

Search for Knowledge is Now Open

By Cassandra Targett, Information Architect, Lucid Imagination

Search is the circulatory system for knowledge. It excels at finding data and information, from which knowledge is constructed. Search can't quite create knowledge—ultimately, that's a human task—but it does help. Yet just as knowledge is not static, search must be more than a static search box.

The best search ensures organizations constantly assimilate new ways of knowing and sharing as they assimilate more data. Of the various alternatives available, open source Apache Lucene/Solr search technology is best suited for integration of search into knowledge management, in two fundamental ways.

First, the more generalizable the search technology, the better it can be adapted to the ongoing needs of knowledge management systems. Whether in processing constantly growing volume of content inputs, variations in source content constructs, or even user query flexibility, search technology that is more open to change will add more long-term value to knowledge management.

Second, to be most effective, knowledge management must focus on end users. The search process and its technology keep access to information fresh and relevant—accounting for differences across organizational contexts, enabling context-specific, organization-specific knowledge management. In fact, one might argue that of all the components of an effective KM system—which may also include document management systems, portals, ontologies, etc.—search is the most actively user-centric.

Indexing Content

The rapidly accelerating supply of content available to any organization requires tremendous flexibility. It's easy to say enterprise search is “not Google or Bing”; it is commonly understood that search in a knowledge management system is bound to the enterprise of a particular organization. But too often, legacy enterprise search platforms fail to embrace that rapid change is the only constant.

A key dimension of change is diversity: email conversations, PowerPoint presentations, product diagrams, bug reports, information about other colleagues, plus external content from database vendors or

other websites—all need to come together in a unified search experience, through an index spanning all of an organization's available information.

Naturally, none of these content types stand still; some change by the day, hour or even minute. Lucene/Solr is uniquely equipped to handle this with speed of index update at near-real-time performance. For example, innovative applications like Twitter use it to index 200 million tweets and process 1.5 billion queries per day.

Knowledge management effort has historically focused on metadata such as author information, location, access control, description, document subject or type, etc., centered on improvements that normalize variations and ease content findability. Open source search takes that one step further: rather than normalizing the metadata, it uses software programming constructs, actively adapting to metadata variations for unmatched flexibility. It's a relatively straightforward matter, for example, to weight various document attributes, to add the results of past searches to future searches, to account for social graphs, to build synonym dictionaries and more—creating a dynamic of learning wherein search feeds knowledge.

Yammer and Jive, two leading innovators in enterprise social collaboration platforms, use Lucene/Solr to understand relationships between users as part of their search experience. Search helps keep content fresh by accounting for metadata variations that may otherwise degrade good results from one repository in favor of poorer results from another one.

The speed, transparency and flexibility of open source Lucene/Solr are essential to managing content used by knowledge management systems to construct better results across repositories, either during indexing or on-the-fly for results.

Focus on Users

Not only is Lucene/Solr adept at getting the right content into a unified index, it helps make the user experience of searching as seamless as possible. Most users in a professional setting are used to searching; the key, however, is to allow them to search in whatever way they are most comfortable.

As full-featured search technology, Lucene/Solr natively supports:

- ◆ Natural language questions (“What is the ...”);
- ◆ Simple keywords with ability to specify which must or must not be included;
- ◆ Proximity (term x within 10 terms of term y);
- ◆ Fielded search (find terms in title field, or documents by specific author); and
- ◆ Support for numbers, hyphens and other special characters.

Combining these techniques delivers extra power when you tune search parameters to organizational needs. For example, a medical equipment company may include researchers in the user community who are accustomed to crafting complex queries of medical literature to retrieve findings specific to their work, or they may want to browse through recent articles in their specialties. Business development staff may be most comfortable with keyword searching to find recent proposals or presentations by their colleagues. The system need not make judgments on which method is superior—delivering quality results however queries are crafted.

Finally, in the context of knowledge management, display of results must serve the user. Results can be complex when information is stored in disparate repositories. For example, “Gladys Knight” might be one of the best results for a search, but her phone number is in the LDAP system while the projects she's worked on is in another database entirely. An effective search for this person would combine the two data points in display of the result to the user.

There are several other options that Lucene/Solr makes available to aid users in finding the information they need. These include:

- ◆ Facets: Derived categories that count up how many results are within a certain category;
- ◆ Predefined top results, where certain queries deliver content that knowledge curators want to lead users to first;
- ◆ Highlighting: Display of search terms in context; and
- ◆ Mixing different content types (people, documents, images) within a single result set.

It's important to avoid cluttering the screen with options; users ultimately want to find the information they need and will not tolerate unnecessary distractions. Again the flexibility of search ensures each organization can find the right balance.

Again, all of these capabilities are well supported in Lucene/Solr. Building a smart search application suited to the constantly evolving, growing supply of content ensures end users can more easily unlock the knowledge within. Open source makes it easier to build knowledge into search to begin with. ■