

 <p>UNIVERSIDADE DE COIMBRA FACULDADE DE CIÊNCIAS E TECNOLOGIA Departamento de Engenharia Informática</p>	<p>Trabalho nº 3 de Algoritmos e Estruturas de Dados 2017-2018 – 2º Semestre LEI</p> <p>CORRETOR ORTOGRÁFICO</p> <p>Data de Entrega Relatório (incluindo submissão do código¹): 28 de Maio de 2018 12:00</p>
<p><i><u>Nota Importante:</u> A fraude denota uma grave falta de ética e constitui um comportamento não admissível num estudante do ensino superior e futuro profissional licenciado. Qualquer tentativa de fraude pode levar a anulação da componente prática tanto do facilitador como do prevaricador.</i></p>	

No presente trabalho pretende-se que o aluno evidencie as capacidades de tomada de decisão adquiridas na disciplina, e fora do espaço da mesma.

Desenvolvimento de um Corretor Ortográfico

Pretende-se desenvolver um corretor ortográfico para a Língua Portuguesa com sugestão da palavra correta quando esta contem erros tipográficos.

Em termos gerais um corretor ortográfico compara as palavras do texto com as palavras num dicionário. A ortografia é aceite se as palavras são encontradas no dicionário, caso contrário, para cada palavra não encontrada, o corretor pergunta ao utilizador se pretende corrigir essa palavra. Em caso afirmativo propõe palavras com maior probabilidade de serem as palavras corretas. Se o utilizador não aceita nenhuma das correções pode indicar que a palavra está correta e esta vai ser incluída no dicionário.

O programa mantém um dicionário e um histórico de correções (tuplos <palavra inicial> <palavra corrigida> <número de ocorrências>). O histórico de correções deve ser usado prioritariamente a qualquer outro modelo de erros tipográficos implementado.

Vamos assumir que o corretor ortográfico a desenvolver não faz análise de contexto.

¹ Proximamente será indicado o procedimento para submissão do código deste trabalho.

Fontes de informação (não exaustivo)

Peter Norvig, How to Write a Spelling Corrector.

<https://norvig.com/spell-correct.html>

Algumas fontes possíveis para alimentar o dicionário inicial a criar pelo aluno:

<https://www.dicio.com.br/lista-de-palavras/>

<http://natura.di.uminho.pt/download/sources/Dictionaries/wordlists/>

Listas de frequência de palavras

<https://www.wordfrequency.info/free.asp?s=y>

<https://www.linguateca.pt/acesso/tokens/formas.ancib.txt>

Entradas / Saídas

O corretor recebe uma lista de palavras terminada pela palavra FIMFIM. Para cada palavra com discrepâncias relativamente ao dicionário vai perguntar ao utilizador se quer corrigir ou guardar no dicionário. Se quer corrigir, o programa vai sugerir uma alteração que o utilizador deve aceitar ou rejeitar. Se rejeitar o corretor passa para a alternativa seguinte até ao máx. de 5 alternativas. Se após as 5 alternativas nenhuma for aceite devolve a palavra CORREÇÃO-NÃO-ENCONTRADA.

Formato do Relatório

O relatório, QUE DEVE SER PREPARADO NUM EDITOR DE TEXTO, deve ter no máximo 3 páginas usando ARIAL 11 pts e incluir os seguintes elementos:

- (MÁX. 1 linha por estrutura) Enumeração das estruturas de dados usadas, com informação do que guardam (ex. BLK – dicionário de palavras).
- (MÁX. 2 linhas por estrutura) Justificação para a escolha da estrutura e alternativas (ex. porque as BLK em condições adequadas têm tempo de acesso sublinear; alternativas BLK95 ou BLK100).
- (MÁX 5 linhas) fontes de dados usadas (ex. dicionário; frequência de palavras; erros tipográficos mais comuns; etc.) incluindo endereços URL.
- (MÁX 1 página) descrição do modelo de erros tipográficos implementado.
- (MÁX 1 página) um exemplo de utilização do programa.

Bom trabalho,

Carlos Lisboa Bento