

Visualisation de corpus documentaires et cartes conceptuelles

Formation doctorale	Informatique
Projet de rattachement	Knowledge and Image analysis for Decision – KID, LGI2P/EMA
Laboratoire d'accueil	Centre de recherche LGI2P de l'école des Mines d'Alès Site EERIE, Nîmes
Equipe d'encadrement	G. Dray, S. Ranwez
Collaboration	Pascal Poncelet, équipe TATOO, LIRMM
Mots-clés	Analyse de corpus documentaire, modélisation de connaissances, <i>clustering</i> , distances sémantiques, cartes de connaissances.

Expression du besoin et contexte

A l'heure où le monde de la recherche se réorganise, en particulier au sein de communautés de chercheurs, la mise en commun de bibliographie ne va pas sans poser plusieurs problèmes :

- motivation du partage et protection face à des concurrents éventuels ;
- hétérogénéité des disciplines et donc du vocabulaire ;
- navigation et recherche dans la base documentaire ;
- détection d'opinion pour l'aide à la décision.

Si le premier point relève plus des sciences sociales et des problèmes liés aux réseaux sociaux, les trois derniers peuvent trouver des solutions dans des technologies logicielles.

Impliqué dans plusieurs projets de plateformes collaboratives, l'équipe KID du centre de recherche LGI2P de l'École des Mines d'Alès propose de traiter la chaîne documentaire dans sa globalité. Avec des techniques d'analyse et de fouille de textes, nous proposons un modèle de connaissance du domaine. Les documents sont ensuite indexés par rapport aux résultats de cette analyse.

Le sujet du stage concerne plus particulièrement l'étape suivante : la navigation et la recherche d'information dans la base documentaire. Le besoin d'outils de recherche d'information performants, rapides et efficaces dans de grandes bases de connaissance suppose à la fois une bonne appréhension globale de la base documentaire par l'utilisateur et des techniques d'indexation et de fouille pour assurer la pertinence des réponses.

Le sujet de stage s'intéressera particulièrement à la visualisation de corpus documentaires, à la notion de distance sémantique entre différentes entités du corpus (pour décider de la pertinence d'une information par rapport à une requête ou bien de la pertinence de certains documents par rapport à un contexte applicatif).

Organisation du stage de recherche

Etat de l'art

Durant ce stage de recherche, le candidat devra se familiariser avec la notion de modèle de connaissances, de distance sémantique, de *clustering* et de visualisation.

A partir des travaux de l'équipe, il devra ensuite approfondir l'état de l'art sur ces notions dans le cadre d'une représentation visuelle (carte de connaissance). Différentes bibliothèques pour la visualisation seront comparées.

Recherche théorique

A partir de cet état de l'art, l'étudiant se familiarisera avec les techniques de fouilles de texte et de détection d'opinion de l'équipe [3][2]. Des mesures de distance sémantique seront mises en place, et la mesure proposée dans [1] pourra être enrichie. Nous verrons comment intégrer ces mesures de distances dans la visualisation de façon à proposer des paysages documentaires sémantiquement enrichis.

Utilisant ces distances sémantiques, la projection sur une carte met en évidence certains regroupements : quel est leur signification ? Disposant d'un modèle de connaissance et de mesures de distances sémantiques entre les éléments de ce modèle, nous proposons d'appliquer différentes techniques de regroupement (clustering hiérarchique, K-Means, etc.) afin de rassembler les différents titres autour de thématiques.

Cette étude sera appliquée dans le cadre de projets industriels et l'étudiant pourra être amené à collaborer avec Pascal Poncelet de l'équipe TATOO du LIRMM.

Contacts

Sylvie Ranwez	Sylvie.Ranwez@ema.fr	04 66 38 70 44
Gérard Dray	Gerard.Dray@ema.fr	04 66 38 70 34

Références

- [1] **Ontological ISA-Distance Measure for Information Visualisation on Conceptual Maps**
Sylvie Ranwez, Vincent Ranwez, Jean Villerd, Michel Crampes. OntoContent Workshop 2006, First International Workshop on Ontology content and evaluation in Enterprise, In conjunction with OnTheMove Federated Conferences (OTM'06), Montpellier, France, 31 Octobre 2006. (Actes à paraître dans un numéro LNCS chez Springer)
- [2] **Défi DEFT07 : Comparaison d'approches pour la classification de textes d'opinion.** *Plantié M., Dray G., Roche M.*, 3 juillet 2007, Grenoble, France
- [3] **Détection d'Opinions : Apprenons les bons Adjectifs !** *A. Harb, G. Dray, M. Plantié, P. Poncelet, M. Roche et F. Troussel.* In Actes de l'Atelier FOuille des Données d'OPinions (FODOP 08), conjointement Conférence INFORSID 08, Fontainebleau, France, Mai 2008.
- [4] **Using Concept Lattices for Visual Navigation Assistance in Large Databases: Application to a Patent Database.** *Jean Villerd, Sylvie Ranwez, Michel Crampes, David Carteret.* In Journal of General Systems, Special Issue on "Concept Lattices and their Applications", to be published.