metasearching at NC State

Kristin Antelman

DLF April 20, 2004
MultiSearch service developed using BlueAngel MetaStar APIs and local JSP/perl scripting
Released August 2002

Two interfaces offered: detailed subject groupings ...
and broad subject groupings w/option to select across groups

<table>
<thead>
<tr>
<th>Agriculture &amp; Life Sciences</th>
<th>Humanities &amp; Social Sciences</th>
<th>Physical Sciences &amp; Engineering</th>
<th>General</th>
</tr>
</thead>
<tbody>
<tr>
<td>Select All</td>
<td>Select All</td>
<td>Select All</td>
<td>Select All</td>
</tr>
<tr>
<td>AIDS and Cancer Research Abstracts</td>
<td>ATLA Religion Database</td>
<td>Aluminium Industry Abstracts</td>
<td>Academic Search Elite</td>
</tr>
<tr>
<td>ASFA Aquaculture</td>
<td>Anthropological Literature</td>
<td>Applied Science and Technology Index</td>
<td>Expanded Academic ASAP</td>
</tr>
<tr>
<td>ASFA Marine Biotechnology</td>
<td>Art Index</td>
<td>Ceramic Abstracts/World</td>
<td>Funk &amp; Wagnalls New World Encyclopedia</td>
</tr>
<tr>
<td>ASFA Aquatic Sciences and Fisheries Abstracts</td>
<td>Art Index Retrospective</td>
<td>Ceramics Abstracts</td>
<td>General Reference Center Gold</td>
</tr>
<tr>
<td>Agriculture</td>
<td>Business source elite</td>
<td>Computer and Information Systems Abstracts</td>
<td>General Science Full Text</td>
</tr>
<tr>
<td>Agricultural and Environmental Biotechnology Abstracts</td>
<td>Communication &amp; Mass Media Complete</td>
<td>Conference Papers Index</td>
<td>InfoTrac OneFile</td>
</tr>
<tr>
<td>Algae, Mycology and Protozoology Abstracts</td>
<td>Dissertation Abstracts</td>
<td>Corrosion Abstracts</td>
<td>MasterFILE Premier</td>
</tr>
<tr>
<td>Animal Behavior Abstracts</td>
<td>ERIC</td>
<td>Dissertation Abstracts</td>
<td>Newspaper Source</td>
</tr>
<tr>
<td>Bacteriology Abstracts</td>
<td>Ecislt</td>
<td>EIS: Digests of Environmental Impact Statements</td>
<td>OmniFile Full Text Mega</td>
</tr>
<tr>
<td></td>
<td>Findex</td>
<td></td>
<td>Readers Guide</td>
</tr>
<tr>
<td></td>
<td>GPO monthly catalog, FastSearch database</td>
<td></td>
<td>Readers' Guide Retrospective</td>
</tr>
</tbody>
</table>

*Note: Not all databases are available on this page. Access additional databases through the subject and alphabetical lists. (More about MultiSearch)*
what we like about it

the ability to customize and post-process results
-- create OpenURLs for citations from databases that are not OpenURL sources
-- embed results in webpages

cross-domain searching
-- support for various metadata schema
**Concerns**

*Collection:* many core databases are not cross-searchable

*Technical:* implementation is expensive
-- poor documentation
-- inconsistent vendor implementations of Z39.50

*Usability:* does the tool meet user needs?
-- are users successful in interacting with it?
-- does it lead users to relevant articles?
key questions

-- what information needs are best met by metasearching?

-- is cross-searching more databases better?

-- which is the better place for a user to start, the smaller or larger result set?
log files

questions:

Is it used?

MultiSearch: ~7000 searches/mo.
Ebsco: ~ 25,000 searches/mo.
questions:

What type of searches do users do?

- any keyword (default): 76%
- subject kw: 9%
- author kw: 8%
- title kw: 5%
questions:

What do the searches look like?

log files

european integration
appreciation for the earth
Sunday v. Stratton Corporation 390 A.2d 398 (Vt. 1978)
assisted housing discrimination

networking
coyote eating habits
plato
television news

habituation of birds and human closeness
log files

questions:

How many terms do users enter?

<table>
<thead>
<tr>
<th>Terms</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10%</td>
</tr>
<tr>
<td>2</td>
<td>35%</td>
</tr>
<tr>
<td>3</td>
<td>25%</td>
</tr>
<tr>
<td>4</td>
<td>20%</td>
</tr>
<tr>
<td>5</td>
<td>10%</td>
</tr>
<tr>
<td>6</td>
<td>2%</td>
</tr>
<tr>
<td>7</td>
<td>1%</td>
</tr>
<tr>
<td>&gt;7</td>
<td>0%</td>
</tr>
</tbody>
</table>

mean=3.12; n=4329
questions:

How many hits are they getting?

log files

n=236; mean > 51,000; stdev > 51,000!
questions:

For large result sets, does it reduce the set size if the search is resent as keyword in subject?

*It can.*

- sets > 10,000: 62% fewer (0% nulls)
- sets > 350: 91% fewer (21% nulls)
- sets < 350: 81% fewer (49% nulls)

n=39, 72, 95
Tasks:
- think of a topic and do a search using MultiSearch
- look at some results and tell us if they “look good to you”

Ask them:
- was it easy or difficult to select the databases you wanted?
- which subject interface do you prefer?
- what did you think was happening?
questions:

Do they think the results are useful?

*Mostly yes.*

- almost anything “works”
- general level of confusion does not affect satisfaction with results
  (also it’s sometimes hard to tell “good” from “bad” in quickly scanning brief records)
questions:

Which is the better subject grouping?

Choose a subject to retrieve a list of databases.

- Select a subject

Accounting
African American Studies
Agricultural and Resource Economics
Agriculture
Animal Science
Anthropology and Archaeology
Art, Architecture and Design
Biochemistry
Biological and Agricultural Engineering
Biological Sciences
Botany
Business and Management
Chemical Engineering
Chemistry
Civil Engineering
Communication and Media
Computer Science
Crop Science
Economics

ncsu libraries
questions:
Which is the better subject grouping?

2. Select Databases
- [ ] Show All | Subject Area (#databases)
  - [ ] Agriculture & Life Sciences (19)
  - [ ] Design (6)
  - [ ] Education (3)
  - [ ] Engineering (11)
  - [ ] Humanities & Social Sciences (25)
  - [ ] Management (4)
  - [ ] Multidisciplinary/General (11)
  - [ ] Physical & Mathematical Sciences (8)
  - [ ] Textiles (4)
  - [ ] Veterinary Medicine (6)
questions:

Can users navigate the interface?
Do they understand what it is doing (or at least sort of)?

Many don’t. They think…
- the databases are subjects
- the databases search all types of materials
- they are searching the full text

Complex interfaces are intimidating
→ serial failure
next steps for NCSU

1. search for a new metasearch tool

2. improve the current interface and experiment with ways to lead searchers to the next step

3. explore other approaches to getting users to appropriate databases
   
e.g., send searches to local data source, such as Ingenta tables of contents