

UPDATE

BA-Insight: Drilling for Information to Drill for Oil

Susan Feldman

IN THIS UPDATE

This IDC update discusses how and why a large oil company decided to implement BA-Insight Longitude to solve the problem of gathering information and sharing it across its research organization. Large oil companies run on information. They must mine it as assiduously as they do their oil reserves. When a major oil company realized its researchers were not able to pull together the information it needed and then share it across its research organization, it looked for a product that would work with its new SharePoint implementation as well as tap into hundreds of scattered repositories.

In Search of Search

Defining the Problem

Drilling for information comes naturally to an oil company. In a business in which time is money and some accurate, comprehensive information is a requirement for technical and R&D departments, streamlining the information gathering process makes sense. By 2007, the company had amassed 2TB of data that its researchers needed to access on a regular basis. It found that pulling together the data from research reports and databases in multiple formats across diverse locations with differing schemas had become a notable time sink. When the research department of a major global oil company realized that it needed to tap into more than 600 different repositories around the world to serve more than 10,000 engineers and researchers in scattered locations, it began its hunt for a solution. Its goal was to create a single virtual knowledgebase for its research organization.

Criteria for Selection

Although the company already had a search engine, it had been heavily customized for its technical community. As it moved to a companywide deployment of Microsoft Office SharePoint Server (MOSS) 2007, it needed to keep custom features such as date range search, alerts, extended search, metadata field displays, and parametric search while tying into MOSS. It also realized that it couldn't afford to devote its developers to recreating research features for drilling down, specialized taxonomies, metadata extraction, or browsing that it had previously built for the old system: It is in the energy business, not the search software business. The solution had to provide a single point of access to multiple repositories. It had to merge access to both structured and unstructured information. The access system had to scale to over 2TB of data that was in multiple formats, schemas, and repositories. It had to map the metadata in databases to the documents that the metadata described. It had to integrate with MOSS and connect to Lotus Notes as well as SQL databases. Finally, using MOSS as the front end, it had to present integrated results in a single, easy-to-

use interface that enabled fuzzy and parametric search, browsing, and tools to help researchers speed up the information gathering phase in order to have time for better analysis. The system had to be ultrasecure, and it had to provide uniform access security across all repositories, even though each repository had its own access rights model. It also had to serve remote users with poor bandwidth but still centralize its document repositories to ensure compliance uniformity of information and security. Because of the many repositories, it had to support real federation and merge results effectively.

Selection and Deployment

Because of the requirement for MOSS integration, as well as the need to provide unified access to structured and unstructured information, and the requirement to scale beyond 2TB, the company found that BA-Insight's Longitude was its first choice. An additional selling point was the fact that BA-Insight had its origins in the energy industry and understood the problems and the terminology. The effective federation it needed was part of the application, and the excellent Notes database connector that comes with Longitude was another key selling point. For a proof of concept, Longitude was set up in a lab environment and tested for performance, security, and to see whether its architecture would fit with the existing software infrastructure.

Once the testing was concluded, the system design was completed and customizations were begun. Most of the work is being done in-house, with configuration help and occasional consulting from BA-Insight. Because the company wanted to provide separate production, development and acceptance, and testing environments, it deployed Longitude on a server farm of 17 servers. The distributed architecture supports this three-tiered environment and also provides redundancy and backup. During the development phase, BA-Insight worked with the company to build the custom features on top of Longitude that the technical community had come to expect. The next step, in February and March 2009, will be to test for user acceptance, although the demos they have been giving have all been enthusiastically received. They will roll the system out to all users at the end of April.

During the entire development, testing, and deployment process, the company's change management team has fine-tuned the relevance, chosen best bets, and implemented taxonomies. And, it has educated users. One feature it likes about Longitude is its ability to use multiple taxonomies for different departments as well as separate indices and metadata.

Challenges

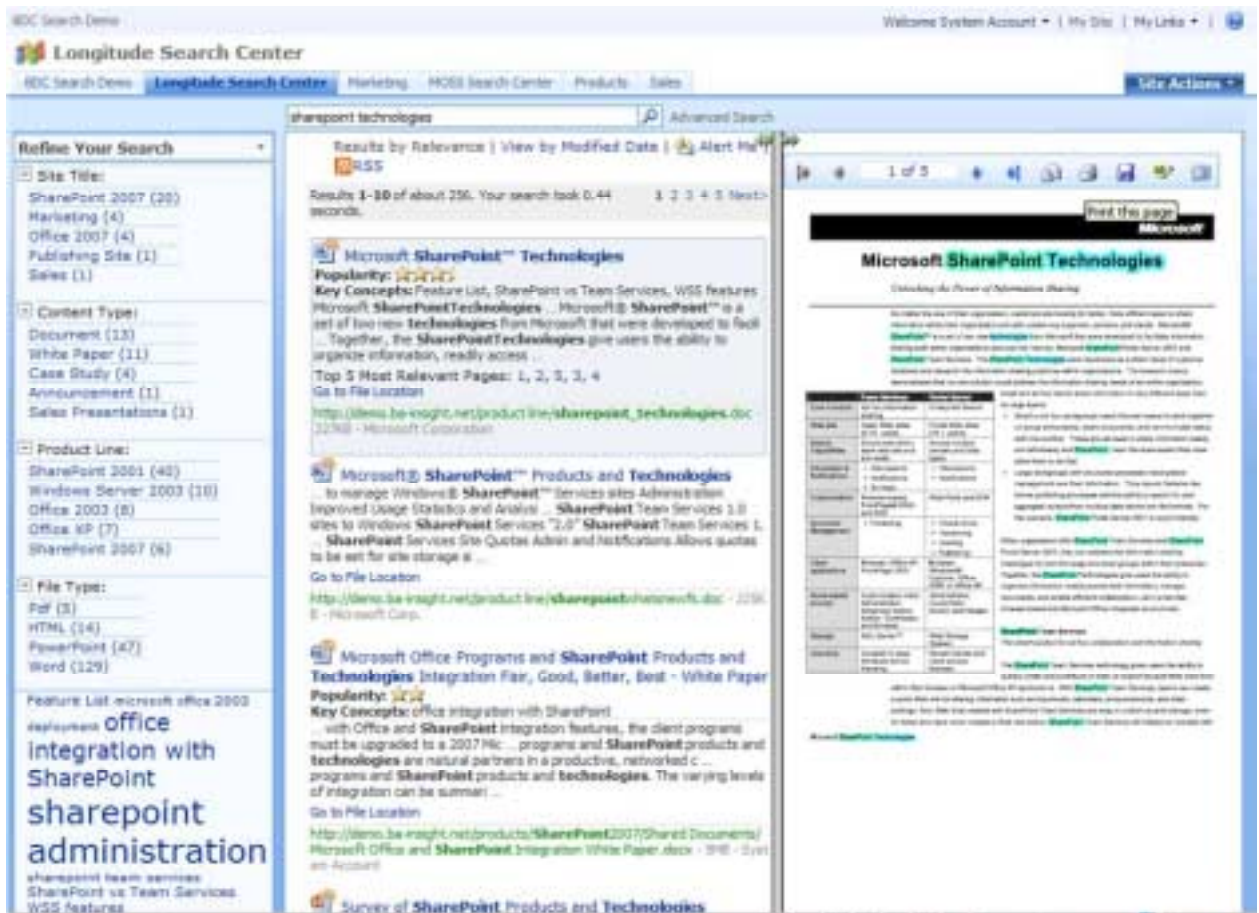
The company quickly found that its biggest challenge was not technical; it was the process of identifying and cleaning up the data and then defining how users would want to use the system. What should the scope be? How would it detect and remove duplicates effectively? How could it create a single secure point of access to multiple repositories if each repository has its own security model, its own schema, and its own data types? Could one system support all those different security models at run time?

Future Outlook and Some Surprises

BA-Insight's Longitude is now in demo mode at the company. Business users are eager to make their information gathering more efficient and to enable it to discover the unexpected. The combination of parametric searches, concept searching, and well-designed browsing, coupled with an easy user interface makes them enthusiastic. They like the preview mode from BA-Insight that lets them look at a document with hit highlighting, without having to leave the search interface to open each document to see what's inside it (see Figure 1).

FIGURE 1

BA-Insight Longitude for SharePoint Interface



Source: BA-Insight, 2009

The project manager was particularly surprised at the lack of technical challenges he encountered. He points out that tuning relevance is critical and is dependent on how

people want to look for information and what they consider to be important. In other words, relevance is in the eye of the beholder. Tools from BA-Insight such as concept clouds, autotagging, popularity ranking, and search logs help to surface what people are looking for, and they provide multiple pathways into the data. He also noted the importance of providing the right type of drill down and using existing metadata, taxonomies, and established terminology. He cited the complexity of the data and the lack of normalization among systems as his major challenges. His word of caution: "You can implement a search engine, but if you don't have repositories with good data, it won't do you much good."

Search can't solve the dirty data problem. At IDC, we have been told over and over again that the most time consuming, difficult part of many deployments is to clean up the data, reconcile schemas, and normalize terminology. While search results can only be as good as the underlying information, what a search engine can do is to unite scattered repositories, reveal data problems, and streamline information foraging. Coupled with careful implementation, these systems become key components that information workers rely on.

Future Plans

While the immediate objective of this project is to roll out enhanced search to the technical community, the project managers would not be surprised if enhanced search was eventually extended to the entire company. In particular, they point to the effective interface design, including the preview pane as a feature that employees see and covet because they immediately understand the time savings they would realize from not having to open each document separately to see what's in it. Oil companies also have far-flung networks of employees, many of whom work at the end of a fairly restricted Internet pipe. Bandwidth is a big problem for them, and the preview pane will solve download time problems as well. Having to click on each document separately, they report, is "untenable."

The company also plans to use the Longitude Notes and database connectors throughout the company to enable other departments to connect their databases to MOSS. It suggests that companies seeking to deploy similar information access systems think carefully about which repositories to include and vet them to make sure that they contain high-quality, final versions of documents that should be shared broadly. It is also planning how to support the information access system as it moves from project development to ongoing application. Because information systems are dynamic, having a strong support group to maintain taxonomies, monitor metadata and schemas, and answer questions will ensure that the system grows with the needs of the company.

Search deployments succeed when there is a strong partnership between the search vendor and the customer. Because of the ease of deployment, extensive and flexible features of Longitude, good customer service from BA-Insight, and the oil company's careful planning, careful cleaning of data, and conservative approach to developing, testing, and rolling out this search system, it looks like this project will succeed and grow, providing the fuel to develop the fuels of the future.

LEARN MORE

Related Research

- ☒ *IDC Predictions 2009: An Economic Pressure Cooker Will Accelerate the IT Industry Transformation* (IDC #215519, December 2008)

- ☒ *Worldwide Search and Discovery Software 2008–2012 Forecast Update and 2007 Vendor Shares: Bloom Amid Economic Gloom* (IDC #214349, October 2008)

- ☒ *How Do Information Workers Look for Information?* (IDC #212580, June 2008)

- ☒ *The Promise of Search* (IDC #211820, April 2008)

- ☒ *The Hidden Costs of Information Work* (IDC #201334, April 2006)

Copyright Notice

This IDC research document was published as part of an IDC continuous intelligence service, providing written research, analyst interactions, telebriefings, and conferences. Visit www.idc.com to learn more about IDC subscription and consulting services. To view a list of IDC offices worldwide, visit www.idc.com/offices. Please contact the IDC Hotline at 800.343.4952, ext. 7988 (or +1.508.988.7988) or sales@idc.com for information on applying the price of this document toward the purchase of an IDC service or for information on additional copies or Web rights.

Copyright 2009 IDC. Reproduction is forbidden unless authorized. All rights reserved.