

Selecting the best search solution for your IBM WebSphere Portal installation

John McPherson

*IBM Distinguished Engineer, Co-Leader IBM Institute of Search and
Text Analysis*

Andreas Prokoph

Lead Architect, Search Technologies, IBM

Joel Waterman

Worldwide Sales Executive, Content Discovery, IBM

Lauren Wendel

Product Manager, WebSphere Portal, IBM

Contents

2 Introduction

4 Embedded search versus stand-alone technology

5 WebSphere Portal Search Engine versus IBM OmniFind Enterprise Edition

6 Product features and descriptions

6 WebSphere Portal—Portal Search Engine

8 OmniFind Enterprise Edition

10 Moving from WebSphere Portal Search to OmniFind Enterprise Edition

11 User interfaces

11 Search Center portlet

12 Search and Browse portlet

13 OmniFind Enterprise Edition sample search application

14 Administration features

Introduction

IBM search technology is available as stand-alone IBM® OmniFind™ Enterprise Edition software or through the embedded Portal Search Engine feature of IBM WebSphere® Portal software. IBM research and development organizations have invested heavily in delivering a common search technology framework, helping to ensure that a best-of-breed approach to search can be used throughout the IBM product portfolio.

The Portal Search Engine feature integrates seamlessly into WebSphere Portal and can also support IBM Lotus® Quickr™ and Web Content Management applications. It is the ideal solution for searching within WebSphere Portal, Lotus Quickr and Web Content Management resources, as well as moderately sized Web content collections.

Using the same underlying technology framework as Portal Search Engine, OmniFind Enterprise Edition provides significant enhancements in scalability and performance. It has enhanced administrative tools as well as the ability to pull content from many more enterprise repositories. These capabilities, paired with additional search features and functions, add to the end user's search experience and make OmniFind Enterprise Edition the strategic IBM enterprise search middleware platform.

This portal search solution guide presents a high-level overview of some technology, selection and deployment issues customers should consider when deploying search functionality in the WebSphere Portal environment. In brief, this guide recommends that Portal Search Engine capabilities embedded in WebSphere Portal are best suited for small- to moderate-sized portal deployments where access to Web-enabled data is a prime consideration.

Customers who need both a scalable search engine—regardless of which portal product they use—and native access to a large set of different data sources should consider OmniFind Enterprise Edition (see Figure 1).

The guide also describes a scenario for large corporations in which user-interface portlets make it transparent to end users that both WebSphere Portal Search and OmniFind Enterprise Edition can coexist to power the search function.

Figure 1: WebSphere Portal Search for small to medium-sized deployments and OmniFind Enterprise Edition for large deployments

Current environment	WebSphere Portal Search Engine	OmniFind Enterprise Edition
<800K documents to search	✓	
>800K documents to search		✓
Small to medium-sized portal	✓	
Enterprise scale portal in a multi-lingual, cross-geo environment		✓
Non-IBM search solution and my contract is up for renewal		✓
Need for a search middleware solution		✓
Only Web-enabled data sources to search	✓	
Multiple data sources, need native access		✓

Embedded search versus stand-alone technology

The IBM search strategy encompasses embedded search components as well as a stand-alone enterprise search solution. The standard version of search is incorporated into WebSphere Portal and the enterprise version is available through OmniFind Enterprise Edition. Both are used by companies in a broad range of industries worldwide.

The search engine within WebSphere Portal offers an entry-level capability, suitable for searching through moderate-sized WebSphere Portal sites and managed content in small to medium-sized WebSphere Portal installations. Because it is embedded in the application—WebSphere Portal, Lotus Quickr or Web Content Management—the search service is not available separately for purchase. The embedded search feature also is designed to search content specific to the application.

OmniFind Enterprise Edition leverages many of the same search technology components found in WebSphere Portal Search Engine. OmniFind Enterprise Edition, a true, enterprise-strength search offering, adds to the services provided in the WebSphere Portal Search Engine and offers scalability, performance and reach into many different data sources that large organizations require. It also includes additional features and functions to further improve the administrator and end-user search experience.

Portal Search Engine technology provides the ability to index and search portal portlets and pages via portal site search, including Web, public IBM Workplace Web Content Management™ and WebSphere Portal Document Manager content and attachments. This technology is also embedded in Lotus Quickr applications and provides search support for team collaboration and document management capabilities.

OmniFind Enterprise Edition is designed to support enterprise-wide search requirements for corporations that have a large variety of content sources. It is easily scalable and supports native document-level security for many enterprise repositories. It is integrated into portals with the same end-user search interfaces used in WebSphere Portal. In addition, OmniFind Enterprise Edition has a browser-based administrative interface and offers a natural enterprise-scale upgrade path from standard WebSphere Portal search.

WebSphere Portal Search Engine versus OmniFind Enterprise Edition

Portal Search Engine offers:

- *Search capabilities for less than 800,000 documents in WebSphere Portal, all stored in portal repositories or Web-enabled content sources*
- *Secure access to information and resources managed by Portal Access Control (Portal pages and portlets, Workplace Web Content Management content and Portal Document Manager documents)*
- *Search of repositories and Web content across a small to medium-sized WebSphere Portal deployment*
- *Fast integration of search features into WebSphere Portal while maintaining the option to migrate easily to an enterprise search environment at a later time*

OmniFind Enterprise Edition offers:

- *Secure access to information stored across the organization*
- *Easier and quicker discovery of relevant information through a single point of access, helping to improve worker productivity*
- *Increased adoption and return on investment for portal applications by delivering a richer search experience, including semantic search capabilities*
- *A scalable enterprise search engine*
- *Search capabilities across multiple data sources that include both Web-enabled and non-Web-enabled content*
- *Search functionality as a separately available middleware layer, allowing different search applications to be built on a single stable platform*
- *Ability to deploy across the enterprise whether or not WebSphere Portal Search is deployed*

Product features and descriptions

Important features and functions of WebSphere Portal Search Engine and OmniFind Enterprise Edition described below can help provide guidance for companies trying to determine the search capability appropriate for their business.

WebSphere Portal Search Engine

WebSphere Portal Search Engine offers many capabilities.

- **High-quality search and document ranking.** Portal Search Engine factors in dozens of variables—such as word statistics or proximity of terms—for each query and goes beyond mere link analysis to determine relevancy.
- **Document summaries.** These summaries are generated automatically when the content is indexed and displayed in search results. Languages available include Danish, Dutch, English, French, German, Italian, Norwegian, Portuguese, Russian, Spanish and Swedish.
- **Highlighted query terms.** The list that results from the query highlights the query terms.
- **Option to limit search results.** Specific languages, file types, Web sites, data sources or specific field values can be specified to help limit search results.
- **Search scope.** The information in all or selected search collections can be segmented to allow users to choose and search for a specific type of information; for example, by domain name, by geography, by job function or by specific repository and so on.
- **Schedule crawls.** The update rate of the content can be controlled.
- **File/attachment index and search support.** More than 250 file types, including HTML, XML, Microsoft® Office (Word, PowerPoint, Excel), Adobe® Portable Document Format (PDF), Adobe PostScript®, WordPerfect, IBM Lotus® SmartSuite® (1-2-3, WordPro, Freelance Graphics), MacWrite and many others are supported.

- **Categorized indexed content.** Content can be categorized using a predefined taxonomy or a rules-based taxonomy. WebSphere Portal Server comes with a predefined taxonomy containing more than 2,300 named categories for subject areas across 12 business disciplines. User-defined categories can also be added. The taxonomy can be further customized with IBM Taxonomy Services, allowing users to take advantage of the entire IBM taxonomy capability, encompassing more than 9,000 business categories focused on 20 different industries. IBM OmniFind Enterprise Edition uses IBM Classification Module for auto-classification and taxonomy management. Both Portal Search Engine and OmniFind Enterprise Edition can also categorize indexed documents according to filter rules—content-based or URL matching filters.
- **Configuration.** Portal Search Engine can be configured to run on the same box as WebSphere Portal Server. For a cluster of portal servers, Portal Search Engine is set up as a remote search service supporting Simple Object Access Protocol (SOAP) or deployed as an Enterprise JavaBean™ (EJB™). It can be installed either on a dedicated computer with only WebSphere Application Server installed on it or on an existing machine in the cluster.
- **Portal site search capability.** This capability supports WebSphere Portal portlet and page searches. The matching of search results to portal access control definitions helps to ensure security of search results that match user access rights.
- **Personalized and secured portal site search results.** The results of searching a WebSphere Portal Server page or a portlet are personalized for that user in the context of the person's unique portal and page configuration.
- **Portal search box.** Portal search services can be accessed from any portal page through the search box in the theme of the pages.
- **Search Center portlet.** This provides a centralized search interface across portal and external Web site content.
- **Remote services.** Portal Search Engine can be deployed as an SOAP- or EJB-based remote service. One or more services can be deployed to enable search in a cluster environment as well as for load-balancing or distributed administration.

OmniFind Enterprise Edition

OmniFind Enterprise Edition includes the Portal Search Engine features listed above as well as additional functionality

Enhanced enterprise features

- **Access to more enterprise data sources.** *These data sources include Microsoft Exchange public folders, IBM DB2® on Linux®, UNIX®, Microsoft Windows®, IBM z/OS®, IBM Informix®, Oracle, Microsoft SQL Server™ and other databases, EMC® Documentum®, IBM FileNet™ P8N, Windows SharePoint® Services and other content management systems, Lotus Notes®/Domino®, Lotus QuickPlace® and Domino Document Manager, and Windows and UNIX file systems. In addition, enterprise search contains a flexible architecture that supports the development of connectors to other data sources.*
- **Enhanced language support.** *More sophisticated language-specific processing is available for 20 languages, including morphological text analysis and overall search capabilities in 50 languages, plus enhanced support for Asian languages.*
- **Enterprise security features.** *A flexible security infrastructure allows OmniFind Enterprise Edition to support native document-level security for many data sources.*
- **Prepared for advanced text analysis.** *OmniFind Enterprise Edition is the strategic platform for the IBM Unstructured Information Management Architecture (UIMA). UIMA is an open-source text analytics application programming interface (API) supporting specific features that can enrich search by, for example, associating company-specific acronyms or terminology to make semantic search a possibility.*

Enhanced search quality features

- **Support for static rank analysis.** These features support link analysis for further improvements when searching enterprise information.
- **Dynamic summarization.** Generated at query time, document summaries accurately reflect those parts of the document that relate directly to the keywords used in the search query.
- **Automatic spell check.** This feature automatically suggests corrections for misspelled search terms.
- **Automated Taxonomy creation.** Integration with IBM Classification Module for automatic classification of documents into taxonomic categories. (For more information about taxonomy management, visit: ibm.com/developerworks/db2/library/techarticle/dm-0702dorre)

Enhanced administration features

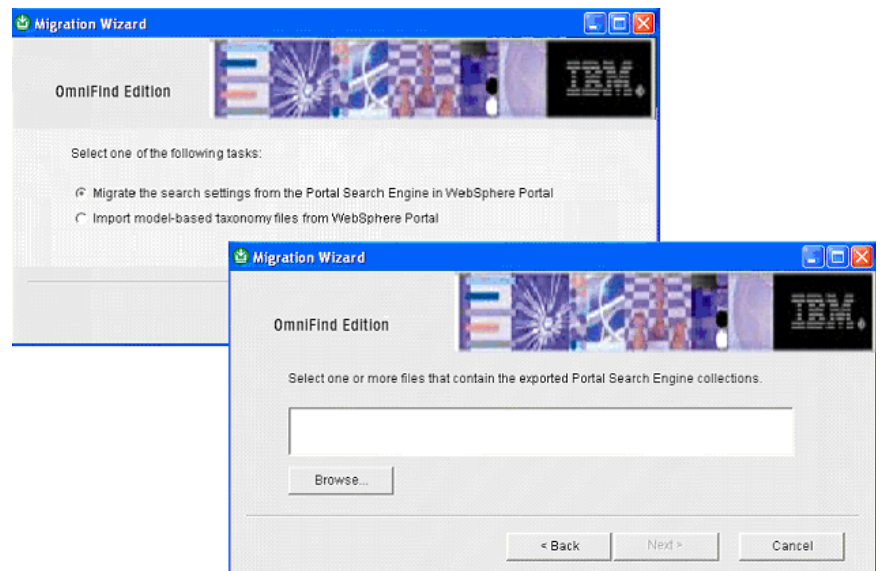
- **Predefined links.** Search results can include predefined links. Keywords are associated with specific links via the administrator interface.
- **Reporting view and exporting.** Daily and hourly reports show top queries, response time, queries served per second and more.
- **URL tracking.** View analysis of all crawled content can quickly identify problematic servers, errors and sources of content.
- **Problem reporting.** Automatic monitoring of crawler performance provides auto-alert e-mail messages to administrator.

Moving from WebSphere Portal Search to OmniFind Enterprise Edition

An easy-to-use WebSphere Portal Migration Wizard (see Figure 2) allows customers to move the Portal Search Engine collection to the OmniFind Enterprise Edition environment in four steps:

1. From the portal, export Portal Search Engine settings to XML files.
2. Copy files to the OmniFind Enterprise Edition index server.
3. Import settings using WebSphere Portal Migration Wizard.
4. Modify using OmniFind Enterprise Edition Admin Console.

Figure 2: Migration wizard helps move Portal Search Engine collection to OmniFind Enterprise Edition environment



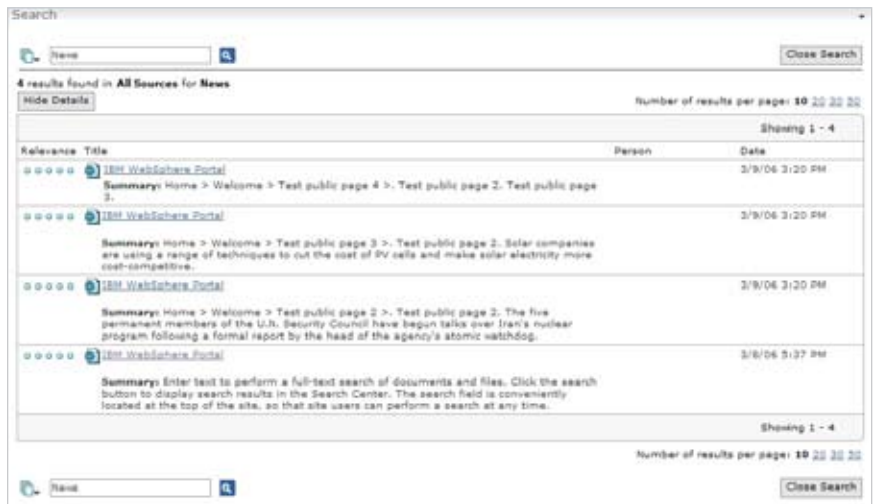
User interfaces

Both Portal Search Engine and OmniFind Enterprise Edition integrate seamlessly with WebSphere Portal via standard search portlets. Because Portal Search Engine is embedded into WebSphere Portal, its use is limited solely to that environment. Customers who are not using WebSphere Portal but are considering IBM search, can use OmniFind Enterprise Edition, which comes with a sample search application and a standards-based application integration method.

Search Center portlet

The Search Center portlet—the central search interface portlet provided in WebSphere Portal—can search various search indexes (see Figure 3). Tabs, which can represent one or more searchable sources, manage the indexes. These can be Portal Document Manager libraries, IBM Workplace Web Content Management sites or Portal Search Engine search collections. When used with OmniFind Enterprise Edition, the Search Center application extends beyond

Figure 3: The Search Center portlet—the central search interface portlet provided in WebSphere Portal

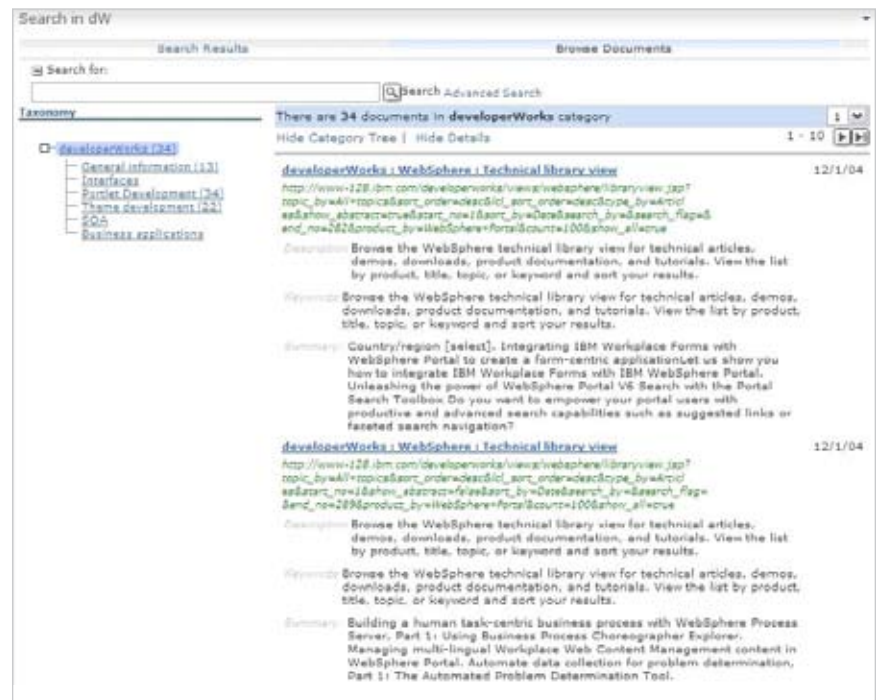


WebSphere Portal sources to include additional enterprise collections or other search services through their native Web interfaces. Search Center can therefore coexist with other search engines, providing the end user with a consistent look and feel while the product in the back-end uses another search engine.

Search and Browse portlet

WebSphere Portal provides another portlet, the Search and Browse portlet, which can be used seamlessly by OmniFind Enterprise Edition. This portlet interface offers advantages over the Search Center portlet when advanced search capabilities are important and when users want the ability to browse through the document collection via an associated taxonomy. One advantage, for example, is the ability to navigate through and search within categorized content (see Figure 4).

Figure 4: Advanced functionality through the Search and Browse portlet



OmniFind Enterprise Edition sample search application

Since Portal Search Engine is embedded in WebSphere Portal and not available separately, companies with other portals or applications that need full enterprise search capabilities should use the flexible Java™ API and the sample search application in OmniFind Enterprise Edition to integrate search functionality into their enterprise. Figure 5 shows a sample search page when using the sample search application.

Figure 5: Flexible Java API and sample search application within OmniFind Enterprise Edition



Administration features

WebSphere Portal Search Engine has a portlet-based administrative interface (see Figures 6 and 7) that allows administrators to create and maintain multiple search services, search collections (search indexes) and search scopes. Each search collection can contain content from multiple sources, such as Web sites.

Figure 6: Portlet-based administrative interface for WebSphere Portal Search Engine for creating and maintaining multiple search services, search collections and search scopes

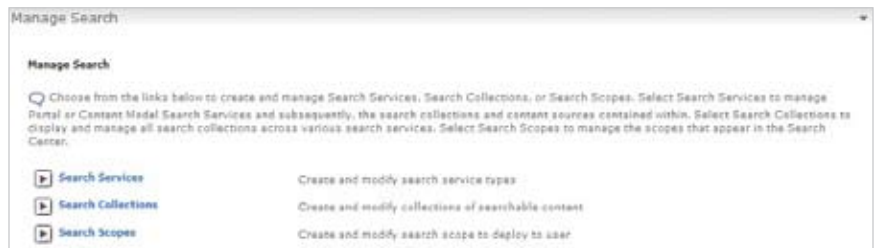
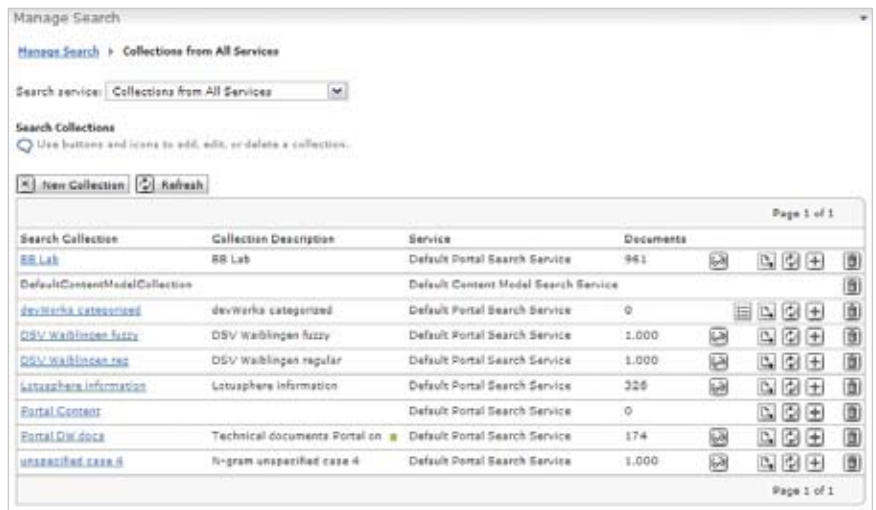


Figure 7: Administrative interface for creating and maintaining multiple search collections



OmniFind Enterprise Edition provides a rich set of tools for the search administrator. An extensive Web-based user interface helps administrators configure, operate and monitor all aspects of the product. Discovery capabilities enable administrators to choose what data sources they want to crawl—all the way down to the field level. All running components can be monitored and reports can be generated to follow activity or fine-tune the configuration, including search response time and top queries. E-mail alerts can be triggered by events. Role-based administration can be used to specify multiple administrators with different privileges. Four different roles are recognized, each associated with different access rights: super administrator, collection administrator, operator and monitor.



For more information

For more information about OmniFind Enterprise Edition and other IBM OmniFind solutions, contact your IBM marketing representative or IBM Business Partner, or visit ibm.com/software/data/enterprise-search/

To learn more about WebSphere Portal, contact your IBM marketing representative or IBM Business Partner, or visit ibm.com/websphere/portal

© Copyright IBM Corporation 2007

IBM Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
4-07

All Rights Reserved

IBM, the IBM logo, DB2, Domino, FileNet, Informix, Lotus Notes, OmniFind, Quickr, QuickPlace, SmartSuite, WebSphere, Workplace, Workplace Web Content Management and z/OS are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries or both.

Adobe, Acrobat, Portable Document Format (PDF), and PostScript are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, other countries or both.

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

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

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

Linux is a trademark of Linus Torvalds in the United States, other countries or both.

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.

The information contained in this document is provided for informational purposes only. While efforts were made to verify the completeness and accuracy of the information contained in this document, it is provided "as is" without warranty of any kind, express or implied. In addition, this information is based on IBM's current product plans and strategy, which are subject to change by IBM without notice. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this document or any other documents. Nothing contained in this document is intended to, nor shall have the effect of, creating any warranties or representations from IBM Software.

TAKE BACK CONTROL WITH **Information Management**