2013 International Nuclear Atlantic Conference - INAC 2013
Recife, PE, Brazil, November 24-29, 2013
ASSOCIAÇÃO BRASILEIRA DE ENERGIA NUCLEAR - ABEN
***ISBN*: 978-85-99141-05-2**

NEW CONCEPTS OF NUCLEAR REACTORS AND FUEL CYCLES: PERFORMING AGILE TECHNOMETRIC STUDIES TO UNDERSTAND THE PROMISES AND THE CURRENT REALITY

José S. B. Reis Junior[1,2] and Antonio C. O. Barroso[1]

[1]Instituto de Pesquisas Energéticas e Nucleares, IPEN-CNEN/SP
Av. Professor Lineu Prestes, 2242
05508-000 São Paulo, SP
jsreis@ipen.br
barroso@ipen.br

[2]Bolsista da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)

ABSTRACT

The progress of previous projects pointed out the need to face some problems of software for detecting emerging research and development trends from databases of scientific publication. Given the lack of efficient computing applications dedicated to this purpose that we consider to be an artifact of great usefulness to better planning R&D programs in institutions, which are obliged to manage and develop, with limited resources and within the realm of complex and multidisciplinary technology fields as is the case of the Brazilian nuclear sector. We performed a review of the currently available software in such a way that we could clearly delineate the opportunity to develop new tools. As a result, we developed a software called Citesnake, which was especially designed to help the detection and study of emerging trends from the analysis of networks of various types extracted from the scientific databases. Using this powerful and stable computational tool, we performed preliminary analyzes of emerging research and development trends in a few thematic fields. The case that concerns this paper is the one devoted to the field of Generation IV Nuclear Power Generation Systems. We analyzed the productivity of authors, co-authorship networks, co-citation networks, development structure and emerging sub-areas of research. The idea was to find what reactors and fuel cycles have evolved more over the past ten years, in such a way to compare the what from the most promising concepts selected from the Generation Four Initiative have better evolved to fulfill some of their promises.

1. INTRODUCTION

The progress of previous projects pointed out the need of softwares for detecting emerging research and development trends from databases of scientific publication. Research in the nuclear field in Brazil unfolded in a rich variety of emerging technologies, opening up a range of many trends of scientific research, some of which have been consolidated as significantly important application areas for the whole society. Among others we may cite: the production of radiopharmaceuticals for diagnostic and therapy; radiation technology for improving the properties of special materials; hydrogen fuel cells; nuclear techniques for industrial process diagnostic. In addition or prior to these applications areas, there have been also significant advances in nuclear technology geared to reactor and fuel cycle development. Each of these areas somehow distinguishes among themselves because of specific characteristics such as product and development frameworks, cycle times and a focused series of specialized journals of preference.

This scenario revealed the opportunity of employment of emphtechnometric methods for detecting emerging trends - also called *scientometric* or *informetric methods* - as a powerful tool for the strategic management of national nuclear sector's research and development, underscoring the need for stable, reliable and powerful softwares able to help managers and scientists, through analysis of databases of scientific publications, towards such goal. But what are technometrics methods? Technometrics is a research field "devoted to quantitative studies of science and technology. It aims at the advancement of knowledge on the development of science and technology, also in relation to societal and to policy questions. A special, but certainly not exclusive emphasis is placed on the role of quantitative, in particular bibliometric (i.e., based on data from scientific and technological literature) methods" [1]. One of the most relevant among such bibliometric methods are techniques of *Citation Analysis*, consisting of "the examination of the frequency, patterns, and graphs of citations in articles and books" [2, 3].

On the other hand, at the beginning of the 2000s, the U.S. Energy Department spearheaded an initiative called Generation IV International Forum (GIF). The GIF intended, through a collective effort, find out which new concepts of nuclear power generation systems would be most promising for a more sustainable future. Under the IAEA, a similar initiative came a few years later, but this one was opened to wider participation of countries and institutions. This was named INPRO - International Initiative on Innovative Reactors and Fuel Cycles. Indeed, there are many reports and articles that discuss how both initiatives and the collaborative research originated by them marched. However, some twelve years later, there are few studies that address, in the point of view of concrete results, what systems seem to be the winners in terms of the possibility of implementation.

We performed a review of currently available software for assessing emerging research and development trends from data on scientific literature, so that we could clearly delineate both the possibilities and impossibilities of such computing applications and also the opportunity to develop new tools. As a result, we developed a software called *Citesnake* [4], which was especially designed to help the detection and study of emerging trends from the analysis of networks of various types coming from the scientific databases. The main features Citesnake, the reasons for its development and advantages and disadvantages with respect to other existing software is precisely the object of Section 2.

Once armed with a powerful and stable computational tool, we performed analyzes of emerging research and development trends in the field of Nuclear Power Generation Systems. By the application of technometric methods, we sought to find, from the lists of GIF and INPRO, the nuclear systems reactors and fuel cycles that evolved more over the past ten years. We also endeavored to identify, among the most evolved, those that could be considered the most appropriate systems for our efforts in R&D, in view of Brazilian needs and its technological base and natural resources. The results our preliminary of these analyzes are presented and widely discussed in Section 3.

2. THE CITESNAKE SOFTWARE

Among the various subfields of Technometrics, the Citation Analysis invariably stands out: this subfield, whose precursors were Derek J. S. Price and Eugene Garfield, has as relevant driving component the examination and study of graphs originated from citations of scientific works: these graphs can be - and indeed are - seen as networks of many types, as co-citation and co-authorship networks. A huge part of the work of a researcher who is engaged in Technometrics refers to analyze such networks. In this scenario, software dedicated to deal with networks originated by citations or by other information on scientific production become a tool which one cannot leave aside.

By 2012, the softwares devoted to Citation Analysis most used and acclaimed by the scientific community were: Publish or Perish [5]; HistCite [6]; CiteSpace [7]; ASE [8] and Sci^2 Tool [9]. However, all these had restrictions to scientific databases that could explore. The software that best address the needs of ongoing projects in our research group was the Sci^2 Tool, developed by scientists at the Indiana University. This tool was able to import data from the most used scientific banks, and allowed temporal, geospatial, topical and network analysis and visualization. Nevertheless, the use of this application brought us problems on consolidating data from different databases, when facing the need to build networks over information from more than one scientific database.

We contacted the developers of Indiana University, asking for clarifications on how the software could be used to cross analyze records from various scientific banks; the intention was to consolidate data of collaboration networks from records in files exported from Web of Science [10] and Scopus [11] databases. It was then explained that, once the collaboration networks were constructed starting from the files of each database, these networks should be merged and, once the formats of author names in each database was stored differently, the algorithm *DetectDuplicateNodes* should be used to collapse nodes representing the same author. We followed the instructions, and immediately some weaknesses of the algorithm *DetectDuplicateNodes* were shown:

- the algorithm collapses nodes whose labels are *similar*; this similarity is defined by a simple similarity calculation among *strings*;
- such similarity calculation requires the user to provide a *treshold* for the program, in order to be collapsed only nodes whose labels have similarity greater than this *treshold*;
- when working with hundreds or thousands of records, it is virtually impossible to guess a *treshold* such that there are no false positives or false negatives results; so the algorithm *DetectDuplicateNodes* becomes *probabilistic* and therefore it has an *accuracy rate*.

Then, a program that could completely nullify the portion of false positives and false negatives results was needed. We so realized that this could be done through the implementation of *translators*: the new application should lead to a common language languages of different databases, having the advantage that, as the language of scientific databases it is not a natural language, there wasn't the problem of ambiguity of meaning. Thereby, aside from previously known pathological cases or wrong information stored, the new software should *always* collapses the correct nodes.

Then, we turned our attention to the implementation fo some scripts in Python language [12], aiming to process the records from the point of view of *formal languages*, ie, by the usage of *regular expressions*. The immediate results were highly satisfactory. In July 2013 the beta 3 version of Citesnake was concluded: it is available for Microsoft Windows and Unix systems. This version has modules that allow the software to process records exported from some of the most important scientific dabases in the world:

- Web of Science;
- Scopus;
- PubMed [13];
- Google Scholar [14] (by making use of the program *Publish or Perish*).

After processing the records from these databases, Citesnake beta 3 exports a meta-network containing five networks:

- Article x Author network;
- Article x Keyword network;
- Article x Reference network;
- Article x Year network;
- Author x Institution network.

These five networks, if properly processed, provide information about scientific collaboration networks by theme, by institution, location and time, as well as several other data of interest to the user. The clustering of co-citation networks acn, for example, lead the user to detect the rise and fall of paradigms within a particular scientific field [15]. Citesnake exports a meta-network in the standard format of ORA [16], a leading software for social network analysis hitherto in use in our research group. Finally, there are virtually no limits to the number of records that can process Citesnake: the limitations are due to the hardware used.

3. PERFORMING AGILE TECHNOMETRIC STUDIES

Thus we proceeded to the use of citesnake in the development of agile technometric studies over the field of advanced nuclear reactors and fuel cicles. Firstly, we took the consolidated list of the Generation IV Forum, which consists of six innovative concepts of nuclear reactors:

- Very High Temperature Reactor (VHTR);
- Sodium-cooled Fast Reactor (SFR);
- Supercritical Water Reactor (SCWR);
- Gas-cooled Fast Reactor (GFR);
- Lead-cooled Fast Reactor (LFR);
- Molten Salt Reactor (MSR).

The goal was to determine the stage of research and development on each one of these six reactor types.

Through expert consultation, we sought to determine terms that would serve as "milestones" of the evolution of the concepts. For purposes of this abbreviated study, we set only ten milestones: simple terms that would denote well-defined significant steps in the progress from R&D to deployment on each type of reactor. The set milestones were:

- "Conceptual design";
- "Operating experience";
- "Technology projects";
- "Test facility";
- "First of a kind engineering (FOAKE)";
- "Fuel performance assessment";
- "Proof of concept demonstration";
- "Phenomena identification and ranking tables (PIRT)";
- "Code scaling, uncertainty, and analysis (CSAU)";
- "High burn-up demonstration".

We then proceeded to search in scientific databases. Among those whose exported results Citesnake can treat, we selected Web of Science, Scopus and Google Scholar; PubMed was discarded because we considered that a database dedicated to the field of Medical and Health Sciences would not contain a relevant amount of articles concerning concepts of advanced nuclear reactors.

3.1. Quantitative Analysis of Publications

In order to first recognize the research field under study, we composed charts of the amount of publications for each type of reactor per year since 2000, when the initiative of generation four reactors was consolidated, until the end of 2012. Publications of the year 2013 were not included in this analysis because, as the year is still in progress, not all publications were held or indexed, and this fact could falsely suggest a decline in the number of publications.

Figure 1 shows the number of publications indexed by Web of Science and Scopus per year and for each type of reactor. This graph shows the number of articles referencing each one of the six types of reactors at the title, abstract or keywords, but without obligation to have any presence of the milestones keywords. This way, we can detect scientific production concerning "pioneer" stages of development of each concept of advanced reactor.

Figure 2 also presents the number of publications indexed by Web of Science and Scopus per year and for each type of reactor, but now those publications that reference at least one of the slected milestones of the evolution of concepts.

Lastly, Figure 3 shows the number of publications indexed by Web of Science, Scopus and Google Scholar, per year and for each type of reactor, that not necessarily reference the any milestone of the evolution of concepts. It is interesting to note that this chart gives us an idea of the stage of development of each reactor concept in a "more informal world": Google Scholar indexes a series of papers and documents that have not yet been published or have not been indexed by the most famous scientific banks. Tools like Google Scholar allow us to identify the state of "hot fronts" of a given area of research.

3.2. Network Analysis of Publications

For this analysis, we focused on the records indexed by Web of Science and Scopus referencing at least one of the six types of reactors at the title, abstract or keywords, and not necessarily referencing some of the ten milestones of the evolution of concepts. Network analysis of this data set already allows an effective presentation of the developed methodology. Indeed, other sets of publications could be analyzed, but it would generate such huge amount of results to be presented that would make the work unsuitable for publication at INAC 2013.

Using the Citesnake we produced, from the set of records mentioned above, a meta-network containing nin networks: the default five networks generated by Citesnake; an special Paper x Reactor concept network, generated through the implementation of an special Citesnake module for this work; and three networks obtained by the *fold* operation:

- Paper x Paper network;
- Author x Author network;
- Keyword x Keyword network.

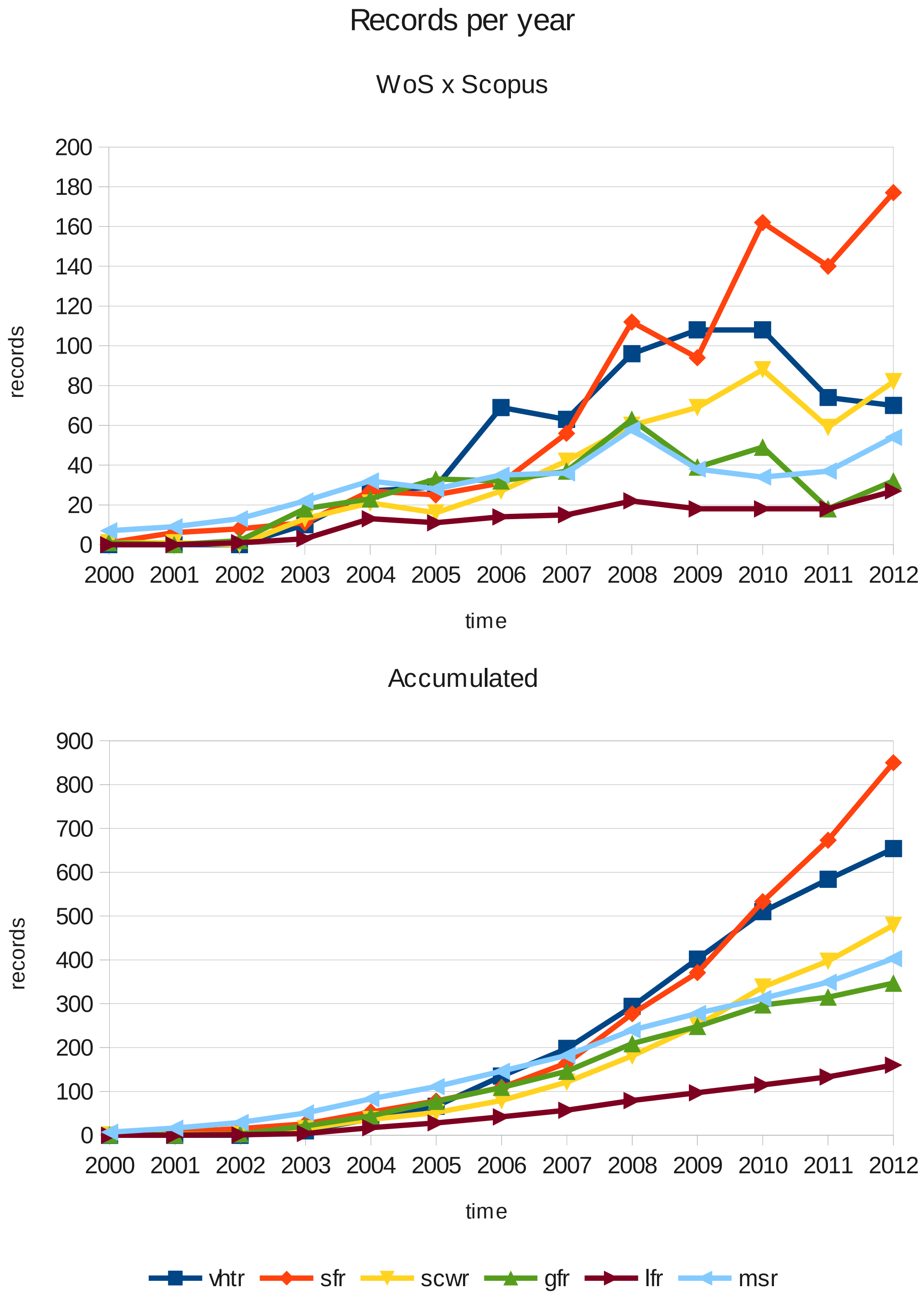

Figure 1: Number of records per year indexed by Web of Science and Scopus referencing each one of the six types of reactors at the title, abstract or keywords, and not necessarily referencing some of the ten milestones of the evolution of concepts.

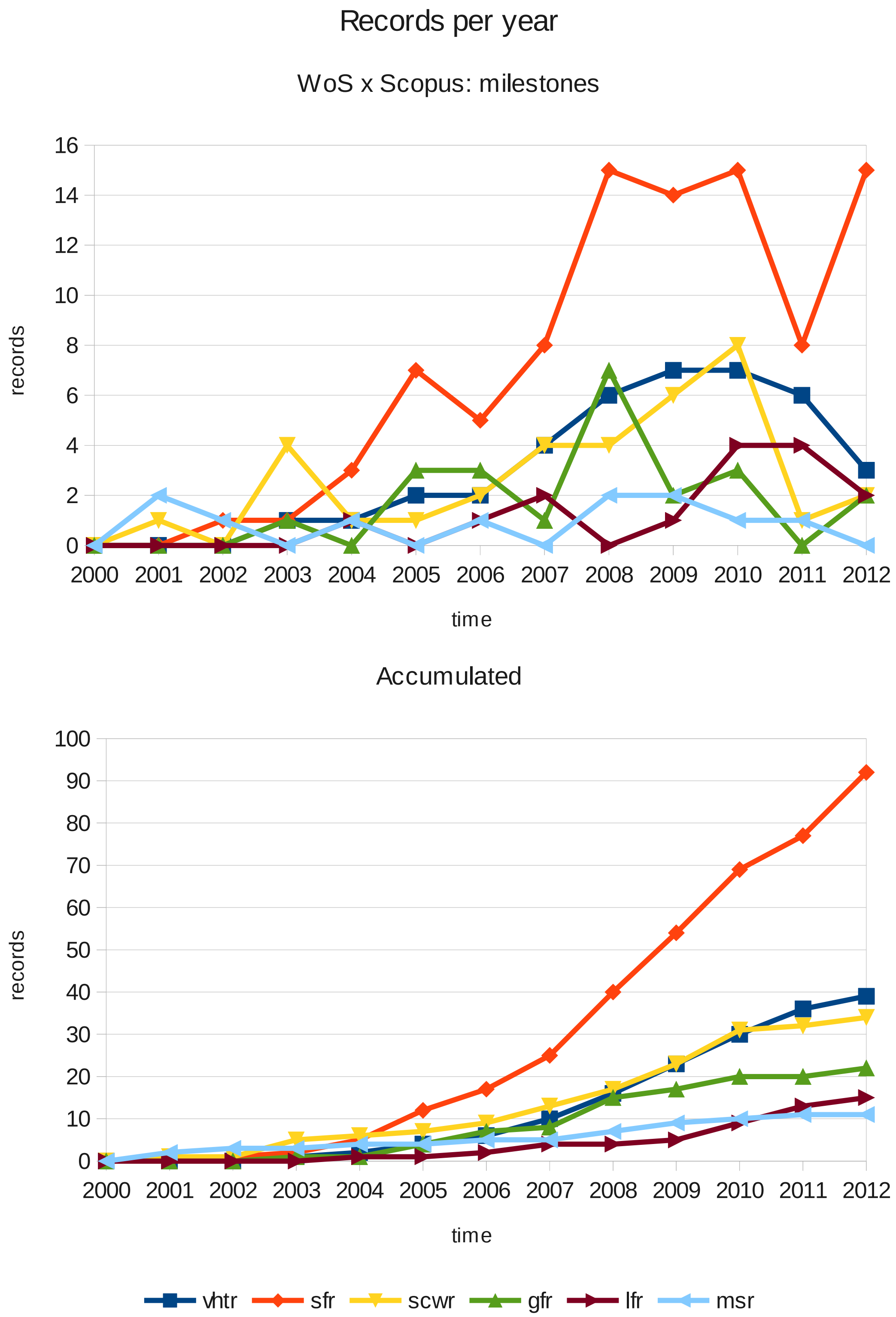

Figure 2: Number of records per year indexed by Web of Science and Scopus referencing each one of the six types of reactors at the title, abstract or keywords, and necessarily citing some of the ten milestones of the evolution of concepts.

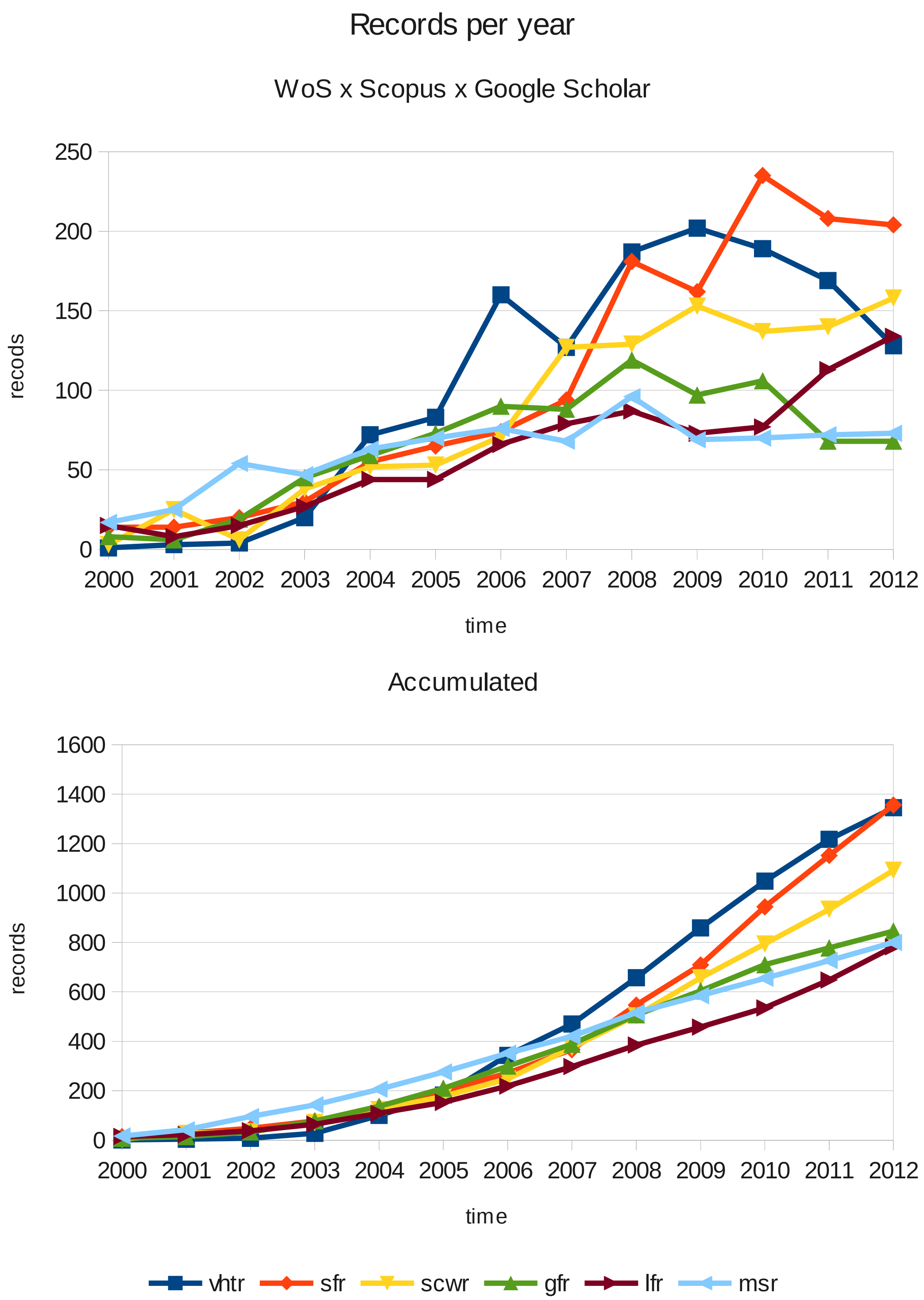

Figure 3: Number of records per year indexed by Web of Science, Scopus and Google Scholar referencing each one of the six types of reactors at the title, abstract or keywords, and not necessarily referencing some of the ten milestones of the evolution of concepts.

Table 1: Top ten institutions by number of active authors

Institutions	Degree
KOREA ATOM ENERGY RES INST	461
JAPAN ATOM ENERGY AGCY	364
CEA	229
INDIRA GANDHI CTR ATOM RES	177
IDAHO NATL LAB	146
XI AN JIAO TONG UNIV	137
MIT	113
ARGONNE NATL LAB	108
PAUL SCHERRER INST	101
UNIV WISCONSIN	97

Once the networks were generated, we calculated the key network measures in order to detect the key entities of each network.

3.2.1. Institution x author

By calculating the degree of each institution, we detected the most active ones, ie, those that group more active authors in the field of advanced reactors. The result is shown in Table 1.

It is interesting to note that all these institutions have published papers dealing with each one of the six innovative reactors concepts of interest, the exception of the institution “INDIRA GANDHI CTR. ATOM. RES.”, which had no paper concerning VHTR and LFR.

3.2.2. Author x paper

Based on the degree of each author, we detected the most active ones, ie, those that have published more papers in the field of advanced reactors, and concepts covered by them. Table 2 shows the result.

We can sort the concepts by coverage by these top-ranked authors, from the the least covered concept for the most covered one: SFR, LFR, SCWR, GFR, VHTR and MSR.

3.2.3. Author x author

Figure 4 shows the authors that are repeatedly top-ranked in the usual network measures. The value shown is the percentage of measures for which the Authors was ranked in the

Table 2: Top ten authors by number of published papers

Author	Degree	Concepts
Y. KIM	56	VHTR, SFR, LFR
K. MIKITYUK	46	SFR, GFR, LFR, MSR
H. OHSHIMA	43	SFR, GFR
S. KIM	42	VHTR, SFR, SCWR, LFR
I. PIORO	40	SFR, SCWR, LFR, MSR
X. CHENG	38	SFR, SCWR
R. CHAWLA	37	SFR, SCWR, GFR
P. HEJZLAR	33	SFR, GFR, LFR
H. YAMANO	30	SFR, LFR
P. V. TSVETKOV	29	VHTR

top three. In simple words, these authors can be seen as the more collaborative ones.

3.2.4. Paper x keyword

The calculation of the degree of each keyword allows us to detect the most popular keywords; these are presented in Table 3.

3.2.5. Papers x references

By calculating the degree of each reference, we get a very interesting result: the most cited references, and the innovative reactor concepts that each one deals with or that the papers citing such reference do. These are shown in Table 4. One should note that the references are in the standard format of Web of Science; oddly enough, this format does not show the title of the paper.

3.2.6. Fragmentation of mono-modal networks

The three mono-modal networks (paper x paper, author x author and keyword x keyword) could have their *fragmentation index* calculated: this index is a number between 0 and 1, where 1 indicates a fully fragmented network (all nodes are isolate). The most fragmented network was the author x author network: 62.2% of the nodes are disconnected. The less fragmented network was the keyword x keyword network: 28.1% of the nodes are disconnected. The paper x paper network presented a fragmentation of 31.8%.

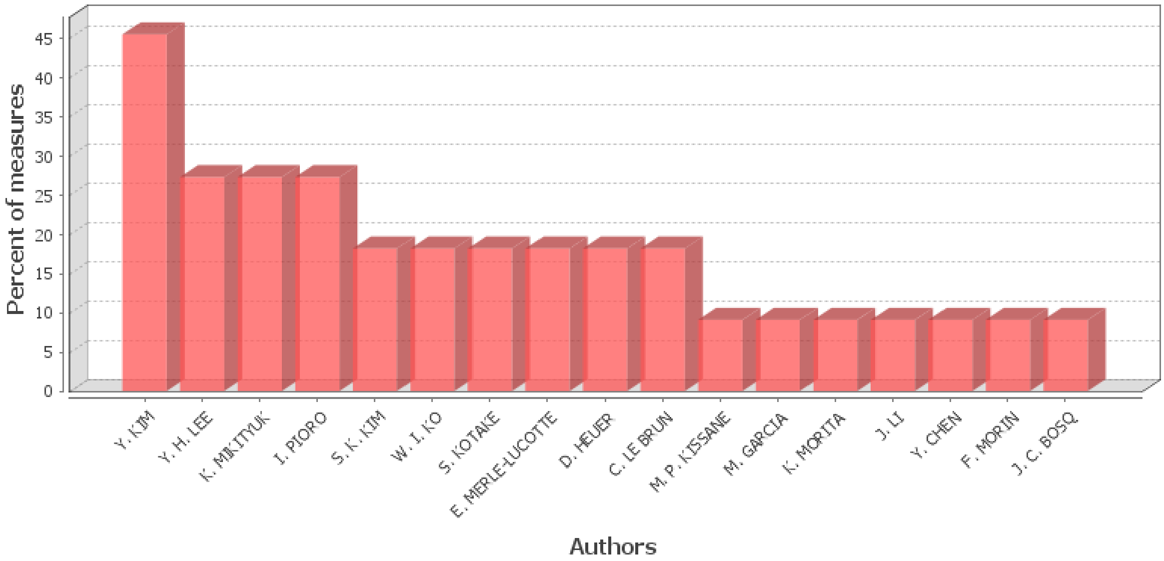

Figure 4: Recurring top ranked authors by usual network measures over colaboration network.

Table 3: Top ten keywords by reference on published papers

Keyword	**Degree**
vhtr	106
sodium-cooled fast reactor	68
molten salt reactor	60
fast reactor	54
scwr	52
hydrogen production	46
corrosion	44
sodium	34
supercritical water	34
heat transfer	33

Table 4: Top ten references by citation on published papers

Reference	Degree	Concept
YAMAGATA K, 1972, INT J HEAT MASS TRAN, V15, P2575, DOI 10.1016/0017-9310(72)90148-2	35	SCWR
CHANG J, 2007, NUCL ENG TECHNOL, V39, P111	30	VHTR
PIORO IL, 2004, NUCL ENG DES, V230, P69, DOI 10.1016/J.NUCENGDES.2003.10.010	23	SCWR
PIORO I. L., 2007, HEAT TRANSFER HYDRAU	23	SCWR
SQUARER D, 2003, NUCL ENG DES, V221, P167, DOI 10.1016/S0029-5493(02)00331-X	23	SCWR
US DOE NUCLEAR ENERGY RESEARCH ADVISORY COMMITTEE AND THE GENERATION IV INTERNATIONAL FORUM, 2002, TECHN ROADM GEN 4 NU	19	all
SHIMAKAWA Y, 2002, NUCL TECHNOL, V140, P1	18	SFR
MIKITYUK K, 2005, ANN NUCL ENERGY, V32, P1613, DOI 10.1016/J.ANUCENE.2005.06.002	18	GFR
MACDONALD P.E., 2003, INEELEXT0300870	18	VHTR
TORGERSON DF, 2006, NUCL ENG DES, V236, P1565, DOI 10.1016/J.NUCENGDES.2006.04.020	18	SCWR

4. CONCLUSIONS

The major objective of this study was to present the software Citesnake and a metodology for its use, by the usage of Citesnake as a tool for conducting technometric studies in the field of advanced nuclear reactors. The robustness of the software became quite clear after processing thousands of records in a few minutes; the variety of generated networks provided enough material for a deep technometric study, as we had to restrict the set of publications to be analyzed so that this work could remain within the scope of INAC 2013. We are working on improving the Citesnake by building modules that allow the recognition of languages other databases. Furthermore, the implementation of a graphical user interface for the software is well advanced.

Interpreting the results presented in Section 3, we must conclude that the SFR is the concept in a more advanced stage of development, followed by the VHTR and SCWR. This conclusion can be drawn by the volume of publications and the recurrence of these terms as keywords, and also because these are the types of reactors studied by leading authors and institutions, becoming the subject of the most significant references. Finally, low fragmentation of the network of keywords and high fragmentation of the network of authors suggests that the researchers involved in the field could work more closely with scientists from other institutions, so that the topics studied could be more diverse.

ACKNOWLEDGMENTS

We acknowledge the National Council for Scientific and Technological Development (CNPq) and the Coordination for the Improvement of Higher Level Personnel (CAPES) for the financial support of our research through the National Institute of Science and Technology of Innovative Nuclear Reactors (INCT-RNI) and through the *Pró-Estratégia* Program.

REFERENCES

1. A. F. J. van Raan, "Scientometrics: State-of-the-art", *Scientometrics*, **38**, pp. 205–218, (1997).
2. R. Rubin. *Foundations of library and information science*, Neal-Schuman Publishers, New York (2010).
3. E. Garfield. *Citation Indexing - Its Theory and Application in Science, Technology and Humanities*, ISI Press, Philadelphia (1983).
4. "Citesnake", http://www.quaero.org.br/citesnake/ (2013).
5. "Publish or Perish", http://www.harzing.com/pop.htm (2013).
6. "Thomson Reuters HistCite", http://interest.science.thomsonreuters.com/forms/HistCite/ (2013).
7. "CiteSpace: visualizing patterns and trends in scientific literature", http://cluster.cis.drexel.edu/∼cchen/citespace/ (2013).
8. "Action Science Explorer (Formerly iOpener Workbench)", http://www.cs.umd.edu/hcil/ase/ (2013).
9. "Sci^2 Tool : A Tool for Science of Science Research and Practice", https://sci2.cns.iu.edu/ (2013).
10. "Web of Science", http://thomsonreuters.com/web-of-science/ (2013).
11. "SCOPUS Database", http://www.scopus.com/scopus/home.url (2013).
12. "Python Programming Language", http://www.python.org/ (2013).
13. "PubMed", http://www.ncbi.nlm.nih.gov/pubmed/ (2013).
14. "Google Scholar", http://scholar.google.com/ (2013).
15. "Mapping Scientific Frontiers", http://www.sciencenews.org/articles/20040313/mathtrek.asp (2013).
16. "*ORA", http://www.casos.cs.cmu.edu/projects/ora/ (2013).