

Understanding Your Options for Searching Oracle WebCenter

KIMBERLY NEGAARD
FISHBOWL SOLUTIONS, INC.



Fishbowl Solutions Notice

The information contained in this document represents the current view of Fishbowl Solutions, Inc. on the issues discussed as of the date of publication. Because Fishbowl Solutions must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Fishbowl Solutions, and Fishbowl Solutions cannot guarantee the accuracy of any information presented after the date of publication.

This Whitepaper is for informational purposes only. FISHBOWL SOLUTIONS MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS DOCUMENT.

Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright, no part of this document may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Fishbowl Solutions Inc. Fishbowl Solutions Inc. may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Fishbowl Solutions, the furnishing of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property.

© 2016 Fishbowl Solutions Corporation. All rights reserved.

Fishbowl Solutions is a registered trademarks or trademarks of Fishbowl Solutions Inc. in the United States and/or other countries. The names of actual companies and products mentioned herein may be the trademarks of their respective owners.

© 2016 Fishbowl Solutions Corporation. All rights reserved.

Contents

1	Introduction – Executive Overview	1
2	Background	2
3	Identifying Your Search Requirements	2
3.1	Understanding the Search Process	2
3.2	Criteria to Consider When Evaluating Search	8
4	Available Search Technologies	9
4.1	Metadata-Only Search	9
4.2	Database Full-Text Search	10
4.3	Oracle Text Search	11
4.4	Oracle Secure Enterprise Search	13
4.5	Google Search Appliance	13
5	Selecting a Search Technology	15
5.1	Side-by-Side Search Technology Summary	16
6	Notes	17

1 Introduction – Executive Overview

Search is a critical part of any effective content management solution. Without it, documents, web pages, policies, and other enterprise resources cannot be easily surfaced to end users. This paper explores the search technologies available to Oracle WebCenter customers including metadata-only and full-text database search, Oracle Text search, and Secure Enterprise Search, as well as the search functionality available with the Google Search Appliance. You will get a side-by-side comparison of these search options covering the pros and cons of each technology and the use cases most suited to their capabilities. Whether you're using Oracle WebCenter to power your website, intranet, or document management system, this paper will help you ask the right questions and understand the differences between each search option in order to determine which one is best for you.

This paper is intended for information technology managers, system administrators, or others involved in selecting a search technology for Oracle WebCenter Content. It focuses on the impact various search features have on user's ability to find content and how the capabilities of each available search technology affects the search experience in Oracle WebCenter.

After reading this paper, the reader will be able to:

- Understand the search process and evaluate search requirements for an organization
- Identify the appropriate search technology for WebCenter based on use cases
- Describe the strengths and weaknesses of each search technology
- Differentiate between WebCenter search metadata-only, database full-text, Oracle Text search, Secure Enterprise Search, and the Google Search Appliance


2 Background

Poor search is a top complaint for Oracle WebCenter Customers. Over the years Stellent and Oracle have offered a number of options for searching the Content Server each with their own set of benefits and drawbacks.

The 11g version of Content Server allows the use of metadata-only search, database full-text search, Oracle Text search or Oracle Secure Enterprise Search (SES). Oracle customers also have the option to integrate their WebCenter Content system with other third-party search products such as Google's enterprise search appliance. As various search options come and go search continues to be an issue for many customers. By understanding your content, users, and search options, you can make the best choice to get the most out of your WebCenter investment.

3 Identifying Your Search Requirements

The initial step in choosing the right technology to search Oracle WebCenter Content is to determine what makes for a good search experience in your organization. The first part of the paper focuses on understanding the search process and search features that can impact your search experience. Each organization is different; features critical to one use case may offer little value to another. Understanding your search requirements will help you prioritize the most important features for your implementation, select the right technology, and empower your users to find the content they need to do their jobs.

About this section: Throughout the following sections you will see bolded questions noted with a checkmark  icon. These questions are intended to help you identify how the search capabilities discussed could impact search for your users and improve the findability of content stored in Oracle WebCenter.

3.1 Understanding the Search Process

We all use search every day. When search works well it often goes unnoticed, but when it doesn't we get frustrated, grumble, and often give up trying to find what we were searching for. And while effective search on the public web often happens so naturally we don't give it much thought, there are a lot of contributing factors that make the difference between finding the information you need and giving up. It's helpful to start the search selection process by taking a step back to look at the pieces at play.

Searching consists of three basic parts: The Query, The Ranking, and The Display. Someone asks a question (the query), the search engine determines what items match that query (the ranking), and the results are displayed back to the user through some type of interface (the display). The various search tools available to Oracle WebCenter customers offer different features to enhance or improve each part of the search process.



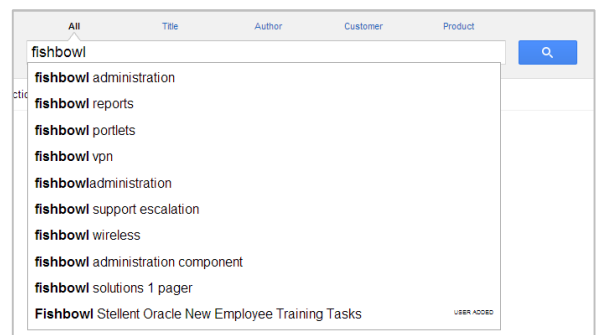
THE QUERY

The first thing that happens when performing a search is that the user enters a query. This query might take the form of a few words typed into a search box such as the WebCenter Content Quick Search, or it might be constructed using a metadata search form like the advanced search page, a profile search page, or a custom interface.

When a search performs poorly it is often caused by a combination of the technology and the user's inability to ask the "right" question (i.e.: input the right query). Search features that help add context to queries make it easier for the user to ask the right question. Query interpretation technology reduces the amount of information we expect a user to know about what they are looking for. The more easily, accurately, and thoroughly we enable users to ask their questions, the more likely they are to find the content they are seeking.

Search Suggestions

Search begins at the point where the user starts typing their query. One of the most common query interpretation capabilities is the ability to provide search suggestions, also referred to as autocomplete or type-ahead search. This feature usually works by aggregating the most common queries from past searches and suggesting them to users by popularity. The average user-query is 1.7 words in length which means a user is only providing one or two words while expecting to find the item they want among collections which often contain millions of documents. By using search suggestions it becomes easier to extract more context from the user. This technique has the added benefit of leveraging the collective knowledge that exists within the enterprise to make search easier for everyone.

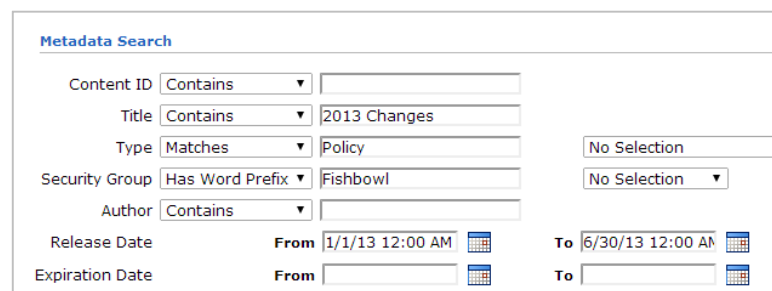


- ☑ **Where might search suggestions be most valuable to users?**
- ☑ **How important is this capability in my WebCenter search deployment?**

Search Syntax and Operators

The search syntax and operators available with each search tool also impact the way users present a query. While search suggestions can be extremely valuable for unstructured keyword searching, the availability of search operators can be equally important especially when your use case requires a more structured approach.

When selecting a search tool think about the types of operators your users will need. If you are searching a website or intranet, users will probably only perform keyword-based searches; however, if you are searching structured, well-tagged content items, the ability to build your search based on date range, custom metadata, and full-text might be required. Each search tool has different capabilities regarding the query syntax and search operators available.



The screenshot shows a 'Metadata Search' form with the following fields and values:

Field	Operator	Value	Additional Info
Content ID	Contains		
Title	Contains	2013 Changes	
Type	Matches	Policy	No Selection
Security Group	Has Word Prefix	Fishbowl	No Selection
Author	Contains		
Release Date	From	1/1/13 12:00 AM	To: 6/30/13 12:00 AM
Expiration Date	From		To:

- Will users prefer to search using a keyword search box, an advanced search form, or both?**
- What query operators should be available to users when performing searches?**
- Do users need to search on partial word matches (substring or wildcard searches)?**

Thesauri, Dictionaries, and Query Expansion

Between the point at which the user enters their query and the search engine computes relevant results, there is one last piece in the search process where search technology can play a role in improving the query. This is often referred to as query expansion. Query expansion describes various ways in which the words the user types can be expanded upon or “understood” by the search engine in order to more accurately represent the original intent and locate the right content.

One mechanism of query expansion is word stemming, which is the process of breaking a word down into its stem or root form and searching across documents using all forms of that word. Stemming allows users to search for “orders” and get results containing “order” “ordered” and “ordering.” It means users don’t have to know whether a word was in past tense, present tense, plural, or singular, in order to find the right content.

- Is the ability to support word stemming important for your search deployment?**
- What languages do you need to support?**

Another way queries can be expanded for better search is through the use of synonyms. Synonyms can be used to set words equal to one-another. Leveraging synonyms means that if

different groups refer to the same topic by multiple names, users don't have to know what variation of the term to use, or do multiple searches in order to find relevant documents. Synonyms can also be used to make abbreviations equal to their full meanings, or set legacy terminology as synonymous with updated names (e.g. UCM = WebCenter Content). Lastly, synonyms can be used to make commonly misspelled words synonyms for the proper spelling.

- ✓ **Does your organization use a lot of custom abbreviations?**
- ✓ **Do various groups or departments have different names for the same concept, process, or facility?**
- ✓ **How often do users misspell words preventing them from finding content?**

The Ranking

After a user submits a query, the search engine must compute which items match the query and in what order to display them. The ranking abilities of the various search tools for Oracle WebCenter Content vary greatly. Typically when results are displayed to users, they are ordered by either relevancy ranking or metadata sorting.

Relevancy Ranking

The relevancy ranking capabilities of a search tool dictate which items appear first in the results based on the query terms provided by the user. In order to identify which tool is right for you, it's important to think about some of the implications for how relevancy is determined, and what that means for your users and your content. Search engineers use the word "signals" to describe the various factors taken into account during the relevancy calculation process. Signals are all the things a relevancy algorithm accounts for when determining which item should be first, second, etc.

A common signal used by search engines is the number of times a word appears in a document's full-text. Documents with more occurrences of a keyword are deemed more relevant. Other signals involve looking at a document's structure where a search engine analyzes the content of HTML pages with an understanding that words the heading tag carry more weight the words in a paragraph tag. In this case, if a user is searching for documents about "Labradors" and two documents both contain the word "Labradors" but one has the word in the heading and another only in the body, the document with the word in the heading would be ranked higher.

As the amount of content in your repository grows it becomes more important to have accurate relevancy calculation capabilities because there is more "clutter" to sort through in order to find the most relevant documents.

- ✓ **What factors influence the relevancy of documents in your repository?**
- ✓ **How important is search relevancy for your implementation?**

Metadata Sorting

Metadata sorting is more straightforward than relevancy ranking. Once a search engine determines which items match the query it returns them in order based on a specified attribute.

The order is not relative to the query term provided. For example, product catalog sites often require the ability to sort results by a number of attributes such as price, model number, product name, rating, or popularity.

- ☑ **What, if any, metadata fields will users benefit from sorting on?**
- ☑ **In general, how should results should be ordered – by relevance or by metadata?**

The Display

The last piece in the search process is the display. Once the most relevant items have been ranked they must be displayed. The user experience depends in part on the capabilities of the search technology you choose, but just as importantly on the effort put into understanding users' needs in selecting or building the right search interface. The search tools for Oracle WebCenter each provide an out-of-the-box search interface. When discussing search for WebCenter-based applications such as websites or intranets the quality of the search experience becomes as much about the thought and effort put into implementing the front-end search experience, as it is about the back-end search tool in use. Take the time to ask questions like "What makes a good snippet of text to summarize a document?", "What metadata or attributes should you present on the page: a title, author, date or icon?", or "How might users want to filter results?" Good presentation is critical to a positive search experience and although partially independent of backend technology, the following search features can contribute to the search presentation.

Search Facets and Navigation

How much time do you have to give a user an accurate result? Many web design experts say it's about four seconds. If a webpage does not provide a user with something useful in four seconds they will leave the page, give up, go to a competitor's site, etc... and by the way, there is a good chance they will spell it wrong and are only willing to give you one or two words to describe what they are actually looking for. Sound familiar?

Effective search presentation and navigation can extend the time a user is willing to spend looking for what they want, or in many enterprise scenarios, what they need to do their jobs effectively. Every click you get another four seconds. Imagine "Joe" starts typing a query. You give Joe search suggestions which allow him to provide a more descriptive query. Then you provide navigation facets which allow Joe to filter his results by metadata. Even if he doesn't initially see what he wants he can use the facets to add more context to his search. Even if he doesn't see the right document on the first search, if he clicks a facet you get another four seconds. Joe feels like he's being listened to and searching becomes more like a human conversation. Effective navigation helps make that possible.

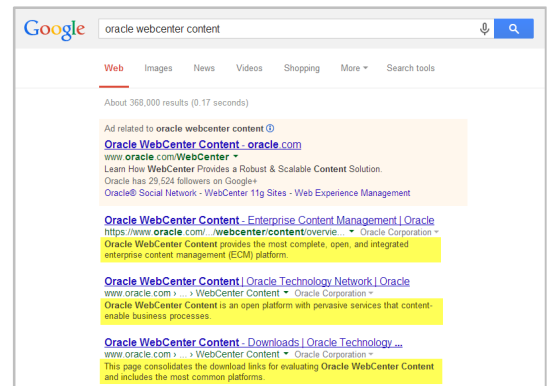
Think about the process a WebCenter user often goes through when doing a search. Many WebCenter users believe search doesn't work because they are asked to fill out lengthy search forms using metadata with which they are unfamiliar. When they don't get the right results they aren't even sure which field they need to use or which one is messing up their search. Contrast that with presenting metadata based navigation *after* performing a search. In this case a user is more inclined to feel like they are finding useful information and paring it down. Oracle Text, Secure Enterprise Search, and the Google Search Appliance provide this type of navigational capability.

- ✓ **Does your content and metadata lend itself to a faceted search approach?**
- ✓ **What content attributes would be most important to filter by?**
- ✓ **Would different user groups need to filter by different attributes?**

Snippets

Snippets are small samples of content from a document or webpage that give users an idea of what is covered in a particular item. Snippets are dynamically generated based on the keywords that were searched, so users will see context from the portion of the document containing their search terms.

For web and document content, snippet text can be very helpful for improving the search experience. This is especially true for website or intranet deployments where users may not know the right document by its title alone. For digital media or transitional data such as pictures, invoices or receipts, snippets can be of no benefit.



- ✓ **Should snippet content be available for your search results?**

experience. This is especially true for website or intranet deployments where users may not know the right document by its title alone. For digital media or transitional data such as pictures, invoices or receipts, snippets can be of no benefit.

Thumbnail and Previews

Many search implementations provide some type of preview or thumbnail to help users identify content and determine which items they need. For document management use cases, previews help users determine which one of several similar documents is best suited to their needs without having to open each one. In digital media use cases, thumbnails are critical to identifying content since no full-text or snippet exists. The Content Server supports thumbnail generation which is not specifically part of the search technology, but can be advantageous when creating a search interface regardless of which technology you choose.

- ✓ **Would thumbnails or previews enhance the search experience for your users?**
- ✓ **Which types on content would be most valuable to preview (images, documents, etc.)?**

Spelling Suggestions

Misspelled words can be addressed either during the query or within the display phase of search. In the case where the correct spelling is assumed and the query is executed as if the correct spelling were input, spelling suggestions can be thought of as part of synonyms – i.e. make an incorrect spelling synonymous with the correct spelling; however, spelling suggestions

can also be presented to the users after searching allowing them to decide if a suggestion is what they intended to search for.

Several years ago Fishbowl Solutions sold a component called Search Statistics which provided Stellent customers with data about the search terms in use in their organization including data about searches that returned zero results. We found that about half of the searches that returned no results were because the user had misspelled the query term. On the public web, Google estimates that between 10% and 20% of queries contain a misspelling.

- ✓ **How often do misspellings prevent users from finding content?**
- ✓ **What types of words are most commonly misspelled – common words, proper names, company terms?**
- ✓ **What languages would spell check need to support?**

3.2 Criteria to Consider When Evaluating Search

Now that we've taken a look at the high-level search process and various features that can be used to create a search experience, there are a few other factors to consider when determining your search requirements. The better you can identify the requirements of successful search for your use case or organization, the better you will be able to choose the right search technology, and design the optimum search experience.

Users

When evaluating a search solution it is critical to consider the types of users who interact with the system to find content. User expectations vary a lot depending on job function, technical experience, and familiarity with your content. For example, casual users of a public website will likely have a low level of familiarity with your content as well as a low level of technical ability.

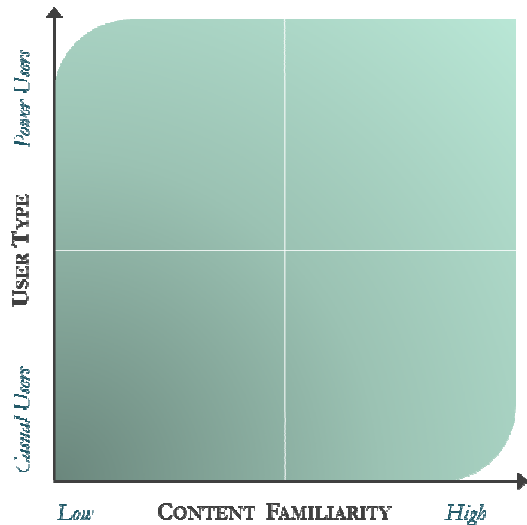
Users with high content familiarity often know what kinds of metadata are applied to documents or have particular content items in mind which they are seeking. These users often want an option to specify these criteria when searching.

If you have users who fit in two or more distinct groups (for example general content consumers and content managers/contributors) it may be effective to create two separate search interfaces tailored to the unique needs of each group.

Content

What type of content do you need to search - documents, images, web pages? Are there any specialty file types you need to index?

The amount of content in your repository impacts search performance and relevancy. A good relevancy algorithm becomes increasingly important as the amount of content grows. With larger collections there are more documents that contain keywords requiring added sophistication to determine which of those are actually the most relevant.



The use of metadata is one of the major differences between searching WebCenter Content and searching the public web. With many Oracle WebCenter implementations metadata is critical to the way content is organized. When designing a search experience good metadata can be a major asset to your users, but when designed incorrectly, metadata can become a barrier to effective search. If users are presented with lengthy search forms and don't know which fields contain which information, it can make finding content very difficult.

4 Available Search Technologies

Note: The search technologies discussed in this paper represent options currently available to Oracle WebCenter Content customers either by Oracle or supported by generally available third-party products. Custom search integrations with technologies such as Microsoft Fast or HP Autonomy are not discussed as no productized integrations with Oracle WebCenter Content are available on the market to date.

4.1 Metadata-Only Search

Metadata-only database search uses the Content Server system database for search and is the default configuration for new Content Server installations. Any Oracle-supported database can be used for metadata-only search. For repositories that do not require full-text indexing, metadata-only search is a much more efficient configuration. Since the full-text contents of the documents are not indexed the search functionality is less comprehensive than with full-text search, however, it is also less process intensive.

STRENGTHS

One of the biggest strengths of metadata-only search is performance because the system does not have to perform full-text indexing. This greatly reduces the load on the Content Server and decreases both indexing and serve time. It also offers better scalability as your repository size increases. Metadata-only database search is the most basic of the WebCenter Content search options and requires minimal administrative effort to install or configure.

Metadata-only search supports complex query structure allowing searching across a combination of fields using ranges and various operators such as contains, substring, matches, and starts-with. This can be useful for power users who are familiar with the content metadata model and prefer the ability to perform advanced searches.

WEAKNESSES

Selecting metadata-only search for your WebCenter Content environment has a number of weaknesses. The most notable are the lack of full-text context for queries and the inability to perform relevancy calculation. Searches executed using metadata-only search cannot be ranked by any type of relevancy or score. They can only be sorted by the indexed metadata values such as date, title, author, etc. The more content in your repository, the more problematic this issue can become—especially if your users do not have a high level of familiarity with your metadata structure.

Metadata-only search does not include any query enhancement features such as synonyms or stemming. It also doesn't include any out-of-the-box navigation capabilities to filter search results using metadata facets, although some customers have created custom search interfaces leveraging metadata for additional browsing and navigation purposes.

While the ability to support and create complex queries was noted as a strength above, the inability to easily perform simple keyword-based searches across all metadata fields can be a disadvantage for casual users who often do not know which fields are used with which values, for example, keywords vs. comments. vs. description.

PRIMARY USE CASES

Metadata-only search is best suited for searching content without full-text contents—for example digital media where the important search attributes might be title, keywords, image type, or category. Similarly, large collections of scanned documents (especially ones which have no searchable text) may be well-suited to metadata search, particularly when users always search by invoice number, customer ID, etc.

For websites, intranets, document management, or similar text-heavy use cases, metadata-only search is not recommended as the lack of full-text context and text-related search features will likely lead to poor findability, lower system adoption, and unhappy users.

4.2 Database Full-Text Search

Like metadata-only search, database full-text search uses the Content Server system database to index content. Database full-text search is available for use with any supported version of Oracle Database or Microsoft SQL server. It has many of the same capabilities as metadata-only search with the added ability to index the full-text content of documents in the repository.

STRENGTHS

The major strength of database full-text search over metadata-only is the addition of full-text indexing. For customers using Microsoft SQL Server database, this provides a full-text search option without needing to install any additional databases or set up additional providers.

Supported File Formats

Oracle: pdf, html, htm, xls, hcsp, text, txt, doc, rtf, ppt

SQL: text, txt, htm, html, doc, msword, ms-word, ms-powerpoint, ppt, ms-excel, xls

WEAKNESSES

The biggest drawback to using database full-text search compared to the other full-text search options is the inability to return results based on relevance or score. In fact, for some clients, this issue is magnified with full-text search because the scope of what's searched is more extensive than with metadata-only. The default sort order for database search is date, which generally puts an overemphasis on new content. If you search for "benefits" under this configuration, you will get the most recently modified items with the word "benefits" anywhere in the document even if many of them are only vaguely related to benefits. For example, a document containing the phrase "low-impact exercise has numerous health benefits..." would appear before a document titled, "Employee Benefits Overview" if it happened to be updated more recently. When searches contain common words this effect can be drastic enough to push the results a user actually needs several pages in, which in most cases means they will never click through that far to find it.

Another major weakness of database-only search is performance and scalability. Compared to Oracle Text search, Secure Enterprise Search, and the Google Search Appliance, database search is less scalable. Customers often encounter problems with searches taking multiple seconds to execute or even timing-out with content volumes exceeding 500,000 items. Oracle cautions full-text searching on repositories with more than 1 million content items.

PRIMARY USE CASES

Many clients who use Microsoft SQL Server as their system database select database full-text search for their WebCenter deployment since it does not require an additional Oracle Database installation as required by Oracle Text search and Secure Enterprise Search. The biggest issue with this solution is the inability to sort results by score.

For customers using Oracle Database 11g, Oracle Text search is recommended over standard database full-text search due to the ability to perform ranking and navigation, as well as improved performance.

4.3 Oracle Text Search

Oracle Text search is a feature of Oracle Database 11g. Oracle WebCenter customers with Oracle Database can use Oracle Text functionality to search the Content Server repository. Oracle Text search can be used with any Content Server databases supported in the Certification Matrix; however, if the system database is not Oracle Database 11g, an external provider for Oracle Text search must be configured. Oracle Text with either 10g or 11g allows for a search collection to be created separate from the Content Server database without having to migrate, upgrade or modify the database used for the Content Server.

STRENGTHS

A major strength of Oracle Text search is that unlike standard database search, it provides relevancy ranking using the built-in Score capability of the Oracle Database. In addition, it allows results to be sorted by any indexed metadata fields. Oracle Text within WebCenter Content can be configured to display up to four navigational facets at the top of the standard search results page enabling metadata-based filtering after a search is performed. Oracle Text also allows snippets to be displayed on the WebCenter Content results page giving users a better idea of the content of a document.

Oracle Text supports a thesaurus file which can be used to associate related terms allowing for query expansion through synonyms.

Oracle Text search supports full-text indexing for 150 file types, significantly more than standard database search which only supports ten. Oracle Text 11g also supports parallel indexing and does not require you to rebuild the search index every time you add a metadata field.

Oracle Text is well suited to search deployments where complex query structure, operators, and sorting are required. Oracle Text search supports flexible use of query operators, nesting, and sorting when building complex queries. These features are often required when searching well-tagged, structured content.

WEAKNESSES

The Oracle Text search relevancy scoring algorithm offers more in terms of relevancy than database search, however, the relevancy computation is still fairly basic. It looks at two signals: the number of times the keywords appears in the document, and the number of times the words appear in all the documents in the index. Based on this relationship documents are given a score. The Oracle Text algorithm does not take into account any other factors such as whether search terms appear in the document's title or other metadata. Oracle Text search does not offer any out-of-the-box weighting configuration options, so modifications to the scoring algorithm must be done as custom development.

In order to take advantage of some of the word stemming and synonym capabilities of Oracle Text, users must be familiar with the proper query syntax. This can be a disadvantage when providing search for untrained or non-technical users. This is even mentioned in Oracle's documentation which states that, "Oracle Text 11g has its own query syntax, which is intended more for use by applications or information professionals rather than casual end-users."

Although Oracle Text search includes a thesaurus file which can be used to set up synonyms, this feature requires users to specify a tilde before their term to include defined synonyms, so they must be trained to use this operator.

There are also limitations to using the snippets generated by Oracle Text search. Oracle warns of the performance impact when enabling snippets for large collections. In addition, customers using WebCenter Content's Site Studio component for Web Content Management, will often find XML code from data files displayed within search result snippets making them confusing and messy for end users.

For clients who are not using Oracle Database 11g, using Oracle Text search requires the installation of a separate database. In some cases this is a weakness because it can add licensing costs and architectural complexity; however, it can also be advantageous as it offloads the search index and can improve performance.

PRIMARY USE CASES

Oracle Text search is recommended over database full-text search for customers using Oracle Database 11g. Its indexing and relevancy capabilities make it a suitable choice for websites,

document management implementations, and other text-heavy use cases especially for smaller sets of content where the limitations of the relevancy computation are likely to be less drastic.

For search implementations targeted at technical users where advanced metadata search is important, Oracle Text search provides a more robust query syntax and is better suited to performing database-like searches than the Google Search Appliance, while still providing a limited number of query interpretation and user experience features. Use cases may include internal catalogs or literature libraries where advanced metadata search is required in conjunction with score-based ranking.

4.4 Oracle Secure Enterprise Search

Prior to 2014, Oracle had positioned Secure Enterprise Search (SES) as their enterprise search solution for indexing and searching content across many systems including WebCenter. In early 2014 Oracle removed SES from their product price list. Although customers with SES licenses can still configure SES as the onboard search engine for WebCenter Content, Oracle recommends doing so only if you are also serving WebCenter content items as part of a wider SES search implementation. Because Oracle Secure Enterprise Search uses Oracle Text search as the underlying search engine, the search experience within Oracle WebCenter Content is effectively the same as it is with Oracle Text. This paper focuses on tools for searching Oracle WebCenter Content; while SES remains among the options, it is no longer being sold as a general enterprise search technology so it is unlikely SES would be a fit for customers in the process of selecting a search technology for Oracle WebCenter Content. Below is Oracle's official statement on the product's direction.

From Oracle.com

As of early 2014, Oracle has decided to cease standalone sales of Secure Enterprise Search (SES) full use- and connector licenses. SES will continue as an Oracle-internal technology to power search over internal content in WebCenter, Oracle Siebel, PeopleSoft, and Fusion Applications and will be available via restricted use license with these products. Customers who have a full use license will continue to be supported for the lifetime of their SES release.

As part of SES becoming a technology to index Oracle application content only, we anticipate that the remaining content connectors will be phased out of support by mid 2014. This includes Microsoft SharePoint (all versions), EMC Documentum & eRoom, Oracle EBusiness Suite, and Microsoft NTFS file shares.

4.5 Google Search Appliance

The Google Search Appliance (GSA) is an on premise hardware-software appliance developed by Google to provide companies with Google search for enterprise content. The search algorithms and machine-learning technology used in the GSA are based on the search capabilities of Google.com with additional features customized for enterprise implementations. The GSA can search websites, databases and document management systems, and supports a wide range of security configurations. Leveraging Google's connector framework, Fishbowl Solutions developed a connector to allow WebCenter Customers to search public and secure content from the Content Server repository with the GSA.

STRENGTHS

The biggest strength of the GSA is its relevancy ranking capability. Google's ability to compute relevancy stems from years of investment in search on the public web. The GSA looks at over 200 signals when calculating relevancy. While Google does not disclose the specifics of their algorithm, their scoring examines many factors such as the presence of a keyword in the document's title and contextual understanding that words in document headings carry more weight than body or footer text. The GSA also evaluates elements like how often the document is modified, the metadata applied to the document, and the proximity of multiple keywords to one another. The GSA allows administrators to further tune the relevancy based on metadata values, content sources, or date. These relevancy adjustments can be applied to users or sites, so each audience sees what is most relevant to them. Relevancy is configured through an admin UI without requiring custom development. In addition, the GSA fine tunes relevance automatically over time by analyzing user behaviors and click patterns, and applying statistical analysis to deliver increasingly precise results.

In addition to advanced relevancy, the GSA also offers a number of features which can enhance the WebCenter search experience. The GSA's query expansion and spelling suggestions include standard dictionary words, as well as words learned from your organization's documents, so that company terminology, proper names, and specialized industry vocabulary can all be included in automatic spelling suggestions and synonyms. The GSA also performs contextual analysis on query text to improve relevancy. For example, it understands that a user searching for "latest apple" probably wants apple computers, while someone searching for "apple calories" is looking for information about a fruit.

Document Previews provide high-resolution preview images of search results which help users determine which document is right for them. The document preview module only supports PDF, Word, and PowerPoint files, so it is best suited for document-heavy search implementations. The GSA's Dynamic Navigation feature enables users to refine results based on content metadata like author, date and location. This is configured through an admin UI and works with existing WebCenter Content metadata.

Lastly, the GSA offers performance and scalability advantages in many scenarios. One GSA can index up to 500 million documents and the GSA hardware is specifically designed for high performance search. When using the GSA with Oracle WebCenter Content the indexing and search load is handed by the search appliance offloading it from the Content Server.

WEAKNESSES

The relevancy the GSA offers when performing keyword-based full-text search is one of its biggest strengths. Conversely, the GSA is weakest when searching content without any full-text such as videos, images, or other binary files. Because this content has no textual context to use for computing relevancy, the GSA ranks these items very low relative to other content. It often takes some careful tuning to get non-text content to appear where users would expect among other text-heavy content items.

The GSA is also not great for searching structured content where advanced query building and operators are required. Version 7.2 introduces support for wildcard search and soft sorting on metadata, however, compared to Oracle Text search, or even database search, the GSA's search operators and advanced query-building features are limited. In addition, it generally

takes 5 to 15 minutes for a document to appear in GSA search results after it is released in the Content Server.

Although the GSA is generally considered easy to administer, it does add a layer of architectural and administrative complexity to your environment. Unlike the onboard WebCenter search options, it is a separate technology which must be configured and maintained. The final drawback to using the GSA for searching Oracle WebCenter Content is the cost; the GSA and WebCenter Content connector require additional licensing not required when using one of the onboard search options.

PRIMARY USE CASES

For search implementations targeted at general end-users where searching is primarily keyword-based such as websites and intranets, the GSA provides superior result relevancy and offers more sophisticated user experience capabilities than the other search technologies available to WebCenter Customers out-of-the-box. It's best suited for document and text heavy implementations where users are searching office documents, web content, policies, project plans, or other text-based content especially when this content is unstructured.

5 Selecting a Search Technology

The table in the following section summarizes the search functionality of each Oracle WebCenter search option. As you compare features, you may also want to consider the level of effort to extend the out-of-the-box capabilities of a particular technology. For example, Oracle SES and the GSA both offer type-ahead search suggestions out of the box; however, WebCenter customers have also created custom components for use with database searching that will capture user queries and provide type-ahead search functionality. A feature like type-ahead suggestions could be added through customizations without an extensive amount of work, whereas modifying or creating an effective relevancy algorithm where one does not exist, is an extremely complex and time consuming endeavor.

5.1 Side-by-Side Search Technology Summary

	Metadata-Only Search	Database Full-Text Search	Oracle Text search	Google Search Appliance
Installation & Architecture Complexity	<i>Low</i>	<i>Low</i>	<i>Medium</i>	<i>Medium</i>
Administration Complexity	<i>Low</i>	<i>Low</i>	<i>Medium</i>	<i>Medium</i>
Full-text Search		✓	✓	✓
Metadata Search	✓	✓	✓	✓
Supported Full-Text File Formats (See Appendix)	0	10	150	200
Type-Ahead Search Suggestions				✓
Word Stemming			✓	✓
Synonyms			✓	✓
Self-learning Scorer				✓
Relevancy Biasing			<i>Requires Development</i>	<i>Built In</i>
Relevancy Signals			2	200
Featured Results				✓
Result Clusters				✓
Related Queries				✓
Faceted Search			✓	✓
Spelling Suggestions				✓
Document Previews				<i>Built-In</i>
Snippets			<i>Low Quality</i>	<i>High Quality</i>
Content Segmentation/Collections				✓
Multiple System Indexing				✓
Sorting by Metadata Values	✓	✓	✓	✓

6 Notes

- ¹Arnold, Stephen E. "Oracle Secure Enterprise Search 11g." *Beyond Search*. N.p., 10 July 2010. Web. 19 Feb. 2014. <http://arnoldit.com/wordpress/2010/07/02/oracle-secure-enterprise-search-11g>.
- ²"Google Enterprise Search Offerings." Fishbowl Solutions, n.d. Web. 19 Feb. 2014. <http://www.fishbowl solutions.com/Google/Search/index.htm>.
- ³"Google Search Appliance: Creating the Search Experience." *Product Documentation 7.0*. Google, Sept. 2012. Web. 19 Feb. 2014. http://www.google.com/support/enterprise/static/gsa/docs/admin/70/gsa_doc_set/admin_search_exp.
- ⁴"Google Search Appliance: Indexable File Formats." *Product Documentation 7.0*. Google, Sept. 2012. Web. 19 Feb. 2014. http://www.google.com/support/enterprise/static/gsa/docs/admin/70/gsa_doc_set/file_formats.
- ⁵"Google Search for Work." *Google Enterprise*. Google, n.d. Web. 19 Feb. 2014. <http://www.google.com/enterprise/search>.
- ⁶"GSA Connector for Oracle WebCenter Content." Fishbowl Solutions, n.d. Web. 19 Feb. 2014. <http://www.fishbowl solutions.com/Google/Search/webcenterconnector/index.htm>.
- ⁷Hatlestad, Kyle. "Adjusting the Score on Oracle Text search Results." Web log post. *Kyle's Blog on WebCenter Content*. Oracle, 10 July 2012. Web. 19 Feb. 2014. https://blogs.oracle.com/kyle/entry/adjusting_the_score_on_oracle.
- ⁸Hatlestad, Kyle. "A Couple of Nice Features When Using OracleTextSearch (Kyle's Blog on WebCenter Content)." Web log post. *Kyle's Blog on WebCenter Content*. Oracle, 16 Feb. 2011. Web. 19 Feb. 2014. https://blogs.oracle.com/kyle/entry/nice_features_oracletextsearch.
- ⁹"Incremental Search." *Wikipedia*. Wikimedia Foundation, 17 Feb. 2014. Web. 19 Feb. 2014. http://en.wikipedia.org/wiki/Incremental_search.
- ¹⁰Mackenthun, Alan. "Effective Strategies for Searching Oracle UCM." Fishbowl Solutions, 23 Apr. 2010. Web. 19 Feb. 2014. http://www.fishbowl solutions.com/fishbowl/groups/public/documents/white_papers/025367.pdf.

- ¹¹"Oracle® Fusion Middleware Administering Oracle WebCenter Content." *Docs.oracle.com*. Oracle, Feb. 2013. Web. 19 Feb. 2014. http://docs.oracle.com/cd/E28280_01/doc.1111/e26692.pdf.
- ¹²"Oracle Secure Enterprise Search 11g." *Oracle Tech Net*. Oracle, n.d. Web. 19 Feb. 2014. <http://www.oracle.com/technetwork/search/oses/overview/index.html>.
- ¹³"The Oracle Text Scoring Algorithm." *The Oracle Text Scoring Algorithm*. Oracle, n.d. Web. 19 Feb. 2014. http://docs.oracle.com/cd/B28359_01/text.111/b28304/ascore.htm.
- ¹⁴"Oracle Text Supported Document Formats." *Oracle® Text Reference 11g Release 2 (11.2)*. Oracle, n.d. Web. 19 Feb. 2014. http://docs.oracle.com/cd/E11882_01/text.112/e24436/afilsupt.htm#CCREF2244.
- ¹⁵"Oracle® WebCenter Content System Administrator's Guide for Content Server." *Docs.oracle.com*. Oracle, Nov. 2011. Web. 19 Feb. 2014. http://docs.oracle.com/cd/E23943_01/doc.1111/e10792.pdf.
- ¹⁶"Searching Oracle Content Server/UCM with Oracle Secure Enterprise Search 11g R1." *Oracle Tech Net*. Oracle, July 2011. Web. 19 Feb. 2014. <http://www.oracle.com/technetwork/search/oses/stellent-white-paper-178229.pdf>.
- ¹⁷"What Words Get Misspelled in Web Searches?" *The Wall Street Journal*. N.p., n.d. Web. 26 Mar. 2014. <http://blogs.wsj.com/digits/2010/06/04/what-words-get-misspelled-in-web-searches/>