

BUSINESSOBJECTS TEXT ANALYSIS

Воспользуйтесь «родными» средствами доступа к неструктурированным текстовым данным, расширяющими возможности BusinessObjects Enterprise и BusinessObjects Data Integrator, для получения более полного представления об эффективности работы Вашего предприятия.

АНАЛИЗ НЕСТРУКТУРИРОВАННЫХ ТЕКСТОВЫХ ДАННЫХ

Качество принимаемых бизнес-решений зависит от наличия актуальной информации. Однако не смотря на то, что значительное количество информации скрыто в текстах (комментарии к обзорам, записи call-центра, данные из Интернета), текстовая информация в основном остается вне BI-приложений и систем управления данными.

Неструктурированные источники текстовых данных, например, электронные письма, документы и контент сайтов, зачастую содержат важную информацию о скрытых рисках и возможностях. Так, например, если целью компании является обеспечение максимального уровня удовлетворенности своих клиентов, то комментарии, содержащиеся в CRM-системе, блогах и электронных письмах, позволяют выявлять причины их недовольства, мнения и отклики.

BusinessObjects™ Text Analysis осуществляет обработку, систематизацию и резюмирование огромных объемов информации, что позволяет получить более полное представление об эффективности работы Вашего предприятия.

Использование BusinessObjects Text Analysis предоставляет следующие возможности:

- **Бизнес-анализ всех текстовых источников на любых языках.** Возможность извлечения информации, содержащейся в неструктурированных источниках данных, и смысла текстов из многочисленных документов в различных форматах на более чем 30 языках.
- **Извлечение, классификация и резюмирование информации.** Быстрое нахождение и распознавание понятий, персон, организаций, географических атрибутов и других данных, которые содержатся только в неструктурированных текстовых источниках.



- **Интеграция с продуктами BusinessObjects.** Нарращивание функциональности BusinessObjects Enterprise, BusinessObjects Data Integrator и корпоративной системы поддержки решений за счет дополнительной возможности анализа неструктурированных текстовых данных.

Рисунок 1. Интеграция BusinessObjects Text Analysis с системой поддержки принятия решений Business Objects.

Поддерживаемые языки

- Арабский
- Каталанский
- Китайский
- Хорватский
- Чешский
- Датский
- Голландский
- Английский
- Фарси
- Финский
- Французский
- Немецкий
- Греческий
- Иврит
- Венгерский
- Итальянский
- Японский
- Корейский
- Норвежский
- Польский
- Португальский
- Румынский
- **Русский**
- Сербский
- Словацкий
- Словенский
- Испанский
- Шведский
- Тайский
- Турецкий

ВИ ПО ДАННЫМ ИЗ ВСЕХ ИСТОЧНИКОВ ТЕКСТОВОЙ ИНФОРМАЦИИ И НА ЛЮБОМ ЯЗЫКЕ

По данным обзора Q2 2007 TDWI в традиционной интеграции данных наблюдается нарастающая тенденция перехода от анализа только структурированных данных к анализу неструктурированных текстовых источников, таких как электронные письма, текстовые файлы, системы управления документацией и контентом, wiki-сайты и веб-страницы.

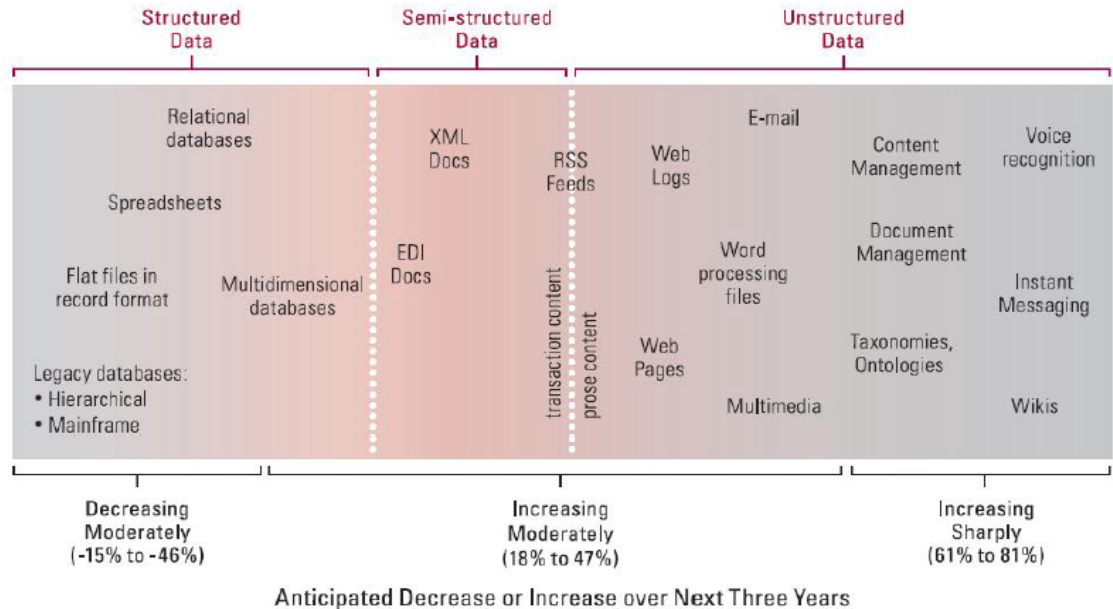


Рисунок 2. Интеграция данных и BI-технологии: переход от анализа структурированных данных к анализу неструктурированных источников текстовой информации. (Источник: TDWI).

Являясь дополнением к средствам анализа структурированной информации, BusinessObjects Text Analysis обеспечивает доступ и производит обработку текстовых файлов с поддержкой более чем 220 форматов, что дает возможность получить полноценное представление о Ваших заказчиках и эффективности работы Вашего предприятия.

Расширенные лингвистические возможности системы позволяют распознавать и производить обработку документов на более чем 30 языках, включая русский (см. колонку слева). Заложенный в BusinessObjects Text Analysis сложный алгоритм обработки естественного языка обеспечивает распознавание взаимосвязей глаголов, существительных и других языковых структур, что позволяет определять контекстный смысл информации, а не только самих слов.

МОЩНЫЕ ВОЗМОЖНОСТИ ИЗВЛЕЧЕНИЯ, КЛАССИФИКАЦИИ И РЕЗЮМИРОВАНИЯ ТЕКСТОВЫХ ДАННЫХ

Основу BusinessObjects Text Analysis составляют мощные средства извлечения данных, производящие структурный анализ больших объемов документов с выявлением таких сущностей как, например, заказчики, продукты, местоположение и финансовая информация, относящихся к Вашей организации. Эта функция дополняется возможностью их таксономической классификации по отдельным компаниям и отраслям промышленности, а также рубрикации и резюмирования, что обеспечивает создание легких для понимания реферативных данных.

Поддерживаемые критерии

- **Кто**
 - Персоналии
 - Должность
 - Номер социальной страховки
- **Что**
 - Компании
 - Организации
 - Финансовые показатели
 - Продукты
- **Когда**
 - Даты
 - Дни недели
 - Праздничные дни
 - Месяцы
 - Годы
 - Время
 - Периоды времени
- **Где**
 - Адреса
 - Города
 - Области
 - Страны
 - Учреждения
 - Адреса в Интернете
 - Номера телефонов
- **Количество**
 - Виды валюты
 - Котировки
- **Понятия**
 - Создание понятий
- **Отношения и мероприятия**
 - Персона-организация
 - Персона-персона
 - Поездки
 - Слияния и поглощения
- **Категории**
 - Бизнес
 - Терроризм
 - Клиентура
- **Настройки**
 - Отрасль промышленности
 - Постоянные покупатели

Мощные возможности извлечения и анализа сути текста

На основе более чем 35 предопределенных в системе сущностей, соотношений и событий BusinessObjects Text Analysis производит автоматическое извлечение ключевой информации по персонам, географическим атрибутам, компаниям, датам, нормативным данным, курсам валют и электронным адресам и выявляет важные взаимосвязи и относящиеся к ним события. Взаимосвязи, возможно выраженные неявно в одном документе или источнике, выявляются за счет того, что BusinessObjects Text Analysis «понимает» что Вы ищете и может сопоставлять неидентичные фрагменты данных.

Кроме того, функциональность Text Analysis может быть расширена для распознавания таких сущностей как коды единиц складского хранения (SKU), названия проектов, взаимосвязи, события, а также отдельные фразы для проведения эмоционального анализа.

The proposed merger between Mega, Inc. and CNA Systems, Incorporated, has been postponed, Mega CEO Joe Smith said in an analyst call. "CNA's 1st quarter revenue dropped by 32%, and they lost 23 million dollars," Smith explained. CNA Systems sources blame weak sales in China. CNA shares (CNAI) fell 47 percent to \$9.84 on May 12, the first trading day after the announcement.	
Company	Mega, Inc., CNA Systems, Incorporated
Date	May 12
Person	Joe Smith
Person Position	Mega CEO
Currency	23 million dollars, \$9.84
Measurement	32%, 47 percent
Country	China
Concept	proposed merger, analyst call, 1st quarter revenue weak sales, first trading day
Event: M&A	The proposed merger between Mega, Inc. and CNA Systems, Inc. has been postponed

Рисунок 3. BusinessObjects Text Analysis извлекает смысловое значение из источников текстовой информации с представлением данных в легком для понимания виде.

Таксономическая классификация документов

BusinessObjects Text Analysis также предоставляет возможность классификации информации на основе содержащейся в документах таксономии. Например, документ может относиться к категории жалоб от заказчиков, если даже слово «жалоба» в тексте при этом не встречается. Text Analysis распознает такие взаимосвязи понятий и автоматически относит эту информацию к соответствующей категории.

Автоматическое резюмирование документов

BusinessObjects Text Analysis также извлекает и анализирует наиболее релевантные предложения с созданием резюме какого-либо документа или комплекта документов. Этот пакет проще всего охарактеризовать как «Автоматический рефератор данных», который считывает информацию и создает «выжимку» из контента. Text Analysis излагает общую суть информации с использованием системы «взвешенных» предложений, обычно с акцентом на первые и последние предложения параграфов, соответствующих параметрам Вашего запроса.

Эта способность извлекать релевантные данные является важной, так как отдельный человек может заниматься изучением погодных условий и интерес для него может представлять «снег», в то время как кто-либо еще может использовать этот же документ для изучения маршрутов перевозок грузов на дальние расстояния. Text Analysis «понимает» Ваши специфические требования и производит оценку значимости реферативных данных в ином ракурсе в зависимости от области Ваших интересов.

ИНТЕГРАЦИЯ С СИСТЕМАМИ НА БАЗЕ BUSINESS OBJECTS

Использование BusinessObjects Text Analysis дает Вашей организации такие возможности как:

- **Получение и анализ текстовых данных** точно теми же средствами, что и структурированных данных – непосредственно из Crystal Reports®, BusinessObjects Web Intelligence®, информационных панелей и других клиентов Business Objects.



Рисунок 4. Интеграция BusinessObjects Text Analysis с продуктами Business Objects: предоставление текстовой информации для принятия бизнес-решений в привычном виде отчетов и метрик.

- **Интеграция с BusinessObjects Enterprise.** Text Analysis автоматически создает юниверсы BusinessObjects, распознавая метаданные непосредственно из извлеченных текстов. Это обеспечивает быструю интеграцию текстовых источников с целостным семантическим слоем в BI-отчеты и метрики, сокращая время представления новой информации для анализа.
- **Интеграция с BusinessObjects Data Integrator.** Data Integrator имеет непосредственный доступ к источникам и заданиям Text Analysis. Это дает возможность быстрого включения текстовых источников в среду управления информацией, что сокращает время получения надежной информации для анализа.