

Information organization and representation by graphic devices: an interdisciplinary approach

Nair Yumiko Kobashi¹, Raimundo N. M. Santos²

¹Universidade de São Paulo. Brasil. nykobash@usp.br

²Puc-Campinas. Brasil. rnsantos@puc-campinas.edu.br

Abstract

We present in this paper the theoretical and methodological challenges faced on the process of building subject maps of Brazilian dissertations produced in two scientific areas: Information science and Nuclear energy. The empirical data are constituted by bibliographic records of dissertations databases. The databases are traditionally used for information retrieval and production of bibliometric indicators. More over, they can also be seen as scientific memory devices of the intellectual works produced within the academic organizations. The study demonstrated that science maps, aroused from bibliographic database exploration, require previous, consistent information organization, based on explicit policies and on adequate terminological instrument for information representation (thesaurus and ontologies). The study also demonstrated that evidencing the multidimensional relations among the bibliographical records elements is fundamental to discover social and cognitive relations which are established within a scientific field and also to represent them through graphic visualization.

Keywords: Bibliometrics, Interdisciplinary studies, Knowledge organization and representation, Science maps.

Resumen

Presentamos en esta comunicación los desafíos teóricos y metodológicos enfrentados en el proceso de construcción de mapas conceptuales de las tesis brasileñas producidas en dos áreas científicas: Ciencia de la Información y Energía Nuclear. Los datos empíricos utilizados han sido los registros bibliográficos de las bases de datos de tesis. Las bases de datos se utilizan tradicionalmente para la recuperación de información y para la producción de indicadores bibliométricos. Asimismo, pueden también ser consideradas como la memoria científica de los trabajos intelectuales producidos dentro de las organizaciones académicas. El estudio demostró que los mapas científicos, derivados de la exploración bibliográfica de las bases de datos, requieren la organización previa y consistente de la información, basada en políticas explícitas y en instrumentos terminológicos adecuados para la representación de

la información (tesauros y ontologías). El estudio también demostró que evidenciar las relaciones multidimensionales entre los elementos de los registros bibliográficos es fundamental para descubrir las relaciones sociales y cognitivas que se establecen dentro de un campo científico y también para representarlas por medio de la visualización gráfica.

Palabras clave: Bibliometría, Estudios interdisciplinarios, Mapas científicos, Organización y representación del conocimiento.

1 Introduction

University institutions' scientific production bibliographic databases are traditionally used for computing researchers' productivity. However, these records, besides their academic management function, may reveal important aspects in social and cognitive institutionalization of research activity. According to Whitley (1974), cognitive institutionalization covers epistemological, theoretical and methodological aspects from scientific research, whereas social institutionalization refers to the formal structures which set scientific community members, their organization level and their integration to scientific activity accreditation' structures.

This paper, as part of a wider research project¹ is limited to presenting some data regarding cognitive institutionalization in Information Science and Nuclear Energy fields. It is an interdisciplinary study, supported by theories and methods of Knowledge Organization and Bibliometrics. Both areas have advanced, theoretically and methodologically, over the last few decades, and have also been benefited by information and communication technology development. According to the point of view adopted herein, gathering Information Organization and Bibliometrics might improve science indicators production studies, as well as their expression by means of maps.

In order to identify cognitive institutionalization aspects, we have elected as empirical data the dissertations produced at Brazilian master and doctoral programs of Information Science and Nuclear Energy. We started from the idea that such production largely expresses the history of scientific activity in these two areas within the nation. This approach is configured as an alternative to traditional studies of scientific production analysis which are usually based on articles published in high impact international magazines. Political and ideological criteria adopted by these magazines, regarding article selection and publication, turn out to be questionable in their representativity for mapping scientific activity in areas, countries or regions. We also bear in mind that master and doctoral programs in Brazil are linked to a national system of master and doctoral education program which usually assesses the quality of programs and dissertations produced; the latter may be considered as reliable scientific study activity throughout the country.

¹ Institutionalization of scientific research in Brazil: thematic and social network cartography by means of bibliometric techniques; this project has been developed since early 2006, with the participation of senior researchers, undergraduate, mastering and doctorate students of Information Science.

2 Methodological procedures

For the empirical preliminary approach, so far of exploratory nature, we have performed the following procedures: a) corpus definition; b) definition of metadata fields to be concerned, in this case, the descriptors field; c) consolidation of corpus data through key word standardization, aiming at terminating dispersions which may commit result analysis and validation; d) definition of technological supporting tools: *Infotrans*, *VantagePoint*, *Dataview*, *Excel/Statistica*, applications, which allow reformatting starting data, transforming them into bibliometric records, accounting them and representing them as matrix and/or distribution by frequency, for statistical treatment and *clusters* automatic generation purposes.

3 Preliminary results

The main obstacle and challenge that comes up when studying a scientific field cognitive institutionalization arises from the subject characterization form on database dissertations. We can observe large variation in descriptors' attribution: they are either too general or too specific, and so there are not in fact any standardization of expression forms for equivalent concepts. There is, on the other hand, excessive or insufficient attribution of descriptors for identifying the subjects approached in the studies. Such a situation prevents the database record direct use. Therefore, in order to obtain valid results, an ad hoc database is necessary, especially organized for studies of such nature. We shall then present and comment the graphs obtained in this study.

Figure 1 shows thematic data representation in Information Science dissertations:

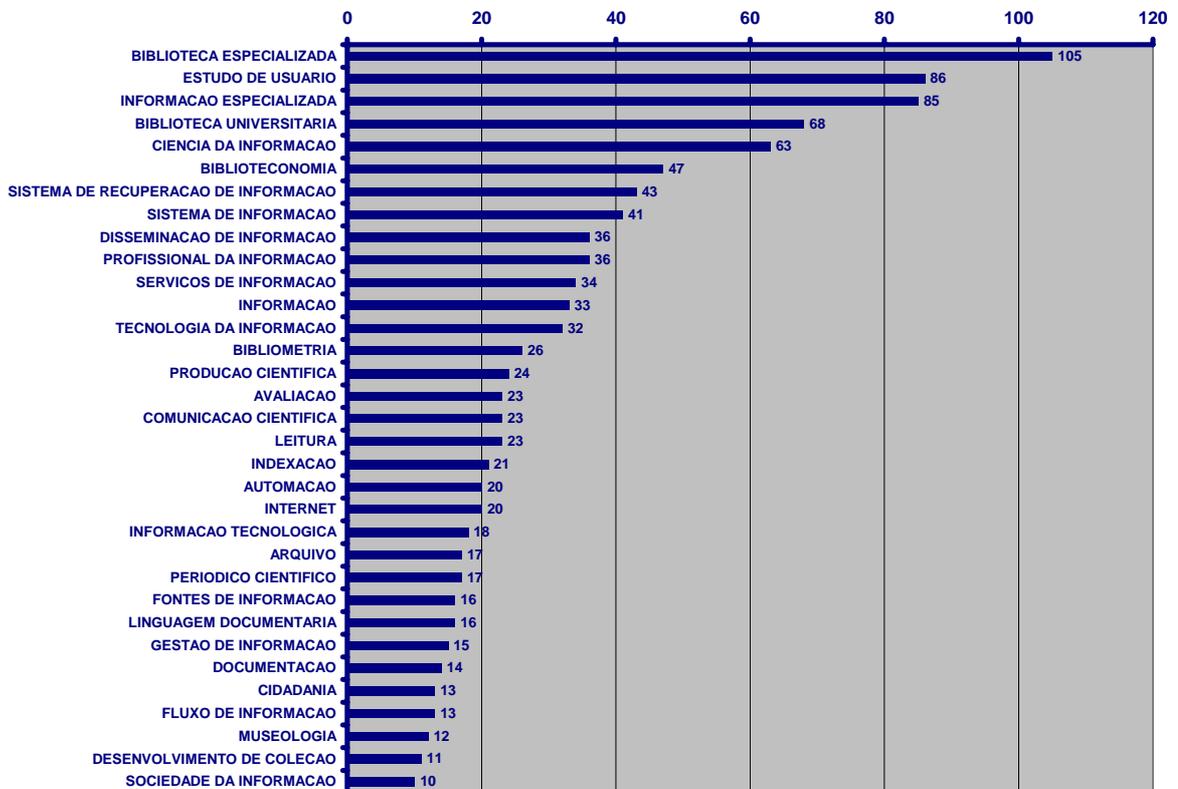


Fig. 1. Frequency of Dissertation subjects in Information Science

The graph above (Fig. 1) has been obtained based on descriptors of the database keyword field. However, they are limited to a subgroup, facing the exploratory aim of this study. We also note that the subjects Libraries, Academic libraries and Special libraries have been largely approached. One may ask whether these subjects do not reflect more a local troubleshooting perspective, by means of case studies, than an effectively research concern to develop the area theoretically and methodologically. On the other hand, we notice that descriptors referring to Information Science and Library Science begin to appear in this group; this perhaps indicates the emergency for epistemological concerns. Likewise, Information Society and Citizenship may indicate the widening of research interests, signaling a possible concern for social and political contextualization of approached subjects. There is also the presence of descriptors such as Archivology and Museology, which may be indicators that these two subjects are identified as research areas integrated to the Information Science field.

The figures below present the Nuclear Energy area. Regarding Figure 2, by ranking the subjects according to their size (based in number of dissertations) in a graphical display, it was possible to obtain an overview of the scientific profile of the Master and Doctoral Program of IPEN (Instituto de Pesquisas Nucleares-Sao Paulo). It strengths the main characteristic of the nuclear area, their interdisciplinary features, as the nuclear techniques could be applied in different knowledge areas, as materials science, physics, chemistry environment and so on. From the 45 subject categories currently used by the INIS/ETDE databases, it could be observed that all the dissertations of this Institution are spreading in 33 of them.

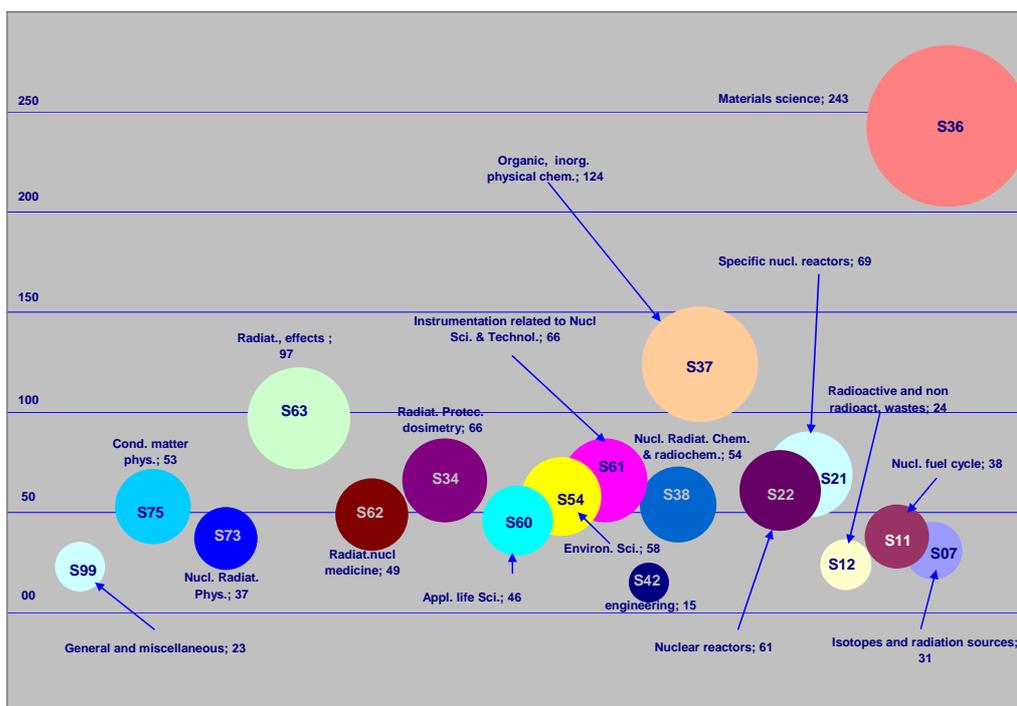


Fig. 2. Ranking of themes in Nuclear energy

Fig 2 shows the current ranking of the major subject categories. The size of the bubbles represents the number of bibliographic issues. Three areas could be considered, today, the Research Front of the Program: Materials Science (243 dissertations), Organic and Inorganic

Chemistry (124 dissertations) and Radiation thermal and pollutant effects in biological materials and living organisms (97dissertations).

Other areas such as: studies in specific and general nuclear reactors, instrumentation related to nuclear science and technology (dosimeters, chambers, detectors) and radiation protection, present a balanced performance.

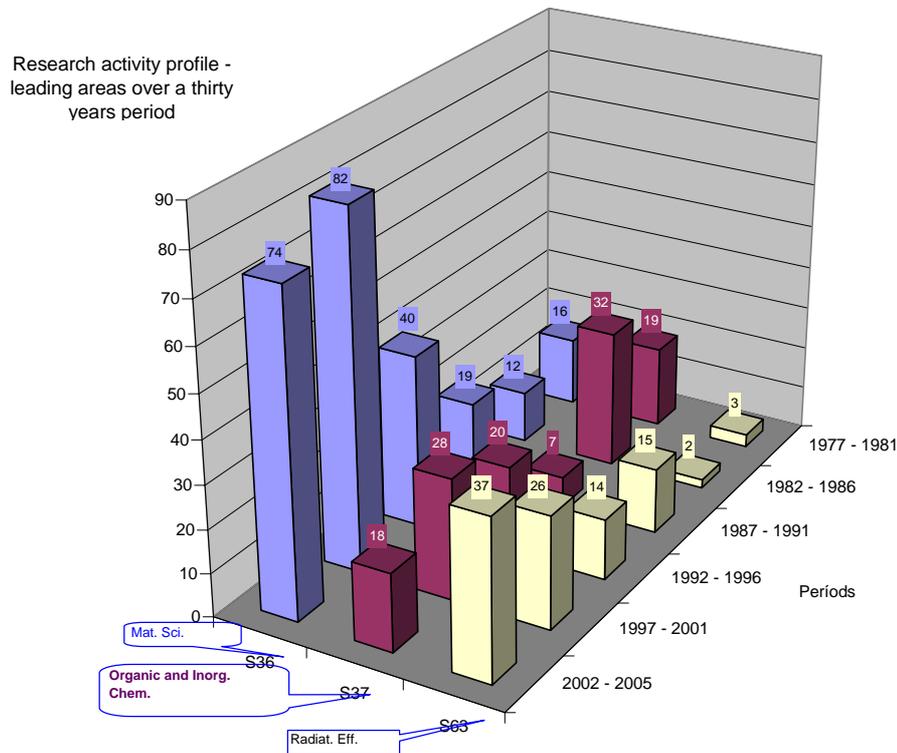


Fig. 3. Development of research areas in Nuclear Energy

Fig 3 shows the development of the three leading areas along the studied period. By a retrospective analysis it could be observed that Materials Science (S36) maintains a linear performance along the period, followed by the Organic and Inorganic Chemistry (S37) which maintain an intermittent ranking position. Concerning the Radiation Effects area (S63) it is remarkable its development along the period.

4 Final Considerations

We have, in this study, attempted to make early impressions about Information Science and Nuclear Energy cognitive institutionalization in Brazil. We have considered in this study the dissertations produced between the years 1977-2001, regarding Information Science, and the years 1976-2005, regarding Nuclear Energy. We have avoided bibliometric approaches laid on merely quantitative criteria. Such criteria, if used isolatedly, are not very efficient for obtaining significant data. In such perspective, we have combined quantity and quality, and that aspect is bound to some considerations.

Bibliometric studies on subjects are based on data classification. Classifying is essentially a descriptive process which consists primarily of choosing empirical data dimensions which

will then be divided into discreet, rankable elements. Therefore, description generates a morphological knowledge which works as evidence of a whole and its parts.

We have assumed herein that qualitative knowledge about subjects may be targeted by quantified relations among descriptors, coming from bibliometric techniques application. Therefore, we can state that qualitative knowledge does not eliminate quantity; it seeks taking the measure as a means of understanding and explaining so as to neutralize dicotomy between quantitative and qualitative means of analysing objects.

Specifically in this study, we needed to neutralize quantitative dispersion so as to form new descriptor groups, i.e., new classes whose approximations have expressed “family ties”, as proposed by Wittgenstein. This “family tie” holds objects related to each other, not by a constellation of characters which transform into severe classes, in which class objects share a whole invariable core. They are, instead, groupings which show some inexactitude and flow, just like the indexing language descriptors. The conceptual abstraction is then the operation which allows corpus data comparison.

Recategorization operations of corpus terminological data, however, gain consistency when they own terminological control instruments such as controlled vocabulary and explicit indexing policies.

This task on Information Science has demanded intense work of descriptor standardization, for dissertation database records are produced in decentralized manner, according to each institutions’ indexing policies. Previous consistency methods of descriptive data are not previewed herein, nor are descriptor standardization for integrating them on a national basis. Descriptors are usually attributed by the authors of dissertations themselves; their standardization, by means of controlled vocabulary, is not a common practice. Nuclear Energy data, however, due to the fact that they are treated sistematically, based on an ever update thesaurus, allow inputting greater accuracy to the data reordering process.

Empirical base reorganization, such as the one performed in this study, seems to be fundamental in providing operationality to semantic nature data, as dissertation subjects are. We have sustained the data reorganization based on the idea that each class foresees a range of characteristics which can be attributed to all class members. Therefore, descriptors which promote excessive dispersion may have their traces neutralized since they are coherently integrated to more general classes.

In a reference database exploration project, for science cognitive institutionalization, empirical base reindexing is a means of restructifying towards remeaning. The information graphic representation is justified by the functionality for obtaining global information visualization as well as to highlight structural and contingent relations among data.

The results we have gotten so far, illustrated in the 3 figures above, indicate that work hypothesis adopted in this research is quite promising; such fact cheers us up to carry on by deepening theoretical and methodological bases which comprehend this current research project.

References

- BRAAM, R. R.; MOED, H. F.; VAN RAAN, A. F. J. Mapping of science by combined co-citation and word analysis. 1. Structural aspects. *Journal of the American Society for Information Science*, 1991, vol. 42, n. 4, p. 233-251.
- COURTIAL, J. P. *Introduction à la scientométrie: de la bibliométrie à la veille technologique*. Paris: Anthropos, 1990.
- HERRERO-SOLANA, V.; HASSAN, Y. Metodologías para el desarrollo de interfaces visuales de recuperación de información: análisis y comparación [electronic resource]. *Information Research*, 2006, vol. 11, n. 3, paper 258. <<http://InformationR.net/ir/11-3/paper258.html>>. [Consulted: 10 feb. 2006]
- KOBASHI, N. Y.; SANTOS, R. N. M. Information Science research institutionalization in Brazil: a bibliometric study. In: *International Conference on Multidisciplinary Information Sciences and Technologies (1st. Mérida, Spain. 2006)*.
- KOBASHI, N. Y.; SANTOS, R. N. M. Arqueologia do trabalho imaterial: uma aplicação bibliométrica a análise de dissertações e teses. In: *Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB) (7th: Marília, Brazil: 2006)*.
- NOYONS, C. M. (ed.). *Bibliometric mapping as a science policy and research management tool*. Thesis, Leiden: DSWO Press, 1999.
- SANTOS, R. N. M. dos [et al.]. Brazil's insertion into nuclear energy research: bibliometric analysis of dissertations developed at the Instituto de Pesquisas Nucleares-São Paulo. In: *International Conference on Multidisciplinary Information Sciences and Technologies: (1st. Mérida, Spain. 2006)*.
- VAN RAAN, A. F. J. *The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments*. Pre-print, 2003.
- WHITLEY, R. Cognitive and social institutionalization of scientific specialities and research areas. In: WHITLEY, R. (ed.) *Social processes of scientific development*. London: Routledge and Kegan, 1974. P. 69-95.