USERS DON'T CARE ABOUT SEARCH ENGINES OR SEARCH SYSTEMS. THEY JUST WANT TO FIND the information they need quickly and easily. How they do it depends on their information foraging style, their experience and education, the type of content and who owns it, and available discovery tools – e.g. indexes,[1] controlled vocabularies, thesauri, content templates, and classification tools.

Even the deceptively simple Google search box is the tip of a huge iceberg that includes millions of Web links and content carefully crafted to boost its rank in search results. Savvy information managers know that the public version of Google doesn't work as well on intranets because the hidden part of the iceberg is lacking or deficient. In other words, the search *engine* may be the same, but the search *system* is different.[2]

Whether SharePoint's search engine is best in class is debatable, but it can hold its own. What distinguishes it from other search programs is the rich and integrated *system* that information managers can customize at all levels: enterprise, department, and work group. Much of the customization required to make a stand-alone search engine like Google work well occurs *within the SharePoint platform* instead of being spread across different (and often incompatible) proprietary applications. For maximum user productivity, SharePoint search needs a systems approach that includes the entire information lifecycle, from creation to disposal or storage (Figure 1).

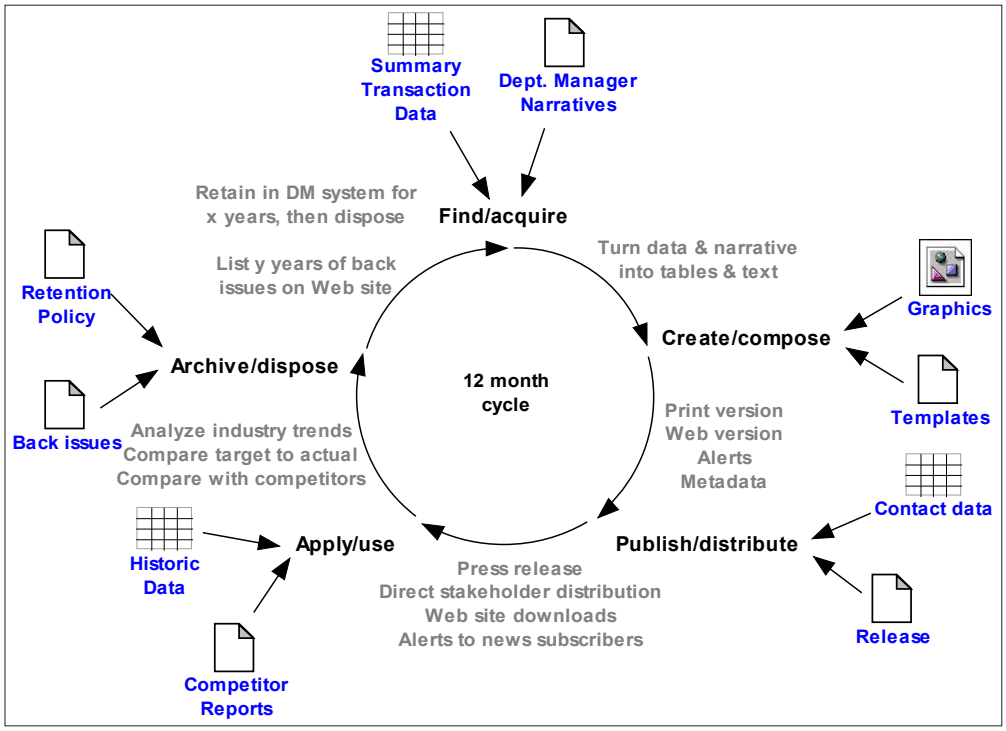

Figure 1: The information lifecycle

SEARCH SYSTEM: USER'S VIEW

Unlike a full-text search engine, which is often used to retrieve a specific known item, a search system responds to the full gamut of the user's information retrieval and processing tasks, such as:

- *Get an overview*. When people pick up a book, they look at the table of contents and maybe the index to find out what it's about. When they visit a Web site, they look at the categories, tabs, and description on its home page.

- *Find frequently used material*. People use physical bookmarks and "dog ears" to go quickly to a specific section. On Web sites, they use electronic bookmarks, tabs, or navigation links.
- *Do a quality check*. To find out whether information is authoritative, current, or popular, people check the publication date, read the author's biography, read a review, or look at who recommended it.
- *Find a known item*. Users retrieve information that they know is available (but they may not know what it's called, who wrote it, or where it's stored).
- *Find an unknown item*. Users look for information that they are not sure exists and may not even know how to go about looking for it.
- *Find an expert*. Users need to find someone who's already solved a similar problem, determine their availability, and assess their level of expertise.
- *Create and update an information "package" for future reuse*. A researcher might check weekly for new press releases issued by a group of competitors.
- *Create an ad hoc information package*. A researcher might compile a one-off report from a variety of sources in response to an executive's request.
- *Find saved items in a personal library*. Typically this involves searching for items saved on a person's desktop computer or personal network shared folder.

To perform all these tasks in most organizations requires multiple applications. One of SharePoint's big selling points is its tight integration with the Windows operating system and the Microsoft Office suite of programs. Operations such as upload, save, and search can be performed without opening another application. Data can flow easily between applications without the need for cut-and-paste or reformatting. Just as important, the integrated system allows nontechnical knowledge stewards to tailor the search system to their unit's workflow.

TWO SEARCH MODELS

A development platform like SharePoint requires us to look holistically at the search needs of users. The following two models — one based on computer science, the other on library science — provide a starting point.

1. *Computer science*. In this model[3], search needs are divided into three categories: informational, navigational, and transactional.

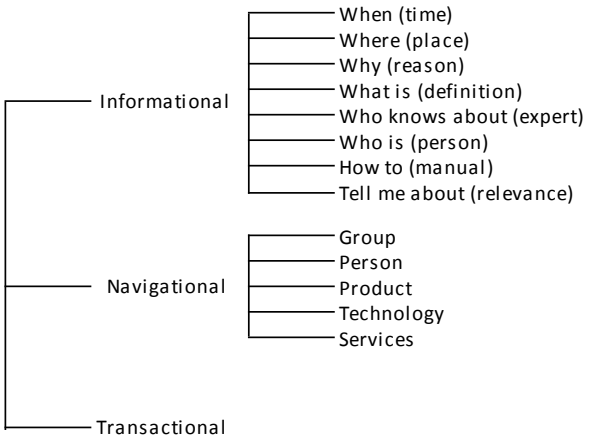

Figure 2: Hierarchy of search needs in the computer science model.

An outgrowth of databases and computer text analysis, the model uses metadata to:

- refine or redirect search results (informational);
- create a browsable "taxonomy" (navigational);
- find, group, summarize, and sort transaction data (e.g. invoices).

Metadata can be entered by a human, supplied by the system, or created using software algorithms.

2. *Library science*. This model[4] includes all the elements of the computer science model but goes beyond it to deal with issues of quality, suitability, and perspective. The library science model covers a broad range of search needs, from facts and statistics to interpretation and analysis. A key part of the model is the reference librarian, who helps users refine their questions, recommends the best sources, and often shows how to use them.

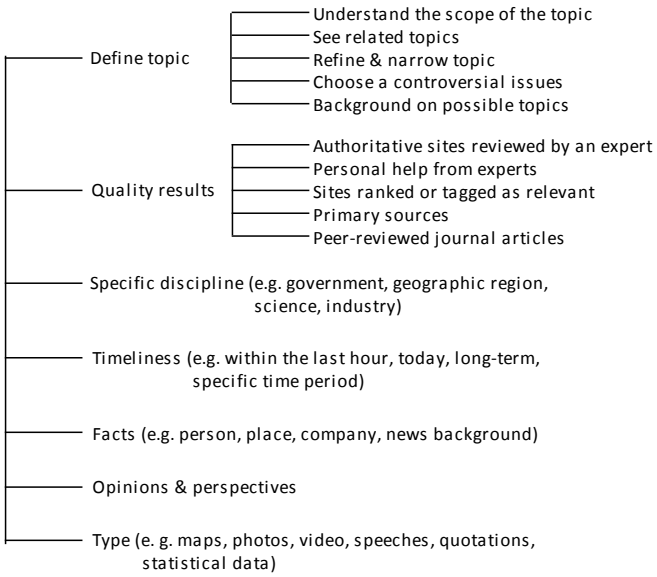

Figure 3: Hierarchy of search needs in the library science model

BUSINESS SEARCH SCENARIOS

It isn't enough to recognize specific kinds of user search needs. To get the most from your SharePoint investment, it's necessary to consider user needs in a specific business context or scenario. Five common scenarios are:

1. *Homogeneous teams*. Members of homogeneous teams tend to have similar backgrounds, speak the same jargon, use a limited number of known sources, and have well-defined work processes. Most of the time, they know what to look for and where to find it. Their search needs track closely with the computer science model of search needs.

2. *Heterogeneous teams*. Heterogeneous teams are often cross-functional, use multiple business or technical vocabularies, access a wide range of both internal and external sources, and sometimes invent their work processes as they go along. Glossaries, cross references, and acronym lists help to bring conceptual gaps among team members. Their search needs are similar to the library science model.

3. *Corporate publishing.* People who write original articles based on research or interviews typically have needs similar to members of a heterogeneous team. Those who simply compile and format data for publication have more in common with homogeneous team members.

4. *Legal and regulatory compliance.* Also called Records Management, Document Management, or eDiscovery, this scenario involves the preservation, control, and dissemination of messages, documents or records. Automated workflow — i.e. routing, saving, categorizing, archiving — is critical, as is accurate metadata. Search needs are similar to those in the computer science model.

5. *Business research.* Business research involves 1) monitoring an organization's vital signs through key performance indicators, 2) keeping tabs on products, partners, and competitors, and 3) gathering and analyzing data for long term planning. The search needs of the first two activities are similar to those of the computer science model. The search needs of the third activity is more closely aligned with the library science model.

Administrators and site owners improve productivity by selecting and customizing SharePoint search system features to match user needs in these and other business scenarios.

SEARCH SYSTEM FEATURES

In subsequent chapters we will explore SharePoint search system features. The most important in terms of user search needs are:

• *Site templates.* SharePoint ships with 25 site templates tailored to different business scenarios. Each template includes one or more search boxes, a navigation system, workflows, and security settings. I found that a large part of the learning curve in the SharePoint search system involves becoming familiar with the different template options, where they occur, and how they interact. For more information, see Chapter 4.

• *Contextual search.* Content searches where the scope varies depending on the user's current ocontext (e.g. whether viewing a site or a list).

• *Social search*. In SharePoint, content search and people search are two separate things — each with its own search box and results page. The relevancy ranking for people search results along with the data displayed for each person depends in part on "social" factors, such as the person's declared expertise and list of colleagues.

• *Metadata filtering.* Users can search on a specific property (e.g. author), refine search results using a topic hierarchy, and view a subset of list or library items.

• *Custom search pages.* A special site template, the Search Center, can be used to customize both the search box and search results pages, including the data displayed for each item and the categories used to refine the search.

• *Desktop integration.* SharePoint 2010 makes it easier than ever to perform a SharePoint search without leaving Windows Explorer, a Web browser, or Microsoft Office.

WHY IT MATTERS

Implementing a search *system* such as SharePoint is typically both a bottom-up and a top-down process. Software installation and maintenance, security systems, and policy development are usually managed centrally by IT, while content selection, design of audience- or function-specific Web sites, and workflow integration are usually performed by departments or business units.

The distinction between search engines and search systems matters because it affects:

- how you approach the implementation process;
- what kind (and how many) support staff are needed;
- where support staff are deployed and how they are supervised;
- how much time and money you spend on customization;
- what policies you put in place to ensure accuracy and consistency.

In Microsoft's free electronic SharePoint book,[5] only 19 of 775 pages are devoted to enterprise search, but the search function is mentioned in the sections on eDiscovery, digital asset management, content management, metadata management, business intelligence, and reporting. In other words, Microsoft is treating search as both a specific software module (the *engine*) and as a component in a *system* designed to address multiple business scenarios.

SEARCH SYSTEM METRICS

Since preparation takes resources, it's wise at the outset to plan metrics that are appropriate for search *systems*. While search *engine* metrics are mostly about **efficiency,** search *system* metrics are about **productivity** — a more complex and comprehensive method of evaluation.

Typically, search *engine* metrics include:

- *success rate* – percentage of time that users find what they want;
- *precision* – ability to find *only* relevant documents;
- *recall* – ability to find *all* relevant documents;
- *time saved* – the amount of time saved in performing a certain task or answering a specific question (requires before and after comparison).

In contrast, examples of search *system* metrics include:

- *customer service performance* – percentage increase in the number of customer service calls that are successfully resolved;
- *publishing turnaround* – percentage decrease in the amount of time required to create a document (e.g. an annual report or Request for Proposal response);
- *sales closure rate* – percentage increase in the number of leads converted from prospects to customers;
- *compliance costs* – percentage decrease in the amount of staff time required to produce required legal documents.

To be sure, the efficiency and time savings attributed to a search engine are part of search system metrics, but the focus should be on metrics that can be clearly shown to affect the bottom line. For more on both kinds of metrics, see Chapter 14.

SHAREPOINT SEARCH SYSTEM ARCHITECTURE

The SharePoint technical architecture reflects a search *system* in which much of the content is stored within program components such as *sites*, *lists,* and *libraries*. The search engine index automatically includes that content, but can also contain pointers to other content on network drives, databases, and external Web sites.

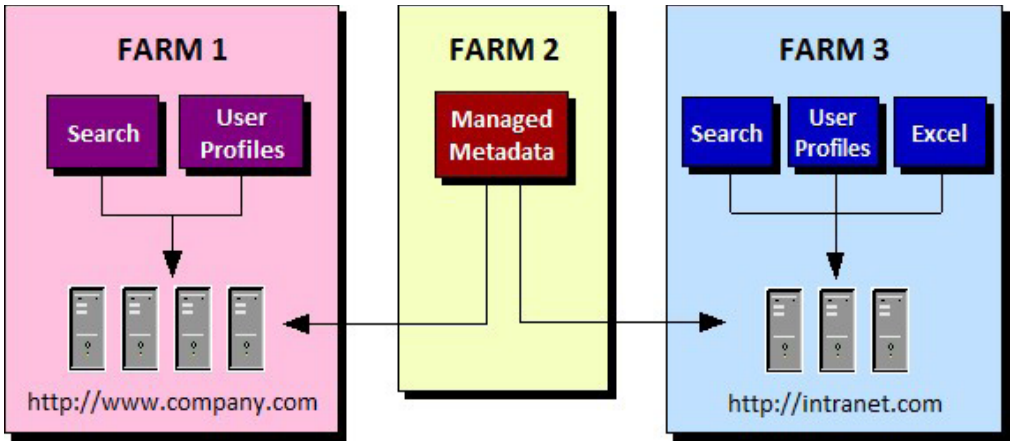

Figure 4: SharePoint 2010 architecture. One or more search and metadata services can be used anywhere in the SharePoint farm.

With most search *engines*, the first task is to decide what content to crawl (index). With SharePoint, the first task is to add user profile data and create content containers such as Site Collections, sites, lists, and libraries. The search engine and metadata management tools exist as separate services that can be used by any site collection or its components.

RECAP

There are two reasons why SharePoint should be approached as a search *system* rather than a search *engine*. First, it includes many of the components that complement full text search, such as built-in navigation, search results security trimming, and authoring templates. Second, it allows end users to mix, match, and tailor these components to meet the needs of their job function or work group. It follows, then, that search system metrics should focus on *productivity*, not just *findability*. The systems approach requires implementors to understand both the user context and SharePoint's architecture before beginning customization tasks.

NOTES

1. In this book "index" has two meanings. In the context of user navigation tools, it means a list of topics arranged in alphabetical order with cross references — similar to a printed back-of-the-book index. In the context of a full text search program, it means a data structure consisting of text snippets along with pointers to the document from which they were extracted — or the processing of "crawling" or scanning content to add values to the index data structure.

2. Search Engine Optimization (SEO), the art and science of increasing the visibility of a Web site on a public search service such as Google, isn't relevant to intranet search. With SEO, a marketer has control only over his/her content and strives to have it rank higher than that of competitors in search results. On an intranet it's possible to control all aspects of the search system and the environment is collaborative rather than competitive.

3. See "A New Approach to Intranet Search Based on information Extraction" <http://research.microsoft.com/en-us/people/hangli/li-etal-cikm05.pdf>.

4. See "Choose the Best Search for Your Information Need" <http://www.noodletools.com/debbie/literacies/information/5locate/adviceengine.html>.

5. See *Planning and architecture for SharePoint Server 2010*, parts 1 and 2 <http://technet.microsoft.com/en-us/library/cc261834.aspx>.