World Patent Information, Vol. 12, No. 4, pp. 222–229, 1990.
Printed in Great Britain.

0172-2190/90$3.00+.00.
Maxwell Online Inc.

Competitive Technology Assessment. Strategic Patent Clusters Obtained with Non-boolean Logic. New Applications of the GET Command

Henri Dou, Parina Hassanaly, Luc Quoniam and Albert La Tela, Centre de Recherche Rétrospective de Marseille, Centre de st Jérome, Université Aix Marseille III, 13397 Marseille cédex 13, France

Summary

Most of the online searches process through combinations made with boolean operators. We wish to show, in this paper, that it is very easy to search online databases differently. This concept involved searches grounded to non-boolean concepts.

Introduction

Online databases are mainly used for bibliographic searches, but other applications are possible. One of these is the usage of online databases in the process of Competitive Technology Assessment.[1, 2]

Today, the competition is so fierce, and the economic considerations are so important, that technology becomes a key factor in organization strategies. Rockart[3] using the Critical Factor Success showed that the information system of an organization is a very important factor in success. This information system deals with various types of information, such as those publicly available or the fuzzier ones difficult to obtain and classify.

The public databases, because they are growing every year, and because they give a very large overview of science and technology, are one of the best means to input data into the strategic information system of an organization.

This prompts us to report in this paper, various non-classical ways to deal with online databases.

What is the Need

The need is to provide for the experts, various clusters of information, where the data are not necessarily linked by the same keywords or index-terms. The clusters must deal with data which are sometimes linked by the same concept, but not necessarily by the same subject terms.

Citations are one of the best ways to build up such sets of references. For instance using the databases USPA or USPB (US patent database, exclusive on Orbit Search Service) it is possible for one set of patents to retrieve all the cited patents or all the fundamental cited references.[4] This will link the set of patents to earlier patents or to fundamental literature. In the same way, using the Science Citation Index database, it will be possible to retrieve for a set of fundamental papers the cited references, and among them the cited patents.

Citations provide a link between fundamental research and technology, and clusters of patents or of fundamental papers may be built up using this technique.[4, 5]

Because we are concerned with strategy trends, it is important to select firms which are dealing with a core of common subjects. This means that we want to retrieve from a patent database for instance the Patent Assignees which are all working in the same core of subjects, but these subjects will not appear at the same time within the same patent or publication. The retrieval of such clusters is not possible with boolean operators.

How to Solve this Problem

Example of information to be obtained

What are the firms dealing with products (or concepts) A, B and C. These products are not necessarily present in the same patent or publication. Figure 1 illustrates this concept.

If we try to solve this problem using boolean operators within the same database, the results show that it is

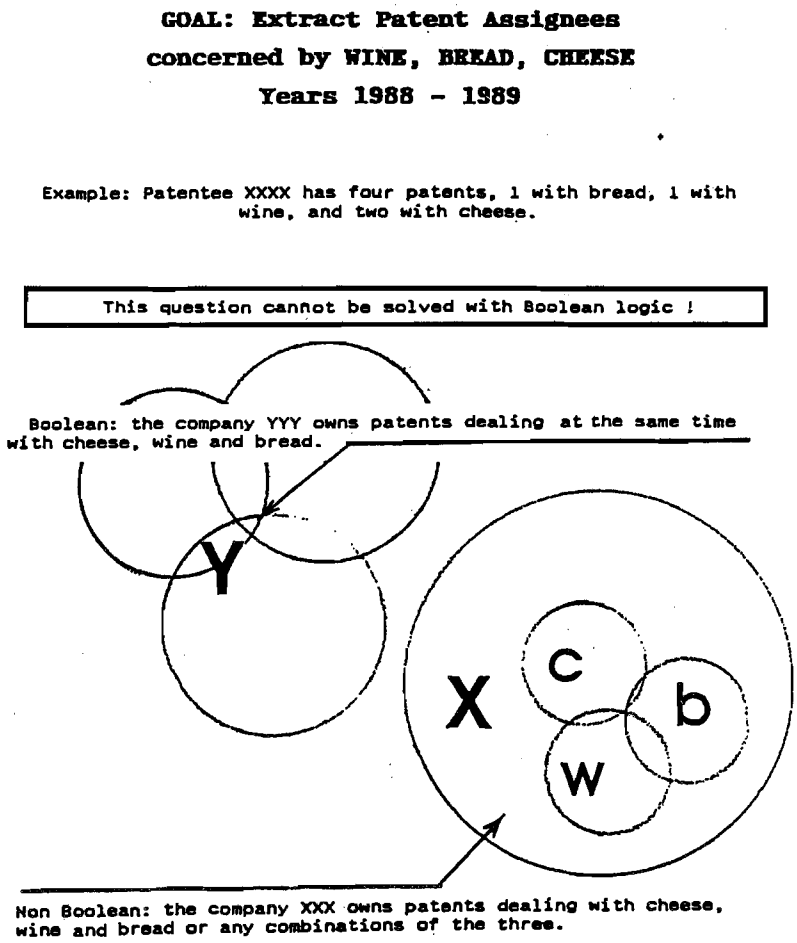


Fig. 1. Stategic clusters from non-boolean searches.

not possible to retrieve the cluster of firms concerned at the same time with the subjects A, B and C since they are not necessarily present in the same reference.

The use of the AND operator gives only references where the data A, B and C are present at the same time, and the OR operator gives a very large set of firms dealing either with one, two or three of these data.

The solution

To solve this problem we are going to use a combination of several features available online, and then to process off line the results.

(i) Select all the references dealing with A, B and C. This gives three sets of posting terms.

(ii) Make a GET[6, 7] on each of the answers. The GET command enables the user to select a field, and to sort and rank each term of this field, in our case the Patent Assignee, in decreasing order. The files obtained with each of the GET will be saved on your computer.

(iii) Make the comparison of the two first GET files to select the common Patent Assignees. Save this file. This comparison is made with DATAGET v 3.0 developed and available from the CRRM.*

(iv) Make the comparison of this former file with the last GET file. This will provide the result. The different steps of the post processing will be illustrated in the following examples.

Examples of Results

Note that the comparisons of one GET file versus another affords the user to obtain in term of frequencies the advantages of one firm versus the other. This comparison in terms of partial or full advantages is common place in business analysis.[9]

*DATAGET is a part of a cluster of softwares designed by the CRRM (Centre de St Jérome, 13397 Marseille, cedex 13, France). These softwares perform various automatic literature analyses.

```
PROG:
ELAPSED TIME ON ORBIT: 0.02 HRS.
YOU ARE NOW CONNECTED TO THE WPI LATEST DATABASE.
COVERS BASICS 1981 - DATE: UPDATED TO
8940/UP;8940/UPEQ,8927/UPA,8912/UPB.


SS 1 /C?
USER:
~bread# and wine# and cheese#

PROG:
SS 1 PSTG (1)

SS 2 /C?
USER:
~prt

PROG:

-1-
AN  - 88-096323/14
XRAM- C88-043414
TI  - Rat attractant - contains alcohol e.g. sake, wine
      etc. impregnated into food
DC  - C03 P14
PA  - (MURA/) MURAKAMI H
NP  - 2
PN  - J63048205-A  88.02.29 (8814) (JP)
      CN8704002-A  88.02.10 (8913)
PR  - 86.08.19 86JP-193677  86.06.03 86JP-U84682  86.06.10
      86JP-135192  86.07.30 86JP-V17378
AP  - 86.08.19 86JP-193677

SS 2 /C?
USER:
~wine# or bread# or cheese#

PROG:
SS 2 PSTG (7568)
```

Fig. 2. Boolean operators are not sufficient to solve the problem.

```
SS 8 /C?
USER:
~bread#

PROG:
SS 8 PSTG (2800)

SS 9 /C?
USER:
~cheese#

PROG:
SS 9 PSTG (2496)

SS 10 /C?
USER:
~wine#

PROG:
SS 10 PSTG (2390)

SS 11 /C?
USER:
~8 and 88-89

PROG:
SS 11 PSTG (691)     ----> GET PA RANK TOP 1000 --> 511 PA

SS 12 /C?
USER:
~9 and 88-89

PROG:
SS 12 PSTG (569)    ----> GET PA RANK TOP 1000 --> 444 PA

SS 13 /C?
USER:
~10 and 88-89

PROG:
SS 13 PSTG (494)    ----> GET PA RANK TOP 1000 --> 479 PA
```

GET

Fig. 3. How to solve the problem.

```
Terms common to files: getwine and getbread
Terms from file: getwine are printed first. Key=  20

 4      TAKEDA CHEMICAL IND KK
 1      TAKEDA CHEMICAL IND KK
 3      QP CORP
 1      QP CORP
 3      TOPPAN PRINTING KK
 1      TOPPAN PRINTING KK
 2      HEINEKEN TEC BEHEER NV
 1      HEINEKEN TEC BEHEER NV
 2      JGC CORP
 1      JGC CORP
 2      KIKKOMAN CORP
 4      KIKKOMAN CORP
 2      MATSUSHITA ELEC IND KK
12      MATSUSHITA ELEC IND KK
 2      MOSC FOOD IND TECHN
 7      MOSC FOOD IND TECHN
 1      AJINOMOTO KK
 4      AJINOMOTO KK
 1      AS SIBE IRKUT ORG CHEM
 1      AS SIBE IRKUT ORG CHEM
 1      GENERAL FOODS CORP
 2      GENERAL FOODS CORP
 1      HOUSE SHOKUHIN KOGY
 2      HOUSE SHOKUHIN KOGY
 1      LIVE INT KK
 1      LIVE INT KK
 1      MAGARACH WINE PROD
 1      MAGARACH WINE PROD
 1      MURAKAMI H
 1      MURAKAMI H
 1      ODESS FOOD IND TECH
 1      ODESS FOOD IND TECH
 1      SAN EI CHEM IND KK
 1      SAN EI CHEM IND KK
 1      SAPPORO BREWERIES
 4      SAPPORO BREWERIES
 1      SHOWA DENKO KK
 1      SHOWA DENKO KK
 1      TAKAHASHI N
 1      TAKAHASHI N
 1      TAKASAGO INT CORP
 1      TAKASAGO INT CORP
```

Fig. 4. Comparison of GETWINE and GETBREAD (WPIL PA 12/06/1989).

(i) *Firms dealing with BREAD, CHEESE and WINE*

The GET command has been processed with the PA (Patent Assignee) field on the WPIL database.

The results are summarized in Figs 2–4.

Figure 2 shows that if you used the boolean operator AND, the result is the selection of only one patent, where the terms BREAD, CHEESE and WINE are present. This is not the answer to our question, since we want to obtain the Patent Assignees which deal with BREAD, CHEESE and WINE, but not necessarily in the same patent. The same figure shows also that if you used the boolean operator OR, you obtain several thousand references. This is not the answer again, since Patent Assignees dealing with one, two or three of the terms are also selected.

To obtain the Patent Assignees concerned with CHEESE, BREAD and WINE it is necessary to use two techniques: the power of the GET command and the post processing of its results. This is explained in Figs 3–5.

Figure 3 shows the number of posting terms selected for each word (the search has been limited to years 1988–1989 for this example). For each answer, the command GET PA RANK TOP 1000† has been performed on line, and the resulting answers have been saved. This leads to 511 PA (Patent Assignee) for the word BREAD, 444 PA for the word CHEESE and 479 PA for the word WINE.

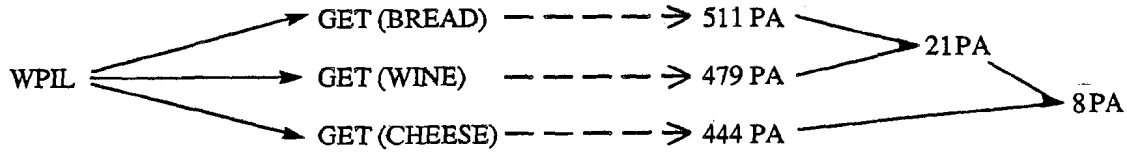
These three files are processed with DATAGET, a software developed in our laboratory and dedicated to the comparison of the GET results. First the file containing the PA dealing with BREAD is compared with the file containing the PA dealing with WINE. The output is a file shown in Fig. 4 (21 Patents Assignees). Note that the results for subjects (BREAD or WINE) are indicated (number of patents).

To obtain the final result, this file (the 21 Patent Assignees) is compared to the file containing the

†The GET command can deal with 4000 unique terms. Here 1000 is only used to be sure that all the PA will be used during the GET process.

Patent Assignees dealing with CHEESE. The result is indicated in Fig. 5. The comparison of the 21 Patents Assignees dealing with BREAD and WINE, to the file containing the 444 Patent Assignees dealing with CHEESE, leads to eight Patent Assignees. These eight Patent Assignees are concerned with BREAD, WINE and CHEESE.

The summary of the process is indicated here:



(ii) *Common subjects to ELF, TOTAL and BP (WPIL database)*

We used the Derwent classes, as an indicator of activities. This means that the GET command has been processed with the DC field of the WPIL database.

We will use here the same concept as for the BREAD, WINE and CHEESE search. But, the application will be made differently. In the present case, we know the name of the Patent Assignees (ELF, BP and TOTAL). But what we have to search is the core of their common research!

ELF, BP and TOTAL have been searched as BP/PA, ELF/PA and TOTAL/PA. This search although not completed, has been used only as an example (see Table 1).

The Derwent Classes are used to identify the areas of research. Other fields may be used [Derwent

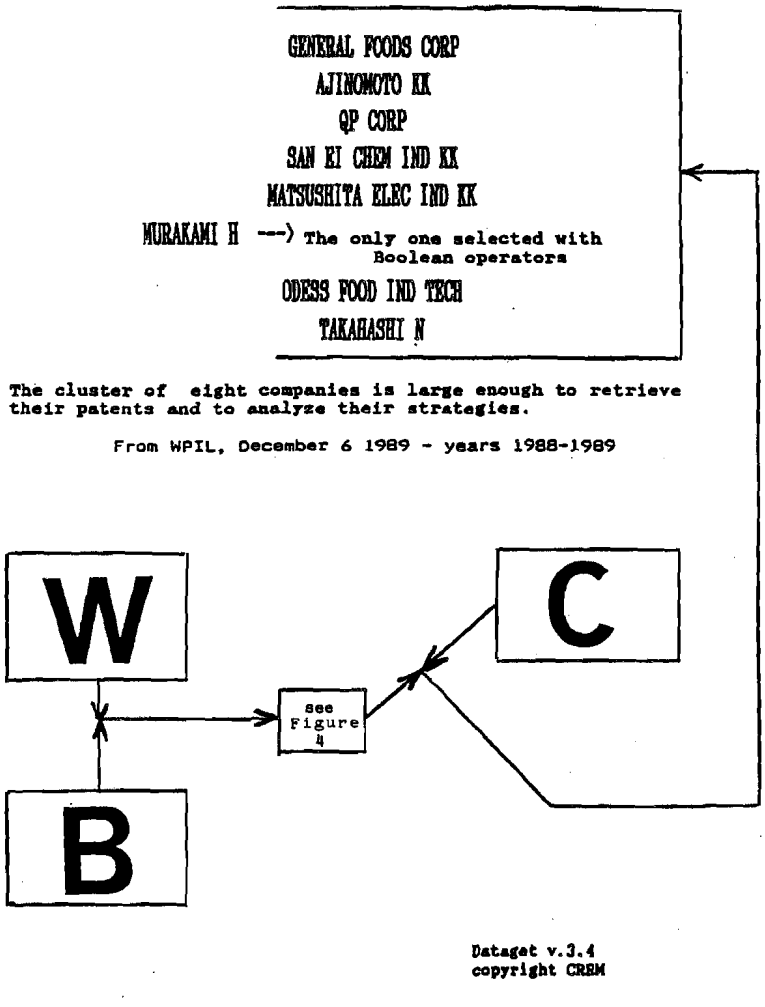


Fig. 5. Companies owning patents dealing with bread, wine and cheese, but not necessarily in the same patent.

Table 1. Results (8 April 1990)

Companies	ELF	BP	TOTAL
Patents	574	390	151
GET files*	213	152	124

* The GET command has been performed on the DC field. The numbers indicated in the table show the numbers of unique DC classes present in the 574, or 390 or 151 patents.

```
get    rank top 1000 ss 4

PROG:
THERE ARE 150 UNIQUE VALUES.
OCCURRENCES      TERM
          8      BINENFELD, ZLATKO
          8      MAKSIMOVIC, MATEJ
          4      BOSKOVIC, BOGDAN
          3      GRANOV, AZRA
          3      KOVACEVIC, VELJKO
          2      BECK, H. C.
          2      BENSCHOP, H. P.
          2      BINENFELD, Z.
          2      BOTER, H. L.
          2      D'AGOSTINO, P. A.
          2      D'AGOSTINO, PAUL A.
          2      DELJAC, VJERA
          2      DURST, H. DUPONT
          2      HAYES, A. W.
          2      MAKSIMOVIC, M.
          2      MINK, WILLIAM H.
          2      NOVKOVIC, STANKO
          2      PHELPS, KIRKMAN R.
          2      PROVOST, L. R.
          2      PROVOST, LIONEL R.
          2      STOCK, THOMAS
          2      STOEHR, RALF
          2      VERWEIJ, A.
          2      VOJINOVIC, MIRJANA
          2      VOJVODIC, VLADIMIR
          1      ALTHOUSE, MARK L.
          1      ARCHER, PHILLIP W.
          1      ARTURSSON, ELISABET
          1      BAJGAR, JIRI
          1      BARRASS, B. C.
          1      BEERS, EVERETTE T.
          1      BERITIC, T.
          1      BEVANDIC, Z.
          1      BLACK, R. M.
          1      BLANCH, J. H.
          1      BLUMBERGS, PETER
          1      BONOMO, M. S.
          1      BOSCHKE, F. L.
          1      BOSKOVIC, B.
          1      BOUCK, JAMES B.
                 .....
```

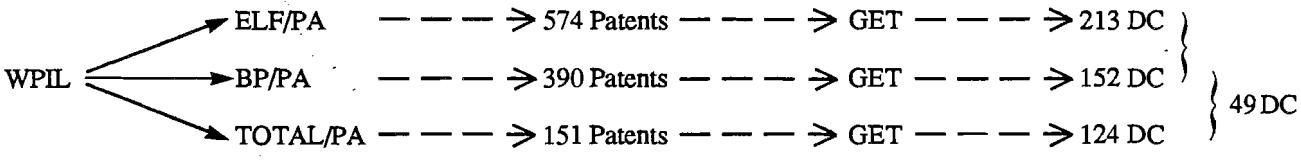
Fig. 6. Example of GET file (author's field). Note that within the same database the authors christian names are written differently.

Manual Codes (MC), or the International Classification (IC)].

First, we determined the amount of patents for each company, second we performed on each result the command GET DC RANK TOP 1000, and we saved the results, summarized above. For instance ELF/PA lead to 574 patents, the GET on the DC field lead to 213 different Derwent classes. Figure 6 also shows the results of the GET on the AU (assignee field).

The comparison of the three GET files (first two of them and the result to the third one), with the DATAGET software allows the selection of 49 Derwent classes common to the three companies. This is the common core of their research.

The summary of the process is indicated here:



Common DC to ELF, TOTAL and BP. To be significant, we have limited the comparison to DC classes $\geqslant 2$. This means at least two patents with this class present for each company.

Using the same steps as for the former example, we have selected the Derwent classes common to these three companies:

H04, H01, Q49, J04, A97, Q67, Q74, S03, L02, Q66, R16, J01, Q35, E19, H06, H07, H09, J08, L03, M13, M25, P26, P73, P81, P85, Q43, Q75, R13, V07, X15, R27, R41, E37, A23, M14, P43, Q42, Q78, P42, Q73, S02, X12, A14, A93, E12, E34, H02, P41, Q39

Total: 49 Derwent classes (to be compared to the values found for each GET file).

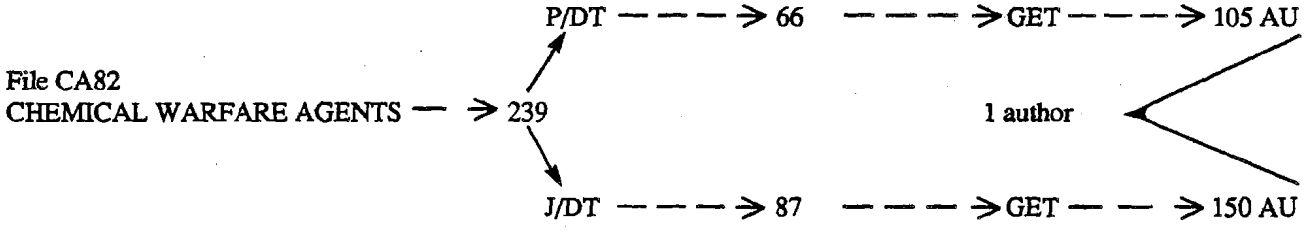
Examples of the meaning of various Derwent classes. The Derwent classes have a format of three digits, one letter and two figures. The meaning of the Derwent classes can be obtained from the Derwent User Manual. For instance:

---------- electric
R16 Measuring, testing. Investigating chem./phys.props.
---------- mechanic
Q49 Building, construction. Mining
---------- chemdoc
E12 General organic. Organometallics
J04 Chemical/Physical processes/apparatus catalysis catalysts colloid-chem.
L02 Refractories, ceramics, cements, soil roads, magnesia, abrasives
---------- general
P73 Pressing, printing. Layered products
---------- plasdoc
A97 Miscellaneous goods
---------- sx-elect
S03 Scientific instrumentation
X12 Power generation

The GET command is not limited to Patent Assignee field or Derwent classes. Inventors, manual codes, index terms, registry numbers, etc. may be used. This provides a way to follow the trends of a technology between two or several firms, or to detect inventors common to various areas of research (e.g. patents and fundamental research).

The resulting GET files are saved and compared to find authors common to patents and journals. The result is summarized in the following table. Note that authors names may have various spellings at the level of the christian names, even in the same database. This difficulty is overcome by using different string lengths to compare the author names.

The results are summarized here (see also Fig. 7):



(iii) *Common points between fundamental and applied research*

In databases dealing at the same time with applied and fundamental research it will be possible to determine the overlap (in term of publication frequencies) between technology and fundamental research.

Using a database such as Chemical Abstracts (CA), we want to select the author(s) which have published in both areas (Journals, and patents) in the field of Chemical Warfare Agents, during the period of 1982 to May 1990.

CHEMICAL WARFARE AGENTS is a CA index term. This allows selection of 239 posting terms. The use of P/DT and J/DT (Patent or Journal as document type) leads to the selection of 66 patents and 87 journals (always as posting terms). The GET command is then performed on both answers, on the Authors fields (in CA Patent inventors are indexed as authors).

This treatment allows the selection of one author over a total of more than 200. This author is interesting since he published fundamental papers and patent.

The Limits of the Method

There are only a few limits to this method:

- the host must offer the best possible facilities at the most competitive price, specially for the use of the GET command;
- the GET command must handle a sufficient number of unique terms. In the case of Orbit, the limit is 4000 and this is sufficient for most of the applications;
- to use the DATAGET software a micro-computer PC compatible, with an arithmetic coprocessor, a speed of at least least 8 Mz and 640 Ko of RAM is necessary. Computers with a lower standard may be used, but the processing will take more time.

```
-----------------------------------------------------------------

Auteurs communs aux brevets et publications MAI 90 - CAS -

     Terms common to files: getP and GETJ
     Terms from file: getP are printed first. Key=  12
-----------------------------------------------------------------

1       SEIDERS, REGINALD P.
1       SEIDERS, REGINALD P.



-----------------------------------------------------------------

Auteurs communs P et J chemical warfare agets (CAS) 239 ref.

     Terms common to files: getp and getJ
     Terms from file: getp are printed first. Key=  6
-----------------------------------------------------------------

1       POWELL, MICHAEL J.
1       POWELL, R. J.
1       SEIDERS, REGINALD P.
1       SEIDERS, REGINALD P.
1       TAYLOR, RICHARD F.
1       TAYLOR, K.

Comparison of the GET results on the author field using
different keys (to be sure to by pass the christian name
problem)
```

Fig. 7. Comparison of the two GET files. Note that the various key lengths allow the selection of true common authors.

Conclusion

The online databases are a very powerful tool when all the command facilities offered by the host are used. The post processing of the GET command results is in our opinion one of the most powerful way to follow strategies of firms, and to select unusual clusters of patents prior to analysis.

Acknowledgement — We would like to thank ORBIT SEARCH SERVICE for its help during the realization of this work.

References

(1) H. Dou, P. Hassanaly and A. La Tela, New aspect of online retrieval. Get more from your downloaded data. Noise can be useful. *Online International Information Meeting, Dec. 1988,* pp. 379–407. Learned Information Editor, London, 1988.

(2) F. Jakobiak, *Maîtriser l'Information Critique.* Les Editions d'Organisation, Paris, 1988.

(3) J. F. Rockart, Chief executives define their own data needs. *Harvard Business Review* March–April, 81–93; 1979.

(4) Narin, Deuxième Colloque sur l'Information Elaborée Société Française de Bibliométrie 1989, Ile Rousse Rev, Française de Bibliométrie.

(5) H. Dou and P. Hassanaly, Strategic management of technologies in chemistry. New usage of online databases. *J. Chem. Inf. Compt. Sci.* Submitted.

(6) J. Terragno, The GET command a powerful new patent searching tool from Pergamon Infoline (now Orbit Search Service). *World Patent Information* **6**, 69–73; 1984.

(7) M. Moureau and A. Girard, Utilisation des bases de données brevets pour les études statistiques. Internal Report 34686, IFP, Rueil Malmaison, France.

(8) M. Reyne, Les Choix Technologiques pour l'Entreprise. Analyse de l'environnement économique. Lavoisier Editor, Paris, 1987.