Bibliometric analysis of patent documents for R & D management

***Quoniam L., **Huot C., *Dou H., *Hassanaly P., *Baldit P.**

* CRRM Fac St Jérôme
13397 Marseille CEDEX 20
France

** CEMAP IBM
68-72 Rives de Seine. Paris
75012 France

Innovation as a necessity

Research and Development (R&D) policy is directly linked to the market. With a good R&D policy, innovation in company can be developed. R&D can be the engine of innovation. Due to internationalisation, companies must develop their competitiveness. Several books and articles relate innovation management necessity [Brou92] [Sale93]. To do this, one of the main tools is innovation. Innovation is considered here in a large acceptation and includes:

-innovation as new products and services sold by a company.
-innovation in terms of process for these products and services.

This innovation which follows research and precedes production must be protected with patents [Jako91]. Numerous are the inventive companies which did not protect their inventions and are spoiled by others. Industrial property is then a tool that goes into R&D services. Industrial property policy helps companies just after invention to protect it. Industrial property is an innovation consequence and gives to companies an invention exploitation monopoly. Due to this fact it gives an answer to the market need to which the R&D policy is linked. Figure 1 shows the links between R&D and Industrial Property.

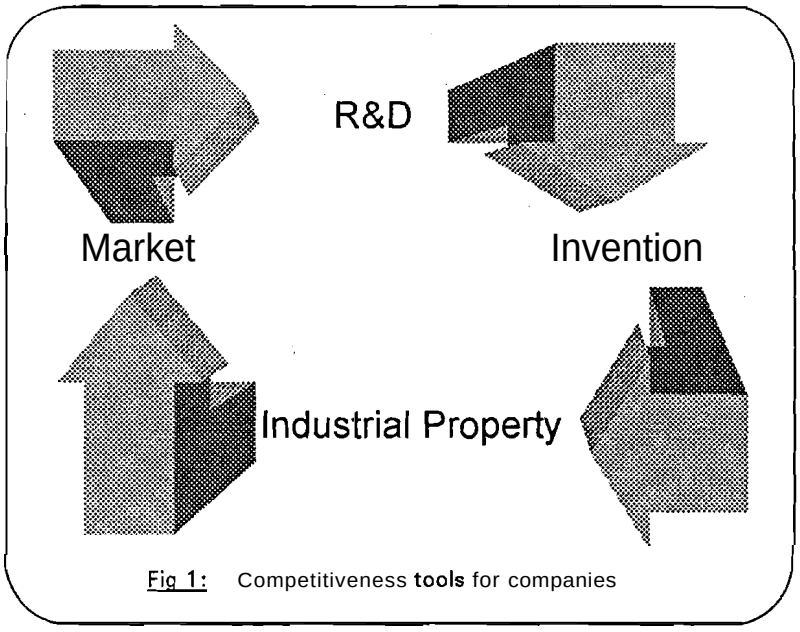

Fig 1: Competitiveness tools for companies

Any-one can see that R&D policy, Invention activity and Industrial property are bound. An active and well adapted industrial property policy can have an effect on invention activity.

Competitive Intelligence

In France, this activity is nowadays well developed. In part, it consists in scanning the external industrial environment. This activity can be divided in various parts. But it is rather easy to understand that a specific Patent Competitive Intelligence System can be efficient to conquer new markets or expand in foreign countries. This kind of activity must be organised, structured and most of all appropriated.

One of the main goals of competitive intelligence system should be to feed the invention process with accessable goal information to:

-apply new technologies for actual product processes and selling
-create new products using improved technologies
-create new products for new markets.

Martinet B. [Mart91] says that 70% of inventions arise from a market need. This leaves 30% that arise from scientific programs. In fact many inventions have a scientific base (more or less old) but most of all, their commercial importance is of a slow adaptation arising to the Science-Market synergy. Patents are a good picture of this synergy. Science allows patenting products and processes, and the market forces the inventors to protect their interests. Big scientific programs can build rupture inventions and market or "science-market" relation induces more adaptive innovations. This represents the "idea to product" chain [Mait92].

Watching scientific activity is not enough. It is very important to watch on the early market needs. It is not always useful to get big scientific programs to increase technologic progress. It may be more financially interesting to use other ways to reach technology [Mart91] [Perr84]. For instance the Reverse Engineering has been widely used and is less expensive even if information has a non negligible price.

Links between innovation, industrial property and Competitive Intelligence systems make those activities almost inseparable. Companies development depends on the strain and the links it decides to give to those activities.

"To invent you must know what your industrial rivals are making. You must also search, find, sell. It is necessary to get an active industrial property policy to be efficient from research to production and market" [Jako92].

Patents as a management tool

Managing an industrial and commercial development means also managing the technologic patrimony. Patenting is one of the best way to protect and safeguard the technological and scientific company's patrimony. First of all this protection is of a juridical form. It is also a "technological memory" that lets you measure and control

the internal and external innovative capability. Patents are a major and ideal tool in competitive intelligence.

Patents information has several specific advantages [Mair88]: Exhaustive, simple and quick access, world-wide diffusion, reliability, precocity, precision, quality, anteriority, translation, classification.

Patents are also strategic weapons and allow different tactical actions [Bern88] [Marq85]: offensive, diversion, dissuasive, defensive and conter-offensive strategies.

Patents of other companies may even constitute the starting point of new ideas and developments.

Patent Competitive Intelligence System is then an active concurrence survey followed by a rational an clever use of the informations issued from this system. It is a part of the global Competitive Intelligence System but needs and uses the same information's exploitation techniques.

Using patent information to analyse the companies search strategy assumes there is a significant correlation between search policy and industrial property strategy. Many authors tried to analyse this correlation [Wise83] [Coor85] [Pavi84] [Cour90]. Patents can be considered as a scientific and technologic indicators of companies activities. One of the common ways to know about concurrent patents is international patent databases. There are many of them and the patent database used in this article is World Patent Information (WPI) produced by Derwent.

Patent bibliometry

Bibliometric analysis of downloaded databases has been widely reported. There are still some important remarks over this kind of analysis of patents:

-Patents databases are built for documentation use but some of the informations when analysed on a statistical or strategic form have no sense or introduce a wrong interpretation. Before a bibliometric analysis, first a conversion to a bibliometric format is necessary. This format has a physical computerised form but also a strategic and statistical interest [Nivo93].
-Almost all the downloaded database fields of a patent may be analysed with a bibliometric method (even if a conversion is first needed). As well as any arrangement of those fields. This provides a large number of bibliometric possible analysis.
-All of the previous analysis may be performed with a large number of statistical methods, increasing the amount of possible analysis.
-All bibliometric analysis are relevant when performed over a large number of documents.

All these remarks introduce other remarks to produce relevant analysis:

-Bibliometric analysis must be perform with a co-operation between several specialists: a decision maker and a studied area expert to help to reduce the

number of analysis with strategic considerations. But also an information expert to help in reducing analysis number, considering analysis validity.

-Bibliometric analysis must be performed on a computerised form. It is impossible to think about statistical analysis of a large number of documents with a large number of statistical methods without a process automation. Unfortunately if softwares are numerous in database and statistical areas, it seems that softwares able to convert databases for numerical applications are not as numerous.

Bibliometric software used

Our laboratory has been developing bibliometric softwares for 15 years and the one used is DATAVIEW developed during H. Rostaing's thesis [Rost93]. Dataview's principles are explained in figure 2 and 3. This software is able to convert any database field (or fields' arrangements) from any database server in an editing format in a relevant number of numeric matrices in statistical commercial softwares formats.

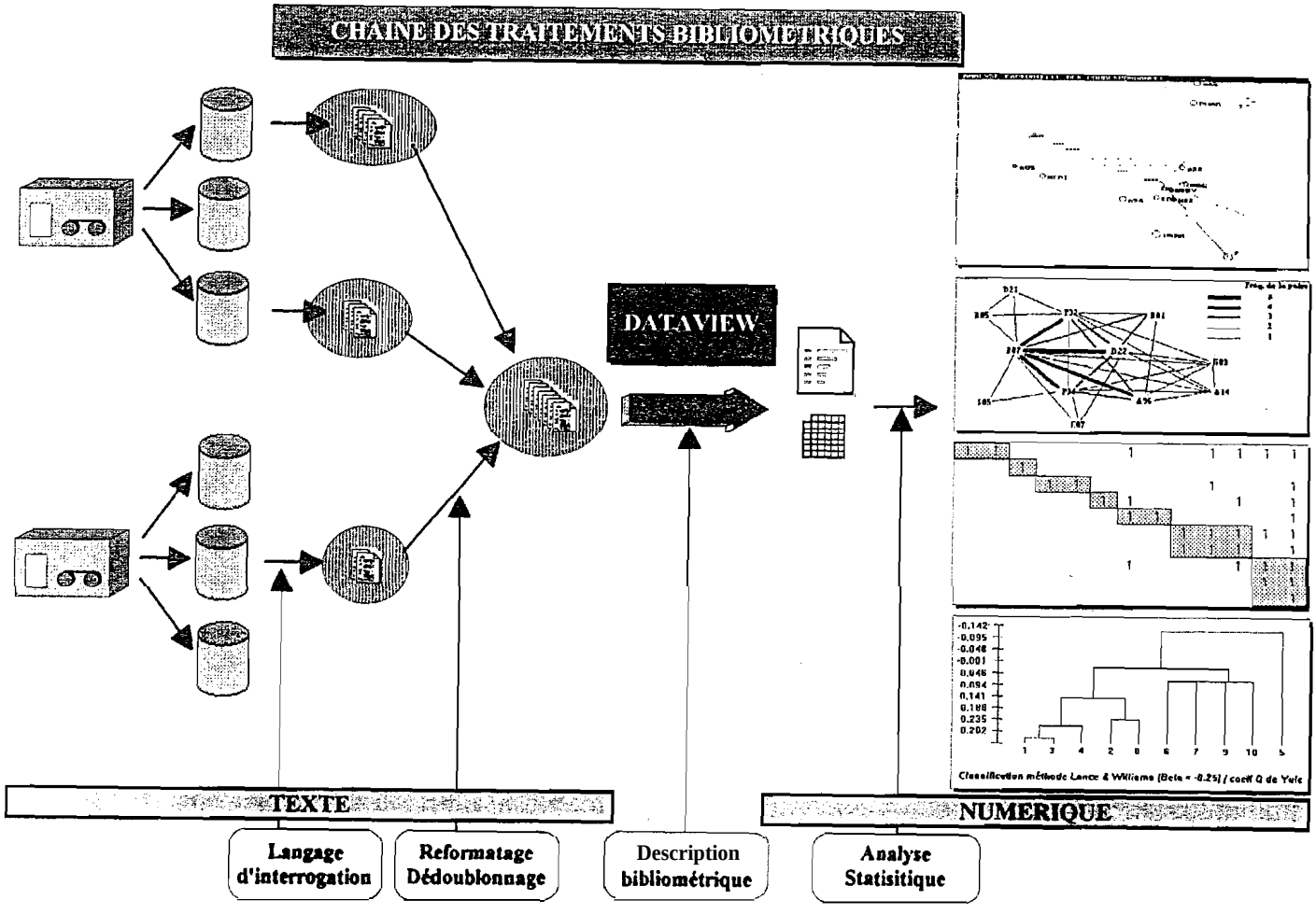

Fig 2: Dataview's position between textual and numerical commercial softwares.

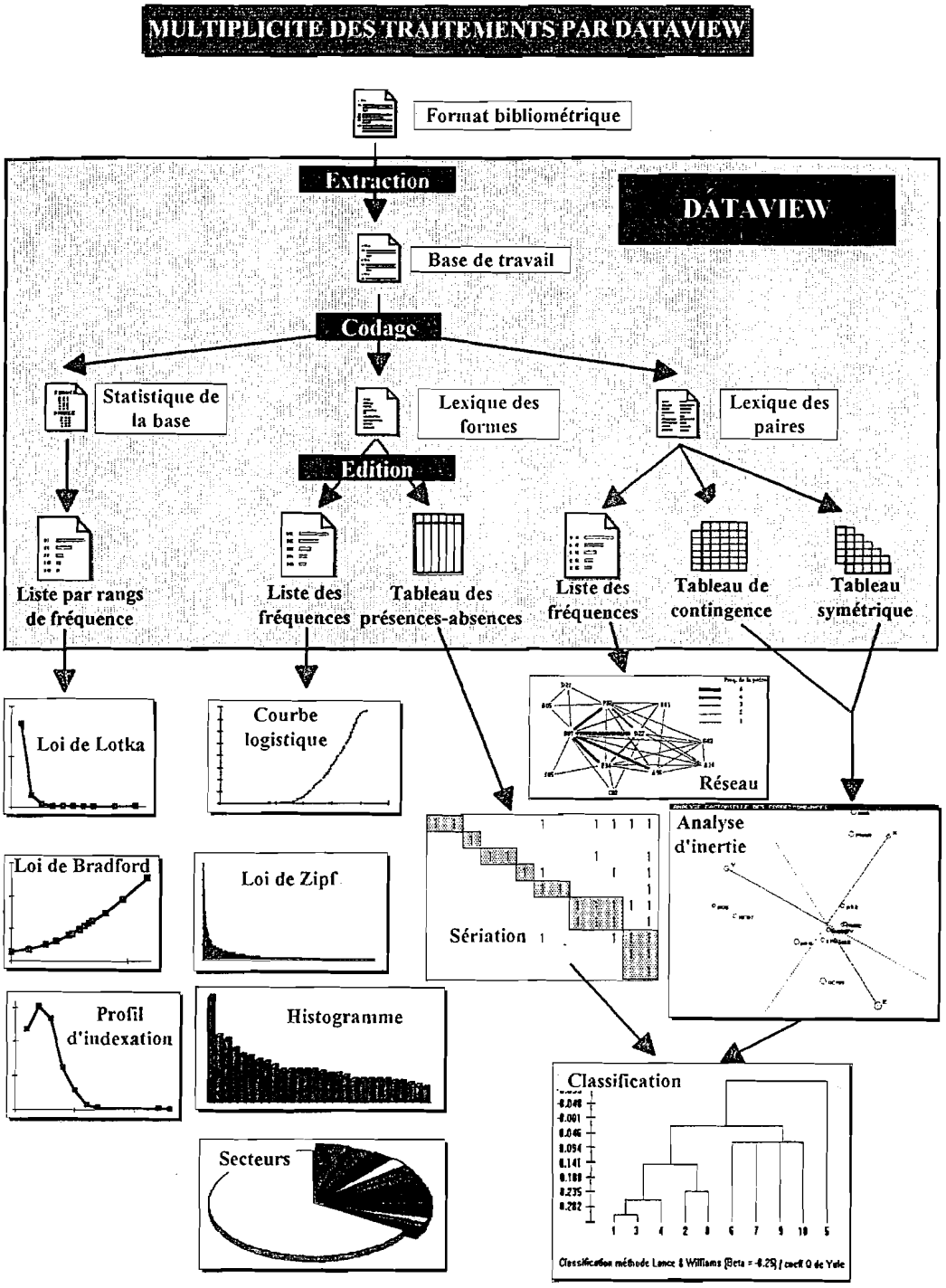

Fig 3: Dataview's treatment multiplicity.

Examples of bibliometric analysis of patent documents for R&D management

Publishing about Bibliometric analysis of patent documents for R&D management is very difficult because of confidentiality problems with companies even if methodologies are already public. This is the reason why we show a non company applied study. Patents in bibliographic databases are indexed with several classifications. The purpose will be to detect complementarities and similarities between three classification having a different structure, signification and use. The goals and case study are given in figure 4 and 5. Figure 6 explains our corpus collection strategy.

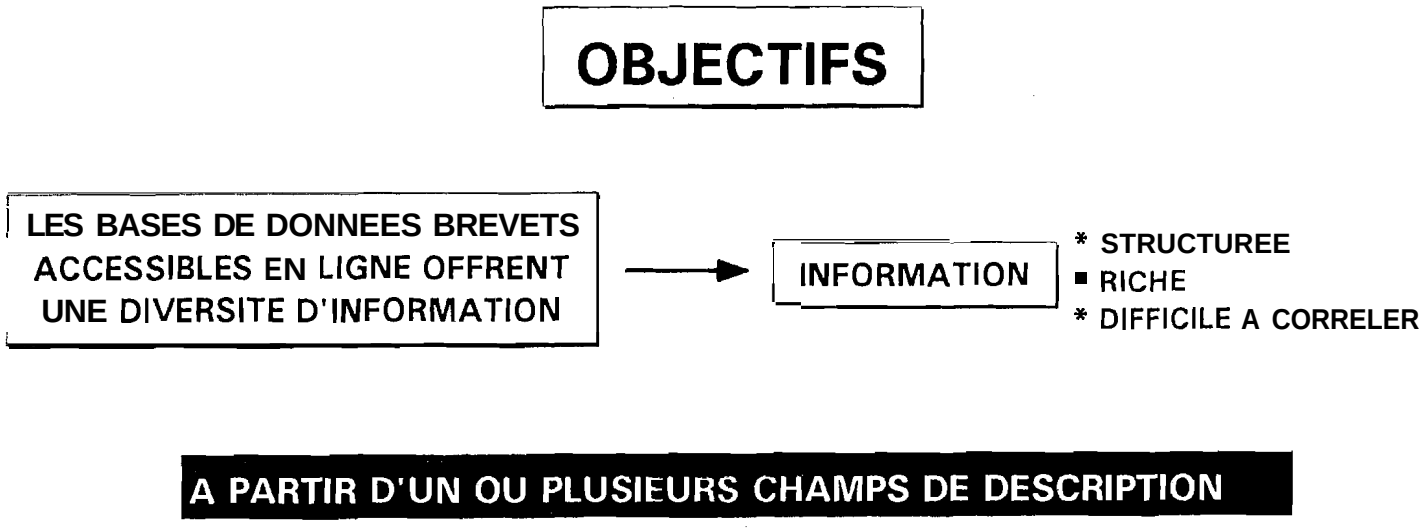

Fig 4: Our study's goals.

LE CAS ETUDIE

LE CORPUS

- LE THEME — SYSTEMES TRANSDERMIQUES THERAPEUTIQUES SOUS FORME DE PATCH (T.T.S.)

  SYSTEMES ADHESIFS QUI ASSURENT UNE DIFFUSION CONTROLEE D'UN PRINCIPE ACTIF PAR VOIE TRANSDERMIQUE

- LA SOURCE — LA BASE BREVETS DERWENT WPIL

LES TRAITEMENTS AUTOMATIQUES DES DONNEES TEXTUELLES

- DATAVIEW (C.R.R.M.)

LES TRAITEMENTS STATISTIQUES

- L'ANALYSE RELATIONNELLE (C.E.M.A.P. - F.MARCOTORCHINO, P.MICHAUD)

Fig 5: Our case study.

FILE WPIL

QUESTION	1	PATCH OR PATCHES OR PATCHS	(2106)
QUESTION	2	1 AND TRANSDERM:	(103)
QUESTION	3	1 AND THERAPEUTIC:	(36)
QUESTION	4	1 AND PERCUTANE:	(15)
QUESTION	5	1 AND (DRUG OR DRUGS)	(102)
QUESTION	6	1 AND MEDICIN:	(14)
QUESTION	7	2 OR 3 OR 4 OR 5 OR 6	(160)
QUESTION	8	7 NOT (CARDIAC# (W) PATCH##)	(159)
QUESTION	9	8 NOT (VASCULAR# (W) GRAFT#)	(158)
QUESTION	10	9 NOT (PATCH## (W) GRAFT#)	(156)
QUESTION	11	10 NOT (FASTENING# (W) PATCH##)	(155)
QUESTION	12	11 NOT (CARRY (W) PATCH##)	(155)
QUESTION	13	12 NOT (CARRIES (W) PATCH##)	(154)
QUESTION	14	13 NOT PROSTHES:	(148)
QUESTION	15	14 NOT CAMERA	(147)
QUESTION	16	15 NOT (TEST# (W) PATCH##)	(146)

PRT AN TI TT AW PR PN AP DS PA IN LA CT IC DC MC AB 1-146

Fig 6: corpus collection strategy.

The three classifications used in our study are:

International Patent Classification (CIB or IP Class) which is an international hierarchical classification of patents attributed by patent offices using the entire patent text.

Manual codes which have a significant meaning in application are attributed by the database producer using the entire patent text.

Derwent codes are abstract reading attributed.

Figure 7 shows the hierarchical structure of those codes

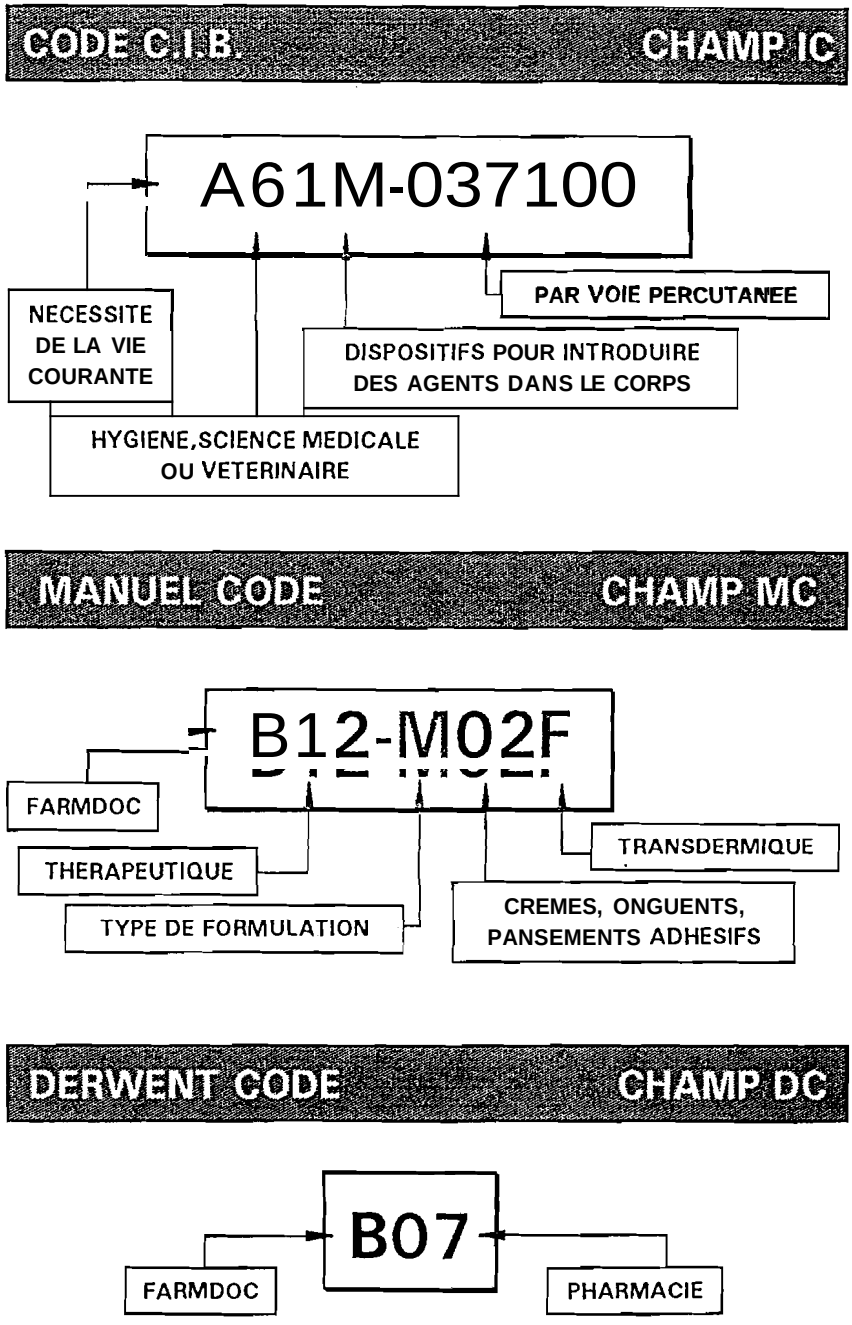

Fig 7: Hierarchical structure of the studied classifications.

In Companies people do not always know which is the relevant classification, and if they are equivalent or not. To perform such an analysis in an automatic way (the best way in bibliometric studies), we do need first to reduce the heterogeneity problem with a hierarchy's code study in our downloaded database. This study is related in figure 8 and allows truncations choices for matrix building.

CHOIX
DES NIVEAUX HIERARCHIQUES
DES CODES

HIERARCHIES DES CODES Nb de digits	NB DE CODES	NB DE CODES APRES ELIMINATION DES CODES A FREQUENCE 1	NB DE BREVETS	NB DE BREVETS APRES ELIMINATION DES CODES A FREQUENCE 1	NB MOYEN DE CODES BREVET	NB DE CLASSES	% DE CLASSES A UN ELEMENT
DC 3	52	32	146	146	3,6	27	30
IC 4	40	21	146	143	2,2	23	22
IC 7	94	35	146	140	2,9	42	36
IC 11	133	36	146	113	2	67	52
MC 3	51	42	143	143	4,6	41	34
MC 5	139	90	143	143	6,9	80	50
MC 7	315	158	142	141	8,7	106	74
MC8	144	69	135	134	4,3	80	55
MC 9	43	12	46	34	1,7	124	90

Fin 8: Hierarchical codes study inducing the future truncations

Then we built a matrix as explained in Figure 9. This binary matrix has been built automatically with DATAVIEW software. Several classifications have been tried over this matrix. The two first classifications were performed in the CEMAP IBM Centre of Paris and are partially shown on figure 10 and 11. These classification techniques have been related in one of the authors thesis [Huot92]. The two last classifications are not explained in this figure but will be explained further in this paper and have been performed in our laboratory.

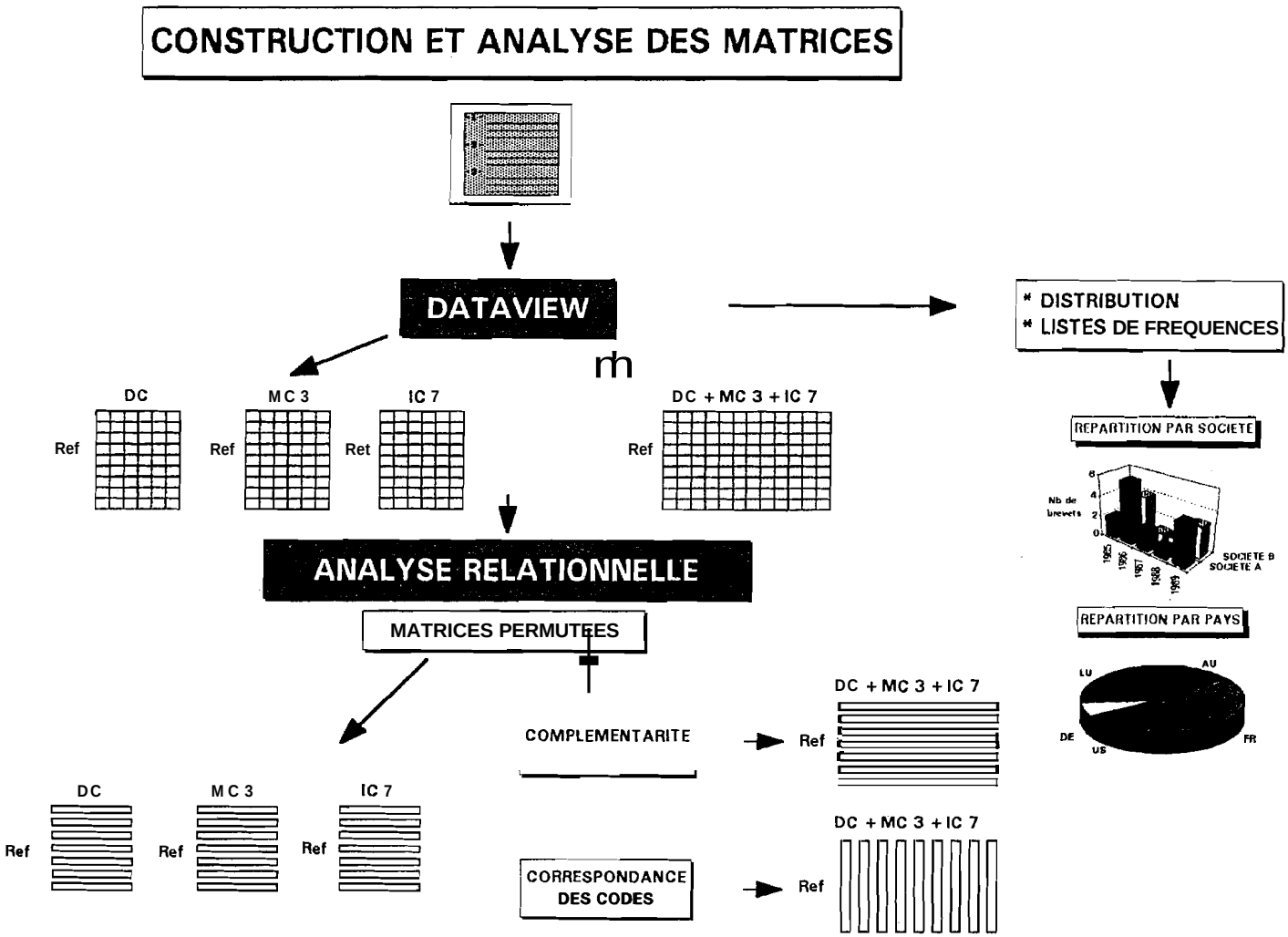

Fig 9: Matrix building and part of classification strategies used

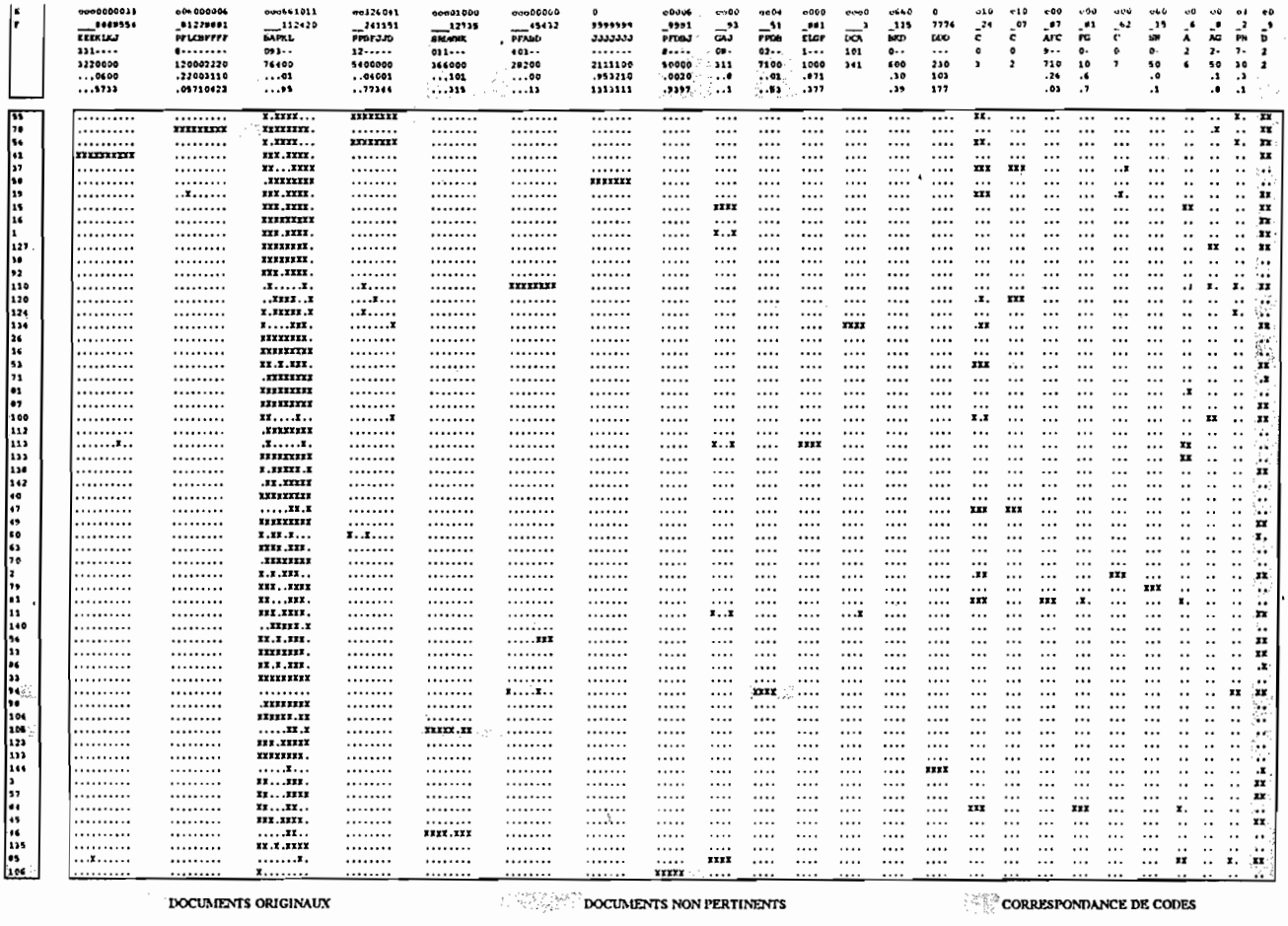

Fig 10: Part of the classification codes classification performed in CEMAP with expert interpretation.

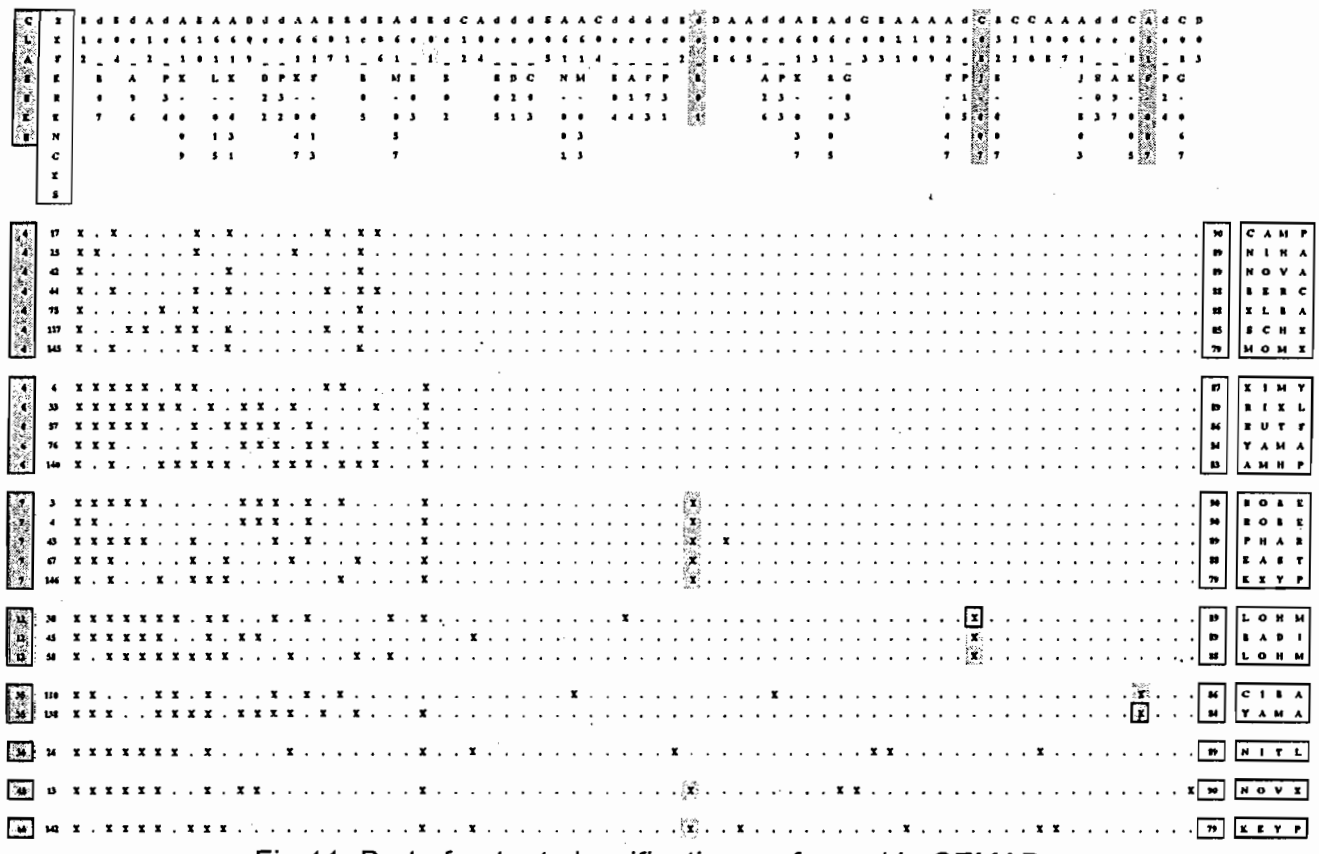

Fig 11: Part of patent classification performed in CEMAP

The idea in our laboratory was to perform a CEMAP techniques adaptation. Our goal is to provide an efficient tool able to be used in micro-computers, for small and middle size companies. Our technique allows a simultaneous codes and patents classification which respects Zipf's law properties. This force a double classification: one only with high frequencies classification and one with medium and low frequencies. Both frequency level and class number are automatically determined. The classification techniques is a non hierarchical method that self determined the best mathematical arrangements of the initial matrix. Those classification techniques could help the expert in detecting general patents (only abstract with high frequency codes) and innovative ones (with small frequency codes) within the entire corpus. Figure 12 and 13 relate those classifications.

Fig 12: High Zipf's law frequency codes classification.

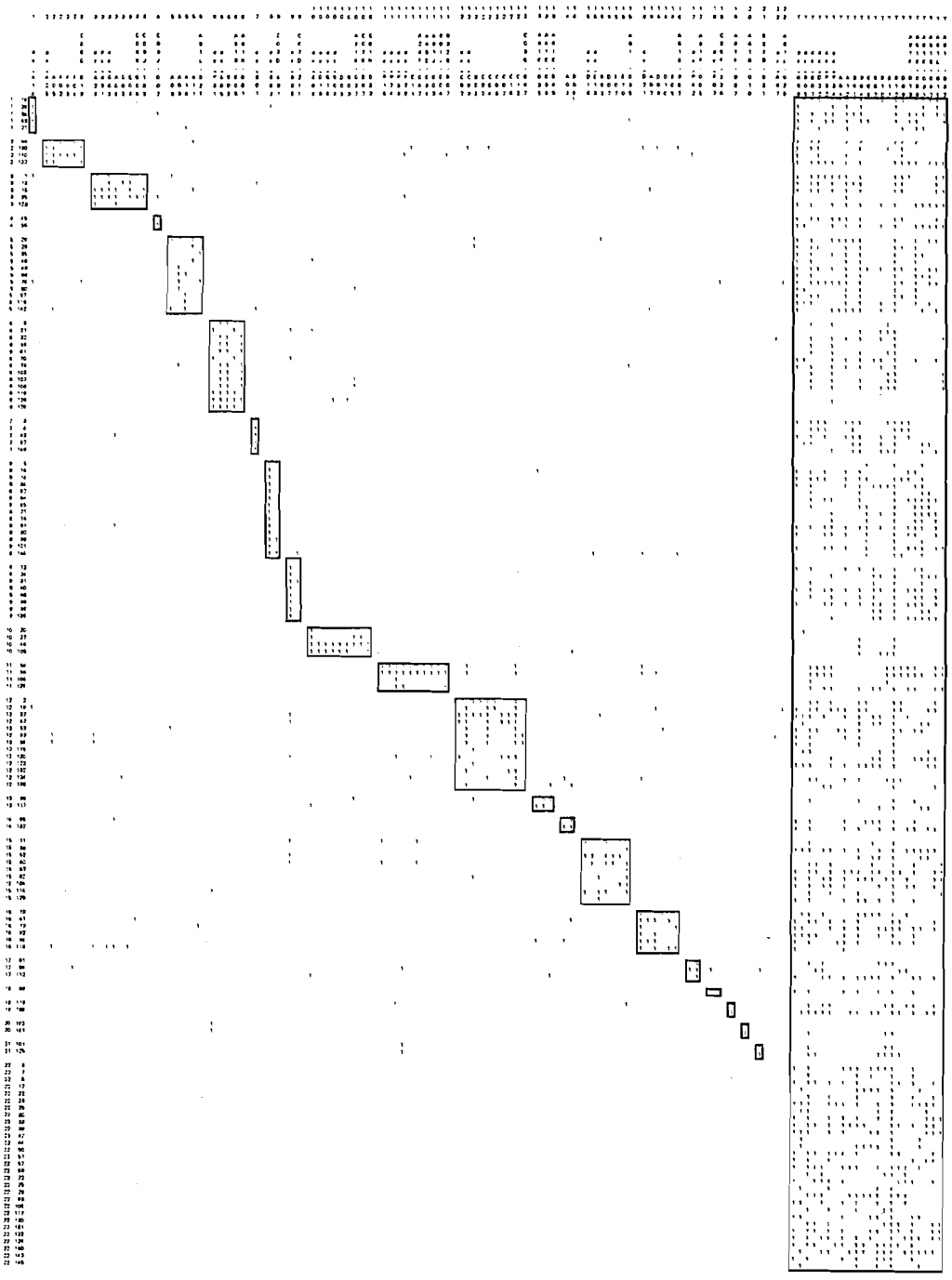

Fig 13: Medium and low Zipf's law frequency codes classification.

Figure 14 shows the classifications interpretation and conclusions which will be used to finalize the report for the decision maker analysis. This figure results from the collaborative work of three persons. An expert in online and patents, a statistical expert, a research studied area expert.

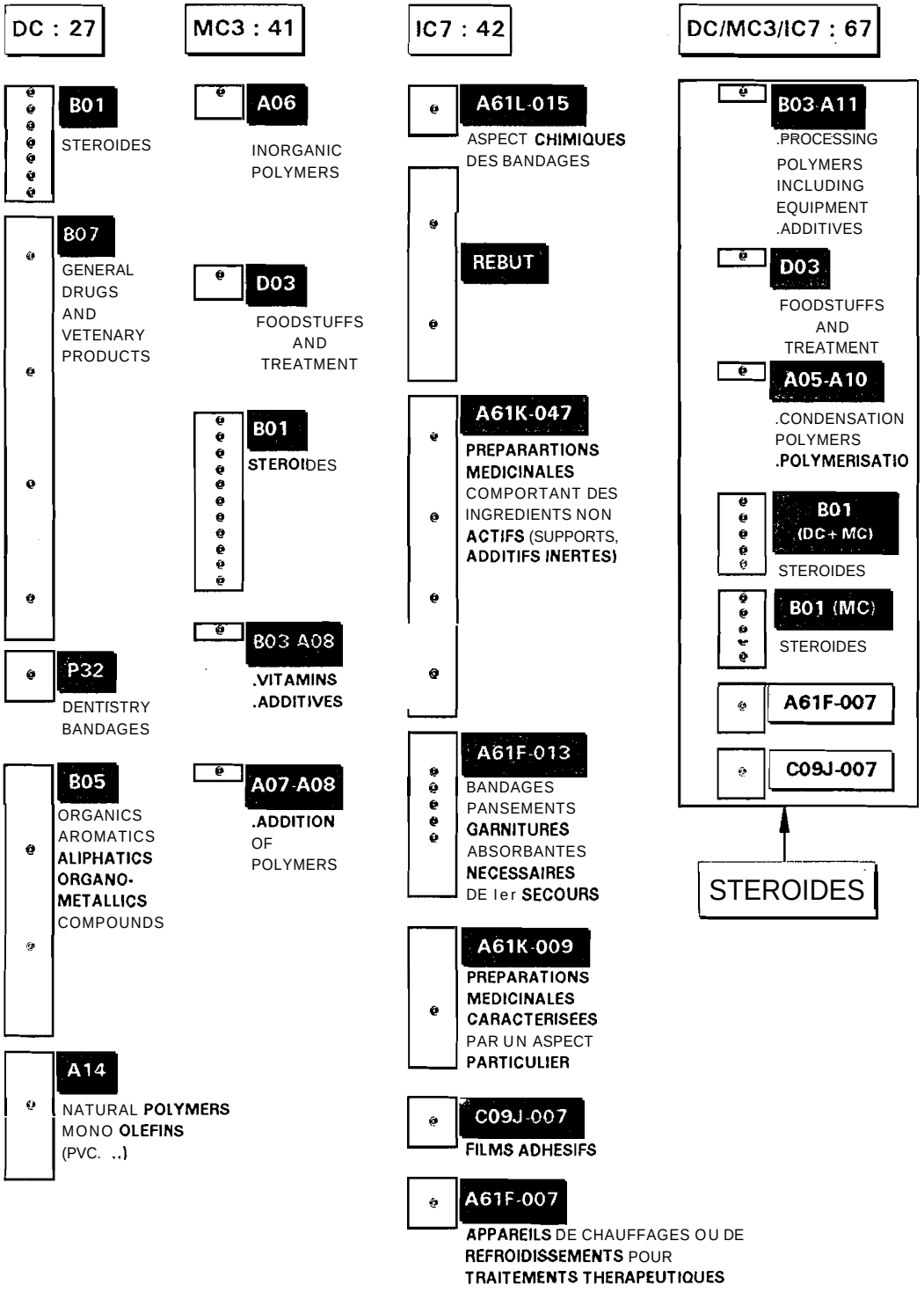

Fig 14: Codes signification and classification interpretation.

Conclusion

We tried to illustrate our purpose with a complete example to show that bibliometric studies of patents may be a relevant way of helping any kind of research activity. What is the most important is that this kind of analysis must be performed with strategic considerations. Patent bibliometric analysis performed in this way is an efficient help in R&D management. The other points are the necessary integration of these analysis in company's competitive intelligence. Without competitive intelligence, bibliometric analysis is poor and vice versa. More important than sophisticated analysis is the relational aspects with strategy, between several experts and communication tools with decision makers. Our experience shows us that this kind of analysis to manage R&D is useful but that it is also very difficult to know exactly the companies usages because of confidentiality.

Bibliography

1. [Bern88]
BRENARD Catherine
La stratégie du brevet
Sciences et Vie économie
Septembre 1988, 50-60

2. [Brou92]
Broustail Joël, Fréry Frederic
Le management stratégique de l'innovation
Précis Dalloz
1992,221 p

3. [Coor85]
COORAY
Knowledge accumulation and technical advance
Research Policy
1985, vol 14,N 2, 83-95

4. [Cour90]
COURTIAL Jean-Pierre
Introduction à la Scientométrie
Editions Anthropos
1990, 436p

5. [Huot92]
Huot Charles
Analyse Relationnelle pour la Veille Technologique
These de doctorat. UIII. Marseille. France
1311211992, 173p

6. [Jako91]
JAKOBIAK François
Pratique de laVerille Technologique
Les Editions d'organisation
1991, 219 p

7. [Jako92]
JAKOBIAK François
Exemples commentés de Veille Technologique
Les Editions d'Organisation
1992, 198 p

8. [Mair88]
MAIRE Georges
La Veille Technologique: les brevets des autres source d'information et d'innovation
Travail et Méthodes
1988, N° 475, 25-29

9. [Mait92]
Maitre Paul, Miquel Didier
De l'idée au produit
Eyrolles
1992, 216 p

10. [Marq85]
MARQUER Francis
Innovation et Management des brevets
Les Editions d'Organisation
1985,260 p

11. [Mart91]
MARTINET Bruno, RIBAULT Jean-Michel, LEBIDOIS Daniel
Le management de technologies
Les Editions d'Organisation
1991,309 p

12. [Nivo93]
NIVOL William
Systemes de surveillance systematique pour le management stratégique de l'entreprise
Thèse de doctorat. UIII. Marseille. France
10/05/1993, 333p

13. [Pavi84]
PAVITT K.
Sectorial patterns of technical change: towards a Taxinomy and a Theory
Research Policy
1984, Vol 13, 323-373

14. [Perr84]
PERRIN Jacques
Les transferts de technologie
Editions La Decouverte
1984, 117 p

15. [Rost93]
ROSTAING Hervé
Veille Technologique et Bibliométrie: Concepts, Outils, Applications
Thèse de doctorat. UIII. Marseille. France
13/01/1993, 353p

16. [Sale93]
Saleh Shoukry, Wang Clément
Mana ement of innovation. Strategy, structure and organizational climate
IEEE Transaction on Engineering Management
1993. Vol 40, N° 1, p16-21

17. [Wise83]
WISEMAN P.
Patenting and intensive activity on synthetic fibre intermediates
Research Policy
1211983. Vol 12, N° 6, 329-339

We would like to thank INFOPRO (ORBIT) for their precious help during this work