

Prospecção e Análise de Dados

2º Sem 2015/2016

Practical Work II a)

(Automatic Extraction of Explicit and Implicit Keywords)

Implement automatic extractor of Explicit and Implicit Keywords from documents, based on the approach studied in the theoretical lectures. Consider the following steps:

- 1) Extract Relevant Expressions (REs) from a set of several documents, by using LocalMaxs extractor you have implemented. Create adequate criteria to select the most informative REs.
- 2) Take the most 10 -15 informative REs of each document and consider them as the Explicit Keywords of the document.
- 3) Calculate the similarity between REs and use that to find the most correlated REs which are semantically close to the Explicit keywords of each document and, although not explicitly written in the document, they may be used as its Implicit Keywords.