

Distances sémantiques et cartes conceptuelles

Formation	Master 2R informatique
Laboratoire d'accueil	Centre de Recherche LGI2P de l'Ecole des Mines d'Alès
Adresse	Site EERIE, Parc. G. Besse, 30 035 Nîmes
Projet de rattachement	Knowledge Discovery for Decision making - KDD
Equipe d'encadrement	<u>S. Ranwez</u> , G. Dray, M. Crampes
Mots-clés	Ontologies, Modélisation de connaissances, clustering, distances sémantiques, visualisation d'ontologies.

Expression du besoin et contexte

Le besoin d'outils de recherche d'information performants, rapides et efficaces dans de larges bases de connaissance s'accompagne souvent du besoin de mesures de similarité sémantique. En effet que ce soit 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, une telle mesure peut être mise en œuvre.

Plusieurs mesures de similarité sémantiques ont été proposées dans la littérature, mais la plupart ne respectent pas les qualités attendues d'une distance (l'inégalité triangulaire, entre autres). Nous avons proposé une distance basée sur la relation "isa" des ontologies [1]. L'objectif de ce stage de recherche est de valider cette distance et de l'étendre à d'autres types de relations.

Cette distance sera mise en œuvre dans différents projets industriels (ANR-RIAM et autre), pour visualiser des ontologies de domaines (musique, brevet, etc.).

Organisation du stage de recherche

Etat de l'art

Durant ce stage de recherche, le candidat devra se familiariser avec la notion d'ontologies.

A partir des travaux de l'équipe, il devra ensuite approfondir l'état de l'art sur les distances sémantiques en insistant particulièrement sur celles qui respectent les propriétés attendues des distances.

Un complément sur les différentes techniques de visualisation des ontologies pourra être demandé.

Recherche théorique

A partir de cet état de l'art, la mesure de distance proposée en [1] devra être validée. En particulier dans le cadre de projets industriels, nous avons développé une application pour manipuler une vaste collection de titres musicaux (navigation, indexation, fouille du paysage) de façon visuelle. Pour donner sens à ce *paysage musical* nous souhaitons projeter l'ontologie du domaine. Or en usant de distance sémantique, cette projection entraîne des regroupements : quel est leur signification ? Disposant de l'ontologie et d'une mesure de distances sémantique entre les concepts de cette ontologie, nous proposons d'appliquer différentes techniques de regroupement (clustering hiérarchique, K-Means, etc.) afin de rassembler les différents titres autour de "familles musicales" (autre que les genres musicaux usuels).

Le stage de recherche a également pour objectif d'étendre la mesure de distance sémantique à d'autres types de relations (méronymie par exemple).

Nous souhaitons, à travers ce projet expérimenter de nouvelles méthodes de calcul sur les ontologies et valider notre mesure de distance.

Pratique

La visualisation des connaissances est réalisée en s'appuyant sur notre environnement de cartographie de connaissance Molage [2], qui permet de faire à la fois des graphes dynamiques et de la projection MDS (MultiDimensional Scaling) combinés. Cet environnement étant écrit en java et les fichiers d'échange en XML, des compétences dans ces technologies seraient appréciées.

Contacts

Sylvie Ranwez	Sylvie.Ranwez@ema.fr	04 66 38 70 44
Gérard Dray	Gerard.Dray@ema.fr	04 66 38 70 34
Michel Crampes	Michel.Crampes@ema.fr	04 66 78 52 32

[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] **An Integrated Visual Approach for Music Indexing and Dynamic Playlist Composition**

Michel Crampes, Sylvie Ranwez, Filip Velickovski, Christopher Mooney and Nicolas Mille. MMCN2006, Thirteenth Annual Multimedia Computing and Networking, January 18-19, 2006, San Jose, California.