Oracle Ultra Search

Architecture
Version 9.2.0 for Oracle9i Database Release 2
Version 9.0.2 for Oracle9i Application Server
February 2002

ORACLE®

Oracle Ultra Search

The proliferation of information has caused chaos inside firewalls, and the resulting difficulty in locating information is causing inefficiencies, and expense.

EXECUTIVE SUMMARY

Oracle Ultra Search is an out-of-the-box search solution that provides search across multiple repositories - Oracle databases, IMAP mail servers, HTML documents served up by a Web server, files on disk and many more. Ultra Search enables a 'Portal' search across the content assets of a corporation, bringing to bear Oracle's core capabilities of scalability and reliability.

This paper gives IT managers and architects an understanding of Ultra Search - its architecture, its main features, its interfaces and the configurations available for its deployment.

1. INTRODUCTION

In the age of the Internet, proliferation of information is causing a new information management crisis for enterprises. Using the World Wide Web, workers become their own information retrieval experts. But searching for the right answers can be more than frustrating:

- Studies predict that by 2006 the amount of information flowing over corporate Intranets will be 200 times what is was in 1998.

- This information is ultimately stored in corporate databases, Web pages, files in various popular document formats and in email or groupware systems. The information that businesses produce, store and use for decision making is scattered across billions of documents and data fragments that reside on many different, and often incompatible, IT servers and systems. Servers are located throughout the country and across the globe.

- Corporate information is distributed across enterprises in both structured and unstructured form - structured relational databases, unstructured Word-processing documents, spreadsheets, presentations.

- As applications demand transactional consistency, coordinated multi-user access, administration and maintenance for content, a natural gradient is created to move more and more information into databases. However, even when multiple databases are involved, searching across databases needs a robust solution.

According to some estimates, the lost time in searching costs companies billions of dollars in lost productivity each year. Bad search can also drive customers to

competitors Web sites. A company can have great products and a terrific looking Web site, but if customers and employees can't find the information they are looking for, they have essentially wasted time and money spent on development and promotion.

Oracle Ultra Search solves the problem of finding relevant information across your company's many disparate repositories of information. Ultra Search is an out-of-the-box application built on Oracle 9*i* and its proven Text technology that provides uniform search-and-locate capabilities over multiple repositories – Oracle databases, other ODBC compliant databases, IMAP mail servers, HTML documents served up by a Web server, files on disk and many more. Ultra Search uses a 'crawler' to index documents; the documents stay in their own repositories, and the crawled information is used to build an index that stays within your firewall in a designated Oracle9*i* database. Ultra Search thus enables a 'Portal' search across the content assets of a corporation without the need for rearchitecting IT topologies, compromising security, or programming against hard-to-use API's.

2. ULTRA SEARCH FOR ENTERPRISE INFORMATION AND CONTENT

Oracle Ultra Search is an out-of-the-box search solution that:

1. Searches text across multiple repositories – Multiple databases, HTML Web pages, Files, IMAP mail servers - and organizes and categorizes the content.

2. Provides the best relevance ranking and globalization support in the industry.

3. Provides value added Portal functionality – crawling, fielded search and metadata extraction.

4. Presents a Web-style interface where users can specify, for example, to indicate "Oracle +Location -France" to indicate they want to retrieve all documents, database records or email containing the terms Oracle and Location, but not the term France.

Enterprises can benefit by using Ultra Search in many different types of applications:

1. **Portal Search** – Ultra Search offers the most powerful search for Enterprise Portals developed with the Oracle Enterprise Portal Framework. Oracle9i Application Server (Oracle 9*i*AS) Portal customers can use Ultra Search through a 'Portlet' (a portlet is a contained area of Portal page that can be rendered in HTML or any other browser-capable technology). The Ultra Search Portlet provides crawling and universal search over all Ultra Search-supported repositories, including the ability to search the contents of Oracle 9iAS Portal.

For organizations who want to build their own portal from scratch, Ultra Search provides a canned, end-user-oriented, web-style search over various corporate databases, HTML pages, IMAP email servers, or filesystem documents. You can either use the 'default' user-interfaces as supplied, or 'embed' Ultra Search in your portal, customizing the look-and-feel to your requirements. Ultra-search also allows you to customize metadata according to the different repositories, and search according to different metadata elements from different repositories.

2. **Web Search for Oracle Text** – Ultra Search is an application built on Oracle Text, Oracle's industry leading text retrieval engine. It provides Oracle Text customers with Web-style searching capabilities without the need for any low-level SQL programming. A significant amount of expertise has gone into translating and tuning web-style queries into underlying SQL-based Oracle Text queries. Ultra Search helps Oracle Text users start that much ahead, for example database applications needing a simple Text search component will find Ultra Search admirably integrated with the Oracle infrastructure.

3. **Library or Archive Search** -- Many organizations with digital libraries, information warehouses or centralized repositories are seeking to convert custom search applications over such repositories to more general, web-based ones. A Library search differs from a Portal search in that the latter seeks a simple search over many dynamically changing sources, whereas the former needs more advanced search over a fewer number of relatively well-defined sources. Ultra Search provides such lower-level API and linguistic access to meet the needs of advanced knowledge workers.

4. **Content Management Search** –. Media organizations creating or publishing content in a collaborative manner need to search across content (Web pages, documents) as it moves through multiple repositories in different stages of the content-management life cycle: from the desktop file of the author to the staged version in a database. Use Ultra Search to build a better search and retrieval system for your documents by integrating Ultra Search with your company's own collaborative content management or document management process. Ultra Search provides both full-text and fielded text retrieval to create a set of indexes tuned for keeping track of your content.

Search can be improved if it can be narrowed down what part of a 'document' a piece of information occurs in - the title, the body, the name of the author and so on. For example, search results for 'London' differ when you look for an author name, versus a title. Generally, different repositories have different such 'metadata' attributes that may be attractive for searching against - databases identify columns and email servers know header/body/attachment.

A flexible metadata mapping methodology that lets you unify diverse repositories in common logical terms for search purposes is one of the big value-adds of Ultra Search. In order to display a uniform set of results ranked by overall relevance,

Ultra Search allows customers to normalize or map the various metadata attributes from various repositories.

The screen shot below shows an example of querying over multiple repositories: A corporate email archive and a database (labeled 'Server Technologies'). The query is narrowed by metadata fields that have been defined to map against both repositories:

- 'Title' maps against the subject line in emails or against a fielded text column in the database.

- 'Author' maps against the sender of an email.

Figure 1: Oracle Ultra Search Query Example.

The next section takes a look at Ultra Search architecture, followed by detailed looks at the important aspects of its functionality.

3. ULTRA SEARCH ARCHITECTURE

Ultra Search builds on the Oracle platform, using existing and proven public interfaces.

Ultra Search integrates proven technologies in Oracle9*i* in a simple yet robust architecture. Ultra Search is entirely an Oracle Text based application, built using the same public interfaces that are available to users of Oracle Text, but enhanced with considerable expertise in aggregating information for indexing, translating queries for the best quality search, and optimizing operations for scalability.

Oracle Text, in its turn, builds on the Oracle 9*i* infrastructure using public interfaces such as SQL and the Oracle Extensibility Architecture ODCI interfaces. Few search engines can search databases effectively, handicapping them for dynamic data. Oracle Text is highly integrated with the Oracle database for best interoperability with dynamic data. One key strength of Ultra Search is it ability to serve search for database-backed web-sites, applications, archives, or content-bases located in a single database or spread across multiple databases.

Ultra Search is a client program to the Oracle server at run time. It can be deployed in two configurations – in the server tier or the mid-tier.
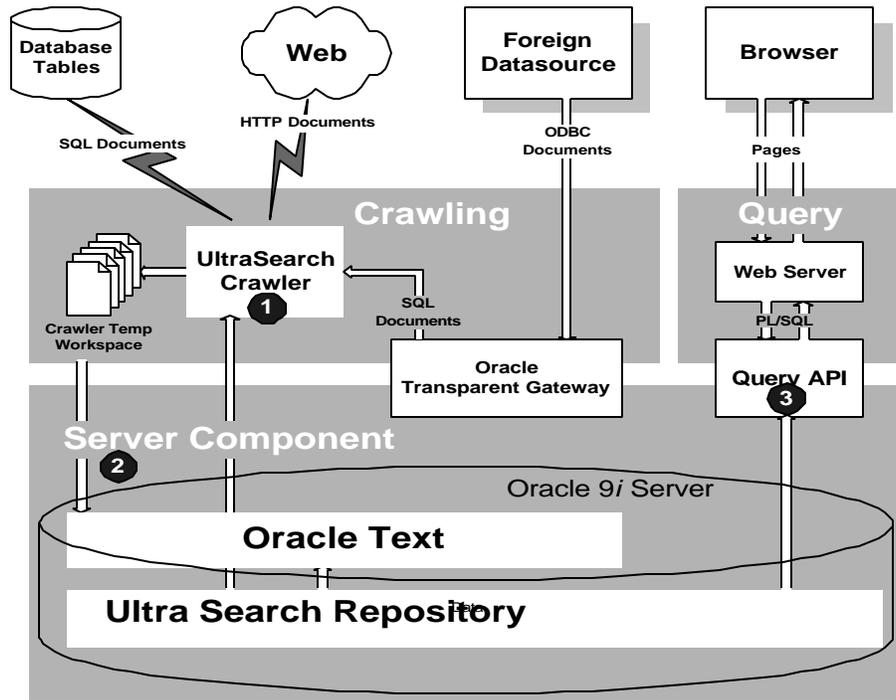

Figure 2: Ultra Search Architecture

Ultra Search is made up of five distinct components:

1. **Ultra Search Crawler** -- The Ultra Search Crawler is a Java process activated by your Oracle server according to a set schedule. When activated, the Crawler spawns a configurable number of processor threads that fetch documents from various data sources and index them using Oracle Text. This index can then be used for querying. The crawler maps link relationships and analyzes them to avoid going in circles and taking wrong turns. The Crawler schedule is integrated with and driven from the DBMS Job queue mechanism. Whenever the Crawler encounters embedded, non-HTML documents during the crawling it uses the Oracle Text filters to automatically detect the document type and to filter and index the document. See section 3 for more details on supported document types and the filtering process.

2. **Ultra Search Server Component** -- The Ultra Search Server Component consists of a Ultra Search repository, and Oracle Text. Oracle Text provides the text indexing and search capabilities required to index and query data retrieved from your data sources such web sites, files, and database tables. This component is not visible to users ; it operates as a "black box" that indexes information from the Crawler and serves up the query results.

3. **Query API & Query Application** -- Ultra Search provides a Java API for querying indexed data. The API returns data with and without HTML markup. The HTML markup can help you build the following search engine web interfaces: Basic Search Form, Advanced Search Form, Query Result Display, Help Page, Feedback Page, Register URL. The Java APIs use JDBC connection pooling for scalability.

   Ultra Search includes a highly functional query application for users to query and display search results. The query application is based on Java Server Pages (JSP) and will work with any JSP1.0 compliant engine.

4. **Ultra Search Administration Tool and Interface** – The Administration Tool is a Java Server Page (JSP) web application you use to configure and schedule the Ultra Search crawler. The administration tool is typically installed on the same machine as your Web server. You can access the Administration Tool from any browser within your intranet. The Administration Tool is independent from the Ultra Search Query Application. Therefore, the Administration Tool and Query Application can be hosted on different machines to enhance security and scalability.

5. **JAVA email API** -- Ultra Search provides Java APIs for accessing archived emails. These APIs are used by the Ultra Search Query Application to display emails. These APIs may also be used when building your own custom query application.

The Ultra Search default query interface and the administration tool run in any HTML browser client. The administration tool relies on certain Java classes in the mid-tier. This logical 'mid-tier' can be the same physical machine as the one that runs the database server, or a different one, running Oracle9*i* AS. Finally the Ultra Search database server component consists of the Ultra Search data dictionary that stores metadata on all the different repositories, as well as the schedules and Java classes needed to drive the crawler. The crawler itself can run either on the database server machine or remotely on another machine.

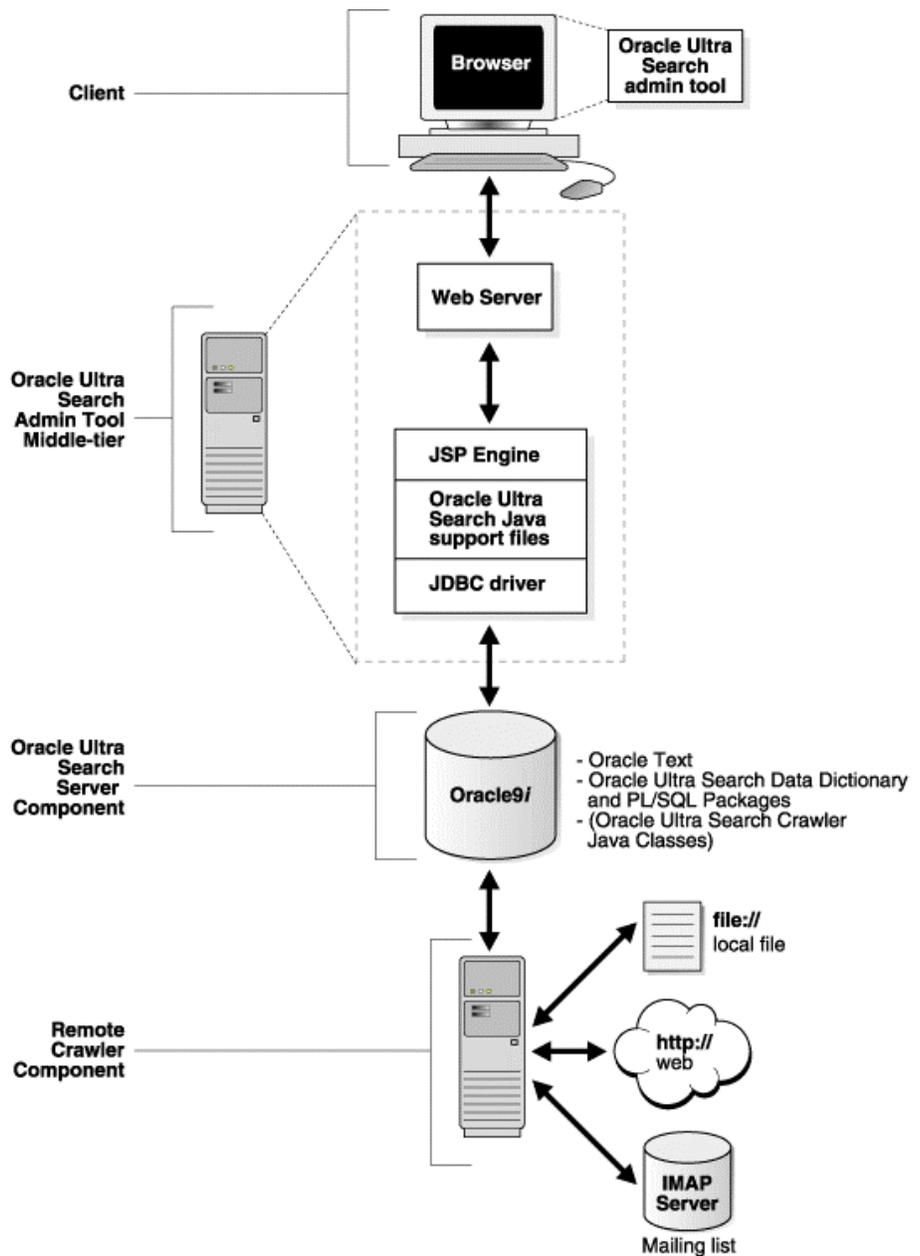The distribution of Ultra Search components is shown in Figure 3.

Figure 3: Overview of Ultra Search Components

3.1.1 The Ultra Search Crawler Component

The Oracle Ultra Search crawler is a multi-threaded Java application responsible for gathering documents from the data sources you specify during configuration. The crawler stores the documents in a local file system cache as a temporary workspace during its crawl. Processing the cached data, Oracle Ultra Search creates the index required for querying.

To crawl different repositories, the Ultra Search crawler allows you to define specific 'data sources' (A data source is a logical construct identifying a repository) You can take a single physical repository, such as a database, and map it to multiple data sources (A data source is also the granularity at which you define metadata). Ultra Search knows the following types of data sources:

- **Web Sites** – Define web sites as a data source with the HTTP protocol.

- **Database Tables** – Ultra Search can crawl Oracle databases and other relational databases that support the ODBC standard. Database tables to be crawled can reside in Ultra Search's own database instance or they can be part of a remote, database accessed over a network. To access remote databases, Ultra Search uses 'database links'. Ultra Search allows the crawling of both full text columns and "fielded text" columns. Fielded text columns allow you to map a database column to an Ultra Search attribute (e.g. AUTHOR, TITLE), creating a set of indexes tuned to the content of your database.

- **Files** – Files can be crawled through the file:// protocol. Files must be accessible by each crawler machine either locally or remote over the network. Ultra Search uses the Oracle Text filters to extract text and metadata from documents and automatically identifies document types. See section 3.2.2 for a description of INSO filters and a n overview of supported file types.

- **Emails** --. This feature is useful for crawling mailing lists. Emails sent to a specific email address can be crawled by creating an IMAP email account that subscribes to a mailing list(s). All messages addressed to the email address / mailing list are indexed. Ultra Search can crawl and open email attachments and 'nested' emails such as in email threads.

To maintain fresh, comprehensive search results, Ultra Search uses *synchronization schedules.* Ultra Search lets you gather from multiple Web sites and data sources, each on a separate schedule. Email search results, for example, can be updated continuously, while published content is gathered on a less frequent schedule. Each synchronization schedule can have one or more data sources attached to it.

To increase crawling performance, you can set up the Ultra Search crawler to run on one or more machines separate from your database. These machines are called 'remote crawlers'. However, each machine must share cache, log, and mail archive directories with the database machine.

To limit the crawling to a specific section of your corporate network or to ensure that crawling does not take wrong turns and follow link relationships that point outside your intranet, Ultra Search lets you specify so-called 'inclusion' and 'exclusion' domains for crawls.

Ultra Search supports '*instance snapshots'* where you create a read-only snapshot of a master Ultra Search instance for query processing or backup purposes. Instance snapshots can be made updatable. This is useful when the master instance is corrupted and you want to use a snapshot as a new master instance.

The Ultra Search crawler can be instructed to collect URLs without indexing them. This *'data harvesting mode'* allows you to examine document URLs and their status, remove unwanted documents, and start indexing.

3.1.2 Extensible Crawler API

The Ultra Search crawler supports the gathering of documents located on the Web, in database table, email servers and file systems. Many users, however, would like to make their own proprietary document repositories searchable. Ultra Search provides an API that can be used to adopt the Ultra Search crawler to any proprietary document repository. Document repositories can include proprietary in-house created systems or third-party software from companies such as Lotus or Documentum. The only condition is that the repository that you would like Ultra Search to adopt needs to offer HTML protocol access to its documents.

The extensible crawler mechanism works as follows:

- Register a new data source with the administrative environment.

- Use the Ultra Search Extensible crawler API to implement your own 'Crawler Agent' in JAVA (a crawler agent is a user-developed program that enables the Ultra Search crawler to access proprietary document repositories). Write your crawling agent so that it streams out documents one-by-one to the Ultra Search crawler, providing the URL and attribute values of every document in your repository. If your repository is password protected, your agent needs to authenticate Ultra Search to access its documents.

- Create a maintenance schedule for the new data source.

The extensible crawler API is also used to make the Ultra Search crawler aware of changes to the documents in your repository -- it calls your agent with a timestamp to instruct it to gather only those URLs that have been updated/inserted recently.

See Figure 4 below to learn how the extensible crawler API works:

1. The Ultra Search crawler calls the user-written crawling agent as soon as the agent has been loaded (labeled "1" in the illustration) and passes important parameters about the new document repository.

2. The agent can now establish a connection to the repository ("2").

3. The crawler obtains a document URL from the crawling agent.

4. The crawler uses the obtained document URL to obtain the corresponding document from the repository.

5. Repeat steps 2 and 3 until all documents have been gathered from the repository.
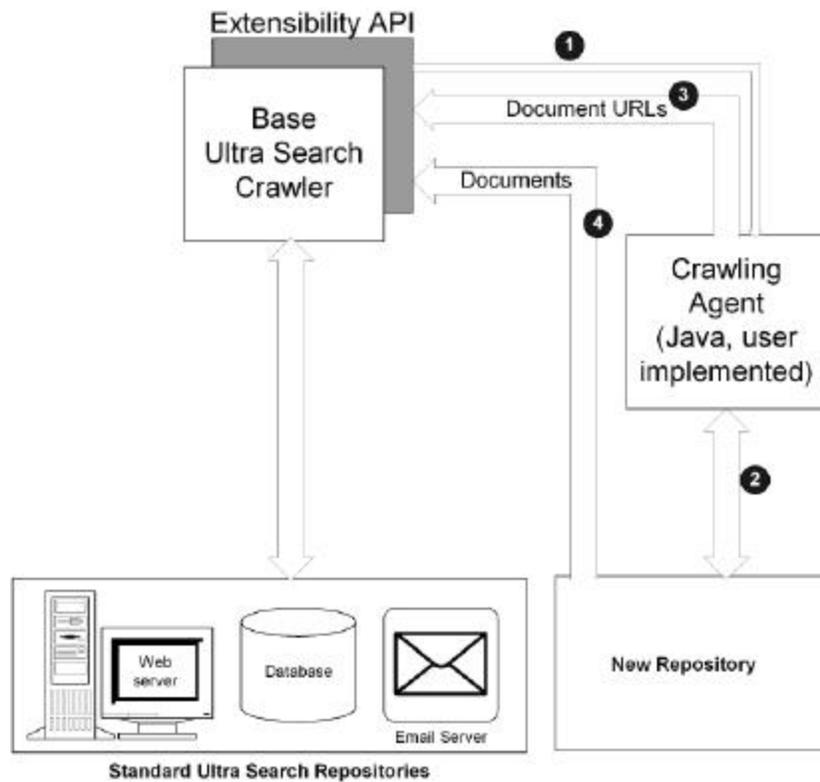
Figure 4: Illustration of Ultra Search extensible crawler API.

3.1.3 The Ultra Search Query API and Query Application

Oracle Ultra Search provides a flexible, easy-to-integrate query framework by means of a set of query APIs. These APIs can be used from Web applications to retrieve and display query results. Ultra Search API's are written in Java. Therefore, they are compatible with a large spectrum of web application servers that support Java Server Pages (JSP version 1.0 and above).

Ultra Search Query APIs enable the inserting Ultra Search query input boxes and result lists in any Web application. In addition, these APIs can be used to customize search screens. To include Ultra Search into Java applications, Ultra Search can be called from Java Server Page (JSP) code.
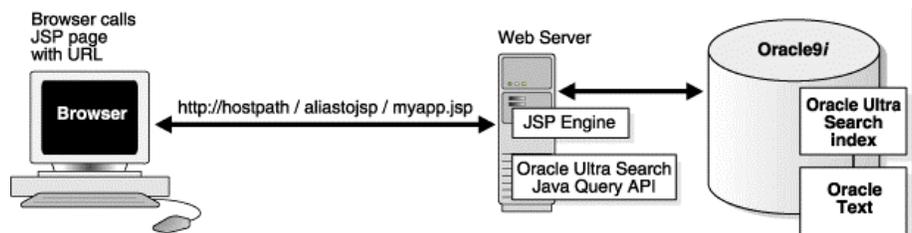

Figure 4: Illustration of Ultra Search Query API

In this illustration, the browser calls a Java server page with the URL (http://hostpath/…) on the Web server. The Web server, which also contains the JSP Engine and the Ultra Search Java Query API, communicates with Oracle9i, the Ultra Search index and Oracle Text.

Ultra Search includes a canned, fully functional query application for users to query and display search results. The query application comes in Java (JSP) and also incorporates an email browser for reading and browsing emails.

The Oracle Ultra Search query API is geared towards easy integration with your applications. It provides functionality for:

1. Customizing the look and feel of your search per your taste - Ultra Search's Query API provides functions for the presentation of HTML code that can be embedded in your Web application.

2. Retrieving data from the Ultra Search instance, including available data sources, languages and metadata attributes (e.g. AUTHOR, TITLE).

To shorten development cycles, it also includes functionality for encapsulating commonly performed Web development tasks.

3.1.4 The Ultra Search Administration Component

The Ultra Search Administration Tool is a web application that allows for:

- **Define Ultra Search instances** – Each Ultra Search instance is identified by name and has its own crawling schedules and index. As many instances as necessary can be created.

- **Manage administrative users** – Ultra Search users can be assigned to manage an instance. Language preferences can be defined for each user.

- **Define crawler parameters** -- Configure and schedule the Ultra Search Crawler

- **Set query options** -- Query options allow users to limit their searches. Searches can be limited to document attributes (e.g. TITLE, AUTHOR) and data groups. Data source groups are logical entities exposed to the search engine user. When entering a query, the search engine user is asked to select one or more data groups to search from. Each data group consists of one or more data sources.

- **Adjust relevancy ranking of the search hit list** -- Ultra Search allows administrators to influence the order that documents are ranked in the search hit list. Use this to promote important documents to higher scores and make them easier to find. This function is new with Ultra Search version 9.0.2.

3.1.5 The Java email API

Oracle Ultra Search enables the retrieving and indexing of emails residing on a server that supports the IMAP4 protocol. Ultra Search defines the concept of an

email source, which derives its content from emails sent to a specific email address. When the Ultra Search crawler crawls an email source, it collects all emails that have the specific email address in any of the "To:" or "Cc:" email header fields.

In portal or content search scenarios, the items of interest are usually found in corporate-wide mailing lists and aliases. For example, all customer support issues for Oracle Corp. may be mailed to support@Oracle.com. In Oracle Ultra Search, you create multiple mail sources, where each mail source represents a public email list to which all searchers are assumed to have access.

Ultra Search email crawling and rendering is built on top of the JavaMail API. This enables Ultra Search to provide a Java API for accessing indexed emails. This API enables the retrieving of information such as:

- email header information

- email body content; and

- attachments of an email.

The Ultra Search email API allows for including browsing functionality into Java Server Pages (JSP) or servlet-based web applications. Ultra Search ships a fully functional JSP web application that directly uses this API to render indexed emails. Since the source code is viewable, it provides an example for building your own customized email browser.

3.1.6 The Display URL Mechanism

This feature allows you to render the results of your database table searches according to an underlying Web application. Use this feature to enhance the presentation of search hits if you have a database with an associated Web application that is used to browse records in the database tables. It allows the Ultra Search administrator to define a 'display URL' that will be sent to your browser and trigger your Web application to bring up the Web screen that is associated with the database record in your search hits.

For example, in your technical service department you might have a service incident database with an associated Web application that is used by service personnel to browse and alter the service records in the database. You would like to use Ultra Search to enable your customer personnel to search across all service records, but keep the familiar Web frontend of your application. The display URL makes this possible. It works like this: Use Ultra Search to create a table data source for the service database. Then define a so-called 'display URL template' (a URL template is a set of user-defined rules that Ultra Search uses to convert the result of database record searches into a URL that your browser can understand). Now when the table data record is in the search hits, Ultra Search will use the display URL template to generate a URL that will point your browser to the respective screen in your Web application. Your service personnel can now see the complete service incident Web screen.

In order to be able to use the display URL feature, your Web application must be written so that there is 'fixed' URL that links a screen in the application with its corresponding records in the underlying database. For example,

```
http://service.mycompany.com/myWebapp.?no=<incident_number>
```

is a fixed URL for accessing the table records from your Web application to get the incident Web screen given an 'incident number'. Ultra Search stores this URL and uses it to bring up the Web screen of your application that corresponds to the serach hit. Please note that URL parameter values cannot be encrypted.

3.2 Ultra Search Methodology

What steps do you need to follow for using Ultra Search ? The Oracle Ultra Search engine follows four logical steps to provide universal search – *gather, analyze, make queryable, and maintain*. These steps are not novel, and are indeed found in most organizations' business process.

Error! No topic specified.

Figure 5: Ultra Search Methodology

3.2.1 The Gather Step

Gathering refers to information that exists in structured relational databases and in unstructured files, Word processing documents, spreadsheets, presentations, e-mail, news feeds, Adobe Acrobat files, and Web pages. Ultra Search gathers this information by "crawling" your corporate intranet and looking through all the information that exists in the various repositories of your company – databases, Web pages, IMAP mail servers and others.

During the gathering process, link relationships are analyzed to avoid going in circles and taking wrong turns. As a result, Ultra Search administrators have an easier time keeping search results complete and up-to-date.
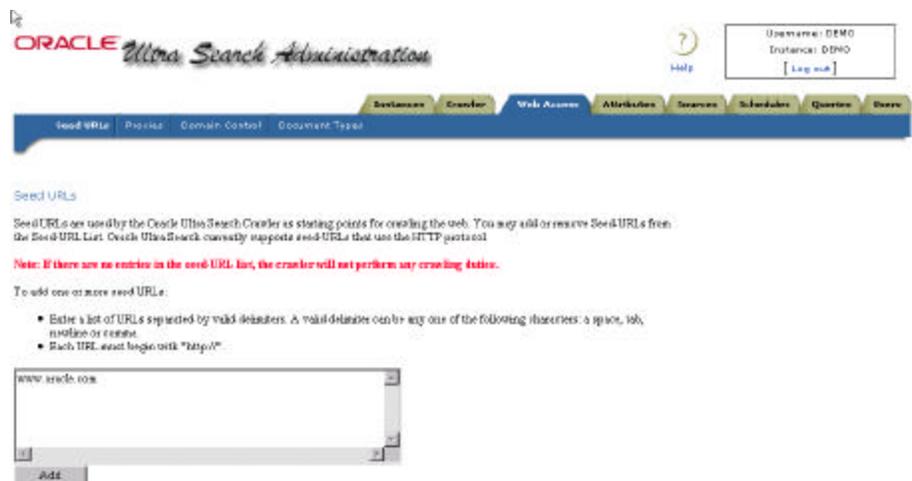

Figure 6: Configuration of Ultra Search Crawler.

The screen shot in the figure above shows the configuring of the Ultra Search crawler for information gathering through the administrator utility.

3.2.2 The Analyze Step

In the analyze phase Ultra Search looks at the meaning and structure of gathered information. In order for information to be searched, it must be indexed. During the analyze phase, Ultra Search uses the Oracle Text engine to extract both meaning and structure from the gathered information by creating an integrated index, effectively "normalizing" both structured and unstructured data. Oracle Text indexes contains a complete wordlist along with other information.

During indexing, text and metadata are extracted from documents by Oracle Text INSO filters. This filtering technology automatically identifies document type, invokes the correct filter and produces indexable text and data. Several predefined metadata fields are supported, including author, date, and title. INSO filters include filters for most (150+) popular file types including:

- Microsoft Office Suite 95/97/2000.

- Spreadsheet documents, such as Microsoft Excel and Lotus 1-2-3.

- Word processing files such as Microsoft Word and Corel Word Perfect, including a PDF filter to index Acrobat PDF files.

- Presentation graphics: Microsoft PowerPoint, Lotus Freehand.

Unlike some document management systems, Ultra Search gathering and analyzing is non-intrusive. Instead of physically moving documents, information and documents are analyzed but reside in their original location under their own name.

In typical Web search technologies, hundreds of hits are returned. As the number of repositories increase, the ability to rank relevance of documents decreases. Ultra Search uses the award winning relevance ranking of Oracle Text to ensure that users consistently find the needle in the haystack.

Figure 7: Configuring Ultra Search Document Filters.

The screenshot above shows how you specify the document types that will be analyzed by Ultra Search and filtered through Oracle Text document filters.

3.2.3 Making Crawling Results Searchable

"Make Searchable" is the function of providing access to all the information that has been indexed in a programmatic fashion. Oracle Ultra Search provides a JAVA API for this purpose. Passing a search term into the query API locates all relevant documents, whether they are stored on Web servers, databases, or in applications. Customers can use Ultra Search APIs to integrate universal search into their own Web pages or applications.

Figure 8 Screen Shot of Example Query Screen.

The Figure above shows an example Query Screen with a search for

"Performance"-related information (Ultra Search relevance rankings appear in red).

3.2.4 The Maintain Step

The maintain step ensures that search results are updated continuously. Ultra Search lets you gather from multiple Web sites and repositories, each on a different schedule. IMAP messaging servers, for example, can be updated continuously, while published content is gathered on a less frequent schedule. Ultra Search maintains content by providing easy, intuitive utilities that provide Administrators with an easy way to keep up with new content that is added through growth or acquisition.

Figure 9: Scheduling the Ultra Search Crawler.

Figure 9 above shows a screenshot of the "Schedule" page of the Ultra Search Administration Utility where maintenance crawling can be configured.

4. ULTRA SEARCH IN ORACLE9I APPLICATION SERVER REL. 9.0.2

Ultra Search has been integrated with Oracle's Internet Application Server and is now being released with both the Oracle9i database and with Oracle9i Application server. This section provides an overview of new Ultra Search features built specifically for the Oracle9i Application Server.

Oracle 9iAS release 9.0.2 includes an integrated version of Oracle Portal and Ultra Search. Ultra Search in 9iAS features:

- New Ultra Search portlets provide search screens wrapped as a 'Portlet' (a portlet is a contained area on a Portal page that provides access to an information source and that outputs HTML). This allows Portal users to utilize Ultra Search directly from their Portal pages.

Ultra Search is shipped as a sub-component of Oracle9I Portal. Add the new search Portlet to your Portal pages.

- A Portal data source allows Oracle Portal customers to go beyond Portals' restricted built-in search function. They can now use Ultra Search for searching outside their Portal installation – over the repositories of multiple Portal installations and in all other Ultra Search-supported repositories.

- Single Sign-On support allows you to log on once for all components of the iAS product suite and never see the Ultra Search administrative login screen a second time.

Please note that although Ultra Search in 9iAS is the same product as in the 9i database release 9.2, the new portlets and single sign-on support are only shipped with 9iAS and Oracle Portal.

4.1 New Search Portlets

Ultra Search provides new sample search screens wrapped as a portlet. Sample search portlets are shipped out-of-the-box with Oracle Portal. For power users, the source code of these sample search applications is written so that they can be easily read and modified. The appearance of the Portlets can be customized by the Portal end-user, for example the number of hits display in each search request can be set.

Ultra Search search portlets are implemented as JSP applications.

Note: Search portlets are not shipped with the DB version, they are only shipped with the iAS Version and Portal.

4.2 New Portal Data Source

Ultra Search can now crawl and index the Oracle Portal. A new data source type '9iAS Portal' allows for crawling one or more installations of Portal. The Ultra Search administrative tool will automatically find and display all page groups of your Portal installation – after registering your Portal, you just need to choose and pick the Page groups that you want to index. Ultra Search is now ready to crawl the page groups you have chosen and the various indexable objects contained in them:

- Pages

- Folders and subfolders

- Text Items

While gathering Text from these objects, the Ultra Search crawler also obtains various attributes including the name of the object, the creator, the date the object was created. This "metadata" can later be used to narrow your searches.

To ensure that your search results keep current with any changes in your Portal page groups, make sure that Ultra Search revisits your new portal source periodically by setting up a respective crawling schedule. Ultra Search operates with Portal in "pull mode"; information to be indexed is periodically polled from Portal by the Ultra Search crawler. The Ultra Search index will not automatically

be closely synchronized with the data in Oracle Portal without a frequent-enough "maintenance crawling" schedule.

Figure 10: Ultra Search 'Portlets'.

Figure 10 above depicts the new Ultra Search Portlet. This Portlet can be customized to show a number of Ultra Search features, including basic or advanced search and number of search hits shown.

4.3 Single Sign-On Authentication

With 9iAS version 9.0.2, Ultra Search delegates the responsibility for user authentication to the 9iAS single-sign-on server. When accessing the Ultra Search administration tool, you will be redirected to the SSO server and asked to authenticate yourself. Authenticated SSO users never see the Ultra Search login screen. Instead, you can immediately choose an instance to manage.

5.   SUMMARY

Companies need to eliminate the chaos inside their firewalls. No solution provider is more focused than Oracle to solve that problem.

In summary, Oracle Ultra Search allows you to reduce the time spent finding relevant documents on your company's IT systems:

- It crawls, indexes, and makes searchable your corporate intranet.

- Provides out-of-the-box, web-style search without the need for coding against hard-to-use low level API. For advanced users, however, APIs are also exposed.

- It organizes and categorizes content from multiple repositories by extracting valuable metadata that can be used in Portal applications.

- It provides effective search by returning more relevant hits - the best relevance ranking in the industry - and finds what you want.

And it provides the best database integration in the industry.

.

ORACLE

Oracle Ultra Search
Architecture White Paper
February 2002
Author: Stefan Buchta
Contributors: Sandeepan Banerjee, Stacy Bruzec

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200
www.oracle.com

Oracle Corporation provides the software
that powers the internet.