

Semaine de formation intensive NooJ
Max Silberztein, ELLIADD, Université de Franche-Comté
INALCO, 12-16 janvier 2015
2, rue de Lille, 75006 Paris

NooJ est un environnement de développement utilisé pour formaliser huit niveaux de phénomènes linguistiques : orthographe et typographie, morphologie flexionnelle et dérivationnelle, syntaxe locale et structurelle, grammaire transformationnelle et analyse sémantique. NooJ contient des outils formels adaptés à chaque type de phénomène (grammaires rationnelles, hors contexte, contextuelles et non restreintes) et permet de développer des ressources à large couverture pour de nombreuses langues.

NooJ est aussi utilisé pour analyser des corpus de textes par des linguistes mais aussi par des chercheurs en sciences sociales pour effectuer des analyses historiques, littéraires, sociologiques ou psychologiques de corpus, et aussi par des entreprises pour extraire et annoter des informations scientifiques et techniques, notamment en intelligence économique. NooJ est gratuit, est disponible en open source JAVA et fonctionne sous Windows, Mac OSX, LINUX et Unix, cf. : www.nooj4nlp.net.

Le stage s'adresse particulièrement aux étudiants de Master, doctorants et chercheurs intéressés par la linguistique et la linguistique de corpus ainsi que par l'analyse automatique de textes dans une perspective de sciences humaines et d'application industrielle. Les étudiants en Master qui peuvent et désirent valider le stage auprès de leur université doivent me prévenir dès leur inscription. Le stage est gratuit et dure une semaine : les matins sont dédiés au cours et aux travaux dirigés ; pendant les après-midis, des chercheurs et doctorants présenteront diverses applications de NooJ.

ATTENTION : les places sont limitées : il faut s'inscrire en envoyant son nom, son statut (étudiant / doctorant / chercheur / industriel) et le nom de son institution à max.silberztein@univ-fcomte.fr.

Lundi 12 janvier

9H-12H30 : **Fouille de texte** : textes et corpus, requêtes et concordances, analyses statistiques.

14H-15H : Denis Le Pesant (Modyco) : Analyse automatique des sentiments.

15H-16H : Héra Fehri (Univ. de Gabès, Tunisie) : Traduction automatique arabe > français

Mardi 13 janvier

9H-12H30 : **Lexique & morphologie** : Unités Linguistiques Atomiques, dictionnaires, morphologie.

14H-15H : Peter Machonis (Univ. Internationale de Floride) : Recensement automatique des verbes à particule de l'anglais.

15H-16H : Xavier Blanco (Univ. Autonome de Barcelone) : Dictionnaires électroniques de l'espagnol.

Mercredi 14 janvier

9H-12H30 : **Syntaxe locale** : requêtes syntaxiques, grammaires locales, annotation automatique de textes, extraction d'entités nommées.

14H-14H40 : Maximiliano Duran (Univ. de Franche-Comté) : Morphologie des verbes quechua.

14H40-15H20 : James Amani (Univ. de Franche-Comté) : Analyse morphologique des verbes du swahili.

15H20-16H : Mourad Aouini (Univ. de Franche-Comté) : Extraction d'entités nommées en Moyen Français

Jeudi 15 janvier

9H-12H30 : **Syntaxe structurelle** : Grammaires hors contexte, contraintes d'accord et arbres.

14H-15H00 : Kamal Naït-Zerrad (INALCO) : Morphologie du kabyle.

15H-16H00 : Maria Pia di Buono (Univ. de Salerne, Italie) : Ontology-based information extraction.

Vendredi 16 janvier

9H-12H30 : **Sémantique** : Analyse transformationnelle et génération de paraphrase automatique, levée d'ambiguïté sémantique, traduction

14H-16H : Philippe Lambert (Institut Jean Lamour CNRS) : Knowledge Management en Nanosciences

15H16H : Slim Mesfar (Institut des Hautes Etudes Commerciales, Tunis) : analyse trilingue d'un corpus de textes culturels.