Implementing strategic information by using APA (Automatic Patent Analysis) - Example in the field of steels for high temperature applications

Henri Dou* and Daniel Massari de Souza Coelho**

*Director of Atelis (Strategic Work Room of the ESCEM Group, France), e-mail douhenri@yahoo.fr

** PhD student, Pontifical *Catholic University of Rio de Janeiro, Brazil, e-mail daniel@coelho.org*

Keywords: Text mining; Automatic patent analysis; Strategic information; High temperature corrosion;

This paper aims to show how patent analysis is a useful information tool to determine the environment of a research subject before the beginning of laboratory experiments and measures. The analysis is focused in materials used in high temperatures where corrosion can limit their lifetime. The Matheo-Patent software - available from the company Matheo-software[1] - and the European Patent Office (EPO) World Patent Database- which covers more than 80 countries - were used for this purpose.

There are several databases available freely from Internet and different tools can be used to query the database, download all the bibliographic data and perform off line an automatic patent analysis (APA). Patents are particularly interesting since they are almost the only documents which build a bridge between the fundamental research and its applications. Moreover, what is published in patent documents is seldom published elsewhere.

The query in the worldwide database from EPO can be done by using words from titles and abstracts – since there is no keywords in patents references-, the assignees or inventors names, the patent numbers or dates, the priority patents and the IPC (International Patent Classification). When a local database is created from the patents downloaded online, this database is formatted to be able to be analyzed automatically. At the same time the significant words from the titles and abstracts are extracted as well as the drawings. All the bibliometrics analysis can be done automatically generating lists, charts, matrices, networks. These different treatments provide implicit information, generally new, that it is impossible to detect by reading or browsing the database, and allow creating groups of patents and further analyzing their content, to select significant information, to benchmark automatically the knowhow of companies and inventors, to determine the inventors or company's networks. The local database created can be updated using the same query or completed by different queries, and the patents will be added to the initial database and the duplicate eliminated. If the user wants to work with a large database, selections (groups) according words, dates, IPC, applicants, countries can be done. These new groups can be analyzed separately.

[1] Matheo Software. Available at: http://www.matheo-software.com . Access on April 2012. (A trial version can be downloaded)

The full text of the patents can be downloaded if necessary, as well as the content of the local database in different formats which can be integrated in more powerful software or in mailing or cooperative platforms. If the users have other patent sources, Matheo-Patent can also work on the two US Patent databases and also import data from Delphion, Patbase, Micropatent and Derwent.

The method used is related to the classical cycle used in Competitive Intelligence. From the initial vision, there is a selection of information or databases and the system will provide the facilities to manage and handle this information and to perform APA. The system provides to decision makers - here the researchers or the person coordinating the research - the recommendations or alert indexes available from the implicit information developed during the analysis. The objective of APA is to get the information available in the group of selected patents to answer the questions: who is doing what, where, when, with whom, what are the competencies of the assignees, their links, etc. This information can be used to select the right patents, to see the trends in technology, eventually the innovative orientations, and to help the users to define the strategy for their company.

Before the development of various queries strategies, it is important to understand the mechanism of the APA, that is, its bibliometrics treatments. The objective is to get from an explicit information (here patent bibliographic data) an implicit information obtained by various statistical treatments. To get this implicit information, the analyzed data should be around the thematic that we want to explore. Thus, the formulation of the query is far away from the classical documentation, where people try to get the best precise query to obtain only the answer they wish. In this condition, the query predetermines the answer and to treat the corpus via bibliometrics systems is useless.

Conclusion

The use of APA (Automatic Patent Analysis) to know the technical situation of a thesis subject is particularly interesting because patent documents make a bridge between fundamental research and applications. The information available in a patent reference allows after a bibliometrics treatment to answer various questions about the main assignees (companies), the new possible entrants, the chronology of the inventions related to the subject, the search for partnerships, the identification of the main competitors, etc. It is also very important to keep in mind in these analysis that what we see (during the interval of time choose for the search) is what the people did in the past or in the recent past. But not what do they do today. Then, if the analysis indicates companies and inventors dealing with research or applications directly linked to the subject, they may be considered as “targets” to be followed to see in which direction they work today using the classical scientific databases (from Dialog, for instance), the Internet and web sites, the international meetings and workshops, etc.