Issues and Opportunities Associated with Federated Searching

Grace Baysinger (graceb@stanford.edu)
Head Librarian & Bibliographer
Swain Chemistry & Chemical Engineering Library
Stanford University Libraries and Academic Information Resources (SULAIR)

Library catalogs and databases contain a wealth of information that is not available to Internet search engines such as Google. It can be difficult for users to identify which research tools to use and time-consuming for them to search each resource one at a time. Federated search tools make it possible to search multiple resources with one query. Several strategies have been developed to provide "one-stop shopping" but those dealing with multiple search interfaces face the biggest challenges. This talk will describe a project underway to develop federated searching prototypes on campus and will cover the viability of providing federated search services as well as the interest level by students and faculty in using them.
In May of 2007, SULAIR launched a *skunkworks* effort to push the envelope of the Libraries’ information discovery tools, services and environments. This effort is known as RaPIDS, for Rapid Prototyping of Intuitive Discovery at Stanford.

The primary objectives of the RaPIDS effort are to:

- Illustrate the features and functions of a cybrary, and
- Increase access to library resources.

This will be accomplished by:

- Rapidly deploying prototypes and gathering feedback, generating demonstrable results and tangible benefits, quickly,
- Focusing on delivering systems that meet patrons’ needs, and
- Helping catalyze the development of new library services.
The effort will be led by representatives from each of the major units of SULAIR: Collections & Services, Technical Services and Academic Computing/Digital Library Systems & Services.

- Grace Baysinger, Head Librarian & Bibliographer, Swain Chemistry & Chemical Engineering Library
- Tom Cramer, Associate Director, Digital Library Systems and Services
- Adan Griego, Curator for Latin American & Chicana/o Studies
- Philip Scheur, Head, Cataloging and Metadata Services

This core team will enlist colleagues throughout SULAIR for their expertise, ideas, content, technical know-how, and reviews.
RaPIDS Initial Projects

- Engaged in a sustained process of experimentation and progress, rather than the implementation of any single product.

- Examples of initiatives planned and underway:
  - Federated search
  - Associative searching
  - Taxonomic browsing

- RaPIDS page on Federated Search
  http://library.stanford.edu/rapids/fedsearch.html
Federated Search

- Ability to simultaneously search multiple resources in real time and pool results in one merged relevancy ranked list.

- Gives scholars a broad view of disparate resources held across many different, isolated systems.

- SULAIR is collaborating with Deep Web Technologies. DWT’s Explorit Research Accelerator federated search engine is used in a number of science, technology and government search portals.
Other Portals Using Explorit
Why Use Federated Tools?

- Users don’t need to have prior knowledge about resources
- Help cope with lack of standardization in interfaces and search protocols for similar tasks
- Eliminates need to repeat same search multiple times
- Reduces time needed to find information
- Way to help users discovery high quality resources acquired by the library and to increase use of them
Three Demonstrations of Federated Search within Stanford Environment

Combined Search Engines Prototypes

Because Google is Just the Tip of the Iceberg, Explore the Depths
Discover Treasures in the Stacks
Encounter the Unexpected

“Top 10” Databases
All Library Catalogs at Stanford
Locally Digitized Collections

“Top 10” Databases
https://deepweb.stanford.edu/cn/

All Library Catalogs at Stanford University
https://deepweb.stanford.edu/catalogs/search.html

Locally Digitized Collections
http://deepweb.stanford.edu/digcolls/search.html

Photo Credit:
christia.n’s photos – Iceberg
http://www.flickr.com/photos/christia.n/terryculley’s photostream – Diver
http://www.flickr.com/photos/7724539@N03/254581229/
decomn – Jelly-Fish
http://www.flickr.com/photos/decomn/497317867/
“Top 10” Databases

RaPIDS Federated Search Prototype - Top 10 Databases

- Select All
- All/Inform Global
- Annual Reviews
- BIOSIS
- Dissertations & Theses - A & I
- Engineering Village
- Expanded Academic ASAP
- Lexis Nexis Academic (News)
- Periodicals Archive Online
- PsychINFO
- Web of Science

© Stanford University. Stanford, CA 94305. (650) 723-2300. Terms of Use | Copyright | Complaints
All Library Catalogs at Stanford
Combined Search of Stanford Digitized Content
Rapid Retrieval, View Some Results While Search is Running
Sort Results – Rank, Source, Date, Title, Author
Ability to Select A Source
Brain Organization for Music Processing

Isabelle Peretz and Robert J. Zatorre

Department of Psychology, University of Montreal, Montreal, Quebec H3C 3J7, email: Isabelle.Peretz@umontreal.ca Montreal Neurological Institute, McGill University, Montreal, Quebec H3A 2B4, email: Robert.Zatorre@McGill.ca

Research on how the brain processes music is emerging as a rich and stimulating area of investigation of perception, memory, emotion, and performance. Results emanating from both lesion studies and neuroimaging techniques are reviewed and integrated for each of these musical functions. We focus our attention on the common core of musical abilities shared by musicians and nonmusicians alike. Hence, the effect of musical training on brain plasticity is examined in a separate section, after a review of the available data regarding musical playing and reading skills that are typically cultivated by musicians. Finally, we address a currently debated issue regarding the putative existence of music-specific neural networks. Unfortunately, due to scarcity of research on the macrostructure of music organization and on cultural differences, the musical material under focus is at the level of the musical phrase, as typically used in Western popular music.

Full Text  PDF
Help for “Top 10” Databases


Enter Search Terms

<table>
<thead>
<tr>
<th>Want to Search</th>
<th>Description</th>
<th>Example</th>
<th>Search Results Will Include</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case sensitive</td>
<td>Please note casesensitive, one type in uppercase or lowercase letters</td>
<td>RAIN falls</td>
<td>Search results will be identical</td>
</tr>
<tr>
<td>Wildcard *</td>
<td>Anything used to match the same character except alphanumeric characters or punctuation</td>
<td>Author search on name* Zoo R OR Zoo K OR Zoo Elephant OR Cheerleader</td>
<td></td>
</tr>
<tr>
<td>Wildcard ?</td>
<td>Anything used to match the same character including alphanumeric characters or punctuation</td>
<td>DOG? OR DOG OR DOG? DOG</td>
<td></td>
</tr>
<tr>
<td>Exact Phrase</td>
<td>Use double quotation marks around terms</td>
<td>“zoo animals”</td>
<td>(Lost exact phrase)</td>
</tr>
<tr>
<td>Multiple terms</td>
<td>When only exact terms used on highlighted keywords, AND is assumed (all terms must be found)</td>
<td>global warming</td>
<td>Global financial markets are warming up</td>
</tr>
<tr>
<td>Boolean Operators</td>
<td>Use AND, OR, or NOT (AND, OR, and NOT terms must be present. Otherwise none present. NOT (within term present)</td>
<td>Global warming AND disasters</td>
<td>Climate is warmer in Alaska due to global changes in the environment</td>
</tr>
<tr>
<td>Parenthesis</td>
<td>Use in groups when using OR as well as AND (parenthesis within parenthesis)</td>
<td>earthquake prediction</td>
<td>Earthquakes predicted according to new earthquake models from new models</td>
</tr>
</tbody>
</table>

Performing Search and Viewing Results

Search

Press “Search” button to search all databases at once.

View More Results

You can view more results while search is still running. To finish search press “View More Results” button.

Results are ranked by relevance. The more stars next to an item, the higher the rank.

Relevance by relevance is determined. To view results in different order, use the pull-down menu.

Search allows you to view items from a specific database. Use “Find” to go to a specific item. There is reference chronological order or most current items are displayed first.

Title order ensures results alphabetically by title.

Full Text Access

Click on title to full text record in database and click to link to full text. Due to design of individual databases, sometimes you have to click on permanent link to right of citation to get to full record and full text.

Session Preferences

After doing an initial search, you may set session “Preferences” on how many results you want per page.

Maximum Number of Records Retrieved Press on Database | Use to limiting retrieval, up to 100 records are retrieved from each database.

If you need to do a more comprehensive search, then search individual databases instead.
# Help “Top 10” Databases

## Combined Search of “Top 10” Databases Data - Description of Databases

https://deepweb.stanford.edu/safe

<table>
<thead>
<tr>
<th>Database</th>
<th>Description</th>
<th>Last Updated</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABI/Inform Global</td>
<td>Provides coverage of Business and Management, including global business company histories, competitive intelligence, and new product development. Over 3,900 journals, 100,000+ articles.</td>
<td>Daily</td>
</tr>
<tr>
<td>Annual Reviews</td>
<td>Publishes authoritative, analytic reviews in 35 focused disciplines within the Business, Physical, and Social Sciences. Annual Reviews publications are among the most highly cited in their respective disciplines. Some articles are published online as soon as they are copy-edited and typeset.</td>
<td>Daily</td>
</tr>
<tr>
<td>Descriptors &amp; Thesaurus (Abstracted &amp; Indexes 1861-2008)</td>
<td>Contains citations to virtually every American dissertation accepted at an accredited institution since 1861. Selected Masters Theses have been included since 1922 and those from foreign countries since 1941. Also includes theses from the United States since 1990 to present. 1.32 million records as of Jan 2008. Updated Monthly.</td>
<td>Daily</td>
</tr>
<tr>
<td>Engineering Village (1884)</td>
<td>The most comprehensive source of Engineering Village is Computer. The online version of Engineering Village includes the most comprehensive bibliographic database of engineering research available today, covering more than 1,000 engineering journals, conferences, and technical reports. Over 9.7 million records as of Jul 2007. Updated Weekly.</td>
<td>Daily</td>
</tr>
<tr>
<td>Expanded Academic ASAP (1990-2023)</td>
<td>Selection includes core Engineering Village (Computer). The online version of Expanded Academic ASAP is a comprehensive full-text database of engineering research available today, covering more than 1,000 engineering journals, conferences, and technical reports. Over 9.7 million records as of Jul 2007. Updated Weekly.</td>
<td>Daily</td>
</tr>
<tr>
<td>Lexis-Nexis Academic (Databases Covered Vary)</td>
<td>This service provides full-text discovery from over 10,000 news, business, legal, medical, and reference publications in a variety of search options. Restrictions include the major legal databases, business databases, international news, and peer-reviewed English language resources. U.S. Federal and state case law, codes, regulations, legal news, law reviews, and international legal information. Updated Weekly. Contains over 8 million records as of Jul 2007.</td>
<td>Daily</td>
</tr>
<tr>
<td>Periodicals Archive Online (1665-1999)</td>
<td>An archive of digital, full-text scholarly journal articles in the humanities and social sciences, this resource provides access to important scholarly literature from the humanities and social sciences disciplines from 1665 to 1995. Currently contains 3.8 million articles from over 500 journals with links to 200,000 full-text articles. Almost 50% of the journals are in non-English languages. Over 8.5 million records as of Mar 2008. Updated quarterly.</td>
<td>Daily</td>
</tr>
<tr>
<td>PsycINFO (1807-2023)</td>
<td>The PsycINFO database provides access to the international literature in psychology and related behavioral and social sciences, including psychology, psychiatry, anthropology, education, pharmacology, and linguistics. PsycINFO contains citations and abstracts for over 600,000 items, books, book chapters, reports, and dissertations. Journal material represents informative articles selected on the basis of relevance to psychology from over 2,700 journals published throughout the world in over 25 languages. Over 1.1 million records as of Jul 2007. Updated Weekly.</td>
<td>Daily</td>
</tr>
<tr>
<td>Web of Science (Science 1900-, Social Science 1973-, Arts &amp; Humanities 1976-)</td>
<td>One world at its source to quickly and interactively search, search, search, and search. Web of Science also provides a unique search method, cited reference searching. A cited reference search allows you to find articles that cite a previously published work. In addition to cited reference searching, you can search these databases by topic, author, source title, and address. Over 20 million records as of Oct 2005. Updated Weekly.</td>
<td>Daily</td>
</tr>
</tbody>
</table>
Federated Search Prototype for the “Top 10” Databases

Database Descriptions Plus Subject Tags

**ARL/Inform Global (1923-)**

**JSTOR**
JSTOR is a digital library that provides access to academic journals, periodicals, and collections in the humanities, social sciences, and arts.

**Academic Search Complete (1950-)**
Academic Search Complete includes full text for more than 1,900 journals and other serials covering a wide range of disciplines.

**Annual Reviews (1952-)**
Annual Reviews publications provide authoritative, comprehensive reviews in 53 focused disciplines within the Biomedical, Physical, and Social Sciences. Annual Reviews publications are among the most highly cited in scientific literature. Size: 27,578 as of October 15, 2007. Site is updated monthly; articles in some series are published online as soon as they are copy-edited and typeset.

**BioOne**
BioOne is the major English-language service providing comprehensive worldwide coverage of research in the biological and biomedical sciences. It includes access to original research from nearly 150 primary journals plus meeting abstracts, reviews, book chapters, notes, letters, and selected reports. U.S. patents are included for 1936-1995, 1986-1995, and 1994 to the present. Size: 18 million records as of Oct 2007. Updated: Weekly.

**Dissertations & Theses (abstracts & Indexes 1861- ; Stanford Full-Text 1989-)**
Includes citations to virtually every American dissertation accepted at an accredited institution since 1861. Selected Masters Theses have been included since 1982 and those from foreign countries since 1988. Also includes full-text access for Stanford Theses from 1989 to present. Size: 2.82 million records as of Jan 2006. Updated: Monthly.

**ProQuest**
ProQuest provides access to a wide range of databases covering all areas of study, including business, economics, education, engineering, fine arts and music, geography and regional planning, genetics, health sciences, history and political science, language and literature, library and information science, mathematics and statistics, philosophy and religion, physics, psychology, and sociology.
Summary Description & Features

- Single search to heterogeneous set of resources
- Federated search is a discovery tool, not for comprehensive searching
- Results in real time from resources, not snapshots
- Fast retrieval and display of some results while search is still being completed
- One ranked list after merging results
  - Up to 100 records per resource
  - Ability to sort several ways, including source
- Link to full record from resource
  - “Permanent Link” is cache of full display
- Change preferences after doing an initial search
Technical Challenges in Using Federated Search Tools

- Speed and performance of other federated search engines
- Rapidly growing number of digital resources
- Type of searchable information in resources varies
- Resources w/ limited numbers of simultaneous users or require individual login before use
- Nearly constant change in interfaces, some changes in URLs
- Names of search fields and formatting of data in them are not standardized
- Sorting order varies for default display of results
- Number of records that can be displayed w/o violating license agreement or slowing search engine too much
- Need to retain “state of session”
- Clearing cache between sessions
Preliminary Feedback

- Almost all students really liked using federated search prototypes and found results to be relevant.

- Librarians want more precision for author searches, made suggestions for improving the interface. Much discussion of choice of “top 10” files. Mixed value of results from library catalogs prototype.

- Strategically, federated search is one of the directions that we want to pursue.
Future Plans

- Offer more resources via federated search
  - eBook & other full-text resources
  - Abstract & Indexing Databases
  - Special Collections Finding Aids
  - Reference resources & web sites

- Implement new interface offered by DWT
  - More search fields
  - Faceted search results
  - Exporting tagged results
  - SFX linking
  - Alerts
Acknowledgements

- Stanford: RaPIDS Members - Tom Cramer, Phil Schreur, Adan Griego; other staff who helped implement the prototypes; plus the many staff and students at Stanford who tested and provided feedback.

- Deep Web Technologies: Abe Lederman, Sol Lederman, Brian Desplain, and other members of the DWT team.