

# EXPLORATION DE DOCUMENTS GRÂCE À LA NAVIGATION CONCEPTUELLE

Les masses documentaires grandissent à une vitesse effarante et il y a fort à parier qu'il en est de même pour celle de votre organisation. Les rapports, courriels, notes et autres documents créés par vous et les membres de votre organisation viennent s'y ajouter quotidiennement. Idéalement, chaque document devrait être annoté dès sa création par le biais de métadonnées pertinentes, cohérentes et complètes. La recherche de documents en serait ainsi facilitée et plus agréable. Mais que faire avec un ensemble de documents, non catégorisés et non annotés ?

Les experts en traitement du langage naturel et en recherche d'information de l'équipe Développement et technologies Internet (DETI) ont développé une approche pour vous aider à découvrir l'information ensevelie dans votre masse documentaire : nous l'appelons la **navigation conceptuelle**.

## LA NAVIGATION CONCEPTUELLE SE DIVISE EN 4 ÉTAPES :

- 1 Extraction de termes de votre masse documentaire.
- 2 Calcul de similarité entre termes et visualisation graphique des termes reliés.
- 3 Recherche de documents contenant des termes particuliers.
- 4 Aperçu rapide du contenu des documents.

La figure 1 montre les 4 étapes telles qu'intégrées dans une interface usager. Notre **extracteur de termes** parcourt votre masse documentaire et en extrait les termes importants (FIGURE 1, EN HAUT À GAUCHE). Notre approche repose sur des méthodes statistiques promues en linguistique computationnelle.

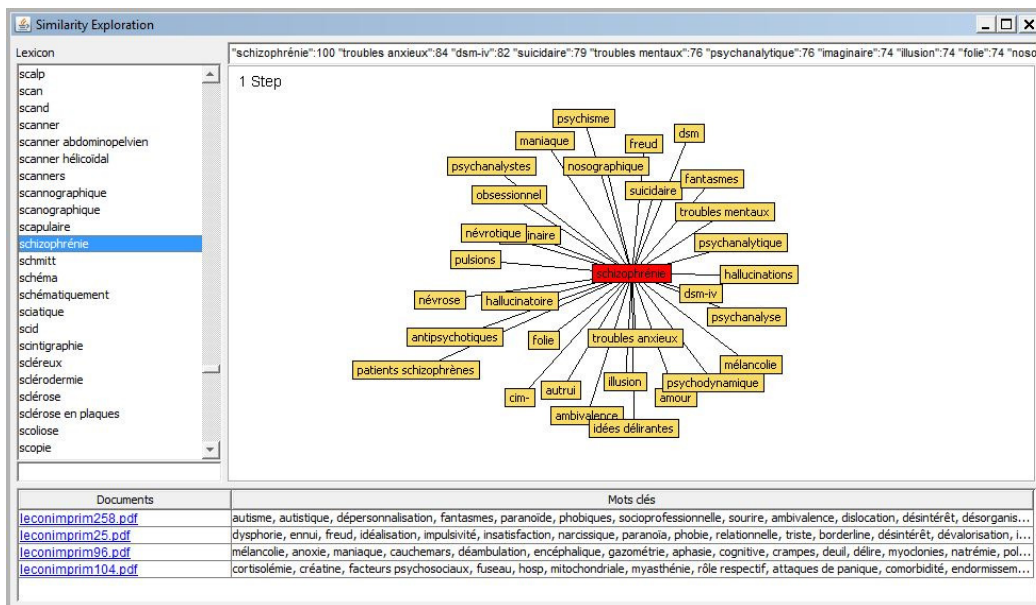


Figure 1  
PROTOTYPE DE  
NAVIGATION  
CONCEPTUELLE.



Notre module de **recherche de mots apparentés** utilise lui aussi des méthodes statistiques permettant d'analyser vos documents et de mesurer quels termes sont similaires dans votre contexte. Cela est différent de listes de synonymes tel que pris dans des dictionnaires, il s'agit d'analyse de similarité en fonction de votre domaine et de vos documents.

La **navigation conceptuelle** est alors possible, car nous représentons graphiquement (FIGURE 1, EN HAUT À DROITE) les termes reliés. Vous pouvez naviguer d'un terme à l'autre à l'aide d'un simple clic.

Vous voyez un terme qui vous intéresse et êtes curieux de voir quels documents contiennent ce terme ? Il suffit d'un clic sur le mot dans l'interface graphique pour **que la liste des documents soit affichée** (FIGURE 1, EN BAS À GAUCHE).

Comme les titres des documents sont souvent peu évocateurs, **un aperçu rapide de votre document** est donné en listant quelques termes pertinents s'y retrouvant (FIGURE 1, EN BAS À DROITE).

Le prototype a été développé dans le but d'illustrer diverses technologies maîtrisées par nos experts. Certes, **ce prototype pourrait être poussé plus loin, et dans diverses directions pour s'adapter à vos besoins.**

- Vous avez déjà une liste de termes ? Cette liste peut remplacer le résultat de l'extracteur.
- Vous avez déjà une liste de synonymes ? Cette liste peut être ajoutée à nos résultats de calculs de similarité.
- Vous trouvez que l'aperçu rapide par mots-clés n'est pas suffisant ? Nous pouvons afficher quelques phrases du document en utilisation des techniques de résumés automatiques.
- Vous désirez une recherche d'information plus poussée avec mots-clés multiples ? Nous pouvons intégrer des techniques plus avancées de recherche d'information.

#### RECHERCHE ET DÉVELOPPEMENT :

Jean-François Lavallée, M. Sc., agent de recherche, équipe DETI, CRIM  
Caroline Barrière, Ph. D., chercheure, équipe DETI, CRIM

#### INFORMATION :

Sacha Leprêtre, directeur de l'équipe DETI, CRIM  
Tél. : 514 840-1238 ou 1 877 840-2746 - [sacha.lepretre@crim.ca](mailto:sacha.lepretre@crim.ca)