PEC 1: Work definition and planning

1. Title proposal

The chosen title for the final work will be “Construction of a Product Data Matching System for an e-Commerce Aggregator”.

2. Keywords

Product Matching, e-Commerce, Machine Learning, Data Mining, Natural Language Processing.

3. Abstract

In this work, we propose the construction of a system using Machine Learning and Natural Language Processing (NLP) techniques to automatically detect matches between products whose data come from different sources.

4. Proposal description and relevance

Nowadays we can find on the Web a bunch of different e-Commerce aggregators, which are websites that collect data of products from different online shops such as Amazon, MediaMarkt or ToysRUs with the aim of showing their customers the best options to buy a certain product. Google Shopping and Idealo are two famous examples of e-Commerce aggregators.

However, the problem of identifying the same products collected across many sources, known as “Product Matching”, must be addressed in order to offer an optimal service. In most cases this is not a trivial task, as data is presented in different formats depending on the source they come from.

5. Personal motivation

There are two main reasons involved in the choice of this proposal:

* Once built, if the system gets good results matching products, it can be used to improve the quality of an e-Commerce aggregator which has been developed in the company the student works for. This is a task which is currently being performed manually by a group of matching operators.
* The problem to be solved induces to implement different tools and techniques which have been seen throughout these studies (Machine Learning, NLP, …) and also during the previous studies of the student (Computer Science Degree).

6. Objectives