

# Une approche collaborative de l'aide à l'indexation thématique d'un livre

## 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 representation and Image Analysis for Decision making – KID

### Collaborations :

- INSERM UMR\_S 872, éq. 20
- CEA

**Equipe d'encadrement** G. Dray (LGI2P), M. Plantié (LGI2P), J. Charlet (STIM), Marie Thérèse Ménager (CEA)

**Mots-clés** Indexation, Terminologies, Analyse de corpus, Acquisition des connaissances à partir de textes, Ontologies, Ontology learning.

## Expression du besoin et contexte

La constitution de l'index d'un ouvrage scientifique peut être considérée comme une activité d'ingénierie des connaissances. Cette activité consiste à identifier, recenser et organiser les termes pertinents d'un domaine, à partir de l'analyse d'un corpus de textes (les pages de l'ouvrage). Ces termes sont ensuite utilisés pour indexer le contenu du corpus. Classiquement, cette tâche est réalisée manuellement par les auteurs et/ou les éditeurs. Plusieurs travaux ont été réalisés pour permettre d'aider à la conception automatique de ces index, notamment par Jean Charlet *et al.* [1], [2]. Des approches complémentaires ont été amorcées concernant *l'ontology learning* et la création automatique d'ontologies à partir d'analyses de textes. Si à l'heure actuelle aucun système n'arrive à atteindre cet objectif, certaines fonctionnalités ont été conçues et peuvent être mises en œuvre [3], [4], [5].

Dans le cadre de ce stage, nous proposons de poursuivre ces travaux pour spécifier et mettre en œuvre sur un cas réel un outil collaboratif d'aide à la conception d'un index.

## Organisation du stage de recherche

### Etat de l'art

Durant ce stage, le candidat devra se familiariser avec les notions d'extraction terminologique, d'indexation, de représentation des connaissances.

A partir des travaux des membres de l'équipe d'encadrants, il devra approfondir l'état de l'art sur les approches d'aide à la construction d'index thématique. L'étude des approches collaboratives sera privilégiée.

### Recherche théorique

A partir de cet état de l'art, le candidat devra spécifier un système d'aide à la conception collaborative d'index thématique. Les verrous méthodologiques, scientifiques et technologiques liés à ce système devront être mis en évidence et des solutions devront être investiguées.

### Application

L'approche proposée sera appliquée sur la création de l'index d'un ouvrage sur la toxicologie nucléaire. L'outil collaboratif devra être conçu pour être intégré dans la plateforme de travail collaboratif ToxNuc.

## Contacts

Gérard Dray Gerard.Dray@ema.fr (04 66 38 70 34)

Michel Plantié Michel.Plantie@ema.fr (04 66 38 70 35)

Jean Charlet Jean.Charlet@spim.jussieu.fr (01 53 10 92 10)

## Références

- [1] Jean Charlet, Touria Aït El Mekki, Didier Bourigault, Adeline Nazarenko, Régine Teulier, and Baruk Toledano - CEDERILIC : constitution d'un livre et d'un index numériques. Actes de la conférence internationale sur le document numérique, 2004.
- [2] Didier Bourigault, Jean Charlet - Construction d'un index thématique de l'Ingénierie des connaissances. Conférence ingénierie des connaissances, Palaiseau , FRANCE 1999 , pp. 107-118
- [3] M. PLANTIÉ, J. MONTMAIN, G. DRAY - " Movies Recommenders Systems: Automation of the Information and Evaluation phases in a Multi-criteria Decision-Making Process "- 16th International Conference on Database and expert Systems Applications (DEXA 2005), Copenhagen, Denmark, 22 - 26 August 2005.
- [4] R. AREZKI, P. PONCELET, G. DRAY and D.W. PERSON. - " PAWebSearch : An Intelligent Agent for Web Information Retrieval " - Proceedings of the IEEE International Conference on Advances in Intelligent Systems - Theory and Applications (AISTA2004), Luxembourg-Kirchberg, November 15-18, 2004
- [5] Ontological ISA-Distance Measure for Information Visualisation on Conceptual Maps Sylvie Ranwez, Vincent Ranwez, Jean Villerd, Michel Crampes. In On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops, Lecture Notes in Computer Science, publisher: Springer Berlin / Heidelberg, Volume 4278/2006, ISBN 978-3-540-48273-4, pp.1050-1061, 2006.