

# Sujet de TER: segmentation de noms de propriétés

## 1 Encadrants

Clémentine Nebut, Mathieu Lafourcade, Jean-Rémy Falleri, Xavier Dolques

## 2 Parcours

*GL, IA.*

## 3 Mots-clefs

*Traitement Automatique du Langage, Ingénierie des Modèles, OWL, Restructuration, Refactoring, Java, Eclipse, Wordnet.*

## 4 Problématique

La plate-forme Galatea est développée au sein de l'équipe D'OC, dans le cadre d'un contrat de recherche avec France Télécom. Elle permet d'analyser et de restructurer des modèles de conception ou du code source de logiciels (UML, Ecore, Java). Une extension de cette plate-forme a été développée pour permettre l'analyse et la restructuration d'ontologies au format OWL. L'analyse effectuée par Galatea est pour le moment uniquement basée sur la structure des modèles. Nous souhaitons l'étendre en utilisant des techniques de traitement du langage pour améliorer les résultats.

## 5 Travail demandé

Le travail demandé pour ce TER se décompose en deux parties :

1. Une phase de familiarisation avec la plate-forme Galatea. Lors de cette phase il sera demandé d'utiliser Galatea sur des modèles OWL et d'interpréter les résultats fournis par l'outil.
2. La spécification et l'implémentation d'un outil de segmentation de nom de propriétés. Les propriétés que nous souhaitons soumettre à une analyse de langue naturelle sont souvent nommées d'une manière qui empêche l'application directe de techniques de traitement du langage. Par exemple un nom de méthode de la forme *uploadFile* n'est pas exploitable directement. Pour pouvoir l'exploiter, il est nécessaire de segmenter ce nom (exemple : *upload file*). La façon de segmenter le texte dépend pour une grande part des conventions de nommage utilisées lors du développement de l'application (exemple alternatif : *upload\_file*). Ainsi il sera demandé d'implémenter un outil de segmentation qui peut être paramétré en fonction des conventions de nommage.

## 6 Technologies utilisées

- Java 1.5
- Eclipse
- Eclipse EMF
- Wordnet

## 7 Bibliographie

- Description du processus mis en oeuvre dans la plate forme Galatea : [1, 2],
- Extension de Galatea sur les ontologies OWL : [3].

## Références

- [1] Arévalo, G., Falleri, J.R., Huchard, M., Nebut, C. : Building abstractions in class models : Formal concept analysis in a model-driven approach. In Nierstrasz, O., Whittle, J., Harel, D., Reggio, G., eds. : MoDELS. Volume 4199 of Lecture Notes in Computer Science., Springer (2006) 513–527
- [2] Falleri, J.R., Arévalo, G., Huchard, M., Nebut, C. : Use of model driven engineering in building generic fca/rca tools. In : Concept Lattices and Applications. (2007)

- [3] Dolques, X., Falleri, J.R., Huchard, M., Nebut, C. : Adaptation d'un processus de construction d'abstraction basé IDM à des modèles bi-niveaux éléments/méta-éléments, application aux logiques de description. In : Langages et Modèles Objets. (2007)