Scopus reviewed and compared

The coverage and functionality of the citation database Scopus, including comparisons with Web of Science and Google Scholar

Jeroen Bosman
Ineke van Mourik
Menno Rasch
Eric Sieverts
Huib Verhoeff

June 2006

Scopus reviewed and compared: the coverage and functionality of the citation database Scopus, including comparisons with Web of Science and Google Scholar

1 Summary

Background

- **Scopus has been launched** by Elsevier; it is a product that introduces competition in a segment of the market for bibliographical databases in which ISI previously had a monopoly position with its product Web of Science. Cost considerations and favourable initial responses mean that a more detailed survey is meaningful.
- **Changing search habits** among students (less emphasis on subject-specific search terms, more often following links than searching systematically) and other rivals (especially Google Scholar, possibly also Windows Live Academic) mean it is important to assess multidisciplinary databases.

Method

- **In depth**; the survey can also be used for subsequent evaluation of other products and is also a test for how this can be approached and what data is available for comparisons.
- **Own research**, because there is little reliable literature (except Jasco 2005 and Pipp 2006).
- Sessions with and input from subject specialists
- Feedback session with manufacturer
- **User input**: interviews and user survey

Research results on coverage

- **Number of records, titles**. Scopus has almost 28 million records; the number of records in our version of WoS, at 19 million, is smaller, but the number in the full WoS (with backfiles stretching back to 1945) is larger, at 37 million. Scopus covers over 15,000 journals, versus 9,000 in WoS. Scopus covers 64% of our digital journals, as against 53% in WoS.
- **Period covered**. Scopus is 5-15% smaller prior to 1996, and 20-45% larger than WoS after 1996 on the basis of the number of records. For publications before 1996, the coverage offered by Scopus for the various subjects is highly uneven.
- **Types of documents**. 95% of the total database of Scopus consists of the records of descriptions of articles in journals. For the years prior to 1996, the number of non-journal articles in Scopus is low, subsequently rising to over 10% in 2005. That means that for recent years the proportion of non-journal articles is significantly higher than in WoS (4%).
- **Subject-specific**. Scopus covers only scientific fields. WoS additionally covers the classics. The coverage provided by Scopus is 4 or more percentage points higher than that of WoS in 16 of the 18 UBU subjects on the basis of the numbers of titles of journals in the range carried digitally by the UBU. The two subjects in which WoS is stronger are both in the arts/humanities. On the basis of a number of searches, Scopus appears to be relatively weak in sociology, physics and astronomy (but

caution is in order here, as further investigation is required), but very good on biomedical and geosciences.

- **Up-to-dateness**. In terms of the inclusion of issues of journals and on the basis of the 'progression percentage' for coverage of the current year, there is hardly any difference between WoS and Scopus as regards the speed with which new publications are included.
- **Nature of data per record**. Scopus has more keywords, for authors but often also from 'controlled vocabulary' (e.g. MeSH). Besides author keywords, WoS has no keywords from controlled vocabulary but it does have Keywords-plus: keywords generated from references.
- **Citation data**. The difference between Scopus and WoS in terms of citation data is comparatively slight, there is a strong overlap. A count on the basis of references to 64 articles from 1995 and 2000 shows that WoS has 6% fewer references to citing articles. The difference between these two and Google Scholar is larger. While Google Scholar has 2% fewer references to these articles than Scopus, it does on average include 5 times as many 'unique' citing publications. For socio-economic sciences in particular, including economics, Google Scholar has many more and more unique citations.

Research result functionality

- **Difference in capabilities**. Scopus is slightly more versatile and has a few clear advantages in functionality in the form of default refine, the table format of results of the Citation Tracker and author identification. WoS has slightly more extensive options for citation analysis for institutions. Note: In June 2006, WoS also included a Refine tool and ISI also announced author identification for WoS.
- **Speed**. There is above all a substantial difference between WoS and Scopus with GS, which produces virtually instant results, and also, depending on the type of search, with the Omega search engine, which is also often very quick. This can (subconsciously) be a major reason for users to choose Google Scholar. While there is little to choose between WoS and Scopus in terms of speed, Scopus is slightly faster.

User ratings

- **Interviews**. Heavy users from the faculties rate the clarity of the Scopus interface and refine and the citation tracker particularly highly. The majority of interviewees values Scopus more highly than Wos, but also 'demands' that JCR has to remain available.
- **Survey**. A survey among 81 users shows that Scopus and WoS are less well-known than Google Scholar, but the results generated by Google Scholar are rated less highly, especially among research trainees/researchers, and among those, largely the scientific disciplines. Scopus is rated best in use, followed closely by Google Scholar. According to the respondents, WoS clearly has some ground to make up

here. In terms of the relevance of the results, Scopus is likewise rated most highly of these three citation databases.

2 Introduction

2004 saw the market launch of a new multidisciplinary database: Scopus. This introduced a measure of competition into a segment of the market for bibliographical databases where it had not existed before (see table 2.1).

Table 2.1 Segments in the market for bibliographical databases

Functionality	Multidisciplinary	Subject-based
Titles + abstracts + full citation functionality	• Scopus • Web of Science	-
Titles + abstracts + limited citation functionality	• EBSCO ASE • Google Scholar (- abs.) • Highwire (+ full text)	• Citeseer • PsycInfo • Pubmed Central • RepEc Econpapers • SciFinder Scholar • SMEALsearch
Titles + abstracts	• DOAJ search • Infotrieve articlefinder • Omega (UBU) • Open J-Gate	• (selection:) • CAB abstracts • Econlit • ERIC • GeoArchive • Geobase • Georef • Pubmed • SocIndex • Sociological Abstracts • TRIS • Zentralblatt MATH
Titles	• Picarta • Windows Live Academic	• (example:) • GeoDOK

The new product, from Elsevier Science, is a direct rival of Web of Science from Thomson-ISI, to which the UBU subscribes. The first reports on the product made it clear that Scopus was attractive in terms of functionality and design, but left many questions unanswered as to its coverage. A group of subject and information specialists undertook to examine to what extent Scopus can truly be seen as a valid alternative to Web of Science. The present document is the report on that investigation.

3 Methodology

The investigation into Scopus was fairly detailed, because of the high costs involved in this database and its rival and the great importance attached to this type of database in academic research. Also, virtually all UU subjects were covered, meaning it is important to have a good basis for weighing up the interests involved. Another reason for reviewing Scopus ourselves in detail is that there are only a few known thorough studies. Only Jascó (2004 and 2005) and Pipp (2006) have compared Scopus and Web of Science in detail, but they did so at a time when Scopus was still largely in the course of being developed or they investigated many aspects, such as the coverage of fields of study and citation data, only to a limited extent. Detailed observations can in fact be found in library weblogs (for instance *One entry to research*), but these usually address only one small aspect of the databases. No investigation is known that compares Scopus and WoS in terms of coverage at the level of titles of journals. Nor does the literature yield any comparison of coverage of subject indices as performed for the present investigation.

The principal basis for this report is a database of journals covered by Scopus, WoS and twenty subject indices and the presence of the journals held digitally by the UBU and the Omega search engine of the UBU. The work carried out for this is usable not only for the evaluation of Scopus and WoS but also for any future evaluations of subject indices.

In addition to the coverage assessment on the basis of titles of journals, many persons also tested the functionality of Scopus for six months and we systematically examined how extensive the citation references in Scopus are and to what extent the database is up to date. We did not only perform our own research but also asked users about their experiences and views. To that end, some twenty researchers were interviewed at the beginning of the study, all of whom are heavy users of Web of Science. Partly on the basis of those sessions, in which they demonstrated how they work with WoS and what was important in that context and also what their initial impressions of Scopus were, we determined which aspects needed to be spotlighted in the investigation. It became clear, for instance, that there was virtual consensus on the ease of use of Scopus, but not on the extent of its coverage. In addition to sessions with heavy users we also conducted a limited web survey among users of Scopus, which also asked many students for their views. Despite the detailed approach, there are a few matters that are only addressed obliquely in this study. The exact effect of automatic polling of Scopus, Keywordsplus or Web of Science and keywords in Scopus is not yet sufficiently clear. In addition, we did not investigate the options for export, alerts and personalization. Nor were we able to test the recently introduced Author Identifier in detail yet. Finally, in view of the rapid development of both Scopus and WoS it is essential to continue to monitor

changes after the completion of this report and to take account of them in purchase decisions.

Note: This report describes research carried out with a view to the needs and circumstances at Utrecht University. The subject classifications applied will often not be matched by those used in other institutions. Additionally, comparisons have been made with products, such as Utrecht's own search engine Omega, that are not available elsewhere. Information about licences and annexes with privacy-sensitive elements has not been included in the public version of this report.

4 Coverage: the contents of the databases

One of the major aspects on which Scopus needs to be assessed is its coverage, the contents of the database. Together with functionality and ease of use, coverage determines the value of the database. This section presents a factual assessment of the coverage, with some aspects weighted by subject specialists, and not a rating by users. This is provided in section 6.

The coverage of a bibliographical and citation database comprises a number of aspects:

- the number of publications and documents covered, specified by subject
- period of coverage for serial publications
- up-to-dateness of the coverage (how rapidly are new publications included)
- nature of the data per document (title, author, abstracts, keywords, references, citations etc.)

Information on coverage in this survey is derived from:

- Sources from the supplier of Scopus
- Comparison of journals covered on the basis of ISSN
- Assessment of missing titles of journals by UBU subject specialists
- Comparison of number of citations
- Interviews with heavy users among the faculty researchers
- Web survey

4.1 The number of documents covered

4.1.1 Total number of documents

The total number of documents covered by Scopus amounted to almost 28 million on 27 June 2006, according to its own count. More than 95% of these are articles in journals. Books make up just under 0.1 per cent of the number of records. This number of records is smaller than the full version of Web of Science (covering 1945-today with 35 million records (Jasco 2005), but substantially more than the UBU version of WoS (covering 1988-today, with 19 million records (estimate based on Jasco 2005, p 1542). More than 50% of the documents indexed in Scopus was published before 1996.

4.1.2 Number of journals covered

With over 15,000 titles, Scopus covers substantially more journals than Web of Science (almost 9,000). In itself information on the millions of articles in those extra journals is of course valuable, but the question obviously is what the nature

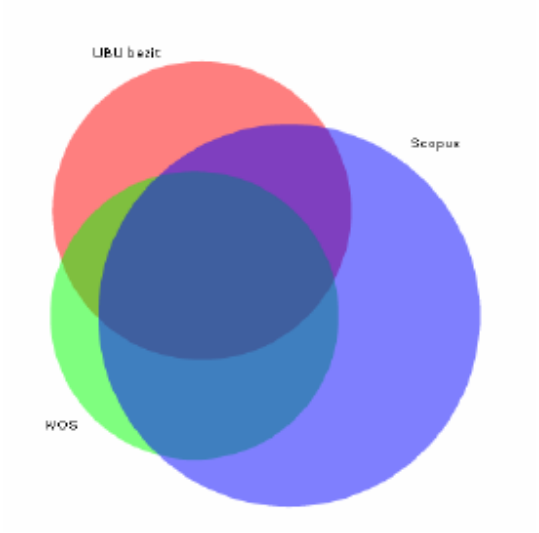
of those extra journals is. For a long time, the corpus of journals covered by WoS has been considered to represent the top segment of journals. This report considers the quality of the journals covered by Scopus but not by WoS, and vice versa. Differences in this respect can be decisive for choosing one of these databases over the other.

The difference between Scopus and WoS as regards indexed journals is interesting in itself because Scopus covers thousands' more titles than WoS in a way (including citation data) that only WoS provides in that quality. Other databases with citation data either provide far fewer professional options (Google Scholar) or only cover one or a few subjects (e.g. Citeseer).

The larger number of journals covered by Scopus is due in large part to the fact that Scopus is oriented internationally. The proportion of journals from the US, Canada, the UK, The Netherlands (Elsevier), Germany (Springer) and Switzerland in WoS is 78% as compared to 67% in Scopus (Pipp 2006). Scopus covers six times more Chinese and three times more Spanish, Russian, Indian, Polish and Italian journals than WoS (Pipp 2006).

It is interesting to see how Scopus and WoS compare as tools to access the holdings of the UBU. Unfortunately we have no comprehensive list of all the journals held by Utrecht University to establish which proportion of it is included in the databases. If we look only at the journals held in *digital* form by UBU (which is around two thirds of the total number of journals subscribed to by UBU) we find that the difference between the products from ISI and Elsevier is still considerable (figure 4.1 and table 4.1). With 53 and 64% respectively, Scopus covers 11 percentage points more of the digital UBU titles. Without the EBSCO journals, this difference narrows considerably, to 4 percentage points. On the other hand, Scopus covers more journals not held digitally by the UBU than WoS. Interviews show that researchers often appreciate this while students usually see it as dead weight.

Figure 4.1 Overlap of digital titles of journals in UBU and titles in Web of Science and Scopus, May 2006.



Source: own research

Table 4.1. Overlap of digital titles of journals in UBU and titles in Web of Science and Scopus, May 2006.

		Number of journals	UBU	WOS	Scopus
including Ebsco	**UBU**	9616		5142 (53%)	6162 (64%)
	WOS	8974	5142 (57%)		7505 (84%)
	Scopus	15785	6162 (39%)	7505 (48%)	
		Number of journals	**UBU**	**WOS**	**Scopus**
excluding Ebsco	**UBU**	7810		58%	62%
	WOS	8974	50%		81%
	Scopus	14191	34%	51%	

Source: own research

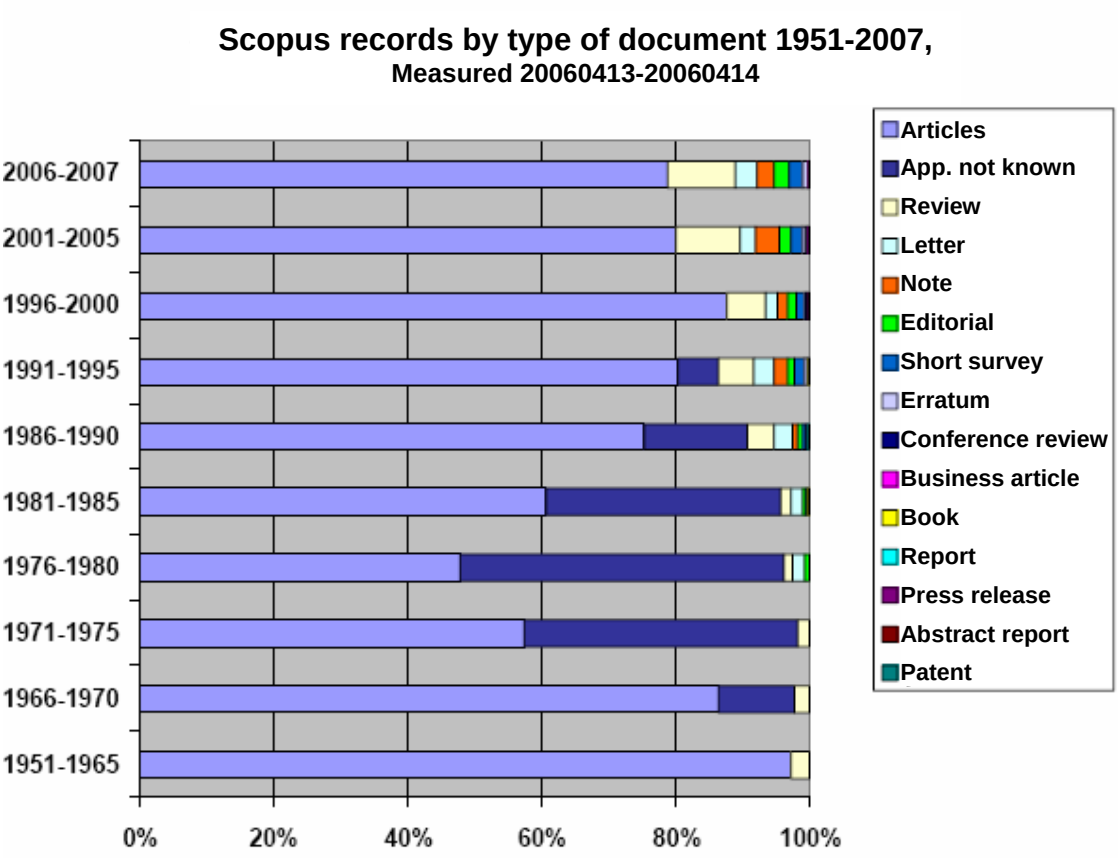
An important question is to what extent Scopus is capable of actually plugging the gaps in its coverage of journals. The literature (Goodman & Deis 2005, Deis & Goodman 2006, Jascó 2006) repeatedly points out missing issues and years. While this also occurs in Web of Science, it appears to occur more often in Scopus. This needs to be analysed and looked into in greater detail, above all by the makers of Scopus themselves.

4.1.3 Documents covered by type of document

Of the 28 million records in Scopus over 90% is a description of an article in a journal. That still means that a few million other kinds of sources are described. For the recent period especially, a relatively large number of reviews, letters, notes and surveys has been included (figure 4.2). Often this material initially sourced from journals. The number of non-journal sources (books, reports, book series, conference papers etc.), at just under thirty thousand, is comparatively small. In this field, some subject indices (Georef, PsycInfo) and Google Scholar

offer far more. The classification by type of publication in refine tools is misleading, however. Because the type of publication is not known for a large number of records from older years (dark blue in figure 4.2), this is strongly underestimated. It is accordingly not advisable to limit searches for material from before 1996 to articles such as type of publication, as this will wrongly exclude millions of records. That is a major shortcoming in the database that requires quick elimination.

Figure 4.2 Scopus records by type of document 1951-2007, measured in April 2006



Source: own research

Comparison with Web of Science (table 4.2) shows that Scopus has significantly more non-article records for a recent year. In both databases, the bulk of these are reviews. Despite these overall figures in Scopus's favour, Pipp (2006, 12) notes that for specific journals, including some important ones, hundreds of articles are lacking in Scopus. Our checks have conformed this, and there is definite room for improvement in Scopus in that regard.

Table 4.2. Document types in WoS and Scopus, publication year 2005, measured on 20060414 by search on *relatively* OR *average*

	WoS	Scopus	WoS	Scopus
	absolute		percentage	
All types	57018	74273	100.00%	100.00%
Article	54650	66824	95.85%	89.97%
Review	1989	6531	3.49%	8.79%
Short survey	0	356	0.00%	0.48%
Note	0	277	0.00%	0.37%
Business article	0	233	0.00%	0.31%
Editorial	201	20	0.35%	0.03%
Letter	58	13	0.10%	0.02%
Conference review	0	12	0.00%	0.02%
Erratum	0	6	0.00%	0.01%
Abstract report	0	1	0.00%	0.00%
Bibliographical item	1	-	0.00%	
Book review	1	-	0.00%	
Correction	24	-	0.04%	
Meeting abstract	51	-	0.09%	
News item	14	-	0.02%	
Reprint	24	-	0.04%	
Software review	2	-	0.00%	

Source: own research

A major shortcoming, in particular for the social sciences, social geography & spatial planning and economics is the absence of book reviews. It is an advantage however that book series can be separately highlighted in the list of sources. Scopus contains, for example, virtually all issues of Nederlandse Geografische Studies – a series comprising doctoral theses in the main – complete with clickable bibliographies.

4.1.4 Documents and journals covered by subject area

It is virtually impossible for the coverage of all subject areas to be even in any multidisciplinary database. A deliberate decision was made not to include the classics in Scopus, as journals are less important in these fields. Only the philosophy of science has been included in Scopus. According to Elsevier itself, the emphasis in the initial development was on STM (Science, Technology, Medicine) and in addition on Social Science (psychology, sociology, economics). To verify to what extent claims made are realised, we made a side-by-side comparison of lists of journals from different bibliographical databases and matched them on the basis of ISSN (table 4.3). This related purely to the occurrence of identical ISSNs, regardless of how many years were included per journal. We compared Scopus with one other multidisciplinary database in this regard, Web of Science, and with 21 subject indices (for the purpose of which we classified EBSCO Academic Search Elite as a subject index for convenience). The subject indices were chosen on the advice of subject specialists, but limited by the avaialability of lists of journals with a full ISSN tag. The Scopus list too,

unfortunately, did not provide an ISSN for each title, as a result of which the coverage in Scopus is sometimes underestimated.

Table 4.3. Coverage of titles from subject-specific databases in WoS and Scopus, with an indication of full text digital availability ('UBU'), as at May 2006.

Database	Number	% in UBU	% in WOS	% in Scopus	In both
CSALISA	477	24.3%	15.7%	25.6%	14.7%
BIOSIS	3244	*	57.9%	82.8%	*
Georef	13345	7.2%	9.4%	13.2%	8.7%
Agricola	2215	28.2%	36.4%	40.5%	35.3%
CSAIPA	371	29.6%	34.5%	59.8%	34.0%
Philosophersindex	1194	25.6%	24.5%	15.7%	11.1%
ATLA	1549	24.2%	21.8%	10.1%	7.7%
Geobase	2065	43.3%	51.8%	87.4%	50.3%
EconLit	1077	41.1%	28.8%	45.4%	27.9%
SociologicalAbstracts	1777	42.7%	37.1%	48.7%	33.0%
PsycInfo	1990	52.8%	54.3%	70.5%	53.0%
CAB	7443	3.4%	38.0%	48.0%	37.2%
INSPEC	8966	18.7%	20.8%	29.3%	20.0%
MathSciNet	2697	22.0%	25.1%	30.1%	23.1%
LLBA	1555	31.9%	28.9%	26.4%	19.2%
Pubmed	7541	*	54.1%	60.3%	*
Embase	4865	47.0%	53.4%	91.6%	52.8%
SocIndex	3703	34.7%	29.6%	38.6%	26.6%
Eric	1264	46.7%	30.9%	40.0%	28.6%
EbscoAcademicSearch	7739	65.8%	56.4%	64.1%	50.2%
BHA	2244	7.8%	11.9%	2.5%	1.6%
MLA	4923	12.9%	14.0%	3.9%	3.0%

Source: own research
* still to be calculated, will be included in a new version of this report expected around 20060820

Scopus provides *higher coverage* than WoS for 18 of the 22 lists of journals of subject indices. Notable features are mainly the extensive coverage in pharmaceutics (CASIPA), geosciences (Geobase), economics (Econlit) and medicine (EMBase). For geosciences and medicine this does not come as a surprise: Geobase as well as EMBase are Elsevier products. It does beg the question, conversely, why no 100% coverage is achieved in Scopus. According to Elsevier itself the overlap of Scopus with EMbase and Compendex is 100%. We suspect the difference is attributable to the missing ISSNs in the Scopus list. It is important to interpret the data in the table only in terms of the comparison between WoS and Scopus. Comparing the coverage of only Scopus or WoS with the various subject indices is difficult because the latter differ strongly by their nature in terms of the total number of indexed serial publications (journals). For Georef, for instance, this means that it also includes thousands of series of reports from geological services. A database such as Econlit by contrast focuses largely on regular journals (and books), as a result of which higher percentages are produced in comparisons with multidisciplinary databases. Finally, the *lower coverage* in Scopus compared with Web of Science in the field of the classics

(and the residual category 'general') is notable. These differences between Scopus and WoS by subject area tally with our findings on the basis of citations in both databases (section 4.6).

Comparison with Google Scholar in this area is difficult, because Google does not publish a list of source journals for the database. Research has demonstrated however that Google Scholar has the best coverage for journals in the medical and biomedical sciences, sharply varying coverage for the non-life sciences (strong on computer science and chemistry, weaker on mathematics and earth sciences), average coverage for the social sciences and economics and relatively poor coverage for the classics (Neuhaus 2006). Google Scholar is however stronger in covering Open Access than non-Open Access, comparatively stronger in English than in other languages, and stronger in covering journals from multidisciplinary large publishers' platforms than of journals in bibliographical databases.

We also carried out a comparison of the coverage of journals available full text digitally in the UU provided by Scopus, WoS and Utrecht University's own search engine Omega (table 4.4). To that end we made use of the subject classification for journals in Omega, which unfortunately is capable of improvement. Again, for each subject only the journals with a primary link to it are included. In other words a journal placed in two categories will only be included in the count for the first. It is important to realise that the underlying data do not always indicate clearly whether a journal is covered in full or selectively in the databases and also that the number years included for each journal does not play a part here.

Table 4.4. Coverage of full text journal titles available in UBU in WoS, Scopus and the Utrecht University search engine Omega, as at May 2006, highest value is marked.

	UBU	**In WOS**		**In Scopus**		**In Omega SE**	
Earth sciences	299	222	74%	242	81%	195	65%
General	298	90	31%	103	35%	179	60%
Biology	1043	768	74%	864	83%	612	59%
Veterinary medicine	114	88	77%	93	82%	42	37%
Economics	659	263	40%	402	61%	429	65%
Pharmaceutics	125	85	68%	101	81%	74	59%
Medicine	2137	1234	58%	1784	83%	1263	59%
Theology	144	47	33%	12	8%	71	49%
Agricultural sciences	82	56	68%	67	82%	53	65%
Language and literature	952	430	45%	171	18%	418	44%
Environmental science	129	62	48%	90	70%	88	68%
Physics and astronomy	464	315	68%	349	75%	364	78%
Law	217	47	22%	75	35%	104	48%
Social geography and spatial planning	156	61	39%	100	64%	88	56%
Chemistry	346	246	71%	261	75%	218	63%
Social sciences	1231	506	41%	666	54%	805	65%
Technology	295	169	57%	218	74%	215	73%
Philosophy	93	51	55%	17	18%	39	42%

Mathematics and computer science	832	402	48%	496	60%	432	52%
Total	**9616**	**5142**	**53%**	**6111**	**64%**	**5689**	**59%**

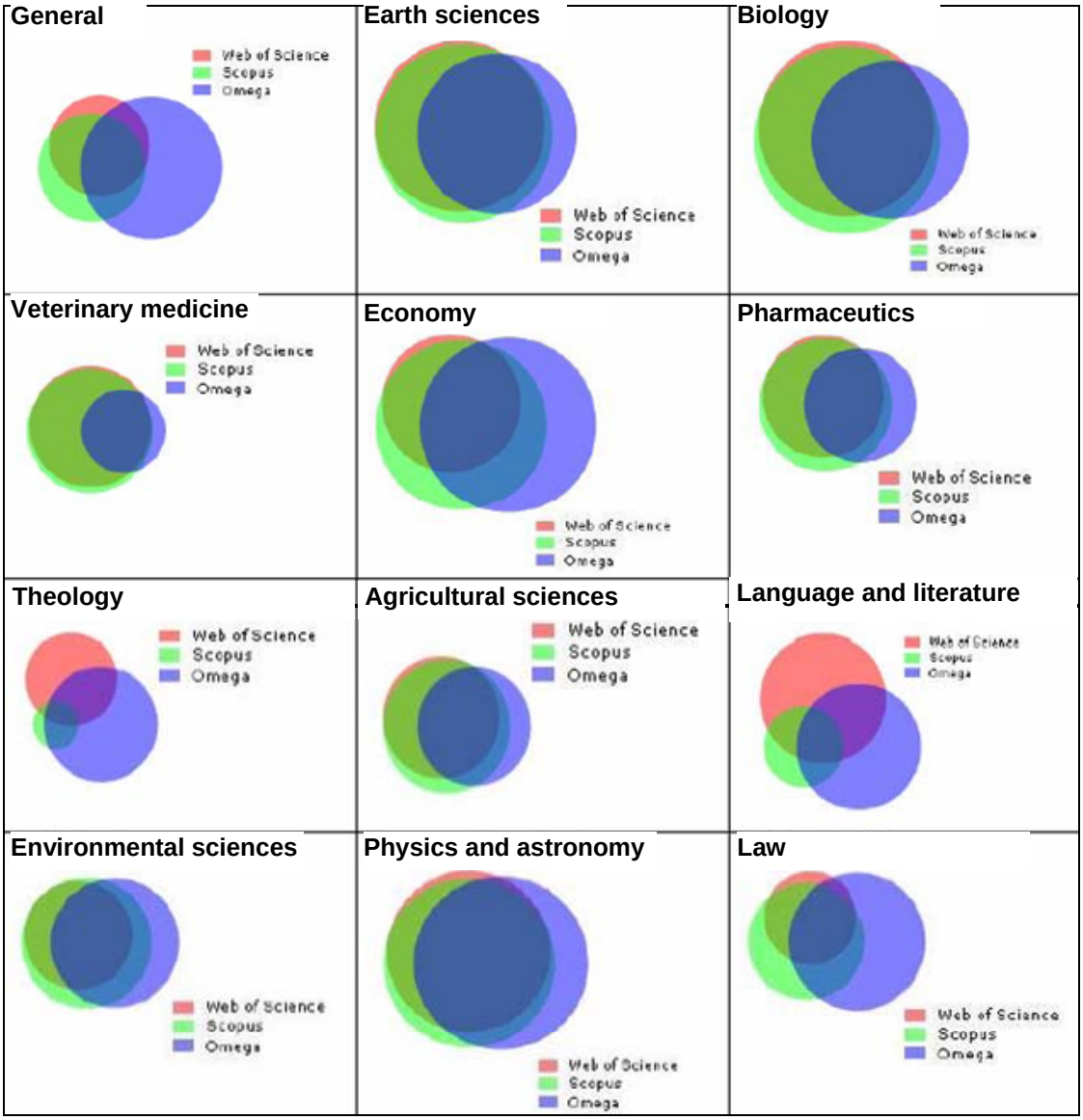
Source: own research

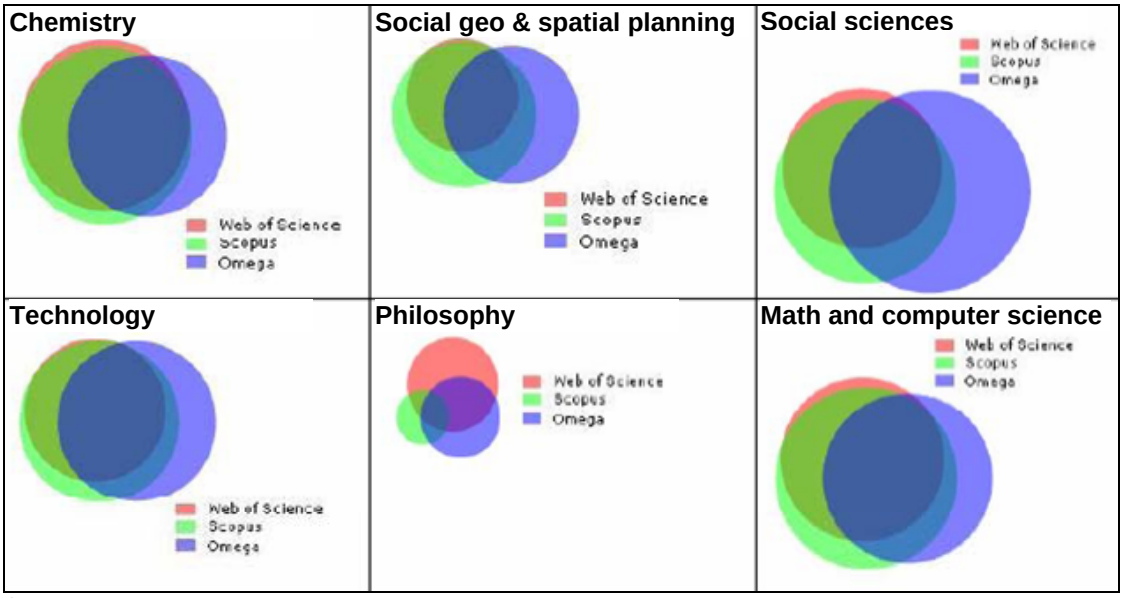
Under the restriction applied here to the journals available full text in the UBU, the more extensive coverage provided by Scopus is clear: overall the coverage in Scopus, at 64%, is more extensive by over 10 percentage points, for our digital journals holdings. With the exception of language and literature, philosophy and theology (according to the Omega classification) Scopus again demonstrates more extensive coverage in all subject areas than WoS. In 5 UBU subject areas (including 'general') the coverage provided by the search engine Omega is the largest of these three multidisciplinary access tools. Note that all journals were included in the full text journals of the UBU, including those with a moving wall for more recent years (JSTOR, PCI) and including comparatively large numbers of non-academic journals from EBSCO ASE. The latter represent a relatively large proportion in social sciences, technology and economics in particular. This entails high scores for the own search engine Omega in those fields, as all full text titles from EBSCO ASE are covered. The own Omega search engine is likewise strong in subject areas where a large proportion of the journals is concentrated with a few of the largest publishers (Elsevier, Springer and Wiley), for which Omega again provides full coverage. This applies for instance to environmental science.

For a good assessment of the coverage of journals in specific subject areas, it is important not only to consider the number of journals but also the overlap of journals covered. If two databases index an equal number of titles in a specific subject area, the degree of overlap will determine the choice for one of the two or for both databases. If databases do not index the same number of titles for a subject area but there is a full overlap, the obvious choice is to opt for the database with the largest number of titles. For most subject areas, however, there will be both an incomplete overlap and a difference in the number of titles covered (figure 4.3).

Overall it is clear that Scopus adds far more value to the Omega search engine in terms of accessing our holdings than WoS. Except in the classics, the number of titles featured only in Scopus is larger in all subject areas than the number unique to WoS. The overlap with the Omega search engine is usually much smaller than that between Scopus and WoS. The smallest overlap between WoS and Scopus is in the classics and the category "general" and in addition in theology, economics and social geography & spatial planning and in the social sciences and mathematics & computer science. In all other subject areas, i.e. physical or life sciences, there is a relatively strong overlap. It is important to determine for each subject area how important the titles in the non-overlapping area between WoS and Scopus are.

Figure 4.3. Overlap of indexation by WoS, Scopus and the Omega search engine of titles of journals held digitally by UBU, by UBU subject area, May 2006.





Source: own research

The exact numbers of journals of the full text UBU holdings that are included only in Scopus, only in WoS or in neither therefore vary strongly per subject area (table 4.5).

Table 4.5 Journals held digitally by UBU that are not included in Scopus, WoS or not in both, by UBU subject area, June 2006.

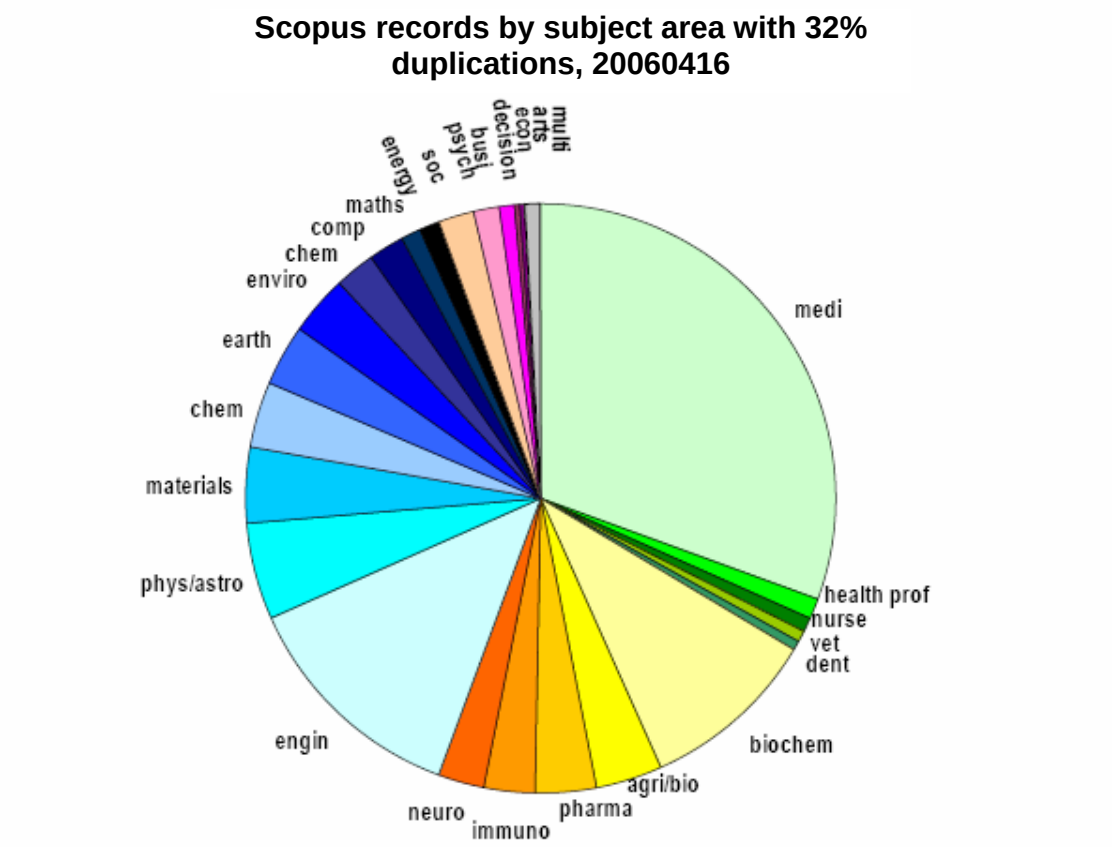
Subject	Not in WoS but in Scopus	Not in Scopus but in WoS	Neither in Scopus nor WoS
Earth sciences	20	5	57
General	14	21	191
Biology	63	29	212
Veterinary medicine	4	2	22
Economics	119	12	227
Pharmaceutics	14	2	26
Medicine	416	21	487
Theology	4	42	93
Agricultural sciences	8	2	18
Language and literature	36	229	486
Environmental science	21	0	46
Physics and astronomy	19	14	129
Law	13	6	157
Chemistry	18	14	82
Social geography and spatial planning	30	1	65
Social sciences	104	33	621
Technology	31	2	95
Philosophy	3	43	38
Mathematics and computer science	66	24	364

Source: own research

Subject specialists from a number of disciplines have looked at these lists with non-overlapping titles from WoS and Scopus and assessed to what extent they include crucial titles (for the titles classed as such by the subject specialists, see annex I). The UBU subject specialists marked crucial titles not in order to assess Scopus and WoS in quantitative terms, but to liaise on missing titles with colleagues in the faculties and with the suppliers of the databases (Elsevier and Thomson-ISI).

Finally, we look at the total number of records per subject area (figure 4.4). This is affected however by some duplication, as a journal will sometimes be classified under more than one subject area by Scopus.

Figure 4.4. Scopus records by subject area



Source: own research

The bias towards STM in the Scopus database is more distinct here than in terms of the number of journals. That strong emphasis on STM compared with quantification by the number of journals is due to the much higher average number of articles per year in STM journals and the fact that for these subjects, the coverage in Scopus extends further back in time and more years have therefore been indexed.

4.2 Coverage on the basis of sample searches

In addition to counts of indexed journals, specific searches provide a good reflection of the size of the two citation databases. This can yield a different view as it involves a count of the number of records, in which the number of indexed articles per journal plays a part. This depends on the number of published articles in these journals and the number of years covered in the database. The results (table 4.6) do in fact generate a different, divergent view compared to the previous results.

Table 4.6 Search results of three searches in default fields, per subject area, numbers of records in Scopus as % of WoS, April 2006

Subject	Exact search string	1988-1995	1996-2005	1988-2005
Earth sciences	geophysic* AND geolog*	342	264	287
Earth sciences	groundwater AND monitoring AND model*	277	193	206
Earth sciences	foraminiferal AND "north Sea"	107	80	88
Biology	(plant* OR animal* OR organism*) AND genera	186	295	269
Biology	learn* and songbirds*	111	135	132
Biology	root pattern OR "root patterns" OR "root patterning"	215	322	294
Veterinary medicine	veterinary	95	144	130
Veterinary medicine	embryogenesis AND bovine	61	44	50
Veterinary medicine	"animal diseases" AND vaccination	383	283	303
Economics	"foreign direct investment" AND competiti*	123	131	130
Economics	("early modern" OR "post-war") AND econom*	188	120	135
Economics	firm* AND merger* AND market*	38	170	144
Pharmaceutics	pharmac* AND receptor*	152	95	109
Pharmaceutics	polymers AND (liposomes OR "drug delivery systems")	264	223	230
Pharmaceutics	"drug targeting" AND "controlled release"	150	164	161
Medicine	cancer AND neuro*	162	96	108
Medicine	"lymphocyte development" AND thymus	151	63	79
Medicine	"endoplasmic reticulum" AND hormone*	110	74	87
Environmental science	enviro* AND pollut*	468	300	334
Environmental science	"food webs"	64	69	68
Environmental science	innovat* AND energ* AND biomass*	170	298	274
P & A: Physics	"string theory"	35	61	55
P & A: Physics	"condensed matter" AND optic*	102	111	110
P & A: Physics	stratocumulus AND "boundary layer"	59	53	54
P & A: Astronomy	telescop* OR asteroid* OR supernova* OR interstellar	59	100	89
P & A: Astronomy	magnetohydrodynamic* AND plasma*	105	124	120
P & A: Astronomy	"stellar winds" AND nebulae	41	91	73
Chemistry	molecular AND aromatic	121	149	142
Chemistry	"protein folding" AND (molecular	164	184	181

	chaperones OR Hsp90 chaperone)			
Chemistry	("phase behaviour" OR "phase behavior") AND (colloids OR rods)	205	109	121
Social geography & spatial planning	(geographical OR spatial) AND (urban OR economic)	324	178	204
Social geography & spatial planning	regional AND evolutionary* AND (business* OR compan* OR econom*)	156	115	119
Social geography & spatial planning	"southern africa" AND develop* AND econom*	383	329	342
Soc.sc: anthropology	anthropo*	119	129	127
Soc.sc: anthropology	(trauma* OR violen*) AND (ethnic* OR ethno* OR societ*)	138	131	132
Soc.sc: anthropology	migrat* AND ethnic*	376	153	194
Soc. sc: psychology	psychol*	152	157	155
Soc. sc: psychology	(aggression OR criminality) AND psycho*	162	110	120
Soc. sc: psychology	neuropsycholog* AND psychopatholog* AND cogniti*	110	86	90
Soc. sc: sociology	sociolog*	58	92	80
Soc. sc: sociology	gender* AND household* AND (labor OR labour OR work*)	118	96	100
Soc. sc: sociology	"life course" OR "life courses"	49	77	72
M & C: Computer science	"computational complexity" AND Bayesian	480	488	487
M & C: Computer science	computational AND geometr* AND virtual	208	622	587
M & C: Computer science	programming AND distributed	613	428	475
M & C: Mathematics	(algebra* OR arithmetic*) AND calculus	87	137	123
M & C: Mathematics	"Lie algebras"	8	68	49
M & C: Mathematics	ocean AND (eigenfunction* OR eigenvector*)	169	210	197

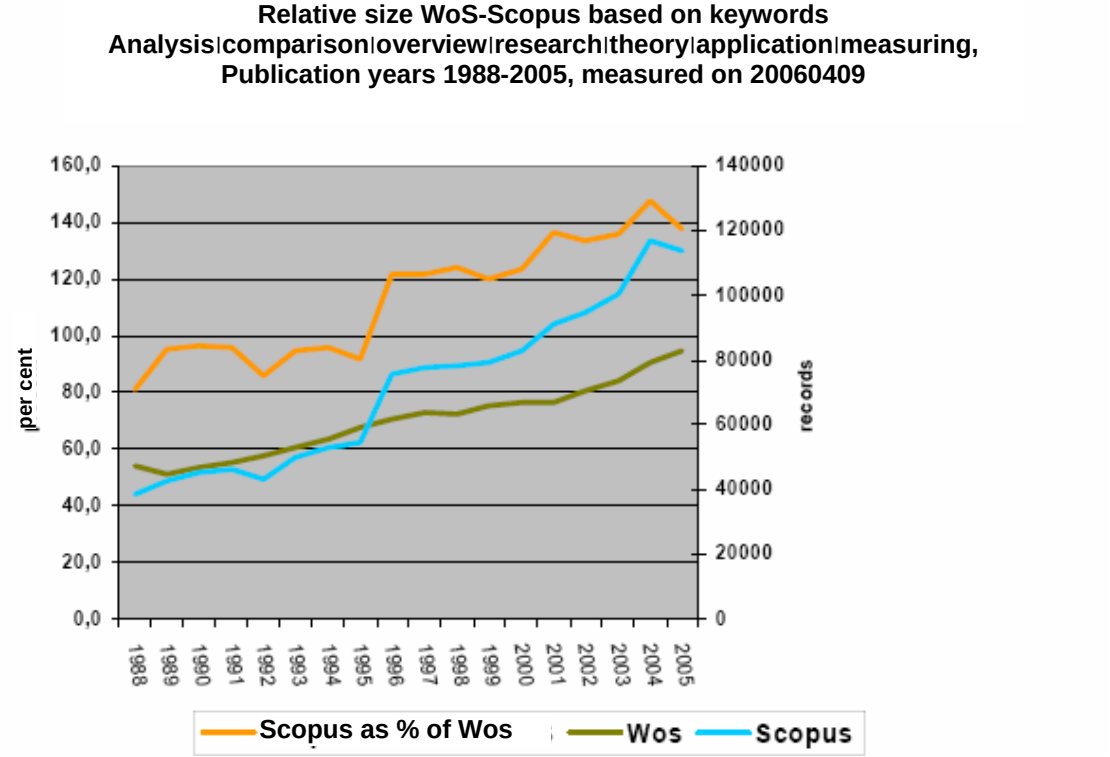
Source: own research; Note. Green = Scopus outscores WoS by more than 10%, red = WoS outscores Scopus by more than 10%, yellow = difference between number of search results of Scopus and WoS was 10 per cent or less.

There are a few notable aspects in the results of the searches. Overall, Scopus clearly produces more records for the majority of the searches than Web of Science. But this view is not complete without pointing out three other notable features. Firstly, there is the ambiguous outcome for veterinary medicine and medicine, where WoS scores better on some searches, and Scopus on others. Secondly, the poorer score achieved by Scopus for sociology, physics and astronomy. Thirdly, the extreme scores, which are not easily explained, of some searches in mathematics and environmental science.
Naturally, the small number of searches per subject area requires a caveat to ward off hasty conclusions. The automatic tuning of Scopus (which means for instance that plural forms are also included) and the Keywords-Plus of WoS have an effect on the numbers that is difficult to eliminate in interpreting results.

4.3 Period of coverage

The total size of Scopus compared to that of WoS measured over time can only be determined indirectly, as no totals are available per year for WoS. Nor does WoS permit searches on the year of publication only. As proxy we worked with a search for general, non-subject specific title words that even for recent years fails to produce a result in WoS of more than 100,000 hits, since WoS does not indicate the exact number in those cases. The result of this search initiative is reflected in figure 4.5. Before 1996 Scopus is smaller than WoS by some 5-15%, after 1996 it is larger by some 20-45%.

Figure 4.5. Relative size of coverage under WoS and Scopus per year, 1988-2005



Source: own research

The fact that Scopus definitely does go further back than 1996 is also clear from figure 4.6. Over 50% of the documents were published before 1996. Admittedly, citation data is shown only for the documents from the publication year 1996 onwards. Older documents do also figure as source documents for citation counts.

Table 4.7 Elsevier databases on which Scopus is based to a significant extent

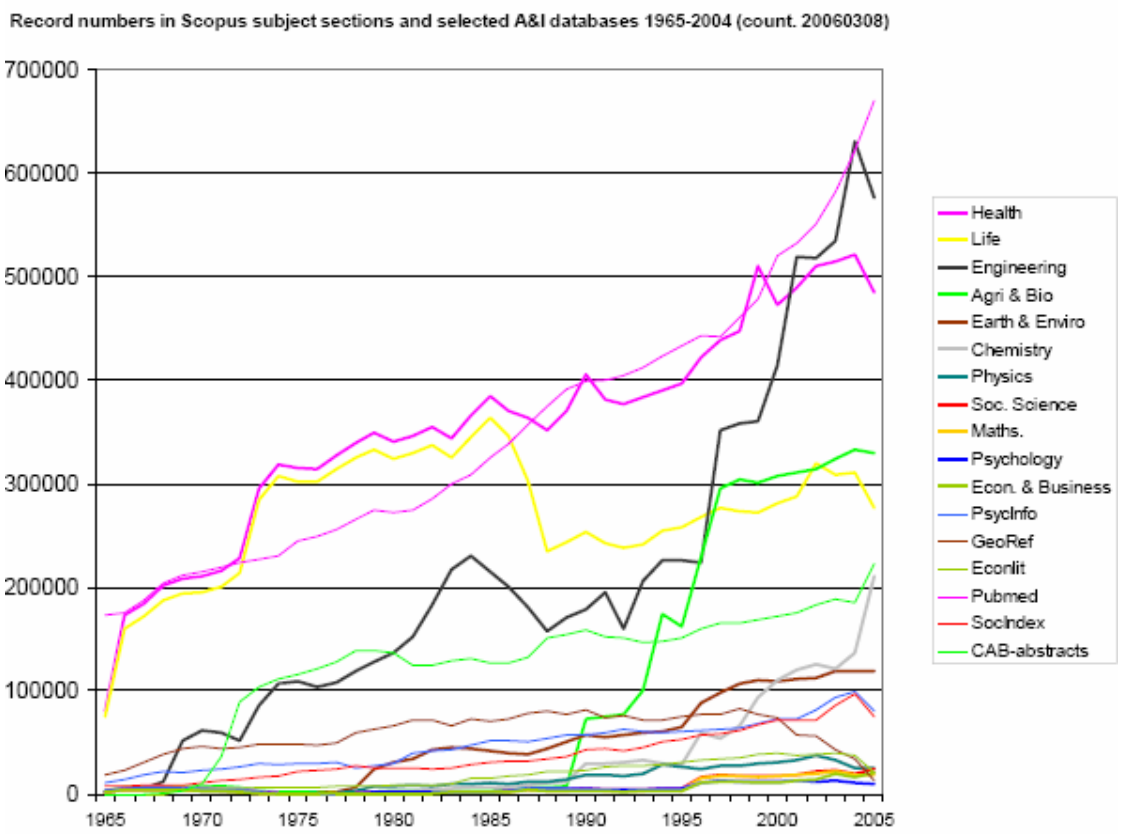
Database	Coverage period
Medline (via EMbase)	1966-…
EMbase	1970-…
Compendex	1970-…
World Textile Index	1970-…
Fluidex	1974-…
Geobase	1980-…
Biobase	1994-…

Source: Goodman & Deis 2005

Coverage before 1996 appears to derive above all from the databases Elsevier already had: EMbase, Biobase, Geobase and Compendex, in addition to smaller databases in the fields of liquids (Fluidex), oceanology (Ocean Base) and textiles (World Textiles) (table 4.6). As a result it provides good coverage for life, health, agri/bio and earth/enviro and also technology (engineering) further back in time. By contrast the coverage for psychology, economics and social sciences, but also for mathematics, is very limited for publications from before 1996. The content of Scopus before 1980 is in fact really largely biomedical.

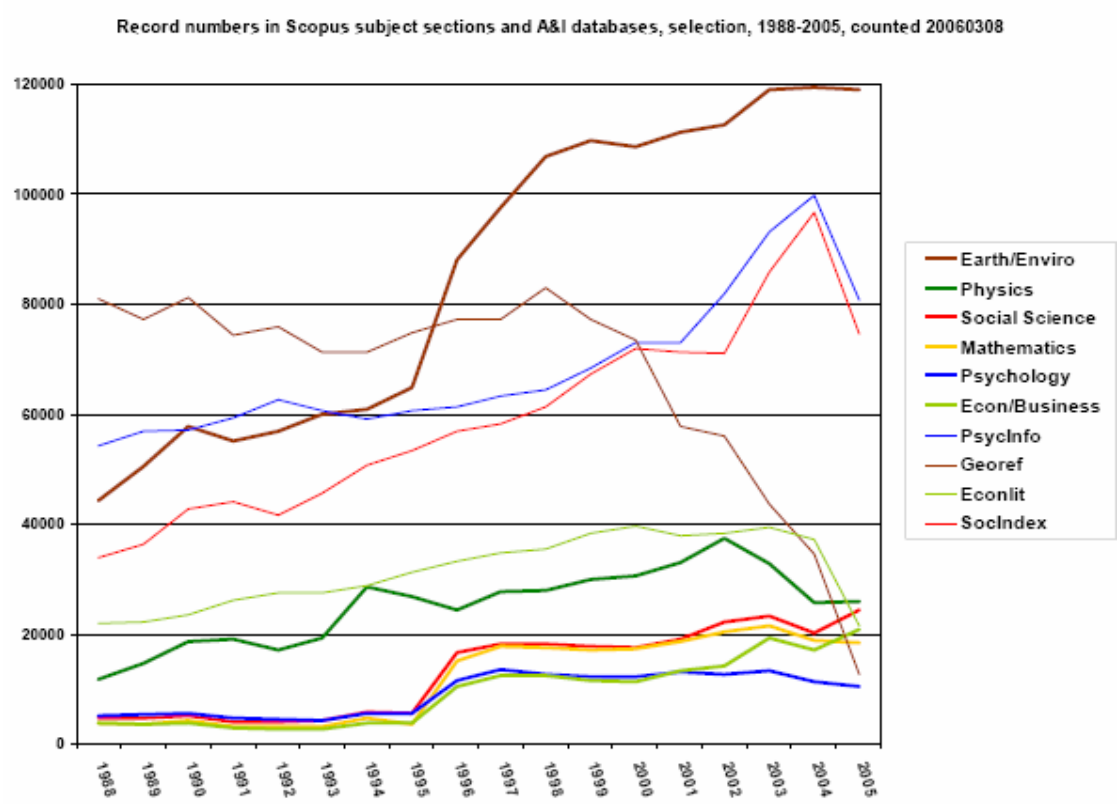
Before 1966, coverage in all subject areas was minimal. This is a difference compared to WoS, which in our version does not stretch back beyond 1988, but in principle (for science coverage) reaches back as far as 1945 and, since recently, even to 1900 for over 200 journals. In Utrecht we do have other ways of digital access to old journals in JSTOR, PCI, at the publishers' platforms Science Direct and SpringerLink, in some subject indices (for instance Georef, PsycInfo and Zentralblatt MATH), for some journals in Online Contents and also for material from some publishers via Omega. The limited coverage for sociology especially in the period prior to 1996 (clearly visible when zooming in on the period 1995-2005 in figure 4.7) finally, is a drawback compared to WoS. Note however that these figures are from March 2006. Scopus has selectively continued its 'backfill' in the period since then.

Figure 4.6. Number of records of the various Scopus subject areas and of selected subject indices, 1965-2005.



Source: own research; NB: since this count was performed the subject classification in Scopus has been refined and modified, as a result of which these counts can no longer be reproduced

Figure 4.7. Number of records of the various Scopus subject areas and of selected subject indices, 1965-2005 (magnified portion of figure 4.6).



Source: own research

An indication of the 'backfill', the inclusion of publications from before 1996, is provided by the ratio of the number of recent records and the number of older records (table 4.8).

Table 4.8. Ratio between records 1986-1995 and 1996-2005 as an indication of 'backfill' of the database, counted on 20060627, on the basis of the new detailed Scopus subject classification (including double counts).

Scopus subject area	Records 1996-2005	Records 1986-1995	1986-1995 as % of 1996-2005
Medicine	3871121	3002118	78
Environmental Science	514086	385134	75
Pharmacology, Toxicology and Pharmaceutics	503491	370771	74
Dentistry	66562	48766	73
Immunology and Microbiology	477611	322546	68
Biochemistry, Genetics and Molecular Biology	1565869	1039927	66
Earth and Planetary Sciences	576363	381365	66
Neuroscience	396825	261343	66
Nursing	125955	78039	62
Health Professions	199752	121077	61
Engineering	1964645	1160747	59
Multidisciplinary	123439	59812	48
Veterinary	131972	57236	43
Psychology	257340	105019	41
Energy	233501	90820	39

Materials Science	921774	328310	36
Agricultural and Biological Sciences	958706	278408	29
Decision Sciences	59071	16850	29
Chemical Engineering	552390	157464	29
Computer Science	509072	145008	28
Physics and Astronomy	1379044	341580	25
Chemistry	935545	222622	24
Social Sciences	492138	114669	23
Arts and Humanities	44248	7460	17
Economics, Econometrics and Finance	115481	14059	12
Business, Management and Accounting	230888	25477	11
Mathematics	405098	39124	10

Source: own research

It is clear that there are comparatively few records in Scopus for subject areas for which there is no underlying database: sociology, social sciences, as well as chemistry, mathematics, and physics and astronomy. Despite the absence of an underlying Elsevier database in the field, psychology has a reasonable backfill, which is probably based on journals (also) included in EMbase.

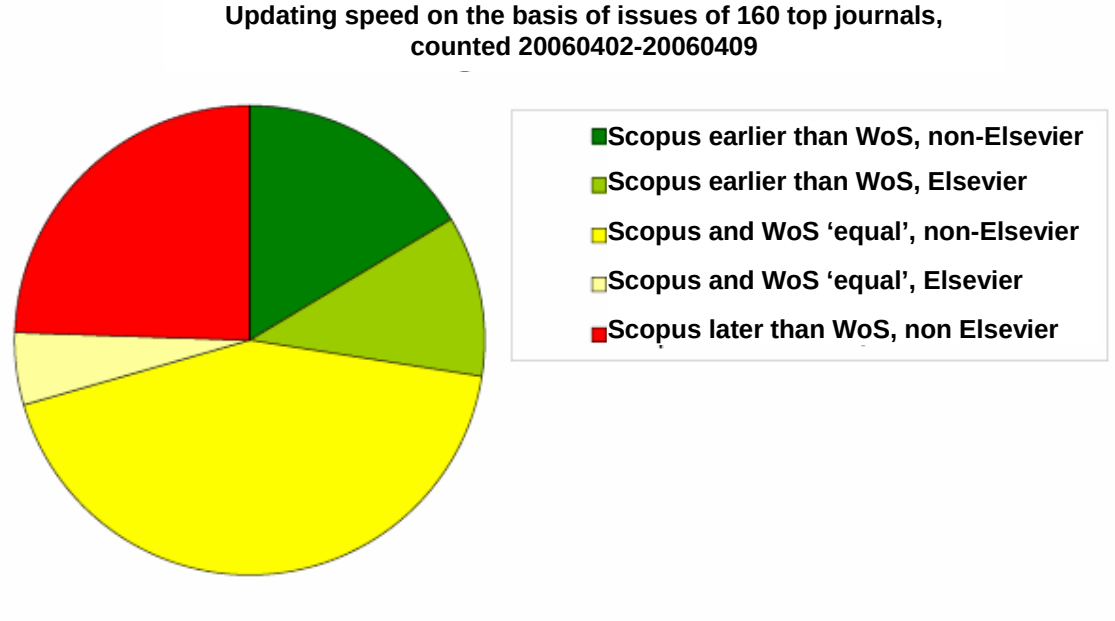
4.4 Updating

The alleged (Goodman & Deis 2005) difference in updating speed of the databases Scopus and WoS, with the latter outscoring the former, is not evident at present from our analysis of indexed issues of 160 top journals. For this analysis, we selected the two titles with the highest impact factor from 80 categories that are important for our university in the Journal Citation Reports (55 science and 25 social science). It is clear that for almost half of these journals Scopus and WoS are evenly matched; for just over a quarter, Scopus is more up-to-date, for just under a quarter WoS is more up-to-date (figure 4.8).

It was to be expected that there would be a disproportionate number of Elsevier titles for the journals on which Scopus is more up to date than WoS.

Pipp (2006, p. 14) has carried out a similar test for a smaller number of journals, but additionally considered how far both WoS and Scopus lag behind the publication of the platform on the publisher's platform (Blackwell Synergy, SpringerLink, Scienc Direct, etc.). She concludes that Scopus is slightly more up to date than Web of Science, but that Scopus lags far behind for a small number of journals. Scopus does not appear to have its workflows in order yet for all titles.

Figure 4.8. Updating speed of Scopus and WoS on the basis of availability of issues of top journals from the Journal Citation Reports.



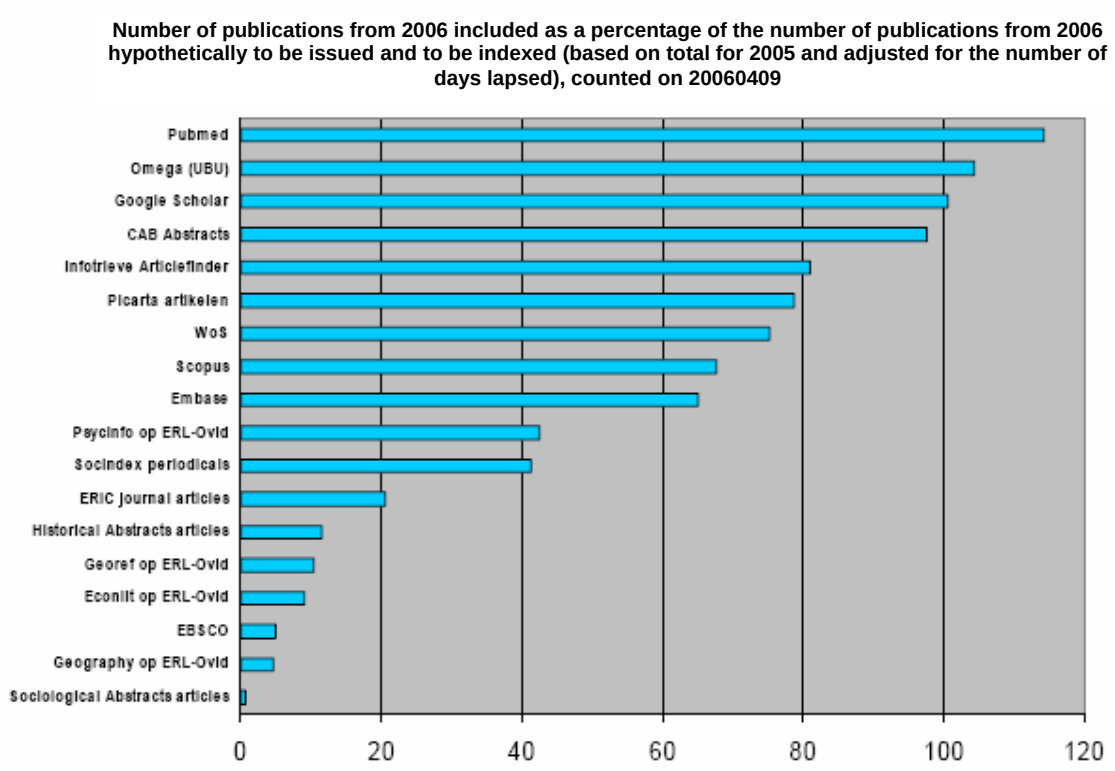
Source: own research

Another way of considering updating speed is to compare the number of records published in the current year with that of the past year, adjusting for the portion of the current year still to come. In view of the time required for processing it is to be expected for instance that after four months one third of the number of titles for that year will only rarely have been included in the data base. At the same time, conversely, a relatively large number of indexed documents can be expected, owing to the growth of the absolute number of publications from year to year. The picture in figure 4.9 therefore must only be interpreted in terms of the comparison. A level of 100 means that if the number of publications in 2006 were equal to that in 2005, the makers of the database are hypothetically exactly on track to have included all publications of this year by the end of the year. The issue is the difference in the extent to which databases deviate from that level of 100. Differences are potentially attributable to effectively slower processing, or a larger proportion of less frequently published titles in the database. Genuine distortion will only be produced if the databases differ in the degree to which they have included new titles starting from 2006 for which they index only 2006. The same would apply if databases differed sharply in the degree to which the indexation of certain titles had been discontinued in 2006.

Among the large databases, Scopus is clearly a mid-ranking performer, slightly lower than WoS. We have already seen that in terms of indexed issues of top journals there is no significant difference in updating speed. It is difficult to conclude whether this is due to slower processing. That is because the lower ranking attained by Scopus is likely to be caused in part by the fact that Scopus, on the basis of its underlying databases for some subject areas, also indexes much less frequently published journals. Conversely, WoS imposes a tight publication regime as a condition for including journals. We can expect the same

influence of less frequently published journals to affect the genuine subject indices, in addition to the fact that some of these bibliographies are also updated less frequently.

Figure 4.9 Updating speed of databases: ratio current year / past year, 2006.



Source: own research

4.5 Nature of the data included per document

In addition to coverage and updating speed, the data included for each reference is likewise a major factor in comparing bibliographical databases. This often offers substantial added value for professional use over simpler search entries such as Google Scholar or Online Contents, for instance address data for authors and (searchable) keywords. A comparison of one record in different databases (table 4.9) provides an instructive first impression of the differences.

Comparing the fields in which the databases differ clearly shows that the large citation databases are the most comprehensive in terms of the quantity of information per record. In both Scopus and WoS the records are more comprehensive than in most other bibliographical databases. That often also applies to the number of keywords. At the same time, there are major differences in the availability of keywords in Scopus and WoS.

In addition to author keywords, Scopus often includes keywords from controlled vocabulary deriving from underlying databases such as Compendex, Geobase and EMbase/Medline. These records are therefore also easy to find for experienced researchers who are used to using this vocabulary. This vocabulary is not individually searchable per term within Scopus. The terms form part of one single keywords field. The Scopus records that do not derive from an own underlying database often only contain author keywords.

Table 4.9 Differences between present fields for 1 article in parallel records of 7 bibliographical databases

Etienne, S (2003) Ecological impact in data-poor systems: a case study on metapopulation persistence in selected databases, 20060414							
Fields showing a difference between the databases	**Biosis** silverpl.	**Embase** silverpl.	**Geobase** silverpl.	**Medline** silverpl.	**Scopus**	**WoS**	**Springer**
Institution, address	1	1	1	1	All*	All*	3
DOI/View at publisher	N	N	N	N	Y	Y	Y
Document type	Y	Y	N	Y	Y	Y	N
Copyright	N	Y	Y	Y	Y	Y	N
Language	Y	Y	Y	Y	Y	Y	N
Number of literature references	N	N	Y	N	Y	Y	N
Literature references	N	N	N	N	Y	Y	Y
Total number of keywords	**44**	**14**	**10**	**6**	**30**	**15**	**5**
Comments on keywords	High number due to biological species thesaurus				5 author and 25 from other sources (Compendex/ EMbase/ Geobase / Medline)	5 author and 15 generated from references (keywordsplus)	

Source: suggested by Van Laarhoven (UB Groningen)
* often only 1-3 addresses for older years; ** authors not linked 1-to-1 to affiliations and often only 1-3 addresses for older years

WoS, on the other hand, does not have terms from controlled vocabulary or specialised word systems in addition to author keywords, but it does have the 'keywords plus' as standard, terms that are generated automatically on the basis of frequently occurring words and concepts in the titles of literature referred to in an article. A comparative study (Qin 2000) appears to indicate that both keywords-plus and the 'controlled vocabulary' (e.g. thesaurus terms) usually cover the main concepts in an article and that the supplementary terms in both offer added value of their own.

Another important aspect of the records is the abstract. This is covered by default by both WoS and Scopus. Evidently, abstracts are not available for some forms of publication, but there are no records for some journals in Scopus either: sometimes because they are not included (in the specialised, 'industry journals'),

and sometimes because Scopus was not able to include them. Jasco (2006) estimates that 20 of the 28 million records contain abstracts. Tests of our own (searching for a|the|an in the abstract) confirm that at least 70% have abstracts: 19.48 million of the total of 27.97 million records (June 2006).

Compared with subject bibliographies, Scopus and WoS often lack specific subject related fields. Examples of data lacking completely are:

- Geographical co-ordinates, as in Georef;
- Molecule structures, as in Chemical Abstracts (in UU via SciFinder Scholar); note that linking with Crossfire-Beilstein is possible;
- Age category, "population group" and methodology used, as in PsycINFO.

Some data are available but are included in the wider keywords field and are hence not separately searchable nor consistently included as in the subject bibliographies referred to above. This applies to:

- Geographical locations (as used in Geography/Geobase, Econlit and CAB);
- Biological species (as used in Geography/Geobase and CAB).

Other data are available and separately searchable, such as CAS registry numbers for chemical compounds (also used in Chemical Abstracts, Pubmed, EMbase and CAB) and genetic sequences.

4.6 Citation data: the coverage of "citing articles"

In addition to the 28 million records, Scopus claims to include another 245 million references to literature from those records. A portion of these refers back to one of the 27 million records. However, data on incoming citations are linked only to records from 1996 onwards. Jascó (2006) puts the number of 'citation enhanced' records at some 9.5 million.

With a view to using Scopus as a citation index it is important to know how the system performs in searching for 'citing articles'. Given the nature of the citation searches, a comparison with other citation indices is the only way to obtain quantitative data on this. Earlier research (Bakkalbassi 2006) has already established that different indices for different subject areas produce divergent results in terms of citation quantity. An added complication is that comparison of figures alone is not sufficient. That is because if Scopus finds 40 citations for an article and another citation does so as well, these will not necessarily be the same articles. That is why a more detailed examination was carried out in which the citation found with different systems were also compared at the level of individual articles. Given the complexity of this method of comparison, only a somewhat limited sample was taken.

The examination was performed as follows:
Citation indices compared: Scopus, Web of Science, Google Scholar.

Number of reference documents compared: 64 articles included in all three systems as articles themselves; to ensure results are sufficiently comparable, we did not consider articles that are only quoted in Web of Science without being included in it.

Other criteria were:

- Given the timeframe covered by Scopus as citation index, 32 articles were selected from 1995 and 32 articles from 2000;
- For each of the 18 in UBU subjects, 4 articles were selected, 2 from 1995 and 2 from 2000; as it was not possible to find enough articles for the subjects theology and philosophy that met the other criteria, those two subjects were eliminated, leaving 16 subject areas;
- the titles of journals from which the articles were sourced were spread evenly across the alphabet;
- in connection with the manual comparison at the title level of the citing articles, 32 articles were selected that are not cited more than 50 times in Scopus and 32 that were not cited more than 50 times in Web of Science;
- to be able to compare a sufficient number of citations, articles were selected in the same manner that are cited at least 30 times.

The quantitative data obtained in this way were summarised (table 4.10 & 4.11). The total number of citing articles per category is cited. In addition to a classification by the 16 subjects, totals are also provided per year of publication and by a broad arts, science and socio-economic science classification.

Table 4.10 Citations of selected articles in Scopus, WoS and Google Scholar, total and per subject area, with overlap data, April 2006.

Number of citing articles	Cumula-tive total	Scopus	Unique Scopus	WoS	Unique WoS	Google Scholar	Unique Google Scholar	Overlap S-W	Overlap S-G	Overlap W-G	Overlap S-W-G
All	4135	2733	242	2581	221	2671	1120	2301	1492	1360	1301
1995	2063	1372	124	1310	120	1273	543	1161	702	643	615
2000	2072	1361	118	1271	101	1398	577	1140	790	717	686
Science	2489	1933	177	1787	88	1501	428	1658	1033	975	935
Socio-economic	1437	667	54	657	118	1038	638	528	388	314	302
Arts/hum	209	133	11	137	15	132	54	115	71	71	64
Earth sc.	216	169	9	163	5	128	39	154	86	83	80
Biology	211	188	9	177	6	154	16	170	137	129	128
Veterinary med.	237	171	8	172	6	145	56	162	85	88	84
Economics	514	205	17	174	8	435	299	164	134	112	110
Pharmaceutics	229	182	22	154	6	139	41	148	98	86	86
Medicine	263	202	29	193	21	156	34	166	116	115	109
Agricultural	204	164	14	149	4	125	32	141	89	84	80
Lang. & Lit.	209	133	11	137	15	132	54	115	71	71	64
Environmen	207	179	16	166	8	116	17	155	96	91	88

tal Sc.											
Physics & Astronomy	204	175	7	181	13	64	13	165	48	48	45
Law	263	98	5	168	76	145	82	85	56	55	48
Chemistry	185	173	4	174	5	98	7	169	91	91	91
Soc. Geogr.	317	170	15	135	10	225	136	126	89	59	59
Soc. Sc.	343	194	17	180	24	233	121	153	109	88	85
Technology	187	151	23	133	7	108	24	121	79	77	72
Mathematics & Comp.	346	179	36	125	7	268	149	107	108	83	72

Table 4.11 Citations of selected articles in Scopus, WoS and Google Scholar, total and per subject area, with overlap data, index (Scopus=1 .00), April 2006.

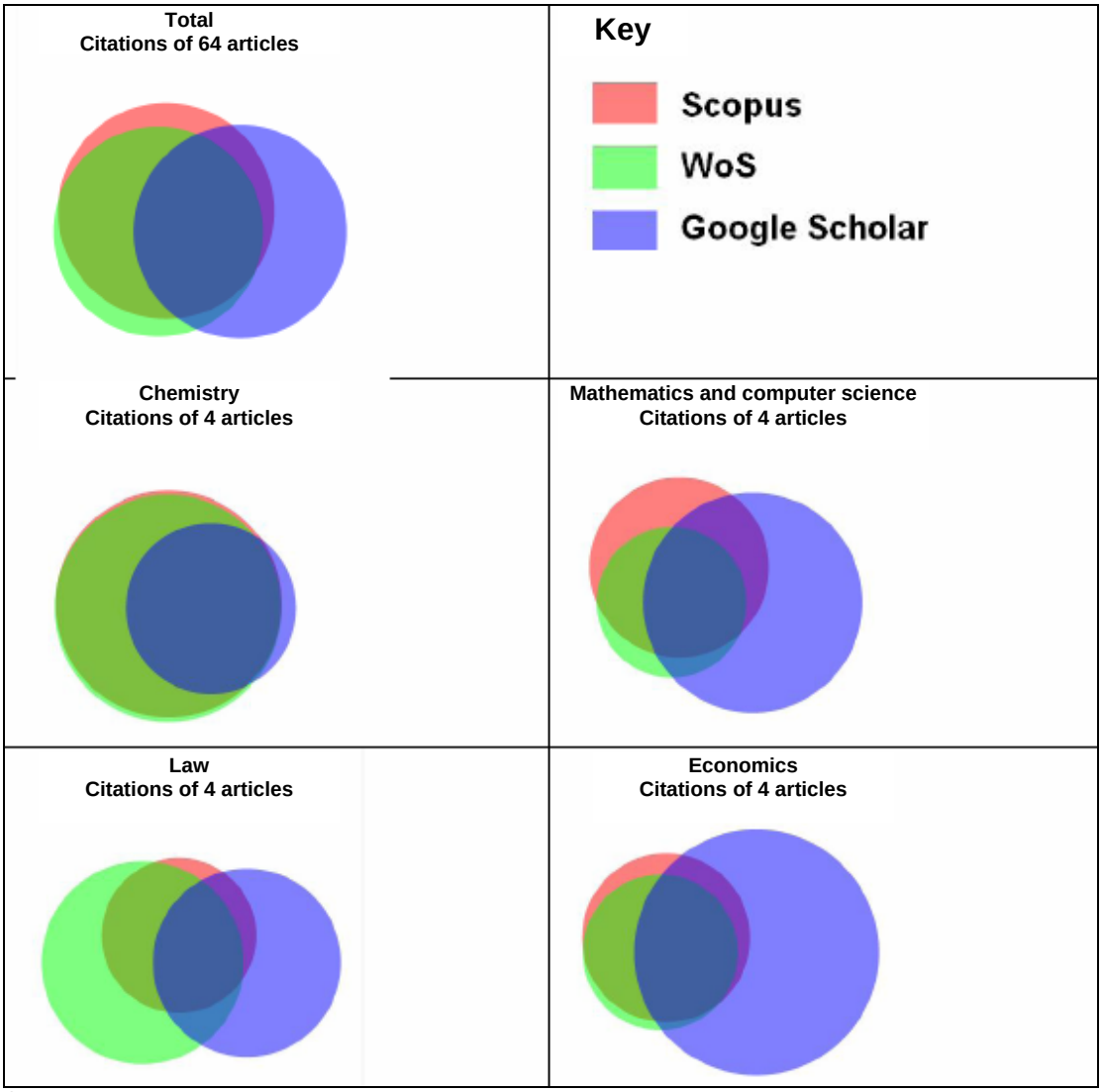
Number of citing articles	Cumula-tive total	Scopus	Unique to Scopus	WoS	Unique to WoS	Google Scholar	Unique to Google Scholar	Overlap S-W	Overlap S-G	Overlap W-G	Overlap S-W-G
All	1.51	1.00	0.09	0.94	0.08	0.98	0.41	0.84	0.55	0.50	0.48
1995	1.50	1.00	0.09	0.95	0.09	0.93	0.40	0.85	0.51	0.47	0.45
2000	1.52	1.00	0.09	0.93	0.07	1.03	0.42	0.84	0.58	0.53	0.50
Science	1.29	1.00	0.09	0.92	0.05	0.78	0.22	0.86	0.53	0.50	0.48
Socio-economic	2.15	1.00	0.08	0.99	0.18	1.56	0.96	0.79	0.58	0.47	0.45
Arts/hum	1.57	1.00	0.08	1.03	0.11	0.99	0.41	0.86	0.53	0.53	0.48
Earth sc.	1.28	1.00	0.05	0.96	0.03	0.76	0.23	0.91	0.51	0.49	0.47
Biology	1.12	1.00	0.05	0.94	0.03	0.82	0.09	0.90	0.73	0.69	0.68
Veterinary med.	1.39	1.00	0.05	1.01	0.04	0.85	0.33	0.95	0.50	0.51	0.49
Economics	2.51	1.00	0.08	0.85	0.04	2.12	1.46	0.80	0.65	0.55	0.54
Pharmaceutics	1.26	1.00	0.12	0.85	0.03	0.76	0.23	0.81	0.54	0.47	0.47
Medicine	1.30	1.00	0.14	0.96	0.10	0.77	0.17	0.82	0.57	0.57	0.54
Agricultural	1.24	1.00	0.09	0.91	0.02	0.76	0.20	0.86	0.54	0.51	0.49
Lang. & Lit.	1.57	1.00	0.08	1.03	0.11	0.99	0.41	0.86	0.53	0.53	0.48
Environmental Sc.	1.16	1.00	0.09	0.93	0.04	0.65	0.09	0.87	0.54	0.51	0.49
Physics & Astronomy	1.17	1.00	0.04	1.03	0.07	0.37	0.07	0.94	0.27	0.27	0.26
Law	2.68	1.00	0.05	1.71	0.78	1.48	0.84	0.87	0.57	0.56	0.49
Chemistry	1.07	1.00	0.02	1.01	0.03	0.57	0.04	0.98	0.53	0.53	0.53
Soc. Geogr.	1.86	1.00	0.09	0.79	0.06	1.32	0.80	0.74	0.52	0.35	0.35
Soc. Sc.	1.77	1.00	0.09	0.93	0.12	1.20	0.62	0.79	0.56	0.45	0.44
Technology	1.24	1.00	0.15	0.88	0.05	0.72	0.16	0.80	0.52	0.51	0.48
Mathematics & Comp.	1.93	1.00	0.20	0.70	0.04	1.50	0.83	0.60	0.60	0.46	0.40

The overall conclusion of the sub-assessment of the citation data is that the differences in terms of their coverage between Scopus and Web of Science are largely very small; the differences between Google Scholar and these two commercial citation indices were substantially wider. Considered in greater detail the following conclusions can be drawn, albeit hesitantly, because of the relatively small sample:

- No difference was found between older (1995) and more recent publications (2000).
- For most of the subject areas, the overlap between Scopus and WoS in citing articles is between 80 and 90% (based on the number of Scopus citations); in veterinary medicine, physics & astronomy and (especially) chemistry the overlap is even greater; in social sciences and Social geography & Spatial planning the overlap is slightly lower, and substantially lower still in Mathematics and Computer science.

- Only for Law is the number of citations found in WoS significantly higher than that in Scopus. Related to the WoS totals, the overlap for that subject area would therefore also have been much smaller.
- For Mathematics & Computer science and to a lesser extent for Social geography & Spatial planning the number of citations found in Scopus is significantly higher than in WoS.
- With regard to Google Scholar – though this is less relevant to the current comparison – it may be concluded that the significantly smaller overlap with the results from the commercial databases is caused mainly by the much higher numbers of publications in languages other than English and other document types that are included in Google Scholar; for economics, law, social geography & spatial planning and mathematics & computer science in particular the numbers of citations found are significantly higher than for Scopus; for most (other) science subjects those numbers are in fact much lower.
- The results of these citation samples for the individual subjects correspond quite closely with the coverage data from section 4.1.3.

Figure 4.10 Overlap of citations for all cited articles between Scopus, WoS and Google Scholar and for articles from a number of selected subject areas with various forms of overlap, April 2006.



Source: own research

5 Search functionality, interface, speed and ease of use

5.1 Search functionality

Scopus and WoS have much in common in several respects, including functionality. Often, the only difference is the design or location of certain items on the screen or within the site. While that can make a difference in terms of look and feel on an individual level, it is difficult to provide a general assessment of these factors. We will therefore focus here on the functionality that differs, i.e. possibilities provided by one of the systems but lacking in the other (table 5.1).

Table 5.1. 'Hard' differences in functionality (as at June 2006)

Possible in Scopus, not in WoS	Possible in WoS, not in Scopus
- default refine (parametric search result)	- classification by country, town, affiliate
- citation table (citation tracker)	- link to Journal Citation Reports
- search for "all"	- link to Crosssearch if 0 results
- Search for casreg numbers (chemical substances)	- link to Current Web Contents
- integration with Beilstein (molecule structures)	- truncate within phrase: neural network
- proximity searches with PRE and W	
- search only in ABS or KEYW of AUTHKEY	
- search for genetic sequences	
- use Author Indifier(expected for WoS)	
- link to Scirus	
- automatic tuning (plural forms)	
- UBUlink direct in list shown of references	
- list of journals browsable per subject	
- list of journals searchable by publisher	

Scopus is slightly more versatile than Web of Science. The standard refine bar is a particular advantage, not only for searching but also (for students) for developing a grasp of a field of research by categorizing the search results by source journal, year of publication, author and overlap of subject areas yielded by a search. A second major advantage of Scopus is the ease with which systematic citation overviews can be created for an author or subject. This option is too difficult to find, however.

The Author Identifier of Scopus, finally, is a long-cherished option to be able to cluster similar and separate dissimilar authors. Using an algorithm applied to factors including affiliation, co-authors and citations, various notations of the name of a single author are grouped together and different authors with exactly identical names and initials are separated. Every author is given a unique ID. The system does work, but is not (yet) able to cluster all author names. It does provide an easy way to give feedback. An added advantage of the author identifier is that where author affiliations are not always known for older articles, affiliations can nonetheless be found via a link to the author data. Web of Science has announced it will be introducing something similar to the Author Identifier.

Web of Science offers slightly more options for advanced citation analysis of entire organisations, especially the options to classify search results by affiliation: the organisation for which the author works and the town and country where this organisation is established. Both companies are currently working hard on improving their search systems. That is in any case a positive effect of the competition introduced by the arrival of Scopus.

5.2 Interface, speed and ease of use

The speed of search systems is an essential element of their 'feel'. Whether or not people like working with a system will depend in part on its speed. Evidently, speed depends on numerous factors: the server, the connection, the browser, other tasks being performed simultaneously by the PC and the specific task given to the search system. In a simple test, we kept as many of these variables constant for four search systems (table 5.2). Both Scopus and WoS have complex interfaces that are relatively slow to materialise, both on starting up the system and for searches. Although the Omega system developed in Utrecht is very fast, it was Google Scholar that proved to be the fastest on most searches. That certainly plays a part in the enthusiasm many people display for this search engine. Naturally, the fact has to be taken into account that the nature of the information shown and the functionality of Scopus and WoS in the phase used to select from the search results can yield time gains as compared to Omega and Google Scholar. The differences between Scopus and WoS are very small.

Table 5.2. Speed of broad article search systems, June 2006, page building in seconds

	Google Scholar	Omega search engine	Scopus	Web of Science
Building search screen after entering URL	<1	5	7	10**
"lymphocyte development" AND thymus	1	2 ***	8	13
trauma AND psychopathology	1	1	10	8
embryogenesis AND bovine	2	1	8	9

NB Performed with MSIE 6.0, all other programmes closed, within 5 minutes of each other. Figures represent the average of two identical searches on different days. Searches in standard fields, number of results per page at 100.
* Number of results per page 50 (maximum); ** Including clicking once from Portal to search screen; *** Omega does not support phrase search.

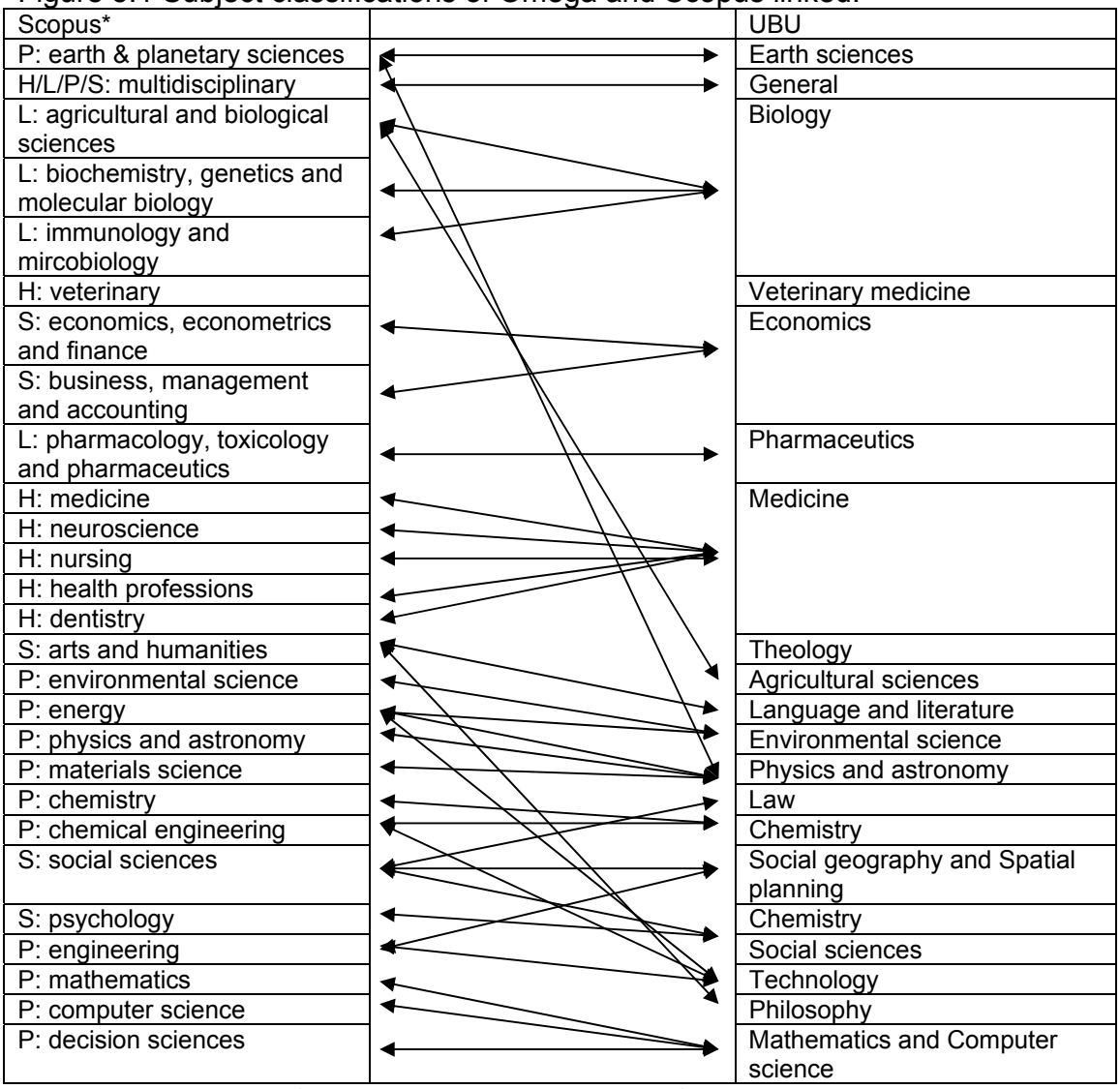
Another factor besides speed that determines ease of use is simplicity in navigation and layout. We did not test these ourselves as part of this study. User ratings do show however that as a rule the set-up for Scopus is felt to be clearer than that for WoS. In view of the rapid development of the user interfaces of both systems, this would not appear to be a critical issue in choosing between them.

5.3 Subject classification

With multidisciplinary databases, it is convenient in some searches to be able to zoom in on documents from a specific subject area. If you are completely certain

this can be entered directly as a limiter in a search, but usually this will only be done in the second instance to restrict search result to a specific subject perspective. For instance one might want to limit articles on genetics to those written in the field of social sciences instead of the medical or biological field. Or the intention could be to study aspects of transport from an economic instead of technical perspective. In those cases it is important to know what the demarcations of the subject classification in a database are and how these relate to known classifications. Scopus applies a classification into 27 subjects, which in turn are clustered into four directions (health, life, physical and social). Following a search, the refine tool will show immediately in which the fields the topic appears to be studied.

Figure 5.1 Subject classifications of Omega and Scopus linked.



* H=health sciences; L=life sciences; P=physical sciences; S=social sciences

The 27 subject areas are formed by classifying journals (as a whole), which means that all articles in that journal are classed under the subject concerned. At the same time, a journal can be classed under more than one subject. This applies to more general journals (Nature, for instance), but also to more specialised ones. In total there is around 30% duplication of records included in Scopus. This is clearly visible in following a search by zooming in on a subject area by using the refine tool. The subject turns out to contain records of other subject areas as well.

In the absence of a generally accepted classification of scientific fields of study, and because there is never a one-on-one fit between classifications, it is useful to examine how the (Omega) classification used in the UBU relates to that used in Scopus (figure 5.1). Most UU subjects are broader than those in Scopus, meaning that selection of relevant subjects in Scopus is comparatively easy and provides good filter options to include only relevant material in the selection without excluding too many relevants. The classification used in Scopus is much broader only for social geography & spatial planning: social sciences. Accordingly the Scopus classification is of little use for people working in these fields.

6 User ratings

6.1 Interviews with researchers performing frequent searches

To understand the way in which demanding researchers use citation databases we conducted interviews with 22 people (see table 6.1). The itemised reports of the interviews are presented in annex III. The interviews were also attended by the UBU subject specialist concerned, in addition to a member of the Scopus research group.

Table 6.1 Interviews with heavy users

Faculty	Number of researchers interviewed
Science	8
Veterinary medicine	3
Humanities	0
Medicine	2
Geosciences	4
Language, Literature & Arts	0
Law	0
Social sciences	5

The overall picture that emerges from the interviews with heavy users is one of substantial variation, both in terms of databases used and in terms of the ways in which they are used. Some researchers are very selective in their searches, directed mainly at known items and citations. Others search to a far greater extent on content terms and also substantively wider, by way of inventory. It also became clear that the UU citation databases are sometimes used for non-subject purposes: as verification for Metis input, as an aid in planning research programmes and as supplier of primary data for research into the innovative potential of countries and regions or in the field of sociology of science.

Scopus is praised virtually unanimously for its interface, clear screen structure and navigation and the refine options visible as standard (parametrical searching, which is also being implemented in Omega). It should be noted in this context that the majority also sees this as an advantage compared with WoS, where people “lose their way” more often or “have to perform roundabout actions”, especially for cited ref searches. Notably, and despite the lauded interface, no one had found the option for the citation overview by themselves. After being shown this, everyone is highly enthusiastic, especially about the clear table-format presentation of citations of articles by a specific author. That mode of presentation has been improved even further in the Spring release 2006 of Scopus by sorting and clicking options. A minority say they will use the option for citation overviews by subject (following a search |select all|save to list|my list|select all|citation overview), even if it could be easier and faster.

Heavy use is made of WoS, especially for the citation searches, for which ISI had a virtual monopoly until recently. There is wide variation in the extent to which WoS is used as the main content database or alternatively as one of the access tools alongside other broad databases (Omega, Picarta, Google Scholar), subject indices (for instance Pubmed, PsycInfo, Georef, Econlit, CAB Abstracts) and publishers’ platforms (e.g. Science Direct). Notably, a number of researchers state that they do not make any use at all of the subject area index in their fields. They basically feel that using WoS/Scopus, Omega and Google Scholar provides sufficient recall.

If asked specifically about how WoS and Scopus compare, they are inclined to opt for Scopus for ease of use but preferences are less clear in terms of coverage. A considerable number of interviewees are affected by uncertainty, unfamiliarity and concern about this. Part of the uncertainty was eliminated during the interview, as coverage in Scopus proved to better or more or less equal in fair comparisons. List of journals were shown to the researchers to give them an idea of differences in coverage. This rarely gave rise to the judgement

that there were major omissions in Scopus compared with WoS. There is continued concern about coverage in social sciences, especially for the period prior to 1996.

The web results and patents are considered to be of interest by a minority, but also of subordinate importance since they are also available outside Scopus. None of the interviewees stated that they valued a comprehensive listing of search results on the web (Scirus) and Scopus itself (though they were not asked about this specifically).

6.2 Web survey among users

6.2.1 Conclusions

An analysis of a user survey among 81 persons provides a basis for a number of instructive conclusions, the most important of which are:

- Google Scholar is more widely known than Scopus or Web of Science.
- Scopus is used mainly by students and researchers in the fields of science and socio-economic sciences, Google Scholar mainly in the arts/humanities; students tend to use Google Scholar more in relative terms, research trainees / researchers, by contrast, make more use of Web of Science.
- Respondents judge that Scopus provides the best hits, followed by Web of Science; only 15% is really satisfied with Google Scholar. Respondents with a science background are all (!) satisfied with the quality of the hits in Scopus.
- For Google Scholar, opinions on the quality of hits are sharply divided: 65% of research trainees/researchers is dissatisfied, while only 22% are dissatisfied among students; 22% of the group in the arts/humanities are dissatisfied, and for those working in science that figure is 61%.

- 84% of the respondents judges Scopus to be easy or fairly easy to use; Google Scholar scores 80%, Web of Science 61%. Only 50% of the students judge Web of Science to be easy or relatively convenient to use, among trainees/researchers that is 76%.

Scopus therefore significantly outscores Web of Science, both in terms of its user friendliness and the quality of hits.

6.2.2 Research objectives

The UBU is considering whether to purchase Elsevier's Scopus database. This database is being trialled. This study has been carried out in addition to other reviews of the coverage and functionality provided by Scopus, in order to develop a better understanding of users' preferences, especially as compared to Web of Science.

The objective of this study accordingly is to provide a statistically substantiated view of the following aspects:

- How often is Scopus used, compared to Web of Science and Google Scholar?
- How do users rate the quality of the search results of Scopus and Web of Science?
- How do users rate the ease of use of Scopus and Web of Science?
- How did they hear of Scopus?

6.2.3 Method

The survey was performed in April and May 2006 using an online tool from www.surveymonkey.com. This offers an easy way of preparing a digital survey whose results can be called up immediately. A news message was placed on various library websites, and the study was also brought to peoples' attention in a number of digital newsletters.

There were 94 respondents in total. Thirteen responses were removed because they did not go beyond question 2; in the end 81 completed questionnaires were therefore analysed.

To maximise the number of respondents, the questionnaire was designed to be capable of being completed in around ten minutes. In fact, almost no one took longer than ten minutes.

6.2.4 Results of digital survey

1. Please state what applies to you:

Bachelor Student	19	23%
Master student	17	21%
Research trainee/postgraduate	17	21%
Researcher	19	23%
Other academic staff	6	7%
Not affiliated with the university	3	4%
Total	**81**	**100%**

All relevant groups are adequately represented in this survey. In the analysis below, this group will be divided into the categories students, research trainees/researchers and others.

2. What is your specialist subject?

Biology	3	4%
Pharmaceutics	8	10%
Medicine	2	2%
Geosciences	7	9%
Theology	1	1%
Language, literature & arts	10	12%
Law / Economics	4	5%
Chemistry	12	15%
Social sciences	27	33%
Mathematics and Computer science	1	1%
Other	6	7%
Total	**81**	**100%**

The social sciences and chemistry are over-represented and accordingly a more active approach was made to these groups to participate in the study. In the analysis below, this group will be summarised as arts/humanities, science (incl. medical), and socio-economic. The following were referred to under “other”:

biomedical sciences (neurosciences), environmental sciences, pedagogy, toxicology and UBU (2x).

3. How often do you use the search engines below for scientific literature?

	Scopus		**Web of Science**		**Google Scholar**	
Several times a week	20	25%	12	15%	15	19%
A few times a month	19	23%	20	25%	22	27%
A few times a year	8	10%	11	14%	10	12%
Rarely or never	11	14%	18	22%	22	27%
I do not know this search engine	23	28%	20	25%	12	15%
Total	**81**	**100%**	**81**	**100%**	**81**	**100%**

Because the study focused on Scopus, it is likely that Scopus users are over-represented in this survey. This means it cannot be concluded that Scopus is more widely used; nor is this supported by the statistics. Notably, comparatively few people do not know Google Scholar (the best score of the three search engines) and it is used more often than Web of Science.

In the analysis below this group is divided (by search engine) into users (who use the search engine at least a few times a year) and non-users (the rest).

Scopus is used more, in relative terms, by people working in the fields of science and socio-economic sciences than in the arts/humanities. Web of Science users comprise comparatively many research trainees / researchers and significantly fewer students. Google Scholar, by contrast, is used more by students and also more, comparatively, by people working in arts/humanities.

4. We want to ask you to use a search question that is relevant to you in Scopus (link via MyUU), Web of Science (link via MyUU) and Google Scholar, i.e. the same question in all three search engines. After doing so, please indicate how you rate the quality of the hits you obtained with the search engine concerned?

	Scopus	**Web of Science**	**Google Scholar**

High quality /relevance	25	49%	13	27%	8	15%
Fair quality/relevance	19	37%	17	35%	21	40%
Mediocre quality/ relevance	5	10%	10	20%	20	38%
Poor quality/ relevance	2	4%	9	18%	3	6%
Do not know or use search engine	9		11		8	
Total	**60**	**100%**	**60**	**100%**	**60**	**100%**

The percentages have been calculated without counting the group that does not know or use the search engine. Interestingly, 21 respondents failed to complete this question, which they evidently felt involved too much effort.
Clearly, Scopus scores substantially better, followed by Web of Science and then Google Scholar.

In the analysis below this group is divided into satisfied users (choosing high or fair relevance) and dissatisfied users (choosing mediocre or poor relevance). The subdivision into students and research trainees /researchers is only significant for Google Scholar: 65% of the latter group is dissatisfied, while only 22% of the students is dissatisfied. There is no substantial difference between these groups for the other search engines.

The division by arts/humanities, science and socio-economic sciences does however produce differences. The entire group in the sciences is satisfied with Scopus (!), in contrast to the arts/humanities group, of which only 63% is satisfied. For Web of Science, the socio-economic sciences group is more satisfied (65%) than the arts/humanities group (50%). The differences are most marked for Google Scholar: 78% in the arts/humanities is satisfied, only 39% in the sciences (the socio-economic sciences do not differ from the average).
Also, users prove to be more satisfied than non-users, but that is of course not surprising.

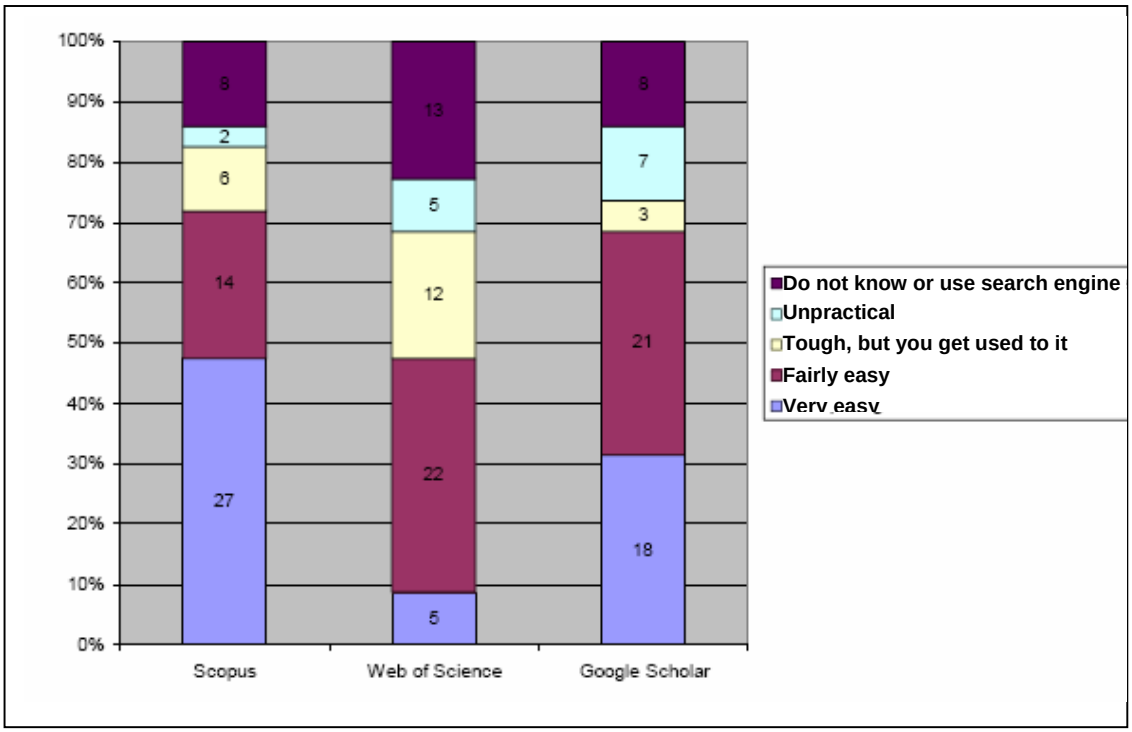
5. On which search terms did you search?

The search terms entered were (quotation marks added subsequently to cluster terms per respondent):

"aggression", "aggressive behavior and information processing", "'American Sign Language' AND agrammatism", "Antichymotrypsin AND Alzheimer Disease", "argument kuhn, d", "authors, working memory, language, development", "author name", "Bayes, diagnosis Psychological", "bijlmermeer", "birth order achievement differences", "bolivia + socialism", "Boschma, R.A.; Competitiveness of Regions from an evolutionary perspective", "cawthon telomere PCR", "chronic low backpain", "computer assisted assessment", "COPD AND collagen", "crowding eye movements", "Down syndrome", "european AND parliament AND portugal", "evidence based practice", "feline hypothyroidism treatment", "fractures bèta blockers", "function cAMP", "gender-mainstreaming", "greenblatt AND shakespeare", "history university utrecht", "iapp and membrane", "McEnrue, M / Human Resource Development Quarterly / Formal Mentoring Programs", "mergers AND R&D AND relatedness", "microbiological terms", "netherlands multiculturalism", "orphan drugs and rare disease", "paclitaxel newborn", "phthalocyanine AND amphiphilic", "protein degradation", "rheumatoid arthritis AND moderat* (in title)", "schizophrenia emotion", "size dependent adsorption", "social learning cumulative cultural evolution", "squaraine", "criminal law", "'task analysis' 'instructional design'", "too many to mention individually ", "'University history" and "F C Donders".

6. Please state how you rate the ease of use of these search engines?

	Scopus		**Web of Science**		**Google Scholar**	
Very easy	27	55%	5	11%	18	37%
Fairly easy	14	29%	22	50%	21	43%
Tough, but you get used to it	6	12%	12	27%	3	6%
Unpractical	2	4%	5	11%	7	14%
Do not know or use search engine	8		13		8	16%
Total	**57**	**100%**	**57**	**100%**	**57**	**116%**



The percentages have been calculated without including the group that does not know or use the search engine. Scopus scores better than Google Scholar and substantially better than Web of Science, for which 38% says that using it is tough or unpractical.

Web of Science is rated as tough or unpractical by 50% of the students, while that percentage is much lower for research trainees / researchers: 24%. For Google Scholar 29% of the science group rates the search engine as tough or unpractical, and in the socio-economic sciences that is only 10%.

7. How were you alerted to start using Scopus?

I do not use Scopus	12	21%
Via the homepage of my subject library	5	9%
Via www.library.uu.nl	11	19%
Via my lecturer	6	11%
Via a fellow student or colleague	12	21%
other, as follows	11	19%
Total	**57**	**100%**

The replies are highly varied; both websites, lecturers and colleagues appear to be important in alerting people to Scopus.
The following was filled in under ‘other’:

- via education rep in my study association
- introductory subject WAR (social geography)
- library newsletter
- via e-mail from library
- via library staff member
- via subject specialist (4x)
- workshop literature search by the Legal Library
- was unable to find Scopus, hidden among bibliographies and reference works

8. Do you have any other comments on Scopus?
The respondents were also asked for free-form comments on Scopus, Web of Science and Google Scholar. Annex IV contains all comments.

7 Summary of subject-specific results

7.1 Overview

Table 7.1 Score for Scopus on various aspects of coverage, per UBU subject area

	(a) % coverage of UBU journals	(b) coverage 1986-1995 as % of 1996-2005	(c) coverage of UBU journals, **compared to WoS**	(d) coverage on basis of searches 1988 and 2005, **compared to WoS**	(e) citations of selected articles 1995 and 2005, **compared to WoS**
EARTH	++	++	+	+	0
BIO	++	+	+	+	+
VET	++	0	+	+	0
ECON	+	-	++	+	++
PHARM	++	++	+	+	++
MED	++	++	++	0	0
THEOL	--	na	--	na	na
AGRI	++	0	+	na	++
LANG ETC	--	-	--	na	0
ENV	+	++	++	+	+
PHYS&ASTR	+	-	+	-	0
LAW	-	na	+		--
CHEM	+	-	0	+	0
SOC.SC&SPAT.PL	+	+	++	+	++
SOC.SC	0	-	+	+	+
TECHN	+	+	++	na	++
PHILO	--	na	--	na	na
MATH&COMP.SC	0	--	+	0	++

(a) --=<20%; -=20-40%; 0=40-60%; +=60-80%; ++=80-100% (table 4.4)
(b) --=0-16%; -=16-32%; 0=32-48%; +=48-64%; ++=64-80% (table 4.8) BIO and SW average of various Scopus subject areas, SG&PL estimate
(c) --=>20% points less; -=5-20% points less; 0=5% points less to 5% points more; +=5-20% points more; ++=>20% points more (table 4.4)
(d) -=Scopus fewer searches with >10% more results than WoS; 0=Scopus has equal number of searches with >10% more results than WoS; Scopus has >10% more results for more searches than WoS (table 4.6); a three point scale was used for this owing to the significance of coincidence in the counts
(e) --=>1 0% fewer; -=5-10% fewer; 0=5% fewer to 5% more; +=5-10% more; ++=>10% more (table 4.10) na= not available

7.2 Comments per UBU subject

A great deal of information on the value of Scopus for specific subject areas has been generated as part of this study. That information is summarised below, by reference to the coverage indicators shown in table 7.1, the availability of specific keywords, subject-specific functionality, the usability of the subject classification of Scopus, comments from heavy users of citation databases and significant differences, if any, within the UBU subject areas.

The coverage of the classics in Scopus is minimal. Elsevier deliberately did not focus on this field in developing the database because the literature needs in these subjects were different (fewer journals, more books; greater need for older material). Nor did Elsevier have any indices of its own in these fields that could have served as a basis. In itself, then, this is a very important drawback compared to the coverage in Web of Science, which has a comprehensive Arts & Humanities section. The weight that is to be attached to this depends on the use made of WoS for these subject areas and on the need to be able to search these fields in a citation database in conjunction with the fields of science and socio-economic sciences.

7.2.1 Earth sciences

The coverage of earth sciences is good to excellent, both in itself and compared to WoS. Scopus covers 81% of the digital UBU titles. The good coverage is

based in the main on the underlying database Geobase, which is included virtually in full in Scopus. Compared to Scopus however, the Georef database includes more non-journal material. On the one hand it is an advantage to be able to simultaneously search journals in Scopus in the field of biology (geobiology) and technology (hydrology). Scopus evidently lacks the genuine thesaurus terms and functionality of Georef, but it has taken over the descriptors from Geobase, including the geographical terms. The Scopus subject classification is readily usable in earth sciences.

7.2.2 Biology

Scopus provides good coverage of biology, both in itself and in comparison with WoS. More than 83% of the digital UBU titles are in Scopus. The underlying database Biobase contributes to that. The UBU does not provide any specific subject indices for biology. Scopus has taken over the keywords from Biobase as well. There is however no species thesaurus with explode function. Biology also benefits from the millions of records from adjacent disciplines such as medicine and chemistry. Scopus contains at least three relevant subject area tags, one of which is shared with agricultural science.

7.2.3 Veterinary medicine

Scopus provides good coverage of veterinary medicine, with 82% of the digital UBU titles. The backfill can still be improved however. Scopus outscores WoS on overlaps with CABAbstracts. Scopus has no specific underlying veterinary medicine database, but veterinary medicine benefits from the good coverage of medicine and MeSH terms from Pubmed that are included in Scopus via EMbase. Scopus does have a separate subject tag ‘veterinary’. Agricultural sciences are likewise covered well in Scopus; a segment of the Scopus classification is shared with biology for this subject area.

7.2.4 Economics

Economics is not covered in Scopus from one of the underlying databases, with the exception of regional economy in Geobase. Nonetheless economics, with 61% coverage of digital UBU titles, is better served than in WoS, and that also

applies to the searches and citation counts made for this report. There is a distinct need however for the inclusion of more older years. The backfill, just shy of 10%, is poor. Scopus provides almost 20% more citing articles than WoS in this field, according to our research. For applied economics and business studies Scopus cannot really compete with the free access to research papers of RepECEconpapers and SMEALsearch.

7.2.5 Pharmaceutics

Scopus covers a large portion of the pharmaceutics journals carried digitally by the UBU (81%). All other indicators for pharmaceutics coverage are likewise very positive, in themselves and by comparison with WoS. This confirms the tests carried out by Schneider (2006). By way of EMbase, pharmaceutics naturally benefits from the presence of Pubmed/Medline records in Scopus and the associated MeSH terms. The subject classification in Scopus easily accommodates the needs of pharmaceuticals.

7.2.6 Medicine

The coverage in Scopus for medicine, at 83% of our digital journals in this field, is excellent, certainly compared to WoS. Only Pubmed offers even wider coverage. Scopus has MeSH terms and includes (almost) all records from EMbase. Specific functionality such as searching for genetic sequences is also available. The question does remain however how to account for the comparatively significant difference between Scopus and Pubmed for recent years. The relatively modest score for backfill (table 4.8) can also be explained by the extraordinarily strong growth in the number of medical publications in the past 10 years. Scopus provides a fair number of relevant subject tags, including a separate tag for nursing.

7.2.7 Theology

As a subject in the field of the classics, theology receives only very minimal coverage in Scopus. No specialised theological journals are covered, although some journals in neighbouring disciplines are included. For example there is some coverage in the field of bioethics (bioethic*=6600, probably mainly from

Medline) but hardly anything else (theolog*=3600). The value of the database for theology is very limited.

7.2.8 Language, Literature and Arts

The broad field of language, literature and arts is given very limited coverage in Scopus, which is a deliberate choice made by its producer. There is reasonable coverage in the field of acoustics, language technology, computer linguistics etc. (probably largely from Compendex), but Scopus is not otherwise relevant to these subjects

7.2.9 Environmental science

Coverage of environmental science is fairly good, with 70% of the UBU titles. This coverage derives from the underlying databases Geobase, Biobase and Compendex. The backfill is therefore good as well. Scopus has a separate section Environmental Science in its subject classification, but depending on the topic, relevant records are also in Earth & Planetary Sciences, Social Sciences, Biological and Agricultural Science as well as Energy and Materials Science. Accordingly it is important not to apply subject limiters by default. For innovation sciences (in the same department as environmental science) an integrated database such as Scopus is ideal. Scopus outscores WoS here owing to its better inclusion of non-US journals and good coverage of technology (on the basis of Compendex) and economics. Scopus provides good coverage of all domains of the innovation sciences (genetic engineering, energy and materials and RO & transport). No dedicated set of controlled vocabulary is available for these disciplines, but they do benefit from the fairly large number of keywords from Geobase, Compendex and Biobase.

7.2.10 Physics and Astronomy

At 75% of the digital journals held by Utrecht, coverage of physics and astronomy is fairly good. Backfill needs to be improved however, especially in view of the large quantity of Open Access material in this field and the availability free of charge elsewhere of databases such as ADS (Astrophysical Data System) and ArXiv. There is backfill from the underlying database Compendex, but that largely

relates to technology (engineering) and only to a limited extent to fundamental research journals (which are however included in Inspec). Numbers of citations for articles selected by us match those in WoS, but a number of the searches carried out for this study were less than convincing. This needs to be looked into in greater detail.
There is a subject area Physics and Astronomy in Scopus, but Materials, Energy and Earth & Planetary Science will often also be relevant.

7.2.11 Law

While both databases offer relatively poor coverage in absolute terms (with the 35% turned in by Scopus still slightly above WoS) for this subject, the comparison between them is instructive. That is because this is the only subject with a major discrepancy between coverage on the basis of titles of journals and citation data (for titles of journals: Scopus/WoS= 1.6 and for citations: Scopus/WoS= 0.6 (and WoS/Scopus= 1.7).
Law is not explicitly covered by Scopus but neither is it expressly excluded, like the classics, from the current aspirations for Scopus. Probably the specific nature of the material (particularly insofar as it is linked to legal practice within specific national contexts) plays a part in this. According to the list of sources, Scopus does cover 187 journals featuring the term law in the title, but the majority of them is specifically directed at the US. In terms of Scopus subjects, journals in the field of law are usually classed under social science.

7.2.12 Chemistry

Coverage of chemistry is relatively good in Scopus: 75% of our digital chemistry holdings. Other coverage indicators are likewise fairly good to good, but backfill needs to be improved. As for physics, chemistry titles providing backfill for the period prior to 1996 via compendex relate mainly to applied chemistry and not to fundamental research titles.
Needless to say, Scopus cannot match the coverage and functionality of Chemical Abstracts or SciFinder Scholar, but Scopus does support searching by

CASREG numbers and linking to records in Crossfire Beilstein to look up reactions and molecule structures.

There are two subject areas in Scopus for chemistry: Chemistry (fundamental research) and Chemical engineering (chemical technology). The segment Materials science will however also often be relevant.

7.2.13 Social geography and Spatial planning

Scopus provides good coverage for social geography and spatial planning, certainly by comparison to WoS. This coverage derives from Geobase. Backfill is therefore likewise good (to 1980), and the availability of keywords is fairly good. Geographical terms from Geobase have also been included in Scopus.

The subject classification represents a problem. Initially, all geoscience journals came under Earth & environmental science, but since the Spring 2006 release physical geography is classed under Earth & planetary science and social geography and spatial planning under Social science. In itself that is an improvement, but many of the journals relevant for SG&PL are nonetheless (erroneously) classed only under Earth & planetary science. Also, the category Social science is too wide to be of use in most searches. That means it is not even advisable to tick any of the four subject clusters in advance. Finally: our study showed that Scopus includes significantly more citations of publications in this subject than WoS.

7.2.14 Social sciences

Coverage for the social sciences is difficult to assess as a whole. Interpretation of the scores in table 7.1 is not straightforward. Overall coverage of journals, at 54%, is fair, better than the 41% offered by WoS. Searches by the key terms of the subjects however reveals poorer coverage than in WoS for the social sciences (table 7.2). This applies both to recent and to older years. For psychology, coverage of older years is better in Scopus , coverage of recent

years is better in WoS. Psychology in Scopus probably benefits from the inclusion of psychological journals in Scopus via EMbase.

Table 7.2 Search results for general terms from the social sciences as title words in Scopus and WoS, total and 1996-2005

	All years			1996-2005		
	Scopus	WoS	Scopus as % of WoS	Scopus	WoS	Scopus as % of WoS
Anthrop*	3518	6051	58	1594	5050	32
Sociolog*	4867	11570	42	2566	5895	44
Psycholog*	44752	45402	99	19706	25336	78

Matters are different for more specific searches within the subject areas (table 4.5). For these, Scopus also scores well on anthropology, but coverage of sociology again is patchy.

Scopus lacks the extensive options and index terms of PsycInfo (quite apart from its coverage), but does provide slightly more functionality for citations. Scopus applies two subject sections for social sciences: Social Science (including social geography and law) for social sciences and Psychology for behavioural sciences. We did not establish under which subject headings Scopus classes disciplines such as pedagogy and educational science.

The situation for the social sciences needs to be looked into in greater detail.

7.2.15 Philosophy

Of the classics, philosophy is the subject with the best coverage, but this does not amount to much in absolute terms. There is some coverage in the fields of philosophy of science, artificial intelligence and philosophical anthropology. A number of journals is also included for these fields.

7.2.16 Mathematics and Computer science

Coverage for Mathematics and Computer science in Scopus is no more than reasonable, though it is better than that offered by WoS in this field. It should be

noted that the pure coverage for this subject is likely to be larger when adjusted for the relatively large number of journals in the field of librarianship included in this UBU subject.

The Backfill for this subject in Scopus is very limited: 10%. Information on older material will still have to be obtained from Zentralblatt MATH for the time being, although this does not provide citation data. Scopus heavily outscores WoS for citations of mathematical articles. These subjects correspond to 3 Scopus subject tags: Mathematics, Computer sciences and Decision sciences.

Literature

- Bakkalbassi, N., K. Bauer, J. Glover & L. Wang (2006) Three options for citation tracking: Google Scholar, Scopus and Web of Science. Biomedical Digital Libraries 3,7. http://www.bio-diglib.com/content/3/1/7
- Deis, L.F. & D. Goodman (2006) Update on Scopus. The Charleston Advisor 7,3. http://www.charlestonco.com/comp.cfm?id=55
- Goodman, D. & L.F. Deis (2005) Web of Science (2004 version) and Scopus. The Charleston Advisor 6,3. http://www.charlestonco.com/comp.cfm?id=43
- Jascó, P. (2004) Scopus [online]. Péter's digital reference shelf, September 2004. http://www.galegroup.com/servlet/HTMLFileServlet?imprint=9999&region=7&fileN ame=reference/archive/200409/scopus.html
- Jascó, P (2005) As we may search - comparison of major features of the Web of Science, Scopus and Google Scholar citation-based and citation-enhanced databases. Current Science 89, pp. 1537-1547. http://www.ias.ac.in/currsci/nov102005/1 537.pdf
- Jascó, P. (2006) Scopus revisited [online]. Péter's digital reference shelf, June 2006. http://reviews.gale.com/index.php/digital-reference-shelf/2006/06/scopus-revisited/
- Neuhaus, Chr., E. Neuhaus, A Asher & C. Wrede (2006) The depth and breadth of Google Scholar: an empirical study. Portal: libraries and the Academy 6, pp. 127-141. http://muse.jhu.edu/journals/portal_libraries_and_the_academy/v006/6.2neuhaus.pdf
- Pipp, E (2006) Vergleich der von Scopus bzw. Web of Sciences erfassten Zeitschriften. Online Mitteilungen 85, pp. 3-17. http://www.univie.ac.at/voeb/php/downloads/om85 .pdf

- Schneider, K. (2006) Scopus - Web of Science: Versuch einer Bewertung aus pharmakognostischer Sicht. Online Mitteilungen 85, pp. 21-24. http://www.univie.ac.at/voeb/php/downloads/om85.pdf
- Qin, J. (2000) Semantic similarities between a keyword database and a controlled vocabulary database: An investigation in the antibiotic resistance literature. Journal of the American Society for Information Science 51, pp. 166-180. http://www3.interscience.wiley.com/cgi-bin/fulltext/6950 113 8/PDFSTART

Annex 6 lists other literature on Scopus that has not been quoted.

Annex VI: Literature on Scopus not quoted

- Burnham, J.F. (2006) Scopus database: a review. Biomedical Digital Libraries 3,1. http://www.bio-diglib.com/content/3/1/1
- Dess, H.M. Database reviews and reports - Scopus. Issues in Science and Technology Librarianship, winter 2006. http://www.istl.org/06-winter/databases4.html
- Fingerman, S. (2005) Scopus: profusion and confusion. Online 29,2, pp.36-38.
  http://www.infotoday.com/Online/mar05/index.shtml
- Goraiz, J. (2006) Web of Science versus Scopus oder das aktuelle Dilemma der
  Bibliotheken. Online Mitteilungen 85, pp. 25-30.
  http://www.univie.ac.at/voeb/php/downloads/om85.pdf
- Grupo SCImago (2006) Análisis de la cobertura de la base de datos Scopus. El professional de la información 15, 2, pp.144-145.
  http://www.ugr.es/~benjamin/EPI-Scopus.pdf
- Kaemper, B-Chr. (2006) A Reader`s Reflection about Scopus: Letter from Bernd-
  Christoph Kaemper. The Charleston Advisor 7,4.
  http://www.charlestonco.com/features.cfm?id=200&type=me
- LaGuardia, C. (2005) ISI Web of Science / Scopus. Library Journal 130,1, pp.40-42.
  http://www.libraryjournal.com/article/CA491154.html
- Roth, D.L. (2005) The emergence of competitors to the Science Citation Index and the Web of Science. Current Science 89,9, pp. 1531-1536.
  http://www.ias.ac.in/currsci/nov102005/1531.pdf
- Wildner, B. (2006) Web of Science - Scopus: auf der Suche nach Zitierungen. Onlin Mitteilungen 85, pp. 18-20.
  http://www.univie.ac.at/voeb/php/downloads/om85.pdf