# Resultado de imagen de uocPEC 2: State of the art

Author: Teguayco Gutiérrez González

Studies: Master’s Degree in Data Science

## 2.1 Introduction

As explained in the introductory section, the problem of matching products coming from different sources brings with it the challenge of identifying identical products whose data may be missing (for instance, universal identifiers such as EAN or UPC codes could not be found) or presented in multiple formats. These difficulties make this matching task to be a non-trivial one and, as a result, a lot of research studies and tools have already arisen, most of them implementing Natural Language Processing (NLP) and Machine Learning techniques.

In this section, a comprehensive review of these previous works will be done based on the approaches they pose to analyse how they implement the solutions for this issue.

## 2.2 Approaches

### 2.2.1 Text similarity

### 2.2.2 Attribute extraction

- Attribute extraction from product titles or descriptions

- Named Entity Recognition Problem

- Conditional Random Fields (NO mencionar StanfordNER; se mencionará su uso en el apartado 3 de la memoria

- the another one used in the same work

### 2.2.3 Image recognition

Matching products identifying these by performing image recognition on them could constitute the hardest approach of all. Some of the

## 2.3 References