The Serbian Citation Index:
Context and content

Pero Šipka

esipka@sbb.co.yu

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

Introduction

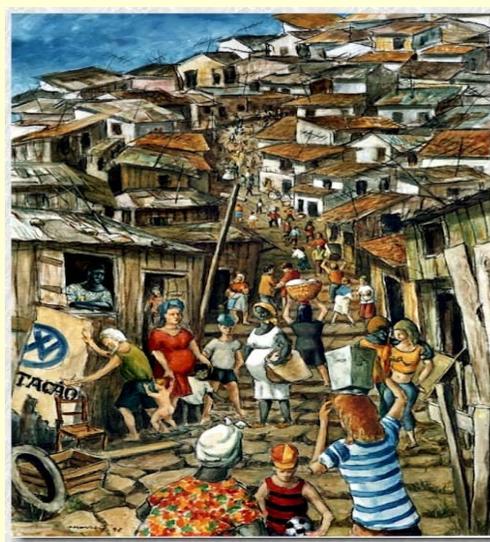Building national citation databases (NCDs) in developing countries to serve as add-ins to the ISI-Thompson citation indexes:

> ⇒ is an old idea,
>
> ⇒ oft contemplated, (a sort of an informetric doctrinal attitude)
>
> ⇒ but rarely realized (only in a few countries).

Perceived reasons (ordinary excuses) for the latter:

> ⇒ NCDs are unnecessary: we have ISI-Thompson products as supranational databases
>
> ⇒ content of locally published journals (LPJs) is not worth digitizing and keeping: developing nations have far more urgent needs in R&D sector
>
> ⇒ NCDs are expensive products

or, metaphorically:

NCDs are just satellite dishes in favellas: it's a sin to put a TV antenna on the roof of a house without running water.

Context

As far as Serbia is concerned, none of these arguments stands:

> ⇒ a NCD is necessary to make a non-arbitrary evaluation possible, since ISI-Thompson products are unable to discriminate performance of Serbian authors or institutions even in the most advanced disciplines
>
> ⇒ in some disciplines and by some classes of authors, such as novices and non-academics (practitioners) LPJs are locally more cited than international journals; and
>
> ⇒ NCDs can be made inexpensive if produced by new technologies and as multi-purpose, large-series product; multiplying purposes can be achieved by making a NCD a secondary publication, full-text archive, journal register, expert debase etc. at the same time, while enlarging number of users can be attained by making a product free and important.

Approach

Serbian NCD was designed as a hybrid product with two-fold purpose:

> (i)   to serve as an evaluation device and
>
> (ii) to ensure indirect (via bibliographic metadata) and direct access (to the full text of articles published in LPJs).

To make a breakthrough in a non-responsive, prejudiced environment and to bring about financial support, we were aware that such a product had to be:

> ⇒ as bibliometric/evaluation tool: reliable (accurate) and comfortable to the extent that makes unnecessary any data cleaning for evaluation purposes;
>
> ⇒ as bibliographic tool: an efficient promotional platform for LPJs, both nationally and internationally, and therefore freely available for all; consequently, its maintenance had to be as inexpensive as possible.

Process

To keep maintenance charges as low as possible, the process was automated:

<u>Pre-processing</u> of data (text entry and spell-checking) was performed by commercial OCRs and MS Word tools, supported by personal dictionaries.
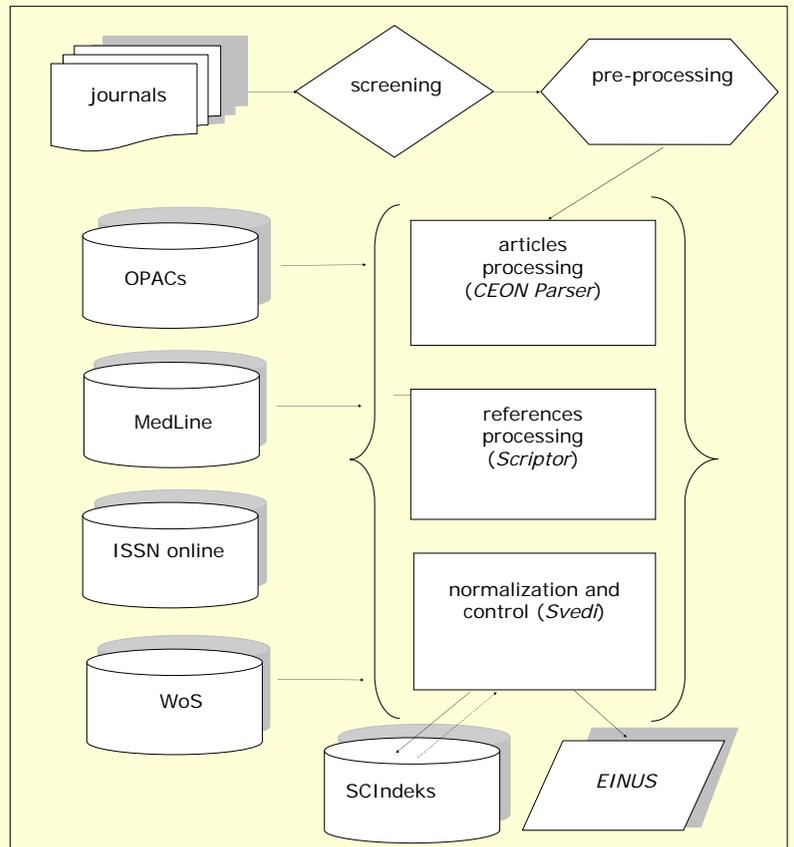
<u>In parsing article metadata</u>, a home-made tool named *CEON Parser* was used.

<u>Parsing references</u> into sub-fields (author name, title, etc.) was automated by *Scriptor*.

<u>Normalization</u> 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 under *Svedi,* an in-house application.

<u>Authority control</u> was performed against various bibliographic resources, such as local OPACs, MedLine, ISSN online, WoS, etc.

<u>External cross-linking</u> of cited international references was carried out via CrossRef Service or Open URLs.

journals → screening → pre-processing

OPACs → articles processing (*CEON Parser*)

MedLine → references processing (*Scriptor*)

ISSN online → normalization and control (*Svedi*)

WoS → SCIndeks / EINUS

Content

The resulting Serbian NCD (promotionally called *SCIndeks*) is

⇒ quite a large database for a small country:

SCIndeks contains about half a million of cited references from all articles published from 2002 to 2004 in 354 listed LPJs; all articles are given as abstracts and some of them as full text.

⇒ a resource of potentially high international visibility:

SCIndeks is given in XML format under the OAI Protocol for Metadata Harvesting, with built-in support for extended Dublin Core.

⇒ supplemented with powerfull ready-to-use evaluation module, (asks for no data cleaning; aggregating local and international bibliometric data), and tailored to the local needs (along with individual authors, institutions, and LPJs, it ranks state-supported projects, academic publishers, and permanent conferences)

Sustainability

Depends on the rate of transition process in the national academic sector:

⇒ to produce *SCIndeks* government support was necessary; conditions for comercialization didn't exist, since the government is the only investor in S&T;

⇒ to maintain and develope it further such support is still needed; government is in the position to secure the authority of the database by <u>tailoring normative acts regulating evaluation</u> to the *SCIndeks* potentials;

⇒ in the long run, along with the enhancement of <u>users familiarity</u>, other parties coming from <u>private sector</u>, media, etc. could be attracted to use it and share expenses of its regular maintenance.

user familiarity — "evaluatization" of academic sector — % of private investments into R&D — 3-D Line 1