Pergamon

World Patent Information, Vol. 19, No. 3, pp. 167–174, 1997

Printed in Great Britain
0172-2190/97 $17.00 + 0.00

PII:S0172-2190(97)00023-9

The function–application relation through a link between classification and indexing

P. Faucompré, L. Quoniam and H. Dou, Centre de Recherche Rétrospective de Marseille, Université Aix-Marseille 3, FR 13397, Marseille Cedex 20, France

The International Patent Classification (IPC) gives a fundamental importance to the distinction between function and application. This distinction constitutes one of the most important classifying rules for patent documents. Most indexing languages do not follow this principle — they deal with this difference by selecting supplementary exact terms or by establishing more definite multiple-term syntagms. End-users have to formulate their searches with relevant terms or a combination of them. With so many different principles, is a connection between scientific information and technical information based only on the IPC still relevant? In this paper, we present a study which attempts to show that the lack of formal distinction between function and application, in the case of thematic indexing, does not create ambiguities. Moreover, at the same time, it allows one to select bibliographic references which only refer to functional aspects and references which only refer to applications. By means of a few examples, we endeavour to show that the distinction between function and application can be advantageously exploited by using both this technical classification and the indexing of scientific databases. Additionally, this distinction provides the opportunity to explore one of the multiple paths between the documentary representation of the scientific literature and the documentary representation of the technical literature.

Introduction

For many reasons which are notably juridical (intellectual property laws), conceptual (notion of novelty) and historical (multi-national origins), the difference between the function of a technical object and its applications has always been considered so important that it has been formalised in the hierarchical structure of the patent classification. In consequence, except for a few cases, the function of an object *per se* and all its applications are classified in different places. However this principle rarely operates in the numerous search languages available for use with bibliographic databases. So can we link with advantage tools which do not make this distinction?

We will only consider this problem from the perspective of the correspondence of the IPC (a universally adopted classifying and search tool for patent documents) with the thematic indexing which has a similar role in scientific bibliographic databases. We will examine whether paths between function and application cannot be considered one of the archetypal relations that we in fact establish or search, between the documentary representation of scientific information and the technical information. For each study, we will successively consider the particular viewpoint of the IPC, and then the viewpoint of indexing principles as used in scientific databases.

Limitations of this study: for classifiers and searchers, the determination of the appropriate classifying at function or application places is sometimes one of the most difficult problems encountered with the IPC.[1] It is also the most discussed aspect of its evolution.[2] We will not treat this problem itself but only through the problematic of paths between a hierarchical classification and an indexing language. The analytical or synthetic characteristics of these tools will be examined in the same manner.

Function and application

On the IPC level

In the IPC, the distinction between function and application is an important principle in the elaboration of its hierarchical structure as a logical division principle and it is a fundamental rule which must be applied in the classification of patents. Since the fifth edition, paragraphs 52 to 54 of the IPC Guide concerning the classifying rules, define clearly that:[3]

52. [..] the classification provides places for classifying:

> (a) a thing "in general", i.e. characterised by its intrinsic nature or function; the thing being either independent of a particular field of use or technically not affected if statements about the field of use are disregarded.
>
> *Eg: valves characterised by constructional or functional aspects, i.e. the structure of the valve*

does not depend on the nature of the fluid passing therethrough or of any system of which the valve may form part (F 16 K).

(b) a thing "specially adapted for" a particular use or purpose, i.e. modified or particularly constructed for the given use or purpose.

Eg: a mechanical valve specially adapted for insertion into a human heart (A 61 F).

(c) the particular use or application of a thing.

Eg: the use of an organic chemical compound as a fertiliser (C 05).

(d) the incorporation of a thing into a larger system.

Eg: the incorporation of a leaf spring into the suspension of a vehicle wheel (B 60 Q).[4]

The first category contains the "function-oriented places" and the following categories refer to "application places". The IPC is known as mainly a *function-oriented classification.* In general, obtaining juridical protection for the function of an invention is strategically more important than gaining protection of any one of its applications. To obtain an equivalent protection for each and every known and potential application is unrealistic. However, despite the importance of the function aspect, we must not forget that many patented inventions are for applications and, for this reason, the IPC uses the two approaches, satisfying the needs of end-users.[5]

At the thematic indexing level. Even if we understand intuitively the relationship between function and application, it is quite difficult to give it a precise definition for documentary purposes. In general, we can say that it is an implicit relationship[6] between associativity (high relativity) and specificity (low relativity). From a strictly documentary viewpoint, it is an anti-symmetrical relationship insofar that all applications presuppose or imply the function but the opposite is not true. However, not all systems use a hierarchical-type semantic relation linking a broader term with a narrower term. In a post-coordinated controlled vocabulary which does not contain particular syntactic procedures and facets allowing this distinction, the same descriptors are used to describe the function and the application. So, this vocabulary can use only two ways to refer to this relationship: either merely to add supplementary descriptors to refer to particular applications, or to use syntagmatic-type linguistic relations allowing the build up of locutions of multiple terms. The set of indexing terms or syntagms allows the user to introduce, by their combination, a sufficient degree of precision to carry out the relevant search.

Consequence of a link between classification and indexing. On the formal level, the relation between function and application, as it has been developed in the IPC, has no simple correspondence in other documentary languages. There are two main reasons for this. First, on a theoretical plane, the concept of viewpoint or facet would be required to correctly represent the function and application aspects of technical objects. This very attractive conceptual approach would not only complicate classifying tasks but also prevent achieving the necessary exhaustiveness of searches. Moreover, the inclusion of these two viewpoints or facets in the logical division may create a semantic polyhierarchy[7] or at least some ambiguities.

The second reason is that to correctly represent the four cases illustrated above by the classification would require weakly symmetrical associative implication relations, partitive relations and hierarchical relations, perhaps together with facets to refer to particular uses or adaptations. In contrast to the IPC, indexing languages used with scientific databases have a natural tendency to give technical objects a functional dimension even where this aspect is not directly and explicitly expressed in documents referring only to their applications. The difference between function and application is not explicit but it is no longer a determining criterion *per se*. The function can always be retrieved over and above all specific applications. The first example above concerning valves and heart valves can be used to illustrate this point (Fig. 1).

It is obviously trite to note that if a vocabulary does not include any distinction between function and application, a search of an object will select both documents concerning this object *per se* and documents referring to its adaptations for a particular use or

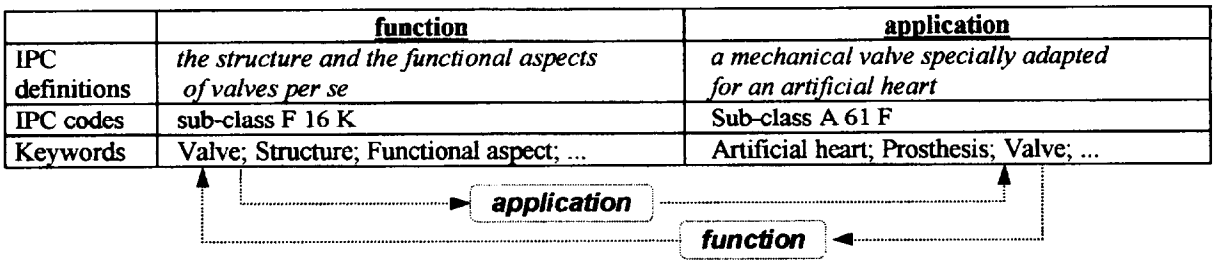


Fig. 1. The path between function and application.

purpose. However, the possibility to go back from one to another is of strategic interest for users who are dealing with researches of well-known or innovative technical solutions.

The implicit and the explicit

On the IPC level. The essential part of a patent document, whether an application or a granted patent, is the scope of the disclosure, especially the scope of the invention claimed. For inventors or patent assignees as manufacturers, the purpose of industrial property departments, and more generally the tasks of patent engineers, is to avoid missing any document which could lead to drastic consequences or to draft a document which fails to adequately protect an invention. Numerous patent assignees have already suffered because of competitors who, without hesitation, systematically exploit these failures when a patented invention represents an important economic and technical asset. So, it is essential to ensure that the implicit has no place in patent classification. It is therefore understandable, that since the claims precisely limit the scope of the juridical protection, that on the documentary level, the claims provide the most essential basis for the classifying of patents.

On the thematic indexing level. In bibliographic databases which cover the scientific literature, indexers have to identify concepts contained in documents. The obvious quality of their indexing, regardless of available tools and producer's strategy, lies in their competency to identify all important concepts: "il est recommandé de ne pas se limiter au contenu explicite du document, et de déceler les concepts *implicites* qui constituent un motif d'intérêt pour le document sans y être directement contenus".[8]

Consequence of a link between classification and indexing. On the methodological plane, those classifying patents and those indexing scientific databases do not follow entirely similar routes. Contrary to common indexing practice, the detailed analytical elements required in the examination of patent applications, are not found at the general disclosure level. The IPC codes applied to patent documents do not usually reflect the finer details of the disclosure since these are not explicitly classified by examiners (here, it does not matter if the classifying symbols are applied before or after the examination). In other words, one cannot explain completely the difference between the final result of these two documentary treatments simply on the analytical characteristics of indexing or synthetic characteristics of classification.[9]

The whole and its parts

On the IPC level. In fact, the loss of a part of the finer detail comes from the logical structure of the IPC itself and the strict respect of the classifying rules by patent examiners. Concerning places provided for technical objects, WIPO explains that

> "The classification attempts to ensure that any technical subject with which an invention is essentially concerned can be classified, as far as possible, as a whole and not by separate classification constituent parts",[10]

the function-oriented classifying reflecting this principle. So, the distinction between function and application follows from this fundamental rule: the obligation to classify the object as a whole and not its different parts. It is itself subordinated to the general principle relative to the unity of invention. The distinction between function and application depends not only on a technical reality but also on a juridical obligation. We have to remember that the third patentability criterion specifies that

> "l'invention technique doit être susceptible d'application industrielle".

If we consider a concrete technical object, be it a product or apparatus, its function or application may be classified and the distinction between function and application will place the document at a particular point within a continuous spectrum of technical disclosures.[11] The point will, in part, depend on the view we adopt as to the effective contribution of the invention to the state of the art in the particular area of technology. The decision to consider an object by its function or only by its application, and the decision to treat this object as a whole or as constituent parts, are relative to a given system and cannot be reduced to only the technical dimension.[12]

On the thematic indexing level. The thematic indexing with keywords depends on a great number of factors such as the richness and the precision of the vocabulary, the indexing policy, the adopted profile of indexing, the specialisation of the source and the needs of end-users. However, despite the importance given to one or another of these factors, the selection of keywords has to consider all relevant data which could be of value in any subsequent search. At this level, the relation of the whole to its constituent parts is subject to the same sort of analysis as that developed for the relation between function and application. An indexing language, without specific structure, will consider this particular case as a partitive type semantic relation, which could be merely generated by the coordination of keywords.

Consequence of a link between classification and indexing. Scientific and the technical bibliographic databases (BDB) and the IPC have numerous common points: they use very similar documentary methods for the information organisation and they share the same objective which is to allow an effective search of the material covered. However, when it comes to dealing with function and application, the BDB and IPC diverge. What is classified with IPC codes on the one

hand, and what is indexed in the BDB with a keyword on the other, can appear on different levels (Fig. 2).

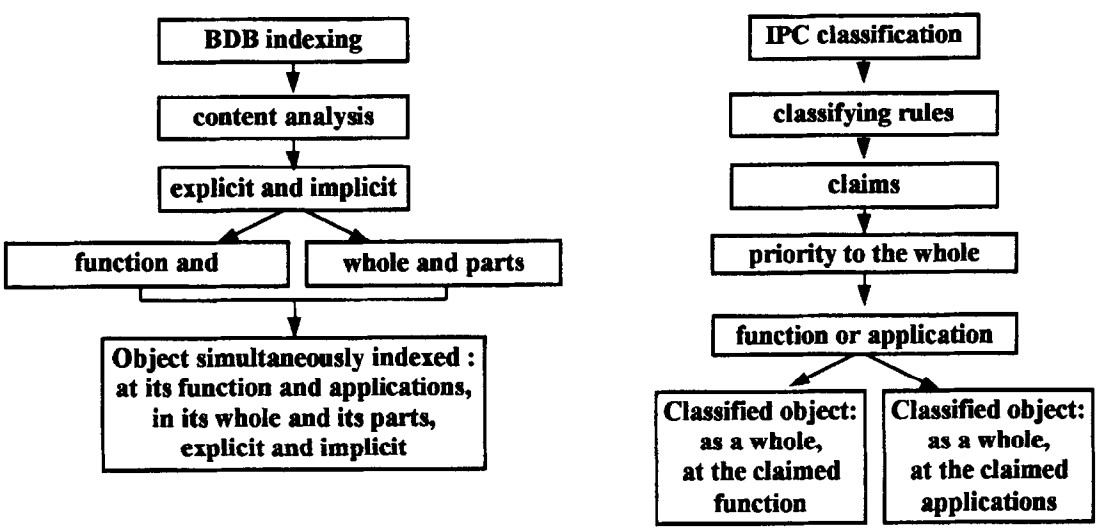
The starting point for a search could be either the IPC codes or keyword indexes: in both cases, a search which attempts to exploit the relationship existing between scientific information and technical information, must set aside the difference of level. We have seen that the distinction between function and application which is inherent in the hierarchical structure of the IPC may be overcome simply by lexical or combinatory means in an indexing language. The consequence is that, for a user who is unfamiliar with searching industrial property, the IPC will have a tendency to reduce recall and relevance whilst for the IPC expert, use of any unfamiliar indexing search language will have the opposite effect of increasing the documentary noise.

The logical link between function and application

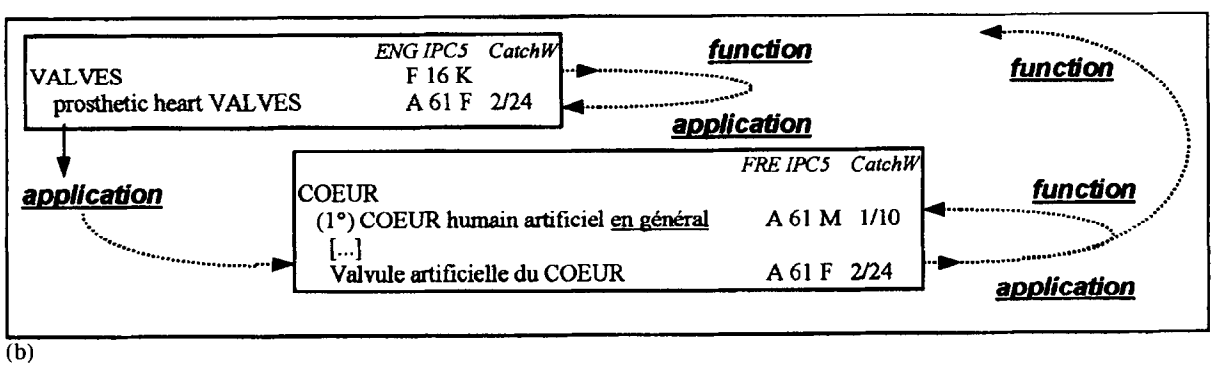
The IPC contains some expressions, well defined terminologically, which refer either to the function, such as "in general" and "*per se*", or refer to applications such as "specially adapted for", "specially designed" and also "specially used". Figure 3 shows, using the previous examples, how these expressions could be of some documentary and industrial interest.

Counting these explicit expressions (from the fifth IPC edition available on the IPC:CLASS CD-ROM) illustrates the main thrust of the classification: in all sections, we found twice as many entries for the function than we did for the applications (Table 1). But, in comparison with the 65,000 classified technical domains, their number seems extremely small. This shows that the distinction between function and application is only included in the internal hierarchical structure.

Do the IPC keyword indexes show this precision? These keywords, called catchwords, provide direct access to the classification system without one having to know the hierarchical order of headings. Certain main or secondary definitions of these catchwords include these expressions which refer to the function or the applications. A similar counting as performed previously shows that catchwords further emphasise functional aspects of technical objects so much that



(a)



(b)

Fig. 2. (a) A few general principles of indexing with keywords and classifying with IPC codes; (b) routes between function and application.

INT. CL[5]	Catchwords	A	B	C	D	E	F	G	H	Σ
function	844	286	635	128	60	207	384	359	211	3114
application	7	160	434	39	28	102	203	121	100	1197

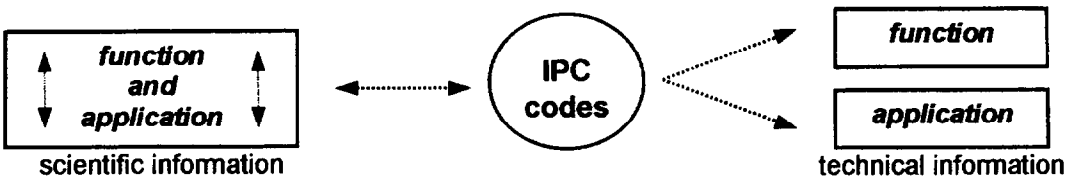


Fig. 3. From the link between indexation/classification to the link between function/application.

Table. 1. Number of expressions referring to the function or the application

Int. Cl. [5]	Catchwords	A	B	C	D	E	F	G	H	Σ
Function	844	286	635	128	60	207	384	359	211	3114
Application	7	160	434	39	28	102	203	121	100	1197

they do not include mention of applications. However, the number of expressions (French edition), in comparison with about the 20,000 main and secondary definitions, is very small. Whether in the body of the classification, notably through the numerous "use notes", or in indexes, the purpose of these expressions is to help the examiner to classify the patent documents. So, the distinction between function and application is almost exclusively based on the hierarchical structure and is directly neither identifiable nor exploitable when only using the technical vocabulary.

A link between science and technology

However, the uses of this vocabulary bring the opportunity to build a relation between the IPC and exclusively technical databases. We have established an automatic path by using correspondence tables between catchwords and compatible indexing languages. These tables allow us to associate each bibliographic reference from target databases with IPC codes. Insofar as automatic paths are only based on the verbal identity of indexing terms, this method can only suggest a presumed relation. However, this presumption is sufficient to explore certain paths, some of which can be very interesting. It allows us to observe how the path between a thematic indexing and a classification, such as the IPC, opens up or uses a second path between function and application as shown by Fig. 3.

To illustrate this picture, we have used the Observatory of the Provence–Alpes–Côte d'Azur regional database which contains all bibliographic documents referencing works carried out by all laboratories located in the Marseille area. The purpose of this observatory is to provide, for firms and for public or private decision makers, a view of the local scientific activity in order to stimulate some collaboration opportunities. For a first experimental stage, the whole database has been re-indexed and, about two thirds of the 5000 references now contain the relevant IPC classification symbols.

For example, a few references have been selected with the "biodegradation" descriptor. This subject illustrates the close links between science and technology because it involves both fundamental biological problems and economically viable solutions, which are created or long awaited by new industrial activities involved in the environmental preservation. Figure 4 refers to work carried out by an oceanology research center. The search dealt with the biological fight against sea-water pollution by hydrocarbons. The first reference referred to the problem of bacterial development and the second to the more specific application of biosurfactants in the fight against pollution. The generic aspect of the descriptor "biodegradation" had selected bibliographic references concerning the bacterial development *per se* and bibliographic references concerning its particular application on hydrocarbons.

The IPC codes, included in the re-indexed references, reflected this two-sided view of the use of such a bacteria. The first document is linked to the IPC subgroup symbol C12N-001/20 (bacteria and culture media therefor included in the main group of microorganisms or enzymes *per se*) whereas the second contains the codes of the subclass C11D-000/00 (detergent compositions) and the main group B01F-017/00 (the use of substances as emulsifying, wetting, dispersing, or foam-producing agents). Then, the Chemical Abstracts database was interrogated with these three IPC classification symbols combined with the truncated keyword "Biodregrad?" to give, for example, the lower abstracts in Fig. 4. There are no common controlled descriptor terms between these two references. In fact, it would be very difficult to obtain the same result without this IPC code because the "biodegradation" keyword would give more than

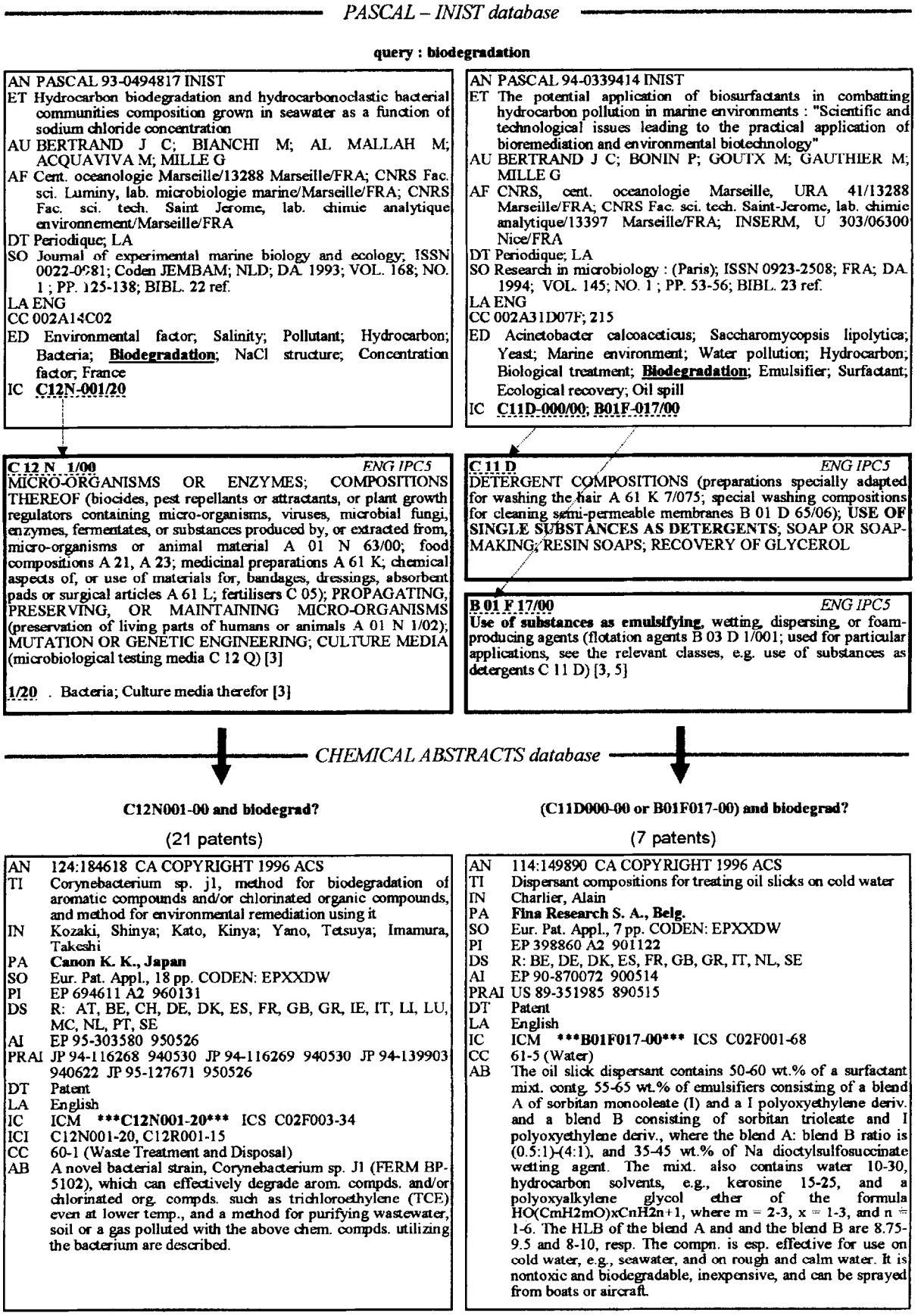


Fig. 4. Function or application classified patents through references linked by IPC codes.

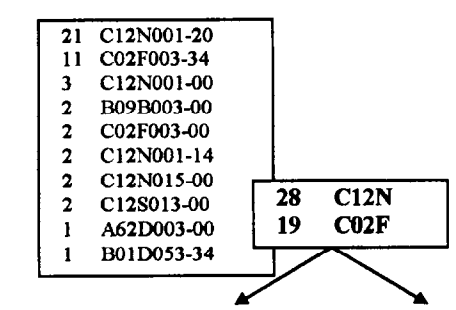
15,000 answers in the Chemical Abstracts database, and since only 3500 patents belong to the micro-organisms domain, this means that more than three-quarters of the publications would not be relevant. This problem is the same on the classical documentary search plane or on a bibliometric analysis plane.

A statistical sorting of the IPC codes from the Chemical Abstracts patent references (in the first case, 21 patents have been considered, and in the second, only the 100 most recent references) showed the divergence between these two series of the most frequently applied codes truncated to four digits. This revealed two clearly different industrial trends. A similar sorting of the patent assignee field showed that, in this biodegradation area, two very different groups of manufacturers were filing patent applications: on biological aspects from USA and Japan, and on chemical aspects from Germany (Fig. 5).

This example shows that classification codes can help to throw light on the context of research and development. These symbols can be used by a laboratory or a firm to find potentially interesting information on emerging new competitive bacterial strains as well as those being studied and also, at the application level, on new formulations of biodegradable substances. These examples show that common descriptors and classification codes can be used to exploit the difference between function and application required by the IPC. A search using merely keywords allows one to retrieve and select simultaneously documents containing IPC codes which are specific to the function and other different IPC codes which only describe the applications. This type of search can be used to demonstrate that while innovative firms invest in certain technological sectors, scientific laboratories carry out more fundamental works in these same sectors. This is a particularly easy approach which allows end-users to differentiate between the fundamental or applied scientific research and industrial research and/or technical inventions protected by the industrial property laws.

Conclusion

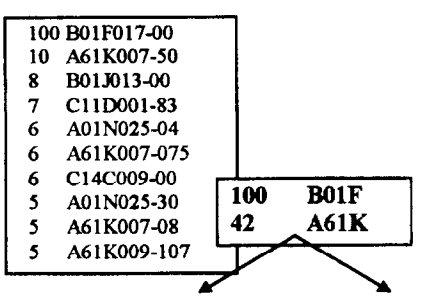
In the IPC, the distinction between function and application is mainly based on its logical division system. However, this classification depends not only on documentary rules but also on well defined legal and technical requirements: its use is to evaluate the so-called novelty and the invention step claimed in a patent application, and to allow the effective and economic search and retrieval of patent documents. In consequence, the purpose of IPC classifying is both documentary and juridical. In most controlled indexing languages, the difference between function and application is only based on the coordination or on the combination of descriptors. This apparent imprecision can be exploited advantageously because it allows end-users using selected keywords either to go from application to the function, or inversely, to retrieve and explore all applications which are linked to a function. In determining research trends, the path from



C 12 N
MICRO-ORGANISMS OR ENZYMES; COMPOSITIONS THEREOF; PROPAGATING, PRESERVING, OR MAINTAINING MICRO-ORGANISMS; MUTATION OR GENETIC ENGINEERING; CULTURE MEDIA

C 02 F
TREATMENT OF WATER, WASTE WATER, SEWAGE, OR SLUDGE

2 CANON K. K., JAPAN/PA
1 BARTL, LUDWIG, CZECH REP./PA
1 CELGENE CORPORATION, USA/PA
1 CIBA-GEIGY A.-G., SWITZ./PA
1 ECOVA CORP., USA/PA
1 FED. REP. GER./PA
1 FELBER, JOSEF/PA
1 FUJITSU LTD, JAPAN/PA
1 GENERAL ELECTRIC CO., USA/PA
1 GOODRICH, B. F., CO., USA/PA
1 HOECHST A.-G., FED. REP. GER./PA
1 IDAHO RESEARCH FOUNDATION, INC., USA/PA
1 KIRIN BEER K. K., JAPAN/PA
1 KOKURITSU KANKYO KENKYUSHO, JAPAN/PA



B 01 F
MIXING, e.g. DISSOLVING, EMULSIFYING, DISPERSING

A 61 K
PREPARATIONS FOR MEDICAL, DENTAL, OR TOILET PURPOSES

9 HENKEL K.-G.A.A., GERMANY/PA
8 HENKEL KGAA, GERMANY/PA
3 BASF A.-G., GERMANY/PA
3 HOECHST A.-G., GERMANY/PA
2 CO OP CHEMICAL CO LTD/PA
2 IFAC GMBH, GERMANY/PA
2 KOGYO GIJUTSUIN, JAPAN/PA
2 NIPPON OILS & FATS CO LTD, JAPAN/PA
2 PEOP. REP. CHINA/PA
2 PULCRA S.A./PA
2 RHONE-POULENC CHIMIE, FR./PA
2 RHONE-POULENC INC., USA/PA
2 RHONE-POULENC SPECIALTY CHEMICALS CO., USA/PA
2 SANDOZ-PATENT-G.M.B.H., GERMANY/PA

Fig. 5. From a scientific keyword to industrial trends.

one to another can offer a very useful search tool, all the more so since it emphasises the difference between science and technology. Through the triple path indexing/classification, function/application and science/technology, the logical precision of the IPC can be combined with generally weakly structured indexing languages to give a new relatively fine retrieval with a high qualitative level.

References

(1) Hansson, B. and Makarov, M. (1995) The sixth edition of the IPC. *World Patent information* **17,** 5–8.
(2) Rozov, B. (1993) About patent classification. *World Patent information* **15,** 217–219.
(3) de Vries, S. (1989) Points of interest concerning the new IPC5. *World Patent Information* **11,** 115–120.
(4) WIPO (1992) IPC:CLASS CD–ROM: English guide, Fifth Edition. WIPO, Geneva.
(5) Claus, P. and Hansson, B. (1980) The International Patent Classification. *World Patent Information* **2,** 13–20.
(6) Maniez, J. (1987) Etude général des langages documentaires. In *Les Langages Documentaires et Classificatoires.* Editions d'Organisation, Paris, p. 271.
(7) Rivier, A. (1990) Construction des langages d'indexation: aspects théoriques. *Documentaliste — Sciences de l'Information* **27**(6), 263–274.
(8) Van Slype, G. (1987) In *Les Langages d'Indexation.* Editions d'Organisation, Paris, p. 155.
(9) J. Maniez, *ibidem,* 163.
(10) WIPO (1990) Classification Internationale des Brevets (5th edn 1989): informations génèrals, 19.
(11) Jenkins, G. (1989) Function and application: is there a difference?. *World Patent Information* **11,** 121–124.
(12) Fen, Z. (1987) Analysis of the properties of the International Patent Classification. *World Patent Information* **9,** 152–156.