

Ecole Doctorale Cultures et Sociétés

Séminaire de l'Ecole Doctorale 2010-2011, 36 heures semestrielles

Enseignant : Pierre. Fiala

Séance mensuelle : samedi 9h 30 - 16h30 ; salles informatique CMC 342

Traitement automatisé des discours et informatique textuelle

Les six séances mensuelles constituent une introduction à l'analyse des corpus textuels et de sémantique discursive. Chaque séance comprend la prise en main d'un logiciel et des problématiques qui s'y rattachent. Des exercices pratiques en linguistique de corpus et en statistique textuelle permettent de tester les fonctionnalités, de répondre aux questions pratiques. Le séminaire s'adresse aux doctorants et aux étudiants en Master 2 de Paris-12, et des autres universités dans la limite des places disponibles, titulaires du C2i ou ayant suivi avec succès un cours d'informatique textuelle. Il nécessite une inscription préalable. Pour les étudiants du Master 2 l'évaluation comprend la participation active à 4 séances et la réalisation d'une analyse informatisée.

Déroulement des séances : 3 heures de cours ; 3 heures de TD. Un accès libre aux postes de travail équipés des logiciels est prévu le mardi suivant chaque séance mensuelle.

Programme et calendrier

- 🔴 Samedi 16 octobre 2010: Etiqueter, quantifier, décrire les textes (Logiciels Cordial, Tree Tagger). Rechercher des formes, des contextes, des schémas syntaxiques, des propriétés sémantiques dans un corpus textuel (Bases textuelles littéraires et historiques, Logiciel Stella et filtres grammaticaux).
- 🔴 Samedi 13 novembre 2010: Construire et analyser des corpus lexicométriques (lexico3). Fonctions documentaires et fonctions statistiques ; typologies génériques Hypothèse, corrélations et causalités interprétatives.
- 🔴 Samedi 11 décembre 2010 : Expérimentations lexicométriques, distance intertextuelle. topographies textuelles, visualisation des résultats (Hyperbase, Astartex).
- 🔴 Samedi 15 janvier 2011 : Analyses lexicométriques et analyses de contenu (Tropes). Cooccurrences dans les énoncés et mondes sémantiques (Alceste).
- 🔴 Samedi 12 février 2011: Voisinages lexicaux ; cooccurrences dans les textes ; tableaux de propriétés textuelles (Weblex et les grammaires MPC).
- 🔴 Samedi 12 mars 2011: Les corpus multilingues : Alignements, traitements lexicométriques de corpus multilingues.

**Mars 2011 : Journées d'études
d'informatique textuelle.**