[Note: to write this success story, I interviewed Associate Publisher and Director John Sack of Highwire Press, a division of Stanford University Libraries.]

Highwire Press

Innovative, not-for-profit online publisher increases researchers' click-thrus on search by 40 percent

Goal: To help researchers find specific information among 15 million+ articles.
Highwire Press, a division of Stanford University Libraries, is always in search of new, cost-effective ways to communicate scholarly knowledge to researchers. Serving what associate publisher and director John Sack calls the crème de la crème in academic publishing, Highwire hosts websites and serves content for not-for-profit societies, university presses, and scholarly publishers of respected peer-reviewed journals in numerous fields. The organization hosts, for example, five of the six top medical journals, including the Journal of the American Medical Association and the New England Journal of Medicine, as well as 100 of the top 200 most frequently cited journals in the world.

Today, Highwire serves its customers by hosting more than 15 million articles, nearly one million of them free and in full-text from close to 800 hosted journals. Highwire, by necessity, also serves its customers' customers as well: Science, medical, and technology researchers whose primary objective is to quickly locate specific information on specialized topics.

Online scholarly search ramps up; stresses on search technology do, too.
Early in 2001, Sack and his team began to observe several issues that, left unaddressed, would hinder Highwire's mission to distribute its stores of scholarly knowledge to academic and research communities. First, the not-for-profit publisher began adding content at a very fast rate—not just new content, but also back

Quick Read

Clustered search results provide more relevant information for greater numbers of users as measured by a 40-percent increase in the search result clickthrough rate.

- Highwire Press, a division of Stanford University Libraries, hosts 15 million+ articles from nearly 800 journals including the New England Journal of Medicine.
- As more journals, many a century old, moved back content online, the quantity of content available to online researchers exploded. People began to browse less and search more to see if a topic in a current issue had been covered before.
- Searches were being performed by a new breed of researchers—undergraduates, journalists—who, lacking fluency in the language of their field, tended to perform less-specific searches and to end up with irrelevant search results.
- Highwire installed the Vivisimo clustering engine on its site and on publishers' sites to cluster researchers' search results on-the-fly, at the time of their queries.
- Using clustered results, organized in categories, researchers clicked on 40 percent more of the links returned by their search queries.

content from existing Highwire-hosted journals. "The journals we work with are 100 and 200 years old," explains Sack. "So when you put all their back content online you suddenly go from having 50,000 articles online to having 500,000 articles online—and the information retrieval problem of finding a needle in a haystack becomes worse and worse."

"When you put all their back content online you suddenly go from having 50,000 articles online to having 500,000 articles online—and the information retrieval problem of finding a needle in a haystack becomes worse and worse."

Sack and his team also became aware that increasing numbers of less specific search queries were being performed by a new breed of researchers: General scientists, undergraduates, professionals in training, and individuals within the social sciences and humanities. "The general scientists and undergrads naturally form less exact queries because they're not yet fluent in the language of the field they're researching," says Sack. "We also started working with more and more social sciences and humanities publishers. Because the languages of those fields are less exacting than are the languages of the sciences, we knew that people were going to be getting search results with too many mixed topics," he explains. "We needed a tool that would summarize the search results for them."

If those two issues were not enough to convince Sack and his team to innovate and head off user satisfaction issues before they became problematic, people had also shifted how they interacted with and used journal websites. "Typically, in the latter half of the 90s, researchers were using online journal sites to look at the latest issue of a journal, so there really wasn't any search involved," says Sack. "But as the amount of online content grew, researchers started saying 'I need to know whether this journal has ever published an article on such and such a topic,' causing them to shift from browsing to searching."

"Researchers were getting larger and larger search results, but without any more time to go through them than they had to begin with."

This combination of issues led Sack and his team to seek change. "We realized that researchers were getting larger and larger search results, but without any more time to go through them than they had to begin with," he says. "We wanted to get ahead of a customer satisfaction problem."

The search for a dynamic search categorization solution begins.

Sack and his team had been looking for solutions that fell under the headings of content searching and search categorization. "We were looking for something like Northern Light had years ago…those little blue folders that categorized search results based on topics," Sack says. "We actually did look at and even purchased a tool that did something like that, but it required us to do a lot of front-end work. Then we learned about Vivisimo's clustering engine, and found that it quickly performed dynamic rather than static categorization."

HighWire Press also runs two different kinds of sites: one is a portal in which all content is brought together in one search engine; the other comprises individual sites for each journal it hosts. “We wanted a tool our publishers, all 800 of them, could put on their individual journal sites as well,” says Sack. “Vivisimo was willing to work with us to develop a licensing model we could offer our publishers.”

Today, diverse researchers perform unique, specific searches with ease.
Highwire implemented Vivisimo’s clustering engine that builds upon an existing taxonomy and uses the taxonomy structure to organize search results on-the-fly, at the time of a researcher’s query. Now when researchers search at either Highwire’s portal or through its publishers’ sites, retrieved documents are clustered based upon common themes and content within them, and dynamically labeled with human-like accuracy. Researchers use clustered results to quickly drill down into the topics that are most relevant to them.

“Because we’re working with a scientific and research-oriented audience, we’re typically working with people who want to be quite specific—like the difference between students and researchers,” Sack says. “Researchers are trying to write something that’s never been written before; students are learning to do research by copying something that’s already been done. That’s why Vivisimo works superbly well for researchers: it lets people very quickly say, ‘I’m not interested in these 50 things; I’m interested in that one there,” explains Sack.

Search satisfaction increases with a 40-percent increase in the clickthrough rate.
“We purchased Vivisimo’s clustering engine on the theory that solving the problems we saw mounting would increase customer satisfaction,” says Sack. “We wanted our customers’ customers to be more satisfied with their user experiences.”

To measure user satisfaction, Sack and his team sought to discover whether researchers were looking at more articles as the result of each search. “Although it was difficult to measure because of how our servers are arranged, we did learn that searchers were examining 40 percent more articles on sites with Vivisimo than on sites without,” Sack says. “And to us, that means success.”

“I don’t know of anyone doing search categorization better than Vivisimo,” Sacks says. “And better means better technologically and better organizationally. By organizationally, I mean that the product is supported very well by the people who built it. We’re a technical organization here at HighWire so we really require and appreciate it when the products we do purchase are supported by organizations like ours that really know what they’re doing,” he explains.

“If, for example, you throw a search query at the product and it comes back with unlikely categories, Vivisimo is interested in hearing about it and improving the product. They

believe in that kind of continuous improvement, and they produce new versions that allow us to solve problems. Vivisimo is an excellent organization to work with," says Sack, happy to have correctly anticipated the shifting search landscape.