

BIBLIOMETRIC ANALYSIS APPLIED TO A SCIENCE POLICY DATABASE

Maria Flaminia RAMOS
Junta Nacional de Investigaçao Cientifica e Tecnologica (JNICT)
Av. D. Carlos 1, 126 -2
1200 Lisboa, Portugal

ABSTRACT

The bibliographic database on science policy developed at JNICT is briefly described as well as the changes made in its structure in order to be able to use it for bibliometric purposes. The first objective was the management of the collection of periodicals received in the library, However as the work was carried on we realised that it could be interesting to consider the following variables : title of publication, edition language, edition country, corporate source, descriptors and subject code. Though the sample is very limited because we only considered 1,532 articles relating to the periodicals received in 1987 and 1988, we are confident that future developments can be interesting.

RESUME

Cet article présente la base de données bibliographiques mise au point par le JNICT ainsi que les changements introduits dans sa structure afin de pouvoir l'utiliser dans le cadre d'études bibliométriques. Le premier objectif était la gestion de l'ensemble des périodiques reçus par la bibliothèque. Les variables suivantes ont été définies : titre de la publication, langue, pays de publication, le pays et l'institution de l'auteur ainsi que des descripteurs et des codes par domaine scientifique. Bien que la taille de l'échantillon soit très limitée (1532 références incluses dans les périodiques reçus en 1987 et 1988) nous pensons que des développements futurs peuvent se révéler intéressants. Nous présentons ici les résultats de ce premier essai.

INTRODUCTION

The use of scientific and technical databases for other purposes than the conventional bibliographic search has been very much developed during the last decade. Though the subject oriented searching continues to be considered as the main use of databases it is also true that the manipulation of big quantities of data rendered possible by automated systems enable to perform studies that are practically impossible to perform by manual methods. The quantitative analysis of the bibliographic characteristics of documents - bibliometrics - is one of these

studies and for its development the automated databases are a powerful tool (1). Barbara Stefaniak (2) points out the various searchable elements that describe the bibliographic characteristics of the documents referred in a database. "Some of them are subject oriented such as classification codes, descriptors, key words, words in the title, while other features point out the type of publication (e.g. Journal papers, conference papers, books, patents, reports), source (e.g. journal title, country of its editor, ISSN number, patent number, and year of publication, volume number of issue, pages), language of publication, name and corporate affiliation of the authors (name of organization, city, country), as well as data on secondary source (year, volume and number of the abstract)."

Besides these characteristics it is also possible to find lists of references that were cited by the authors of particular papers and this is what makes the difference of the databases developed at the Institute of Scientific Information in Philadelphia. The citation indexes prepared from those lists opened new areas of research in scientometrics - areas dealing with all the quantitative aspects of science of science - which has been very much developed since the publication of Science Citation Index (SCI) in 1961 (3). Through the citation analysis and in spite of all criticism involving citations above all when small countries or non English speaking countries are in question (4), many interesting aspects of the science of science have been studied such as science mapping, interdisciplinary relations, detection of new areas of research, the obsolescence of scientific literature, etc. But it is as a source for the evaluation of scientific production that the SCI has been usually used.

The study of science through the analysis of a set of scientific documents contained in a bibliographic database can be made in different directions. HAWKINS (5) described some of them: identification of competing research organizations or authors; comparisons of journal coverage; identification of leading Journals in a given field; finding neglected areas of research.

The identification of the leading Journals in a given field is very interesting in librarianship as it makes easier the management of the library collections. This is the aspect that motivated the study presented in this paper concerning the structure of the database on science policy "C&T" developed at the Scientific and Technical Information Division of the National Board for Scientific and Technological Research. The database contains interdisciplinary information distributed by four main fields :

- scientific and methodological foundations of science and technology policies;
- scientific and technological potential;
- formulation, implementation and monitoring of science and technology policies;
- sectoral science and technology policies.

The database began in 1976 and contains articles of the main journals specialised in the area of science policy, monographs, reports and papers to congresses and expert meetings. The total number of references is around 15.000 from Portuguese, foreign and international origin. The process of automation started in 1987 and the database contains all the documents received since then

and also those referring to 1986. For the purpose of the present study were only considered the article of the journals entered in the base during 1987 and 1988, in the total number of 1,532 references distributed by 119 journals.

OBJECTIVE OF THE STUDY

As previously mentioned the first objective was to contribute to a better management of the collection of journals received in the library by knowing the core of journals which contributed the most to the database. But as the work unfolded it was decided to take into consideration other variables taking advantage of the changes that meanwhile have been made in the structure of the base.

METHODOLOGY

The "C&T" database has been developed with the Unesco Mini-micro CDS-ISIS software adapted by Portuguese National Library PORBASE. As that version didn't include some of the variables we were interested to search some changes in the structure of the base have been made and consequently in the input of data. But it was also necessary to harmonize some criteria and standardize some data. As a matter of fact the use of the base for bibliometric purposes had not been foreseen before and therefore the detection of some lack of standardization was only possible when the search was needed in certain fields.

The wide variation and great inconsistency that exists in bibliographic databases in representing the same journal titles, author names, corporate organizations etc. is usually taken as a constraint for their easy utilization for bibliometric purposes. Various authors have dedicated their attention to the aspects of harmonization and standardization mainly as far as authors and corporate sources are considered. (6 and 7)

In the definition of the base the most important chance was the introduction of the field corporate source. Considering the experimental phase of our study this field was only fulfilled when it referred to Portuguese authors. The following variables were introduced:

- Title of the publication (TI)
- Edition language (LG)
- Edition Country (OR)
- Descriptor (AS)
- Subject Code (CD)
- Corporate Source:
 - Country of the author (PA)
 - Institution (IA)
 - Department (DA)
 - City (CA)

For descriptors and subject categories we are using the UNESCO Spines Thesaurus (Science Policy Information System).

RESULTS

Dispersion of articles

Journals were ranked according to their decreasing productivity so that the journal relating to science policy and contributing the most articles to the base was ranked in the first place, the journal with the next greatest number of articles was ranked number two, and so on. In the present sample we find that the largest number of articles contributed by a single title was sixty three. On the other hand there were twenty six journals contributing only one article each. We can still observe that 18 journals (15,1%) contributed with half of the articles

Language distribution

Table 1. Language Distribution

Language	Number of articles	%
English	1,036	63,0
French	285	17,3
Portuguese	285	17,3
Spanish	38	2,3

We can observe that more than fifty per cent of the documents are of English origin and that Portuguese has a small share, which as a matter of fact is still smaller because all the documents written in Portuguese are included here. This means that they can stem from any Portuguese speaking country.

Distribution by edition country

When we compare the distribution of country of edition with the distribution by languages we see, as far as Portuguese is concerned, that there is a decrease in the number of documents edited in Portugal since sixty eight of them are edited in Brazil. This confirms the observations made above (Table 2).

Table 2. Distribution by editing country

USA	258
FRANCE	256
HOLLAND	254
PORTUGAL	215
SWITZERLAND	79
BRAZIL	68
OTHERS	505
TOTAL	1635

Corporate source

As previously mentioned this field was created in the scope of the present study. Therefore the input of data concerning the institution, the department and the city of the author was made always when those data were mentioned in the article but only if they concerned Portuguese authors.

Distribution by Spines subject categories

The category *BO3 Scientific and technical information* is the one which is gathering the greatest number of articles. We have to admit that this is due to a certain ambiguity that exists in this category because it includes all the theoretical aspects of information science, scientific and technological information, information processing documentation software, factual data and statistics ... etc. (cf. Table 3).

Table 3. Distribution by SPINES Subject categories

SPINES subject categories	No. of articles
A00 - Foundations of S&T policy-making of which Sociology of Science = 96	306
B00 - Science and Technology Ressources of which .Human Ressources = 91 .S&T information = 324	513
C00 - Practice of S&T policy-making of which .Elaboration, implementation and monitoring of S&T policy = 136 .S&T forecasting and assesment = 250 . Organization and management = 227	735
D00 - General content and results of S&T plans, programmes and projects	269
TOTAL	1581

Language distribution by subject category

As an example we show below the language distribution concerning the category A04 *Sociology of Science and Technology* and once again we can see the great representativity of the English language: 73,9%

The total number of articles is 96 distributed as follows

English	71
French	10
Portuguese	13
Spanish	2

Other indicators can be used as for exemple the number of authors by country in each SPINES subject category on the frequency of some descriptors in relation to the journal titles.

Final remarks

According to Moravcsik (8) bibliometric indicators as an evaluation method are less problematic and less questionable when they are applied to big quantities of documents. In the present case we consider that the sample is too small to achieve conclusive results and therefore they must be considered as provisional.

Future developments of the 'C&T' database aiming at bibliometric studies can supply important indicators for decision making concerning the management of library collection of journals. But they can also contribute to a better knowledge of the scientific production in the area of science policy, identify gaps, research groups or research centres

It could also be possible to analyse research trends by following the scientific production in a certain area for a given period of time.

More detailed studies, namely those concerned with co-word analysis would only be possible with the aid of powerful cumputerised means and the support of specialized human ressources. It is a very active research line for example in France and in United Kingdom. It will be very helpful for future oriented and evaluation studies but it is far from our present capacities.

For a small country like Portugal where the representation of scientific production in International databases is not very accurate it seems that the development of local databases enabling the measurement of S&T outputs can be an useful tool for the evaluation of research and for the decision making. It is not the only method but when used with peer review or other evaluation methods, it makes the judgement process much more independent.

References

- (1) HAWKINS Donald T. - Unconventional uses of on-line information retrieval systems: on-line bibliometric studies. *Journal of the American Society for Information Science* 28 (January 1977). pp 13-18.
- (2) STEFANIAK, Barbara. Use of Bibliographic Databases for Scientometric Studies. *Scientometrics* 12(3-4),1987 pp.149-161.
- (3) GARFIELD E.. Citation Analysis as a Tool in Journal Evaluation. *Science* 178(4060), 1972, pp.471-79.
- (4) MAcROBERTS Michael H. and MAcROBERTS Barbara R. Problems of Citation Analysis: a Critical Review. *Journal of the American Society for Information Science* 40(5),1989 pp.342-349.
- (5) HAWKINS op.cit.
- (6) GOMEZ I and GALBAN C. Lack of Standardization in the Corporate Source Field of Different Databases (pre-print).
- (7)HAWKINS op. cit.
- (8) MORAVCSIK Michael J. trad. ALVAREZ-OSORIO, Jose Ramon Perez. Como evaluar la ciencia y a los científicos? *Revista Espanola de Documentacion Cientifica*, Madrid 12 (3) 1989, pp.313-325.