

AN EFFECTIVE LINK BETWEEN SCIENCE AND TECHNOLOGY

P. FAUCOMPRÉ, L. QUONIAM, H. DOU

*Centre de Recherche Rétrospective de Marseille, Université d'Air-Marseille 3
FR 13397 Marseille Cedex 20 (France)
E-mail: cmm@cmm.univ-mrs.fr*

(Received August 30, 1997)

The link between science and technology represents a major strategic stake, so the relation between scientific bibliographic references and technical bibliographic references can be of very important documentary interest. To set up this link, International Patent Classification catchwords have been used as a switching language. A previous feasibility study had shown the possibilities of such a full automatic correspondence and its obvious inadequacies. We present here the most important modifications brought to this correspondence, in particular the consideration of multilingual indexes which allow to link several indexation fields with one of the most complete representation of patent classification. The major evolution of our project affects the correspondence mechanism which now generates a global reindexation of bibliographic reference with classification codes. We also discuss the concept of correspondence itself which must be interpreted as a simple presumption of the link. There are some consequences due to these developments: first, insofar as there is not an univocal relation, end users do not have to select switching keywords which generate concordances. They can directly use codes which symbolize the industrial property classification. Next, main documentary indicators do not seem to be adapted to measure the **performance** evaluation of this new field. It has the single role of suggesting trails that can be explored. Lastly, it seems that only end users should be able to supply a complete validation and we show that a documentary validation is not sufficient.

Introduction

Firms which deal with technological innovation are faced with the difficulty of quickly articulating a research of technical information to research of scientific information. For some economic actors, the documentary representation of these two universes, and the multiplication of information systems to access them, weaken the visibility of effective or sought links between scientific research and industrial research. The path from one to another allows the manufacturers to detect more quickly the potential opportunities induced by scientific activity and allows the laboratories to

perceive the needs of technical solutions which come from firms. This link could bridge the gap between the research plane, the R&D plane, and production plane described by (Ziman 1992). To accelerate this path, we have built a correspondence system between these two domains leading to the projection of the International Patent Classification to referenced scientific literature. Each bibliographic reference reindexed by this system contains a new field which contains one or more classification symbols used in the industrial property. In a previous communication (Faucomprk 1995), we had shown a few feasibility principles of this correspondence. After a brief recall of the general context of this project, we will expose the primary results and their limits. We will especially emphasize the evolution of the correspondence concept.

A strategic relation between science and technology

Scientific information and technical information are essential components of the transfer between scientific research and technical development. Their readability contributes to accelerate exchange (*Conseil Economique et Social* 1994), in particular where collaboration networks within technological innovation processes are involved. Firms which participate in these networks are particularly sensible to actual or sought links between scientific information and technical information. However, industrial research permanently redefines established borders between academic research and state of the art techniques. Industrial research is one of the main competitive advantage factors. The consideration, as far as possible, of needed information for technological innovation becomes a determining factor for strategic decision making in the firm. With this emergence, a purely linear and mechanist viewpoint of flows between research, development and production, becomes less relevant: each plane advances according to its specific logic (*Dou* 1993 and *Ziman* 1992). In this common elaboration (*Guellec* 1993) and in these continuously redefined configurations, all actors tie or untie links with another, contributing to bring together or take away the different planes (*Schmidt-Tiedmann* 1982). So, in this context, each partner has to control a wider unsettled environment and the relations that he has to link to a set of technical or economic actors. He must also capture and analyze very diversified information, to sound out multiple sources and to resort to extremely varied research tools and methods. That is, in the perception of this complexity, that firms modify their approach of technical development (*Iribarne* 1987) and their attitude in front of the research (*Retourna* 1995). They appropriate the technology (*Lainke* 1991) as knowledge, competence or goods with successful, only with the limits of their technical culture. For these firms, the readability of required information by these collective constructions is a major

component of all collaboration. Without this readability, they could turn away technical information and their access procedures, or they could not identify technical solutions and the innovative potential that scientific development could supply. Essential relationships of collaboration between firms, or between manufacturers and research laboratories, are developed preferentially at an informal level through very close personal networks. These relations bring out a high degree of credibility, confidentiality and confidence, all indispensable to the projects in which the firm invests an important part of its resources and, sometimes, its future. But beyond this insurance, the formalized exchanges also allow to reduce the part of inherent uncertainty in these collaborations. The documentary information generally has an important role in these exchanges, and, in particular, redirects the work of the information specialist.

An automatic link

Because of the proliferation of documentary systems and their research tools, our project leans on two systems that we could describe as canonical in their respective area: the technical information on one hand, and the scientific literature on the other.

The representation of technical information

A way between science and technology is not feasible by a direct and simple link. So, it is indirectly built between bibliographic references from scientific literature and bibliographic references from technical literature, the patent documents, and this by using the International Patent Classification (IPC). This classification is used by examiners of industrial property offices to decide if the object presented as novelty can be considered as a technical invention and consequently merits the title of industrial property. Its main purpose is to allow a comparison between technical objects, it carries out classification of the technical object in the domain to which it belongs (Fen 1987 and World Intellectual Property Organization 1990). At the most detailed level, more than sixty-five thousand classification codes describe all the patentable technology. The IPC is the comprehensive documentary representation of the judicial protection system of technical objects lawfully applied by 75 national or regional patent offices. Therefore, more than 32 billion available patent all over the world are classified with these international codes. The interest of this classification, commonly used by manufacturers, is the most important since 70% to 80% of technical information is exclusively contained in patent documents (Lawson 1979 and Jakobiak 1994). These all

published patents are referenced by more 90 databases. On the strict documentary plane, only an enumerative and monohierarchical classification as the IPC is able today to insure the exhaustiveness of documentary searches.

The representation of scientific information

On the level of scientific information, the PASCAL database of INIST presents some advantages. Its 11 billion bibliographic references, of which nearly 85% come from the fundamental literature, cover all main scientific domains and their peripheries with many related sectors. Its position as a database for first bibliographic approaches, resolutely multidisciplinary, multilingual and international, gives it an exploratory dimension and thus becomes an ideal experimental database for our project. On the documentary level, its purpose naturally imposes a very analytical indexing language characterized by an extremely voluminous but weakly structured vocabulary.

A switching language

The two selected documentary languages are obviously highly heterogeneous in their nature, operation and purpose, indeed absolutely opposed. It seems that the realization of direct links between them is impossible with tools built on such distant principles. The PASCAL database owns an effective classification scheme, but it is well known that it is very difficult, indeed impossible (*Maniez 1987*), to set up a satisfying correspondence between two classifications which are built around such different logical division principles. Thus a compatible and intermediary language with these two representations is required. This classification contains keyword indexes called catchwords. These IPC catchwords offer to end-users a direct access to classification, without presupposing the slightest preliminary knowledge of the hierarchical structure, and bring to patent examiners an eventual aid in delimiting a technical subject in the preparation stage of documentary research. The choice of catchword indexes is essential for three main reasons. First, they avoid the elaboration of a new tool. And, they offer a natural and validated path between IPC controlled terms and keywords of other documentary resources. Lastly, they allow to free from a direct link between text classification itself.

An automatic link

Once material is selected, a mechanism must now be found to correlate the elements of these two systems and go from one system to the other. So, "*the following suggestions are intended to provide methods for setting up of a compatibility instrument, a black box*" (Dahlberg 1981). A manual processing cannot be considered because of six main quantitative and qualitative reasons:

- the very great volume of data that must be treated;
- the financial cost of such a realization (nearly ten work years);
- the training at double competence (classifying and indexing tasks);
- the difficulty to control inevitable drifts for the duration of a such project;
- the cost of maintenance and update of this system;
- the necessity to make this process repetitive.

Considering these constraints, the principle of a such correspondence can only accept very low options. Thus, the link is set up on the lexical coincidence between the enhanced definition of catchwords and uniterm or multiterms descriptors of the target database. It is important to underline that this simplicity allows to supply a relative robustness to the correspondence and provides some realism to the new established links between the two documentary representation systems. To reinforce the legitimacy of these links exclusively built on the strict identity of words, the processing of correspondence, our *black box*, requires a succession of filters. They use numerical values giving the measure of a few parameters as very common and classical bibliometric indicators. For example, the proximity between indexation terms and catchwords is quantified through a simple inclusion degree which is very similar to a Dice indicator. Another example: a hierarchical frequency indicator, based on the counting of occurrences, allows to select associated IPC codes. All of these filters have a double function. On the qualitative plane, they allow to select the nearest expressions on a verbal level when situations of competing choices appear. This last case is the most frequent. On the quantitative plane, they avoid the explosion of number of the links.

The evolution of the concept of correspondence

The project of correspondence, because of possibilities and inadequacies underscored by the feasibility study, has progressively evolved more at the conceptual level than at the technical level. Paradoxically, this evolution is marked by the consideration of new documentary parameters and, at the same time, by certain distancing of the strict documentary plane.

The widening of the technical vocabulary

To cover the greatest possible number of domains according to the idea that "*The greater the relative specificity and vocabulary size of the input language, the better the switching performance*" (Horsnell 1974), the system of correspondence uses not only French catchwords but also the vocabularies of 5 national patent offices (Germany, Austria, France, Great-Britain and Switzerland). In this manner, the very European perspective of this classification is confirmed. These 5 indexes allow a vocabulary giving the most complete state of the techniques and respect the diversity of documentary resources which can support such a correspondence. However the most important element is not the absolute number of definitions but the representation that these 160,000 entries give of the classification. In comparison with the 65,000 codes of the whole of the IPC, the three indexes cover nearly two thirds of classified objects and technical domains. If we consider one of the hierarchical levels which is particularly important for initial research, the 613 present sub-classes cover 99.51% of symbols. The weight of the German index, used in its English translation, is obviously determining.

Table 1
The number of considered catchwords and correspondence with IPC codes

Index source	Language	Definitions	Codes	Cover
France	French	19,881	14,328	22%
Great-Britain	English	18,029	9,663	15%
Germany, Switzerland, Austria	English	122,811	38,402	59%
Σ		160,721	40,456	62%

The partial consideration of indexation

In order for the best respect of the effective indexation of bibliographic references, the same method has been applied to the PASCAL database. First, the three vocabulary types, corresponding to the indexation of 3 years entries into the database, have been considered. In their whole, that represents more than 1.5 billion bibliographic references. But their analysis revealed that the free indexation presents an inadequate characteristic with a correspondence process: in a very pronounced distribution Zipf, only 12% of free terms occur once a year.

Table 2
Structure of the PASCAL keywords

Extracted vocabularies	Field	Entries	Doublons	Threshold	Processed
Controlled descriptors	FD	96,433	1,388	–	95,045
	ED	99,046	3,153	–	95,893
Table of equivalence	F/E	128,717	3,508	–	125,210
Candidate descriptors	FD	27,404	315	–	27,087
	ED	29,155	563	–	28,590
Free words	FD	537,929	–	>2	65,508
	ED	7,041	35	>2	477

(FD = French descriptor, ED = English descriptor).

In other words, 88% of free words occur less than 3 times for the 3 years of extraction, that induces a quasi null probability of link. As a consequence, only entries which appear at least once a year have been selected. This threshold also presents the advantage of a reduction in the same proportion of the volume of terms which must to be processed.

The recognition of the multilingual aspect of indexes

The consideration of indexes available in several languages follows from the same widening of the base of vocabularies. In fact, the reason for multilingualism is different in the case of indexation and in the case of IPC. Concerning the PASCAL database, the presence of a particular index term at the same time in French and in English indexation fields increases the potential number of links, especially with the German index. This can also avoid some anomalies. In the case of catchwords, the simultaneous presence of an entry in two or more multinational indexes significantly reinforces the probability of its relevance. If several competitive entries claim to same link, only the entries which have the highest frequency in two or more different indexes are selected. In this very frequent case, the final link corresponds to a multiple index object.

The removal of an apparent interface

In the initial version, elaborated for the feasibility study, some information was presented to the end user: the catchwords linked with specific descriptors to bibliographic references, and, associated with each of these catchwords, a numerical indicator showing a relevance degree and one or several classification codes. Then, the

end user had to select the most interesting catchword, eventually in function of the proposed correspondence degree. In a third and last time, he had to choose the most relevant classification symbol. However this system was too heavy because of the number of intermediary stages and the number of selections. The correspondence system imposed too many accesses. Instead of further the final readability, it only increased the overall opaqueness of links. Our project underestimated that end users should not be interested by keywords lists which were supposed common but never used. The relational tool has to become completely transparent. Only obtained classification codes must be apparent. Setting aside mediation tools, the end-user can directly understand familiar documentary representation without considering production conditions of codes symbolizing the industrial property.

The concept of global reindexing

In the previous interface, only one index and one indexation field were used. So, each established link between a descriptor and a catchword was presented separately. Classification codes, only available through the selection of a particular catchword, were completely independent. The necessity, more conceptual than ergonomic, to present only classification codes generates a closer correspondence with the common profile of symbols applied to classified patent documents. Considering that there is no realistic possibility to further one descriptor more than another, and to take into account that there is no obvious legitimacy to compute some values measuring the distance or the proximity between two descriptors (*Risjbergen 1978*), no synthesis or automatic classification of keywords is able to select one or more classification codes. Furthermore, this hypothetical synthesis or this automatic classification would not allow to work the common point between all these codes which is their appartenance to the same hierarchy. The treatment associating classification codes to references now considers the hierarchical frequency of selected codes for the final reindexation, without being concerned with intermediary catchwords which become useless.

Correspondence tables between indexing vocabularies

The link with PASCAL database

Next, the three target vocabularies of the PASCAL database can be automatically linked. This treatment is filly reproducibile with the indexation vocabulary of any other bibliographic database. The correspondence system has generated a total of 8,175 links.

This result can be estimated as relatively weak regarding initial data. However, we have to remember that, on one hand, each vocabulary comes from of a very different universe, and, on the other hand, documentary tools designed for more than 30 years, IPC excepted, contain from 4,000 to 6,000 entries. Beyond the particular result for these three types of vocabulariy, it is important to underline their respective contribution to the final score of obtained links. After a simple dedoubling, the three vocabularies have a total of 3,126 French linked descriptors and 5,049 English linked descriptors. However it is very interesting to remark that only controlled descriptors supply about 84% of French links and about 96% of English links. This fact reveals the very power contribution of candidate descriptors and of free term whereas their respective volume corresponds to 30% and to 68% of controlled descriptors. The consideration of these two specific vocabularies, can be without doubt discussed. As a last remark: the linked 499 sub-classes through the covered codes by established links allow to cover more than 81% of the hierarchical structure.

Table 3
Number of links between vocabularies and catchwords

Target vocabularies	Field	Linked entries
Controlled descriptors	FD	2,619
	ED	4,843
Candidate descriptors	FD	360
	ED	770
Free terms	FD	1,589
	ED	16

An example of indexing of bibliographic references

The generated correspondence tables have been applied to the Observatory of the Regional Provence-Alpes-Côte-d'Azur database which today contains about 5,000 bibliographic references. This database is full available on CRRM Web host (CRRM 1997). It has been created by an extraction of all references which satisfy the geographical criteria corresponding to the Marseille area for the most recent years (1993–1995) from the PASCAL database. The purpose of this observatory is to provide firms and public or private decision-makers, a view of the activity of laboratories located in the Region and to stimulate some collaboration opportunities. For the experimental stage, the whole database has been reindexed. Today about two thirds of the references contain IPC classification symbols. Figure 1 shows the primary results of

this reindexation. To measure the information brought by our system, we present, by descending order, a few PASCAL classification codes and a few of the most frequent descriptors included in all databases.

Below these two series, we show some of the most frequent IPC codes (fifth English edition) with the main abridged title of their respective definition. In this picture, we can observe that the view given by IPC codes and the view given by PASCAL classification codes are appreciably different. It is interesting to note that the correspondence does not seem to introduce obvious aberrations in comparison with the key words attributed by the indexers of INIST (this seems logical because from same descriptors). This new data is not a substitute for any existing information, and does not correct any other information. On the contrary, its quality depends not only on generating treatment but above all on indexation elements. As simple probability of link, it only makes up present information and enriches the interpretation that we can give to a bibliographic reference by adding an opening towards other technical domains.

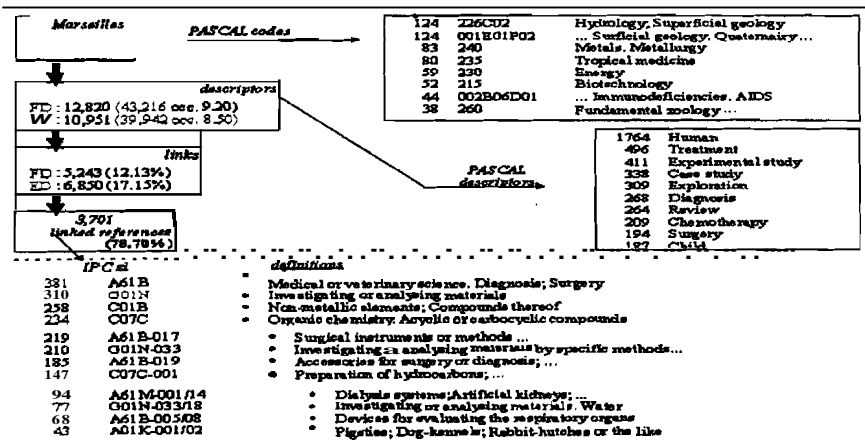


Fig.1. The reindexing codes of the Observatory of Provence-Alpes-Côte-d'Azur database

For example, in this database, several publications deal with highly acute renal pathology. Discussions are mainly oriented to surgical, symptomatological or other methodological aspects. Indexation terms represent these aspects very well but only on the scientific plane. However, if this precision is required at a fundamental level it does

not bring an immediate visibility to end users, e.g. small and middle-sized firms who are not very familiar with this type of vocabulary. In examining the associated IPC codes to these references, we can observe that these symbols belong to the domain of dialysis techniques and artificial kidneys. Obviously, this does not mean that dialysis or artificial kidneys are the response which must be given to publications previously evoked. This only means that a possible relation, perhaps very naive, occurs between this information. This suggested path between medical publications, not translatable in patent terms, and the world of industrial property, is presented with a direct assimilated form by manufacturers. On the documentary plane, it is important to remark that the native indexation of these references does not include an *artificial* kidney descriptor or a similar concept. The Figure 2 shows a complete reference with its associated IPC code field which attempts to supply a new bridge between scientific literature and technical information.



<p>AN 70011949 - (C) CNRS FT Kyste hydatidique du rein de l'enfant ET (Hydatid cyst of the kidney in children) AU PANUEL M; GOLDSTEIN P; DEVRED P; GUYS J M; FAURE F; GARNIER J M AF CHU La Timone, serv. radiologie pédiatrique/13385 Marseille/FRA; CHU La Timone, clin. chirurgicale infantile/13385 Marseille/FRA; CHU La Timone, serv. pédiatrie/13385 Marseille/FRA DT Périodique; LA SO Pédiatrie : (Marseille); ISSN 0031-4021; Coden PEDRAN; FRA; DA. 1992; VOL. 47; NO. 11 ; PP. 779-783; ABS. fre:eng; BIBL. 10 ref. LA FRE CC 002B05E03A3 IC A61M-001/14; A61B-017/00; A61B-019/00 FD : Tumeur; Exploration; Traitement; Enfant; Exploration ultrason; Echographie; Chirurgie; : Kyste hydatidique; Rein ED: Tumor; Exploration; Treatment; Child; Sonography; Echography; Surgery; Hydatid cyst; : Kidney</p>	
<p>ENG IPC5 StichW } kidney, artificial A 61 M 1/14 }</p>	
<p>ENGIPC5 CatchW A 61 M 1/14 . Dialysis systems; Artificial kidneys; Blood oxygenators (processes of separation using semi-permeable membranes B 01 D 61/00; semi-permeable membranes characterised by the material manufacturing processes thereof B 01 D 71/00) [d]</p>	

Fig.2. A link between scientific information and technical information

All references containing this supplementary field bring a new opportunity to reinterrogate other databases, specialized or not, in patent documents, with these associated codes. They only have to include an IPC codes field. The example of Fig.3, with abridged references, shows that such a link which can be easily set up between a scientific publication referenced by the PASCAL database and a patent document referenced by Chemical Abstracts by using only a classification code. There is no commonly controlled indexation terms between these two references. In fact, it should be very difficult to obtain the same result without this IPC code because a biodegradation indexation term refers to more than 15,000 publications in Chemical Abstracts. Insofar as only 3,500 patents belong to the micro-organisms domain, more than three-quarters of the publications would not have relevance. This problem is the same on the classical documentary search plane or a bibliometric analysis plane. The first publication presents the work of an oceanology laboratory of Marseille. This publication shows the effective impact of salt concentration on the development of bacteria that can degrade hydrocarbons. The second reference is a Canon patent document which protects a novel, and in consequence, a competitive bacterial strain which can also degrade certain aromatic compounds polluting water or soil.

AN PASCAL 93-049417 INIST
 ET Hydrocarbon biodegradation and hydrocarbonoclastic bacterial communities composition vs sodium chloride concentration
 AU BERTRAND J C; BIANCHI M; AL MALLAH M; ACQUAVIVA M; MILLE G
 AF Cent. oceanologie Marseille/13288 Marseille/FRA; CNRS Fac. sci. Luminy, lab. microbiolo CNRS Fac. sci. tech. 5 sirt Jexxre, lab. chimie analytique etvironnement/Marseille/FRA
 DT Periodique; LA
 SO Journal of experimental marine biology and ecology, ISSN 0022-0981; Coden JEMBAM; N
 PP 125-138; BIBL. 22 sef.
 LA ENG
 CC 002A14C02
 ED Environmental factor, Salinity, Pollutant, Hydrocarbon, Bacteria, Biodegradation, NaCl str
 IC C12N 00120

C12N 100

1/00 Micro-organisms per se, e.g. protozoa; Compositions thereof (medicinal purposes); micro-organisms A 61 K 35/06; preparing medicinal bacterial antigen or vaccines, A 61 K 39/00; Processes of propagating, maintaining or preserving them; Processes of preparing or isolating a composition containing a micro-organism

1/20 . Bacteria; Culture media transfer [3]

AN 1241846B CA COPYRIGHT 1996 ACS
 TI Corynebacterium sp. 31, method for biodegradation of aromatic compounds and/or chlorinate method for environmental remediation using it
 IN Kozaki, Shigeo; Kato, Kiryo; Yano, Teiyo; Imanura, Takeshi
 PA Canon K K Japan

Fig.3. The way from a fundamental publication to a patent

Discussion and further developments

A new documentary field

Today, the correspondence system creates a new field which includes classification codes. Specifically, end-users can once again use supplied codes to further research about patent resources without be very familiar with the symbolic of this technical classification. On the other hand, a manufacturer can explore more fundamental publications by using only the classification codes without well controlling the academic vocabularies. More generally, this field can be treated as any available information included in a bibliographic reference and it can be combined with any other documentary field. However, the obvious limits of this automatically generated link must also lay down the limitations of all interpretations that we could get out of it, notably through its combination with other information, and more particularly from bibliometric perspectives.

The concept of a simple presumption link

In the world of industrial property, the concept itself of concordance between IPC and the other national patent classifications is particularly problematic. The conditions which are linked to patents are not easily accessible "*They imply a degree of complexity which can in no way be covered by fine level of concordances*" (Knoll 1983). In our project, the evolution of correspondence concept is, in part, the renunciation of the concordance idea as it is. This correspondence must not be interpreted from a term to term relation between a keyword and an other key word. It is only the sign of a possible relation between scientific publications and patents. If the suggested link does not seem to be an aberration and even though it can be particularly relevant, its interpretation cannot clear the step between the probability of a link and an absolute certainty.

The relevance of documentary appraisal indicators

By limiting the definition of this relation as a simple presumption of a link, a few indicators, otherwise very useful and required for the appraisal of documentary information systems, lose in part of their relevance. Considering the most used, famous, reproducible and, at least, the less discussed indicators as recall or precision rates, this relation cannot be more easily measured with usual values. In general, only the

documentary noise indicator remains relevance. However, this indicator cannot show the opening degree supplied by this system, the only useful parameter of its interest degree and of its effectiveness.

Multiple expert evaluations and end users

How can we correctly appreciate this system? Above, a strict documentary control of new suggested relations cannot bring any certainty of relevance. This control can only note that common terms or common locations have effectively been found in the two vocabularies. Concerning the two linked documentary languages, they have been appraised for a long time. Below, each actor implicated by this correspondence excels in its evaluation domain. However it cannot individually appraise the established links. In consequence, a joint validation of multiple evaluations must be elaborated in a common dialogue and a contradictory debate between experts from different domains. Every day the relation between science and technology is established by numerous end-users. This way, always present in technological innovation processes, is their own competency. The main place belongs to end-users: their experience is in the heart of our automatic system. Only this competency can ensure a final consistency in the multiple intermediary appraisals and expert evaluations.

Present perspectives

The third stage of this project would concern a first industrial validation. The last five years of the PASCAL database, more than 2,5 billions bibliographic references, should be treated.

Conclusion

It is well known that the relation between scientific information and technical information is intellectually and strategically complex. The realism of an automatic solution, per se highly reductive, imposes a correspondence mechanism which must be, in the same time, very selective and only based on low level data. So, it is necessary at the same time to avoid an explosion in the number of links and to supply a minimal statistical quality for the new relations. A succession of algorithms allows to select the most frequent vocabularies and classification codes which are hierarchically the most significant. This method resolves this apparent contradiction. On the strict statistic

plane, the concept of the simple presumption of links departs without doubt from the classical notion of relevance. Moreover, it wanders from the bibliometric analysis, otherwise very powerful, to the high and low frequencies. If the simple probability of link generates an obvious documentary noise, this must be drastically reduced. However, it does not exist any justifiable statistical reason, other than cost consideration, to give a greater place to a particular threshold of frequency. So, the interpretation of suggested relations and their selections depend entirely on the decision of end users: are these relations potentially interesting? Such correspondences only allow end-users to explore uncertain paths traced between two conceptually separate universes. But in fact, for other documentary fields commonly included in bibliographic references, do end users have such different criteria and choice procedures? The most important progress is, beyond formal links between scientific references and technical references and through processes based on bibliometrics indicators, the new opportunity to more quickly articulate scientometrical analysis and technometrical analysis together.

References

- Conseil économique et Social, Charzat M. (1994), Les transferts de technologie en matière de recherche industrielle. J.O. Avis et rapport du Conseil économique et Social, 19 : 68.
- CRRM (1997), L'Observatoire – Marseille – Publications. (Online). URL address: <http://crrm.univ-mrs.fr/orps/inist/orps-inist.html>.
- DAHLBERG I. (1981), Towards establishment of compatibility between indexing languages, *International Classification*, 8(2): 86–91.
- DOU H. (1993), Intelligence sociale et veille technologique, *Humanisme et entreprise*, 17:53–64.
- FAUCOMPRÉ P., BALDIT P., DOS SANTOS R., QUONIAM L., DOU H. (1995), Bibliometric tools for bibliographic databases: technological and methodological aspects for relational use of bibliographic databases. In: *Proc. of the Fifth Int. Conf. of the Int. Soc. for Scientometrics and Informetrics*, June 7–10, 1995, River Forest, IL. Medford, NJ: Learned Information, 157–166.
- FEN Z. (1987), Analysis of the properties of the International Patent Classification, *World Patent Information*, 9(3):152–156.
- GUELLEC D. (1993), *Innovation technologique, croissance et compétitivité*, Paris: INSEE.
- HORSNELL V. (1974), *Intermediate Lexicon for Information Science: A Feasibility Study*, London: Polytechnic of North London, School of Librarianship.
- IRIBARNE A. (1987), Les PME face au changement technologique, in: A SILEM (Ed.), *Diffusion des nouvelles technologies, stratégies et effets: Colloque international*, Lyon, 18–19 sept. 1986. Paris: CNRS. 185–208.
- JAKOBIAK F. (1994), *Le brevet: source d'information*, Paris: Dunod.
- KNOLL H. (1983), Observations on concordances, *World Patent Information*, 5(1):19–34.
- LAINÉE F. (1991), *La veille technologique: de l'amateurisme au professionnalisme*, Paris: Eyrolles.
- LAWSON W. (1979), The use of patent information as an indicator of technological output. OCDE: 2nd Workshop on the measurement of R and D output. Paris.

- MANIEZ J. (1987), *Les langages documentaires et classificatoires*, Paris: Ed. d'Organisation.
- RETOURNA C (1995), *Analyse de cas concrets d'innovations dans les PME/PMI: problématiques et discussions*, Marseille: Thèse Univ Aix-Marseille 3.
- RISJBERGEN C.J.VAN. (1978), Automatic classification in information retrieval, *Drexel Library Quaterly*, 14(2): 75–89.
- SCHMIDT-TJEDMANN K. (1982). A new model of the innovation process, *Research Management*, 25: 18–21.
- World Intellectual Property Organisation (1990), *International Patent Classification*. (fifth ed. 1989): general information. Geneva: WIPO.
- ZIMAN J. (1992) A neural net model for innovation, *Science and Public Policy*, 18:65–75.