

Un outil de navigation dans un espace sémantique

Yann Vigile Hoareau*, Murat Ahat**
David Medernach*** Marc Bui****

*Université Paris 8, vigilehoareau@gmail.com

**Ecole Pratique Des Hautes Etudes, murat.ahat@etu.epeh.sorbonne.fr

***Université Paris 8, david.medernach@gmail.com

****Ecole Pratique Des Hautes Etudes, marc.bui@ephe.sorbonne.fr

1 Introduction

L'outil que nous allons présenter a pour objectif d'exploiter les propriétés des espaces sémantiques et des méthodes de visualisations de graphes pour proposer un moteur de recherche disposant des fonctionnalités suivantes :

1. recherche à partir de mot-clés
2. recherche à partir de documents entiers
3. recherche dite "sémantique" qui n'est plus basé sur la croisement des mots-clés de la requête avec les mots qui apparaissent dans les documents, mais sur le croisement du sens des mots qui compose la requête avec le sens des mots qui composent les documents. À titre d'illustration, contrairement à la recherche classique par mot-clé, la recherche "sémantique" serait capable de retrouver le document "les troubles à Bangkok" à partir de la requête "les émeutes en Thailand".
4. visualisation globale d'une grande collection de documents (plusieurs milliers de documents) pour permettre à l'utilisateur d'appréhender "la structure" de la collection de documents. Si l'on prend l'exemple d'un centre de recherche scientifique donné, on pourrait saisir instantanément les thèmes de recherches les plus populaires ainsi que les documents les plus centraux.
5. visualisation locale de la similarité sémantique d'un nombre réduit de documents (plusieurs dizaines) afin de dépasser la simple liste ordonnée de documents pour offrir à l'utilisateur une représentation visuelle exprimant intuitivement les relations de proximités sémantiques parmi les résultats retournés à partir de la requête.

Notre outil s'appuie sur un modèle d'espace sémantique appelé Random Indexing (Kanerva et al., 2000) pour représenter les connaissances. Notre méthode de visualisation est une alternative à la méthode classique de visualisation d'espace sémantique par la méthode de *multi-dimensional scaling*. Elle consiste à représenter l'espace vectoriel obtenue à partir de RI sous la forme d'un graphe qui garde les propriétés de l'espace sémantique.