

Ficha 7

Scripting no Processamento de Linguagem Natural

25 de Abril de 2015

Escreva um módulo que lhe permita fazer o reconhecimento de nomes e entidades em texto. O módulo deverá:

- Ser configurável através de classes de palavras como por exemplo preposições, títulos, lugares, nomes próprios, apelidos, tempo, etc;
- Possibilitar o uso de estruturas de dados hierárquicas que relacionem palavras para melhorar o seu desempenho;
- Ter formas de configurar heurísticas que possam ser utilizadas no reconhecimento de nomes e entidades;
- Permitir definir novas entidades a serem reconhecidas e disponibilizar a forma de apresentá-las;
- Definir situações de excepção;
- Disponibilizar ferramentas auxiliares como por exemplo formas de partir texto por frases e palavras.

Contexto

O seu módulo deverá ser testado numa aplicação que pegue em notícias de jornal e que reconheça os nomes e as entidades aí presentes. A título de exemplo, considere a seguinte notícia:

Bruxelas diz que taxa turística de Lisboa contraria leis europeias **Inês Boaventura**

Numa resposta a uma eurodeputada do PSD, a Comissão Europeia lembra que a legislação em vigor "proíbe a discriminação em razão da nacionalidade". Cláudia Aguiar conclui que a "taxa de Costa é ilegal", o que "obriga" o município a retirá-la.

A Comissão Europeia afirma, quando questionada sobre a legalidade da taxa turística criada pela Câmara de Lisboa, que "a legislação da União Europeia (UE) proíbe a discriminação em razão da nacionalidade". Isentar quem tem domicílio fiscal em Portugal do seu pagamento pode, alerta a instituição, "colocar numa situação de particular desvantagem" quem não cumprir esse critério.

Em resposta a uma pergunta feita pela eurodeputada portuguesa Cláudia Monteiro de Aguiar, a Comissão Europeia "recorda que, em termos gerais, a legislação da UE proíbe a discriminação em razão da nacionalidade". Algo que se aplica às "discriminações ostensivas em razão da nacionalidade", mas também "às formas dissimuladas de discriminação" que, aplicando critérios como o da residência fiscal, "levam ao mesmo resultado, uma vez que são susceptíveis de afectar em maior grau os cidadãos da UE do que os nacionais e de os colocar numa situação de particular desvantagem".

A Comissão Europeia lembra ainda, na resposta à que o PÚBLICO teve acesso, que uma "consulta aberta" realizada a actores do sector do turismo, entre Dezembro de 2013 e Março de 2014, "identificou a proliferação de taxas turísticas, incluindo as taxas locais, como excessivas e com efeitos negativos sobre a competitividade da indústria do turismo".

No caso concreto de Lisboa, a câmara aprovou a aplicação de uma taxa turística, no valor de um euro, a cada passageiro que desembarcar no Aeroporto Internacional de Lisboa. De acordo com o Regulamento Geral de Taxas publicado no fim de 2014, ficam isentos desse pagamento "os passageiros com domicílio fiscal em território nacional".

Face à resposta da Comissão Europeia, Cláudia Aguiar conclui que a "taxa de Costa é ilegal" e defende que tal "obriga" o município de Lisboa a abandoná-la. A eurodeputada do PSD, que na pergunta que dirigiu à Comissão Europeia sustentava estar em causa "uma clara violação da liberdade de circulação de pessoas na União", aproveita para lançar uma farpa ao ex-presidente da câmara e actual secretário-geral do PS, responsabilizando-o pela criação de "taxas ilegais que afectam directamente o crescimento e a competitividade do país".

"A Câmara de Lisboa não tem conhecimento da opinião de nenhuma instituição e muito menos de nenhuma decisão judicial contrária ao modelo definido", informa o presidente da autarquia, Fernando Medina, através de um elemento do seu gabinete.

Em comunicado de imprensa, Cláudia Aguiar sublinha ainda "os efeitos perversos" que a criação da taxa turística poderá ter "numa altura em que o sector turístico tem vindo a mostrar uma enorme resiliência e que tem crescido acima da média Europeia".

Desde que o então presidente da Câmara de Lisboa António Costa anunciou, em Setembro de 2014, que pretendia relançar o debate em torno da criação de uma taxa turística (uma ideia que tinha sido lançada em 2010 mas que acabou por não ir avante nessa altura por ter sido mal recebida pelo sector do turismo), este tem sido um dossier envolto em polémica.

Uma das principais vozes de oposição foi a do ministro da Economia, Pires de Lima, que desafiou António Costa a "resistir à tentação" de criar "taxas e taxinhas" na área do turismo. Um apelo que caiu em saco roto: com a apresentação do orçamento da câmara para 2015, ficou a saber-se que a taxa turística era mesmo para avançar e que o município pretendia cobrar um euro a quem chegasse a Lisboa por via aérea ou marítima e um euro por cada noite de dormida num hotel da cidade.

Depois de muitas críticas da oposição na câmara e de algumas alterações ao calendário inicial, ficou estabelecido que as taxas de dormida e de chegada por via marítima só começariam a ser cobradas a 1 de Janeiro de 2016, enquanto a taxa de chegada por via aérea sê-lo-ia a partir de 1 de Abril de 2015. A dois dias dessa data, a câmara e a ANA assinaram um protocolo que prevê a assunção pela gestora aeroportuária do pagamento da taxa pelo menos durante o ano de 2015, substituindo-se, assim, aos passageiros.

Mais uma vez as críticas não se fizeram esperar: a oposição na câmara veio exigir esclarecimentos sobre o porquê de a ANA ter decidido pagar um valor entre 3,6 e 4,4 milhões de euros, e Porto, Maia e Faro apressaram-se a vir dizer que também queriam uma taxa turística paga pela gestora aeroportuária.

No caso da Câmara do Porto, Rui Moreira já fez saber que considera que a introdução de taxas de aeroporto e de dormidas não deve ser encarada como um "tabu", mas sublinhou que esta é uma medida que deve ser avaliada com "uma prudência infinita". "Temos de ter cuidado e não matar a galinha dos ovos de ouro. É preciso ver como o mercado se comporta e como isso pode induzir desconfiança nos pequenos operadores privados", disse o autarca no fim de 2014.

Neste texto temos vários tipos de entidades dos quais se mostram os seguintes exemplos:

- Aeroporto Internacional de Lisboa
- ANA
- António Costa
- Câmara de Lisboa
- Câmara do Porto
- Cláudia Aguiar
- Cláudia Monteiro de Aguiar
- Comissão Europeia
- Fernando Medina
- Inês Boaventura
- ministro da Economia
- município de Lisboa
- Pires de Lima
- presidente da Câmara de Lisboa
- PS
- PSD
- PÚBLICO

- Regulamento Geral de Taxas
- Rui Moreira
- UE
- União Europeia

Serão dadas bonificações se o módulo identificar corretamente conceitos como pessoas, locais (e.g., Lisboa, Porto, Maia, Faro), instituições, organizações, datas, e conseguir perceber relações como por exemplo que o presidente da Câmara de Lisboa é o António Costa, o ministro da Economia é o Pires de Lima, que Cláudia Aguiar é a mesma pessoa que Cláudia Monteiro de Aguiar ou que UE é abreviatura para União Europeia.

Avaliação

Quem Grupos de duas pessoas

Quando Até ao fim do dia **18 de Maio de 2015** (não serão avaliadas entregas posteriores)

O quê Deverá ser entregue a distribuição do módulo (i.e., usando o **make dist**) contendo o seguinte material:

- Código do módulo
- Código da aplicação que utiliza o módulo para o reconhecimento de entidades em Português
- Testes (e.g., usando **Test::More**)
- Documentação (usando **PerlPod**)
- Relatório (usando **PerlPod**)

Como A avaliação do módulo irá utilizar textos recolhidos em jornais Portugueses para averiguar a capacidade da aplicação. Para além disto, a avaliação irá também olhar para a qualidade do código escrito, dos testes, da documentação e do relatório produzido.