Develop an Enterprise Search Strategy

April 2014

Abstract

A decision to implement enterprise search cannot be taken lightly. Many companies end up frustrated with high priced products that failed to live up to their expectations. However, these companies typically put little effort into creating a compelling search experience especially given the potential productivity gains effective search can bring. To avoid the failed search experience, information and knowledge management professionals should follow these steps to maximize the impact of their search investments, while minimizing the risk of over-investing in the technology.

Content Growth Demands Better Information Access Tools

The past 10 years have seen both the volume and diversity of digital content within enterprises grow at unprecedented rates. Increased use of departmental file shares, collaboration tools, content management systems, messaging systems with file attachments, corporate blogs and wikis, and databases has turned corporate networks into a virtual mix of content where useless work-in-process, duplicate, and untraceable documents are mixed with valuable information needed to get work done.

Despite IT efforts to control content through the use of content management systems, only a small percentage of content that gets created makes it into a managed repository like an enterprise portal or content management system. This results in:

- **Hard to use, "system of record" document management repositories.** Many of these systems have proven so complex to use that they have been relegated to storing specific authoritative information like business records, parts manuals, and formal methods and procedures that changes infrequently or must be retained for extended time period of time for legal reasons.

- **Out of control, duplicate file storage.** Meanwhile, unmanaged repositories, such as file systems, continue to get used heavily.

Most users who are trying to manage their own "stuff" don't want a repository that requires them to diligently tag their documents for subsequent retrieval. It's just too time-consuming and information and knowledge management professionals face a battle if they expect users to keep track of where they put things and accurately tag their documents. What users want is really good and simple search which looks across all their information and finds just the right information when they need it.

Given this environment, it's no wonder enterprise search software vendors continue to see double-digit revenue growth rates, offering the promise of simple, lightening-fast retrieval of information from disparate content, database, and business applications. As overall content growth shows no signs of slowing, it's also no surprise that major software infrastructure vendors like IBM, Microsoft, Oracle, and SAP are clamoring to market their own flavors of enterprise search products.

Web Search Sets the Bar for Users Expectations

Expectations for quality search continue to rise driven largely by the search experience users get on the Internet through services like Google, MSN, and Yahoo!. "Just give me Google" is a common phrase among impatient users who can't help but frown when they can't find that slide or proposal they know is buried somewhere in the depths of their corporate network behind multiple passwords, unintuitive interfaces, and unfamiliar Intranet site navigation. Users find this when searching web search:

- **Web search is simple to use.** Internet search requires little or no formal training. The wide, white boxes, simple

tabbed interfaces, and generously sized results set text contribute to an experience that makes both novice and expert searchers productive information seekers.

- **It is extremely fast.** On average, Internet searchers execute five separate queries per session adjusting the keywords they use until the results look acceptable. This fact alone is enough to understand why a search inside an enterprise that takes minutes to execute is utterly unacceptable to most users. Yet some search functions embedded in applications still suffer from poor performance.

- **It is relevant.** Relevance is an extremely complex topic, but from an end user's perspective, it's a near magical experience finding almost precisely what they are looking for on a public search site. But relevance inside enterprises tends to be more contextual associated with specific tasks or processes than general-purpose Internet search.

But Enterprise Search is a Different than Internet Search

These differences between web and enterprise search are worth considering when formulating a search strategy:

- **Business content lacks context, text, and links.** Businesses reward brevity in communications. Summarizing ideas, forecasts, proposals, plans, and contracts into bullets, graphs, and key takeaways is the norm. Thus, keyword-based search matching algorithms start at a disadvantage inside enterprises relative to the Web, which is an enormous collection of text and hyperlinks. Further, techniques that use hyperlinks to improve relevance in Web search have little value when applied to enterprises in which business people communicate with popular Microsoft Office formats, not HTML.

- **Users must store and access information securely.** Most Internet information is open for everyone to see. But inside an enterprise, the information must be secured. There is no place for letting information like employee pay rates, financial information, or confidential communications end up in the wrong person's hands. Companies in defense contracting, healthcare, and consumer finance have strict policies and federal regulations force them to secure and strictly audit all use of sensitive information. Thus, careful design of security within enterprise search strategies is a must.

- **Users need the best quality, not the most popular, information.** The ranking algorithms used in Web search are largely a popularity contest among carefully crafted and content-rich Web sites. Conversely, enterprises are littered with work-in-process, fast-changing content that holds only short-term value to anyone but the creator and a small circle of collaborators. As a result, poorly planned enterprise search systems end up returning too many useless results rather than high-quality, final deliverables containing authoritative information that can be trusted.

DEVELOP AN ENTERPRISE SEARCH STRATEGY

Both price and complexity have decreased for enterprise search applications as competition has heated up, but making search work well requires a lot more than just plugging in and turning on an application. To deliver a successful search project, information professionals should focus on seven steps when developing an enterprise search strategy.

Step 1: Define Specific Objectives For Your Search Tools

Users don't search for the sake of searching. They search because they are looking to find and use information to get their jobs done. Use methods like persona creation and scenario design to answer three questions:

1. **Who is searching?** Which roles within the organization are using the search function, and what requirements do they have? For example, a corporate librarian is likely familiar with writing long-form Boolean search queries and using advanced search forms, while a layperson searcher likely prefers a simple search box. A sales professional may need instant access to past proposals for an upcoming meeting, but compliance professionals conducting investigations can often let searches run overnight as a deep search gets executed across massive message archiving and records management systems.

2. **What categories of information are they looking for?** Define the big buckets of information that are most relevant to different roles. And realize that not all roles need all information. Part of why desktop search tools are popular is they inherently define a bucket called "stuff on my machine." Defining categories for searching project information, employee information, sales tools, and news helps searchers formulate the right query for the right type of search.

3. **What are they likely to do with the information when they find it?** After defining broad information categories, work to understand context and answer the question: why are people searching? For example, if a marketer is collecting intelligence on a particular competitor by searching on the company's name, it is often useful to expand that query to include related information, like other competitors in the industry, specific business units or product lines, pricing information, past financial performance, and more. Related information can be included in search experiences through a variety of methods, including the search results themselves or methods like faceted navigation.

Obviously, it's impossible to account for every type of information that users may be looking for, but defining broad user roles, like solution sales professionals or market researchers, and identifying their most common search scenarios is a great way to define the scope of a search project. Use proven methods such as persona design, interview future users to validate assumptions about what processes they are involved in, and identify the information that is the most useful to support those processes.

Step 2: Define Logical Types of Searches

While it is impossible to predict and account for everything people search for, it is possible to organize the search experience so it is intuitive to use. Start with defining logical types of searches. For example:

- **People search.** Searching for fellow employees has gained acceptance as a valuable type of search within enterprises for finding expertise on a subject. A search for people, whether it is a simple name look-up or more advanced expertise search, requires attention to everything from how the query gets processed to how results appear in the interface. For example, searchers typically want to see an alphabetical list of names in a people search as opposed to results ranked by relevance.

- **Product search.** A search for products frequently needs to include product brand names (e.g. Trek), concepts, and terms related to the product (e.g., bike, bicycle, road race, touring), product description, and specific product attributes like frame size, material, and color. Knowing where all this information is stored and how it should be optimally presented to users is essential.

- **Customer search.** It is now possible to search for virtually any item in an enterprise like orders, customers, products, and places. Information professionals should tap into sources like enterprise data warehouses, ERP systems, order histories, and others to create a full picture of the item that is searched.

By categorizing types of searches into logical categories, you can improve the quality of searches. Several methods include applying type-specific thesauri, taxonomies, and other query expansion techniques. Also, you can influence the relevance algorithm in a way that returns the right information the right way like weighting hits in a product description more heavily than a product attribute field. This categorization will help with defining search requirements.

Step 3: Define the Desired Scope And Inventory Repositories

When using the search function built into a particular content management system, this function itself constrains the scope of the search to whatever is stored locally in this content management system. Conversely, third-party search engines, like Autonomy, Endeca, Exalead, Google, Vivisimo, and others will search across multiple content management systems and databases. Increasingly, portal products and collaboration platforms from companies like IBM, Microsoft, Oracle, and Open Text will also search content that is stored inside and outside of content management systems.

- **Create an inventory of required repositories.** When creating your inventory, document the name of each repository, its ownership, description of its content, an assessment of the quality of its content, and the quantity and rate of growth of content in each repository. Also document the technology products used as well as any specific security access policies in place.

- **Match roles and search categories to relevant content sources.** Search requirements often include multiple repositories, such as document libraries, file systems, databases, etc. These repositories usually consist of multiple technology products, such as Lotus Notes, EMC Documentum, Microsoft SharePoint, and others. Using the roles and types of searches you are looking to support, work content repositories owners and your IT department to identify all of the relevant repositories necessary to achieve your desired search scope.

- **Use search applications to reach outside the confines of a single repository.** Cross-repository search becomes essential when companies use different content repositories for different purposes. For example, Vivisimo can be used to index content that is stored in Interwoven and then present the results via SharePoint portal. This approach can drastically improve the search experience, reducing the time it takes to return results.

- **Consider a phased rollout and select simple but telling data source candidates for kick-off.** When rolling out a search strategy project that involves disparate sources and complex UIs, a phased rollout may be preferable depending upon factors such as resource constraints and time-to-launch pressure.

By approaching the project in phases, you can test the process and workflow while familiarizing users with the objectives. Inventory and prioritize the sources at the start of the project. You may want to start with the collections that will have a big impact without being overly complex. For example, basic queries into a CRM system can add a lot of value while remaining relatively straightforward. Throughout this process, it is important to set expectations with your users, since this approach may lengthen their engagement with the project.

Documenting your repositories lets software vendors effectively size and bid on your project. Most search software gets priced based on the number of documents (or data items) in the index plus additional fees for premium connectors that ingest content from repositories like content management systems. For example, strategies that require a limited set of commodity connectors are priced altogether differently than those with premium connectors for content management systems and enterprise applications. Thus, knowing which repositories are relevant and understanding the rate of content growth within them can help to avoid unnecessary spending.

Step 4: Evaluate and Select the Best Method for Enriching Content

When addressing content with very little descriptive text and metadata, like PowerPoint documents, image and video files, and spreadsheets, evaluate several methods for enriching the content to improve the search experience. Methods range from manual application of metadata vocabularies to automatic categorization. Some companies use a mix of both methods. It is important to use the method that best fits the specific situation because each requires trade-offs and very different effort levels.

Step 5: Define Requirements, List Products, and Vendors to Consider

After specifying a search scope, focus on defining requirements for search products. It is not important to not get distracted by irrelevant features, but instead focus on products that adequately meet the organization's requirements over a specified time period. Consider factors like ease of implementation, product strategy, and market presence in any product evaluation. Finally, score and select vendors on criteria that are relevant for your organization needs. For example, the requirements around searching a product catalog are likely to be very different from the requirements for searching a collection of long-form research documents.

Consider these points:

- **There are many vendors to choose from.** There is a wide variety of medium to small-size enterprise search products vendors to choose from, including: Autonomy, Coveo Solutions, Endeca Technologies, Exalead, Google Search Appliance, ISYS Search Software, Recommind, Thunderstone Software, Vivisimo, and many more. Also, large software providers IBM, Microsoft, Oracle, and SAP have one or more search products on the market.

- **Prices have fallen.** Prices have declined significantly in the past 10 years since Verity offered its first product. It used to be that buying any enterprise search product meant six figures. It is not the case any more. However, full-featured, high-scale, high-security search implementations can still have a high price.

- **There is a wide range of features.** Product capabilities range from highly sophisticated, large-scale, secure searches that mix advanced navigation and filtering, to basic keyword searches across file systems. Products differ a great deal depending on whether the content being searched consists primarily of data. For example, high-end search companies like Endeca offer robust tools for searching structured data from databases, while small-scale basic file system search needs can be met with products like the Google Search Appliance.

Step 6: Plan for a Relevant User Experience

Recognize that not all search experiences should be the same. Google, Yahoo!, and Bing popularity on the Web have generated strong interest in offering simple-to-use wide search boxes and tabbed interfaces within the enterprise. But in the enterprise, it is often helpful to use more advanced interface techniques to clarify what users are looking for, including:

- **Faceted navigation that adds precision to search.** Faceted navigation exposes attributes of the items returned to users directly into the interface. For example, a search through a product information database for "electrical cables" might return cables organized by gauge, casing materials, insulation, color, and length giving an engineer cues to help find exactly what he is looking for.

- **Statistical clustering methods remove ambiguity.** Methods like statistical clustering automatically organize search results by frequently occurring concepts. Clusters provide higher level groupings of information than the individual results can provide, and can make lists of millions of documents easier to scan and navigate.

- **Best bets drive users to specific information they need.** Creating best bets is the process of writing a specific rule that says something like: "when a person enters the term '401K allocation' into the corporate intranet, they should see a link to the 401K allocation page on the intranet." Additionally, products like Google OneBox and SAP's Enterpise Search Appliance enable retrieval of frequently searched facts, such as sales forecast data, dashboards, and partner information from back-end ERP systems. Best bets help users avoid a lot of irrelevant results and are very effective for frequently executed queries.

Use basic interface mock-ups and pilot efforts to test, refine, and make these concepts useful for users in your organization. Consider testing out search user interface concepts and tools in a small audience prior to exposing them more broadly to the entire company.

Step 7: Implement, Monitor, and Improve

Search initiatives typically require fewer resources overall than implementation of a content management system.

For initial implementation, gather a team consisting of a project manager, an application architect, a search product specialist, a representative set of users, and partial involvement of content repository owners. Large enterprise-wide search projects will require additional specialized experts in metadata, taxonomy, and ontologies.

For large projects, allow ample time for change management. IT has administrative responsibility for specific repositories throughout the organization. Thus, it is essential to keep IT informed of product evaluation and selection plans so that the final implementation supports security and regulatory policies that are in place for these systems. Anticipate significant investments in change management effort as IT will have to maintain the interface between the search engine and all of its back-end content sources.

For ongoing operations, anticipate reserving a small team for ongoing maintenance of search indexing processes and exceptions. Consider creating search reports for the organization. Monthly reports that include most frequent searches performed, zero results searches, and overall usage of the search function can help you troubleshoot existing implementation and drive future decisions on how to enhance the search experience over time. Enhancements typically include adding types of searches to the experience, further enriching content assets for better retrieval, and incorporating new, valuable content into the overall experience.

Final Recommendations

Focus On Search: More Content Drives the Need for Great Search

The evidence is clear: digital content is growing at double-digit rates. Information professionals should recognize that search is an important issue and make plans to tackle it. Do this by developing a strategy for accessing content from multiple repositories to support your organization. But don't simply buy an off-the-shelf product and hope it meets your needs. Instead, formulate a comprehensive strategy around better information access.

- **Recognize that Web search and enterprise search are different.** Search on the Web returns results from vast numbers of publicly accessible Web pages in a fraction of a second. But enterprise search differs in important ways. Most importantly, enterprise search requires quality content, security, and innovative approaches to retrieving the most relevant results.

- **Design search for simplicity, speed, and relevance.** Despite differences, information professionals should follow what has worked on the Web: simplicity, speed, and relevance. Complex, difficult to learn interfaces, sluggish search results, and thousands of irrelevant or duplicative links are unacceptable to a workforce trained to search on the Web.

- **Follow seven steps toward a better enterprise search experience.** Ultimately, information professionals' success in enterprise search will depend a great deal on their experiences with search products to date. Plan to build a strategy around search that maximizes its impact on the company and minimizes the risk of over-spending on irrelevant capabilities.