



FAST DATA SEARCH

SEARCH PORTLET FOR IBM WEBSPHERE PORTAL SERVER 5.0

version 3.2.1

MODULE GUIDE

Copyright

Copyright © 1997-2004 by Fast Search & Transfer, Inc. and its associated companies and licensors. All rights reserved. Fast Search & Transfer may hereinafter be referred to as FAST.

Information in this document is subject to change without notice. The software described in this document is furnished under a license agreement. The software may be used only in accordance with the terms of the agreements. No part of this document may be reproduced, stored in a retrieval system, or transmitted in any form or any means, electronic or mechanical, including photocopying and recording, for any purpose other than the purchaser's use, without the written permission of FAST.

Trademarks

FAST is a registered trademark of Fast Search & Transfer. All rights reserved.

FAST Search, and FAST Data Search are trademarks of Fast Search & Transfer. All rights reserved.

Sun, Sun Microsystems, all SPARC trademarks, Java and Solaris are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States and other countries. All rights reserved.

Netscape is a registered trademark of Netscape Communications Corporation in the United States and other countries.

Microsoft, Windows, Visual Basic, and Internet Explorer are registered trademarks of Microsoft Corporation.

Red Hat is a registered trademark of Red Hat, Inc. All rights reserved.

Linux is a registered trademark of Linus Torvalds. All rights reserved.

UNIX is a registered trademark of The Open Group. All rights reserved.

AIX is a registered trademark of International Business Machines Corporation. All rights reserved.

HP and the names of HP products referenced herein are either trademarks and/or service marks or registered trademarks and/or service marks of HP and/or its subsidiaries.

All other trademarks and copyrights referred to are the property of their respective owners.

Restricted Rights Legend

Software and accompanying documentation are provided to the U.S. government in a transaction subject to the Federal Acquisition Regulations with Restricted Rights. Use, duplication, or disclosure of the software by the government is subject to restrictions as set forth in FAR 52.227-19 Commercial Computer Software-Restricted Rights (June 1987).

Contents

| | | |
|------------------|--|-----------|
| | FAST Support..... | v |
| | About this Guide..... | vii |
| Chapter 1 | Product Overview | 1 |
| | Introduction..... | 1 |
| | Supported Features..... | 1 |
| | Supported Versions..... | 1 |
| Chapter 2 | Installation | 3 |
| | Installing the Portlet..... | 3 |
| | Contents of the WAR File..... | 4 |
| | Adding the Portlet to a Web Application..... | 4 |
| Chapter 3 | Configuration | 7 |
| | Modifying the portlet properties..... | 7 |
| | Customizing the portlet..... | 9 |
| Chapter 4 | Operation and Verification | 11 |
| | Simple search..... | 12 |
| | Advanced search..... | 12 |
| | Example search..... | 14 |

FAST Support

Website

Please visit us at:

<http://www.fastsearch.com/>

Contacting FAST

Fast Search & Transfer Inc.

Wellesley Gateway

93 Worcester Street

Wellesley, MA 02481-9181 USA

Tel: +1 (781) 304-2400 (8:30am - 5:30pm EST)

Fax: +1 (781) 304-2410

Technical Support and Licensing Procedures

E-mail: fds-support@fastsearch.com

Product Training

E-mail: fastuniversity@fastsearch.com

Sales

E-mail: sales@fastsearch.com

About this Guide

Purpose of this Guide

This guide describes how to install and use the Search Portlet for IBM WebSphere. It includes the following chapters:

- Chapter 1 *Product Overview* provides a brief introduction to the product and its features.
- Chapter 2 *Installation* provides steps for installing the software and using the portlet in the WebSphere environment.
- Chapter 3 *Configuration* describes the portlet properties and discusses briefly how to .
- Chapter 4 *Operation and Verification* describes the available search features of the portlet and provides an example of a search.

Audience

This guide provides information for web developers who need to use the Search Portlet for IBM WebSphere.

Conventions

This guide uses the following textual conventions:

- Terminal output, contents of plaintext ASCII files will be represented using the following format:
Answer yes to place the node in the known_hosts file.
- Terminal input from operators will be in the same but bold format:
chmod 755 \$HOME
- Input of some logic meaning, will be enclosed in <> brackets:
setup_<OS>.tar.gz

where <os> represents a specific operating system that must be entered.

- URL's, directory paths, commands and the names of files, tags, and fields in paragraphs appear in the following format:

The default home directory is the *C:\DataSearch* directory.

- User Interface page/window texts, buttons, and lists appear in the following format:
Click **Next** and the **License Agreement** screen is displayed.
- *\$FASTSEARCH* (UNIX) or *%FASTSEARCH%* (Windows) refer to an environment variable set to the directory where FAST Data Search is installed.

Product Overview

Introduction

The Search Portlet for IBM WebSphere Portal Services 5.0 allows a WebSphere web application to query and process search results from FAST Data Search. You can use the portlet as provided or modify the user interface to comply with your company preferences and policies.

Supported Features

- Simple installation for use in WebSphere portal administration console.
- More than 20 configurable parameters to customize the portlet.
- Access to most search features in FAST Data Search.
- Java Server Pages (JSPs) that can be customized for the application.

Supported Versions

- IBM WebSphere Portal Server 5.0
- FAST Data Search 3.2

Chapter 2

Installation

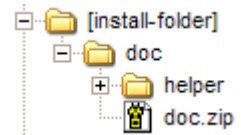
Installing the Portlet

The Search Portlet for IBM WebSphere Portal Services 5.0 is distributed in a zip file named **FDSIBMPortlet.zip**. To install the portlet:

- 1 Unzip the file to a location accessible from your administration console, in this example **[install-folder]**.

The install folder then contains:

- a Java Web application container named **FASTDataSearch.war** as described in *Contents of the WAR File* on page 4.
- a folder named **doc** containing:



WebSphere_Portlet_V321A.pdf This guide

helper Folder containing Javadoc for helper classes. Open a Web browser to **index.html**.

doc.zip Archive file containing full Javadoc for the FAST Data Search Java API

- 2 (Optional) To use the Fast Search API Javadoc, unzip the file **doc/doc.zip** into **[install-folder]**. The **doc** folder then contains:

dsapi Folder containing Javadoc for the FAST Data Search Java API. Open a Web browser to **index.html**.

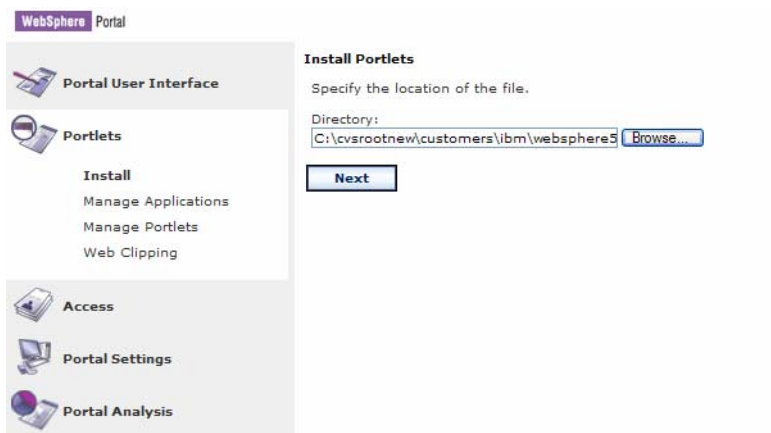
Contents of the WAR File

The contents of the **FASTDataSearch.war** file are as follows:

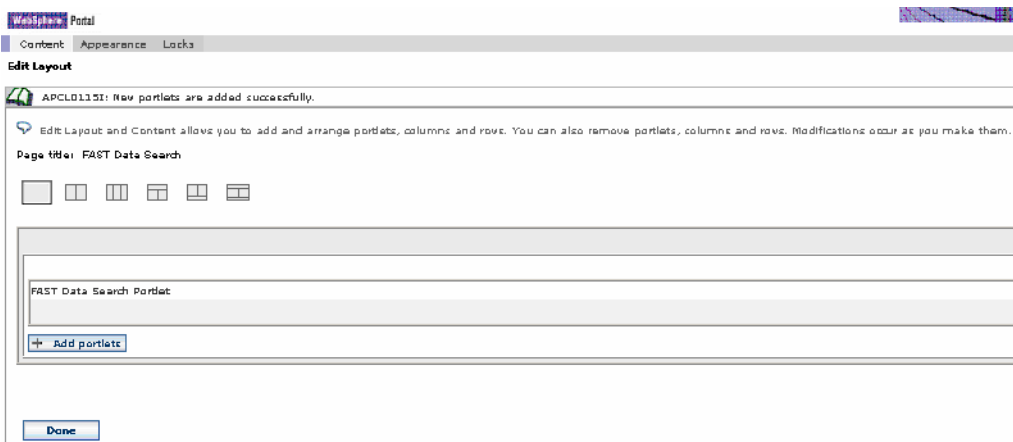
| Folder/File | Description |
|------------------------------|---|
| WEB-INF/portlet.xml | Portlet configuration file |
| WEB-INF/web.xml | Web application configuration file |
| WEB-INF/classes | Contains main portlet class |
| WEB-INF/lib/xmlrpc-1.1.jar | Library used for communicating with a Data Search for config info |
| WEB-INF/lib/dsapi.jar | Main API for performing a search and traversing the results |
| WEB-INF/lib/fast_portlet.jar | Helper classes for the search page |
| jsp/search.jsp | Main JSP of the portlet |
| jsp/inc/*.jsp | File that displays top concepts of the retrieved documents |
| include/style.css | Style sheet used by the search JSP |
| images/*.gif | Various images for the GUI |
| js/*.gif | Various images for tree-based navigation of categories |
| js/*.js | Javascripts used with the search page |

Adding the Portlet to a Web Application

- 1 **Copy the WAR** file to a location accessible from the portal administration console.
- 2 In the administration GUI, choose **Portlets > Install**
- 3 Enter the **path to the WAR** file, as shown below.



- 4 Click **Next**.
- 5 Click **Install** to install the portlet.
- 6 Set the **DataSearch host name** and **port number** configuration parameters as described in Chapter 3 *Configuration*.
- 7 Select **Portal User Interface > Manage Pages** to place the new portlet on a page. You can add a new page or select an existing page.
- 8 After having selected a page, choose **Edit page layout** (the pencil icon) for that page. The figure below shows a page named FAST Data Search that has been added to the portal MyPortal, and the new portlet added to that page.



- 9 Click **Done** and go to the page where you placed the portlet. If the portlet preferences are set with default values, the portlet is displayed on the screen as follows:

The screenshot shows a web browser window displaying the FAST Data Search Query Toolkit. The browser's address bar shows "WebSphere Portal". The page has a navigation menu with items: Documents, Welcome, My Work, My Finances, FAST Data Search (highlighted), My Newsroom, and My Favorites. Below the navigation is a sub-header "FAST Data Search Portlet". The main content area features the "fast Query Toolkit" logo. There is a search input field, a "Submit Query" button, and two dropdown menus labeled "Sort By:" (with "Desc" selected) and "Collection:". Below these is an "Advanced" button.

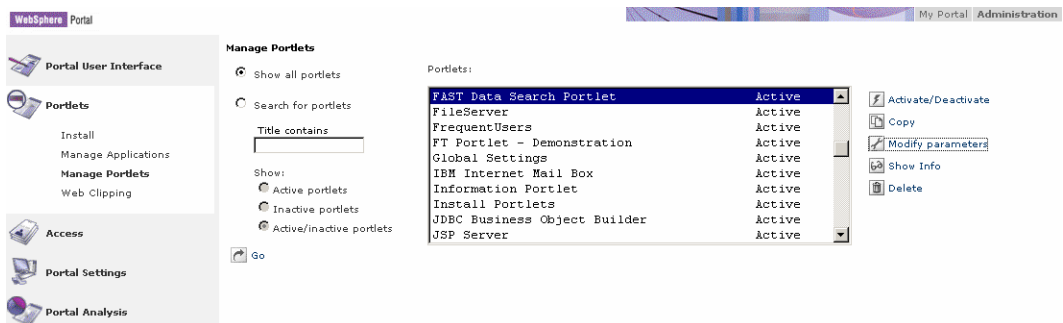
Chapter 3

Configuration

Modifying the portlet properties

In order to perform searches with the portlet, you need access to a FAST Data Search installation. If you want to change this or other configurable properties of the portlet:

- 1 Open the **WebSphere Administration interface**.
- 2 Choose **Portlets > Manage Portlets**.



- 3 Select the **FAST Data Search Portlet**.
- 4 Choose **Modify Parameters**.

Manage Portlets

Portlet Name : FAST Data Search Portlet

Edit Parameters :

| Parameter | Value |
|---|------------------------------|
| <input type="checkbox"/> DefaultCollection | <input type="text"/> |
| <input type="checkbox"/> ShowLeafs | false |
| <input type="checkbox"/> DataFields | title,body,url,modified,size |
| <input type="checkbox"/> ConceptOrderByName | false |
| <input type="checkbox"/> ConceptDocCount | 20 |
| <input type="checkbox"/> ShowAdvanced | true |
| <input type="checkbox"/> AllowSorting | true |
| <input type="checkbox"/> Cluster | webcluster |
| <input type="checkbox"/> SortFields | ,modified,size |
| <input type="checkbox"/> Logo | fast_qt.gif |
| <input type="checkbox"/> ShowConcepts | true |
| <input type="checkbox"/> DataSearchPort | 26005 |
| <input type="checkbox"/> ShowCategories | true |
| <input type="checkbox"/> ConceptMinTerms | 2 |
| <input type="checkbox"/> ShowDrillDowns | true |

The following is a list of preferences that can be modified for the search portlet in order to modify its behavior. .

| Preference | Description | Default value |
|----------------|---|----------------------------------|
| DataSearchHost | The fully qualified host name of the FAST Data Search installation. | localhost |
| DataSearchPort | The port number for accessing the Config server. | 16005 |
| Logo | The name of an image file to use as logo. Must be located in the image directory. | fast_qt.gif |
| DataFields | A list of fields to be displayed with each search result. | title, body, url, modified, size |
| MetaFields | A list of meta fields to display with each search result. | rank, collection |
| ShowCategories | If true, categories of the result set are displayed. | true |
| CategoryField | The field to search for categories. | urls |
| AllowSorting | If true, the user can specify sorting criteria with a search. | true |

| Preference | Description | Default value |
|------------------------|--|-----------------|
| SortFields | A list of fields that is sortable. Comma first makes sure no field is selected by default. | ,modified, size |
| ShowAdvanced | If true, the advanced search options are available. | true |
| Cluster | The search engine cluster to search. | webcluster |
| AllowSelectCollection | If true, the user can select which collection to search in. | true |
| DefaultCollection | The name of the collection to search in | "" |
| ShowLeafs | If true, also show nodes of the category tree that only has a single document. | true |
| ShowConcepts | If true, show top concepts of the search results. | true |
| ConceptMinTerms | The minimum number of terms in identified concepts. | 2 |
| ConceptDocCount | The maximum number of search results to use for identifying concepts. | 20 |
| ConceptMaxLength | The maximum character length of concepts. | 100 |
| ConceptOrderByName | If true, order the concepts alphabetically, rather than score. | false |
| ShowDrillDown | If true, drill down navigators are displayed. | true |
| AutoResubmitQuery | If true, resubmits query when there are 0 hits and available query transformations. | false |
| ShowActualNavDocCounts | If true, shows the number of documents for each drill down link. | false |

Customizing the portlet

In addition to the portlet preferences, you can modify the portlet by changing the source JSP to obtain a different look and feel, or behavior. The JSPs and their content are as follows:

`search.jsp`

The main JSP that includes the search form and includes other JSPs to be displayed for the various parts of the search results.

| | |
|----------------------------------|---|
| <code>searchAdvanced.jsp</code> | The input fields of the search form |
| <code>categories2.jsp</code> | For displaying results from supervised clustering/categorization. |
| <code>clusters.jsp</code> | For displaying top concepts from unsupervised clustering. |
| <code>debug.jsp</code> | For displaying debug information. |
| <code>footer.jsp</code> | Footer of the portlet. |
| <code>navigation.jsp</code> | For displaying drill down links. |
| <code>queryTransforms.jsp</code> | For displaying query transformations, e.g., spell checks. |
| <code>results.jsp</code> | Main results display. |
| <code>resultsdata.jsp</code> | Document summaries for each result document. |

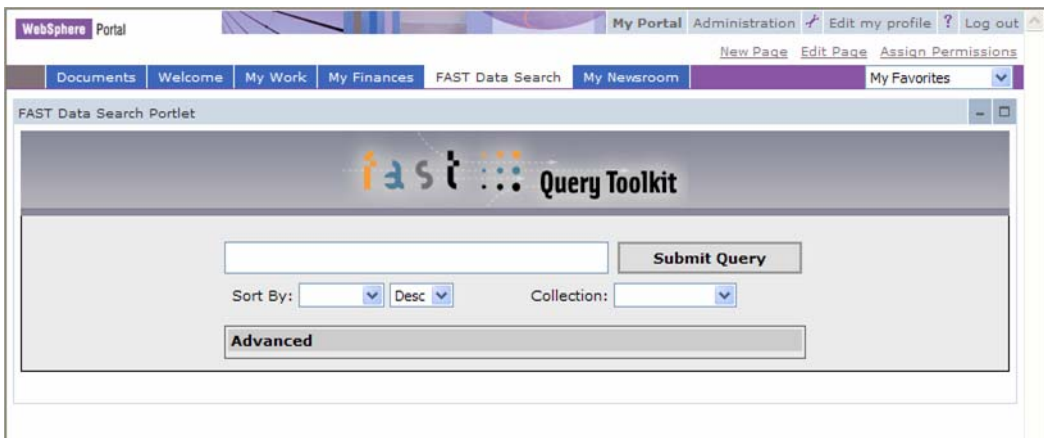
Javadoc is provided for the following:

| | |
|---|---|
| <code>[install-folder]\doc\helper\index.html</code> | helper classes for the portlet |
| <code>[install-folder]\doc\dsapi\index.html</code> | the full search API of FAST Data Search |

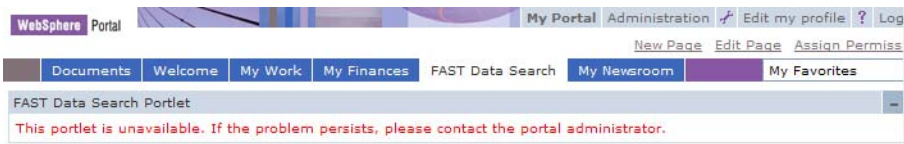
Chapter 4

Operation and Verification

If the portlet preferences are set with default values, the portlet is displayed on the screen as follows:



Otherwise, you see this screen:



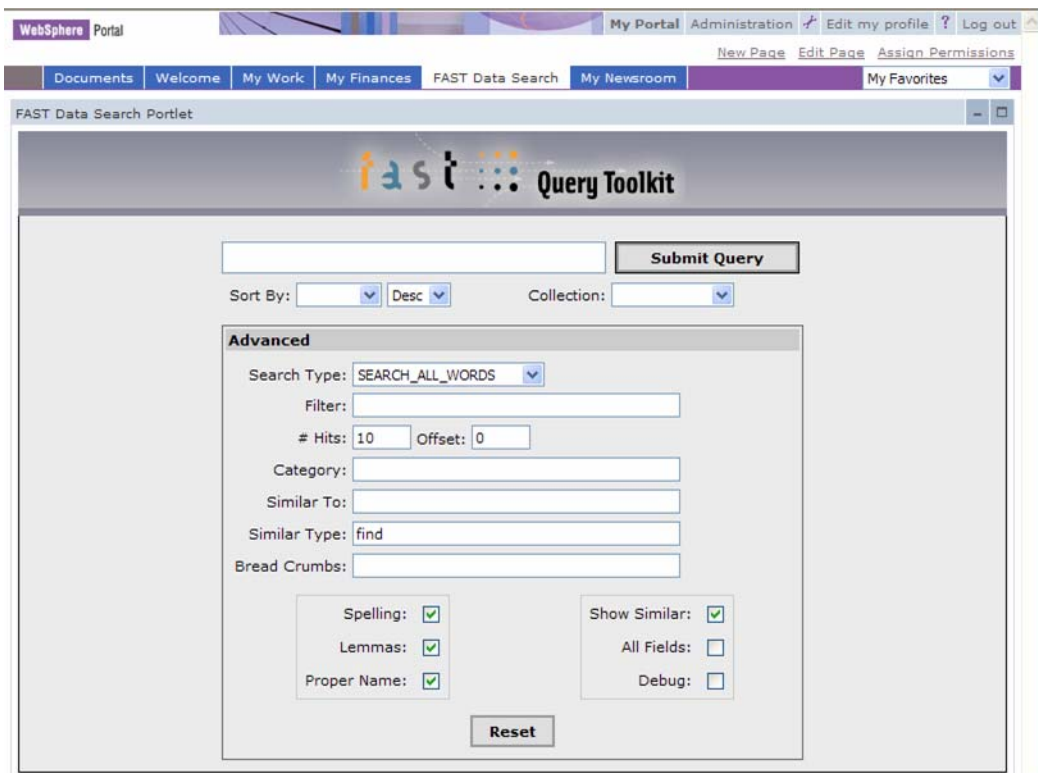
Simple search

To perform a simple search, do the following:

- 1 Enter one or more search terms in the main input field, replacing the '*'.
- 2 If desired, select a Sort by criterion to sort results by values of certain fields. Also choose descending (Desc) or ascending sort order.
- 3 If desired, select a particular Collection in which to search. By default, all available collections are searched.
- 4 Click Submit Query.

Advanced search

If you want more control of the search, click on the Advanced link to access additional search features. The image below shows the search portlet with Advanced features visible.



You can use the following features to use to modify your query:

Table 4-1

| | |
|------------------------------------|--|
| Search type | Decide whether you want to perform an AND'ed, OR'ed, phrased, or advanced search query. The latter allows you to use logical operators like AND, OR, ANDNOT and parenthesis with your search. |
| Filter | You can add filter terms to your query. These will limit the search results as ordinary query terms, but will not be subject to linguistic analysis such as spell checking. You can add '+' or '-' in front of filter terms to specify whether or not the terms will be included in the retrieved documents. |
| Hit and offset | You can specify how many hits you want returned, and the offset into the results set. |
| Category | A category that the retrieved documents should belong to. |
| Similar to and Similar type | A document vector that the retrieved documents should be similar to. Consult the FAST Data Search Integration Guide for details on this feature. 'Similar type' may be either 'find', 'exclude', or 'refine'. |
| Spelling | If checked, a spell check is performed on the search terms and spelling corrections are suggested. |

Table 4-1

| | |
|---------------------|---|
| Lemmas | Perform search in the lemmatized version of a content field if checked. |
| Proper name | Proper name recognition is performed on search terms if checked. |
| Show Similar | Provide 'Show similar' search links with each search result if checked. |
| All fields | Display all fields with each search result if checked. |
| Debug | Display debug information if checked. |

Example search

Searching with the query term 'basketball' gives the results as displayed below.

The screenshot displays a search interface with the following components:

- Categories:**
 - Taxonomy:** A list of domains including shopping.yahoo.com, uk.news.yahoo.com, beta.news.yahoo.com, www.cnn.com, ca.dir.yahoo.com, us.cnn.com, us.news.yahoo.com, search.news.yahoo.com, www.aol.com, and cf.us.biz.yahoo.com.
 - Navigators:** Language set to 'en', character set dropdown, and document size dropdown.
 - Clusters:** A list of top concepts such as 'buyer protection (4.99)', 'protection review (4.99)', 'top stories (4.93)', 'full coverage (3.66)', and 'team sports (2.19)'.
- Main Results Area:**
 - Header:** 113 Results found in 0.512s (1-10 >>)
 - Result 1:** Rank: 3181, collection: endrez. Title: Basketball - Team Sports - Sports and Outdoors - Compare Prices, Read Reviews at Yahoo! Shopping. Body: ...bikes, ski, fishing, bowling Team Sports 6508 'Basketball' from 368 stores Show me: 1 product per store...Finder: Sport: Product: Brand: Price: \$ to \$ Basketball-NCAA Portable Basketball Hoop NCAA Portable Basketball Hoop Get a game... Modified: Tue Jan 06 13:50:22 CET 2004. Size: 72604. Similar: Find Refine Exclude.
 - Result 2:** Rank: 3178, collection: endrez. Title: Basketball - Fan Gear and Memorabilia - Sports and Outdoors - Compare Prices, Read Reviews at Yahoo! Shopping. Body: ...bowling Fan Gear and Memorabilia 3916 'Basketball' from 224 stores Show me: 1 product...Sport: Product: Brand: Price: \$ to \$ Basketball-NBA Replica Jersey NBA Replica Jersey...Buyer Protection Review this merchant Basketball:Reebok USA Basketball Customized Replica... Modified: Tue Jan 06 16:38:57 CET 2004. Size: 80085. Similar: Find Refine Exclude.
 - Result 3:** Rank: 2375, collection: endrez. Title: Basketball Game Bombing Kills 10. Body: ...Science Oddly Enough Sport Reuters | AFP | Sky News | Photos Sunday January 4, 08:21 PM Basketball Game Bombing Kills 10 A bomb explosion at a basketball game in the southern Philippines has killed 10 people and wounded at least 40 others, including... Modified: Tue Jan 06 07:09:27 CET 2004. Size: 15253. Similar: Find Refine Exclude.
 - Result 4:** Rank: 1228, collection: endrez. Title: NBA - Action Figures - Toys by Category - Toys and Baby - Compare Prices, Read Reviews at Yahoo! Shopping. Body: ...Protection Review this merchant more results from this store (5) compare NBA:Swish the Basketball Player Swish the Basketball Player Swish the Basketball Player's jumps shots and talented swishes make him king of the court. Bendable action... Modified: Tue Jan 06 16:56:35 CET 2004. Size: 55560. Similar: Find Refine Exclude.

There are four main areas of the results page:

Document summaries

These show the information retrieved about the top hits. Each result comes with a set of meta fields and a set of data fields, as previously explained. Meta fields are to the left, and data fields to the right. In this example, the meta fields are 'rank' and 'collection'. The rank is a computed relevancy and is determined from factors such as where the search terms were found in the document, or how many times they occurred. Below the data fields are three links that can be clicked to find similar/dissimilar documents to the one in focus.

Categories

It is possible to browse a tree of categories that were found for the top N search results. By clicking on a link in the tree, a new search is performed, limiting the results to documents belonging to that part of the category tree.

Navigators

These are drop down boxes that offer drill downs in the results set. For instance, there is one drop down box for 'Language'. The box contains languages found for the top N hits (N is configurable) with the percentage of each language, and when the user selects a language, a new search is performed with this an additional search criterion. This way, the user can restrict the results based on analysis of a gradually smaller result set, until the desired document(s) turn up.

Top concepts

This is based on the distinguishing terms found from analyzing the top N hits. The most distinguishing concepts are listed first. By clicking on the links, a new search is performed with the phrased terms as an additional search term.

