

➡ NeuroNav™

Un environnement complet d'analyse statistique de contenus textuels

NeuroNav™ est un logiciel de Text Mining consacré à l'analyse et au parcours de corpus textuels. Dédié au back office, il constitue un environnement complet d'analyse statistique de contenus textuels avec des outils de navigation et de classification performants. Par leur emploi bien maîtrisé, les outils de NeuroNav™ peuvent générer des gains de productivité appréciables pour le gestionnaire de l'information qui souhaite opérer des extractions pertinentes dans un corpus documentaire volumineux, en maîtriser l'indexation, en tirer des analyses d'ensemble ou partielles et les diffuser sur l'Intranet de son entreprise.

NeuroNav™ facilite le contrôle éditorial de l'indexation des textes, l'appréhension du corpus documentaire et l'accès aux documents. Il donne à l'utilisateur une vision de la structure thématique du corpus documentaire à chaque étape de son travail d'appropriation.

- **Navigation dans le corpus.**

Elle se fait selon trois axes : mots, documents et thèmes. Dans cette navigation "triangulaire", une liste de mots résulte d'une combinaison booléenne de documents et/ou de thèmes, une liste de documents résulte d'une combinaison booléenne de mots et/ou de thèmes.

- **Contrôle de l'indexation et des documents indexés**

NeuroNav™ intègre un moteur d'indexation performant qui construit une première indexation de départ. A tout moment au cours de la navigation, l'utilisateur peut modifier l'indexation : suppression, modification, création, fusion de mots d'index. Le logiciel permet aussi la suppression et la modification des unités documentaires et même la création "ex nihilo" de nouvelles unités.

- **Accès lexical « flou » au vocabulaire d'indexation.**

La saisie d'un mot par l'utilisateur engendre la liste des mots d'index du corpus les plus proches au plan lexical. Cette fonctionnalité apporte la nécessaire tolérance aux variantes d'écriture d'un même mot ou aux fautes d'orthographe, qu'elles viennent de l'utilisateur ou du corpus lui-même.

- **Expansion sémantique.**

Le mécanisme d'expansion sémantique engendre la liste des documents proches sémantiquement d'un ou plusieurs documents donnés (les documents qui partagent le plus de termes significatifs communs), ainsi que la liste des mots d'index proches sémantiquement d'un ou plusieurs mots donnés (les mots le plus souvent associés dans les documents).

- **Cartographie sémantique.**

Une classification automatique des documents selon la méthode des K-Means axiales résume le corpus en ses composantes thématiques principales. Ces composantes évoluent au fil des modifications opérées par l'utilisateur sur l'indexation du corpus documentaire. A chacun des thèmes sont associés une liste de documents, ordonnés par importance décroissante et une liste de mots clés également ordonnés.

Ces listes dont l'homogénéité signe la qualité des clusters fournissent la matière nécessaire à la compréhension fine des regroupements effectués.

Présentation cartographique des thèmes.

Les listes des mots caractéristiques des thèmes identifiés sont analysées selon une méthode d'analyse en composantes principales pour être présentées dans un plan factoriel.

DIATOPIE

Ingénierie
documentaire

Les proximités géométriques entre les thèmes sont renforcées et nuancées par le dessin des liaisons entre les thèmes proches.

- **Accès au document.**

A une étape quelconque de la navigation l'utilisateur peut accéder à un document, visualiser son contenu, ses mots d'index ainsi que la liste globale de l'index.

- **Construction de plusieurs « vues » sur le corpus**

NeuroNav™ est conçu pour conduire un travail d'indexation et d'analyse selon plusieurs approches parallèles. Une approche est matérialisée par une vue sur la base associée à une carte sémantique. Les retours en arrière sont rendus possibles en passant d'une vue sauvegardée à l'autre de façon à faciliter le cheminement vers la vision de la base la mieux adaptée à la problématique de chaque utilisateur. La duplication de vue, le zoom sur un thème et la comparaison de cartes sémantiques sont assurés.

- **Gestion d'un « panier » et des anti-dictionnaires**

La collecte de termes au fur et à mesure des navigations est rendue possible grâce au « panier », réceptacle de mots autonome et transversal à l'ensemble des vues. Le panier est exportable vers d'autres applications.

Les anti-dictionnaires sont des fichiers « métier » contenant des listes de mots à ne pas retenir dans l'indexation d'un corpus. L'application d'un anti-dictionnaire à une vue se fait de façon contrôlée, c'est à dire que l'utilisateur peut rejeter certains mots de la liste des mots de l'anti-dictionnaire.

- **Publication de vues sur un site web**

NeuroNav™ permet de publier sur un site Web tout ou partie des vues sur un corpus textuel construites en back-office. Celles-ci peuvent alors être consultées via Internet / Intranet grâce à l'applet Java *CartoWeb*, extension Web de NeuroNav™. *CartoWeb* est un outil de consultation en front office L'utilisateur final dispose des fonctions de navigation triangulaire et les cartes sémantiques sont visualisables.

Le concept de navigation triangulaire entre mots, documents et thèmes, la méthode des K-means axiale permettant la classification floue des documents et mots associés, la méthode de calcul de la similarité entre documents ou mots pour l'expansion sémantique ainsi que la méthode d'accès lexical flou au vocabulaire d'indexation sont issus des travaux d'Alain Lelu, enseignant-chercheur d'abord à l'Université de Paris 8 puis à l'Université de Franche Comté / LASELDI.