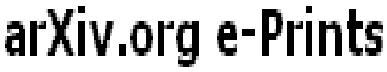
Science-Specific Search: Bridging the Gap between Dissemination & Access to Information

Presented by: Joris van Rossum, Head of Scirus
Event: IATUL 2007

Content

- How has content provision changed?
- How has information retrieval changed?
  - Lessons from Elsevier products
  - How can librarians improve information retrieval?
- Future Trends

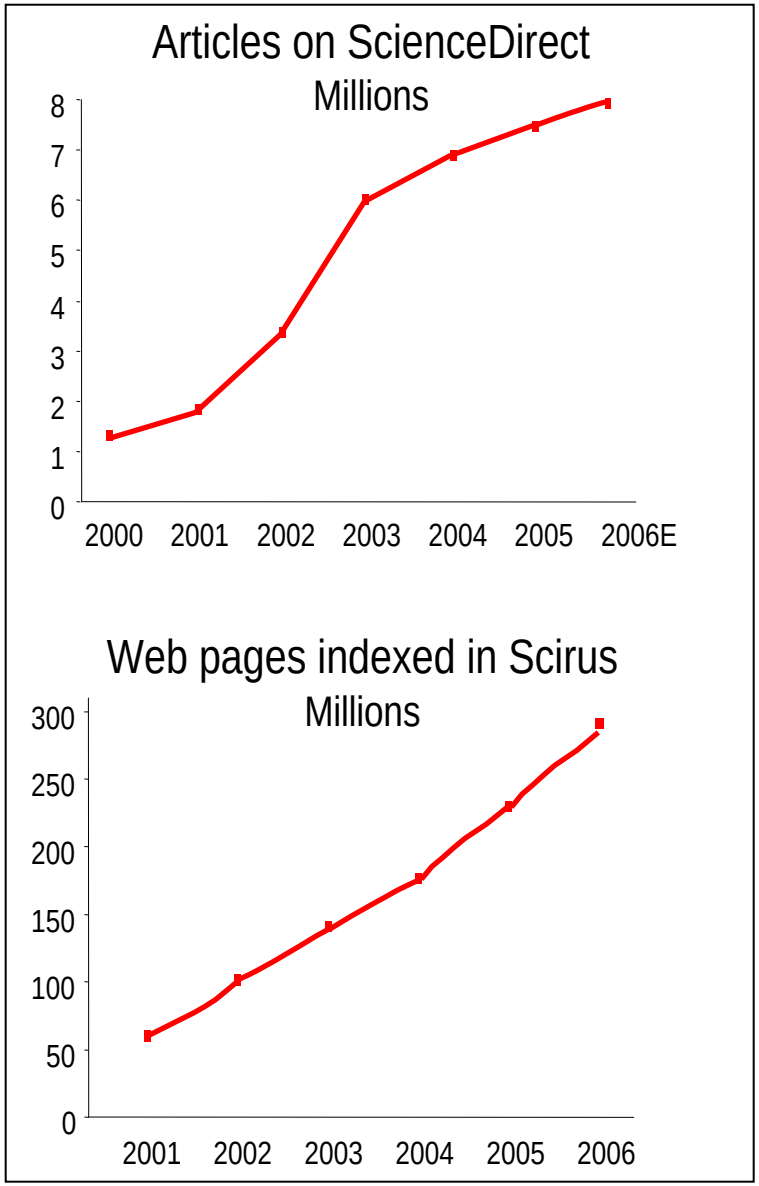
How did scientific content change with Web?



Internet has made journal publishing just one of many options

Increasing amount of content available on-line

- High amount of published content
  - Scopus has 30 million abstracts
  - ScienceDirect has 8 million articles
  - # of articles published per year by Elsevier increased from 160K in 2000 to 250K+ in 2005
- Amount of scholarly Web content even higher
  - Scirus currently indexes over 400 million scholarly Web pages
- Size of general Web has exploded
  - Aug '05, Yahoo indexed over 19 Billion pages
  - Google says it indexes 3x more than closest competitor*



*http://battellemedia.com/archives/001954.php

Different content discovery methods

- Browsing
- Linking
- Alerting
- Searching
- User collaboration/sharing

Covered in Following Slides

Different content discovery methods

- Browsing
- Linking
- Alerting
- Searching
- User collaboration/sharing

Covered in Following Slides

Browsing remains effective type of content discovery

- Journal browsing remains important for content discovery
  - Journal browsing is used to keep abreast of latest developments in subject area
- 31% of all full text article use on ScienceDirect is a result of journal browsing
- Users that start on a journal home page download, on average, 1.9 articles in a single session
- Users are provided with the option to list favourite journals and receive Journal Issue alerts (Table of Contents)

Different content discovery methods

- Browsing
- Linking
- Alerting
- Searching
- User collaboration/sharing

Covered in Following Slides

Linking articles

- Reference linking and ‘cited-by’ links are very effective content discovery methods
- Publishers are collaborating to ensure correct reference linking (CrossRef)
- 8% of all full text article use on ScienceDirect comes from reference linking. Same is expected from ‘cited-by’ links
- Next to reference and cited-by links in official literature there is
  - Web references and cited-bys
  - Patent references and cited-bys
  - Clustering
  - Author linking

Different content discovery methods

- Browsing
- Linking
- Alerting
- Searching
- User collaboration/sharing

Covered in Following Slides

Different content discovery methods

- Alerting
  - Journal Issue alerts (RSS)
  - Top articles alert
  - Citation alert
  - Search alert

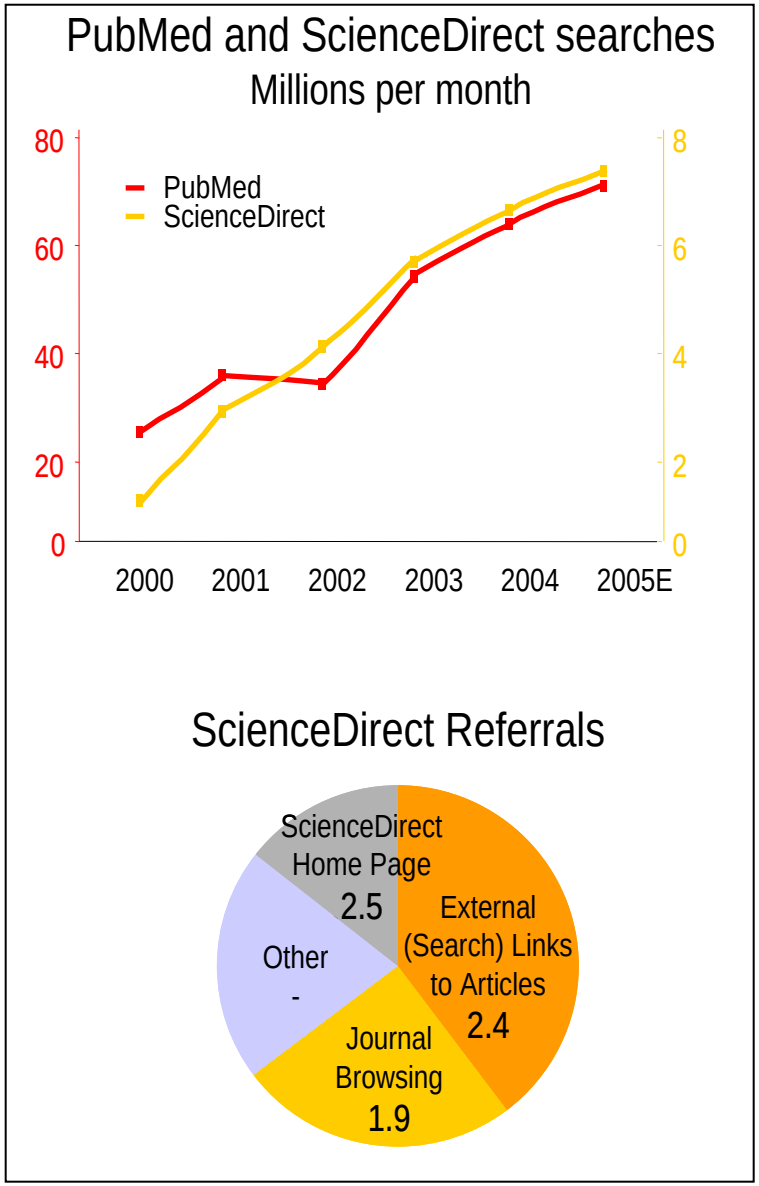
Different content discovery methods

- Browsing
- Linking
- Alerting
- Searching
- User collaboration/sharing

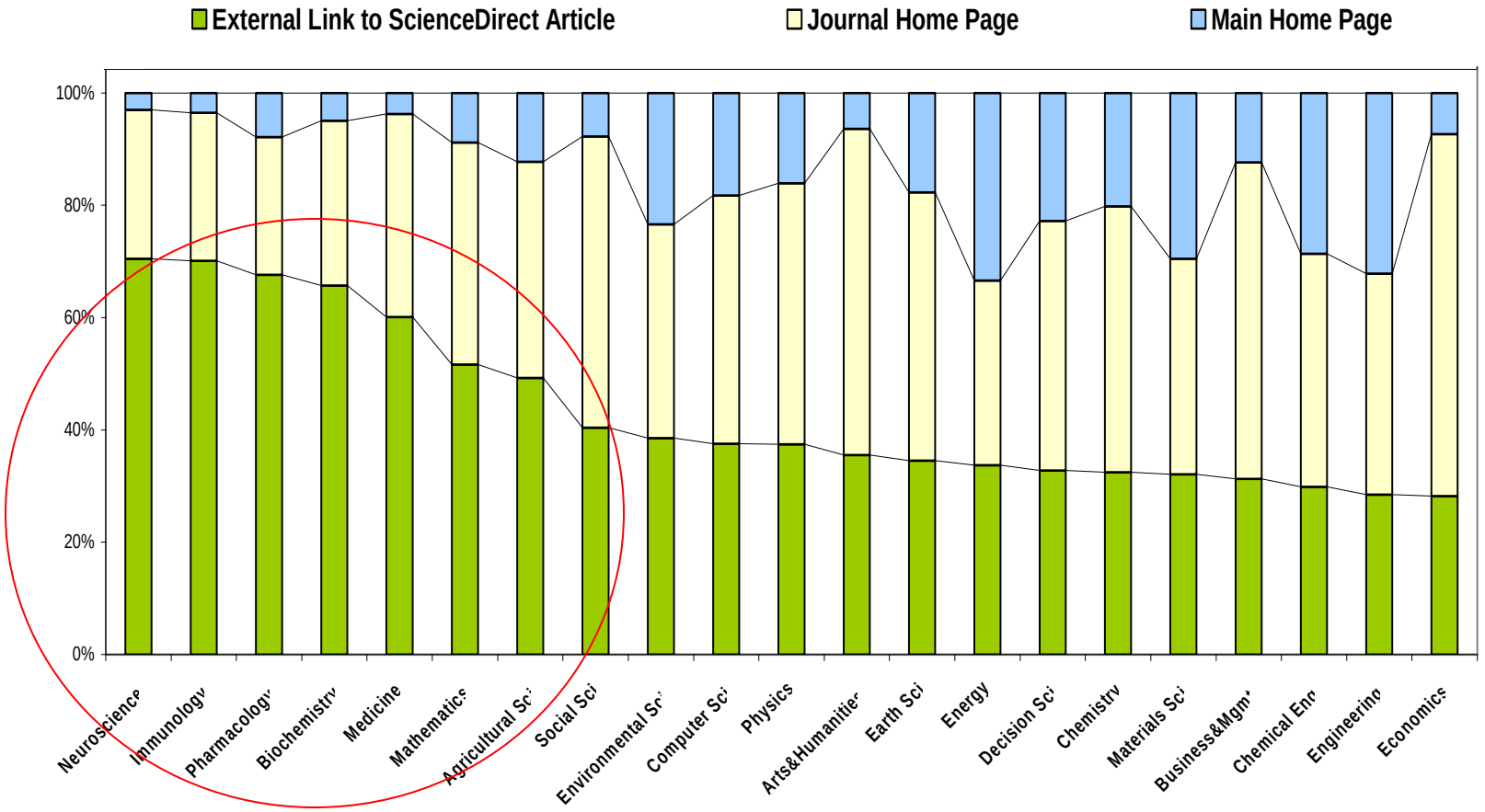
Covered in Following Slides

Search is main driver of journal article use

- Exponential growth of PubMed and ScienceDirect searches
  - Growth rates between 20 – 110% from 2001 to 2005
- Search has overtaken Browsing
  - In 2005 39% of Full Text Article Downloads on ScienceDirect resulted from sessions starting in External Search platforms
- Search proven more effective than browsing
  - 2.4 full text articles downloads per session starting from external search platform vs. 1.9 per session starting from journal homepage



Contribution Search differs per subject area



- In Life Sciences 60-70% of users link directly to a ScienceDirect article via an External Search platform
- % lower for subject areas lacking strong subject-specific search platform

Search yields more than just journal results!

Subject Specific Search platforms remain important

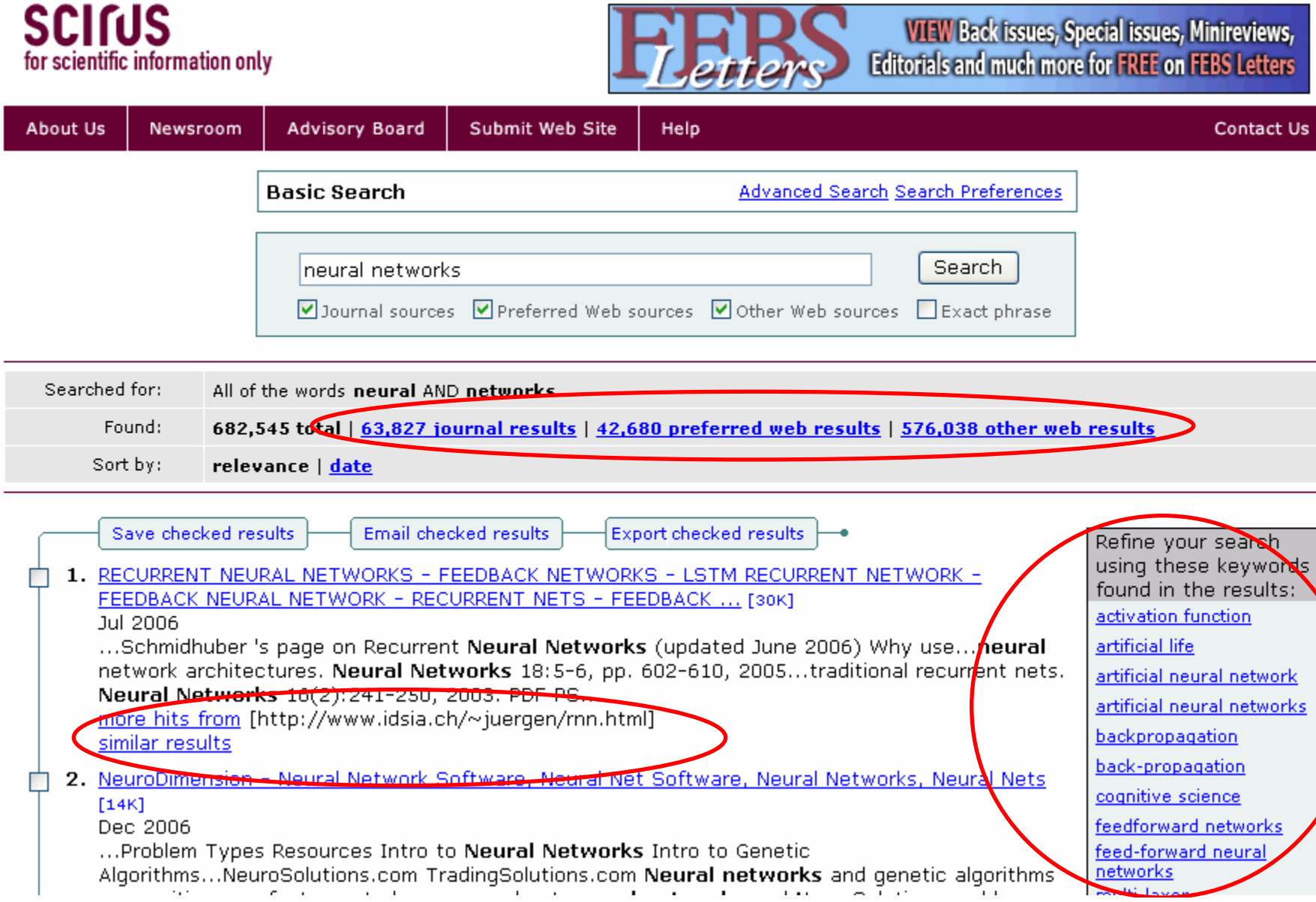
- General Web search engines often 1st choice information tool for scientists (66%) and physicians (55%)*
- Subject-specific search platforms remain important
  - Professionals are becoming increasingly dissatisfied with general search**
  - Average # of full text article downloads per session initiated from PubMed is 3. With alternative platforms this is as low as 1.5

* Outsell report for RE Search group ** Outsell report – iMarket – Vertical Search Engines Deliver What Big Search Engines miss

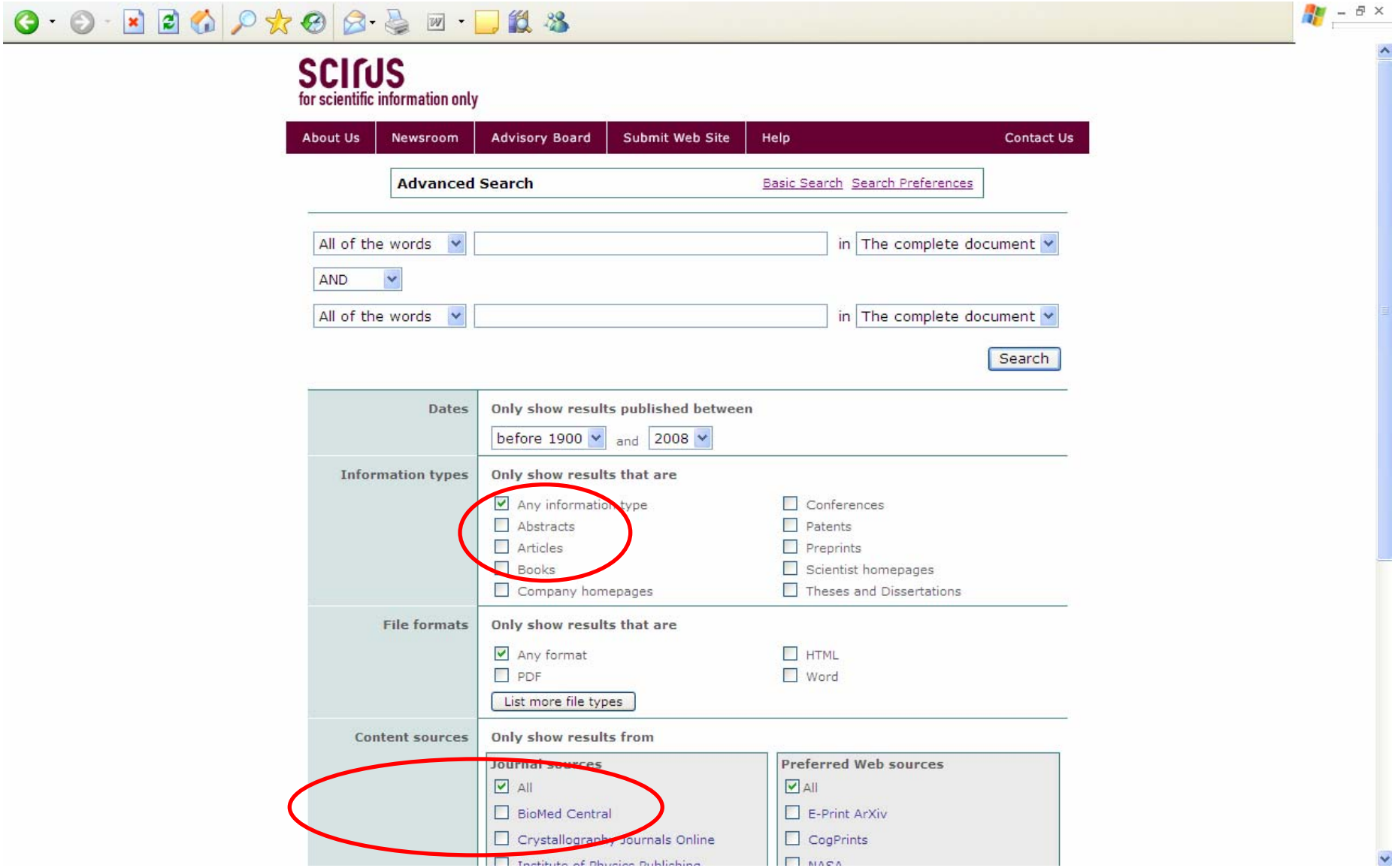
Important elements in Searching

- Indexing
  - Ensuring completeness (including only relevant sources)
  - Capturing existing structures and adding automatic structures and information
- Ranking
  - Date, term frequency and location (title, reference, abstract, full text)
    - Abstract (A&I Databases) vs. full text searching (Web search engines)
  - Link analysis
    - Formal citations (Web of Science, Scopus – journal and patents)
    - Web citations (Google Scholar, Scopus)
    - Web links and use of anchor text (most Web Search engines)
  - Clustering
    - Results from one source (most Web Search engines)
    - Versions (Google Scholar)
- Result classification
  - Keyword or Thesauri (specialised A&I)
  - Subject classification (e.g. math or physics); journal; author etc.
  - Sources and document type classification (e.g. journal article, patent, pre-print or web document) (Scopus, Scirus)
  - Similar results (Amazon, Scirus)

Scirus search results

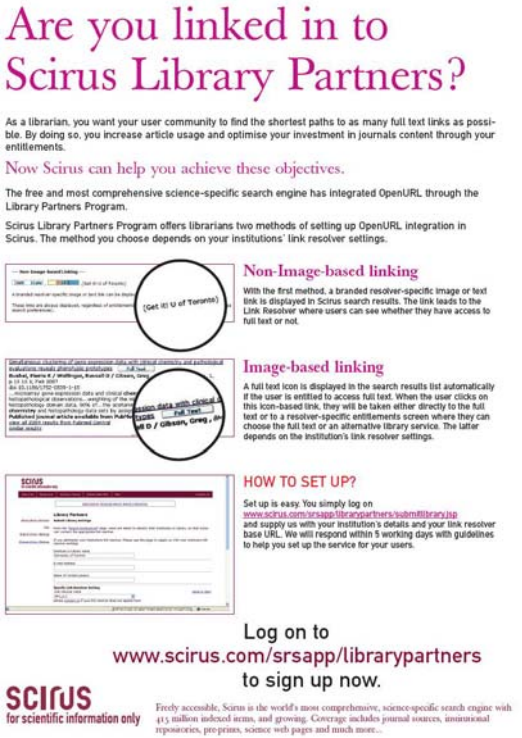


Scirus search results



How can librarians improve search efficiency?

- Education
  - There are more options available than Google…
  - …and they are more focused and superior
- Library Integration
  - Search on Library Homepage
  - Open URL integration



Different content discovery methods

- Browsing
- Linking
- Alerting
- Searching
- User collaboration/sharing

Covered in Following Slides

What will the future bring

- Discovery through communities and networks
  - Community-driven sources selection
  - Social network searching
  - Reviewing and commenting
  - Ranking & classification by users
    - » Active and inactively (through usage & behaviour)

Combining Browsing, Linking, Alerting and Search in a Community and Network-Driven system!

Thank you!