



# A Glimpse at the Future of Enterprise Search

**Bob Lewis**

**E**nterprise search is at the other end of the spectrum from the Internet search tools with which we are all so familiar. The differences between the two forms of search are so great that we can use several spectrums to compare enterprise search with Internet search. In fact, the contrast between the two forms of search is so fundamental that many insist on referring to enterprise search as *information access* to emphasize the differences.

## **PRECISION**

One spectrum for viewing and comparing search performance has precision at one end and recall at the other. The enterprise category is more interested in precision, where relatively few documents are returned in the search results but their relevance is high. Recall is what we all generally experience on the Internet—large search results with the documents or files that we are interested in buried in pages of search results.

*Search by a customized and dynamic SOA can provide the greater precision, relevancy, semantic awareness, and adaptivity the enterprise requires.*

All three of the theme articles in this special enterprise search issue of *IT Professional* approach search from the precision end of the spectrum—but with differing perspectives and technologies.

## **SEARCH SYSTEMS**

Another spectrum spans from the one extreme of a basic search tool to the other of an integrated

search system composed of several diverse search technologies as well as multiple forms of input and output. Enterprise search is at the system end of that spectrum whereas the state of today's Internet search is more tool-like with indexing at its core, and usually the only search technology applied.

Enterprise search systems can include numerous different forms of output, going well beyond the list of files or documents we expect to see during an Internet search. The most common approach taken to make the search results more useful is to operate on those results using extraction and visualization techniques. In other words, the system approach to enterprise search continues processing at the point where Internet search stops by showing the user, in different ways, what content the documents contain and the returned pages. Search systems do this by extracting all places in the search results and mapping them for the user to check if there is interest. Such systems can also extract all the people, organizations, objects, and other entities contained in the search results and let users select those that are of interest. In addition, search results can be displayed as heat maps, digraphs, neighborhood maps, and by using other visualization techniques that permit users to navigate within each graphical depiction to locate, with precision, the most relevant information.

The three theme articles in this issue each apply a different technology—different enough to be complementary. And complementary enough to be integrated into a system providing synergistic qualities that together yield high precision to users in the enterprise.

## DIVERSE SEARCH TECHNOLOGIES

David Bean's article describes search in terms of facts, events, and actions (subject, verb, object or actor, action, object), often referred to as *triples*, so that unlike conventional noun-focused Internet search, the verb becomes prominent. This capability of viewing search results as triples adds another dimension to search within organizations. It not only brings greater precision but also can turn search results into a story as the facts, events, and actions are displayed in time sequence. This display of triples also makes including collaborative sessions and email strings as search sources more practical since information overload in the enterprise is somewhat mitigated by the reduced—but higher quality and logical arrangement—of the search results.

The article by Fritz Knabe and Dan Tunkelang describes the faceting capability available to an enterprise in advanced search tools. This approach to high precision presents users with multiple perspectives of the same entity, person, place, or thing. The perspectives displayed depend on the entity that is the object of the search but might include time, cost, quality, effectiveness, manufacturer, or location. Users can then navigate along any of these paths. With each narrowing choice, more microfacets are displayed that can lead to the most relevant and pertinent search results.

## INFORMATION EXTRACTION

Yet another spectrum that distinguishes enterprise search from Internet search has document retrieval at one end and information extraction at the other. In other words, instead of search results that return a set of files or documents for users to open and study, the relevant information in those files is extracted and presented as phrases, sentences and paragraphs with highlighting.

Stephen Buxton's article on XML content servers describes the unique capabilities of this form of repository system and the extreme precision and information extraction that it can achieve. The server's content of unstructured text is richly tagged, usually by inflow entity extractors or taxonomies. This provides a high degree of semantic quality and makes high relevancy search and disambiguation possible. Search, as well as other applications, can be developed to sit atop the server and take full advantage of the metadata. In this way, the enterprise can benefit from true information extraction in search as well as in other applications requiring high precision and a degree of semantic awareness.

## FEDERATED ACCESS

Another important spectrum for comparing Internet and enterprise search is federation or federated access versus single-source access. An enterprise generally requires more than crawlers that simply index static Web pages. It also needs to send out sophisticated queries that can extract information from dynamic Web sites, those that generate a Web page only after being queried. An enterprise must also extract information from relational databases, particularly

when those databases contain unstructured text in the form of character large objects. Users might also want the search results to include saved collaboration sessions and email streams. This federated access capability for applying an assortment of connectors and techniques to data extraction from diverse forms of repositories and databases is important to enterprise search whereas Internet search provides satisfactory performance, crawling only static Web pages and following links.

## COLLABORATION AND EXPERTISE LOCATION INTEGRATION

Enterprise search systems are also distinguished from Internet search tools in the way they relate to their organizational setting. Rather than a standalone function, we expect enterprise search to integrate tightly with other knowledge management functions such as collaboration and expertise location. This tight integration can yield some striking synergies. For example, search results might include information on people in the organization who are experts in the topics the search covers in addition to documents and information extractions. As we've seen, search results can also draw from saved collaboration sessions and email streams. In this way, enterprise search becomes a mechanism that continually reinforces the reality that an organization's knowledge resides in its people and their experience and training—the tacit—as well as in documents and email—the explicit.

## ENTERPRISE SEARCH AS A SYSTEM OF SERVICES

Enterprise search is basically a three-way capabilities extension of Internet search where input becomes federated, processing is expanded to include other search technologies, and output is the object of continued analysis in the quest for greater precision and relevance. We can view each of these capability extensions and their underlying functions as separate and individual services available to a basic search core that resembles an index-based Internet search tool with simple crawlers. Organizations can then select, architect, and orchestrate the capability extensions or services that best fit their business processes and culture.

**T**he current direction of enterprise search is an SOA built on an indexing core. That core will be complemented by services such as the diverse search technologies proposed in this theme issue and future issues of *IT Professional*. It will also be supported by the federated services of crawling and querying techniques and the many forms of analyses services operating on the basic list-of-documents output. ■

*Bob Lewis is chief technologist for Knowledge Management at Lockheed Martin. Contact him at [bernard.lewis@lmco.com](mailto:bernard.lewis@lmco.com).*