

## Miniprojecto 2 de Língua Natural

### MEIC-Tagus

### 2015/2016

## 1 Objectivos gerais

Melhorar o sistema anterior usando os conhecimentos adquiridos na disciplina de LN.

## 2 Dados de entrada

O aluno tem de novo à sua disposição os mesmos 2 ficheiros do projecto anterior, PerguntasPosSistema.txt e AnotadoAll.txt.

## 3 Tarefas detalhadas

As tarefas a realizar são as seguintes:

- a) Implementar, em Python, um programa que, dada uma frase, devolva uma resposta, tendo por base estratégias que:
  - a. Tenham em conta semelhanças entre answers para a escolha da mais frequente. Pelo menos três medidas de semelhança devem ser implementadas.
  - b. Dado um user input, tenham em conta triggers semelhantes ao user input e não apenas iguais. Pelo menos três medidas de semelhança devem ser implementadas (podem ser as mesmas que as usadas em a), desde que façam sentido). De notar que **apenas deve ser accionada esta hipótese se nenhum trigger igual for encontrado**.
  - c. Usem informação morfológica no cálculo das semelhanças anteriores. Ou seja, as medidas anteriormente implementadas devem funcionar com e sem esta informação. Fica ao critério dos alunos usar esta informação como bem entenderem. Recomenda-se fortemente que usem as ferramentas do NLTK para a análise morfológica.

Podem optar por responder “Não percebi” caso o sistema não tenha confiança para responder.

Nome da função: sss

Parâmetro 1: ficheiro PerguntasPosSistema.txt

Parâmetro 2: pergunta

Resultado: resposta escolhida pelo sistema ou “Não percebi”.

- b) Para avaliar os resultados obtidos com cada uma das técnicas implementadas, usar a função de avaliação implementada no MP1.

Nome da função: myAvalia

Parâmetro 1: ficheiro AnotaAll.txt

Parâmetro 2: ficheiro com perguntas

Resultado: accuracy do sistema.

## 4 Entrega

### 4.1 Quando?

Este mini-projecto deverá ser entregue até às 23h59 horas do dia 13 de Novembro de 2015 (não serão aceites trabalhos a partir dessa data).

### 4.2 Onde?

As entregas deverão ocorrer via Fénix, num zip (e não rar) constando o número do grupo no nome do ficheiro entregue (será descontado 1 valor, caso este requisito não seja satisfeito).

### 4.3 O quê?

Material a entregar:

- Código (por favor, nada de paths locais)
- Readme a explicar EXACTAMENTE como correr as duas funções
- Relatório (com um máximo de 5 páginas + potencial bibliografia) que consistirá nas seguintes secções:
  - Capa: identificação clara do grupo e dos autores (NÃO façam uma capa à parte, por favor);
  - Introdução: breve introdução ao problema que vão tratar;
  - Arquitectura proposta: pode ser ilustrada com uma figura que explicita que funções/módulos (ex: normalizador) são chamados e onde, qual o fluxo de dados, etc.
  - Experimental Setup: breve descrição dos corpora usados, explicação do que usaram para desenvolvimento, teste, etc., definição da medida de avaliação;
  - Resultados: apresentação das várias técnicas implementadas, resultados e comentários aos resultados obtidos;
  - Conclusões e eventual bibliografia.

Sugestão: não serão avaliados por isto, mas recomendo que façam o relatório em Latex, em especial se nunca o usaram. É maravilhoso.

## 5 Avaliação do Projecto

- Qualidade do relatório (organização de acordo com o pedido, clareza do texto, correcção do português, etc.): 5 valores;
- Qualidade da proposta em si, em especial respeitando a metodologia estudada: 12 valores.
- Qualidade do readme (se realmente conseguir pôr tudo a correr sem problemas no meu computador) e do código (em relação a este último só vou ver se está bem comentado e com nomes compreensíveis): 3 valores