



Proposition de sujet de stage

Indexation de documents

(Méthode structurelle)

Lieu: Laboratoire L3I, Université de La Rochelle

Contact: Romain Raveaux et Jean-Marc Ogier {romain.raveaux01, jean-marc.ogier}@univ-lr.fr

Mots clés :

- Langages : Java
- Concepts : Graphes, Mesure de similarité entre graphes.
- Outils : XML, Eclipse.
- Filière : Master 2, Images et Calculs.

Contexte:

Ce stage s'intègre dans le contexte d'un projet appelé : « ALPAGE¹ ». Ce projet traite des plans cadastraux couleurs de l'espace urbain parisien suivant différentes époques, 14^{ème} et 19^{ème} siècle. Dans ce cadre, nous donnons d'une part une interprétation sémantique à des pixels couleurs (amas de données inertes) et d'autre part, une capacité d'indexation de documents.

Description du sujet:

Le but de ce stage est d'étudier la problématique de la recherche d'information dans une collection de documents. Une fois l'information extraite des plans cadastraux, une représentation structurelle est générée, chaque plan est donc conceptualisé dans un formalisme de graphe et la question tourne donc, en un problème d'extraction de connaissance dans une masse de données structurelles. Cette approche sera implantée dans un *framework* existant opérant sur des graphes (MOGL pour MultiObjective Graph Learning [RAV06]). L'un des points importants sera la mise en œuvre d'une mesure de similarité entre graphes en adéquation avec notre problème [SOR03] et de comparer cette distance à celles déjà implémentées dans MOGL.

Pour en discuter :

Romain Raveaux et Jean-Marc Ogier {romain.raveaux01, jean-marc.ogier}@univ-lr.fr. Bureau 117.

Bibliographie :

[SOR03] : Sébastien Sorlin, mesurer la similarité de graphes étiquetés (JNPC'03).

[RAV 06] Raveaux R., Barbu E. « Graph Classification Using Genetic Algorithm and Graph Probing : Application to Symbol Recognition ».

¹ <http://alpageproject.free.fr>