SECOND QUARTER 2007
TDWI BEST PRACTICES REPORT

BI SEARCH AND TEXT ANALYTICS

New Additions to the BI Technology Stack

By Philip Russom

tdwi
THE DATA WAREHOUSING INSTITUTE

Research Sponsors

Business Objects

Cognos

Endeca

FAST

Hyperion Solutions Corporation

Sybase, Inc.

BI SEARCH AND TEXT ANALYTICS

New Additions to the BI Technology Stack

By Philip Russom

Table of Contents

Research Methodology and Demographics 3
Introduction to BI Search and Text Analytics 4
Defining BI Search 5
Defining Text Analytics 5
The State of BI Search and Text Analytics 6
Quantifying the Data Continuum 7
New Data Warehouse Sources from the Data Continuum 9
Ramifications of Increasing Unstructured Data Sources 11
Best Practices in BI Search 12
Potential Benefits of BI Search 12
Concerns over BI Search 13
The Scope of BI Search 14
Use Cases for BI Search 15
Searching for Reports in a Single BI Platform
Searching for Reports in Multiple BI Platforms
Searching Report Metadata versus Other Report Content
Searching for Report Sections
Searching non-BI Content along with Reports
BI Search as a Subset of Enterprise Search
Searching for Structured Data
BI Search and the Future of BI 18
Best Practices in Text Analytics 19
Potential Benefits of Text Analytics 19
Entity Extraction 20
Use Cases for Text Analytics 22
Entity Extraction as the Foundation of Text Analytics
Entity Clustering and Taxonomy Generation as Advanced Text Analytics
Text Analytics Coupled with Predictive Analytics
Text Analytics Applied to Semi-structured Data
Processing Unstructured Data in a DBMS
Text Analytics and the Future of BI 25
Vendor Products for BI Search and Text Analytics 26
BI Vendors 27
Search Vendors 28
Database Vendors 29
Pure-Play Text Analytics Vendors 30
Recommendations 31

About the Author

PHILIP RUSSOM is the senior manager of research and services at The Data Warehousing Institute (TDWI), where he oversees many of TDWI's research-oriented publications, services, and events. Prior to joining TDWI in 2005, Russom was an industry analyst covering BI at Forrester Research, Giga Information Group, and Hurwitz Group. He also ran his own business as an independent industry analyst and BI consultant, and was contributing editor with *Intelligent Enterprise* and *DM Review* magazines. Before that, Russom worked in technical and marketing positions for various database vendors. You can reach him at prussom@tdwi.org.

About TDWI

The Data Warehousing Institute, a division of 1105 Media, Inc., is the premier provider of in-depth, high-quality education and training in the business intelligence and data warehousing industry. TDWI is dedicated to educating business and information technology professionals about the strategies, techniques, and tools required to successfully design, build, and maintain data warehouses. It also fosters the advancement of data warehousing research and contributes to knowledge transfer and the professional development of its Members. TDWI sponsors and promotes a worldwide Membership program, quarterly educational conferences, regional educational seminars, onsite courses, solution provider partnerships, awards programs for best practices, resourceful publications, an in-depth research program, and a comprehensive Web site.

About TDWI Research

TDWI Research provides research and advisory services to business intelligence and data warehousing (BI/DW) professionals worldwide. TDWI Research analysts focus exclusively on BI/DW issues and team up with industry practitioners (TDWI faculty members) to deliver both a broad and deep understanding of the business and technical issues surrounding the deployment of BI/DW solutions. TDWI Research delivers commentary, reports, and inquiry services via TDWI's worldwide Membership program and provides custom research, benchmarking, and strategic consulting services to both user and vendor organizations.

Acknowledgments

TDWI would like to thank many people who contributed to this report. First, we appreciate the many users who responded to our survey, especially those who responded to our requests for phone interviews. Second, our report sponsors, who diligently reviewed outlines, survey questions, and report drafts. Finally, we would like to recognize TDWI's production team: Jennifer Agee, Bill Grimmer, Denelle Hanlon, Deirdre Hoffman, and Marie McFarland.

Sponsors

Business Objects, Cognos, Endeca, FAST, Hyperion Solutions Corporation, and Sybase, Inc. sponsored the research for this report.

Research Methodology

Report Scope. This report is designed for technical executives who wish to understand the many options that are available today—though still not used that much—for integrating unstructured data into data warehouse and business intelligence databases and tools. The report describes technical approaches to unstructured data—especially various forms of search and text analytics—and how they may be applied to BI purposes.

Survey Methodology. This report's findings are based on a survey run in late 2006, as well as interviews with data management practitioners, consultants, and software vendors. In November 2006, TDWI sent an invitation via e-mail to the data management professionals in its database, asking them to complete an Internet-based survey. The invitation also appeared on several Web sites and newsletters, and 401 people completed all of the survey's questions. From these, we excluded the 31 respondents who identified themselves as academics or vendor employees, leaving the completed surveys of 370 respondents as the data sample for this report.

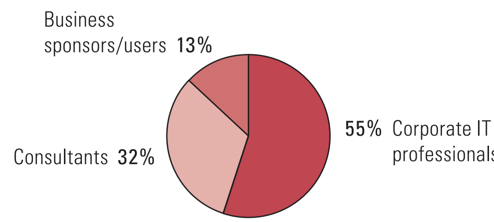
TDWI also conducted telephone interviews with numerous technical users and their business sponsors, and received product briefings from software vendors with products related to the best practices under discussion.

Survey Demographics. The wide majority of survey respondents are corporate IT professionals (55%), whereas the remainder consists of consultants (32%) or business sponsors/users (13%). Judging by how they answered survey questions, it's likely that most of the survey respondents have experience with data warehousing. But note that the best practices discussed in this report are new, so it's unlikely that respondents have had direct, hands-on experience with them.

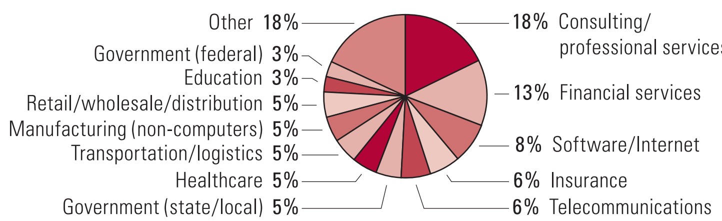
The financial services and IT consulting industries (31% combined) dominate the respondent population, followed by software/Internet (8%), insurance (6%), telecommunications (6%), and other industries (5% or less). We asked consultants to fill out the survey with a recent client in mind. By far, most respondents reside in the US (44%) and Europe (23%). Respondents are fairly evenly distributed across companies of all sizes.

Demographics

Position

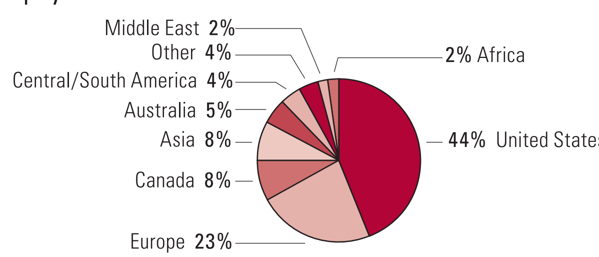


Industry

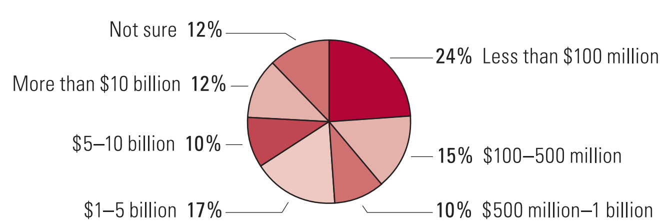


(The "other" category consists of industries with less than 3% of respondents.)

Geography



Company Size by Revenue



Based on 370 Internet-based respondents.

Introduction to BI Search and Text Analytics

The technology stack for business intelligence (BI) and data warehousing (DW) is currently expanding to accommodate two relatively new additions, namely BI search and text analytics. Although each stands ably on its own, the two are related in that they tend to operate on unstructured data. In fact, the growing use of BI search and text analytics is part of a larger trend toward leveraging unstructured data in BI and DW, fields that previously have relied almost exclusively on structured data. Another way to put it is that unstructured data is playing a larger role in BI and DW over time, and that role is today supported largely by tools and techniques for BI search and text analytics.

BI search and text analytics extend existing BI and DW investments.

This trend is not revolutionary or even evolutionary; it's accretive. In other words, BI search and text analytics certainly won't replace the traditional BI/DW technology stack. And it's unlikely that they will replace any components of the stack. Instead, BI search and text analytics are being added to BI/DW infrastructure to accommodate unstructured data (via text analytics) and related techniques (like search). Thus, most user organizations should have a mature BI/DW implementation in place before attempting to add BI search and text analytics to it.

BI search and text analytics impact different layers of the BI/DW technology stack:

- **BI Search.** Configurations vary, but most index the reports of one or more BI platforms to help end users more easily find whole reports and sections of reports. Hence, BI search affects one end of the BI/DW technology stack, especially how a BI platform enables report access at the GUI level.
- **Text Analytics.** Again, configurations vary, but most parse text containing human language and convert entities found there into some form of structured data. Hence, text analytics affects the other end of the BI/DW stack, especially how unstructured data is transformed into BI-friendly data structures at the database level.
- **Intersections of the two.** BI search and text analytics are largely complementary, but not mutually exclusive. For example, in some configurations, BI search may index databases in the data warehousing layers of the stack, and text analytics may create data structures that enable better-targeted search results in the business intelligence layers. (See Figure 1.)

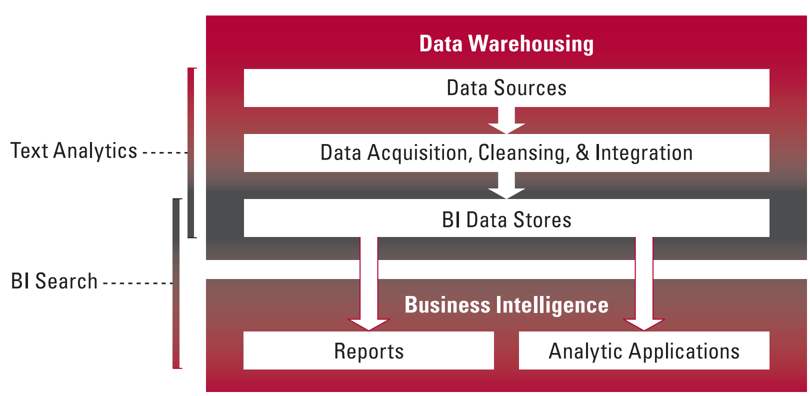


Figure 1. Layers of the DW/BI technology stack affected by BI search and text analytics.

This report defines BI search and text analytics, identifies their drivers, and describes several plausible use cases for each. This should help user organizations discover creative ways to extend their existing BI/DW technology stacks, as well as recognize business and technical requirements that are key to successful planning for such extensions.

Defining BI Search

The current definition of search. Search technologies have broadened considerably in recent years, as they have converged with related tools for text mining, taxonomy generation, topic clustering, portal infrastructure, and so on. When you shop on an e-commerce Web site, search is an easy-to-use interface to the *structured* data in the seller's product database. Furthermore, most search engines and their tools have expanded into new capabilities for natural language processing, entity extraction, data visualization, and various types of indexing. Most search engines today (whether for Internet or enterprise) support an array of capabilities that go far beyond the simple keyword indexing of the first search engines of the early 1990s. Hence, in this report, "search" refers to a multi-purpose technology that can access, index, and retrieve a range of data types (not just unstructured data) and present and manage that data in multiple ways.

A definition of BI search. We all associate search tools with unstructured data (typically in documents or files), because this is what they usually give us access to. This is true of "Internet search," where Web sites like Google and Yahoo! provide access to Web pages and other content on the Internet. It's also true of "enterprise search," which provides access to corporate documents like spreadsheets, presentations, and word-processing files.

But search tools can also extend the structured world of a BI solution by giving BI users a view into corporate documents beyond a single BI platform. Or the search engine could be turned inward on the BI platform to help users find just the right report, generate a query, or suggest related reports to peruse. The scope of search could reach across multiple BI platforms for a single access point into multiple bodies of reports. And some search tools can index metadata and metrics, giving BI users an easy-to-use equivalent of ad hoc query. These and other applications of search technologies in a BI context are called "BI search."

Defining Text Analytics

The ironic role of text analytics. BI as a business practice strives to be fact based, and usually number based. Hence, it's not enough to discover facts; you have to quantify them, too. This helps explain why BI—for most of its life—has focused almost exclusively on structured data, as expressed in reports and analytic data models. Furthermore, most BI tools for reporting and analysis operate exclusively on data in specific structures—like relational tables and multidimensional cubes—and most tools access these only through SQL.

While documents containing unstructured data can contribute to the decision making of BI, they cannot participate directly in its data-driven reports and analyses—unless facts discovered in unstructured data are extracted and transformed into structured data that's conducive to reporting and analysis. As precedence, we assume that data extraction, transformation, and load (ETL) are part and parcel of integrating structured data into a data warehouse or similar BI data store.

Unstructured data must be transformed into structured data before participating fully in BI.

We need now to extend that assumption to also encompass unstructured data and semi-structured data. They also require extraction to locate relevant entities and their facts—followed by transformation into appropriate data structures—before they can be loaded into a data warehouse and be useful for the traditional accoutrements of BI, like standard reports, multidimensional analyses, and statistical analyses. The curious irony is that this data is unstructured or semi-structured in its source form, yet must be transformed into structured data—via some kind of text analytics—before participating fully in BI.

A definition of text analytics. Text analytics is where a software tool 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."[1] 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.

Text analytics always involves the transformation of unstructured data into some kind of data structure. It may also take the next step, which is to provide an analysis of the resulting structure. Or another tool may effect the analysis, based on the data structure created by text analytics.

The State of BI Search and Text Analytics

Incorporating unstructured data and search technologies into BI sounds good. But do user organizations think it's worthwhile? Is anyone really doing it?

TDWI asked survey respondents to rate the business value of BI search and text analytics as they understand them today. With eager zeal, respondents rated the business value as high or very high for both BI search (62%) and text analytics (51%). (See Figures 2 and 4, respectively.) This ebullient bubble bursts when we consider their response to the next question: "What's the status of your organization's use of search in BI" and text analytics? (See Figures 3 and 5, respectively.) Very few have deployed either BI search (9%) or text analytics (6%). And relatively few have committed to a design or implementation phase for BI search (15%) or text analytics (7%). It appears that these practices are still very new to BI, so few user organizations have deployed or started building systems for them.

TDWI suspects that adoption of both BI search and text analytics—though rarely deployed today—will increase over five years, until they are as commonplace as Web GUIs and dashboards are today (although these were rare five years ago). Of the two, BI search is currently in the lead and will most likely enter the mainstream first.

USER STORY

BI users are aware of BI search and look forward to using it.

"We're like a lot of companies, in that we're looking into BI search, but we haven't done anything yet," said the general manager of a performance management group at a large financial services firm. "For years, we've been using two major BI platforms; one for general enterprise use and the other for financial analysis. Both have BI search embedded, so our plan is to start there with a pilot program and see what happens. I bet that once we turn this on, users will love it so much that we'll have to give them more."

[1] Source: Wikipedia at http://en.wikipedia.org/wiki/Natural_language_processing.

Rate the business value of BI search, as you understand it today.

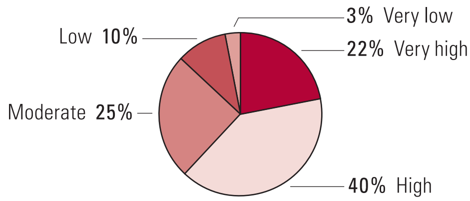


Figure 2. Based on 370 respondents.

What's the status of your organization's use of search in BI?

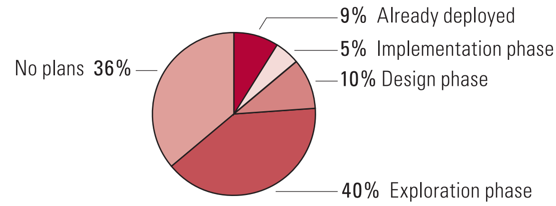


Figure 3. Based on 370 respondents.

Rate the business value of text analytics, as you understand it today.

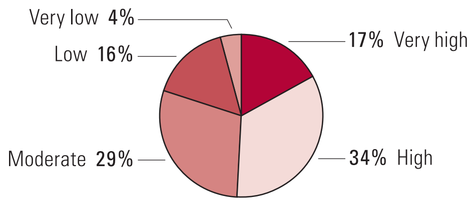


Figure 4. Based on 370 respondents.

What's your organization's status in the use of text analytics?

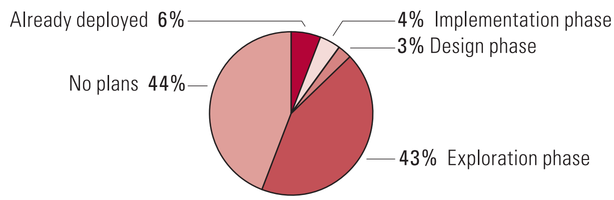


Figure 5. Based on 370 respondents.

Quantifying the Data Continuum

Before drilling into BI search and text analytics, we need to review the spectrum of available data sources. After all, the "data continuum" has direct import on the scope of reports and other documents indexed by search or mined by text analytics.

The data continuum breaks into three broad areas.

- **Structured data**. At one extreme of the data continuum, structured data is commonly found in database management systems (DBMSs) of various types.
- **Unstructured data.** The other extreme includes documents of mostly natural-language text, like word-processing files, e-mail, and text fields from databases or applications.
- **Semi-structured data.** The area between the two extremes includes semi-structured data in spreadsheets, flat files in record format, RSS feeds, and XML documents. Many of these media are used with cross-enterprise data-exchange standards like ACORD, EDI, HL7, NACHA, and SWIFT.

Some data sources are hybrids that are hard to categorize. Despite the three broad types of data sources, the continuum includes sources that can manage both structured and unstructured data. For example, a row in a database table has a well-defined record structure that defines fields of mostly numeric data types. Yet, the same record may also have fields that are character data types, like text fields or binary large objects (BLOBs). Likewise, a report may contain structured data (or a query that fetches structured data), as well as report metadata and text in headings that can be searched. RSS feeds are especially problematic, since they can transport a variety of information, ranging from prose (unstructured) to transactions (semi-structured).

In recent years, market research conducted by various software vendors and consulting firms has attempted to quantify the relative percentage split between structured and unstructured data in the average user organization. Most estimates name unstructured data the unqualified winner at 80–85%, leaving structured data in a distant second place at 15–20%.

Unstructured data volume is not as big as previously thought.

However, TDWI Research finds that unstructured data is not as overwhelming in volume as previously thought. In an Internet survey conducted in late 2006, TDWI asked each respondent to estimate "the approximate percentages for structured, semi-structured, and unstructured data across your entire organization." (See the top bar in Figure 6.) Averaging the responses to the survey puts structured data in first place at 47%, trailed by unstructured (31%) and semi-structured data (22%). Even if we fold semi-structured data into the unstructured data category, the sum (53%) falls far short of the 80–85% mark claimed by other research organizations. The discrepancy is probably due to the fact that TDWI surveyed data management professionals who deal mostly with structured data and rarely with unstructured data. All survey populations have a bias, as this one does from daily exposure to structured data. Yet, the message from TDWI's survey is that unstructured data is not as voluminous as some claim.

Unstructured and structured data in warehouses today.

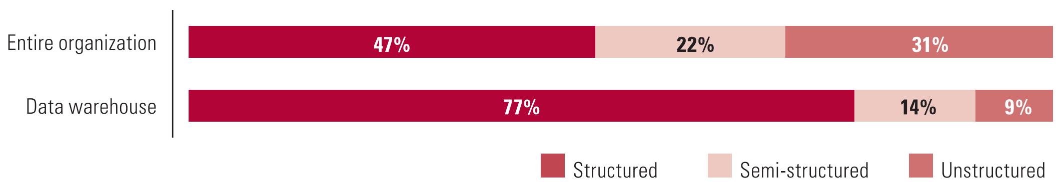


Figure 6. Little unstructured or semi-structured data makes its way into data warehouses today. Based on 370 respondents.

Unstructured data needs more representation in BI/DW.

Now that we have a new and different quantification of the unstructured segment of the data continuum, what should we do about it? We should all pare down our claims about unstructured data volumes, but we should *not* change our conclusions about what needs to be done. In other words, regardless of how the numbers add up, we all know that the average user organization has a mass of textual information that BI and DW technologies and business processes are ignoring. And this needs to change.

Why can't data warehousing professionals go on ignoring unstructured data? Among the many good reasons, two stand out:

- The view of corporate performance seen from a data warehouse is incomplete unless it represents (in a structured way) facts discovered in unstructured and semi-structured data.
- BI platforms today commonly manage thousands of reports, and techniques borrowed from unstructured data management (i.e., search) can make reports a lot more accessible.

To quantify the situation, TDWI asked each survey respondent to estimate "the approximate percentages for structured, semi-structured, and unstructured data feeding into your organization's data warehouse or BI processes." (See the bottom bar in Figure 6.) The survey responses reveal that structured data accounts for a whopping 77% of data in the average data warehouse or other BI data store, darkly overshadowing semi-structured (14%) and unstructured data (9%). Indeed, little data originating in unstructured or semi-structured form makes its way into data warehouses today, despite large quantities of it elsewhere in an organization. (Figure 6 compares these.)

The dearth of unstructured data in the warehouse isn't surprising, considering that almost all best practices in data warehouse modeling demand structured data. Likewise, we analyze and report off of data warehouse data using tools that see data only through the eyes of SQL, which in turn demands data in relational or multidimensional structures. As we'll see in detail later in this report, you have to impose structure on unstructured data before it's usable with a BI/DW technology stack.

Data is usable for BI only when structured.

New Data Warehouse Sources from the Data Continuum

As we've seen, the data continuum divides into three broad segments for structured, semi-structured, and unstructured data. In turn, each of these segments is populated by various types of systems, files, and documents that can serve as data sources for a data warehouse or other BI solution. These range from flat files, to databases, to XML documents, to e-mail, and so on.

To understand which of these are feeding data into data warehouses today—and in the near future—TDWI asked, "Which types of data and source systems feed your data warehouse?" Survey respondents selected those in use today, as well as those they anticipate using in three years. Figure 7 charts survey responses for both today and the future; it also calculates the expected rate of change (or "delta"). Judging by users' responses to this question, the kinds of data sources for the average data warehouse will change dramatically in the next few years:

- **Unstructured data sources will soon be more common for data warehouse feeds.** The survey predicts the greatest increases with technologies that convey natural language information in text (aka unstructured data), like voice recognition (up 81% in three years), wikis (81%), RSS feeds (68%), taxonomies (70%), instant messaging (69%), and document management systems (61%). Admittedly, some of these show a high rate of change because they're starting from almost nothing, as with voice recognition and wikis (11% and 12% today).

Prepare to extract and formalize data from unstructured data sources.

- **Semi-structured data sources will increase moderately.** This includes stalwarts like XML and EDI documents (up 32% and 18% in three years, respectively). The new kid on the block is the RSS feed, which contains both semi- and unstructured data. Most RSS feeds transport prose (unstructured data as text), but are beginning to carry transactions as semi-structured data in markup documents. Either way, 22% of survey respondents claim that their data warehouse accepts RSS feeds today, and 90% anticipate integrating data from RSS feeds in three years. This makes sense, because RSS feeds operate in near real time over the Web, and many organizations are looking for faster and broader ways to deliver alerts, time-sensitive data, and transactions.
- **Miscellaneous unstructured sources will increase moderately, too.** These are mostly files containing text, like e-mail (up 47% in three years), word-processing files (35%), Web pages (35%), and Web logs (27%). Their increase will be moderate because they're already established.

Structured data sources aren't going away—they're being joined by more unstructured and semi-structured ones.

- **Some sources of structured data may decline, but the category will keep its hegemony.** Survey respondents anticipate reducing data extraction from various older types of database management systems (DBMSs), namely those that are hierarchical (-15% in three years), mainframe (-30%), legacy (-46%), and flat files in record format (-31%). Indeed, these are legacy platforms that are ripe for retirement or migration. But survey respondents also anticipate extracting less data from spreadsheets (-21% in three years) and relational DBMSs (-22%). While the decline of legacy databases as data warehouse sources seems plausible, TDWI Research is deeply skeptical about the decline in relational databases and spreadsheets claimed by survey respondents. Since these are so deeply ingrained in BI and in IT in general—and are spawning new instances constantly—their decline seems very unlikely.

Which types of data and source systems feed your data warehouse? (Select all that apply for both today and in three years.)

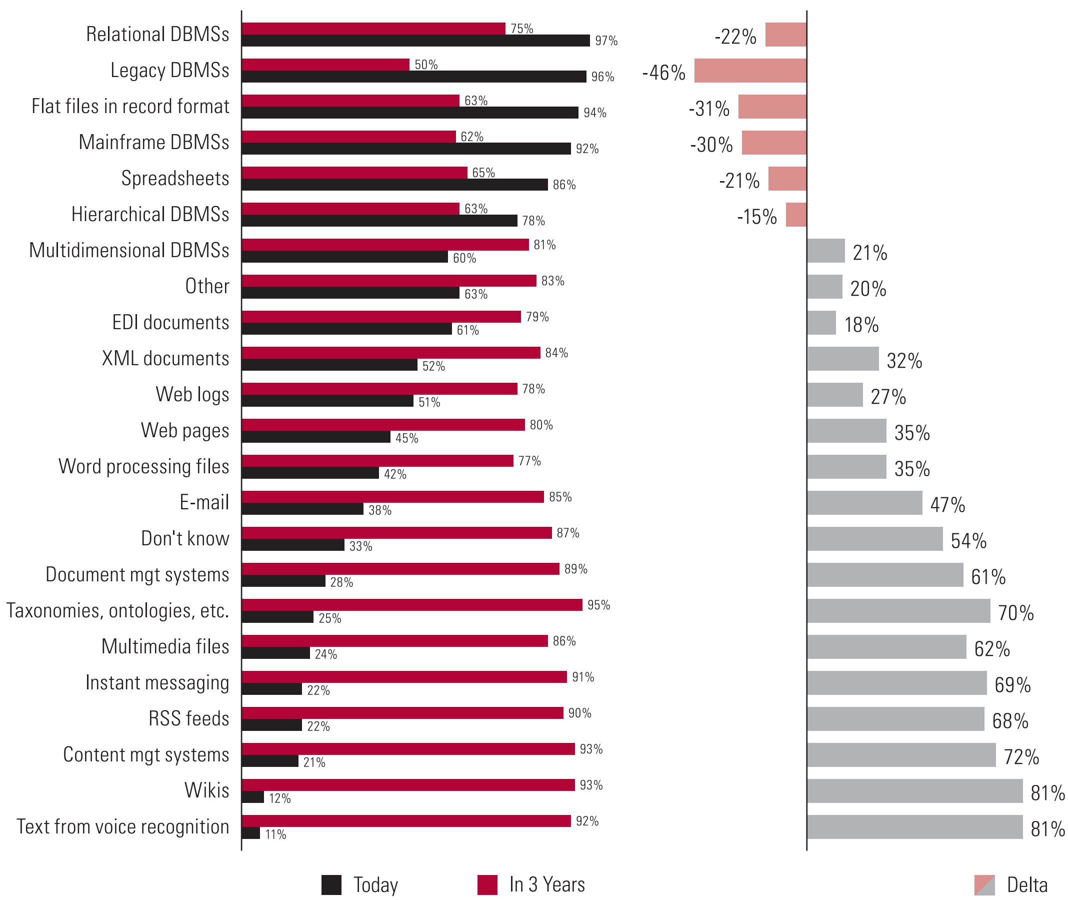


Figure 7. Based on 370 respondents.

The general trend—toward more unstructured data sources. Survey responses show that priorities along the data continuum will soon shift relative to data warehouse sources, with some data sources declining and others rising. Although respondents may have been overly optimistic about the rate of change they will embrace, the survey clearly signals a shift toward using more semi-structured and—especially—unstructured data sources. The trend is plotted conceptually in Figure 6, and the shift can be visualized as an increase in the types of data sources plotted in the middle or on the right side of the graph. Another way to see it is that the wide majority of data warehouse feeds today come from the left end of the graph. These won't go away, but instead will be joined incrementally by more data sources toward the right end.

Data and source types plotted on the data continuum

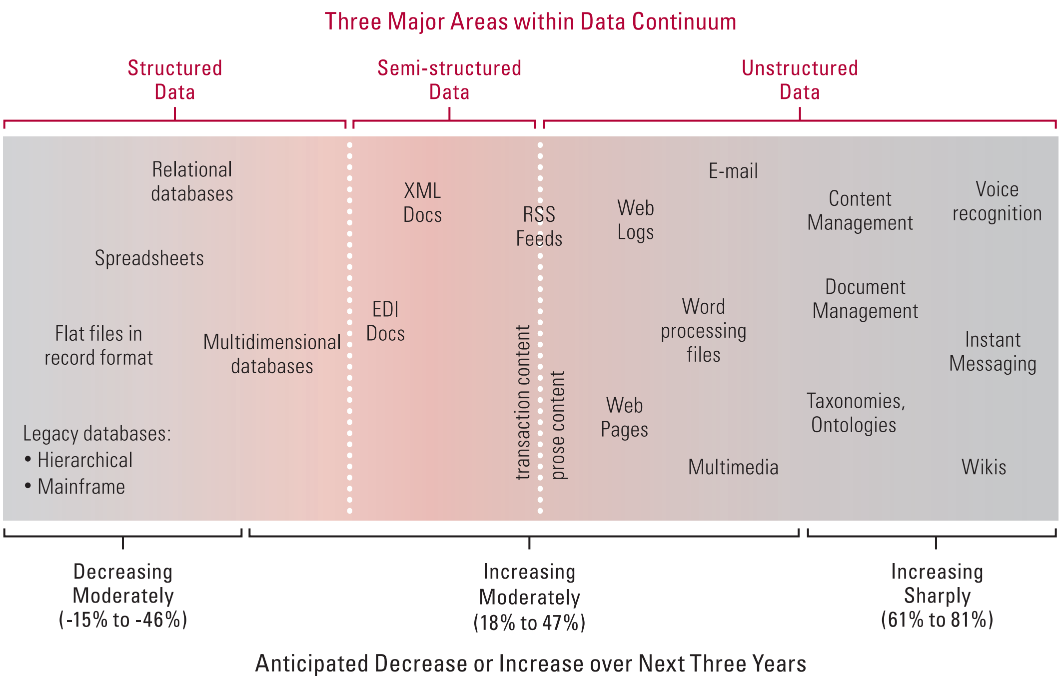


Figure 8. The data clearly signals a shift toward using more semi-structured and—especially—unstructured data sources.

Ramifications of Increasing Unstructured Data Sources

The evolving list of data sources means changes for DW/BI practices. Data warehousing professionals should be aware of these and prepare for them:

- **Unstructured and semi-structured data must be transformed into structured data.** Note that sources of unstructured and semi-structured data will be increasingly tapped for data warehousing, but that doesn't mean that much of this raw data will actually go into a data warehouse. In most cases, this source data will need to be parsed for entity extraction or otherwise transformed into structures that are meaningful in a data warehouse or to a reporting tool.

- **Data integration will need to change substantially.** The wide majority of data integration routines for data warehousing today interface with structured data sources and transform the data accordingly before loading it into the data warehouse. Assuming that unstructured sources will increase, data integration for the data warehouse will need to reinvent itself in the next few years.

- **Data modeling could face a similar transformation, but not as extreme.** A few data models in data warehouses will require adjustments to accommodate the structured data coming from unstructured data sources. Since the data is usually structured by the time it arrives in the data warehouse environment, adjustments should be slight. Similar adjustments are required when users want to copy unstructured data into a warehouse.
- **Training—and learning—are in order.** Data warehousing professionals currently have little or no experience with unstructured or semi-structured data sources. Likewise, experience is rare with search and text analytic tools. So additional training is needed, and—due to minimal experience—the learning curve will be long and flat.

Best Practices in BI Search

Potential Benefits of BI Search

There are many good reasons for implementing BI search:

- **Self service for report consumers is the best reason for BI search.** This conclusion is corroborated by users' responses to the survey question, "What are the potential benefits of BI search?" (See Figure 9.) Self-service report and information discovery (68%) rose to the top of the list. Furthermore, in the user interviews TDWI conducted, technical users identified self service for report consumers as their primary reason for implementing BI search.
- **BI search is mostly about finding pre-built reports, not creating new ones.** Relatively few respondents (26%) felt that BI search would result in less report authoring, although a few vendor offerings now enable users to construct new reports from data or report sections discovered through BI search.
- **BI search unearths more facts for decision making.** A high percentage of respondents thought BI search would lead to decisions based on a broader range of information (57%). After all, BI is about making fact-based decisions, and BI search can reveal more facts, regardless of where those facts are stored. Depending on how BI search is configured, it may also enable a broader range of possible questions (37%).
- **Users associate ease of use with BI search.** As a wider range of end users joins the BI community, BI platforms are increasingly used by new or casual users who are not expert in where reports are stored. Even when they know the system, navigating the complex taxonomy in which reports are organized in a portal or BI platform can be daunting. For these situations, BI search gives end users a simple user interface (44% in Figure 9) that helps them find reports and related content. When BI search indexes BI data stores (which is rare today), it may provide an easy-to-use substitute for ad hoc query tools, as cited by 39% of respondents. Given the technical skills required of ad hoc query, any simplification is welcome. If simplification of BI search is effective, it might result in a greater user adoption of BI (55% in Figure 9).
- **BI search can be a poor man's BI portal.** Deploying a search engine on a corporate intranet is an easy task, and integrating one with a body of reports isn't much more difficult. Yet users surveyed don't see ease of deployment (10% in Figure 9) or ease of maintenance (11%)

as potential benefits. And half of respondents fear that BI search is complex (49% in Figure 10). Even so, one user interviewed in this research felt strongly that BI search was a far less complex and costly path than building a full-blown BI portal. In fact, he referred to BI search as "a poor man's BI portal."

Despite its apparent benefits, BI search won't cure all ills. "BI search is great when you don't know what you're looking for, or can't find what already exists," said Wayne Eckerson, TDWI's director of research and services. "But if the average business user needs a search tool to find basic information or reports about the processes they manage on a daily basis, there is something wrong with the business!" In other words, managers should know exactly what information will help them run a line of business and receive this information customized for them in a performance dashboard. If they don't know what they need or where to find it, this is a symptom of a grander malaise than BI search can cure.[2]

EXPERT COMMENT

BI search may be a symptom of a grander malaise.

What are the potential benefits of BI search? (Select five or fewer.)

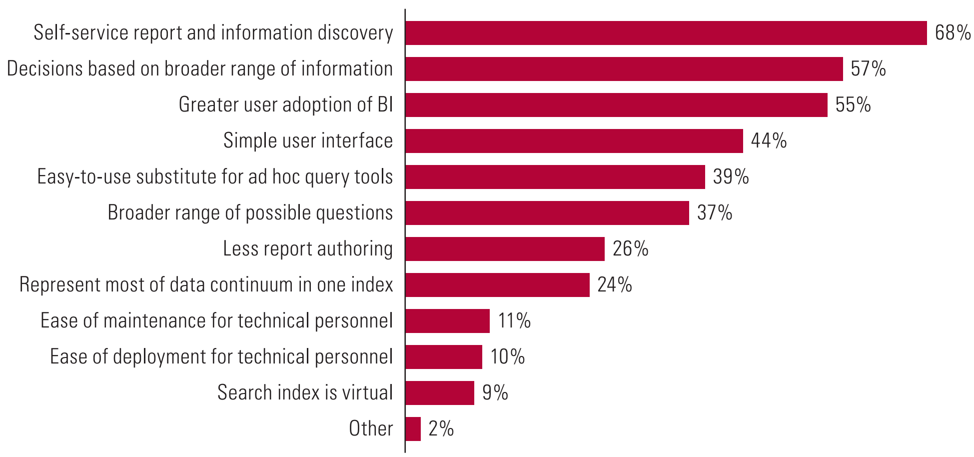


Figure 9. 1,411 responses from 370 respondents. (3.8 responses per respondent)

Concerns over BI Search

Although users perceive benefits from BI search, they also have valid concerns. (See Figure 10.)

- **Most concerns center on the quality of BI search's result set.** And rightfully so. In simple implementations of search, the value of the result set depends on human interpretation (56%), the relevance of the result set varies (51%), and the result set lacks structure for relations and analysis (44%). But these problems can be allayed to various degrees in advanced implementations. Some BI search tools can group reports and other items listed in the result set, thereby adding structure to the list that can guide the user to the most relevant groups. When topic or document clustering is applied to the result set, the clusters and their relations *are* an interpretation (and possibly an analysis) of the result set. Groups and clusters are usually identified visually in the user interface, and the visualization may serve double duty as a facile navigation mechanism. Hence, these features can alleviate many of the problems of BI search's result sets. Alas, to get these, you'll probably need to acquire

[2] See Wayne Eckerson, "The Real Value of BI Search," *What Works* Volume 22, November 2006, online at: www.tdwi.org/publications/display.aspx?id=8095

a fairly advanced (and more costly) tool, and then devote far more development time than a simple implementation would require.

- **Complexity (49%) and cost (29%) are common concerns.** Simple implementations of an enterprise search tool (where you install the engine on a server and point its crawlers at files and other sources) are straightforward and cheap, compared to other enterprise software implementations. It's the advanced implementations of search (which require custom development of lexicons, rules, and interfaces) that incur high complexity and cost. According to users interviewed in this research, integrating a BI platform with a third-party search engine is one of the highest risks. Therefore, users should look to their BI vendor for guidance, since many of the functions discussed here are available now (or will be soon) directly from the BI platform. If integrating an external search engine is desirable or inevitable, users should stick to the ones that the BI platform supports explicitly through a dedicated interface.

What reservations do you have about applying search in a BI context? (Select five or fewer.)

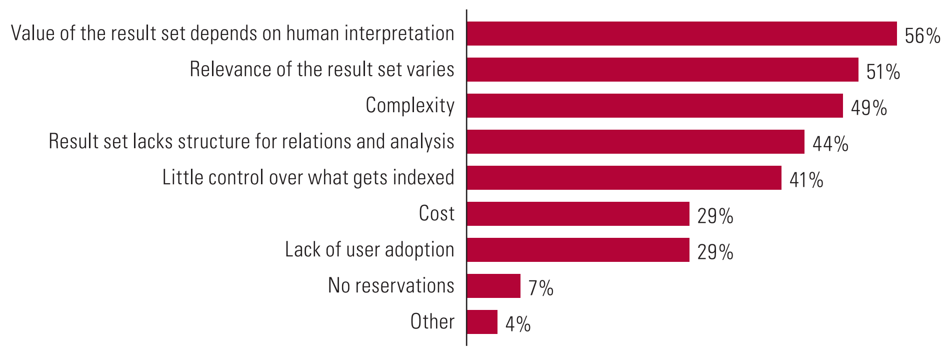


Figure 10. 1,417 responses from 370 respondents. (3.1 responses per respondent)

The Scope of BI Search

Defining the scope of searchable content is basic to all search implementations, including BI search.

The list of files, documents, reports, and systems (and their component parts) indexed by a BI search implementation constitute its scope. The scope of the implementation determines many things, including what's visible through the search index (all else is invisible) and what elements of the indexed source are included in the index (which influences how users search and how relevant result sets are). This definition of scope applies to both simple and advanced search implementations, and scope in a well-planned implementation will start small and grow over time. Hence, defining scope for multiple project phases is one of the most important steps in planning any implementation of BI search.

With scope in mind, TDWI Research asked, "What reports and documents should be searchable?" (See Figure 11.)

- **Users want to integrate multiple BI systems via BI search.** An overwhelming majority of survey respondents said they'd like to index and search reports within *all* BI platforms across an enterprise (76%), while considerably fewer would settle for just the reports within a single BI platform (29%). This is contrary to current practice, since most BI search implementations today index a single brand of BI platform, usually just a single instance. In organizations that have multiple brands of BI platforms, an independent implementation of BI search could provide a single view into multiple bodies of reports, so that no facts are missed before making a decision.

- **Report metadata (69%) is more desirable than report content (52%) for BI search.** In most BI platforms today, report metadata takes the form of properties defined for a report object. Some properties may define data sources for queries, while others are text fields that can be indexed as keywords. Indexing the full text of each report catches keywords and phrases from headings in the report that may not appear in report metadata. When possible, both metadata and full text should be indexed. In a heavily object-oriented BI platform, the two may be indistinguishable.
- **Half of users want to index databases (55%).** This is about indexing database metadata and properties, not the actual data. The point is to enable data discovery and possibly an equivalent of ad hoc query.
- **Demand is low for non-report documents in the scope of BI search.** Survey respondents ranked relatively low the prospect of indexing all non-report documents in the enterprise (29%) or selected bodies of non-report documents (28%).

When applying a search tool in a BI context, what reports and documents should be searchable? (Select 5 or fewer.)

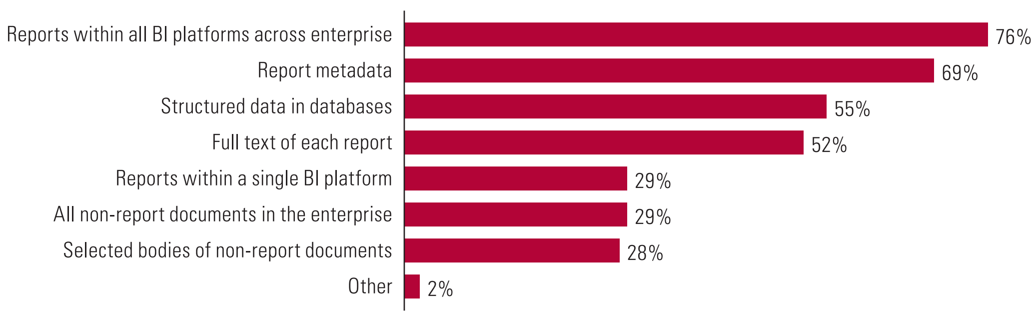


Figure 11. 1,261 responses from 370 respondents. (3.4 responses per respondent)

Use Cases for BI Search

These use cases are largely configurations of BI search, based on the scope of search, plus integration with other systems like BI platforms, enterprise search tools, intranets, portals, databases, data warehouses, and so on. The list gives users and their business counterparts examples of how they can leverage BI search, as well as technical requirements and business reasons for each use case.

Single-platform BI search is the place to start, though not the place where most users want to stay.

Searching for Reports in a Single BI Platform

This is the simplest configuration of BI search (because there's only one data source), which explains why it's the most common one today. Most BI platforms embed a search capability that can be turned on quickly and easily. In case the embedded function isn't enough, these platforms also support interfaces to popular third-party search engines, making their integration with the BI platform relatively fast and risk-free (though not license-free!). Furthermore, searching a single BI platform is an attainable goal that serves well as a first-phase deliverable for a BI search implementation.

This technical solution addresses a number of problems that report consumers suffer:

- **BI search is better for some end users than navigating a hierarchical directory.** The portal-like user interface through which you access reports in most BI platforms presents directories, folders, and reports in a hierarchical arrangement designed by IT. Getting organized this way is necessary, but it becomes complex when thousands of reports are cataloged in hundreds of folders. Plus, some reports covered multiple topics, so they should be cataloged in multiple folders, though they seldom are. Hence, navigating a hierarchical directory (sometimes called a taxonomy or catalog) is so cumbersome (and so dependent on users' memories) that it regularly leads to a dead end. This is when BI search can come to the rescue by giving report consumers an alternate way of finding the report that has eluded them. But don't think that BI search will replace taxonomies: you're better off with both.
- **Self service requires less assistance from IT.** BI professionals devote a big chunk of their time to helping report consumers find data and reports. BI search helps reduce this resource drain, and (when successful) helps end users find reports faster than waiting for assistance.
- **Finding reports reduces report writing.** Finding relevant reports (whether an end user knew of their existence or not) helps avoid the creation of new (and possibly redundant) reports. But BI search will not get IT out of the report-writing business, which many wish to flee. BI search is mostly for finding pre-built reports quickly—seldom for creating new ones.

USER STORY

BI search helps casual users find 10,000+ reports.

"Once we built up to over ten thousand reports, it got near impossible to find a report without knowing its exact location in the BI platform's catalog," said the BI manager at a large Internet service provider. "About the time users started complaining, we upgraded to a version of our BI platform that has search embedded. So, we flipped on BI search and got great results immediately. Now all of our BI users rely on search. But the casual users get the most out of it, because they don't have the catalog structure memorized, like the power users do."

Searching for Reports in Multiple BI Platforms

Although companies have worked hard to consolidate BI platform brands and instances in recent years, many still have multiple platforms. When the same business entities and processes are represented in reports from multiple BI platforms, this is a barrier to gaining a complete view of corporate performance. Business analysts and other report consumers must hit multiple systems one at a time, hoping that they've found all the reports that are relevant. This task is simpler, faster, and less prone to error when BI search encompasses all BI platforms. Furthermore, this configuration of BI search helps end users associate related reports, regardless of their points of origin. And BI search might be a compelling stop-gap alternative to an expensive and disruptive BI consolidation project.

Multi-platform BI search is the most desirable—and hardest—use case.

Searching for reports across multiple BI platforms is an advanced configuration with a lot of challenges. The search functions embedded in a vendor's BI platform support only that vendor's brand. So, indexing multiple BI platforms will probably require a third-party enterprise search engine. In order for the complex system integration to work, that engine must support all the BI platforms involved, or have interfaces that can be adapted. Other sticky details involve security, scheduling crawlers, and the permission of platform owners. Despite the challenges, this use case provides a unique solution for organizations with multiple BI platforms.

Searching Report Metadata versus Other Report Content

Most BI platforms follow an object-oriented methodology that defines a report as a container object that (when instantiated) inherits multiple smaller objects. Each object has properties, and these range from query definitions to metadata to formatting options to section headings. As long as the properties are clean and consistent, they're a good source for indexing. Older systems that follow a template or other non-object strategy may separate metadata from report text. The point here is that (ideally) BI search should index all report metadata and text, plus distinguish container objects from others, so end users can search for and reuse whole reports or just sections.

Searching for Report Sections

Distinguishing reports from report sections—and representing them individually in search results—is critical to a use case where the end user finds report sections (not whole reports) and either stitches them together into a new report or (more likely) pastes them into other files like Microsoft Excel, PowerPoint, and Word. Just as a search's result set lists a report by its title, the result set might also list a section of a report by its section title, object title, or whatever identifier is available. Ideally, the user interface should enable the end user to click on a report section listed in the result set and drag it into a new report or other document. If the section involves a query, it may need to fetch fresh data. Note that this is a *very* advanced feature of BI search, so it will be a while before the vendor community supports this consistently.

Searching Non-BI Content along with Reports

Decisions are not just based on structured database records and reports populated from them. A variety of documents and files containing unstructured data also influence decisions, especially spreadsheets, word-processing files, presentations, and Web pages. For this reason, some companies deliver reports and analyses via a corporate portal, so that non-BI documents appear alongside reports at the user interface level. In fact, performance dashboards combine information from many sources this way, along with the usual accoutrements of portals, like news, weather, alerts, and so on.

One of the facets of BI portals is that their scope is almost always just a single BI platform, thereby excluding non-BI content. BI search suffers the same limitation in most configurations. So, when decision makers demand non-BI content intermingled with reports—as is typical with performance dashboards—consider implementing BI search as part of a larger corporate portal or enterprise search implementation (as explained in the next section).

"We need to search every repository, although priorities have to come first," said the enterprise data architect at a large financial services firm. "Enterprise content management and workflow repositories are obvious places to start, because they contain high-value information, and the highest business demand is to improve visibility and access for these. File services are next. Somewhere down the line, we'd get to BI report repositories and data stores. In 2007, I'll devote a lot of my bandwidth to evangelizing these use cases."

USER STORY

BI search is an unlikely priority when content systems have no search.

BI Search as a Subset of Enterprise Search

Like a lot of data-driven practices, enterprise search is currently struggling to become enterprise in scope. Search today is still deployed mostly in technology silos, like systems for content management, document management, collaboration, and corporate intranet. In many companies, this has led to a plague of search engine brands and instances that are barriers to the leading goal of enterprise search: a central index through which most corporate content is accessible.

There are different ways to address the swarm of search engines plaguing most companies:

- **Search consolidation.** This practice reduces the number of brands and instances, but stirs up turf wars and scalability issues, because you have to take search engines away from departments and replace them with a larger one controlled at the corporate level. It's also contrary to best-of-breed practices for tool selection.
- **Federated search.** This leaves individual search implementations in place when they are critical in some sense (not all are), then integrates their indices into a central, enterprise one. Federated search has a lot of advantages over consolidation. It allows 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 (though controlled locally) is still visible and accessible through a central implementation.
- **Hybrid search.** To achieve enterprise search while respecting departmental requirements, organizations need to selectively choose which search implementations to ignore, which to consolidate into a large central implementation, and which to leave as is but integrate with a central implementation through federation.

BI search can fit into any of these scenarios involving enterprise search. The catch, however, is that BI search has special requirements beyond those of enterprise search, like interfacing with BI platforms, indexing report metadata and text, and (in advanced cases) indexing query and database metadata. So, when implementing BI search with a general-purpose enterprise search product, be sure it can satisfy these special requirements.

USER STORY

Multiple, simultaneous approaches to search demand a multi-purpose enterprise search platform.

"I think BI search would be really useful, but I doubt that I can sell it to my colleagues," said the enterprise data architect at a large financial services firm. "They want to go Google, because they've heard the brand name, and it's dirt cheap. I'd rather have a multi-purpose platform that lets me start with enterprise-scope search, consolidate lots of search silos, then satisfy specific requirements like BI search, and later grow sidewise into related technologies, like text analytics and taxonomy generation."

Searching for Structured Data

Reports are templates or objects that contain no data until they're run. So, there's seldom data in a report for a search engine to index. Even if you could, indexing numbers and other values would be non-productive, since they change constantly, and a value out of context is meaningless. Instead, BI search indexes references to data structures in report metadata and text. Hence, BI search can help users find structured data in multiple ways:

- **Search for data structures referenced in report metadata or text.** For instance, a BI user may know which database table column contains the desired information, but cannot remember or does not know which reports fetch data from it. BI search can help the user find all reports referencing this data structure. This assumes the user is a power user who understands data structures and knows which one he/she needs.
- **Fetch structured data directly through BI search.** Even if a user finds a data structure referenced in a report, the user usually needs to run the report to fetch the freshest data. But some BI platforms enable BI search to fetch the data (and more will soon), so you should expect to use this feature regularly in the future. Hence, BI search has the potential to serve as the equivalent of a query tool.

There's already precedence for this outside of BI. For instance, when you search an e-commerce Web site looking for a product, you are usually searching the structured and unstructured data of a product catalog that resides in a database. Search provides a user-friendly interface to an otherwise inaccessible database. Depending on the configuration, search may create database queries (by interpreting the search string), retrieve data via a keyword index, or a combination of both.

BI Search and the Future of BI

Looking into our crystal ball for a moment, we see a possible future in which most BI will be operated through the search box. BI search can already find and deliver reports or sections of them, fetch data, and paste found items into new reports. If it catches on, BI search stands a good change of significantly changing report authoring and delivery. This would be a radical simplification of these two complex BI tasks, which in turn could help more new users embrace BI and older users be more productive.

Best Practices in Text Analytics

Potential Benefits of Text Analytics

There are many good reasons for implementing text analytics. To sort these out, TDWI Research's survey asked, "How would BI users benefit from data extracted from text sources?"

- **Leading benefits relate directly to BI dogma.** Given that BI is a fact-based methodology, it's not surprising that survey respondents feel that the leading benefits would be more facts for better decision making (68%) and broader information access and data discovery (66%). (See Figure 12.) Likewise, BI strives to assemble facts that tell the whole truth, so it's predictable that the next areas of anticipated benefits are a more complete view of the business (60%) and a more complete view of each customer (58%). The survey aside, this second group of benefits is the one that users pointed out in the interviews conducted for this report.
- **Corporate performance and regulatory reporting gain less from text analytics.** In recent years, BI methodology has moved steadily into corporate performance management. But survey respondents aren't confident that text analytics could lead to a better assessment of corporate performance (27%), possibly because extracted entities don't relate directly to performance metrics. Likewise, more accurate regulatory reporting (23%) scored low, despite the heavy use of text analytics in the automotive industry for compliance with the TREAD Act.
- **Text analytics for security and governance is coming.** A possible use case that interviewees discussed is to turn text analytics on e-mail and other corporate communications to find fraudulent activity (like insider trading) or data usage that's contrary to privacy or governance policies. These use cases scored low in the survey (16%), so they are not yet established.

How would BI users benefit from data extracted from text sources? (Select five or fewer.)

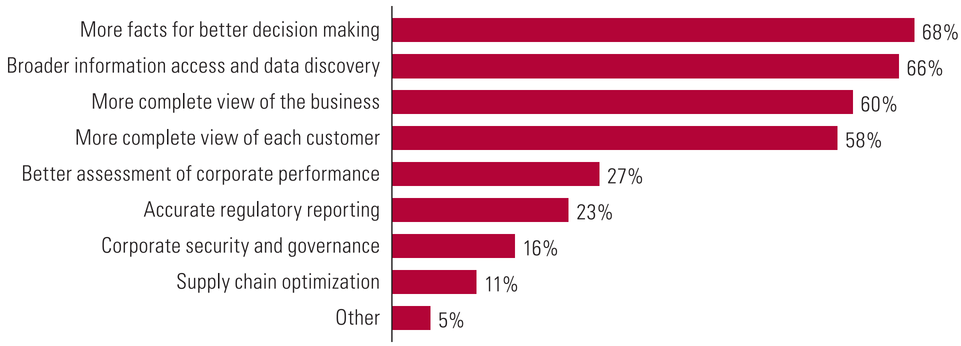


Figure 12. 1,241 responses from 370 respondents. (3.3 responses per respondent)

EXPERT COMMENT

Text mining's goal is knowledge, not data structures.

"Text analytics is about making human communications comprehensible to computers," said Seth Grimes, founder of Alta Plana Corporation and a renowned BI consultant. "Despite the inapt 'unstructured' label, textual documents have linguistic structure that is easily comprehended by people. Text-mining tools apply linguistic (natural language) techniques to parse this structure and model it in ways that computers can understand."

But that's not all, says Grimes. "Mining is merely a means to an end, and creating structure is not the goal. The goal is to discover knowledge—categorizations, relationships, predictive rules—and apply the knowledge to processing documents via clustering, routing, and analysis on extracted information. Put it all together, and text analytics becomes both technology and process—a mechanism for knowledge discovery applied to documents and a means of finding value in text."

Entity Extraction

Not surprisingly, the subjects commonly extracted via text analytics are the leading entities of any corporation, namely: customers, products, locations, and financials.

- **Customers are popular subjects for text analytics.** For example, when unstructured data comes from a call center application or service reports, text analytics parses text looking for customers' names (49%) or company names (21%) in a business-to-business scenario. (See Figure 13.) Once a customer is spotted, facts about this entity are gathered from the text, like customers' complaints (57%), customers' praise (33%), products acquired by the customer (33%), and failures in products/parts (36%). This information is then folded into a record for each instance of the customer found in the unstructured data source. These records are written to a row in a database table, added to an index for search, or passed directly to another tool.
- **Products are popular entities, too.** Imagine the example given in the previous bullet, but focused on product information. Text analytics could recognize instances of products (33%) and collect facts about them like parts of products (14%), failures in products/parts (36%), customers' complaints (57%), and customers' praise (33%). For example, automotive vehicle manufacturers parse service records to identify parts and failures of affinity, in an effort to improve quality and safety of products. When entities and their facts relate to each other this way (as in the interaction between customers and products), the NLP parser may need

to make multiple passes, and the data model of the target database or search index may need to be multidimensional.

- **Location data is important in some industries.** Twenty-nine percent of survey respondents expressed interest in mining text for data about locations. For example, insurance claims involve a lot of text describing a loss, like an automobile collision. Text analytics can parse for an entity like accident or policy holder, then associate location facts with it, like street name, house number, town name, and so on. When analyzed, this data could identify dangerous intersections, geographical areas prone to specific loss types (like burglary or flooding), or general demographics about losses. In turn, this could enlighten fraud detection, risk management, and actuarial assessments. Likewise, location data is useful in logistics, utilities, retail merchandizing, and military intelligence.

Which facts and entities are the most important for extraction from text sources? (Select five or fewer.)

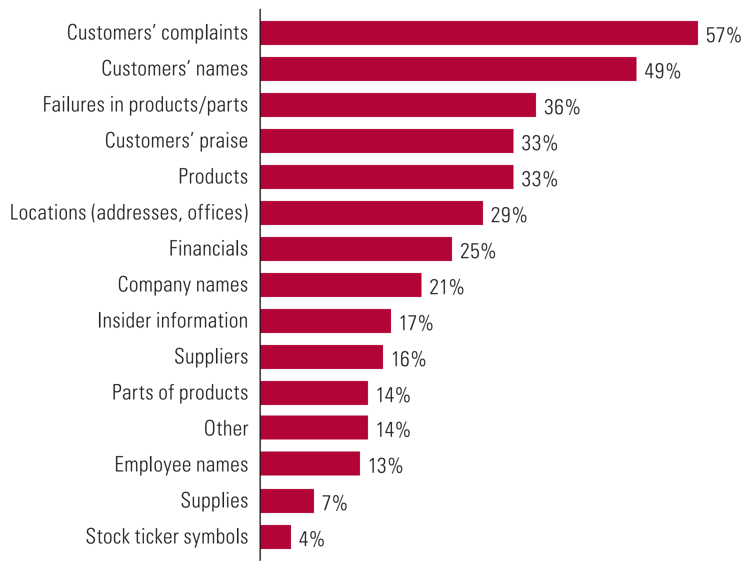


Figure 13. Based on 1,369 responses from 370 respondents (3.7 responses per respondent).

Text mining and text analytics are similar but have different uses.

Text mining versus text analytics. Text analytics is often confused with text mining because the two operate similarly, produce similar results, and are sometimes performed by the same vendor product. Although both are useful, text analytics is more appropriate to data warehousing and BI. Therefore, it's worth distinguishing the two. In a nutshell, text mining helps you discover the evolving list of entities that are mentioned in a text source, whereas text analytics extracts information about a constant list of entities. Furthermore, text mining tends to output a taxonomy that organizes the entities found, whereas text analytics outputs structured data as records, tables, databases, and so on. Both use some form of NLP and access text in files, documents, repositories, databases, content management systems, and so on. Table 1 compares various attributes of text mining and text analytics.

Text Mining and Text Analytics Compared and Contrasted

	Text Mining	Text Analytics
Approach	Both employ natural language processing (NLP) algorithms to parse text and record entities found there.	
When to use	When you don't know what's in the text or what you need to extract from it.	When you know what's in the text and what you need to extract from it.
Types of sources	With either, text can be anywhere accessible, including various types of files, repositories, databases, or content management systems.	
Quantity of sources	Most configurations operate on large numbers of documents, possibly across many repositories, to understand what topics these documents mention.	Most configurations operate on one or a small number of unstructured data sources, the content of which you understand well.
Output	A taxonomy of entities (with links to their sources). The taxonomy is analytic in nature, since it has classified the content it parsed and even organized the taxonomy by how found entities relate.	A stream of records, each describing an entity instance. These records usually go into a database table or file, less often into a search index.
Use of output	The taxonomy may double as a navigation medium through which users access source documents.	The resulting database or index is itself not analytic per se, but it's used as a source for standard reports or statistical analysis.
Applications	Classifying content in large bodies of documents (possibly in content management systems or portals) in media (especially publishing), research (especially pharma), federal intelligence agencies, etc.	Risk analysis, fraud detection, call center analysis, product service reports, regulatory reporting, customer base segmentation, other customer analytics, etc.
Synonyms	Content classification, topic clustering, taxonomy generation	Entity extraction, concept extraction, concept search

Table 1. Text mining helps you discover the evolving list of entities that are mentioned in a text source, whereas text analytics extracts information about a constant list of entities.

Use Cases for Text Analytics

These use cases are largely configurations of text analytics, plus integration with other systems, like data warehouses, integration tools, and predictive analytics applications. The list gives users and their business counterparts examples of how they can leverage text analytics, as well as technical requirements and business reasons for each use case.

Entity Extraction as the Foundation of Text Analytics

Entity extraction is the foundation of text analytics, upon which are built intelligent applications for entity clustering or taxonomy generation.

An absolute requirement of text analytics is entity extraction, which takes priority over secondary functions like classifying the extracted entities (via taxonomy generation) or analyzing them (via entity clustering). In fact, as we'll see from many of its use cases, text analytics is more like a data integration task than a data analysis task. This report has already discussed the types of entities this process reveals, as well as why collecting this data is important. Let's now look at how different use cases handle extracted entities:

- **Text analytics produces a record for each entity instance.** When a text analytics tool recognizes an entity in text, it looks for related facts (and possibly related entities) in surrounding text. It packages what it found (plus metadata about the data source, etc.) into a record structure that describes the entity instance. Hence, the entity record is structured data, which analytic tools require, although derived from seemingly unstructured data.

- **Entity records have to be stored somewhere.** Multiple storage options are available, depending on how entity records will be processed after extraction:
  - **Database.** Most configurations write entity records to a database management system (DBMS) so that multiple tools can operate on the data and a history is kept. The database may be modeled as a data warehouse or an operational data store.
  - **Flat file.** Instead of a DBMS, the text analytic tool may write entity records to a flat file. This makes sense when the output of text analytics becomes the input of tools that require or prefer flat files, as many predictive analytics and data integration tools do.
  - **Virtual database.** In some advanced configurations, the text analytics tool hands each entity record directly to a predictive analytics tool, which may temporarily cache the data in server memory as a virtual database.
  - **Search index.** When text analytics is performed by a search platform, entity records may be added to an index optimized for search or taxonomy generation.
  - **Exceptions.** In some cases, a text analytics tool inserts a tag in a source document instead of outputting an entity record. For instance, when the tool parses a document and finds a reference to a person, it inserts an XML or HTML tag into the document at that spot, and the tag contains metadata about the person. These tags improve the relevancy of search results, but don't necessary fulfill a BI purpose.
- **Data modeling is a big part of the work with text analytics.** There are many ways to use data from text analytics, and each use has its own data modeling requirements and challenges:
  - **Entity record models.** You can create a record for every instance of an entity, then group and merge these records later. Or, you can create a single record for each unique entity and fold instance facts into it. The former is verbose and generates a lot of data, but is very flexible later, if you have the resources to process and transform it.
  - **Multidimensional models.** Some text analytic tools can parse for many entities simultaneously and recognize that related entities are involved with the same fact (and vice versa). When these relations are described in a database or index, a multidimensional data model results. Similarly, a multidimensional data model may be required when end users with OLAP-based tools wish to analyze the output of text analytics.
  - **Reconciling with a data warehouse model.** The data model of the extracted entities inevitably varies from data structures in a data warehouse. So, when the entity database is a source for a data warehouse, its content requires additional processing to transform it into whatever's required of warehouse models. This is not necessarily easy, of course. When unstructured data is a new source for a data warehouse, it has to go through the usual process of consensus building around its model, usefulness, and accuracy. Many users handle mined data in a separate technology stack for a long time before figuring out how to reconcile it with the enterprise data warehouse and BI technology stacks.
  - **Application-specific models.** The structure of mined data must be conducive to the type of analysis or reporting that is the ultimate output of the task. For instance, text analytics applied to regulatory reporting produces relational tables for reporting. Text analytics and predictive analytics that perform clustering (as with customer base

analysis) require mined and extracted data in a format (typically a flat file in record format) that the tool's algorithms can understand. Many applications that rely on data from text analytics involve complex business models that are hard to represent in data models, like predictive analytic applications for risk and fraud. Each application has nuances for modeling.

- **Data flow.** Further processing of entity records is usually required when multiple tools (with multiple data model needs) will operate on it. In these cases, the usual best practices of data integration apply, such that entity records and related data may be subject to a data flow that stages and transforms data along the way.
- **Scalable storage.** Technical users interviewed for this report explained that when unstructured data is converted to structured data, the volume of storage increased by a factor of 2 to 10, due to the overhead of metadata and indexing. Be sure to run tests to quantify this before attempting capacity planning for storage and server throughput.

Entity Clustering and Taxonomy Generation as Advanced Text Analytics

Entity extraction is fundamental to producing the structured data that a BI technology stack demands. But text analytics tools may also perform more intelligent tasks, like entity clustering or taxonomy generation, which have their place in managing unstructured data for a BI purpose.

For example, entity clustering is very useful for the discovery of entities in unstructured data. Clustering entities shows you hot spots among entities found in the mined text sources, like recurring customer complaints and product problems. The clusters are analytic by nature, in that they reveal in what proportions entities are referenced in a body of text sources. This, in turn, helps you decide which unstructured data sources to eventually mine via text analytics, as well as which entities and entity relationships to parse for.

Taxonomy generation is likewise useful in advanced configurations of BI search. When applied to search results, taxonomy generation improves relevancy by organizing the result set. Entity clustering might also be applied to a BI search result set, producing an analytic view of it.

Text Analytics Coupled with Predictive Analytics

When interviewing users for this report, TDWI Research encountered a few users who've developed applications that include both text analytics and predictive analytics, although they sometimes call them text mining and data mining. Consider the following use case, told in the user's own words:

USER STORY

Analytics based on mining applies to both structured and unstructured data.

"A lot of claims data is unstructured, and you can learn a lot by mining it," said a risk manager at a large insurance company. "For example, when a loss is reported to our call center, a lot of text is collected, mostly rudimentary facts about the loss. This is logged in a transactional system, and activity logs collect more information as the claim is processed. Eventually, all data about a claim—whether structured or unstructured—goes into a claim master record or CMR. For years, people inside and outside the company have been able to view CMRs as they work a claim. Recently, we started mining CMRs to learn more about our customers and internal efficiencies, as well as to watch for risk and fraud. We spent a year figuring out the real value of the CMR and building reports. Now we're mining the CMRs' structured and unstructured data for predictive analytics.

"Our technical team uses the same tools and similar techniques, whether mining structured or unstructured data," the risk manager continued. "That's because both are simply data preparation steps that have to be done before we can run reports and analyses. Text mining—or text analytics or whatever you want to call it—is key to converting information in text to structured data that our analytic tools require. The output of text analytics is often an input for predictive analytics, showing that the two work well in tandem. Without text analytics, we'd be performing predictive analytics on half as much data, and getting only half the picture. Our primary business goals are to decrease the funds reserved for payouts and increase the funds acquired through subrogation. We also hope to improve fraud models and regulatory compliance. I feel confident that text analytics and predictive analytics will help us get there."

Text Analytics Applied to Semi-structured Data

This report focuses on unstructured data and how it can contribute to traditional BI and DW technology stacks. But semi-structured data is likewise useful for BI when processed by text analytics, as seen in the following user story. The story also shows that the natural progression from traditional data warehousing to predictive analytics to text analytics defines a low-risk phased approach to project implementation:

"We depend heavily on EDI for transactions and communication both inside our company and outside with partners," said the managing director of business information at a mortgage securities firm. "EDI documents contain a mish-mash of structured, semi-structured, and unstructured data, and all of it's useful for determining risk, so we mine all of it. Before we got into mining, we already had a mature data warehouse practice as a foundation. We started with data mining, which told us a lot, but not everything. So we added text mining to it, such that facts deduced from text feed into the data mining implementation we'd already deployed. The order of steps we took—from data warehousing to data mining to text mining—turned out to be a safe, phased approach."

USER STORY

Start with a data warehouse, then predictive analytics, and finally text analytics.

Processing Unstructured Data in a DBMS

Leading database management systems (DBMSs) have functions for processing unstructured data within a database, typically stored as text fields or binary large objects (BLOBs). Depending on the DBMS brand, these functions do keyword indexing for search, text analytics for entity extraction or clustering, and data mining for predictive analytics.

A prominent use case for this configuration involves e-mail. TDWI has encountered several user organizations that dump e-mail text into DBMSs periodically, then index it for search and entity analysis, looking for legal and compliance infractions like insider trading and data privacy violations. Since most DBMSs can create multiple indices of various types, the data can be accessed by various query, reporting, search, and predictive analytics tools.

Hence, a DBMS may qualify as a tool for text analytics or search. Oddly enough, this makes sense, because DBMS servers have diverse indexing capabilities and scalable performance. Assuming that these functions meet user requirements, they can be compelling to organizations that wish to leverage their DBMS investment or to manage unstructured content in a DBMS.

Text Analytics and the Future of BI

Text analytics has the potential to double the volume and breadth of data that is applied to BI. It's not so much the volume as the breadth that's significant. Organizations embracing text analytics all report having an epiphany moment when they suddenly knew more than before, which helped them to gain more precision in fact-based decision making. After all, without representation from unstructured data, a data warehouse is a single truth, but not the whole truth. In another direction, BI is heading strongly into performance management, which assumes accurate measurement; text analytics can quantify performance in areas that weren't quantifiable before, especially text-laden processes like call center, claim processing, patient chart, and so on. If it gains broad adoption, text analytics will transform data warehouses from a single truth to the whole truth and performance management from guesswork to precision measurement.

Vendor Products for BI Search and Text Analytics

BI search is the most dynamic area in BI functionality today.

Software vendor activity around BI search is fast and furious at the moment, driven by multiple trends and events:

- **BI vendors and users alike realize that BI search could be a game-changing event.** Infusing BI platforms with search capabilities—especially advanced ones like indexing data for an ad hoc query equivalent—stands a good chance of changing dramatically how end users find, consume, and create reports and analyses (as discussed earlier in this report).
- **BI search is a new feature for vendors to compete on.** In a domino effect, once a few vendors announced enhancements around search, others felt compelled to follow suit as a competitive measure. Numerous software vendors are currently rolling out new products for BI search or enhancing those that already exist.
- **Enterprise search implementations are finally reaching true enterprise-scope.** This is achieved by federating indexes of many corporate sources to make all content visible from a single search term box. BI content (mostly within BI platform repositories) is an important source for a federated enterprise search implementation (along with other application-specific sources for ERP, CRM, etc.). Hence, BI and search vendors are furiously partnering to ensure that BI content can be federated into larger enterprise search implementations.
- **Google's recent actions have upped the ante.** Google is aggressively expanding and marketing its enterprise offerings. It introduced Google OneBox in 2006, which expands its already established product Google Search Appliance by adding interfaces to enterprise databases, packaged applications, and—significant for this report—BI platforms.

Vendor offerings are a moving target at the moment, even more than usual.

With so much activity now—and more to come in the next 18 months—it's difficult to say which vendors have what and which claims will result in actual products in the future. This chapter tries to hit this moving target by describing software vendor products that support BI search and text analytics. Most of the space here is devoted to BI and search tools, since vendor products and partnerships for these are far more difficult to sort out than those for text analytics.

Vendors and products mentioned here are representative, and this chapter does not attempt to be a comprehensive description of the entire vendor community. (See Table 2 for a list.) Based on the products mentioned here, a technical user should be able to identify user requirements, understand what combination of tools and technologies are needed, and draft an evaluation list of vendor products that map credibly to user requirements.[3]

[3] The sections "Use Cases for BI Search" and "Use Cases for Text Analytics" earlier in this report describe most of the product configurations with which this section associates vendor products. So the reader may wish to review those sections before reading this chapter.

BI Vendors

When evaluating the BI search and text analytics capabilities of a BI platform, you need to look at capabilities built into the platform (both today and in the future) and those that come from third-party products. For example, most BI platforms already support rudimentary BI search in the sense of keyword indexing for reports in the platform's repository, while advanced search-based functions for ad hoc query and report creation are coming in the next year or so. Advanced configurations (like mixing BI content with other enterprise content) require a third-party enterprise search product.

Most BI vendors will not build a text analytics or enterprise search product, but will instead rely on partner vendors for this. Evaluate partnerships carefully: partnerships come and go quickly and few have meat behind them. Look for partnerships that have reference customers in common and involve product enhancements that ensure partners' products interoperate deeply and reliably. Furthermore, look for BI-to-search integrations that employ the security features of the BI platform to keep unauthorized users from accessing sensitive BI content. Note that most of the new product functions and partnerships mentioned in this chapter assume that an organization has upgraded to the most recent release of the BI platform and related partner products.

Many configurations force you to evaluate and acquire products from multiple vendors.

Business Objects

Business Objects has announced an aggressive road map for delivering product features in 2007 that relate to BI search and text analytics. In early 2007, the Q1 Productivity Pack for BusinessObjects XI introduces new tools for BI repository search and search as query. A tool for search-based data discovery is coming later in the year. To develop its partner ecosystem for text analytics, enterprise search, and related professional services, Business Objects has launched the Open Search Initiative, and the current list of partners includes: Accenture, Attensity, Autonomy, BearingPoint, Clarabridge, ClearForest, Endeca, FAST, Google, Inxight, IBM, and Oracle. Hence, Business Objects has a well-defined product road map and a large partner ecosystem for BI search and text analytics.

Business Objects and Cognos both have many useful partnerships with search and text analytics vendors.

Cognos

Developed in-house and introduced in early 2006, the new Cognos 8 Go! Search enables BI search for the single metadata store under Cognos 8 BI. Cognos 8 Go! Search supports all BI content, including reports, analysis, metrics, events, scorecards, dashboards, and packages. It records where keywords occur (column or row headings, prompts, data points, etc.) and factors this into the ranking algorithm, so that the end user can distinguish reports, report sections, and data references (like query parameters). Cognos Go! also provides BI content for enterprise search implementations when integrated with products by Autonomy, FAST, Google, and IBM. These are deep partnerships, in that Go! provides a module for easy integration with Google OneBox, supports the IBM UIMA framework for accessing any kind of data, and FAST can analyze data coming from Cognos 8. Cognos provides text analytics via integration with Autonomy IDOL Server, ClearForest Text Analytics Platform, and IBM WebSphere Information Integrator OmniFind. Hence, Cognos already has a search service native to its BI platform which it has integrated with leading enterprise search and text analytics products.

Hyperion

Like other leading BI platforms, Hyperion System 9 supports Google OneBox for Enterprise, a feature of the Google Search Appliance. As with other BI-to-Google integrations, corporate end users can search via Google to hit all enterprise sources, including BI repositories. The Google-powered Hyperion solution is accessible through the Hyperion System 9 Workspace, providing a single point of entry to the various Hyperion System 9 modules and report repositories, as well as data sources known to the BI platform, like relational databases, Essbase, SAP BW, and financial management applications. Besides finding reports, end users can get direct answers to search questions that require a numeric answer.

Information Builders (IBI)

IBI WebFOCUS Intelligent Search integrates with the Google Search Appliance and Google OneBox to gain the usual benefits of BI content via enterprise search. But IBI's approach "one ups" other BI-to-Google integrations by including the special application integration and data integration capabilities of connectors and other interfaces from iWay Software, an IBI subsidiary. iWay's connectors extend the reach of search to almost any structured, semi-structured, or unstructured data source imaginable. WebFOCUS Magnify is a brand new product that screens messages on a bus, enhances selected messages, then sends enhanced messages directly to the search indexing mechanism, which in turn makes information accessible via search within seconds of its origin.

Search Vendors

Before looking for a search engine to use for BI purposes, decide which configuration you want. In many configurations that integrate a BI platform with an enterprise search platform, BI content becomes one small brick in the larger brick wall of enterprise search content. This is usually the case with Google Search Appliance or IBM OmniFind. Just about any search engine (say, from market leaders Autonomy and Verity) will work in a low-end configuration where BI search merely indexes the reports of a single BI repository. Another route is to employ a search product that itself includes analytic and data integration capabilities that complement those of a BI platform or data warehouse, as with Endeca and FAST.

Google

Thanks to OneBox, Google suddenly has influence on BI search.

Google Search Appliance (GSA) is an enterprise search solution consisting of purpose-built hardware and software bundled into a server blade. Early in 2006, Google released Google OneBox, a new interoperability feature that extends GSA's reach to several BI platforms (Business Objects, Cognos, Hyperion, Information Builders, SAS) and operational applications (Employease, Netsuite, Oracle, Salesforce.com). Google's sudden and influential appearance in BI has led some industry observers to call this development the "Googlization of BI." But let's not forget that BI platforms can also integrate with other enterprise search tools, most commonly those from Autonomy, Endeca, FAST, and IBM. GSA is known for its low price ($30,000 entry level), speedy deployment (within hours), and easy expansion (just plug in more blades).

IBM

IBM's software products are known for their support of federation, so it's no surprise that federated search is part of IBM's enterprise search platform, WebSphere Information Integration OmniFind. Furthermore, OmniFind supports Unstructured Information Management Architecture (UIMA), an IBM-sponsored open standard for indexing enterprise data—whether structured, unstructured, or semi-structured—regardless of data's origins, storage, or model. Between federation and UIMA, OmniFind can index and federate just about anything, and it's known for high scalability. OmniFind supports UIMA-based text analytics, and it also integrates with pure-play text analytic products from Attensity and Clearforest. The list of vendors supporting OmniFind and UIMA includes Attensity, Clearforest, Cognos, Endeca, Inxight, SAS, and SPSS.

Federation is becoming a norm for enterprise search architecture.

Endeca

The Information Access Platform from Endeca is a search engine that can analyze its search results. It applies a kind of text analytics or text mining to identify entities and concepts in an index or result set, then cross references these to describe relationships among them. This imposes a structure on the output, so that the end user gains analytic insight, not just a list of documents, which in turn helps the end user locate information of the greatest relevance. Endeca is known for its stunningly pleasant user experience, multidimensional search parameters, and high relevance of search results. But achieving these benefits demands a fair amount of development time and cost.

Fast Search & Transfer (FAST)

Adaptive Information Warehouse (AIW), a family of search-based BI products from FAST, fits between the BI platform and its data sources, ranging from traditional data warehouses to bodies of unstructured documents. This way, BI search stays focused on BI content, and BI content is enriched by text-based information. AIW can crawl and index databases and data warehouses, and its text analytics can feed both databases and search indices. AIW even has a data cleansing tool. Hence, from the BI platform, searches, ad hoc queries, and reports have access to a broad, friendly, clean, and high-performance index that represents the full data continuum. As with Endeca, getting to this desirable goal takes development effort, but not as much as a traditional data warehouse or data integration project.

Database Vendors

All modern RDBMSs have some kind of search indexing built in, and sometimes related features for text mining, text analytics, data mining, text management, and XML management. Most of these features in most RDBMSs require that unstructured and semi-structured data being indexed or mined reside in the RDBMS alongside structured data. As exceptions, IBM DB2 8 and Oracle Data Server 10g are known for operating on data outside the RDBMS through federation. RDBMS-based search and text analytics functions are rarely best-of-breed, but the robust maturity of an RDBMS yields other benefits, like a familiar platform, a single store for all data, and scalable data storage and processing speed.

RDBMS-based search has advantages, like familiarity, scalability, and indexing options.

Sybase

Data Fusion is an add-on product for Sybase IQ that parses text and indexes it in various ways. The resulting index can look like a search index or a SQL-accessible relational or multidimensional database. This is possible due to Sybase IQ's inherent powers for indexing data of any type in multiple ways, so that the index is optimized for rapid retrieval, while minimizing data redundancy. The text can be in binary large objects (BLOBs) or relational text fields in the RDBMS, and Data Fusion has special functions for loading e-mails, file attachments, and other text data sources. In a nutshell, you can configure Sybase IQ Data Fusion to be a search index on steroids. One customer has indexed 3.4 trillion records, proving its scalability. It's a great choice for users wanting deep search and fast lookup capabilities built into the data warehouse platform.

EII Vendors

Most platforms for enterprise information integration (EII) can parse unstructured and semi-structured data, and represent entities and facts from these in an internal XML-based index, along side data culled from structured sources. The resulting index—sometimes called a virtual database—is a database view through which various tools (most commonly reporting tools) can access data in a federated manner in real time. Representative EII tools are available from Composite Software, IBM, Ipedo, and Skytide. EII platforms are a good choice when you need a real-time, SQL-accessible virtual database that represents sources from the entire data continuum.

Pure-Play Text Analytics Vendors

Text analytics tools can feed data into their own analytic applications or those based on other vendor products.

Attensity and ClearForest are the leading independent software vendors that focus exclusively on text analytics (as defined earlier in this report). Both provide development tools, servers, and analytic applications based around text analytics. Both can output extracted entities and facts in relational, XML, and other data models for analysis by tools from other vendors. But both also provide numerous applications that provide industry and department-specific analyses, and each provides professional services and training for all products.

Attensity

Attensity Text Analytics is a suite of tools and analytic applications that includes integration technology, knowledge engineering tools, a scalable server platform, patented extraction engines, and several analytic applications. Attensity can be installed at a customer site or used as a hosted solution. The technology extracts facts and then creates output in XML or a structured relational data format that is fused with existing structured data for analysis using Attensity's applications or other business intelligence applications. Attensity has partnerships with Business Objects, Cognos, and IBM, and Teradata resells Attensity in their 'root cause' analytic application.

ClearForest

The ClearForest Text Analytics solution is comprised of an advanced tagging and extraction engine, a development environment, and multiple analytic applications. The rules-based semantic tagger identifies and categorizes basic entities, facts, noun groups, and keyword relationships within one or more documents and marks them up for extraction. It then outputs the data in XML or database file format. Once structured, this information can drive ClearForest's stand-alone analytics applications or feed into data warehouses and combine with structured data to provide more comprehensive business intelligence using other tools. ClearForest partners with Business Objects, Cognos, Endeca, IBM, Information Builders, and others.

Vendor Products Related to BI Search and Text Analytics

Vendor Name	Vendor Product	Vendor Web Site
BI VENDORS		
Business Objects	Q1 Productivity Pack for BusinessObjects XI	www.businessobjects.com
Cognos	Cognos 8 Go! Search Service	www.cognos.com
Hyperion	Hyperion System 9 Workspace	www.hyperion.com
Information Builders	WebFOCUS Intelligent Search	www.informationbuilders.com
SEARCH VENDORS		
Google	Google Search Appliance, Google OneBox	www.google.com
IBM	WebSphere Information Integration OmniFind	www.ibm.com
Endeca	Information Access Platform	www.endeca.com
Fast Search & Transfer	Adaptive Information Warehouse	www.fastsearch.com
DATABASE VENDORS		
Sybase	Sybase IQ Data Fusion	www.sybase.com
EII Vendors	Miscellaneous	NA
PURE-PLAY TEXT ANALYTICS VENDORS		
Attensity	Attensity Text Analytics Suite	www.attensity.com
ClearForest	ClearForest Text Analytics Suite	www.clearforest.com

Table 2. The list is representative and not meant to be comprehensive.

Recommendations

- **Recognize that BI search and text analytics are additions to the BI technology stack.** At the risk of stating the obvious, you can't add them to a BI stack until you have one in place. For most organizations, BI search and text analytics are later-phase deliverables, although it's possible to include them in the first phase of a BI platform implementation. As an exception, text analytics may be deployed as a silo, the way that predictive analytics often is.
- **Represent more unstructured and semi-structured data in the data warehouse.** Three-quarters of data in the average data warehouse is structured, although only half of data across the rest of the enterprise is structured. Most organizations need to close this gap; otherwise, the data warehouse will remain a single version of the truth, but not the whole truth.

The mix of data feeding a data warehouse is set to change, so get ready.

- **Prepare for an unstructured data deluge.** Semi-structured and unstructured data sources for data warehouses will increase dramatically in the next three years (up 21–81% depending on source type).
- **Rethink data integration and warehouse data models.** Structured data sources are declining but not disappearing, so you must maintain what you've built for them. In addition, you'll need new interfaces to unstructured data, which text analytic tools can provide. Expect a lot of head-scratching over how to reconcile the output of text analytics with data structures in your data warehouse, plus a lot of discussion and confidence building before this happens.

BI search is a big bang for the buck, since it helps all BI users and IT.

- **BI search is a baby step, so just take it.** Simple configurations that index the reports of a single BI platform are relatively cheap and easy, especially when your BI platform has the capability built in. Given that all users benefit from BI search—even IT people whose load is lessened by its self service—it makes a big bang for the buck.
- **Start with one platform, but try to incorporate others into the scope of BI search.** This applies to organizations with multiple BI brands. Multi-platform BI search eases the pain of ping-ponging among platforms to gather related reports. And it can be a compelling stop-gap alternative to expensive and disruptive platform consolidations.

BI search assists with many BI tasks, but won't replace tools.

- **Don't expect BI search to replace anything.** BI search assumes a BI platform is already in place. Search is an alternate way of finding reports, but it can't replace the BI platform's taxonomy. It can find data, but it's not a full-featured query tool. It can find report sections and paste them into new reports, but it won't replace report authors or authoring tools.
- **Recognize that text analytics requires a fair amount of development time.** It's not as fast and easy to deploy as BI search. You need to decide what entities and facts to look for. Plus, you must work out data models for extracted entity records and the target database or index they'll go into. Plus, you'll probably need ETL jobs to transform extracted data into the data models that your data warehouse and other applications need.
- **Expect vendor products to evolve.** Software vendors have barely scratched the surface in terms of tying search and text analytics directly into business intelligence tools and data warehouses. The flurry of vendor partnerships and promised products seen in 2006 is a good start. But this is a moving target that will continue moving for years, as partnering vendors deepen interoperability and new products finally see the light of day. In the mean time, users must investigate vendor offerings for their current as well as future capabilities, plus the depth and ease of integration in multi-product configurations.

Business Objects
3030 Orchard Parkway
San Jose, CA 95134
408.953.6000
Fax: 408.953.6001
www.businessobjects.com

Business O bjects is the world's leading BI software company, helping organizations gain better insight into their business, improve decision making, and optimize enterprise performance. The company's business intelligence (BI) platform, BusinessObjects™ XI, offers the industry's most advanced and complete platform for reporting, query and analysis, performance management, and enterprise information management including data integration, data quality, and metadata management. BusinessObjects XI includes Crystal Reports®, the industry standard for enterprise reporting. Business Objects also has the industry's strongest and most diverse partner community, with more than 3,000 partners worldwide. In addition, the company offers consulting and education services to help customers effectively deploy their BI projects.

THE NEXT LEVEL OF PERFORMANCE™

Cognos
3755 Riverside Drive
PO Box 9707, Station T
Ottawa, ON K1G 4K9, Canada
613.738.1440
Fax: 613.738.0002
information@cognos.com
www.cognos.com

Cognos, the world leader in business intelligence and corporate performance management, delivers software and services that help companies drive, monitor, and understand corporate performance. Cognos delivers the next level of competitive advantage—corporate performance management (CPM)—achieved through the strategic application of BI on an enterprise scale. Our integrated CPM solution helps customers drive performance through planning; monitor performance through scorecarding; and understand performance through business intelligence. Cognos serves more than 23,000 customers in over 135 countries. Cognos enterprise business intelligence and performance management solutions and services are also available from more than 3,000 worldwide partners and resellers. For more information, visit the Cognos Web site at www.cognos.com.

Endeca
101 Main Street
Cambridge, MA 02142
617.674.6000
http://endeca.com/contact.html
www.endeca.com/

Endeca's information access technology combines zero-training search, discovery and analytics to inform daily decision making across the enterprise. It unites disparate data sources—regardless of structure, format or schema—to provide a 360-degree view of related customer, product and financial information. As a result, organizations can tackle problems that historically required separate database, search and OLAP technologies through one versatile platform. Hundreds of customers like Bank of America, Boeing, IBM, The Home Depot, John Deere, Nike and Wal-Mart rely on Endeca to support information-intensive processes, from compliance and human capital management to strategic sourcing and customer relationship management.

FAST
117 Kendrick Street
Needham, MA 02494
781.304.2400
Fax: 781.304.2410
info@fastsearch.com
www.fastsearch.com

FAST creates the real-time search and business intelligence solutions that are behind the scenes at the world's best known companies with the most demanding information challenges. FAST's flexible and scalable integrated technology platform and personalized portal connects users, regardless of medium, to the relevant information they need.

Hyperion Solutions Corporation
5450 Great America Parkway
Santa Clara, CA 95054
408.588.8000
www.hyperion.com

Hyperion Solutions Corporation is the global leader in business performance management software. More than 11,000 customers rely on Hyperion software to provide visibility into how their businesses are performing and to help them plan and model to improve that performance. Using Hyperion software, customers collect data, organize and analyze it, then communicate it across the enterprise. Hyperion offers the industry's only business performance management solution that integrates financial management applications with a business intelligence platform into a single system. Hyperion serves global customers in 45 countries and has a network of more than 600 partners who provide the company's innovative and specialized solutions and services. Hyperion is traded under the NASDAQ symbol HYSL. For more information, please visit www.hyperion.com.

Sybase, Inc.
One Sybase Drive
Dublin, CA 94568
800.8.SYBASE
infobi@sybase.com
www.sybase.com

Sybase is the largest global enterprise software company exclusively focused on managing and mobilizing information from the data center to the point of action. Our highly optimized analytics server, Sybase IQ, is designed specifically to deliver dramatically faster results for mission-critical business intelligence, analytics, and reporting solutions on standard hardware and software platforms. Sybase IQ delivers unsurpassed query performance and storage efficiency for structured and unstructured data. Sybase IQ combines extraordinary speed and agility with low total cost of ownership, enabling enterprises to perform analysis and reporting that was previously impossible, impractical, or cost prohibitive.

TDWI RESEARCH

TDWI Research provides research and advice for BI professionals worldwide. TDWI Research focuses exclusively on BI/DW issues and teams up with industry practitioners to deliver both broad and deep understanding of the business and technical issues surrounding the deployment of business intelligence and data warehousing solutions. TDWI Research offers reports, commentary, and inquiry services via a worldwide Membership program and provides custom research, benchmarking, and strategic planning services to user and vendor organizations.

tdwi
THE DATA WAREHOUSING INSTITUTE

1201 Monster Road SW
Suite 250
Renton, WA 98057

T 425.277.9126
F 425.687.2842
E info@tdwi.org

www.tdwi.org