

## INFOGRAPHIC ANALYTICAL TOOLS FOR DECISION MAKERS

### ANALYSIS OF THE RESEARCH PRODUCTION IN SCIENCES. APPLICATION TO CHEMISTRY, COMPARISON BETWEEN MARSEILLE AND MONTPELLIER (FRANCE)

H. DOU, P. HASSANALY, L. QUONIAM

*Centre de Recherche Retrospective de Marseille UA126 CNRS,  
Universite Aix-Marseille III - 13397 Marseille Cedex 13 (France)*

(Received September 14, 1988)

Most of the scientific and technical databases contain codes. These codes divide the area of the database field in subfields. These divisions can be used to map automatically the research network of a subject, and to provide its main research poles. The present paper explains the methodology, and applies it to *Chemical Abstracts*, and to the analysis of the research production of various academic institutions. The method is general and can be used with other databases such as Inspec, WPI-WPIL, etc. . .

#### Introduction

During the process of decision in sciences, the production of a scientific community, of an institution, is often analyzed in various ways. But, when the volume of this production is too important (e.g. several hundreds of papers) it is very difficult to obtain a fast and realistic view of all the results.

We developed in our laboratory various softwares to analyze the scientific literature from downloaded references coming from different databases. An infographic analysis dealing with the chemical production of two important university communities in France will be presented in this paper.

#### Methodology

When an expert group, or a peer review, is concerned with several thousands of items (e.g. the scientific production of a large group, or of several institutions located in the same town, . . .) their feelings follow very often an irrational point of view, where rumors, self-knowledge, polical interest, lobbies, play a large part.<sup>1</sup>

This attitude is very general, because it is almost impossible to have a synthetic view of all the subjects, and to be able, in a minimum amount of time to make a global, reliable, and unquestionable comparisons.

This is the reason we prompt us to develop a method which will fill this gap. Interested in chemistry and physics, we choose the *Chemical Abstracts* database as a source of informations for the purpose of this paper. This database indexes all the scientific publications, patents, reports, . . . and references are available online, they contain several fields such as authors, titles, indexterms, codes, etc.

The following example shows two references dealing with the oxidation of coal, obtained from the host ORBIT INFORMATION TECHNOLOGIES.<sup>2</sup>

AN – CA05-87508(10)

TI – Electrochemical synthesis of humic acid and other partially oxidized carbonaceous materials

AU – Vaughan, Ronald J.

SO – U.S. 4592814, 12 pp. Cont.-in-part of U.S. Ser. No. 496, 799, abandoned., 86.06.03 PAT APP = 653664, 84.09.21, PAT PRY = 496799 (IN US), 83.05.23, PCL 204/78 PCL (IPC) C25B-003/02A

OS – Chevron Research Co., USA

DT – P (Patent)

CC – SEC72-7; SEC5; SEC19; SEC49

AN – CA05-85796(10)

TI – Mechanism of potassium-catalyzed carbon/carbon dioxide reaction

AU – Sams, D. A.; Shadman, Farhang

SO – AICHE J. (AICEAC), V 32 (7), p. 1132–7, 1986, ISSN 00011541

OS – Univ. Arizona, Dep. Chem. Eng., Tucson (AZ), USA, 85721

DT – J (Journal)

CC – SEC67-2

In this references, the CC field (category code field), is very interesting, since in *Chemical Abstracts* the papers indexed are allocated in sections. There are 80 sections to cover all the chemistry. For the period of time analyzed the division remains constant.<sup>3,4</sup> This field may contain one section or several sections. The first, primary section deals with the main part of the work, and the others, secondary or cross reference sections deal with side applications of the subject.

These fields (e.g. the CC) may be automatically analyzed, and will lead to:

- the main poles of research, using primary sections,
- the network of these poles with the other part of chemistry, using the secondary sections.

The list of the 80 *Chemical Abstracts* sections are indicated in the Appendix.

The presentations of the results are made in 3D for the main research poles, and in 2D for the research network of these poles. Because the network can be rather complicated in some cases, a zoom function allows to isolate the network of one pole if necessary.<sup>5,6</sup>

### Results

The references have been selected using the query: MARSEILLE/CI with limitations by date, and the same for MONTPELLIER. The production of these two Centres are indicated in Table 1.

Table 1  
Comparison of the scientific productions of Marseille  
and Montpellier between 1982 and 1986

Years	Amount of references	
	Marseille	Montpellier
1982	498	489
1983	566	495
1984	556	536
1985	551	563
1986	535	599

#### *The research poles:*

For the clarity of the presentation, we will present only the comparison between year 1985 for Marseille and Montpellier. To illustrate the differences occurring with time, the year 1978 will be also used for Marseille.

The Figures 1 and 2 represent the research poles of Marseille, (years 1978 and 1985) and the Figs 3 and 4 the research poles for Montpellier using two thresholds.

In the graphs, the sections are presented always at the same place, on a chess board where section 1 is at the upper left, and section 80 at the lower right. The frequency of the items (e.g. references with the same primary section) is the number of publications appearing in the time interval. This is the height of the cylinders.

It is possible to use a frequency filter, and to consider various thresholds. Figures 3 and 4 show for the town of Montpellier the same analysis with a threshold above 0.5% and 2.5% of the overall number of references analyzed. This threshold can be

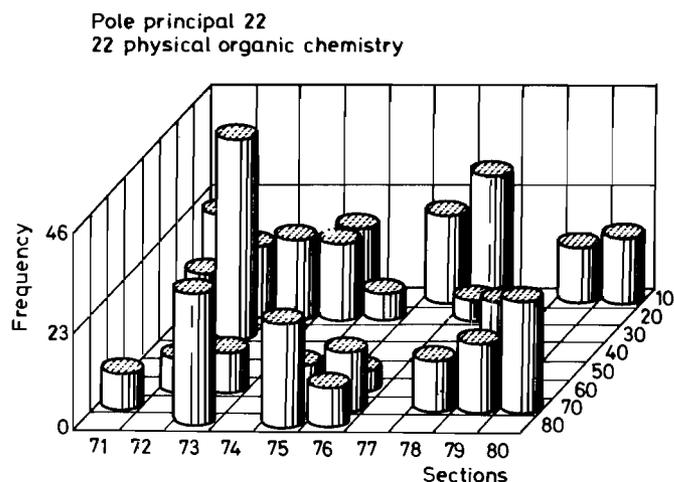


Fig. 1. Research poles at Marseille, year 1978, using a threshold of 5

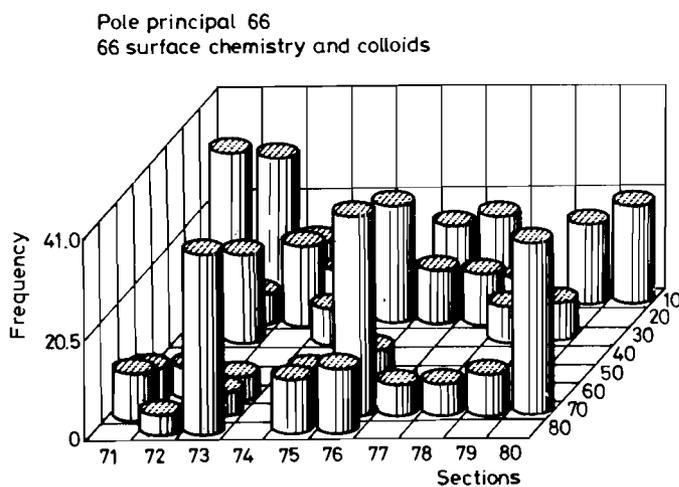


Fig. 2. Research poles at Marseille, year 1985, using a threshold of 5

determined in various ways, for instance in per cent of the main research pole, etc . . . but in our opinion, working in per cent of the total production seems to be the best. This give a good indication of the scattering of the subjects, and of the perception of these later by outside observers.

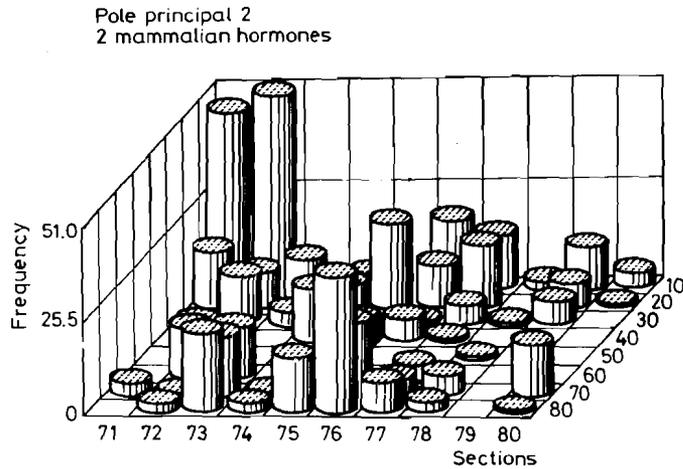


Fig. 3. Research poles at Montpellier, year 1985, using a threshold of 2

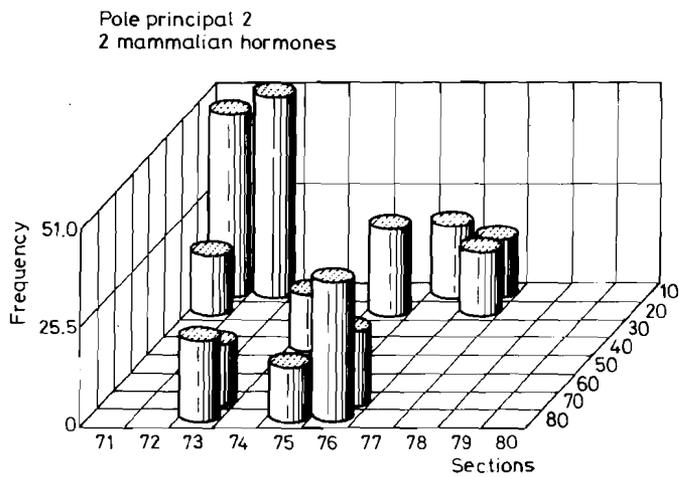


Fig. 4. Research poles in Montpellier, year 1985, using a threshold of 15

*The research network:*

The network, when all the sections are considered, is too complicated to give valuable informations. It can only be used to spot the main knots of the system and to have an overview of the multidisciplinary of the researchs. The use of the zoom function is more accurate. Once a significant section is identified by its number,

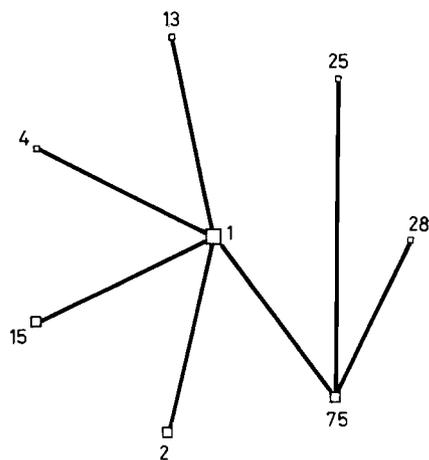


Fig. 5. Zoom of pole 1 (Marseille, 1982, pharmacology)

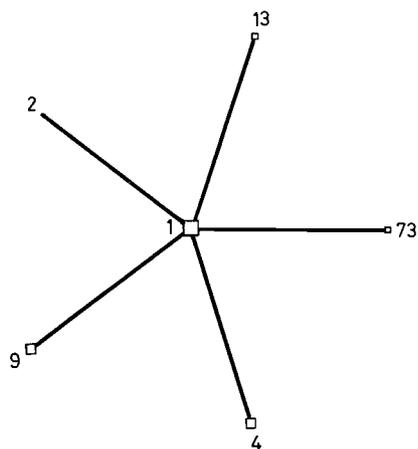


Fig. 6. Zoom of pole 1 (Marseille 1985, pharmacology)

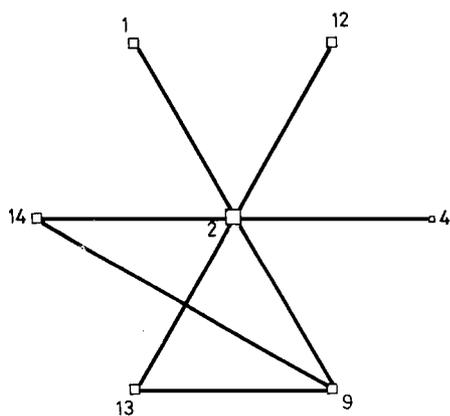


Fig. 7. Zoom of pole (Marseille 1985, Mammalian hormones)

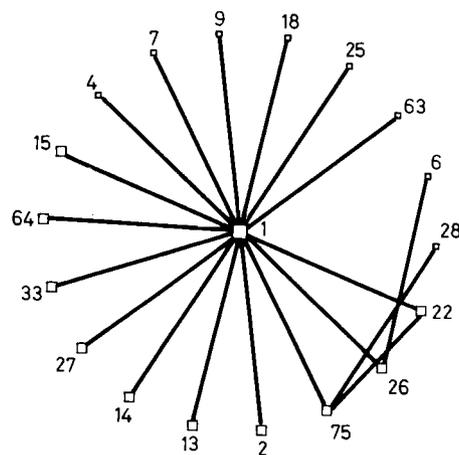


Fig. 8. Zoom of pole 1 (Montpellier 1985, pharmacology)

either from the preceding graph or as a main research pole, it is possible to see the network in which this pole is inserted. We analyzed automatically all the references which have in the CC field two or more sections, and the program generate all the bounds between sections. In the zoom function, when a pole is considered, only references with this pole present as primary section are analyzed.

Results are presented in Figs 5 to 9 where Figs 5 to 7 represent the zoom of poles 1 (pharmacology) and 2 (mammalian hormones) for Marseille for years 1982, 1985, and Figs 8 and 9 the zooms of poles 1 (pharmacology) and 2 (mammalian hormones) for Montpellier, for year 1985.

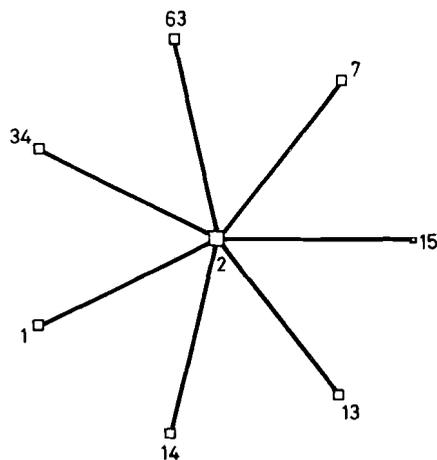


Fig. 9. Zoom of pole 2 (Montpellier 1985, mammalian hormones)

In the software there is also a frequency filter which allows to select the research poles above a certain threshold. In this paper, the threshold of 2 has been used, this means that only primary sections (with one or more secondary sections present) appearing at least twice are considered to build the network.

### Discussion

The comparisons between the research poles is a mean to follow the trends in research in Marseille and Montpellier. If the changes are small in short time intervals (1, 3 years), they can be very drastic for a longer period. The examples of year 1978 and 1985 for Marseille are very significant. The analysis suggests the following questions:

- Why a very strong pole such as Physical Organic Chemistry, disappeared between 1978 and 1985? is this due to the lack of equipment, labour force, incentive, . . . ?
  - Why poles 1 and 2 have had during the same period such a fast growth?
  - What will be the consequences of this growth on the other parts of chemistry?
- (When the networks are concerned, we can see that the number of bounds for pole

1 and 2 in Marseille are lower than for Montpellier. This can be explained because a strong tradition in these 2 disciplines (pharmacology and mammalian hormones) exists in Montpellier.

The fundamental question is to know if and when pole 1 and 2 in Marseille will have a better integration in the scientific network of chemistry. Analysis of years 1986 and 1987 begins to provide an answer, since the density of the network increases.

Analysis of all the research poles, as well as comparisons can be made if necessary. The goal of this paper is not to present all the possible applications of this method, but to reinforce the role of softwares which will help decision makers, committees, councils, to analyze in term of network, expertise, . . . the output of databases providing meaningful codes. In this frame, various analysis have been performed in our laboratory using databases such as WPI-WPIL (Derwent codes, Patents), Biosis (Biocodes), Predicast (Economy), Inspec (Physic).

### Appendix

#### *List of Chemical Abstracts sections*

1. pharmacology
- 2 mammalian hormones
- 3 biochemical genetics
- 4 toxicology
- 5 agrochemical bioregulators
- 6 general biochemistry
- 7 enzymes
- 8 radiation biochemistry
- 9 biochemical methods
- 10 microbial biochemistry
- 11 plant biochemistry
- 12 nonmammalian biochemistry
- 13 mammalian biochemistry
- 14 mammalian pathological biochemistry
- 15 immunochemistry
- 16 fermentation and bioindustrial chemistry
- 17 food and feed chemistry
- 18 animal nutrition
- 19 fertilizers, soils, and plant nutrition

- 20 history, education, and documentation
- 21 general organic chemistry
- 22 physical organic chemistry
- 23 aliphatic compounds
- 24 alicyclic compounds
- 25 benzene, its derivatives, and condensed benzoid compounds
- 26 biomolecules and their synthetic analogs
- 27 heterocyclic compounds (one hetero atom)
- 28 heterocyclic compounds (more than one hetero atom)
- 29 organometallic and organometalloidal compounds
- 30 terpenes and terpenoids
- 31 alkaloids
- 32 steroids
- 33 carbohydrates
- 34 amino acids, peptides, and proteins
- 35 chemistry of synthetic high polymers
- 36 physical properties of synthetic high polymers
- 37 plastics manufacture and processing
- 38 plastics fabrications and uses
- 39 synthetic elastomers and natural rubber
- 40 textiles
- 41 dyes, organic pigments, fluorescent brighteners, and photographic sensitizers
- 42 coating, inks, and related products
- 43 cellulose, lignin, paper, and other wood products
- 44 industrial carbohydrates
- 45 industrial organic chemicals, leather, fats, and waxes
- 46 surface-active agents and detergents
- 47 apparatus and plant equipment
- 48 unit operations and processes
- 49 industrial inorganic chemicals
- 50 propellants and explosives
- 51 fossil fuels, derivatives, and related products
- 52 electrochemical, radiational, and thermal energy technology
- 53 mineralogical and geological chemistry
- 54 extractive metallurgy
- 55 ferrous metals and alloys
- 56 nonferrous metals and alloys
- 57 ceramics
- 58 cement, concrete, and related building materials

- 59 air pollution and industrial hygiene
- 60 waste treatment and disposal
- 61 water
- 62 essential oils and cosmetics
- 63 pharmaceuticals
- 64 pharmaceutical analysis
- 65 general physical chemistry
- 66 surface chemistry and colloids
- 67 catalysis, reaction kinetics, and inorganic reaction mechanisms
- 68 phase equilibria, chemical equilibria and solutions
- 69 thermodynamics, thermochemistry, and thermal properties
- 70 nuclear phenomena
- 71 nuclear technology
- 72 electrochemistry
- 73 optical, electron, and mass spectroscopy and other related properties
- 74 radiation chemistry, photochemistry, and photographic and other reprographic processes
- 75 crystallography and liquid crystals
- 76 electric phenomena
- 77 magnetic phenomena
- 78 inorganic chemicals and reactions
- 79 inorganic analytical chemistry
- 80 organic analytical chemistry

#### References

1. British-French seminar: The indicators for science and technology policy. December 1986, London.
2. Orbit Information Technologies, Achilles House, Park Royal, Western av 30UA, London, England. We like to thank very much the Orbit Information Technologies for its help in providing our laboratory with online chemical informations datas.
3. *Subject Coverage and Arrangement of Abstracts by Sections in Chemical Abstracts*, Editor American Chemical Society, 1982.
4. K. SHINICHIRO, Changes in sections of Chemical Abstracts, *Shikoku Kokenkai Ho*, 33 (1982) 2-6.
5. H. DOU, P. HASSANALY, . MESSIANE, *Le traitemets des codes par micro-ordinateurs*, Symposium on "Les systemes d'information elaboree", Ile Rousse, September 1987, France.
6. H. DOU, P. HASSANALY, Mapping the Scientific network of patent and non-patent documents from chemical abstracts for a fast scientometric analysis, *World Patent Information*, 2 (1988, Juin).