



CorText Plateforme Digitale de l'IFRIS - <http://www.cortext.fr/>

Invitation à une Formation-Atelier le 8 et 9 Décembre 2014

Marne La Vallée, Bâtiment Bois de l'Etang, Salle 219

Contexte. La plateforme numérique de l'IFRIS combine plusieurs compétences scientifiques et techniques : le traitement du langage naturel, l'extraction d'information, gestion des connaissances, analyse des réseaux complexes, la scientométrie, conception de sites Web et l'informatique. La plateforme CorText a développé une solution logicielle pour répondre aux besoins des chercheurs menant des études empiriques dans les domaines des études sociales de la science. Cette plateforme logicielle est une capacité d'appui méthodologique, de construction de jeux de données et d'analyse avec des fonctionnalités de visualisation des données. Outils et logiciels, conçus et développés au sein de la plate-forme, sont mis à la disposition des chercheurs via un service Web.

Invitation. CorText propose aux chercheurs de l'IFRIS, et notamment aux doctorants et aux post-doctorants, de participer à une Formation-Atelier le 8 et 9 décembre 2014 (début à 10:00 - fin à 18:30, avec un plateau repas sur la méridienne et une pause d'une heure). Cette Formation-Atelier est proposée par l'équipe CorText qui animera ces journées avec l'appui de collègues de la plateforme. Merci donc d'exprimer votre intérêt et votre engagement sur ces journées à l'adresse suivante : equipe@cortext.fr. Le programme est proposé ci-après.

Objectif de la Formation-Atelier. La plateforme CorText offre un large éventail d'outils et de méthodes: Statistique / analyse exploratoire de données, outils d'analyse textuels/ text mining, la Cartographie du Web, la scientométrie, outils d'analyse de données qualitatives, la cartographie hétérogène et analyse structurale des réseaux sociaux et sociotechniques, l'analyse des séries temporelles. La plate-forme vous propose ainsi : de découvrir et de vous familiariser avec certains des modules de collecte et d'analyse de données hétérogènes issues des sources variés, et de les mettre en œuvre sur des jeux de données qui sont les vôtres ou qui vous seront apportés.

Données concernées.

- Les bases de données de l'information scientifique et technologique (publications, brevets, projets de recherche)
- Les archives numérique des médias traditionnels (presse / TV / radio) et des sources comme blogs / sites Web et forums
- Des bases de données Spécifiques - des projets de recherche, entreprises, produits, etc.
- Les données recueillies par des méthodes d'enquête et des entretiens qualitatifs

Prérequis. Pour la réussite de cette opération, il est nécessaire de réunir un nombre limité de participants (maximum une dizaine), aussi, en fonction des engagements à participer la Plateforme CorText donnera la priorité aux chercheurs qui n'ont pas encore eu la possibilité de travailler avec elle, et par ordre d'arrivée des candidatures sinon. Bien entendu cette initiative sera renouvelée. Il est demandé aux participants de venir avec une courte présentation de leur projet ou de leurs intentions et éventuellement un sujet à traiter avec possiblement un jeu de données.

J'en profite pour vous rappeler que tous les Jeudis, sur place à Marne La Vallée, vous êtes les bienvenu(e)s pour présenter vos projets, vos idées, vos envies. Prendre contact à : equipe@cortext.fr

Programme

Jour 1 – 8 décembre (10:00 – 18:30)

(Repas fourni sur place)

09:30 Accueil avec un café

10:00 – 13:00 - Introduction à la Formation

Tour de table

Présentation des participants de leurs projets et intentions

Présentation du programme de la formation

Exposés, activités pratiques et discussion

La Collecte des données

- Présentation des sources de données, leurs interfaces et les options.
- Stratégies de recherche de données et de délimitation de corpus.
- L'extraction des données, formats de données, transformation de données.

L'analyse des données

- Exploration de données en utilisant un module de statistiques descriptives. Description des séries temporelles, leur définition et leur utilisation.
- Analyse des données textuelles. Champs textuels dans les sources de données. Principes et étapes de base de l'extraction de termes. Indexation automatisée de données textuelles, lemmatisation et le filtrage statistique.

13:00 – 14:00 - Repas sur place

14:00 – 17:00 - Conférence

Conférence de Noortje Marres de l'université de Londres sur les humanités numériques (<http://www.gold.ac.uk/sociology/staff/marres/>) propose dans le cadre du Séminaire LISIS. Les participants à la formation rejoindront la Conférence

17:00 – 18:30 – Suite Exposés, activités pratiques et discussions

La Visualisation de graphes

- Cartographie des réseaux hétérogènes. Dynamique des réseaux socio-sémantiques. Description mésoscopique de données, clusterisation et détection des communautés.

Jour 2 – 9 décembre (10:00 - 18:30)

(Repas fourni sur place)

Atelier pratique toute la journée : reprise des acquis de la veille, mobilisés sur une question et un corpus ou des corpus porté par chaque usager, en trois étapes :

Etape 1. La préparation de corpus : mise en œuvre et partage sur les difficultés, et discussion

Etape 2. Analyse des données : mise en œuvre et partage sur les difficultés, et discussion

Etape 3. Présentation des résultats : mise en œuvre et présentation, discussion sur les rendus et leurs sens par rapport au projet de connaissance des usagers

Discussion générale et retour d'expérience sur la journée.