

Séminaire MIDI : Claudio Lucchese

26 Mai 2015, 14:00 – 15:30

Titre du séminaire et orateur

Big Data and Web Search.

[Claudio Lucchese](#), ISTI, Pise (Italie).

Date et lieu du séminaire

Mardi 26 mai 2015, 14h.

Université de Cergy-Pontoise, site de St-Martin 1, bât. A, 5ème étage, salle 570.

Abstract

Data mining tools are essential in many information search and filtering tasks, e.g., document ranking in web search. This talk explores a few opportunities and challenges in mining big data for Web Search.

We first introduce the pattern mining task, and discuss a recent algorithm for mining the most interesting patterns from binary datasets according to the Minimum Description Length principle. Then, we discuss some results about fast scoring of web documents with complex and expensive machine learning models, i.e., thousands of gradient boosted regression trees. Finally, we illustrate how user generated structured content such as Wikipedia, can be exploited to annotate text and improve its analysis.

We show a case study where text annotation is exploited together with Twitter data to improve personalised news recommendations.