

## **MAPI Thesis topic:**

**Creating a personalized alert system focusing on novel topics with the help of web/text mining and data stream techniques**

## **Context and Objectives**

The text mining techniques that are available nowadays have advanced a great deal and so it is relatively easy to construct systems that not are able to classify documents as relevant (or not), extract relevant pieces of information to satisfy user's needs or analyze similarities among given documents.

In this project the aim is to apply the techniques to articles or tests in a chosen information source (journal(s) / website(s) / social network (e.g. twitter)), use web/text mining techniques to examine these periodically and return pieces of information that are relevant to the user (alerts), while taking into account his profile or a particular query (as in Information Retrieval). The profile could thus provide a focus on a particular type of activity or a particular area (e.g. economics, finance or management).

We note that a particular topic may often be described or discussed in a series of related articles or texts, while their focus and contents may evolve with time. The aim of this thesis project is to explore not only how to present this information in a compact form using informative words (keywords) or automatic summaries, but also how to detect that there has been a shift of focus or some novel concepts have arisen that could be relevant to the user.

## **Supervision:**

Pavel Brazdil (pbrazdil@inescporto.pt) and João Gama (jgama@fep.up.pt)

## **R&D Unit / Unidade de investigação proponente:**

LIAAD INESC TEC

## **References**

Feldman, R., Sanger, J.: *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, New York (2007)

Weiss, S. M., Indurkha, N., Zhang, T., Damerou, F. J. *Text Mining: Predictive Methods for Analyzing Unstructured Information*. Springer, New York (2005)

J. Gama, A. Carvalho, C. Lorena, K. Faceli, M. Oliveira (2012), *Extração de Conhecimento de Dados Silabo*

J. Gama: *Knowledge Discovery from Data Streams (2010)*, Chapman & Hall/CRC Press

Pavel Brazdil: Classification of Documents using Text Mining package "tm", Slides of ECDII

Luis Trigo: Análise de proximidade entre investigadores de alguns centros de I&D da Universidade do Porto usando *text mining* sobre bases de dados bibliográficas, Tese de mestrado ADSAD, defendida em Jan. 2011.

Kaustubh Patil, Pavel Brazdil: SumGraph: Text Summarization using Centrality in the Pathfinder Network, International Journal on Computer Science and Information Systems, Volume 2, Number 1, page 18--32 - 2007