

Stanford University Libraries Launches New Federated Search Prototype to Discover Scholarly Content from the “Deep Web”

Press Release

March 27, 2008

If you think everything you’re looking for can be found through Internet search engines, think again. The “crawl” technology of search engines like Google identifies and indexes only a fraction of all the information that is available on the Web. The “invisible web” is made up of thousands of databases and searchable sources that contain highly targeted and valuable information, and whose content is inaccessible and therefore neither seen nor indexed by traditional search engines.

These online resources are often the bread and butter for today’s scholars: news archives, e-journal aggregators, indexing services, proprietary databases, subscription services are generally invisible to the casual internet search. As an example, Stanford University Libraries and Academic Information Resources (SULAIR) subscribes to thousands of scholarly journals and hundreds of licensed databases. And while these represent a trove of valuable resources, scholars have traditionally had no easy mechanism to search across multiple sources simultaneously.

For Michael Keller, University Librarian, giving scholars effective tools to discover digital information is an over-riding priority: “The digital revolution and the explosion in the number of relevant online resources are fantastic opportunities, but also a great challenge. Stanford students and faculty need a way to reduce the amount of time they spend searching, and to make it easier for them to find precisely relevant information on multidisciplinary research topics.”

To address these needs, SULAIR has partnered with Deep Web Technologies (www.deepwebtech.com/) to release a prototype federated search service tailored to the environment of an academic research library. A federated search simultaneously queries a number of online resources held across many different, isolated systems and returns the results in one merged list. This gives scholars a quick and broad view across a number of possible sources, letting them identify the most promising areas and resources in which to extend their search.

Three demonstrations of federated search within the Stanford environment are now online at <http://library.stanford.edu/rapids/fedsearch.html>:

1. **All Library Catalogs at Stanford** (Socrates, Lane Medical, Jackson Business, Stanford Linear Accelerator, and the Health Library of the Stanford Hospitals and Clinics). (*Openly accessible*)
2. **“Top 10” Databases at Stanford** (ABI/Inform, Annual Reviews, Biosis, Dissertations & Theses-A&I, Engineering Village, Expanded Academic ASAP, Lexis Nexis Academic, Periodicals Archive Online, PsycINFO, and Web of Science). (*Limited to current Stanford students, faculty, and staff*)
3. **Stanford Digitized Content** available from collections.stanford.edu (*Openly accessible*)

Abe Lederman, president and founder of Deep Web Technologies, said that: “I’m very excited to be working with Stanford on this project. Deep Web Technologies has been very successful in building search tools for the science communities and national research communities, and the opportunity to work with SULAIR lets us explore how we can extend our services into one of the world’s leading academic research libraries.”

Unlike a standard search engine, Deep Web Technologies' federated search product, Explorit Research Accelerator, does not rely upon a stored index built in advance. Instead, it operates in real time, replicating the query and broadcasting into multiple databases. Lederman explains: "The Deep Web search engine immediately reaches out to relevant databases at various sites, drilling down into these information centers all at once, organizing the info and merging the results in a ranked priority list—in real time. So, not only are you delving into databases, you're getting the most precise and current results."

Employing a familiar, intuitive search interface, Explorit quickly presents high-value results from a common, single point of access. Explorit is currently powering a number of science, technology and government search portals, including National Digital Library for Agriculture, Science.gov, Scitopia, and WorldWideScience.org. Explorit is also being used for federated search at Intel's corporate library.

Links for More Information

RaPIDs Group, Stanford University Libraries & Academic Information Resources
<http://library.stanford.edu/rapids/>

Deep Web Technologies
<http://deepwebtech.com/>

Whitepaper: "How to Maximize Your Strategic Investment in Federated Search"
<http://www.deepwebtech.com/PDFs/Whitepaper.pdf>

Federated Search Blog
<http://federatedsearchblog.com/>

Combined Search Prototypes Developed by Deep Web Technologies and SULAIR:

All Library Catalogs at Stanford
<http://deepweb.stanford.edu/catalogs/search.html>

"Top 10" Databases at Stanford
<https://deepweb.stanford.edu/su/>

Stanford Digitized Content
<http://deepweb.stanford.edu/digcolls/search.html>

Other Combined Search Sites Developed by Deep Web Technologies:

National Digital Library for Agriculture (NDLA)
<http://www.nal.usda.gov/ndla>

Science.gov (Gateway to U.S. Government Information, incl. R&D Results)
<http://www.science.gov>

Scitopia (Portal containing documents from 15 sci-tech societies plus patents and government data)
<http://www.scitopia.org>

WorldWideScience.org (Global Science portal developed by U.S. Dept of Energy and British Library)
<http://worldwidescience.org/>

Information Outlook: Federated Search at the Intel Library (article available courtesy of SLA)
<http://www.deepwebtech.com/PDFs/SLA-Hill-Intel-0710.pdf>