# SCIENCE*SPRING*DAY

DEPARTAMENTO DE INFORMÁTICA

Carlos Gonçalves

PhD Student FCT-UNL Doctoral Program in Informatics (Supervisor: Prof. José Cardoso e Cunha; Co-supervisor: Prof. Joaquim Ferreira Silva)

- (2002) MSc Informatics Engineering
- (2011-) Adjunct Professor at ISEL

Statistical and Distributed Text Mining

COMPUTER SYSTEMS

CHi CENTER FOR INFORMATICS AND INFORMATION TECHNOLOGIES

Objectives

- Propose a computational model for statistical-based text mining and its implementation in a parallel/distributed environment

- Design a Problem Solving Environment capable of extract relevant expressions form a large language corpus.
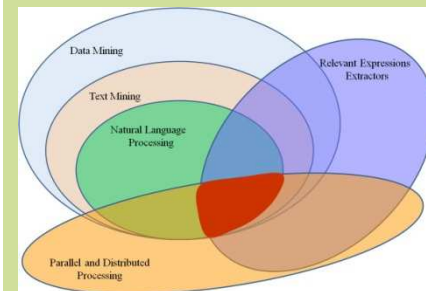
Methodology

- Analyze the limitations of the existing state-of-the-art solutions

- Develop parallelization and mapping strategies

- Experimental evaluation with text mining applications
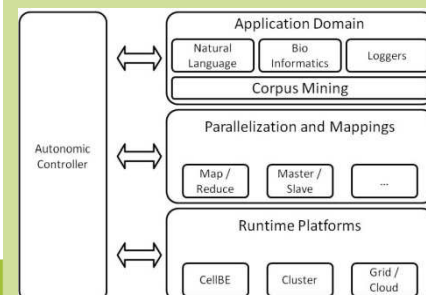
Expected Results

- A Problem Solving Environment, based on an abstract machine providing the basic operators to work over a language corpus. Provide support for different metrics for extracting relevant expressions.
- Publications:

[1] C. Goncalves, L. Assuncao, and J. C. Cunha, "Data analytics in the Cloud with Flexible MapReduce Workflows," in *Cloud Computing Technology and Science (CloudCom), In IEEE 4th International Conference on*, Dec. 2012, pp. 427 –434.

[2] L. Assuncao, C. Goncalves, and J. C. Cunha, "Autonomic Activities in the Execution of Scientific Workflows: Evaluation of the AWARD Framework," In IEEE *9th International Conference on Autonomic Trusted Computing (ATC 2012)*, Sept. 2012, pp. 423 –430.

[3] L. Assuncao, C. Goncalves, and J. C. Cunha, "On the Difficulties of Using Workflow Tools to Express Parallelism and Distribution - A Case Study in Geological Sciences," Proceedings of the International Workshop on Workflow Management of the International Conference on Grid and Pervasive Computing. IEEE Computer Society, 2009, pp. 104–110.