

# TURNING UNSTRUCTURED TEXT INTO INSIGHT

Extending Business Intelligence With Text Analysis and Search

## EXECUTIVE SUMMARY

While traditional business intelligence (BI) has transformed business using structured information from operational applications and transaction systems, there is a huge source of information that has by and large been ignored—people's thoughts and opinions, found in communications such as emails, web pages, reports, surveys, customer relationship management (CRM) note fields, contracts, blogs, wikis, and reports. Whether it's customer complaints, employee feedback, analyst opinions, or competitors' intentions, this valuable information lies hidden in unstructured text sources.

Just as multiple BI techniques have emerged for maximizing the insights found in structured data, such as reporting on operational transactions, or using management dashboards to view trends and monitor key performance indicators, two distinct capabilities have emerged to address maximizing insight from unstructured text sources:

1. *Search*, for finding individual or sets of documents and files
2. *Text analysis*, which extracts meaning from hundreds to hundreds of thousands of unstructured text documents for BI consumption.

Until now, search and text analysis existed without native BI and data warehouse integration. While these standalone applications have merits and benefits, there are tremendous benefits of integrating search, text analysis, and unstructured data with your BI and data warehousing deployments. Benefits include lower total cost of ownership (TCO), more complete and trusted BI, and increased IT control.

There's now a way to use this unstructured text and get a more complete view of the business. Business Objects has delivered the first and only text analysis and search capabilities that integrate directly with end-to-end BI. This is especially interesting to existing Business Objects customers as these new capabilities integrate directly with their existing BusinessObjects data integration and BI deployments, making text-based insight available to users from the same familiar reports and metrics they already use today.

*Author: Jason Kuo*

*Audience: This paper is intended for IT and business managers who wish to learn more about the value of complementing quantitative structured information with qualitative unstructured information found in text sources.*

# STRUCTURED VERSUS UNSTRUCTURED TEXT DATA SOURCES

Traditional BI examines transactions, processes, and events from information stored in structured systems, such as operational systems and data warehouses. BI deployments have predominantly focused on optimizing your business operations and driving related decision-making. And while today's BI has transformed business using structured data, it has, as Dr. Ronen Feldman notes, "fallen short of providing visibility into 'cause, complaint, correction'" information.<sup>1</sup>

Organizations are now learning that a more complete view of their business—including people's opinions, thoughts, and ideas—lies untapped in their unstructured text sources. Text sources include emails, documents, notes fields, and web content. Some practical examples include customer call records, analyst reports, market blogs, news sites, CRM comments, wikis, and financial documents.

Feldman describes how leveraging unstructured text can provide a more complete view of the business:

"Unstructured data provides situational context around an event or set of events that answers the questions 'why' and 'how', essentially filling in the 'cause, complain, correction' knowledge cycle. Knowing the 'why' and 'how' empowers organizations to uncover hidden relationships, evaluate events, discover unforeseen patterns and facilitate problem identification for rapid resolution. Utilizing intelligence extracted from unstructured data enables organizations to avoid loss of profit margins due to preventable write-offs, customer churn, legal settlements, warranty claims, or inefficient product development cycles."<sup>2</sup>

Including unstructured text into your organization's decision-making framework isn't an academic topic for the distance future. Both line-of-business and IT departments project a sharp increase in using text to drive better decision-making. A recent survey by The Data Warehousing Institute (TDWI) clearly signals a shift toward organizations using more unstructured data sources.<sup>3</sup> Survey respondents say that over the next three years, data warehouse source priorities will shift from traditional structured sources, such as database systems and mainframes, to a variety of unstructured sources.

---

<sup>1</sup> Feldman, Dr. Ronen. "Unified Business intelligence: Part I, Voices of BI," *DM Review*, February 2005.

<sup>2</sup> Ibid.

<sup>3</sup> Russom, Philip. "BI Search and Text Analytics," The Data Warehousing Institute, May 2007.

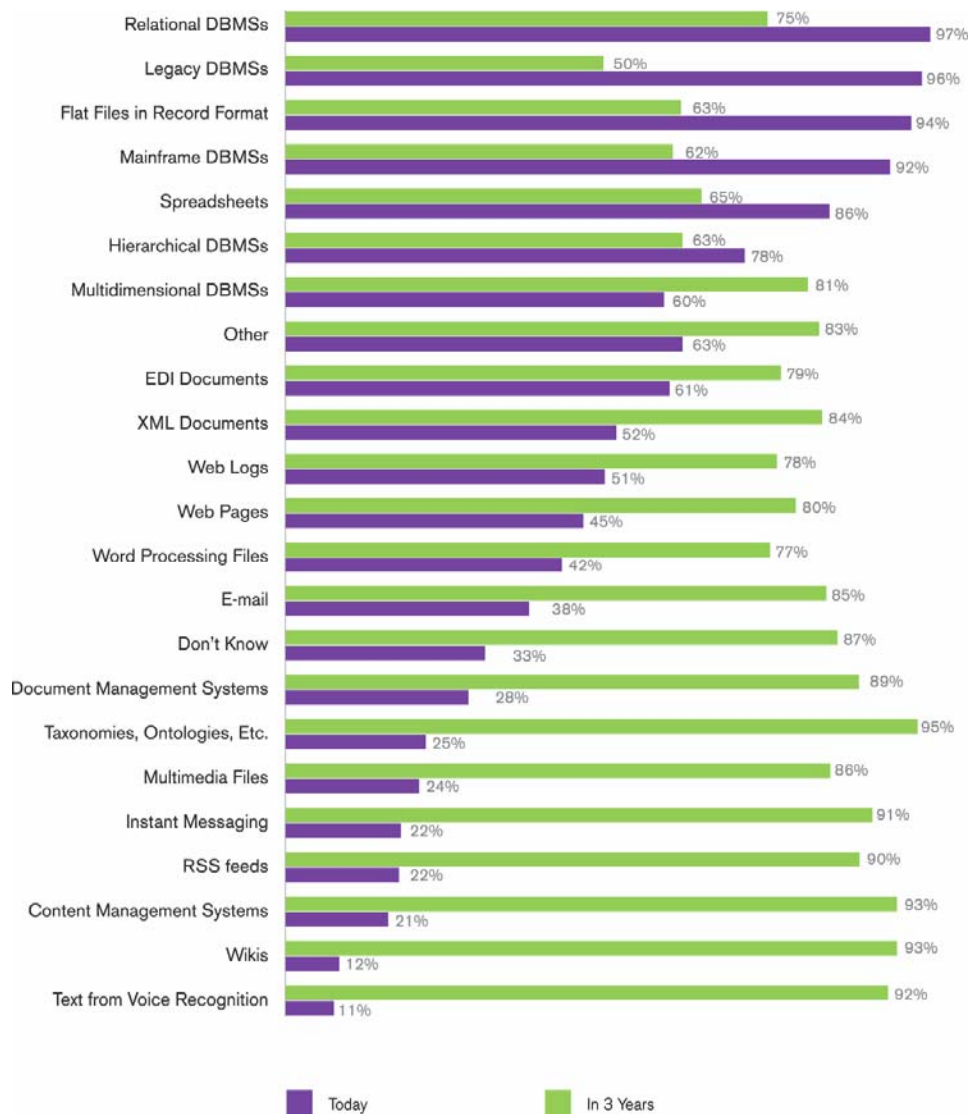


Figure 1. Organizations anticipate a gigantic increase unstructured data as sources for data warehouses. [Source: TDWI].

In a separate Ventana Research survey, 87% of respondents rate unstructured and semi-structured data types as somewhat important or very important.<sup>4</sup> Clearly, organizations have a high interest in incorporating untapped text data into their decision-making and performance management frameworks.

“Having access to both unstructured and structured information from the same application means businesses can finally get a complete view of their organizations,” says Sue Feldman, vice president for content technologies at IDC. “Unified access lets them make use of all the information they have, and find relationships across information silos. By mining otherwise untapped, invaluable information, they understand their customers better, lower their risks, and discover new business opportunities.”<sup>5</sup>

<sup>4</sup> “Business Intelligence and Search: Combining technologies to foster better business decisions,” Ventana Research, March 2007.

<sup>5</sup> “Business Objects Turns Text Into Insight,” [http://www.businessobjects.com/news/press\\_release.asp?id=20070924\\_006494](http://www.businessobjects.com/news/press_release.asp?id=20070924_006494), September 2007.

# CAPITALIZING ON TEXT WITH SEARCH AND TEXT ANALYSIS

Depending on your goal or the questions you're asking, you may drive your organizational performance and make decisions based on a small or large set of information. For example, your account manager wants to see what products a particular customer purchased last quarter, and your business unit general manager wants to know aggregate revenue by product and geography over multiple quarters. You need information available at both a granular, transaction level and at an aggregated level—using the same information aggregated with a large quantity of other transactions and granular data for broad summaries that show trends, relationships, and performance over time.

Organizations need both capabilities, and use end-to-end BI platforms, with usage-specific capabilities, to address their range of information requirements. For example, enterprise reporting deployments often focus on operational transactions while query-and-analysis and dashboard deployments often provide aggregated views of business performance over time. Consequently, a long-established requirement is for these various BI capabilities to be integrated and interoperable, allowing you to move from one usage type to another. For example, you may start with a dashboard showing a trend, and then perform root-cause analysis, drilling down to examine different slices of data, with the ability to drill down on individual records.

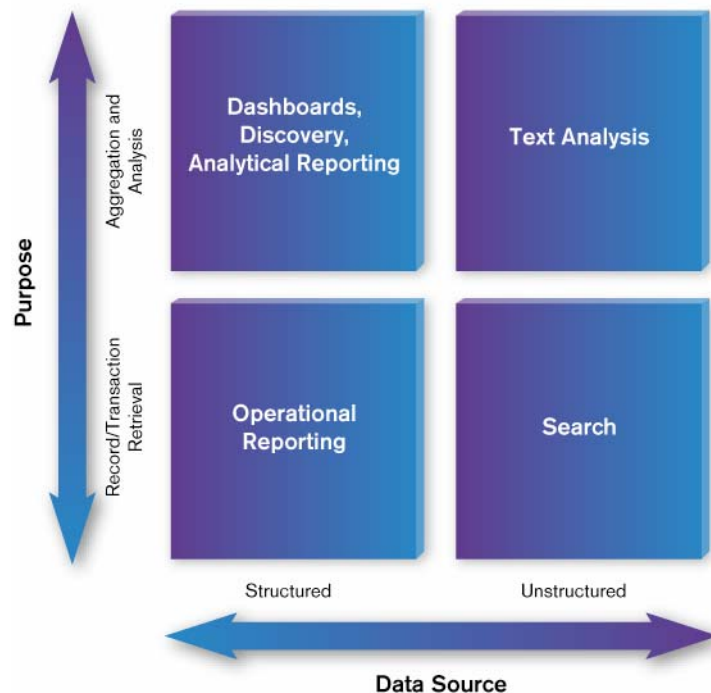


Figure 2. Both structured and unstructured data have their own distinct capabilities for record retrieval, aggregation, and analysis.

This same need for detail, aggregate trending, and interoperability applies to text sources as well. As people need the right tool to address the appropriate type of question, two distinct capabilities have emerged to address both of these requirements—search, for finding individual or sets of documents and files, and text analysis, which extracts meaning from hundreds to hundreds of thousands of unstructured text documents for BI consumption.

## Tapping Into Data Assets With Federated Search

While the ease and ubiquity of internet search has transformed the way we all access the web, numerous constraints limit the adoption of search within the enterprise.<sup>6</sup> These include a single search engine's ability to address enterprise security, complex dictionaries, and multilingual environments. In addition, search is mostly deployed in silos, with separate systems for content management, document management, BI systems, and corporate intranets. To address the need of finding individual or sets of documents across disparate sources, federated search has emerged as a way of simultaneously searching multiple internal and external sources via a single search query.

As described by TDWI Research, federated search "allows a best-of-breed search for individual systems and departments. Search indices can be isolated for security and performance. It avoids the cost and disruption of rip-and-replace. And information from individual search implementations is still visible and accessible through a central implementation."<sup>7</sup> In addition, security, dictionary, and multilanguage text processing have emerged as key requirements and considerations in implementing search of all sources and content type.

## Defining Text Analysis

Text analysis software "parses text and extracts facts (addresses, parts, complaints) about key entities (customers, products, accounts). Recognizing entities and facts about them involves natural language processing (NLP), which is a subfield of artificial intelligence that converts samples of human language into more formal representations that are easier for computer programs to manipulate. The facts and entities extracted via text analytics may be stored in a file, database, or search tool's index. Hence, text analytics (sometimes called entity extraction) imposes structure on information found in unstructured data sources and sometimes semi-structured ones."<sup>8</sup> After imposing structure, you can load the data into a data warehouse or data mart for BI consumption.

---

<sup>6</sup> Hodgkinson, Grant. "Six barriers to enterprise search," ITWeb, July 2007, <http://www.itweb.co.za/sections/industryinsight/contentmanagement/hodgkinson070705.asp>

<sup>7</sup> Russom, Philip. "BI Search and Text Analytics," TDWI, May 2007.

<sup>8</sup> Ibid.

# USAGE CASES

There are as many potential use cases for making decisions from unstructured text as there are for making decisions from structured data. Because so much is communicated in text and because so many text sources remain disparate and unavailable for decision-making, search and text analysis have proven to be very proficient at discovering unseen risks and unrealized opportunities. Here are some common usage scenarios:

## Customer

Whether it's finding customer records or looking at customer sentiment across large data sets, the insights you find in text can help your organization better understand customers—their likes, dislikes, and what motivates them. Search and text analysis can help you answer these customer-centric questions:

- How successful is the current marketing campaign?
- How satisfied are your existing customers?
- Is there customer churn?
- What are people saying about your brands, products, or services?
- What do consumers want that's not available today?

## Products

The same customer-centric questions apply to products as well:

- What product or parts are failing?
- What products or features receive criticism?
- What products or features receive praise?

## Risk

Threats come in various forms depending on the industry and situation. Here is a list of risk-based usages where search and text analysis is used:

- Regulatory compliance
- Communications risk assessment
- Fraud detection
- Intelligence and counterterrorism

# INTEGRATING BUSINESS INTELLIGENCE WITH TEXT ANALYSIS, SEARCH, AND UNSTRUCTURED DATA

In a recent Ventana Research survey, 90% of CIOs and IT management believe it's important to use the same vendor for search and BI, and for text analysis and BI.<sup>9</sup> Until now, search and text analysis have existed without native BI and data warehouse integration. While standalone applications have merits and benefits, there are tremendous benefits of having search, text analysis, and unstructured data integrated with BI and data warehousing deployments. Integration and single-vendor benefits include:

- **Unified data quality.** Cross data source cleansing, matching, and consolidation ensures trust across all data assets.
- **Common metadata.** Cross data source business semantics ensures consistent, shared business definitions across all data assets.
- **Common administration.** Cross data source system management including data processing and auditing ensures ease of administration and low TCO.
- **Familiar reporting and analysis across all data.** Access and analysis of unstructured data from the same traditional reporting, analysis, and dashboard eliminates the need for user training and ensures shorter time to insight.
- **Combined BI on structured and unstructured data.** Reports and metrics that combine structure and unstructured data enable powerful analysis, trending, and insight that neither traditional BI nor text analysis alone could provide.
- **Search across all sources, including multiple BI deployments.** Fast, secure search across all internal and external sources, including the unique ability to search across multiple BI deployments.
- **Common decision-making framework.** A single, common platform and point of entry for quantitative and qualitative decision-making.
- **Single trusted solution provider.** Besides reducing costs in every aspect of evaluating, purchasing, implementing, and maintaining BI tools, a single trusted solution increases user satisfaction, business insight, and IT control.

---

<sup>9</sup> "Business Intelligence and Search: Combining technologies to foster better business decisions," Ventana Research, March 2007.

# BUSINESS OBJECTS—TURNING UNSTRUCTURED TEXT INTO INSIGHT

Business Objects delivers the first and only text analysis and search capabilities that integrate directly with end-to-end BI.

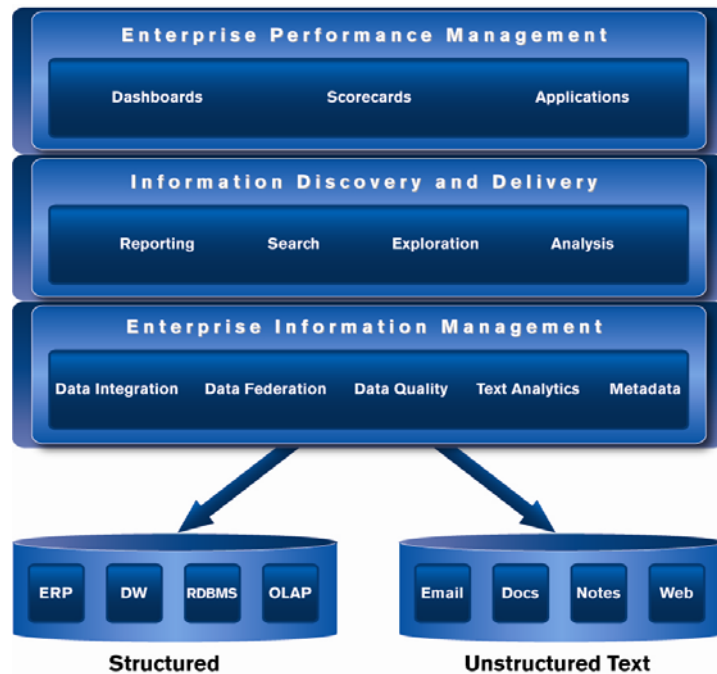


Figure 3. Get more complete and trusted BI, lower TCO, and increased IT control with the first and only BI suite that integrates text analysis, search, and unstructured data with end-to-end BI.

## DERIVING MEANING FROM UNSTRUCTURED TEXT FOR A MORE COMPLETE BUSINESS VIEW

BusinessObjects™ Text Analysis unlocks the information in text sources to provide a more complete view of your company's customers, markets, and potential risks. Powered by a deep understanding of over 220 text formats and more than 30 languages, BusinessObjects Text Analysis provides an automated way to extract, categorize, and summarize vast amounts of text information. These text analysis capabilities are designed specifically to complement and integrate with BusinessObjects Data Integrator and BusinessObjects Enterprise and deployments—ensuring unified trust and administration across all data assets.

Your users can access and analyze information contained in text in the same manner as they do their structured data—directly from the Crystal reports, BusinessObjects Web Intelligence® documents, and dashboards they use today, giving a more complete view of your business without needing to invest in entirely new systems or retrain users.

## **BUILDING INSIGHT FROM HUNDREDS OF HIGH-VALUE INFORMATION SOURCES THROUGH A SINGLE INTERFACE**

BusinessObjects Intelligent Search enables your users to search external and internal resources such as the web, subscription services, company content, search engines, and BI systems with one secure search box—allowing them to spend more time making decisions and less time looking for disparate files or snippets of information. Once this information has been acquired, the BusinessObjects Intelligent Search extraction capabilities cluster and enable users to filter results by the most relevant people, places, events, and concepts—making it easier to find relevant documents buried in search results, and turn that data into meaningful business information.

To learn more about BusinessObjects Text Analysis and Intelligent Search visit [http://www.businessobjects.com/products/platform/articles/072509\\_text.asp](http://www.businessobjects.com/products/platform/articles/072509_text.asp)

© 2007 Business Objects. All rights reserved. Business Objects and the Business Objects logo, BusinessObjects, Crystal Reports, Crystal Xcelsius, Crystal Decisions, Intelligent Question, Desktop Intelligence, Crystal Enterprise, Crystal Analysis, Web Intelligence, Rapid Marts, and BusinessQuery are trademarks or registered trademarks of Business Objects in the United States and/or other countries. All other names mentioned herein may be trademarks of their respective owners.

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

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

© 2007 Business Objects. All rights reserved. Part # WP3113-A

