An Oracle White Paper
November 2011

Content Organization and Search Optimization for Oracle Knowledge Management Products

ORACLE®

Introduction

The Intelligent Search feature of Oracle Knowledge for Contact Center and Oracle Knowledge for Web Self Service is a search application that can search content coming from different sources, such as content in databases, folder structures, the Web, and Information Manager (the component that manages the Oracle Knowledge databases). Intelligent Search uses a dictionary, which determines how search is to be conducted. Intelligent Search does not search directly against the content, but rather constructs an index out of the documents that have been crawled and conducts the search against that index.

To make it possible to maintain content that is captured appropriately for users of Intelligent Search and other similar search technology, this white paper presents best practices for enterprise content creation and organization as part of a larger knowledge management system. In addition, it offers technical tips for search results optimization.

Intelligent Search

Intelligent Search performs three major steps in returning the best answers:

- **Create a semantic index.** Even before the user enters a question, Intelligent Search creates a semantic index, which organizes all available content along with the key concepts, meta-information, and all existing relationships. The engine does this by importing all content in its original format from the various datasources (such as Websites, PDFs, and Word documents). It then parses the documents and retrieves concepts from the dictionary to create an index of the content. When new content is created or existing content is updated, the engine iterates the process to add the new content into the index.

- **Build the ideal answer criteria.** When a customer asks a question, Intelligent Search analyzes the language patterns via the concepts and rules in the dictionary. The engine also takes into consideration the context of the question by looking at the profile and session information submitted with the question. Once it extracts the language pattern and the contextual information, the engine then develops the criteria needed to match the ideal answer to the question.

- **Retrieve the ideal answer.** After building the ideal answer criteria, and with the semantic index created, Intelligent Search is ready to retrieve the best answer—or, rather, the best answer available in the content. For this step, the engine ranks all possible answers, using its own scoring formulas, and then retrieves the best ones from the semantic index. Intelligent Search formats them according to the display parameters defined by the ideal answer criteria, and sends them to the user in a response overview page.

Before the ideal answer criteria are constructed, Intelligent Search must index the content to be searched. Indexing consists of three basic steps:

- **Crawl content.** Intelligent Search goes to the content sources you have configured and imports native formats (Word, PDF, HTML, or from databases) into the Intelligent Search application.

- **Convert (preprocess) content.** Intelligent Search then parses the content structure and, using the dictionaries, reformats the crawled content into InQuira-defined XML (IQXML).

- **Index content.** Intelligent Search then creates an index of all the content included in those IQXML documents.

When an end user enters a search, Intelligent Search identifies matches in its index of content and returns links to the corresponding original source content in its original location.

It is crucial that, for this process to be successful, administrators ensure that the following occur:

- All needed content must be correctly crawled.

- Content must be created and organized following specific guidelines.

Content Creation

Content is now created in a dizzying number of varieties, with a rainbow of file format and display options available to users. However, if content that is created is either unsupported by search technology or invisible to search technology, it is not much use to either the user or the content owner.

Supported Documents

The Intelligent Search feature supports a total of 255 document types and file extensions. Among these are the following:

- General text documents

  Document Types: application content, American Standard Code for Information Interchange (ASCII) text, HTML, XML, Microsoft Word and Rich Text Format (RTF), newsgroups, and unstructured text from Oracle's Siebel applications within databases

  File Extensions: .htm, .html, .xml, .doc, .rtf, .txt

- Portable Document Format (PDF)

  Document Type: PDF

  File Extension: .pdf

- Microsoft Excel

  Document Type: Microsoft Excel

  File Extension: .xls

- Microsoft PowerPoint

  Document Type: Microsoft PowerPoint

  File Extension: .ppt

Guidelines for Content Creation

Listed below are best practices in content creation:

Document Titles

It is important to provide clear, meaningful, and specific document titles (and subtitles, if appropriate), as well as to avoid naming all the pages identically. For example, do not repeat "FAQ" as the title for all the FAQ documents or use only the company name in all the titles of your documents.

When browsing search results, most users only scan the first three to five words of each title. Therefore, the first three to five words of your title determine the success of your content— particularly, how much traffic your content will generate—so make sure that you choose titles carefully and keep them short. To achieve this goal, use sentence fragments and make use of gerunds, as in "Planning your sabbatical" or "Changing call plan." In addition, make sure your title is meaningful and

can be understood out of context. Remember, you can get better search results when you create titles that match the most commonly searched phrases and keywords for your topic.

Business Writing Guidelines

General business writing suggestions can also apply to the creation of your content. Before creating your content, you should determine your business goals and identify your audience. While writing your documents, consider the level of technical expertise of your readers, and tailor the content accordingly. Avoid acronyms—instead, spell out words (for example, write *service request* instead of *SR*) and define all technical terms the first time they are used in the document.

When possible, avoid creating long documents. It is best to create short documents and to attach the long documents to the content as Word documents, PDFs, or some other commonly readable format. This is because the majority of knowledge access is rushed, causing long documents to be rejected because they take too long to process. It is best to think of knowledge in small discrete units, the combination of which creates the ideal answer.

Technical Guidelines

Keep in mind that well-organized content makes for cleaner and easier content categorization and content acquisition. Use a directory/link structure that logically separates different topics and access levels. Notice that if a logical directory structure is not feasible, you can consider using metatags to logically organize the content.

Note also the following technical guidelines:

- For your HTML documents, use anchors where appropriate—particularly if the HTML documents are long. Make sure that you use well-formed HTML, ideally XML-compliant HTML.

- Consistency is key. So, make sure you give documents a consistent HTML structure. Use the correct encoding parameters: for example, if you write your page in Universal Character Set Transformation Format—8-bit (UTF-8), don't list the encoding as Latin-1.

- Use both an HTML site map page and an XML site map file.

- It is advisable to use descriptive document links—for example, title the link "Content Creation Guidelines" instead of simply the phrase "Click Here."

- Avoid putting important text in images and CID fonts in PDFs because these cannot be indexed. It is likewise advisable to avoid the use of JavaScript for generating links and the use of code-based output such as JavaScript document.write( ).

- PDFs should contain internal outlines and titles, and it is best to make use of bookmarks whenever possible.

- Make sure to spell out a useful description for any 404 http error.

- Bear in mind that if some text or content (outside of Information Manager) is meant to be available for some users and not for others, you should separate those pieces of content or text into two separate documents.

When Using Information Manager

If you are using Information Manager to access the information stored in the databases of Oracle Knowledge products, it is also useful to take advantage of the meta-area where you can include information that is shared across different locales and versions. Finally, when using Information Manager, place in the image library any shared images that are used across channels.

Content Organization

The organization of enterprise content is more than just the placement of files and folders, or of text on the page. Meticulously organized content is accounted for under the company's knowledge management system—which aims to ensure successful content governance, content lifecycles, and learning and change management.

Content Governance

Any discussion of content governance must take into consideration all governance aspects of a knowledge management program. By so doing, the full scope of content governance can be clarified and any potential gaps can be identified.

Successful and sustainable knowledge management programs require attention to three key areas: business processes, content, and change management (including learning). This is true whether the knowledge management program is within a single line of business or across multiple lines of business. The criticality grows as more business areas begin to participate within the knowledge management program and share a single knowledge management solution.

Business Processes Standards and Guidelines

Within a shared knowledge management environment, there must be agreed-upon standards and guidelines for various business processes. Best practices indicate agreement on common core process steps for receiving, processing, maintaining, and archiving knowledge assets (content). This does not mean every business unit has to have identical workflows. It does mean that the enterprise should identify what degree of commonality there is among the business units and agree that all existing and new participants will adopt those steps as the minimum standard process. Guidelines can be developed to direct business units toward best practices. Although these practices may not be enforced by the knowledge management system in use, there may be business benefits to following them anyway. For example, quality review processes are among recommended best practices and ensure that content is accurate, brand compliant, and consistent in tone and style, but the knowledge management system can make the review step either mandatory or optional.

Content Consistency and Quality

A critical aspect of developing a consistent and high-quality user experience is ensuring the consistency and quality of the content. The enterprise must agree on the authoritative sources for various types of content, develop clear content ownership guidelines, adopt and comply with a content lifecycle, and ensure the provision of basic knowledge asset management capabilities, as described below.

- **Authoritative content sources.** There must be clearly identified sources for content on various topics. This does not mean that only these sources may publish on the topic, but it does mean that if there is a conflict, the user will know which source can be presumed to be accurate.

- **Content ownership.** There must be clearly defined ownership rights to content so that decisions can be made regarding the disposition of the content—so that, for example, archiving can be scheduled or items can be deleted in the rare circumstances when that is called for. The business owner of the content is also the single point of contact for agreement on changes to such protocol as editorial standards and review processes. The best practice is to minimize the number of content owners by identifying people in key roles within business units who have sufficient domain knowledge and authority to speak on behalf of the business unit or division within the business unit.

- **Content lifecycle.** There should be a standard lifecycle defined for all content, although the length of time it takes to pass through the lifecycle will vary by content type and (potentially) by business unit. All enterprises should have retention policies that dictate the full lifespan of all types of content. A typical lifecycle starts with the creation of content and runs through to the final disposition of the content (archiving or deleting it). Each stage in the lifecycle should have well-defined standards and guidelines that drive uniformity and consistency across the enterprise and mitigate the risk of noncompliance with regulatory obligations.

- **Asset management capabilities.** In order to be able to obtain accurate and relevant metrics as well as control assets as they move through the lifecycle, the knowledge management system must support workflow, the development and application of key metadata, the ability to archive and restore content, and the ability to capture audit trails that reflect changes to the assets. The best practice is to capture the following at the minimum:

  - Roles and names (IDs) of all actors

  - Status of content (submitted, in process, in review, in revision, approved, published, archived, deleted)

  - History of changes (including dates, nature of the change, and actor causing the change)

System-enforced workflow should be a balance between the minimum effort needed to ensure consistency in content quality (and to mitigate risk) across the enterprise and the effort needed to publish the content. The more the required steps, the longer it takes to make critical content available to the business. The more content that is required to pass through full workflow, the more resources that are required to meet publishing service-level agreements (SLAs). Business-critical content should pass through workflow requiring subject matter expert (SME) review prior to publication. Content that is less critical should potentially skip this step in order to minimize the quantity of content (and potential backlog) handled by the SMEs. Best practice is to utilize a hybrid staffing approach involving a centralized team with in-depth knowledge of editorial policies, publishing tools, and the content base to handle the majority of the work, with SMEs contributing specialized content or reviewing content as necessary based on domain knowledge.

Content Lifecycle

The important characteristics of knowledge are the following:

- Knowledge is information upon which we can act.

- Knowledge is acquired through continuous learning, interaction, and experience.

- Knowledge is never 100 percent accurate or complete.

- Knowledge is not reviewed; we gain confidence by using it.

And yet, in spite of these dynamics and ambiguity, our knowledge is extremely valuable. Given the characteristics of knowledge as outlined above, consider the diagram below about how knowledge content is created:

Strategy
- Determine business ownership areas and taxonomy
- Establish content program and production process
- Content sourcing scope and plan

Audit
- Continuous improvement

Plan
- Staffing recommendations
- Metadata requirements
- Content migration plan

Maintain
- Analytics to determine gaps, success, needs for improvement
- Impact on search and retrieval

Create
- Content asset architecture
- Content asset production
- Governance model for overall process

Figure 1. In an efficient knowledge management system, content is created and maintained according to the above five-phase circular process.

Even though the content lifecycle is a circular process, we identify the Strategy step as its initial step. This step comprises the following activities: determining the business ownership areas and related taxonomy, establishing a program for the content as well as a process for the production of such content, and determining both the scope and a plan for content sourcing. The next step is the Plan step, which consists of recommending choices for staffing, establishing the requirements for the

metadata, and determining the plan for migrating content. Once this plan is in place, we move to the Create step. The main activities of this step include designing the content asset architecture, engaging in the content asset production, and establishing a governance model for the overall content creation process. Following content creation is the Maintain step, during which maintenance of the content and the search system is achieved. This step includes the use of analytics to determine gaps in the content, measure success of the system, and identify areas where improvement is needed. This step includes researching the impact of the created or modified content on both search and retrieve document tasks. Finally, the whole system is subject to the Audit step to ensure continuous improvement.

The established strategy for content creation goes hand in hand with a strategy for managing the lifecycle of content. The first step of the content lifecycle is to engage in an analysis of the content requirements, identifying the involved people and the possible scenarios and looking into content governance and required budget. Following the Analyze step, the business can engage in the Collect step, which comprises the activities of acquiring the content, authoring and editing the same, localizing it if necessary, and adding the metadata. Once this step is completed, it is time to for the Manage step. This is done by modeling the content and establishing its architecture; determining its configuration and components; and designing an appropriate structure, valid and up-to-date standards, and the repository. Finally, the business can engage in the Publish step. This step consists of aggregating the content, as well as transforming it if necessary. It also includes presenting and publishing the content simultaneously in different venues and revising it.

It is also crucial that out-of-date content be detected and eliminated. Along the same lines, it is important to look for duplicates in the content and to eliminate them in a timely manner.

Learning and Change Management

Without the participation of people, a knowledge management solution will never deliver business value. Learning and change management are therefore critical aspects of governing a knowledge management program. To ensure a sustainably successful adoption and ongoing use of the solution, the following activities are considered to be best practices:

- Conduct a *readiness assessment*, which includes an assessment of the attitude of the staff toward change as well as an assessment of the change in the skills the staff will need.

- Identify the various audiences (such as stakeholders, end users, and content creators).

- Identify messages and materials appropriate for and specific to those audiences, considering the results of the readiness assessment. Develop a schedule for delivery of the materials.

- Establish ongoing mechanisms for gathering and responding to user feedback and suggestions.

- Work closely with application manager(s) to coordinate communications regarding any enhancements to the knowledge management solution.

Search Results Optimization

The following are sets of general best practices intended to achieve search results optimization. These guidelines should be kept in mind when making decisions on how to structure the interaction of the consumers with the Website.

Navigation

Once users have become accustomed to searching for their information rather than navigating around to a specific piece of knowledge, many user communities will be able to interact with the system in their own "language" rather than by using a predetermined navigation structure. However, there are cases when navigation is required.

Whenever navigation is involved, it is best practice to ensure the use of a topic-based navigation structure. It is preferable to use navigation based on topics rather than content types, because this ensures that relevant content types are not eliminated prematurely. When designing the navigation structure, one should also strive to ensure that a user can get to a piece of content within four clicks.

The Guided Navigation functionality in Oracle Knowledge products includes the following advantages:

- Users are directed through a multistep troubleshooting process.

- End goals could be another service channel.

- Process wizards enable simple step-by-step navigation.

- Guided flows support more-complex flows with integration into back-end systems.

Graphical User Interface Layout

When designing any graphical user interface (GUI), it is important to think about users interacting with that interface and what key tasks they are looking to complete. Having a user-centric design will ensure that users have a natural tendency to use the tool to complete their search.

Portlets that accelerate rather than detract from the users' workflow should be exposed. A "news" portlet, which does not provide information related to the task at hand, should be put in a place where it is accessible for offline browsing.

It is strongly recommended that the Website have, at a minimum, a document-level feedback mechanism to capture requested content changes. This is extremely beneficial to the users' experience, and the data can be gathered for analysis using the analytics features of Oracle Knowledge for Contact Center and Oracle Knowledge for Web Self Service.

Usability Features

There are a few Website features that can enhance usability. These fall into four major categories:

Search Experience Enhancers

The following features enhance the users' search experience in various ways:

- Process wizards

- Intent responses/search disambiguation:

  Leverage intents and intent responses where appropriate to help classify problem types. To direct the user to the right location, offer the following helpful actions:

  - If the query is too broad, ask for clarification: "Did you mean…?"

  - If the query is related to a product, redirect to that product landing page.

- Promotions

- Glossary

- Facets

Portlets on Document Pages (Information Manager Client)

Portlets on document pages can take users to other useful information, such as

- Related documents

- Related discussions

Issue Resolution

Issues can be resolved through many Website development tools.

- The answer may lie in multiple channels (such as peer-to-peer discussion forums or guided troubleshooting).

- The user may need some help clarifying the issue to ensure resolution.

Cross-Sell and Up-Sell Tools

It is possible to encourage users to buy if the right tools are used. For example, Websites can

- Offer promotions alongside search results

- Offer promotions in other portions of the Website

- Couple intent and intent responses with specific search portlets to drive users to a buy action

These features should be evaluated to see what the end-user benefits will be. Implementing features without a solid understanding of the benefit will just increase total cost of ownership and the complexity of the solution.

Eliciting Natural Language Questions

There are several guidelines recommended to elicit natural language questions from users. These guidelines include the following:

- Make the question box long enough for a full sentence.

- Put a limit on the amount of text that can be typed to prevent paragraphs (200 characters).

- The form submission button should read "Ask" or "Go" instead of "Search."

- There should be a sample natural language question near the box. A sample question could also be prepopulated inside the search box as an alternative.

- A tips link should provide more-detailed instructions with the kinds of questions that produce the best results.

Different factors can influence the correctness of search results. For example, depending on how unique the search question itself is, it may be harder to return appropriate results when compared to a set of queries. The uniqueness of the solution title and the solution content also plays a role in a successful search. The more unique the title and the solution document are, the better the search results will be.

Dictionary Maintenance

From the dictionary side, it is important to remember that terms or phrases that belong to the same concepts as synonyms will be treated as synonymous and, given the same query, provide the user with the same set of results. This must be kept in mind when concepts are created, and when changes are applied to the dictionary ontology.

In order for the search engine to return the best results, it is important that the dictionary be maintained properly. To this aim, the dictionary manager should regularly engage in the following activities:

- **Dictionary tuning.** This involves grading quarterly all existing test profiles in the quality monitor, followed by implementing the necessary changes in the dictionary.

- **Dictionary updating.** As the enterprise introduces new products, all dictionary objects that are product-sensitive need to be updated. These could be concepts, alias lists, intents, managed answers, and search component rules. The intent library should be updated by checking the use of the intents with analytics functionality and by adding intent responses if needed.

- **Analytics reporting.** The client and Oracle collaborate to define analytics reports to identify which objects of the dictionary need to be modified.

Conclusion

There was a time when enterprise knowledge was so limited in scope and technology that organizing it and ensuring access to it was a fairly straightforward, if still daunting, task. Today, the range of formats in which content can be found is extremely broad and only getting broader. Further, such content is no longer simply organized as a file cabinet might be, but requires complex governance.

Recommended best practices for content creation include recommended document types and file extensions, protocol for document titles, general business writing tips, and technical suggestions. Guidelines for content organization include standards for business processes, tips on how to maintain consistency and quality in content, a five-step process for content lifecycle management, and the need to continue to encourage and manage learning and change. Finally, to optimize search results, the key areas of best practice recommendations are navigation, GUI layout, usability features, natural language search, and dictionary maintenance.

The natural outgrowth of blooming content and in-depth organization is the need to tailor content for search technology as the primary means for users to access enterprise knowledge. Oracle Knowledge for Contact Center and Oracle Knowledge for Web Self Service offer the search application and feature Intelligent Search, the next generation of search technology with which knowledge management systems must keep pace if they are to make the best use of enterprise knowledge.

ORACLE®

Content Organization and Search Optimization
for Oracle Knowledge Management Products
November 2011

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

oracle.com

Oracle is committed to developing practices and products that help protect the environment

Hardware and Software, Engineered to Work Together