A NEW METHOD FOR ANALYSING DOWNLOADED DATA
FOR STRATEGIC DECISION

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

* IBM CEMAP (Centre Européen de Mathématiques Appliquées),
68-76 quai de la rapée, 75592 Paris CEDEX 12 (France)
** CRRM Faculté de St Jérôme, 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
Rockart1 the "Critical Success Factors" (CSF). As soon as defined, these "CSF" must
be overlooked by specialists of information retrieval. To do this they need access 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
increases2 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>
<thead>
<tr>
<th>Field</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>AN</td>
<td>Accession Number in the database</td>
</tr>
<tr>
<td>TI</td>
<td>Title of the patent</td>
</tr>
<tr>
<td>DC</td>
<td>Derwent Codes</td>
</tr>
<tr>
<td>PA</td>
<td>Patent Owner</td>
</tr>
<tr>
<td>IN</td>
<td>Inventor</td>
</tr>
<tr>
<td>PN</td>
<td>Patent Number</td>
</tr>
<tr>
<td>PR</td>
<td>Priority Number</td>
</tr>
</tbody>
</table>

Table 1  
Constitutive fields of a bibliographic reference and their meanings

280 Scientometrics 25 (1992)