International Journal of Competitive Intelligence, Strategic, Scientific and Technology Watch
SciWatch Journal 1 (2008), Issue 1, 85 - 86

REVIEW OF ARTICLE: "Competitive intelligence and patent analysis in drug discovery: Mining the competitive knowledge bases and patents"

Grandjean, N., Charpiot, B., Pena, C.A., Peitsch. M.C (2005)

(Review by M. Welling Flensborg)
e-mail: m.welling.flensborg@gmail.com

The article begins by trying to establish the relation that exists between Patent Analysis and Competitive Intelligence in the pharmaceutical industry, particularly in the drug discovery process. The text analyzes how patent analysis methods and data exploration techniques are applied and how these methods can help in the task of getting information on competitor's activities, R+D activities, potentially emerging fields, collaborations that are being done, etc…

First of all, the article exposes that the existence of an 18 month lag between the filing date of an innovation and the publication done by the patent offices proofs there is an important delay between real research being done and what patent offices are publishing. Even though this delay can alter the results of patent analysis methods, patents are still the most reliable source of information on R+D activities (along with the other resources that Competitive Intelligence, as a whole, has). Therefore, the quality of data from patents becomes increasingly more important in the Competitive Intelligence field, especially when it comes to understanding a company's strategy by analyzing its patent portfolio. The main issue with patent data is the lack of precision and the general intelligibility of the texts (due mostly to its complexity) in the patent's primary archives. On the other hand, the secondary sources that exist are generally filled with poor descriptions of the contents of patent documents.

The process of extracting useful information from patents can be either simple (parsing and normalized) or, in the majority of cases, more complex. Complex extraction requires extensive processing procedures to obtain a "data set" suitable for further analysis, exploration and interpretation. In these more complex cases, sources are characterized for needing context-depending information extraction and categorization: (e.g. the concept "precise therapeutic targets" can be simply extracted from the patent document or either processed and categorized in a parallel data archive). Another good example of complex extraction methods could be concepts such as "companies", "inventors" or "models of action" that are also shared by other second level data archives. Therefore these concepts can be used to integrate and bridge different databases with each other.

The different steps in the process of extracting useful content from patents are:

1. **Zoning**: tagging relevant contexts that are identified by the recognition of scientifically meaningful strings found on the text.
2. **Parsing**: extraction of strings to a data file.
3. **Normalization of the concepts**: can be done either by applying a set of rules or by applying a preferred term to a synonym.
4. **Lexical extraction**: Extraction of the concepts taking into account the different lexical forms that concept can take (synonyms, polysemy) and considering the context in which is found.
5. **Categorical coding**: splitting values into categories or bins.

6. **Categorization**: In order to navigate through the data set, a taxonomy system is applied, so the information can be displayed in different ways according to the search criteria.
7. **Relate concepts**: Establish relationships between the concepts using syntactic analytic rules.

This process can be used to create databases that complement the existent commercial databases and it also allows the building of new customized views of the documents. These new views are more in line with current scientific context, and should always be scientifically coherent, mathematically valid and suitable for data exploration.

Exploring and analyzing databases can be done using exploratory data analysis techniques. These techniques can show the different relations that exist between different variables, and they usually work together with graphical tools that help show those relations, and are also capable of identifying unexpected new patterns or emerging information from the variables.

The article then exposes the different commercial options that are available for both patent analysis methods and data exploration techniques.