



Системы классификации, поиска и анализа информации

Information Management software

Расширенные возможности по работе с информацией

Организация
Неструктурированного
Контента

IBM Classification

Бизнес анализ

IBM Content Analytics

Решения
По Раскрытию
Информации

Улучшение доступа к
корпоративному контенту

OmniFind™ Search



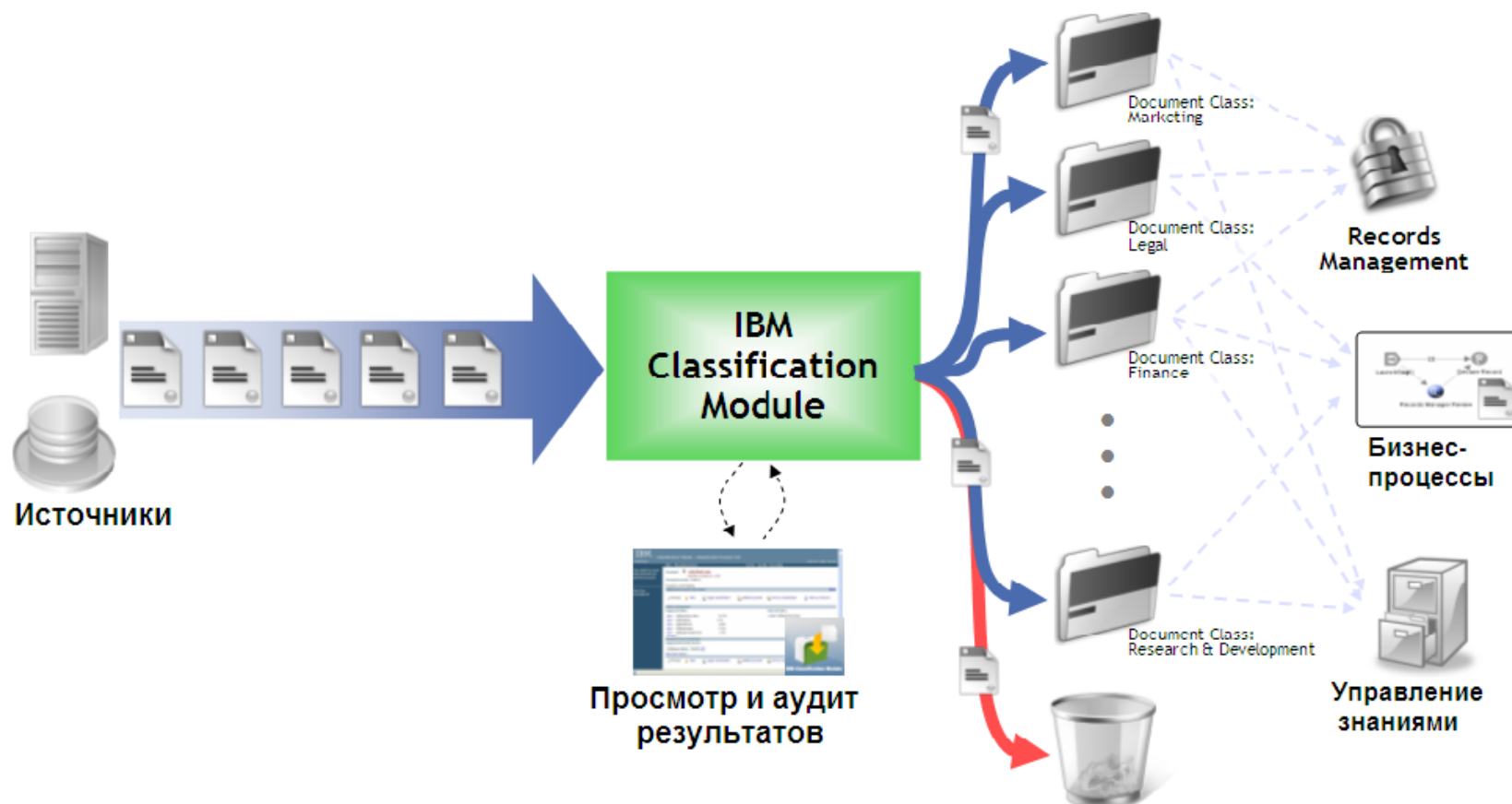
IBM Classification Module



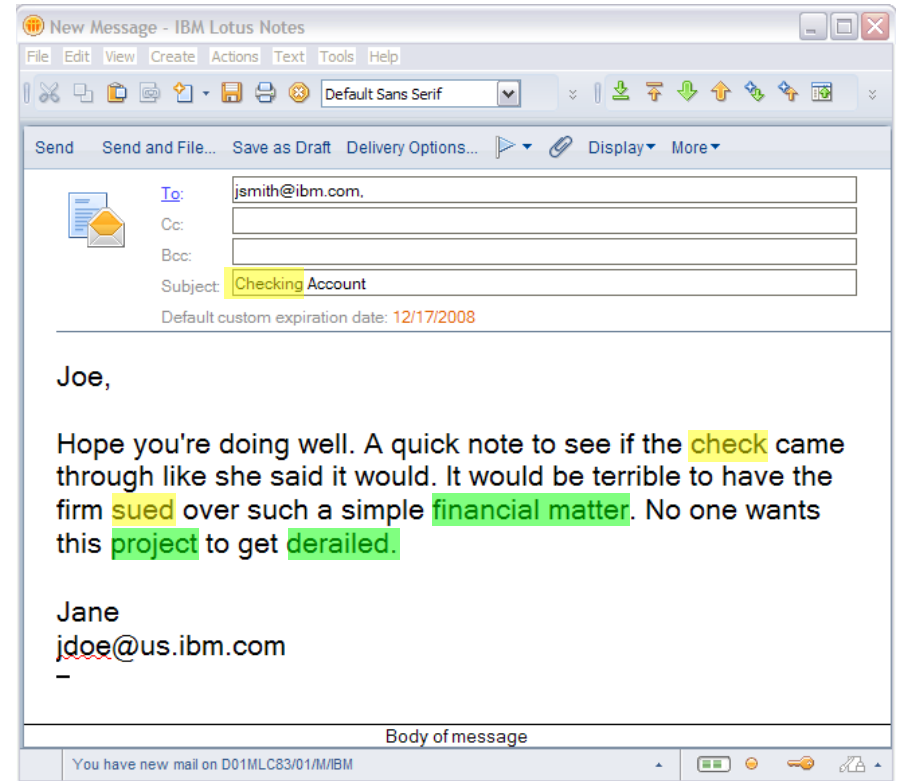
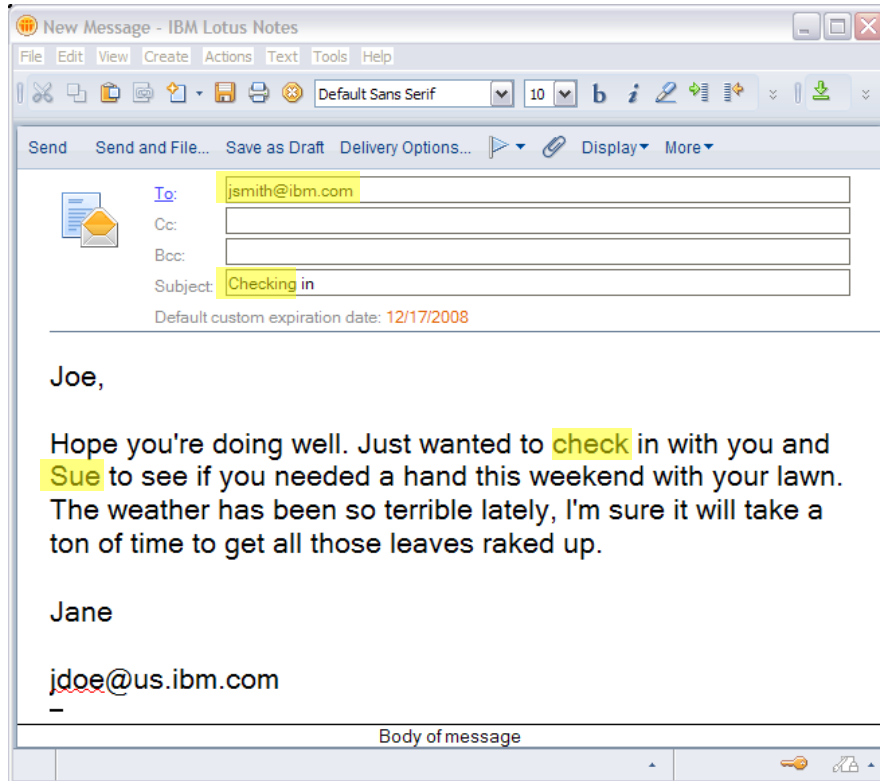
Применение

- Классификации могут подвергаться любые типы неструктурированной информации, включая документы, электронные письма, факсы, формы и т.п.
- Классификация проводится как на основе содержания документа, так и путем анализа большого количества атрибутивной информации, что обеспечивает очень высокое качество получаемых результатов.
- IBM Classification Module способен понимать лингвистику, семантику и контекст информации, что позволяет эффективно классифицировать неструктурированную и слабо структурированную информацию на различных языках.
- На этапе анализа позволяет применять к каждой группе правила анализа текстов, специфические только для документов данной логической группы
- На этапе поиска наличие классификации позволяет осуществлять более эффективную навигацию по результатам выполнения поисковых запросов

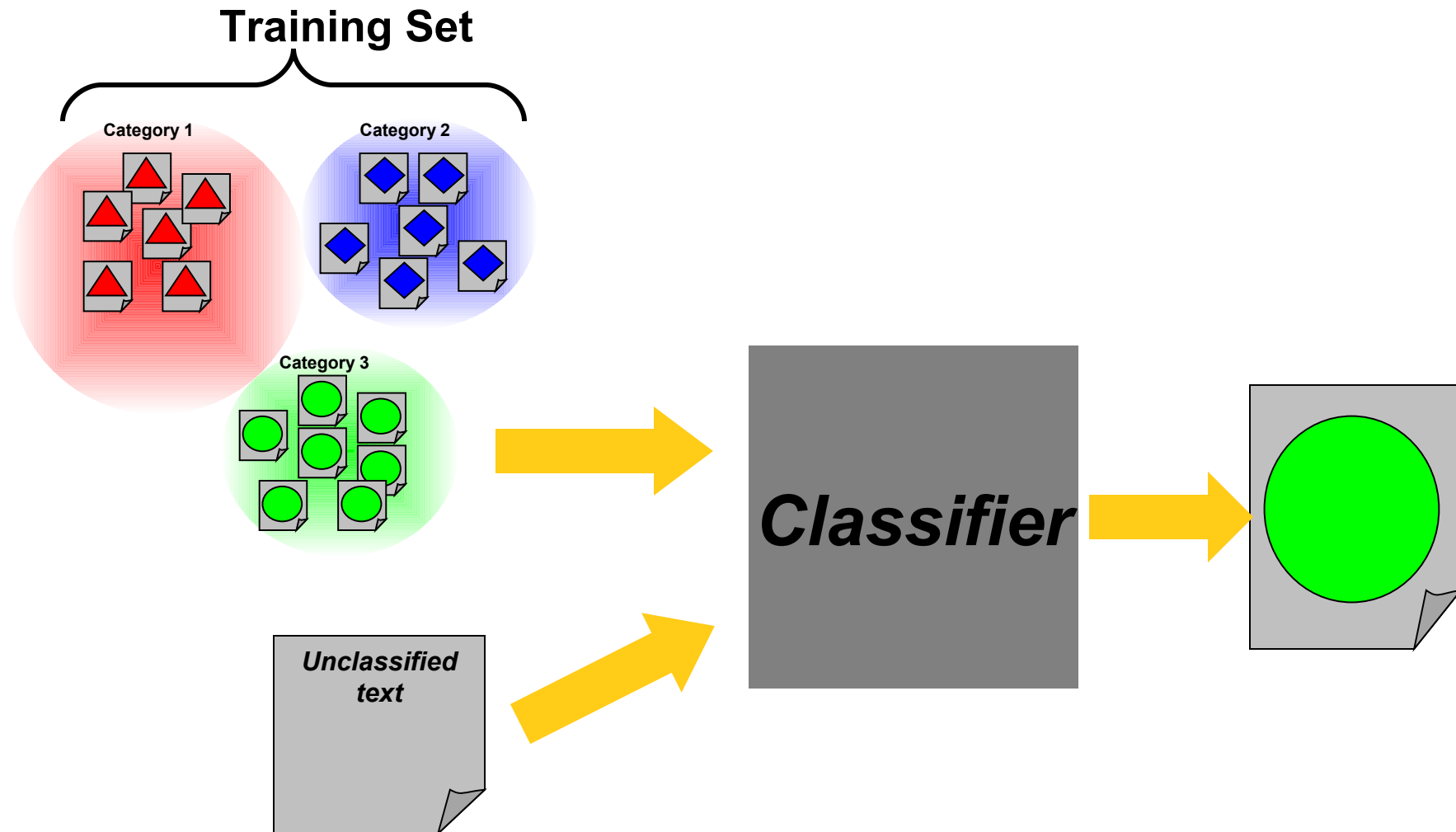
Применение



Анализ текста



Самообучаемость



Как это работает

- Построение статистических профилей для каждой категории
- Сопоставление нового документа с существующими категориями
- Категоризация и, если это необходимо, проверка правильности категоризации



IBM OmniFind и UIMA – расширенные возможности поиска



Базовые сведения

- высокая степень релевантности результатов поиска
- масштабируемость до десятков миллионов документов и десятков тысяч пользователей
- также высокая скорость обработки поисковых запросов
- встроены расширенные функции обработки лингвистической информации для двух десятков языков (включая русский язык)
- средства разграничения доступа к информации в зависимости от прав конкретных пользователей

Результаты поиска в OmniFind

- Фильтрация по типу файла, источнику, языку, метаданным ...
- Динамическое формирование аннотации, выделение текста
- Группировка результатов
- Проверка написания
- Использование метаданных (автор, даты, место и т.п.)
- Расширяемая модель ранжирования
- Сортировка по дате/релевантности
- Синонимы
- Параметрический поиск
- Поиск с использованием символов подстановки (*, ?)
- Категоризация
- Поиск в XML

DB2. Information Integrator enterprise search

Search for:

Search results for "how to change thinkpad power-on password"
 Result page: 1 2 3 4 5 6 7 8 9 10 ... Next

Quick links:
[Intranet Password](#)

[Setting Power-On, Hard-Disk & Screen Saver Passwords](#)
 ...From left to right the icons will be for **Power On** Hard Disk and Supervisor **passwords**. To set the **Power on password** click the **Power On** icon. ...Hint If you select the same **password** for both your **power on** and hard disk **password** you will only be ...Select **Change Power on password** and press Enter. Enter your new **power on password** where prompted. ...
<http://w3-5.ibm.com/services/emea/north/3ntgs.nsf/c1256b5e00254cb9c12568ea0046aed9/2637cbc8d31d5a3180256abc003f8a56?OpenDocument>

[IBM Technology Group | BTE | News](#)
 ...Before removing the **power on** or hard drive **password** all sensitive or IBM Confidential information or **passwords** should be removed. ...Follow the directions for deleting or **changing** the **power on password**. ...**Power off** your **ThinkPad** and wait 5 seconds to reboot. When requested to enter the **power on password** type in your current **password** and hit the space bar. ...To delete or **change** the **power on password** when it is different from the hard drive **password**. ...
<http://w3-3.ibm.com/technology/bte/news/1217-hddpw.html>

[IBM Technology Group | BTE | \(Hard Drive\) Password Rules](#)
 ...If you choose a different **password** for the hard drive the first **password** prompt when **powering** up your **thinkpad** is ...This is also true of **power on passwords** on **thinkpads**. Setting the Hard Drive **Password** on IBM **Thinkpads**. ...Continue holding down F1 until the **thinkpad power on password** prompt appears. ...For more information on hard drive **passwords** see IBM **ThinkPad** General - How to set remove or **change** a hard ...
<http://w3-3.ibm.com/technology/bte/security/harddrivepwrules.html>
[\[More results from w3-3.ibm.com\]](#)

[Transforming IBM : Using your Technology - Thinkpad](#)
 ...Whether you are new to IBM or a longtime employee your new **ThinkPad** will **change** the way you do your ...**Power On Password**. ...The **ThinkPad Power On password** may not exceed 7 characters. If you create an 8 character **password** your next **Power On** will fail and lock you out. ...When you are near a **power** source it wise to use your **ThinkPad** plugged in. ...
<http://w3-1.ibm.com/transform/dio.nsf/printview/1DBECEEF168891B685256ABE005E68641>

Динамическое резюме

Преднастроенный результат

Лексическая сущность

Лемматизация

«Свертывание» сайта

Исключение стоп-слов

Search for:

Filter by file type: [All](#) [doc](#) [html](#) [pdf](#) [ppt](#) [txt](#) [xls](#) [xml](#) | Filter by source type: [All](#) [DB2](#) [Notes](#) [UNIX file system](#) [Web](#)

Sorted by relevance | [Sort by date](#)

256 search results returned | 256 documents match all query terms

		Show Details	1-10 of 256
92.45%	7/24/05	<p>iiysr.htm</p> <p> /opt/IBM/es/docs/en_US</p> <p> iiysr.htm</p> <p>...fp2IBM WebSphere Application Server Deployment Manager Version 5.1 with Fix Pack 1 5.1.1 Deployment Manager is required to allow WebSphere ...DB2 Content Manager Version 8.2 or 8.3 Accessed with the Content Manager crawler. ...For enterprise search on AIX and Windows the Content Manager crawler uses the Java TM connector for Content Manager Version ...</p>	
92.42%	7/24/05	<p>iiysr.txt</p> <p> /opt/IBM/es/docs/en_US</p> <p> iiysr.txt</p> <p>...IBM WebSphere Application Server Deployment Manager Version 5.1 with Fix Pack 1 5.1.1 Deployment Manager is required to allow WebSphere ...DB2 Content Manager Version 8.2 or 8.3 Accessed with the Content Manager crawler. ...For enterprise search on AIX and Windows the Content Manager crawler uses the Java TM connector for Content Manager Version ...</p> <p>[More results from the same source]</p>	
92.23%	5/6/05	<p>Guidance section - not for release to the client</p> <p> Model SOW for OmniFind PoC</p> <p> Model OmniFind SOW.doc</p> <p> WW OmniFind Teamroom - IBM Confidential</p> <p>...Prepare and submit Periodic Status Reports to the CUSTOMER NAME Project Manager. Administer the Project Change Control Procedure with the CUSTOMER NAME Project Manager. ...Maintain project communications through the FSG CUST_NAME Project Manager. ...</p>	
94.57%	1/17/05	<p>GERMAN version of 2004 prz</p> <p> GERMAN version of 2004 prz</p> <p> IBM DB2 Info Integrator OmniFind Tech Overview_GERMAN VERSION.ppt</p> <p> WW OmniFind Teamroom - IBM Confidential</p> <p>...Exchange Public Folder File System Content Management System DB2 Content Manager mit DB2 Oracle Documentum Filenet Open Text</p>	

Интеграция с WebSphere Portal

The screenshot shows the IBM WebSphere Portal interface. At the top, there is a navigation bar with links for 'My Portal', 'Administration', 'Edit my profile', and 'Log out'. Below this is a secondary navigation bar with 'New Page', 'Edit Page', 'Assign Permissions', and a search box. The main content area is titled 'OmniFind Portlet 8.3.006' and contains a search form with 'IBM' entered. To the right of the search box are links for 'Search', 'Advanced Search', 'Category Tree', 'Preferences', and 'My Profile'. Below the search form, there are filters for 'File type filter' and 'Source type filter'. The search results are sorted by relevance and show 454 results. The first two results are listed below.

Search for: [Search](#) [Advanced Search](#) [Category Tree](#) [Preferences](#) [My Profile](#)

File type filter: [All](#) [doc](#) [html](#) [pdf](#) [ppt](#) [txt](#) [xls](#) [xml](#)
Source type filter: [All](#) [Content Edition](#) [Web](#) [Windows file system](#)
Sorted by **relevance** | [Sort by date](#) 454 search results returned | 454 documents match all query terms

[Show Details](#) 1-10 of 454

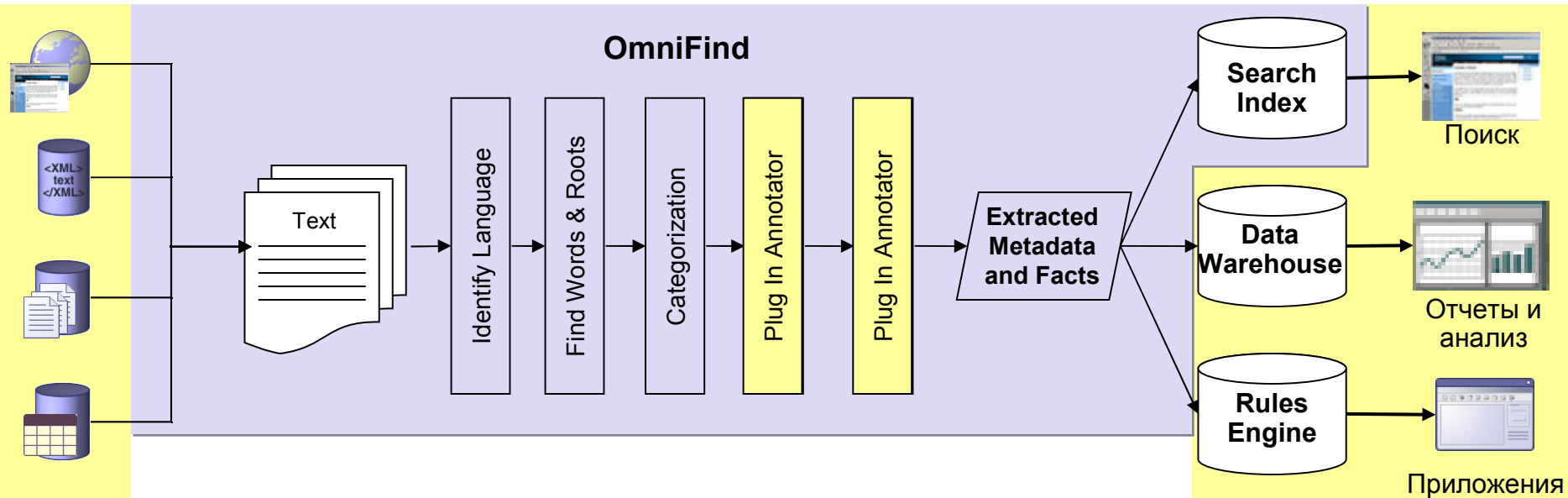
100.00%	3/27/06	IBM United States http://www.ibm.com/us/ IBM United States. ...IBM Certified Used Equipment. ...IBM Business Partners. ...Jobs at IBM. IBM Research. ...
100.00%	3/27/06	About IBM - United States http://www.ibm.com/ibm/us/ ...IBM Venture Capital Group. ...At IBM we strive to lead in the invention development and manufacture of the industry 's most advanced information technologies ...IBM Corporation 1133 Westchester Avenue White Plains New York 10604 United States. ...For answers to general questions about IBM from within the continental United States call 800-IBM-4YOU from outside the continental United ...Enter the name of an IBM employee to find his or her e-mail address and telephone number. ... [More results from the same source]

Особенности

- Поиск с учетом морфологии русского языка.
- Поиск по содержанию и атрибутам документов.
- Возможности использования запросов на естественном языке и запросов с использованием логических выражений.
- Учет прав доступа при индексировании и поиске документов (ACL – access control list, списки управления доступом).
- Поддержка сервисно ориентированной архитектуры (SOA).
- Проверенная масштабируемость (различные варианты развертывания в зависимости от индексируемых объемов).
- Возможность создания нескольких поисковых приложений (гибкая настройка поиска по одной или нескольким коллекциям).
- Уникальная возможность интеграции с различными источниками.
- Интеграция с порталами различных производителей.

Структура решения

- Plug and play компоненты для анализа текста
- Платформа увеличивает отдачу от инвестиций
 - Доставляет контент из внешних источников в платформу для анализа
 - Проводит различный анализ для различных коллекций
 - Позволяет широкому кругу приложений использовать результаты анализа



Особенности

- Готовые модули извлечения информации из различных систем (поддерживаются практически все известные БД, системы электронного документооборота, системы коллективной работы, web, файловые системы и т.д.). Допускается создание собственных модулей извлечения информации для решения нестандартных задач.
- Поддержка обработки более 200 различных форматов документов.
- Поддержка архитектуры UIMA – Unstructured Information Management Architecture.
- Настраиваемые и расширяемые аннотаторы – модули разбора и структурирования текста. Наличие возможности подключения внешних аннотаторов, учитывающих как отраслевую специфику, так и предназначенных для решения узкоспециальных задач.
- Наличие готовых модулей разбора (определение кодовой страницы и языка документа, **извлечение информации на основе регулярных выражений**).
- Возможность ведения нескольких поисковых коллекций под различные задачи (у каждой коллекции свои модули извлечения и модули анализа текста).



Анализ текста - UIMA

- UIMA – Unstructured Information Management Architecture
- Разработана в IBM и передана сообществу open-source
- Базируется на понятии «Аннотатор»
 - Извлечение сущностей (люди, даты, места, продукция, виды лицензий и т.д.)
 - Определение темы (категория, настроение)
 - Взаимосвязи
 - и т.д.
- Omnifind реализует исполнительную подсистему для выполнения аннотаторов



Расширенные возможности по изучению контента

Nhanced Search Tool - Microsoft Internet Explorer

Address: http://206.162.161.125:40000/BlogsNews/index.do

DB2. Information Integrator OmniFind Edition

Find it! Search Blogs News Blogs

You searched for: honda civic / fuel economy / honda civic hybrid
We found: 23 documents

One-Click Access

- Makes & Models +
- Organizations +
- Places +
- Concepts +

Mine results ...

powered by Stein

Search Results BI Enablers

Tone

Tone	Percentage
Positive	30%
Negative	57%
Neutral	13%

Hot Topics

Hot Topic	Frequency
fuel economy, mpg city	100
suvs, toyota prius, fuel	66
engines, fuel economy, transmissions	58
environmental protection agency, fuel economy, toyota prius	50
sedans, honda accord, toyota echo	45

Document Types

Point of View

Мониторинг публичного мнения: Поиск по блогам, новостям по ключевым словам «хонда цивик гибрид» предоставляет наиболее популярные мнения и иллюстрирует общественный настрой

Примеры решений на базе UIMA

Решение

Управление проблемами

Мониторинг
общественного мнения

Анализ

Опросы клиентов, звонки в call центр, отзывы на сайте и e-mail, анализ отказов, BLOG, аналитические сайты, публикации

BLOG, аналитические сайты, публикации, группы новостей

Понимание

Беспокойство клиентов, проблемы с продуктами, возможности продуктов, причины оттока клиентов

восприятие рынка, проблемы потребителей, продукты и функции, что важно, тон высказываний

Для кого

Automotive	Durable Goods	High Tech Complex Goods	Consumer Goods	Automotive	Durable Goods	High Tech Complex Goods	Consumer Goods
Hospitality	Telecom	Financial Services	Insurance	Hospitality	Telecom	Financial Services	Insurance

**Бизнес
выгоды**

- Создание и управление лояльностью клиентов
- Понимание потребностей
- Минимизация потерь прибыли

- Понимание общественного восприятия
- Открытость для изменений
- Более эффективный маркетинг
- Защита репутации



Примеры решений на базе UIMA

Решение

Know Your Customer

Fraud & Risk Assessment

Анализ

Звонки в call центр, заметки в CRM, опросы, данные клиентов, финансовые отчеты, контракты, e-mails, открытые архивы

заявления, полисы, финансовые отчеты, данные клиентов, открытые архивы

Понимание

люди, организации, связи, аномалии, продукты/услуги, шаблоны/тенденции

«частые гости», роли, связи, группы, странные заявления, шаблоны/тенденции

Для кого

Financial Services

Insurance

Financial Services

Insurance

Бизнес выгоды

- 360° View of Customers
- Сокращение рисков
- Максимизация прибыли (\$ Potential)
- Повышение удовлетворенности клиентов

- Сокращение рисков
- Выявление мошенничества
- Сокращение затрат/времени на выявление мошенничества

Примеры решений на базе UIMA

Решение

Анализ проблем с продукцией (Quality Early Warning)

Служба поддержки клиентов

Анализ

Заявления по гарантии, отзывы клиентов, технические заметки, звонки в call центр, анализ e-mail

Контакты с Call центром, отчеты о проблемах, e-mails, база знаний, заметки, FAQ, документация по поддержке

Понимание

Упомянутые продукты/компоненты, проблемы, условия/симптомы, действия, тенденции

Выявление проблем, связанные проблемы, вовлеченные продукты, взаимосвязь продуктов, решения, лучшие практики

Для кого

Automotive	Durable Goods	High Tech & Complex Goods
------------	---------------	---------------------------

Utilities	Durable Goods	High Tech Complex Goods	Consumer Goods
Hospitality	Telecom	Financial Services	Insurance

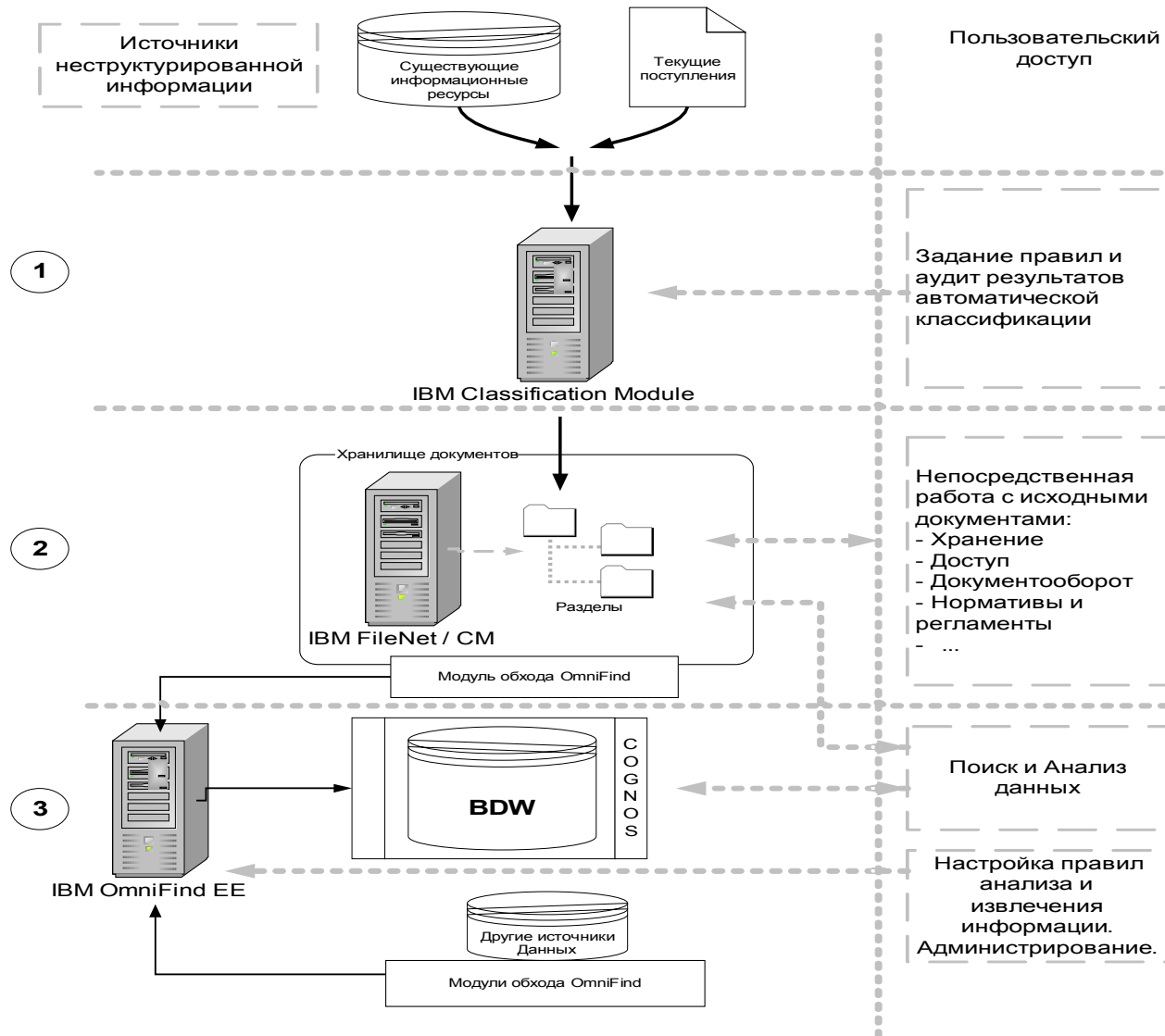
Бизнес выгоды

- Повышение качества продукции
- Сокращение затрат на поддержку
- Повышение удовлетворенности клиентов

- Сокращение проблем поддержки
- Быстрое решение вопросов
- Повышение удовлетворенности клиентов



Пример интегрального решения





30 customers deployed or deploying