# 1 Objectivos - Uso e implementação de Tabelas de Hash

Utilização de tabelas de dispersão e outras estruturas de dados, para implementar um corrector ortográfico. O aluno deve providenciar juntamente com os ficheiros pedidos a implementação de Tabelas de dispersão, tal como foram tratadas nas aulas práticas. Deve apresentar a implementação de tabelas de acesso linear e de acesso quadrático. Para a construção da tabela que contem o dicionário deve usar uma das referidas tabelas de (linear, quadrática ou de dispersão dupla), para a tabela que contem as sugestões deve usar uma tabela de tipo diferente da que usou para o dicionário.

# 2 O trabalho

## 2.1 Descrição

Pretende-se a criação e utilização dum corrector ortográfico, que seja usado sobre um determinado texto. A sua aplicação deve ler o texto e palavra a palavra e verificar a ortografia no dicionário. Se a palavra lida é encontrada no dicionário, assume-se que está correcta, caso contrário assume-se uma grafia incorrecta. As palavras com grafia incorrecta são guardadas e encontradas sugestões para a sua correção. Neste trabalho pretende-se que construa um corrector ortográfico, que receba um texto, identifique as palavras incorrectas no texto(i.e. mal escritas) e que faça sugestões de correcção. O output do trabalho deverá ser um ficheiro de texto que contem as palavras mal escritas do documento original e as sugestões de correcção, uma por linha. O desempenho do corrector pode ser melhorado, guardando numa outra tabela de hash, as palavras incorrectas já encontradas, por forma a não repetir a procura de sugestões se tal trabalho já tiver sido realizado, p.e. no mesmo texto aparece o mesmo erro mais do que uma vez.

## 2.2 Implementação

São fornecidos para a realização do seu trabalho dois ficheiros em Português:

- um, com o texto a verificar
- outro, com uma lista de palavras existentes no português. Esta lista, vulgarmente designada por dicionário, encontra-se no formato de uma palavra por linha. Existem outros dicionários mais completos, mas assumir-se-á a correcção da ortografia, por presença da palavra no dicionário que lhe é fornecido. Isto é, se uma palavra não se encontra neste dicionário, assume-se que está mal escrita.

Deverá providenciar o ficheiro spellChecker.c que deverá conter as funções com os nomes e funcionalidades seguintes:

- `lerDicionario` - lê o ficheiro com o dicionário e armazena-o numa tabela de hash

- `spellCheck` - lê o ficheiro de texto e gera a tabela com os erros e sugestões. Esta função deverá ser responsável por ler o ficheiro de texto e gerar uma tabela de dispersão com os erros encontrados e respectivas sugestões de correcção. O ficheiro deve processar-se do seguinte modo:
  - Para cada palavra do ficheiro, verifique-se se existe no dicionário, se sim, não se faz nada, senão verifica-se se existe na tabela de sugestões, se sim, não faz nada (já lá está a palavra e as sugestões!), senão adicione-se à tabela a palavra mal escrita e as sugestões possíveis. As sugestões possíveis são geradas pelas seguintes regras:

    1. Adicionar um caracter à palavra
    2. Remover um caracter da palavra
    3. Trocar caracteres adjacentes. Se após a aplicação de qualquer das regras a palavra obtida existe no dicionário, é uma palavra válida, deve adicionar-se às sugestões possíveis, i.e. à tabela de sugestões, na entrada correspondente à palavra mal escrita, a nova sugestão de correção.

- `geraOutput` - que gera o ficheiro de output nas condições enunciadas. Este ficheiro poderá ter o formato que entender mas um par (palavra mal escrita → sugestão de correção) por linha no ficheiro, é perfeitamente aceitável

- `main` - Esta função deverá ter duas variáveis *fileDic* e *fileTex*, que definem o nome completo dos respectivos ficheiros(dicionário e texto). A execução do programa será realizada por chamada das funções anteriormente indicadas.

Deverá também submeter as implementações das tabelas de hash e outras estruturas que use no seu programa. Como habitualmente os ficheiros .h e c. devem ser submetidos.

### 2.3 Entrega

O trabalho será realizado em grupos de 2 elementos(+/- 1). A data limite para a submissão do trabalho é 30 de Junho de 2024, sendo realizada a submissão pelo moodle, nos moldes habituais. Todos os ficheiros deverão ser "zipados" e submetidos num único ficheiro com o(s) número(s) do(s) aluno(s) que realizaram o trabalho e o nome do trabalho(ex: "NNNNN-MMMM-corrector.zip"). Não necessita submeter o dicionário, mas se quiser enviar o ficheiro de usou para os seus testes, pode submetê-lo. Os trabalhos serão apresentados (4 e 5 de Julho). Todos os membros do grupo têm de realizar a apresentação.

Trabalho Final EDA1

hashlinear.h

hashlinear.c

spellChecker.h

spellChecker.c

lerDicionário → ler o ficheiro "portugese.txt";
atribuir cada palavra lida numa tabela de hash.

spellCheck → ler o ficheiro de texto de entrada;
gerar uma tabela com os erros e sugestões.
Para cada palavra do ficheiro:
Verificar se existe no dicionário:
-Sim:
Não se faz nada
-Não:
Verificar se existe na tabela de sugestões:
- Sim:
Não se faz nada
-Não:
Adiciona-se à tabela de erros e procura-se sugestões de correção.
Regras para as sugestões:
1. Adicionar um caractere à palavra;
2. Remover um caractere à palavra;
3. Trocar caracteres adjacentes à palavra;
(Se após qualquer aplicação for formada uma palavra existente no dicionário é adicionada às sugestões)