



**IBM OmniFind Enterprise Edition:
A strategic platform for enterprise
search and text analytics**

Table of Contents

Executive Summary	2
Introduction: Maximizing the value of enterprise information	2
<i>Information as a Service</i>	2
<i>IBM Content Discovery: Search and Beyond</i>	3
<i>IBM OmniFind Enterprise Edition</i>	6
<i>IBM OmniFind Enterprise Starter Edition</i>	7
The leading enterprise search and text analytics platform	7
<i>A flexible architecture to deliver highly relevant results</i>	7
<i>Highly available and highly scalable</i>	15
<i>Open framework for extensibility and interoperability</i>	16
<i>Flexible administration</i>	17
<i>Robust set of security features</i>	18
The search engine of choice for Lotus Notes/Domino environments	20
Extend the reach and capability of IBM WebSphere Portal	22
Building a world-class corporate intranet: IBM On Demand Workplace	25
OmniFind Enterprise Edition as a text analytics platform	29
<i>Application examples</i>	31
Conclusion	34

Executive Summary

IBM OmniFind Enterprise Edition delivers breakthrough enterprise search capabilities to help get the right information to the right people at the right time. Built on a flexible, open architecture, it offers a unique and powerful combination of performance, security, scalability, enterprise reach, and openness for applying advanced linguistic processing. Leveraging native UIMA^[1] support, OmniFind Enterprise Edition also provides a rich platform for processing unstructured information to support enterprise search and business intelligence solutions. OmniFind Enterprise Edition is well-adapted to work in the heterogeneous and multi-vendor world of the modern enterprise. Lotus Notes/Domino and WebSphere Portal customers, in particular, will benefit from the secure best-in-class integrations OmniFind Enterprise Edition offers to them. An entry-level offering, OmniFind Enterprise Starter Edition, is also available providing the same functional capabilities as OmniFind Enterprise Edition in a scaled-down implementation. This white paper will explore the OmniFind Enterprise Edition architecture, the key capabilities it delivers, and how it can be used to power key business applications.

Introduction: Maximizing the value of enterprise information

Information as a Service

Companies have invested in numerous technologies to manage and control enterprise information, and the tactical and widespread nature of these investments has left silos and redundancies in its wake. As a result, key information assets are scattered across the enterprise, isolated and inaccessible to people, applications and processes. As information sources and the way these sources are consumed change, organizations need a flexible architecture that can decouple information from its source, and deliver it as a service to any person, business process or application, on demand.

^[1] UIMA (Unstructured Information Management Architecture) is an open source, standards based framework for processing unstructured information through text analytics.

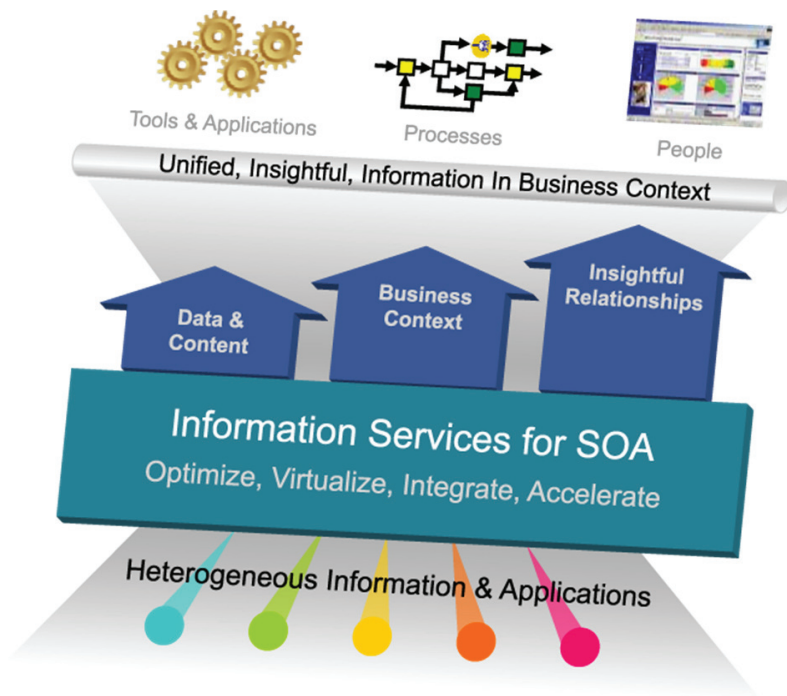


Figure 1: Information as a Service enables organizations to move from a project-based to a flexible, service-oriented architecture.

Information as a Service is about delivering a new level of services that help add value to information. IBM Information Management software provides the most comprehensive set of services that deliver value above and beyond traditional data management. This includes services that integrate information – both data and content, regardless of location – to provide a unified view; services that add business context to raw information, enabling transaction data to take the form of a policy, for example; and services that expose insightful relationships buried in the information through advanced analysis, to enable better decision making. To achieve this, Information as a Service must leverage standards such as XML, XQuery, JSR170 and Web Services.

IBM Content Discovery: Search and Beyond

IBM's Content Discovery platform and solutions enable Information as a Service by providing the delivery of the right information to the right people at the right time, on demand. Content Discovery puts information to work by providing insight into structured and unstructured information assets, and empowering the user to take action. The Content Discovery portfolio provides a full spectrum of capabilities, ranging from entry-level enterprise search to highly-scalable, available and secure search for intranets and portals, to sophisticated line-of-business solutions incorporating rich, easily-managed search experiences and real-time understanding of user intent.

The Content Discovery portfolio consists of IBM OmniFind offerings that are available in a tiered fashion, addressing the various search needs of users inside and outside the enterprise. They are specifically designed to maximize the value of your intranet, Portal, Lotus Domino, Self-Service, Commerce, Contact Center, eDiscovery and Business Intelligence applications.

- **IBM OmniFind Analytics Edition** enables unprecedented insight into unstructured content, utilizing sophisticated linguistic understanding and trend analysis.
- **IBM OmniFind Enterprise Edition** provides full-scale secure enterprise search and is best suited for secure intranet deployments. It also leverages UIMA, an open, extensible framework for processing unstructured information through text analytics for extracting insight from unstructured content to enable capabilities such as semantic queries, navigation of business intelligence reports and custom analytics applications.
- **IBM OmniFind Enterprise Starter Edition** provides secure enterprise search for smaller departmental-scale implementations. It brings all of the flexibility and rich functionality of OmniFind Enterprise Edition to smaller-scale intranet projects.
- **IBM OmniFind Discovery Edition** provides a managed search experience for customer- and partner-facing search solutions. This includes packaged solutions for Commerce, Self-Service and Contact Center search.
- **IBM Classification Module** extends the sophisticated contextual understanding capabilities of the core products by automatically cataloguing long-form text, such as online chat, case management notes or emails, into content “folders”, making it more efficiently and accurately accessible.
- **IBM OmniFind Yahoo! Edition** provides a simple, no-charge enterprise search solution for basic intranet and site search.

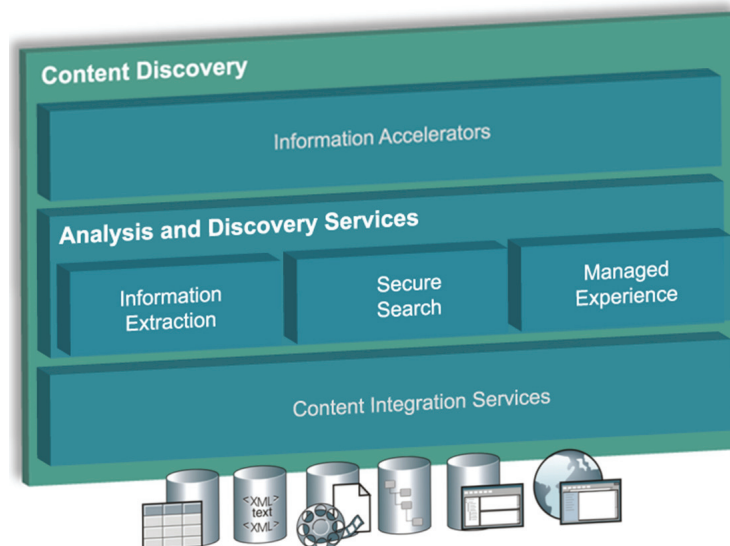


Figure 2: The IBM Content Discovery platform provides a broad range of information access capabilities for delivering meaningful, high-value search solutions that can address real business needs.

IBM customers, and IBM itself, have deployed a number of high-value solutions based on the IBM Content Discovery portfolio.

Secure Intranet Search

Chicago Bridge and Iron, a global engineering, procurement, and construction company, deployed a secure corporate intranet spanning 60 locations worldwide and 11,000 employees using OmniFind Enterprise Edition. The content was scattered across file systems, Web servers, Lotus Domino databases, and other repositories. By unifying all of this information into a single, searchable index, Chicago Bridge and Iron is maximizing the value of existing information assets to increase employee efficiency and reduce project delivery costs.

In 2003, *IBM* deployed a world-class corporate intranet (“w3”)^[2] using OmniFind Enterprise Edition. Today, more than 70% of IBM’s 300,000 employees rely on w3 as the primary and trusted source of company information. The application also features new search capabilities such as social bookmarking and expertise location to deliver even more relevant search results. The IBM intranet makes it easier for employees to find the information they need to perform more effectively and to make critical business decisions.

Commerce Search

Johnny’s Selected Seeds (Johnny’s), a mail-order seed producer and merchant, deployed a Web site powered by OmniFind Discovery Edition to convert shoppers into buyers and increase online sales. By providing advanced search and navigation capabilities, OmniFind Discovery Edition makes it much easier for customers to find the products that match their specific needs. As a direct result of this implementation, Johnny’s realized a 32% increase in online sales, a 17% increase in average order value, and exceeded previous year’s sales in just 6 months.

Campmor, an outdoor equipment specialty retailer, significantly increased revenue derived from online searches by deploying OmniFind Discovery Edition. After deployment, Campmor realized a 64% increase in online revenue; a 9% increase in orders initiated from customer Web site searches; a click-through rate increase from 16% to 25%; and their average search order size was 14% larger than non-search orders.

Self-Service and Contact Center Search

Micrel, Inc., a leading manufacturer of integrated circuit solutions for enterprise, consumer, industrial, mobile, telecommunications, automotive and computer markets, is leveraging OmniFind Discovery Edition to improve their site search and increase customer acquisition and retention. By delivering accurate and relevant information to customers quickly, Micrel increased portal usage by 300%, and realized a 25% increase in retention rate of new site visitors. Micrel is also gaining insight into customer needs to enhance product development.

^[2] To learn more about the IBM intranet, please refer to the section entitled, “IBM On Demand Workplace.”

Cloudmark provides award-winning software products and services that offer real-time protection for service providers, consumers and businesses from messaging threats such as spam, phishing and fraud. *Cloudmark* is using *OmniFind Discovery Edition* to improve agent productivity and reduce the cost of human intervention by deflecting 40% of customer e-mails and Web-based inquiries, and delivering highly accurate automated responses. After implementing the solution, *Cloudmark* realized a 100% payback in less than 3 months, and increased customer retention by 15%.

Business Analytics and Semantic Search

A leading healthcare provider in the southeast US is leveraging an IBM solution to gain greater insight into their network of service providers. The solution is based on *OmniFind Enterprise Edition* and *UIMA*, and provides a single view of revenue data, identifies and categorizes customer service issues, details service provider demographics and offerings, facilities upgrades, and provides access to all customer contracts. This insight allows the healthcare provider to deliver better customer service and aid sales when re-negotiating provider agreements.

A non-profit organization representing pharmaceutical associations from around the world is using *OmniFind Enterprise Edition* and the *UIMA* framework to deliver cross-language semantic search for a *WebSphere Portal*-based *Clinical Trials Portal*. This enables non-advanced end users, such as doctors and patients, to search for worldwide clinical trials (and their locations) by disease, symptom or drug name, using the terminology they are familiar with (scientific terms or common names), in their language of choice.

IBM OmniFind Enterprise Edition

OmniFind Enterprise Edition delivers high-quality, scalable and secure free-form text search that finds the most relevant corporate information for employees, suppliers, business partners, and clients. By simply entering a keyword or phrase, users can quickly search intranets, corporate public Web sites, relational database systems, file systems, and content repositories, and obtain meaningful results. Users also have the ability to go beyond standard full-text search to perform parametric and semantic queries that dramatically improve the relevancy of search results.

OmniFind Enterprise Edition is the first commercially available *UIMA*-based platform for processing text-based information. *UIMA* enables seamless integration of text analytics components that analyze documents, extract knowledge and identify higher-level concepts such as people, places, organizations, products, problems and other "entities" that are buried within unstructured data. This knowledge can be used to create an enhanced index for searching, or routed to a traditional data mart or data warehouse for use in business intelligence and analysis applications.

Additionally, *OmniFind Enterprise Edition* is an archetype of the service-oriented architecture (SOA) paradigm. By searching information scattered across multiple sources throughout the enterprise, enhancing its value by enriching metadata through text analytics, and making it available as a service to people, processes, and applications that can take advantage of it, *OmniFind Enterprise Edition* delivers a core function for any SOA environment.

IBM OmniFind Enterprise Starter Edition

IBM also offers OmniFind Enterprise Starter Edition, an entry-level enterprise search offering that brings all of the flexibility and rich functionality of OmniFind Enterprise Edition to smaller departmental-scale implementations. This cost-effective option lets companies start by solving a tactical search problem, and enables them to grow the solution over time as business requirements evolve. Customers who start with a small implementation can seamlessly move their application to OmniFind Enterprise Edition with the purchase of a trade-up license.

The leading enterprise search and text analytics platform

A flexible architecture to deliver highly relevant results

The three primary components of OmniFind Enterprise Edition are the **crawlers**, the **index server** and the **search runtime**. The crawlers extract content from different sources; the index server parses and analyzes the documents, and then builds a collection (index); the search runtime processes search requests, finds the most relevant documents in the index, and returns the results with subsecond response times.

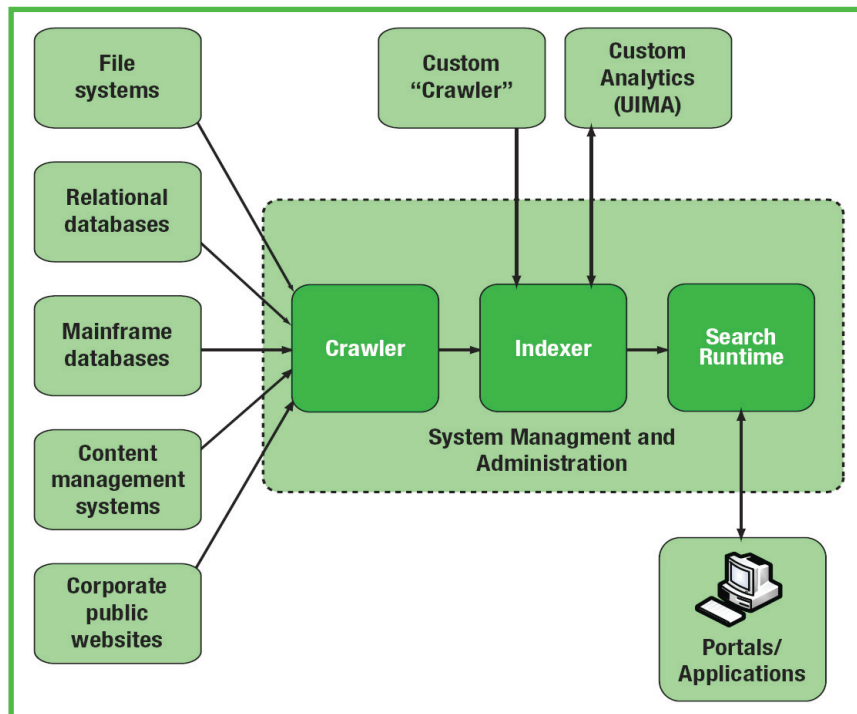


Figure 3: OmniFind Enterprise Edition provides an extensible architecture for the enterprise.

The crawlers access a wide variety of back-end data sources; alternatively, content can be pushed directly to the next step in the process, which includes breaking the document into tokens and appropriate linguistic analysis. Then, if so desired, content can be categorized with taxonomic rules. The indexing step brings in static ranking and duplicate deletion. Finally, at search time, queries can be corrected for spelling against a list of terms that appear in the collection compared against custom dictionaries for synonyms or domain-specific vocabularies, and results are ranked dynamically based on query terms.

During indexing, text is extracted and analyzed using highly sophisticated native linguistic processing. Further processing in the index server analyzes the link structure for intranet content, removes duplicate content and performs other processing on the collection of documents to improve overall search quality. Leveraging UIMA, additional third party and custom document analysis can be performed to generate even more metadata, which once added to the index, allows advanced capabilities, such as searching for concepts or facts buried within unstructured information. The option of deploying two search runtime servers provides redundancy to ensure that search is always available, as well as load balancing.

In addition to this core capability, the architecture is open and extensible to easily support a wide variety of industry or enterprise-specific search applications. For example, using the Search and Indexing Admin API, it is possible to create crawlers to customized and non-standard applications, so that their content can be reflected in the index. This allows IT departments to add their homegrown applications to the searchable content. This API can also be used to “push” content to the search index at intervals that are determined by business processes, such as publication of time-sensitive materials like press releases or new product information on external websites.

The application interface to the search engine is a well-defined Java API for ease of deployment into existing enterprise applications. In addition, there is a Web services interface for easing deployment in non-Java environments. By leveraging UIMA, OmniFind Enterprise Edition also provides the foundation to plug in custom analytics to enable domain-specific search capabilities, and more.

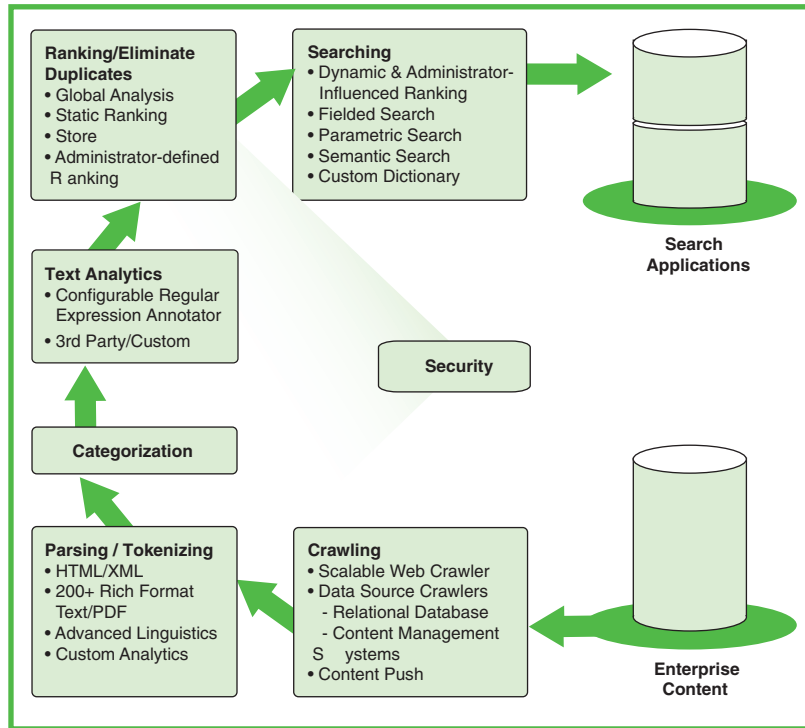


Figure 4: The flow of content through the OmniFind Enterprise Edition system.

State-of-the-art, language-specific relevancy

OmniFind Enterprise Edition uses sophisticated ranking algorithms developed by IBM Research and advanced linguistic analysis based on IBM LanguageWare® to deliver highly relevant search results. Not only does it evaluate and rank documents during preprocessing (before the index is built), it also includes a patented feature to rank results based on the type of query the search user enters. In other words, when search users enter their search terms, the system can dynamically analyze those queries and adjust ranking factors accordingly. Ranking factors are different, for example, for the query “ibm” versus the query “how do I change my password.” This technology has significantly improved overall search quality in blind surveys conducted using the IBM intranet.

Custom dictionaries for synonym support

The use of custom synonym dictionaries, or “thesaurus”, improves search results and assists the user by expanding query terms with a list of synonyms. Since the quality of the result depends on the quality of the query, it is vital to keep the quality level of each

query high. Synonyms provide a way to enrich the query with related keywords that may not have occurred to the user instantly but were intended and perhaps implicit. Synonyms are particularly useful for product names and for domain or even company-specific vocabulary. Synonyms are also extremely powerful to search acronyms and their expanded forms in one query.

For example, in a call center environment, these expanded searches can be very valuable. A Customer Service Representative can search for the term “car”, but this time, with the help of synonyms, the result set will cover a broader range of documents. OmniFind Enterprise Edition can automatically expand the search with a list of synonyms that have been defined for cars. Searching for “cars” now also means searching for “vans”, “pickups”, “vehicles”, “convertibles”, and so on. The user types in one of the synonym terms but probably had in mind the whole group of words that describe a car. With the help of synonyms, OmniFind Enterprise Edition assists the user by understanding and anticipating user intent.

Ability to influence ranking

Ranking algorithms can be modified by administrators to provide additional visibility for important content:

- **Create custom stop-word dictionaries.** Stop words are “noise” words that are not useful for query terms (such as ‘a’ or ‘the’ in English). This feature enables you to extend the linguistic stop words provided with OmniFind Enterprise Edition with stop words specific to your collections.
- **Create custom boosted terms dictionaries** to increase or decrease the weights of the specified boosted terms. OmniFind Enterprise Edition by default assigns weights to query terms, based on the relative frequency of those terms in the collection. For example, words like ‘the’, ‘of’ are very popular in all collections, but their importance is very low, so they will have very little effect on a document’s score. This feature enables you to influence the weights of terms that might otherwise get assigned a low or high score that is not appropriate for your particular collection.
- **Assign arbitrary fields relative importance scores.** This enables you to boost query terms according to the fields they occur in, such as a keyword or description field.
- **Boost the scores of documents that match a specified URI pattern.** URI patterns represent document type and location specific information in a hierarchical form – the deeper the paths, the more specific or localized the information to a specific subject. This feature enables you to boost specific pages, specific areas of interest, or specific types of documents, such as database tables or Lotus Notes.

Multi-lingual support

In today’s global enterprise, a high-quality search solution must seamlessly support multiple languages. This can only be achieved through advanced search engine design and subsequent integration of linguistic processing technology. Such linguistic technology is not only required as the mechanism with which to provide high-quality multi-lingual search, but also as the basic building block on which more advanced search capabilities can ultimately be built.

Let us consider just a few of the challenges that result from some of the linguistic subtleties of the various languages.

- **Word normalization** (often called lemmatization) is important for any language which inflects, but it is absolutely critical to achieve reasonable recall for highly inflected languages, such as Russian. In general, lemmatization cannot be reliably substituted by stemming, which mechanically strips endings to create a stem that approximates a lemma. These stems are frequently non-words which have no linguistic or semantic meaning, or worse, provide completely different meaning and thereby negatively impact search quality. See Table 1 for an example of how lemmatization provides superior results by understanding the close relationship between “organizing” and “organize,” and not including “organ,” a more distant term. Simple string manipulation would not understand that relationship.
- **Text segmentation** is required to deliver high precision for non-white-spaced languages, such as Chinese and Japanese.
- **Word segmentation** is needed to break words into their semantic units and is important for compounding languages, such as German.
- **Variant spellings** are common in many languages and also need to be taken into consideration. For example, in German there are differences due to traditional and reformed spelling rules; in English, there is U.S. versus U.K. variants; in Chinese, there are Traditional and Simplified versions; and in Japanese, you frequently see phonetic spelling variants.
- **Prefixes and postfixes** can mask the semantic component of words for languages, such as Arabic.
- **Character normalization**, which identifies character combinations that may be considered interchangeable for search, is another important language characteristic, and includes such things as Unicode normalization, transformations for non-ascii characters (such as ù, ú, û, ü), mappings between Traditional and Simplified Chinese characters, and more.

OmniFind Enterprise Edition delivers this inherently multi-lingual search solution through its integration of IBM's corporate linguistic technology, IBM LanguageWare.

IBM LanguageWare, the result of years of research and development involving language specialists across four continents, is a highly sophisticated linguistic technology. Unlike many other linguistic technologies that prescribe exactly how they believe language should be analyzed, and subsequently indexed, LanguageWare is open, extensible and customizable and allows this definition to be modified according to the needs of a particular enterprise. The result is a multi-lingual search application that understands the subtleties of the languages, as well as the vocabularies used in a corporation, and indexes them appropriately.

To learn more about globalizing your On Demand Business, visit: <http://www-306.ibm.com/software/globalization/index.jsp>

Surface form	Simple stemming	IBM linguistics
organ	organ	organ
organs	organ	organ
organizing	organ	organize
organized	organ	organize
organizations	organ	organization
organic	organ	organic

Table 1. IBM linguistic analysis engine provides more meaningful results using lemmatization.

Finding key sources quickly

To make it easier for users to see more results from different sites, OmniFind Enterprise Edition supports the ability to suppress additional results from the same site in the top results, enabling a wider variety of results to appear in the user’s browser. For example, results can be configured so that no more than two highly relevant search results are returned from each site.

Every organization has a set of key sites that users need to access frequently and fast, such as medical forms, HR policies, travel restrictions and business partner guidelines. No matter how inappropriately titled or how poor the meta data might be, administrators have the ability to assign predefined results for common queries, that are presented along with the most relevant search results. This Quick Links feature can produce substantial time savings for employee self-service inside the organization and improve customer satisfaction on corporate public Web sites. For example, queries such as “401K,” “pension” and “retirement plan” can be defined by Quick Links to take employees to the HR department home page.

Relevant, dynamic summaries

Document titles can be misleading or not provide enough information. For this reason, users need the ability to “peek” at the contents of each item in the returned results list without going through the time-consuming process of clicking on each result. OmniFind Enterprise Edition uses the words in the search request to dynamically generate document summaries based on the phrases that best represent the concepts of the document for which the user is searching. Dynamic summaries are further enhanced by search-term highlighting. The highlighted summaries can then be displayed in the result list so that users can easily determine the content and likely relevance of the document. Unlike most search solutions, which deliver static summaries extracting a specified, pre-stored summary from the document, OmniFind Enterprise Edition creates intelligent and dynamic summaries from document content.

Top Results Analysis

OmniFind Enterprise Edition includes a Top Results Analysis feature for analyzing the top 500 results based on metadata attributes. In addition to generating a standard list of search results, a navigation pane with dynamic bar charts and drop-down menus is displayed for refining the query by category, language, source, or other metadata. Clicking on a bar chart item issues a new search with the selected item added as a new fielded search term. This feature helps users find the right information faster.

Top Results Analysis also includes a custom HTML chart, which can be extended to search any repository or index and create an ordered list of results. A sample has been written in which the custom HTML chart returns a list of results from Dogear (IBM's enterprise social tagging tool, part of the Lotus Connections product). The number and type of charts is easily configurable using a graphical layout tool.

Query correction based on your data

The use of IBM LanguageWare in OmniFind Enterprise Edition extends beyond indexing and query processing. Imagine a query interface that was smart enough to recognize when a query was suboptimal and was able to suggest intelligent alternatives. IBM is working to make this feature a reality by integrating the linguistic engine into the query interface. This integration will help ease the frustration that occurs when users mistakenly spell a word wrong and get no intelligent results back.

OmniFind Enterprise Edition can generate alternative queries when it recognizes misspellings within a query. OmniFind Enterprise Edition also uses the LanguageWare dictionary customization features to ensure that suggested corrections are tailored to the frequency of terms that appear in your collection and will make suggestions that make sense for your user community.

Flexible and extensible categorization support

A taxonomy is a way to categorize content to make it easier for search users to find relevant information. OmniFind Enterprise Edition includes a rules-based classifier, which provides a simple, flexible way to set up rules used to create the categories in a taxonomy. A sample rule is "Find any document that contains the term 'Automobile' but not the word 'Dealers' and place it in a category called 'Automotive Industry'." Rule-based classifiers are quick, easy to set up, and work best with relatively small numbers of categories.

Also, OmniFind Enterprise Edition can be integrated with IBM Classification Module for OmniFind Discovery Edition to provide automated document classification and taxonomy management capabilities. Taxonomies can be imported, edited and exported, and documents can be automatically tagged based on the taxonomy. Through training and continual learning, IBM Classification Module for OmniFind Discovery Edition provides a flexible way to group and tag enterprise content. For more information on this topic visit: <http://www-128.ibm.com/developerworks/db2/library/techarticle/dm-0702dorre/>

OmniFind Enterprise Edition also integrates well with enterprise taxonomy management tools such as SchemaLogic's SchemaServer. For a discussion of this topic, see this [article on "Leveraging taxonomies for enterprise search"](#) on IBM developerWorks.

Query support for new users and power users

Everyone is familiar with the ease of finding information with free-form text search. Free-form text search offers an easy and intuitive way of searching. In free-form searches, users simply enter a word, phrase or sentence and the search engine finds documents that are relevant to the query. This concept has become widely used in Internet search engines.

However, power users and the development of specialized applications will need a richer and more effective query language. Advanced query operators let you identify specific document attributes in your query, such as the document language, type or source.

In addition, OmniFind Enterprise Edition understands the structure of documents and can take advantage of this structure to specifically constrain the search to a section of the document or for specific metadata associated with the document, or limit the scope to a particular site. The search query must include the field name. Otherwise, the capability can be built into the search application. Some examples:

- For a Lotus Notes database, search in specific fields.
- For an XML document, search inside a specific tag (author, title, subject). For an HTML document, search on specific tags (META tags, title, URL)—for example, find all documents with "OmniFind" in the title on the [ibm.com](#) site.
- For newsgroups, search in a particular group.
- For MS Office or IBM SmartSuite® documents, search on document properties, such as author, title or document description.

In addition, OmniFind Enterprise Edition can perform comparative or evaluative queries on numeric and date fields and metadata (parametric search). For example, with the UNIX® file system, you can search for documents that are of a certain size or written after a certain date. For a database source, you can specify a search based on fields in the database, such as documents in which the price for an item is greater than a particular value.

Semantic search

Although keyword search can be a powerful tool in discovering documents based on the metadata and keywords in the document itself, it cannot by itself enable searches for information about higher-level concepts such as people's names, telephone numbers, parts or conditions. Semantic search applications extend user queries beyond simple keywords to find higher-level concepts, such as people, places, organizations, products, problems and other "entities" that may appear in the text. Moreover, these applications enable you to specify relationships between these concepts, such as problems "caused by" a certain machine part. They do this by the ability to search on the enhanced metadata that text analysis provides to a corpus

Supported data sources

- DB2 Content Manager Multiplatform V8.2, V8.3
- DB2 Content Manager for z/OS V8.3 FP5
- DB2 Document Manager V8.3 (via Content Manager connector)
- DB2 Universal Database for Linux,UNIX, Windows® V8.1, V8.2
- DB2 Universal Database for z/OS V7, V8
- DB2 Universal Database for IBM System i V5.3, V5.4
- EMC/Documentum 4.2x, 5.2.5, and 5.3
- IBM FileNet Content Services 5.3, 5.4
- IBM FileNet Content Manager 3.0, 3.5
- Hummingbird™ DM 5.1.0.5 with SR4
- Informix Dynamic Server Version 9
- Lotus Domino Document Manager V6.5.1, V7.0
- Lotus Notes/Domino Server V5.0.9a, or later, V6.0.2CF2, V7.0, V8.0 (V6.0.2 or later required to configure native security)
- Lotus Domino QuickPlace V6.5.1, V7.0
- Lotus Quickr Services for Domino V8.0
- Lotus Quickr Services for WebSphere Portal V8.0
- Microsoft Exchange Server 2000, 2003
- Microsoft SQL Server 2000
- Newsgroup (NNTP)
- OpenText Livelink Enterprise Server 9.2, 9.5
- Oracle® 9i, 10g
- Sybase® 11.9.2, 12.0, 12.5
- UNIX file systems
- Windows® file systems
- Web (HTTP or HTTPS)
- WebSphere Portal V5.1.0.1, V6.0 Web pages
- WebSphere Portal V5.1.0.1, V6.0 Document Manager
- Windows SharePoint Services, Service Pack 2, for Windows 2003
- Workplace Web Content Management V2.5, V5.1, V6.0 (with WebSphere Portal)
- JDBC databases accessible through the DBC crawler (IBM DB2 Universal Database V8.2, Oracle 9i, Oracle 10g, Microsoft SQL Server 2000, and Microsoft SQL Server 2005 databases through Type 4 JDBC drivers)
- Mainframe sources accessible through separate purchase of IBM WebSphere Information Integrator Classic Federation for z/OS (VSAM, IMS, CA-Datcom, Software AG Adabas)

For a complete and current list of data sources and required client software, visit <http://www.ibm.com/software/data/enterprise-search/omnifind-enterprise/>
Additional connectors can be built to other data sources.

Optimal performance with some content management sources may require some additional configuration services.

of information. Semantic search is the ability to understand and navigate through that enhanced metadata, providing users with more relevant results and with the ability, for example, to find entities regardless of how they are described in the text (for example, searching for documents that mention a person, but the person's name is unknown, or that contains a date, but the date is unknown).

OmniFind Enterprise Edition makes semantic search easier to realize by using synonym expansion as a way to bring conceptual understanding to keyword search rather than having to rely on special XML query syntax. OmniFind Enterprise Edition ships with a configurable annotator that can detect concepts in unstructured text based on regular expressions. A sample configuration to detect phone numbers, URLs and e-mail addresses is included together with a matching synonym dictionary. The sample can be tailored to meet a specific customer's needs, e.g., to detect part numbers or product names, in a few simple steps.

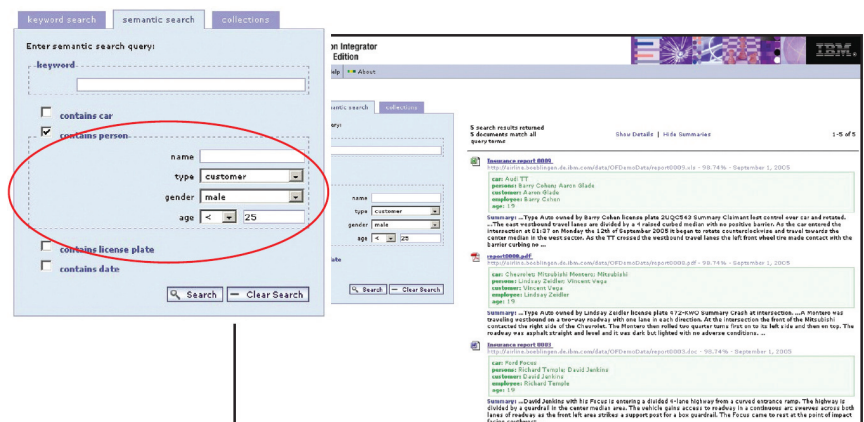


Figure 5: OmniFind Enterprise Edition and UIMA enable users to go beyond keyword search to find concepts and facts buried in text.

Highly available and highly scalable

In addition to providing high-quality results, OmniFind Enterprise Edition is designed to provide superior performance and scalability. A single index server can currently scale to more than 20 million documents, and federation of multiple index servers is supported to handle even larger environments.

For high availability and additional query throughput, OmniFind Enterprise Edition provides both a two-server and four-server configuration. The two-server configuration allows an administrator to install the crawler, the indexer and the search runtime components on one server, and an additional search runtime on a second server. By making the search component available on two different servers, this option provides high search availability.

When additional capacity is required for very large document collections, the four-server configuration provides both search runtime redundancy and the ability to split crawling from parsing and indexing.

OmniFind Enterprise Edition has proven its robustness and scalability on one of the most challenging intranets in the world: the IBM intranet, w3.ibm.com , serving a user community of more than 300,000 people.

Open framework for extensibility and interoperability

The OmniFind Enterprise Edition architecture is an open and extensible framework for interoperability of third party and custom developed enhancements. It also provides a rich set of tools and interfaces to simplify and accelerate integration into a wide variety of industry or enterprise-specific search applications:

- **A sample search application** built upon standard technology and easily configurable via a graphical user interface to match any look and feel. Use it for quick out-of-the box testing or as a foundation upon which to build a customized application. Includes simple and advanced search options, the ability to browse a category tree, top results analysis using dynamic bar charts and graphs, content fetching, search within results, and more;

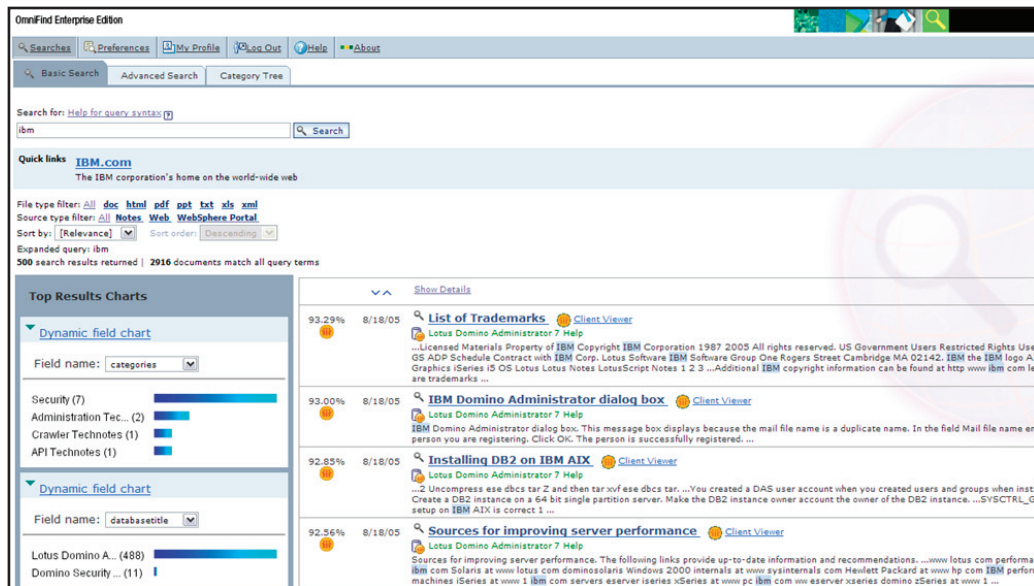


Figure 6: The OmniFind Enterprise Edition Search Application is a rich, out-of-the-box application that can be easily configured to match any look and feel.

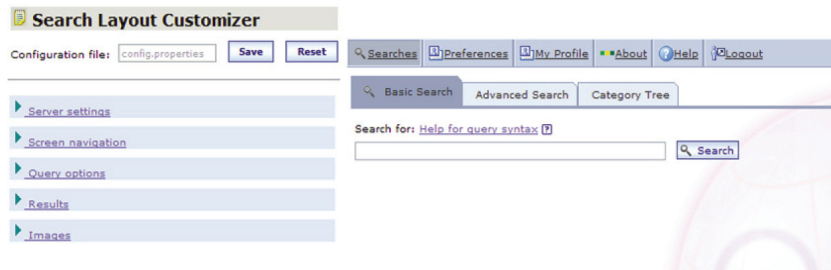


Figure 7: The OmniFind Enterprise Edition Search Application is easily customizable by making selections from within a graphical user interface.

- **Automated deployment script and sample portlet** for easy deployment into WebSphere Portal applications;
- **Integration with WebSphere Portal Universal Search Center** (including WebSphere Portal 6), enabling Search Center users to include OmniFind Enterprise Edition collections, and administrators to add or remove such collections from the Search Center;
- **Integration with leading desktop search vendors**, including Yahoo!/X1 Desktop Search, enabling enterprise search from within familiar desktop search environments;
- **A rich text analytics platform** based on the Unstructured Information Management Architecture, enabling developers to plug in advanced, domain-specific text analysis engines, and create new interoperable text analytics modules; and,
- **A well-defined Java™ API** for ease of deployment into existing enterprise applications.

The Java API is built around the IBM strategic interface for indexed search solutions (the Search and Indexing API), enabling developers to build a search application that can work with any incumbent search product, such as WebSphere Portal and OmniFind Enterprise Edition. This API also includes administrative capabilities for administering collections and for adding documents to collections. Additionally, this API provides a way to enable applications with time-sensitive information to “push” the latest updates to the data into the search engine instead of waiting for the system to recrawl the data source, thus making sure that the latest information is always available for searches. It also allows for the creation of custom crawlers.

Flexible administration

The OmniFind Enterprise Edition administration console provides a one-stop destination for installation, configuration, monitoring and other administration tasks. The easy-to-use Web-based interface provides the administrator with all the tools needed to keep the search environment up-and-running 24x7 and to help the administrator keep the system in top-running shape.

Analysis features are pre-integrated and under-the-covers to simplify the task of getting high-quality search results without extensive tuning.

- You only need to specify what to search and when to index to get started.
- Searched content can be expanded dynamically without rebuilding the search index.
- Administrators can monitor search activity with status summaries, which include performance, coverage, errors, query rate, response time, top queries and recent queries.

The system can send alerts to any e-mail address to help the busy administrator stay on top of system performance. For example, the system can determine if the remaining disk space for one of the search engines is less than 100 megabytes, or the response time of the search engines is more than the specified value, or when the crawl rate of the crawler is less than the specified value.

Within a typical enterprise, it is likely that multiple collections will be created for different applications. A separate collection might be created for the finance group (with its own set of users searching that collection), and another for the human resources group. Collections may contain documents from multiple data sources. Each application and its associated collection may require the expertise of different individuals in the organization to aid in the setup and configuration of the collection.

OmniFind Enterprise Edition supports this differentiation by allowing the administrator to assign roles to one or more specific collections. Therefore, individual collections can have their own administrators and operators, ensuring that the people who know and understand the content have the rights to administer it.

Robust set of security features

Enterprise search removes barriers to finding information, but it also places a burden on IT departments to ensure that sensitive content is accessible only to authorized users. OmniFind Enterprise Edition addresses this need for security through authentication and authorization at two separate levels of access.

Authentication is the process by which the system verifies that users are who they say they are. OmniFind Enterprise Edition has been engineered to work with the existing authentication components and does not require a separate login process for the end users. When OmniFind Enterprise Edition requires the identity of the logged-in user, it interacts with the host environment (e.g., WebSphere Portal) or the application to obtain the user's credentials. This approach allows OmniFind Enterprise Edition to be seamlessly integrated with the enterprise's existing authentication policies without requiring a separately maintained user registry.

The ability to verify authorization with the owning repository at query time is supported for the following data sources:

- Lotus Notes/Domino
- WebSphere Portal Document Manager
- WebSphere Portal pages
- Domino Document Manager
- Lotus QuickPlace®
- Lotus Quickr Services for Domino
- Lotus Quickr Services for WebSphere Portal
- DB2 Content Manager Multiplatform
- IBM FileNet Content Services
- IBM FileNet Content Manager
- Hummingbird DM
- Windows File System
- Windows SharePoint Services, Service Pack 2, for Windows 2003
- Documentum
- Open Text Livelink
- WebSphere Portal Web Content Management V6.0

Authorization is the mechanism by which a system grants or revokes the right to access some data, or perform some action. OmniFind Enterprise Edition employs several levels of access control which can be used independently or together to provide increasing levels of authorization. Access to collections can be restricted to only the search applications that have been granted access by the administrator.

For example, many instances of the OmniFind Enterprise Edition search portlet can be created in the WebSphere Portal server and placed on different portal pages. Each instance of the search portlet can be configured in such a way that only certain collections are available for searching through that particular instance of the search portlet. Portal Access Control can then be used to control who has access to which search portlet. Consequently, one instance of the search portlet could be set up to serve the finance organization and configured appropriately for that organization in terms of which collections are available for search.

OmniFind Enterprise Edition also controls access to documents to ensure that end users can only access certain specified documents. Document-level access control is accomplished by allowing the administrator to associate one or more security tokens with each document at crawl time, and optionally verifying authorization with the original data source at query time.

The following methods can be used for specifying security tokens:

- The value can be obtained from the native ACL of backend repositories for which this feature is supported (see Sidebar).
- The value can be specified by the administrator using the Administration Console.
- The value can be extracted from an administrator-designated field in the crawled document.
- The value can be determined by a user-defined Java routine, via the OmniFind Enterprise Edition “security token plug-in” API.

Customer Success – Chicago Bridge and Iron

Chicago Bridge and Iron, a global engineering, procurement, and construction company, deployed a secure corporate intranet spanning 60 locations worldwide and 11,000 employees using OmniFind Enterprise Edition. The content was scattered across file systems, Web servers, Lotus Domino databases, and other repositories. By unifying all of this information into a single, searchable index, Chicago Bridge and Iron is maximizing the value of existing information assets to increase employee efficiency and reduce project delivery costs.

At search time, the search application ensures that the current search user is properly authenticated and retrieves the correct security token for that user. When users submit a query, their security token must match the security token for the stored document to ensure that they do not see any documents for which they do not have access. If native ACLs are used as the security token, a runtime check against the native backend will pick up any changes in authorization since crawl time.

In addition, OmniFind Enterprise Edition supports single sign-on authentication through WebSphere Application Server LTPA for the following data source types:

- Domino Document Manager, Lotus Notes, and Lotus QuickPlace databases. The crawlers must use Domino Internet Inter-ORB Protocol (DIIOP) to connect to the Domino servers.
- Portal Document Management repositories accessible by the WebSphere Information Integrator Content Edition connector.
- When users use the OmniFind Enterprise Edition search portlet to search collections from within WebSphere Portal, single sign-on security is also provided for documents crawled by the Web Content Management and WebSphere Portal crawlers.

The search engine of choice for Lotus Notes/Domino environments

OmniFind Enterprise Edition is designed to be the search engine of choice for Lotus Domino customers. It scales to millions of documents, honors the most current Domino security settings, supports a broad range of non-Domino repositories, delivers highly relevant search results, and provides advanced search features such as synonym expansion, spell correction, dynamic summaries, and site collapse for a richer end-user experience. OmniFind Enterprise Edition provides secure, best-in-class support for Lotus Domino, including support on Linux, UNIX, Windows, and System i platforms.

OmniFind Enterprise Edition helps Lotus Domino customers to:

- **Maximize the value of Lotus Domino applications** by delivering a richer search experience and extending reach to non-Domino repositories;
- **Optimize Domino performance** by moving search processing cycles and disk consumption from busy Domino servers to OmniFind Enterprise Edition;

- **Minimize risk and total cost of ownership** by implementing a search engine that is designed to integrate with Lotus Domino;
- **Maximize knowledge worker productivity** by making it easier and quicker to find content that is scattered across Lotus Domino databases and other sources.

OmniFind Enterprise Edition ships with a dedicated Lotus Notes/Domino crawler, which can be configured directly from within the OmniFind Enterprise Edition Web Administration client. It also ships with a sample search application, which can open Lotus Notes databases and documents directly with the Lotus Notes client, or in a Web browser. Additionally, the OmniFind Enterprise Edition crawlers can be configured to search targeted Lotus Notes views set up by Lotus Notes administrators.

OmniFind Enterprise Edition uses native Domino interfaces (either NotesRPC or Domino IIOPI); understands the structure of Lotus Notes databases from fields to attachments; and comes with pre-built crawlers for crawling Lotus Notes/Domino, Lotus QuickPlace, Lotus Quickr Services for Domino, and Lotus Domino Document Manager using native security.

OmniFind Enterprise Edition offers a unique approach to document-level security. In addition to indexing security data about what users and groups have access to a given document, it provides the option of verifying access privileges with the owning Lotus Notes/Domino server prior to returning a result set. This means it always honors the most current Domino security settings, regardless of how fast access privileges change. OmniFind Enterprise Edition also supports single sign-on authentication through WebSphere Application Server LTPA for Lotus Notes/Domino, Lotus QuickPlace, and Domino Document Manager.

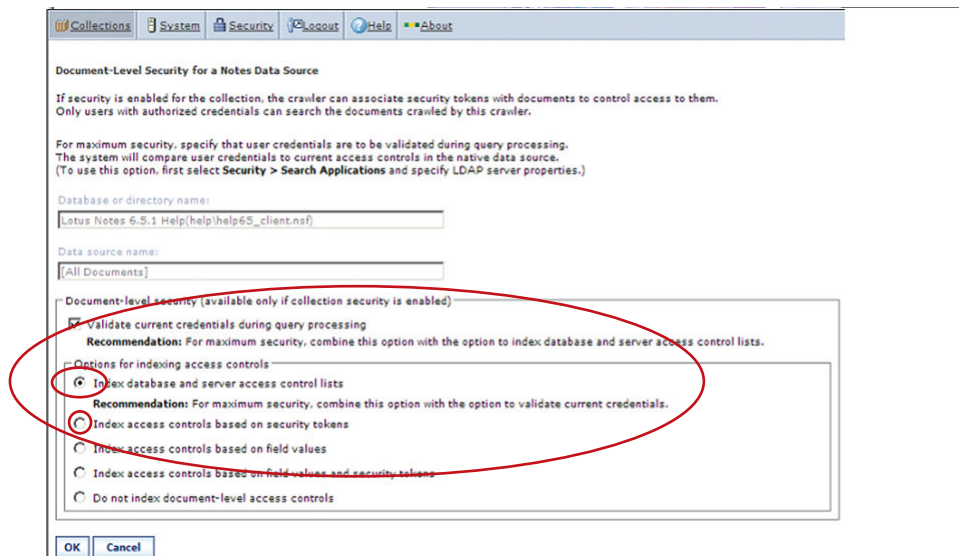


Figure 8: If the option to validate security credentials is selected, OmniFind Enterprise Edition will consult the Lotus Notes/Domino Access Control Lists (ACLs) during query processing to make sure the user still has access to the specified documents.

OmniFind Enterprise Edition can be integrated into a Lotus Notes application on a user's desktop using the SI-API and LotusScript® programming. LotusScript can be used to invoke SI-API, which is the Java interface into the backend OmniFind Enterprise Edition server. This means that all of the rich search functionality offered by OmniFind Enterprise Edition can be made available to users directly from within a Lotus Notes workspace. Additionally, a plug-in is provided for integrating OmniFind Enterprise Edition V8.5 into the Lotus Notes V8.0 client, allowing users to issue enterprise searches directly from within the Lotus Notes search bar.

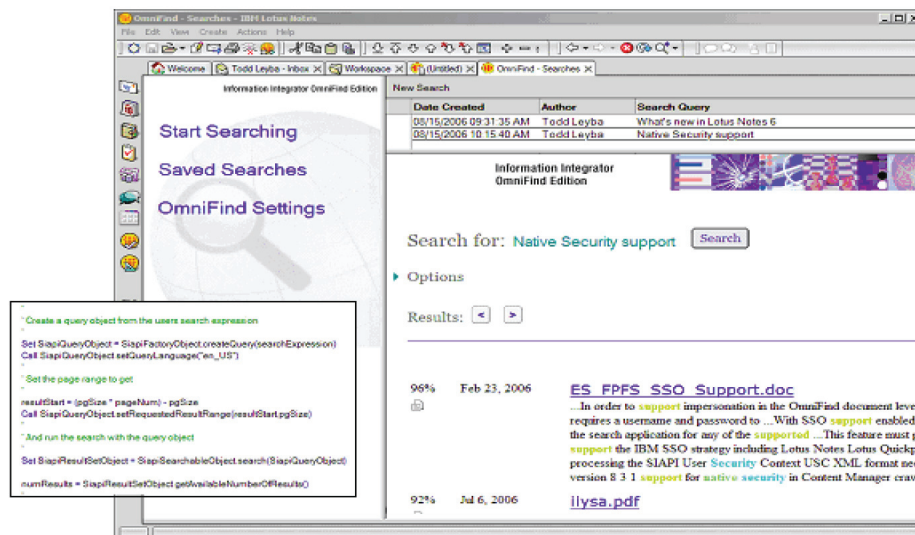


Figure 9: A sample Lotuscript Domino application shows how OmniFind Enterprise Edition can be integrated into a Lotus Notes workspace via SI-API.

Extend the reach and capability of IBM WebSphere Portal

OmniFind Enterprise Edition is designed to integrate with WebSphere Portal to provide scalable and secure enterprise search directly from within the familiar WebSphere Portal interface. This integration provides several advantages over using the embedded Portal Search Engine, including the ability to scale to millions of documents, secure reach to non-Web enterprise content sources, and all of the rich search functionality available in OmniFind Enterprise Edition including synonym expansion, Quick Links, dynamic summaries, and more.

OmniFind Enterprise Edition helps WebSphere Portal customers to:

- Increase adoption and ROI for portal applications by delivering a richer search experience and exposing all relevant enterprise content through a single portal search interface;

- Minimize risk and total cost of ownership by implementing a search engine that is designed to integrate with WebSphere Portal;
- Maximize knowledge worker productivity by making it easier and quicker to find relevant information through a single point of access.

Search OmniFind Enterprise Edition Collections from the Portal Search Center

Portal Search Center is a unified portlet for performing searches across a broad set of information sources. OmniFind Enterprise Edition supports the new WebSphere Portal 6 Search Center, making OmniFind Enterprise Edition collections searchable through the Search Center portlet. The Administrative Portlet is provided to create tabs in the Portal Search Center and register one or more OmniFind Enterprise Edition collections with each tab. When there is more than one collection, Portal Search Center will federate across them. Portal Search Center is available via the WebSphere Portal Portlet catalog.

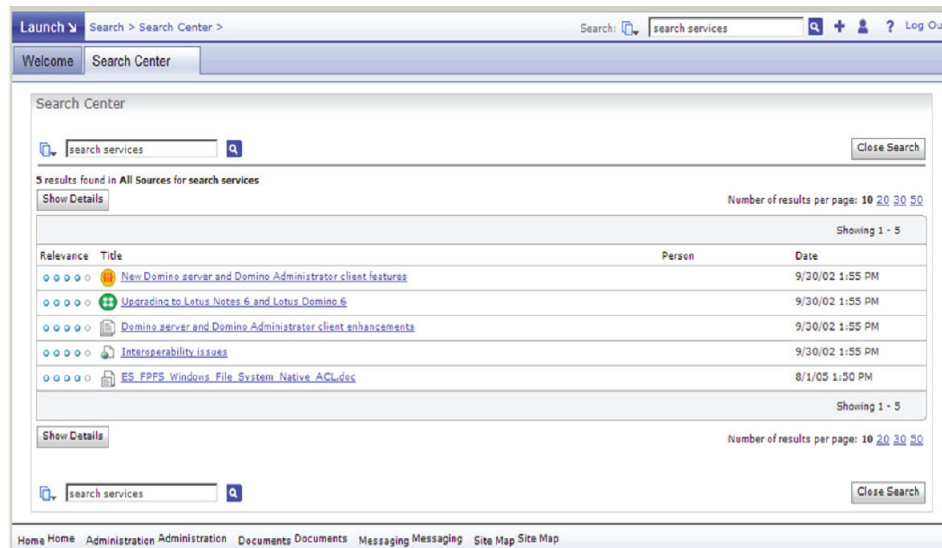


Figure 10: OmniFind Enterprise Edition provides secure access to a broad range of enterprise content from within the WebSphere Portal Search Center portlet.

Sample OmniFind Enterprise Edition search portlet

This portlet is similar in look, feel and function to the search portlet that comes with WebSphere Portal, but it also exposes all of the advanced features/settings available through OmniFind Enterprise Edition, such as control of display options, categorization, the number of results per page, whether to use site collapse, how the results are to be sorted, and so on. The OmniFind Enterprise Edition search portlet can be deployed quickly and easily through an automated command-line script. Also, queries can be redirected from the WebSphere Portal Search bar or Search Center to the OmniFind Enterprise Edition search portlet.

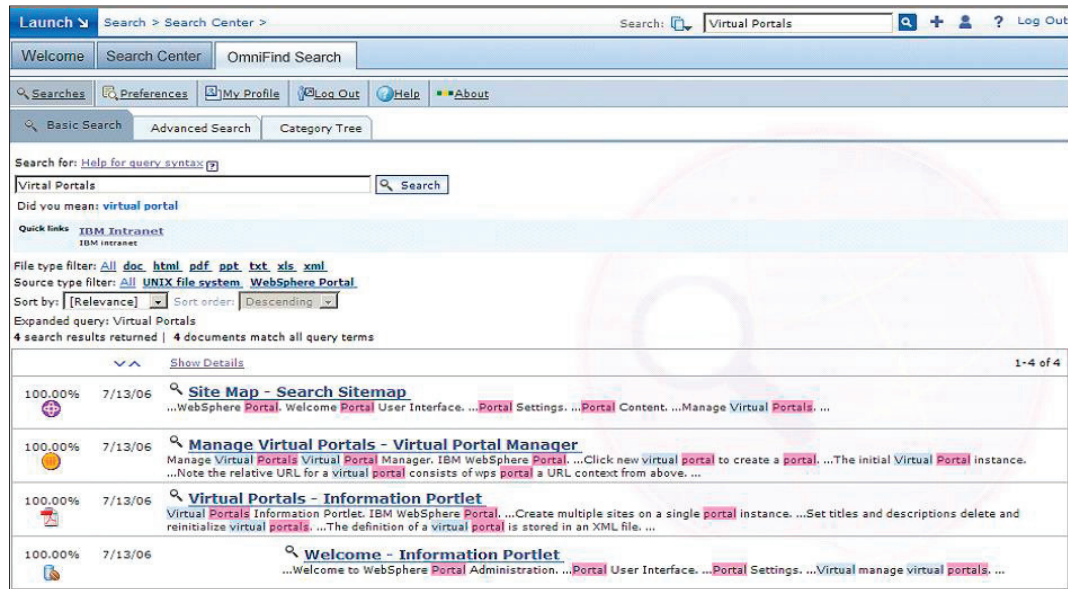


Figure 11: The OmniFind Enterprise Edition search portlet exposes all of the advanced search functionality available in OmniFind Enterprise Edition directly through the Portal.

Additional integration features include:

- A utility for migrating WebSphere Portal search collections to OmniFind Enterprise Edition;
- The ability to crawl WebSphere Portal Document Manager, WebSphere Portal pages, Workplace Web Content Management, and Lotus Quickr Services for WebSphere Portal out-of-the-box using native security;
- Proxy, SSO/SSL support when accessing WebSphere Portal.

To help customers experience first-hand how OmniFind Enterprise Edition can help them extend the reach, scope and value of their WebSphere Portal applications, WebSphere Portal customers are entitled to a limited use license of OmniFind Enterprise Edition (one processor for production and one processor for non-production per enterprise, for a total of 200 Value Units). This enables customers to build small search applications using OmniFind Enterprise Edition to access sources such as Lotus Domino, WebSphere Portal Document Manager, Windows File System, Lotus Quickr, and Workplace Web Content Management. The limited use license of OmniFind Enterprise Edition is applicable to WebSphere Portal Enable and WebSphere Portal Extend, including the limited use versions of these offerings.

Building a world-class corporate intranet: IBM On Demand Workplace

The IBM intranet, internally referred to as “w3”, was established in 1995 as an information source for IBM employees. With the advent of OmniFind Enterprise Edition, w3 has evolved into what is now referred to as the IBM On Demand Workplace, the definitive source for company information, news, and access to critical applications. Today, more than 70% of IBM’s 300,000 employees rely on w3 as the primary and trusted source of company information.

To date, two applications with very different requirements have been implemented:

- Very large scale general intranet search of over 10,000 Web sites and 25 million URLs. Over 6 million unique documents are indexed with a turnaround of four hours for updating the index. This application has been in 24x7 production since September 2003.
- An Employee Profile application for expertise location that searches over 500,000 XML records and makes up to 20,000 updates daily with two- to three-hour turnaround. This application has been in production since March 2004.

The search portlet on w3

While employees can access search from multiple sites on w3, the most popular access point is the search portlet on the w3 home page:

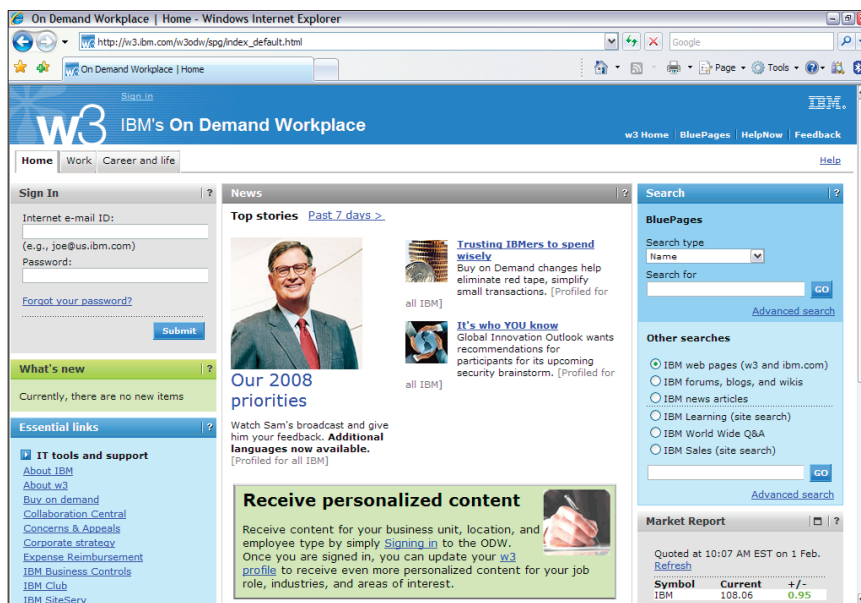


Figure 12: Search is one of the key capabilities of the w3 On Demand Workplace, and can be invoked from multiple sites, including a portlet on the primary page.

The w3 search portlet consolidates several information source searches into a simple and intuitive user interface, using a consistent set of radio buttons. The search options include:

- BluePages – The main IBM employee directory allows users to search for information about people, projects, teams, and interests
- IBM Web pages – This includes www.ibm.com and other external IBM web sites
- IBM Intranet pages – Users can search all publicly accessible content and documents on the IBM intranet, including blogs, wikis, and forums
- Knowledge Assets – This includes various resources for IBM employees, including a Learning Portal, market intelligence reports, FAQs, etc.
- News – Users search current and archived articles from internal and external news sources
- Tagged pages – Users can search on pages that have been tagged by other IBM employees

Behind the scenes, w3 has defined OmniFind Enterprise Edition “collections” to create separate indexes for each of these sources.

Advanced Search

Users can also access an advanced search page from the search portlet by choosing a radio button and then clicking “Advanced search.” Advanced Search offers users the ability to hone in on a result by specifying where in the document a query term should be found. Additionally, users can specify that certain words not be in a document, to help differentiate between similar topics. Users can also refine searches based on URL and language. While w3 architects use the OmniFind Enterprise Edition query scoping to define the advanced search syntax behind the scenes, users can also use this syntax directly at the search prompt.

The Search Results page

The most frequently used page, however, is the search results page:

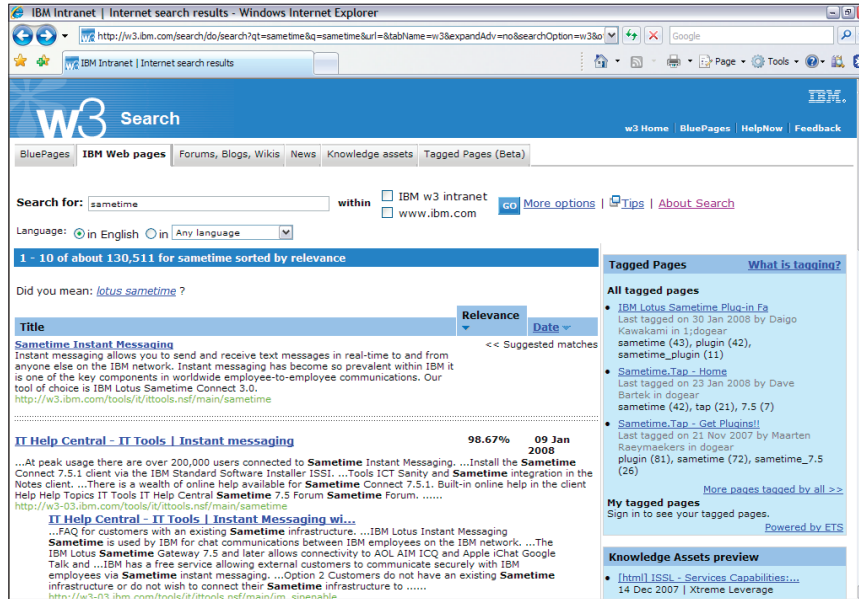


Figure 13: The w3 search results page delivers a number of unique features, including synonym expansion, social bookmarking and expertise location.

This page was developed as a result of user-centered design analysis, and analysis of end user feedback. It is supported by the OmniFind Enterprise Edition Search and Indexing API. The page features:

- **Suggested matches** – Administrators can easily define “don’t miss” items or “best bets” for specific keyword searches to ensure key information is always easy to find.
- **Dynamic summaries** – OmniFind Enterprise Edition uses the key words in the search request to dynamically generate intelligent document summaries. This includes search-term highlighting and helps users to easily determine the content and likely relevance of the document.
- **“Other meanings”** – The system compares the user’s query against a dictionary and suggests other potentially relevant queries or spelling corrections.

- **Site collapse** – Similar results from the same source are collapsed and indented to the right. This prevents any one source from dominating the results. Users can click “More results from...” links to see more results from the same source.
- **Social tagging** – Users can save their own keywords (“tags”) for pages they found useful. The tags are shared across users. Users can also search on pages that have been tagged.
- **Expert locator** – “BluePages matches” identify the appropriate subject matter experts associated with a given query term, and list their name, title and photo.
- **Search resources** – This includes a listing of the most popular w3 searches, and links to the w3 Search Wiki and the IBM corporate taxonomy.

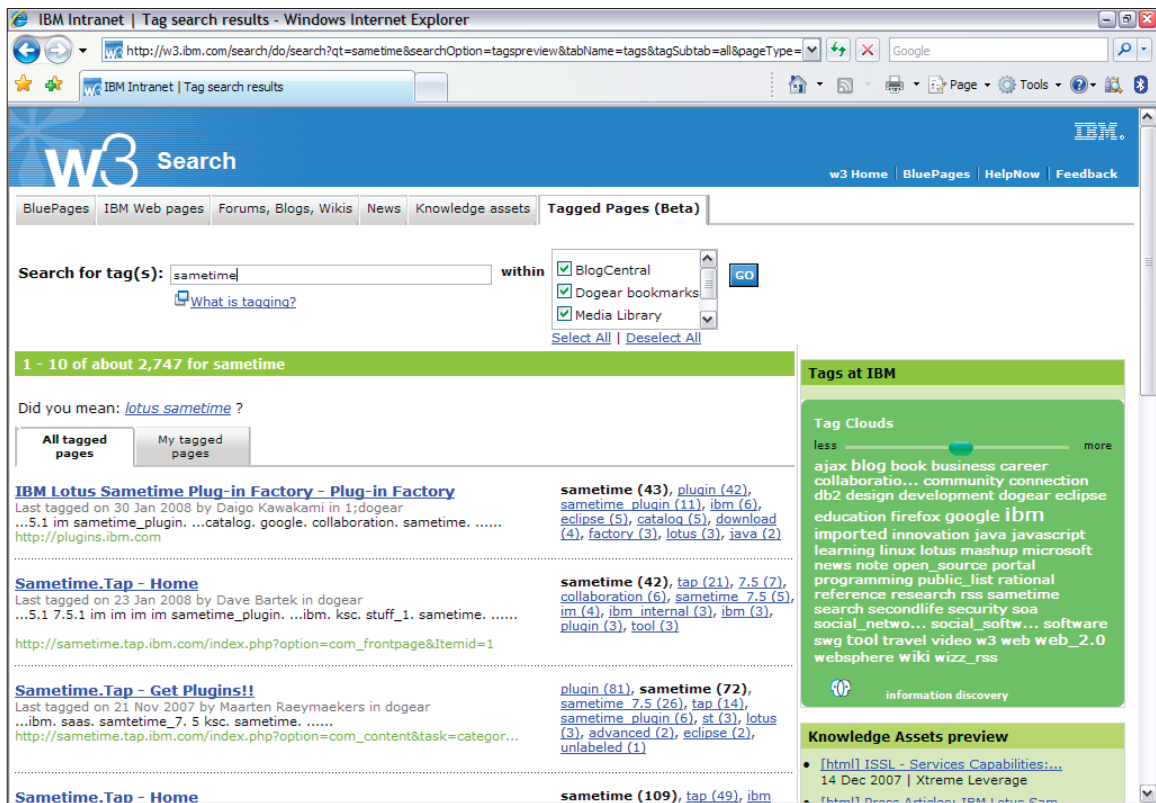


Figure 14: On the IBM Global Intranet, OmniFind Enterprise Edition allows users to search on enterprise content, as well as pages that have been tagged by other IBM employees.

OmniFind Enterprise Edition as a text analytics platform

In addition to providing a robust enterprise search infrastructure, OmniFind Enterprise Edition is the first commercially available UIMA-based platform for processing text-based information. UIMA (Unstructured Information Management Architecture) is an open analysis framework developed by IBM, donated to the open source community (currently an Apache incubator project), and submitted for standardization via OASIS.

Improving the quality of search results

UIMA is an open, standard framework for processing unstructured content to generate more descriptive and semantically rich metadata. UIMA enables seamless integration of text analytics components that analyze documents and extract additional knowledge to create an enhanced index for searching. This can be used to identify higher-level concepts such as people, places, organizations, products, problems, and other “entities” that are buried within unstructured content.

The quality of results is greatly enhanced by enabling users to search on the more detailed metadata. The accuracy and relevancy of search results can be even more dramatically improved by leveraging the relationships and facts that have been discovered. This enables users to more easily find the specific documents relevant to a relationship or fact of interest.

A rich text analytics platform

OmniFind Enterprise Edition provides the ability to collect and parse content from various sources, process that content to apply text analytics through the UIMA framework, index the enhanced metadata for advanced searching capabilities, and send the extracted information to databases, data warehouses, ontologies, and other external knowledgebases. OmniFind Enterprise Edition includes some core text analytic capabilities, such as advanced linguistics, extraction of word roots and thesaurus support for identifying synonyms.

Third party and custom text analytic components can be plugged in and configured through standard administration tools to generate additional metadata. In addition, IBM provides customers with sample entity extractors that demonstrate how to enhance content indexing with people, places, product names and other entities identified within unstructured content. Many of these samples can even be customized to address company specific concepts. Finally, OmniFind Enterprise Edition enables users to perform keyword or semantic searches to find the most relevant content.

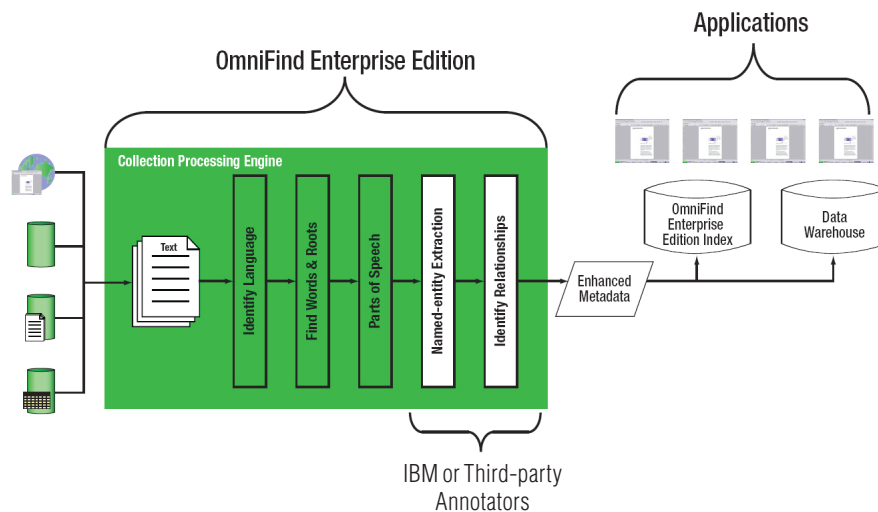


Figure 15: *OmniFind Enterprise Edition is the first commercially available UIMA-based platform for processing text-based information.*

The enhanced metadata extracted from unstructured content can also be sent to external data sources, such as databases or data warehouses, where standard reporting tools can be used to perform analysis on the new set of knowledge. This enables companies to more quickly spot inconsistencies, identify outliers and discover trends that were previously buried in documents. The resulting faster discovery of information can have direct business benefits, such as reducing costs through quicker problem detection and resolution, minimizing the associated risks, and speeding time to market for new offerings.

Maximizing the value of existing text analytics implementations

Most companies that have implemented text analytics have done so as part of very specific applications or initiatives. They have invested a significant amount of time and money to capture and codify the methods and terminologies related to their business, and extract a deeper set of knowledge from various sets of unstructured content. This has also typically required significant integration work. After all of this effort, companies have limited the use of this enhanced set of information to a small minority of their employees solely for deep analytic purposes.

However, by leveraging a broader platform for deployment of these same text analytics capabilities, in combination with enterprise search engines and integration with their business applications, companies can start taking greater advantage of these investments and realize much broader benefits from the technology. The same information that is being extracted as part

of existing processes, along with additional related content, can be funneled into an enterprise search index in parallel. This will provide a larger set of users across the organization with the ability to more easily search for and leverage the knowledge that is being created through simplified search interfaces that provide sub-second response times.

Application examples

Online Media Analysis

The proliferation of blogs, news feeds, consumer review sites, newsgroups and articles published daily on the Web represents a tremendous opportunity for today's companies. These sources are filled with insight from consumers, experts and competitors that can be analyzed and used by businesses to manage reputational risk, and make better decisions on products, services and business strategies. The Online Media Analysis Solution combines OmniFind Enterprise Edition and text analytics modules from IBM to enable organizations to extract valuable meaning from text-based information, and track, in real-time, how they are perceived by customers, competitors, analysts, investors, stakeholders and other constituents.

The solution extends the OmniFind Enterprise Edition search application to include an analysis dashboard and dynamic charts to show Media Presence for understanding the importance of a given topic; Tonality to understand whether the mood is positive, negative, neutral, or ambivalent; Hot Terms, i.e., the most prevalent terms in the set of documents, signifying emerging topics or concepts; Association Analysis for understanding which pre-defined concepts occur frequently with a given search term; and, Timeline Analysis for monitoring topics over time to understand shifts in media presence and tonality. The solution can be customized to meet specific business needs.

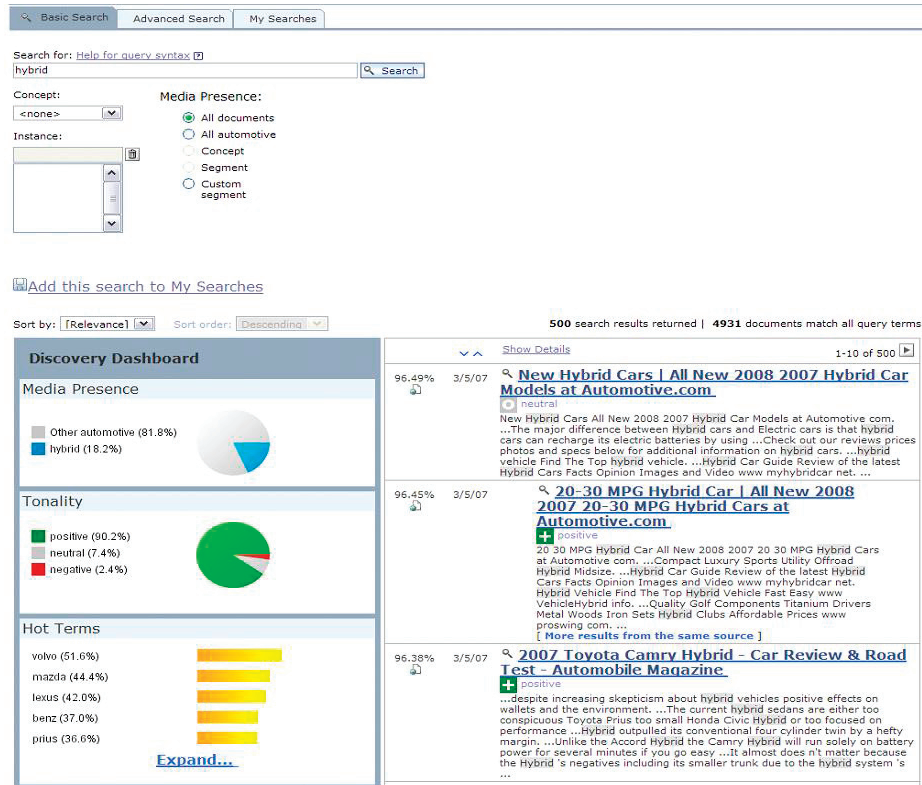


Figure 16: The Online Media Analysis Solution provides real-time insight into public perception and sentiment.

Most business intelligence applications report on and analyze structured data, but fail to take advantage of unstructured content, such as contracts, memos, customer correspondence, or free-form data such as comment, description and note fields stored in databases. This prevents organizations from getting a complete picture of operations, opportunities and risks. Applying text analytics can enable business intelligence tools to extract facts from unstructured content assets and incorporate this knowledge into reporting and analysis for a more complete and accurate picture of an organization, including its products, services, customers and suppliers.

The unique combination of OmniFind Enterprise Edition and UIMA empowers a new class of text-analysis-based applications. Some of these solutions are being delivered in conjunction with IBM business partners.



Figure 17: Issues Analytics and Quality Early Warning Applications can deliver richer reports that identify other parts involved in a problem and provide an understanding of the actual conditions that cause failures.

Issues Analytics and Quality Early Warning systems

Most organizations can generate problem reports based on the actual problem codes, but once a pervasive problem is identified, they cannot easily determine the actual conditions that caused the problem, other parts and components involved, and all of the actions that may have been taken. The Quality Early Warning solution combines OmniFind Enterprise Edition and UIMA to extract this knowledge from unstructured information sources, such as technician comments and customer complaints, to enable the creation of rich reports that deliver a greater understanding of product failures. Analysis of service and maintenance records provides early insight into product defects and service issues before they become widespread, enabling quicker resolution and lower after-market service and recall costs. Customers can even set up alerts to be automatically notified about frequent occurrences of certain symptoms, and expose warranty and repair information through more simplified search interfaces.

Advanced intelligence for anti-terrorism and law enforcement

Most of the information available to intelligence analysts exists in unstructured formats such as internal documents, surveillance reports, audio from phone taps, satellite imagery, local media Web sites, transcripts, and e-mail. OmniFind Enterprise Edition and UIMA empower analysts to analyze and interpret this information to connect the dots, assess threats and respond accordingly. Insight mined from unstructured data helps analysts to uncover hidden patterns, understand potential implications behind events or activities, and identify potential criminal or terrorist activity.

Conclusion

Today's companies need to maximize the value of enterprise information to improve business processes and enable more informed decision making. This can be achieved by implementing a strategic, SOA-compliant enterprise search architecture that can offer rich text analytic capabilities to uncover hidden intelligence and value from structured and unstructured data. Leveraging its deep-seated expertise in text analytics and semantic search technology, and its leadership in Information on Demand, IBM is uniquely positioned to deliver a strategic enterprise search and analysis platform to help companies improve productivity and gain unprecedented levels of business insight. OmniFind Enterprise Edition provides a robust, highly scalable and secure enterprise search architecture to support enterprise portal and business intelligence initiatives. It also provides a rich, UIMA-based text analytics foundation, helping organizations to get the most out of their information assets.

Appendix A - Supported languages

OmniFind Enterprise Edition supports the searching of over 50 languages. Advanced linguistic analysis is provided for the following languages:

Arabic	Chinese (Traditional, Simplified)	Czech
Danish	Dutch	English
Finnish	French (Continental, Canadian)	German
Greek	Hebrew (via Services engagement)	Italian
Japanese	Korean	Norwegian (Nynorsk, Bokmal)
Portuguese (Continental, Brazilian)	Russian	Spanish
Swedish		

The following languages can be automatically detected, but base-form computation is not currently supported:

Hungarian	Turkish
-----------	---------

Basic linguistic support in OmniFind Enterprise Edition means that the language is segmented using either white space information or n-grams. No language limitation is built into OmniFind Enterprise Edition, so a wider set of languages are searchable than the list below. IBM has tested the following list of additional languages:

Simple text languages

Albanian	Azerbaijani (Latin)	Bulgarian	Byelorussian
Catalan	Croatian	Estonian	Hungarian
Icelandic	Indonesian	Kazakh	Latvian
Lithuanian	Macedonian	Malay	Romanian
Serbian (Cyrillic, Latin)	Slovak	Slovenian	Turkish
Ukrainian			

Complex text languages

Bengali	Gujarati	Hindi	Kannada
Malayalam	Marathi	Oriya	Punjabi
Tamil	Telugu	Thai	Vietnamese



© Copyright IBM Corporation 2009

IBM
Route 100
Somers, NY 10589
U.S.A.

Produced in the United States of America
02-08
All Rights Reserved

- 1 UIMA (Unstructured Information Management Architecture) is an open source, standards based framework for processing unstructured information through text analytics.
- 2 For more information, please refer to the section entitled, "IBM On Demand Workplace."

DB2, DB2 Universal Database, Domino, IMS, Informix, iSeries, LanguageWare, Lotus, Lotus Notes, Notes, OmniFind, QuickPlace, SmartSuite, WebSphere and z/OS are trademarks of International Business Machines Corporation in the United States, other countries or both.

Java and all Java-based trademarks are trademarks Sun Microsystems, Inc. in the United States, other countries or both.

Microsoft, SharePoint and Windows are registered trademarks of Microsoft Corporation in the United States, other countries or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Other company, product or service names may be trademarks or service marks of others.

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates. Offerings are subject to change, extension or withdrawal without notice.

All statements regarding IBM future direction or intent are subject to change or withdrawal without notice and represent goals and objectives only.

For more information about OmniFind, visit ibm.com/software/data/enterprise-search