

BusinessObjects™ Text Analysis software extracts key business information in more than 30 languages from unstructured text sources such as e-mails and Web-based and customer documents. This can improve analysis of customers, root causes, links, shareholder value, counterterrorism, and employee satisfaction.



## A COMPREHENSIVE APPROACH TO TEXT ANALYSIS

### TAPPING A KEY SOURCE OF INFORMATION

Better information leads to better business decisions. Yet a considerable amount of text information – in the form of customer comments, call center records, or Web content – is not readily accessible through most information management solutions.

Such untapped text sources can contain critical information about hidden business risks and opportunities. Customer feedback may remain buried in customer relationship management solutions, blogs, and e-mail, for example. Key compliance information may be missed in various documents, records, and contracts.

BusinessObjects™ Text Analysis software from Business Objects, an SAP company, can help you examine text in more than 30 languages and convert that information into useful business intelligence. This complete, fully integrated set of tools has helped more than 450 businesses, government organizations, and OEM customers to find, extract, cleanse, analyze, and visualize critical information in both SAP® and non-SAP enterprise applications (see Figure 1). Such support is simply not possible using standard search engines or traditional business intelligence and document management software.

With BusinessObjects Text Analysis, you can:

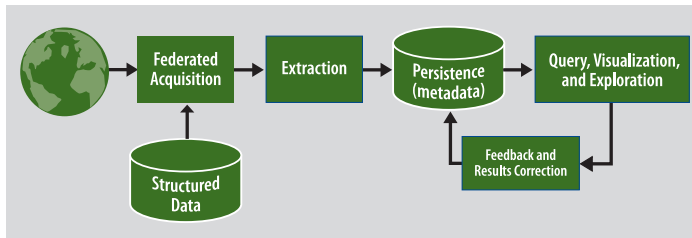
- Power new value-added products and revenue streams
- Reduce compliance and litigation costs
- Gain a better understanding of your customers and their internal operations
- Reduce development costs and speed time to market for information-intensive OEM software applications

Based on more than 25 years of research, BusinessObjects Text Analysis offers an unparalleled depth and breadth of features for finding, using, and understanding information. Whether the source of that information is the Internet, internal information stores, subscription content, or deep Web repositories, BusinessObjects Text Analysis can also alert you to new or changing information as it develops. Once information is collected, the extraction and analysis tools in the software let you automatically navigate key relationships, concepts, and timelines that you might otherwise overlook.

# TRANSFORMING TEXT INTO ACTIONABLE INFORMATION

## A DEEPER UNDERSTANDING OF LANGUAGE

Figure 1: A Complete Solution for Text Analysis



Designed to handle the extremely high throughput and accuracy demands of today's information-intensive enterprise environments, BusinessObjects Text Analysis is available as part of a server-based solution or as individual software development kits (SDKs). An open and flexible application interface simplifies integration of technology from Inxight Software Inc. into your own applications to give you a distinct competitive edge.

BusinessObjects Text Analysis structures language into its most basic parts through automatic language and character encoding identification, document analysis, word segmentation (tokenization), stemming, normalization, decomposing, part-of-speech tagging, and noun phrase extraction.

These functions power downstream operations such as entity and fact extraction, categorization, and summarization – which structure text for more effective use in business intelligence, search, data mining, legal discovery, storage, and routing solutions.

### Powerful Entity Extraction and Analysis

With more than 35 preconfigured data items, BusinessObjects Text Analysis can automatically extract key information – including people, places, companies, dates, measurements, currency figures, and e-mail addresses – and reveal important relationships and events among those data items (see Figure 2). The software can put disparate items together even when relationships are not explicit.

You can extend BusinessObjects Text Analysis to recognize custom entities and entity types – such as SKU numbers, project names, special relationships and events, or specific phrases for sentiment analysis.

### Taxonomy-Based Document Categorization

BusinessObjects Text Analysis also provides taxonomy-based, document-level classification of information that is not explicit within documents. For example, a document may relate to a customer complaint category – even though the word “complaint” never appears in the text. By identifying such types of conceptual relationships, the software can automatically place your information in the proper category.

### Automatic Document Summarization

BusinessObjects Text Analysis can also summarize the most relevant sentences in a document or set of documents. An automated abstract creator will prepare an executive report on the content. The software can interpret the general substance of information by giving more weight to the first and last sentences of a paragraph, for example, along with the specific query parameters of your search.

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

Figure 2: Providing Meaning from Text Sources in a Readily Understood Format

# BETTER INFORMATION FOR BETTER DECISIONS

BusinessObjects Text Analysis can also weigh text information according to each user's specific area of interest. Consider, for example, a search related to the word "snow" in a document. One user might be interested in a specific weather phenomenon. Another might want to investigate weather-related transportation patterns of long-haul trucks.

## Easy Integration

BusinessObjects Text Analysis was designed for simple integration into any environment – whether SAP or non-SAP. Available as an easily administrated, distributable Simple Object Access Protocol (SOAP) server or a C++ SDK set, BusinessObjects Text Analysis can:

- Integrate with business intelligence platforms, either at the data integration or "universe" layer
- Provide output into databases, such as Oracle Express, that can then be leveraged for downstream data-mining operations
- Tightly integrate with applications for storage, discovery, or link analysis

## Improve Revenue and Operational Efficiency

Text comprises 85% of the typical organization's digital assets, according to industry estimates. Including these assets in your strategies for data management and business intelligence provides the comprehensive insight into your customers, markets, and internal operations you need for effective business decisions.

With BusinessObjects Text Analysis, you gain a more complete view of your organization's performance and customer relationships. Access to text information related to service and sales will help you understand why products fail or sales are lost. Examining information hidden in customer surveys and e-mails will provide useful guidance or increasing customer satisfaction. A daily summary of the mentions of your enterprise on key blogs and Web sites can help you create more effective marketing messages.

Better access to your text assets can also help ensure compliance with company directives and various regulatory requirements. You can more easily monitor internal company communications, identify potential leaks of customer information, and find information for legal queries.

## Optimize Your Text Assets

The large amount of information available to your business today is only useful to the extent that you can find, structure, and understand it. BusinessObjects Text Analysis can extract, categorize, and summarize text information from a wide range of document types – in a variety of languages. This powerful tool can reveal the hidden relationships and events that lie untapped in both internal and external text materials – providing the information you need for better decision making and compliance.

To learn more, please visit our Web site at [www.businessobjects.com/product/catalog/text\\_analysis/](http://www.businessobjects.com/product/catalog/text_analysis/).

## Extractable Languages

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

---

## Summary

BusinessObjects™ Text Analysis software can extract key enterprise information in more than 30 languages from unstructured text sources such as e-mails, Web-based documents, and customer documents. This information – unavailable through standard search engines or traditional business intelligence and document management software – can help you more reliably meet your compliance requirements and make better business decisions.

---

## Challenges

- Improve business decisions and compliance through better information
- Find, structure, and analyze untapped text in sources such as e-mails, documents, notes fields, and Web content
- Integrate data from these text sources into various information management solutions

---

## Supported Business Processes and Software Functions

- **Information collection and structure** – Use business intelligence, search, data-mining, legal discovery, storage, and routing solutions more effectively
- **Extraction and analysis** – Stay apprised of new or changing information as it develops and, once information is collected, automatically navigate key relationships, concepts, and timelines that might otherwise be overlooked
- **Custom formatting** – Improve recognition of information such as SKU numbers, project names, special relationships, or specific phrases through customized data extraction and analysis
- **Taxonomy-based categorization** – Properly classify even nonexplicit information, such as customer complaints
- **Automatic document summarization** – Simplify creation of executive content reports with automatic summaries of the most relevant sentences in a document or set of documents

---

## Business Benefits

- New value-added products and revenue streams
- Reduced costs for development, compliance, and litigation
- Improved understanding of customers and internal operations
- Faster time to market for information-intensive applications

---

## For More Information

Call your SAP representative or visit us online at [www.businessobjects.com/product/catalog/text\\_analysis/](http://www.businessobjects.com/product/catalog/text_analysis/).

50 088 690 (08/03)

©2008 by SAP AG

All rights reserved. SAP, R/3, xApps, xApp, SAP NetWeaver, Duet, PartnerEdge, ByDesign, SAP Business ByDesign, and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP AG in Germany and in several other countries all over the world. All other product and service names mentioned are the trademarks of their respective companies. Data contained in this document serves informational purposes only. National product specifications may vary.

These materials are subject to change without notice. These materials are provided by SAP AG and its affiliated companies ("SAP Group") for informational purposes only, without representation or warranty of any kind, and SAP Group shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP Group products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.