MarkLogic®

Innovative Publisher Cuts Costs and Improves Time to Market

American Psychological Association replaces open source with MarkLogic to create subsecond search results and reduce server count from 10 to 3

AMERICAN PSYCHOLOGICAL ASSOCIATION

The American Psychological Association (APA) was founded in 1892 to advance the creation, communication, and application of psychological knowledge to benefit society and improve people's lives. The APA has 150,000 members and 54 divisions in the subfields of psychology. One of the association's primary objectives has been the exchange of scientific information. To that end, the Office of Publications and Databases, a major division of the APA, publishes an array of scholarly journals, dozens of books annually, and several academic databases including PsycINFO and PsycARTICLES.

Summary

Customer
American Psychological Association

Industry
Publishing

Challenge
Slow content delivery, inconsistent search speed, and challenges in maintaining Lucene

Customer Use Case
Online Content Publishing Portal

Benefits
Rapid content delivery, subsecond response times, streamlined architecture, and more effective use of resources

Highlights

- Founded in 1892, APA is a major academic publisher of journals and books.

- APA moved its content to MarkLogic.

- APA replaced Lucene with MarkLogic to speed up content delivery and search results.

- APA developed subsecond search results across complicated datasets and now pushes new content live in a few minutes instead of two days.

- APA reduced costs and increased business opportunities as a result of the search replacement.

Challenge

As the premier provider of information in the discipline of psychology, APA's website gets more than one million hits a day. In 2008, APA moved its content from a legacy system to MarkLogic. Currently, the single operational database stores a complicated collection of Big Data including 160,000 full journal articles, 3 million abstracts, and 54 million cited references dating back as far as the late 19th century. The most challenging aspect of leveraging this content is the variety and complexity of the documents. The articles often have ambiguous connections that must be linked together for researchers to properly perform their jobs.

While APA acknowledges Lucene is a popular open source search engine and a good technology, there were several challenges in using it. One challenge was that Lucene was simply an additional layer of architecture which required additional code, infrastructure, and deployment tasks – none of which are necessary with an all-in-one database and search solution. In addition, search response times were inconsistent which created a disappointing user experience. However, the biggest challenge was in the timely delivery of content to APA users. With the Lucene-based implementation, depending on the size and complexity of the content package, it could take up to 48 hours to prepare, deliver, and post content on APA websites.

Case Study

Since APA is the authoritative source for this content, timely online availability is important to the authors, readers, and business partners who depend on it. The situation was unacceptable to Beverly Jamison, senior director of IT Architecture and Publishing Solutions, "Search has a measurable impact on how our customers use our digital properties. The business benefits of having a product that goes to market faster are quite significant."

Solution

Jamison needed a solution that could quickly deliver. The APA development team would have to invest significant time and resources to tune, test, and retune Lucene in order to improve performance, but even then the improvement would not be as thorough and sustainable as what APA would have with MarkLogic.

APA didn't have to look far. Already familiar with the speed and flexibility of MarkLogic as a database, Jamison knew several major companies relied on MarkLogic for search and analysis as well. Beyond that, she had come to believe in MarkLogic and its community, "The culture really helped to back me up. I have never met a bad engineer from MarkLogic."

Early proofs of concept showed MarkLogic would quickly alleviate APA's pain points. Query speeds consistently hit under one second in testing. Less than 4 months later, APA had completely moved search from Lucene to MarkLogic with no downtime. Customers are already praising the faster response times. Index maintenance is now a small portion of one person's job which has freed up two developers to focus on new features and performance improvements instead of keeping up a legacy system. That enabled the team to add spell suggest and typeahead features just a few weeks after deploying the search. The APA also was also able to reduce the server count from ten to three.

The biggest impact is simply the benefit of faster, more reliable searches. Because 70% of APA's revenue stream comes from publishing, a "first to information" advantage is crucial. MarkLogic reduced the turnaround time for new content from up to two days to a few minutes and cut search response times to less than a second. In Jamison's words, "Time to market has a huge impact when we talk to potential customers or business partners. With MarkLogic, it is taking one hour to stage content, a few minutes to push live, and less than a second to run queries and analysis. This is a tremendous coup for our business as well as our technology."

Next Steps

Jamison says MarkLogic's architecture and the new infrastructure at APA has continuing benefits. Foremost, the development team has newfound freedom to build and test new applications and front-end features in as little as a few weeks. The APA is now focusing on adding semantics services for customers and building new visualization options. It is also continuously simplifying its architecture as it takes down legacy systems one at a time to create better book and journal workflows.

Conclusion

By switching from Lucene to MarkLogic for search, APA not only met its initial goal of getting content out to customers more quickly, it also cut maintenance and hardware costs, and is consistently delivering subsecond search response. The end result is an edge in negotiations for new business as well as a more adept technology infrastructure in tune with the innovative work of the APA.