

Réflexion sur les visualisations en sciences humaines, quels apports pour la textométrie?

Séminaire du CEDITEC,
vendredi 24 mai 2013,
10h30-17h30

Salle Keynes
Bâtiment des Sciences Economiques et de Gestion
Mail des Mèches
61 avenue du Général de Gaulle
94010 Créteil

Métro Créteil Université, Ligne 8

Séance coordonnée par
JM Leblanc - UPEC - Céditec (EA 3119) - Chantier
Textopol

[jean-marc.leblanc @u-pec.fr](mailto:jean-marc.leblanc@u-pec.fr)

<http://ceditec.u-pec.fr>
<http://textopol.u-pec.fr>

10h30

Introduction (JM Leblanc)

«Modèles tridimensionnels pour la représentation de l'état des connaissances»

Marie Pérès: Docteure en histoire ancienne, plasticienne, (UPEC)

Proposera une contribution portant sur les problèmes posés par l'interprétation des données dans le cadre de la modélisation des sites archéologiques, et présentera des productions artistiques interrogeant les modalités de la représentation des données.

«Extraction automatisée et analyse de données à partir des pages web, quelles perspectives ?»

**Jérôme Baray, PR géomarketing, (UPEC),
Chercheur à l'Institut de Recherche en Gestion**

Evoquera les problèmes de visualisation qui se posent dans le domaine des S.I.G (Systèmes d'Information Géographique) et proposera par ailleurs un outil d'extraction automatique de données à partir des pages web pour analyser le rayonnement des territoires, la notoriété et la valeur financière des marques ou encore établir un classement plus rationnel des revues de recherche.

L'intervention permettra en outre de poser les jalons d'un échange entre visualisation en SIG et textométrie et d'évoquer quelques perspectives de développement.

14h30

«Nuages arborés et analyse textuelle. Présentation de l'outil TreeCloud»

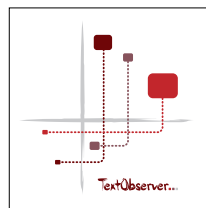
**Philippe Gambette, MCF informatique, LIGM,
Université Paris-Est Marne la Vallée**

Présentera la visualisation de textes en nuages arborés et ses applications en analyse textuelle, sur un corpus de théâtre et un corpus de presse, puis fera une démonstration du logiciel libre TreeCloud, téléchargeable sur treecloud.org, pour construire ces visualisations.

«Opérations sur graphes lexicographiques issus de corpus à des fins de visualisation sémantique»

**Romain Loth, MoDyCo - UMR 7114
Université Paris Ouest Nanterre**

Présentera des opérations sur graphes lexicaux permises par les librairies R et autres outils (définition matricielle, clustering puis travail sur les zones : sélection de parangon, intersections de zones,...), de ce que cela permet d'envisager comme visualisations de données et des pistes que cela alimente en retour sur la recherche lexicographique/sémantique .



«Propositions de visualisations pour l'analyse des corpus textuels et multimodaux. TextObserver/WebObserver, état d'avancement des développements et perspectives»

**JM Leblanc - MCF Linguistique informatique
UPEC Céditec**

Nous ferons plus particulièrement état des développements de TextObserver outil d'exploration des données textuelles multidimensionnelles et multimodales, conçu par notre équipe de recherche.

TextObserver est fondé initialement sur des préoccupations ergonomiques. Il propose des fonctionnalités originales sur le plan de la visualisation, rendues explicites par l'interactivité, et du traitement dynamique des données et des résultats textométriques. Il rend possible l'intégration de données textuelles diversifiées dans un cadre multimédia.

Il répond en temps réel aux questionnements expérimentaux comme les facteurs de la variation discursive.

Nous en présenterons les principes essentiels en prenant pour terrain d'expérimentation un corpus constitué de discours politiques rituels, (vœux de présidents de la cinquième République, de 1959 à 2012)

<http://textopol.u-pec.fr/textobserver>

«TextObserver/WebObserver: L'utilisation d'agents communicants pour l'extraction et la visualisation de corpus textuels.»

**Amani Daknou docteure en informatique,
post-doc Ceditec**

Nous proposons un système multi agent dynamique et ouvert qui vise à la recherche, la classification et l'extraction des informations nécessaires et pertinentes de pages web pour créer un corpus évolutif interrogeable en temps réel. Cette approche présente un grand avantage dans la mesure où elle déploie une collecte coopérative d'informations à base d'agents. Le système proposé permet donc de prendre en compte des contextes de recherche en considérant des regroupements de pages Web liés à des domaines particuliers. Ce système, couplé à la visualisation dynamique des résultats propre à TextObserver, constituera une fonctionnalité importante du dispositif.