The Scholarly Database (SDB)

Nianli Ma, Russell Duhon, Katy Borner
for the Digital Library Brown Bag
Nov 5th, 2008

Data Sets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#Records</th>
<th>Years Coverage</th>
<th>updated</th>
<th>Restricted Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medline</td>
<td>16,053,495</td>
<td>1898-2008</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>PhysRev</td>
<td>398,005</td>
<td>1893-2006</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>PNAS</td>
<td>16,167</td>
<td>1997-2002</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>USPTO</td>
<td>3,710,952</td>
<td>1976-2007</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>NSF</td>
<td>174,835</td>
<td>1985-2003</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>NIH</td>
<td>1,043,804</td>
<td>1972-2002</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>21,456,336</td>
<td>1893-2008</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>
Records Per year

![Graph showing records per year](https://nwb.slis.indiana.edu/community/?n=ScientometricsDatasets.HomePage)

Detailed Information

https://nwb.slis.indiana.edu/community/?n=ScientometricsDatasets.HomePage
Challenges

Need to interlink
- Grants and papers/patents.
- Grants/papers/patents and their PIs/authors/inventors, etc.

Use resulting networks to
- Count #papers, #citations, etc.
- Determine strength of co-PI/author/inventor relations, etc.

SDB Architecture
SDB Architecture (continued)

- **Solr**: full-text search server
  - Open source
  - Uses the Lucene search library

- Interface developed in Django
  - Open source
  - Particularly suited for content-focused web applications

SDB Interface [http://sdb.slis.indiana.edu](http://sdb.slis.indiana.edu)
Number of Medline Publications

US Patent by Geo-location (U.S. only)
Future Work

- Provide full-text search on science news
- Add JCR impact factor to publications
- Discover user interests through what they search and download

References


Thanks

- You!
- National Science Foundation under grants nos. IIS-0238261 and IIS-0513650
- James S. McDonnell Foundation grant in the area of Studying Complex Systems