

# **Séminaire ETIS : Julien Longhi**

01 Octobre 2013, 11:00 – 12:00

## **Titre du séminaire et orateur**

Des données aux corpus : problèmes et enjeux pour « fouiller » des textes et « extraire » des connaissances.

Julien Longhi, UCP (IUT), CRTF-Pôle Lascod

## **Date et lieu**

Mardi 1er octobre 2013, 11h.

Université de Cergy-Pontoise, site de St-Martin, salle à préciser.

## **Résumé**

La disponibilité de données issues du numérique, l'ouverture croissante des sources d'informations au public (Données Ouvertes), et la diffusion de contenus et d'informations sur les réseaux, donnent accès à des nouvelles sources et formes de connaissances. Face à la demande sociale, la constitution de corpus numériques est la voie privilégiée pour interpréter et exploiter les données textuelles. Parallèlement, l'impact des innovations permanentes dans le volet informatique et le traitement qualitatif des données doit être considéré.

En tant que linguiste, amorcer des interactions avec des chercheurs en informatique, dont le travail sur le traitement des données est reconnu et efficace, mais dont la dimension symbolique et sémiotique des données textuelles n'est pas le centre d'attention (car la sémantique est prise en compte par le biais de l'utilisation des taxonomies, des ontologies) me semble un enjeu tant scientifique qu'institutionnel et social. La réflexion se portera sur des thèmes tels que les concepts et les connaissances, les données et les métadonnées, le data mining, ou encore le web sémantique et la sémantique du web.