A Case of Improving Content Flow—From Storage to Access

Company: R.R. BOWKER

R.R. Bowker is an ISBN (international standard book number) agency for the U.S. The company receives data feeds from publishers, wholesalers, and distributors. New Providence, N.J.-based Bowker, which has served the publishing industry for more than 130 years, sells its information in data feeds, web services, and websites, as well as through other custom solutions.

www.bowker.com

BUSINESS CHALLENGE

Bowker certainly handles a lot of information—from data being entered into its systems to data being requested by customers. Bowker’s Global Books in Print database contains bibliographic information on millions of book titles, with more than 20 million documents accessible to a customer base comprising the entire book industry, from publishers and libraries to bookstores and distributors. Those millions of documents equal approximately 600GB of information in a database that conducts 600,000–700,000 daily updates. Query volume averaged four queries per second. Bowker needed an extremely effective way to store and provide access to a whole lot of data.

VENDOR OF CHOICE: MARK LOGIC

Founded in 2001, Mark Logic is an enterprise software company based in San Carlos, Calif. The company’s main product is MarkLogic Server, an XML content server. Mark Logic focused initially on the publishing market, assisting customers with various content-related challenges, such as providing more integrated search functionality for their end users, as well as improved overall access to that content. The company is now entering additional markets, such as government and financial services.

www.marklogic.com

THE PROBLEM IN DEPTH

Demand for the vast amount of information housed in the Books in Print database was high—and the company’s technology solution of Verity search capabilities working with an Oracle database were not effectively meeting that demand. The company recognized that a new search solution was necessary to help Bowker customers retrieve relevant content from its enormous database. “With Verity, we were getting content out to our customers, but we were anywhere from six to eight days behind when a particular document would come in because of the processing nature and volume,” says Bowker CIO Mark Heinzelman. “One of the [things I thought about], when I started looking for the new tool, was focusing on the idea: ‘I have to do 600,000 to 1 million updates a day while it’s being searched and while the content is being updated.’”

THE SOLUTION

The Mark Logic XML content server essentially combines full-text search with the W3C-standard XQuery language. The platform can load, query, manipulate, and render content. When content is loaded into the server, it is automatically converted into XML. Employing the MarkLogic Server enabled Bowker to improve its search capabilities through a combination of XML element query, XML proximity search, and full-text search. Mark Logic’s XQuery interface searches the content and the structure of the XML data, making that XML content more easily accessible. It took only about 4 to 5 months for Mark Logic and Bowker to develop the solution and implement it.

Beyond helping Bowker solve its immediate need for a better search engine, Mark Logic also assisted the company
I now control when the content gets out to the customers.

THE OUTCOME

Using Mark Logic technology has helped Bowker improve the overall functionality of Books in Print, providing quicker updates and access to information to end users. The content can be updated and searched simultaneously, and users can receive the information more efficiently—with a subsecond response time, according to Heinzelman. “I now control when the content gets out to the customers,” he says. “Now we run a day behind, but that’s a choice we made; not because we were limited by the tool. The search response time is subsecond; where the search response time in the Verity world was around two and a half to three seconds.”

Heinzelman says that the increased speed of information searches and updates is more important to customers who are using Books in Print as a purchasing tool and that need the most up-to-date book prices. Customers just using the database as a research tool aren’t as interested in the tool’s efficient properties. “But overall, we find that most of our customer base is asking for it to act as quickly as possible,” says Heinzelman. “Especially when you’re coming out with the next best-seller, like the Harry Potter books, you have all of that content coming in, and customers want to be able to find...
it very quickly. From our standpoint, competitors to us are like an Amazon. So we’ve got to be in that world.”

Adds Kreisa: “We see that as a general trend for publishers and information providers wanting more flexibility and to more quickly provide information to their customers so they can better compete.”

Another key benefit is the cost savings Bowker has realized as a result of the initiative. Bowker needed a full-time employee on staff to manage Verity. Now, the company has an employee who spends, at best, one-quarter of his time managing the current infrastructure. “We save on the infrastructure side internally and our customers get the content more quickly,” says Heinzelman.

“One of the things I realized is that the days are getting shorter. Everyone wants to be in a 24/7 world,” adds Heinzelman. “To process data in that time, you have to touch it as little as possible. My ultimate goal is to put the content repository out on the edge so there is one repository that our data services group and our editorial group uses to enter the data as it gets searched at the same time. To date, I don’t know of anyone else that can really do that.”

Heinzelman says the flexibility of the Mark Logic solution has Bowker already contemplating next steps; steps that will include making the most of full book content. They plan to use the technology “as a tool to mine the content and create tools to come up with ways to sell relevancy,” says Heinzelman. One of the problems he sees with many search engines is “you conduct a search and get pages and pages of results. We want to reduce the amount of time people need to look. We think we can do that by mining content and selling information and metadata around relevancy; and with the sales data that we have, be able to tie that into it too.” Heinzelman says the Verity platform solution did not provide the flexibility to allow for full book content. As long as Bowker can collect such data, the Mark Logic server will be able to handle it. Heinzelman hopes that Bowker will have a product offering based around the full book content by the end of the year. Heinzelman adds that he expects the company to be completely XML-based by 2010.

MARI MCCLORE (MARIMCCLORE@SBCGLOBAL.NET) IS A FREELANCE WRITER BASED IN CONNECTICUT.

COMMENTS? EMAIL LETTERS TO THE EDITOR TO ECLETTERS@INFOTODAY.COM.