

Searching for KM

ROBERT J. BOERI

Knowledge Management is a term that goes in and out of favor. Whatever name you give it though, consistently capturing and reusing intellectual assets is a critical endeavor within an organization that values information-sharing. In addition to the social and organizational aspects of KM, a short list of systems supporting KM includes search, collaboration, and distance learning. Whether called KM or not, most companies have used email and calendaring for some time and, in a limited sense, already use a KM system: basic collaboration.

Today—as far as KM in practice goes—enterprise search and retrieval seems to be giving organizations the most trouble. “But wait,” you say, “our email system, intranets, and document management systems all *have* search systems.” And that is precisely why enterprise search is becoming so important. It is almost a sure bet that your information repositories—email, intranet, document management system—use different search systems not because anyone chose them, but instead because each repository came bundled with some OEM edition of a search product. Users find it difficult to learn how to use these different search systems given their differing capabilities and syntax. This difficulty violates a cardinal principle of KM: usability. Yet even that problem pales before the fact that each OEM search system focuses on only one repository’s content. You can’t search in several places at once, performing the equivalent of a “joined query” to find key information common to documents in different repositories. To search several repositories with one query, you must upgrade to a single enterprise edition of some vendor’s search solution. Those two problems, usability and single-repository search, are only the beginning in resolving the enterprise-wide search problem.


The volume and types of content also keep increasing. To cope with ever-increasing volumes of content, vendors continuously add features (or acquire companies) to increase the relevance of searches. Vendors’ systems start with fundamental designs differentiating them and constraining their capabilities in subtle ways. Lastly, enterprises themselves are developing corporate KM strategies aiming to enhance their use and management of knowledge assets. In this column, I’ll focus on basic considerations in search systems; later I’ll cover other KM search issues.

In all search systems, results improve when queries provide more restrictions. Search for “cell” and you might get references

both to biology and to terrorist organizations. Include “clone” in the search—whether as an additional search element or as a keyword attribute that may not appear literally in text—and results will likely exclude al-Qaeda and hone in on the biological results. The key point is how each system delivers the most useful search results without omitting important ones, leveraging the strengths of its system design. Verity uses a strong Boolean base with keywords and thesauruses to yield precise, controlled searching. Autonomy’s core strength is Bayesian statistical techniques. Its search system can learn what users consider relevant, simplifying and improving search results over time, but requires well-selected samples to learn from. Convera’s design is based on fuzzy logic

and neural network technology. When natural language is important or searching scanned documents with their inevitable optical character recognition spelling errors, Convera may have the edge. Each vendor has its own spin on the advantage of its enterprise solution. Ron Kolb, Autonomy’s director of technical strategy, told me “The inherent flaw with keyword technology is keyword technology itself. Autonomy’s approach is instead done proactively by

linking to content.” Emphasizing the importance of keywords, Verity’s senior vice president of development and new business activities, Dr. Ashok Chandra, told me that Verity’s “...advanced search capabilities include parametric searching to find single or groups of documents with specific attributes, federated search to return results from many sources in a single query, and category drill-down to let users browse through categories and subcategories.” Lastly, Convera’s spokesperson says it will continue enhancing RetrievalWare’s “...natural language processing capabilities as well as the ability to manage and integrate its product in a highly scalable manner across the enterprise.”

These basic design directions provide a starting point for selecting an enterprise search solution, but only that. My next column will list and explore another set of important differentiators. Is your organization struggling with an enterprise search solution? Email me your thoughts and I’ll share the best in an upcoming column. 

ROBERT J. BOERI (bboeri@ieee.org) is a knowledge management analyst for a Boston-area biopharmaceutical firm.
Comments? Email letters to the editor to ecletters@infotoday.com.