UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
Bacharelado em Ciência da Computação e
Engenharia da Computação
INF 01203 – Estruturas de Dados

TRABALHO FINAL – ETAPA 01
Identificação de Erros Ortográficos em Textos

1 Objetivo

Este trabalho tem por objetivo proporcionar aos alunos a oportunidade de aplicar os conhecimentos adquiridos e as estruturas de dados desenvolvidas em aula na solução de um problema que utilize várias dessas estruturas.O trabalho prático da disciplina está dividido em **duas** etapas. A primeira etapa, descrita neste enunciado, envolve a utilização de estruturas de dados do tipo lista e/ou árvores.

2 Especificação da Aplicação

A aplicação modelada corresponde à ***identificação de erros ortográricos*** mais frequentes em um dado texto. Serão fornecidos um dicionário e alguns textos. Para cada texto, a aplicação deverá listar os *k* erros mais frequentes encontrados no texto.

Dados de Entrada:

- arquivo texto referente ao dicionário;
- arquivo texto a ser analisado;
- constante K (zero indica que todos os erros devem ser exibidos).

Dados de Saída:

- tempo gasto para carregar o dicionário para a estrutura de dados;
- listagem dos *k* erros mais frequentes, por ordem decrescente de frequência e, dentro desta, por ordem lexicográfica. A listagem deverá ter o seguinte formato (exemplo com k=4)

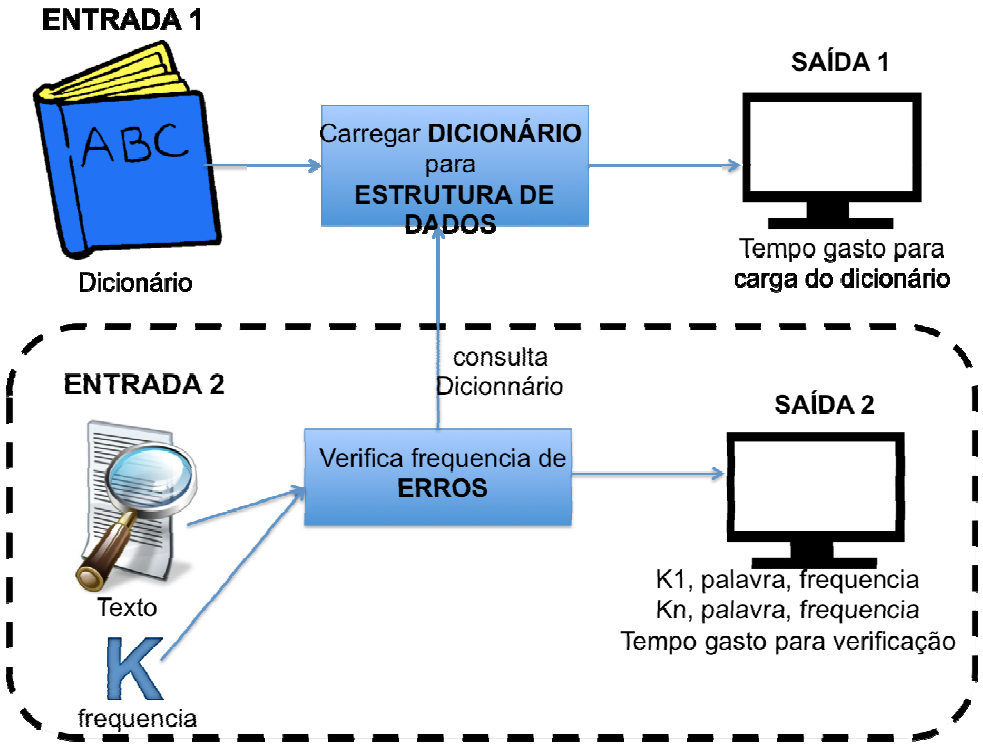
Número (k)	Erro	Frequencia
1	futibol	10
2	extadio	7
3	internacinal	7
4	gremiu	5

- tempo gasto para buscar os erros de um texto.

Restrições para a aplicação:

1. não há limite para a quantidade de palavras no dicionário;
2. não há limite para a quantidade de palavras nos textos;
3. o dicionário deve ser carregado no inicio da aplicação. Em seguida, diversos textos podem ser verificados para um mesmo dicionário carregado.

A Figura a seguir ilustra resumidamente o fluxo de atividades da aplicação. A execução da aplicação pode ser dividida em duas etapas. Na primeira etapa, a aplicação recebe como entrada um dicionário e o carrega para a estrutura de dados definida. O tempo gasto para a carga do dicionário na base de dados deve ser exibido na tela. Na segunda etapa, a aplicação recebe como entrada um texto e um valor *k* e processa a quantidade de erros do texto, classificados por frequência. Note que a etapa dois pode ser realizada quantas vezes o usuário desejar. Isto significa que para um mesmo dicionário, o usuário pode submeter inúmeros textos para verificação. Sendo que cada texto é processado por vez.



Principais definições que serão utilizadas no trabalho:

- palavra – é uma sequência de letras. Todos os outros caracteres deverão ser considerados como espaços separadores de palavras. Deve-se desprezar diferenças entre letras maiúsculas e minúsculas;
- erro ortográfico – palavra que não esteja contemplada no dicionário de palavras fornecido;
- ordem lexicográfica – é análoga à ordem das palavras em um dicionário. Ela se baseia na ordenação dos caracteres estabelecida na tabela ASCII, da mesma forma que a ordem das palavras em um dicionário se baseia na ordenação das letras no alfabeto. Para comparar duas strings `s` e `t`, procura-se a primeira posição, digamos `k`, em que as duas strings diferem. Se `s[k]` vem antes `de t[k]` na tabela ISO então `s` é lexicograficamente menor que `t`.

3 Requisitos e Avaliação

O trabalho deve ser feito em **duplas.** A linguagem de programação aceita é **C** (Não é **C++** nem **C#**).

IMPORTANTE: O aluno pode utilizar o tipo de estrutura de dados que julgar mais apropriada ou uma combinação de várias estruturas. Porém, terá que justificar sua escolha, sendo que esta justificativa também será considerada para a nota. Cabe ressaltar, que a estrutura escolhida poderá não ser eficiente para todas as operações (carga do dicionário e verificação de erros). A justificativa deve conter as vantagens e limitações da estrutura escolhida, constituindo esta justificativa um importante item de avaliação. A escolha da estrutura deve demonstrar conhecimento teórico e prático buscando a melhor combinação que atinja os resultados satisfatoriamente. Esse trabalho não avalia apenas o desempenho, mas a capacidade do aluno de criar estruturas elegantes e fáceis de serem mantidas. Para avaliar esse critério, é muito importante que o aluno DESCREVA COM RIQUEZA DE DETALHES a estrutura utilizada no programa.

A avaliação será baseada em uma apresentação oral, e no material entregue.

- Cada dupla terá entre 5 e 10 minutos para apresentação.
- No dia da apresentação, cada grupo deverá submeter (via *Moodle*) uma descrição do trabalho realizado, justificando as escolhas feitas para cada uma das estruturas de dados, e um arquivo compactado com os programas fonte (com comentários)

Itens para avaliação

- Organização e documentação do código.
- Justificativa da escolha das estruturas de dados.
- Capacidade de defender o código-fonte durante a apresentação e explicar com detalhes o funcionamento do software.
- Implementação de todas as funções necessárias.

4 Observações

- Será disponibilizado um arquivo com os dados permitindo o teste da aplicação. Este arquivo demonstra como a entrada deve ser realizada, e como o programa será testado, mas não tem como objetivo testar com cobertura a aplicação realizada. Este teste está a cargo dos alunos.
- No dia da apresentação, um novo arquivo de dicionário e novos textos (com os mesmos formatos dos arquivos de teste) deverão ser processados pelo sistema.
- Este trabalho deve refletir a solução **da dupla** para o problema proposto. **Casos de plágio serão tratados com severidade e resultarão na reprovação dos alunos envolvidos.** Para detectar o plágio, usaremos o software MOSS (http://theory.stanford.edu/~aiken/moss/).
- Dúvidas sobre o trabalho devem ser postadas no **Fórum de Dúvidas - Trabalho Final** para que, ao serem respondidas, sejam do conhecimento de todos os alunos.

5 Datas Importantes

Turmas A e B

16/11: Entrega e apresentação da Parte 1 do trabalho (descrita neste enunciado)

Turma C

17/11: Entrega e apresentação da Parte 1 do trabalho (descrita neste enunciado)

*****Somente as apresentações nestas datas serão avaliadas*****