

# Construction d'une ontologie du programme ToxNuc-E

## Formation Master 2R

**Laboratoire d'accueil** INSERM UMR\_S 872, éq. 20

**Adresse** 11, rue de l'École de médecine, 75006 Paris

## Collaborations :

- Centre de Recherche LGI2P de l'Ecole des Mines d'Alès
- CEA

**Equipe d'encadrement** J. Charlet (UMR\_S872), G. Dray (LGI2P), S. Ranwez (LGI2P), Marie Thérèse Ménager (CEA)

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

## Expression du besoin et contexte

Une plateforme de partage des connaissances a besoin que les documents qu'elle contient soient indexés. Cette indexation peut être, dans un premier abord « plein texte » mais il est rapidement nécessaire, pour des usages scientifiques d'avoir une indexation conceptuelle à base d'ontologies.

La communauté francophone d'ingénierie des connaissances travaille depuis plus d'une dizaine d'années sur le problème de la construction de ressources terminologiques ou ontologiques (RTO) à partir de corpus. Elle a produit des résultats, tant théoriques que méthodologiques et logiciels, qui ont été éprouvés dans un certain nombre de projets applicatifs. Elle est maintenant reconnue comme étant à la pointe de cette problématique au niveau international [1, 4, 5].

Dans le cadre de ce stage, nous proposons de poursuivre ces travaux pour spécifier et mettre en œuvre sur un cas réel la construction d'une ontologie sur la toxicologie nucléaire.

## Organisation du stage de recherche

### Etat de l'art

Durant ce stage, le candidat devra se familiariser avec les notions d'ontologie, de logique formelle (OWL), de représentation des connaissances. A partir des travaux de l'équipe d'accueil [2, 6, 7], il devra approfondir l'état de l'art sur les approches de construction d'ontologies [3, 8], en particulier à partir de corpus textuels et de schémas de bases de données. L'étude des approches à partir de corpus sera privilégiée.

### Recherche théorique

A partir de cet état de l'art, le candidat devra préciser le protocole de construction de l'ontologie. Les verrous méthodologiques, scientifiques et technologiques liés à ce système devront être mis en évidence et des solutions devront être proposées, en particulier autour de l'usage des ontologies pour la recherche d'information et des rapports entre ontologies et thésaurus classiquement utilisés en recherche d'information.

### Application

L'ontologie devra être construite et appliquée à la plateforme de travail collaboratif ToxNuc.

### Compétences

Le candidat devra avoir des compétences en programmation JAVA et PERL.

## Contacts

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

Sylvie Ranwez Sylvie.Ranwez@ema.fr 04 66 38 70 44

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

## Références

- [1] Bachimont B. engagement sémantique et engagement ontologique: conception et réalisation d'ontologies en ingénierie des connaissances. In : Ingénierie des connaissances. J.Charlet, M.Zacklad, G.Kassel, D.Bourigault (eds). Eyrolles:Paris - Collection technique et scientifique des télécommunications, 2000; pp305-323
- [2] Baneyx A., Charlet J. et Jaulent M.-C. Building an ontology of pulmonary diseases with natural language processing tools using textual corpora. International Journal of Medical Informatics, 76 :208–15, 2007.
- [3] Biebow B., Szulman S. 1999. Terminae : A linguistic-based tool for building of a domain ontology. In Proc. of the 11th European Workshop, Knowledge Acquisition, Modelling and Management (EKAW'99), Juan-les-Pins, France, 329-334.
- [4] Bourigault D., Fabre C., Approche linguistique pour l'analyse syntaxique de corpus, Cahiers de Grammaires, n° 25, Université Toulouse - Le Mirail, 2000; pp. 131-151
- [5] Charlet J. The management of medical knowledge : between non-structured documents and ontologies. Annales des télécommunications, 62(7-8) :808–26, 2007.
- [6] Charlet J., Ingénierie des connaissances : développements, résultats et perspectives pour la gestion des connaissances médicales. Mémoire d'HDR dans la spécialité Informatique, Paris 6, 2002.
- [7] Le Moigno S, Charlet J, Bourigault D, Degoulet P, Jaulent M-C. Terminology extraction from text to build an ontology in surgical intensive care. Proc AMIA Symp. 2002;:430-4.
- [8] Maedche A., Staab S. Mining Ontologies from Text. In Proc. of the 12th European Workshop, Knowledge Acquisition, Modelling and Management (EKAW 2000), 189-202.