Scientometrics, Vol. 25. No. 2 (1992) 279-294

A NEW METHOD FOR ANALYSING DOWNLOADED DATA FOR STRATEGIC DECISION

CH. HUOT*, L. QUONIAM**, H. DOU**

* *IBM CEMAP (Centre Européen de Mathematiques Appliquées), 68-76 quai de la rapée, 75592 Paris CEDEX 12 (France)*
** *CRRM Faculte de St Jérome, 13397 Marseille CEDEX 13 (France)*

(Received July 16, 1991)

Technology assessment survey is nowadays a specific and scientific subject that any manufacture needs for increasing productivity. This function was initially reserved to experts of the studied field. But the increase of information volume has called for a change. Now, we need specialists of technology assessment survey which know about sophisticated methods to extract strategic information from downloaded data. We will explain how to build strategic information. We present here a new and original method of data analysis. This Factorial Relational Analysis is born after 15 years of IBM France mathematics research center works on qualitative data analysis. The method is based on Relational Analysis. The particularity of this method is to work with sparse matrices and to obtain the best classification without any *a priori fixation of number of classes.* Relational Analysis is used in other sectors than the analysis of matrices issued from downloaded data. For example it is also used in computational lexicography or in credit scoring or in any domain where classification is concerned. Here we choose to present an example of an application in patent analysis.

Introduction

To face the "technological war" that has begun all around the world, without any exempted country, it is crucial for the chief executives to be always very well informed on the few subjects that could have great consequences about their decisions to be always more competitive. These few subjects have been called by *Rockart*[1] the "Critical Success Factors" (CSF). As soon as defined, these "CSF" must be overlooked by specialists of information retrieval. To do this they need acess to international specific databases. As soon as they have found the database, they can query and obtain a lot of bibliographic references that would provide the requested information. The problem nowadays is that the amount of available information increases[2] in such a rate that a lot of problems occur during information retrieval.

– First there are problems concerning the construction and use of databases.

– Second, after getting information, it is more and more difficult to analyze the information with accuracy (the brain ability is rather constant for a continual increase of the amount of information).

In this paper the second point will be developed and an application to a case study concerning patents analysis will be treated. We choose patents information because of its major importance in manufacture strategy. The database used was the WPIL (World Patents Index Latest) which contains patents issued after 1981. The request was about the problem of cleaning the contact lens by a chemical process (more specially with papain which is a proteolitic enzyme). The exact query formulation was:

"CONTACT LENS AND PAPAIN?"

The answer gave 4 patents and we added the 9 patents that contained one of the 4 formers in their citation field (all the patents provide information upon the technology and the technologies which use it). The analysis treats those 13 patents. We chose an application with few patents because of space problem but the method we use could be applied to a large number of patents (> 1000).

Basic information

Once the query is formulated on the host, it is possible to download the data: it means that one gets the data in your laboratory, on a microcomputer. Afterwards, in the laboratory you may analyze the bibliographic data the way you want (Post Processing of Online search: PPOS Concept[3]). The bibliographic data you get are divided into fields, each of them having a specific mean as shown on Table 1.

Table 1
Constitutive fields of a bibliographic reference and their meanings

-1-		
AN	-89-354640/48	Accession Number in the database
TI	-Wall panel - has slits joining inter-pane space to interior, made in top of window frame, and slit joining it to ventilation cavity	Title of the patent
DC	-Q44 Q48	Derwent Codes
PA	-(EVEN/) EVENTOV V S	Patent Owner
IN	-EVENTOV VS	Inventor
PN	-SU1479589-A 89.05.15 (8948)	Patent Number
PR	-86.05.08 86SU-075778	Priority Number

There are two ways to look at the references:[4]

– To consider them as bibliographic tools and then whatever combination of the fields you make, the information will always be a bibliographic reference.

– To consider them as a sum of specific strategic information and in this case combination and analysis of constitutive fields provide strategical information (like the frequency of patents owners). Strategic information is a part of information for IHDS[5] (Interactive Helps for Decision Systems, SIAD in French).

Many authors working on bibliometrics (reviewed in Ref. 6) and scientometrics all around the world already use this concept with different methods to analyze separated fields.

In our case study, we developed the Derwent codes field analysis. We chose this field for several reasons:

– Codes are interesting to study because they provide short, concise information.

– Figures processing is easier and faster than word processing.

– More important, codes are independent from linguistic vogues, almost not related to from period considerations (changes in code signification are uncommon) and independent from space considerations (the codes affectation is the same for US patents and USSR patents).

– Codes are quite "objective" when you consider that the person who abstracts a large number of patents has surely the best overview on a subject.

So they are very powerful for quick analysis.

The Derwent code classification is made of 330 codes which can be divided into 8 non equal parts (cf Table 2).

Table 2
Number of codes by section

Section	Number of codes
PLASDOC	41
FOODDET	13
ELECTRIC	44
FARMAG	11
CHEMDOC	56
GENERAL	46
MECHANIC	70
SX-ELECT	49

The detailed exhaustive list of the codes for this analysis is given in the Table 4.

Different ways to analyze a specific field

To analyze a specific field there are many methods that we are going to overview in an increasing complexity order.

Frequency analysis

First of all it is possible to count the frequency of each constitutive code. The result concerning our data is given in Table 3.

Table 3
Frequency of each Derwent codes in our downloaded data

Derwent code	Frequency
D22	12
D16	9
P34	7
A96	6
P81	5
E36	4
P43	4
R16	3
E17	2
E19	2
E37	1
D25	1
S05	1
A97	1
P31	1

This result provides a scan over three different zones:

1. *Evident information*: this is information which is present in almost all references (high frequency codes). It just gives information about the subject we work on.
2. *Potentially innovative information*: medium frequency codes which provide information only present in some references and that could be specific to some new technical consideration.
3. *Noisy information*: so low frequency codes that it is impossible to say if they are accidental data or it is real innovation. Usually this part is the most important but not in our sample because of the few patents we used.

In our case we could divide in codes D22 to P81; E36 to E19; E37 to P31. It is possible to determine graphically the frontiers between zones. Many authors worked on these aspects of information retrieval that could be named "Zipf or Bradford or Lotka or Informetrics" distributions.[7-9]

Presentation of downloaded data

Table 4
Derwent codes

Letter	Signification
A	Codes PLASDOC
D	Codes FOODDET
E	Codes CHEMDOC
P	Codes GENERAL
R	Codes ELECTRIC
S	Codes SX-ELECT

D22	Desinf, deter., dental, sterilizing, bandages, sutures, plaster casts, prostheses (lens).
P34	Health, amusement, sterilizing, syringes, electrotherapy
D16	Food, fermentation industry, brewing, yeast, pharmaceuticals alcohol
P81	Optics, photography, general. optics
A96	Veterinary, medical, dental.
E36	General inorganic, none-metallic elements
R16	Measuring, testing, investigating chem./ phys. prosp.
E17	General organic, other aliphatics
D25	Desinf. deter. soap. including metal salt and fatty acids used in soaps
P43	Separating, mixing. sorting, cleaning
S05	Electromedical
A97	Miscellaneous goods
E19	General organic, other organic compounds general
P31	Health, amusement. diagnosis, surgery
E37	General inorganic, mixtures of many components

Pairing techniques

A second aspect of the exploration of a specific field is to analyze the relations between the different constituents. More than knowing which codes are present, it is important to see the connections between the different techniques. It introduces a multidimensional view of the problem. Links may be studied in different ways. The simplest one consists of counting the links and presenting them on a graph as shown in Fig. 1.

The permuted pairs are equivalent. This graph is easy to construct but has two major drawbacks:

– It is not easy to represent the differences in pairs frequencies.

– The pairs frequencies are not relevant to the importance of the pairs.

Let us examine the example in Table 5.

Table 5
Frequency of pairs

Code	Frequency	Pairs	Frequency
A	20	AD	20
B	100	BC	30
C	200		
D	25		

In the example the (BC) pair seems to be more important (higher frequency) but in fact the (AD) pair is much more important (quasi 100% of the constitutive codes are engaged in the pair formation).

Another inconvenience is the difficulty to draw the graph for a very large amount of downloaded references that leads us to cut and draw only high frequency pairs, but we will explain later that the low frequency pairs are not inevitably interesting. These inconveniences are largely balanced by the easiness and rapidity of the method (few minutes for automatic treatment of 1000 references on a micro computer). The results concerning our case study are shown in Fig 1. On the graph, we see the different links between codes. We can divide these links in several parts:

– Links between the main frequency codes (that we defined as evident information) which are not necessarily interesting (seem to be evident links). They explain the subject we work on.

– Links between high frequency and medium frequency codes (that we define as innovative information) or between the medium frequency codes and themselves which are relevant to interesting links (strong or not) because of their potential innovative aspects.

– Links between any frequency codes and low frequency codes which are generally not drawn due to the fact that they complicate the graph with potentially noisy information (constitutive codes may be noisy).

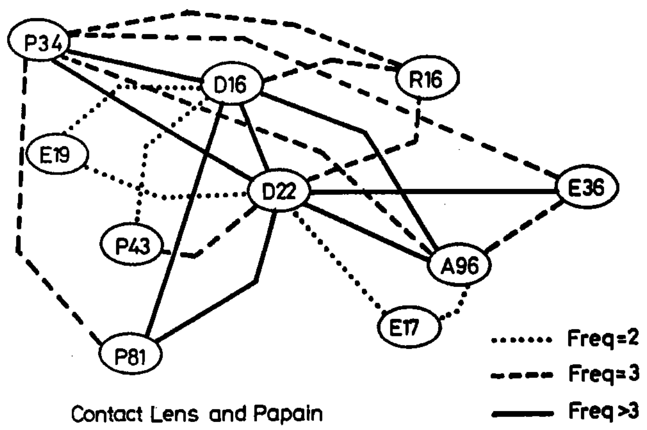

Fig. 1. Links between Derwent codes in our downloaded database

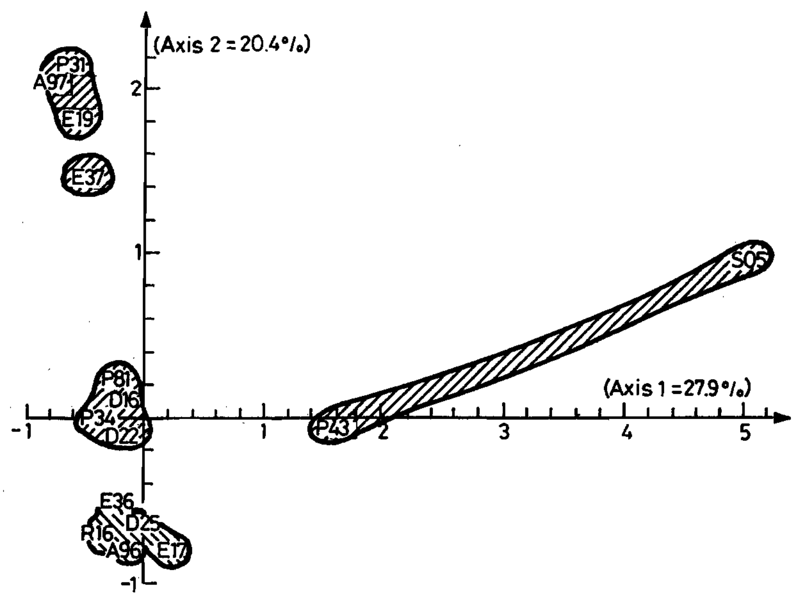

Fig. 2. Graph of the factorial relational analysis applied to our downloaded data

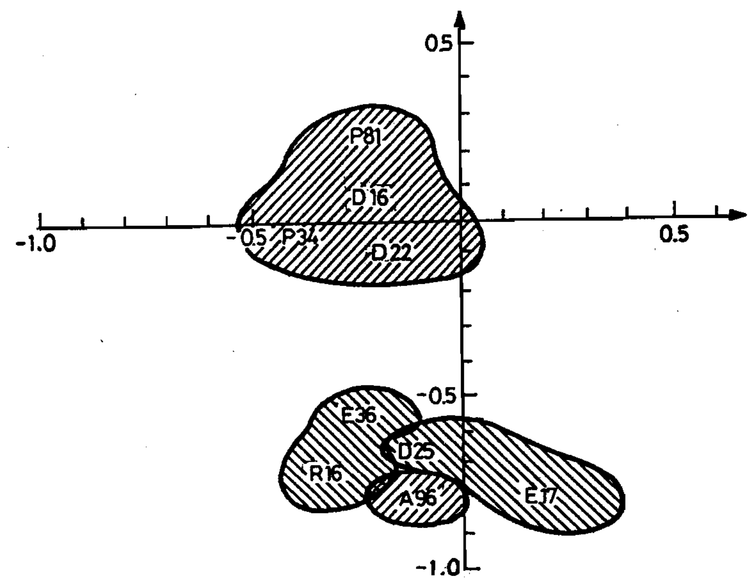

Fig. 3. Zoom of the graph central part

Matrix building

To overcome the drawbacks of pairing analysis, many authors build up matrices and analyze the result with classical data analysis methods, which provide metric solutions.

The results dependant, in part, on the matrix analyzed:

– *"Co-frequencies" matrices*: are matrices that contain in each row column intersection the of co-frequency of two concepts, one is a row header, the other a column header. This kind of matrix is symmetric if row header are equal to column headers.

– *"Frequencies" matrices*: are built with the concepts of column headers and the reference numbers as row headers. If just considering the presence of a concept, the matrix will only contain 1 for presence and 0 for absence (in our case study, we built up such a matrix which is given in Table 6). This former matrix form is called a "presence/absence matrix" and is always used for code analysis (each code is present only once by reference).

It is important to notice that a symmetric co-frequencies matrix is computable from a frequencies matrix (it is simply a Burt matrix[10]). The bibliographic references matrices are very "sparsed," i.e. they contain few numbers (due to information properties developed when we described codes frequencies). This consideration is very important to understand the problems in using classical data analysis.[11] Building a matrix must always be an automatic process regarding the amount of data and possible human errors.

Classical methods of data analysis

There are two families of data analysis which are complementary and often consecutively used.

(1) *Inertia analysis*: they represent the whole matrix in a reduced space obtained by computation of eigenvalues and eigenvectors of a calculated distance matrix. It introduces a simplification of the information but an increase of significance. The problem is that the reduced space must be small. The sparser the matrix, the more the reduced space is similar to the original one; so graphs are numerous and analysis difficult to understand[12] (dispersion concept). In spite of this major restriction several authors use this type of analysis to provide strategic information from downloaded data.[13,14] Some others overcome this drawbacks by grouping variables[15] but we think it induces too much important loss of information to be an efficient method. Factorial analysis is an example of inertia analysis. An other example of recent inertia analysis is the quasi-correspondence analysis (QCA) used in Ref. 16 and described in Refs 17 and 18.

(2) *Classification analysis*: are techniques that classify either rows or columns of a matrix using an aggregation criterion over a computed distance. They are used separately[19] or as a complementary tool for the interpretation of inertia analysis[20] (multidimensional scaling and cluster analysis).[21] In this former case the classification is calculated with the position in the reduced space. At the end, classes repartition is available with a major restriction: you must specify the amount of classes before standing the process. Ascendant Hierarchical Classification (AHC) is an example.

All the restrictions mentioned, induced the co-word approach.[22,23]

(3) *Co-word analysis method*: The authors calculate an AHC but instead of working in a matrix space they use a chained-word notion. They solve the problem of class number determination with a special representation of the results. According to us the problem seems to be about the classification criterion choice which is not the most impressive.[24,25]

Relation Analysis

The mathematic department of the IBM Scientific Center of Paris has been working on a new method of qualitative data classification for 15 years. This method named Relational Analysis was developed by *Marcotorchino* and *Michaud*.[26-28] Relational Analysis groups together a pull of techniques to modelize and solve problems defined like:

"Find a structured relation Y which is the closest to a set of any relation R".

The method keeps data under a relational form and modelize the problem by using linear programming.

Data representation, basic tables

To build the different matrices used by the methodology we define

1. N = number of objects,
2. M = number of variables,
3. P = total number of modalities.

With these elements we define 3 tables.

(1) The complete disjonctive table K

The table K (dimension = $N \cdot P$) has for general term k_{ij} with:

$$k_{ij} = 1 \text{ if } i \text{ "is relation with" } j$$

$$k_{ij} = 0 \text{ otherwise}$$

$$\sum_{j=1}^{P} k_{ij} = M$$

$$\sum_{i=1}^{N} \sum_{j=1}^{P} k_{ij} = M \cdot N$$

Factorial Relational Analysis is a part of the Relational Analysis developed by *Marcotorchino* since 1989.[29] This method does not use the original *"Condorcet Criterion"* but weighted criteria. These criteria were introduced for different notions of classification, and to create a bridge, mathematically validated, between Multiple Correspondence Factorial Analysis and Relational Analysis.

(2) The weighted Condorcet table

The weighed criteria allow for the introduction of weights in the object comparisons; they are based on the traditional "presence-rareness index."

The presence-rareness index:

It is because of the data nature we need that we used weighted criteria. The whole data are very poor in information (there are many 0 in the matrix), so we need a measure which reflects this phenomenon of "presence-rareness." By "presence-rareness" index we mean a similarity index S(x,y), between two objects x and y:

$$S(x,y) = N(x,y)/D(x,y) \text{ varying from 0 to 1.}$$

The numerator equals 0 or 1 depending if x is similar or not with y, the denominator D(x,y) equals number of objects y with:

$$N(x,y) = 1, \text{ so } D(x,y) = |\{y|N(x,y) = 1\}| \text{ (cardinal set)}$$

$$\text{So } s(x,y) = 1, \text{ if x is alone and } s(x,y) \leq 1 \text{ if } N(x,y) = 1 \text{ et } D(x,y) > 0.$$

If $N(x,y) = 0$: $\forall$ the denominator value, $S(x,y) = 0$ That is the origin of it's name "presence-rareness" index, because it measures the presence of a similarity according to it's rarity. The principle of a such measure is to consider two objects all the more similar than they are rare in the studied population.

We consider that two objects are very similar when they share a characteristic which is rare in the population to classify.

The matrix $\hat{C}$ of pairwise comparisons between objects (using "presence-rareness index") has as general term:

$$\hat{c}_{ii'} = \sum_j k_{ij}k_{i'j}/k_{\bullet j}$$

where

$$k_{\bullet j} = \sum_i k_{ij} = \text{number of objects which possess the form } j.$$

The similarity $\hat{c}_{ii'}$ between two objects i and i' will be as greatest as they share (k_{ij} $k_{i'j}$) rare forms (divided by $k_{\bullet j}$).

The weighted Condorcet criterion, based on this similarity, takes the following form:

$$\hat{C}(X) = \sum_{i\in I} \sum_{i\in I} (\hat{c}_{ii'}-\bar{\hat{c}}_{ii'})x_{ii'} + \sum_{i\in I} \sum_{i'\in I} \bar{\hat{c}}_{ii'}$$

with

$$\bar{\hat{c}}_{ii'} = \hat{c}_{ii}+\hat{c}_{i'i'}/2 - \hat{c}_{ii'}$$

(3) The weighted Burt table

When analyzing the modalities of classification using the presence rareness index the data matrix we process is the B matrix, defined as follows:

$$b_{jj'} = \Sigma \quad k_{ij}k_{ij'}/k_i\bullet$$

where $k_i\bullet = \Sigma \ k_{ij}$ = number of modalities of the object $i = M$ (number of variables).

So the weighted Burt criterion takes the form:

$$B(Y) = \Sigma \ \Sigma \ (b_{jj'}-\bar{b}_{jj'})y_{jj'} + \ \Sigma \ \Sigma \ \bar{b}_{jj'}$$

with

$$\bar{b}_{jj'} = b_{jj}+b_{j'j'}/2 - b_{jj'}$$

It is precisely on the use of both weighted Burt criterion and weighted Condorcet criterion Relational Factorial Analysis is based on.

Factorial Relation Analysis (Figs 2 and 3)

Unlike to other non hierarchical clustering method based upon inertial criterion, the relational analysis methodology does not oblige to fix *a priori* the number of classes of the solution.

According to the Huyghens principle, we know that the total inertia of a partition P noted I_t is the sum of its within cluster inertia $I_W(P)$ and its between clusters inertia $I_B(P)$:

$$I_T = I_B(P) + I_W(P) \forall P$$

This is a well known result to maximize $I_B(P)$ or to minimize $I_W(P)$. Also a trivial solution exist if we do not place constraints upon the number of clusters, the solution consist of a partition in N clusters, where all the objects are isolated. In the case of qualitative data and for a "trivial" partition $P_{iso,}$ solution of the inertia maximization problem, without constraints upon the number of clusters, the between cluster inertia keeps the value P/M-1, with:

$$I_T = I_B(P_{iso}) = P/M - 1 \text{ and } I_W(P_{iso}) = 0$$

This result, well-known the AFCM's specialists (Analyse Factorielle des Correspondances Multiples), explains why it is necessary to fix number of clusters when using an inertial index in non hierarchical cluster analysis. That is the reason why *Marcotorchino*[30,31] proposed to keep the inertial properties to define a criterion with "natural intuitive properties" but which does not require the fixation *a priori* the number of clusters.

The within and between cluster inertia can simply be written with relational notions, as follows.

$$I_B(P) = \sum_i \sum_{i'} \hat{c}_{ii'}/M\, x_{ii'}/x_i\bullet\text{-}1$$

and

$$I_W(P) = P/M\text{-}1\text{-}I_B(P)$$

where X is the binary relational matrix of the equivalence relation searched (unknown partition) and *C* the weighted Condorcet matrix.

In multiple correspondence factorial analysis the processed matrix, (*AF*), has for general term:

$$AF_{ii'} = \sum_i k_{ij}k_{i'j}/\sqrt{k_i\bullet}\sqrt{k_{i'}\bullet}k\bullet_j = 1/M \sum_j k_{ij}k_{i'j}/k\bullet_j$$

So

$$AF_{ii'} = \hat{c}_{ii'}/M$$

The weighted Condorcet matrix is the basic matrix of the multiple correspondence factorial analysis. In practice a multiple correspondence factorial analysis consists in projecting points on the first factorial axis (r = 1, 2, 3,...) and generally making the interpretation for small values of r. One qualitative indicator of an analysis is the percentage of inertia explained by the primary axis:

$$\lambda_1+\lambda_2+\lambda_3+\ldots\lambda_r/ \sum \lambda_i$$

In AFCM, $I_T = \sum_{j=1}^{q} \lambda_i$ (where q is the number of eigen-values non equal to 0) and if we note $I_F(r)$ the inertia explained by the r first axis we would like to compare $I_F(r)$ and $I_B(P)$. If $I_B(P)$ is greater than $I_F(r)$ we are in a configuration where the partition P provides "more information" on the data structure than the AFCM does.

Maximizing $I_B(P)$ is equivalent to maximize $\sum_{i}^{N} \sum_{i'}^{M} \hat{c}_{ii'}/M\, x_{ii'}/x_{i}\bullet\text{-}1$, and this inertia (if we do not add constraints) is maximal and equal to I_T when all the points are separated.

To avoid the simple response presented previously it is necessary to find a partition without fixing the number of clusters, and which is the nearest from the inertia research. This explain the use of the weighted Condorcet criterion, because it generally gives an inertia greater than $I_F(r)$ when r is small. So we obtain a clustering without fixing hypothesis "a priori," and we can surround, on the factorial graph, realist clusters given by the classification, and compatible with the factorial data analysis, because the cost $(\hat{c}_{ii'}/M - \bar{\hat{c}}_{ii'}/M)x_{ii'}$ of the weighted Condorcet Criteria works on the same data as the AFCM (Table 6).

Table 6
Initial matrix

	D22	P34	D16	P81	A96	E36	R16	E17	D25	P43	S05	A97	E19	P31	E37	TOTAL
US4749511	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	4
US7757014	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	6
US4609493	1	0	1	1	1	0	0	1	1	0	0	0	0	0	0	6
US4614549	1	1	1	1	0	0	0	0	0	1	0	0	0	0	0	5
US4872965	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2
US4839082	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2
US4832754	1	0	1	0	0	1	0	0	0	1	0	0	0	0	0	4
US4808239	1	0	0	0	1	0	0	1	0	1	0	0	0	0	0	4
US4767559	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	3
US4784790	1	1	1	0	0	0	0	0	0	0	0	1	1	1	0	6
US4829001	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	6
US4855234	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	6
US4710313	1	0	1	1	0	0	0	0	0	0	0	0	1	0	1	5
Total	12	7	9	5	6	4	3	2	1	4	1	1	2	1	1	59

With references we can build a matrix crossing the patent numbers fields and the Derwent codes fields. This matrix is the basic table called K in our relational notations. This table is used to create the $\hat{B}$ and $\hat{C}$ table. Our purpose is not to analyze the result of the analysis, because it is the job of experts in contact lens and chemical or enzymatical treatment. But we can say, when we look at the graphics that the partition we have on our projection, does have a logical explication when analyzing primary data. For example, codes P43 and S05 are in the same cluster because only one patent has in its description the couple (P43, S05). And if we just analyze the AFC projection in the first two axes it is impossible to reach this conclusion.

Conclusions

This case study was developed to show how difficult the analysis of downloaded information can be. This is ue to information properties and to the final user of the information which is usually an expert of the analyzed subject or a chief executive but not an expert in information science. This point confirms that technology assessment is a real interface science which needs experts in information science whose job is to look for sophisticated methods, test and use them. The results must then be presented to the chief executive in a literal form. The other important point is to show that nowadays new computerized tools exist for solving what was considered a problem not so long ago. We do hope to go on with such operations.

*

The authors would like to thank the referees for useful suggestions and help concerning references, and they thank Orbit Information Technologies for providing access to various database files.

References

1. J.F. Rockart, Chief executive define their own data needs, *Harvard Business Review*, (1979) 03-04.
2. H. Dou, P. Hassanaly, How to use online databases as a tool for forecasting fundamental and applied research. 7th International Online Information Meeting, London England, *Learned Information*, Oxford, 1983, p. 175-183.
3. H. Dou, P. Hassanaly, L. Quoniam, A. La Tela, Post processing of online search, *Bladen voor de documentatie*, Bruxelles, 3 (1990) 51-70.
4. H. Dou, P. Hassanaly, L. Quoniam, A. La Tela, Veille technologique et information documentaire, *Documentaliste*, Paris, 27 (3), 05-06 (1990) 132-141.
5. A La Tela, *Systemes Interactifs d'Aide à la Décision (S.I.A.D.)*, Doctoral Thesis, Marseille, 1987, pp. 96.
6. H.D. White, K.W. Mc Cain, Bibliometrics, *Annual Review of Information Science and Technology*, 24 (1989) 119-186.
7. L. Egghe, The duality of informetric systems with applications to the empirical laws, *Journal of Information Science*, 16 (1989) 17-27.
8. T. Lafouge, Une autre approche de la circulation de l'information, *Revue Française de Bibliométrie*, 7 (1990) 1-13.

9. B.C. Brookes, Ranking techniques and the Empirical Log Law, *Information Processing and Management*, 20 (1984) 37-46.
10. J.P. Benzecri, *L'analyse des Données: L'analyse des correspondances*, Dunod, Paris, 1980, pp. 616.
11. L. Quoniam, H. Dou, C. Huot, Les Méthodes d'analyse des données face à l'information stratégique et l'innovation, *Colloque International de Sériation par Bloc*, Strasbourg, 1990, p. 04.
12. B. Escoffier, J. Pages, *Analyses factorielles simples et multiples*, Dunod, Paris, 1988, pp. 241.
13. C. Dutheuil, Du corpus documentaire à l'interprétation des résultats de l'analyse des données, *Revue Française de Bibliométrie*, 6 (1990) 90-104.
14. J.C. Dore, J. Gilbert, J.F. Miquel, A. Deroulede, C. Dutheuil, Banques de données et analyses multivariées, *Congrés CNIC*, Paris, France, 1986.
15. C. Paoli, P. Billard, P. Blanchet, C. Longuevialle, Apport de l'analyse factorielle et de la classification ascendante hiérarchique dans l'analyse des banques de données bibliographiques *Congrés S.F.B.A.*, Les systémes d'information élaborée, Ile Rousse, France, 1987, p. 65-75.
16. L. Leydesdorff, R. Zaal, Co-word and citations relations between document sets and environments, *First International Conference on Bibliometrics and Theoretical Aspects of Information Retrieval*, Luc, Belgium, 1987, pp. 31.
17. A.F.J. Van Raan, H.P.F. Peters, Dynamics of a scientific field analysed by co-subfield structures, *Scientometrics*, 15 (1989) 607-620.
18. R.J.W. Tijssen, J. De Leew, A.F.J. Van Raan, Quasi-correspondence analysis on scientometric transaction matrices, *Scientometrics*, 11 (1987) 351-366.
19. R.J.W. Tijssen, J. De Leew, A.F.J. Van Raan, A method for mapping bibliometric relations based on field-classifications and citations of articles, in: *Informetrics 87/88*, L. Egghe, R. Rousseau (Eds), Amsterdam, Elsevier Science Publishers, 1988.
20. R. Todorov, M. Winterhager, Mapping Australian geophysics: a co-heading analysis, *Scientometrics*, 19 (1990) 35-56.
21. C. Dutheuil, Analyse factorielle et classification automatique appliquées aux brevets, *Colloque CNIC*, Paris, France 1987, pp. 20.
22. A. Rip, *Mapping of Science, Possibilities and Limitation, Handbook of Quantitative Studies of Science and Technology*, North-Holland, Amsterdam, 1988, p. 253-273.
23. J.P. Courtial, Introduction à la scientométrie, *Anthropos*, Paris, France, 1990, p. 141.
24. R. Rousseau, Classification ascendant hiérarchique et décomposition en blocs diagonaux, *Colloque International de Sériation par Bloc*, Strasbourg, France, 1990.
25. G. Celeux, E. Diday, G. Govaert, Y. Lechevalier, H. Ralambondrainy, *Classification automatique des données*, Dunod, Paris, 1989, pp. 285.
26. J.F. Marcotorchino, P. Michaud, *Modéles d'optimisation en analyse des données relationnelles*, Gauthier-Villars, Paris, 1979, pp. 312.
27. J.F. Marcotorchino, A unified approach of the block seriation problems, *Journal of Applied Stochastic Models and Data Analysis*, 3 (2) (1987).
28. C. Bédécarrax, *Classification automatique en analyse relationnelle: La quadri-décomposition et ses applications*, Doctoral Thesis, Paris VI, France, 1989.
29. J.F. Marcotorchino, Liaison Analyse Factorielle-Analyse Relationnelle (I) Dualité Burt-Condorcet, *Etude du Centre Scientifique IBM France N° F142'*, pp. 54, (1989)
30. J.F. Marcotorchino, L'Analyse Factorielle Relationnelle (Partie I et II), *Etude du CEMAP IBM France N° MAP-003'*, (1991).
31. N. El Ayoubi, *Liaison Analyse Factorielle-Analyse Relationnelle: Extension*, Doctoral Thesis, Paris, France, 1990.