# Development of a distributed recommender system using the Hadoop Framework

Raja Chiky, Renata Ghisloti, Zakia Kazi Aoul

LISITE-ISEP
28 rue Notre Dame Des Champs
75006 Paris
firstname.lastname@isep.fr

**Abstract.** Producing high quality recommendations has become a challenge in the recent years. Indeed, the growth in the quantity of data involved in the recommendation process pose some scalability and effectiveness problems. These issues have encouraged the research of new technologies. Instead of developing a new recommender system we improve an already existing method. A distributed framework was considered based on the known quality and simplicity of the MapReduce project. The Hadoop Open Source project played a fundamental role in this research. It undoubtedly encouraged and facilitated the construction of our application, supplying all tools needed. Our main goal in this research was to prove that building a distributed recommender system was not only possible, but simple and productive.

## 1  Introduction

The amount of information in the web has greatly increased in the past decade. This phenomenon has promoted the advance of the recommender systems research area. The aim of Recommender Systems is providing personalized recommendations. They help users by suggesting useful items to them, usually dealing with enormous amounts of data. Amazon, for example, that has incorporated recommender systems to personalize the online store for each user, has recorded in 2003 more than 29 million users and several million catalog items Linden et al. (2003).

Many recommender systems approaches have been developed in the past decade, but a considerable amount of them were constructed and evaluated with small datasets. Furthermore, the volume of web information has greatly increased in the last years, and for that, several recommender systems suffer from performance and scalability problems when dealing with larger datasets.

Our main goal in this paper is describing a method to possibly overcome these issues. We propose a distributed recommender system, and we intend to demonstrate that it could be easily developed and present good results. We chose The Slope One Lemire and Maclachlan (2005) as recommender algorithm and we study the MapReduce Dean and Ghemawat project to construct such distributed system.