Science indicators revisited – Science Citation Index versus SCOPUS: A bibliometric comparison of both citation databases

Rafael Ball * and Dirk Tunger
Forschungszentrum Juelich, Central Library, 52425 Juelich, Germany

Abstract. Although the evaluation of knowledge in the form of a quantitation of scientific output is not uncontroversial, it is a widely practised form of science evaluation. For more than 30 years, the Science Citation Index (SCI) has been alone in fulfilling this purpose. But since 2005 Scopus is a direct competitor to the SCI-databases. Comparing the two databases should help to answer questions that could have repercussions for the future generation of bibliometric analyses. The results of the comparison will allow us to more reliably rate Scopus, as a new data source, against the established SCI. In future, people who generate bibliometric analyses must be able to justify why they chose to use one database and not the other. It will not be enough to simply claim that SCI is the established source.
Keywords: Bibliometrics, Science Citation Index, Scopus, citationdatabase, science evaluation

1. Introduction

For a long time, the Science Citation Index (SCI)[1] was the only multidisciplinary database that could be used to quantitatively determine the response to scientific publications. This method of measuring response (as the number of citations per article) became increasingly important as a decisive factor in the evaluation of scientific output, which no longer consisted solely of unspecific, personal assessments by experts, but rather – immersed in the pool of performance indicators from the field of economics – included quantitative parameters in the evaluation. This development was recognised early on by the founder of the Science Citation Index at the Institute of Scientific Information, which led to the creation of a database that holds an unattested monopoly in the market decades later as the only benchmark for the quantitative evaluation of scientific output [3]. This era came to a close in 2005. The large scientific concern "Reed Elsevier" placed a second multidisciplinary database (SCOPUS) on the market in 2004.[2] On the one hand, this product was to serve as a database of research literature for scientists, and therefore corresponds to a "normal" bibliographic database like those provided by libraries in every possible discipline. On the other hand, the fact that it records the citations of scientific articles put it in direct

---

*Corresponding author: Rafael Ball, Forschungszentrum Juelich, Central Library, 52425 Juelich, Germany. Tel.: +49 2461 613013; Fax: +49 2461 616103; E-mail: r.ball@fz-juelich.de.
[1]http://portal.isiknowledge.com.
[2]www.SCOPUS.com.

competition with Thomson Scientific's existing Science Citation Index. The introduction of a new product would not have warranted attention from an economic point of view had the second multidisciplinary research database of scientific reception not thrown the scientific community into a state of emergency when it came to deciding which database should now be used as the international benchmark for scientific rankings. If SCOPUS has not yet managed to establish itself among the scientific community, it is certainly not something that the general public is familiar with. Computer science and centres of excellence for bibliometrics, in particular, have now begun to analyse and compare both systems with regard to usability of results for quantitative scientific evaluations.

A comprehensive comparison was recently published in the Austrian "Online-Mitteilungen" brought out by the University of Vienna [4,7], while other isolated papers have dealt with this topic [1,2,6]. We detected a research desideratum here – the question of a global, internationally recognised benchmark for the evaluation of scientific output is far too important to simply leave it up to one product or the other without further investigation or debate.

Then the one thing that is to be expected is that the results of citation analyses and the rankings derived from them will be different for both databases for the same random sample. The consequences, particularly in the area of science management with regard to the allocation of funds for science and research based on performance, are far too grave to be left to random selection by the reference database.

The following comparison of the databases is based on the selection of some interesting points in the content of the databases and merely marks the beginning of a comparison of the databases that is to continue. We do not take aspects of ease of use and usability into account here. Findings from the view point of information specialists are already available on this [5].

SCOPUS claims that it is a database built on the thorough analysis of over 15,000 scientific journals. These journals come from the STM sector (science, technology, medicine) and the social sciences.

With over 15,000 journals that are regularly evaluated, the number of periodicals covered is almost twice as high as that of SCI. SCI has around 7,500 journals covering the natural and social sciences.[3]

Our Austrian colleagues have also examined both databases with a fine tooth comb. Juan Gorraiz describes the advantages and disadvantages of both databases from a bibliometric point of view. He comes to the conclusion that we cannot say with certainty which of the two databases is better since it also depends to a certain extent on the subject area in question.

2. Key issues

The strategies employed by each database are very different. In SCI, only a selection of approx. 7500 scientific journals are evaluated. The selection is over proportionately based on the average citation rate of articles in the journals. Limiting themselves to a set volume of journals is considered a quality criterion for SCI. The selection of titles is continuously checked and adjusted. If a journal falls below the assessment threshold, then it is dropped from the index, while journal titles that rise above the threshold are incorporated into the index.

SCOPUS on the other hand favours the "as much as you can" strategy. This means that a much greater number of journal titles are taken into account.[4] This alone illustrates that the citation analyses in SCI and SCOPUS, each based on different data sets, produce very different results that are NOT comparable.

---

[3]For more details, see Journal Citation Report (JCR) in the Web of Knowledge http://portal.isiknowledge.com.

[4]http://info.SCOPUS.com/detail/what/.

The current paper is primarily concerned with whether the higher number of journals (journal volume) is accompanied by an increase in qualitative content. We want to answer this question with the aid of selected bibliometric analyses that were carried out simultaneously in SCI and SCOPUS.

In more concrete terms, we did this by conducting four identical investigations in both databases for identical periods of time. The four issues investigated were:

1. *Number of articles covered in each database*
   SCOPUS claims that it evaluates a larger volume of journals. Does this increase in the underlying basis actually lead to a significantly higher number of articles in general and, more importantly, to a higher number of relevant articles?

2. *Number of non-cited articles*
   A certain number of articles are never cited, even years after their initial publication. This parameter does not allow direct conclusions to be drawn on the actual reception of an article but it does give us indirect information as to whether a larger underlying basis of articles is useful for the scientist or whether it simply "clogs up" the search results.

3. *Areas of interest in the databases*
   Each database develops a certain focus in terms of content, language and region. The user can orientate him/herself based on this and justify his/her choice of database. Describing the areas of interest means that the user is given an idea as to what he/she can actually find and in what volume.

4. *Specialised search on the topic of "fuel cells"*
   An important aspect in bibliometrics is citation analysis. Indicators are developed for many assessments in science, the main focus of which is geared towards the number of citations of a set of articles. Information generated in both databases when the topic of "fuel cells" is being processed is compared.

3. Methods

For the issues outlined above, identical searches were conducted in June 2006 in each of the databases at the same time, taking into account the usability typical of the databases. In both instances the databases were accessed via the web. This is the typical "entry point" that is available to most users.

For issues 1 (number of articles covered), 2 (number of non-cited articles) and 4 (search on fuel cells), time parameters were set and questions investigated. This had the advantage of illustrating changes over time and allowing further conclusions to be drawn as a result.

For issue 3 (areas of interest in the databases), a percentage was determined using the data set as at June 2006, which represents the proportion of a discipline in terms of the entire database. The journals are classified according to discipline.

The observation period for individual issues varied and will be outlined together with the presentation of results in Section 4.

With regard to the analyses conducted, use was made of the full database in each case. Constraints were applied for the number of articles covered discussed in relation to issue 1: since SCI only covers scientific publications in the STM sector, parity was achieved between the databases in that only those articles of a similar scientific nature in SCOPUS were used.

In order to determine the proportion of non-cited articles for issue 2, we had to limit the search to three disciplines in SCOPUS for technical reasons. The topics chosen were computer science, physics (Germany only) and mathematics.

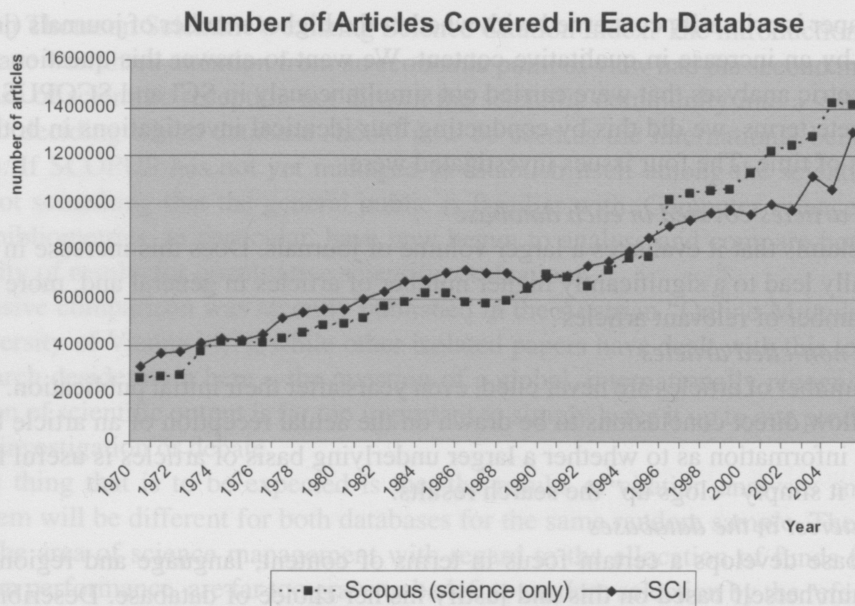**Number of Articles Covered in Each Database**

Fig. 1. Direct comparison of the quantitative development of content in both databases.

Issues 3 and 4 were investigated using the complete databases, although this was of little significance for searches on the topic of fuel cells. The majority of hits were of a scientific nature. Here, it is interesting to apply standard bibliometric indicators, such as the citation rate, to the search results and to compare them over time. Since the keywords used for the searches were identical, tailored to the query language of the database and the period investigated was also the same, the results can be directly compared.

4. Results

Despite an allegedly substantially bigger database, SCOPUS only shows a larger data pool than ISI's Science Citation Index from 1996 onwards. The number of publicly available journals evaluated in SCOPUS is misleading in that it should be almost twice as high. In the years before 1996, SCOPUS failed to equal the number of articles held by SCI. This analysis was based on all of the articles published by institutions in the STM sector.

In relation to this, it is interesting to note that references (citations) are only fully available in SCOPUS from 1996 onwards. Before this date, the database remains incomplete in this regard. SCI is different: here bibliographic references and citations currently stretch back to 1945 and plans are afoot to extend coverage back to 1900 with the launch of the "Century of Science".

Another aspect is the analysis of non-cited documents. In general, a scientific publication is perceived within a period of 1–2 years and then cited in new publications. Literature databases however also contain non-cited publications. Approximately 40% of all articles from 1996 in the field of computer science have not yet been cited ten years later. It can be said that the differences between the two databases are very small here and the same is also true of mathematics.

However, if we analyse physics articles with German involvement for example, variations between the two databases are detected for the entire analysis period. For part of this period (1999–2005), these
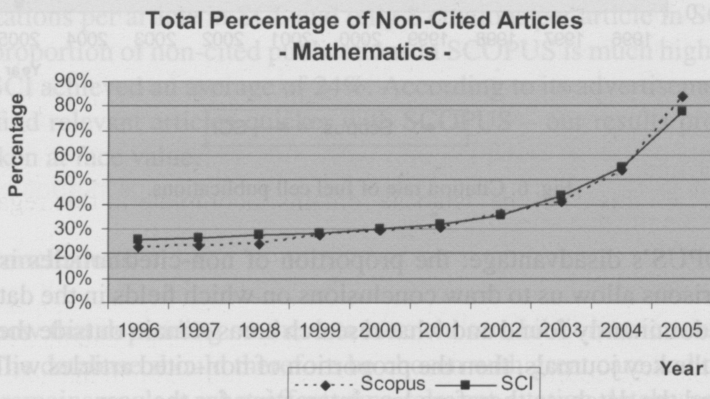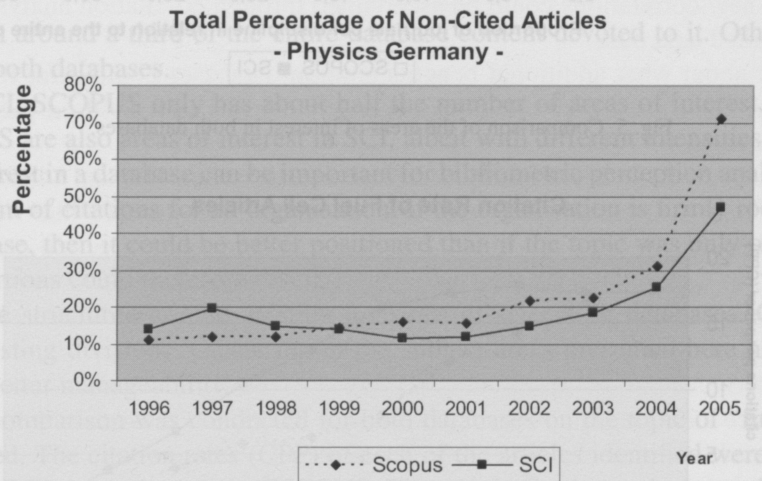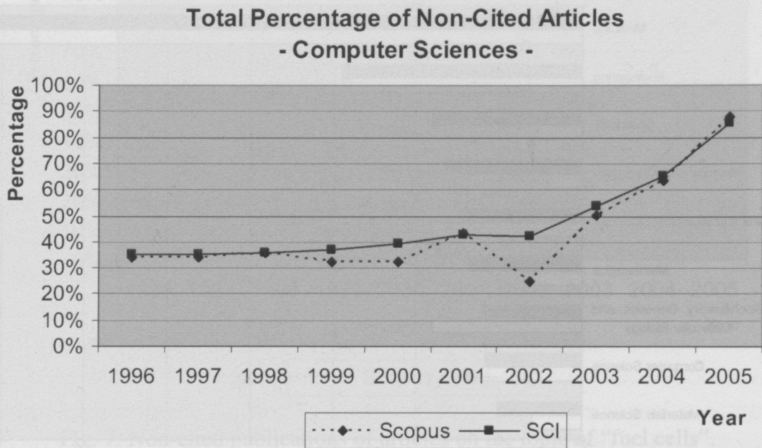
**Total Percentage of Non-Cited Articles
- Computer Sciences -**

Percentage — 100%, 90%, 80%, 70%, 60%, 50%, 40%, 30%, 20%, 10%, 0% — Years 1996–2005 — Year — Scopus · SCI

**Total Percentage of Non-Cited Articles
- Physics Germany -**

Percentage — 80%, 70%, 60%, 50%, 40%, 30%, 20%, 10%, 0% — Years 1996–2005 — Year — Scopus · SCI

**Total Percentage of Non-Cited Articles
- Mathematics -**

Percentage — 90%, 80%, 70%, 60%, 50%, 40%, 30%, 20%, 10%, 0% — Years 1996–2005 — Year — Scopus · SCI

Fig. 2–4. Proportion of non-cited publications in relation to time for different disciplines.

heavy interest ... have ... weeks devoted to it. Other topics of interest trail far behind in both databases.

Compared to SCOPUS, SCI only has about half the number of areas of interest. All of the areas of interest in SCOPUS are also listed in ... in SCI, albeit with difficult in allocation.

The areas of interest in a database can be important for bibliometric perception analyses in determining response in the form of ... areas for ... If a ... area ... is more devoted in the thematic focus of the database, then ... results are better positioned than if the topic was one that appeared on the margins. Distortion ...

Subject areas are ... Clusters are formed based on these existing ... has the advantage of transparency and be ...

A bibliometric comparison ... of "fuel cells". The same keywords were used. The citation rates ... were found to be higher for all years in the ... than in SCI ... the analysis on the same topic had a citation rate of 12 citations per ... article in SCOPUS.

It is clear that the proportion of non-cited ... in SCOPUS is much higher for all years and lies at 43% on average. SCI ... at an average of 24%. According to ... Elsevier promises users that they will find ... articles ... with SCOPUS ... results prove that this is a claim that should not be taken ... value.

variations are to SCOPUS's disadvantage: the proportion of non- ... is ... significantly higher. These types of comparisons allow ... from ... high field ... the database key journals of a discipline can be ... in ... journals in ... SCOPUS moves too far beyond the ... journals in ... will increase ... articles will be lost relevant ... the database historical ... for ... the lower ... of ... subject areas, whose proportion of journals in relation to the respective total number of journals covered is no less than 5%. Multi-classifications are possible both in SCI and SCOPUS ... medicine is an area of ...

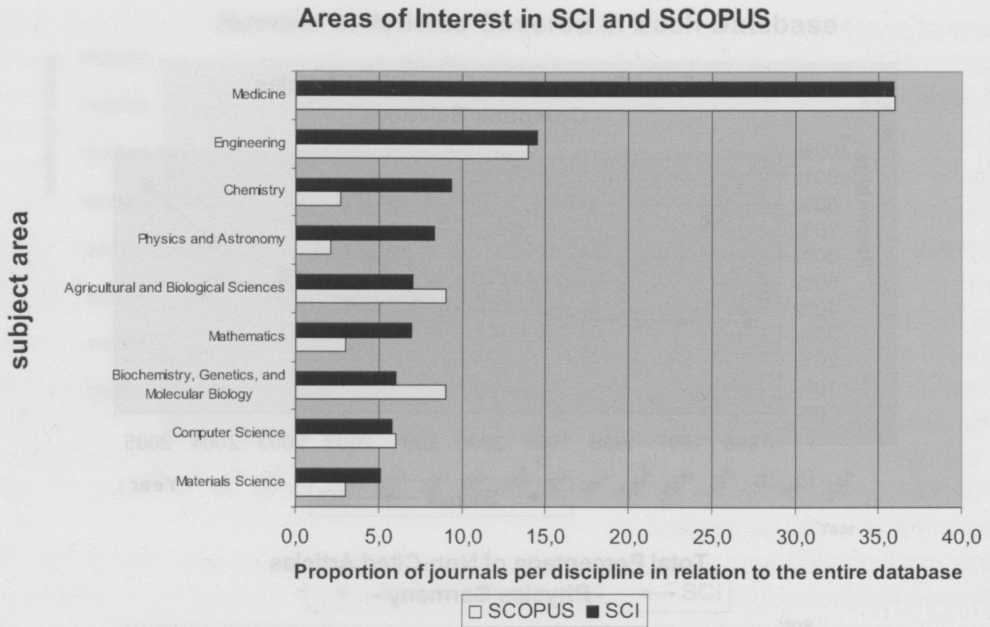## Areas of Interest in SCI and SCOPUS

Fig. 5. Comparison of the areas of interest in both databases.
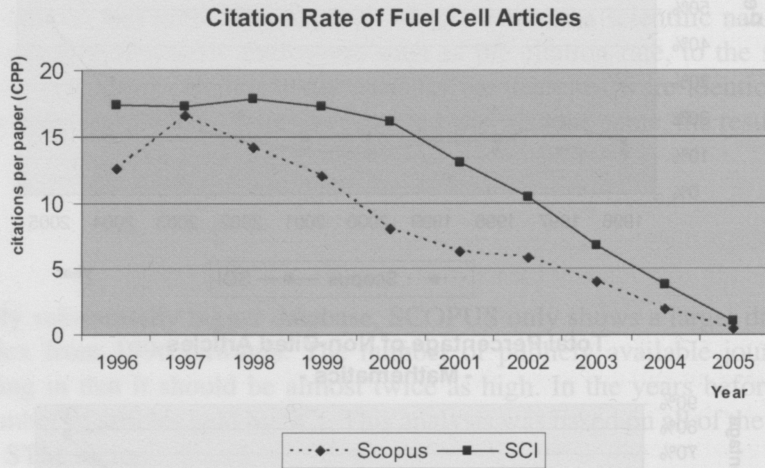

Fig. 6. Citation rate of fuel cell publications.

variations are to SCOPUS's disadvantage: the proportion of non-cited articles is significantly higher. These types of comparisons allow us to draw conclusions on which fields in the database key journals of a discipline can be predominantly found and what research areas remain outside these fields. If SCOPUS moves too far beyond the key journals, then the proportion of non-cited articles will increase, the articles will be less relevant and the database therefore less interesting for the user.

Both of the databases have developed different areas of interest. We consider areas of interest to be all subject areas, whose proportion of journals in relation to the respective total number of journals covered is no less than 5%. Multi-classifications are possible here. In SCI and SCOPUS, medicine is an area of

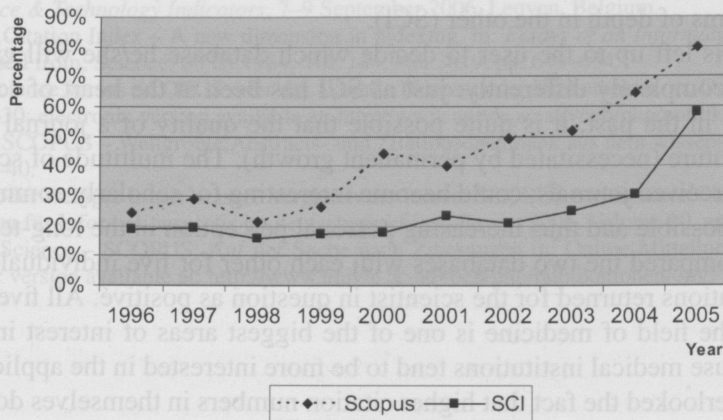**Total Percentage of Non-cited Fuel Cell Articles**

Fig. 7. Non-cited publications of articles on the topic of "fuel cells".

heavy interest with around a third of the entire database content devoted to it. Other topics of interest trail far behind in both databases.

Compared to SCI, SCOPUS only has about half the number of areas of interest. All of the areas of interest in SCOPUS are also areas of interest in SCI, albeit with different intensities.

The areas of interest in a database can be important for bibliometric perception analyses in determining response in the form of citations for an organisation. If the organisation is firmly rooted in the thematic focus of the database, then it could be better positioned than if the topic was only one that appeared on the margins. Distortions could therefore result.

Subject areas are structured around existing divisions of science in databases. Clusters are formed based on these existing divisions. Clustering in the subject areas presented here has the advantage of transparency and better manageability.

A bibliometric comparison was conducted for both databases on the topic of "fuel cells". The same keywords were used. The citation rates (CPP) of each of the articles identified were found to be higher for all years in the SCI database than in SCOPUS. The articles in the analysis on the same topic had a citation rate of 12 citations per article in SCI, and only 8 citations per article in SCOPUS.

It is clear that the proportion of non-cited publications in SCOPUS is much higher for all years and lies at 43% on average. SCI achieved an average of 24%. According to its advertisements, Elsevier promises users that they will find relevant articles quicker with SCOPUS – our results prove that this is a claim that should not be taken at face value.[5]

5. Discussion and conclusions

We were able to show that depending on the data base chosen, bibliometric analyses provide very different information. The database should therefore be chosen with great care and on the basis of content. Furthermore, it still remains unclear what effects two relevant citation databases with similar contents will have on the process of scholarly communication. The current strategies employed by the two data-

---

[5]See also: http://www.info.scopus.com/overview/what/

base providers are conflicting: the number of journals covered are expanded in terms of breadth in one (SCOPUS) and in terms of depth in the other (SCI).

At the moment, it is left up to the user to decide which database he/she will give preference to. In future, this could be completely differently: just as SCI has been at the heart of scientific communication with its journals in the past, it is quite possible that the quality of a journal will not be the sole crucial factor in the future (necessitated by permanent growth). The multitude of scientific publications, even in less highly perceived journals, could become interesting for scholarly communication in making permanent presence possible and thus increasing personal perception in the long-term.

Brigitte Wildner compared the two databases with each other for five individuals and considered the higher number of citations returned for the scientist in question as positive. All five people chosen were medical scientists. The field of medicine is one of the biggest areas of interest in both SCOPUS and Web of Science because medical institutions tend to be more interested in the application of quantitative analyses. Wildner overlooked the fact that higher citation numbers in themselves do not actually contain any information. Due to a larger number of journals, the average citation rate for the discipline could plausibly be higher in SCOPUS than in SCI. This is not striking in itself. It could be of benefit to medical science however, if new correlations could be ascertained through co-citation analyses, for example, as a result of the expanded data base.

We need to think about what additional benefits citation databases could offer their users and what additional parameters scientometrics could use in order to paint as realistic a picture as possible of scholarly communication.

The decision in future will lie between strict qualitative selection in SCI and a more extensive breadth of coverage (SCOPUS). Strict qualitative selection also means that journals that were accepted at one point in time could be dropped again if they no longer meet the strict selection criteria. This puts the publishers in question under pressure to ensure constant high perception of their journal titles. The scientist also profits in turn from this as he/she can be assured that his/her work will receive particular attention and the seal of quality.

The selection criteria for the SCOPUS database appear to be less strict; significantly more publishers have managed to place their journals in the citation index. This could be due to the fact that the selection criteria that could lead to a journal being dropped are well-known. However, growth in the breadth of coverage could bring advantages with it: research areas would be included and recorded in the database at an earlier stage. This would not only be beneficial for literature searches but also as proof of scientific work that can be found in the first instance and then cited.

Both databases show evidence of limitations with regard to their Internet interfaces, which make bibliometric analyses more difficult: for example, constraints on the number of articles that can be downloaded. To make matters worse, in SCOPUS only 2000 articles per set can be displayed on the screen. This constraint prevents us from determining things like how high the proportion of non-cited articles is for a large subject domain or for a country.

Despite everything, bibliometrics did reveal the nuts and bolts of the two databases and will conduct other comparative analyses with these data archives. The SCI benchmark is still number one worldwide. It remains to be seen whether SCOPUS will ever be in a position to take over.

References

[1] R. Ball and D. Tunger, Science indicators revisited, in: *Book of Abstracts – 9th International Conference on Science & Technology Indicators*, 7–9 September 2006, Leuven, Belgium.

[2] J. Bar-Illan, Which h-Index? Comparison of WoS, SCOPUS & Google Scholar, in: *Book of Abstracts – 9th International Conference on Science & Technology Indicators*, 7–9 September 2006, Leuven, Belgium.

[3] E. Garfield, Science Citation Index – A new dimension in indexing, in: *Essays of an Information Scientist*, Vol. 7, 1984, pp. 525–535; Reprinted from: Science, 08.05.1984, Vol. 144, No. 3619, pp. 649–654.

[4] J. Gorraiz, Web of Science versus SCOPUS oder das aktuelle Dilemma der Bibliotheken, in: Online-Mitteilungen No. 85, March 2006, pp. 25–30; electronic version available at: http://www.univie.ac.at/voeb/php/publikationen/om/.

[5] V. Trkulja, Elseviers SCOPUS – Weltgrößte Abstracts- und Zitationsdatenbank aus dem wissenschaftlichen Web, in: password 5/2005, pp. 37–40.

[6] D. Tunger, Ist mehr oder weniger Inhalt besser? Vergleich der Inhalte SCOPUS und Science Citation Index, in: Password: Nachrichten & Wissen für Informationsprofis im deutschsprachigen Raum, 2005, Volume 10, pp. 17–19.

[7] B. Wildner, Web of Science – SCOPUS: Auf der Suche nach Zitierungen in: Online-Mitteilungen No. 85, March 2006, pp. 18–20; electronic version available at: http://www.univie.ac.at/voeb/php/publikationen/om/.