

The Serbian Citation Index: Context and content

Pero Šipka

esipka@sbb.co.yu

Centre for Evaluation in Science and Education, Juriša Gagarina 111, 11070 Beograd and
University of Novi Sad, Stevana Musića 24, 21000 Novi Sad (Serbia)

Introduction

Building national citation databases (NCDs) in developing countries to serve as necessary add-ins to ISI citation indexes is an old idea, almost as old as bibliometrically-driven evaluation itself. However, results in applying this doctrinary attitude worldwide remained somehow skinny. Only recently, a serious breakthrough seems to be made by Chinese citation indexes (Xin-ning, Xin-ming & Xin-ning, 2001; Yishan W. et al., 2004). Even countries firmly integrated into the core science, such as Japan and to same extent Taiwan, launched their own NCDs (Negishi, Sun, & Shigi, 2004; Kuang-hua, 2004). In Europe, just a few sporadic, basically experimental attempts to build a NCD can be traced. One of such efforts was made in Serbia. *SocioFakt: The Serbian Social Science Citation Index* was been developing since 1990 on, got operational in 1995, and made online in 2001 (Kosanović & Šipka, 1996). Recently, *SocioFakt* was extended to the *Serbian Citation Index (SCIndeks)*, covering all of the research fields practiced in the country.

Context

Sharing unfortunate destiny of the society as a whole, Serbian science suffered a visible decline during the two previous decades. Now, at the exit of an era of extreme political instability, Serbia is lagging behind the countries in the region of once similar R&D output. This made authorities in the new democratic government to set up an ambitious strategy of fast catching-up. The strategy is aiming at raising the quality and fertility of research. The core of the problem is in a long-lasting low motivation of academics, resulting in a huge brain drain and low performance (Šipka, 2001). This can hardly be solved without introducing robust, non-arbitrary evaluation, including impact indicators. In applying this, ISI citation indexes are known to be of only partial usefulness, due to their inability to discriminate among entities belonging to a low-performance and/or isolated academic communities. A strong contributing factors to the low Serbian performance in R&D were found to be low level of international cooperation and low visibility of locally published journals (LPJs), underrepresented in international databases.

Approach

The purpose of *SCIndeks* was (i) to serve as a ground for an effective, rigorous evaluation and (ii)

to ensure access to articles published in LPJs. In shaping *SCIndeks* as an evaluation tool we ensured the compatibility with ISI citation indexes and *WoS Essential Science Indicators*. In working out the bibliographic design, we applied the concept of cross-referencing, as employed in modern full-text databases. Also, we wanted to have the full-text made fully searchable. Finally, we opted for the idea of implementing open access.

The resulting hybrid model turn out to be highly demanding, especially when it comes to the processing of references. These are given in the full format and are parsed into more sub-fields than in ISI products. More importantly, they all pass the process of normalization and authority control. Such a treatment ensures more related records, thus providing more powerful search. It also makes calculating fractional citation rate of authors possible. Finally, it ensures evaluation of some additional entities, such as academic publishers and state-supported permanent conferences.

Process

In initial selection of journals a few fairly liberal criteria were applied. Final selection based on quality was left to the impact factor to be generated later by joining the data from ISI citation indexes and *SCIndeks* itself. Pre-processing of data (text entry and spell-checking) was normally performed by combining commercial OCRs and MS Word, supported by personal dictionaries. In parsing article's (meta)data, a home-made tool named *CEON Parser* was used. Parsing references into sub-fields (author name, title, etc.) was automated by *Scriptor*, a semi-intelligent application described elsewhere (Pajić, Šipka, & Kosanović, 2002). Normalization of references and imputation for missing data was performed against compiled journal lists, research staff registers, publisher and conferences lists, and existing normalized *SCIndeks* records. Again, an in-house application (*Svedi*) had to be developed for the purpose.

Authority control was performed under *Svedi* against various bibliographic resources, such as local *OPACs*, *MedLine*, *ISSN online*, *WoS*, etc. External cross-linking of cited international references was carried out via *CrosRef Service* or *Open URLs*. Finally, the article full-texts in pdf. format were linked to their citations under *Svedi*. This is provided

only for selected number of journals which publishers agreed upon full presentation. A simplified overview of the whole process of *SCIndeks* preparation is given in figure 1.

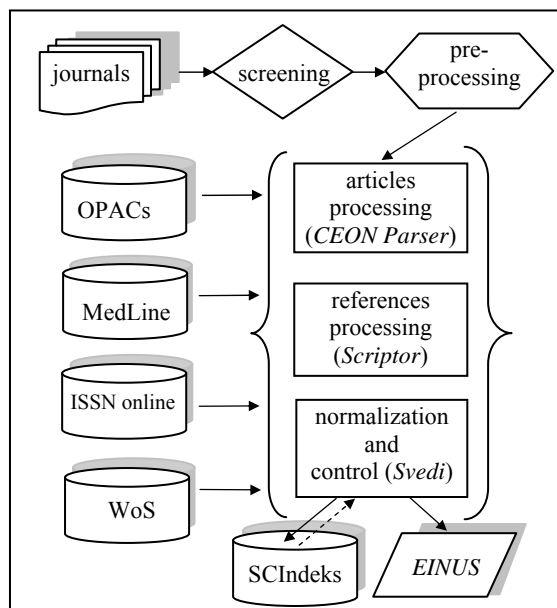


Figure 1: *SCIndeks* process model

Content

Promotion version of *SCIndeks* contains all articles published from 2002 to 2004 in 354 listed LPJs. Social science journals, 60 of them, are fully covered from 1991 on. Previous volumes of arts and humanities journals are planned to be appended to the database retrospectively.

SCIndeks is given in XML format and supports OAI Protocol for Metadata Harvesting, with built-in support for extended Dublin Core. It provides most search options commonly available in modern bibliographic databases. Search is possible within all fields including cited title, so far restricted to Serbian and English references. Search results are offered in *brief format*, leading to various article metadata or to the full text when available.

SCIndeks is supplemented by an analysis&report module named *Electronic Report on Research Output in Serbia (EINUS)*, operating as a stand-alone program. As a research tool it is intended at preparing data for bibliometric analysis. As a tool for practical evaluation it provides indicators for individual authors, institutions, research projects, journals, publishers, and permanent conferences.

Rankings of all entities by various indicators (row fractional, or weighted), are given in *EINUS* within 'natural' groups of members having about the same expected values for indicators. Groups may be formal, such as 'all law departments', or formed ad-hoc, e.g. 'all authors dealing with marketing'.

Indicators are generated for each publication year, as well as for the whole period, thus enabling analysis of trends. *EINUS* aggregates data from both *SCIndex* and *WoS*, showing overall performance of entities.

Sustainability

The usefulness of *SCIndeks* is not at stake. Thanks to both availability of most entrance materials in electronic form and the efficiency of our data processing tools, it is also a relatively cheap product, although the expenses are not negligible for a small, poorly funded academic community. Hence, the questions is not the cost-benefit ratio of the project, but rather who should pay the bill.

In a developing country such as Serbia, a NCD practically cannot be produced and maintained on the commercial basis. It should be entirely under care of the Ministry of science, at least in transition period, i.e. as long as the government is almost the only investor in S&T. In the long run, other parties coming from private sector, media, etc. could be attracted to use it and share expenses of its regular maintenance. By tailoring normative acts regulating evaluation to the potentials of *SCIndeks*, Ministry is in the position to secure the authority of the database, giving it a chance to become at least a self-sustained, if not a commercial product.

References

- Kosanović, B. & Šipka, P. (1996). SocioFakt - Jugoslovenska baza za društvene činjeničke nauke [SocioFakt: The Yugoslav database for social factual sciences]. In P. Kostić (Ed.), *Merenje u psihologiji* (pp. 85-95). Beograd: IKSI i Centar za primenjenu psihologiju.
- Kuang-hua, C. (2004). The construction of the Taiwan Humanities Citation Index. *Online Information Review*, 28, 410-419.
- Negishi, M., Sun, Y. & Shigi, K. (2004). Citation database for Japanese papers: A new bibliometric tool for Japanese academic society. *Scientometrics*, 60, 333-351.
- Pajić, D., Šipka, P. & Kosanović, B. (2002). Scriptor: Program za parsiranje bibliografskih informacija [Scriptor: A program for parsing bibliographic information]. *Infoteka*, 3, 13-21.
- Šipka, P. (2001). *Nauka u Srbiji: u susret evaluativnoj državi* [Science in Serbia: Towards an evaluative state]. CEON. Retrieved January 30, 2005 from: http://www.ceon.org.yu/erd_0101.htm
- Xin-ning, S., Xin-ming, H. & Xin-ning, H. (2001). Developing the Chinese Social Science Citation Index. *Online Information Review*, 25, 365-369.
- Yishan, W., Yuntao, P., Yuhua, Z., Zheng, M., Jingan, P., Hong, G., Bo, X. & Zhiqing, Y. (2004). China Scientific and Technical Papers and Citations (CSTPC): History, impact and outlook *Scientometrics*, 60, 385-397.