



BEA WebLogic Portal™

Getting Started with Autonomy Search

Copyright

Copyright © 2005 BEA Systems, Inc. All Rights Reserved.

Restricted Rights Legend

This software and documentation is subject to and made available only pursuant to the terms of the BEA Systems License Agreement and may be used or copied only in accordance with the terms of that agreement. It is against the law to copy the software except as specifically allowed in the agreement. This document may not, in whole or in part, be copied, photocopied, reproduced, translated, or reduced to any electronic medium or machine readable form without prior consent, in writing, from BEA Systems, Inc.

Use, duplication or disclosure by the U.S. Government is subject to restrictions set forth in the BEA Systems License Agreement and in subparagraph (c)(1) of the Commercial Computer Software-Restricted Rights Clause at FAR 52.227-19; subparagraph (c)(1)(ii) of the Rights in Technical Data and Computer Software clause at DFARS 252.227-7013, subparagraph (d) of the Commercial Computer Software--Licensing clause at NASA FAR supplement 16-52.227-86; or their equivalent.

Information in this document is subject to change without notice and does not represent a commitment on the part of BEA Systems. THE SOFTWARE AND DOCUMENTATION ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND INCLUDING WITHOUT LIMITATION, ANY WARRANTY OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. FURTHER, BEA Systems DOES NOT WARRANT, GUARANTEE, OR MAKE ANY REPRESENTATIONS REGARDING THE USE, OR THE RESULTS OF THE USE, OF THE SOFTWARE OR WRITTEN MATERIAL IN TERMS OF CORRECTNESS, ACCURACY, RELIABILITY, OR OTHERWISE.

Trademarks or Service Marks

BEA, BEA JRockit, BEA Liquid Data for WebLogic, BEA WebLogic Server, Built on BEA, Jolt, JoltBeans, SteelThread, Top End, Tuxedo, and WebLogic are registered trademarks of BEA Systems, Inc. BEA AquaLogic, BEA AquaLogic Data Services Platform, BEA AquaLogic Enterprise Security, BEA AquaLogic Service Bus, BEA AquaLogic Service Registry, BEA Builder, BEA Campaign Manager for WebLogic, BEA eLink, BEA Manager, BEA MessageQ, BEA WebLogic Commerce Server, BEA WebLogic Enterprise, BEA WebLogic Enterprise Platform, BEA WebLogic Enterprise Security, BEA WebLogic Express, BEA WebLogic Integration, BEA WebLogic Java Adapter for Mainframe, BEA WebLogic JDriver, BEA WebLogic JRockit, BEA WebLogic Log Central, BEA WebLogic Personalization Server, BEA WebLogic Platform, BEA WebLogic Portal, BEA WebLogic Server Process Edition, BEA WebLogic WorkGroup Edition, BEA WebLogic Workshop, and Liquid Computing are trademarks of BEA Systems, Inc. BEA Mission Critical Support is a service mark of BEA Systems, Inc. All other company and product names may be the subject of intellectual property rights reserved by third parties.

All other trademarks are the property of their respective companies.

Contents

Autonomy Search

| | |
|---|---|
| Introducing the Autonomy Components | 2 |
| Locating the Autonomy Product. | 2 |
| Licensing Autonomy Modules | 3 |
| Determining the Number of CPUs for Your Search Needs | 3 |
| Using Supported Repositories and Databases | 3 |
| Using Supported Operating Systems and Platforms | 6 |
| Searching the Oracle Database on UNIX | 6 |
| Using the Autonomy Search Engine | 6 |
| Configuring the Autonomy Search Engine. | 7 |
| Exposing Data from the Virtual Content Repository to Search | 7 |
| Using Other Search Portlets | 7 |
| Localizing Autonomy Portlets | 7 |
| Choosing a Search API. | 7 |
| Integrating Autonomy with Other Search Engines. | 8 |
| Upgrading and Getting Support | 8 |
| Upgrading Autonomy | 8 |
| Contacting Support | 8 |

Autonomy Search

BEA WebLogic Portal ships with a license of Autonomy search capabilities. Common use cases include integration with content management systems, relational databases, and external web sites. These sources of information can be exposed for search using prepackaged portlets, and developers can author new portlets and business logic for integrating search as well.

This guide covers the capabilities of the licensed Autonomy products and information on additional Autonomy capabilities which can be resold.

This chapter includes the following sections:

- [Introducing the Autonomy Components](#)
- [Using Supported Repositories and Databases](#)
- [Using Supported Operating Systems and Platforms](#)
- [Using the Autonomy Search Engine](#)
- [Upgrading and Getting Support](#)

Introducing the Autonomy Components

The Dynamic Reasoning Engine (DRE) server has been renamed Intelligent Data Operating Layer (IDOL). IDOL is Autonomy's core engine that processes incoming content to enable search functionality.

Autonomy provides Fetch modules to connect sources of information to search. The BEA license includes the following modules:

- The Autonomy HTTPFetch™ module for external web sites
- The Autonomy ODBCFetch™ module for relational databases
- The Autonomy FileSystemFetch™ module for files on the file system

You can integrate the Autonomy HTTPFetch module to create applications that retrieve content from the Internet or an intranet using the HTTP protocol.

The following Autonomy features can be used with WebLogic Portal:

- Multiple socket connections for optimal bandwidth usage
- Simultaneous retrieval of multiple remote and local web site pages
- Conditional retrieval based on date ranges and configurable date formats
- Conditional retrieval based on configurable inclusive and exclusive wild card lists, optionally checked in the URL, header and body
- Optional politeness delay between retrievals and retrieval rate
- Support for HTTP authentication, HTML forms, login, and HTML cookie authentication
- Insertion of meta tags for import or index processes
- Insertion of original reference to allow viewing of retrieved pages
- Support for proxy servers and firewalls

Locating the Autonomy Product

Autonomy is bundled with the WebLogic Portal and WebLogic Platform installers. The files are located in the *WeblogicHome\portal\thirdparty\search* directory.

See <http://www.autonomy.com/content/Products/Connectors/> for a complete list of Autonomy components.

Licensing Autonomy Modules

In BEA WebLogic Portal 8.1, customers can install one instance each on production, development, and a failover instance on their portal. This license is not a CPU-by-CPU basis; the license model is one portal to one Autonomy server. For more information, see [“Licensing Autonomy Modules” on page 3](#). The license is paper-based, and all development instances ship with a full production version of Autonomy. A production instance of Autonomy is included with the WebLogic Portal installation.

Determining the Number of CPUs for Your Search Needs

The number of CPUs that you need for a production instance varies with the number and type of documents you are exposing, as well as the way they are exposed (for example, automated searching, user driven, and so on).

A single instance of one CPU can potentially support tens of thousands of users and millions of documents. Contact your BEA or Autonomy sales representative for additional licenses, if needed.

Using Supported Repositories and Databases

Autonomy can supply standard connectors (also called Fetches) for proprietary data repositories and different file formats.

[Table 3](#) lists the required supported file system repositories and connectors.

Table 1 File System Repositories and Connectors

| Supported Repository | Required Connector |
|----------------------|--------------------|
| NT File Systems | AutoIndexer |
| Network File Systems | AutoIndexer |
| UNIX File Systems | AutoIndexer |

[Table 2](#) lists the required supported relational database repositories and connectors.

Table 2 Relational Database Repositories and Connectors

| Supported Repository | Required Connector |
|----------------------|--------------------|
| DB2 | ODBCFetch |
| Microsoft SQL | ODBCFetch |
| ODBC Databases | ODBCFetch |
| Oracle Database | OracleFetch |

[Table 3](#) lists the required supported document repositories and connectors for external web sites.

Table 3 Document Repositories and Connectors for External Web Sites

| Supported Repository | Required Connector |
|----------------------|--------------------|
| ATG Dynamo | ODBCFetch |
| Intershop | ODBCFetch |
| Sybase | ODBCFetch |
| Documentum | DocumentumFetch |
| Exchange Server | ExchangeFetch |
| FileNet | FileNetFetch |
| Lotus Quickplace | NotesFetch |
| Oracle 9i | OracleFetch |
| OpenText LiveLink | OpenTextFetch |
| PCDocs-DocsOpen | PCDocsFetch |
| Microsoft SharePoint | SharePointFetch |
| Siebel 2000 | SiebelFetch |
| Vignette | OracleFetch |
| Lotus Notes | NotesFetch |

[Table 4](#) lists the required supported Internet services repositories and connectors.

Table 4 Internet Services Repositories and Connectors

| Supported Repository | Required Connector |
|-----------------------|--------------------|
| Edgar Online | EdgarOnlineFetch |
| FTP Servers | FTPFetch |
| HTTP Servers | HTTPFetch |
| MoreOver Web Newsfeed | MoreOverFetch |
| NNTP News Servers | NNTPFetch |
| NewsEdge Newsfeed | FTPFetch |
| SiteMinder Netegrity | HTTPFetch |

[Table 5](#) lists the required supported e-mail server repositories and connectors.

Table 5 E-Mail Server Repositories and Connectors

| Supported Repository | Required Connector |
|----------------------|--------------------|
| Microsoft Exchange | ExchangeFetch |
| POP3 Servers | POP3Fetch |
| Lotus Notes | NotesFetch |

[Table 6](#) lists the required supported repositories and connectors for other repositories and databases.

Table 6 Other Supported Repositories and Connectors

| Supported Repository | Required Connector |
|----------------------------------|--------------------|
| Other Repositories and Databases | OmniFetch |

Using Supported Operating Systems and Platforms

[JOSH: Is it Ok to refer to this IDOL doc? I don't see it in the list of 3rd party docs]

To determine the operating systems that Autonomy supports, see the *IDOL Server Version 5* documentation (the IDOL server version 4.x is the version that ships, but version 5 is available to all customers).

Supported platforms for Autonomy include the following:

- Microsoft Windows NT4, 2000, XP, and 2003
- LINUX (all versions) kernel 2.2, 2.4, and 2.6
- Sun Solaris for SPARC versions 5 through 9
- Sun Solaris for Intel version 9
- AIX version 4.3, 5, and 5.1
- HP-UX for PA-RISC version 10, 11, and 11i
- HP-UX for Itanium version 11i
- Tru64 version 5.1

Note: If you are installing the IDOL server on Solaris, you need the `libiconv` library file which you can download from <http://www.gnu.org/software/libiconv/>. The IDOL server also supports other POSIX UNIX versions on request.

Searching the Oracle Database on UNIX

The BEA WebLogic Portal software comes with a license for ODBCFetch, which you can use to connect to a database over the ODBC protocol. ODBC is traditionally utilized on Windows environments, and Autonomy recommends OracleFetch with Oracle DBMS because it is pre-configured to work with Oracle.

OracleFetch is available at an additional cost. ODBCFetch also works with the Oracle database after you configure it. See the *Autonomy ODBCFetch Administrator's Guide* at <http://e-docs.bea.com/wlp/docs81/pdf/ODBC2.2.pdf> for more information.

Using the Autonomy Search Engine

After you configure the Autonomy search engine, you can search data in the BEA Virtual Content Repository, and perform other searches.

Configuring the Autonomy Search Engine

For instructions on configuring the Autonomy search engine, consult the documentation located at <http://e-docs.bea.com/wlp/docs81/index.html>, in the [Included Third-Party Technologies](#) section. In addition to the Service Dashboard portlet, you can modify the actual configuration files to control the behavior of the search engine itself.

Exposing Data from the Virtual Content Repository to Search

[JOSH: Should I remove this section and references to TRIP?]

You can write a series of dynamic pages to publish content information from the Virtual Content Repository to a set of stand-alone web pages that can then be indexed by Autonomy's HTTPFetch. Contact BEA Technical Support for assistance.

Using Other Search Portlets

The following enhanced Autonomy Portlets can provide added functionality:

- **The `setProfile` portlet** – Allows you to pass a valid option to the Autonomy Search engine. You can search using Boolean logic, provide sort order to your search, choose which Autonomy database to search, and choose case sensitive searches.
- **The `autonomySearch` portlet** – Allows you to place your content anywhere in your file system. This content represents the actual content that Autonomy will index. The portlet currently supports a limited number of file types: PDF, JPG, MS Word, MS Excel, MS PowerPoint, Simple ASCII, and HTML.

Localizing Autonomy Portlets

Over 60 languages are supported for reading text provided by a Fetch. See a complete list at <http://www.autonomy.com/content/Products/Connectors/Multilingual.html>. The actual portlets can be localized to different environments, but support for English is provided when you purchase the software.

Choosing a Search API

It is recommended that you use the Autonomy's Search API.

For more information on the Autonomy Search API, see <http://e-docs.bea.com/wlp/docs81/javadoc/autonomy/index.html>.

For more information on the BEA Search API, see <http://e-docs.bea.com/wlp/docs81/javadoc/index.html>.

Upgrading and Getting Support

You can upgrade to newer versions of Autonomy and get technical support.

Upgrading Autonomy

BEA WebLogic Portal 8.1 customers can file a trouble ticket with BEA Support to request an updated version of an Autonomy product or a version for a different operating system. Support specialists log into Autonomy's Automater system and download the most recent version for you.

Contacting Support

The Autonomy documentation is located at <http://e-docs.bea.com/wlp/docs81/index.html>, under **Included Third-Party Technologies**.

BEA provides front-line support for Autonomy components—contact the BEA Support Department. BEA Support will contact Autonomy for additional back-line support as needed.