkennung wählen, werden Sie zum letzten Schritt geführt.

Abbildung 6-17
 "Vordefinierte Kategorien importieren", Schritt "Importoptionen"



- ▶ Wenn eine oder mehrere Zeilen Spaltentitel oder andere irrelevante Informationen enthalten, wählen Sie in der Option Import beginnen bei Zeile die Zeilennummer aus, bei der Sie mit dem Import beginnen möchten. Wenn Ihre Kategorienamen beispielsweise in Zeile 7 beginnen, müssen Sie die Ziffer 7 für diese Option eingeben, damit die Datei korrekt importiert wird.
- ▶ Wenn Ihre Datei Kategoriecodes enthält, wählen Sie die Option Enthält Kategoriecodes. Damit unterstützen Sie den Assistenten bei der korrekten Erkennung Ihrer Daten.
- ▶ Überprüfen Sie die farbkodierten Zellen und die Legende, um sicherzustellen, dass die Daten korrekt identifiziert wurden. Etwaige in der Datei erkannte Fehler werden in Rot angezeigt und erhalten einen Verweis unter der Formatvorschautabelle. Wenn das falsche Format ausgewählt wurde, gehen Sie zurück und wählen Sie ein anderes Format aus. Falls Sie Korrekturen in Ihrer Datei durchführen müssen, korrigieren Sie die Datei und starten Sie dann den Assistenten neu, indem Sie die Datei erneut auswählen. Sie müssen alle Fehler korrigieren, bevor Sie den Assistenten beenden können.
- ▶ Klicken Sie auf Weiter, um das Set an Kategorien und Unterkategorien zu sehen, das importiert wird, und um zu definieren, wie Deskriptoren für diese Kategorien erstellt werden.

Abbildung 6-18
Dialogfeld "Vordefinierte Kategorien importieren", Schritt "Vorschau"



- ▶ Prüfen Sie das Kategorienset, das in die Tabelle importiert wird. Wenn Sie die als Deskriptoren erwarteten Stichwörter nicht sehen, wurden diese eventuell beim Import nicht erkannt. Stellen Sie sicher, dass sie über das korrekte Präfix verfügen und in der korrekten Zelle erscheinen.
- ▶ Wählen Sie, wie bereits existierende Kategorien in Ihrem Projekt gehandhabt werden.
 - Alle vorhandenen Kategorien ersetzen. Diese Option löscht alle existierenden Kategorien. Anschließend werden die importierten Kategorien alleine an deren Stelle verwendet.
 - An vorhandene Kategorien anhängen. Diese Option importiert die Kategorien und führt alle häufigen Kategorien mit den vorhandenen Kategorien zusammen. Beim Hinzufügen zu vorhandenen Kategorien müssen Sie festlegen, wie Duplikate behandelt werden sollen. Eine Option (Zusammenführen) ist es, alle importierten Kategorien mit bestehenden Kategorien zusammenzuführen, wenn sie einen gemeinsamen Kategorienamen haben. Eine andere Option (Aus Import ausschließen) ist es, den Import von Kategorien zu untersagen, wenn eine Kategorie mit demselben Namen vorhanden ist.
- ▶ Stichwörter als Deskriptoren importieren ist eine Option für den Import der Stichwörter, die in Ihren Daten als Stichwörter für die verknüpfte Kategorie identifiziert werden.
- ▶ Kategorien durch abgeleitete Deskriptoren erweitern ist eine Option, die Deskriptoren aus den Wörtern erzeugt, die den Namen der Kategorie oder Unterkategorie und/oder die Wörter repräsentieren, aus denen die Anmerkung besteht. Wenn die Wörter extrahierten Ergebnissen entsprechen, werden diese als Deskriptoren zur Kategorie hinzugefügt. Diese Option erzeugt die besten Ergebnisse, wenn die Kategorienamen oder Anmerkungen lang und beschreibend

sind. Dies ist eine schnelle Methode, um Kategoriedeskriptoren zu erzeugen, mit deren Hilfe die Kategorie Datensätze erfassen kann, die diese Deskriptoren enthalten.

- Im Feld Von können Sie wählen, von welchem Text die Deskriptoren abgeleitet werden, den Namen der Kategorien und Unterkategorien und/oder den Wörtern in den Anmerkungen.
 - Im Feld As können Sie wählen, ob diese Deskriptoren in der Form von Konzepten oder TLA-Mustern erstellt werden. Wenn keine TLA-Extrahierung erfolgt ist, sind die Optionen für Muster in diesem Assistenten deaktiviert.
- Klicken Sie auf Fertigstellen, um die vordefinierten Kategorien in den Bereich “Kategorien” zu importieren.

Einfaches Listenformat

In diesem einfachen Listenformat befindet sich nur eine erste Ebene von Kategorien ohne Unterkategorien oder Unternetzen. Kategoriennamen befinden sich in einer einzelnen Spalte.

Abbildung 6-19
Beispiel für ein einfaches Listenformat

	A	B	C
1	1	Unterkunft	
2	11	Unterkunft/Ausstattung	Kommentare über die Einrichtung
3		_gut eingerichtet	
4	12	Unterkunft/Sauberkeit	
5	13	Unterkunft/Größe	
6	14	Unterkunft/Ruhe	
7		_ruhig	
8	15	Unterkunft/Komfort	Kommentare die darauf hinweisen, dass der Ort bequem ist
9		_klimaanlage	
10		_spa	
11	2	Lage	Bemerkungen in Bezug auf die Nähe zur Stadtmitte, zum Strand, Meer oder Angabe, dass der Ort ideal gelegen ist
12	3	Bedienung	Kommentare über das Personal
13		_freundlich	
14		_hilffreich	
15		_höflich	
16	4	Essen	Bemerkungen in Bezug auf Essen, Küche
17		_regionalküche	
18		_büffet	
19		_frühstück	
20	5	Preis	Kommentare über die Kosten
21		_preis-Leistungsverhältnis	

Die folgenden Informationen können sich in einer Datei dieses Formats befinden:

- Die optionale Spalte **Codes** enthält numerische Werte, die eine Kategorie eindeutig identifizieren. Wenn Sie angeben, dass die Datendatei Codes enthält (Option Enthält Kategoriecodes im Schritt Inhaltseinstellungen), muss eine Spalte mit eindeutigen Codes für jede Kategorie in der Zelle direkt links neben dem Kategoriennamen vorhanden sein. Wenn Ihre Daten keine Codes enthalten, Sie jedoch später Codes erstellen möchten, können Sie diese jederzeit tun (Kategorien > Kategorien verwalten > Codes automatisch erzeugen). Sie können Codes zu einem späteren Zeitpunkt bearbeiten, indem Sie die Optionsfolge Anzeigen > Kategoriecode wählen; die Codes werden im Fensterbereich “Kategorie” in einer Codespalte angezeigt, wo Sie sie manuell abändern können.
- Eine *erforderliche* Spalte **Kategoriennamen** enthält alle Namen der Kategorien. Diese Spalte ist erforderlich, um einen Import mit diesem Format durchzuführen.

- Optionale **Anmerkungen** in der Zelle direkt rechts neben dem Kategorienamen. Diese Anmerkung besteht aus Text, der Ihre Kategorien/Unterkategorien beschreibt.
- Optionale **Stichwörter** können als Deskriptoren für Kategorien importiert werden. Damit sie erkannt werden, müssen diese Stichwörter in der Zelle direkt unter dem verknüpften Kategorie- bzw. Unterkategorienamen vorhanden sein und der Liste der Stichwörter muss ein Unterstrich (_) vorangestellt sein, wie beispielsweise “_Feuerwaffen, Waffen/Gewehre”. Die Stichwortzelle kann ein oder mehrere Wörter enthalten, die zur Beschreibung jeder Kategorie verwendet werden. Diese Wörter werden als Deskriptoren importiert oder ignoriert, abhängig von Ihrer Angabe im letzten Schritt des Assistenten. Später werden Deskriptoren mit den extrahierten Ergebnissen aus dem Text verglichen. Liegt eine Übereinstimmung vor, wird der Datensatz bzw. das Dokument der Kategorie zugewiesen, die diesen Deskriptor enthält.

Tabelle 6-3

Einfaches Listenformat mit Codes, Stichwörtern und Anmerkungen

Spalte A	Spalte B	Spalte C
Kategoriecode (<i>optional</i>)	Kategorienname	Anmerkung
	_Deskriptor-/Stichwortliste (<i>optional</i>)	

Kompaktes Format

Das kompakte Format ist ähnlich strukturiert wie das einfache Listenformat, allerdings wird das kompakte Format mit hierarchischen Kategorien verwendet. Daher ist eine Spalte für die Codeebene erforderlich, um die hierarchische Stufe jeder Kategorie und Unterkategorie zu definieren.

Abbildung 6-20

Beispiel einer kompakten, vordefinierten Kategoriedatei in Microsoft Excel

	A	B	C	D	E
1	1	1	Unterkunft		
2	2	11	Ausstattung	Kommentare über die Einrichtung	
3			_gut eingerichtet		
4	2	12	Sauberkeit		
5	2	13	Größe		
6	2	14	Ruhe		
7			_ruhig		
8	2	15	Komfort	Kommentare die darauf hinweisen, dass der Ort bequem ist	
9			_Himaanlage		
10			_spa		
11	1	2	Lage	Bemerkungen in Bezug auf die Nähe zur Stadtmitte, zum Strand, Meer oder Angabe, dass der Ort ideal gelegen ist	
12	1	3	Bedienung	Kommentare über das Personal	
13			_freundlich		
14			_hilfreich		
15			_höflich		
16	1	4	Essen	Bemerkungen in Bezug auf Essen, Küche	
17			_regionalküche		
18			_buffet		
19			_frühstück		
20	1	5	Preis	Kommentare über die Kosten	
21			_preis-leistungsverhältnis		

Die folgenden Informationen können sich in einer Datei dieses Formats befinden:

- Die *erforderliche* Spalte **Code-Ebene** enthält Zahlen, die die hierarchische Position für die nachfolgenden Informationen in dieser Zeile enthalten. Beispiel: Wenn die Werte 1, 2 oder 3 angegeben sind und Sie sowohl Kategorien als auch Unterkategorien haben, steht 1 für Kategorien, 2 für Unterkategorien und 3 für Unter-Unterkategorien. Wenn Sie nur Kategorien und Unterkategorien haben, steht 1 für Kategorien und 2 für Unterkategorien. Das kann beliebig weitergeführt werden bis zur gewünschten Kategorientiefe.
- Die optionale Spalte **Codes** enthält Werte, die eine Kategorie eindeutig identifizieren. Wenn Sie angeben, dass die Datendatei Codes enthält (Option Enthält Kategoriecodes im Schritt Inhaltseinstellungen), muss eine Spalte mit eindeutigen Codes für jede Kategorie in der Zelle direkt links neben dem Kategorienamen vorhanden sein. Wenn Ihre Daten keine Codes enthalten, Sie jedoch später Codes erstellen möchten, können Sie diese jederzeit tun (Kategorien > Kategorien verwalten > Codes automatisch erzeugen). Sie können Codes zu einem späteren Zeitpunkt bearbeiten, indem Sie die Optionsfolge Anzeigen > Kategoriecode wählen; die Codes werden im Fensterbereich "Kategorie" in einer Codespalte angezeigt, wo Sie sie manuell abändern können.
- Eine *erforderliche* Spalte **Kategorienamen** enthält alle Namen der Kategorien und Unterkategorien. Diese Spalte ist erforderlich, um einen Import mit diesem Format durchzuführen.
- Optionale **Anmerkungen** in der Zelle direkt rechts neben dem Kategorienamen. Diese Anmerkung besteht aus Text, der Ihre Kategorien/Unterkategorien beschreibt.
- Optionale **Stichwörter** können als Deskriptoren für Kategorien importiert werden. Damit sie erkannt werden, müssen diese Stichwörter in der Zelle direkt unter dem verknüpften Kategorie- bzw. Unterkategorienamen vorhanden sein und der Liste der Stichwörter muss ein Unterstrich (_) vorangestellt sein, wie beispielsweise "_Feuerwaffen, Waffen/Gewehre". Die Stichwortzelle kann ein oder mehrere Wörter enthalten, die zur Beschreibung jeder Kategorie verwendet werden. Diese Wörter werden als Deskriptoren importiert oder ignoriert, abhängig von Ihrer Angabe im letzten Schritt des Assistenten. Später werden Deskriptoren mit den extrahierten Ergebnissen aus dem Text verglichen. Liegt eine Übereinstimmung vor, wird der Datensatz bzw. das Dokument der Kategorie zugewiesen, die diesen Deskriptor enthält.

Tabelle 6-4

Beispiel für kompaktes Format mit Codes

Spalte A	Spalte B	Spalte C
Hierarchische Code-Ebene	Kategoriecode (<i>optional</i>)	Kategoriename
Hierarchische Code-Ebene	Unterkategoriecode (<i>optional</i>)	Unterkategoriename

Tabelle 6-5

Beispiel für kompaktes Format ohne Codes

Spalte A	Spalte B
Hierarchische Code-Ebene	Kategoriename
Hierarchische Code-Ebene	Unterkategoriename

Eingerücktes Format

Im Dateiformat “Eingerückt” ist der Inhalt hierarchisch strukturiert, d. h., die Datei enthält Kategorien und eine oder mehrere Ebenen von Unterkategorien. Darüber hinaus ist die Struktur ihrer Hierarchie entsprechend eingerückt. Jede Zeile in der Datei enthält entweder eine Kategorie oder eine Unterkategorie, dabei sind Unterkategorien weiter eingerückt als die Kategorien, Unter-Unterkategorien sind weiter eingerückt als die Unterkategorien usw. Sie können diese Struktur manuell in Microsoft Excel erstellen oder eine Struktur verwenden, die aus einem anderen Produkt exportiert und in einem Microsoft Excel-Format gespeichert wurde.

Abbildung 6-21
Beispiel einer eingerückten Kategorie in Microsoft Excel

- **Kategoriecodes und Kategorienamen der ersten Ebene** belegen die Spalten A bzw. B. Falls keine Codes vorhanden sind, befindet sich der Kategorienamen in Spalte A.
- **Unterkategoriecodes und Unterkategorienamen** belegen die Spalten B bzw. C. Falls keine Codes vorhanden sind, befindet sich der Unterkategorienamen in Spalte B. Die Unterkategorie gehört einer Kategorie an. Unterkategorien können nur vorhanden sein, wenn Kategorien in der ersten Ebene vorhanden sind.

Tabelle 6-6
Eingerückte Struktur mit Codes

Spalte A	Spalte B	Spalte C	Spalte D
Kategoriecode (optional)	Kategorienamen		
	Unterkategoriecode (optional)	Unterkategorienamen	
		Unter-Unterkategoriecode (optional)	Unter-Unterkategorienamen

Tabelle 6-7
Eingerückte Struktur ohne Codes

Spalte A	Spalte B	Spalte C
Kategorienname		
	Unterkategorienname	
		Unter-Unterkategorienname

Die folgenden Informationen können sich in einer Datei dieses Formats befinden:

- Optionale **Codes** müssen Werte sein, die jede Kategorie oder Unterkategorie eindeutig identifizieren. Wenn Sie angeben, dass die Datendatei Codes enthält (Option Enthält Kategoriecodes im Schritt Inhaltseinstellungen), muss in der Zelle direkt links neben dem Kategorie-/Unterkategorienamen ein eindeutiger Code für jede Kategorie bzw. Unterkategorie vorhanden sein. Wenn Ihre Daten keine Codes enthalten, Sie jedoch später Codes erstellen möchten, können Sie diese jederzeit tun (Kategorien > Kategorien verwalten > Codes automatisch erzeugen). Sie können Codes zu einem späteren Zeitpunkt bearbeiten, indem Sie die Optionsfolge Anzeigen > Kategoriecode wählen; die Codes werden im Fensterbereich "Kategorie" in einer Codespalte angezeigt, wo Sie sie manuell abändern können.
- Ein *erforderlicher Name* für jede Kategorie und Unterkategorie. Unterkategorien müssen um eine Zelle nach rechts unter den Kategorien in einer separaten Zeile eingerückt sein.
- Optionale **Anmerkungen** in der Zelle direkt rechts neben dem Kategorienamen. Diese Anmerkung besteht aus Text, der Ihre Kategorien/Unterkategorien beschreibt.
- Optionale **Stichwörter** können als Deskriptoren für Kategorien importiert werden. Damit sie erkannt werden, müssen diese Stichwörter in der Zelle direkt unter dem verknüpften Kategorie- bzw. Unterkategorienamen vorhanden sein und der Liste der Stichwörter muss ein Unterstrich (_) vorangestellt sein, wie beispielsweise "_Feuerwaffen, Waffen/Gewehre". Die Stichwortzelle kann ein oder mehrere Wörter enthalten, die zur Beschreibung jeder Kategorie verwendet werden. Diese Wörter werden als Deskriptoren importiert oder ignoriert, abhängig von Ihrer Angabe im letzten Schritt des Assistenten. Später werden Deskriptoren mit den extrahierten Ergebnissen aus dem Text verglichen. Liegt eine Übereinstimmung vor, wird der Datensatz bzw. das Dokument der Kategorie zugewiesen, die diesen Deskriptor enthält.

Wichtig: Wenn Sie auf einer Ebene einen Code verwenden, müssen Sie einen Code für jede Kategorie und Unterkategorie angeben. Andernfalls schlägt der Importvorgang fehl.

Exportieren von Kategorien

Sie können die vorhandenen Kategorien in einem geöffneten Projekt in ein Microsoft Excel-Dateiformat (*.xls, *.xlsx) exportieren. Die exportierten Daten stammen im Wesentlichen aus dem aktuellen Inhalt des Bereichs "Kategorie" bzw. aus den Kategorieeigenschaften. Daher wird ein erneutes Scoring empfohlen, wenn Sie auch den Score-Wert Docs. exportieren möchten.

Immer exportieren...

- Kategoriecodes, falls vorhanden
- Kategorienamen (und Unterkategorienamen)
- Codeebenen, falls vorhanden (*Flaches/Kompaktes* Format)
- Spaltenüberschriften (*Flaches/Kompaktes* Format)

Optional exportieren...

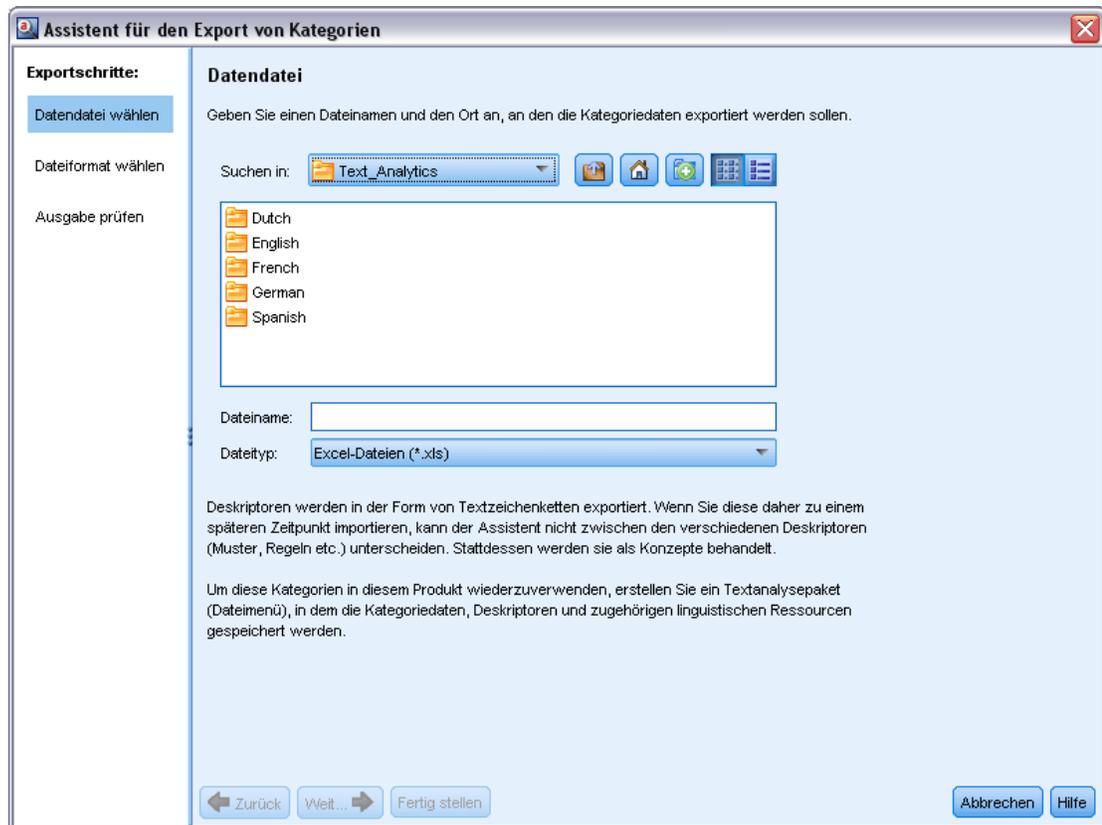
- Docs.-Scores
- Kategorieanmerkungen
- Deskriptornamen
- Deskriptoranzahl

Wichtig: Wenn Sie Deskriptoren exportieren, werden diese in Textzeichenfolgen umgewandelt und ihnen wird ein Unterstrich vorangestellt. Wenn Sie diese erneut in dieses Produkt importieren, kann nicht mehr zwischen Deskriptoren, bei denen es sich um Muster, Kategorieregeln oder reguläre Konzepte handelt, unterschieden werden. Wenn Sie diese Kategorien in diesem Produkt erneut verwenden möchten, wird dringend empfohlen, stattdessen eine Textanalysepaket- (TAP-) Datei zu erstellen, da das TAP-Format alle aktuell definierten Deskriptoren sowie all Ihre Kategorien, Codes und die verwendeten linguistischen Ressourcen speichert. TAP-Dateien können sowohl in IBM® SPSS® Text Analytics als auch in IBM® SPSS® Text Analytics for Surveys verwendet werden. [Für weitere Informationen siehe Thema Verwendung von Text Analysis Packages in Kapitel 3 auf S. 43.](#)

So exportieren Sie vordefinierte Kategorien:

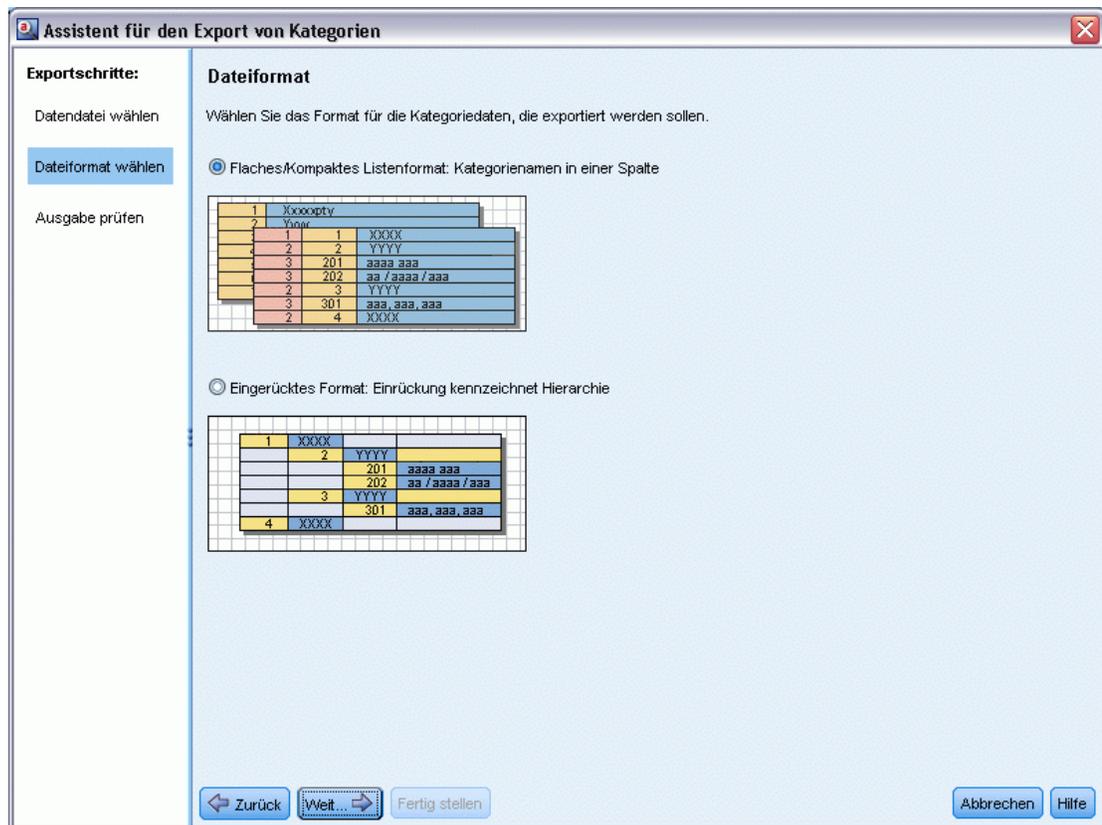
- Wählen Sie in den Menüs die Optionsfolge Kategorien > Kategorien verwalten > Kategorien exportieren. Ein Assistent für den Export von Kategorien wird angezeigt.

Abbildung 6-22
Assistent "Kategorien exportieren," Schritt 1



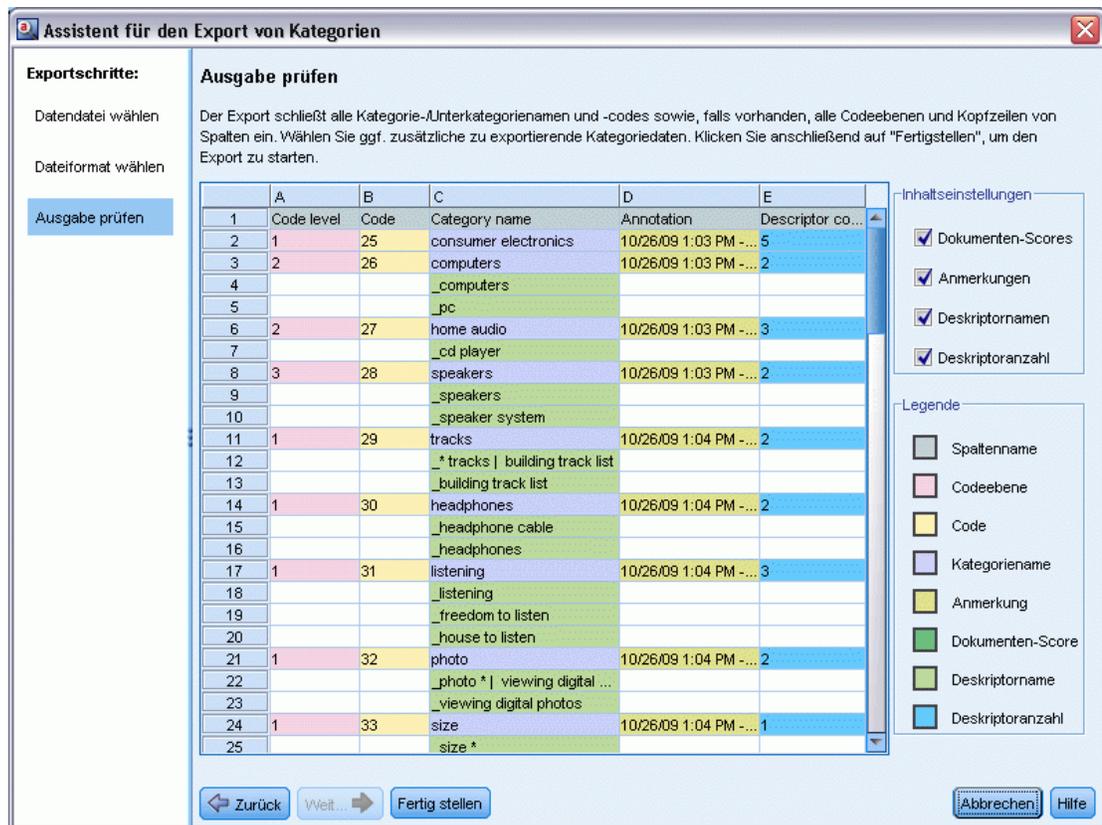
- Wählen Sie den Speicherort aus und geben Sie den Namen der Datei ein, die exportiert wird.
- Geben Sie einen Namen für die Ausgabedatei in das Textfeld "Dateiname" ein.
- Klicken Sie auf Weiter, um das Format zu wählen, in das Sie Ihre Kategoriendaten exportieren möchten.

Abbildung 6-23
Assistent "Kategorien exportieren," Schritt 2



- ▶ Wählen Sie das Format aus den folgenden Optionen:
 - Einfaches oder kompaktes Listenformat: [Für weitere Informationen siehe Thema Einfaches Listenformat auf S. 141.](#) Einfache Liste enthält keine Unterkategorien. [Für weitere Informationen siehe Thema Kompaktes Format auf S. 142.](#) Kompaktes Listenformat enthält hierarchische Kategorien.
 - Eingerücktes Format: [Für weitere Informationen siehe Thema Eingerücktes Format auf S. 144.](#)
- ▶ Klicken Sie auf Weiter, um mit der Auswahl des zu exportierenden Inhalts zu beginnen und die vorgeschlagenen Daten zu prüfen.

Abbildung 6-24
Assistent "Kategorien exportieren," Schritt 3



- ▶ Prüfen Sie den Inhalt für die exportierte Datei.
- ▶ Aktivieren oder deaktivieren Sie zusätzliche zu exportierende Inhaltseinstellungen wie Anmerkungen oder Deskriptornamen..
- ▶ Klicken Sie auf Fertigstellen, um die Kategorien zu exportieren.

Verwenden von Kategorieregeln

Es gibt verschiedene Methoden zum Erstellen von Kategorien. Eine dieser Methoden ist die Definition von Kategorieregeln, die Ideen ausdrücken. Kategorieregeln sind Anweisungen, mit denen Datensätze auf Basis eines logischen Ausdrucks mithilfe von extrahierten Konzepten, Typen und Mustern sowie von Boole'schen Operatoren automatisch einer Kategorie zugewiesen werden. Sie könnten beispielsweise einen Ausdruck schreiben, der bedeutet *Schließe alle Datensätze, die das extrahierte Konzept Botschaft enthalten, nicht jedoch Argentinien, in diese Kategorie ein.*

Während manche Kategorieregeln beim Aufbau von Kategorien mithilfe von Gruppierungstechniken wie *Kookkurrenz* und *Konzeptwurzelableitung* (Kategorien > Aufbaueinstellungen > Erweiterte Einstellungen: Linguistik) automatisch erzeugt werden, können Sie Kategorieregeln auch manuell im Regeleditor erstellen, indem Sie Ihr Kategorieverständnis der

Daten und des Kontexts zu Rate ziehen. Jede Regel wird an eine einzelne Kategorie angehängt, so dass jedes jeder Datensatz, der mit dieser Regel übereinstimmt, über das Scoring dieser Kategorie zugewiesen wird.

Kategorieregeln helfen dabei, die Qualität und Produktivität Ihrer Text-Mining-Ergebnisse und weiterer quantitativer Analysen zu verbessern, indem sie es Ihnen ermöglichen, Antworten mit höherer Spezifität zu kategorisieren. Ihre Erfahrung und Ihr geschäftliches Wissen ermöglichen Ihnen unter Umständen ein spezifisches Verständnis Ihrer Daten und des Kontexts. Dieses Verständnis können Sie nutzen, um dieses Wissen in Kategorieregeln umzusetzen, um so Ihre Datensätze noch effizienter und genauer zu kategorisieren, indem Sie extrahierte Elemente mit Boole'scher Logik kombinieren.

Die Möglichkeit zur Erstellung dieser Regeln verbessert die Kodierungsgenauigkeit, Effizienz und Produktivität, indem sie es Ihnen ermöglicht, Ihr Geschäftswissen in die Extrahierungstechnologie des Produkts einzubringen.

Anmerkung: Beispiele für die Übereinstimmung von Regeln mit Text finden Sie unter [Beispiele für Kategorieregeln auf S. 155](#)

Kategorieregelsyntax

Während manche Kategorieregeln beim Aufbau von Kategorien mithilfe von Gruppierungstechniken wie *Kookkurrenz* und *Konzeptwurzelableitung* (Kategorien > Aufbaueinstellungen > Erweiterte Einstellungen: Linguistik) automatisch erzeugt werden, können Sie Kategorieregeln auch manuell im Regeleditor erstellen. Jede Regel ist ein Deskriptor einer einzelnen Kategorie, daher wird jedes jeder Datensatz, der mit dieser Regel übereinstimmt, automatisch über das Scoring dieser Kategorie zugewiesen.

Anmerkung: Beispiele für die Übereinstimmung von Regeln mit Text finden Sie unter [Beispiele für Kategorieregeln auf S. 155](#)

Beim Erstellen oder Bearbeiten einer Regel muss diese im Regeleditor geöffnet sein. Sie können Konzepte, Typen oder Muster hinzufügen oder Platzhalter verwenden, um die Übereinstimmungsmöglichkeiten zu vergrößern. Wenn Sie extrahierte Konzepte, Typen und Muster verwenden, haben Sie den Vorteil, dass alle verwandten Konzepte gefunden werden.

Wichtig: Um häufig vorkommende Fehler zu vermeiden, sollten Sie die Konzepte direkt per Drag-and-Drop aus dem Fensterbereich "Extrahierungsergebnisse", Bereichen zur Textverknüpfungsanalyse oder dem Bereich "Daten" in den Regeleditor übertragen oder sie über die Kontextmenüs hinzufügen.

Wenn Konzepte, Typen und Muster erkannt werden, erscheint neben dem Text ein Symbol.



Extrahiertes Konzept



Extrahierter Typ



Extrahiertes Muster

Regelsyntax und Operatoren

Die folgende Tabelle enthält die Zeichen, mit deren Hilfe Sie Ihre Regelsyntax definieren. Verwenden Sie diese Zeichen zusammen mit den Konzepten, Typen und Mustern, um Ihre Regel zu erstellen.

Tabelle 6-8
Unterstützte Syntax

Zeichen	Beschreibung
&	Der Boole'sche Operator "und" Beispielsweise enthält $a \ \& \ b$ sowohl <i>a als auch b</i> wie etwa bei: - Invasion & Vereinigte Staaten - 2016 & Olympiade - gut & Apfel
	Der Boole'sche Operator "oder" ist eingeschlossen. Wenn also ein Element oder alle Elemente gefunden werden, erfolgt eine Übereinstimmung. Beispielsweise enthält $a \ \ b$ entweder <i>a oder b</i> wie etwa bei: - Angriff Frankreich - Eigentumswohnung Apartment
! ()	Der Boole'sche Operator "nicht" Beispielsweise enthält $!(a)$ nicht <i>a</i> wie etwa bei $!(\text{gutes} \ \& \ \text{Hotel})$
*	Ein Platzhalter, der je nach Verwendung für alles zwischen einem einzelnen Zeichen und einem ganzen Wort stehen kann. Für weitere Informationen siehe Thema Platzhalter in Kategorieregeln auf S. 153.
()	Ein Trennzeichen für Ausdrücke. Jeder Ausdruck innerhalb der Klammern wird zuerst ausgewertet.
+	Der Musterkonnektor, der zur Bildung eines reihenfolgespezifischen Musters verwendet wird. Wenn vorhanden, müssen eckige Klammern verwendet werden. Für weitere Informationen siehe Thema Verwenden von TLA-Mustern in Kategorieregeln auf S. 151.
[]	Das Trennzeichen für Muster wird benötigt, wenn Sie nach Übereinstimmungen auf der Basis eines extrahierten TLA-Musters innerhalb einer Kategorieregel suchen. Der Inhalt in den eckigen Klammern verweist auf TLA-Muster und wird niemals mit Konzepten oder Typen auf der Basis einfacher Kookkurrenz übereinstimmen. Wenn Sie dieses TLA-Muster nicht extrahiert haben, ist keine Übereinstimmung möglich. Für weitere Informationen siehe Thema Verwenden von TLA-Mustern in Kategorieregeln auf S. 151. Verwenden Sie keine eckigen Klammern, wenn Sie nach Übereinstimmungen von Konzepten und Typen anstelle von Mustern suchen. <i>Anmerkung:</i> In älteren Versionen wurden mithilfe von Kategorieerstellungsmethoden generierte Kookkurrenz- und Synonymregeln in eckige Klammern eingefasst. In allen neuen Versionen zeigen eckige Klammern das Vorhandensein eines TLA-Musters an. Stattdessen stehen durch die Kookkurrenztechnik erzeugte Regeln und Synonyme in runden Klammern, z. B. $(\text{Lautsprecher} \ \ \text{Lautsprechersysteme})$.

Die Operatoren $\&$ und $|$ sind kommutativ, das heißt $a \ \& \ b = b \ \& \ a$ und $a \ | \ b = b \ | \ a$.

Maskieren von Zeichen mit einem umgekehrten Schrägstrich

Falls Sie ein Konzept haben, das ein Zeichen enthält, das auch ein Syntaxzeichen ist, müssen Sie diesem Zeichen einen umgekehrten Schrägstrich voranstellen, damit die Regel korrekt interpretiert wird. Mit dem umgekehrten Schrägstrich (\backslash) werden Zeichen maskiert, die anderenfalls eine besondere Bedeutung hätten. Bei Drag-and-Drop-Verschiebungen in den Editor werden die umgekehrten Schrägstriche automatisch hinzugefügt.

Den folgenden Regelsyntaxzeichen muss ein umgekehrter Schrägstrich vorangestellt werden, wenn diese nicht als Regelsyntax behandelt werden sollen:

$\& \ ! \ | \ + \ < \ > \ (\) \ [\] \ *$

Da beispielsweise das Konzept $r \ \& \ d$ den Operator "und" enthält ($\&$), ist der umgekehrte Schrägstrich bei einer Eingabe in den Regeleditor erforderlich, also: $r \ \backslash \ \& \ d$.

Verwenden von TLA-Mustern in Kategorieregeln

Text Link Analysis-Muster können explizit in Kategorieregeln definiert werden, damit Sie sogar noch spezifischere und kontextabhängige Ergebnisse erhalten. Wenn Sie ein Muster in einer Kategorieregeln definieren, umgehen Sie die einfacheren Konzeptextrahierungsergebnisse und vergleichen Dokumente und Datensätze auf der Basis der extrahierten Ergebnissen von Text Link Analysis-Mustern.

Trennung durch eckige Klammern. Ein TLA-Muster muss in eckigen Klammern [] stehen, wenn Sie es innerhalb einer Kategorieregeln verwenden. Das Trennzeichen für Muster wird benötigt, wenn Sie nach Übereinstimmungen auf der Basis eines extrahierten TLA-Musters suchen. Da Kategorieregeln Typen, Konzepte oder Muster enthalten können, verdeutlichen die Klammern für die Regel, dass der Inhalt in den Klammern auf extrahierte TLA-Muster verweist. Wenn Sie dieses TLA-Muster nicht extrahiert haben, ist keine Übereinstimmung möglich. Falls Sie ein Muster ohne Klammern sehen, z. B. Apfel + gut im Bereich “Kategorien”, bedeutet das höchstwahrscheinlich, dass das Muster außerhalb des Kategorieregleditors der Kategorie direkt hinzugefügt wurde. Wenn Sie beispielsweise ein Konzeptmuster aus dem Bereich “Extrahierungsergebnisse” direkt einer Kategorie hinzufügen, erscheint es nicht in eckigen Klammern. Wird ein Muster jedoch innerhalb einer Kategorieregeln verwendet, müssen Sie das Muster innerhalb der Kategorieregeln in eckige Klammern setzen, z. B. [Banane + !(gut)].

Verwendung des Pluszeichens in Mustern. In IBM® SPSS® Text Analytics for Surveys sind zweiteilige Muster möglich. Wenn die Reihenfolge wichtig ist, verwenden Sie das +-Zeichen, um die Elemente miteinander zu verbinden. Falls die Reihenfolge unwichtig ist, können Sie stattdessen den Boole’schen Operator & verwenden. In den folgenden zwei Texten ist die Position des Wortes *besser* wichtig: “*Brokkoli ist besser*” und “*es ist besser als Brokkoli*”.

Angenommen, Sie hätten die folgenden Ausdrücke: “*Ich mag Ananas*” und “*Ich mag keine Ananas. Aber ich mag Erdbeeren*”. Der Ausdruck *mag & Ananas* weist eine Übereinstimmung mit beiden Texten auf, da es sich um einen Konzeptausdruck und nicht um eine Textlinkregel (nicht in Klammern eingeschlossen) handelt. Der Ausdruck *Ananas + mag* entspricht nur “*Ich mag Ananas*”, da im zweiten Text das Wort *mag* stattdessen mit *Erdbeeren* verknüpft ist.

Gruppierung mit Mustern. Sie können Ihre Regeln mit Ihren eigenen Mustern vereinfachen. Angenommen, Sie möchten die folgenden drei Ausdrücke festhalten: *Roter Pfeffer + mag*, *Grüner Pfeffer + mag* und *Pfeffer + mag*. Sie können sie zu einer einzigen Kategorieregeln gruppieren, z. B. [** Pfeffer & mag*]. Falls Sie einen anderen Ausdruck hätten, *scharfer Pfeffer + gut*, könnten Sie alle vier mit einer Regel wie [** Pfeffer + <Positive>*] gruppieren.

Reihenfolge in Mustern. Zur besseren Organisation von Ausgaben versuchen die Text Link Analysis-Regeln aus den mit dem Produkt installierten Vorlagen, die grundlegenden Ausgabemuster in derselben Reihenfolge unabhängig von der Wortfolge im Satz auszugeben. Wenn beispielsweise ein Datensatz den Text “*gute Präsentationen.*” und ein anderer Datensatz den Text “*die Präsentationen hatten gute Inhalte*” enthält, werden beide Texte durch dieselbe Regel erfasst und in derselben Reihenfolge wie *Präsentation + gute* in swN Konzeptmusterergebnissen ausgegeben, nicht als *Präsentation + gute* und gleichzeitig *gute + Präsentation*. Und in Zwei-Slot-Mustern wie im Beispiel werden die Konzepte,

die Typen in der Meinungsbibliothek zugeordnet sind, standardmäßig in der Ausgabe als letzte präsentiert, wie z. B. Apfel + schlecht.

Tabelle 6-9

Mustersyntax und Verwendung von Boole'schen Operatoren

Ausdruck	Resultiert in der Übereinstimmung mit Datensätzen, auf die Folgendes zutrifft:
[]	Enthält ein beliebiges TLA-Muster. Das Trennzeichen für Muster wird <i>in Kategorieregeln</i> benötigt, wenn Sie nach Übereinstimmungen auf der Basis eines extrahierten TLA-Musters suchen. Der Inhalt in den Klammern verweist auf TLA-Muster und nicht auf einfache Konzepte und Typen. Wenn Sie dieses TLA-Muster nicht extrahiert haben, ist keine Übereinstimmung möglich. Wenn Sie eine Regel erstellen wollten, die keinerlei Muster enthält, könnten Sie Folgendes verwenden: !([]).
[a]	Enthält ein Muster, in dem mindestens ein Element a ist, unabhängig von seiner Position im Muster. Beispielsweise kann [Geschäft][Geschäft+ gut] oder nur [Geschäft + .] als Übereinstimmung finden.
[a + b]	Enthält ein Konzeptmuster. Zum Beispiel [Geschäft + gut]. <i>Anmerkung:</i> Wenn Sie nur dieses Muster festhalten und keine weiteren Elemente hinzufügen möchten, ist es empfehlenswert, das Muster direkt Ihrer Kategorie hinzuzufügen, anstatt eine Regel damit zu erstellen.
[<A> +]	Enthält ein Muster vom Typ <A> im ersten Slot und vom Typ im zweiten Slot, wobei es genau zwei Slots gibt. Das +-Zeichen zeigt an, dass die Reihenfolge der übereinstimmenden Elemente wichtig ist. Zum Beispiel [<Budget> + <Negative>]. <i>Anmerkung:</i> Wenn Sie nur dieses Muster festhalten und keine weiteren Elemente hinzufügen möchten, ist es empfehlenswert, das Muster direkt Ihrer Kategorie hinzuzufügen, anstatt eine Regel damit zu erstellen.
[<A> &]	Enthält ein Typmuster mit Typ <A> und Typ . Zum Beispiel [<Budget> & <Negative>]. Dieses TLA-Muster wird nie extrahiert; bei dieser Schreibweise gleicht es jedoch tatsächlich [<Budget> + <Negative>][<Negative> + <Budget>]. Die Reihenfolge der übereinstimmenden Elemente ist unwichtig.
[a + .]	Enthält ein Muster, in dem a das einzige Konzept ist und keine anderen Slots für dieses Muster einen Inhalt haben. Zum Beispiel [deal + .] findet als Übereinstimmung das Konzeptmuster, bei dem die einzige Ausgabe das Konzept Geschäft ist. Wenn Sie das Konzept Geschäft als Kategoriedeskriptor hinzufügen, erhalten Sie alle Datensätze mit "Geschäft" als Konzept, einschließlich positive Aussagen über ein Geschäft. Bei Verwendung von [Geschäft + .] werden jedoch nur die Datensatzmusterergebnisse abgeglichen, die Geschäft präsentieren und keine anderen Beziehungen oder Meinungen, d. h. sie würden nicht mit Geschäft + großartig übereinstimmen. <i>Anmerkung:</i> Wenn Sie nur dieses Muster festhalten und keine weiteren Elemente hinzufügen möchten, ist es empfehlenswert, das Muster direkt Ihrer Kategorie hinzuzufügen, anstatt eine Regel damit zu erstellen.
[<A> + <>]	Enthält ein Muster, in dem <A> der einzige Typ ist. [<Budget> + <>] findet das Muster als Übereinstimmung, dessen einzige Ausgabe ein Konzept vom Typ <Budget> ist. <i>Anmerkung:</i> Sie können das <> nur dann verwenden, um einen leeren Typ zu kennzeichnen, wenn Sie es hinter das +-Symbol im Typmuster stellen, z. B. [<Budget> + <>], nicht jedoch [price + <>]. <i>Anmerkung:</i> Wenn Sie nur dieses Muster festhalten und keine weiteren Elemente hinzufügen möchten, ist es empfehlenswert, das Muster direkt Ihrer Kategorie hinzuzufügen, anstatt eine Regel damit zu erstellen.

Ausdruck	Resultiert in der Übereinstimmung mit Datensätzen, auf die Folgendes zutrifft:
[a + !(b)]	Enthält mindestens ein Muster, das das Konzept a einschließt, nicht jedoch das Konzept b. Muss mindestens ein Muster einschließen. Zum Beispiel [Preis + !(hoch)] oder für Typen [!(<i><Obst></i> <i><Gemüse></i>) + <i><Positive></i>]
!([<A> &])	Enthält kein spezifisches Muster. Zum Beispiel !([<Budget> & <Negative>]).

Anmerkung: Beispiele für die Übereinstimmung von Regeln mit Text finden Sie unter [Beispiele für Kategorieregeln auf S. 155](#)

Platzhalter in Kategorieregeln

Platzhalter können Konzepten in Regeln hinzugefügt werden, um die Übereinstimmungsmöglichkeiten zu erweitern. Der Platzhalter Stern (*) kann vor und/oder nach ein Wort gestellt werden, um anzuzeigen, wie nach Übereinstimmungen in Konzepten gesucht wird. Es gibt zwei Arten der Verwendung von Platzhaltern:

- **Affix-Platzhalter.** Diese Platzhalter werden einer Zeichenfolge direkt voran- oder nachgestellt, ohne dass ein Leerzeichen zwischen der Zeichenfolge und dem Stern steht. Übereinstimmungen für *operat** können beispielsweise *operat*, *Operator*, *Operation*, *Operationen*, *operativ* usw. sein.
- **Wort-Platzhalter.** Diese Platzhalter stehen vor oder nach einem Konzept, wobei ein Leerzeichen zwischen dem Konzept und dem Stern steht. Übereinstimmungen für ** operation* können beispielsweise *Operation*, *medizinische Operation*, *geschäftliche Operation* usw. sein. Außerdem kann ein Wort-Platzhalter neben einem Affix-Platzhalter verwendet werden, z. B. ** operat* **, wobei mögliche Übereinstimmungen *Operation*, *medizinische Operation*, *erfahrener Operateur*, *operative Maßnahme* usw. sein können. Wie das letzte Beispiel zeigt, sollte man mit Platzhaltern vorsichtig umgehen, um zu vermeiden, dass ein zu weites Feld abgedeckt wird und ungewollte Übereinstimmungen entstehen.

Ausnahmen!

- Ein Platzhalter kann nie alleine stehen. (*Apfel | **) zum Beispiel wäre nicht möglich.
- Ein Platzhalter kann nie für Übereinstimmungen mit Typnamen verwendet werden. *<Negative*>* stimmt mit überhaupt keinen Typnamen überein.
- Sie können bestimmte Typen durch Filtern nicht aus der Suche nach Übereinstimmungen in Konzepten ausschließen, die über Platzhalter gefunden wurden. Der Typ, dem das Konzept zugewiesen ist, wird automatisch verwendet.
- Ein Platzhalter kann sich nie in der Mitte einer Wortfolge befinden, weder am Ende noch am Beginn eines Worts (*Konto* eröffnen*) oder als eigene Komponente (*Konto * eröffnen*). Sie können Platzhalter außerdem nicht in Typnamen verwenden. Beispiel: *word* word* wie *Apfel* Rezept* entspricht nicht *Apfelcreme-Rezept* und auch keinem anderen Begriff. Jedoch würde *Apfel* ** den Einträgen *Apfelcreme-Rezept*, *Apfel im Schlafrock*, *Apfel* usw. entsprechen. In einem anderen Beispiel stimmt *word * word* wie *Apfel * Toast* nicht mit *Apfel auf Toast* und auch keinem anderen Begriff überein, da das Sternchen zwischen zwei anderen Wörtern erscheint. Jedoch würde *Apfel ** den Einträgen *Apfel auf Toast*, *Apfel*, *Apfel im Schlafrock* usw. entsprechen.

Tabelle 6-10
Verwendung von Wildcards

Ausdruck	Resultiert in der Übereinstimmung mit Datensätzen, auf die Folgendes zutrifft:
*apfel	Enthält ein Konzept, das mit der angegebenen Zeichenfolge endet, jedoch beliebig viele Buchstaben als Präfix haben kann. Beispiel: *apfel endet mit den Buchstaben <i>apfel</i> , kann aber beispielsweise folgende Präfixe enthalten: <ul style="list-style-type: none"> - Apfel - Granatapfel - Augapfel
apfel*	Enthält ein Konzept, das mit der angegebenen Zeichenfolge beginnt, jedoch beliebig viele Buchstaben als Suffix haben kann. Beispiel: apfel* beginnt mit den Buchstaben <i>apfel</i> , kann aber beispielsweise folgende Suffixe (oder kein Suffix) enthalten: <ul style="list-style-type: none"> - Apfel - Apfelkompott - Apfelwein. <p>Apfel* & !(Birne* Quitte) beispielsweise enthält ein Konzept, das mit den Buchstaben <i>Apfel</i> beginnt, jedoch kein Konzept, das mit den Buchstaben <i>Birne</i> beginnt und auch nicht das Konzept <i>Quitte</i>. Daher ergäbe Folgendes KEINE Übereinstimmung: Apfel & Quitte Folgendes ergäbe hingegen durchaus eine Übereinstimmung:</p> <ul style="list-style-type: none"> - Apfelkompott - Apfel & Orange
produkt	Enthält ein Konzept mit der Buchstabenfolge produkt, das aber beliebig viele Buchstaben als Präfix, Suffix oder beides aufweisen kann. Beispiel: Übereinstimmungen für *produkt* wären zum Beispiel: <ul style="list-style-type: none"> - Produkt - Nebenprodukt - unproduktiv
* Darlehen	Enthält ein Konzept mit dem Wort Darlehen, das aber mit einem anderen Wort, das davor steht, eine Zusammensetzung bilden kann. Übereinstimmungen für * Darlehen wären zum Beispiel: <ul style="list-style-type: none"> - Darlehen - zinsloses Darlehen - nicht gewährtes Darlehen. <p>Übereinstimmende Konzeptmuster für [* Lieferung + <Negative>], das an erster Stelle ein Konzept enthält, das mit dem Wort Lieferung endet und an zweiter Stelle einen Typ <Negative>, wären zum Beispiel:</p> <ul style="list-style-type: none"> - erste Lieferung + langsam - wichtige Lieferung + spät
Veranstaltung *	Enthält ein Konzept mit dem Wort Veranstaltung, das aber mit einem anderen Wort, das dahinter steht, eine Zusammensetzung bilden kann. Übereinstimmungen für * Veranstaltung wären zum Beispiel: <ul style="list-style-type: none"> - Veranstaltung - Veranstaltung Berlin - Veranstaltung am Fluss
* Apfel *	Enthält ein Konzept, das mit einem beliebigen Wort beginnen kann, nach dem das Wort Apfel steht, auf das ein anderes Wort folgen kann. * bedeutet 0 oder n, daher stimmt es auch mit Apfel überein. Übereinstimmungen für * Apfel * wären zum Beispiel: <ul style="list-style-type: none"> - warmer Apfelkompott - Granny Smith Apfel marktfrisch - roter Apfel kandiert. - Apfel <p>Übereinstimmende Konzeptmuster für [* Reservierung* * + <Positive>], das an erster Stelle ein Konzept enthält, das das Wort Reservierung enthält (unabhängig davon, an welcher Stelle es im Konzept steht) und an zweiter Stelle einen Typ <Positive>, wären zum Beispiel:</p> <ul style="list-style-type: none"> - Reservierung Hotel + gut - Online Reservierung + gut

Anmerkung: Beispiele für die Übereinstimmung von Regeln mit Text finden Sie unter [Beispiele für Kategorieregeln auf S. 155](#)

Beispiele für Kategorieregeln

Werfen Sie einen Blick auf folgende Beispiele, um zu veranschaulichen, wie Regeln mit Datensätzen verglichen werden, die auf unterschiedliche Weise auf der Syntax basieren, die für deren Ausdruck verwendet wird.

Beispieldatensätze

Angenommen, Sie hatten zwei Datensätze:

- **Datensatz A:** *“Als ich in meinen Geldbeutel sah, stellte ich fest, dass mir fünf Euro fehlten.”*
- **Datensatz B:** *“Ich fand die fünf Euro im Picknickbereich, aber die Decke fehlte.”*

Die folgenden zwei Tabellen zeigen mögliche extrahierte Konzepte und Typen sowie Konzept- und Typmuster.

Aus dem Beispiel extrahierte Konzepte und Typen

Tabelle 6-11

Aus Beispiel extrahierte Konzepte und Typen

Extrahiertes Konzept	Art des Konzepts
Geldbeutel	<Unbekannt>
fehlend	<Negativ>
5 Euro	<Währung>
Decke	<Unbekannt>
Picknickbereich	<Unbekannt>

Aus dem Beispiel extrahierte TLA-Muster

Tabelle 6-12

Aus Beispiel extrahierte TLA-Musterausgabe

Extrahierte Konzeptmuster	Extrahierte Typmuster	Aus Datensatz
Picknickbereich + .	<Unbekannt> + <>	Datensatz B
Geldbeutel + .	<Unbekannt> + <>	Datensatz A
Decke + fehlend	<Unbekannt> + <Negativ>	Datensatz B
5 Euro + .	<Währung> + <>	Datensatz B
5 Euro + fehlend	<Währung> + <Negativ>	Datensatz A

Mögliche Übereinstimmung von Kategorieregeln

Die folgende Tabelle enthält Syntaxeinträge, die im Kategorieregeleditor eingegeben werden könnten. Nicht alle hier aufgeführten Regeln funktionieren und nicht alle stimmen mit denselben Datensätzen überein. Beachten Sie, wie sich die unterschiedliche Syntax auf die verglichenen Datensätze auswirkt.

Tabelle 6-13
Beispielregeln

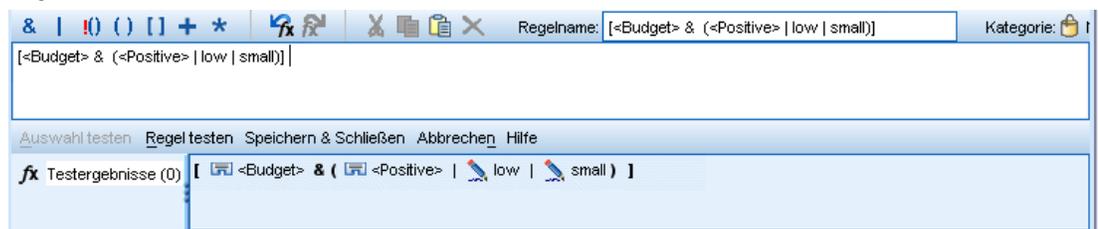
Regelsyntax	Ergebnis
5 Euro & fehlend	Übereinstimmung mit Datensatz A und B, da beide das extrahierte Konzept <i>fehlend</i> und das extrahierte Konzept <i>5 Euro</i> enthalten. Dies entspricht: (5 Euro & fehlend)
fehlend & 5 Euro	Übereinstimmung mit Datensatz A und B, da beide das extrahierte Konzept <i>fehlend</i> und das extrahierte Konzept <i>5 Euro</i> enthalten. Dies entspricht: (fehlend & 5 Euro)
fehlend & <Währung>	Übereinstimmung mit Datensatz A und B, da beide das extrahierte Konzept <i>fehlend</i> und ein Konzept enthalten, das mit dem Typ <Währung> übereinstimmt. Dies entspricht: (fehlend & <Währung>)
<Währung> & fehlend	Übereinstimmung mit Datensatz A und B, da beide das extrahierte Konzept <i>fehlend</i> und ein Konzept enthalten, das mit dem Typ <Währung> übereinstimmt. Dies entspricht: (<Währung> & fehlend)
[5 Euro + fehlend]	Übereinstimmung mit A, aber nicht mit B, da Datensatz B keine TLA-Musterausgabe mit 5 Euro + fehlend erzeugt hat (siehe vorherige Tabelle). Dies entspricht folgender TLA-Musterausgabe: 5 Euro + fehlend
[fehlend + 5 Euro]	Weder Übereinstimmung mit A noch mit B, da kein extrahiertes TLA-Muster (siehe vorherige Tabelle) mit der hier ausgedrückten Reihenfolge mit <i>fehlend</i> an erster Stelle übereinstimmt. Dies entspricht folgender TLA-Musterausgabe: 5 Euro + fehlend
[fehlend & 5 Euro]	Übereinstimmung mit A, aber nicht mit B, da kein derartiges TLA-Muster aus Datensatz B extrahiert wurde. Das Symbol & zeigt an, dass die Reihenfolge beim Vergleich unwichtig ist; aus diesem Grund sucht diese Regel nach einer Musterübereinstimmung entweder mit [fehlend + 5 Euro] oder [5 Euro + fehlend]. Es gibt nur eine Übereinstimmung bei [5 Euro + fehlend] aus Datensatz A.
[fehlend + <Währung>]	Weder Übereinstimmung mit Datensatz A noch mit Datensatz B, da kein extrahiertes TLA-Muster mit dieser Reihenfolge übereinstimmt. Hier gibt es keine Entsprechung, da eine TLA-Ausgabe nur auf Begriffen (5 Euro + fehlend) oder auf Typen (<Währung> + <Negativ>) basiert, Konzepte und Typen jedoch nicht vermischt.
[<Währung> + <Negativ>]	Übereinstimmung mit Datensatz A, aber nicht B, da kein TLA-Muster aus Datensatz B extrahiert wurde. Dies entspricht folgender TLA-Ausgabe: <Währung> + <Negativ>
[<Negativ> + <Währung>]	Weder Übereinstimmung mit Datensatz A noch mit Datensatz B, da kein extrahiertes TLA-Muster mit dieser Reihenfolge übereinstimmt. Standardmäßig belegt in der Vorlage <i>Meinungen</i> , wenn ein <i>Thema</i> mit einer <i>Meinung</i> gefunden wird, das <i>Thema</i> (<Währung>) die erste Slot-Position und die <i>Meinung</i> (<Negativ>) die zweite Slot-Position.

Erstellen von Kategorienregeln

Beim Erstellen oder Bearbeiten einer Regel muss diese im Regeleditor geöffnet sein. Sie können Konzepte, Typen oder Muster hinzufügen oder Platzhalter verwenden, um die Übereinstimmungsmöglichkeiten zu vergrößern. Wenn Sie erkannte Konzepte, Typen und Muster verwenden, haben Sie den Vorteil, dass alle verwandten Konzepte gefunden werden. Wenn Sie beispielsweise ein Konzept verwenden, werden alle damit verbundenen Begriffe, Pluralformen und Synonyme ebenfalls mit der Regel in Übereinstimmung gebracht. Ebenso werden, wenn Sie einen Typ verwenden, all seine Konzepte ebenfalls von der Regel erfasst.

Sie können den Regeleditor öffnen, indem Sie eine vorhandene Regel bearbeiten oder mit der rechten Maustaste auf den Namen der Kategorie klicken und die Option Regel erstellen wählen.

Abbildung 6-25
Regeleditorbereich



Sie können Kontextmenüs und Drag-and-Drop verwenden oder Konzepte, Typen und Muster manuell in den Editor eingeben. Danach kombinieren Sie diese mit Boole'schen Operatoren (&, !(), |) und Klammern, um Ihre Regelausdrücke zu formulieren. Um häufig vorkommende Fehler zu vermeiden, sollten Sie die Konzepte direkt per Drag-and-Drop aus dem Fensterbereich "Extrahierungsergebnisse" oder "Daten" in den Regeleditor übertragen. Achten Sie genau auf die Syntax dieser Regeln, um Fehler zu vermeiden. [Für weitere Informationen siehe Thema Kategorieregelsyntax auf S. 149.](#)

Anmerkung: Beispiele für die Übereinstimmung von Regeln mit Text finden Sie unter [Beispiele für Kategorieregeln auf S. 155](#)

So erstellen Sie eine Regel:

- ▶ Wenn Sie noch keine Daten extrahiert haben oder Ihre Extrahierung veraltet ist, tun Sie es jetzt. [Für weitere Informationen siehe Thema Daten extrahieren in Kapitel 5 auf S. 85.](#)
- ▶ Wählen Sie im Fensterbereich "Kategorien" die Kategorie aus, die Ihrer Regel hinzugefügt werden soll.
- ▶ Wählen Sie in den Menüs die Optionsfolge Kategorien > Regel erstellen. Im Fenster wird der Kategorienregel-Editorbereich geöffnet.
- ▶ Geben Sie im Feld "Regelname" einen Namen für Ihre Regel ein. Wenn Sie keinen Namen angeben, wird der Ausdruck automatisch als Name verwendet. Sie können diese Regel später umbenennen.

- ▶ Im größeren Textfeld für Ausdrücke haben Sie folgende Möglichkeiten:
 - Geben Sie Text direkt in das Feld ein oder verschieben Sie ihn per Drag-and-Drop aus einem anderen Bereich dorthin. Verwenden Sie nur extrahierte Konzepte, Typen und Muster. Wenn Sie beispielsweise das Wort *Katzen* eingeben, im Fensterbereich “Extrahierungsergebnisse” aber nur die Singularform, *Katze*, angezeigt wird, kann der Editor *Katzen* nicht erkennen. In diesem letzten Fall könnte die Singularform automatisch den Plural einschließen, ansonsten könnten Sie einen Platzhalter verwenden. [Für weitere Informationen siehe Thema Kategorieregelsyntax auf S. 149.](#)
 - Wählen Sie die Konzepte, Typen oder Muster aus, die Sie Regeln hinzufügen möchten, und verwenden Sie die Menüs.
 - Boole’sche Operatoren hinzufügen, um Elemente in Ihrer Regel miteinander zu verknüpfen. Über die Schaltflächen der Symbolleiste können Sie Ihrer Regel die Boole’schen Operatoren “und”, &, “oder”, | und “nicht” !() sowie runde Klammern () und eckige Klammern für Muster [] hinzufügen.
- ▶ Klicken Sie auf die Schaltfläche Regel testen, um zu überprüfen, ob Ihre Regel korrekt formuliert ist. [Für weitere Informationen siehe Thema Kategorieregelsyntax auf S. 149.](#) Die Anzahl der gefundenen Datensätze wird in Klammern neben dem Text Testergebnis angezeigt. Rechts neben dem Text können Sie die Elemente in Ihrer Regel sehen, die erkannt wurden, bzw. etwaige Fehlermeldungen. Ein rotes Fragezeichen neben dem Typ, Muster oder Konzept zeigt an, dass das Element mit keinen bekannten Extrahierungen übereinstimmt. Falls dies der Fall ist, wird die Regel keine Datensätze finden.
- ▶ Um einen Teil der Regel zu testen, wählen Sie den betreffenden Teil aus und klicken Sie auf Auswahl testen.
- ▶ Nehmen Sie ggf. die erforderlichen Änderungen vor und testen Sie die Regel erneut, wenn Sie Probleme festgestellt haben.
- ▶ Klicken Sie abschließend auf Speichern & Schließen, um Ihre Regel erneut zu speichern und den Editor zu schließen. Der neue Regelname wird in der Kategorie angezeigt.

Abbildung 6-26
Regel im Fensterbereich “Kategorien”

Kategorie	Anzahl D...	Anzahl Tr...
Alle Datensätze	-	405
Nicht kategorisiert	-	282
Keine Konzepte extrahiert	-	4
songs+<Positive>	3	9
fx [<Budget> & (<Positive> low small)]		1
fx [songs + like] [songs + meets needs] [songs		7
drag & dropped songs from their site+<Positive>	1	1
fx [drag \& dropped songs from their site + eas		1
memory	2	2
fx [memory + easy to use]		1
fx [storage + able]		1
battery+<Contextual>	2	3
design+<Positive>	2	7
store+<Positive>	2	3
skipping+<Contextual>	2	2

Bearbeiten und Löschen von Regeln

Nachdem Sie eine Regel erstellt und gespeichert haben, können Sie sie jederzeit bearbeiten. [Für weitere Informationen siehe Thema Kategorieregelsyntax auf S. 149.](#)

Wenn Sie eine Regel nicht mehr benötigen, können Sie sie löschen.

So bearbeiten Sie Regeln:

- ▶ Wählen Sie im Fensterbereich “Kategorien” die Regel aus, die Sie bearbeiten möchten.
- ▶ Wählen Sie in den Menüs die Optionsfolge Kategorien > Regel bearbeiten oder doppelklicken Sie auf den Regelnamen. Der Editor wird geöffnet und die ausgewählte Regel wird angezeigt.
- ▶ Nehmen Sie Änderungen an der Regel über Extrahierungsergebnisse und die Schaltflächen der Symbolleiste vor.
- ▶ Testen Sie die Regel erneut, um sicherzustellen, dass sie die erwarteten Ergebnisse liefert.
- ▶ Klicken Sie auf Speichern & Schließen, um Ihre Regel erneut zu speichern und den Editor zu schließen.

So löschen Sie eine Regel

- ▶ Wählen Sie im Fensterbereich “Kategorien” die Regel aus, die Sie löschen möchten.
- ▶ Wählen Sie im Menü Bearbeiten > Löschen. Die Regel wird aus der Kategorie gelöscht.

Bearbeiten und Verfeinern von Kategorien

Nachdem Sie einige Kategorien erstellt haben, sollten Sie diese untersuchen und Anpassungen vornehmen. Neben der Verfeinerung der linguistischen Ressourcen sollten Sie Ihre Kategorien überprüfen, indem Sie nach Möglichkeiten suchen, die zugehörigen Definitionen zu kombinieren oder zu bereinigen, und einige der kategorisierten Datensätze prüfen. Außerdem können Sie die Datensätze in einer Kategorie überprüfen und Anpassungen vornehmen, sodass die Kategorien so definiert sind, dass Nuancen und Unterschiede erfasst werden.

Sie können die integrierten, automatisierten Kategorieerstellungungsverfahren zum Erstellen Ihrer Kategorien verwenden; Sie werden jedoch wahrscheinlich noch einige Optimierungen an diesen Kategorien vornehmen wollen. Nach der Verwendung eines oder mehrerer Verfahren wird eine Reihe neuer Kategorien im Fenster angezeigt. Anschließend können Sie die Daten in einer Kategorie überprüfen und Anpassungen vornehmen, bis Sie mit den Kategoriedefinitionen zufrieden sind. [Für weitere Informationen siehe Thema Erläuterung von Kategorien auf S. 110.](#)

Hier sind einige Optionen zur Verfeinerung Ihrer Kategorien, von denen die meisten auf den folgenden Seiten beschrieben werden:

- Bearbeitung von Kategorieeigenschaften (Umbenennung, Hinzufügen von Bezeichnungen, Hinzufügen von Anmerkungen)
- Hinzufügen von Deskriptoren zu Ihren Kategorien
- Bearbeiten von Kategorien

- Verschieben von Kategorien
- Glätten hierarchischer Kategorien
- Zusammenführen von Kategorien
- Hinzufügen von Textabgleich
- Erzwungenermaßen in Kategorien eingeschlossene Antworten
- Kopieren und Wiederverwenden von Kategorien
- Löschen von Kategorien
- Vornehmen von Änderungen an den linguistischen Ressourcen und erneutes Extrahieren
- Visualisieren, wie die Kategorien zusammenarbeiten, und Vornehmen von Anpassungen. [Für weitere Informationen siehe Thema Visualisierung von Diagrammen in Kapitel 7 auf S. 172.](#)

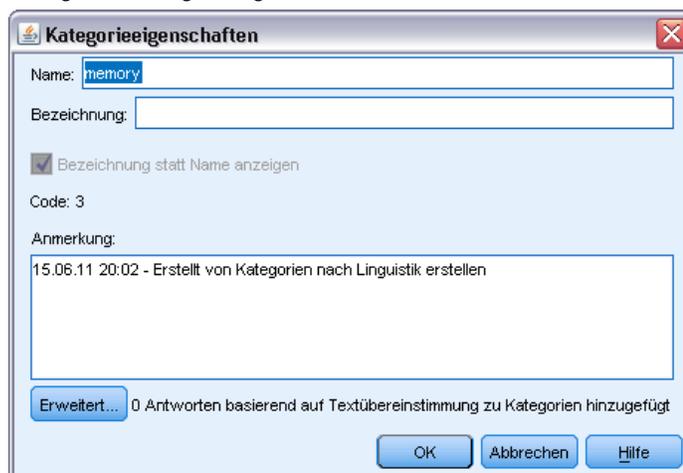
Bearbeiten von Kategorieeigenschaften

Wie viele andere Elemente in IBM® SPSS® Text Analytics for Surveys können Sie auch die Eigenschaften Ihrer Kategorien bearbeiten: Namen, Bezeichnungen, Anmerkungen und erweiterte Textübereinstimmungseinträge. [Für weitere Informationen siehe Thema Kategorieeigenschaften auf S. 111.](#) Neben den bearbeitbaren Eigenschaften können Sie auch die Anzahl der Elemente sehen, die in der Kategoriedefinition enthalten sind, also die Anzahl von Begriffen, Typen, TLA-Mustern oder Kategorieregeln, die diese Kategorie bilden. Die Codenummer wird ebenfalls angezeigt und entspricht dem Codewert im Coderahmenmanager.

So bearbeiten Sie Kategorieeigenschaften:

- ▶ Wählen Sie in den Menüs die Optionsfolge Kategorien > Kategorieeigenschaften. Das Dialogfeld “Kategorieeigenschaften” wird geöffnet.

Abbildung 6-27
Dialogfeld “Kategorieeigenschaften”



- ▶ Benennen Sie bei Bedarf die Kategorie um, indem Sie einen neuen Namen ins Feld “Name” eingeben.

- ▶ Ändern Sie Kategorienname oder -bezeichnung.
- ▶ Um die Bezeichnung auf der Benutzeroberfläche, beispielsweise im Fensterbereich “Kategorie”, anstatt des Kategorienamens zu verwenden, wählen Sie die Option Bezeichnung statt Name anzeigen aus.
- ▶ Bei Bedarf geben Sie eine Anmerkung im Feld “Anmerkung” ein.
- ▶ Um ein Wort oder eine Phrase in die Kategoriedefinition zu zwingen, klicken Sie auf Erweitert und geben Sie Ihre Textabgleiche in die Tabelle ein. [Für weitere Informationen siehe Thema Textübereinstimmungen in Kategorien auf S. 167.](#)
- ▶ Klicken Sie auf OK, um die Änderungen anzuwenden.

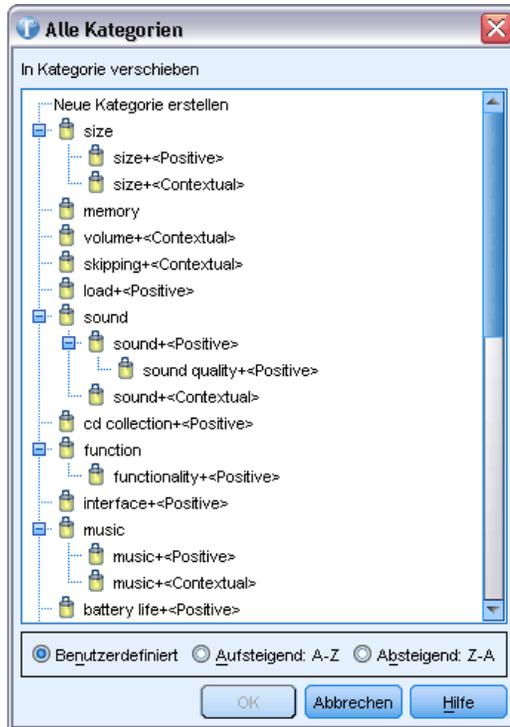
Hinzufügen von Deskriptoren zu Kategorien

Nach der Verwendung von automatisierten Verfahren liegen Ihnen mit großer Wahrscheinlichkeit noch Extrahierungsergebnisse vor, die in keiner der Kategoriedefinitionen verwendet wurden. Sie sollten die betreffende Liste im Fensterbereich “Extrahierungsergebnisse” durchgehen. Wenn Sie Elemente finden, die Sie in eine Kategorie verschieben möchten, können Sie diese zu einer bestehenden oder zu einer neuen Kategorie hinzufügen.

So fügen Sie ein Konzept bzw. einen Typ zu einer Kategorie hinzu:

- ▶ Wählen Sie in den Fensterbereichen “Extrahierungsergebnisse” und “Daten” die Elemente, die Sie einer neuen oder einer bestehenden Kategorie hinzufügen möchten.
- ▶ Wählen Sie in den Menüs Kategorien > Zu Kategorie hinzufügen. Im Dialogfeld “Alle Kategorien” wird das Kategorie-Set angezeigt. Wählen Sie die Kategorie aus, zu der die ausgewählten Elemente hinzugefügt werden sollen. Wenn Sie die Elemente zu einer neuen Kategorie hinzufügen möchten, wählen Sie Neue Kategorie. Im Fensterbereich “Kategorien” wird eine neue Kategorie angezeigt, für die der Name des ersten ausgewählten Elements verwendet wird.

Abbildung 6-28
Dialogfeld "Alle Kategorien"



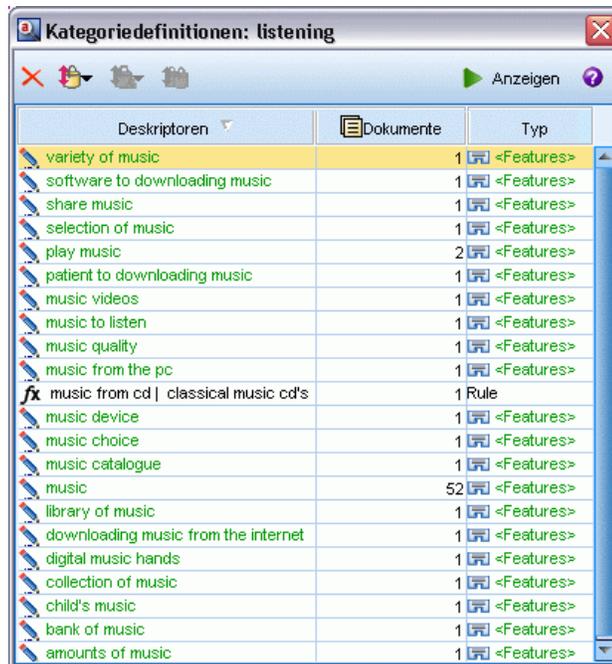
Bearbeiten von Kategoriedeskriptoren

Nachdem Sie einige Kategorien erstellt haben, können Sie die einzelnen Kategorien öffnen, um alle Deskriptoren anzuzeigen, aus denen die Definition der jeweiligen Kategorie besteht. Im Dialogfeld "Kategoriedefinitionen" können Sie eine Reihe von Bearbeitungsschritten an den Kategoriedeskriptoren vornehmen. Außerdem können Sie mit ihnen dort arbeiten, wenn Kategorien im Kategoriebaum angezeigt werden.

So bearbeiten Sie eine Kategorie:

- ▶ Wählen Sie die zu bearbeitende Kategorie im Fensterbereich "Kategorien" aus.
- ▶ Wählen Sie Ansicht > Kategoriedefinitionen. Das Dialogfeld "Kategoriedefinitionen" wird geöffnet.

Abbildung 6-29
Dialogfeld "Kategoriedefinitionen"



- Wählen Sie den zu bearbeitenden Deskriptor aus und klicken Sie auf die entsprechende Schaltfläche auf der Symbolleiste.

In der folgenden Tabelle sind die einzelnen Symbolleistenschaltflächen beschrieben, mit denen Sie die Kategoriedefinitionen bearbeiten können.

Tabelle 6-14
Symbolleistenschaltflächen und Beschreibungen

Symbole	Beschreibung
	Löscht die ausgewählten Deskriptoren aus der Kategorie .
	Verschiebt die ausgewählten Deskriptoren in eine neue oder bereits vorhandene Kategorie.
	Verschiebt die ausgewählten Deskriptoren in Form einer &-Kategorieregel in eine Kategorie. Für weitere Informationen siehe Thema Verwenden von Kategorieregeln auf S. 148 .
	Verschiebt die einzelnen ausgewählten Deskriptoren jeweils als eigene neue Kategorie
	Aktualisiert die Anzeige des Daten- und des Visualisierungsbereichs in Funktion der ausgewählten Deskriptoren

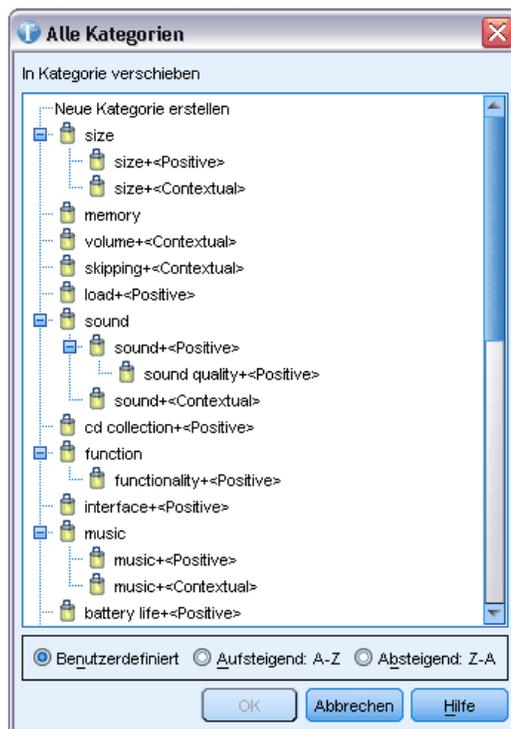
Verschieben von Kategorien

Wenn Sie eine Kategorie in eine andere vorhandene Kategorie einordnen oder Deskriptoren in eine andere Kategorie verschieben möchten, können Sie dies tun.

So verschieben Sie eine Kategorie:

- ▶ Wählen Sie im Fensterbereich “Kategorien” die Kategorien oder Deskriptoren aus, die Sie in eine andere Kategorie verschieben möchten.
- ▶ Wählen Sie in den Menüs Kategorien > In Kategorie verschieben. Im Menü wird eine Menge von Kategorien angezeigt, wobei sich die zuletzt erstellte Kategorie ganz oben in der Liste befindet. Wählen Sie den Namen der Kategorie aus, in die die ausgewählten Konzepte verschoben werden sollen.
 - Wenn Sie den gesuchten Namen gefunden haben, wählen Sie ihn aus und die ausgewählten Elemente werden zu der betreffenden Kategorie hinzugefügt.
 - Wenn er nicht angezeigt wird, wählen Sie Weitere aus, um das Dialogfeld “Alle Kategorien” anzuzeigen, und wählen Sie die Kategorie in der Liste aus.

Abbildung 6-30
Dialogfeld “Alle Kategorien”

**Zusammenführen bzw. Kombinieren von Kategorien**

Um zwei oder mehr vorhandene Kategorien in einer neuen Kategorie zu kombinieren, können Sie sie zusammenführen. Wenn Sie Kategorien zusammenführen, wird eine Kategorie mit einem generischen Namen erstellt. Alle Konzepte, Ausdrücke und Muster, die in den Kategoriedeskriptoren verwendet werden, werden in diese neue Kategorie verschoben. Sie können die Kategorie später durch Bearbeiten der Kategorieeigenschaften umbenennen. [Für weitere Informationen siehe Thema Bearbeiten von Kategorieeigenschaften auf S. 160.](#)

So führen Sie eine Kategorie bzw. einen Teil einer Kategorie zusammen:

- ▶ Wählen Sie im Fensterbereich “Kategorien” die Elemente aus, die zusammengeführt werden sollen.
- ▶ Wählen Sie in den Menüs Kategorien > Kategorien zusammenführen. Das Dialogfeld “Kategorieeigenschaften” wird angezeigt und Sie können einen Namen für die neu erstellte Kategorie eingeben. Die ausgewählten Kategorien werden als Unterkategorien in die ausgewählte Kategorie zusammengeführt.

Einschließen von Antworten in Kategorien erzwingen

Das Erzwingen des Einschlusses von Antworten in Kategorien bzw. des Ausschlusses von Aufnahmen aus Kategorien ermöglicht es Ihnen, die durch automatische Kategorieerstellungsmethoden erstellten Kategoriedefinitionen zu überschreiben, ohne die tatsächliche Kategoriedefinition zu ändern. Obwohl die Antwort Begriffe enthält, die für die Definition einer bestimmten Kategorie verwendet werden, sollte die Antwort selbst unter Umständen nicht in dieser Kategorie enthalten sein. In solchen Fällen können Sie den Ausschluss der Antwort aus dieser Kategorie erzwingen, ohne die Begriffe aus der Kategoriedefinition entfernen zu müssen.

Erzwingen wird in besonderen Fällen verwendet, wenn eine Antwort zu einer Kategorie passt (bzw. nicht dazu passt), aber aus bestimmten Gründen (weil sie zum Beispiel einen bestimmten Begriff enthält) dieser Kategorie zugewiesen (bzw. nicht zugewiesen) wird. Im Normalfall tritt dies auf, wenn eine befragte Person in ihrer Antwort Sarkasmus verwendet, z. B. “*Die Pizza war hervorragend. Ich bin sicher, dass jeder Mensch verbrannte, kalte Pizza liebt.*” Falls Sie nun eine Kategorie Pos: [`<Essen>` + `<Positive>`] haben, um positive Meinungen zum Essen eines Restaurants festzuhalten, wird diese Antwort unter Umständen dieser Kategorie zugewiesen. In diesem Fall könnten Sie erwägen, den Ausschluss dieser Antwort aus dieser Kategorie zu erzwingen.

So werden Antworten in Kategorien eingeschlossen bzw. aus Kategorien ausgeschlossen

- ▶ Wählen Sie im Bereich “Daten” die Antwort, deren Einschluss in eine bestimmte Kategorie (bzw. deren Ausschluss aus einer bestimmten Kategorie) Sie erzwingen möchten.
- ▶ Wählen Sie in den Menüs die Optionsfolge Kategorien > Treffer erzwungenermaßen einschließen oder Kategorien > Treffer erzwungenermaßen ausschließen. Die Liste von Kategorien, aus denen Sie auswählen können, wird in einem Untermenü angezeigt.
- ▶ Wählen Sie die Kategorie aus, auf die Sie den erzwungenen Einschluss (bzw. Ausschluss) der Antwort anwenden möchten. Falls Sie viele Kategorien erstellt haben, werden einige unter Umständen im Untermenü nicht angezeigt.
 - Wählen Sie in diesem Fall Mehr am Ende des Untermenüs. Das Dialogfeld “Alle Kategorien” öffnet sich und Sie können darin die gewünschte Kategorie auswählen und auf OK klicken, um die Änderung anzuwenden.
 - Wenn Sie den Einschluss der Antwort in eine neue Kategorie erzwingen möchten, wählen Sie Neue leere Kategorie. Im Kategoriebaum wird eine neue Kategorie angezeigt, für die ein generischer Name verwendet wird.

Immer wenn eine Kategorie eine oder mehrere erzwungene Antworten enthält, wird unterhalb des Kategorienamens im Baum die Pseudokategorie “Erzwungenermaßen einschließen” bzw. “Erzwungenermaßen ausschließen” angezeigt. Sie können auch erkennen, welche Antworten erzwungenermaßen in eine Kategorie eingeschlossen (bzw. aus einer Kategorie ausgeschlossen) wurden, indem Sie die Spalten “Erzwungenermaßen einschließen” und “Erzwungenermaßen ausschließen” im Bereich “Daten” aufrufen. [Für weitere Informationen siehe Thema Der Fensterbereich “Daten” auf S. 100.](#)

So löschen Sie einen “Erzwungenen Treffer“-Status

- ▶ Wählen Sie im Bereich “Daten” die Antwort, deren Einschluss in eine bestimmte Kategorie (bzw. deren Ausschluss aus einer bestimmten Kategorie) Sie nicht mehr beibehalten möchten.
- ▶ Wählen Sie in den Menüs Kategorien > Treffer erzwungenermaßen einschließen oder Kategorien > Treffer erzwungenermaßen ausschließen. Neben den Kategorien, in die die Antwort erzwungenermaßen eingeschlossen wird (bzw. aus denen die Antwort erzwungenermaßen ausgeschlossen wird), wird ein Häkchen angezeigt.

Abbildung 6-31

Einschluss und Ausschluss von Antworten aus dem Bereich “Daten” heraus erzwingen

	ID	Antwort	Kategorien
1	1	Ziemlich teuer	
2	2	Schlechtes Personal	personal
3	3	n/a	
4	4	Parkplatz zu klein	parkplatz
5	5	nd des Schimmbades	badezimmer
6	6	ngsreich, die Getränke	essen getränke
7	7		
8	8		
9	9		
10	10		
11	11		
12	12		
13	13		
14	14		buffet
15	15	kleines Zimmer	
16	16	die Öffnungszeiten des Fitness-Cent	

- ▶ Wählen Sie im Untermenü die markierte Kategorie aus, für die Sie den erzwungenen Ein- bzw. Ausschluss entfernen möchten. Das Häkchen wird entfernt und der Ein- bzw. Ausschluss der Antwort wird nicht mehr erzwungen.

So löschen Sie einen “Erzwungenen Treffer“-Status

So löschen Sie alle erzwungenen Antwortstatus:

- ▶ Wählen Sie aus den Menüs die Optionsfolge Kategorien > Alles löschen > Erzwungenermaßen einschließen oder Kategorien > Alles löschen > Erzwungenermaßen ausschließen. Der

“Erzwingen”-Status der Antworten wird gelöscht und sie werden nicht mehr erzwingenermaßen in die Kategorien eingeschlossen bzw. aus ihnen ausgeschlossen.

Textübereinstimmungen in Kategorien

Falls Sie versucht haben, die Extrahierung eines Worts oder einer Wortfolge durch die linguistischen Ressourcen zu erzwingen, und dieses wegen anderer linguistischer Handhabungsregeln immer noch nicht extrahiert ist, können Sie einen Textübereinstimmungseintrag erstellen, um alle Kategorien, die diesen Text enthalten, einer bestimmten Kategorie zuzuweisen, ohne die Extrahierungsergebnisse zu verwenden.

Wenn Sie einer Kategorie Textübereinstimmungseinträge (ein Wort oder eine Wortfolge) hinzufügen, weist IBM® SPSS® Text Analytics for Surveys automatisch alle Antworten, die das Wort oder die Wortfolge enthalten, dieser Kategorie zu. Textübereinstimmung sollte nur verwendet werden, wenn Sie zuvor bereits versucht haben, dieses Wort den linguistischen Ressourcen hinzuzufügen, da Sie auf diese Weise den meisten Nutzen aus dieser Definition ziehen können.

[Für weitere Informationen siehe Thema Erzwingen von Fachausdrücken in Kapitel 10 auf S. 231.](#)

Immer wenn eine Textübereinstimmung zu einer Kategoriedefinition hinzugefügt wird, wird unterhalb der Kategorie im Kategoriebaum die Pseudokategorie “Textübereinstimmung” angezeigt. [Für weitere Informationen siehe Thema Der Fensterbereich “Daten” auf S. 100.](#)

Sie können Einträge aus dieser Tabelle auch löschen, indem Sie die Zeilen, die Sie löschen möchten, auswählen und auf die Schaltfläche Löschen klicken.

So fügen Sie einer Kategoriedefinition einen Textübereinstimmungseintrag hinzu:

- ▶ Wählen Sie im Bereich “Daten” das Wort oder die Wortfolge aus, deren Einschluss in eine Kategoriedefinition Sie erzwingen möchten.
- ▶ Wählen Sie im Bereich “Kategorien” die Kategorie aus, für die Sie den erzwungenen Einschluss des Worts oder der Wortfolge vornehmen möchten.
- ▶ Wählen Sie in den Menüs die Optionsfolge Kategorien > Kategorieeigenschaften. Das Dialogfeld “Kategorieeigenschaften” wird geöffnet.
- ▶ Klicken Sie auf Erweitert. Das Dialogfeld “Textübereinstimmung” wird geöffnet.

Abbildung 6-32
Dialogfeld "Textübereinstimmung"



- ▶ Geben Sie in der Tabelle das Wort oder die Wortfolge in die erste Zelle der Spalte "Text" ein.
- ▶ Wählen Sie aus, wie im Text der Antworten nach Übereinstimmungen für das Wort oder die Wortfolge gesucht werden soll. Um Übereinstimmungen zu finden, die exakt dem Wort oder der Wortfolge entsprechen, die Sie eingegeben haben, wählen Sie das ganze Wort oder die ganze Wortfolge aus. Um das Wort oder die Wortfolge für Übereinstimmungen mit längeren Wortfolgen zu verwenden, wählen Sie Teilzeichenfolge.
- ▶ Falls bei der Übereinstimmung mit dem Wort oder der Wortfolge, die Sie eingeben, Groß- und Kleinschreibung beachtet werden müssen, wählen Sie Groß-/Kleinschreibung beachten in der Spalte "Fall".
- ▶ Klicken Sie auf OK, um die Änderungen zu speichern und das Dialogfeld zu schließen. Die Anzahl der Antworten, die der Kategorie aufgrund von Textübereinstimmungseinträgen zugeordnet werden, wird aktualisiert und im Dialogfeld angezeigt.
- ▶ Klicken Sie auf OK, um die Änderungen anzuwenden.

Kopieren von Kategorien

Wenn Sie dieselben oder ähnliche Fragen in mehreren Umfragen verwenden, kann die Wiederverwendung von Kategoriedefinitionen zu erheblichen Zeiteinsparungen führen. Sie können die Kategorien von einer Frage in eine andere im selben Projekt kopieren. Bei der Wiederverwendung von Kategorien müssen Sie eine erneute Extrahierung vornehmen, um die Kategorien mit den Antwortdaten abzugleichen. Bevor Sie eine erneute Extrahierung vornehmen, werden die Kategorien im Fensterbereich "Kategorien" mit einem Fragezeichen für den Häufigkeitswert angezeigt.

Anmerkung: Um Kategorien in einem anderen Projekt wiederzuverwenden, sollten Sie ein Text Analysis Package (TAP) mit den Kategorien und Ressourcen aus Ihrem Projekt erstellen und bei der Erstellung Ihres neuen Projekts im Assistenten verwenden. [Für weitere Informationen siehe Thema Verwendung von Text Analysis Packages in Kapitel 3 auf S. 43.](#)

So kopieren Sie Kategorien innerhalb desselben Projekts:

- ▶ Rufen Sie die Frage auf, deren Kategorien Sie kopieren möchten.
- ▶ Wählen Sie im Baum im Fensterbereich “Kategorien” alle Kategorien aus.
- ▶ Wählen Sie in den Menüs Bearbeiten > Kopieren, um die Kategorien zu kopieren.
- ▶ Rufen Sie die Frage auf (Ansicht > Frage), in die Sie die Kategorien einfügen möchten.
- ▶ Wählen Sie in den Menüs Bearbeiten > Einfügen, um die Kategorien einzufügen. Die Kategorien werden zum Fensterbereich hinzugefügt. Es sind keine Häufigkeitswerte bekannt, da Sie keine erneute Extrahierung vorgenommen haben. Daher werden die Zählungen als Fragezeichen (?) angezeigt.

Drucken von Kategorien

Sie können die Baumansicht im Fensterbereich “Kategorien” drucken.

So drucken Sie die Baumansicht für Kategorien:

- ▶ Im Baum im Fensterbereich “Kategorien” können Sie die Baumelemente erweitern, reduzieren bzw. sortieren, je nachdem, was gedruckt werden soll.
- ▶ Wählen Sie in den Menüs die Optionsfolge Datei > Drucken > Kategorien drucken. Das Dialogfeld “Druckvorschau” wird geöffnet.

Abbildung 6-33
Dialogfeld "Druckvorschau"

Kategorie	Konzept	Anzahl	Wert
Alle Datensätze		-	405
Nicht kategorisiert		-	282
Keine Konzepte extrahiert		-	4
songs+<Positive>		3	9
fx [listening to songs + like]			1
fx [songs + like] [songs + meets needs]			7
drag & dropped songs from their site+<Po		1	1
fx [drag & dropped songs from their si			1
memory		2	2
fx [memory + easy to use]			1
fx [storage + able]			1
battery+<Contextual>		2	3
design+<Positive>		2	7
store+<Positive>		2	3
skipping+<Contextual>		2	2
look+<Positive>		2	3
feature+<Positive>		2	3
software+<Positive>		1	3
function		2	2
playlists+<Positive>		1	2
battery life+<Positive>		1	5
headphones+<Positive>		1	2
load+<Positive>		1	2
friends+<Contextual>		1	2
interface+<Positive>		1	4
cd collection+<Positive>		1	2
volume+<Contextual>		1	2
unit+<Positive>		1	1
music		12	28
music+<Positive>		9	24
music+<Contextual>		3	5
sound		7	23
sound+<Positive>		5	23

- Klicken Sie auf die Schaltfläche "Drucken", um die Ansicht so zu drucken, wie sie im Dialogfeld angezeigt wird.

Löschen von Kategorien

Wenn Sie eine Kategorie nicht mehr benötigen, können Sie sie löschen. Wenn Sie eine Kategorie löschen, werden alle Konzepte, die nicht in einer anderen Kategorie verwendet werden, auf der Registerkarte "Nicht verwendete Extrahierungen" im Bereich "Extrahierungsergebnisse" angezeigt.

So löschen Sie eine Kategorie:

- ▶ Wählen Sie im Fensterbereich “Kategorien” die zu löschende(n) Kategorie(n) aus.
- ▶ Wählen Sie im Menü Bearbeiten > Löschen.

Visualisierung von Diagrammen

Bei der Erstellung von Kategorien ist es wichtig, die Definitionen der Kategorien, die darin enthaltenen Antworten und die Überschneidung von Kategorien sorgfältig zu überprüfen. Im Visualisierungsbereich können Kategorien von etlichen Perspektiven betrachtet werden. Der Visualisierungsbereich befindet sich rechts oben im Fenster "Frage". Sollte er noch nicht sichtbar sein, können Sie auf diesen Bereich über das Menü "Ansicht" (Ansicht > Bereiche > Visualisierung) zugreifen.

In dieser Ansicht bietet der Visualisierungsbereich drei Darstellungen der Gemeinsamkeiten bei der Kategorisierung von Antworten. Mit den Grafiken und Diagrammen in diesem Bereich können Sie Ihre Kategorisierungsergebnisse analysieren und Kategorien oder Berichte feiner abstimmen. Bei der Verfeinerung von Kategorien können Sie in diesem Bereich Ihre Kategoriedefinitionen überprüfen, um Kategorien aufzudecken, die zu ähnlich sind (z. B. über 75 % ihrer Antworten gemein haben) oder zu unterschiedlich sind. Wenn zwei Kategorien zu ähnlich sind, kann es sinnvoll sein, die beiden Kategorien zu kombinieren. Alternativ können Sie auch die Kategoriedefinitionen ausarbeiten, indem Sie bestimmte Deskriptoren von der einen oder anderen Kategorie entfernen. Sie können die Ergebnisse in diesem Bereich kopieren, einfügen und drucken, um Ihre Analyse zu unterstützen oder Berichte zu drucken.

Abbildung 7-1
Bereiche "Kategorie" und "Visualisierung"



Je nach Auswahl im Bereich "Extrahierungsergebnisse" oder "Kategorien" können Sie die entsprechenden Interaktionen zwischen Antworten und Kategorien auf allen Registerkarten in diesem Bereich anzeigen. Auf jeder Registerkarte werden ähnliche Informationen dargestellt, jedoch auf unterschiedliche Weise oder unterschiedlichen Detailebenen. Falls erforderlich können Sie die Farben der Grafiken und Diagramme im Dialogfeld "Optionen" anpassen. [Für weitere Informationen siehe Thema Optionen: Registerkarte "Anzeige" in Kapitel 2 auf S. 21.](#)

Anmerkung: Sie können auch Auswertungsdiagramme wie ein Balkendiagramm der Top-5-Kategorien erzeugen. Diese Diagramme, die in HTML exportiert werden, können dann in Präsentationen verwendet werden. [Für weitere Informationen siehe Thema Exportieren von Zusammenfassungsdiagrammen in Kapitel 4 auf S. 60.](#)

Der Visualisierungsbereich bietet die folgenden Diagramme:

- **Kategoriebalkendiagramm.** Eine Tabelle und ein Balkendiagramm zeigen gemäß Ihrer Auswahl und den assoziierten Kategorien die Überschneidungen zwischen den Antworten an. Das Balkendiagramm zeigt auch das jeweilige Verhältnis der Antworten in den Kategorien

zu der Gesamtzahl der Antworten an. Gegebenenfalls können Sie auch eine importierte Referenzvariable auswählen, um die Werte der Referenzvariablen der Datensätze in jeder Kategorie zu vergleichen. [Für weitere Informationen siehe Thema Kategoriebalkendiagramm auf S. 173.](#)

- **Kategorienetzdiagramm.** Dieses Diagramm zeigt die Überschneidung der Antworten der Kategorien an, denen die Antworten gemäß der Auswahl in den anderen Bereichen zugeordnet sind. [Für weitere Informationen siehe Thema Kategorienetzdiagramm auf S. 174.](#)
- **Tabelle für Kategorienetzdiagramm.** Auf dieser Registerkarte werden dieselben Informationen wie auf der Registerkarte "Kategorienetzdiagramm" dargestellt, allerdings als Tabelle. Die drei Spalten in der Tabelle können durch Klicken auf den Spaltentitel sortiert werden. [Für weitere Informationen siehe Thema Tabelle für Kategorienetzdiagramm auf S. 175.](#)

[Für weitere Informationen siehe Thema Kategorisieren von Textdaten in Kapitel 6 auf S. 96.](#)

Kategoriebalkendiagramm

Auf dieser Registerkarte werden eine Tabelle und ein Balkendiagramm angezeigt, die gemäß Ihrer Auswahl und den assoziierten Kategorien die Überschneidungen zwischen den Antworten anzeigen. Das Balkendiagramm zeigt auch das jeweilige Verhältnis der Antworten in den Kategorien zu der Gesamtzahl der Antworten an. Das Layout dieses Diagramms kann nicht geändert werden.

Sie können mithilfe der Kontextmenüs in diesem Balkendiagramm Spalten sortieren, die Diagrammfarben ändern, den Diagramminhalt auswählen, den Inhalt kopieren sowie die Legende ein- und ausblenden.

Das Diagramm enthält die folgenden Spalten:

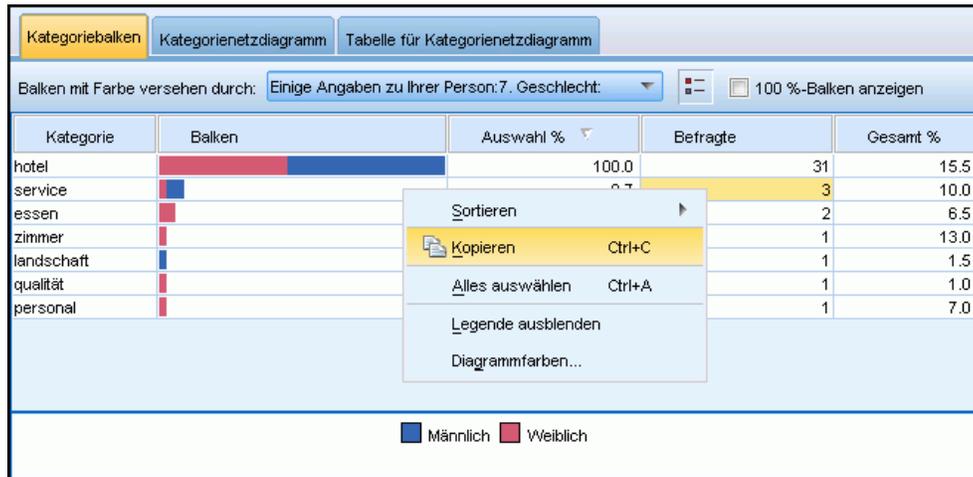
- **Kategorie.** In dieser Spalte wird der Name der Kategorien in Ihrer Auswahl angezeigt. Standardmäßig wird die häufigste Kategorie in Ihrer Auswahl an erster Stelle aufgeführt.
- **Balken.** Diese Spalte zeigt auf visuelle Art das jeweilige Verhältnis der Datensätze in einer Kategorie zu der Gesamtzahl der Datensätze an.
- **Auswahl %.** Diese Spalte zeigt einen Prozentsatz auf der Basis des jeweiligen Verhältnisses der Datensätze in einer Kategorie zu der Gesamtzahl der in der Auswahl enthaltenen Datensätze an.
- **Antworten** In dieser Spalte wird die Anzahl der Antworten in einer Auswahl für die jeweilige Kategorie angezeigt.
- **Gesamt %.** Diese Spalte zeigt einen Prozentsatz auf der Basis des jeweiligen Verhältnisses der Datensätze in einer bestimmten Kategorie zu der Gesamtzahl der Datensätze für diese Frage an (nicht in der Auswahl).

Sie können auch eine verfügbare Referenzvariable aus der Dropdown-Liste wählen, um ihre Werte zu vergleichen. Wenn Sie eine Referenzvariable auswählen, werden die Balken in der Tabelle gemäß den Werten in Ihren Referenzvariablen unterteilt und farbkodiert. Durch Klicken auf den jeweiligen farbkodierten Referenzwert in einem Balken wird der Bereich "Daten" aktualisiert und er zeigt die Unterauswahl der Antworten entsprechend dem Referenzvariablenwert an. Klicken Sie auf die Symboleleistenschaltfläche "Legende", um eine Legende der Referenzvariablenwerte anzuzeigen.

Abbildung 7-2
Symbolleistenschaltfläche "Legende"



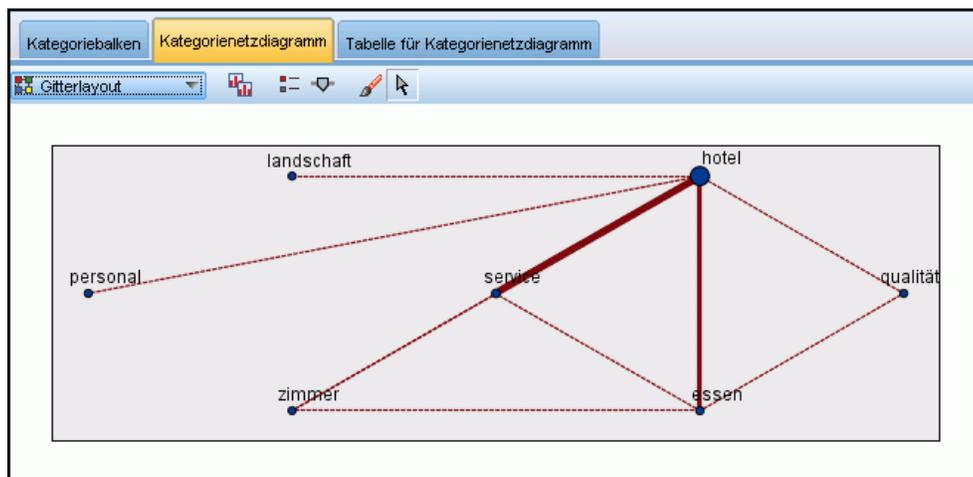
Abbildung 7-3
Visualisierungsbereich: Kategorieleiste



Kategorienetzdiagramm

Auf dieser Registerkarte wird ein Kategorienetzdiagramm angezeigt. Das Netzdiagramm zeigt die Überschneidung der Antworten der Kategorien an, denen die Antworten gemäß der Auswahl in den anderen Bereichen zugeordnet sind. Vorhandene Kategoriebezeichnungen werden im Diagramm angezeigt. Sie können über die Symbolleistenschaltflächen in diesem Bereich das Layout des Diagramms auswählen (Netz-, Kreis-, gerichtetes oder Gitterlayout).

Abbildung 7-4
Kategorienetzdiagramm, Gitterlayout



Im Netzdiagramm stellt jeder Knoten eine Kategorie dar. Sie können mit der Maus innerhalb des Bereichs Knoten auswählen und verschieben. Die Größe des Knotens stellt die relative Größe auf der Basis der Anzahl der Datensätze der Kategorie in Ihrer Auswahl dar. Die Stärke und Farbe der Linie zwischen zwei Kategorien gibt die Anzahl der Datensätze an, die in beiden Kategorien vorhanden sind. Wenn Sie im Untersuchungsmodus den Mauszeiger auf einen Knoten platzieren, zeigt eine QuickInfo den Namen (oder die Bezeichnung) der Kategorie und die Gesamtanzahl an Datensätzen in der Kategorie an.

Anmerkung: Standardmäßig ist der Untersuchungsmodus für die Diagramme aktiviert, in denen Sie Knoten verschieben können. Sie können das Layout Ihrer Diagramme im Bearbeitungsmodus bearbeiten, u. a. Farben und Schriftarten, Legenden usw. [Für weitere Informationen siehe Thema Verwenden von Diagrammsymboleisten und Paletten auf S. 175.](#)

Table für Kategorienetzdiagramm

Auf dieser Registerkarte werden dieselben Informationen wie auf der Registerkarte “Kategorienetzdiagramm” dargestellt, allerdings als Tabelle. Die drei Spalten in der Tabelle können durch Klicken auf den Spaltentitel sortiert werden:

- **Anzahl.** In dieser Spalte wird die Anzahl der in beiden Kategorien vorhandenen oder gemeinsamen Datensätze angezeigt.
- **Kategorie 1.** Diese Spalte zeigt den Namen der ersten Kategorie an, gefolgt von der – in Klammern eingeschlossenen – Gesamtzahl der darin enthaltenen Datensätze.
- **Kategorie 2.** Diese Spalte zeigt den Namen der zweiten Kategorie an, gefolgt von der – in Klammern eingeschlossenen – Gesamtzahl der darin enthaltenen Datensätze.

Abbildung 7-5
Visualisierungsbereich: Tabelle für Kategorienetzdiagramm

Anzahl	Kategorie 1	Kategorie 2
1	laken(2)	lärm(2)
1	zimmer(16)	laken(2)
1	zimmer(16)	lärm(2)
1	zimmer(16)	klimaanlage(3)
1	zimmer(16)	badetuche im zimmer(1)
1	zimmer(16)	fön im zimmer(1)
1	zimmer(16)	geldschrank im zimmer...

Verwenden von Diagrammsymboleisten und Paletten

Für das Kategorienetzdiagramm steht eine Symboleiste zur Verfügung, über die Sie schnell auf häufig verwendete Paletten zugreifen können, mit denen Sie zahlreiche Aktionen mit Ihren Diagrammen durchführen können. Sie können zwischen den folgenden Ansichtsmodi wählen: *Untersuchen* oder *Bearbeiten*.

Während Sie im Sondierungsmodus die durch die Visualisierung dargestellten Daten und Werte erforschen, gestattet Ihnen der Bearbeitungsmodus, das Layout und Aussehen der Visualisierung zu ändern. Sie können beispielsweise die Schriftarten und die Farben so ändern, dass sie den Stilvorgaben Ihres Unternehmens entsprechen. In diesen Modus wechseln Sie, indem Sie Ansicht > Bereich "Visualisierung" > Bearbeitungsmodus aus den Menüs wählen (oder auf das entsprechende Symbol in der Symbolleiste klicken).

Im Bearbeitungsmodus stehen mehrere Symbolleisten zur Verfügung, mit denen sich die verschiedenen Aspekte des Visualisierungslayouts beeinflussen lassen. Wenn Sie einige der Symbolleisten nicht benötigen, können Sie sie ausblenden und im Dialogfeld, in dem die Grafik angezeigt wird, mehr Platz schaffen. Sie blenden Symbolleisten aus bzw. deaktivieren sie, indem Sie im Menü "Ansicht" auf den entsprechenden Symbolleisten- oder Palettenamen klicken.

Tabelle 7-1
Schaltflächen der Symbolleiste "Textanalyse"

Schaltfläche/Liste	Beschreibung
	Aktiviert den Bearbeitungsmodus. Wechseln Sie in den Bearbeitungsmodus, wenn Sie das Aussehen des Diagramms ändern möchten, z. B. die Schrift vergrößern, die Farben Ihrer Unternehmensrichtlinien anpassen oder Bezeichnungen und Legenden entfernen möchten. Für weitere Informationen siehe Thema Bearbeiten von Visualisierungen auf S. 176.
	Aktiviert den Untersuchungsmodus. Standardmäßig ist der Untersuchungsmodus aktiv; das bedeutet, dass Sie Knoten im Diagramm verschieben und ziehen und auch mit der Maus über Diagrammobjekte fahren können, um weitere QuickInfos anzuzeigen.
	Wählen Sie die Art der Netzanzeige für die Diagramme. <ul style="list-style-type: none"> ■ Kreislayout. Allgemeines Layout, das für alle Diagramme verwendet werden kann. Bei diesem Layout wird das Diagramm unter der Voraussetzung erstellt, dass alle Links ungerichtet sind, und alle Knoten werden gleich behandelt. Knoten werden nur um den Kreisumfang herum platziert. ■ Netzlayout. Allgemeines Layout, das für alle Diagramme verwendet werden kann. Bei diesem Layout wird das Diagramm unter der Voraussetzung erstellt, dass alle Links ungerichtet sind, und alle Knoten werden gleich behandelt. Knoten werden frei innerhalb des Layouts platziert. ■ Gerichtetes Layout. Layout, das nur für gerichtete Diagramme verwendet werden sollte. In diesem Layout werden baumförmige Strukturen von den Stammknoten bis zu den Blattknoten erstellt und nach Farben sortiert. ■ Gitterlayout. Allgemeines Layout, das für alle Diagramme verwendet werden kann. Bei diesem Layout wird das Diagramm unter der Voraussetzung erstellt, dass alle Links ungerichtet sind, und alle Knoten werden gleich behandelt. Knoten werden nur an Gitterpunkten innerhalb dieses Bereichs platziert.
	Eine Umschaltfläche, mit der die Legende angezeigt wird. Wenn diese Schaltfläche nicht gedrückt ist, wird die Legende nicht angezeigt.
	Eine Umschaltfläche, mit der der Links-Schieberegler unter dem Diagramm angezeigt wird. Sie können die Ergebnisse durch Verschieben des Pfeils filtern.

Bearbeiten von Visualisierungen

Zur Bearbeitung einer Visualisierung im **Bearbeitungsmodus** sind mehrere Optionen verfügbar. Sie haben folgende Möglichkeiten:

- Bearbeiten und formatieren Sie den Text.
- Ändern Sie die Füllfarbe, die Transparenz und das Muster von Rahmen und Grafikelementen.

- Ändern Sie die Farbe und das Strichmuster von Rahmen und Linien.
- Drehen und ändern Sie die Form und das Verhältnis von Punktelementen.
- Ändern Sie die Größe von Grafikelementen (beispielsweise Balken und Punkte).
- Passen Sie den Platz um Objekte an, indem Sie Ränder und Abstände verwenden.
- Geben Sie Formate für Zahlen an.
- Ändern Sie die Einstellungen für Achsen und Skalierungen.
- Sortieren Sie Kategorien auf einer kategorialen Achse oder schließen Sie Kategorien aus oder blenden Sie sie aus.
- Legen Sie die Orientierung von Feldern.
- Weisen Sie einem Koordinatensystem Transformationen zu.
- Ändern Sie Statistiken, Grafikelementtypen und Kollisionsmodifikatoren.
- Ändern Sie die Position der Legende.
- Weisen Sie Visualisierungs-Stilvorlagen hinzu.

Die folgenden Themen beschreiben, wie diese verschiedenen Aufgaben ausgeführt werden. Es wird auch empfohlen, dass Sie die allgemeinen Regeln zur Bearbeitung von Grafiken lesen.

So wechseln Sie in den Bearbeitungsmodus:

- ▶ Wählen Sie in den Menüs folgende Befehlsfolge:
Ansicht > Bearbeitungsmodus

Allgemeine Regeln zur Bearbeitung von Visualisierungen

Bearbeitungsmodus

Alle Bearbeitungen finden im Bearbeitungsmodus statt. Wählen Sie zum Aktivieren des Bearbeitungsmodus die folgenden Befehle aus den Menüs aus:

Ansicht > Bearbeitungsmodus

Auswahl

Die verfügbaren Optionen für die Bearbeitung hängen von der Auswahl ab. Abhängig von der Auswahl sind auch unterschiedliche Optionen der Symbolleiste und der Werkzeugpalette aktiviert. Nur die aktivierten Objekte gelten für die aktuelle Auswahl. Wenn beispielsweise eine Achse ausgewählt ist, stehen im Fenster "Eigenschaften" die Registerkarten "Skala", "Hauptteilstriche" und "Hilfsteilstriche" zur Verfügung.

Hier finden Sie einige Tipps zur Auswahl von Objekten in der Visualisierung:

- Klicken Sie auf ein Objekt, um es auszuwählen.

- Wählen Sie ein Grafikelement (z. B. Punkte in einem Streudiagramm oder Balken in einem Balkendiagramm) mit einem einzigen Klick aus. Klicken Sie nach der anfänglichen Auswahl erneut, um die Auswahl auf Gruppen von Grafikelementen oder ein einziges Grafikelement einzuschränken.
- Drücken Sie Esc, um die Auswahl vollständig aufzuheben.

Paletten

Wenn ein Objekt in der Visualisierung ausgewählt ist, werden die verschiedenen Paletten passend zur Auswahl aktualisiert. Die Paletten enthalten Steuerungen für die Bearbeitung der Auswahl. Paletten können Symbolleisten oder ein Fenster mit mehreren Steuerungen und Registerkarten sein. Paletten können ausgeblendet sein, stellen Sie also sicher, dass die erforderliche Palette für die jeweilige Bearbeitung angezeigt ist. Prüfen Sie im Menü "Ansicht", welche Paletten derzeit angezeigt werden.

Sie können die Paletten umpositionieren, indem Sie in den leeren Bereich in der Symbolleistenpalette oder an der linken Seite anderer Paletten klicken und ihn an eine andere Stelle ziehen. Die Stellen, an denen Sie die Palette andocken können, werden visuell gekennzeichnet. Für Paletten, die keine Systemleisten sind, können Sie auch auf das Schließfeld klicken, um die Palette auszublenden, und auf die Schaltfläche zum Loslösen, um die Palette in einem separaten Fenster zu zeigen. Klicken Sie auf die Schaltfläche "Hilfe", um Hilfe zur jeweiligen Palette zu erhalten.

Automatische Einstellungen

Einige Einstellungen bieten die Option -auto-. Das zeigt an, dass automatische Werte zugewiesen sind. Welche automatischen Einstellungen benutzt werden, hängt von der jeweiligen Visualisierung und den Datenwerten ab. Sie können einen Wert eingeben, um die automatische Einstellung zu überschreiben. Wenn Sie die automatische Einstellung wiederherstellen möchten, löschen Sie den aktuellen Wert und drücken Sie die Eingabetaste. Die Einstellung zeigt wieder -auto- an.

Ein- und Ausblenden von Elementen

Sie können zahlreiche Elemente in der Visualisierung ein- oder ausblenden. Sie können beispielsweise die Legende oder Achsenbeschriftung ausblenden. Sie löschen ein Element, indem Sie es auswählen und die Löschtaste drücken. Wenn das Element nicht gelöscht werden darf, geschieht nichts. Wenn Sie ein Element versehentlich löschen, drücken Sie Strg+Z, um den Löschvorgang zu widerrufen.

Status

Einige Symbolleisten reflektieren den Status der aktuellen Auswahl, andere nicht. Das Fenster "Eigenschaften" reflektiert den Status immer. Wenn eine Symbolleiste den Status *nicht* reflektiert, wird dies im Thema zur Beschreibung der Symbolleiste vermerkt.

Bearbeiten und Formatieren von Text

Sie können Text an seiner Position bearbeiten und die Formatierung eines ganzen Textblocks ändern. Beachten Sie, dass Sie keinen Text bearbeiten können, der direkt mit Datenwerten verknüpft ist. Sie können beispielsweise eine Teilstrichbeschriftung nicht bearbeiten, da der Inhalt der Beschriftung aus den zugrunde liegenden Daten ermittelt wird. Sie können jedoch jeden Text in der Visualisierung formatieren.

So bearbeiten Sie Text direkt:

- ▶ Doppelklicken Sie auf den Textblock. Dadurch wird der vollständige Text ausgewählt. Alle Symbolleisten werden deaktiviert, da Sie während der Bearbeitung von Text keinen anderen Bereich der Visualisierung ändern können.
- ▶ Schreiben Sie neuen Text, um den bestehenden Text zu ersetzen. Sie können auch erneut in den Text klicken, um einen Cursor anzuzeigen. Positionieren Sie den Cursor wie gewünscht und geben Sie den zusätzlichen Text ein.

So formatieren Sie Text

- ▶ Wählen Sie den Rahmen aus, der den Text enthält. Doppelklicken Sie nicht auf den Text.
- ▶ Formatieren Sie den Text mithilfe der Symbolleiste "Schriftarten". Wenn die Symbolleiste nicht aktiviert ist, stellen Sie sicher, dass nur der *Rahmen*, der den Text enthält, ausgewählt ist. Wenn der Text ausgewählt ist, ist die Symbolleiste deaktiviert.

Abbildung 7-6
Symbolleiste "Schriftart"



Folgende Schriftänderungen sind möglich:

- Farbe
- Schriftfamilie (z. B. Arial oder Verdana)
- Größe (Die Einheit ist Punkt, sofern Sie keine andere Einheit angeben.)
- Stärke
- Ausrichtung in Relation zum Textrahmen

Die Formatierung gilt für den gesamten Text im Rahmen. Die Formatierung einzelner Buchstaben oder Wörter in einem bestimmten Textblock lässt sich nicht ändern.

Ändern von Farben, Mustern, Strichmustern und Transparenz

Viele verschiedene Elemente in einer Visualisierung verfügen über eine Füllung und einen Rahmen. Ein deutliches Beispiel ist ein Balken in einem Balkendiagramm. Die Farbe der Balken ist die Füllfarbe. Sie können auch von einem durchgehenden schwarzen Rahmen eingefasst sein.

In der Visualisierung gibt es weniger deutliche Beispiele mit Füllfarben. Wenn die Füllfarbe transparent ist, wissen Sie womöglich gar nicht, dass eine Füllung vorhanden ist. Betrachten Sie beispielsweise den Text in einer Achsenbeschriftung. Es sieht so aus, als wäre es ein “fließender” Text, aber in Wirklichkeit befindet er sich in einem Rahmen mit einem transparenten Füllmuster. Sie können den Rahmen sehen, wenn Sie die Achsenbeschriftung auswählen.

Jeder Rahmen in einer Visualisierung kann über ein Füllmuster und eine Rahmenart verfügen, auch der Rahmen um die gesamte Visualisierung. Zudem ist mit jedem Füllmuster ein Grad der Lichtdurchlässigkeit/Transparenz verbunden, der sich anpassen lässt.

So ändern Sie Farben, Muster, Strichmuster und Transparenz:

- ▶ Wählen Sie das Element aus, das Sie formatieren möchten. Wählen Sie beispielsweise die Balken in einem Balkendiagramm oder einen Textrahmen aus. Wenn die Visualisierung durch eine kategoriale Variable oder ein Feld getrennt ist, können Sie auch die Gruppe auswählen, die einer individuellen Kategorie entspricht. Hiermit können Sie die Standardformatierung für diese Gruppe ändern. Sie können beispielsweise die Farbe von einer der Stapelgruppen in einem gestapelten Balkendiagramm ändern.
- ▶ Verwenden Sie die Symbolleiste “Farbe”, um die Füllfarbe, die Rahmenfarbe oder das Füllmuster zu ändern.

Abbildung 7-7
Symbolleiste “Farbe”

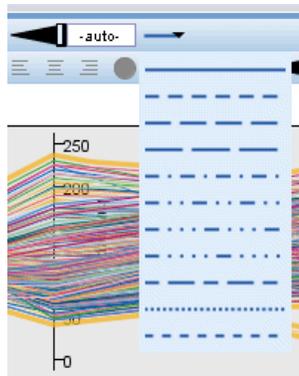


Anmerkung: Diese Symbolleiste reflektiert nicht den Status der aktuellen Auswahl.

Sie können eine Farbe oder ein Füllmuster ändern, indem Sie auf die Schaltfläche klicken, um die angezeigte Option auszuwählen, oder auf den Dropdown-Pfeil, um eine andere Option zu wählen. Beachten Sie, dass eine der Farboptionen weiß mit einer roten diagonalen Linie aussieht. Dies ist die transparente Farbe. Sie können sie beispielsweise verwenden, um den Rahmen von Balken in einem Histogramm auszublenden.

- Die erste Schaltfläche steuert die Füllfarbe.
 - Die zweite Schaltfläche steuert die Rahmenfarbe.
 - Die dritte Schaltfläche steuert das Füllmuster. Das Füllmuster verwendet die Rahmenfarbe. Daher ist das Füllmuster nur sichtbar, wenn es eine sichtbare Rahmenfarbe gibt.
 - Die vierte Steuerung besteht aus einem Schieberegler und einem Textfeld und dient zur Steuerung der Lichtundurchlässigkeit der Füllfarbe und des Füllmusters. Ein geringerer Prozentsatz bedeutet weniger Undurchlässigkeit und mehr Transparenz. 100 % ist völlig undurchlässig (keine Transparenz).
- ▶ Um das Strichmuster eines Rahmens oder einer Linie zu ändern, verwenden Sie die Linien-Symbolleiste.

Abbildung 7-8
Linien-Symbolleiste



Anmerkung: Diese Symbolleiste reflektiert nicht den Status der aktuellen Auswahl.

Wie bei der anderen Symbolleiste können Sie auf die Schaltfläche klicken, um die angezeigte Option auszuwählen, oder auf den Dropdown-Pfeil, um eine andere Option zu wählen.

Drehen und Ändern der Form und des Verhältnisses von Punktelementen

Sie können Punktelemente drehen, ihnen eine andere vordefinierte Form zuweisen oder das Größenverhältnis (Breite zu Höhe) ändern.

So ändern Sie Punktelemente:

- ▶ Wählen Sie die Punktelemente aus. Sie können die Form und das Seitenverhältnis einzelner Punktelemente nicht drehen oder ändern.
- ▶ Verwenden Sie die Piktogramm-Symbolleiste, um die Punkte zu ändern.

Abbildung 7-9
Piktogramm-Symbolleiste



- Mit der ersten Schaltfläche können Sie die Form der Punkte ändern. Klicken Sie auf den Dropdown-Pfeil und wählen Sie eine vordefinierte Form aus.
- Mit der zweiten Schaltfläche können Sie die Punkte an eine bestimmte Kompassposition drehen. Klicken Sie auf den Dropdown-Pfeil und ziehen Sie dann die Nadel an die gewünschte Position.
- Mit der dritten Schaltfläche können Sie das Seitenverhältnis ändern. Klicken Sie auf den Dropdown-Pfeil und ziehen Sie dann das angezeigte Rechteck in die gewünschte Form. Die Form des Rechtecks repräsentiert das Seitenverhältnis.

Ändern der Größe von Grafikelementen

Sie können die Größe der Grafikelemente in der Visualisierung ändern. Hierzu gehören Balken, Punkte, Linien usw. Wenn die Größe des Grafikelements durch eine Variable oder ein Feld eingestellt wird, handelt es sich dabei um die *Mindestgröße*.

So ändern Sie die Größe von Grafikelementen:

- ▶ Wählen Sie die Grafikelemente aus, deren Größe Sie ändern möchten.
- ▶ Verwenden Sie den Schieberegler oder geben Sie eine bestimmte Größe für die Option in der Piktogramm-Symboleiste ein. Die Einheit ist Pixel, sofern Sie keine andere Einheit festlegen. (Eine vollständige Liste der Einheitenabkürzungen siehe unten.) Sie können auch einen Prozentwert (z. B. 30 %) eingeben, um festzulegen, dass ein Grafikelement den angegebenen prozentualen Anteil des verfügbaren Platzes einnimmt. Der verfügbare Platz hängt vom Typ des Grafikelements und der spezifischen Visualisierung ab.

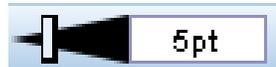
Tabelle 7-2

Gültige Abkürzungen für Maßeinheiten

Abkürzung	Einheit
cm	Zentimeter
in	Inch
mm	Millimeter
pc	Pica
pt	Punkt
px	Pixel

Abbildung 7-10

Größensteuerung an der Piktogramm-Symboleiste



Festlegen von Rändern und Abständen

Wenn um oder im Rahmen in der Visualisierung zu viel oder zu wenig Platz ist, können Sie seine Rand- und Abständeinstellungen ändern. Der **Rand** ist die Menge an Platz zwischen dem Rahmen und benachbarten anderen Elementen. Der **Abstand** ist die Menge an Platz zwischen der Begrenzungslinie des Rahmens und dem *Inhalt* des Rahmens.

So legen Sie Ränder und Abstände fest:

- ▶ Wählen Sie den Rahmen aus, für den Sie Ränder und Abstände festlegen möchten. Dabei kann es sich um einen Textrahmen, den Rahmen um eine Legende oder sogar den Datenrahmen handeln, der die Grafikelemente (z. B. Balken und Punkte) anzeigt.
- ▶ Geben Sie die Einstellungen in der Registerkarte "Ränder" im Fenster "Eigenschaften" an. Alle Größen sind in Pixel angegeben, sofern Sie keine andere Einheit wählen (z. B. cm oder in).

Abbildung 7-11
Registerkarte "Ränder"



Formatieren von Zahlen

Sie können das Format für Zahlen in Teilstrichbeschriftungen auf einer fortlaufenden Achse sowie bei Datenwertelabels mit Zahlen angeben. Beispielsweise können Sie festlegen, dass an den Hauptteilstrichen angegebene Zahlen in Tausendern gezeigt werden.

So geben Sie Zahlenformate an:

- ▶ Wählen Sie die Teilstrichbeschriftungen der fortlaufenden Achse oder die Datenwertelabels aus, wenn sie Zahlen enthalten.
- ▶ Klicken Sie im Fenster "Eigenschaften" auf die Registerkarte Format.

Abbildung 7-12
Registerkarte "Format"



- ▶ Wählen Sie die gewünschten Zahlenformatoptionen aus.

Präfix. Ein Zeichen, das vor der Zahl angezeigt werden soll. Geben Sie beispielsweise ein Dollarzeichen (\$) ein, wenn es sich bei den Zahlen um Gehälter in US-Dollar handelt.

Suffix. Ein Zeichen, das nach der Zahl angezeigt werden soll. Geben Sie beispielsweise ein Prozentzeichen (%) ein, wenn es sich bei den Zahlen um Prozentsätze handelt.

Min. Ganzzahlstellen. Mindestanzahl an Stellen, die im ganzzahligen Teil der Dezimaldarstellung angezeigt werden sollen. Wenn der tatsächliche Wert nicht über die Mindestanzahl an Stellen verfügt, wird der ganzzahlige Teil des Wertes mit Nullen aufgefüllt.

Max. Ganzzahlstellen. Höchstanzahl an Stellen, die im ganzzahligen Teil der Dezimaldarstellung angezeigt werden sollen. Wenn der tatsächliche Wert die Höchstanzahl an Stellen überschreitet, wird der ganzzahlige Teil des Wertes durch Sternchen ersetzt.

Min. Dezimalstellen. Mindestanzahl an Stellen, die im Dezimalteil der Dezimal- oder wissenschaftlichen Darstellung angezeigt werden sollen. Wenn der tatsächliche Wert nicht über die Mindestanzahl an Stellen verfügt, wird der Dezimalteil des Wertes mit Nullen aufgefüllt.

Max. Dezimalstellen. Höchstanzahl an Stellen, die im Dezimalteil der Dezimal- oder wissenschaftlichen Darstellung angezeigt werden sollen. Wenn der tatsächliche Wert die Höchstanzahl an Stellen überschreitet, wird der Dezimalwert auf die passende Anzahl an Stellen gerundet.

Wissenschaftlich. Ob Zahlen in wissenschaftlicher Notation angezeigt werden sollen. Die wissenschaftliche Notation ist für sehr große oder sehr kleine Zahlen sinnvoll. -auto- überlässt es der Anwendung zu entscheiden, ob wissenschaftliche Notation angemessen ist.

Skalierung. Ein Skalierungsfaktor, der eine Zahl ist, durch die der Originalwert dividiert wird. Verwenden Sie einen Skalierungsfaktor, wenn die Zahlen groß sind, Sie jedoch nicht möchten, dass die Teilstrichbeschriftung durch die Anzeige der großen Zahl übermäßig lang wird. Wenn Sie das Zahlenformat der Teilstrichbeschriftungen ändern, bearbeiten Sie auch den Achsentitel, um anzugeben, wie die Zahl zu interpretieren ist. Nehmen wir an, auf der Skalenachse werden Gehälter angezeigt und die Labels lauten 30.000, 50.000 und 70.000. Hier können Sie einen Skalierungsfaktor von 1.000 eingeben, um 30, 50 und 70 anzuzeigen. In diesem Fall wäre es sinnvoll, den Skalenachsensentitel zu bearbeiten, um den Text in Tausend einzuschließen.

Klammern für -ve. Ob negative Werte eingeklammert werden sollen.

Gruppierung. Ob ein bestimmtes Zeichen zwischen Zifferngruppen angezeigt werden soll. Das aktuelle Gebietschema Ihres Computers bestimmt, welches Zeichen zur Zifferngruppierung verwendet wird.

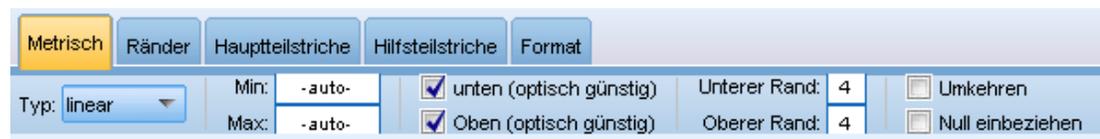
Ändern der Einstellungen für Achsen und Skalen

Für das Ändern von Achsen und Skalen stehen mehrere Optionen zur Verfügung.

So ändern Sie Einstellungen für Achsen und Skalen:

- ▶ Wählen Sie einen beliebigen Teil der Achse aus (z. B. die Achsenbeschriftung oder Teilstrichbeschriftungen).
- ▶ Verwenden Sie die Registerkarten “Skala”, “Hauptteilstriche” und “Hilfsteilstriche” im Fenster “Eigenschaften”, um die Einstellungen für Achsen und Skalen zu ändern.

Abbildung 7-13
Fenster “Eigenschaften”



Registerkarte “Skala”

Typ. Legt fest, ob die Skala linear oder transformiert ist. Skalentransformationen sind hilfreich beim Verständnis der Daten und bei den für Statistiken erforderlichen Annahmen. Bei Streudiagrammen können Sie eine transformierte Skala verwenden, wenn die Beziehung zwischen den unabhängigen und den abhängigen Variablen oder Feldern nicht linear ist. Skalentransformationen können außerdem verwendet werden, um ein schiefes Histogramm symmetrischer zu machen, sodass es eher wie eine Normalverteilung aussieht. Beachten Sie, dass Sie nur die Skala transformieren, auf der die Daten angezeigt werden. Die tatsächlichen Daten werden nicht transformiert.

- **linear.** Gibt eine lineare, nicht transformierte Skala an.
- **log.** Gibt eine mit dem Logarithmus zur Basis transformierte Skala an. Um Nullwerte und negative Werte zu berücksichtigen, verwendet diese Transformation eine modifizierte Version der Logarithmusfunktion. Diese "sichere" Log-Funktion ist als $\text{sign}(x) * \log(1 + \text{abs}(x))$ definiert. Also ist `safeLog(-99)` gleich:

$$\text{sign}(-99) * \log(1 + \text{abs}(-99)) = -1 * \log(1 + 99) = -1 * 2 = -2$$
- **power.** Gibt eine exponententtransformierte Skala mit einem Exponenten von 0,5 an. Um negative Werte zu berücksichtigen, verwendet diese Transformation eine modifizierte Version der Potenzfunktion. Diese "sichere" Potenzfunktion ist als $\text{sign}(x) * \text{pow}(\text{abs}(x), 0.5)$ definiert. Also ist `safePower(-100)` gleich:

$$\text{sign}(-100) * \text{pow}(\text{abs}(-100), 0.5) = -1 * \text{pow}(100, 0.5) = -1 * 10 = -10$$

Min/Max/Guter unterer Wert/Guter oberer Wert. Gibt den Bereich für die Skala an. Wenn Guter unterer Wert und Guter oberer Wert ausgewählt sind, kann die Anwendung auf der Basis der Daten eine geeignete Skala wählen. Es handelt sich um "gute" untere und obere Werte, da sie in der Regel um einige Werte größer oder kleiner als die Höchst- und Mindestdatenwerte sind. Wenn die Daten beispielsweise von 4 bis 92 reichen, kann ein guter unterer und oberer Wert für die Skala 0 bzw. 100 sein, anstatt die tatsächlichen Mindest- und Höchstwerte aus den Daten zu übernehmen. Achten Sie darauf, dass Sie keinen Bereich festlegen, der zu klein ist und wichtige Elemente verdeckt. Beachten Sie auch, dass Sie keinen expliziten Mindest- und Höchstwert festlegen können, wenn die Option Null einschließen aktiviert ist.

Unterer Rand/Oberer Rand. Erstellen Sie Ränder am unteren und/oder oberen Ende der Achse. Der Rand erscheint senkrecht zur ausgewählten Achse. Die Größeneinheit ist Pixel, sofern Sie keine andere Einheit wählen (z. B. cm oder in). Wenn Sie beispielsweise den oberen Rand für die vertikale Achse als 5 festlegen, verläuft ein horizontaler Rand in der Breite von 5 Pixel am oberen Rand des Datenrahmens.

Umkehren. Legt fest, ob die Skala umgekehrt ist.

Null einschließen. Gibt an, dass die Skala 0 enthalten muss. Diese Option wird häufig für Balkendiagramme verwendet, um sicherzustellen, dass die Balken bei 0 beginnen und nicht bei einem Wert in der etwaigen Höhe des kleinsten Balkens. Wenn diese Option aktiviert ist, sind Min und Max deaktiviert, da Sie keinen benutzerdefinierten Mindest- oder Höchstwert für den Skalenbereich festlegen können.

Registerkarten "Hauptteilstriche" und "Hilfsteilstriche"

Teilstriche oder **Einteilungsstriche** sind die Striche auf einer Achse. Sie geben Werte in bestimmten Intervallen oder Kategorien an. **Hauptteilstriche** sind die Einteilungsstriche mit Beschriftung. Sie sind auch länger als andere Einteilungsstriche. **Hilfsteilstriche** sind Einteilungsstriche zwischen den Hauptteilstrichen. Einige Optionen hängen von der Art der Einteilungsstriche ab, aber die meisten Optionen sind für Hauptteilstriche und Hilfsteilstriche verfügbar.

Teilstriche anzeigen. Gibt an, ob in einem Diagramm Haupt- oder Hilfsteilstriche angezeigt werden sollen.

Gitterlinien einblenden. Gibt an, ob an den Haupt- oder Hilfsteilstrichen Gitterlinien angezeigt werden. **Gitterlinien** sind Linien, die auf einem ganzen Diagramm parallel zu den Achsen verlaufen.

Position. Legt die Position der Teilstriche relativ zur Achse fest.

Länge. Legt die Länge der Teilstriche fest. Die Größeneinheit ist Pixel, sofern Sie keine andere Einheit wählen (z. B. cm oder in).

Basis. *Gilt nur für Hauptteilstriche.* Gibt den Wert an, bei dem der erste Hauptteilstrich erscheint.

Delta. *Gilt nur für Hauptteilstriche.* Gibt den Abstand zwischen Hauptteilstrichen an. Das bedeutet, dass ein Hauptteilstrich bei jedem n . Wert erscheint, wobei n den Delta-Wert angibt.

Unterteilungen. *Gilt nur für Hilfsteilstriche.* Gibt die Anzahl der Hilfsteilstriche zwischen Hauptteilstrichen an. Die Anzahl der Hilfsteilstriche ist um 1 weniger als die Anzahl der Unterteilungen. Nehmen Sie einmal an, dass sich Hauptteilstriche an den Positionen 0 und 100 befinden. Wenn Sie 2 als Anzahl der Hilfsteilstrichunterteilungen eingeben, befindet sich *ein* Hilfsteilstrich an der Position 50 und teilt den Bereich 0–100 in *zwei* Unterteilungen.

Bearbeiten von Kategorien

Sie können die Kategorien auf einer kategorialen Achse auf mehrere Arten bearbeiten:

- Die Sortierfolge für die Anzeige der Kategorien kann geändert werden.
- Bestimmte Kategorien können ausgeschlossen werden.
- Sie können Kategorien hinzufügen, die nicht im Daten-Set angezeigt werden.
- Kleinere Kategorien lassen sich in einer Kategorie zusammenfassen oder kombinieren.

So ändern Sie die Sortierfolge von Kategorien:

- ▶ Wählen Sie eine kategoriale Achse aus. Das Fenster “Kategorien” zeigt die Kategorien auf der Achse an.

Anmerkung: Wenn das Fenster nicht sichtbar ist, stellen Sie sicher, dass Sie es aktiviert haben.

- ▶ Wählen Sie im Fenster “Kategorien” eine Sortieroption aus der Dropdown-Liste:

Benutzerdefiniert. Sortierung der Kategorien in der Reihenfolge, in der sie im Fenster angezeigt werden. Verschieben Sie die Kategorien mithilfe der Pfeilschaltflächen an den Beginn der Liste, nach oben, nach unten oder an das Ende der Liste.

Daten. Sortierung der Kategorien in der Reihenfolge, in der sie im Daten-Set angezeigt werden.

Name. Sortierung der Kategorien in alphabetischer Reihenfolge nach den Namen, die im Fenster angezeigt werden.

Wert. Sortierung der Kategorien nach dem zugrunde liegenden Datenwert unter Verwendung der Werte, die im Fenster eingeklammert angezeigt werden. Nur Datenquellen mit Metadaten (z. B. IBM® SPSS® Statistics Datendateien) unterstützen diese Option.

Statistik. Sortierung von Kategorien auf der Basis der berechneten Statistik für jede Kategorie. Zu den Statistiken zählen beispielsweise die Anzahl, Prozentsätze und Mittelwerte. Diese Option ist nur verfügbar, wenn im Diagramm eine Statistik verwendet wird.

So fügen Sie eine Kategorie hinzu:

Standardmäßig sind nur Kategorien verfügbar, die im Daten-Set angezeigt werden. Sie können der Visualisierung bei Bedarf eine Kategorie hinzufügen.

- ▶ Wählen Sie eine kategoriale Achse aus. Das Fenster “Kategorien” zeigt die Kategorien auf der Achse an.

Anmerkung: Wenn das Fenster nicht sichtbar ist, stellen Sie sicher, dass Sie es aktiviert haben.

- ▶ Klicken Sie im Fenster “Kategorien” auf die Schaltfläche “Kategorie hinzufügen”:

Abbildung 7-14
Schaltfläche “Kategorie hinzufügen”



- ▶ Geben Sie im Dialogfeld “Neue Kategorie hinzufügen” einen Namen für die Kategorie ein.
- ▶ Klicken Sie auf OK.

So schließen Sie bestimmte Kategorien aus:

- ▶ Wählen Sie eine kategoriale Achse aus. Das Fenster “Kategorien” zeigt die Kategorien auf der Achse an.

Anmerkung: Wenn das Fenster nicht sichtbar ist, stellen Sie sicher, dass Sie es aktiviert haben.

- ▶ Wählen Sie im Fenster “Kategorien” in der Liste “Einschließen” einen Kategorienamen aus und klicken Sie dann auf die Schaltfläche “X”. Um die Kategorie wieder zurückzuverschieben, wählen Sie ihren Namen in der Liste “Ausgeschlossen” aus und klicken dann auf den Pfeil rechts neben der Liste.

So kombinieren Sie kleinere Kategorien bzw. fassen sie zusammen:

Sie können Kategorien zusammenfassen, die so klein sind, dass sie nicht gesondert angezeigt werden müssen. Bei einem Kreisdiagramm mit vielen Kategorien könnten beispielsweise alle Kategorien mit einem Prozentsatz von weniger als 10 zusammengefasst werden. Das Zusammenfassen ist nur für additive Statistiken verfügbar. Sie können beispielsweise keine Mittelwerte addieren, da Mittelwerte nicht additiv sind. Daher ist das Kombinieren bzw. Zusammenfassen von Kategorien mithilfe eines Mittelwerts nicht möglich.

- ▶ Wählen Sie eine kategoriale Achse aus. Das Fenster “Kategorien” zeigt die Kategorien auf der Achse an.

Anmerkung: Wenn das Fenster nicht sichtbar ist, stellen Sie sicher, dass Sie es aktiviert haben.

- ▶ Wählen Sie im Fenster “Kategorien” die Option Zusammenfassen aus und geben Sie einen Prozentsatz an. Alle Kategorien, deren prozentualer Anteil die angegebene Zahl unterschreitet, werden in eine Kategorie zusammengefasst. Der Prozentsatz basiert auf der im Diagramm angezeigten Statistik. Die Zusammenfassung ist nur für häufigkeitsbasierte und Summenstatistiken verfügbar.

Ändern der Orientierung von Feldern

Wenn Sie in Ihrer Visualisierung Felder verwenden, können Sie deren Orientierung ändern.

So ändern Sie die Orientierung der Felder:

- ▶ Klicken Sie auf einen beliebigen Bereich der Visualisierung.
- ▶ Klicken Sie im Fenster “Eigenschaften” auf die Registerkarte Felder.

Abbildung 7-15
Registerkarte “Felder”



- ▶ Wählen Sie eine Option aus Layout:

Tabelle. Ordnet Felder wie eine Tabelle an, indem jedem einzelnen Wert eine Zeile oder Spalte zugewiesen wird.

Transponiert. Ordnet Felder wie eine Tabelle an, vertauscht aber auch die ursprünglichen Zeilen und Spalten. Diese Option ist nicht identisch mit dem Transponieren des Diagramms. Beachten Sie, dass die x -Achse und die y -Achse unverändert bleiben, wenn Sie diese Option auswählen.

Liste. Ordnet Felder wie eine Tabelle an, in der jede Zelle eine Kombination von Werten repräsentiert. Spalten und Zeilen sind nicht mehr einzelnen Werten zugeordnet. Diese Option ermöglicht, dass die Felder bei Bedarf umbrechen.

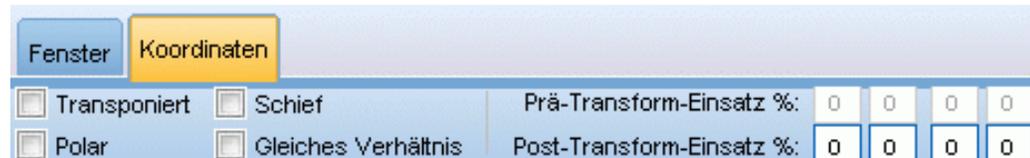
Transformieren des Koordinatensystems

Viele Visualisierungen werden in einem flachen, rechteckigen Koordinatensystem angezeigt. Sie können das Koordinatensystem wie erforderlich umformen. Sie können dem Koordinatensystem beispielsweise eine polare Transformation zuweisen, schräge Schatteneffekte hinzufügen und die Achsen transponieren. Sie können diese Transformationen auch rückgängig machen, wenn sie der aktuellen Visualisierung bereits zugewiesen sind. Beispiel: Ein Kreisdiagramm ist in einem Polarkoordinatensystem gezeichnet. Nach Wunsch können Sie die Polartransformation rückgängig machen und das Kreisdiagramm als einzelnes gestapeltes Balkendiagramm in einem rechteckigen Koordinatensystem anzeigen.

So transformieren Sie das Koordinatensystem:

- ▶ Wählen Sie das zu transformierende Koordinatensystem aus. Sie wählen das Koordinatensystem aus, indem Sie den Rahmen um das individuelle Diagramm auswählen.
- ▶ Klicken Sie im Fenster "Eigenschaften" auf die Registerkarte Koordinaten.

Abbildung 7-16
Registerkarte "Koordinaten"



- ▶ Wählen Sie die Transformationen aus, die Sie dem Koordinatensystem zuweisen möchten. Sie können die Auswahl einer Transformation auch aufheben, um sie rückgängig zu machen.

Transponiert. Das Ändern der Achsenorientierung wird **Transponieren** genannt. Es gleicht dem Vertauschen der vertikalen und horizontalen Achsen in einer 2D-Visualisierung.

Polar. Eine Polartransformation zeichnet die Grafikelemente in einem bestimmten Winkel und Abstand vom Mittelpunkt des Diagramms. Ein Kreisdiagramm ist eine 1D-Visualisierung mit einer Polartransformation, die einzelne Balken in bestimmten Winkeln zeichnet. Ein Radardiagramm ist eine 2D-Visualisierung mit einer Polartransformation, die Grafikelemente in bestimmten Winkeln und Abständen von der Mitte des Diagramms zeichnet. Eine 3D-Visualisierung würde eine zusätzliche Tiefendimension umfassen.

Schräg. Eine schräge Transformation fügt den Grafikelementen einen 3D-Effekt hinzu. Diese Transformation verleiht den Grafikelementen mehr Tiefe, die aber rein dekorativen Zwecken dient. Sie wird durch keine bestimmten Datenwerte beeinflusst.

Gleiches Verhältnis. Wenn Sie dasselbe Verhältnis zuweisen, repräsentiert derselbe Abstand auf jeder Skale denselben Abstand in den Datenwerten. Beispielsweise repräsentieren 2 cm auf beiden Skalen einen Abstand von 1.000.

% Einsatz vor der Transformation. Wenn Achsen nach der Transformation abgeschnitten sind, sollten Sie dem Diagramm Einsätze hinzufügen, bevor Sie die Transformation zuweisen. Die Einsätze schrumpfen die Abmessungen um einen bestimmten Prozentsatz, bevor dem Koordinatensystem etwaige Transformationen zugewiesen werden. Sie können die unteren x -, oberen x -, unteren y - und oberen y -Abmessungen in dieser Reihenfolge steuern.

% Einsatz nach der Transformation. Wenn Sie das Seitenverhältnis eines Diagramms ändern möchten, können Sie dem Diagramm Einsätze hinzufügen, nachdem die Transformation zugewiesen wurde. Die Einsätze schrumpfen die Abmessungen um einen bestimmten Prozentsatz, nachdem dem Koordinatensystem etwaige Transformationen zugewiesen wurden. Diese Einsätze können auch zugewiesen werden, wenn dem Diagramm keine Transformation zugewiesen wird. Sie können die unteren x -, oberen x -, unteren y - und oberen y -Abmessungen in dieser Reihenfolge steuern.

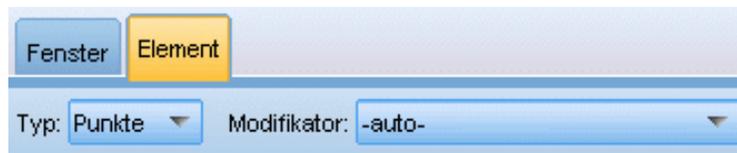
Ändern von Statistiken und Grafikelementen

Sie können ein in einen anderen Typ umwandeln, die zum Zeichnen des Grafikelements verwendete Statistik ändern oder den Kollisionsmodifikator angeben, der bestimmt, was geschieht, wenn sich Grafikelemente überlagern.

So konvertieren Sie ein Grafikelement:

- ▶ Wählen Sie das Grafikelement aus, das Sie konvertieren möchten.
- ▶ Klicken Sie im Fenster “Eigenschaften” auf die Registerkarte Element.

Abbildung 7-17
Registerkarte “Element”



- ▶ Wählen Sie einen neuen Grafikelementtyp aus der Liste “Typ”.

Grafikelementtyp	Beschreibung
Punkt	Ein Marker, der einen spezifischen Datenpunkt identifiziert. Ein Punktelement wird in Streudiagrammen und anderen verwandten Visualisierungen verwendet.
Intervallskala (MDS Model)	Eine rechteckige Form, die an einem bestimmten Datenwert dargestellt wird und den Platz zwischen Ursprung und einem anderen Datenwert ausfüllt. Ein Intervallelement wird in Balkendiagrammen und Histogrammen verwendet.
Linie	Eine Linie, die Datenwerte verbindet.
Pfad	Eine Linie, die Datenwerte in der Reihenfolge ihres Auftretens im Daten-Set verbindet.
Fläche	Eine Linie, die Datenelemente mit der Fläche zwischen der Linie und einem Ursprung verbindet.
Polygon	Eine mehrseitige Form, die eine Datenregion umgibt. Ein Polygonelement könnte in einem klassifizierten Streudiagramm oder in einer Karte verwendet werden.
Schema	Ein Element, das aus einem Rechteck mit Whiskern und Markern besteht, die Ausreißer anzeigen. Ein Schemaelement wird für Boxplots verwendet.

So ändern Sie die Statistik:

- ▶ Wählen Sie das Grafikelement aus, dessen Statistik Sie ändern möchten.
- ▶ Klicken Sie im Fenster “Eigenschaften” auf die Registerkarte Element.
- ▶ Wählen Sie aus der Dropdown-Liste “Auswertung” eine neue Statistik. Beachten Sie, dass bei der Auswahl einer Statistik die Daten zusammengefasst werden. Wenn die Visualisierung stattdessen nicht zusammengefasste Daten anzeigen soll, wählen Sie (keine Statistik) aus der Liste “Auswertung”.

Aus stetigen Feldern berechnete Auswertungsstatistiken

- **Mittelwert.** Ein Lagemaß. Das arithmetische Mittel, d. h. die Summe geteilt durch die Anzahl der Fälle.
- **Median.** Wert, über und unter dem jeweils die Hälfte der Fälle liegt; 50. Perzentil. Bei einer geraden Anzahl von Fällen ist der Median der Mittelwert der beiden mittleren Fälle, wenn diese auf- oder absteigend sortiert sind. Der Median ist ein Lagemaß, das gegenüber Ausreißern unempfindlich ist (im Gegensatz zum Mittelwert, der durch wenige extrem niedrige oder hohe Werte beeinflusst werden kann).
- **Modalwert.** Der am häufigsten auftretende Wert. Wenn mehrere Werte gleichermaßen die größte Häufigkeit aufweisen, ist jeder von ihnen ein Modalwert.
- **Minimum.** Der kleinste Wert einer numerischen Variablen.
- **Maximum.** Der größte Wert einer numerischen Variablen.
- **Bereich.** Die Differenz zwischen dem Mindest- und Höchstwert.
- **Spannweitenmitte.** Die Mitte der Spannweite, d. h. der Wert, dessen Differenz vom Mindestwert identisch ist mit der Differenz vom Höchstwert.
- **Summe.** Die Summe der Werte über alle Fälle mit nichtfehlenden Werten.
- **Kumulierte Summe.** Die kumulierte Summe der Werte. Jedes Grafikelement zeigt die Summe für eine Untergruppe plus die Gesamtsumme aller vorherigen Gruppen.
- **Prozentsumme.** Der Prozentsatz innerhalb jeder Untergruppe basierend auf einem summierten Feld im Vergleich zur Summe aller Gruppen.
- **Kumulierte Prozentsumme.** Der kumulierte Prozentsatz innerhalb jeder Untergruppe basierend auf einem summierten Feld im Vergleich zur Summe aller Gruppen. Jedes Grafikelement zeigt den Prozentsatz für eine Untergruppe plus den Gesamtprozentsatz aller vorherigen Gruppen.
- **Varianz.** Ein Maß der Streuung um den Mittelwert. Es ist gleich dem Quotienten aus der Summe der quadrierten Abweichung vom Mittelwert und der um 1 verringerten Fallanzahl. Die Maßeinheit der Varianz ist das Quadrat der Maßeinheiten der Variablen.
- **Standardabweichung.** Ein Maß für die Streuung um den Mittelwert. Bei einer Normalverteilung liegen 68 % der Fälle im Bereich von einer Standardabweichung um den Mittelwert und 95 % der Fälle im Bereich von zwei Standardabweichungen. Wenn beispielsweise für das Alter der Mittelwert 45 und die Standardabweichung 10 beträgt, liegen bei einer Normalverteilung 95 % der Fälle im Bereich zwischen 25 und 65.
- **Standardfehler.** Ein Maß für die Abweichung des Werts einer Teststatistik zwischen Stichproben. Dies ist die Standardabweichung der Stichprobenverteilung einer Statistik. So ist z. B. der Standardfehler des Mittelwerts die Standardabweichung des Stichprobenmittelwerts.
- **Kurtosis.** Ein Maß dafür, wie sich die Beobachtungen um einen zentralen Punkt gruppieren. Bei einer Normalverteilung ist der Wert der Kurtosis gleich 0. Positive Kurtosis gibt an, dass sich die Beobachtungen in Relation zur Normalverteilung mehr um den Mittelpunkt der Verteilung gruppieren und deren Flanken zu den Extremwerten der Verteilung hin dünner werden. An diesem Punkt sind die Flanken der leptokurtischen Verteilung breiter als bei einer Normalverteilung. Negative Kurtosis gibt an, dass sich die Beobachtungen in Relation zur Normalverteilung weniger um einen zentralen Punkt gruppieren und ihre Flanken zu den

Extremwerten der Verteilung hin breiter werden. An diesem Punkt sind die Flanken der platykurtischen Verteilung schmäler als bei einer Normalverteilung.

- **Schiefe.** Ein Maß für die Asymmetrie einer Verteilung. Die Normalverteilung ist symmetrisch, ihre Schiefe hat den Wert 0. Eine Verteilung mit einer deutlichen positiven Schiefe läuft nach rechts lang aus (lange rechte Flanke). Eine Verteilung mit einer deutlichen negativen Schiefe läuft nach links lang aus (lange linke Flanke). Als Faustregel kann man verwenden, dass ein Schiefe-Wert, der mehr als doppelt so groß ist wie sein Standardfehler, für eine Abweichung von der Symmetrie spricht.

Die folgende Regionsstatistik kann mehr als ein Grafikelement pro Untergruppe ergeben. Bei Verwendung der Intervall-, Flächen- oder Randgrafikelemente ergibt eine Regionsstatistik ein Grafikelement, das den Bereich angibt. Alle anderen Grafikelemente ergeben zwei separate Elemente. Eines markiert den Anfang des Bereichs und eines das Ende des Bereichs.

- **Region: Bereich.** Der Wertebereich zwischen dem Mindest- und Höchstwert.
- **Region: 95%-Konfidenzintervall des Mittelwerts.** Ein Wertebereich, der mit einer Wahrscheinlichkeit von 95 % den Mittelwert der Stichprobe enthält.
- **Region: 95%-Konfidenzintervall des Einzelfalls.** Ein Wertebereich, der mit einer Wahrscheinlichkeit von 95 % den vorhergesagten Wert für den Einzelfall enthält.
- **Region: 1 Standardabweichung über/unter dem Mittelwert.** Ein Wertebereich zwischen 1 Standardabweichung über und unter dem Mittelwert.
- **Region: 1 Standardfehler über/unter dem Mittelwert.** Ein Wertebereich zwischen 1 Standardfehler über und unter dem Mittelwert.

Häufigkeitsbasierte Auswertungsstatistik

- **Anzahl.** Die Anzahl der gültigen Zeilen/Fälle.
- **Kumulierte Anzahl.** Die kumulierte Anzahl der gültigen Zeilen/Fälle. Jedes Grafikelement zeigt die Anzahl für eine Untergruppe plus die Gesamtsumme aller vorherigen Gruppen.
- **Prozentsatz der Anzahl.** Der Prozentsatz an Zeilen/Fällen in jeder Untergruppe im Vergleich zur Gesamtanzahl an Zeilen/Fällen.
- **Kumulierter Prozentsatz der Anzahl.** Der kumulierte Prozentsatz an Zeilen/Fällen in jeder Untergruppe im Vergleich zur Gesamtanzahl an Zeilen/Fällen. Jedes Grafikelement zeigt den Prozentsatz für eine Untergruppe plus den Gesamtprozentsatz aller vorherigen Gruppen.

So geben Sie den Kollisionsmodifikator an:

Der Kollisionsmodifikator bestimmt, was geschieht, wenn sich Grafikelemente überlagern.

- ▶ Wählen Sie das Grafikelement aus, für das Sie den Kollisionsmodifikator angeben möchten.
- ▶ Klicken Sie im Fenster "Eigenschaften" auf die Registerkarte Element.
- ▶ Wählen Sie aus dem Dropdown-Listefeld "Modifikator" einen Kollisionsmodifikator aus. -auto- überlässt es der Anwendung zu bestimmen, welcher Kollisionsmodifikator für den Grafikelementtyp und die Statistik geeignet ist.

Überlagert. Zeichnet Grafikelemente mit demselben Wert übereinander.

Stapeln. Stapelt Grafikelemente, die einander normalerweise überdecken würden, wenn sie dieselben Datumswerte besitzen.

Ausweichen. Verschiebt Grafikelemente neben andere Grafikelemente, die am gleichen Wert erscheinen, anstatt sie übereinanderzudecken. Die Grafikelemente werden symmetrisch angeordnet. D. h. die Grafikelemente werden an entgegengesetzten Seiten einer zentralen Position verschoben. Ausweichen ist dem Clustering sehr ähnlich.

Übereinander. Verschiebt Grafikelemente neben andere Grafikelemente, die am gleichen Wert erscheinen, anstatt sie übereinanderzudecken. Die Grafikelemente werden asymmetrisch angeordnet. D. h., die Grafikelemente werden schräg übereinander angeordnet, wobei das untere Grafikelement an einem bestimmten Wert auf der Skala positioniert ist.

Streuen (normal). Positioniert Grafikelemente am selben Datenwert anhand einer Normalverteilung zufällig um.

Streuen (uniform). Positioniert Grafikelemente am selben Datenwert anhand einer Gleichverteilung zufällig um.

Ändern der Position der Legende

Wenn ein Diagramm eine Legende umfasst, wird diese in der Regel rechts neben dem Diagramm angezeigt. Sie können diese Position bei Bedarf ändern.

So ändern Sie die Position der Legende:

- ▶ Wählen Sie die Legende aus.
- ▶ Klicken Sie im Fenster “Eigenschaften” auf die Registerkarte Legende.

Abbildung 7-18
Registerkarte “Legende”



- ▶ Wählen Sie eine Position aus.

Kopieren von Visualisierungen und Visualisierungsdaten

Die Palette “Allgemein” enthält Schaltflächen zum Kopieren der Visualisierung und ihrer Daten.

Abbildung 7-19
Schaltfläche “Visualisierung kopieren”



Kopieren der Visualisierung. Diese Aktion kopiert die Visualisierung als Bild in die Zwischenablage. Mehrere Bildformate stehen zur Verfügung. Wenn Sie das Bild in eine andere Anwendung einfügen, können Sie eine “Einfügen-Spezial”-Option wählen, um eines der verfügbaren Bildformate zum Einfügen festzulegen.

Abbildung 7-20
Schaltfläche “Visualisierungsdaten kopieren”



Kopieren der Visualisierungsdaten. Diese Aktion kopiert die zugrundeliegenden Daten, anhand denen die Visualisierung gezeichnet wird. Die Daten werden als Standardtext oder HTML-Text in die Zwischenablage kopiert. Wenn Sie die Daten in eine andere Anwendung einfügen, können Sie eine “Einfügen-Spezial”-Option wählen, um eines dieser Formate zum Einfügen festzulegen.

Tastaturkombinationen

Tabelle 7-3
Tastaturkombinationen

Tastenkürzel	Funktion
Strg+Leerzeichen	Umschalten zwischen Sondierungs- und Bearbeitungsmodus
Löschen	Löschen eines Visualisierungselements
Strg+Z	Rückgängig
STRG + Y	Wiederholen
F2	Anzeigen des Umrisses zur Auswahl von Elementen im Diagramm

Teil III:
Ressourceneditor

Vorlagen und Ressourcen

IBM® SPSS® Text Analytics for Surveys erfasst und extrahiert Schlüsselkonzepte aus Textdaten schnell und genau. Dieser Extrahierungsprozess beruht erheblich darauf, dass durch die linguistischen Ressourcen festgelegt wird, wie Informationen aus Textdaten extrahiert werden. [Für weitere Informationen siehe Thema Wie Extrahierung funktioniert in Kapitel 1 auf S. 4.](#) Sie können die Feinabstimmung für diese Ressourcen in der Resource Editor-Ansicht vornehmen.

Bei der Installation der Software erhalten Sie außerdem eine Reihe von spezialisierten Ressourcen. Diese im Lieferumfang enthaltenen Ressourcen ermöglichen es Ihnen, von jahrelangen Forschungen zu profitieren und Feinabstimmungen für spezifische Sprachen und Anwendungen vorzunehmen. Da die mitgelieferten Ressourcen ggf. nicht perfekt an den Kontext Ihrer Daten angepasst sind, können Sie diese Ressourcenvorlagen bearbeiten oder sogar benutzerdefinierte Bibliotheken erstellen und verwenden, die spezifisch auf die Daten Ihrer Organisation abgestimmt sind. Diese Ressourcen liegen in verschiedenen Formaten vor und jede davon kann in Ihrem projekt verwendet werden. Ressourcen finden Sie an folgenden Stellen:

- **Ressourcenvorlagen.** Vorlagen setzen sich aus einer Reihe von Bibliotheken, Typen und einigen erweiterten Ressourcen zusammen, die ein spezialisiertes Ressourcen-Set bilden, das auf eine bestimmte Domäne (wie z. B. Meinungen zu einem Produkt) ausgerichtet ist.
- **Text Analysis Packages (TAP).** Zusätzlich zu den Ressourcen, die in einer Vorlage gespeichert sind, bündeln TAPs auch ein oder mehrere spezielle Kategoriensets, die mithilfe jener Ressourcen erstellt wurden, so dass die Kategorien und die Ressourcen gemeinsam gespeichert werden und wiederverwendbar sind. [Für weitere Informationen siehe Thema Verwendung von Text Analysis Packages in Kapitel 3 auf S. 43.](#)
- **Bibliotheken.** Bibliotheken werden als Bausteine für TAPs und Vorlagen verwendet. So können Ressourcen innerhalb Ihres Projekts auch einzeln hinzugefügt werden. Jede Bibliothek besteht aus mehreren Wörterbüchern zur Definition und Verwaltung von Typen, Synonymen und Ausschlusslisten. Während Wörterbücher auch einzeln bezogen werden können, werden sie in Vorlagen und TAPs vorab gemeinsam verpackt. [Für weitere Informationen siehe Thema Mit Bibliotheken arbeiten in Kapitel 9 auf S. 209.](#)

Anmerkung: Im Laufe der Extrahierung werden auch einige kompilierte interne Ressourcen verwendet. Diese kompilierten Ressourcen enthalten eine große Anzahl von Definitionen, die die Typen in der Core Library ergänzen. Diese kompilierten Ressourcen können nicht bearbeitet werden.

Der Resource Editor ermöglicht den Zugriff auf eine Reihe von Ressourcen, die für die Gewinnung von Extrahierungsergebnissen (Konzepten, Typen und Mustern) verwendet werden. Im Resource Editor können Sie eine Vielzahl von Aufgaben durchführen, darunter:

- **Arbeiten mit Bibliotheken**[Für weitere Informationen siehe Thema Mit Bibliotheken arbeiten in Kapitel 9 auf S. 209.](#)
- **Erstellen von Typ-Wörterbüchern**[Für weitere Informationen siehe Thema Erstellen von Typen in Kapitel 10 auf S. 225.](#)

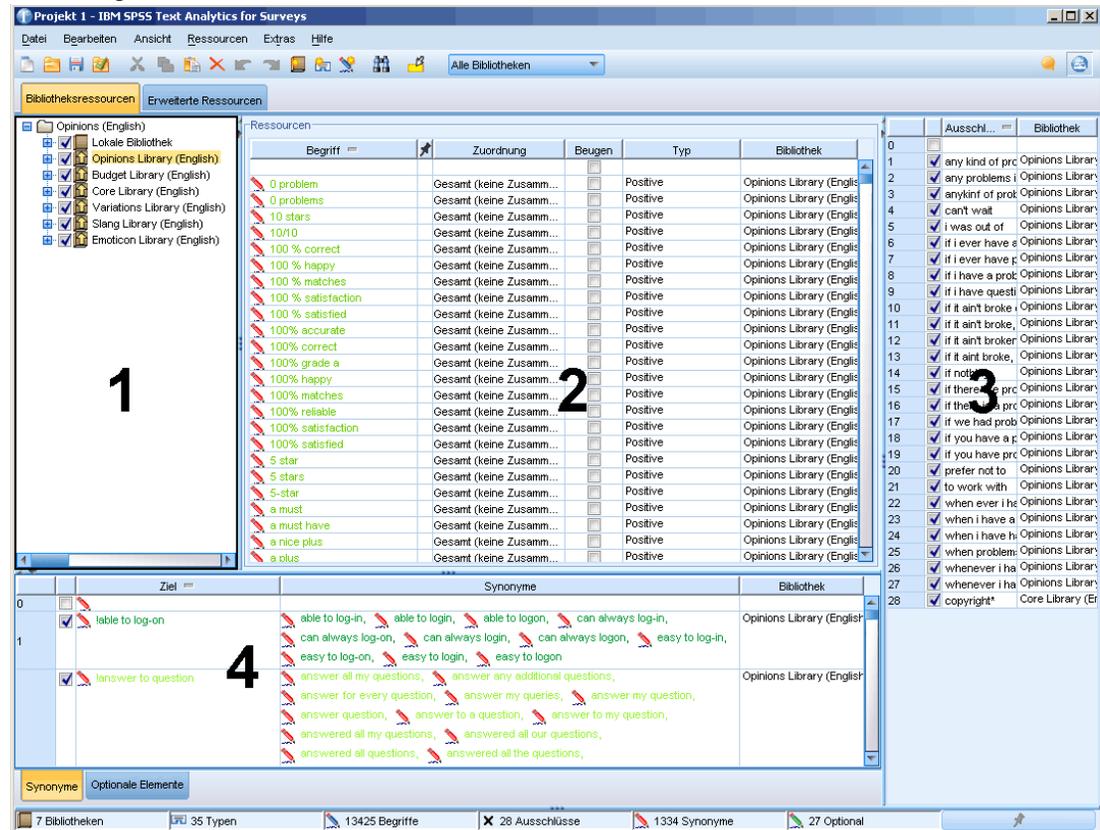
- **Hinzufügen von Fachausdrücken zu Wörterbüchern**Für weitere Informationen siehe Thema [Hinzufügen von Fachausdrücken in Kapitel 10 auf S. 227.](#)
- **Erstellen von Synonymen**Für weitere Informationen siehe Thema [Definieren von Synonymen in Kapitel 10 auf S. 236.](#)
- **Aktualisierung der Ressourcen in TAPs.** Für weitere Informationen siehe Thema [Aktualisierung von Text Analysis Packages in Kapitel 3 auf S. 45.](#)
- **Erstellen von Vorlagen.** Für weitere Informationen siehe Thema [Erstellen und Aktualisieren von Vorlagen auf S. 200.](#)
- **Importieren und Exportieren von Vorlagen**Für weitere Informationen siehe Thema [Importieren und Exportieren von Vorlagen auf S. 203.](#)
- **Veröffentlichen von Bibliotheken**Für weitere Informationen siehe Thema [Bibliotheken veröffentlichen in Kapitel 9 auf S. 219.](#)

Die Editoroberfläche

Bei den im Resource Editor durchgeführten Operationen geht es um die Verwaltung und Feinabstimmung der linguistischen Ressourcen. Diese Ressourcen werden in Form von Vorlagen und Bibliotheken gespeichert. [Für weitere Informationen siehe Thema Typ-Wörterbücher in Kapitel 10 auf S. 223.](#)

Registerkarte "Bibliotheksressourcen"

Abbildung 8-1



Die Oberfläche gliedert sich in die vier folgenden Bereiche:

1. Fensterbereich "Bibliotheksbaum". Dieser Plan befindet sich links oben und zeigt eine Baumstruktur mit den Bibliotheken. Sie können die Bibliotheken in diesem Baum aktivieren bzw. deaktivieren sowie die Ansichten in den anderen Fensterbereichen filtern, indem Sie eine Bibliothek im Baum auswählen. In diesem Baum können Sie mithilfe der Kontextmenüs eine Vielzahl von Operationen durchführen. Wenn Sie eine Bibliothek im Baum erweitern, wird das darin enthaltene Typen-Set angezeigt. Sie können diese Liste auch über das Menü Ansicht filtern, wenn Sie sich nur auf eine bestimmte Bibliothek konzentrieren möchten.

2. Bereich für Fachausdruckslisten von Typen-Wörterbüchern. Dieser Fensterbereich befindet sich rechts neben dem Bibliotheksbaum und zeigt die Fachausdruckslisten der Typ-Wörterbücher für die im Baum ausgewählten Bibliotheken an. Ein **Typ-Wörterbuch** ist eine Zusammenstellung von Fachausdrücken, die unter einer Bezeichnung (bzw. einem Typ oder einem Namen) gruppiert werden sollen. Beim Lesen der Textdaten vergleicht die Extrahierungsengine die im Text gefundenen Wörter mit den Fachausdrücken in den Typ-Wörterbüchern. Wenn ein extrahiertes Konzept als Fachausdruck in einem Typ-Wörterbuch vorkommt, wird der entsprechende Typname zugewiesen. Sie können das Typ-Wörterbuch als eigenes Wörterbuch mit Fachausdrücken betrachten, die etwas gemeinsam haben. Der Typ <Location> (Ort) in der Core Library beispielsweise enthält Konzepte wie *new orleans*, *great britain* und *new york*. Diese

Fachausdrücke stehen jeweils für geografische Orte. Eine Bibliothek kann ein oder mehrere Typ-Wörterbücher enthalten. [Für weitere Informationen siehe Thema Typ-Wörterbücher in Kapitel 10 auf S. 223.](#)

3. Fensterbereich "Ausschlusswörterbuch". Dieser Fensterbereich befindet sich auf der rechten Seite und zeigt die Sammlung der Fachausdrücke, die aus den endgültigen Extrahierungsergebnissen ausgeschlossen werden. Die Fachausdrücke in diesem Ausschlusswörterbuch werden nicht im Fensterbereich "Extrahierungsergebnisse" angezeigt. Ausgeschlossene Fachausdrücke werden in der Bibliothek Ihrer Wahl gespeichert. Jedoch zeigt der Bereich "Ausschlussbibliothek" alle ausgeschlossenen Fachausdrücke für alle Bibliotheken an, die im Bibliotheksbaum sichtbar sind. [Für weitere Informationen siehe Thema Ausschlusswörterbücher in Kapitel 10 auf S. 240.](#)

4. Fensterbereich "Substitutionswörterbuch". Dieser Fensterbereich befindet sich links unten und zeigt Synonyme und optionale Elemente in einer jeweils eigenen Registerkarte an. Synonyme und optionale Elemente helfen bei der Gruppierung ähnlicher Fachbegriffe unter einem Haupt- oder Zielkonzept in den endgültigen Extrahierungsergebnissen. Das Wörterbuch kann bekannte Synonyme, benutzerdefinierte Synonyme und Elemente sowie häufig vorkommende Rechtschreibfehler zusammen mit der korrekten Schreibung enthalten. Synonyme und optionale Elemente können in einer Bibliothek Ihrer Wahl gespeichert werden. Jedoch zeigt der Bereich "Substitutionswörterbuch" den Inhalt für alle Bibliotheken an, die im Bibliotheksbaum sichtbar sind. Während dieser Bereich alle Synonyme und optionale Elemente aus allen Bibliotheken anzeigt, werden die Substitutionen für alle Bibliotheken in der Baumstruktur zusammen in diesem Fensterbereich angezeigt. Eine Bibliothek kann nur ein einziges Substitutionswörterbuch enthalten. [Für weitere Informationen siehe Thema Substitutions-/Synonymwörterbücher in Kapitel 10 auf S. 234.](#)

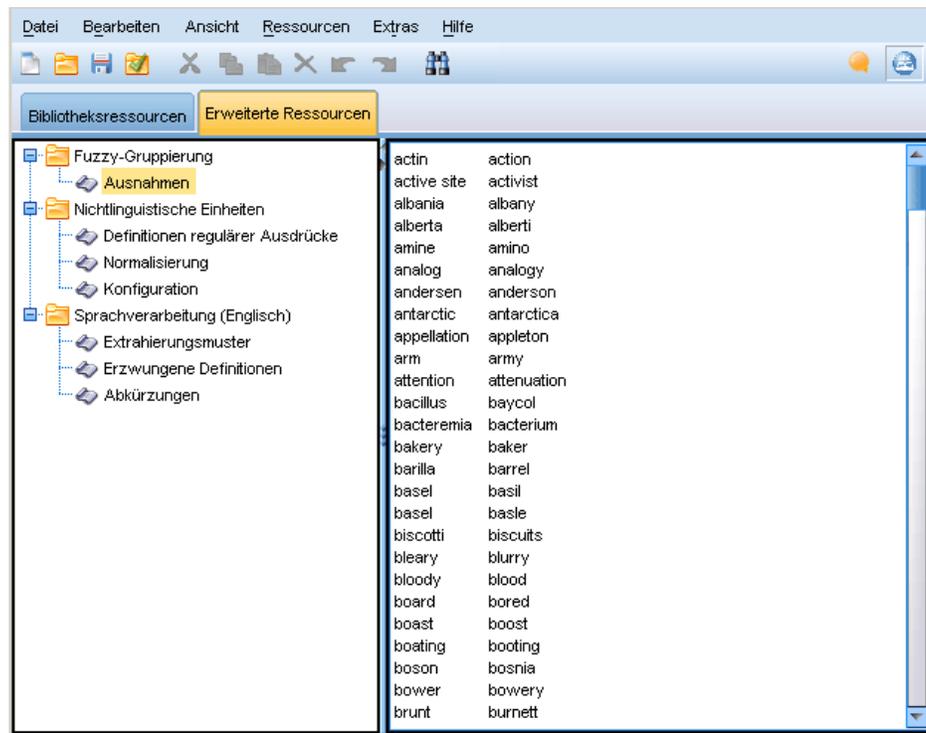
Anmerkung:

- Wenn Sie einen Filtervorgang durchführen möchten, sodass nur die Informationen angezeigt werden, die zu einer einzigen Bibliothek gehören, können Sie die Bibliotheksansicht mithilfe der Dropdown-Liste in der Symbolleiste ändern. Sie enthält auf der obersten Ebene den Eintrag Alle Bibliotheken sowie einen zusätzlichen Eintrag für jede einzelne Bibliothek. [Für weitere Informationen siehe Thema Bibliotheken anzeigen in Kapitel 9 auf S. 213.](#)

Registerkarte "Erweiterte Ressourcen"

Die erweiterten Ressourcen stehen in der zweiten Registerkarte der Editoransicht zur Verfügung. Sie können die erweiterten Ressourcen in dieser Registerkarte überprüfen und bearbeiten. [Für weitere Informationen siehe Thema Informationen zu erweiterten Ressourcen in Kapitel 11 auf S. 242.](#)

Abbildung 8-2
Erweiterte Ressourcen



Erstellen und Aktualisieren von Vorlagen

Wenn Sie Änderungen an Ihren Ressourcen vornehmen und sie in Zukunft erneut verwenden möchten, speichern Sie die Ressourcen als Vorlage. Dabei können Sie die Vorlage unter einem bestehenden Vorlagennamen oder unter einem neuen Namen speichern. Wenn Sie künftig diese Vorlage laden, erhalten Sie dieselben Ressourcen.

Anmerkung: Sie können Ihre Bibliotheken auch veröffentlichen und mit anderen Benutzern gemeinsam nutzen. Für weitere Informationen siehe [Thema Bibliotheken gemeinsam nutzen in Kapitel 9 auf S. 217](#).

Abbildung 8-3
Dialogfeld "Ressourcenvorlage erstellen"



So erstellen (oder aktualisieren) Sie eine Vorlage:

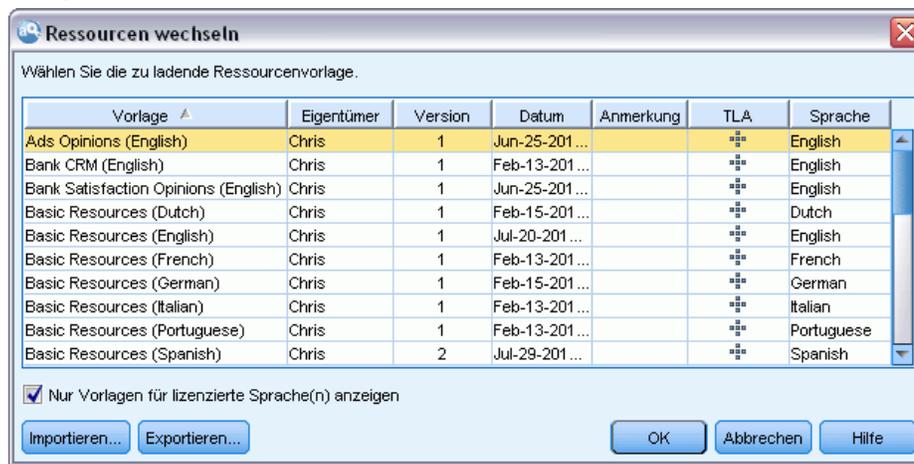
- ▶ Wählen Sie in den Menüs der Resource Editor-Ansicht die Optionsfolge Ressourcen > Ressourcenvorlage erstellen. Das Dialogfeld "Ressourcenvorlage erstellen" wird geöffnet.
- ▶ Geben Sie einen neuen Namen in das Feld "Vorlagenname" ein, wenn Sie eine neue Vorlage erstellen möchten. Wählen Sie eine Vorlage in der Tabelle aus, um eine vorhandene Vorlage mit den derzeit geladenen Ressourcen zu überschreiben.
- ▶ Klicken Sie auf Speichern, um die Vorlage zu erstellen.

Wechseln von Ressourcenvorlagen

Wenn Sie die aktuell geladenen Ressourcen durch eine Kopie der Ressourcen einer anderen Vorlage ersetzen möchten, können Sie zu diesen wechseln. Dadurch werden die .

Sie können die Vorlage auswählen, deren Inhalt Sie in den Resource Editor kopieren möchten, und dann auf OK klicken. Dadurch werden die Ressourcen in dieser Projekt

Abbildung 8-4
Dialogfeld "Ressourcen wechseln"



So wechseln Sie Ressourcen:

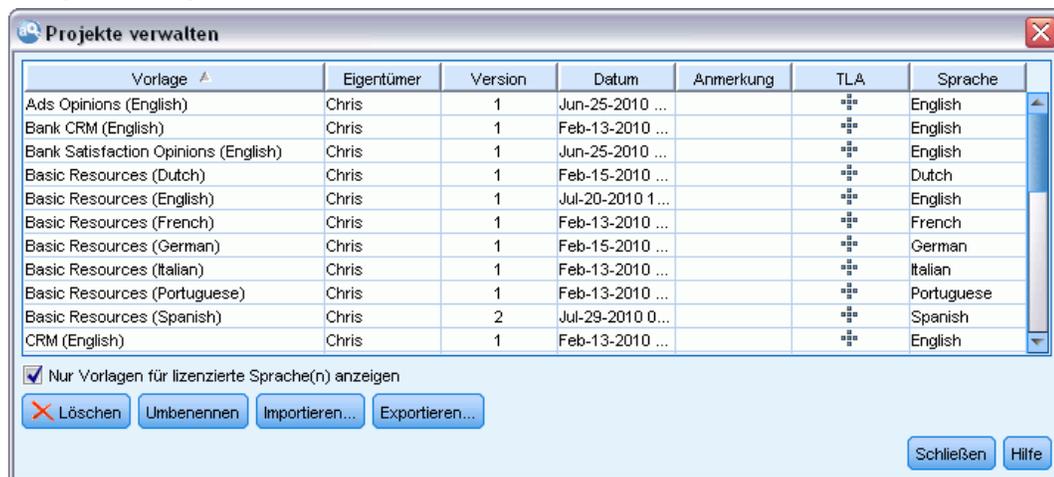
- ▶ Wählen Sie in den Menüs der Resource Editor-Ansicht die Optionsfolge Ressourcen > Ressourcenvorlage wechseln. Das Dialogfeld "Ressourcenvorlagen wechseln" wird geöffnet.
- ▶ Wählen Sie die gewünschte Vorlage aus der Tabelle aus.
- ▶ Klicken Sie auf OK, um die derzeit geladenen Ressourcen zu entfernen und stattdessen eine Kopie der Ressourcen der ausgewählten Vorlage zu verwenden. Wenn Sie Änderungen in Ihren Ressourcen vorgenommen haben und Ihre Bibliotheken für künftige Verwendungen speichern möchten, können Sie sie vor dem Wechseln veröffentlichen, aktualisieren und für die gemeinsame Nutzung freigeben. [Für weitere Informationen siehe Thema Bibliotheken gemeinsam nutzen in Kapitel 9 auf S. 217.](#)

Vorlagen verwalten

Gelegentlich sollten Sie einige grundlegende Verwaltungsmaßnahmen an Ihren Vorlagen durchführen, z. B. das Umbenennen der Vorlagen, Importieren und Exportieren von Vorlagen oder Löschen veralteter Vorlagen. Diese Aufgaben werden im Dialogfeld "Vorlagen verwalten" durchgeführt. Durch Importieren und Exportieren von Vorlagen können Sie Vorlagen mit anderen Benutzern gemeinsam nutzen. [Für weitere Informationen siehe Thema Importieren und Exportieren von Vorlagen auf S. 203.](#)

Anmerkung: Mit diesem Produkt installierte (oder mitgelieferte) Vorlagen können nicht umbenannt oder gelöscht werden. Wenn Sie eine Vorlage umbenennen möchten, können Sie die installierte Vorlage öffnen und eine neue mit einem Namen Ihrer Wahl erstellen. Ihre benutzerdefinierten Vorlagen können Sie löschen. Wenn Sie jedoch versuchen, eine mitgelieferte Vorlage zu löschen, wird sie auf die installierte Version zurückgesetzt.

Abbildung 8-5
Dialogfeld "Vorlagen verwalten"



So benennen Sie eine Vorlage um:

- ▶ Wählen Sie in den Menüs die Optionsfolge Ressourcen > Ressourcenvorlagen verwalten aus. Das Dialogfeld "Vorlagen verwalten" wird geöffnet.
- ▶ Wählen Sie die Vorlage aus, die Sie umbenennen möchten, und klicken Sie auf Umbenennen. Das Namensfeld wird in der Tabelle editierbar.
- ▶ Geben Sie einen neuen Namen ein und drücken Sie die Eingabetaste. Es wird ein Bestätigungs-Dialogfeld geöffnet.
- ▶ Wenn Sie die Namensänderung anwenden möchten, klicken Sie auf Ja. Wenn Sie sie verwerfen möchten, klicken Sie auf Nein.

So löschen Sie eine Vorlage:

- ▶ Wählen Sie in den Menüs die Optionsfolge Ressourcen > Ressourcenvorlagen verwalten aus. Das Dialogfeld "Vorlagen verwalten" wird geöffnet.
- ▶ Wählen Sie im Dialogfeld "Vorlagen verwalten" die zu löschende Vorlage aus.
- ▶ Klicken Sie auf Löschen. Es wird ein Bestätigungs-Dialogfeld geöffnet.
- ▶ Klicken Sie auf Ja, um zu löschen, oder auf Nein, um den Anforderungsvorgang abzubrechen. Wenn Sie auf Ja klicken, wird die Vorlage gelöscht.

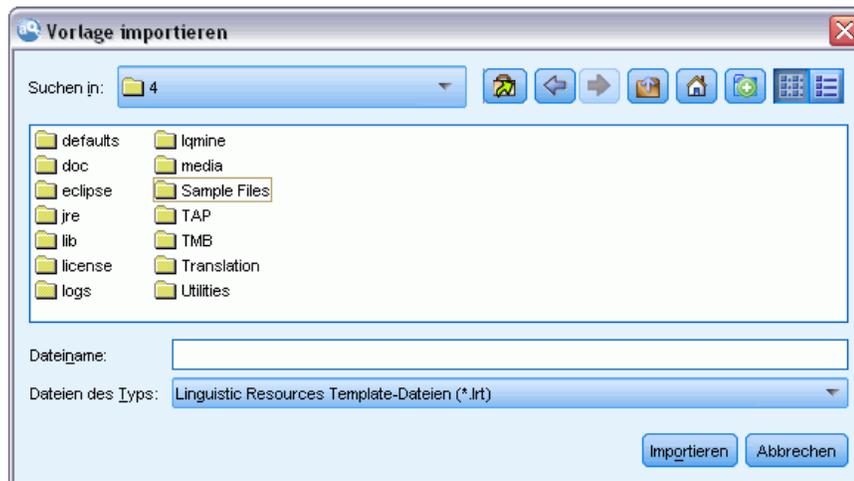
Importieren und Exportieren von Vorlagen

Sie können Vorlagen mit anderen Benutzern oder Computern gemeinsam nutzen, indem Sie sie importieren und exportieren. Vorlagen werden in einer internen Datenbank gespeichert, können jedoch als *.lvt-Dateien auf die Festplatte exportiert werden. Im Dialogfeld "Vorlagen verwalten" im Resource Editor können Sie Vorlagen importieren und exportieren.

So importieren Sie eine Vorlage:

- ▶ Klicken Sie im Dialogfeld auf Importieren. Das Dialogfeld “Vorlage importieren” wird geöffnet.

Abbildung 8-6
Dialogfeld “Vorlage importieren”

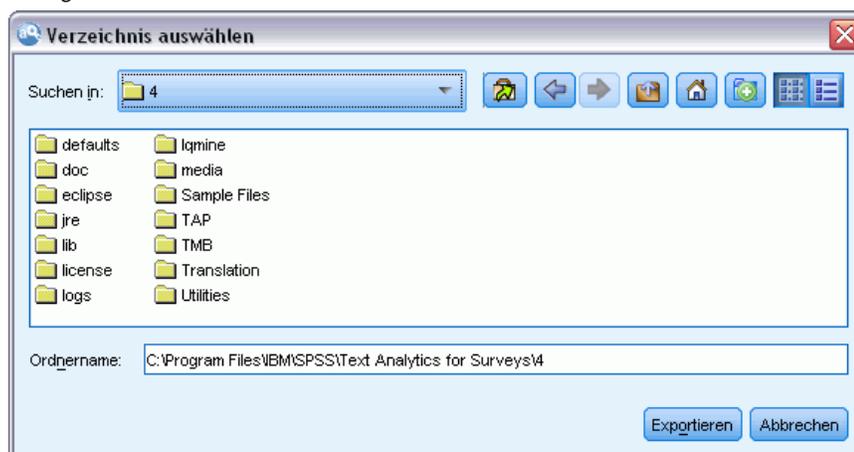


- ▶ Wählen Sie die zu importierende Ressourcenvorlagendatei (*.lrt) aus und klicken Sie auf Importieren. Sie können die zu importierende Vorlage unter einem neuen Namen speichern oder die vorhandene Vorlage überschreiben. Das Dialogfeld wird geschlossen und die Vorlage wird nun in der Tabelle angezeigt.

So exportieren Sie eine Vorlage:

- ▶ Wählen Sie im Dialogfeld die zu exportierende Vorlage aus und klicken Sie auf Exportieren. Das Dialogfeld “Verzeichnis auswählen” wird geöffnet.

Abbildung 8-7
Dialogfeld “Verzeichnis auswählen”



- ▶ Wählen Sie das Verzeichnis aus, in das exportiert werden soll, und klicken Sie auf Exportieren. Das Dialogfeld wird geschlossen und die Vorlage wird mit der Dateierweiterung (*.lrt) exportiert.

Ressourcen sichern

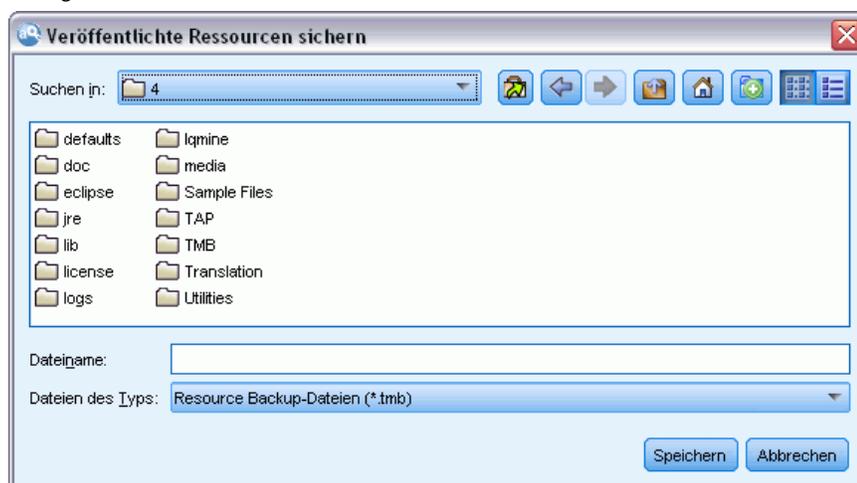
Sicherheitshalber sollten Sie Ihre Ressourcen gelegentlich sichern.

Wichtig: Bei der Wiederherstellung wird der gesamte Inhalt Ihrer Ressourcen gelöscht und nur der Inhalt der Sicherungsdatei ist in dem Produkt verfügbar. Dies gilt auch für geöffnete Arbeiten.

So sichern Sie die Ressourcen:

- Wählen Sie in den Menüs die Optionsfolge Ressourcen > Sicherungstools > Ressourcen sichern. Das Dialogfeld "Sicherung" wird geöffnet.

Abbildung 8-8
Dialogfeld "Ressourcen sichern"

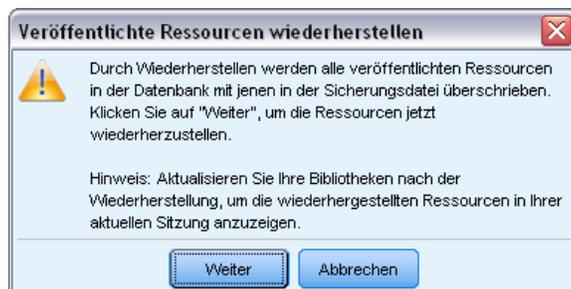


- Geben Sie einen Namen für Ihre Sicherungsdatei ein und klicken Sie auf Speichern. Das Dialogfeld wird geschlossen und die Sicherungsdatei erstellt.

So stellen Sie die Ressourcen wieder her:

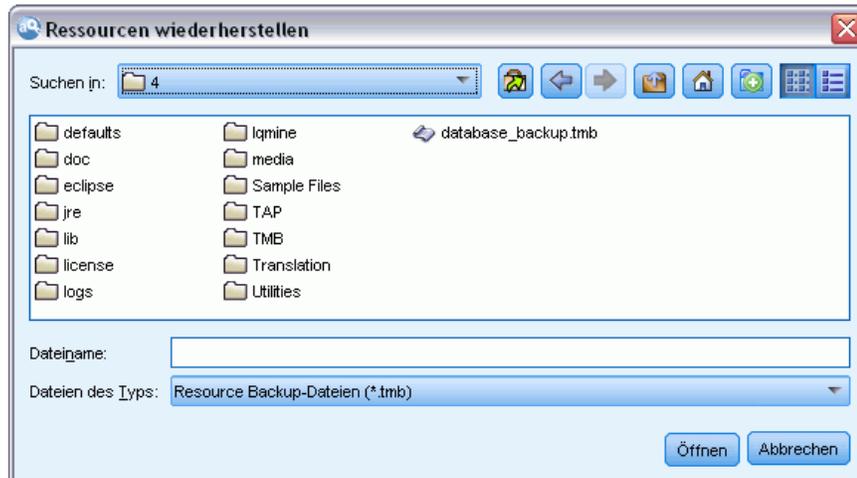
- Wählen Sie in den Menüs die Optionsfolge Ressourcen > Sicherungstools > Ressourcen wiederherstellen. Ein Warnhinweis informiert Sie darüber, dass bei einer Wiederherstellung der Inhalt Ihrer Datenbank überschrieben wird.

Abbildung 8-9
Warnhinweis bei Überschreibung



- Klicken Sie auf Ja, um fortzufahren. Falls ein Projekt geöffnet ist, wird es nicht gelöscht, da es sich im Speicher befindet; Sie müssen es jedoch erneut speichern, um es in die wiederhergestellte Datenbank zu übernehmen. Das Dialogfeld wird geöffnet.

Abbildung 8-10
Dialogfeld "Ressourcen wiederherstellen"



- Wählen Sie die Sicherungsdatei aus, die Sie wiederherstellen möchten, und klicken Sie auf Öffnen. Das Dialogfeld wird geschlossen und die Ressourcen werden in der Anwendung wiederhergestellt.

Wichtig: Bei der Wiederherstellung wird der gesamte Inhalt Ihrer Ressourcen gelöscht, und nur der Inhalt der Sicherungsdatei ist in dem Produkt verfügbar. Dies gilt auch für geöffnete Arbeiten.

Ressourcendateien importieren

Wenn Sie Änderungen direkt in Ressourcendateien außerhalb dieses Produkts vorgenommen haben, können Sie sie in eine ausgewählte Bibliothek importieren, indem Sie diese Bibliothek auswählen und mit dem Import fortfahren. Wenn Sie ein Verzeichnis importieren, können Sie auch alle unterstützten Dateien in eine bestimmte geöffnete Bibliothek importieren. Es können nur *.txt-Dateien importiert werden.

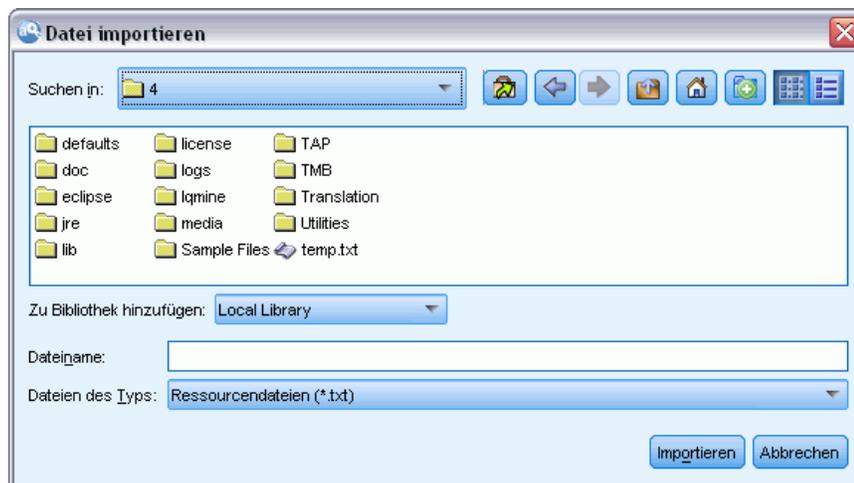
Jede importierte Datei darf nur einen Eintrag pro Zeile enthalten und wenn die Inhalte folgendermaßen strukturiert sind:

- Eine Liste mit Wörtern oder Phrasen (eine pro Zeile). Die Datei wird als Fachausdrucksliste für ein Typ-Wörterbuch importiert. Dabei übernimmt das Typ-Wörterbuch den Namen der Datei ohne die Erweiterung. .
- Als Liste von Einträgen wie *Fachausdruck1*<TAB>*Fachausdruck2*. In diesem Fall wird die Datei als Liste von Synonymen importiert. Dabei ist *Fachausdruck1* das Set des zugrunde liegenden Begriffs und *Fachausdruck2* ist der Zielausdruck.

So importieren Sie eine einzelne Ressourcendatei:

- Wählen Sie in den Menüs die Optionsfolge Ressourcen > Dateien importieren > Einzeldatei importieren. Das Dialogfeld "Datei importieren" wird geöffnet.

Abbildung 8-11
Dialogfeld "Datei importieren"

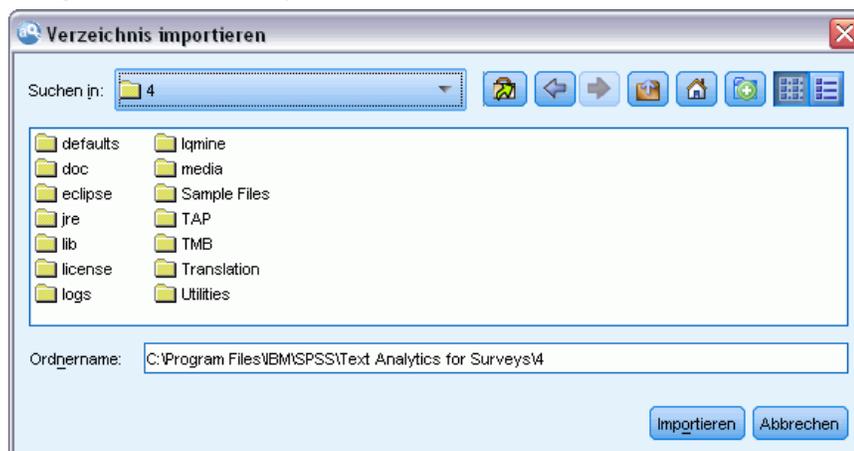


- ▶ Wählen Sie die zu importierende Datei aus und klicken Sie auf Importieren. Der Inhalt der Datei wird in ein internes Format umgewandelt und zu Ihrer Bibliothek hinzugefügt.

So importieren Sie alle Dateien eines Verzeichnisses:

- ▶ Wählen Sie in den Menüs die Optionsfolge Ressourcen > Dateien importieren > Ganzes Verzeichnis importieren. Das Dialogfeld "Verzeichnis importieren" wird geöffnet.

Abbildung 8-12
Dialogfeld "Verzeichnis importieren"



- ▶ Wählen Sie die Bibliothek, in die alle Ressourcendateien importiert werden sollen, aus der Liste Importieren aus. Wenn Sie die Option Standard auswählen, wird eine neue Bibliothek mit dem Namen des Verzeichnisses erstellt.
- ▶ Wählen Sie das Verzeichnis aus, aus dem die Dateien importiert werden sollen. Es werden keine Unterverzeichnisse gelesen.

- ▶ Klicken Sie auf Importieren. Das Dialogfeld wird geschlossen und der Inhalt der importierten Ressourcendateien wird im Editor als Verzeichnisse und erweiterte Ressourcendateien aufgeführt.

Mit Bibliotheken arbeiten

Die Ressourcen, die von der Extrahierungseingabe zum Extrahieren und Gruppieren von Fachausdrücken verwendet werden, enthalten immer eine oder mehrere Bibliotheken. Die Bibliotheken werden im Bibliotheksbaum im oberen linken Bereich des Resource Editors aufgeführt. Sie bestehen aus zwei Arten von Wörterbüchern: Typ, Substitution und Ausschluss. [Für weitere Informationen siehe Thema Informationen zu Bibliothekwörterbüchern in Kapitel 10 auf S. 223.](#)

Die ausgewählte Ressourcenvorlage oder die Ressourcen aus dem TAP enthalten verschiedene Bibliotheken, die Ihnen die unverzügliche Extrahierung von Konzepten aus Ihren Textdaten ermöglichen. Sie können jedoch auch eigene Bibliotheken erstellen und zur Wiederverwendung auch veröffentlichen. [Für weitere Informationen siehe Thema Bibliotheken veröffentlichen auf S. 219.](#)

Angenommen, Sie arbeiten z. B. häufig mit Textdaten zur Automobilindustrie. Nach der Analyse Ihrer Daten möchten Sie benutzerdefinierte Ressourcen zur Verarbeitung von branchenspezifischem Wortschatz oder von Fachsprache erstellen. Mit dem Resource Editor können Sie eine neue Vorlage und darin eine Bibliothek zum Extrahieren und Gruppieren von Begriffen aus der Automobilindustrie erstellen. Da Sie die Informationen in dieser Bibliothek erneut benötigen werden, veröffentlichen Sie Ihre Bibliothek in einem zentralen Archiv, das Sie in dem Dialogfeld **Bibliotheken verwalten** aufrufen können, sodass sie unabhängig in verschiedenen Projekten wiederverwendet werden kann.

Angenommen, Sie möchten außerdem Fachausdrücke aus verschiedenen Unterbranchen, z. B. zu elektronischen Geräten, Motoren, Kühlsystemen, oder sogar Fachausdrücke eines bestimmten Herstellers oder Markts gruppieren. Sie können für jede Gruppe eine Bibliothek erstellen und sie dann so veröffentlichen, dass sie mit mehreren Textdatensätzen verwendet werden kann. So können Sie Bibliotheken hinzufügen, die dem Kontext Ihrer Textdaten am besten entsprechen.

Anmerkung: Zusätzliche Ressourcen können in der Registerkarte "Erweiterte Ressourcen" konfiguriert und verwaltet werden. Einige beziehen sich auf alle Bibliotheken und verwalten nicht linguistische Elemente, Ausnahmen bei der Fuzzy-Gruppierung usw. [Für weitere Informationen siehe Thema Informationen zu erweiterten Ressourcen in Kapitel 11 auf S. 242.](#)

Mitgelieferte Bibliotheken

Es werden standardmäßig mehrere Bibliotheken mit IBM® SPSS® Text Analytics for Surveys installiert. Mit diesen vorformatierten Bibliotheken können Sie auf Tausende vordefinierter Fachausdrücke und Synonyme sowie auf viele verschiedene Typen zugreifen. Diese mitgelieferten Bibliotheken sind auf verschiedene Domänen abgestimmt und in verschiedenen Sprachen verfügbar.

Es gibt eine Vielzahl von Bibliotheken, am häufigsten werden jedoch die folgenden verwendet:

- **Local Library.** Zum Speichern von benutzerdefinierten Wörterbüchern. Dies ist eine leere Bibliothek, die standardmäßig allen Ressourcen hinzugefügt wird. Sie enthält außerdem ein leeres Typ-Wörterbuch. Diese ist besonders bei Änderungen oder Feineinstellungen nützlich, die direkt über das Fenster “Textanalyse” vorgenommen werden (z. B. Hinzufügen eines Worts zu einem Typ). In diesem Fall werden diese Änderungen und Verfeinerungen automatisch in der ersten Bibliothek im Bibliotheksbaum im Resource Editor gespeichert – standardmäßig ist dies die *Local Library*. Diese Bibliothek können Sie nicht veröffentlichen, weil sie für die projekt-Daten bestimmt ist. Um ihren Inhalt zu veröffentlichen, müssen Sie die Bibliothek zunächst umbenennen.
- **Core Library.** Wird in den meisten Fällen verwendet, da sie die fünf integrierten Grundtypen, d. h. Personen, Orte, Organisationen, Produkte und Unbekannt, umfasst. Obwohl möglicherweise nur wenige Fachausdrücke in einem ihrer Typ-Wörterbücher aufgeführt werden, sind die in der Core Library dargestellten Typen tatsächlich Ergänzungen zu den robusten Typen in den internen, kompilierten Ressourcen, die im Lieferumfang Ihres Text-Mining-Produkts enthalten sind. Diese internen, kompilierten Ressourcen enthalten Tausende von Fachausdrücken für jeden Typ. Daher kann ein Fachausdruck dennoch extrahiert und mit einem Kerntyp versehen werden, auch wenn Sie ihn nicht in der Fachausdruckliste des Typ-Wörterbuchs sehen. Dies erklärt, wie Namen, z. B. *George*, extrahiert werden und den Typ <Person> erhalten können, wenn nur *John* in dem Typ-Wörterbuch <Person> in der Core Library aufgeführt wird. Wenn Sie die Core Library nicht einschließen, werden diese Typen ggf. trotzdem in den Extrahierungsergebnissen aufgeführt, da die kompilierten Ressourcen, die diese Typen enthalten, weiterhin von der Extrahierungsengine verwendet werden.
- **Opinions Library.** Am häufigsten zum Extrahieren von Meinungen und Stimmungen aus Textdaten verwendet. Diese Vorlage beinhaltet Tausende von Wörtern für Einstellungen, Vermerke und Präferenzen, die—wenn sie zusammen mit anderen Fachausdrücken verwendet werden—eine Meinung über ein Thema ausdrücken. Diese Bibliothek enthält eine Reihe von integrierten Typen, Synonymen und Ausschlüssen. Sie enthält außerdem eine große Menge an Musterregeln für die Textlinkanalyse.
- **Budget Library.** Wird zum Extrahieren von Begriffen zum Thema “Kosten” verwendet. Diese Bibliothek enthält zahlreiche Wörter und Ausdrücke, die Adjektive, Vermerke und Entscheidungen zu den Themen “Preis” oder “Qualität” darstellen.
- **Variations Library.** Wird verwendet, um Fälle einzuschließen, in denen bestimmte Sprachvarianten zur richtigen Gruppierung Synonymdefinitionen erfordern. Diese Bibliothek enthält nur Synonymdefinitionen.

Obwohl einige der mitgelieferten Bibliotheken außerhalb der Vorlagen dem Inhalt einiger Vorlagen ähneln, sind die Vorlagen speziell auf bestimmte Anwendungen abgestimmt und enthalten zusätzliche erweiterte Ressourcen. Es wird empfohlen, dass Sie es mit einer speziellen Vorlage für die Art der verwendeten Textdaten versuchen und Ihre Änderungen an diesen Ressourcen ausführen anstatt einer allgemeineren Vorlage nur individuelle Bibliotheken hinzuzufügen.

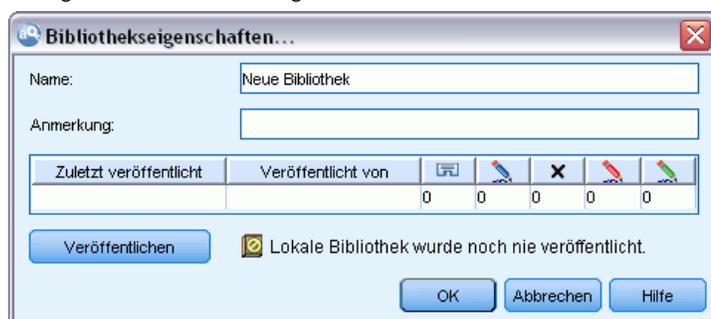
Es sind außerdem kompilierte Ressourcen im Lieferumfang von SPSS Text Analytics for Surveys enthalten. Sie werden bei jedem Extrahierungsprozess verwendet und enthalten eine große Anzahl ergänzender Definitionen zu den integrierten Wörterbüchern in den Standardbibliotheken. Da diese Ressourcen kompiliert sind, können sie nicht angezeigt oder

bearbeitet werden. Sie können jedoch die Aufnahme eines Fachausdrucks, der einen Typ durch diese kompilierten Ressourcen erhalten hat, in ein anderes Wörterbuch erzwingen. [Für weitere Informationen siehe Thema Erzwingen von Fachausdrücken in Kapitel 10 auf S. 231.](#)

Bibliotheken erstellen

Sie können beliebig viele Bibliotheken erstellen. Nach dem Erstellen einer neuen Bibliothek können Sie Typ-Wörterbücher in dieser Bibliothek erstellen und Fachausdrücke, Synonyme und Ausschlüsse eingeben.

Abbildung 9-1
Dialogfeld "Bibliotheken-Eigenschaften"



So erstellen Sie eine Bibliothek:

- ▶ Wählen Sie in den Menüs die Optionsfolge Ressourcen > Neue Bibliothek. Das Dialogfeld Bibliothek zu Projekt hinzufügen wird geöffnet.
- ▶ Geben Sie im Textfeld "Name" einen Namen für die Bibliothek ein.
- ▶ Bei Bedarf geben Sie einen Kommentar im Textfeld "Anmerkungen" ein.
- ▶ Klicken Sie auf Veröffentlichen, wenn Sie diese Bibliothek nun veröffentlichen möchten, bevor Sie Einträge in die Bibliothek vornehmen. [Für weitere Informationen siehe Thema Bibliotheken gemeinsam nutzen auf S. 217.](#) Sie können sie auch später zu einem beliebigen Zeitpunkt veröffentlichen.
- ▶ Klicken Sie auf OK, um die Bibliothek zu erstellen. Das Dialogfeld wird geschlossen und die Bibliothek wird in der Baumansicht aufgeführt. Wenn Sie alle Bibliotheken in dem Baum anzeigen, wird ein leeres Typ-Wörterbuch angezeigt, das automatisch in die Bibliothek aufgenommen worden ist. In dieses können Sie sofort Fachausdrücke aufnehmen. [Für weitere Informationen siehe Thema Hinzufügen von Fachausdrücken in Kapitel 10 auf S. 227.](#)

Öffentliche Bibliotheken hinzufügen

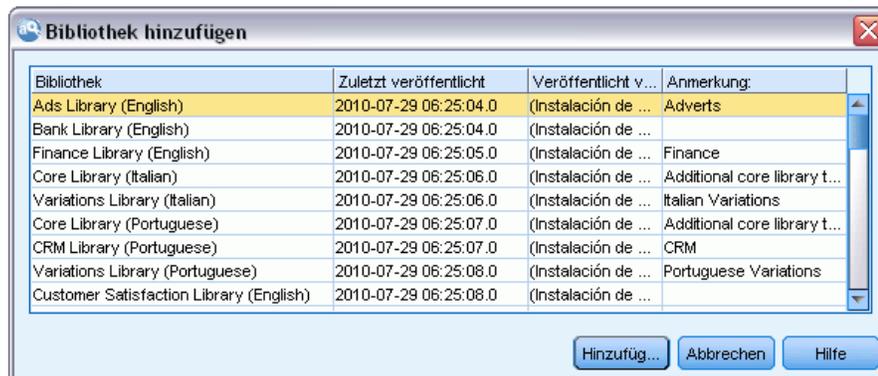
Um eine Bibliothek aus anderen Projektdaten wiederzuverwenden, fügen Sie sie Ihren aktuellen Ressourcen hinzu, vorausgesetzt, es handelt sich um eine Public Library. Eine **Public Library** ist eine Bibliothek, die veröffentlicht worden ist. [Für weitere Informationen siehe Thema Bibliotheken veröffentlichen auf S. 219.](#)

Wenn Sie eine Public Library hinzufügen, wird eine **lokale** Kopie in Ihre Projektdaten eingebettet. Sie können Änderungen an dieser Bibliothek vornehmen, müssen jedoch die öffentliche Version der Bibliothek erneut veröffentlichen, um die Änderungen mit anderen gemeinsam nutzen zu können.

Beim Hinzufügen einer Public Library kann das Dialogfeld “Konflikte auflösen” angezeigt werden, wenn Konflikte zwischen den Fachausdrücken und Typen in einer Bibliothek und den anderen Local Libraries festgestellt werden. Lösen Sie diese Konflikte auf oder akzeptieren Sie die vorgeschlagenen Lösungen, um den Vorgang abzuschließen. [Für weitere Informationen siehe Thema Konflikte auflösen auf S. 220.](#)

Anmerkung: Wenn Sie Ihre Bibliotheken jedesmal aktualisieren, wenn Sie ein Projekt öffnen, oder veröffentlichen, wenn Sie ein Projekt schließen, ist die Gefahr geringer, dass Bibliotheken asynchron werden. [Für weitere Informationen siehe Thema Bibliotheken gemeinsam nutzen auf S. 217.](#)

Abbildung 9-2
Dialogfeld “Bibliothek hinzufügen”



So fügen Sie eine Bibliothek hinzu:

- ▶ Wählen Sie in den Menüs die Optionsfolge Ressourcen > Bibliothek hinzufügen. Das Dialogfeld “Bibliotheken hinzufügen” wird geöffnet.
- ▶ Wählen Sie die Bibliotheken aus der Liste aus.
- ▶ Klicken Sie auf Hinzufügen. Wenn Konflikte zwischen den neu hinzugefügten Bibliotheken und bereits vorhandenen Bibliotheken auftreten, werden Sie aufgefordert, die Konfliktauflösungen vor Abschluss des Vorgangs zu bestätigen oder zu ändern. [Für weitere Informationen siehe Thema Konflikte auflösen auf S. 220.](#)

Fachausdrücke und Typen suchen

Sie können mit der Suchfunktion in den verschiedenen Bereichen im Editor suchen. Im Editor können Sie die Optionsfolge Bearbeiten > Suchen in den Menüs wählen, woraufhin die Symbolleiste “Suchen” angezeigt wird. Mit dieser Symbolleiste können Sie jeweils ein Vorkommnis suchen. Durch erneutes Klicken auf Suchen können Sie nachfolgende Vorkommnisse Ihres Suchausdrucks suchen.

Bei der Suche durchsucht der Editor nur die in der Dropdown-Liste der Symbolleiste “Suchen” aufgeführten Bibliotheken. Wenn Alle Bibliotheken ausgewählt ist, durchsucht das Programm alle Bibliotheken im Editor.

Eine Suche beginnt in dem fokussierten Bereich. Sie wird in Schleifen durch alle Abschnitte fortgesetzt, bis sie zur aktiven Zelle zurückkehrt. Mit den Richtungspfeilen können Sie die Suchrichtung umkehren. Sie können auch wählen, ob die Suche Groß-/Kleinschreibung unterscheiden soll.

So finden Sie Zeichenketten in der Ansicht:

- ▶ Wählen Sie im Menü Bearbeiten > Suchen. Die Symbolleiste “Suchen” wird angezeigt.
- ▶ Geben Sie die zu suchende Zeichenfolge ein.
- ▶ Klicken Sie auf die Schaltfläche Suchen, um die Suche zu starten. Das nächste Vorkommnis des Fachausdrucks oder Typs wird hervorgehoben.
- ▶ Klicken Sie erneut auf die Schaltfläche, um zu den nachfolgenden Vorkommnissen zu navigieren.

Bibliotheken anzeigen

Sie können den Inhalt einer bestimmten Bibliothek oder aller Bibliotheken anzeigen. Dies kann hilfreich sein, wenn Sie mit vielen Bibliotheken arbeiten oder wenn Sie den Inhalt einer bestimmten Bibliothek vor der Veröffentlichung überprüfen möchten. Eine Änderung der Ansicht beeinflusst nur die Anzeige in der Registerkarte “Bibliotheksressourcen”, deaktiviert aber keine Bibliotheken bei der Extrahierung. [Für weitere Informationen siehe Thema Local Libraries deaktivieren auf S. 214.](#)

Die Standardansicht ist Alle Bibliotheken, die alle Bibliotheken im Baum und ihren Inhalt in anderen Bereichen anzeigt. Sie können diese Auswahl mit der Dropdown-Liste auf der Symbolleiste oder durch eine Menüauswahl (Ansicht > Bibliotheken) ändern. Wenn eine einzelne Bibliothek angezeigt wird, werden alle Elemente in anderen Bibliotheken aus der Ansicht entfernt, aber trotzdem bei der Extrahierung gelesen.

So ändern Sie die Bibliothekenansicht:

- ▶ Wählen Sie in den Menüs der Registerkarte “Bibliotheksressourcen” die Optionsfolge Ansicht > Bibliotheken. Es wird ein Menü mit allen Local Libraries geöffnet.
- ▶ Wählen Sie die anzuzeigende Bibliothek aus oder wählen Sie die Option Alle Bibliotheken, um den Inhalt aller Bibliotheken anzuzeigen. Der Inhalt der Ansicht wird entsprechend Ihrer Auswahl gefiltert.

Local Libraries verwalten

Local Libraries sind im Gegensatz zu Public Libraries die Bibliotheken innerhalb Ihres -Projekts oder innerhalb einer Vorlage. [Für weitere Informationen siehe Thema Public Libraries verwalten auf S. 215.](#) Sie sollten außerdem einige grundlegende Maßnahmen zur Verwaltung der Bibliotheken durchführen, u. a.: Umbenennen, Deaktivieren oder Löschen einer Local Library.

Local Libraries umbenennen

Sie können Local Libraries umbenennen. Wenn eine Local Library umbenannt wird, wird sie von der öffentlichen Version getrennt, falls eine solche vorhanden ist. In diesem Fall können nachfolgende Änderungen nicht mehr in der öffentlichen Version gemeinsam genutzt werden. Sie können diese Local Library unter ihrem neuen Namen erneut veröffentlichen. Dies bedeutet außerdem, dass Sie in der ursprünglichen öffentlichen Version keine Änderungen aktualisieren können, die Sie in der lokalen Version vornehmen.

Anmerkung: Eine Public Library kann nicht umbenannt werden.

- ▶ Wählen Sie in den Menüs die Optionsfolge Bearbeiten > Bibliotheken-Eigenschaften. Das Dialogfeld “Bibliotheken-Eigenschaften” wird geöffnet.

Abbildung 9-3
Dialogfeld “Bibliotheken-Eigenschaften”



So benennen Sie eine Local Library um:

- ▶ Wählen Sie in der Baumansicht die Bibliothek aus, die Sie umbenennen möchten.
- ▶ Geben Sie im Textfeld “Name” einen neuen Namen für die Bibliothek ein.
- ▶ Klicken Sie auf OK, um den neuen Namen der Bibliothek zu akzeptieren. Das Dialogfeld wird geschlossen und der Name der Bibliothek wird in der Baumansicht aktualisiert.

Local Libraries deaktivieren

Um eine Bibliothek vorübergehend aus dem Extrahierungsprozess auszuschließen, heben Sie die Auswahl des Kontrollkästchens links von dem Namen der Bibliothek in der Baumansicht auf. Dies signalisiert, dass die Bibliothek beibehalten, aber der Inhalt bei der Konfliktprüfung und der Extrahierung ignoriert werden soll.

So deaktivieren Sie eine Bibliothek:

- ▶ Wählen Sie im Bibliotheksbaum die Bibliothek aus, die Sie deaktivieren möchten.
- ▶ Drücken Sie die Leertaste. Das Kontrollkästchen links neben dem Namen wird deaktiviert.

Local Libraries löschen

Sie können eine Bibliothek entfernen, ohne die öffentliche Version der Bibliothek zu löschen, und umgekehrt. Durch Löschen einer Local Library werden die Bibliothek und ihr gesamter Inhalt nur aus dem projekt gelöscht. Wenn Sie eine lokale Version einer Bibliothek löschen, wird diese Bibliothek nicht aus anderen Projekten entfernt und auch die öffentliche Version wird nicht entfernt. [Für weitere Informationen siehe Thema Public Libraries verwalten auf S. 215.](#)

So löschen Sie eine Local Library:

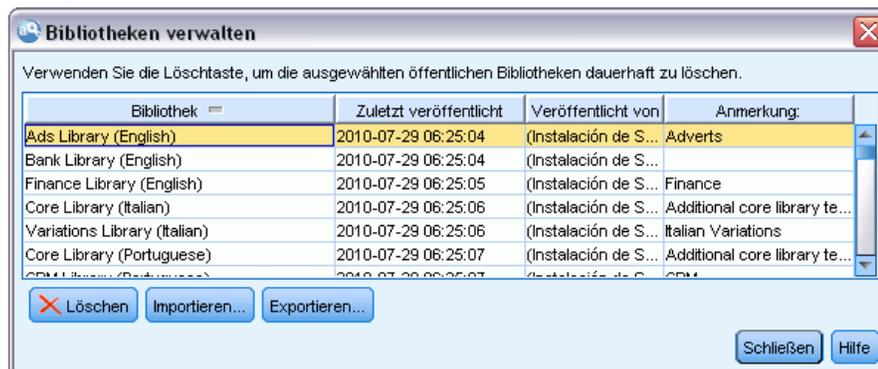
- ▶ Wählen Sie in der Baumansicht die zu löschende Bibliothek aus.
- ▶ Wählen Sie in den Menüs die Optionsfolge Bearbeiten > Löschen, um die Bibliothek zu löschen. Die Bibliothek wird entfernt.
- ▶ Wenn diese Bibliothek zuvor nicht veröffentlicht worden ist, werden Sie aufgefordert zu entscheiden, ob Sie die Bibliothek löschen oder beibehalten möchten. Klicken Sie auf Löschen, um fortzufahren, oder auf Beibehalten, um die Bibliothek beizubehalten.

Anmerkung: Es muss immer eine Bibliothek bestehen bleiben.

Public Libraries verwalten

Um Local Libraries wiederzuverwenden, können Sie sie veröffentlichen und sie dann über die Dialogbox “Bibliotheken verwalten” (Ressourcen > Bibliotheken verwalten) anzeigen und so mit ihnen arbeiten. [Für weitere Informationen siehe Thema Bibliotheken gemeinsam nutzen auf S. 217.](#) Zu den grundlegenden Maßnahmen zur Verwaltung von Public Libraries gehören das Importieren, Exportieren oder Löschen einer Public Library. Eine Public Library kann nicht umbenannt werden.

Abbildung 9-4
Dialogfeld “Bibliotheken verwalten”



Public Libraries importieren

- ▶ Klicken Sie im Dialogfeld “Bibliotheken verwalten” auf Importieren.... Das Dialogfeld “Bibliothek importieren” wird geöffnet.

Abbildung 9-5
Dialogfeld "Bibliothek importieren"



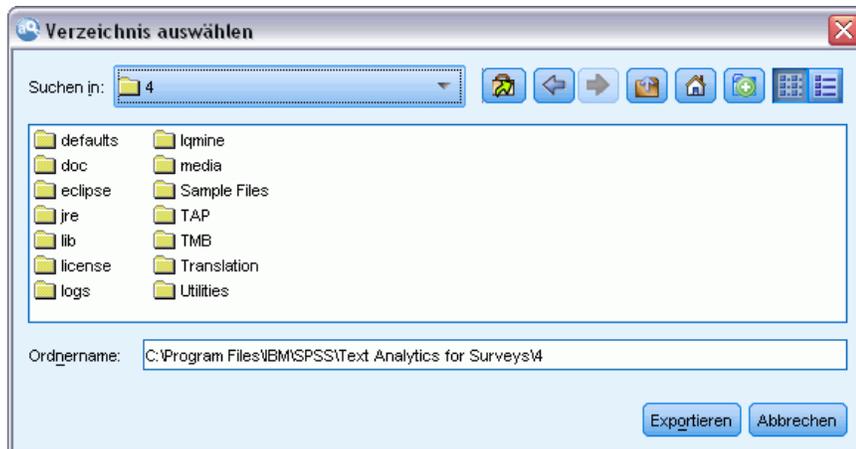
- ▶ Wählen Sie die zu importierende Bibliotheksdatei (*.lib) und wählen Sie Bibliothek zu aktuellem Projekt hinzufügen, wenn Sie diese Bibliothek außerdem lokal hinzufügen möchten.
- ▶ Klicken Sie auf Importieren. Das Dialogfeld wird geschlossen. Wenn bereits eine Public Library mit demselben Namen vorhanden ist, werden Sie aufgefordert, die zu importierende Bibliothek umzubenennen oder die aktuelle Public Library zu überschreiben.

Public Libraries exportieren

Sie können Public Libraries in das .lib-Format exportieren, sodass Sie sie mit anderen Benutzern gemeinsam nutzen können.

- ▶ Wählen Sie im Dialogfeld "Bibliotheken verwalten" die Bibliothek, die in die Liste exportiert werden soll.
- ▶ Klicken Sie auf Exportieren. Das Dialogfeld "Verzeichnis auswählen" wird geöffnet.

Abbildung 9-6
Dialogfeld "Verzeichnis auswählen"



- ▶ Wählen Sie das Verzeichnis aus, in das exportiert werden soll, und klicken Sie auf Exportieren. Das Dialogfeld wird geschlossen und die Bibliothekendatei (*.lib) wird exportiert.

Public Libraries löschen

Sie können eine Local Library entfernen, ohne die öffentliche Version der Bibliothek zu löschen, und umgekehrt. Wenn die Bibliothek jedoch aus diesem Dialogfeld entfernt wird, kann sie nicht mehr zu Projekten hinzugefügt werden, bis wieder eine lokale Version veröffentlicht wird.

Wenn eine Bibliothek gelöscht wird, die zusammen mit dem Produkt installiert wurde, wird die ursprünglich installierte Version wiederhergestellt.

- ▶ Wählen Sie im Dialogfeld “Bibliotheken verwalten” die zu löschende Bibliothek aus. Sortieren Sie die Liste, indem Sie auf den entsprechenden Titel klicken.
- ▶ Klicken Sie zum Löschen der Bibliothek auf Löschen. IBM® SPSS® Text Analytics for Surveys bestätigt, ob die lokale Version der Bibliothek mit der öffentlichen Bibliothek übereinstimmt. In dem Fall wird die Bibliothek ohne Warnhinweis entfernt. Wenn sich die Versionen der Bibliothek unterscheiden, wird ein Warnhinweis geöffnet, der Sie auffordert zu entscheiden, ob Sie die öffentliche Version beibehalten oder entfernen möchten.

Bibliotheken gemeinsam nutzen

Mithilfe von Bibliotheken können Sie Ressourcen leicht in mehreren Projekten gemeinsam nutzen. Bibliotheken können in zwei Status oder Versionen vorliegen. Bibliotheken, die einem bestimmten Projekt zugeordnet sind, werden als **Local Library** bezeichnet. Während der Arbeit in einem Projekt können Sie zahlreiche Änderungen, z. B. in der *Vegetables*-Bibliothek, vornehmen. Wenn Ihre Änderungen auch für andere Daten nützlich sein können, können Sie diese Ressourcen durch Erstellen einer **Public Library**-Version der *Vegetables*-Bibliothek verfügbar machen. Wie der Name andeutet, ist eine Public Library für alle anderen Projekten verfügbar.

Die Public Libraries werden im Dialogfeld “Bibliotheken verwalten” angezeigt. Sobald diese öffentliche Version der Bibliothek vorhanden ist, können Sie sie zu den Ressourcen in anderen Kontexten hinzufügen, sodass diese benutzerdefinierten linguistischen Ressourcen gemeinsam genutzt werden können.

Die mitgelieferten Bibliotheken sind anfänglich Public Libraries. Sie können die Ressourcen in diesen Bibliotheken bearbeiten und dann eine neue öffentliche Version erstellen. Die neuen Versionen sind dann in anderen neuen Projekten verfügbar.

Wenn Sie die Arbeit mit Ihren Bibliotheken fortsetzen und Änderungen vornehmen, werden die Versionen der Bibliothek asynchron. In einigen Fällen kann eine lokale Version aktueller als die öffentliche Version sein und in anderen Fällen kann die öffentliche Version aktueller als die lokale Version sein. Es ist auch möglich, dass sowohl die öffentliche als auch die lokale Version Änderungen enthalten, die nicht in der anderen enthalten sind, wenn die öffentliche Version aus einem anderen Projekt aktualisiert wurde. Wenn Ihre Bibliotheken asynchron werden, können Sie sie neu synchronisieren. Zum Synchronisieren der Versionen einer Bibliothek werden Local Libraries erneut veröffentlicht und/oder aktualisiert.

Wenn Sie ein Projekt öffnen oder schließen, werden Sie aufgefordert, alle Bibliotheken zu synchronisieren, die aktualisiert oder erneut veröffentlicht werden sollten. Sie können außerdem den Synchronisierungsstatus ihrer Local Libraries leicht an dem Symbol erkennen, das neben dem Namen der Bibliothek in der Baumansicht angezeigt wird, oder indem Sie das Dialogfeld “Bibliotheken-Eigenschaften” anzeigen. Dies ist auch zu einem beliebigen Zeitpunkt durch Menüauswahl möglich. Die folgende Tabelle beschreibt die fünf möglichen Status und die ihnen zugeordneten Symbole.

Tabelle 9-1
Status der Synchronisierung von Local Libraries

Symbol	Statusbeschreibung von Local Libraries
	Unveröffentlicht—Die Local Library ist nie veröffentlicht worden.
	Synchronisiert—Die lokale und die öffentliche Version der Bibliothek stimmen überein. Dies gilt auch für die <i>Local Library</i> , die nicht veröffentlicht werden kann, da sie nur projektspezifische Ressourcen enthalten soll.
	Veraltet—Die öffentliche Version der Bibliothek ist aktueller als die lokale Version. Sie können die lokale Version mit den Änderungen aktualisieren.
	Neuer—Die lokale Version der Bibliothek ist aktueller als die öffentliche Version. Sie können die lokale Version erneut als öffentliche Version veröffentlichen.
	Asynchron—Sowohl die Local Library als auch die Public Library enthalten Änderungen, die in der anderen nicht enthalten sind. Sie müssen die Local Library entweder aktualisieren oder veröffentlichen. Bei einer Aktualisierung verlieren Sie die Änderungen, die Sie seit der letzten Aktualisierung oder Veröffentlichung vorgenommen haben. Bei einer Veröffentlichung überschreiben Sie die Änderungen in der öffentlichen Version.

Anmerkung: Wenn Sie Ihre Bibliotheken jedesmal aktualisieren, wenn Sie ein Projekt öffnen, oder veröffentlichen, wenn Sie ein Projekt schließen, ist die Gefahr geringer, dass Bibliotheken asynchron werden.

Sie können eine Bibliothek zu einem beliebigen Zeitpunkt erneut veröffentlichen, zu dem die Änderungen in der Bibliothek anderen Projekten, die diese Bibliothek ebenfalls enthalten, zugute kommen. Wenn die Änderungen anderen Projekten zugute kommen, können Sie die lokalen Versionen in diesen Projekten aktualisieren. So können Sie Projekte für jeden Kontext oder jede Domäne erstellen, die zu Ihren Daten passen, indem Sie neue Bibliotheken erstellen und/oder Ihren Ressourcen eine beliebige Anzahl von Public Libraries hinzufügen.

Wenn eine öffentliche Version einer Bibliothek gemeinsam genutzt wird, ist die Wahrscheinlichkeit größer, dass Unterschiede zwischen der lokalen und der öffentlichen Version entstehen. Wenn Sie ein Projekt öffnen oder ein Projekt schließen und aus diesen Bibliotheken veröffentlichen, wird eine Meldung angezeigt, damit Sie Bibliotheken veröffentlichen und/oder aktualisieren können, die nicht mehr synchron mit denjenigen im Dialogfeld “Bibliotheken verwalten” sind. Wenn die öffentliche Version der Bibliothek aktueller als die lokale Version ist, wird ein Dialogfeld geöffnet, das Sie auffordert zu entscheiden, ob Sie aktualisieren möchten. Wählen Sie, ob Sie die lokale Version unverändert beibehalten möchten, statt sie mit der öffentlichen Version zu aktualisieren, oder ob Sie die Aktualisierungen in der Local Library zusammenführen möchten.

Bibliotheken veröffentlichen

Wenn eine bestimmte Bibliothek noch nie veröffentlicht worden ist, gehört zur Veröffentlichung die Erstellung einer öffentlichen Kopie Ihrer Local Library in der Datenbank. Bei der erneuten Veröffentlichung einer Bibliothek wird der Inhalt der öffentlichen Version durch den Inhalt der Local Library ersetzt. Nach der erneuten Veröffentlichung können Sie diese Bibliothek in allen anderen Projekten aktualisieren, sodass ihre lokalen Versionen synchron mit der öffentlichen Version sind. Obwohl eine Bibliothek veröffentlicht werden kann, wird immer eine lokale Version in dem Projekt gespeichert.

Wichtig: Wenn Sie Änderungen an Ihrer Local Library vornehmen und in der Zwischenzeit auch die öffentliche Version der Bibliothek geändert wurde, gilt Ihre Bibliothek als asynchron. Sie sollten zunächst die lokale Version mit den öffentlichen Änderungen aktualisieren, danach die gewünschten Änderungen vornehmen und dann Ihre lokale Version erneut veröffentlichen, damit beide Versionen übereinstimmen. Wenn Sie zuerst Änderungen vornehmen und die Bibliothek veröffentlichen, überschreiben Sie die Änderungen in der öffentlichen Version.

Abbildung 9-7
Dialogfeld "Bibliotheken veröffentlichen"



So veröffentlichen Sie Local Libraries in der Datenbank:

- ▶ Wählen Sie in den Menüs die Optionsfolge Ressourcen > Bibliotheken veröffentlichen. Das Dialogfeld "Bibliotheken veröffentlichen" wird geöffnet, wobei alle Bibliotheken, die veröffentlicht werden sollten, standardmäßig ausgewählt sind.
- ▶ Wählen Sie das Kontrollkästchen links von jeder Bibliothek, die Sie veröffentlichen oder erneut veröffentlichen möchten.
- ▶ Klicken Sie auf Veröffentlichen, um die Bibliotheken in der Datenbank "Bibliotheken verwalten" zu veröffentlichen.

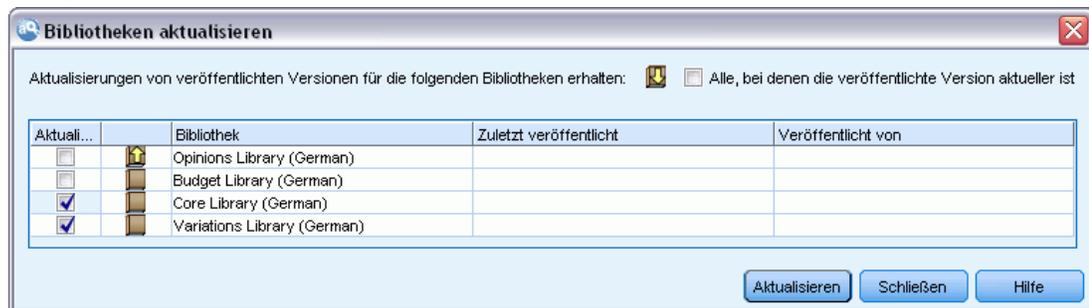
Bibliotheken aktualisieren

Wenn Sie ein Projekt öffnen oder schließen, können Sie alle Bibliotheken aktualisieren oder veröffentlichen, die nicht mehr synchron mit den öffentlichen Versionen sind. Wenn die öffentliche Version der Bibliothek aktueller als die lokale Version ist, wird ein Dialogfeld geöffnet,

das Sie auffordert zu entscheiden, ob Sie die Bibliothek aktualisieren möchten. Wählen Sie, ob Sie die lokale Version beibehalten möchten, statt sie mit der öffentlichen Version zu aktualisieren oder die lokale Version durch die öffentliche zu ersetzen. Wenn eine öffentliche Version einer Bibliothek aktueller als Ihre lokale Version ist, können Sie die lokale Version aktualisieren, um ihren Inhalt mit dem der öffentlichen Version zu synchronisieren. Aktualisieren bedeutet, die Änderungen in der öffentlichen Version in die lokale Version zu integrieren.

Anmerkung: Wenn Sie Ihre Bibliotheken jedesmal aktualisieren, wenn Sie ein Projekt öffnen, oder veröffentlichen, wenn Sie ein Projekt schließen, ist die Gefahr geringer, dass Bibliotheken asynchron werden. [Für weitere Informationen siehe Thema Bibliotheken gemeinsam nutzen auf S. 217.](#)

Abbildung 9-8
Dialogfeld "Bibliotheken aktualisieren"



So aktualisieren Sie Local Libraries:

- ▶ Wählen Sie in den Menüs die Optionsfolge Ressourcen > Bibliotheken aktualisieren. Das Dialogfeld "Bibliotheken aktualisieren" wird geöffnet, wobei alle Bibliotheken, die veröffentlicht werden sollten, standardmäßig ausgewählt sind.
- ▶ Wählen Sie das Kontrollkästchen links von jeder Bibliothek, die Sie veröffentlichen oder erneut veröffentlichen möchten.
- ▶ Klicken Sie zum Aktualisieren der Local Library auf Aktualisieren.

Konflikte auflösen

Konflikte zwischen Local und Public Libraries

Wenn Sie ein Projekt öffnen, führt IBM® SPSS® Text Analytics for Surveys einen Vergleich der Local Libraries mit den im Dialogfeld "Bibliotheken verwalten" aufgeführten Bibliotheken durch. Falls Local Libraries in Ihrem projekt nicht mit den veröffentlichten Versionen synchron sind, wird das Dialogfeld "Warnung zur Synchronisierung der Bibliothek" geöffnet. Wählen Sie eine der folgenden Optionen, um die Versionen der Bibliotheken auszuwählen, die Sie hier verwenden möchten:

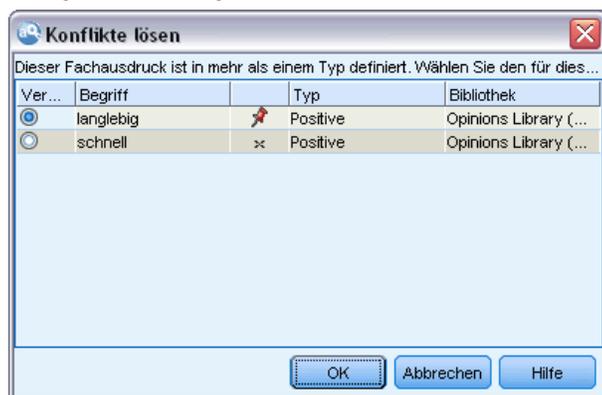
- **Alle Bibliotheken, die in der Datei als lokal eingestuft werden.** Mit dieser Option werden alle Local Libraries unverändert beibehalten. Sie können Sie später erneut veröffentlichen oder aktualisieren.

- **Alle auf diesem Computer veröffentlichten Bibliotheken.** Diese Option ersetzt die angezeigten Local Libraries durch die Versionen in der Datenbank.
- **Alle aktuelleren Bibliotheken.** Diese Option ersetzt die älteren Local Libraries durch die aktuelleren öffentlichen Versionen in der Datenbank.
- **Andere.** Mit dieser Option können Sie die gewünschten Versionen manuell in der Tabelle auswählen.

Konflikte durch erzwungene Fachausdrücke

Durch Hinzufügen einer Public Library oder Aktualisierung einer Local Library können Konflikte oder doppelte Einträge unter den Fachausdrücken und Typen in diesem Wörterbuch und den Fachausdrücken und Typen in den anderen Wörterbüchern in Ihren Ressourcen aufgedeckt werden. In diesem Fall werden Sie aufgefordert, die vorgeschlagenen Konfliktauflösungen vor Abschluss des Vorgangs in dem Dialogfeld “Erzwungene Fachausdrücke bearbeiten” zu verifizieren oder zu ändern. [Für weitere Informationen siehe Thema Erzwingen von Fachausdrücken in Kapitel 10 auf S. 231.](#)

Abbildung 9-9
Dialogfeld “Erzwungene Fachausdrücke bearbeiten”



Das Dialogfeld “Erzwungene Fachausdrücke bearbeiten” enthält alle Paare der in Konflikt stehenden Fachausdrücke oder Typen. Die Konfliktpaare werden durch abwechselnde Hintergrundfarben optisch unterschieden. Diese Farben können im Dialogfeld “Optionen” geändert werden. [Für weitere Informationen siehe Thema Optionen: Registerkarte “Anzeige” in Kapitel 2 auf S. 21.](#) Das Dialogfeld “Erzwungene Fachausdrücke bearbeiten” enthält zwei Registerkarten:

- **Doppelte Einträge.** Diese Registerkarte enthält die in den Bibliotheken ermittelten doppelten Einträge. Wenn ein Reißzwecken-Symbol hinter einem Fachausdruck angezeigt wird, ist das Vorkommen dieses Fachausdrucks erzwungen worden. Wenn ein schwarzes X-Symbol angezeigt wird, wird das Vorkommen dieses Fachausdrucks bei der Extrahierung ignoriert, weil er an anderer Stelle erzwungen worden ist.
- **Benutzerdefiniert.** Diese Registerkarte enthält eine Liste der Fachausdrücke, die manuell im Fachausdrucksbereich des Typ-Wörterbuchs und nicht durch Konflikte erzwungen worden sind.

Anmerkung: Das Dialogfeld “Erzwungene Fachausdrücke bearbeiten” wird geöffnet, nachdem Sie Einträge zu einer Bibliothek hinzugefügt oder die Bibliothek aktualisiert haben. Wenn Sie den Vorgang in diesem Dialogfeld abbrechen, brechen Sie nicht die Aktualisierung der Bibliothek oder das Hinzufügen ab.

So lösen Sie Konflikte auf:

- ▶ Wählen Sie im Dialogfeld “Erzwungene Fachausdrücke bearbeiten” die Optionsschaltfläche in der Spalte “Verwenden” für den zu erzwingenden Fachausdruck.
- ▶ Nach Abschluss der Auswahl klicken Sie auf OK, um die erzwungenen Fachausdrücke anzuwenden und das Dialogfeld zu schließen. Klicken Sie auf Abbrechen, um den Änderungsvorgang in diesem Dialogfeld abzubrechen.

Informationen zu Bibliothekswörterbüchern

Die zum Extrahieren von Textdaten verwendeten Ressourcen werden in Form von Vorlagen und Bibliotheken gespeichert. Eine Bibliothek kann aus drei Wörterbüchern bestehen.

- Das **Typ-Wörterbuch** enthält eine Zusammenstellung von Fachausdrücken, die unter einer Bezeichnung oder einem Typnamen gruppiert sind. Beim Lesen der Textdaten vergleicht die Extrahierungsengine die im Text aufgefundenen Wörter mit den Ausdrücken, die in den Typ-Wörterbüchern definiert sind. Bei der Extrahierung werden die gebeugten Formen der Fachausdrücke und Synonyme eines Typs unter einem Zielausdruck zusammengefasst, der als Konzept bezeichnet wird. Extrahierte Konzepte werden dem Typ-Wörterbuch zugewiesen, in dem sie als Ausdrücke erscheinen. Sie können Ihre Typ-Wörterbücher in den oben links und in der Mitte gelegenen Bereichen des Editors verwalten — in der Bibliotheksstruktur und im Ausdrucksbereich. [Für weitere Informationen siehe Thema Typ-Wörterbücher auf S. 223.](#)
- Das **Substitutionswörterbuch** enthält eine Zusammenstellung von Wörtern, die als Synonyme oder optionale Elemente definiert sind und zur Gruppierung ähnlicher Ausdrücke unter einem Zielausdruck verwendet werden, dem sogenannten “Konzept” in den Extrahierungsergebnissen. Sie können Ihre Substitutionswörterbücher im unten links gelegenen Bereich des Editors über die Registerkarte “Synonyme” und die Registerkarte “Optional” verwalten. [Für weitere Informationen siehe Thema Substitutions-/Synonymwörterbücher auf S. 234.](#)
- Das **Ausschlusswörterbuch** enthält eine Sammlung von Fachausdrücken und Typen, die aus den endgültigen Extrahierungsergebnissen entfernt werden. Sie können Ihre Ausschlusswörterbücher im ganz rechts gelegenen Bereich des Editors verwalten. [Für weitere Informationen siehe Thema Ausschlusswörterbücher auf S. 240.](#)

[Für weitere Informationen siehe Thema Mit Bibliotheken arbeiten in Kapitel 9 auf S. 209.](#)

Typ-Wörterbücher

Ein **Typ-Wörterbuch** besteht aus dem Typnamen bzw. der Überschrift und einer Liste von Fachausdrücken. Typ-Wörterbücher werden im oberen linken und im mittleren Bereich der Registerkarte “Bibliotheksressourcen” im Editor verwaltet. Sie können diese Ansicht mit der Optionsfolge Ansicht > Ressourceneditor in den Menüs aufrufen.

Wenn die Extrahierungsengine Ihre Textdaten liest, vergleicht sie die im Text gefundenen Wörter mit den in Ihren Typ-Wörterbüchern definierten Fachausdrücken. Fachausdrücke sind Wörter oder Ausdrücke in den Typ-Wörterbüchern in Ihren linguistischen Ressourcen.

Wenn ein Wort mit einem Fachausdruck übereinstimmt, wird es dem Typnamen für diesen Fachausdruck zugeordnet. Wenn die Ressourcen während des Extrahierungsprozesses gelesen werden, werden die im Text gefundenen Fachausdrücke einigen Verarbeitungsschritten unterzogen, bevor sie Konzepte im Bereich “Extrahierungsergebnisse” werden. Wenn mehrere Fachausdrücke, die zum selben Typen-Wörterbuch gehören, von der Extrahierungsengine als

synonym eingestuft werden, werden sie unter dem am häufigsten auftretenden Fachausdruck zusammengefasst und im Bereich “Extrahierungsergebnisse” als **Konzept** zusammengefasst. So könnten beispielsweise die Fachbegriffe *Frage* und *Abfrage* letzten Endes unter dem Konzeptnamen *Frage* zusammengefasst werden.

Abbildung 10-1
Bibliotheksbaum und Fachausdrucksbereich

Begriff	Zuordnung	Beugen	Typ	Bibliothek
riesengross	Gesamt und Start	<input checked="" type="checkbox"/>	Contextual	Opinions Library (German)
riesengroß	Gesamt und Start	<input checked="" type="checkbox"/>	Contextual	Opinions Library (German)
riesenhalt	Gesamt und Start	<input checked="" type="checkbox"/>	Contextual	Opinions Library (German)
riesig	Gesamt und Start	<input checked="" type="checkbox"/>	Contextual	Opinions Library (German)
ruecklaufig	Gesamt und Start	<input checked="" type="checkbox"/>	Contextual	Opinions Library (German)
rücklaufig	Gesamt und Start	<input checked="" type="checkbox"/>	Contextual	Opinions Library (German)
sardonisch	Gesamt und Start	<input checked="" type="checkbox"/>	Contextual	Opinions Library (German)
satirisch	Gesamt und Start	<input checked="" type="checkbox"/>	Contextual	Opinions Library (German)
schamhaft	Gesamt und Start	<input checked="" type="checkbox"/>	Contextual	Opinions Library (German)
schamhaftigkeit	Gesamt und Start	<input checked="" type="checkbox"/>	Contextual	Opinions Library (German)
schaulustig	Gesamt und Start	<input checked="" type="checkbox"/>	Contextual	Opinions Library (German)
scherzhaft	Gesamt und Start	<input checked="" type="checkbox"/>	Contextual	Opinions Library (German)
schlicht	Gesamt und Start	<input checked="" type="checkbox"/>	Contextual	Opinions Library (German)
schlichtheit	Gesamt und Start	<input checked="" type="checkbox"/>	Contextual	Opinions Library (German)
schrankenlos	Gesamt und Start	<input checked="" type="checkbox"/>	Contextual	Opinions Library (German)
schwer zu übertrumpfen	Gesamt (keine Zusammens...	<input type="checkbox"/>	Contextual	Opinions Library (German)
selbstgemacht	Gesamt und Start	<input checked="" type="checkbox"/>	Contextual	Opinions Library (German)
selbstgenuegsam	Gesamt und Start	<input checked="" type="checkbox"/>	Contextual	Opinions Library (German)
selbstgenügsam	Gesamt und Start	<input checked="" type="checkbox"/>	Contextual	Opinions Library (German)
selbstsicher	Gesamt und Start	<input checked="" type="checkbox"/>	Contextual	Opinions Library (German)
selbstverstaendlich	Gesamt und Start	<input checked="" type="checkbox"/>	Contextual	Opinions Library (German)
selbstverständlich	Gesamt und Start	<input checked="" type="checkbox"/>	Contextual	Opinions Library (German)
seltsam	Gesamt und Start	<input checked="" type="checkbox"/>	Contextual	Opinions Library (German)
seltsamkeit	Gesamt und Start	<input checked="" type="checkbox"/>	Contextual	Opinions Library (German)

Die Liste der Typ-Wörterbücher wird links im Bereich des Wörterbuchbaums angezeigt. Der Inhalt der einzelnen Typ-Wörterbücher wird im mittleren Bereich angezeigt. Typ-Wörterbücher enthalten mehr als nur eine Liste mit Fachausdrücken. Die festgelegte Abgleichsoption bestimmt, wie die Übereinstimmung der in Ihren Textdaten enthaltenen Wörter und Wortfolgen mit den in den Typ-Wörterbüchern definierten Fachausdrücken ermittelt wird. Eine **Abgleichsoption** legt fest, wie ein Fachausdruck hinsichtlich eines in den Textdaten enthaltenen möglichen Worts oder einer möglichen Wortfolge verankert ist. [Für weitere Informationen siehe Thema Hinzufügen von Fachausdrücken auf S. 227.](#)

Darüber hinaus können Sie die Fachbegriffe in Ihrem Wörterbuch ausweiten, indem Sie angeben, ob Sie im Wörterbuch automatisch gebeugte Formen der Fachbegriffe generieren und hinzufügen möchten. Indem Sie gebeugte Formen erzeugen, werden im Typ-Wörterbuch automatisch Pluralformen für im Singular angegebene Fachausdrücke, Singularformen von Fachausdrücken und Adjektive hinzugefügt, die im Plural angegeben sind. [Für weitere Informationen siehe Thema Hinzufügen von Fachausdrücken auf S. 227.](#)

Anmerkung: Konzepte, die in keinem Typen-Wörterbuch gefunden, aber aus dem Text extrahiert werden, erhalten automatisch den Typ <Unbekannt>.

Integrierte Typen

IBM® SPSS® Text Analytics for Surveys wird mit einem Satz linguistischer Ressourcen geliefert, die als Bibliotheken und kompilierte Ressourcen vorliegen. Die mitgelieferten Bibliotheken enthalten einen Satz integrierter Wörterbücher: <Location>, <Organization>, <Person> und <Product>

Diese Typwörterbücher werden von der Extraktionsengine benutzt, um den extrahierten Begriffen Typen zuzuweisen, z. B. den Typ <Location> zum Begriff Paris. Obwohl in den integrierten Wörterbüchern eine Vielzahl von Fachausdrücken definiert ist, decken diese nicht alle Möglichkeiten ab. Sie können daher Fachausdrücke hinzufügen oder eigene Wörterbücher anlegen. Eine Beschreibung der Inhalte eines der mitgelieferten Typ-Wörterbücher finden Sie in den Anmerkungen, die das Dialogfeld "Typeigenschaften" aufführt. Wählen Sie den Typ im Baum aus und wählen Sie die Optionsfolge Bearbeiten > Eigenschaften aus dem Kontextmenü.

Anmerkung: Neben den mitgelieferten Bibliotheken enthalten die kompilierten Ressourcen (die ebenfalls von der Extrahierungsengine verwendet werden) eine Vielzahl von Definitionen, die die integrierten Wörterbücher ergänzen. Ihre Inhalte sind im Produkt allerdings nicht einsehbar. Sie können jedoch erzwingen, dass ein Fachausdruck, dessen Typ durch die kompilierten Wörterbücher bestimmt wurde, in ein anderes Wörterbuch übertragen wird. [Für weitere Informationen siehe Thema Erzwingen von Fachausdrücken auf S. 231.](#)

Erstellen von Typen

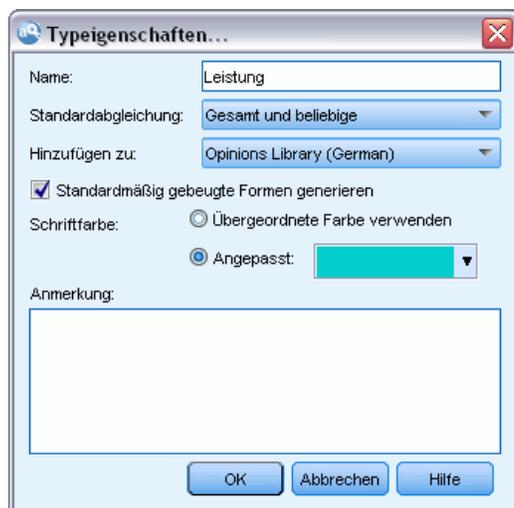
Sie können Typ-Wörterbücher erstellen, mit denen ähnliche Fachausdrücke gruppiert werden können. Wenn während des Extrahierungsprozesses in diesen Wörterbüchern vorhandene Fachausdrücke erkannt werden, wird diesen der entsprechende Typname zugewiesen und sie werden unter einem Konzeptnamen extrahiert. Wenn Sie eine Bibliothek erstellen, enthält diese automatisch eine leere Typbibliothek. Sie können daher sofort mit der Eingabe von Fachausdrücken beginnen.

Wenn Sie Text über Nahrungsmittel analysieren und Fachausdrücke zum Thema Gemüse gruppieren möchten, können Sie Ihr eigenes Typ-Wörterbuch <Gemüse> erstellen. Dort können Sie dann Begriffe wie Karotte, Brokkoli und Spinat hinzufügen, sofern Sie der Meinung sind, dass es sich um wichtige Fachausdrücke handelt, die im Text vorkommen. Wenn dann während der Extrahierung einer dieser Begriffe gefunden wird, wird dieser als Konzept extrahiert und dem Typ <Gemüse> zugeordnet.

Es ist nicht erforderlich, dass Sie alle Formen eines Worts oder eines Ausdrucks definieren, Sie können die gebeugten Formen der Begriffe stattdessen generieren lassen. Wenn Sie diese Option wählen, erkennt die Extrahierungsengine automatisch die Singular- bzw. Pluralformen der Fachausdrücke sowie andere zum Begriff gehörende Formen. Diese Option ist besonders hilfreich, wenn Ihr Typ überwiegend Substantive enthält, da es eher unwahrscheinlich ist, dass Sie gebeugte Formen von Verben oder Adjektiven erzeugen möchten.

Wichtig: Wir empfehlen dringend, keine neuen Typen in der Opinions Library zu erstellen, da sie sonst bei der Verarbeitung nicht einbezogen werden. Der Inhalt der Opinions Library wird anders gehandhabt als andere Bibliotheken, da er für die Erstellung von Mustern verwendet wird. Arbeiten Sie stattdessen in den Typen, die bereits in der Bibliothek vorhanden sind, oder fügen Sie einer anderen Bibliothek in Ihrem Projekt neue Typen hinzu.

Abbildung 10-2
Dialogfeld "Typ-Eigenschaften"



Name. Der Name des Typ-Wörterbuchs, das sie gerade erstellen. Es wird empfohlen, keine Leerzeichen in Typnamen zu verwenden, insbesondere wenn zwei oder mehr Typnamen mit demselben Wort beginnen.

Standardabgleich. Das Attribut "Standardabgleich" legt fest, wie die Extrahierungseingine die Übereinstimmung dieses Fachausdrucks mit den Textdaten ermittelt. Jedes Mal, wenn Sie einen Fachausdruck zu diesem Typ-Wörterbuch hinzufügen, wird diesem automatisch dieses Abgleichsattribut zugewiesen. Die Abgleichseinstellung können Sie in der Fachausdrucksliste jederzeit manuell ändern. Zu den Optionen gehören: Gesamter Begriff, Start, Ende, Beliebig, Start oder Ende, Gesamt und Start, Gesamt und Ende, Gesamt und (Start oder Ende) und Gesamt (keine Zusammensetzungen). [Für weitere Informationen siehe Thema Hinzufügen von Fachausdrücken auf S. 227.](#)

Hinzufügen zu. Dieses Feld gibt die Bibliothek an, in der Sie Ihr neues Typ-Wörterbuch erstellen.

Gebeugte Formen standardmäßig generieren. Diese Option weist die Extrahierungseingine an, die grammatikalische Morphologie zu verwenden, um ähnliche Formen der Fachausdrücke, die Sie zu diesem Wörterbuch hinzufügen, zu erfassen und zusammenzufassen, wie Singular- oder Pluralformen des Worts. Diese Option ist besonders hilfreich, wenn Ihr Typ überwiegend Substantive enthält. Wenn Sie diese Option auswählen, besitzen alle neu zu diesem Typ hinzugefügten Fachausdrücke automatisch diese Option, die Sie in der Liste allerdings manuell ändern können.

Schriftfarbe. Über dieses Feld können Sie ein Unterscheidungsmerkmal für diesen Typ in Bezug auf die anderen auf der Benutzeroberfläche dargestellten Typen festlegen. Wenn Sie die Option Übergeordnete Farbe verwenden auswählen, wird für dieses Typ-Wörterbuch die Standardtypfarbe verwendet. Die Standardfarbe wird im Dialogfeld "Optionen" festgelegt. [Für weitere Informationen siehe Thema Optionen: Registerkarte "Anzeige" in Kapitel 2 auf S. 21.](#) Wenn Sie Benutzerdefiniert auswählen, wählen Sie aus der Dropdown-Liste eine Farbe aus.

Anmerkung. Dieses Feld ist optional und wird für Kommentare und Beschreibungen verwendet.

So erstellen Sie ein Typ-Wörterbuch:

- ▶ Wählen Sie eine Bibliothek aus, in der Sie ein neues Typ-Wörterbuch erstellen möchten.
- ▶ Wählen Sie im Menü Extras > Neuer Typ aus. Das Dialogfeld “Typ-Eigenschaften” wird geöffnet.
- ▶ Geben Sie den Namen für Ihr Typ-Wörterbuch in das Textfeld Name ein und wählen Sie die gewünschten Optionen aus.
- ▶ Klicken Sie auf OK, um das Typ-Wörterbuch zu erstellen. Der neue Typ wird im Bibliotheksbaum und im mittleren Bereich angezeigt. Sie können sofort beginnen, Fachausdrücke hinzuzufügen. Weitere Informationen finden Sie unter [Hinzufügen von Fachausdrücken](#).

Anmerkung: Diese Anweisungen zeigen Ihnen, wie Sie in der Resource Editor-Ansicht Änderungen vornehmen können. Beachten Sie, dass Sie solche Feinabstimmungen auch direkt in den Bereichen “Extrahierungsergebnisse”, “Daten”,.. [Für weitere Informationen siehe Thema Extrahierungsergebnisse verfeinern in Kapitel 5 auf S. 88.](#)

Hinzufügen von Fachausdrücken

Im Bereich des Bibliotheksbaums werden die Bibliotheken angezeigt. Diese können erweitert werden, um die darin enthaltenen Typ-Wörterbücher anzuzeigen. Im mittleren Bereich zeigt eine Liste je nach Auswahl im Baum die in der ausgewählten Bibliothek oder im ausgewählten Typ-Wörterbuch enthaltenen Fachausdrücke an.

Abbildung 10-3
Ausdrucksbereich

Begriff	Zuordnung	Beugen	Typ	Bibliothek
		<input type="checkbox"/>		
a.g	Ende	<input type="checkbox"/>	Organization	Core Library (German)
a.g.	Ende	<input type="checkbox"/>	Organization	Core Library (German)
ag	Ende	<input type="checkbox"/>	Organization	Core Library (German)
co	Ende	<input type="checkbox"/>	Organization	Core Library (German)
co.	Ende	<input type="checkbox"/>	Organization	Core Library (German)
corp	Ende	<input type="checkbox"/>	Organization	Core Library (German)
corp.	Ende	<input type="checkbox"/>	Organization	Core Library (German)
gbh	Ende	<input type="checkbox"/>	Organization	Core Library (German)
gmbh	Ende	<input type="checkbox"/>	Organization	Core Library (German)
inc	Ende	<input type="checkbox"/>	Organization	Core Library (German)
inc.	Ende	<input type="checkbox"/>	Organization	Core Library (German)
kg	Ende	<input type="checkbox"/>	Organization	Core Library (German)
kgaa	Ende	<input type="checkbox"/>	Organization	Core Library (German)
l.l.c	Ende	<input type="checkbox"/>	Organization	Core Library (German)
l.l.c.	Ende	<input type="checkbox"/>	Organization	Core Library (German)
llc	Ende	<input type="checkbox"/>	Organization	Core Library (German)
lly	Ende	<input type="checkbox"/>	Organization	Core Library (German)
ltd	Ende	<input type="checkbox"/>	Organization	Core Library (German)
ltd.	Ende	<input type="checkbox"/>	Organization	Core Library (German)
mbh	Ende	<input type="checkbox"/>	Organization	Core Library (German)
plc	Ende	<input type="checkbox"/>	Organization	Core Library (German)
s.a	Ende	<input type="checkbox"/>	Organization	Core Library (German)
s.a.	Ende	<input type="checkbox"/>	Organization	Core Library (German)
s.c.a	Ende	<input type="checkbox"/>	Organization	Core Library (German)
s.c.a.	Ende	<input type="checkbox"/>	Organization	Core Library (German)
sa	Ende	<input type="checkbox"/>	Organization	Core Library (German)
sca	Ende	<input type="checkbox"/>	Organization	Core Library (German)
spss deutschland gmbh	Gesamter Begriff	<input type="checkbox"/>	Organization	Core Library (German)

Im Resource Editor können Sie Fachausdrücke zu einem Typ-Wörterbuch direkt im Fachausdrucksbereich oder über das Dialogfeld “Neue Fachausdrücke hinzufügen” hinzufügen. Bei den hinzugefügten Fachausdrücken kann es sich um einzelne Wörter oder um Wortfolgen handeln. Am Anfang der Liste befindet sich stets eine leere Zeile, in die Sie ein neues Wort eingeben können.

Anmerkung: Diese Anweisungen zeigen Ihnen, wie Sie in der Resource Editor-Ansicht Änderungen vornehmen können. Beachten Sie, dass Sie solche Feinabstimmungen auch direkt in den Bereichen “Extrahierungsergebnisse”, “Daten”, „. [Für weitere Informationen siehe Thema Extrahierungsergebnisse verfeinern in Kapitel 5 auf S. 88.](#)

Ausdrucksspalte

Geben Sie in diese Spalte ein Wort oder eine Wortfolge in die Zelle ein. In welcher Farbe der Fachausdruck angezeigt wird, hängt von der Farbe des Typs ab, in dem der Ausdruck gespeichert wurde. Die Typfarben können Sie im Dialogfeld “Typeigenschaften” ändern. [Für weitere Informationen siehe Thema Erstellen von Typen auf S. 225.](#)

Erzwingen-Spalte

Wenn Sie in dieser Spalte in dieser Zelle ein Reißzweckensymbol hinzufügen, weisen Sie die Extrahierungsengine an, das Vorkommen desselben Ausdrucks in allen anderen Bibliotheken zu ignorieren. [Für weitere Informationen siehe Thema Erzwingen von Fachausdrücken auf S. 231.](#)

Übereinstimmungsspalte

Wählen Sie in dieser Spalte eine Übereinstimmungsoption aus, um festzulegen, wie die Extrahierungsengine die Übereinstimmung dieses Fachausdrucks mit den Textdaten ermittelt. Beispiele finden Sie in der Tabelle. Sie können den Standardwert ändern, indem Sie die Typeigenschaften bearbeiten. [Für weitere Informationen siehe Thema Erstellen von Typen auf S. 225.](#) Wählen Sie im Menü Bearbeiten > Abgleich ändern. Folgende Beispiele sind grundlegende Abgleiche, da Kombinationen aus ihnen ebenfalls möglich sind:

- **Start.** Dieser Typ wird zugewiesen, wenn das im Wörterbuch gefundene Konzept mit dem Anfang eines aus dem Text extrahierten Fachausdrucks übereinstimmt. Wenn Sie beispielsweise `Apfel` eingeben, gilt `Apfelkuchen` als Übereinstimmung.
- **Ende.** Dieser Typ wird zugewiesen, wenn das im Wörterbuch gefundene Konzept mit dem Ende eines aus dem Text extrahierten Fachausdrucks übereinstimmt. Wenn Sie beispielsweise `Apfel` eingeben, gilt `Klarapfel` als Übereinstimmung.
- **Beliebig.** Dieser Typ wird zugewiesen, wenn das im Wörterbuch gefundene Konzept mit irgendeinem Teil eines aus dem Text extrahierten Fachausdrucks übereinstimmt. Wenn Sie z. B. `Apfel` eingeben, sorgt die Option `Alle` dafür, dass `Apfelkuchen`, `Klarapfel` und `Klarapfelkuchen` demselben Typ zugeordnet werden.
- **Gesamter Begriff.** Wenn das gesamte aus dem Text extrahierte Konzept exakt mit dem in der Bibliothek vorhandenen Ausdruck übereinstimmt, wird dieser Typ zugewiesen. Das Hinzufügen eines Fachausdrucks als `Gesamter Begriff`, `Start` und `Ende`, `Gesamt` und `Ende`, `Gesamt` und `Beliebig` oder `Gesamt` (keine Zusammensetzungen) erzwingt die Extrahierung eines Fachausdrucks.

Da der Typ `<Person>` außerdem nur zweiteilige Namen wie *edith piaf* oder *mohandas gandhi* extrahiert, sollten Sie gegebenenfalls die Vornamen in dieses Typ-Wörterbuch aufnehmen, wenn Sie versuchen, einen Vornamen zu extrahieren, wenn kein Nachname angegeben ist. Wenn Sie beispielsweise alle Instanzen von *edith* als Name erfassen möchten, sollten Sie `edith` dem Typ `<Person>` über `Gesamter Begriff` oder `Gesamt` und `Start` hinzufügen.

- **Gesamt (keine Zusammensetzungen).** Wenn das gesamte aus dem Text extrahierte Konzept exakt mit dem Fachausdruck im Wörterbuch übereinstimmt, wird dieser Typ zugewiesen und die Extrahierung wird beendet, um die Extrahierung aufgrund einer Übereinstimmung des Fachausdrucks mit einem längeren zusammengesetzten Ausdruck zu verhindern. Wenn Sie beispielsweise `Apfel` eingeben, wird der Begriff durch die Option `Gesamt` (keine Zusammensetzung) als Übereinstimmung für `Apfel` gehandhabt, nicht jedoch für den zusammengesetzten Ausdruck `Grüner Apfel`, sofern dies nicht an anderer Stelle erzwungen wird.

Die folgende Tabelle geht von einem Fachausdruck `Apfel` in einem Typ-Wörterbuch aus. Je nach Abgleichsoption zeigt diese Tabelle an, welche Konzepte extrahiert und mit Typen versehen würden, wenn sie im Text gefunden würden.

Tabelle 10-1
Abgleichsbeispiele

Abgleichsoptionen für den Fachausdruck:  Apfel	Extrahierte Konzepte			
	Apfel	ApfelKuchen	reif Apfel	<i>selbst gemachter ApfelKuchen</i>
Gesamter Begriff	✓			
Start		✓		
Ende			✓	
Start oder Ende		✓	✓	
Gesamt und Start	✓	✓		
Gesamt und Ende	✓		✓	
Gesamt und (Start oder Ende)	✓	✓	✓	
Beliebig		✓	✓	✓
Gesamt und Beliebig	✓	✓	✓	✓
Gesamt (keine Zusammensetzungen)	✓	<i>nie extrahiert</i>	<i>nie extrahiert</i>	<i>nie extrahiert</i>

Flektierungsspalte

Wählen Sie in dieser Spalte aus, ob die Extrahierungsengine während der Extrahierung gebeute Formen dieses Fachausdrucks erzeugen soll, so dass sie alle zusammengefasst werden. Der Standardwert für diese Spalte ist in den Typeigenschaften definiert. Sie können diese Option direkt in der Spalte, aber für jeden einzelnen Fall ändern. Wählen Sie im Menü Bearbeiten > Beugung ändern.

Typspalte

Wählen Sie in dieser Spalte ein Typ-Wörterbuch aus der Dropdown-Liste aus. Die Liste der Typen wird anhand der von Ihnen im Bibliotheksbaum getroffenen Auswahl gefiltert. Als erster Typ wird in der Liste immer der im Bibliotheksbaum ausgewählte Standardtyp angezeigt. Wählen Sie im Menü Bearbeiten > Typ ändern.

Bibliotheksspalte

In dieser Spalte wird die Bibliothek angezeigt, in der Ihr Fachausdruck gespeichert wird. Sie können einen Begriff mit der Maus auf einen anderen Typ im Bibliotheksbaum ziehen, um seine Bibliothek zu ändern.

So fügen Sie einen einzelnen Fachausdruck zu einem Typ-Wörterbuch hinzu:

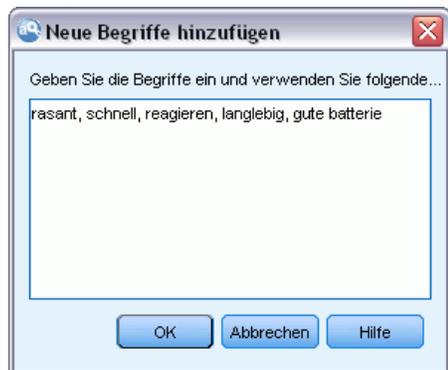
- ▶ Wählen Sie im Bibliotheksbaum das Typ-Wörterbuch aus, zu dem Sie den Fachausdruck hinzufügen möchten.
- ▶ Geben Sie in der mittleren Liste der Fachausdrücke Ihren Fachausdruck in die erste verfügbare leere Zelle ein und wählen Sie die gewünschten Optionen für diesen Fachausdruck aus.

So fügen Sie mehrere Fachausdrücke zu einem Typ-Wörterbuch hinzu:

- ▶ Wählen Sie im Bibliotheksbaumbereich das Typ-Wörterbuch aus, dem Sie die Fachausdrücke hinzufügen möchten.
- ▶ Wählen Sie im Menü Extras > Neue Begriffe aus. Das Dialogfeld “Neue Fachausdrücke hinzufügen” wird geöffnet.

Abbildung 10-4

Dialogfeld “Neue Fachausdrücke hinzufügen”



- ▶ Geben Sie die Begriffe ein, die Sie zum ausgewählten Typ-Wörterbuch hinzufügen möchten. Sie können die Begriffe eingeben oder mehrere Begriffe kopieren und einfügen. Wenn Sie mehrere Fachausdrücke eingeben, müssen Sie diese mit dem Trennzeichen, das im Dialogfeld “Optionen” definiert ist, trennen oder jeden Begriff in einer neuen Zeile eingeben. [Für weitere Informationen siehe Thema Festlegen von Optionen in Kapitel 2 auf S. 19.](#)
- ▶ Klicken Sie auf OK, um die Begriffe zum Wörterbuch hinzuzufügen. Als Abgleichsoption wird automatisch die Standardoption für dieses Typ-Wörterbuch festgelegt. Das Dialogfeld wird geschlossen und die neuen Begriffe erscheinen im Wörterbuch.

Erzwingen von Fachausdrücken

Wenn Sie möchten, dass ein Fachausdruck einem bestimmten Typ zugewiesen wird, können Sie ihn zum entsprechenden Typ-Wörterbuch hinzufügen. Wenn jedoch mehrere Fachausdrücke mit demselben Namen vorliegen, muss die Extrahierungsengine wissen, welcher Typ verwendet werden soll. Sie werden daher aufgefordert, den zu verwendenden Typ auszuwählen. Dies wird als **Erzwingen** der Zuordnung eines Fachausdrucks zu einem Typ bezeichnet. Diese Option ist vor allem beim Überschreiben der Typenzuordnung aus einem kompilierten (internen, nicht bearbeitbaren) Wörterbuch hilfreich. Generell empfehlen wir, doppelte Fachbegriffe von vornherein zu vermeiden.

Das Erzwingen *entfernt* die anderen Stellen, an denen der Fachausdruck vorkommt, nicht. Diese werden stattdessen von der Extrahierungsengine ignoriert. Sie können später ändern, welche Fundstelle verwendet wird, indem Sie das Erzwingen eines Fachausdrucks festlegen oder aufheben. Es kann außerdem erforderlich sein, dass Sie erzwingen, dass ein Fachausdruck in ein Typ-Wörterbuch eingefügt wird, wenn Sie eine öffentliche Bibliothek hinzufügen oder aktualisieren.

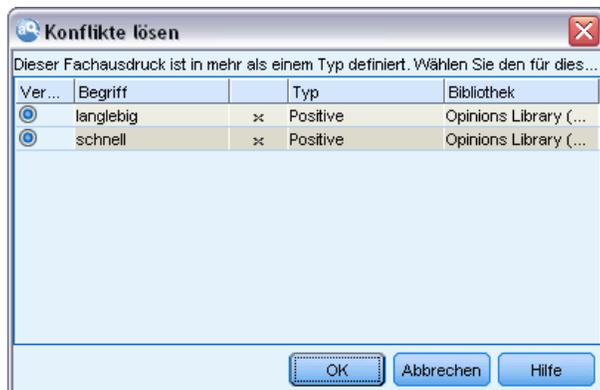
Abbildung 10-5
Symbole für den Erzwingungsstatus

Begriff	Zuordnung	Beugen	Typ	Bibliothek
\$\$	Gesamter Begriff	<input type="checkbox"/>	Budget	Budget Library (English)
amount of money	Gesamter Begriff	<input type="checkbox"/>	Budget	Budget Library (English)
arepriced	Gesamter Begriff	<input type="checkbox"/>	Budget	Budget Library (English)
award	× Gesamt und (Start oder Ende)	<input checked="" type="checkbox"/>	Budget	Budget Library (English)
balance	Gesamt und (Start oder Ende)	<input checked="" type="checkbox"/>	Budget	Budget Library (English)
bank charge	× Gesamter Begriff	<input checked="" type="checkbox"/>	Budget	Budget Library (English)
bill	Gesamt und Beliebige	<input checked="" type="checkbox"/>	Budget	Budget Library (English)

In der Erzwingungsspalte (2. Spalte im Fachausdrucksbereich) sehen Sie, welche Ausdrücke erzwungen oder ignoriert werden. Wenn ein Reißzweckensymbol angezeigt wird, bedeutet dies, dass dieses Vorkommen des Ausdrucks erzwungen wurde. Wenn ein schwarzes X-Symbol angezeigt wird, bedeutet dies, dass dieses Vorkommen des Ausdrucks während des Extrahierens ignoriert wird, da er anderweitig erzwungen wurde. Wenn Sie einen Fachausdruck erzwingen, wird er außerdem in der Farbe des Typs angezeigt, für den er erzwungen wurde. Dies bedeutet, dass ein Ausdruck, der sowohl in Typ 1 als auch in Typ 2 vorkommt und den Sie als Typ 1 erzwungen haben, im Fenster immer in der für Typ 1 definierten Schriftfarbe angezeigt wird.

Um den Status zu ändern, können Sie auf das Symbol doppelklicken. Wenn der Fachausdruck an einer anderen Stelle vorkommt, wird das Dialogfeld “Konflikte auflösen” geöffnet, in dem Sie auswählen können, welche Fundstelle verwendet wird.

Abbildung 10-6
Dialogfeld “Konflikte auflösen”



Umbenennen von Typen

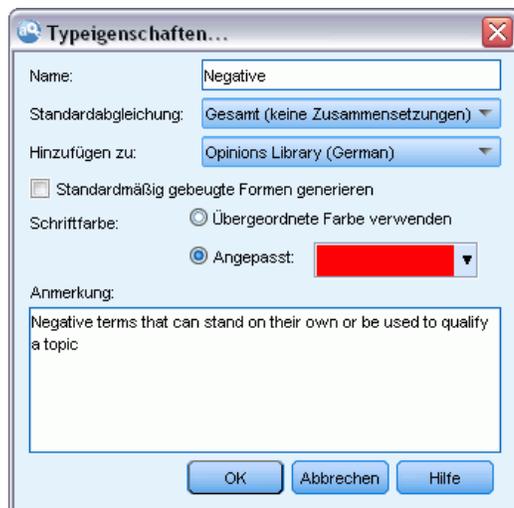
Sie können ein Typ-Wörterbuch umbenennen oder andere Wörterbucheinstellungen ändern, indem Sie die Typeigenschaften bearbeiten.

Wichtig: Es wird empfohlen, keine Leerzeichen in Typnamen zu verwenden, insbesondere wenn zwei oder mehr Typnamen mit demselben Wort beginnen. Es ist außerdem empfehlenswert, die Typen in der Core Library und Opinions Library nicht umzubenennen oder ihre Standard-Übereinstimmungsattribute zu ändern.

So benennen Sie einen Typ um:

- ▶ Wählen Sie im Bereich des Bibliotheksbaums das Typ-Wörterbuch aus, das Sie umbenennen möchten.
- ▶ Klicken Sie mit der rechten Maustaste und wählen Sie im Kontextmenü Typeigenschaften aus. Das Dialogfeld “Typ-Eigenschaften” wird geöffnet.

Abbildung 10-7
Dialogfeld “Typ-Eigenschaften”



- ▶ Geben Sie den neuen Namen für Ihr Typ-Wörterbuch in das Textfeld “Name” ein.
- ▶ Klicken Sie auf OK, um den neuen Namen zu übernehmen. Der neue Name wird im Bibliotheksbaum angezeigt.

Verschieben von Typen

Sie können ein Typ-Wörterbuch mit der Maus an eine andere Stelle innerhalb einer Bibliothek oder in eine andere Bibliothek im Baum ziehen.

Anmerkung: Es ist empfehlenswert, die integrierten Typen nicht zu verschieben.

So ordnen Sie einen Typ in einer Bibliothek neu an:

- ▶ Wählen Sie im Bereich des Bibliotheksbaums das Typ-Wörterbuch aus, das Sie verschieben möchten.
- ▶ Wählen Sie im Menü Bearbeiten > Nach oben verschieben aus, um das Typ-Wörterbuch im Bibliotheksbaum um eine Position nach oben zu verschieben, bzw. Bearbeiten > Nach unten verschieben, um es um eine Position nach unten zu verschieben.

So verschieben Sie einen Typ in eine andere Bibliothek:

- ▶ Wählen Sie im Bereich des Bibliotheksbaums das Typ-Wörterbuch aus, das Sie verschieben möchten.

- ▶ Klicken Sie mit der rechten Maustaste und wählen Sie im Kontextmenü Typeigenschaften aus. Das Dialogfeld “Typ-Eigenschaften” wird geöffnet. (Sie können den Typ auch mit der Maus in eine andere Bibliothek ziehen.)
- ▶ Wählen Sie im Feld “Hinzufügen zu” die Bibliothek aus, in die Sie das Typ-Wörterbuch verschieben möchten.
- ▶ Klicken Sie auf OK. Das Dialogfeld wird geschlossen und der Typ befindet sich in der von Ihnen ausgewählten Bibliothek.

Deaktivieren und Löschen von Typen

Wenn Sie ein Typ-Wörterbuch vorübergehend entfernen möchten, können Sie im Bibliotheksbaum das Kontrollkästchen links neben dem Namen des Wörterbuchs deaktivieren. Dies bewirkt, dass das Wörterbuch in Ihrer Bibliothek bleibt, dass die Inhalte bei der Prüfung von Konflikten und während des Extrahierungsvorgangs aber ignoriert werden.

Sie können Typ-Wörterbücher auch dauerhaft aus einer Bibliothek entfernen.

Anmerkung: Es ist empfehlenswert, die integrierten Typen in der Core Library oder Opinions Library nicht zu löschen. Es ist besser, sie stattdessen zu deaktivieren.

So deaktivieren Sie ein Typ-Wörterbuch:

- ▶ Wählen Sie im Bereich des Bibliotheksbaums das Typ-Wörterbuch aus, das Sie deaktivieren möchten.
- ▶ Drücken Sie die Leertaste. Das Kontrollkästchen links neben dem Typnamen wird deaktiviert.

So löschen Sie ein Typ-Wörterbuch:

- ▶ Wählen Sie im Bereich des Bibliotheksbaums das Typ-Wörterbuch aus, das Sie löschen möchten.
- ▶ Wählen Sie im Menü Bearbeiten > Löschen aus, um das Typ-Wörterbuch zu löschen.

Substitutions-/Synonymwörterbücher

Ein **Substitutionswörterbuch** ist eine Sammlung von Begriffen, mit deren Hilfe ähnliche Begriffe unter einem Zielbegriff gruppiert werden. Substitutionswörterbücher werden im unteren Bereich der Registerkarte “Bibliotheksressourcen” verwaltet. Sie können diese Ansicht mit der Optionsfolge Ansicht > Ressourceneditor in den Menüs aufrufen.

In diesen Wörterbüchern können Sie zwei Formen von Substitutionen definieren: **Synonyme** und **optionale Elemente**. Sie können auf die Registerkarten in diesem Bereich klicken, um zwischen ihnen zu wechseln.

Nachdem Sie eine Extrahierung Ihrer Textdaten durchgeführt haben, finden Sie möglicherweise mehrere Konzepte, bei denen es sich um Synonyme oder um gebeugte Formen anderer Konzepte handelt. Indem optionale Elemente und Synonyme identifiziert werden, können Sie erzwingen, dass die Extrahierungseingabe diese auf einen einzigen Zielbegriff abbildet.

Die Substituierung der Verwendung von Synonymen und optionalen Elementen verringert die Anzahl an Konzepten im Bereich “Extrahierungsergebnisse”, indem sie in sinnvollere, aussagekräftigere Konzepte mit einer höheren Häufigkeit von Anzahlen zusammengefasst werden.

Abbildung 10-8
Bereich des Substitutionswörterbuchs

	Ziel	Synonyme	Bibliothek
0	<input type="checkbox"/>		Lokale Bibliothek
1	<input checked="" type="checkbox"/>	klug, intelligent, kenntnisreich	Opinions Library (German)
2	<input checked="" type="checkbox"/>	lausflügen, lausflug, lausflüge, lausflügen, lausflüger, lausflüglern, lausflüglers	Opinions Library (German)
3	<input checked="" type="checkbox"/>	lausfällig, ausfaellig , ausfallend	Opinions Library (German)
4	<input checked="" type="checkbox"/>	ifamilien, ifamilien	Opinions Library (German)
5	<input checked="" type="checkbox"/>	ifrein, ifreinheit	Opinions Library (German)
6	<input checked="" type="checkbox"/>	iflach, iflachheit	Opinions Library (German)
7	<input checked="" type="checkbox"/>	ifläche, ifläche, iflächen	Opinions Library (German)
8	<input checked="" type="checkbox"/>	ifruh, ifrueh	Opinions Library (German)
	<input checked="" type="checkbox"/>	lgross, lgeröss, lgrrosse, lgrassem, lgrossen, lgrösser, lgrösseres, lgrösses, lgroß, lgroße, lgroßem, lgroßen, lgroßer, lgroßes, lgrössere,	Opinions Library (German)

Synonyme

Synonyme verknüpfen zwei oder mehr Wörter mit derselben Bedeutung. Mithilfe von Synonymen können Sie außerdem Fachausdrücke mit ihren Abkürzungen gruppieren oder auch falsch geschriebene Wörter mit dem Fachausdruck in der richtigen Schreibweise. Sie können diese Synonyme auf der Registerkarte “Synonyme” definieren.

Eine Synonymdefinition setzt sich aus zwei Teilen zusammen. Der erste Teil ist ein Zielausdruck, d. h. der Ausdruck, unter dem die Extrahierungsengine alle Synonyme zusammenfassen soll. Wenn dieser Zielausdruck nicht als Synonym eines anderen Zielausdrucks verwendet wird oder er ausgeschlossen wird, ist es wahrscheinlich, dass dieser Ausdruck das Konzept wird, das im Bereich “Extrahierungsergebnisse” angezeigt wird. Der zweite Teil besteht aus einer Liste von Synonymen, die unter dem Zielbegriff zusammengefasst werden.

Wenn Sie beispielsweise *Automobil* durch *Fahrzeug* ersetzen möchten, dann ist *Automobil* das Synonym und *Fahrzeug* der Zielausdruck.

Sie können jedes beliebige Wort in die Spalte Synonym setzen, wenn das Wort jedoch bei der Extrahierung nicht gefunden wird und der Fachausdruck eine Ableichsoption mit *Gesamt* hatte, kann keine Substitution vorgenommen werden. Der Zielausdruck muss jedoch nicht extrahiert werden, damit die Synonyme unter diesem Begriff zusammengefasst werden.

Abbildung 10-9
Substitutionswörterbuch, Registerkarte “Synonyme”

	Ziel	Synonyme	Bibliothek
0	<input type="checkbox"/>		Core Library (English)
1	<input checked="" type="checkbox"/>	a.g	Product Satisfaction Library (
2	<input checked="" type="checkbox"/>	look, lookin, the way it looks	Product Satisfaction Library (
3	<input checked="" type="checkbox"/>	advertisement, ad, advert, advertasing, advertise, advertising, advertisement	Product Satisfaction Library (
4	<input checked="" type="checkbox"/>	aftertaste, after-taste, a.g	Product Satisfaction Library (
5	<input checked="" type="checkbox"/>	anti-spam, antispam, control, anti-spam	Product Satisfaction Library (
6	<input checked="" type="checkbox"/>	appearance	Product Satisfaction Library (
7	<input checked="" type="checkbox"/>	authorisation, authorise, authorising, authorization, authorize, authorizing	Product Satisfaction Library (
8	<input checked="" type="checkbox"/>	battery, abbtery, batery, batti, battery life, battrey	Product Satisfaction Library (
9	<input checked="" type="checkbox"/>	call-waiting, call waiting	Product Satisfaction Library (
10	<input checked="" type="checkbox"/>	characteristic, attribute, charatceristic, properties	Product Satisfaction Library (
11	<input checked="" type="checkbox"/>	comfort	Product Satisfaction Library (
12	<input checked="" type="checkbox"/>	communicate, communicate, amount of mail, communication, communication,	Product Satisfaction Library (

Optionale Elemente

Optionale Elemente kennzeichnen optionale Wörter in einem zusammengesetzten Begriff, die während der Extrahierung ignoriert werden können, um ähnliche Begriffe auch dann zusammenzuhalten, wenn sie im Text leicht unterschiedlich auftreten. Optionale Elemente sind einzelne Wörter, die, wenn sie aus einem zusammengesetzten Ausdruck entfernt werden, eine Übereinstimmung mit einem anderen Ausdruck darstellen können. Diese einzelnen Wörter können an einer beliebigen Stelle innerhalb des zusammengesetzten Ausdrucks vorkommen – am Anfang, in der Mitte oder am Ende. Sie können optionale Elemente auf der Registerkarte “Optional” definieren.

Um z. B. die Ausdrücke `ibm` und `ibm corp` zu gruppieren, müssen Sie festlegen, dass `corp` in diesem Fall als optionales Element behandelt werden soll. Wenn Sie andererseits festlegen, dass `access` ein optionales Element ist und wenn bei der Extrahierung sowohl `internet access speed` als auch `internet speed` gefunden werden, dann werden diese unter dem Ausdruck gruppiert, der am häufigsten vorkommt.

Abbildung 10-10
Substitutionswörterbuch, Registerkarte “Optional”

Optionale Elemente	Bibliothek
<input checked="" type="checkbox"/>	Local Library
<input checked="" type="checkbox"/>	Product Satisfaction Library (English)
<input checked="" type="checkbox"/>	Opinions Library (English)
<input checked="" type="checkbox"/>	Budget Library (English)
<input checked="" type="checkbox"/>	Core Library (English)
<input checked="" type="checkbox"/>	Variations Library (English)
<input checked="" type="checkbox"/>	Variations Library (German)

<input checked="" type="checkbox"/>	<code>a.g.</code> , <code>a.g.</code> , <code>ag.</code> , <code>co.</code> , <code>co.</code> , <code>corp.</code> , <code>corp.</code> , <code>corporation.</code> , <code>gmbh.</code> , <code>gmbh.</code> , <code>inc.</code> , <code>inc.</code> , <code>incorporated.</code> , <code>kgaa.</code> , <code>l.l.c.</code> , <code>l.l.c.</code> , <code>llc.</code> , <code>ltd.</code> , <code>ltd.</code> , <code>org.</code> , <code>plc.</code> , <code>s.a.</code> , <code>s.a.</code> , <code>s.c.a.</code> , <code>s.c.a.</code> , <code>sa.</code> , <code>sca</code>
-------------------------------------	---

Synonyme Optionale Elemente

Definieren von Synonymen

Auf der Registerkarte “Synonyme” können Sie in die Leerzeile am Anfang der Tabelle eine Synonymdefinition eingeben. Beginnen Sie, indem Sie den Zielausdruck und seine Synonyme definieren. Sie können auch die Bibliothek auswählen, in der diese Definition gespeichert werden soll. Während der Extrahierung werden alle Fundstellen der Synonyme für die endgültige Extrahierung unter dem Zielausdruck gruppiert. [Für weitere Informationen siehe Thema Hinzufügen von Fachausdrücken auf S. 227.](#)

Wenn Ihre Textdaten viele Fachausdrücke aus der Telekommunikation enthalten, liegen beispielsweise folgende Fachausdrücke vor: `cellular phone`, `wireless phone` und `mobile phone`. In diesem Beispiel sollten Sie `cellular` und `mobile` als Synonyme für `wireless` definieren. Wenn Sie diese Synonyme definieren, werden alle extrahierten Fundstellen von `cellular phone` und `mobile phone` als derselbe Ausdruck wie `wireless phone` behandelt und zusammen in der Liste der Fachausdrücke ausgegeben.

Wenn Sie Ihre Typ-Wörterbücher erstellen, können Sie einen Ausdruck eingeben, für den Ihnen drei oder vier Synonyme einfallen. In diesem Fall könnten Sie alle Ausdrücke und anschließend Ihren Zielausdruck in das Substitutionswörterbuch eingeben und dann die Synonyme durch Ziehen übertragen.

Die Substitution von Synonymen wird auch auf gebeugte Formen (wie Pluralformen) der Synonyme angewendet. Je nach Kontext sollten Sie Einschränkungen für die Substitution der Fachausdrücke festlegen. Sie können bestimmte Zeichen verwenden, um Einschränkungen dafür festzulegen, wie weit die Substitution durchgeführt wird:

- **Ausrufezeichen (!).** Wenn das Ausrufezeichen direkt vor dem Synonym steht `!Synonym`, bedeutet das, dass keine gebeugten Formen des Synonyms durch den Zielausdruck ersetzt werden. Ein Ausrufezeichen direkt vor dem Zielausdruck `!Zielausdruck` bedeutet hingegen, dass kein weiterer Bestandteil dieses zusammengesetzten Fachausdrucks oder Varianten davon substituiert werden sollen.
- **Sternchen (*).** Ein direkt nach dem Synonym stehendes Sternchen (*) (`Synonym*`) bedeutet, dass dieses Wort durch den Zielausdruck ersetzt werden soll. Wenn Sie beispielsweise den Quellausdruck `manage*` als Synonym und `management` als Zielausdruck definiert haben, dann wird `associate managers` durch den Zielausdruck `associate management` ersetzt. Sie können nach dem Wort auch ein Leerzeichen und ein Sternchen (*) einfügen (`Synonym *`), wie z. B. `internet *`. Wenn Sie den Zielausdruck `internet` sowie die Synonyme `internet * *` und `web *` definiert haben, dann werden `internet access card` und `web portal` entsprechend durch `internet` ersetzt. Ein Wort oder eine Zeichenfolge kann in diesem Wörterbuch nicht mit dem Sternchen (*) als Platzhalterzeichen beginnen.
- **Winkelzeichen (^).** Ein vor einem Synonym stehendes Winkelzeichen und ein Leerzeichen (`^ Synonym>`) bedeuten, dass die Synonymgruppierung nur dann durchgeführt wird, wenn der Fachausdruck mit dem Synonym beginnt. Wenn Sie beispielsweise `^ wage` als das Synonym und `income` als den Zielausdruck definiert haben und beide Ausdrücke extrahiert werden, dann werden beide unter dem Ausdruck `income` gruppiert. Werden dagegen `minimum wage` und `income` extrahiert, erfolgt keine Gruppierung, weil `minimum wage` nicht mit `wage` beginnt. Zwischen diesem Symbol und dem Synonym muss ein Leerzeichen eingefügt werden.
- **Dollarzeichen (\$).** Ein nach einem Synonym stehendes Leerzeichen und ein Dollarzeichen (`Synonym $`) bedeuten, dass die Synonymgruppierung nur dann durchgeführt wird, wenn der Fachausdruck mit dem Synonym endet. Wenn Sie beispielsweise `cash $` als das Synonym und `money` als den Zielausdruck definiert haben und beide Ausdrücke extrahiert werden, dann werden beide unter dem Ausdruck `money` gruppiert. Werden dagegen `cash cow` und `money` extrahiert, erfolgt keine Gruppierung, weil `cash cow` nicht auf `cash` endet. Zwischen diesem Symbol und dem Synonym muss ein Leerzeichen eingefügt werden.
- **Winkelzeichen (^) und Dollarzeichen (\$).** Wenn Winkel- und Dollarzeichen zusammen verwendet werden (`^ Synonym $`), stimmt ein Ausdruck nur bei exakter Übereinstimmung mit dem Synonym überein. Dies bedeutet, dass im extrahierten Ausdruck vor oder nach dem Synonym keine Wörter stehen dürfen, damit die Synonymgruppierung stattfindet. Beispiel: Sie definieren `^ van $` als Synonym und `truck` als Zielausdruck, sodass nur `van` mit `truck` gruppiert wird, während `marie van guerin` unverändert bleibt. Wenn Sie ein Synonym mit dem Winkel- und dem Dollarzeichen definieren und wenn das entsprechende Wort an einer Stelle des Quelltexts auftritt, wird das Synonym außerdem automatisch extrahiert.

Abbildung 10-11
Substitutionswörterbuch, Registerkarte "Synonyme" mit Beispiel

	Ziel	Synonyme	Bibliothek
0			
1	look	look, lookin, the way it looks	Core Library (English)
2	advertisement	ad, advert, advertasing, advertise, advertising, advertisement	Product Satisfaction Library (
3	aftertaste	after taste, after-taste, a.g	Product Satisfaction Library (
4	anti-spam	anti spam, antispam, control, anti-spam	Product Satisfaction Library (
5	appearance	appearance	Product Satisfaction Library (
6	authorisation	authorise, authorising, authorization, authorize, authorizing	Product Satisfaction Library (
7	battery	abattery, batery, batt, battery life, batttery	Product Satisfaction Library (
8	call-waiting	call waiting	Product Satisfaction Library (
9	characteristic	attribute, charatceristic, properties	Product Satisfaction Library (
10	comfort	confort	Product Satisfaction Library (
11	communication	^ communicate, ^ communicate, amount of mail, commuication, communciation,	Product Satisfaction Library (
12			

Synonyme Optionale Elemente

So fügen Sie einen Synonymeintrag hinzu:

- ▶ Klicken Sie im Substitutionsbereich oben links auf die Registerkarte Synonyme.
- ▶ Geben Sie in der leeren Zeile am Anfang der Tabelle Ihren Zielausdruck in die Zielspalte ein. Der von Ihnen eingegebene Zielausdruck wird farblich angezeigt. Die Farbe stellt den Typ dar, als der der Ausdruck erscheint oder in den er erzwungen wird, sofern dies der Fall ist. Wenn der Ausdruck schwarz dargestellt wird, bedeutet dies, dass er in keinem der Typ-Wörterbücher vorkommt.
- ▶ Klicken Sie auf die zweite Zelle rechts neben dem Zielausdruck und geben Sie die Synonyme ein. Trennen Sie die einzelnen Einträge mithilfe des globalen Trennzeichens, das im Dialogfeld "Optionen" definiert ist. Für weitere Informationen siehe Thema Festlegen von Optionen in Kapitel 2 auf S. 19. Die von Ihnen eingegebenen Ausdrücke werden farblich angezeigt. Die Farbe stellt den Typ dar, in der der Ausdruck angezeigt wird. Wenn der Ausdruck schwarz dargestellt wird, bedeutet dies, dass er in keinem der Typ-Wörterbücher vorkommt.
- ▶ Klicken Sie auf die letzte Zelle, um die Bibliothek auszuwählen, in der die Synonymdefinition gespeichert werden soll.

Anmerkung: Diese Anweisungen zeigen Ihnen, wie Sie in der Resource Editor-Ansicht Änderungen vornehmen können. Beachten Sie, dass Sie solche Feinabstimmungen auch direkt in den Bereichen "Extrahierungsergebnisse", "Daten",. Für weitere Informationen siehe Thema Extrahierungsergebnisse verfeinern in Kapitel 5 auf S. 88.

Definieren optionaler Elemente

Auf der Registerkarte "Optional" können Sie für beliebige Bibliotheken optionale Elemente definieren. Diese Einträge werden für alle Bibliotheken miteinander gruppiert. Sobald eine Bibliothek zum Bibliotheksbaum hinzugefügt wird, wird auf der Registerkarte "Optional" eine leere Zeile für optionale Elemente hinzugefügt.

Alle Einträge werden automatisch in Kleinschrift umgewandelt. Die Extrahierungseingine ordnet Wörter, die im Text in Groß- oder Kleinbuchstaben vorkommen, zu.

Abbildung 10-12
Substitutionswörterbuch, Registerkarte "Optional"

Optionale Elemente	Bibliothek
<input checked="" type="checkbox"/>	Local Library
<input checked="" type="checkbox"/>	Product Satisfaction Library (English)
<input checked="" type="checkbox"/>	Opinions Library (English)
<input checked="" type="checkbox"/>	Budget Library (English)
<input checked="" type="checkbox"/> a.g., a.g., ag, co., co., corp, corp., corporation, gbh, gmbh, inc, inc., incorporated, kga, llc, llc., llc, ltd, ltd., org, plc, s.a, s.a., s.c.a, s.c.a., sa, sca	Core Library (English)
<input checked="" type="checkbox"/>	Variations Library (English)
<input checked="" type="checkbox"/>	Variations Library (German)

Synonyme Optionale Elemente

Anmerkung: Ausdrücke werden mithilfe des Trennzeichens getrennt, das im Dialogfeld "Optionen" definiert ist. Für weitere Informationen siehe Thema Festlegen von Optionen in Kapitel 2 auf S. 19. Wenn das von Ihnen eingegebene optionale Element das Trennzeichen als Teil des Ausdrucks enthält, müssen Sie diesem einen umgekehrten Schrägstrich (\) voranstellen.

So fügen Sie einen Eintrag hinzu:

- ▶ Klicken Sie im Editor im Substitutionsbereich oben links auf die Registerkarte "Optional".
- ▶ Klicken Sie in der Spalte "Optionale Elemente" auf die Zelle der Bibliothek, zu der Sie diesen Eintrag hinzufügen möchten.
- ▶ Geben Sie das optionale Element ein. Trennen Sie die einzelnen Einträge mithilfe des globalen Trennzeichens, das im Dialogfeld "Optionen" definiert ist. Für weitere Informationen siehe Thema Festlegen von Optionen in Kapitel 2 auf S. 19.

Deaktivieren und Löschen von Substitutionen

Sie können einen Eintrag vorübergehend entfernen, indem Sie ihn in Ihrem Wörterbuch deaktivieren. Deaktivierte Einträge werden bei der Extrahierung ignoriert.

Sie können zudem veraltete Einträge aus Ihrem Substitutionswörterbuch löschen.

So deaktivieren Sie einen Eintrag:

- ▶ Wählen Sie in Ihrem Wörterbuch den Eintrag aus, den Sie deaktivieren möchten.
- ▶ Drücken Sie die Leertaste. Das Kontrollkästchen links neben dem Eintrag wird deaktiviert.

Anmerkung: Sie können auch links neben dem Eintrag auf das Kontrollkästchen klicken, um es zu deaktivieren.

So löschen Sie einen Synonymeintrag:

- ▶ Wählen Sie in Ihrem Wörterbuch den Eintrag aus, den Sie löschen möchten.
- ▶ Wählen Sie in den Menüs die Optionen Bearbeiten > Löschen oder drücken Sie die Taste Entfernen auf Ihrer Tastatur. Der Eintrag wird aus dem Wörterbuch entfernt.

So löschen Sie einen Eintrag für ein optionales Element:

- ▶ Doppelklicken Sie in Ihrem Wörterbuch auf den Eintrag, den Sie löschen möchten.
- ▶ Löschen Sie diesen Ausdruck manuell.
- ▶ Drücken Sie die Eingabetaste, damit die Änderung wirksam wird.

Ausschlusswörterbücher

Ein **Ausschlusswörterbuch** ist eine Liste von Wörtern, Ausdrücken oder Teilzeichenfolgen. Ausdrücke, die mit einem Eintrag im Ausschlusswörterbuch übereinstimmen, werden ignoriert oder aus der Extrahierung ausgeschlossen. Ausschlusswörterbücher werden im rechten Bereich des Editors verwaltet. In der Regel handelt es sich bei den zu dieser Liste hinzugefügten Ausdrücken um Füllwörter oder Wortfolgen, die im Text verwendet werden, aber keine wichtigen Informationen enthalten und daher die Extrahierungsergebnisse stören können. Wenn Sie solche Wörter zum Ausschlusswörterbuch hinzufügen, stellen Sie sicher, dass sie nie extrahiert werden.

Ausschlusswörterbücher werden im oberen rechten Bereich der Registerkarte “Bibliotheksressourcen” im Editor verwaltet. Sie können diese Ansicht mit der Optionsfolge Ansicht > Ressourceneditor in den Menüs aufrufen.

Abbildung 10-13
Ausschlusswörterbuchbereich

	Ausschlussliste	Bibliothek
0	<input type="checkbox"/>	
1	<input checked="" type="checkbox"/> any kind of problem	Opinions Library (English)
2	<input checked="" type="checkbox"/> any problems i have	Opinions Library (English)
3	<input checked="" type="checkbox"/> anykinf of problem	Opinions Library (English)
4	<input checked="" type="checkbox"/> as usual	Opinions Library (English)
5	<input checked="" type="checkbox"/> can't wait	Opinions Library (English)
6	<input checked="" type="checkbox"/> i was out of	Opinions Library (English)
7	<input checked="" type="checkbox"/> if i ever have a problem	Opinions Library (English)
8	<input checked="" type="checkbox"/> if i ever have problems	Opinions Library (English)
9	<input checked="" type="checkbox"/> if i have a problem	Opinions Library (English)
10	<input checked="" type="checkbox"/> if i have questions	Opinions Library (English)
11	<input checked="" type="checkbox"/> if there are problems	Opinions Library (English)
12	<input checked="" type="checkbox"/> if there is a problem	Opinions Library (English)

In das Ausschlusswörterbuch können Wörter, Wortfolgen oder Teile von Zeichenfolgen in die Leerzeile am Anfang der Tabelle eingegeben werden. Sie können Zeichenfolgen zu Ihrem Ausschlusswörterbuch als eines oder mehrere Wörter hinzufügen oder auch als Wortteile, indem Sie das Sternchen als Platzhalter verwenden. Die im Ausschlusswörterbuch erfassten Einträge werden verwendet, um Konzepte von der Extrahierung auszuschließen. Falls ein Eintrag auch an anderer Stelle deklariert ist, beispielsweise in einem Typ-Wörterbuch, wird er in den anderen Wörterbüchern durchgestrichen dargestellt, was anzeigt, dass er zur Zeit ausgeschlossen ist. Diese Zeichenfolge muss nicht in den Textdaten vorkommen oder als Teil eines Typ-Wörterbuchs deklariert sein, um angewendet zu werden.

Anmerkung: Wenn Sie ein Konzept zum Ausschlusswörterbuch hinzufügen, das in einem Synonymeintrag als Zielausdruck verwendet wird, dann werden auch das Ziel und alle Synonyme ausgeschlossen. [Für weitere Informationen siehe Thema Definieren von Synonymen auf S. 236.](#)

Verwenden von Platzhaltern (*)

Sie können das Sternchen als Platzhalter verwenden, wenn der Ausschlusseintrag als Teilzeichenfolge behandelt werden soll. Alle Ausdrücke, die von der Extrahierungsengine aufgefunden werden und ein Wort enthalten, das mit einer im Ausschlusswörterbuch angegebenen Zeichenfolge beginnt oder endet, werden von der endgültigen Extrahierung ausgeschlossen. Es gibt jedoch zwei Fälle, in denen ein Platzhalterzeichen nicht zulässig ist:

- Bindestrich (-), vor dem ein Sternchen (*) als Platzhalterzeichen steht (*-)
- Apostroph (‘), vor dem ein Sternchen (*) als Platzhalterzeichen steht (*’s)

Tabelle 10-2

Beispiele für Ausschlusseinträge

Aufnahme (Cox Regression)	Beispiel	Ergebnisse
Wort	<i>next</i>	Es werden keine Konzepte (bzw. die darin enthaltenen Ausdrücke) extrahiert, die das Wort <i>next</i> enthalten.
Wortfolge	<i>for example</i>	Es werden keine Konzepte (bzw. die darin enthaltenen Ausdrücke) extrahiert, die den Ausdruck <i>for example</i> enthalten.
Teilweise	<i>copyright*</i>	Schließt alle Konzepte (bzw. die darin enthaltenen Ausdrücke) aus, die mit Variationen des Worts <i>copyright</i> übereinstimmen bzw. diese enthalten, wie <i>copyrighted</i> , <i>copyrighting</i> , <i>copyrights</i> oder <i>copyright 2010</i> .
Teilweise	<i>*ware</i>	Schließt alle Konzepte (bzw. die darin enthaltenen Ausdrücke) aus, die mit Variationen des Worts <i>ware</i> übereinstimmen bzw. diese enthalten, wie <i>freeware</i> , <i>shareware</i> , <i>software</i> , <i>hardware</i> , <i>beware</i> oder <i>silverware</i> .

So fügen Sie Einträge hinzu

- ▶ Geben Sie in die Leerzeilen am Anfang der Tabelle einen Ausdruck ein. Der von Ihnen eingegebene Ausdruck wird farbig angezeigt. Die Farbe stellt den Typ dar, in der der Ausdruck angezeigt wird. Wenn der Ausdruck schwarz dargestellt wird, bedeutet dies, dass er in keinem der Typ-Wörterbücher vorkommt.

So deaktivieren Sie Einträge

Sie können einen Eintrag vorübergehend entfernen, indem Sie ihn in Ihrem Ausschlusswörterbuch deaktivieren. Deaktivierte Einträge werden bei der Extrahierung ignoriert.

- ▶ Wählen Sie in Ihrem Ausschlusswörterbuch den Eintrag aus, den Sie deaktivieren möchten.
- ▶ Drücken Sie die Leertaste. Das Kontrollkästchen links neben dem Eintrag wird deaktiviert.

Anmerkung: Sie können auch links neben dem Eintrag auf das Kontrollkästchen klicken, um es zu deaktivieren.

So löschen Sie Einträge

Sie können nicht mehr benötigte Einträge aus Ihrem Ausschlusswörterbuch löschen.

- ▶ Wählen Sie in Ihrem Ausschlusswörterbuch den Eintrag aus, den Sie löschen möchten.
- ▶ Wählen Sie im Menü Bearbeiten > Löschen. Der Eintrag wird aus dem Wörterbuch entfernt.

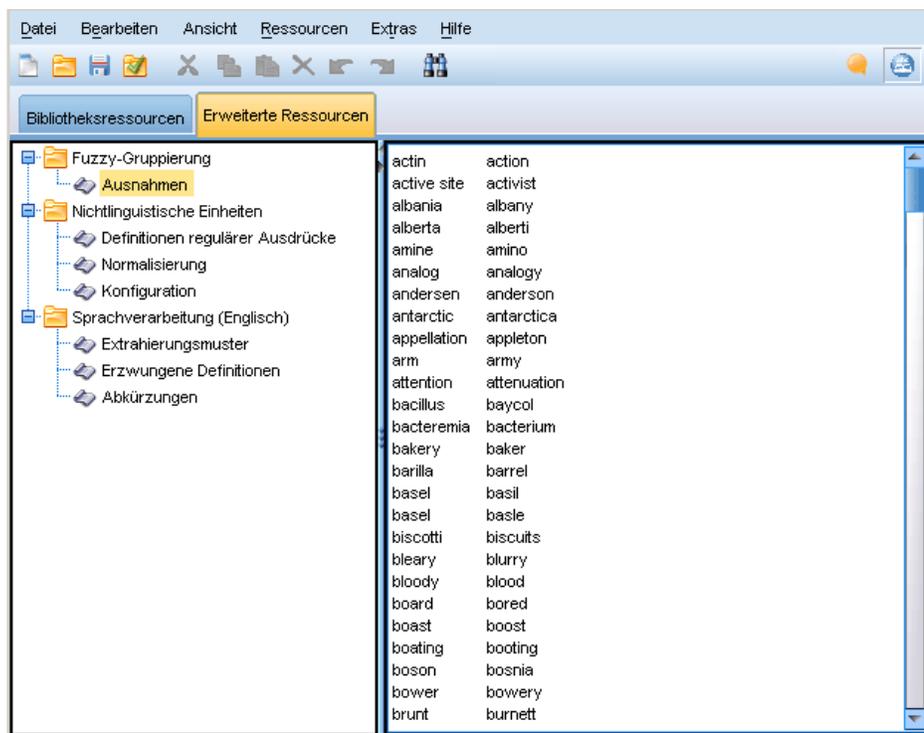
Informationen zu erweiterten Ressourcen

Ergänzend zu den Typ-, Ausschluss- und Substitutionswörterbüchern können Sie auch mit einer Vielzahl erweiterter Ressourceneinstellungen arbeiten, z. B. Einstellungen für unscharfe Gruppierung und nicht-linguistische Typdefinitionen. Sie können in der Registerkarte “Erweiterte Ressourcen” im Resource Editor mit diesen Ressourcen arbeiten. Sie können Ihre Änderungen auch standardmäßig für alle Projekte speichern oder zum ursprünglichen Inhalt zurückkehren.

Wenn Sie die Registerkarte “Erweiterte Ressourcen” aufrufen, können Sie folgende Informationen bearbeiten:

- **Unscharfe Gruppierung (Ausnahmen).** Hiermit werden Wortpaare aus dem Algorithmus für die unscharfe Gruppierung ausgeschlossen (Rechtschreibfehlerkorrektur). [Für weitere Informationen siehe Thema Unscharfe Gruppierung auf S. 245.](#)
- **Nicht linguistische Entitäten.** Hiermit wird aktiviert bzw. deaktiviert, welche linguistischen Elemente extrahiert werden und welche regulären Ausdrücke und Normalisierungsregeln bei ihrer Extrahierung angewendet werden. [Für weitere Informationen siehe Thema Nicht linguistische Elemente auf S. 246.](#)
- **Sprachbehandlung.** Darüber wird festgelegt, wie Sätze strukturiert werden (Extrahierungsmuster und erzwungene Definitionen) und wie Abkürzungen für die ausgewählte Sprache verwendet werden. [Für weitere Informationen siehe Thema Sprachbehandlung auf S. 252.](#)

Abbildung 11-1
Erweiterte Ressourcen



Anmerkung: Mit der Symbolleiste zum Suchen und Ersetzen können Sie Informationen schnell auffinden oder in einem Abschnitt identische Änderungen durchführen. [Für weitere Informationen siehe Thema Ersetzen auf S. 244.](#)

So bearbeiten Sie erweiterte Ressourcen:

- ▶ Finden Sie den Ressourcenabschnitt, den Sie bearbeiten möchten, und wählen Sie ihn aus. Der Inhalt wird im rechten Fensterbereich angezeigt.
- ▶ Über die Menübefehle oder die Schaltflächen der Symbolleiste können Sie Inhalte ausschneiden, kopieren oder einfügen.
- ▶ Bearbeiten Sie die Dateien, die Sie ändern möchten, mithilfe der Formatierungsregeln in diesem Abschnitt. Ihre Änderungen werden direkt gespeichert. Mithilfe der in der Symbolleiste angezeigten Pfeile zum Rückgängigmachen bzw. Wiederholen können Sie Ihre Änderungen rückgängig machen.

Suchen

In manchen Fällen ist es erforderlich, Informationen in einem bestimmten Abschnitt schnell aufzufinden. Mit der Suchfunktion können Sie eine spezifische Regel schnell auffinden. Für die Suche nach Informationen in einem Abschnitt können Sie die Symbolleiste "Suchen" verwenden.

Abbildung 11-2
Symbolleiste "Suchen"



So verwenden Sie die Suchfunktion:

- ▶ Suchen Sie den Ressourcenabschnitt, den Sie durchsuchen möchten, und wählen Sie ihn aus. Der Inhalt wird im rechten Bereich des Editors angezeigt.
- ▶ Wählen Sie im Menü Bearbeiten > Suchen. Oben rechts im Dialogfeld "Erweiterte Ressourcen bearbeiten" wird die Symbolleiste "Suchen" angezeigt.
- ▶ Geben Sie die Wortfolge, nach der Sie suchen möchten, in das Textfeld ein. Mit den Schaltflächen der Symbolleiste können Sie festlegen, ob zwischen Groß- und Kleinschreibung unterschieden wird, ob eine teilweise Übereinstimmung zulässig ist und in welche Richtung die Suche durchgeführt wird.
- ▶ Klicken Sie auf Suchen, um die Suche zu starten. Wenn eine Übereinstimmung gefunden wird, wird der Text im Fenster markiert.
- ▶ Klicken Sie erneut auf Suchen, um nach der nächsten Übereinstimmung zu suchen.

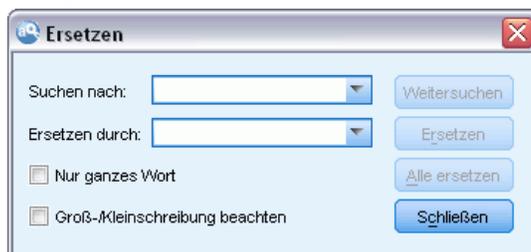
Ersetzen

Manchmal ist es erforderlich, erweiterte Ressourcen umfangreich zu aktualisieren. Mit der Funktion "Ersetzen" können Sie Ihren Inhalt einheitlich aktualisieren.

So verwenden Sie die Funktion "Ersetzen":

- ▶ Suchen Sie den Ressourcenabschnitt, in dem Sie suchen und ersetzen möchten, und wählen Sie ihn aus. Der Inhalt wird im rechten Bereich des Editors angezeigt.
- ▶ Wählen Sie im Menü Bearbeiten > Ersetzen. Das Dialogfeld "Ersetzen" wird geöffnet.

Abbildung 11-3
Dialogfeld "Ersetzen"



- ▶ Geben Sie in das Textfeld Suchen nach das Wort ein, nach dem Sie suchen möchten.
- ▶ Geben Sie in das Textfeld Ersetzen durch die Zeichenfolge ein, die Sie anstelle des gefundenen Texts verwenden möchten.

- ▶ Wählen Sie die Option Nur ganzes Wort aus, wenn Sie nur vollständige Wörter suchen und ersetzen möchten.
- ▶ Wählen Sie Groß-/Kleinschreibung beachten, wenn Sie nur Wörter suchen oder ersetzen möchten, deren Schreibweise exakt übereinstimmt.
- ▶ Klicken Sie auf Weitersuchen, um nach einer Übereinstimmung zu suchen. Wenn eine Übereinstimmung gefunden wird, wird der Text im Fenster markiert. Wenn Sie eine gefundene Übereinstimmung nicht ersetzen möchten, klicken Sie so oft erneut auf Weitersuchen, bis Sie einen Treffer erhalten, den Sie ersetzen möchten.
- ▶ Klicken Sie auf Ersetzen, um den ausgewählten Treffer zu ersetzen.
- ▶ Klicken Sie auf Ersetzen, um alle im Abschnitt gefundenen Treffer zu ersetzen. Anschließend wird eine Meldung mit der Anzahl der durchgeführten Ersetzungen angezeigt.
- ▶ Wenn Sie Ihre Ersetzungen durchgeführt haben, klicken Sie auf Schließen. Das Dialogfeld wird geschlossen.

Anmerkung: Wenn Sie beim Ersetzen einen Fehler gemacht haben, können Sie diesen rückgängig machen, indem Sie das Dialogfeld schließen und im Menü Bearbeiten > Rückgängig auswählen. Dies müssen Sie für jede Änderung, die Sie rückgängig machen möchten, wiederholen.

Unschärfe Gruppierung

Wenn Sie im Dialogfeld “Extrahierungseinstellungen” Rechtschreibung korrigieren für eine minimale Anzahl an Stammzeichen von: auswählen, haben Sie den Algorithmus für die unscharfe Gruppierung aktiviert.

Mit der unscharfen Gruppierung können Sie Wörter einfacher gruppieren, die häufig falsch oder ähnlich geschrieben werden, indem Sie vorübergehend alle Vokale (außer den ersten) und doppelt/dreifach auftretende Konsonanten aus den extrahierten Wörtern entfernen und anschließend einen Vergleich dieser Wörter durchführen. Die Funktion für die unscharfe Gruppierung wird während des Extrahierungsprozesses auf die extrahierten Ausdrücke angewendet und die Ergebnisse werden dann verglichen, um eventuelle Übereinstimmungen zu ermitteln. Wenn dies der Fall ist, werden die ursprünglichen Begriffe in der endgültigen Extrahierungsliste gruppiert. Die Gruppierung erfolgt unter dem Ausdruck, der am häufigsten in den Daten vorkommt.

Anmerkung: Wenn die zwei miteinander verglichenen Begriffe unterschiedlichen Typen zugeordnet werden (ausschließlich des Typs <Unknown>), wird die unscharfe Gruppierung auf dieses Paar nicht angewendet. Die Begriffe müssen also zum selben Typ oder zum Typ <Unknown> gehören, damit diese Methode angewendet wird.

Wenn Sie diese Funktion aktiviert haben und feststellen, dass zwei Wörter mit einer ähnlichen Schreibweise fälschlicherweise gruppiert wurden, können Sie diese Wörter aus der unscharfen Gruppierung ausschließen. Hierzu geben Sie die fälschlicherweise als übereinstimmend eingestuft Wortpaare auf der Registerkarte “Erweiterte Ressourcen” im Abschnitt “Ausnahmen” ein. [Für weitere Informationen siehe Thema Informationen zu erweiterten Ressourcen auf S. 242.](#)

Das folgende Beispiel veranschaulicht, wie eine unscharfe Gruppierung durchgeführt wird. Wenn die unscharfe Gruppierung aktiviert ist, werden die folgenden Wörter als identisch eingestuft:

```
color -> colr          mountain -> montn
colour -> colr         montana -> montn

modeling -> modlng    furniture -> furntr
modelling -> modlng   furnature -> furntr
```

In diesem Beispiel würden Sie sicherlich verhindern wollen, dass `mountain` und `montana` miteinander gruppiert werden. Hierzu können Sie dies wie folgt im Abschnitt “Ausnahmen” angeben:

```
mountain      montana
```

Wichtig: In manchen Fällen werden trotz definierter Ausnahmen für unscharfe Gruppierung zwei Wörter gepaart, weil bestimmte Synonymregeln angewendet werden. Versuchen Sie in diesem Fall, Synonyme mit einem Ausrufezeichen (!) als Platzhalter einzugeben, um zu verhindern, dass Wörter in der Ausgabe als synonym betrachtet werden. [Für weitere Informationen siehe Thema Definieren von Synonymen in Kapitel 10 auf S. 236.](#)

Formatierungsregeln für Ausnahmen bei der unscharfen Gruppierung

- Definieren Sie nur je ein Ausschlusspaar pro Zeile.
- Geben Sie einzelne Wörter oder zusammengesetzte Wörter an.
- Geben Sie die Wörter nur in Kleinbuchstaben ein. Wörter in Großbuchstaben werden ignoriert.
- Trennen Sie die Wörter in einem Paar durch ein Tabulator-Zeichen.

Nicht linguistische Elemente

Bei der Arbeit mit bestimmten Datenarten sind Datumsangaben, Sozialversicherungsnummern, Prozentsätze und andere nicht linguistische Elemente von Interesse. Diese Elemente werden explizit in der Konfigurationsdatei deklariert, in der Sie auch die Elemente aktivieren bzw. deaktivieren können. [Für weitere Informationen siehe Thema Konfiguration auf S. 250.](#) Um die Ausgabe der Extrahierungsengine zu optimieren, wird die Eingabe aus der nicht linguistischen Verarbeitung so normalisiert, dass ähnliche Elemente gemäß den vordefinierten Formaten gruppiert werden. [Für weitere Informationen siehe Thema Normalisierung auf S. 250.](#)

Anmerkung: Sie können die Extrahierung nicht linguistischer Elemente in den Extrahierungseinstellungen aktivieren und deaktivieren.

Verfügbare nicht linguistische Entitäten.

Die in der folgenden Tabelle aufgeführten nicht linguistischen Elemente können extrahiert werden. Der Typname wird in Klammern angegeben.

Adressen (<Address>)	Organisationen (<Organization>)
Aminosäuren (<Aminoacid>)	Prozente (<Percent>)

Währungen (<Currency>)	Produkte (<Product>)
Datumsangaben (<Date>)	Proteine (<Gene>)
Verspätung (<Delay>)	Telefonnummern (<PhoneNumber>)
Stellen (<Digit>)	Uhrzeiten (<Time>)
E-Mail-Adressen (<email>)	Sozialversicherungsnummer (USA) (<SocialSecurityNumber>)
HTTP/URL-Adressen (<url>)	Gewichte und Maßangaben (<Weights-Measures>)
IP-Adresse (<IP>)	

Bereinigen von Text für die Verarbeitung

Bevor die Extrahierung nicht-linguistischer Einheiten beginnt, wird der Eingabetext bereinigt. Bei diesem Schritt werden die folgenden temporären Änderungen durchgeführt, damit nichtlinguistische Einheiten identifiziert und als solche extrahiert werden können:

- Jede Folge von zwei oder mehr Leerzeichen wird durch ein einzelnes Leerzeichen ersetzt.
- Tabulatorzeichen werden durch Leerzeichen ersetzt.
- Einzelne Zeilenende-Zeichen oder Sequenzzeichen werden durch ein Leerzeichen ersetzt, während mehrere Zeilenende-Zeichen als Absatzende markiert werden. Das Zeilenende kann durch Zeilenschaltungen (Carriage Return, CR) und/oder Zeilenvorschub (Line Feed, LF) gekennzeichnet werden.
- HTML- und XML-Tags werden temporär unterdrückt und ignoriert.

Definitonen regulärer Ausdrücke

Beim Extrahieren nicht linguistischer Elemente sollen ggf. die Definitionen für reguläre Ausdrücke bearbeitet oder ergänzt werden, mit denen diese regulären Ausdrücke erkannt werden. Dies erfolgt im Abschnitt Definitionen regulärer Ausdrücke auf der Registerkarte "Erweiterte Ressourcen". [Für weitere Informationen siehe Thema Informationen zu erweiterten Ressourcen auf S. 242.](#)

Die Datei ist in mehrere Abschnitte gegliedert. Der erste Abschnitt trägt die Bezeichnung [macros]. Neben diesem Abschnitt kann jeweils ein zusätzlicher Abschnitt für jedes nicht linguistische Element vorliegen. Sie können weitere Abschnitte in diese Datei aufnehmen. Die Regeln in den einzelnen Abschnitten sind jeweils nummeriert (*regex1*, *regex2* usw.). Die Regeln müssen sequenziell von 1 bis *-n* nummeriert werden. Eine Unterbrechung der Nummerierung führt dazu, dass diese Datei überhaupt nicht verarbeitet wird.

In bestimmten Fällen ist ein Element sprachabhängig. Ein Element gilt dann als sprachabhängig, wenn es für den Sprachparameter in der Konfigurationsdatei einen Wert ungleich 0 besitzt. [Für weitere Informationen siehe Thema Konfiguration auf S. 250.](#) Ist ein Element sprachabhängig, muss die Sprache dem Abschnittsnamen vorangestellt werden, z. B. [english/PhoneNumber]. Wenn Sie dem Element `PhoneNumber` den Wert 2 für die Sprache zuweisen, enthält dieser Abschnitt bestimmte Regeln, die nur für englische Telefonnummern gelten.

Wichtig: Wenn Sie an dieser oder einer anderen Datei im Editor Änderungen vornehmen und wenn die Extrahierungseengine anschließend nicht mehr wie gewünscht arbeitet, dann verwenden Sie die in der Symbolleiste vorhandene Option Auf Original zurücksetzen, um die Datei in den Zustand zurückzusetzen, den sie bei der Installation hatte. Für diese Datei sollten Sie mit regulären

Ausdrücken vertraut sein. Falls Sie weitere Hilfe in diesem Bereich benötigen, wenden Sie sich an IBM Corp..

Sonderzeichen. [] {} () \ * + ? | ^ \$

Alle Zeichen entsprechen sich selbst, mit Ausnahme der folgenden Sonderzeichen, die für einen bestimmten Zweck in Ausdrücken benutzt werden: . [{ () \ * + ? | ^ \$ Um diese Zeichen als solche zu verwenden, muss ihnen in der Definition ein umgekehrter Schrägstrich (\) vorangestellt werden.

Beispiel: Wenn Sie versuchen, Internet-Adressen zu extrahieren, ist der Punkt sehr wichtig für die Entität, daher müssen Sie ihn folgendermaßen mit einem umgekehrten Schrägstrich versehen:

```
www\. [a-z]+ \. [a-z]+
```

Wiederholungsoperatoren und -Quantifikatoren ? + * {}

Damit flexiblere Definitionen möglich sind, können Sie mehrere Platzhalterzeichen verwenden, die für reguläre Ausdrücke Standard sind. Dabei handelt es sich um * ? +

- Ein *Sternchen* * gibt an, dass *null oder mehr* Instanzen der vorangehenden Zeichenfolge vorhanden sind. Beispiel: `ab*c` findet “*ac*”, “*abc*”, “*abbc*” usw. als Übereinstimmung.
- Ein *Pluszeichen* + gibt an, dass *eine oder mehrere* Instanzen der vorangehenden Zeichenfolge vorhanden sind. Beispiel: `ab+c` findet “*abc*”, “*abc*”, “*abbc*”, aber nicht “*ac*” als Übereinstimmung.
- Ein *Fragezeichen* ? gibt an, dass *null oder eine* Instanz der vorangehenden Zeichenfolge vorhanden ist. Beispiel: `modell?ing` findet sowohl “*modelling*” als auch “*modeling*”.
- *Beschränken von Wiederholungen mit Klammern* {} gibt die Grenzen der Wiederholung an. Zum Beispiel
 - ▶ `[0-9]{n}` entspricht einer Ziffer, die exakt *n* Mal wiederholt wird. Beispiel: `[0-9]{4}` findet “*1998*”, aber weder “*33*” noch “*19983*”.
 - ▶ `[0-9]{n}` entspricht einer Ziffer, die *n Mal oder öfter* wiederholt wird. Beispiel: `[0-9]{3,}` findet “*199*” oder “*1998*”, aber nicht “*19*”.
 - ▶ `[0-9]{n,m}` entspricht einer Ziffer, die zwischen *n und m Mal (inklusive)* wiederholt wird. Beispiel: `[0-9]{3,5}` findet “*199*”, “*1998*” oder “*19983*”, aber nicht “*19*” und nicht “*199835*”.

Optionale Leerzeichen und Bindestriche

In manchen Fällen ist es erforderlich, ein optionales Leerzeichen in eine Definition einzufügen. Wenn Sie beispielsweise Währungen extrahieren möchten wie “*uruguayan pesos*”, “*uruguayan peso*”, “*uruguay pesos*”, “*uruguay peso*”, “*pesos*” oder “*peso*”, müssen Sie sich damit auseinandersetzen, dass zwei Wörter vorhanden sein können, die durch ein Leerzeichen getrennt sind. In diesem Fall könnte die Definition als `(uruguayan |uruguay)?pesos?` geschrieben werden. Da auf *uruguayan* bzw. *uruguay* ein Leerzeichen folgt, wenn es mit *pesos/peso* verwendet wird, muss das optionale Leerzeichen innerhalb der optionalen Sequenz `(uruguayan |uruguay)` definiert werden. Wenn das Leerzeichen nicht in der optionalen Sequenz wie `(uruguayan|uruguay)? pesos?` vorhanden wäre, würde es nicht “*pesos*” oder “*peso*” finden, da das Leerzeichen erforderlich wäre.

Wenn Sie eine Reihe von Einträgen einschließlich eines Bindestrichs (-) in einer Liste suchen, muss der Bindestrich als Letztes definiert werden. Beispiel: Wenn Sie ein Komma (,) oder einen Bindestrich (-) suchen, verwenden Sie [, -], aber auf keinen Fall [-,].

Reihenfolge von Zeichenfolgen in Listen und Makros

Sie sollten stets die längste Sequenz vor einer kürzeren definieren. Andernfalls wird die längste Sequenz nie gelesen, da die Übereinstimmung schon mit der kürzeren gefunden wird. Beispiel: Wenn Sie die Zeichenfolgen “billion” oder “bill” suchen, muss “billion” vor “bill” definiert werden, also (billion|bill) und nicht (bill|billion). Dies gilt auch für Makros, da Makros Listen von Zeichenfolgen sind.

Reihenfolge von Regeln im Definitionsabschnitt

Definieren Sie nur je eine Regel pro Zeile. Die Regeln in den einzelnen Abschnitten sind jeweils nummeriert (*regexp1*, *regexp2* usw.). Die Regeln müssen sequenziell von 1 bis $-n$ nummeriert werden. Eine Unterbrechung der Nummerierung führt dazu, dass diese Datei überhaupt nicht verarbeitet wird. Um einen Eintrag zu deaktivieren, geben Sie am Anfang jeder Zeile, in der der reguläre Ausdruck definiert ist, eine Raute (#) ein. Soll ein Eintrag aktiviert werden, entfernen Sie das Nummernzeichen (#) am Anfang der zugehörigen Zeile.

In jedem Abschnitt müssen die spezifischsten Regeln vor den allgemeinsten Regeln definiert werden, um eine korrekte Verarbeitung sicherzustellen. Wenn Sie beispielsweise ein Datum in der Form “Monat Jahr” und in der Form “Monat” suchen, muss die Regel “Monat Jahr” vor der Regel “Monat” definiert werden. Die Definition sollte wie folgt aussehen:

```
#@# January 1932
regexp1=$(MONTH),? [0-9]{4}

#@# January
regexp2=$(MONTH)
```

und nicht:

```
#@# January
regexp1=$(MONTH)

#@# January 1932
regexp2=$(MONTH),? [0-9]{4}
```

Verwenden von Makros in Regeln

Wenn eine bestimmte Sequenz in mehreren Regeln vorkommt, können Sie ein Makro verwenden. Wenn Sie dann die Definition dieser Sequenz ändern müssen, müssen Sie sie nur einmal ändern und nicht in allen Regeln, die darauf verweisen. Hier ein Beispiel: Angenommen, es liegt folgendes Makro vor:

```
MONTH=( (january|february|march|april|june|july|august|september|october|
november|december) | (jan|feb|mar|apr|may|jun|jul|aug|sep|oct|nov|dec) (\.)?)
```

Bei jedem Verweis auf den Namen des Makros muss dieser von \$ () umgeben sein, z. B.:

```
regexp1=$(MONTH)
```

Alle Makros müssen im Abschnitt [macros] definiert werden.

Normalisierung

Beim Extrahieren nicht linguistischer Elemente werden die aufgefundenen Elemente so normalisiert, dass ähnliche Elemente gemäß den vordefinierten Formaten gruppiert werden. Währungssymbole und die zugehörigen Währungskürzel werden beispielsweise als gleich behandelt. Die Normalisierungseinträge werden im Abschnitt Normalisierung auf der Registerkarte “Erweiterte Ressourcen” gespeichert. [Für weitere Informationen siehe Thema Informationen zu erweiterten Ressourcen auf S. 242.](#) Die Datei ist in mehrere Abschnitte gegliedert.

Wichtig: Diese Datei richtet sich nur an fortgeschrittene Benutzer. Änderungen an dieser Datei fallen eher selten an. Falls Sie weitere Hilfe in diesem Bereich benötigen, wenden Sie sich an IBM Corp..

Formatierungsregeln für die Normalisierung

- Fügen Sie nur je einen Normalisierungseintrag pro Zeile hinzu.
- Beachten Sie genauestens die Abschnitte in dieser Datei. Es ist nicht möglich, neue Abschnitte hinzuzufügen.
- Um einen Eintrag zu deaktivieren, geben Sie am Anfang der entsprechenden Zeile eine Raute (#) ein. Soll ein Eintrag aktiviert werden, entfernen Sie das Nummernzeichen (#) am Anfang der zugehörigen Zeile.

Konfiguration

Sie können die nicht linguistischen Elementtypen, die Sie extrahieren möchten, in der Konfigurationsdatei der nicht linguistischen Elemente aktivieren bzw. deaktivieren. Wenn Sie die nicht benötigten Elemente deaktivieren, wird die Verarbeitung beschleunigt. Dies erfolgt im Abschnitt Konfiguration auf der Registerkarte “Erweiterte Ressourcen”. [Für weitere Informationen siehe Thema Informationen zu erweiterten Ressourcen auf S. 242.](#) Wenn die Extrahierung nicht linguistischer Elemente aktiviert ist, liest die Extrahierungsengine diese Konfigurationsdatei beim Extrahieren, um zu ermitteln, welche Typen nicht linguistischer Elemente extrahiert werden sollen.

Für diese Datei gilt die folgende Syntax:

```
#Name<TAB>Sprache<TAB>Code
```

Tabelle 11-1
Syntax für Konfigurationsdatei

Spaltenüberschrift	Beschreibung
#Name	Wortlaut, mit dem die anderen beiden Dateien, die für die Extrahierung nicht linguistischer Elemente erforderlich sind, auf diese nicht linguistischen Elemente verweisen. Bei den hier angegebenen Namen wird zwischen Groß- und Kleinschreibung unterschieden.
Sprache	Die Sprache der Datensätze. Im Idealfall sollten Sie die jeweilige Sprache angeben; alternativ steht die Option Alle zur Verfügung. Mögliche Optionen: 0 = Alle, was verwendet wird, wenn ein regexp nicht sprachspezifisch ist und in mehreren Vorlagen mit verschiedenen Sprachen benutzt werden kann, z. B. IP/URL/E-Mail-Adressen; 1 = Französisch; 2 = Englisch; 4 = Deutsch; 5 = Spanisch; 6 = Niederländisch; 8 = Portugiesisch; 10 = Italienisch.

Spaltenüberschrift	Beschreibung
Code	Code für die Wortklasse. Die meisten Elemente erhalten den Wert "s", bis auf wenige Ausnahmefälle. Mögliche Werte: s = Stoppwort; a = Adjektiv; n = Nomen. Wenn diese Option aktiviert ist, werden zuerst die nicht linguistischen Elemente extrahiert und dann wird die Rolle dieser Elemente mithilfe der Extraktionsmuster in einem größeren Zusammenhang ermittelt. Prozentsätze erhalten beispielsweise den Wert "a" Angenommen, 30 % wird als nicht linguistisches Element extrahiert. Dieses Element wird als Adjektiv erkannt. Wenn der Text dann die Wortgruppe "30% salary increase" enthält, erfüllt das nicht linguistische Element "30%" das Wortklasse-Muster "an" (Adjektiv-Nomen-Nomen).

Reihenfolge bei der Definition von Elementen

Die Reihenfolge, in der Sie die Elemente in dieser Datei deklarieren, ist von Bedeutung und wirkt sich auf die Extrahierung aus. Die Einträge werden in der angegebenen Reihenfolge angewendet. Wenn Sie die Reihenfolge ändern, ändern sich auch die Ergebnisse. Die spezifischsten nicht linguistischen Elemente müssen vor den allgemeineren definiert werden.

Beispielsweise wird das nicht linguistische Element "Aminosäure" definiert durch:

```
regexp1=$(AA)-?$(NUM)
```

Dabei entspricht \$(AA) den Werten

"(ala|arg|asn|asp|cys|gln|glu|gly|his|ile|leu|lys|met|phe|pro|ser)",
bei denen es sich um bestimmte Folgen aus drei Buchstaben handelt, die bestimmten Aminosäuren entsprechen.

Andererseits wird das nicht linguistische Element "Gen" allgemeiner und wird definiert durch:

```
regexp1=p[0-9]{2,3}
regexp2=[a-z]{2,4}-?[0-9]{1,3}-?[r]
regexp3=[a-z]{2,4}-?[0-9]{1,3}-?p?
```

Wenn im Abschnitt "Konfiguration" das Element "Gen" vor "Aminosäure" definiert wird, wird für "Aminosäure" nie eine Übereinstimmung gefunden, da regexp3 von "Gen" immer die erste Übereinstimmung erzielt.

Formatierungsregeln für die Konfiguration

- Trennen Sie die Einträge in einer Spalte durch ein Tabulator-Zeichen.
- Löschen Sie keine Zeilen.
- Halten Sie sich an die Syntax, die in der vorangehenden Tabelle angegeben ist.
- Um einen Eintrag zu deaktivieren, geben Sie am Anfang der entsprechenden Zeile eine Raute (#) ein. Soll ein Element aktiviert werden, entfernen Sie die Raute (#) am Anfang der zugehörigen Zeile.

Sprachbehandlung

Alle modernen Sprachen besitzen eigene Systeme, um Ideen auszudrücken, Sätze zu strukturieren und Abkürzungen zu verwenden. Im Abschnitt Sprachbehandlung können Sie Extraktionsmuster bearbeiten, Definitionen für diese Muster erzwingen und Abkürzungen für von Ihnen in der Dropdown-Liste “Sprache” ausgewählte Sprachen deklarieren.

- Extraktionsmuster
- Erwungene Definitionen
- Abkürzungen

Extraktionsmuster

Beim Extrahieren von Daten aus Ihren Datensätzen wendet die Extrahierungseingine eine Reihe von Wortklasse-Mustern (Part of Speech) auf einen “Stapel” von Wörtern im Text an, um so Kandidaten (Wörter und zusammengesetzte Ausdrücke) für die Extrahierung zu erkennen. Sie können die Extraktionsmuster hinzufügen und bearbeiten.

Die Wortarten (Part of Speech) bestehen aus grammatischen Elementen, z. B. Nomen, Adjektive, Partizip Präteritum, Determinatoren, Präpositionen, Koordinatoren, Vornamen, Initialen und Partikel. Eine Reihe dieser Elemente bildet ein Wortklasse-Extraktionsmuster. In IBM Corp.-Text-Mining-Produkten ist jede Wortart (Part of Speech) mit einem einzelnen Buchstaben gekennzeichnet, so dass Sie die Muster leichter definieren können. Ein Adjektiv ist beispielsweise am Kleinbuchstaben *a* erkennbar. Die unterstützten Codes werden standardmäßig am Anfang jedes Abschnitts für Standardextraktionsmuster aufgeführt, zusammen mit einer Reihe von Mustern und Beispielen für die Muster, mit denen die verwendeten Codes erläutert werden.

Formatierungsregeln für Extraktionsmuster

- Ein Muster pro Zeile.
- Um ein Muster zu deaktivieren, geben Sie am Anfang der entsprechenden Zeile eine Raute (#) ein.

Die Reihenfolge, in der Sie die Extraktionsmuster aufführen, ist von großer Bedeutung, weil eine gegebene Wortfolge nur einmal in der Extraktionsengine gelesen und dann dem ersten Extraktionsmuster zugewiesen wird, für das die Engine eine Übereinstimmung erkennt.

Erzwungene Definitionen

Beim Extrahieren von Daten aus Ihren Datensätzen scannt die Extrahierungseingine den Text und erkennt dabei die Wortart (Part of Speech) für jedes gefundene Wort. In einigen Fällen kann ein Wort verschiedene POS-Rollen annehmen, je nach Kontext. Im Bereich Erzwungene Definition der Registerkarte “Erweiterte Ressourcen” können Sie eine bestimmte Rolle für ein Wort erzwingen oder dieses Wort komplett aus der Verarbeitung der Wortklasse ausschließen. [Für weitere Informationen siehe Thema Informationen zu erweiterten Ressourcen auf S. 242.](#)

Um eine Wortklassen-Rolle für ein bestimmtes Wort zu erzwingen, müssen Sie in diesem Bereich eine Zeile mit der folgenden Syntax hinzufügen:

Fachausdruck:Code

Tabelle 11-2
Beschreibung der Syntax

Aufnahme (Cox Regression)	Beschreibung
Fachausdruck	Der Name eines Fachausdrucks
Code	Ein einstelliger Code für die Wortklassen-Regel. Sie können bis zu sechs verschiedene Wortklassen-Codes pro Uniterm angeben. Mit dem Code <i>s</i> (Kleinbuchstabe "s") können Sie außerdem die Extrahierung eines bestimmten Wortes in zusammengesetzte Worte bzw. Wortfolgen unterbinden, z. B. <i>additional:s</i> .

Formatierungsregeln für erzwungene Definitionen

- Eine Zeile pro Wort.
- Doppelpunkte sind in Fachausdrücken nicht zulässig.
- Mit dem Kleinbuchstaben *s* als Code für die Wortart geben Sie an, dass ein Wort überhaupt nicht extrahiert werden soll.
- Geben Sie bis zu sechs Wortklassen-Codes pro Zeile ein. Unterstützte Wortklassen-Codes werden im Abschnitt "Extraktionsmuster" angezeigt. [Für weitere Informationen siehe Thema Extraktionsmuster auf S. 252.](#)
- Das Sternchen (*) am Ende einer Zeichenfolge dient als Platzhalter für teilweise Übereinstimmungen. Wenn Sie beispielsweise *add* : s* eingeben, werden Wörter wie *add*, *additional*, *additionally*, *addendum* und *additive* weder als Fachausdruck noch als Bestandteil eines zusammengesetzten Worts extrahiert. Ist eine Wortübereinstimmung jedoch explizit als Fachausdruck in einem kompilierten Wörterbuch oder in den erzwungenen Definitionen deklariert, wird diese dennoch extrahiert. Wenn Sie beispielsweise sowohl *add* : s* als auch *addendum : n* eingeben, wird *addendum* auf jeden Fall aus dem Text extrahiert.

Abkürzungen

Im Allgemeinen behandelt die Extrahierungseingine bei der Verarbeitung von Text jeden Punkt als ein Zeichen dafür, dass der Satz beendet ist. In der Regel ist das auch richtig, doch gilt diese Behandlung von Punktzeichen nicht, wenn Abkürzungen im Text enthalten sind.

Wenn Sie Ausdrücke aus Ihrem Text extrahieren und feststellen, dass bestimmte Abkürzungen falsch behandelt wurden, sollten Sie die jeweilige Abkürzung in diesem Bereich explizit deklarieren.

Anmerkung: Wenn die Abkürzung bereits als Synonymdefinition erscheint oder im Typ-Wörterbuch als Ausdruck definiert ist, ist es nicht erforderlich, die Abkürzung hier einzutragen.

Formatierungsregeln für Abkürzungen

- Definieren Sie nur je eine Abkürzung pro Zeile.

Ausnahmen bei japanischem Text

Für Text in japanischer Sprache erfolgen Verarbeitung und Mining zwar ähnlich wie für andere Sprachen in IBM® SPSS® Text Analytics for Surveys, jedoch gibt es eine Reihe von Abweichungen. Die geringfügigeren Änderungen werden in dieser Dokumentation direkt mit den Anleitungen für die anderen Sprachen beschrieben. Jedoch werden einige der größeren Unterschiede in diesem Anhang erläutert.

Extrahieren und Kategorisieren von japanischem Text

Der Mining-Vorgang für Text in japanischer Sprache gleicht dem Vorgang für andere unterstützte Sprachen. [Für weitere Informationen siehe Thema Informationen zum Text-Mining in Kapitel 1 auf S. 3.](#) Jedoch gelten für die japanische Sprache die im Folgenden erläuterten Unterschiede.

Wie Extrahierung funktioniert

Während der Extrahierung von Schlüsselkonzepten und -begriffen aus Ihren Antworten wird bei IBM® SPSS® Text Analytics for Surveys die linguistisch basierte Textanalyse angewendet. Diese Methode bietet die Geschwindigkeit und Kosteneffektivität von statistisch basierten Systemen. Sie ermöglicht jedoch ein weitaus höheres Maß an Genauigkeit, während ein deutlich geringeres Maß an Eingriffen seitens des Benutzers erforderlich ist. Linguistisch basierte Textanalyse baut auf einem Forschungsgebiet namens Verarbeitung natürlicher Sprache auf, das auch als Computerlinguistik bekannt ist.

Für Text in japanischer Sprache lässt sich der Unterschied zwischen statistikbasierten und linguistikbasierten Methoden beim Extrahierungsvorgang am Wort 沈む als ein Beispiel illustrieren. Mit diesem Wort können wir Ausdrücke wie 日が沈む finden, übersetzt als *Sonne geht unter*, oder 気分が沈む, übersetzt als *sich traurig fühlen*. Wenn Sie nur statistische Techniken einsetzen, werden 日 (übersetzt als *Sonne*), 気分 (übersetzt als *fühlen*) und 沈む (übersetzt als *traurig*) separat extrahiert. Bei Einsatz der Stimmungsanalyse, die linguistische Verfahren verwendet, werden jedoch nicht nur 日, 気分 und 沈む, sondern auch 気分が沈む (übersetzt als *sich traurig fühlen*) extrahiert und dem Typ <悪い - 悲しみ全般> zugewiesen. Die Verwendung von linguistischen Verfahren durch die Stimmungsanalyse ermöglicht die Extrahierung von bedeutungsvolleren Ausdrücken. Die Analyse und Erfassung von Stimmungen beseitigt die Mehrdeutigkeit von Texten, weshalb das linguistisch basierte Text-Mining definitionsgemäß den verlässlicheren Ansatz darstellt.

Wenn Sie verstehen, wie der Extrahierungsprozess funktioniert, fällt es Ihnen leichter, bei der Feinabstimmung Ihrer linguistischen Ressourcen (Bibliotheken, Typen, Synonyme und anderer) zentrale Entscheidungen zu treffen. Der Extrahierungsprozess umfasst folgende Schritte:

- Konvertieren von Quelldaten in ein Standardformat
- Ermittlung von Kandidaten
- Ermittlung von Äquivalenzklassen und Integration von Synonymen

- Zuweisung eines Typs
- Indexerstellung und, falls gewünscht, Musterabgleich mit einem Sekundäranalysator

Schritt 1. Konvertieren von Quelldaten in ein Standardformat

Im ersten Schritt werden die importierten Daten in ein einheitliches Format konvertiert, das für weitergehende Analysen genutzt werden kann. Diese Konvertierung erfolgt intern. Ihre Ausgangsdaten werden dabei nicht geändert.

Schritt 2. Ermittlung von Kandidaten

Es ist wichtig zu verstehen, welche Rolle die linguistischen Ressourcen während der linguistischen Extrahierung bei der Ermittlung von Kandidaten spielen. Linguistische Ressourcen kommen jedes Mal zum Einsatz, wenn ein Extrahierungsvorgang ausgeführt wird. Sie liegen in Form von Vorlagen, Bibliotheken und zusammengestellten Ressourcen vor. Bibliotheken bestehen aus Wortlisten, Beziehungen und weiteren Informationen, die eingesetzt werden, um die Extrahierung abzustimmen oder zu spezifizieren. Die zusammengestellten Ressourcen können nicht angezeigt oder bearbeitet werden. Die übrigen Ressourcen können jedoch im Resource Editor bearbeitet werden.

Zusammengestellte Ressourcen sind interne Kernkomponenten der Extraktor-Engine in SPSS Text Analytics for Surveys. Diese Ressourcen umfassen ein allgemeines Wörterbuch, in dem eine Liste von Grundformen mit einem Code für die Wortart (Part of Speech) enthalten ist (Nomen, Verb, Adjektiv usw.). Die Ressourcen beinhalten auch belegte integrierte Typen, die verwendet werden, um den folgenden Typen eine Vielzahl von extrahierten Termen zuzuweisen: <地名>, <組織> oder <人名>. Für weitere Informationen siehe [Thema Verfügbare Typen für japanischen Text auf S. 265](#).

Zusätzlich zu diesen zusammengestellten Ressourcen sind auch mehrere Bibliotheken im Lieferumfang enthalten. Diese können verwendet werden, um die Typen und Konzeptdefinitionen der zusammengestellten Ressourcen zu ergänzen und Synonyme zu liefern. Diese Bibliotheken—sowie sämtliche benutzerdefinierte Bibliotheken, die Sie erstellen—bestehen aus mehreren Wörterbüchern. Diese umfassen Typ-Wörterbücher, Synonymwörterbücher sowie Ausschlusswörterbücher. Für weitere Informationen siehe [Thema Bearbeitungsressourcen für japanischen Text auf S. 260](#).

Sobald die Daten importiert und konvertiert wurden, beginnt die Extrahierungseingine, Kandidaten für die Extrahierung zu identifizieren. Kandidaten sind Wörter oder Wortgruppen, die verwendet werden, um Konzepte im Text zu ermitteln. Bei der Verarbeitung des Texts werden einzelne Wörter (**Uniterms**) und zusammengesetzte Wörter (**Multiterms**) über Extraktoren auf der Grundlage von Wortklasse-Mustern (Part of Speech) ermittelt. Der Multiterm 青森りんご, der dem Wortklasse-Muster <地名> + <名詞> entspricht, besteht beispielsweise aus zwei Komponenten. Anschließend werden mithilfe der Stimmungstextlinkanalyse Kandidaten für Stimmungstichwörter identifiziert.

Angenommen, Sie verfügen über den folgenden japanischen Text: 写真が新鮮で良かった. In diesem Fall würde die Extrahierungseingine den Stimmungstyp 良い - 褒め・賞賛 zuweisen, nachdem (品物) + が + 良い mithilfe einer der Stimmungstextlinkregeln abgeglichen wurde.

Anmerkung: Die Terme aus dem oben genannten zusammengestellten allgemeinen Wörterbuch stellen eine Liste aller Wörter dar, die als Uniterms wahrscheinlich uninteressant sind oder sprachliche Mehrdeutigkeiten aufweisen. Diese Wörter werden von der Extrahierung

ausgeschlossen, wenn die Uniterms ermittelt werden. Sie werden jedoch erneut ausgewertet, wenn Wortarten bestimmt oder längere zusammengesetzte Wörter (Multiterms) als Kandidaten geprüft werden.

Schritt 3. Ermittlung von Äquivalenzklassen und Integration von Synonymen

Im Anschluss an die Ermittlung von Uniterms und Multiterms, die als Kandidaten infrage kommen, werden über ein Normalisierungswörterbuch der Software Äquivalenzklassen ermittelt. Bei einer Äquivalenzklasse handelt es sich um eine Grundform eines Ausdrucks oder einer einzelnen Form von zwei Varianten desselben Ausdrucks. Ausdrücke werden Äquivalenzklassen zugeordnet, damit beispielsweise die Begriffe *Nebenwirkung* und 副作用 nicht als unterschiedliche Konzepte betrachtet werden. Um festzustellen, welches Konzept für die betreffende Äquivalenzklasse als Hauptterm verwendet wird, *Nebenwirkung* oder 副作用, werden die folgenden Regeln in der aufgeführten Reihenfolge durch die Extrahierungsengine angewendet:

- Die vom Benutzer festgelegte Form in einer Bibliothek.
- Die häufigste Form, wie von vorkompilierten Ressourcen definiert.

Schritt 4. Zuweisen eines Typs

Anschließend werden den extrahierten Konzepten Typen zugewiesen. Bei einem Typ handelt es sich um Konzepte, die nach semantischen Gesichtspunkten gruppiert werden. Für diesen Schritt werden sowohl zusammengestellte Ressourcen als auch die Bibliotheken verwendet. Zu den Typen gehören beispielsweise übergeordnete Konzepte, positive und negative Wörter, Vornamen, Orte, Organisationen und anderes. [Für weitere Informationen siehe Thema Typ-Wörterbücher in Kapitel 10 auf S. 223.](#)

Japanische Textressourcen verfügen über ein anderes Set an Typen. [Für weitere Informationen siehe Thema Verfügbare Typen für japanischen Text auf S. 265.](#)

Schritt 5. Indexerstellung und Musterabgleich mit Ereignisextrahierung

Die gesamte Reihe der Datensätze wird indiziert. Dazu wird ein Zeiger zwischen einer Textstelle und dem bezeichnenden Term für jede Äquivalenzklasse erstellt. Das setzt voraus, dass sämtliche gebeugten Formen, die als Kandidaten für ein Konzept vorkommen, als Grundform für den Kandidaten indiziert werden. Die globale Häufigkeit wird für jede Grundform berechnet.

Mit SPSS Text Analytics for Surveys können nicht nur Typen und Konzepte sondern auch Beziehungen ermittelt werden, die zwischen diesen bestehen. Mit diesem Produkt stehen mehrere Algorithmen und Bibliotheken zur Verfügung, über die Beziehungsmuster der Textlinkanalyse zwischen Typen und Konzepten extrahiert werden können. Sie sind besonders nützlich beim Versuch, bestimmte Meinungen zu entdecken (z. B. Reaktionen auf ein Produkt).

Wie sekundäre Extrahierung funktioniert

Beim Ausführen einer Extrahierung aus japanischem Text erhalten Sie automatisch Konzepte aus den grundlegenden Stichwörtern und acht Basistypen: 人名, 地名, 組織名, 名詞, 形容詞, 動詞, 形容動詞 und その他. Um jedoch die Standardressourcen für japanischen Text optimal zu nutzen, müssen Sie eine der folgenden sekundären Analysen wählen: Stimmung oder Abhängigkeit.

Durch die Wahl eines Sekundär-Analysators können Sie Textlinkanalysemuster extrahieren und Beziehungen zwischen den Begriffen im Text aufdecken.

Sekundäre Analyse. Beim Start einer Extrahierung werden anhand des Standardsatzes an Typen grundlegende Stichwörter extrahiert. [Für weitere Informationen siehe Thema Verfügbare Typen für japanischen Text auf S. 265.](#) Wenn Sie jedoch eine sekundäre Analyse wählen, erhalten Sie mehr und vielfältigere Konzepte, da der Extraktor nun Partikel und Hilfsverben als Teil des Konzepts berücksichtigt. Nehmen Sie beispielsweise an, der Satz 肩の荷が下りた wurde in “*Mir wurde eine große Last von den Schultern genommen*”. In diesem Beispiel kann die grundlegende Stichwortextrahierung jedes Konzept wie folgt separat extrahieren: 肩 (*Schultern*), 荷 (*Last*), 下りる (*wurde genommen*), jedoch wird die Beziehung zwischen diesen Wörtern nicht extrahiert. Wenn Sie jedoch die Stimmungsanalyse anwenden, können Sie vielfältigere Konzepte hinsichtlich eines Stimmungstyps extrahieren, z. B. das Konzept =肩の荷が下りた, das als “*eine große Last von den Schultern nehmen*”, was dem Typ <良い-安心> zugeordnet ist. Im Fall der Stimmungsanalyse wird auch eine große Anzahl an zusätzlichen Typen eingeschlossen. Des Weiteren ermöglicht Ihnen die Wahl einer sekundären Analyse, auch Textlinkanalysis-Ergebnisse zu generieren.

Anmerkung: Wenn ein Sekundär-Analysator genannt wird, dauert der Extrahierungsprozess länger. [Für weitere Informationen siehe Thema Wie sekundäre Extrahierung funktioniert auf S. 257.](#)

- **Abhängigkeitsanalyse.** Die Wahl dieser Option führt zu erweiterten Partikeln für die Extrahierungskonzepte aus der grundlegenden Typ- und Stichwortextrahierung. Sie können auch vielfältigere Musterergebnisse aus einer Abhängigkeits-Textlinkanalyse (TLA) gewinnen.
- **Stimmungsanalyse.** Bei dieser Analyse werden zusätzliche Konzepte und – wann immer möglich – TLA-Musterergebnisse extrahiert. Neben den grundlegenden Typen erhalten Sie auch den Vorteil von über 80 Stimmungstypen, z. B. 嬉しい, 吉報, 幸運, 安心, 幸福 usw. Mithilfe dieser Typen werden Konzepte und Muster im Text durch den Ausdruck von Emotionen, Stimmungen und Meinungen aufgedeckt. Es gibt drei Optionen, die den Fokus für die Stimmungsanalyse festlegen: Alle Stimmungen, Nur repräsentative Stimmung und Nur Schlussfolgerungen.

Optionen für Stimmungsanalyse

Beim Arbeiten mit japanischem Text können Sie nach Wunsch mithilfe der Stimmungsanalyse zusätzliche Konzepte und Typen extrahieren. Diese Analyse umfasst mehr als 80 zusätzliche Typen, mit denen Sie Meinungen, Gefühle und Emotionen aus Ihren Textdaten extrahieren können. Zusätzlich müssen Sie bei der Wahl der Stimmungsanalyse als sekundäre Analyse eine der folgenden Optionen wählen, die der Extrahierungsengine mitteilen, welche Stimmungen extrahiert werden sollen:

- **Alle Stimmungen**
- **Nur repräsentative Stimmung**
- **Nur Schlussfolgerungen**

Bei der Extrahierung beginnt die Stimmungsanalyse durch Zerlegen eines Datensatzes in Klauseln, von denen jede ein Prädikat enthält. Zum Beispiel wird der Text “4月になったがまだ寒い。”, der mit “*Es ist April, aber es ist immer noch kalt.*” übersetzt wurde, vom Analysator als 2 Klauseln interpretiert, obwohl er nur ein Satzendzeichen (。) enthält. Jede Klausel wird dann von der Extrahierungsengine untersucht, um festzustellen, ob sie Ihrer gewählten Option entspricht.

Untersuchen wir anhand des folgenden Beispieltextes die drei Optionen: “案内してくれた仲居さんは無愛想だったが、部屋は広くて申し分なかった。夕食も満足。”. Dieser Text wird übersetzt mit: “Eine Bedienung war nicht freundlich, aber das Zimmer war groß und zufriedenstellend. Auch mit dem Abendessen war ich zufrieden”. Bei der Extrahierung wird der Originaltext in die folgenden Klauseln zerlegt:

- 案内してくれた仲居さんは無愛想だったが、, übersetzt als “Eine Bedienung war nicht freundlich, aber”
- 部屋は広くて申し分なかった。 , übersetzt als “Das Zimmer war groß und zufriedenstellend”
- 夕食も満足。 , übersetzt als “Auch mit dem Abendessen war ich zufrieden.”

Alle Stimmungen

Diese Option extrahiert alle Stimmungen, Meinungen und Emotionen, die den Ressourcen und Stimmungstextlinkregeln entsprechen. In unserem Beispiel würden die folgenden Konzepte aus dem Beispieltext extrahiert.

Tabelle A-1

Mögliche Ausgabe für das Beispiel mit der Option “Alle Stimmungen”

Konzept	Typ
仲居さんは無愛想だった	<悪い - 対応が不親切>
部屋は広くて	<良い - 満足>
申し分なかった	<良い - 満足>
満足	<良い - 満足>

Anmerkung: In der vorangehenden Tabelle zeigen die zweite und dritte Zeile, wie der Extraktor zwei Konzepte aus derselben Klausel beziehen kann.

Nur repräsentative Stimmung

Diese Option extrahiert nur die repräsentativeren Meinungen oder Emotionen, die in jeder Klausel ausgedrückt werden. Wenn der Text mehrere Meinungen oder Emotionen enthält, wird ein Algorithmus angewendet. Dieser Algorithmus versucht, die Wichtigkeit der ermittelten Stimmungen und die Position der Wörter in einer Klausel zu bestimmen. In einigen Fällen, in denen zwei Stimmungswörter mit derselben Wichtigkeit gefunden werden, wird das Stimmungswort an der letzten Position in der Klausel extrahiert.

部屋は広くて (übersetzt als *Das Zimmer war groß*) wird nicht aus dem Text extrahiert, da der zweite Begriff, *申し分なかった*, gegenüber *部屋は広くて* in dieser Klausel als wichtiger erachtet wird, wenn der interne Algorithmus und die Wortposition angewendet werden.

Tabelle A-2

Mögliche Ausgabe für den Text mit der Option “Nur repräsentative Stimmung”

Konzept	Typ
仲居さんは無愛想だった	<悪い - 対応が不親切>
申し分なかった	<満足>
満足	<満足>

Nur Schlussfolgerungen

Diese Option zwingt den Extraktor, ein Stimmungsstichwort zu identifizieren und zu extrahieren, das die Schlussfolgerung des gesamten Datensatzes repräsentiert. Nicht sämtlicher Text verfügt über eine Schlussfolgerung, in einigen Fällen wird also für einen bestimmten Text mit dieser Option nichts extrahiert. Außerdem ist es für die Analyse bei einem längeren Datensatz zunehmend schwieriger, die Hauptschlussfolgerung zu identifizieren. Obwohl dies selten auftritt, können manchmal auch mehrere Schlussfolgerungen extrahiert werden.

満足 (übersetzt als *zufrieden*) wird als die wichtigste Schlussfolgerung der im Text ausgedrückten Gefühle angesehen.

Tabelle A-3

Mögliche Ausgabe für den Text mit der Option "Nur Schlussfolgerungen"

Konzept	Typ
満足	<満足>

Wie die Kategorisierung funktioniert

Sie haben die Wahl zwischen verschiedenen Methoden, um Kategorien zu erstellen. Da jeder Datensatz seine besonderen Eigenheiten aufweist, kann die Zahl der Methoden und die Reihenfolge, in der sie angewendet werden, gegebenenfalls variieren. Da sich Ihre Interpretation der Ergebnisse möglicherweise von der einer anderen Person unterscheidet, müssen Sie gegebenenfalls ein wenig mit den unterschiedlichen Methoden experimentieren, um zu erkennen, mit welcher Sie die besten Ergebnisse für Ihre Textdaten erzielen.

In diesem Handbuch bezieht sich **Kategorieerstellung** auf die Erstellung von Kategoriedefinitionen und Klassifikation über eine oder mehrere integrierte Methoden und **Kategorisierung** auf das Scoring oder Labeling, einen Prozess, bei dem den Kategoriedefinitionen für jeden Datensatz unverwechselbare IDs (Name/ID/Wert) zugewiesen werden.

Während der Kategorieerstellung werden die extrahierten Konzepte und Typen als Bausteine für Ihre Kategorien verwendet. Bei der Erstellung von Kategorien werden den Kategorien automatisch die Datensätze zugewiesen, die Text enthalten, der einem Element der jeweiligen Kategoriedefinition entspricht.

IBM® SPSS® Text Analytics for Surveys bietet Ihnen mehrere automatisierte Kategorieerstellungsmethoden, mit denen Sie Ihre Datensätze schnell kategorisieren können. Die einzelnen Verfahren sind für bestimmte Datentypen und Situationen jeweils sehr gut geeignet, doch ist es oftmals nützlich, bei einer Analyse mehrere Verfahren miteinander zu verbinden, um die Datensätze vollständig zu erfassen. Sie können ein Konzept in mehreren Kategorien erkennen oder redundante Kategorien vorfinden.

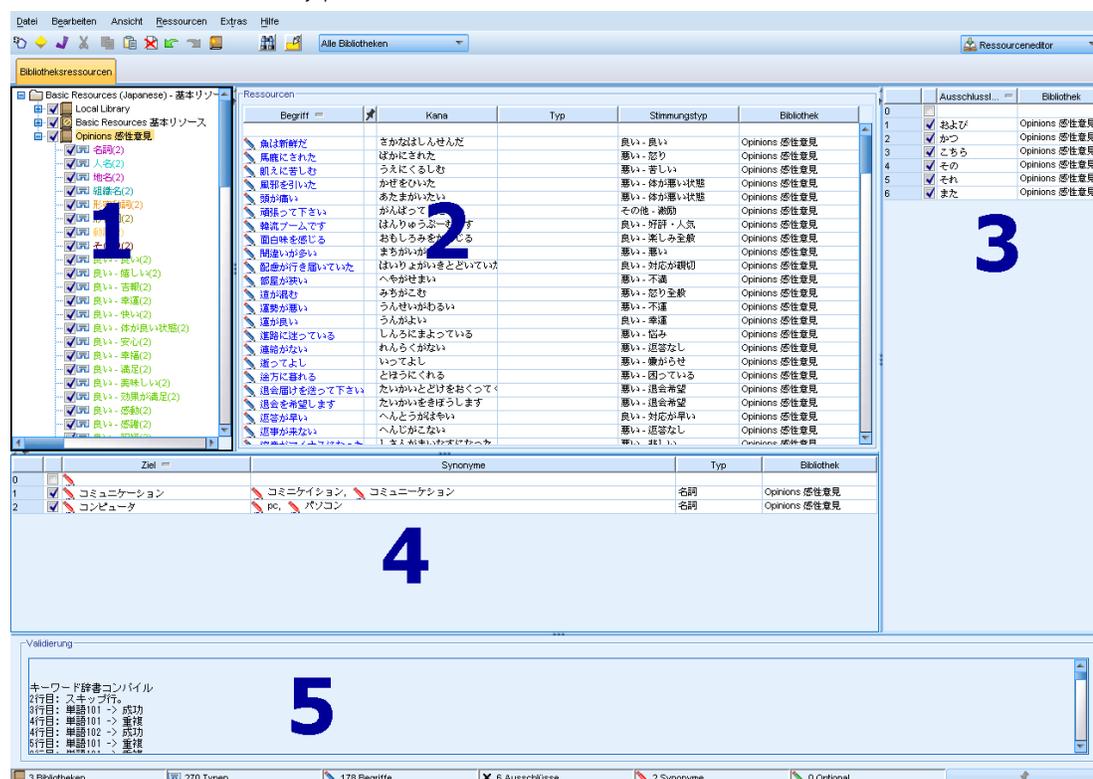
Bearbeitungsressourcen für japanischen Text

Ab IBM® SPSS® Text Analytics for Surveys Version 4 stehen eine neue Vorlage und ein neues Textanalysepaket (TAP) für Texte in japanischer Sprache zur Verfügung. Sie können Änderungen an den Ressourcen vornehmen, indem Sie Begriffe hinzufügen und bearbeiten und diese dann an Ihre Daten anpassen. Das Text Analysis Package enthält auch ein

Kategorie-Set, das aus Kategorien besteht, die positive Stimmungen, negative Stimmungen und kontextbezogene/allgemeine Stimmungen darstellen.

Sie können Ihre Ressourcen im Resource Editor bearbeiten. Editoren funktionieren für alle Textsprachen in ähnlicher Weise. Es gibt jedoch einige wichtige Unterschiede für japanischsprachige Texte, die im Folgenden erläutert werden.

Abbildung A-1
Resource Editor-Ansicht für japanischen Text



Die folgenden Punkte umreißen einige der Hauptunterschiede bei der Arbeit mit Ressourcen für japanischen Text. Eine allgemeine Beschreibung der vier Hauptbereiche auf der Registerkarte “Bibliotheksressourcen” finden Sie unter [Die Editoroberfläche](#) auf S. 197.

1. Fensterbereich “Bibliothek”. Dieser Bereich befindet sich links oben und funktioniert im Wesentlichen wie für andere Sprachen. Jedoch gibt es wenige Abweichungen, z. B. können keine neuen Typen erstellt oder bestehende Typen umbenannt werden. [Für weitere Informationen siehe Thema Mit Bibliotheken arbeiten in Kapitel 9 auf S. 209.](#)

2. Fensterbereich “Fachausdruck” für Typen-Wörterbücher. Dieser Bereich befindet sich rechts neben dem Bibliotheksbaum und unterscheidet sich für japanischen Text deutlich. Neben dem Ausdrucksnamen können Sie auch den Kana-Namen hinzufügen sowie einen oder zwei Typen wählen, mit denen Sie den Fachausdruck assoziieren können. Jedoch können Sie keine flektierten Formen von Fachausdrücken generieren oder Abgleichsoptionen für japanische Fachausdrücke zuordnen, wie dies für andere Sprachen möglich ist. [Für weitere Informationen siehe Thema Japanischer Bibliotheksbaum, Typen und Fachausdrucksbereich auf S. 262.](#)

3. Fensterbereich "Austausch-/Synonymwörterbuch". In japanischen Textressourcen finden Sie eine Registerkarte "Synonym", in der Sie alle Synonyme für Ihre Ressourcen definieren können. In der Registerkarte "Synonym" befindet sich die zusätzliche Spalte "Typ", in der Sie für die eingegebenen Synonyme einen Typ festlegen müssen. [Für weitere Informationen siehe Thema Verwenden des Synonymwörterbuchs für japanischen Text auf S. 269.](#) *Anmerkung:* Da die Registerkarte "Optionale Elemente" nicht für japanischen Text gilt, ist sie nicht verfügbar.

4. Fensterbereich "Wörterbuch ausschließen". Hier gibt es keine Abweichungen für japanische Textressourcen, außer dass die Verwendung des Platzhalterzeichens * nicht unterstützt wird.

5. Fensterbereich "Validierung". Für japanischen Text gibt es einen zusätzlichen Validierungsbereich, in dem Sie Ihre Ressourcen vor der Extrahierung prüfen können. Beim Extrahieren von japanischem Text kompiliert die Extrahierungsebene die Ressourcen automatisch neu, wenn vor Beginn des Extrahierungsvorgangs Änderungen erkannt werden. Um potenzielle Fehler bei der Extrahierung zu vermeiden, können Sie die Ressourcen vor dem Extrahieren neu kompilieren und validieren, damit Sie etwaige auftretende Fehler korrigieren können. [Für weitere Informationen siehe Thema Validieren und Kompilieren von japanischen Ressourcen auf S. 271.](#)

Anmerkung: Es gibt keine bearbeitbaren erweiterten Ressourcen oder Textlinkregeln für japanischen Text, deshalb sind diese Registerkarten nicht verfügbar.

Japanischer Bibliotheksbaum, Typen und Fachausdrucksbereich

Die Arbeit mit Bibliotheken und Typen erfolgt für japanische Ressourcen sehr ähnlich wie für andere Sprachen. [Für weitere Informationen siehe Thema Typ-Wörterbücher in Kapitel 10 auf S. 223.](#)

Jedoch gibt es die folgenden wesentlichen Unterschiede:

- Japanische Textressourcen verfügen über ein anderes Set an Typen. [Für weitere Informationen siehe Thema Verfügbare Typen für japanischen Text auf S. 265.](#)
- Typen können nicht erstellt oder umbenannt werden, jedoch lassen sich ihre Eigenschaften bearbeiten. [Für weitere Informationen siehe Thema Bearbeiten von japanischen Typeigenschaften auf S. 269.](#)
- Sie können Fachausdrücke hinzufügen und bearbeiten, einschließlich der Angabe eines Kana-Namens für einen Fachausdruck sowie der Zuweisung zu einem Typ und einem sekundären Stimmungstyp. [Für weitere Informationen siehe Thema Japanischer Bibliotheksbaum, Typen und Fachausdrucksbereich auf S. 262.](#)

Der Bibliotheksbaum zeigt die Bibliotheken sowie ihre Typ-Wörterbücher an. Wenn Sie eine Bibliothek oder einen Typ im linken Bereich auswählen, zeigt ein Bereich rechts die Fachausdrücke für die ausgewählten Bibliotheken bzw. Typ-Wörterbücher an. Sie können Fachausdrücke zu einem Typ-Wörterbuch direkt im Fachausdrucksbereich oder über das Dialogfeld Fachausdrücke hinzufügen hinzufügen. Bei den hinzugefügten Fachausdrücken kann es sich um einzelne Wörter oder um Wortfolgen handeln. Am Anfang der Liste befindet sich stets eine leere Zeile, in die Sie ein neues Wort eingeben können.

Wenn Sie einen Fachausdruck in einem Typ-Wörterbuch definieren, wird dieser standardmäßig als Nomen betrachtet und erhält automatisch den Typ <名詞>. Sie können jedoch den Typ in einen anderen Basistyp wie <動詞>, <形容詞>, <地名> usw. ändern. Wenn die Extrahierungsengine diesen Fachausdruck als dieselbe Wortklasse wie den Typen findet, dem Sie ihn in der Spalte Typ zugeordnet haben, wird er diesem Typ zugeordnet und extrahiert. Sie können den Fachausdruck zusätzlich einem der Stimmungstypen in der Spalte Stimmungstyp zuweisen. Wenn Sie dann die sekundäre Stimmungsanalyse verwenden, wird Text ein zweites Mal verarbeitet, um Fachausdrücke zu finden und den Stimmungstyp zuzuordnen. Wenn Sie darüber hinaus sowohl einen Stimmungstyp als auch einen Basistyp definieren und die Extrahierungsengine findet, dass dieser Fachausdruck beiden Typen entspricht, wenn ebenfalls eine sekundäre Stimmungsanalyse durchgeführt wurde, hat der Stimmungstyp Vorrang und wird im Fensterbereich "Extrahierungsergebnisse" und in den Textlinkanalyse-Ergebnissen angezeigt. Beispiel: Wenn ein Verb als Verbtyp <動詞> sowie als positive Art von Typ wie "beliebt" extrahiert wurde, würde die Oberfläche diesen Fachausdruck dem positiven Typ zugehörig anzeigen, da die Erfassung von Stimmungen meistens interessanter ist als nur eine Wortklasse.

Abbildung A-2

Fensterbereiche "Bibliothek" und "Fachausdruck" für japanische Ressourcen

Begriff	Kana	Typ	Stimmungstyp	Bibliothek
魚は新鮮だ	さかなはしんせんだ		良い - 良い	Opinions 感性意見
風塵にされた	ぼかにされた		悪い - 怒り	Opinions 感性意見
帆えに苦しむ	うえにくるしむ		悪い - 苦しい	Opinions 感性意見
風邪を引いた	かぜをひいた		悪い - 体が悪い状態	Opinions 感性意見
頭が痛い	あたまがいたい		悪い - 体が悪い状態	Opinions 感性意見
頑張ってください	がんばってください		その他 - 激励	Opinions 感性意見
機嫌ブームです	ほんりゆうぶームです		良い - 好評・人気	Opinions 感性意見
面白味を感じる	おもしろみを感じる		良い - 楽しみ全般	Opinions 感性意見
間違が多い	まちがいがおおい		悪い - 悪い	Opinions 感性意見
配慮が行き届いていた	はいりよがいきとどいてい		良い - 対応が親切	Opinions 感性意見
部屋が狭い	へやがせまい		悪い - 不満	Opinions 感性意見
運が悪む	みちがわる		悪い - 怒り全般	Opinions 感性意見
運勢が悪い	うんせいがわるい		悪い - 不運	Opinions 感性意見
運が良い	うんがよい		良い - 幸運	Opinions 感性意見
運路に迷っている	しんろにまよっている		悪い - 悩み	Opinions 感性意見
連絡がない	れんらくがない		悪い - 返答なし	Opinions 感性意見
運ってよし	いってよし		悪い - 嫌がらせ	Opinions 感性意見
塗方に書れる	どほうに書れる		悪い - 困っている	Opinions 感性意見

Tabelle A-4

Beschreibungen der Spalten im Fachausdruckbereich

Spaltenname	Spaltenbeschreibung
Begriff	Geben Sie ein Wort oder eine Wortfolge in die Zelle ein. In welcher Farbe der Fachausdruck angezeigt wird, hängt von der Farbe des Typs ab, in dem der Ausdruck gespeichert wurde. Die Typfarben können Sie im Dialogfeld "Typeigenschaften" ändern. Für weitere Informationen siehe Thema Bearbeiten von japanischen Typeigenschaften auf S. 269 . Im Allgemeinen wird der Fachausdruck in Kanji geschrieben, kann aber auch Kana umfassen. Wichtig: Die Eingabe von Verben mit Katakana-Zeichen wird nicht unterstützt.
Erzwingen	Wenn Sie auf ein Reißzweckensymbol klicken und es in dieser Zelle hinzufügen, weisen Sie die Extrahierungs-Engine an, das Vorkommen desselben Ausdrucks in allen anderen Bibliotheken zu ignorieren. Für weitere Informationen siehe Thema Erzwingen von Fachausdrücken in Kapitel 10 auf S. 231 . Dies ist für alle Sprachen gleich.
Kana	Geben Sie die Kana-Schreibung des Kanji-Fachausdrucksnamens ein.
Typ	Wählen Sie den Basistypnamen, dem der Fachausdruck zugeordnet werden soll. Für weitere Informationen siehe Thema Verfügbare Typen für japanischen Text auf S. 265 .

Spaltenname	Spaltenbeschreibung
Stimmungstyp	Wenn eine sekundäre Analyse ausgeführt wird, wählen Sie den Stimmungsnamen, dem der Ausdruck zugewiesen werden soll. Für weitere Informationen siehe Thema Verfügbare Typen für japanischen Text auf S. 265.
Bibliothek	Wählen Sie die Bibliothek aus, in der Ihr Fachausdruck gespeichert wird. Sie können einen Begriff mit der Maus auf einen anderen Typ im Bibliotheksbaum ziehen, um seine Bibliothek zu ändern.

So fügen Sie einen einzelnen Fachausdruck zu einem Typ-Wörterbuch hinzu:

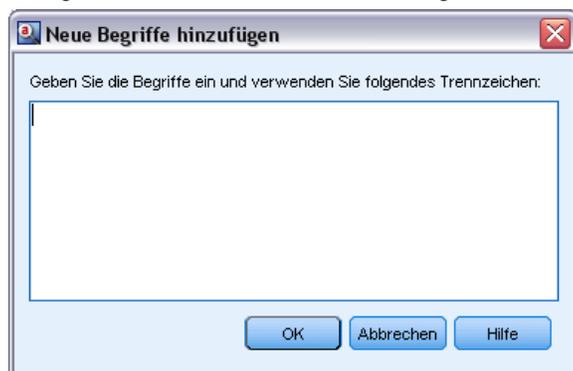
- ▶ Wählen Sie im Bibliotheksbaum das Typ-Wörterbuch aus, zu dem Sie den Fachausdruck hinzufügen möchten.
- ▶ Geben Sie in der mittleren Liste der Fachausdrücke Ihren Fachausdruck in die erste verfügbare leere Zelle ein und wählen Sie die gewünschten Optionen für diesen Fachausdruck aus.

So fügen Sie mehrere Fachausdrücke zu einem Typ-Wörterbuch hinzu:

- ▶ Wählen Sie im Bibliotheksbaumbereich das Typ-Wörterbuch aus, dem Sie die Fachausdrücke hinzufügen möchten.
- ▶ Wählen Sie im Menü Extras > Neue Begriffe aus. Das Dialogfeld “Neue Fachausdrücke hinzufügen” wird geöffnet.

Abbildung A-3

Dialogfeld “Neue Fachausdrücke hinzufügen”



- ▶ Geben Sie die Begriffe ein, die Sie dem ausgewählten Typ-Wörterbuch hinzufügen möchten. Sie können die Begriffe eingeben oder mehrere Begriffe einfügen. Wenn Sie mehrere Fachausdrücke eingeben, müssen Sie diese mit dem Trennzeichen, das im Dialogfeld “Optionen” definiert ist, trennen oder jeden Begriff in einer neuen Zeile eingeben. [Für weitere Informationen siehe Thema Festlegen von Optionen in Kapitel 2 auf S. 19.](#)
- ▶ Klicken Sie auf OK, um die Begriffe zum Wörterbuch hinzuzufügen. Das Dialogfeld wird geschlossen und die neuen Begriffe erscheinen im Wörterbuch.

Verfügbare Typen für japanischen Text

Sie können den japanischen Ressourcen keine neuen Typen hinzufügen, jedoch können Sie ihnen Fachausdrücke hinzufügen oder von ihnen entfernen. Die folgenden Tabellen umfassen das Set der aktuell verfügbaren japanischen Typen.

Typen für grundlegende Extrahierung

Bei jedem Start einer Extrahierung werden die nachstehenden Typen verwendet.

Tabelle A-5

Typen für grundlegende Extrahierung

Typen	Beschreibung
名詞	Wörter, die sich auf Dinge beziehen, z. B. "Auto" und "Film". Personennamen, Ortsnamen und Organisationsnamen werden jedoch separat kategorisiert.
人名	Nomen, die den Namen bestimmter Personen entsprechen, z. B. "Tokugawa" und "Ieyasu". Kombinationen von Vor- und Nachnamen wie "Tokugawa Ieyasu" sind ebenfalls Personennamen.
地名	Nomen wie "Tokyo" und "London", die sich auf bestimmte Orte beziehen.
組織名	Nomen, die bestimmte Firmen und Organisationen bezeichnen, z. B. "The Federation of Economic Organizations".
形容動詞	Wörter wie "still (shizuka)", die die Charakteristik oder Bedingung eines Gegenstands beschreiben und in Phrasen wie "nicht [Adjektiv] (~ de nai)" und "ein [Adjektiv] Gegenstand (~ na koto)" beschrieben werden können.
形容詞	Wörter wie "lustig (tanoshii)", die die Charakteristik oder Bedingung eines Gegenstands beschreiben und in Phrasen wie "[Adjektiv] werden (~ku naru)" und "ein [Adjektiv] Gegenstand (~i koto)" verwendet werden können.
動詞	Wörter, die Bewegungen oder Aktionen beschreiben, einschließlich Verben des Typs I (Konsonantenstamm), Verben des Typs II (Vokalstamm) sowie unregelmäßige Verben (sagyou henkaku und kagyou henkaku).
その他	Wörter wie Adverbien, Adjektivkomplemente vor Nomina, Konjunktionen und Interjektionen; Beispiele sind "ganz", "was auch immer", "dann" und "danke".

Typen für Stimmungsanalyse

Immer wenn Sie die sekundäre Analyse zur Stimmungsextrahierung verwenden, erhalten Sie eine große Anzahl an Typen zusätzlich zu den acht Basistypen.

Tabelle A-6

Typen für Stimmungsanalyse

Typen	Beschreibung
良い - 良い	Ausdrücke für generell positive Dinge, die sich als "gut" klassifizieren lassen.
良い - 嬉しい	Beschreibt ein wünschenswertes Ereignis, das eine angenehme Stimulation erzeugt.
良い - 吉報	Beschreibt ein angenehmes Ereignis, das nur durch beträchtlichen Aufwand ermöglicht werden kann.
良い - 幸運	Beschreibt ein glückliches Ereignis, das nur durch Zufall oder ein bemerkenswertes Zusammentreffen ermöglicht werden kann.
良い - 快い	Legt nahe, dass etwas als Stimulus oder Umgebung zur Auslösung einer angenehmen physiologischen Empfindung dient.

Typen	Beschreibung
良い - 体が良い状態	Beschreibt einen Zustand, in dem der Körper frei von Krankheit, Verletzung oder Müdigkeit ist, oder einen Zustand, in dem sich die physische Verfassung verbessert.
良い - 安心	Legt nahe, dass man ruhig ist und kein Risiko von Unglück oder Schaden besteht.
良い - 幸福	Gibt an, dass man durch eigene Aktionen oder die Umstände der Geburt besonders vorteilhafte Bedingungen oder Zuneigung erlangt hat.
良い - 満足	Beschreibt ein wünschenswertes Ereignis, das den Geist beruhigt.
良い - 美味しい	Gibt an, dass das Essen einen angenehmen Geschmack hat.
良い - 効果が満足	Impliziert, dass ein bestimmter Gegenstand die erwartete Wirkung erzeugt hat.
良い - 感動	Legt nahe, dass die Signifikanz, die Bedeutung oder der Wert von etwas erstaunlich positiv ist.
良い - 感謝	Legt nahe, dass man die Aktionen eines anderen auf positive Weise anerkennt.
良い - 祝福	Drückt die Ansicht aus, dass die Situation einer anderen Person vorteilhaft ist (in einem Maß, das für den Sprecher akzeptabel ist).
良い - 喜び全般	Andere positive Ereignisse oder positive Ereignisse mit wenig Verbindung zum Sprecher.
良い - 楽しい	Weist auf Aktivitäten wie Kameradschaft, Unterhaltung und Erholung hin bzw. erwartet diese.
良い - 可笑しい	Bedeutet, dass etwas eine humorvolle Qualität hat, die eine angenehme Stimulation verursacht.
良い - 笑い	Drückt ein Lächeln oder Lachen aus, das durch eine gute und/oder humorvolle Angelegenheit verursacht wurde.
良い - 期待	Sagt voraus, dass in der Zukunft ein positives Ereignis eintreten wird.
良い - 楽しみ全般	Andere erfreuliche Ereignisse und/oder positive Aktivitäten/Verhalten mit wenig Verbindung zum Sprecher.
良い - 金額への賞賛	Deutet darauf hin, dass etwas vom Käuferstandpunkt gesehen einen wünschenswerten Geldwert hat.
良い - 対応が早い	Legt nahe, dass ein Service in der erwarteten Zeit geliefert oder abgeschlossen wurde.
良い - 対応が親切	Legt nahe, dass die Haltung oder das Verhalten eines Serviceanbieters gewissenhaft war.
良い - 説明が良い	Drückt die Idee aus, dass der Typ und/oder die Quantität an Information und/oder die Methode der Informationsbereitstellung angemessen ist.
良い - 対応への賞賛	Ansichten, die von den obigen abweichen, die den Anbieter eines Service loben.
良い - 褒め・賞賛	Ansichten, die von den obigen abweichen, die die Merkmale, die Fähigkeiten und/oder den Betrieb eines bestimmten Gegenstands loben.
良い - 好き	Drückt den Wunsch aus, einen bestimmten Gegenstand zu besitzen oder diesem näherzukommen.
良い - 入会希望	Beschreibt den Wunsch, einer bestimmten Gruppe anzugehören bzw. Teil dieser Gruppe zu bleiben.
良い - 買いたい	Impliziert, dass man Geld für den Bezug eines bestimmten Gegenstands aufbringen möchte.
良い - 好評・人気	Gibt an, dass die Anzahl der Personen, die einen bestimmten Gegenstand möchten oder schätzen, ein bestimmtes Ziel überschritten hat.
良い - 売れた	Weist darauf hin, dass die Präsenz von Personen, die einen bestimmten Gegenstand kaufen oder die Anzahl oder der Wert der Käufe ein bestimmtes Ziel überschritten hat.
悪い - 悪い	Ausdrücke für generell negative Dinge, die sich als "schlecht" klassifizieren lassen.

Typen	Beschreibung
悪い - 怒り	Ein deutliches Gefühl des Ärgers, wenn etwas nicht wie geplant läuft.
悪い - 批判	Drückt den Gedanken aus, dass eine andere Person nicht die passende Wahl getroffen hat.
悪い - お叱り	Wörter oder Aktionen, die eine Person so einschüchtern, dass sie den Absichten einer anderen Person folgen.
悪い - 誹謗・中傷	Wörter, mit denen eine äußerst schlechte Meinung von einer anderen Person demonstriert wird.
悪い - 軽蔑	Impliziert, dass der Charakter, die Fähigkeiten und/oder sonstigen Qualitäten einer anderen Person zu wünschen übrig lassen.
悪い - 恨み	Drückt Ärger oder Verstimmung für einen Nachteil aus, der durch eine andere Person verursacht wurde.
悪い - 嫌がらせ	Wörter, die verwendet werden, um Kommunikation zu unterbinden.
悪い - 不満	Ein unangenehmes Gefühl, das sich aus der Unfähigkeit ergibt, den gewünschten Gegenstand oder Zustand zu erreichen.
悪い - 不味い	Gibt an, dass das Essen einen schlechten Geschmack hat.
悪い - 効果が不満	Impliziert, dass etwas nicht die erwartete Wirkung erzeugt hat.
悪い - 金額が不満	Deutet darauf hin, dass etwas vom Käuferstandpunkt gesehen einen nicht erwünschten Geldwert hat.
悪い - 対応への不満	Legt nahe, dass der Anbieter eines Service im Unrecht ist.
悪い - 対応が遅い	Legt nahe, dass ein Service nicht rechtzeitig ausgeführt/abgeschlossen wurde oder dass der Service noch abgeschlossen werden muss.
悪い - 対応が不親切	Bezeichnet ein unangenehmes Gefühl, das durch die Haltung oder das Verhalten eines Serviceanbieters verursacht wurde.
悪い - 説明が悪い	Drückt die Idee aus, dass der Typ und/oder die Quantität an Information und/oder die Methode der Informationsbereitstellung nicht angemessen ist.
悪い - 返答なし	Gibt an, dass der Anbieter eines Service nicht geeignet reagiert, obwohl die Situation eine entsprechende Reaktion verlangt.
悪い - 不快	Legt nahe, dass etwas als Stimulus oder Umgebung zur Auslösung einer negativen physiologischen Empfindung dient.
悪い - 怒り全般	Gefühle des Ärgers, die von den obigen abweichen. Allgemeiner Ärger in der Organisation oder Firma des Sprechers oder Beschreibungen von Ereignissen, die durch den erwähnten Ärger verursacht wurden.
悪い - 悲しい	Ein deutlich unangenehmes Gefühl, wenn jemand etwas verliert oder nicht bekommt.
悪い - 凶報	Drückt den Gedanken aus, dass ein bestimmtes Ziel trotz erheblichem Aufwand nicht erreicht werden konnte.
悪い - 不運	Gibt ein negatives Ergebnis an, das durch ein unglückliches Zusammentreffen oder Pech und nicht durch einen eigenen Fehler verursacht wurde.
悪い - ショック	Legt nahe, dass jemand durch eine unverhergesehene, negative Sache oder ein Ereignis aufgeregt ist und keine geeignete Antwort finden kann.
悪い - 残念	Ein unglückliches Gefühl, das entsteht, wenn ein erwartetes Geschehnis nicht eintritt.
悪い - 落胆	Ein Zustand, in dem einen ein unglückliches, enttäushtes Gefühl überkommt.
悪い - 諦め	Legt nahe, dass eine negative Sache, die dem Sprecher oder einer anderen Person widerfahren ist, nicht verbessert werden kann.
悪い - 後悔	Drückt den Gedanken aus, dass man in der Vergangenheit versäumt hat, die korrekte Wahl zu treffen, obwohl diese als Option verfügbar war.
悪い - 謝罪	Gibt die Erkenntnis des Sprechers an, dass dieser einer anderen Person Schaden zugefügt hat.

Typen	Beschreibung
悪い - 淋しい	Drückt den Gedanken aus, dass Kontakt mit anderen selten stattfindet oder dass nur mit wenigen anderen Personen Kontakt möglich ist.
悪い - 哀れみ	Drückt den Gedanken aus, dass die Situation einer anderen Person erheblich schlechter als die des Sprechers ist.
悪い - 悩み	Gibt an, dass eine Auswahl getroffen werden muss, man aber nicht in der Lage ist, aus den verfügbaren Optionen zu wählen.
悪い - 困っている	Drückt den Gedanken aus, dass es keine wirksame Methode gibt, auf eine Situation zu reagieren, die eine Aktion erfordert.
悪い - 苦しい	Drückt einen unangenehmen psychologischen Zustand aus, in dem man aufgrund von externen Ursachen oder eigener Fehler nicht normal handeln kann.
悪い - 体が悪い状態	Beschreibt einen Zustand, in dem der Körper krank, verletzt und/oder müde ist, oder einen Zustand, in dem sich die physische Verfassung nicht verbessert.
悪い - 不安	Drückt den Gedanken aus, dass etwas nicht in seinem wünschenswerten Zustand fortfährt oder die Erwartungen nicht erfüllt.
悪い - 恐怖	Legt nahe, dass eine bestimmte Sache wahrscheinlich Schaden oder Verletzungen verursacht.
悪い - 悲しみ全般	Von den obigen abweichende Gefühle der Traurigkeit, z. B. allgemeine Traurigkeit über eine nicht spezifizierte Sache.
悪い - 嫌い	Gibt an, dass man eine Sache auf Abstand halten oder sich von einer Sache entfernen möchte.
悪い - 退会希望	Beschreibt den Wunsch, eine bestimmte Gruppe zu verlassen bzw. dieser Gruppe nicht beizutreten.
悪い - 買いたくない	Legt nahe, dass man eine bestimmte Sache nicht wünscht oder nicht beabsichtigt, für diese Sache zu bezahlen.
悪い - 不評・不人気	Gibt an, dass die Anzahl an Personen, die eine bestimmte Sache mögen, nicht einem bestimmten Ziel entspricht bzw. dass viele Personen der erwähnten Sache gegenüber negative Gefühle hegen.
悪い - 売れていない	Weist darauf hin, dass die Abwesenheit von Personen, die einen bestimmten Gegenstand kaufen, oder die Anzahl oder der Wert der Käufe ein bestimmtes Ziel nicht erreicht hat.
その他 - 疑問	Ausdrücke, die Informationen anfordern, zu denen die andere Person weitere Untersuchungen oder Gedanken aufbringen muss.
その他 - 問い合わせ	Ausdrücke, die Informationen anfordern, die bereits im Besitz der anderen Person sind.
その他 - 要望	Ausdrücke, die der anderen Person befehlen, ein Problem zu lösen (wenn die andere Person direkt Schuld ist oder eine niedrigere Position als der Sprecher hat).
その他 - 提案・忠告	Ausdrücke, die der anderen Person befehlen, sich besser zu benehmen (wenn die andere Person direkt Schuld ist oder eine niedrigere Position als der Sprecher hat).
その他 - お願い	Ausdrücke, die der anderen Person befehlen, ein Problem zu lösen (wenn die andere Person nicht Schuld ist oder keine niedrigere Position als der Sprecher hat).
その他 - 激励	Ausdrücke, die eine andere Person ermutigen, oder Beschreibungen eines ermutigenden Verhaltens.
その他 - 勧誘	Ausdrücke, die einer anderen Person befehlen, gemeinsam mit dem Sprecher etwas zu unternehmen.
その他 - 驚き	Drückt den Gedanken aus, dass die Plötzlichkeit oder der Grad eines Ereignisses über ein rationales Urteil/Verständnis hinausgeht.
評価なし - 評価なし	Kein Ausdruck der Bewertung.

Bearbeiten von japanischen Typeigenschaften

Sie können zwar in japanischen Ressourcen keine Typen erstellen, jedoch können Sie Typeigenschaften anzeigen und bearbeiten. Beachten Sie, dass die Optionen wie die Abgleichoption und flektierte Formen nicht für japanischen Text gelten.

Abbildung A-4

Dialogfeld "Typeinstellungen" für japanische Textressourcen



Name. Der Name des Typwörterbuchs.

Hinzufügen zu. Dieses Feld gibt die Bibliothek an, in der Sie Ihr neues Typ-Wörterbuch erstellen.

Schriftfarbe. Über dieses Feld können Sie ein Unterscheidungsmerkmal für diesen Typ in Bezug auf die anderen auf der Benutzeroberfläche dargestellten Typen festlegen. Wenn Sie die Option Übergeordnete Farbe verwenden auswählen, wird für dieses Typ-Wörterbuch die Standardtypfarbe verwendet. Die Standardfarbe wird im Dialogfeld "Optionen" festgelegt. [Für weitere Informationen siehe Thema Optionen: Registerkarte "Anzeige" in Kapitel 2 auf S. 21.](#) Wenn Sie Benutzerdefiniert auswählen, wählen Sie aus der Dropdown-Liste eine Farbe aus.

Anmerkung. Dieses Feld ist optional und wird für Kommentare und Beschreibungen verwendet.

So können Sie die Typeigenschaften anzeigen bzw. bearbeiten:

- ▶ Wählen Sie die Typeigenschaften aus, die Sie sehen möchten.
- ▶ Klicken Sie mit der rechten Maustaste und wählen Sie im Kontextmenü Typeigenschaften aus. Das Dialogfeld "Typ-Eigenschaften" wird geöffnet.
- ▶ Nehmen Sie alle erforderlichen Änderungen vor.
- ▶ Klicken Sie auf OK, um die Änderungen im Typ-Wörterbuch zu speichern.

Verwenden des Synonymwörterbuchs für japanischen Text

Für japanischen Text enthält das Substitutionswörterbuch nur eine Registerkarte zur Verwaltung Ihrer Synonyme, nämlich die Registerkarte "Synonyme". Synonyme verknüpfen zwei oder mehr Wörter mit derselben Bedeutung. Mithilfe von Synonymen können Sie außerdem Fachausdrücke

mit ihren Abkürzungen gruppieren oder auch falsch geschriebene Wörter mit dem Fachausdruck in der richtigen Schreibweise.

Abbildung A-5
Synonymeinträge für japanischen Text

	Ziel	Synonyme	Typ	Bibliothek
0				
1	☑ コミュニケーション	☑ コミュニケーション, ☑ コミュニケーション	名詞	Opinions 感性意見
2	☑ コンピューター	☑ pc, ☑ パソコン	名詞	Opinions 感性意見

Eine Synonymdefinition setzt sich aus zwei Teilen zusammen. Der Zielausdruck ist der Ausdruck, unter dem die Extrahierungsengine alle Synonyme zusammenfassen soll. Wenn dieser Zielausdruck nicht als Synonym eines anderen Zielausdrucks verwendet wird oder er ausgeschlossen wird, ist es wahrscheinlich, dass dieser Ausdruck das Konzept wird, das im Bereich “Extrahierungsergebnisse” angezeigt wird. Die Liste der Synonyme sind die Ausdrücke, die unter dem Zielbegriff zusammengefasst werden.

Auf der Registerkarte “Synonyme” können Sie in die Leerzeile am Anfang der Tabelle eine Synonymdefinition eingeben. Beginnen Sie, indem Sie den Zielausdruck und seine Synonyme definieren. Sie können auch die Bibliothek auswählen, in der diese Definition gespeichert werden soll. Während der Extrahierung werden alle Fundstellen der Synonyme für die endgültige Extrahierung unter dem Zielausdruck gruppiert. [Für weitere Informationen siehe Thema Hinzufügen von Fachausdrücken in Kapitel 10 auf S. 227.](#)

Wenn Sie Ihre Typ-Wörterbücher erstellen, können Sie einen Ausdruck eingeben, für den Ihnen drei oder vier Synonyme einfallen. In diesem Fall könnten Sie alle Ausdrücke und anschließend Ihren Zielausdruck in das Substitutionswörterbuch eingeben und dann die Synonyme durch Ziehen übertragen.

Wichtig: Platzhalter- und Sonderzeichen werden für japanische Textsynonyme nicht unterstützt.

So fügen Sie einen Synonymeintrag hinzu:

- ▶ Geben Sie in der leeren Zeile am Anfang der Tabelle in der Registerkarte “Synonym” im Bereich “Substitution” Ihren Zielausdruck in die Zielspalte ein. Der von Ihnen eingegebene Zielausdruck wird farbig angezeigt. Die Farbe stellt den Typ dar, als der der Ausdruck erscheint oder in den er erzwungen wird, sofern dies der Fall ist. Wenn der Ausdruck schwarz dargestellt wird, bedeutet dies, dass er in keinem der Typ-Wörterbücher vorkommt.
- ▶ Klicken Sie auf die zweite Zelle rechts neben dem Zielausdruck und geben Sie die Synonyme ein. Trennen Sie die einzelnen Einträge mithilfe des globalen Trennzeichens, das im Dialogfeld “Optionen” definiert ist. Alle eingegebenen Synonyme sollten denselben Typ aufweisen. [Für weitere Informationen siehe Thema Festlegen von Optionen in Kapitel 2 auf S. 19.](#) Die von Ihnen eingegebenen Ausdrücke werden farbig angezeigt. Die Farbe stellt den Typ dar, in der der Ausdruck angezeigt wird. Wenn der Ausdruck schwarz dargestellt wird, bedeutet dies, dass er in keinem der Typ-Wörterbücher vorkommt.
- ▶ Legen Sie in der dritten Spalte, der Spalte “Typ”, einen Typ für diese Synonyme fest. Das Ziel übernimmt jedoch den Typ, der bei der Extrahierung zugewiesen wurde. Wenn allerdings das

Ziel nicht als ein Konzept extrahiert wurde, wird der in dieser Spalte aufgelistete Typ dem Ziel in den Extrahierungsergebnissen zugewiesen.

- ▶ Klicken Sie auf die letzte Zelle, um die Bibliothek auszuwählen, in der die Synonymdefinition gespeichert werden soll.

Anmerkung: Diese Anweisungen zeigen Ihnen, wie Sie in der Resource Editor-Ansicht Änderungen vornehmen können. Beachten Sie, dass Sie solche Feinabstimmungen auch direkt in den Bereichen “Extrahierungsergebnisse”, “Daten”,. [Für weitere Informationen siehe Thema Extrahierungsergebnisse verfeinern in Kapitel 5 auf S. 88.](#)

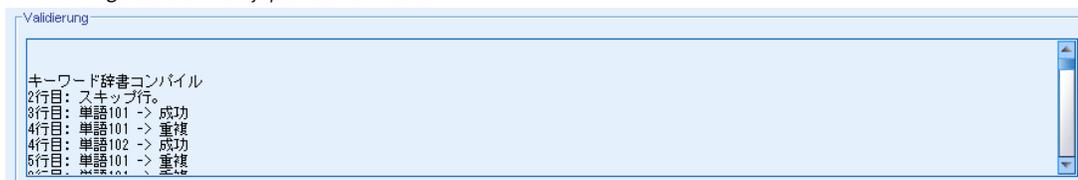
Validieren und Kompilieren von japanischen Ressourcen

Für japanischen Text gibt es einen zusätzlichen Validierungsbereich, in dem Sie Ihre Ressourcen vor der Extrahierung prüfen können. Bevor der Extrahierungsprozess für japanischen Text beginnt, kompiliert die Extrahierungseingine automatisch die Ressourcen, wenn vor dem Start des Extrahierungsprozesses Änderungen erkannt wurden. Wenn bei der Extrahierung ein Fehler gefunden wird, kann der Vorgang möglicherweise nicht korrekt abgeschlossen werden.

Zur Vermeidung von Kompilierungsfehlern wird empfohlen, dass Sie Ihre Ressourcen nach dem Durchführen von Änderungen im Resource Editor validieren und kompilieren. Wenn Fehlermeldungen auftreten, können Sie Korrekturen ausführen und die Validierung erneut versuchen.

Abbildung A-6

Validierungsbereich für japanischen Text



So validieren Sie Ressourcen:

- ▶ Wählen Sie in den Menüs Extras > Ressourcen-Validierung. Der Validierungsbereich wird geöffnet, um Kompilierungs- und Fehlermeldungen anzuzeigen.

Weitere Ausnahmen für Japanisch

Interne Ressourcen, die benutzerdefinierte Ressourcen überschreiben

Für japanischen Text umfassen die Standardressourcen einige vorkompilierte, interne Basisressourcen. Diese internen Ressourcen sind nicht editierbar. Dennoch können Sie Resource Editor verwenden, um einige Änderungen und Feineinstellungen vorzunehmen. In beinahe allen Fällen haben die Fachausdrücke, Synonyme und Ausschlusslisteneinträge, die Sie in Ihren Ressourcen definieren, Vorrang vor den vorkompilierten internen Ressourcen. Jedoch gibt es mehrere Ausnahmen, wie in einigen der folgenden Beispiele illustriert.

- Es gibt Instanzen, in denen das Hinzufügen von Fachausdrücken zu einem bestimmten Typ keine Wirkung auf die Extrahierungsergebnisse hat. Dies tritt am wahrscheinlichsten auf, wenn die Daten lange Sätze mit mehreren morphologischen Elementen, Interpunktion oder Symbolen enthalten. Da in japanische Textressourcen zudem eine große Anzahl an häufigen Fachausdrücken vorkompiliert ist, gibt es einige häufige Wörter, die stets in eine bestimmte linguistische Definition gezwungen werden.
- Eventuell können Sie Fachausdrücke wie ある, いる oder なる nicht ausschließen, da die Extrahierungseingine die Extrahierung dieser Fachausdrücke immer erzwingt.
- Es ist zwar möglich, den Typ des Fachausdrucks 東京 von <地名> in <名詞> zu ändern, jedoch ignoriert die Extrahierungseingine Ihre Änderung, wenn Sie versuchen, mit dem Stichwort- (Typ-) Wörterbuch den Typ eines Fachausdrucks von <地名> in <動詞> oder in <形容詞> zu ändern.
- Gelegentlich ist es möglich, dass Änderungen, die Sie im Resource Editor oder im vornehmen, die Extrahierungsergebnisse von einem Satz beeinflussen, jedoch nicht von einem anderen Satz, da der Extrahierungsvorgang durch Verweisen auf die Kookkurenzwörter in jedem Satz endet.

Probleme mit Katakana-Anzeige in halber Breite

Katakana-Zeichen in halber Breite werden bei der Extrahierung intern in Katakana-Zeichen in voller Breite umgewandelt, erscheinen aber dennoch im Datenbereich in der Benutzeroberfläche. Beachten Sie, dass Katakana-Zeichen in halber Breite im Datenbereich nicht markiert werden können. Konvertieren Sie zur Vermeidung dieses Problems alle Ihre Datensätze vor der Verarbeitung in Katakana in voller Breite.

Verwendung von Klein- und Großschreibung

Großbuchstaben werden beim Einlesen in die Anwendung temporär in Kleinbuchstaben umgewandelt. Jedoch zeigt der Datenbereich den Text stets in derselben Schreibung wie im Originaltext an. Klein- und Großbuchstaben werden in diesem Produkt als identisch behandelt.

Hinweise

Diese Informationen wurden für weltweit angebotene Produkte und Dienstleistungen erarbeitet.

IBM bietet die in diesem Dokument behandelten Produkte, Dienstleistungen oder Merkmale möglicherweise nicht in anderen Ländern an. Informationen zu den in derzeit in Ihrem Land erhältlichen Produkten und Dienstleistungen erhalten Sie bei Ihrem zuständigen IBM-Mitarbeiter vor Ort. Mit etwaigen Verweisen auf Produkte, Programme oder Dienste von IBM soll nicht behauptet oder impliziert werden, dass nur das betreffende Produkt oder Programm bzw. der betreffende Dienst von IBM verwendet werden kann. Stattdessen können alle funktional gleichwertigen Produkte, Programme oder Dienste verwendet werden, die keine geistigen Eigentumsrechte von IBM verletzen. Es obliegt jedoch der Verantwortung des Benutzers, die Funktionsweise von Produkten, Programmen oder Diensten von Drittanbietern zu bewerten und zu überprüfen.

IBM verfügt möglicherweise über Patente oder hat Patentanträge gestellt, die sich auf in diesem Dokument beschriebenen Inhalte beziehen. Durch die Bereitstellung dieses Dokuments werden Ihnen keinerlei Lizenzen an diesen Patenten gewährt. Lizenzanfragen können schriftlich an folgende Adresse gesendet werden:

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785, U.S.A.

Bei Lizenzanfragen in Bezug auf DBCS-Daten (Double-Byte Character Set) wenden Sie sich an die für geistiges Eigentum zuständige Abteilung von IBM in Ihrem Land. Schriftliche Anfragen können Sie auch an folgende Adresse senden:

Intellectual Property Licensing, Legal and Intellectual Property Law, IBM Japan Ltd., 1623-14, Shimotsuruma, Yamato-shi, Kanagawa 242-8502 Japan.

Der folgende Absatz gilt nicht für Großbritannien oder andere Länder, in denen derartige Bestimmungen nicht mit dem dort geltenden Recht vereinbar sind. INTERNATIONAL BUSINESS MACHINES ÜBERNIMMT FÜR DIE VORLIEGENDE DOKUMENTATION KEINERLEI GEWÄHRLEISTUNG IRGENDWELCHER ART, WEDER AUSDRÜCKLICH NOCH STILLSCHWEIGEND, EINSCHLIESSLICH, JEDOCH NICHT DARAUF BEGRENZT, DER STILLSCHWEIGENDEN GEWÄHRLEISTUNGEN IN BEZUG AUF DIE NICHTVERLETZUNG VON RECHTEN DRITTER, AUF HANDELSÜBLICHKEIT ODER DIE EIGNUNG FÜR EINEN BESTIMMTEN ZWECK. Einige Staaten lassen bei bestimmten Transaktionen keine Ausschlussklauseln ausdrücklicher oder stillschweigender Gewährleistungen zu, sodass diese Erklärung möglicherweise nicht auf Sie zutrifft.

Diese Informationen können technische Ungenauigkeiten oder typografische Fehler enthalten. An den hierin enthaltenen Informationen werden in regelmäßigen Abständen Änderungen vorgenommen, die in spätere Ausgaben der Publikation eingearbeitet werden. IBM kann jederzeit ohne Vorankündigung Verbesserungen und/oder Veränderungen an den in dieser Publikation beschriebenen Produkten und/oder Programmen vornehmen.

Alle in diesen Ausführungen enthaltenen Verweise auf Websites, die nicht zu IBM gehören, dienen lediglich der Information. Die Nennung bedeutet nicht, dass IBM den Inhalt dieser Websites unterstützt. Das Material auf diesen Websites ist kein Bestandteil des Materials für dieses IBM-Produkt. Sie verwenden diese Websites auf eigenes Risiko.

IBM ist berechtigt, die von Ihnen bereitgestellten Informationen in jeglicher Form zu verwenden bzw. weiterzugeben, die dem Unternehmen geeignet erscheint, ohne dass ihm daraus Verbindlichkeiten Ihnen gegenüber entstehen.

Lizenznehmer dieses Programms, die Informationen dazu benötigen, wie (i) der Austausch von Informationen zwischen unabhängig erstellten Programmen und anderen Programmen und (ii) die gegenseitige Verwendung dieser ausgetauschten Informationen ermöglicht wird, wenden sich an:

IBM Software Group, Attention: Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA.

Derartige Informationen stehen ggf. in Abhängigkeit von den jeweiligen Geschäftsbedingungen sowie in einigen Fällen der Zahlung einer Gebühr zur Verfügung.

Das in diesem Dokument beschriebene lizenzierte Programm und sämtliche dafür verfügbaren lizenzierten Materialien werden von IBM gemäß dem IBM-Kundenvertrag, den internationalen Nutzungsbedingungen für Programmpakete der IBM oder einer anderen zwischen uns getroffenen Vereinbarung bereitgestellt.

Alle in diesem Dokument enthaltenen Leistungsdaten wurden in einer kontrollierten Umgebung ermittelt. Daher können die unter anderen Betriebsumgebungen erzielten Ergebnisse erheblich abweichen. Einige Messungen wurden möglicherweise an Systemen im Entwicklungsstadium vorgenommen und es besteht keine Garantie, dass spätere allgemein verfügbare Systeme dieselben Messwerte aufweisen. Außerdem wurden einige Messwerte möglicherweise mittels Extrapolation geschätzt. Die tatsächlichen Ergebnisse können abweichen. Die Benutzer dieses Dokuments sollten die entsprechenden Daten für ihre jeweilige Umgebung überprüfen.

Informationen zu Drittanbieterprodukten stammen von den Herstellern dieser Produkte, ihren veröffentlichten Verlautbarungen oder aus anderen öffentlich verfügbaren Quellen. IBM hat diese Produkte nicht getestet und kann daher die Richtigkeit der Angaben zu Leistung und Kompatibilität oder anderer Behauptungen in Bezug auf Drittanbieterprodukte nicht bestätigen. Fragen zu den Funktionen von Drittanbieterprodukten sind an die Hersteller dieser Produkte zu richten.

Diese Informationen enthalten Beispiele für Daten und Berichte, die in alltäglichen Betriebsabläufen verwendet werden. Um sie möglichst umfassend darzulegen, enthalten die Beispiele Namen von Einzelpersonen, Unternehmen, Marken und Produkten. Alle diese Namen sind frei erfunden und jegliche Ähnlichkeit mit den von einem tatsächlichen Handelsunternehmen verwendeten Namen und Adressen ist rein zufällig.

Bei der Anzeige dieser digitalen Informationsversion sind die Fotografien und Farbillustrationen möglicherweise nicht sichtbar.

Marken

IBM, das IBM-Logo, ibm.com und SPSS sind Marken von IBM Corporation, die in vielen Ländern weltweit eingetragen sind. Eine aktuelle Liste der IBM-Marken finden Sie im Internet unter <http://www.ibm.com/legal/copytrade.shtml>.

Adobe, das Adobe-Logo, PostScript und das PostScript-Logo sind eingetragene Marken oder Marken von Adobe Systems Incorporated in den USA und/oder anderen Ländern.

Intel, das Intel-Logo, Intel Inside, das Intel Inside-Logo, Intel Centrino, das Intel Centrino-Logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium und Pentium sind Marken oder eingetragene Marken von Intel Corporation oder seinen Tochtergesellschaften in den USA und anderen Ländern.

Linux ist eine eingetragene Marke von Linus Torvalds in den USA und/oder anderen Ländern.

Microsoft, Windows, Windows NT und das Windows-Logo sind Marken von Microsoft Corporation in den USA und/oder anderen Ländern.

UNIX ist eine eingetragene Marke von The Open Group in den USA und anderen Ländern.

Java und alle Java-basierten Marken und Logos sind Marken von Sun Microsystems, Inc. in den USA und/oder anderen Ländern.

Screenshots von Microsoft-Produkten nachgedruckt mit Genehmigung durch Microsoft Corporation.

Weitere Produkt- oder Servicenamen können Marken von IBM oder anderen Unternehmen sein.

Index

- ! ^ * \$ Symbole in Synonymen, 237
- & | !() Regeloperator, 158

- Abgeschlossen-Flag, 76
- Abgleichsoption, 223, 225, 227, 229–230, 269
- Abhängigkeitsanalyse, 257–258
- Abkürzungen, 252–253
- Abschnitte über Sprachverwendung, 242, 252
 - Abkürzungen, 252–253
 - Erzwungene Definitionen, 252
 - Extraktionsmuster, 252
- Adressen (nicht linguistisches Element), 246
- Aktivieren nicht linguistischer Elemente, 250
- Aktualisieren
 - Bibliotheken, 217, 219
 - Diagramme, 172
 - Vorlagen, 200
- Aktualisieren von Diagrammen, 172
- Alle Dokumente, 98
- Aminosäuren (nicht linguistisches Element), 246
- Ändern
 - Datenquelle, 63
 - Vorlagen, 201
- Anmerkungen
 - für Kategorien, 111, 160
 - für Projekte, 50
- Ansicht “Frage”, 12–13
- Ansicht “Gesamtes Projekt”, 12, 15
- Ansichten
 - Ansicht “Frage”, 13
 - Fenster “Ressourceneditor”, 16
 - Gesamtes Projekt, 15
 - Textanalysefenster, 12
- Anti-Links, 120
- Antworten, 100
 - Als abgeschlossen markieren, 76
 - Einschließen in Kategorien erzwingen, 165
 - Kennzeichnen, 76
- Antworten ohne Inhalt, 28
- Anzeigen
 - Bibliotheken, 213
 - Daten, 51
 - Kategorien, 172
- Anzeigen (Schaltfläche), 99
- Ausnahmen für unscharfe Gruppierung, 86, 242, 245
- Ausrufezeichen (!), 237
- Ausschließen
 - Aus Kategoriezusammenhängen, 120
 - Bibliotheken deaktivieren, 214
 - Deaktivieren von Ausschlusseinträgen, 241
 - Deaktivieren von Wörterbüchern, 234, 239
 - Konzepte von der Extrahierung ausschließen, 94
 - von der Fuzzy-Gruppierung, 245
- Ausschlusswörterbuch, 209, 240–241
- Ausweichen, 193
- auto-, Einstellungen, 178

- Automatisches Speichern von Projekten, 19

- Balkendiagramme, 172
- Bearbeiten
 - Eigenschaften, 160
 - Extrahierungsergebnisse verfeinern, 88
 - Kategorien, 159–160, 162
 - Kategorienregeln, 159
- Bearbeiten von Visualisierungen, 176
 - Abstand, 182
 - Achsen, 184
 - Ausschließen von Kategorien, 186
 - Auswahl, 177
 - automatische Einstellungen, 177
 - Farben und Muster, 179
 - Fenster, 188
 - Hinzufügen von 3-D-Effekten, 188
 - Kategorien, 186
 - Kombinieren von Kategorien, 186
 - Legendenposition, 193
 - Punktform, 181
 - Punktrotation, 181
 - Punktverhältnis, 181
 - Ränder, 182
 - Reduzieren von Kategorien, 186
 - Regeln, 177
 - Skalen, 184
 - Sortieren von Kategorien, 186
 - Strichmuster, 179
 - Text, 179
 - Transformieren von Koordinatensystemen, 188
 - Transparenz, 179
 - transponieren, 188
 - Zahlenformate, 183
- Bearbeitungsmodus, 176
- Begriffskomponentenbildung, 122
- Benennen
 - Bibliotheken, 214
 - Kategorien, 111, 160
 - Typ-Wörterbücher, 232
- Benutzerdefinierte Farben, 21
- Berichte und Zusammenfassungsdiagramme, 60
- Berücksichtigen
 - Interpunktionsfehler, 86
 - Rechtschreibfehler, 86
- Bezeichnungen für Kategorien, 111, 160
- Bibliotheken, 16, 209, 223
 - Aktualisieren, 219
 - Anzeigen, 213
 - Benennen, 214
 - Budget Library, 224
 - Core Library, 224
 - Deaktivieren, 214
 - Erstellen, 211
 - Exportieren, 216
 - Gemeinsam benutzen, 76

- Gemeinsam nutzen und veröffentlichen, 217
- Hinzufügen, 211
- Importieren, 215
- Local Libraries, 217
- Löschen, 215, 217
- Mitgelieferte (Standard-) Bibliotheken, 209
- Opinions Library, 224
- Public Libraries, 217
- Synchronisieren, 217
- Umbenennen, 214
- Verknüpfen, 211
- Veröffentlichen, 53, 219
- Warnung zur Synchronisierung der Bibliothek, 217
- Wörterbücher, 209
- Bibliotheken filtern, 213
- Bibliotheken gemeinsam nutzen, 217
 - Aktualisieren, 219
 - Public Libraries hinzufügen, 211
 - Veröffentlichen, 53, 219
- Bibliotheken synchronisieren, 217, 219
- Boole'sche Operatoren, 158
- Budget Library, 224
- Budgettyp-Wörterbuch, 224

- cluster, 193
- Coderahmen, 136
- Core Library, 224

- Dateiwiederherstellung, 19
- Daten
 - Aktualisieren, 74
 - Ändern der Datenquelle, 63
 - Anzeigen, 51
 - Bearbeiten von Variableneigenschaften, 52
 - Datenquellenauswahl, 30, 64
 - Ergebnisse verfeinern, 88
 - Exportieren, 55
 - Extrahieren, 85
 - IBM SPSS Data Collection, 35, 56, 69
 - IBM SPSS Statistics.sav-Dateien, 31, 55–56, 65
 - Kategorieerstellung, 7, 116, 118, 121, 129
 - Kategorisierung, 96, 112, 134
 - Microsoft Excel.xls-/xlsx-Dateien, 32, 55, 58, 66
 - ODBC, 34, 68
 - sortieren, 52
- Datenbereich, 100
 - Anzeigen (Schaltfläche), 99
- Datensätze, 100
- Datumsangaben (nicht linguistisches Element), 246
- Deaktivieren
 - Ausschlusswörterbücher, 241
 - Bibliotheken, 214
 - Nicht linguistische Elemente, 250
 - Statusleisten, 77
 - Substitutionswörterbücher, 239
 - Synonymwörterbücher, 245
 - Typ-Wörterbücher, 234
- Deaktivieren nicht linguistischer Elemente, 250
- Definitionen, 105, 111
- Deskriptoren, 98
 - Auswahl der besten, 107
 - Bearbeiten in Kategorien, 162
 - Kategorien, 105, 111
- Diagramm/Tabelle für Kategorienetz, 172, 174–175
- Diagramme, 172
 - Aktualisieren, 172
 - Bearbeiten, 176
 - Exportieren von Zusammenfassungsdiagrammen, 60
 - Größe von Grafikelementen, 182
 - Kategorienetzdiagramm, 172
 - Untersuchungsmodus, 175
- Diagramme bearbeiten
 - Größe von Grafikelementen, 182
- Dokumente (Spalte), 98–99
- Dollarzeichen (\$), 237
- Drucken von Kategorien, 169

- E-Mail (nicht linguistisches Element), 246
- Eigenschaften
 - für japanische Typen, 269
 - Kategorien, 111, 160
 - Projekte, 50
 - Variablen, 52
- Einfaches Listenformat, 141
- Eingerücktes Format, 144
- Einstellungen, 19, 21, 23
- Ergebnisse verfeinern
 - Erstellen von Typen, 91
 - Extrahierungsergebnisse, 9, 88
 - Kategorien, 9, 159
 - Konzepte ausschließen, 94
 - Konzepte zu Typen hinzufügen, 91
 - Konzeptextrahierung erzwingen, 95
 - Synonyme hinzufügen, 89
- Erstellen
 - Ausschlusswörterbucheinträge, 241
 - Bibliotheken, 211
 - Kategorien, 104, 112, 135
 - Kategorien mit Regeln, 149
 - Kategorienregeln, 148–149, 157
 - Optionale Elemente, 238
 - Projekte, 29
 - Synonyme, 88–89, 236
 - Synonyme für Japanisch, 269
 - Typ-Wörterbücher, 225, 269
 - Typen, 91
 - Vorlage aus Ressourcen, 200
- Erstellen von Kategorien, 6–7, 112, 114, 116, 118, 120–129, 131, 134, 260
 - Gruppierungstechniken Peer/Geschwister, 131
- Erstellen von Vorlagen über Ressourcen, 200
- Erweitern von Kategorien, 129
- Erweiterte Ressourcen, 242
 - Suchen und Ersetzen im Editor, 243–244

- Erzwingen
 Antworten, 165
 Erzwungenermaßen ausschließen, 100
 Erzwungenermaßen einschließen, 100
 Fachausdrücke, 231
 Konzeptextrahierung, 95
 Spalten anzeigen, 100
 Wörter in Kategorien, 167
- Erzwungene Definitionen, 252
- Exportieren, 55
 als Microsoft Excel-Dateien (.xls/.xlsx), 58
 als IBM SPSS Statistics.sav-Dateien, 56
 Ausgabeformate, 55
 für IBM SPSS Data Collection, 56
 Kategorisierungsergebnisse, 55–56, 58
 Public Libraries, 216
 Vordefinierte Kategorien, 145
 Vorlagen, 203
 Zusammenfassungsdiagramme, 60
- Extrahieren, 3–4, 80, 85–86, 209, 223, 255
 Ergebnisse, 27
 Ergebnisse speichern, 88
 Ergebnisse verfeinern, 9, 88
 Konzepte, Typen und Muster, 80
 Nicht linguistische Elemente, 87
 Uniterms, 5, 87, 256
 Wörter erzwingen, 95
- Extrahierungsergebnisse speichern, 88
- Extraktionsmuster, 252
- Fachausdrücke
 Abgleichsoptionen, 223
 Einschluss von Wörtern in Kategorien erzwingen, 167
 Erzwingen von Fachausdrücken, 231
 Farbe, 226, 269
 Gebeugte Formen, 223
 Hinzufügen zu japanischen Typen, 262
 Hinzufügen zum Ausschlusswörterbuch, 241
 Im Editor suchen, 212
 Zu Typen hinzufügen, 227
- Fachausdrücke und Typen suchen, 212
- Farben
 Ausschlusswörterbuch, 241
 Festlegen von Farboptionen, 21
 für Typen und Fachausdrücke, 226, 269
 für Zusammenfassungsdiagrammbalken, 63
 Synonyme, 238
- Frage mit geschlossenem Ende, 2
- Gebeugte Formen, 122, 223, 225–227, 269
 Gebeugte Formen erzeugen, 223, 225–227, 269
 Gemeinsame Projektbenutzung, 76
 Gewichte/Maße (nicht linguistisch), 246
 Globales Trennzeichen, 19
- Grafikelemente
 Ändern, 190
 Kollisionsmodifikatoren, 192
- Konvertieren, 190
 Typen, 190
- Häufigkeit, 127
- Hinzufügen
 Deskriptoren, 107
 Fachausdrücke zu japanischen Typ-Wörterbüchern, 262
 Fachausdrücke zu Typ-Wörterbüchern, 227
 Klänge, 21, 23
 Konzepte zu Kategorien, 161
 Liste der auszuschließenden Ausdrücke, 241
 Optionale Elemente, 238
 Public Libraries, 211
 Synonyme, 89, 236
 Synonyme für Japanisch, 269
 Typen, 91
- HTTP/URL (nicht linguistisch), 246
- IBM SPSS Data Collection, 35, 69
 Ändern der Datenquelle, 63
 Exportieren, 56
- IBM SPSS Statistics.sav-Dateien, 31, 65
 Ändern der Datenquelle, 63
 Ausgabeformat, 55
 Exportieren, 56
- ID-Variablen, 27–28, 35, 69
- Importieren
 Aktualisieren von Daten, 74
 Daten vorbereiten, 28
 Eingabedaten, 35, 69
 Neuimportieren von Daten, 63
 ODBC, 33, 67
 Public Libraries, 215
 Vordefinierte Kategorien, 136
 Vorlagen, 203
- Interpunktionsfehler, 86
- IP-Adressen (nicht linguistisches Element), 246
- Japanisch, 255, 260
 Resource Editor, 260
 Typ-Eigenschaften, 269
 Typen, 265, 271
- Kategoriebalkendiagramm, 172–173
- Kategorieerstellung, 6, 112, 116, 260
 Ausnahmen für Klassifizierungszusammenhänge, 120
 Konzeptinbeziehungsverfahren, 132
 Konzeptwurzelableitungsverfahren, 7, 118, 132
 Kookkurrenzregelverfahren, 7, 118, 132
 mithilfe von Verfahren, 7, 118
 Semantische Netze (Verfahren), 7, 118, 132
- Kategorien, 27, 96–97, 110, 159
 Anmerkungen, 111, 160
 Antworten erzwingen, 165
 Aufbauen, 7, 112, 116, 118, 121, 131
 Bearbeiten, 159, 162

- Bezeichnungen, 111, 160
- Deskriptoren, 105, 107, 111
- Drucken, 169
- Eigenschaften, 111, 160
- Ergebnisse verfeinern, 9, 159
- Erstellen, 104, 127, 135
- Erstellen neuer, leerer Kategorien, 134
- Erweitern, 121, 129
- Exportieren, 55
- Gemeinsamkeitsdiagramme, 172
- Hinzufügen zu, 161
- kopieren, 168
- Löschen, 170
- Manuelle Erstellung, 134
- names, 111, 160
- Netzdiagramm, 172
- Relevanz, 101–102
- Scoring, 99
- Strategien, 104
- Text Analysis Packages, 43, 45
- Umbenennen, 134
- Verschieben, 163
- Wörter erzwingen, 167
- Zusammenführen, 164
- Kategorienname, 98
- Kategorienbereich, 97
- Kategorienregeln, 148–149, 155, 157–159
 - Aus der Konzept-Kookkurrenz, 7, 118, 122, 126, 132
 - aus synonymen Wörtern, 7, 116, 118, 121–122, 129, 132
 - Beispiele, 155
 - Kookkurrenzregeln, 116, 121, 129
 - Syntax, 149
- Kategorisierung, 6, 96, 260
 - Häufigkeitsverfahren, 127
 - Konzepteinbeziehung, 116, 121, 123
 - Konzeptwurzelableitung, 116, 121–122
 - Kookkurrenzregeln, 116, 121, 126
 - Linguistische Verfahren, 112, 129
 - Manuell, 134
 - Methoden, 104
 - mithilfe von Verfahren, 121
 - Semantische Netze, 116, 121, 124
 - Verwenden von Gruppierverfahren, 116
- Kennzeichnen von Antworten, 76
- Klangoptionen, 23
- Kollisionsmodifikatoren, 190
- Kombinieren von Kategorien, 164
- Kompaktes Format, 142
- Komponentenbildung, 122
- Konzepte
 - beste Deskriptoren, 107
 - Erstellen von Typen, 88
 - Extrahieren, 80
 - Extrahierung erzwingen, 95
 - Hinzufügen zu Kategorien, 105, 111, 161
 - in Kategorien, 105, 111
 - Von der Extrahierung ausschließen, 94
 - Zu Typen hinzufügen, 91
- Konzepte ignorieren, 94
- Konzepteinbeziehungsverfahren, 7, 116, 119, 121, 123, 129
- Konzeptwurzelableitungsverfahren, 116, 121–122, 129, 132
- Kookkurrenzregelverfahren, 7, 116, 119, 121–122, 126, 129, 132
- Koordinatensysteme
 - transformieren, 188
- kopieren
 - Kategorien, 168
- Kopieren von Visualisierungen, 193
- Language Weaver, 24, 37, 72
- leere Antworten, 28
- Legende
 - Lage, 193
 - *.lib, 215
- Linguistische Ressourcen, 27, 209
 - Ressourcenvorlagen, 196
 - Text Analysis Packages, 43, 45
- Linguistische Verfahren, 3, 7–8, 118
- Link-Ausnahmen, 120
- Löschen
 - Ausgeschlossene Einträge, 241
 - Bibliotheken, 215, 217
 - Bibliotheken deaktivieren, 214
 - Kategorien, 170
 - Kategorienregeln, 159
 - Optionale Elemente, 239
 - Ressourcenvorlagen, 202
 - Synonyme, 239
 - Typ-Wörterbücher, 234
- Marken, 274
- Maximal zu erstellende Anzahl an Kategorien, 120
- Maximaler Suchabstand, 119, 125, 132
- Microsoft Excel.xls-/xlsx-Dateien, 32, 58, 66
 - Ändern der Datenquelle, 63
 - Ausgabeformat, 55
 - Exportieren von vordefinierten Kategorien, 145
 - Importieren vordefinierter Kategorien, 136
- Minimaler Zusammenhangswert, 120
- Mitgelieferte (Standard-) Bibliotheken, 209
- Muster, 80, 83
- natürliche Sortierung, 52
- Netzdiagramme, 172
- Neue Funktionen, 1
- Neue Kategorien, 134
- Neuheiten, 1
- Neuimportieren von Daten, 63
- NICHT (Regeloperator), 158
- Nicht kategorisiert, 98

- Nicht linguistische Elemente, 87
- Adressen, 246
 - aktivieren und deaktivieren, 250
 - Aminosäuren, 246
 - Datumsangaben, 246
 - E-Mail-Adressen, 246
 - Gewichte und Maßangaben, 246
 - HTTP-Adressen/URLs, 246
 - IP-Adressen, 246
 - Normalisierung, *NonLingNorm.ini*, 250
 - Proteine, 246
 - Prozente, 246
 - Reguläre Ausdrücke (*RegExp.ini*), 247
 - Sozialversicherungsnummer (USA), 246
 - Telefonnummern, 246
 - Währungen, 246
 - Zeitangaben, 246
 - Ziffern, 246
- Normalisierung, 250
- ODBC-Datenquelle, 33–34, 63, 67–68
- ODER (Regeloperator), 158
- Offene Frage, 2
- Öffnen von Projekten, 49
- Operatoren in Regeln & |() , 158
- Opinions Library, 224
- Optionale Elemente, 234
- Definition von, 236
 - Hinzufügen, 238
 - Löschen von Einträgen, 239
 - Ziel, 238
- Optionen, 19
- Anzeige, 21
 - Klang, 23
 - System, 19
 - Übersetzung, 24
- Organisationstyp-Wörterbuch, 224
- Paletten
- Anzeigen, 178
 - Ausblenden, 178
 - Verschieben, 178
- Part of Speech, 252
- Permutationen, 87
- Personentyp-Wörterbuch, 224
- Pluralformen, 226
- Polarkoordinaten, 188
- Produkttyp-Wörterbuch, 224
- Projekte, 27, 49
- Auswählen von Kategorien und Ressourcen, 39
 - Datenquelle, 30, 64
 - Eigenschaften, 50
 - Erstellen, 29
 - Gemeinsam benutzen, 76
 - öffnen, 49
 - Optionen für Bibliotheken, 19
 - Speichern, 53
- Statusleiste, 77
 - Text Analysis Packages, 39
 - Übersetzung, 37, 72
 - Umbenennen, 53
 - Variablenauswahl, 35, 69
 - Wiederverwenden von Kategorien, 168
- Proteine (nicht linguistisches Element), 246
- Prozentsätze (nicht linguistisches Element), 246
- Rechtliche Hinweise, 273
- Rechtschreibfehler, 28, 86, 245
- Referenzvariablen, 27–28, 35, 69
- Regeln
- Bearbeiten, 159
 - Boole'sche Operatoren, 158
 - Erstellen, 157
 - Kookkurrenzregelverfahren, 126
 - Löschen, 159
 - Syntax, 149
- Relevanz der Antworten und Kategorien, 101–102
- Ressourcen
- Bearbeiten von erweiterten Ressourcen, 242
 - Mitgelieferte (Standard-) Bibliotheken, 209
 - Sichern, 205
 - Wechseln von Vorlagenressourcen, 201
 - Wiederherstellen, 205
- Ressourcen durch eine Vorlage ersetzen, 201
- Ressourcen sichern, 205
- Ressourcen wiederherstellen, 205
- Ressourceneditor, 12, 200–201, 242
- Aktualisieren von Vorlagen, 200
 - Erstellen von Vorlagen, 200
 - für Japanisch, 260
 - Globales Trennzeichen, Option, 19
 - Wechseln von Ressourcen, 201
- Ressourcenvorlagen, 4, 196, 256
- Schriftfarbe, 226, 269
- Score (Schaltfläche), 99
- Scoring, 99
- Sekundäre Analyse
- Abhängigkeitsanalyse, 257–258
 - Stimmungsanalyse, 257–258
- Semantische Netze (Verfahren), 7, 116, 118, 121–122, 124, 129, 132
- Sortieren von Daten und Variablen, 52
- Sozialversicherungsnummer (USA) (nicht linguistisch), 246
- Spalten im Datenbereich anzeigen, 100
- Spalten im Kategoriebereich anzeigen, 98
- Spaltenumbruch, 21
- Speichern
- Automatische Speicherung von Projekten, 19
 - Extrahierungsergebnisse, 88
 - Projekte, 53
 - Ressourcen, 205
 - Ressourcen als Vorlagen, 200

- Standardbibliotheken, 209
 Standorttyp-Wörterbuch, 224
 Stapeln, 193
 Statistik
 Bearbeiten in Visualisierungen, 190
 Beschreibungen, 191
 Statusleisten, 77
 Sternchen (*)
 Ausschlusswörterbuch, 241
 Synonyme, 237
 Stimmungsanalyse, 257–258
 Optionen, 258
 Streuen, 193
 Stummschalten der Klänge, 23
 Substitutionswörterbuch, 209, 234, 238–239
 Suchen und Ersetzen (erweiterte Ressourcen), 243–244
 Synonyme, 88, 234
 ! ^ * \$ Symbole, 237
 Ausnahmen für unscharfe Gruppierung, 86, 245
 Definition von, 235
 Farben, 238
 für japanischen Text, 269
 Hinzufügen, 89, 236, 269
 Löschen von Einträgen, 239
 Zielausdrücke, 236, 269
- Tabelle für Netzdiagramm, 172
 *.tap Text Analysis Packages, 39, 43–45, 48
 Telefonnummern (nicht linguistisch), 246
 Text Analysis Packages, 39, 43–45
 Text-Mining, 3
 Textanalyse, 3, 8, 12
 Texttrennzeichen, 19
 Textübereinstimmung, 111, 160, 167
 Textvariablen, 27–28, 35, 69
 Titel für das Exportzusammenfassungsdiagramm, 62
 TLA, 201
 Trennzeichen, 19
 Typ-Wörterbuch, 209
 Deaktivieren, 234
 Erstellen von Typen, 225, 269
 Erzwingen von Fachausdrücken, 231
 Hinzufügen von Fachausdrücken, 227
 Hinzufügen von Fachausdrücken für Japanisch, 262
 Integrierte Typen, 224
 Löschen, 234
 Optionale Elemente, 223
 Synonyme, 223
 Umbenennen, 232
 Verschieben, 233
 Typen, 223
 Erstellen, 225, 269
 Extrahieren, 80
 für Japanisch, 265, 269, 271
 Im Editor suchen, 212
 Integrierte Typen, 224
 Konzepte hinzufügen, 88
 Standardfarbe, 21, 226, 269
 Typenhäufigkeit, 127
 Wörterbücher, 209
 Typenhäufigkeit, 127
- Übersetzung, 24, 74
 Einstellung “Übersetzungsgenauigkeit”, 38, 73, 75
 Optionen, Dialogfeld, Übersetzung, Registerkarte, 24
 Übersetzen ins Englische, 37, 72, 74
 Übersetzungseinstellungen, Dialogfeld, 74
 Überwachen, 76
 Umbenennen
 Bibliotheken, 214
 Kategorien, 134, 160
 Projekte, 53
 Ressourcenvorlagen, 202
 Typ-Wörterbücher, 232
 Umfragedaten, 2, 27–28, 35, 69
 UND (Regeloperator), 158
 Uniterms, 87
 Untersuchungsmodus, 175
- Variablen
 Abgleichen, 71
 Aktualisieren, 74
 Ändern der Datenquelle, 63
 Bearbeiten von Eigenschaften, 52
 Exportieren, 55
 ID-Variablen, 27–28, 35, 69
 Importieren, 31, 33, 65, 67
 Referenzvariablen, 27–28, 35, 69
 Textvariablen, 27–28, 35, 69
- Verfahren
 Häufigkeit, 127
 Konzepteinbeziehung, 116, 121, 123, 129
 Konzeptwurzelaableitung, 116, 121–122, 129
 Kookkurrenzregeln, 116, 121, 126, 129
 Semantische Netze, 116, 121, 124, 129
 Ziehen und Ablegen, 135
 Verfolgen von Antworten, 76
 Veröffentlichen, 53, 219
 Bibliotheken, 217
 Public Libraries hinzufügen, 211
 Verschieben
 Kategorien, 163
 Typ-Wörterbücher, 233
 Verwalten
 Kategorien, 159
 Local Libraries, 213
 Public Libraries, 215
 Visualisierungen
 Abstand, 182
 Achsen, 184
 Bearbeiten, 176
 Farben und Muster, 179
 Fenster, 186, 188

- Kategorien, 186
- kopieren, 193
- Legendenposition, 193
- Punktform, 181
- Punktrotation, 181
- Punktverhältnis, 181
- Ränder, 182
- Skalen, 184
- Strichmuster, 179
- Text, 179
- Transformieren von Koordinatensystemen, 188
- Transparenz, 179
- transponieren, 186, 188
- Zahlenformate, 183
- Visualisierungsbereich
 - Diagramme aktualisieren, 172
 - Kategorienetzdiagramm, 172
- Vorbereiten Ihrer Daten, 28
- Vordefinierte Kategorien, 136, 145
 - Einfaches Listenformat, 141
 - Eingerücktes Format, 144
 - Kompaktes Format, 142
- Voreinstellungen, 19, 21, 23
- Vorlagen, 4, 196, 256
 - Aktualisieren und speichern als, 200
 - Erstellen über Ressourcen, 200
 - Importieren und Exportieren, 203
 - Löschen, 202
 - Sichern, 205
 - TLA, 201
 - Umbenennen, 202
 - Vorlagen wechseln, 201
 - Wiederherstellen, 205

- Währungen (nicht linguistisches Element), 246
- Wichtig-Flag, 76
- wiederhergestellte Dateien, 19
- Wiederverwenden
 - Kategorien, 168
- Winkelzeichen (^), 237
- Wörterbuch der negativen Typen, 224
- Wörterbuch der positiven Typen, 224
- Wörterbuch der unbekanntenen Typen, 224
- Wörterbuch der ungeklärten Typen, 224
- Wörterbücher, 16, 223
 - Ausschlüsse, 209, 223, 240
 - Substitutionen, 209, 223, 234
 - Typen, 209, 223

- `.xls-/.xlsx`-Dateien, 55, 58

- Zeitangaben (nicht linguistisches Element), 246
- Ziehen und Ablegen, 135
- Zielausdrücke, 238
- Ziffern (nicht linguistisches Element), 246
- Zusammenfassungsdiagramme, 60
- Zusammenführen von Kategorien, 164
- Zuverlässigkeit, 8
- Zuweisen von Flags, 76