Contents

Research Evaluation

Volume 3 Number 1
ISSN 0958-2029

April 1993

Summaries of articles are repeated on the back cover

Special issue on bibliometrics, informetrics and scientometrics
Guest editors
Wolfgang Glänzel and Hildrun Kretschmer

Continued from December 1992 issue

Research *Evaluation*, volume 3, number 1, April 1993, pages 13-18, Beech Tree Publishing, 10 Watford Close, Guildford, Surrey GUI 2EP, England.

Patent analysis

Bibliometric analysis of patent documents for R&D management

L Quoniam, P Hassanaly, P Baldit, H Rostaing and H Dou

The links between innovation, competitive intelligence, patents and bibliometric analysis are briefly described, mainly to introduce the reasons for our bibliometric example. Results of an original statistical classification of patents are presented to show the complementarity between documentary classifications and how patent data (and not patent citation) can provide strategic information.

L Quoniam, P Hassanaly, P Baldit, H Rostaing and H Dou are at CRRM Fac St Jérôme, 13397 Marseille CEDEX 20, France.
The authors would like to thank INFOPRO (ORBIT) for their valuable help during this study.

RESEARCH AND DEVELOPMENT (R&D) policy is often directly linked to the market. A good R&D policy can lead a company to successful innovations, and be the engine of innovation. Given the international nature of many markets, companies must develop their competitiveness. Several books and articles relate innovation to the need for good management.[1,2] Innovation is defined here in a broad sense, and includes:

- innovation as new products and services sold by a company;
- innovation in the processes used to supply these products and services.

Innovations are often protected by patents.3 There are numerous instances of inventive companies not protecting their inventions, and being deprived by others of their just rewards. Industrial and intellectual property protection is a tool that allows the company a monopoly over the exploitation of these assets, and provides a link between R&D policy and the satisfaction of market needs.

R&D policy, inventive activity and innovation, and industrial intellectual property are interdependent. An active and well adapted property policy can have an effect on innovation.

Competitive intelligence

Briefly, competitive intelligence consists of scanning the external industrial environment. It can be divided into various functions, but it is quite easy

0958-2029/93/010013-6 US$08.00

to see that a specific 'Patent Competitive Intelligence System' can be efficient for gaining new markets or expanding in foreign countries. This kind of activity must be organised, structured and, primarily, appropriate.

One of the main goals of a competitive intelligence system should be to feed the innovation process with accessible information, about:

- new technologies, for existing products and processes;
- new products, for existing markets; and
- new products for new markets.

Martinet[4] says that around **70%** of inventions arise from a market need (with, of course, many variations from area to area). This leaves perhaps **30%** in which scientific programmes may play an originating role. But whatever their source, any commercial importance they have comes from a slow adaptation arising from the science-market synergy. For many sectors, patents seem to form a good picture of this synergy. Science generates items which can be patented as products and processes, and the market forces inventors to protect their interests.[5]

Research programmes do not necessarily lead to any hoped-for, commercially successful innovations. It is very important to be aware of market needs at an early stage, and other ways of accessing technologies may be more financially rewarding.[4,6] For instance, reverse engineering has been widely used in some places, and can be less expensive.

Innovation, industrial property, and competitive intelligence systems are almost inseparable. Successful corporate development may depend on the importance placed on the links between these activities.

> "To invent, you must know what your industrial rivals are making. You must also search, find, sell. It is necessary to follow an active industrial property policy to be efficient from research to production and market."[7]

Patents as a management tool

> "Patent indicators provide a very useful forecasting tool for decision makers in the public and private sectors. These tools are usable for R&D planning, for competition analyses, and for analytical studies of how technologies emerge, mature and pass away."[8]

Patenting provides legal protection. It is also a 'technological memory' that allows the company to measure and control, to some degree, its innovative capability. Patents are significant and useful tools in competitive intelligence because they are good indicators of technological activities,[9] with advantages in their thoroughness, simple and quick access, world-wide diffusion, reliability, early warning, precision, quality, timeliness, interpretation, and classification.[10]

Patents reflect strategic moves and weapons, and allow different tactical actions:[11,12] offensive, diversive, dissuasive, defensive, and counter-offensive.

Patents of other companies may be the starting point of new ideas and developments. The Patent Competitive Intelligence System is thus an active survey of the current competitive situation, together with a rational and intelligent use of the information gained. It is part of the global competitive intelligence system.

Using patent information to analyse company research strategies assumes there is a significant correlation between research policies and intellectual property strategies. Many authors have tried to analyse this correlation.[13,14,15,16]

One of the common ways to obtain current patent information is from international patent databases, which are numerous. The database used in this article is the World Patent Information database (WPI), produced by Derwent.

Patent bibliometry

Bibliometric analysis of patent databases has been widely reported.[17,18] Narin's work on the statistical analysis of patent citations is widely known,[19,20] but this approach is only possible with databases including citations. Such analyses are perhaps more oriented to the sociology of science and technology citations than to analysis of the subjects of the patents.

There are some important preliminary points to make about statistical analysis of patents:

- Patents databases are built for documentation purposes, so some of the data, when analysed on a statistical or strategic basis, make little sense, or lead to wrong interpretation. Before conducting a bibliometric analysis, conversion to a bibliometric format is necessary. This format has a physical computerised form, but is also of strategic and statistical interest.[21]
- Almost all the downloaded database fields of a patent may be used in a bibliometric method (even if a conversion is first needed), as well as any arrangement of those fields. This provides a large number of possible analyses.
- All these analyses may be performed with a large number of statistical methods, increasing the range of options and the workload.
- All bibliometric analyses are capable of producing information of some interest when they are performed for a sufficiently large number of documents.

To produce relevant bibliornetric analyses of patent databases there must be co-operation among a decision maker, an expert in the field and a data expert, and the calculations must be performed using a computer

This list leads to the operational need to focus on what constitutes a relevant analysis:

- Bibliometric analysis must be performed with co-operation among several specialists. Using both a decision-maker and an expert in the study area may help to reduce the number of analyses with strategic considerations. An information expert is needed to reduce the number of analyses necessary to produce a useful result.
- Bibliometric analysis needs a computer. It is unrealistic to think about doing a statistical analysis of a large number of documents with a large number of statistical methods without automating the process. Unfortunately, despite the amount of software available for database and statistical analysis, there is a shortage of software to convert databases data into a form suitable for numerical analysis.

Collaborating with several companies to improve any theoretical scientific considerations by introducing real data, forces the search for technical solutions to automated procedures. Otherwise, bibliometric and statistical analysis remain theories that are unable to help in any kind of R&D development.

This is why our laboratory has been developing bibliometric software for 15 years. That software fills the gap in commercial software, and allows applied bibliometric research in companies.

Example of bibliometric analysis

Data collection and study goal

Publishing bibliometric analyses of patent documents for R&D management is very difficult because of confidentiality problems with companies. This is why we are presenting a case study that was not applied in a company. We choose to apply our methodology over a set of patents dealing with 'therapeutic transdermic systems' — more commonly known as patches. The main reason for this choice was that we could enlist the help of an expert in this field.

Analyses of patent keywords have been obtained[1] and give, in our opinion, a better description of patents than citation analysis. The main problem in patent descriptor analysis is the choice of appropriate descriptors. Anyone could choose 'home-made' descriptors from the abstract or the full patent text, but automated methods of text analysis are not easy to use.[23] The WPI database contains a normalised title which is a lexically reduced title, but is not necessarily relevant for indexing the patent meaningfully.

Patents in bibliographic databases are indexed with several documentary classifications. These are very interesting because they are not attributed by the same office, nor with the same goal, but are created by reading different parts of the patent. There may also be a great deal of information in a classification subset, much more than in the individual words of the normalised title.

The WPI database has more descriptors than other bibliographic databases (three documentary classifications, title (normalised or not), abstract, use, and advantages). Statistical analysis of all these descriptors have been performed individually.[24,25] In companies, people do not always know which is the relevant classification, and whether or not the classifications are equivalent. The purpose here will be to automatically and statistically detect com-

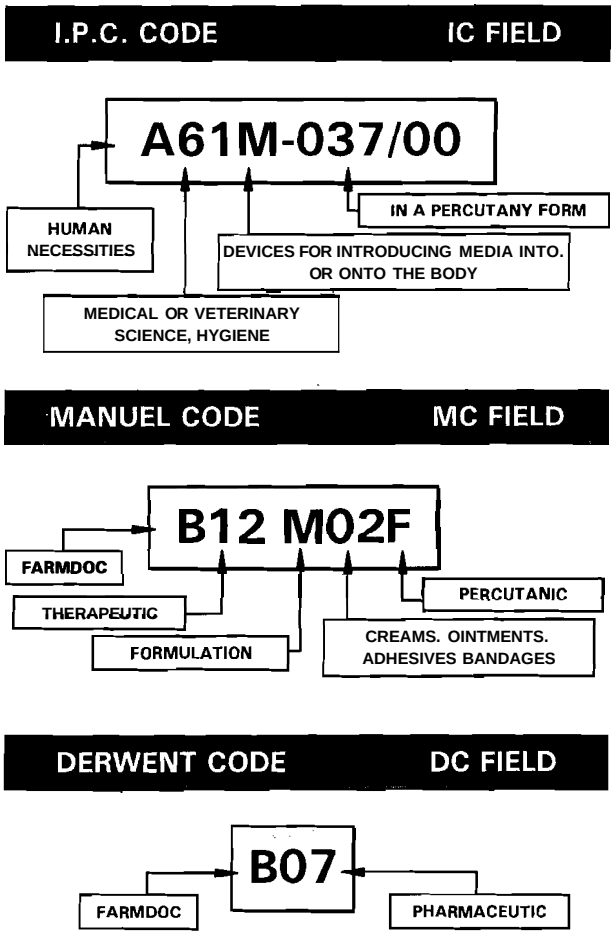

Figure 1. Hierarchical structure of the studied classifications

Hierarchical codes levels choice

Computed by CEMAP

Hierarchical code digit number	Codes number	Codes number after frequency one codes deletion	number	Patents number after frequency one patents deletion	Middle number of codes patents	Classes number	Percent of classes with one element
DC 4	52	32	146	146	3.6	27	30
IC 4	40	21	146	143	2.2	23	22
IC 7	84	35	146	140	2.8	42	36
IC 11	133	36	146	113	2	67	52
MC 3	81	42	143	143	4.6	41	24
MC 5	139	90	143	143	6.9	80	50
MC 7	315	158	142	141	8.7	106	74
MC 8	144	69	135	134	4.3	80	55
MC 9	43	12	46	34	1.7	124	90

Figure 2. Hierarchical codes study inducing the future truncations

plementarities and similarities among three classifications having different structures, significance and use. The three classifications used are:

- International Patent Classification (CIB or IP Class or IC), which is an international hierarchical classification of patents attributed by patent offices using the entire patent text. It is probably the most powerful documentary classification that has been in existence for a long time.
- Manual codes (MC), which have a significant meaning in the application, are attributed by the database producer using the entire patent text.
- Derwent codes (DC) are attributed by reading the abstract.

Figure 1 shows the hierarchical structure of these codes.

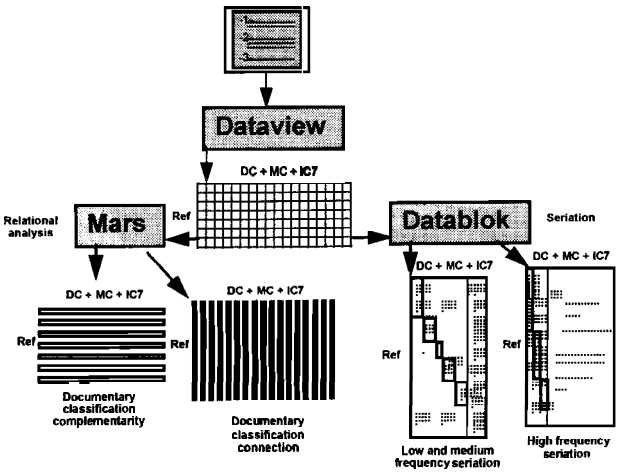

Figure 3. Matrix construction and analysis

Hierarchical heterogeneity problem

In order to perform an analysis automatically, we need first to reduce the heterogeneity problem within the hierarchical structure in our downloaded database. We can find in the same data either a generic or a specific description of a patent. If complementarity and similarity between codes are to be analysed, a first homogeneity comparison must be performed to consider equivalent levels in documentary classifications. This allows truncation choices for matrix building, and simplified later interpretation. This study is summarised in Figure 2.

Next we built automatically a binary matrix (presence/absence of truncated codes in patents), with DATAVIEW software[26] (Figure 3).

Statistical analysis

Several statistical clustering techniques have been used on this matrix, as shown in Figure 3. All these classifications are non-hierarchical, and automatically determine the number of clusters. The first two were performed in the CEMAP IBM Centre in Paris and were presented at a conference.[27] These classification techniques have been reported in a doctorate thesis[28] and a paper.[29] They cluster either the codes (columns) or references (rows) with the 'relational analysis', thus showing the connection or complementarity of the documentary classifications.

The goal in our laboratory was to adapt the CEMAP technique to provide an efficient tool (DATABLOC) capable of being used in microcomputers, for small and middle-sized companies. Our technique allows a simultaneous code and patent classification ('block seriation') which re-

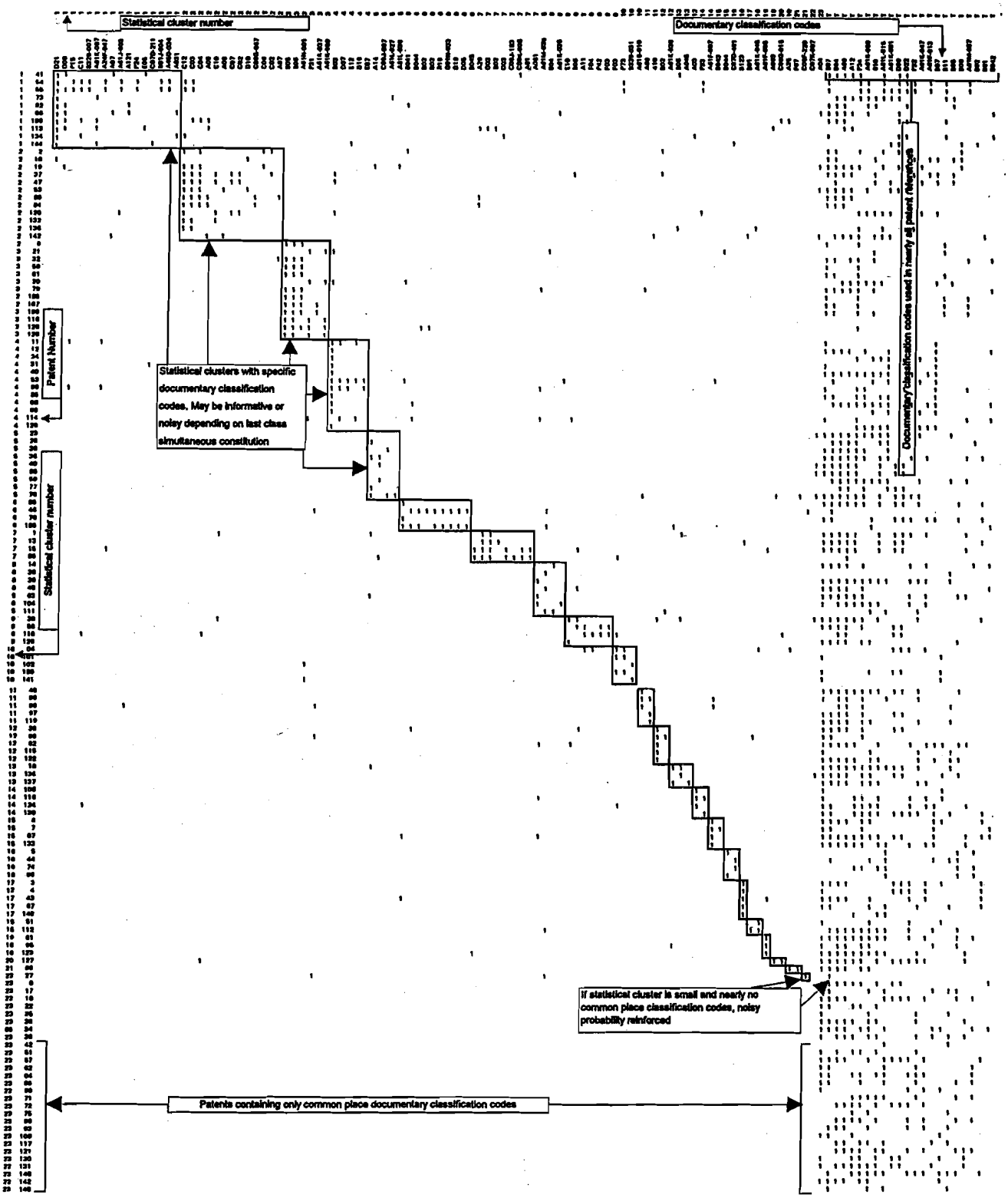

Figure 4. Medium and low Zipf's law frequency code classification

Note: **Cluster number is automatically determined: width and class density are specified by the user**

spects Zipfs law. Several papers[30,31] provide an explanation of the derivation and application of this law. It is taken into account here not only with highly used codes, but also with many low frequency codes which can be used for innovation detection or special-application patents.

Our algorithm determines for itself the high frequency limit, not only in frequency considerations but also for relational purposes. Then we perform two separate classifications: one with only high frequencies, and one with medium and low frequencies. Both frequency level and class number are automatically determined. The algorithm determines the best mathematical arrangements of

the initial matrix, placing the highest density in the diagonal, and underlines groups.

Because statistical techniques cannot replace an expert in the area of study, we decided to provide the results in two forms: a rearranged binary matrix, and the initial references in an hypertext format with links built according to the clustering technique, which here is just a 'reading guide'. Figure 4 shows low frequency clustering.

Area-expert conclusions

The simultaneous row and column classification helps in viewing the group structure within those two spaces. Those clustering techniques help the expert in detecting general patents (only abstract with high-frequency codes), innovative ones (small frequency code groups with high frequency codes), and some 'noise-level' patents (small frequency code groups without high frequency codes) within the entire body of data.

A particular part of the patches technology was only apparent by combining both of the documentary classifications.

The clustering technique helps in building up a reading plan that is useful in finalising a report for the decision maker's analysis, but it could not do away with patent reading.

Conclusion

This paper is the result of the collaborative work of three persons: an expert in on-line computing and patents, a statistical expert, and a research expert in-the application area. We tried to illustrate our cause with a complete example to show that bibliometric studies of patents may be a relevant way of helping many kinds of research activity.

The most important conclusion is that this type of analysis must be performed with strategic considerations. Then, patent bibliometric analyses are an efficient way of helping R&D management.

Other points to note are that it is necessary to integrate these analyses with the company's competitive intelligence — both are weakened without the other. More important than sophisticated analysis are the relations with strategy, the experts, and the communication tools to and from decision makers. Our experience shows that this kind of analysis is useful in managing R&D, but that it is also very difficult to know exactly how companies would use it, because of confidentiality.

References

1. J Broustail and F Frery, Le management *stratégique* de *l'in*novation (Précis Dalloz, 1992).
2. S Saleh and C Wang, "Management of innovation. Strategy, structure and organizational climate", IEEE Transaction on Engineering Management, 40(1), 1993, pages 16-21.
3. F Jakobiak, Pratique de la Veille Technologique (Les Editions d'organisation), 1991).
4. B Martinet, J M Ribault and D Lebidois, Le management de technologies (Les Editions d'organisation, 1991).
5. P Maitre and D Miquel, De *l'idée* auproduit (Eyrolles, 1992).
6. J Perrin, Les transferts de *technologie* (Editions La Découverte, 1984).
7. F Jakobiak, Exemples *commentés* de Veille Technologique (Les Editions d'organisation, 1992).
8. R S Campbell, "Patent trends as a technological forecasting tool", World Patent Information, 5(3), 1983, pages 137-143.
9. P Patel and K Pavitt, "The importance of the technological activities of the world's largest firms", World Patent *Information*, 12(2), 1990, pages 89-94.
10. G Maire, "La veille technologique: les brevets des autres source d'information et d'innovation", Travail et *Méthodes*, 475, 1988, pages 25-29.
11. C Brenard, "La stratbgie du brevet", Sciences et *Vie économie*, September 1988, pages 50-60.
12. F Marquer, Innovation et Management des brevets (Les Editions d'organisation), 1985).
13. P Wiseman, "Patenting and intensive activity on synthetic fibre intermediates", Research Policy, 12(6), 1211983, pages 329-339.
14. Cooray, "Knowledge accumulation and technical advance", Research Policy, 14(2), 1985, pages 83-95.
15. K Pavitt, "Sectorial patterns of technical change: towards a taxonomy and a theory", Research Policy, 13, 1984, pages 323-373.
16. J P Courtial, Introduction à la Scientometrie (Editions Anthropos, 1990).
17. Z Griliches, "Patent statistics as economic indicators", Journal of Economic Literature, 28, 1990, pages 1661-1707.
18. H Grupp, U Schmoch and U Kuntze, "Patents as potential indicators of the utility of European Communities programs", Scientometrics, 21(3), 1991, pages 417-445.
19. F Narin, E Noma and R Perry, "Patents as indicators of corporate technological strength", Research Policy, 16(2-4), 1987, pages 143-155.
20. F Narin and D Olivaro, "Technology indicators based on patents and patent citations", in A F J Van Raan (editor), Handbook of Quantitative Studies in Science and Technology (Elsevier, 1988) pages 465-507.
21. W Nivol, "Systemes de surveillance systematique pour le management stratbgique de l'entreprise", doctorate thesis, UIII, Marseille, France, 10 May 1993.
22. R R Braam, H F Moed and A F J Van Raan, "Mapping of science by combined co-citation and word analysis", parts I and II, Journal of the American Society for Information Science, 42(4), 1991, pages 233-266.
23. W Lehnert and B Sundheim, "Performance evaluation of text-analysis technologies", *AI* Magazine, 12(3), 1991, pages 81-94.
24. W G Vijvers, "The International Patent Classification as a search tool", World Patent Information, 12(1), 1990, pages 26-30.
25. H Dou, P Hassanaly and L Quoniam, "Easy mapping classifications classifications of patent references with microcomputers", The Montreux International Chemical conference proceedings, 1989, pages 283-310.
26. H Rostaing, "Veille Technologique et Bibliométrie: Concepts, Outils, Applications", doctorate thesis, UIII, Marseille, France, 13 January 1993.
27. H Rostaing, W Nivol, L Quoniam, C Bedecarrax and C Huot, "L'exploitation systématique des bases de donnbes, Journée d'btude ADEST, June 1992.
28. C Huot, "Analyse Relationnelle pour la Veille Technologique", doctorate thesis, UIII, Marseille, France, 13 December 1992.
29. C Huot, L Quoniam and H Dou, "A new method for analysing downloaded data for strategic decision", *Scientometrics*, 25(2), 1992, pages 279-294.
30. B C Brookes, "The derivation and the application of the Bradford-Zipf distribution", Journal of Documentation, 24(4), 1968, pages 247-267.
31. L Egghe, "The exact place of Zipf's and Pareto's law amongst the classical informatrics laws", Scientometrics, 20(1), 1991, pages 93-106.