

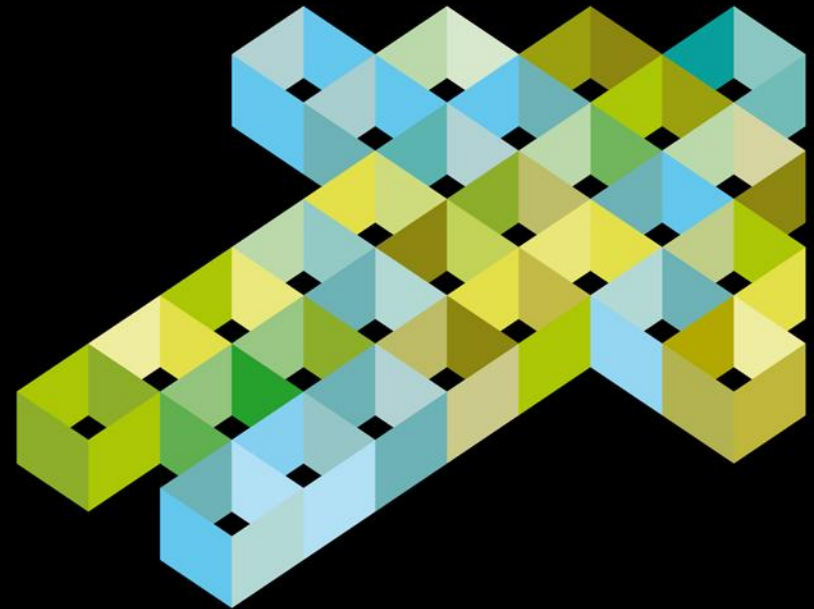
# Vom Content zu Wissen - IBM Content Analytics und Suche

Wir schlagen die Brücke vom System Watson über das Quiz Jeopardy hinaus in reale Kundensituationen mit dem Produkt IBM Content Analytics

Andreas Schneider

Uschi Bosch

01.Juni 2012



# IBM ECM Portfolio Übersicht

*Nutzen Sie den Wert unstrukturierter Informationen!*



# Agenda

- Die Herausforderungen der Informationsflut
- IBM Watson
- Ausgewählte Kundensituationen und Szenarien mit IBM Content Analytics
- Lizenzierung, Preise, Services
- Zertifizierung und Ausbildung



# Die Herausforderungen der Informationsflut



***“Es scheint, als ob Sie alle Daten hätten – was ist also das Problem?!”***





# Agenda

- Die Herausforderungen der Informationsflut
- **IBM Watson**
- Ausgewählte Kundensituationen und Szenarien mit IBM Content Analytics
- Lizenzierung, Preise, Services
- Zertifizierung und Ausbildung



# IBM Watson – Die Herausforderung



-  Am 16. Februar 2011 schrieb das IBM Watson System Geschichte
-  Watson fordert uns Menschen im eigenen Terrain heraus – der Quizshow Jeopardy!
-  Watson gewinnt gegen Ken Jennings und Brat Rutter - die erfolgreichsten Teilnehmer die jemals bei Jeopardy! mitgespielt haben
-  *Watson kann schneller und besser antworten als 26'000 US Bürger*





# IBM Watson – ein System was Antworten auf Fragen gibt

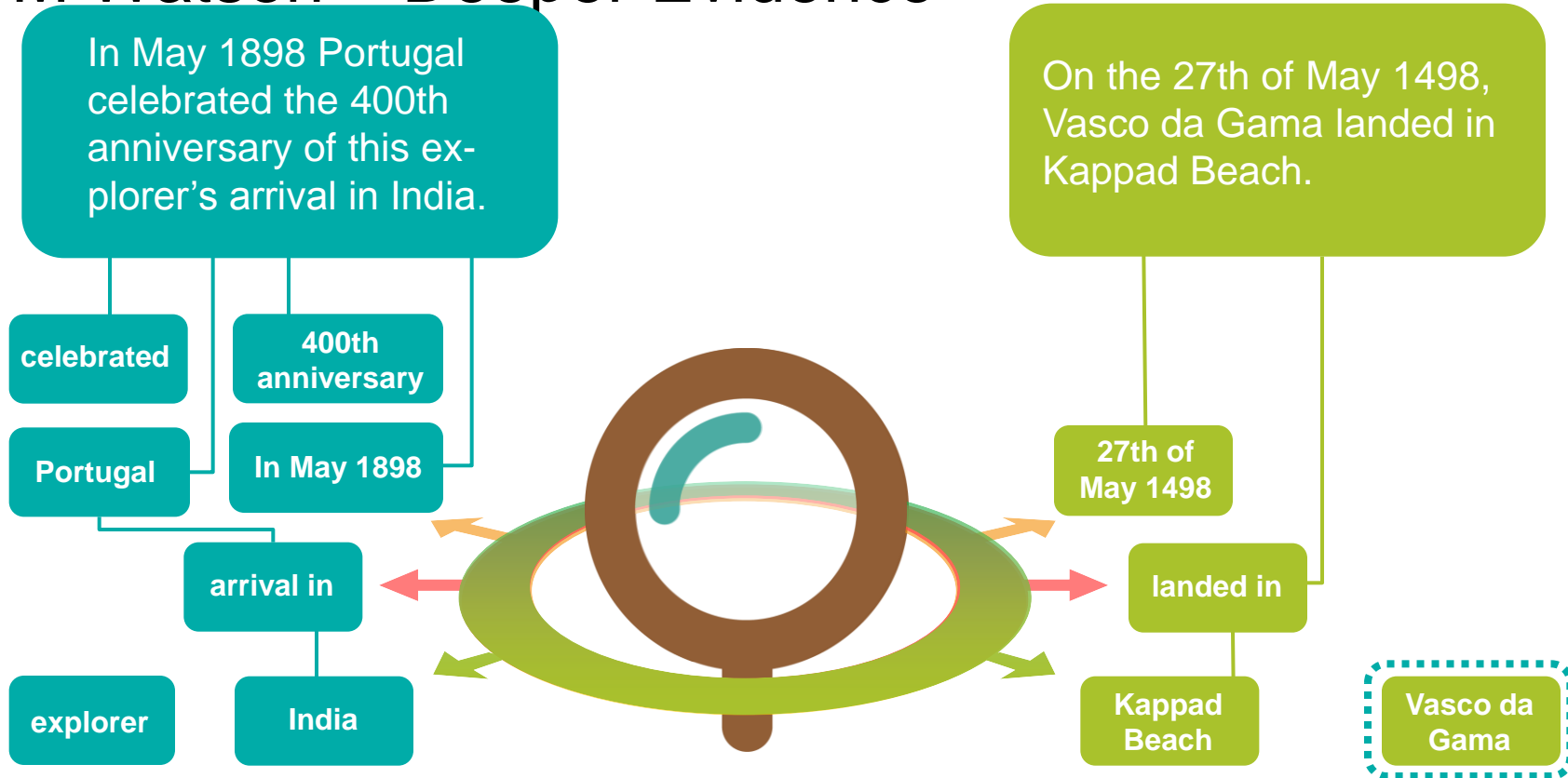


Selbst wenn Fragen:

- indirekt
- hochgradig kontextabhängig
- doppeldeutig und
- unscharf sind?

Was wäre, wenn ein System **nicht nur Zugriff** auf alle Informationen hätte, **sondern auch Antworten** auf unsere Fragen geben könnte?

# IBM Watson – Deeper Evidence



Date Math



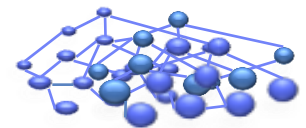
Temporal Reasoning

Phrases



Statistical Phrasing

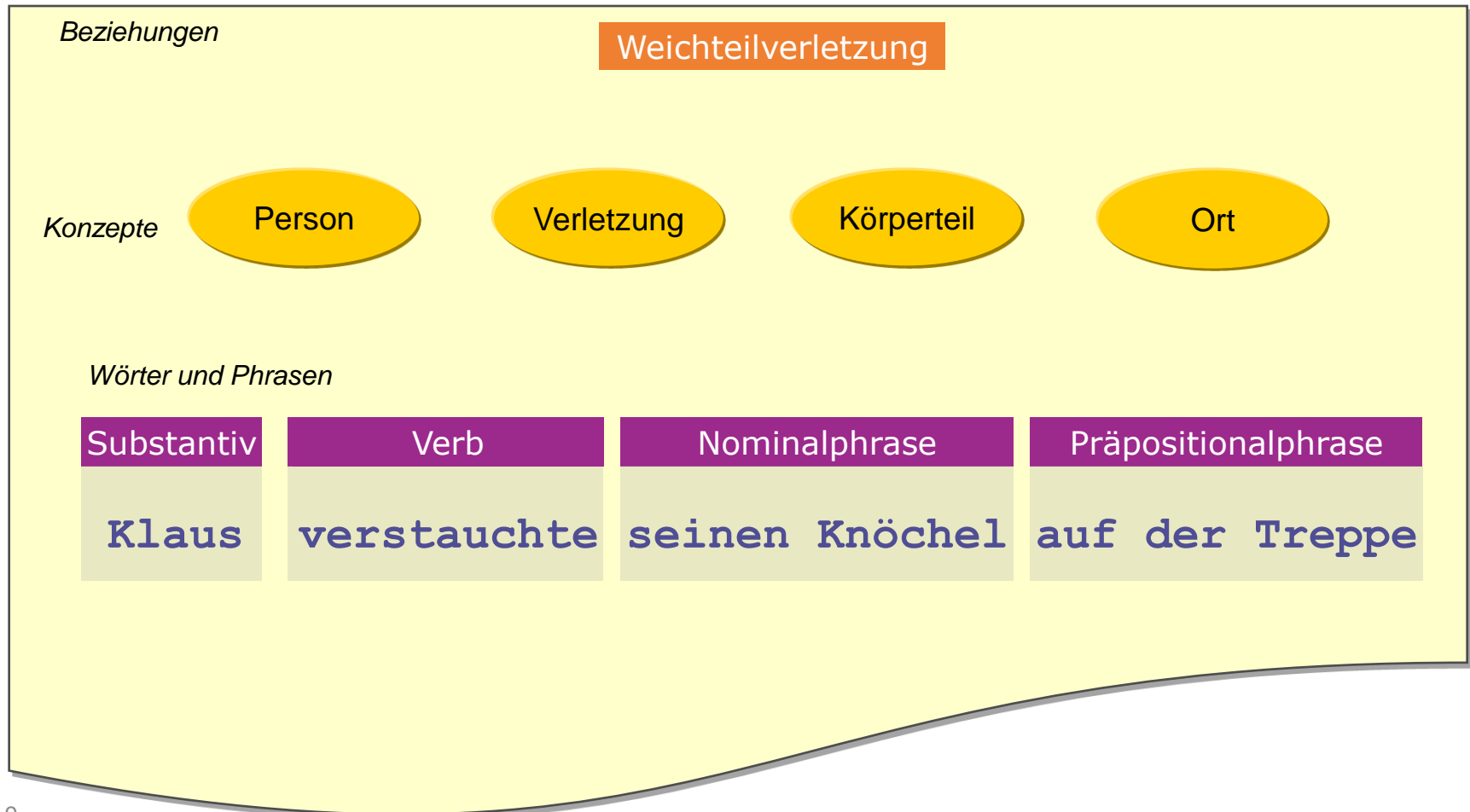
Geo



Geospatial Reasoning



# IBM Watson – UIMA und LanguageWare Textanalyse



# Wie kommt die Intelligenz in den Text?

Spracherkennung ▶ Segmentierung ▶ Normalisierung ▶ Anreicherung ▶ Wörterbücher ▶ Regeln

- deutsch
  - Einzelne Wörter:
    - z.B. Ford
  - Sätze:
    - z.B. Nun ist aber die Klimaanlage kaputt.
- Wagen = Auto
- gekauft = kaufen
- Reperatur = Nomen
- kaufen = Verb
- Fachwörter:
  - z.B. Klimaanlage, Topf
- Defekte:
  - z.B. Klimaanlage ist kaputt, Topf, der das Kondenswasser auffängt defekt

Hallo,  
 ich habe vor etwa 5 Monaten eine gebrauchten Ford Focus beim Händler gekauft. Nun ist  
 aber die Klimaanlage kaputt. Ich habe den Wagen erst zum nächsten Fordhändler gebracht.  
 Er hat dann auch bemerkt, dass der Topf, der das Kondenswasser auffängt defekt ist.  
 Summar summarum würde es sich auf 313 Euro belaufen. Nun ist meine  
 Frage: Wäre diese Reparatur auch in der Gebrauchtwagengarantie enthalten, oder bräuchte  
 ich es erst gar nicht versuchen?

# IBM Watson – Die Brücke zu ECM Analytics (1)

- Unstructured Information Management Architecture (UIMA)
  - Wide range of analytic algorithms, e.g.
    - Paraphrasing
    - Temporal Reasoning
    - Geographical Reasoning
  - **IBM Content Analytics**
  - IBM InfoSphere BigInsights
- 
- IBM POWER 750
    - 15 Terabyte RAM
    - 2880 processor cores
    - 80 TeraFLOP/s



**IBM Watson**



# IBM Watson – Die Brücke zu ECM Analytics (2)

Von IBM Watson genutzt

## ECM Analytics

Natural Language Processing  
and content analysis  
leveraging UIMA

- Natural Language Processing
- Aufbereitung der Daten zur Analyse
- Wo sind Korrelationen?
- Drill-Down in Texten

## IBM Content Analytics



## IBM Watson – Die Brücke zu ECM Analytics (3)

- Kernkomponenten von Watson sind seit Jahren in ECM Produkten enthalten (Lucene Index, UIMA Textanalyseframework, LanguageWare Workbench zur Entwicklung von Wissensbausteinen)
- Weitere werden folgen (Hadoop für große Datenmengen)
- ECM Analyticsprodukte
  - IBM OmniFind
  - **IBM Content Analytics (ICA)**
  - eDiscovery Analyzer (eDA)
  - IBM Classification Module (ICM)
  - UIMA (mit ICA/ OF/ eDA integriert, gebündelt)
  - LanguageWare Workbench (mit ICA/ OF integriert, gebündelt)

# IBM Watson – Die Brücke zu ECM Analytics (4)

## *Mitbewerber (intern und extern)*

- IBM Content Analytics (ICA im Vergleich zu anderen Herstellern
  - EMC Intelligence Service ist nur Klassifizierung von XML Dateien
  - EMC xPlorer ist reine Index-/ Suchmaschine
  - OpenText Nstein: erstes Analytics Play beschränkt sich auf Web Content Management
  - Autonomy IDOL: schwer zu implementieren, bekannte Skalierungsprobleme, kein dynamisches UI
  - Google Search Appliance: Fokus auf Webseiten und Suche
  - Microsoft FAST: Fokus auf Suche
- IBM: SPSS/ Cognos
  - Fokus auf strukturierte Daten in DBs
  - Sehr gut erweiterbar um ICA: ICA liefert Inhalte der unstrukturierten Daten zu
  - Cognos ist in ICA zur Generierung von Reports integriert
- IBM: Cognos Consumer Insights (CCI)
  - Unterstützt „nur“ den UseCase „Voice of the Customer“ auf Internetdaten vom Dienstleister Boardreader
  - Analysiert keine internen Daten
  - Kann sich mit ICA ergänzen, wenn interne Daten wichtig sind
- Nischenanbietern/ Agenturen
  - Verarbeiten meist keine internen Daten, Fokus auf Web
  - Keine Integration in Kundenprozesse
  - Meist kein anpassbares, dynamisches UI

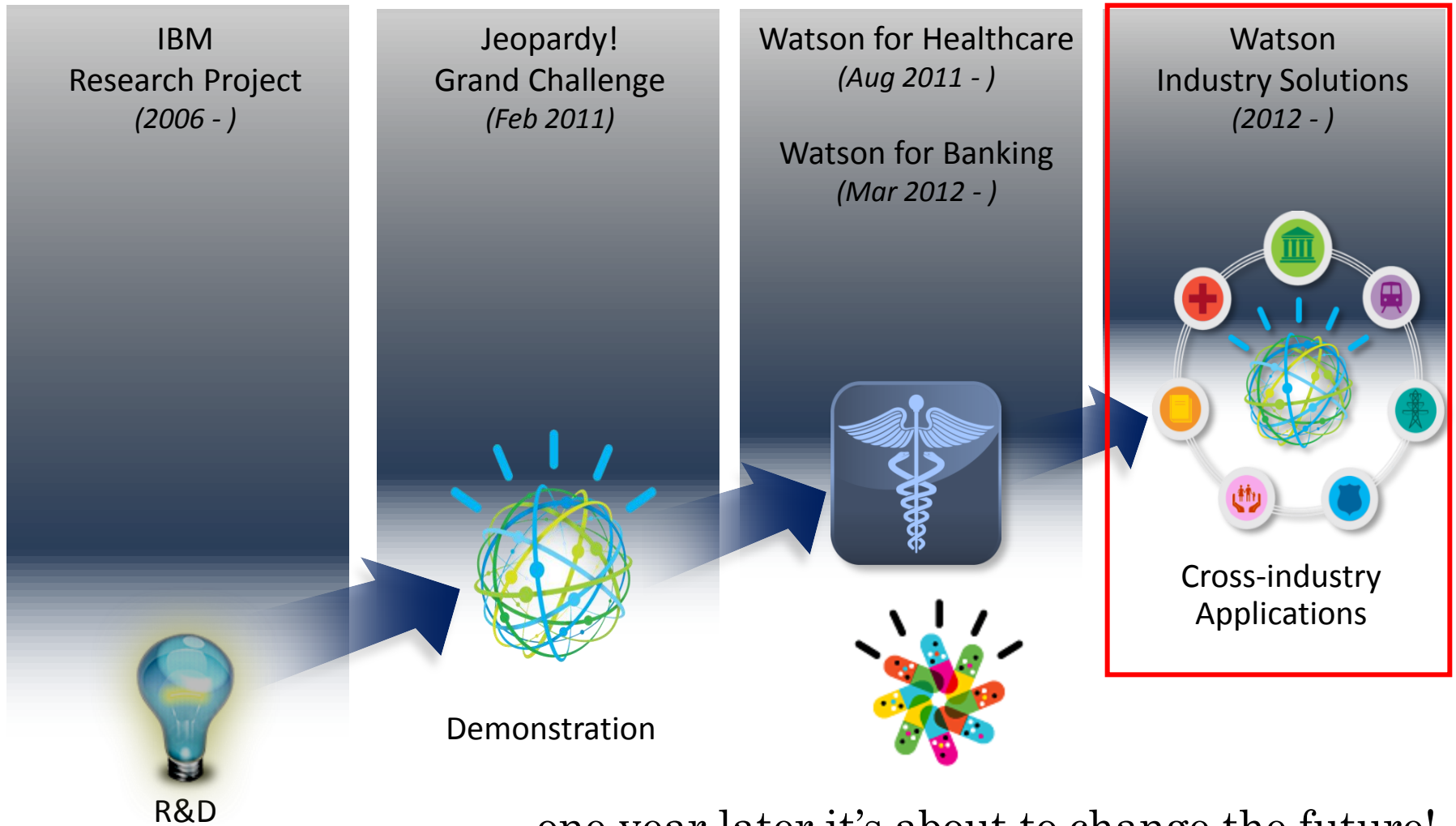
# Agenda

- Die Herausforderungen der Informationsflut
- IBM Watson
- **Ausgewählte Kundensituationen und Szenarien mit IBM Content Analytics**
- Lizenzierung, Preise, Services
- Zertifizierung und Ausbildung





# Ausgewählte Kundensituationen und Szenarien



... one year later it's about to change the future!

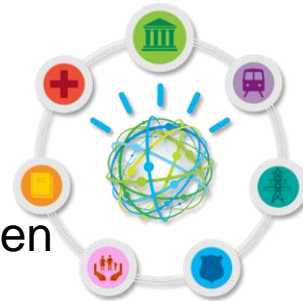
# Ausgewählte Kundensituationen und Szenarien

*Aktuelle Kunden, Opportunities und Projekte mit IBM Content Analytics (Auszug)*

- **Pressemonitoring, Wissensmanager für Steuerberater & Rechtsanwälte**
  - Erkennung der Kundenmeinung (Sentiment) und Zitate-Ranking
  - Strukturierung/ Wissensanreicherung von Webinhalten → neuer Dienst für Kunden
  - Erkennen von Kunden, die von Gesetzesänderungen betroffen sind
- **Banken und Versicherungen**
  - Service-Center: schnellere und hochwertigere Beratung
  - Up-/ Cross Selling: z.B. spezifische Vertragskonditionen eines Kunden erkennen
- **Verteidigung/ Öffentliche Hand**
  - Automatische Erkennung hoch sensibler Daten
  - Analyse der Gefahrenlage in Krisengebieten
- **Automobilbranche**
  - Fahrzeugannahme: schnelles Finden der Lösung (z.B. „vorne rechts klapperts“)
  - Sprachsteuerung: Funktionsauswahl auf Basis des Fahrerkommandos (z.B. „mir ist kalt“)

# Ausgewählte Kundensituationen und Szenarien

- **Marketing / Vertrieb**
  - Auswertung von Kundenaussagen
  - Bewertung von Aussagen über das eigene Unternehmen & Produkte
  - Erkennen von neuen Themen & Trends
- **Einkauf**
  - Vertragsmanagement (Finden nicht mehr benötigter / ähnlicher Verträge, ...)
  - Auswertung von Angeboten & Aufträgen
- **Personalwesen**
  - Auswertung von Bewerberprofilen (gezielte Suche nach Skills)
  - Auswertung von Stellenbörsen (welche Skills werden am Markt nachgefragt)
- **Entwicklung**
  - Frühwarnung bei Qualitätsproblemen
- **Projektgeschäft**
  - Suche nach ähnlichen Projekten, Angeboten & Kostenkalkulationen
  - Suche nach Spezialisten zu Projekten & Themen
- **Betrugserkennung**



# Ausgewählte Kundensituationen und Szenarien

## *Beispiel 1 aus der Automobilbranche*

Aufbau eines **Decision Support Systems** im Automotivesektor auf Basis von IBM Content Analytics, welches...

- zu einem Serviceanliegen passende Lösungsdokumente suchen und vorschlagen kann
- sinnvolle Hinweise und Empfehlungen anzeigt
- Auswertungen über die Reparaturfälle im Backend ermöglicht

### **Dies führt zu...**

- einem schnelleren Auffinden des eigentlichen Problems
- einer größeren Kundenzufriedenheit
- einer Reduktion der Standzeiten und damit einhergehenden Kosten in einer Werkstatt



# Ausgewählte Kundensituationen und Szenarien

## *Beispiel 2 aus der Automobilbranche*

Aufbau eines **Quality Management/Early Warning Systems** im Automotivesektor mit IBM Content Analytics, welches...

- schnellstmöglichst das Erkennen von Qualitätsproblemen erlaubt
- Ursachenforschung durch Einzoomen auf Problembereiche ermöglicht
- automatisiert Schlüsselbegriffe im Kontext von Qualitätsproblemen automatisch anzeigt

**Dies führt zu...**

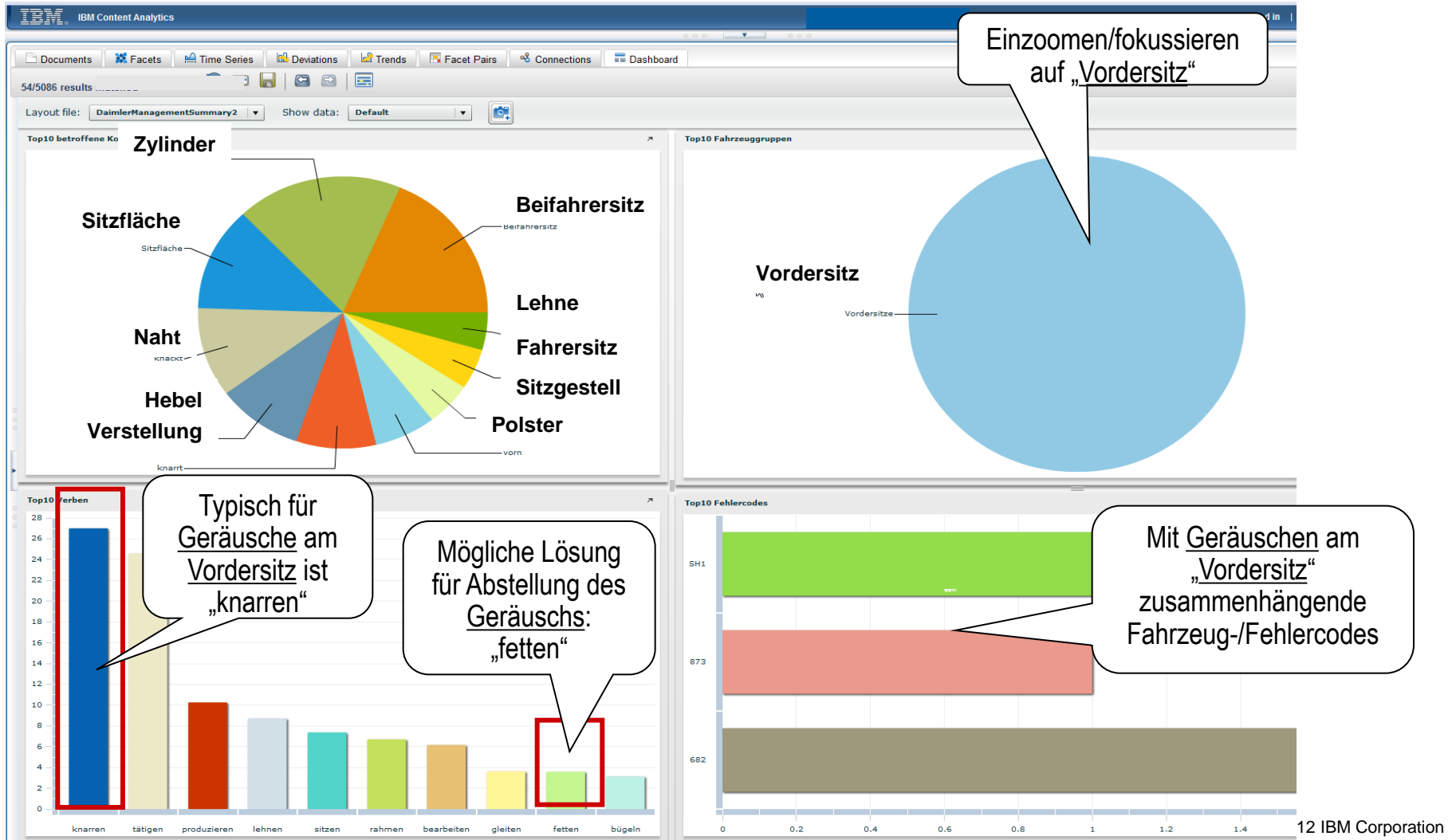
- Vermeidung teurerer Rückrufaktionen
- frühzeitigem Erkennen von Qualitätsproblem um Gegenmaßnahmen einzuleiten
- einer größeren Kundenzufriedenheit





# Ausgewählte Kundensituationen und Szenarien

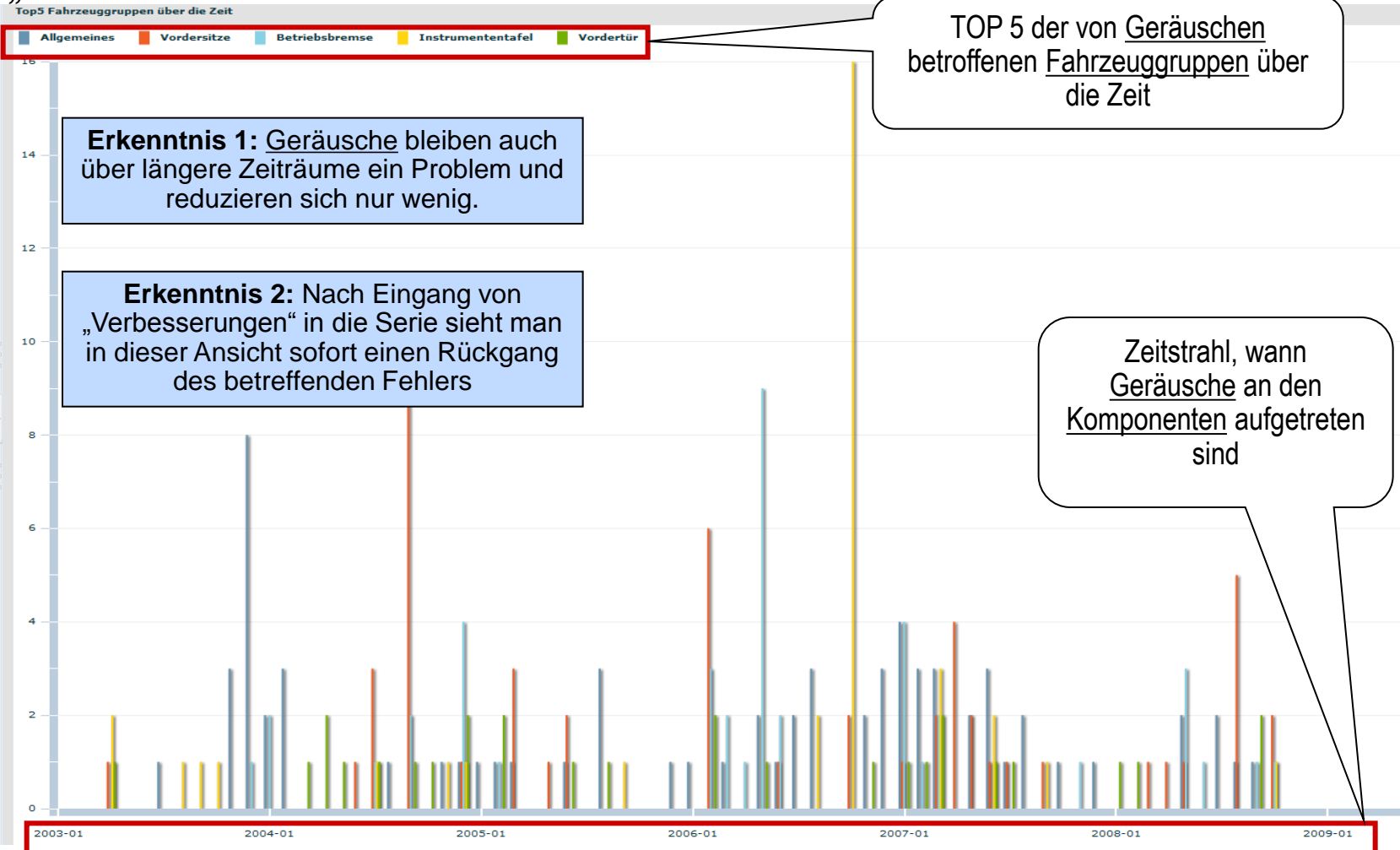
## Ursachenforschung und Lösung für „Geräusch“ am „Vordersitz“





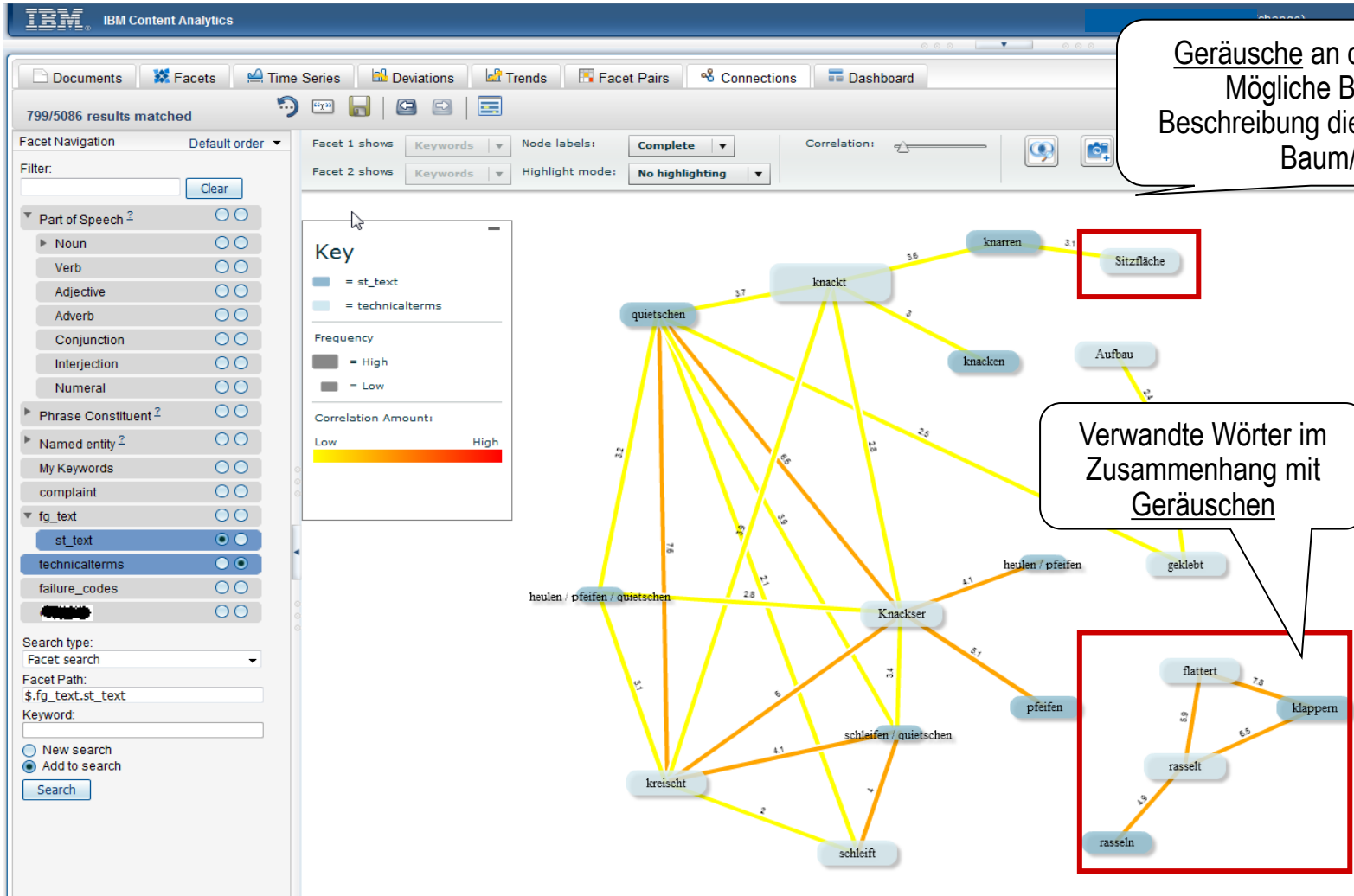
# Ausgewählte Kundensituationen und Szenarien

*Verbesserungsüberwachung: reduzieren sich die Fehler im Zusammenhang mit „Geräuschen“?*

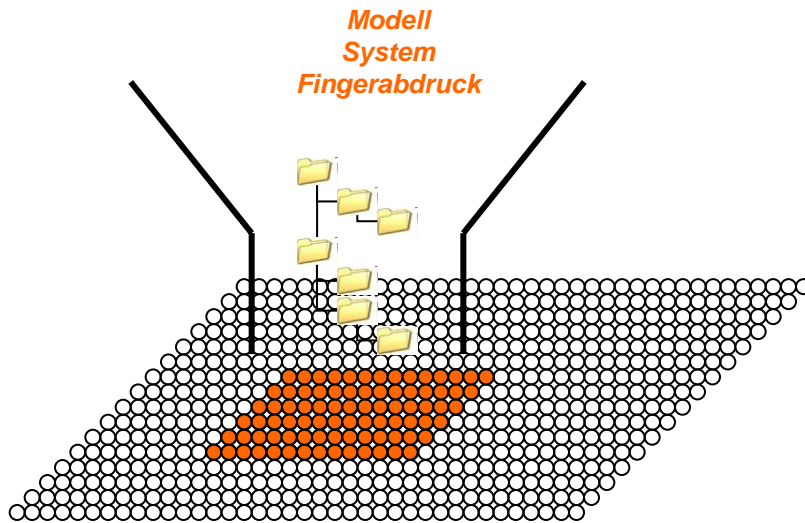


# Ausgewählte Kundensituationen und Szenarien

## Abhängigkeiten/ Verwandtschaften im Zusammenhang mit „Geräuschen“



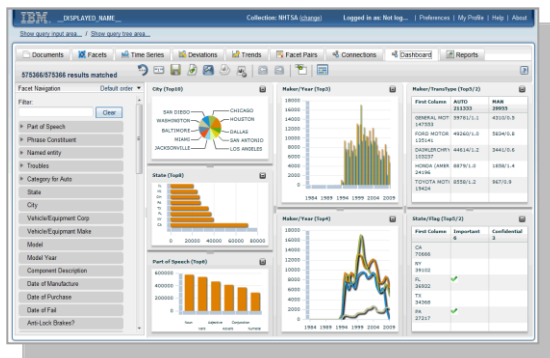
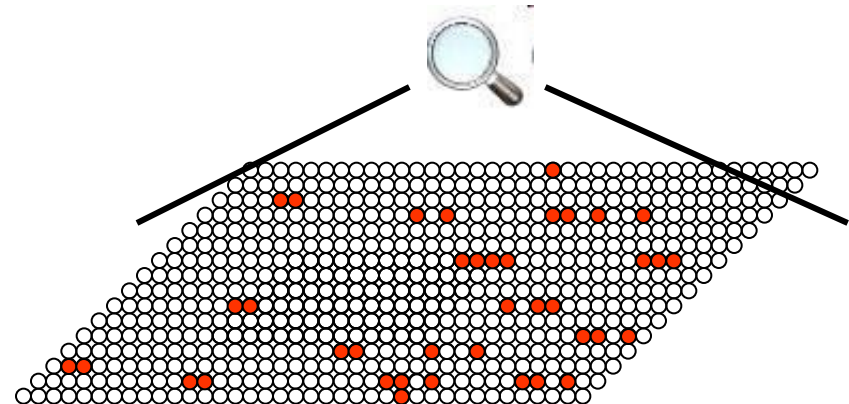
Neuer Zugriff auf Informationen  
Strukturiert – Unstrukturiert – Kombiniert



- Effektive Suchen
  - Zusammenhänge erkennen
  - Unterstütztes Finden
  - Slice & Dice parallel zur Suche
  - Einfach & Schnell
  - ...
- → Die richtigen Lösungen zu meinen Symptomen



**„Klimaanlage muffig“**



# Ausgewählte Kundensituationen und Szenarien

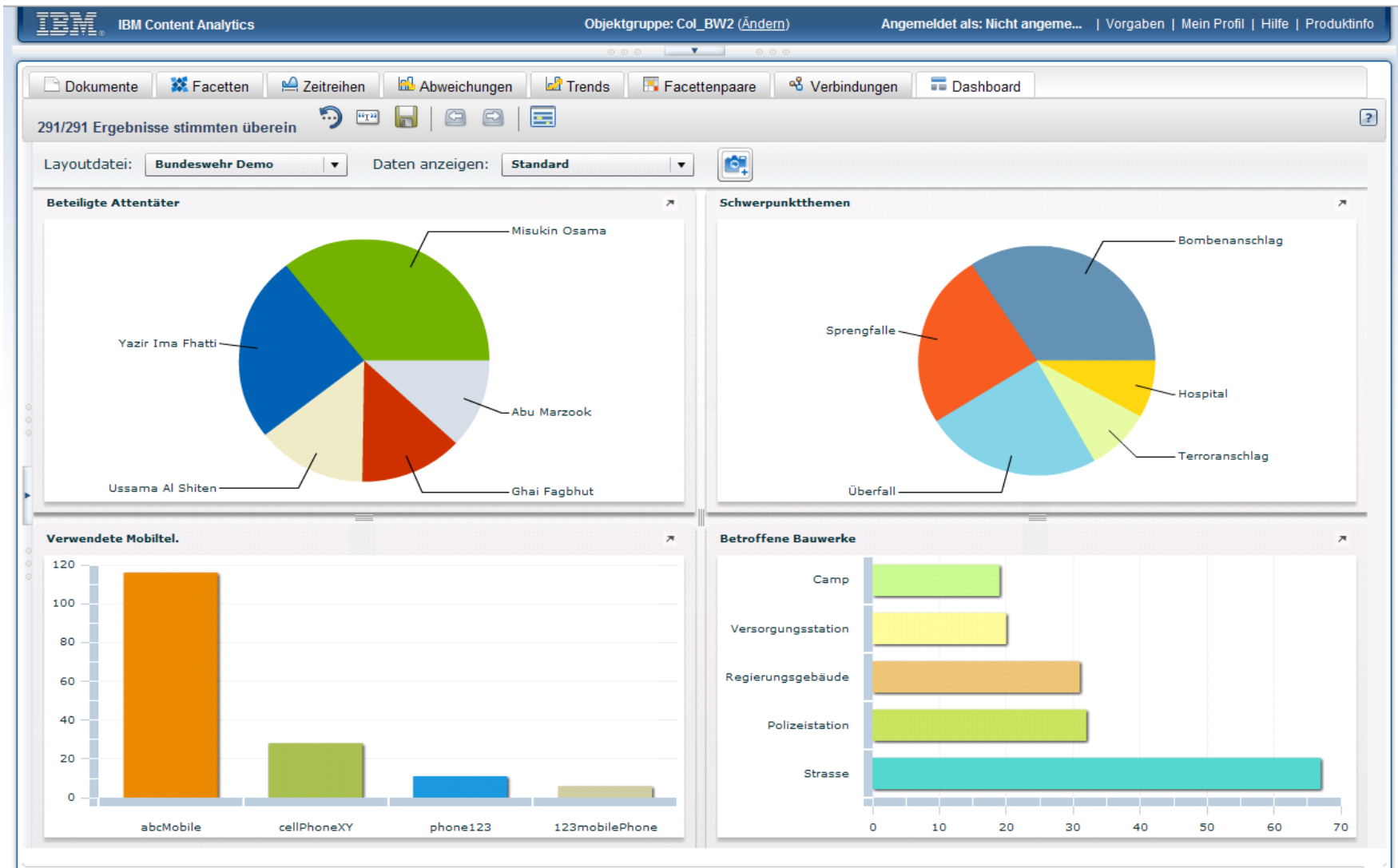
## *Beispiel aus dem Militär*

- **Welche Daten wurden analysiert?**
  - Aufklärungsberichte aus Krisenregionen
- **Was wurde in den Daten analysiert?**
  - Abweichungsanalysen im Zeitverlauf um das Verhalten von Attentätern zu beobachten
  - Korrelationsanalysen von wichtigen Zusammenhängen bzgl. Anschlagsarten, Attentätern, verwendeten Hilfsmitteln (z.B. Mobiltelefon), betroffene Regionen und Städte
  - Auffinden von Zusatzinformationen in den Texten, wie Frequenzbereiche und Geokoordinaten
- **Welche Erkenntnisse konnten gewonnen werden?**
  - Management summary auf Knopfdruck (z.B. über die Gefahrenlage in einer Region)
  - Schaffen eines Verständnisüberblicks über unbekannte, militärische Daten
  - Finden von Detailinformationen mit regionalem Bezug über Geoinformationen
  - Einschätzen der Gefahrenlage in Wiederaufbauregionen



# Ausgewählte Kundensituationen und Szenarien

## Management Summary: Aktuelle Gefahrenlage



# Kundensituationen und Szenarien

Analysebausteine liefern Anhaltspunkte zum Dokumenteninhalt

The screenshot shows the IBM Content Analytics search interface. At the top, a search bar contains the query `":*:* url:demodaten"`. Below the search bar, a list of search results is displayed, each with a document title, date, and a snippet of text. On the left side, a 'Facettenstruktur' (Facet Structure) panel is visible, showing various categories and their counts. On the right side, there are three map thumbnails showing geographical locations like Tyros and Kabul.

**Callout 1:** Wir wissen noch nicht, was in den Texten steht → Suche ohne Einschränkung

**Callout 2:** Analysebausteine: Verarbeitung von unstrukturierten Texten, als wären es Metadaten in einer Datenbank

**Callout 3:** „Kategorie/Themenauswahl“ zu den Dokumenten und Anzahl

**Callout 4:** Wenn Städte im Dokument genannt sind, erfolgt eine Kartendarstellung

**Callout 5:** Welche Städte sind betroffen?



# Kundensituationen und Szenarien

## Verarbeitung von Geoinformationen

Ergebnisse 1-10 von 291  
(291 Ergebnisse stimmten überein)

Sie suchten nach: "":\* url:demodaten" /

Ihre Ergebnisse eingrenzen

Facettenstruktur

- ▼ Dokumente mit Schutzzinhalten
  - schutzwürdig(291)
  - ISAF(95)
  - geschützt(95)
  - 69° 10' 0" E.(77)
  - 34° 32' 0" N(77)
  - 33° 20' 0" N(24)
  - 44° 23' 0" E.(24)
  - 34° 57' 59.99" N(19)
  - 69° 17' 33.99" E.(19)
  - 35° 0' 36" N(15)
  - 69° 10' 12" E.(15)
  - 42° 26' 0" E.(14)
  - 11° 48' 0" N(14)
  - 68° 38' 30" E.(13)
  - 34° 26' 0" N(13)
- Mehr...
- ▼ Dokumentenstatus
  - schutzwürdig(291)
  - geschützt(95)
- ▶ Erstelldatum
- ▶ Frequenz\_MHz\_per
- ▶ Koordinaten\_Breitengrad
- ▶ Koordinaten\_Laengengrad
- ▶ Koordinaten\_per\_Breitengrad\_OSM

B: 34,533€ L: 69,166€

0km 100km 200km 300km

Nach Dateityp anzeigen: 10 - ALLE -

Ergebnisse pro Seite: 10 - ALLE -

Darstellung von Städten über ihre Geokoordinaten

Kartendarstellung von Städten

Kabul

Kabul International Airport

Bagrami

© 2012 IBM Corporation



# Ausgewählte Kundensituationen und Szenarien

*Anwaltskanzlei Baker & McKenzie (Smarter Planet Referenz)*

- **Welche Daten wurden analysiert?**
  - Gerichtsakten
- **Was wurde in den Daten analysiert?**
  - Auffinden von Schlüsselpersonen
  - Wer kommuniziert mit wem?
  - Abhängigkeiten zwischen Orten und kriminellen Aktivitäten
- **Welche Erkenntnisse konnten gewonnen werden?**
  - Schaffen eines Überblicks über unbekannte, gerichtliche Daten
  - Finden von Detailinformationen über Personen, Orte und kriminelle Handlungen
  - Graphische Aufbereitung der Daten für gerichtliche Verfahren



# Zusammenfassung

- **ICA Zielgruppen:** IT & Fachbereiche, die unstrukturierte Daten in Geschäftsentscheidungen miteinbeziehen
  - Analyse, die über das klassische BI hinausgeht
  - Analyse, die über SUCHE hinausgehen
  - Semantische Analyse von Text
- **Herausforderung:** Neue Erkenntnisse gewinnen, nach denen man evtl nicht gesucht hat: Trends, Muster, Beziehungen, Zusammenhänge
  - Produkt-/Servicequalität, Fehlererkennung
  - Kundenmeinung, Kundenzufriedenheit
  - Compliance & Haftung, Betrugserkennung

# Agenda

- Die Herausforderungen der Informationsflut
- IBM Watson
- Ausgewählte Kundensituationen und Szenarien mit IBM Content Analytics
- Lizenzierung, Preise, Services
- Zertifizierung und Ausbildung



# Preise & Lizenzierung

- **Lizenzierung nach PVU**
  - **D0IMBLL** IBM CONTENT ANALYTICS WITH ENTERPRISE SEARCH PROCESSOR VALUE UNIT (PVU)
    - 1.364 EUR / Staffel BL \*
  - **D0IMJLL** IBM CONTENT ANALYTICS WITH ENTERPRISE SEARCH **STARTER PACK** PROCESSOR VALUE UNIT (PVU)
    - Bis 280 PVU
    - 683 EUR / Staffel BL \*
- **Typische Services vom Partner**
  - Installation des Systems, Konfiguration, Programmierung
  - Aufbau und Anpassung von Annotatoren (iterativer Prozess)
  - Validierung und Test
  - Ca. 15-20 Servicetage für einen einfachen PoC

# Zertifizierungen & Kurse

## Zertifizierung

- Technical Sales Zertifizierungen:
  - M90 IBM Content Analytics Theory Technical Professional v1
  - M91 IBM Content Analytics Product Fundamentals Technical Professional v1
  - M73 OmniFind Technical Sales Professional v1
- IBM Professional Zertifizierung:
  - Test 583: IBM Certified Specialist - Content Analytics and Search V2.2

## Training

- Business Partner Enablement (kostenlos):
  - **SAVE THE DATE:** Juli 2012, Ehningen, 2 Tage Proof-of-Technology
  - Bei Interesse Kontakt [ursula.bosch@de.ibm.com](mailto:ursula.bosch@de.ibm.com)
- IBM Training – 4 Tages Training (kostenpflichtig)
  - IBM Content Analytics Administration Workshop

[https://www-304.ibm.com/services/learning/content/ites.wss/us/en?pageType=course\\_description&courseCode=F1890](https://www-304.ibm.com/services/learning/content/ites.wss/us/en?pageType=course_description&courseCode=F1890)

# Kontakt Daten

**Uschi Bosch**

Business Partner Representative



Mobile 0172 837 4484  
Email [Ursula.bosch@de.ibm.com](mailto:Ursula.bosch@de.ibm.com)

**Andreas Schneider**

IT Specialist Content Analytics



Mobile 170 6347830  
Email [schneider@de.ibm.com](mailto:schneider@de.ibm.com)



**Peter Becker**

Business Partner Representative



Mobile 0160-909 84318  
Email [peter.becker@de.ibm.com](mailto:peter.becker@de.ibm.com)