

Battle of the giants: a comparison of Web of Science, Scopus ... & Google Scholar ©

Gary Horrocks

Research & Learning Liaison Manager,
Information Systems & Services
King's College London

gary.horrocks@kcl.ac.uk

Focus

- Content
- Currency
- Keyword searching
- Author & affiliation searching
- Citation searching & browsing
- Google Scholar
- Summary & conclusions



Content: WoS

- One component of the Web of Knowledge Platform
- 8, 700 current international and high impact titles 6,000 STM
- Multidisciplinary resource
- Conferences in ISI Proceedings
- Patents in Derwent Innovations Index
- Integration possible with range of additional tools including Journal Citation Reports, Medline, BIOSIS, Web Citation Index



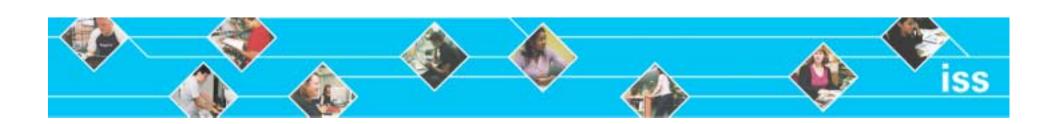
Scopus: a 'one stop shop'

- 15, 000 current international titles
- Essentially STM and social science
- Conferences included
- Patents searched separately
- Includes web search SCIRUS: web sites, theses and e-prints



Content: WoS

- Essentially designed as a citation index
- SCI 1900 (abstracts from 1991 and keywords from 1991)
- SSCI 1956 (abstracts from 1992 and keywords from 1991)
- AHCI 1975 (abstracts from 2000 and keywords from 1991)
- Index Chemicus 1993- Current Chemical Reactions 1986 -
- Retrospective access dependent on backfile purchase; but timescales affect any notion of size

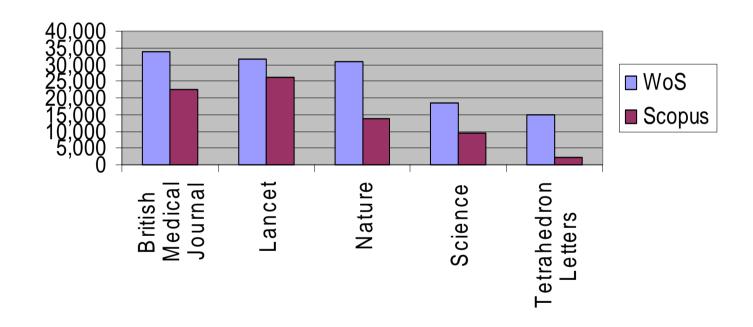


Content: Scopus

- Publisher-submitted records 1996 -
- Citation enhancements from 1996 -
- PubMed records from 1966-
- Content from other Elsevier databases: including Embase (1970-), Biobase (1994-) and Compendex (1970-)
- Notion of broad-based STM is exaggerated. Pre-1996: heavy health & life, engineering focus. Medicinal chemistry & chemical engineering wellserved
- 2007 back files 7m records from publishers including Elsevier, Springer, Nature, RSC - will include abstracts and focus on chemistry, physics and social science

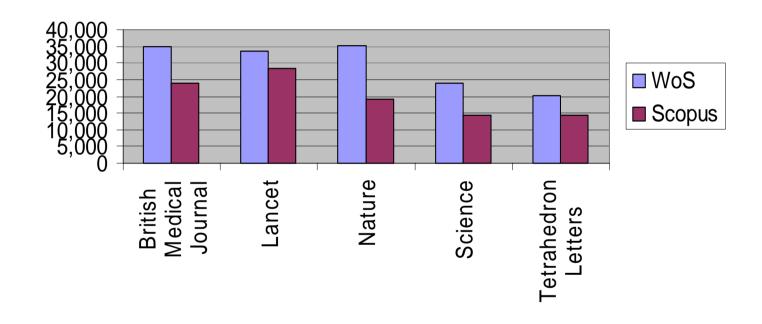


Journal title search 1976-1985



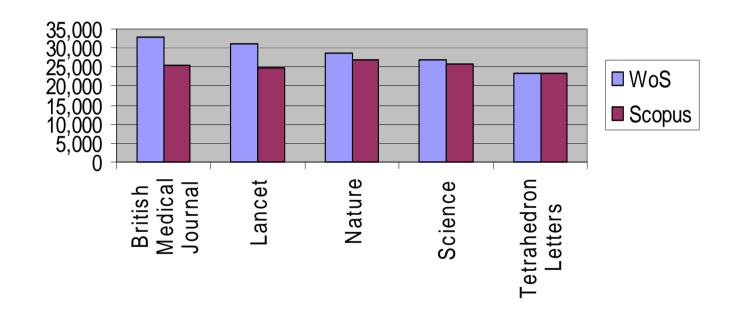


Journal title search 1986-1995





Journal title search 1996-2005





Currency

Journal	Publisher	WoS latest	Scopus
	latest issue	issue	latest issue
ВМЈ	7572	7568	7568
Lancet	9544	9541	9535
Nature	7112	7111	7110
Science	5797	5796	5794
Tetrahedron Letters	46	38	45



Currency

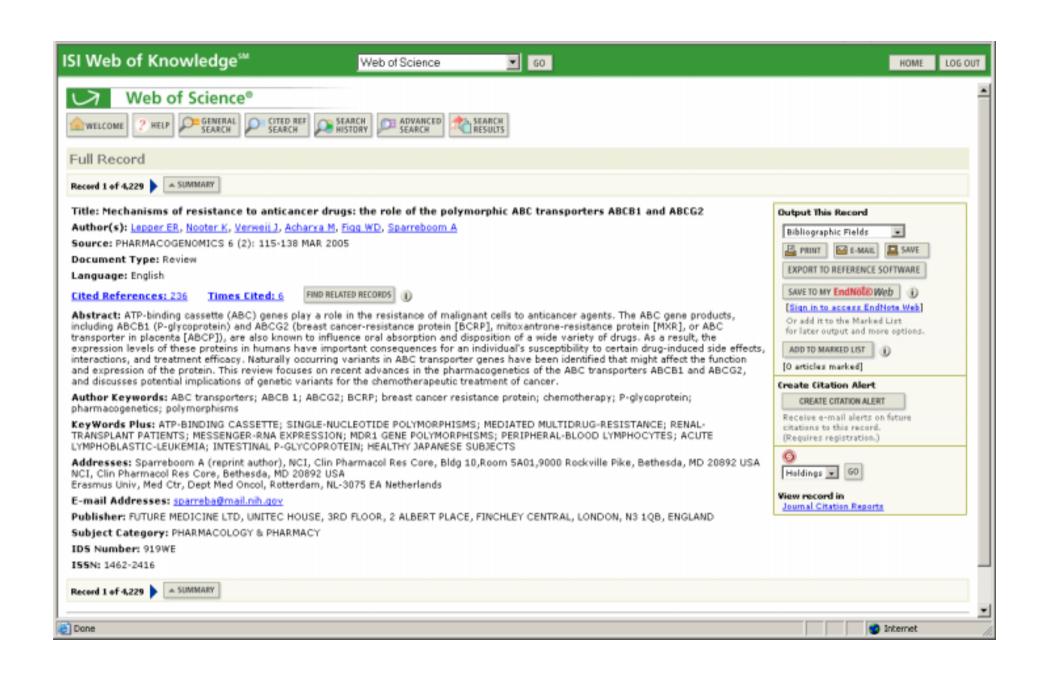
Journal	Publisher	WoS latest	Scopus
	latest issue	issue	latest issue
J.A.C.S	42	38	38
Optics Letters	21	18	15
Chem. Comms.	40	38	38
Pharm J.	7422	7404	7419
N.E.J.M	15	14	13



Keyword searching

- Both databases are free text; neither offers linguistic support tools
- WoS has automated indexing that assigns author keywords & 'keywords plus' from 1991 onwards
- Scopus adds author keywords and some index terms from Pubmed (MeSH) and Elsevier databases (including EMTREE)
- While they are derived from controlled vocabularies, there is none of the functionality
- Harvesting records in Scopus and refining searches has great potential
- However, ambiguity and inconsistency in the process and more clarity on indexing rules is required
- Best to utilise quality of indexing in PubMed, Embase, SciFinder







ABC transporters; ABCB1; ABCG2; BCRP; Breast cancer resistance protein; Chemotherapy; P-glycoprotein; Pharmacogenetics; Polymorphisms

Index Keywords

EMTREE drug terms: antineoplastic agent; breast cancer resistance protein; glycoprotein P

EMTREE medical terms: malignant neoplastic disease; pharmacogenetics

Molecular Sequence Numbers

GENBANK, M14758(referenced)

Chemicals and CAS Registry Numbers

4 [3 (4 benzyl 1 piperidinyl)propionyl] 2,3,4,5 tetrahydro 7 methoxy 1,4 benzothiazepine, 145903-06-6; 6 formylamino 12,13 dihydro 1,11 dihydroxy 5h indolo[2,3 a]pyrrolo[3,4 c]carbazole 5,7(6h) dione 13 glucoside, 151069-12-4; 7 ethyl 10 hydroxycamptothecin, 86639-52-3; 7 hydroxystaurosporine, 112953-11-4; biricodar, 174254-13-8; cyclosporin, 79217-60-0; dactinomycin, 1402-38-6, 1402-58-0, 50-76-0; daunorubicin, 12707-28-7, 20830-81-3, 23541-50-6; dexniguldipine, 113145-70-3, 120054-86-6; dicloxacillin, 13412-64-1, 3116-76-5, 343-55-5; diethylstilbestrol, 30498-85-2, 55-3; difflomotecan, 220997-97-7; digoxin, 20830-75-5, 57285-89-9; DNA topoisomerase, 80449-01-0; docetaxel, 114977-28-5; doxorubicin, 23214-92-8, 25316-40-9; efavirenz, 154598-52-4; elacridar, 143664-11-3; epirubicin, 56390-09-1, 56420-45-2; estradiol, 50-28-2; etoposide, 33419-42-0; exatecan, 144008-87-7, 169869-90-3, 171335-80-1, 197720-53-9; flavopiridol, 146426-40-6; gefitinib, 184475-35-2, 184475-55-6, 184475-56-6, 59-05-2, 7413-34-5; midazolam, 59467-70-8; mitoxantrone, 65271-80-9, 70476-82-3; n [4 (3 chloro 4 fluoroanilino) 7 (3 morpholinopropoxy) 6 quinazolinyl]acrylamide, 267243-28-7, 338796-35-3; nelfinavir, 159989-64-7, 159989-65-8; nortriptyline, 72-69-5, 894-71-3; novobiocin, 1476-53-5, 303-81-1, 39301-00-3, 4309-70-0; paclitaxel, 33069-62-4; phenytoin, 57-41-0, 630-93-3; prednisone, 53-03-2; reserpine, 50-55-5, 8001-95-4; rifampicin, 13292-46-1; ritonavir, 155213-67-5; talinolol, 57460-41-0; tamoxifen, 10540-29-1; tariquidar, 206873-63-4; teniposide, 29767-20-2; testosterone, 58-22-0; tipifarnib, 192185-72-1; topotecan, 119413-54-6, 123948-87-8; tsukubaenolide, 104987-11-3; valspodar, 121584-18-7; vinblastine, 865-21-4; vincristine, 57-22-7; warfarin, 129-06-6, 2610-86-8, 3324-63-8, 5543-58-8, 81-81-2; zosuquidar, 167354-41-8, 167465-36-3, 312905-17-2, 474276-97-6

Tradenames

Drug tradename: bn 80915, ci 1033, dx 8951f, gf 120918, j 107088, jtv 519, ly 335979, nb 506, psc 833, r 101933, r 115777, sn 38, ucn 01, vx 710, xr 9576, zarnestra.

References (238)

First 60 references displayed (view all references)





⇔rkeγworus

On the Web based on





Search options

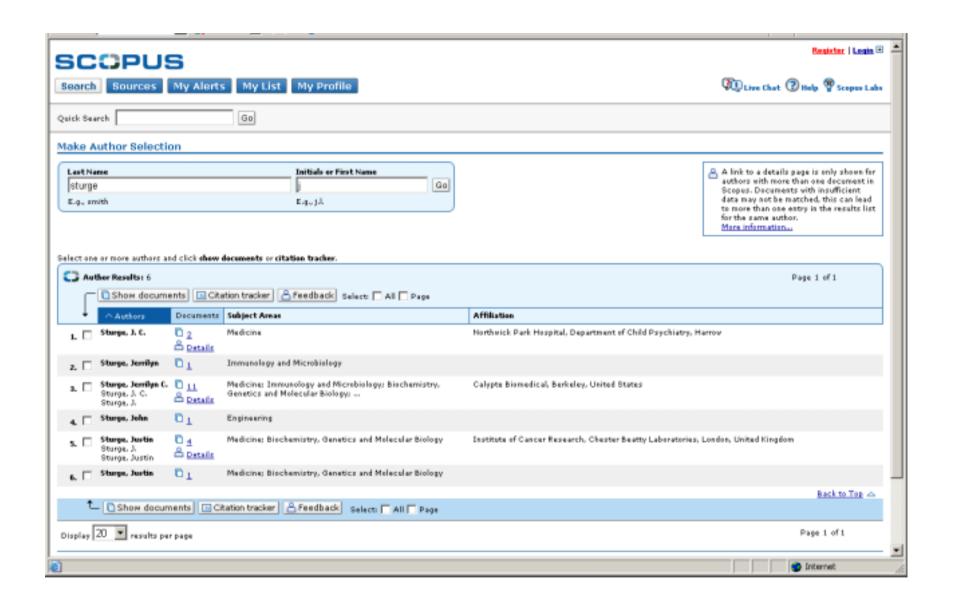
- Both offer basic and advanced functions
- Boolean, proximity operators, wild cards...
- Limits
- Both databases have embraced Google syntax and default to AND searching unless you use ""
- Approach free text with caution



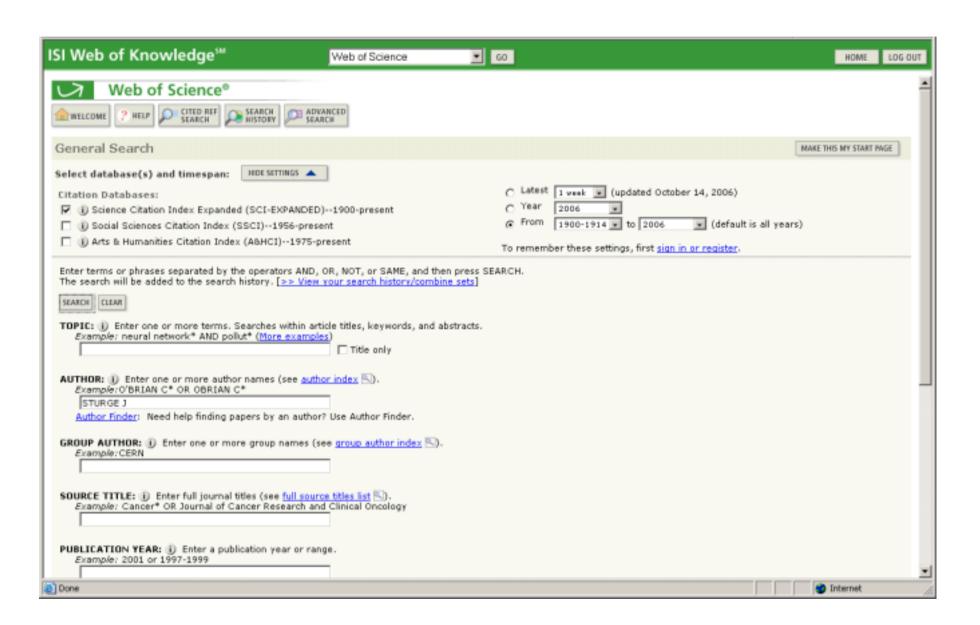
Author & affiliation searching

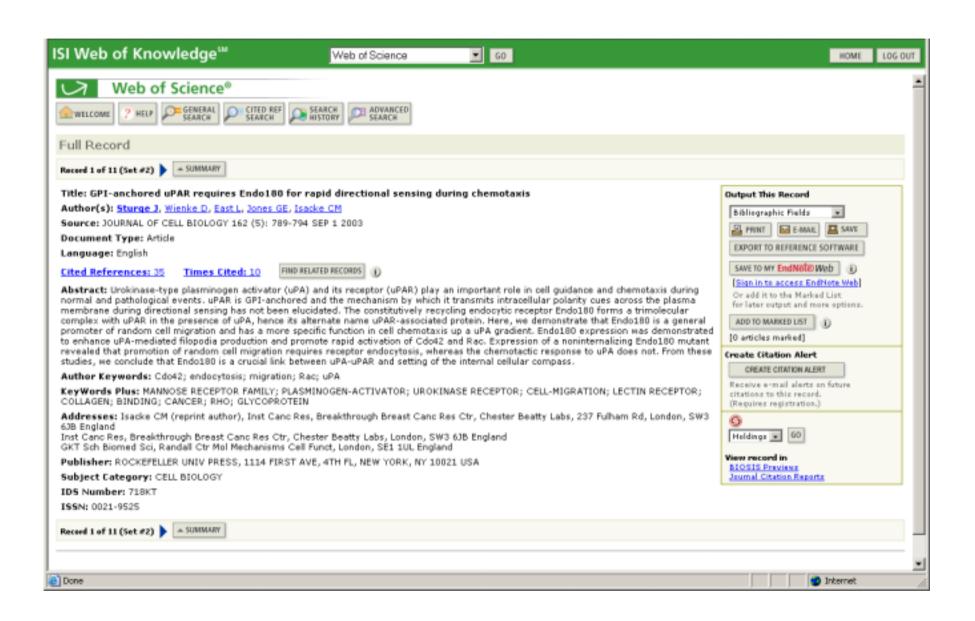
- Matching authors to affiliation information is crucial to disambiguate
- WoS indexes all authors and affiliations provided in the source
- Varying levels of consistency in Scopus: pre-96 1st author/correspondence address; 1996-2001 80% all; 2003 - all
- Standardisation problems partly rest with the authors and institutions!
- WoS offers consistency via an authority file for addresses, including postcode and offers an 'Author Finder' option
- Scopus enables you to select an author from a list of authors with same name and initials

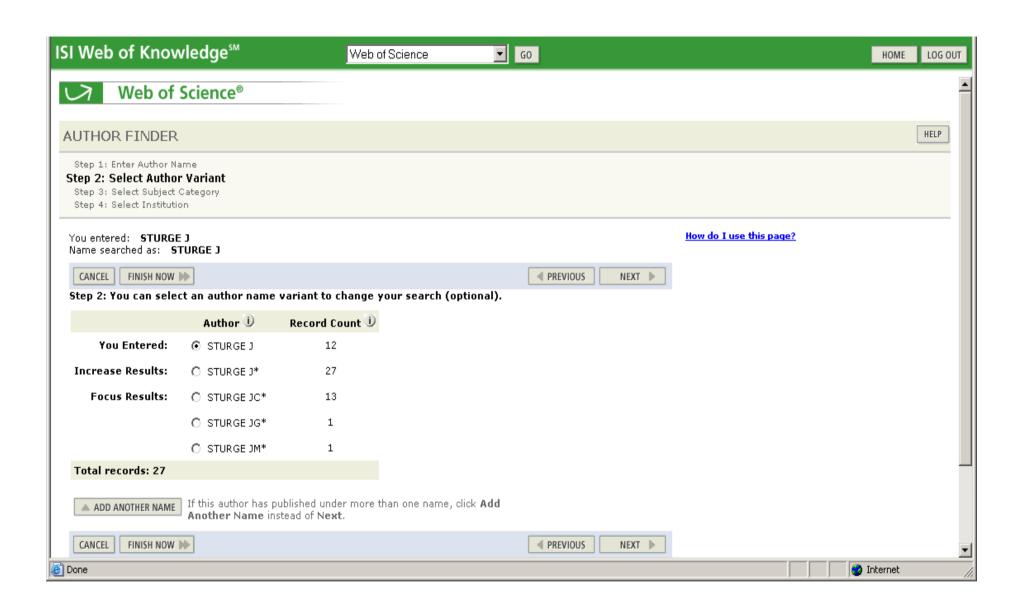


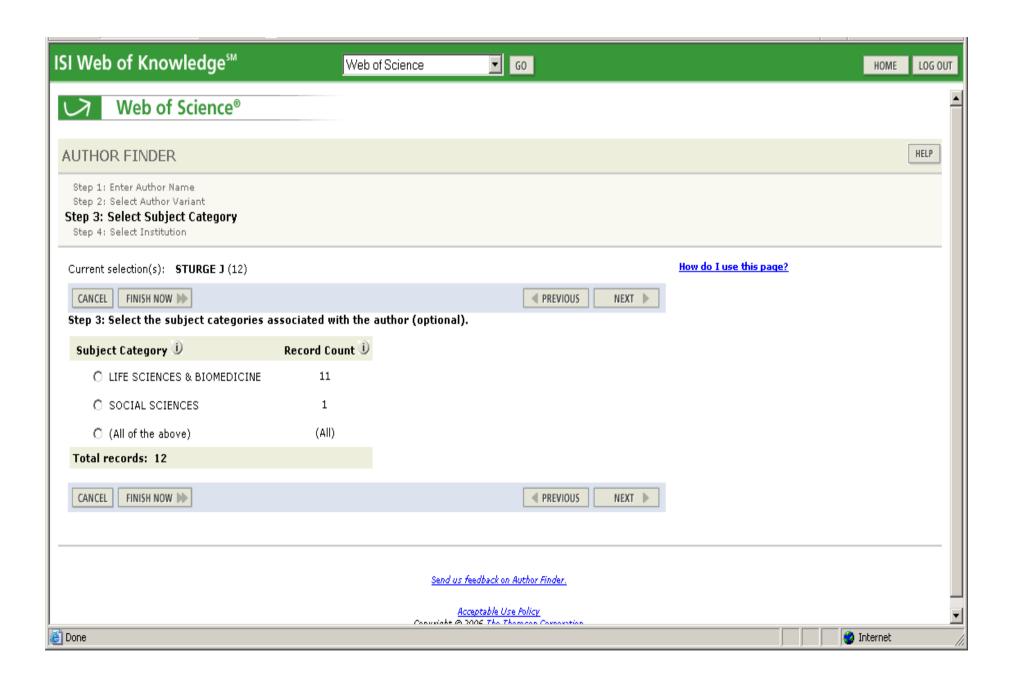


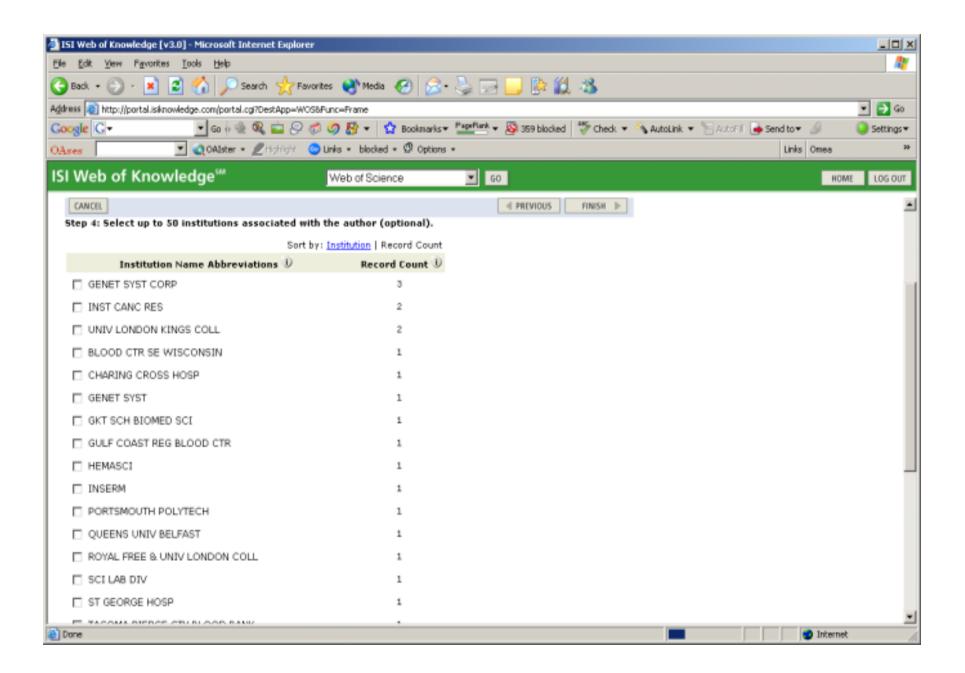












Citation searching

- Key performance management tool
- ...a degree of academic vanity ©
- WoS retrospective coverage enhances power of citation searching (100+ years)
- Scopus has added some 245 million references from 1996 -
- Both offer citation browsing and alerting facilities



Citation searching

King, MC and Wilson, AC 1975. Evolution at two levels in humans and chimpanzees Science: 188, 107-116

WoS: 819 citations (including 10 mis-citations)

Scopus: 232

WoS consistently higher for older articles

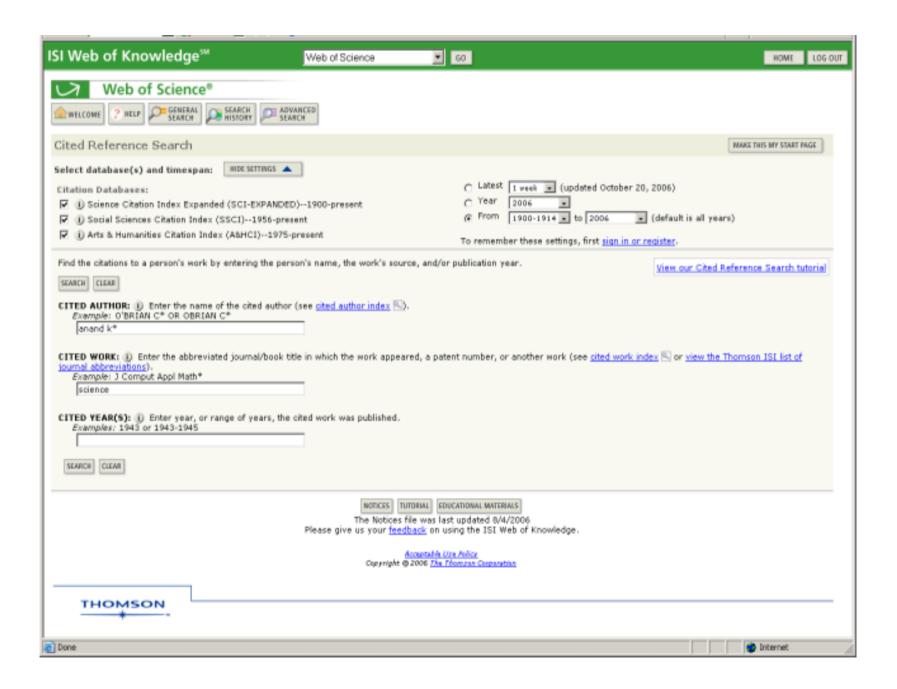
So: chronology does matter in the size debate

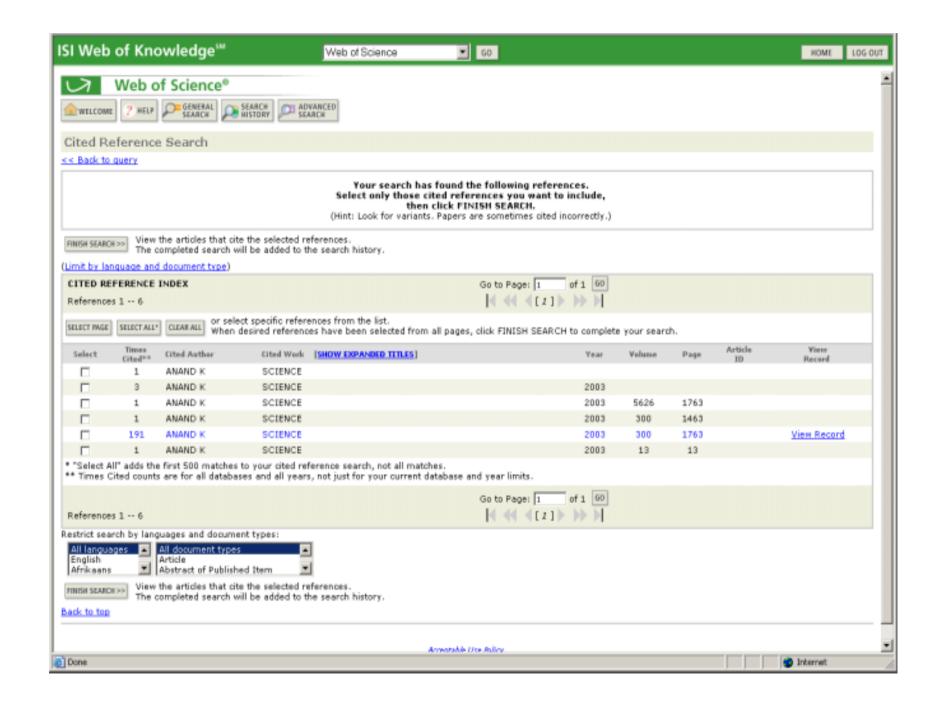


WoS Citation searching

- Offers a 'cited reference search' option
- Where there is a full record for the cited work in the database all authors are searchable
- Where there is not, you require the first author
- A related records option is based on an analysis of citations
- The citation display screen provides a global overview of the citation process; including miscitations ©



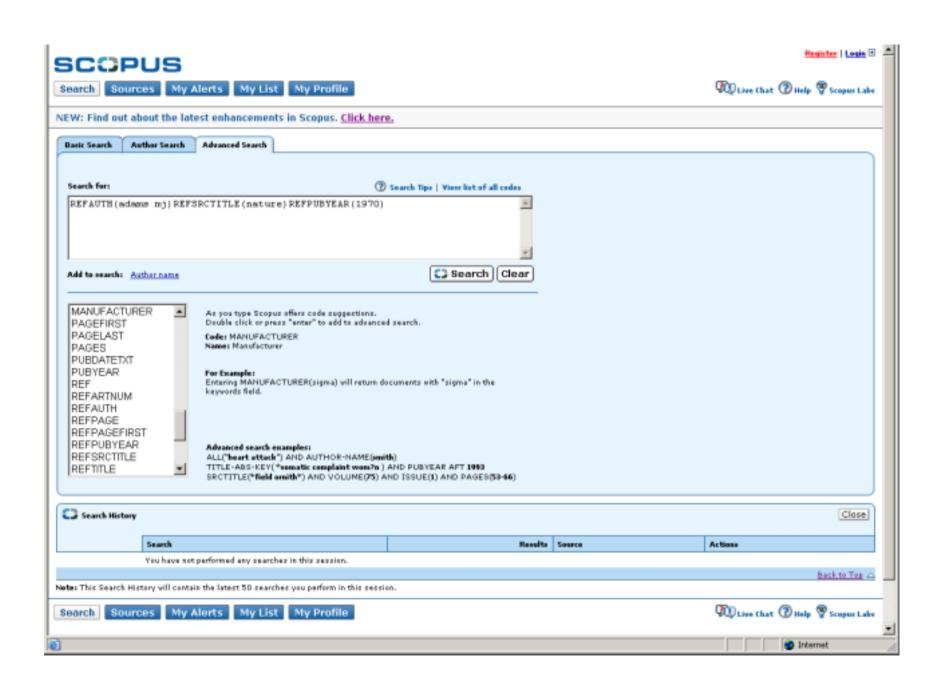


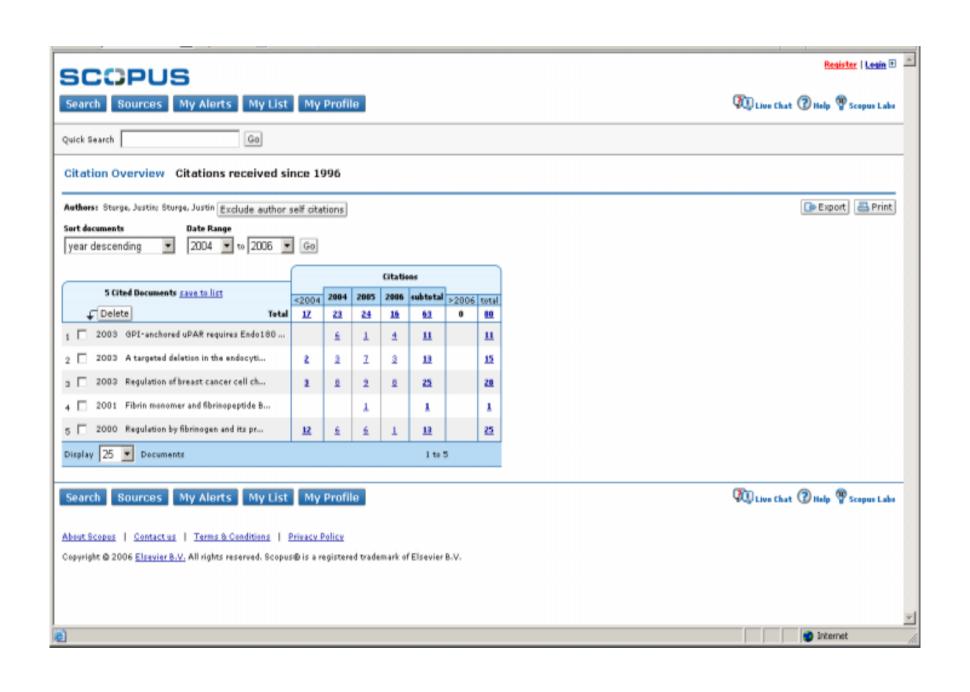


Scopus Citation searching

- A related documents option is available, based on citations, authors or keywords
- 'Citation Tracker' generates a global overview of articles and citation rates
- It's more complex to find citations to items not included in Scopus ...
- No simple 'cited reference search' option you have to use advanced search



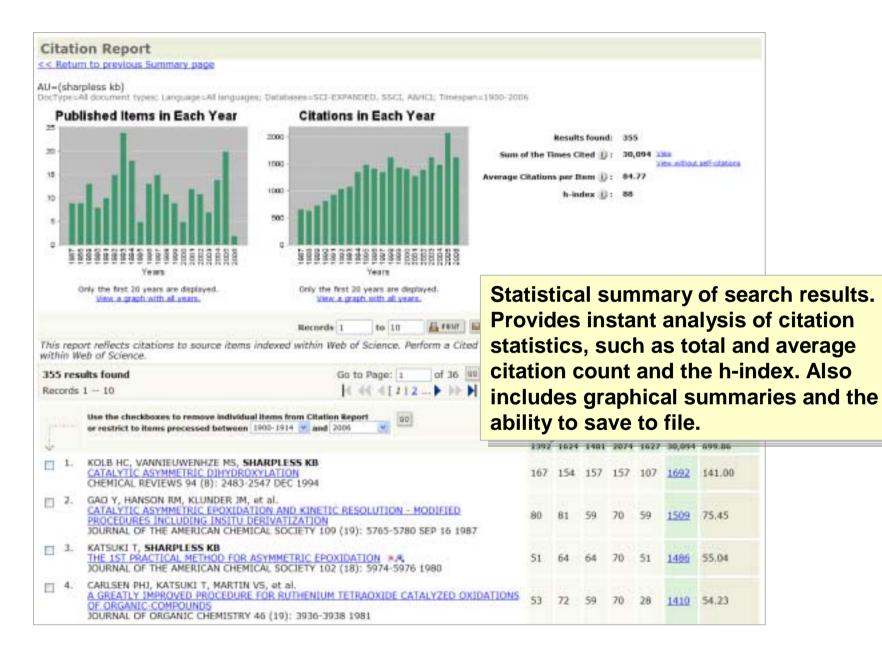




Citation searching

- Scopus back files will NOT include cited references
- WoS is about to launch a Citation Report tool: statistical & graphical summaries





Data Management

- Print, email, save
- Both offer integrated workflow options
- Export to bibliographic software
- WoS EndnoteWeb
- RSS alerting



Summary

- As an amalgam of databases there are considerable variations in data entry in Scopus
- Scopus needs to address gaps and inconsistencies in pre-96 coverage...a work in progress
- Scopus interface is clearer and more intuitive
- WoS is good for currency and depth of coverage
- Scopus is good for author/affiliation searching
- Scopus is good for keyword searching
- WoS is good for citation searching, particularly for pre-1996 articles
- Detailed and subject-specific analysis required



Google Scholar

- Free, but lots of content requires authentication
- Still in beta and still an unknown quantity
- Comprises peer-reviewed papers, abstracts, theses, books, e-prints, technical reports
- Trawls publisher sites, professional societies, institutional repositories, full text documents and cited references – but how??
- No clarity on size, content, selection criteria or time span
- No clarity on data gathering or ranking algorithms



Google Scholar ©

- It's more up to date and identifies material not yet indexed by WoS or Scopus
- It identifies lots of unique material
- It's fast
- It's increasing collaboration with publishers and libraries ensures links to 'appropriate copy' and holdings information is available
- Google search options including "" phrase searching are easy to learn
- It has an 'advanced' search option, with the opportunity to restrict to broad subject areas
- It includes citation data, and there is a 'cited by' function



Google Scholar 🕾

- We've no idea what's included or excluded or even how it works
- Not all publishers will play ball, so PubMed is used as a proxy index
- Google is constantly crawling the web, but how often is GS updated?
- It lacks any of the sophisticated search functionality of Scopus and WoS - but what would you expect for free!
- Data is inconsistent and there are no efforts at standardisation
- Does it add to info glut?





Conclusions

- Budgetary restraint: fee or free?
- JISC's academic database assessment tool may help
- Demands for Google search functionality will see further interface developments
- Local user requirements crucial in any comparison
- Recall versus precision: does size matter?
- We are not comparing like with like: Comparison of Scopus with accumulated WoS, ISI Proceedings, Medline, Biosis and Web Citation Index would be better



Conclusions

- Estimated critical mass of STM journals 30-50K...so the Google Scholar option is an attractive 'mop up' solution: potentially huge critical mass of supplementary data
- Got to be mean to keep them keen ©
- Intense competition has bred innovation in author profiling and citation analysis
- No perfect answers



Acknowledgements

Linda Humphreys, Science Faculty Librarian,
 University of Bath and Chair, JIBS User Group



- Roger A. Mills, Biosciences Librarian, Oxford University Library Services
- Andria McGrath, Information Specialist, King's College London