

## ➡ NeuroNav™

### A comprehensive software environment for statistical analysis of large amounts of textual content

NeuroNav™ is a text mining software designed for analysis and skimming of textual work. Dedicated to documentary corporate back-office use, NeuroNav™ is a software environment for statistical analysis of textual content, navigation and classification.

The software is designed to spare a lot of precious time when meaningful extractions are needed to be used locally or to be published onto a corporate Web site. With the NeuroNav™ tools the indexing process is fully controlled. At each step of the work, NeuroNav™ makes it easy to modify the index, to quickly get an overview to the work and to understand its semantic structure.

- **Navigation**

The user navigates the contents along cross-referenced entities: words, titles and topics. A list of words can result from a boolean combination of titles or topics. A list of titles can result from a boolean combination of words or topics, and so on.

- **Controlling the indexing process**

NeuroNav™ includes a powerful indexing engine which produces a first state full text index that can then be modified. The user can create, merge, modify or delete keywords according to his final scope. The elementary documents themselves can be deleted or modified or even created and indexed from scratch.

- **Fuzzy lexical access to the indexing vocabulary**

By entering some words, the user generates from the indexing vocabulary a short list of words that are lexically similar. This is very useful to get results in spite of spelling variations or orthographic errors residing inside documents or made by the user.

- **Semantic expansion**

The semantic expansion process generates the list of entities, words or documents, which are semantically close to the words and/or documents selected by the user.

- **Semantic map**

An automated classification process based upon the techniques of *axial K-means* summarizes the textual corpus to its main topics. If the indexing vocabulary changes, the topics can also become different. Each topic comes with a list of related words and a list of related documents sorted by decreasing importance. These lists are generally very homogeneous and bring a clear understanding of the reason why the documents have been clustered and therefore of the topic.

- **Displaying the topics on the screen**

The lists of words related to the topics are statistically processed with classical PCA (Principal Component Analysis) so that the topics can be graphically displayed onto the screen as a semantic map. Proximities are reinforced by the drawing of links with light and shade between the neighboring topics.

- **Accessing the document**

While navigating the user can always access a particular document and look at its content. He can also always access its related keywords, modify them and display the whole indexing vocabulary.

- **Building views onto the textual work**

NeuroNav™ is designed for building several views of the work and allows the user to work on each one independently. A view can be duplicated to save one interesting index state. The user can then swap to another view or start a new one. Thus different views and their semantic maps can be compared. Zooming in on a topic is also possible.

- **« Caddie » and anti-dictionaries**

Collecting terms while navigating is made possible with a “caddie” that crosses all the views and into which the user puts the words he has picked up. The “caddie” can be exported as an ASCII file. Anti-dictionaries are ASCII files containing words to be rejected from the indexing vocabulary. Applying an anti-dictionary to a view will trigger the deletion of its words from the related indexing vocabulary. The user keeps full control of the process and can prevent some words from being deleted.

- **Publishing views on a web site**

Views built up with NeuroNav™ in the corporate documentary back-office can be published on a Web site for further consultation through Internet / Intranet.

# DIATOPIE

Ingénierie  
documentaire

*CartoWeb* is the java applet allowing a client to access a view. He can thus navigate through words, titles, topics, and see the related semantic map.

*Alain Lelu, professor and researcher in the French University of Franche Comté, has created the concept of symmetric navigation through words, titles, topics, the axial k-means method allowing fuzzy classification, the similarity calculation method used for semantic expansion functionality and the word lexical approximation method.*