

Big data conference day and challenge awards in Toulouse
January 26 2017 - 9h30/17h
TSE - Amhi Colloque, Manufacture des Tabacs,
21 allée de Brienne, 31000 Toulouse.

The registration is reserved for Master 2 students. I

In the morning, talks will be in English but the afternoon will be all in French.

Program

09h30 - 09h45: Opening of the day

09h45 - 10h30: Eric Gaussier (Ismag, Grenoble)

Title : **Classification and retrieval in large-scale text collections**

Summary : Web-scale textual collections have changed the way people retrieve documents and classify them. I'll review in this presentation standard techniques for indexing and retrieving documents in such large collections as well as approaches to classification in web-scale taxonomies, as Wikipedia or DMOZ. In particular, I'll discuss some of the basic properties of textual collections and will show that such properties influence the behavior of the methods one can adopt for classification.

10h30 -11h00: Break

11h00 - 11h45: Aurélie Vanheuverzwyn (Médiamétrie, Paris)

Title : **Audience measurement in the era of big data**

Summary : The explosion of digital data offers new opportunities to see and analyze the world, especially the media. In this context, audience measurement, which relies on samples, is often criticized because of the apparent precision of big data. Is it relevant to contrast these two approaches? For Médiamétrie, the conviction is rather that these two worlds must enrich each other.

11h45 - 12h30: Bruno Goutorbe (Cdiscount, Bordeaux)

Title : **Application of big data and data science to e-commerce**

Summary : The objective of this talk is to show what represent concretely "big data" and "data science" at Cdiscount, a leading e-commerce company in France, and how mathematical algorithms applied to big data help answer specific needs encountered in the e-commerce industry. The characteristics of the data at our disposal as well as the big data architecture will be succinctly described, and I will illustrate the range covered by algorithms written by data scientists through practical examples in production on the internet website of Cdiscount. I will show that user session data and information associated with products allow to optimize algorithmically features as diverse as relevance of the search engine, recommendation systems or traffic acquisition.

12h30 – 14h: Lunch break

14h00 -14h45: Céline Castets-Renard (UT1C, Toulouse)

Title : Law and big data

Summary : L'objectif de la présentation est de présenter comment le droit appréhende le big data. Le big data n'est pas réglementé en tant que tel. Seules certaines données le sont. Des dispositions spécifiques s'appliquent ainsi à des éléments du big data : selon la nature des données en vue de le restreindre (données personnelles, données protégées par le droit d'auteur, données protégées par le secret des affaires), selon la structuration des données (droit sui generis des bases de données) ou encore selon certains usages des données (open data). Les objectifs du législateur peuvent donc être contradictoires : encourager le secret et l'exclusivité sur les données ou au contraire contraindre leur ouverture.

Si le big data n'est pas réglementé en tant que tel, le législateur exige désormais que les traitements sur les données soient transparents et loyaux en consacrant des principes de loyauté et transparence algorithmique à l'égard de l'administration et des plateformes numériques.

14h45 - 15h00: Challenge and Awards presentation by the organizers

15h00 - 15h30: Break

15h30 - 17h00: Students presentations and discussion

Space is limited so please fill the pre-registration form on <https://ecole.tse-fr.eu/en/registration-form-2017-01-26>

Your registration will be confirmed by email.



THE JEAN-JACQUES LAFFONT
DIGITAL CHAIR



