

# BUSINESSOBJECTS TEXT ANALYSIS

*For a more complete view of organizational performance, access your unstructured text data with the first and only native text analysis capabilities designed specifically to complement your BusinessObjects Enterprise and BusinessObjects Data Integrator deployments.*

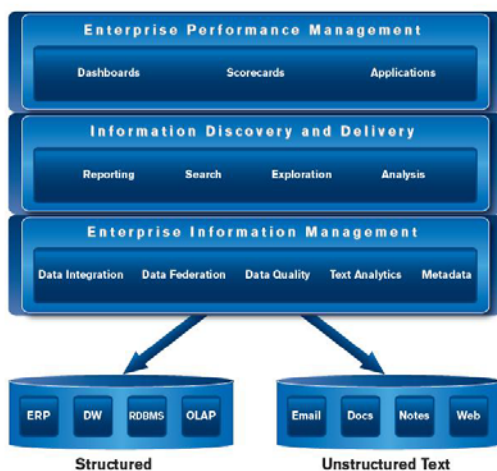
## COMPLEMENTING STRUCTURED DATA WITH THE INSIGHT IN TEXT

Better business decisions are made when more actionable information is available. Yet, while there is a considerable amount of information hidden in text—such as survey comments, call center records, and the web—text information remains predominantly outside of business intelligence (BI) deployments and their information management systems.

Untapped text sources such as emails, documents, notes fields, and web content often contain critical information about hidden risks and opportunities. For example, while customer satisfaction remains a company goal, customer frustrations, opinions, and feedback remains hidden in customer relationship management (CRM) system comments, blog sites, and email. And while organizations try to address compliance regulations, textual information that includes compliance risk-related issues remains hidden in documents, records, and contracts.

BusinessObjects™ Text Analysis processes, classifies, and summarizes vast amounts of text information—helping you gain better insight into your business. With BusinessObjects Text Analysis, you get a more complete view of organizational performance with the first and only native text analysis capabilities designed specifically to complement your data integration and BI deployments. BusinessObjects Text Analysis gives you the following:

- **BI on all your text sources, and from any language.** Complement structured BI with the wisdom buried in unstructured text sources by accessing and deriving meaning from hundreds of thousands of text documents—in a variety of file formats and in more than 30 major languages.
- **Powerful extraction, categorization, and summarization.** Quickly identify and understand the concepts, people, organizations, places, and other information that only exists in unstructured text sources.



- **Integration with your BusinessObjects deployments.** Easily add sophisticated text analysis capabilities to your BusinessObjects Enterprise and BusinessObjects Data Integrator deployments, and your organization's decision-making framework.

Figure 1. BusinessObjects Text Analysis integrates with your BusinessObjects decision-making framework.

**Extractable Languages**

- Arabic
- Catalan
- Chinese (Simplified)
- Chinese (Traditional)
- Croatian
- Czech
- Danish
- Dutch
- English
- Farsi
- Finnish
- French
- German
- Greek
- Hebrew
- Hungarian
- Italian
- Japanese
- Korean
- Norwegian (Bokmal)
- Norwegian (Nynorsk)
- Polish
- Portuguese
- Romanian
- Russian
- Serbian
- Slovak
- Slovenian
- Spanish
- Swedish
- Thai
- Turkish

## BI ON ALL YOUR TEXT SOURCES—AND FROM ANY LANGUAGE

A Q2 2007 TDWI survey illustrates the burgeoning shift from traditional data integration and BI using exclusively structured sources towards organizations increasing their focus on incorporating unstructured text sources,<sup>1</sup> such as emails, word processing files, document and content management systems, wikis, and web pages.

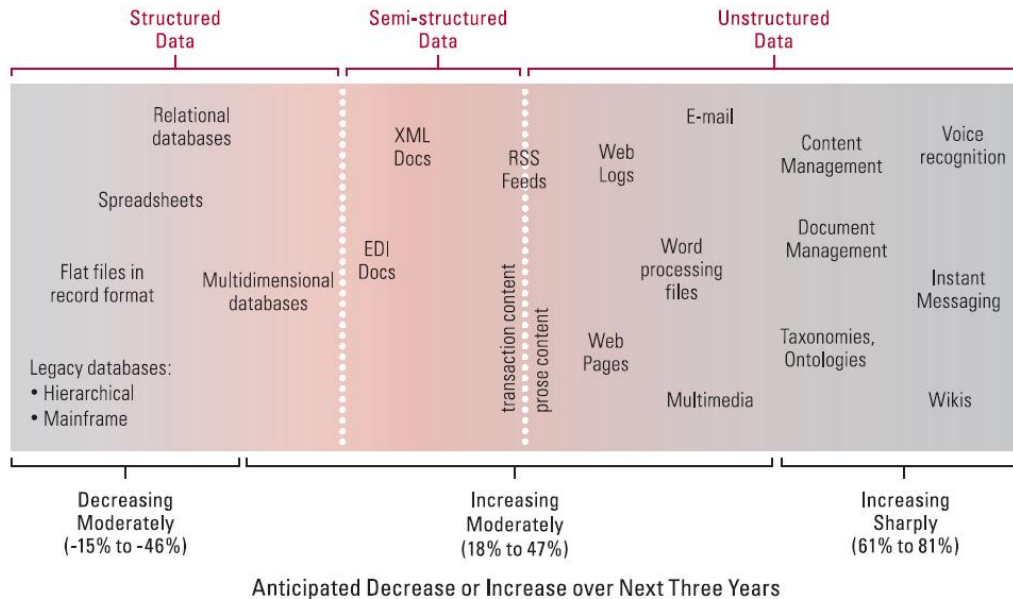


Figure 2. Data integration and BI technologies are shifting from using exclusively structured sources to rapidly incorporating unstructured text sources. (Source: TDWI)

BusinessObjects Text Analysis can access and apply its sophisticated natural language processing to over 220 file formats—complementing structured information and creating a full-spectrum view of your customers and organizational performance.

The product's advanced linguistics capabilities read and “understand” documents in more than 30 major languages (see sidebar). Through sophisticated natural language processing (NLP), BusinessObjects Text Analysis knows how verbs, nouns, and other language structures interact. In essence, it understands the *meaning* and *context* of information—not just the words themselves.

## POWERFUL EXTRACTION, CATEGORIZATION, AND SUMMARIZATION

The foundation of BusinessObjects Text Analysis is its entity-extraction capabilities. Powerful extraction parses large volumes of documents, identifying “entities” such as customers, products, locations, and financial information relevant to your organization. Entity extraction is complemented by categorization capabilities that can apply company- or industry-specific taxonomies to the text for subject-level classification, and summarization capabilities that create abstracts that are readily understood.

### Powerful Entity Extraction and Analysis

With more than 35 out-of-the-box entities, relations, and events preconfigured with the system, BusinessObjects Text Analysis automatically extracts key information—including people, places, companies, dates, measurements, currency figures, and email addresses—and

<sup>1</sup> Russom, Philip. “BI Search and Text Analytics,” TDWI, May 2007.

**Extractable Entities**

- Who
  - People
  - Job title
  - Social Security numbers
- What
  - Companies
  - Organizations
  - Financial indexes
  - Products
- When
  - Dates
  - Days
  - Holidays
  - Months
  - Years
  - Times
  - Time periods
- Where
  - Addresses
  - Cities
  - States
  - Countries
  - Facilities
  - Internet addresses
  - Phone numbers
- Quantity
  - Currencies
  - Measures
- Concepts
  - Concept creation
- Relations and events
  - Person-org
  - Person-person
  - Travel
  - M&A
- Categories
  - Business
  - Terrorism
  - Custom
- Customizable
  - Industry
  - Custom

reveals important relations and events occurring between those data items. Relationships that may not be implicit in one document or source are uncovered because BusinessObjects Text Analysis understands what you are looking for and can put the disparate bits and pieces together.

With its deep understanding of natural language, Text Analysis can also be extended to recognize custom entities and entity types, such as stock keeping unit (SKU) numbers, project names, special relationships and events, and specific phrases for sentiment analysis.

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.	
<b>Company</b>	Mega, Inc., CNA Systems, Incorporated
<b>Date</b>	May 12
<b>Person</b>	Joe Smith
<b>Person Position</b>	Mega CEO
<b>Currency</b>	23 million dollars, \$9.84
<b>Measurement</b>	32%, 47 percent
<b>Country</b>	China
<b>Concept</b>	proposed merger, analyst call, 1st quarter revenue weak sales, first trading day
<b>Event: M&amp;A</b>	The proposed merger between Mega, Inc. and CNA Systems, Inc. has been postponed

Figure 3. With its understanding of natural language, BusinessObjects Text Analysis extracts meaning from text sources and makes these facts available to the business in a readily understood format.

**Taxonomy-Based Document Categorization**

BusinessObjects Text Analysis also provides taxonomy-based, document-level classification of information that isn't implicit within documents. For example, your document may belong to a customer complaint category—even though the word “complaint” never appears in the text. But any document that talks about dissatisfaction, for instance, is certainly a complaint-related document, regardless of whether “complaint” occurs in the document. Text Analysis understands these types of conceptual relationships and automatically classifies your information in the right category.

**Automatic Document Summarization**

BusinessObjects Text Analysis also extracts and analyzes the most relevant sentences to create a summary of any document or set of documents. Think of it as an “automated abstract creator” that reads the information and creates an executive report on the content. Text Analysis digests the general substance of the information by using a system of weighted sentences, generally giving more weight to the first and last sentences of a paragraph, for example, along with the specific query parameters of your search.

This capability to extract relevant data is important because one person may be investigating weather phenomenon and be interested in “snow,” whereas someone else may use the same document to investigate transportation patterns of long-haul trucks. Text Analysis understands your specific requirements and will weight the summary differently depending on your area of interest.

## INTEGRATION WITH YOUR BUSINESSOBJECTS DEPLOYMENTS

With BusinessObjects Text Analysis, your organization can:

- **Access and analyze insight found in text** in the same manner as structured data—directly from the same Crystal Reports®, BusinessObjects Web Intelligence®, dashboards, and other Business Objects clients you use today.

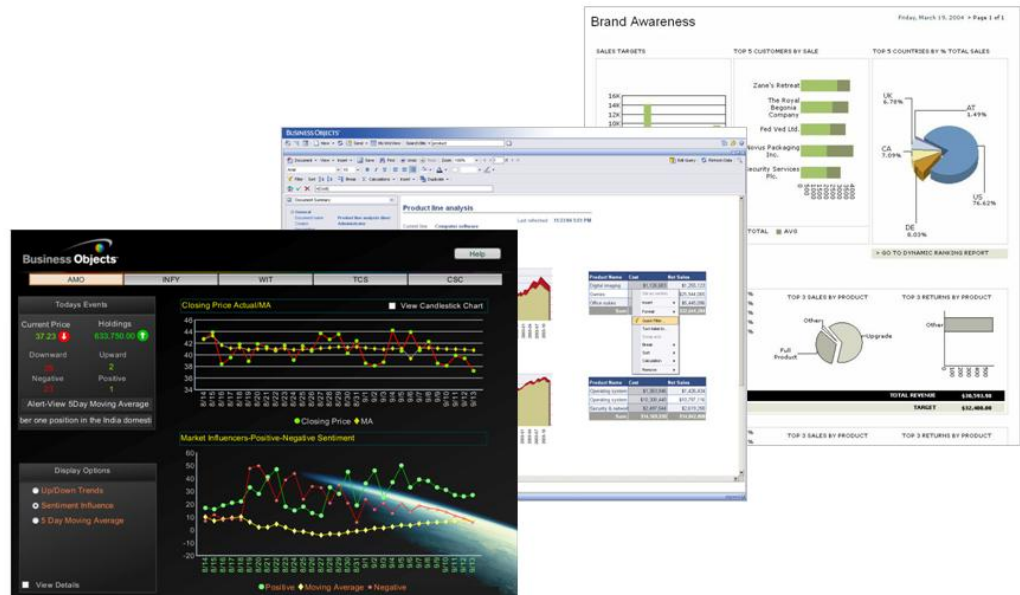


Figure 4. BusinessObjects Text Analysis integrates directly with BusinessObjects deployments and makes text information available to your business via the same reports and metrics it uses today.

- **Integrate with BusinessObjects Enterprise.** Text Analysis automatically generates metadata-aware BusinessObjects universes directly off of recently extracted text. This makes for quick integration of text sources with consistent metadata semantics into BI reports and metrics, giving you faster time to new insight.
- **Integrate with BusinessObjects Data Integrator.** Data Integrator directly accesses Text Analysis sources and jobs. This makes for quick incorporation of text sources into the broader information management environment—which means faster time to trusted insight.

[businessobjects.com](http://businessobjects.com)

[insight.businessobjects.com](http://insight.businessobjects.com)