

# EADW - Political Personalities

68206 - Artur Balanuta  
68210 - Dário Nascimento  
Instituto Superior Técnico

## Abstract

*Traditional standalone file systems aren't enough for current users demand because they need to store thousands of media files and access them everywhere at anytime. Recent file system must solve this problem being distributed and providing high-availability, replication and being reliable. We describe our PADI-FS design which provides a distributed file system interface. This file system is based on three metadata servers and hundreds of data servers which are controlled by client applications.*

## 1. Introduction

## 2 Functional Requirements

### 2.1 News Collection and Storage

### 2.2 News Search

### 2.3 Extraction of Named Entities

A extração de nomes de entidades no nosso sistema é suportada por uma lista nomes de personalidades previamente conhecida. O objectivo é que novos nomes sejam adicionados a esta lista e que nomes já existentes sejam reconhecidos com boa confiança.

Com base numa lista de entidades previamente conhecida, gerámos uma tabela de nomes próprios designada de "ProperNounTable" e uma

tabela de Entidades com uma popularidade pré-definida e a popularidade adquirida nas notícias políticas designada de "EntitiesTable".

Para cada nome que o NLTK indentifica como potencial nome próprio, verificamos:

- Se o nome não pertence à lista de nomes próprios errados conhecida ("O", "A", "Desde", entre outros)
- Se o nome pertence à lista de nomes próprios conhecidos. Caso pertença, é concatenado com o nome em formação. Caso contrário é definido como candidato a nome próprio. Se o nome seguinte for também um nome próprio, então ambos são adicionados à "ProperNounTable", expandindo a lista de nomes próprios conhecida.
- Se os nomes próprios constituintes do nome da entidade desconhecida correspondem em mais de 60% ao nome de uma das entidades conhecidas. De todas as entidades conhecidas candidatas a este nome, é seleccionada a que tem maior reputação nas notícias e a que tem maior reputação pré-definida. Caso não haja nenhuma entidade conhecida com o nome, este nome é adicionado como nova entidade.

Ao nível do processamento de texto, obtámos por unificar todos os nomes sem acentos graves ou agudos de modo a que palavras como: "Luís" ou "Luis" fossem equivalentes. Esta situação permitiu aumentar em 3 terços o número

de entidades detectadas. Em particular o número de ocorrências do ministro "Vitor Gaspar" aumentou imenso.

#### **2.4 Sentiment Analysis**

#### **2.5 Other Functionalities**

#### **2.6 Other Functionalities**

#### **2.7 Other Functionalities**

### **3. Conclusions and Future Work**

### **Referências**

- [1] S. Ghemawat, H. Gobioff, and S.-T. Leung. The Google file system. *ACM SIGOPS Operating Systems Review*, 37(5):29, 2003.
- [2] L. Lamport. The part-time parliament. *ACM Transactions on Computer Systems*, 16(2):133–169, 1998.