

Laboratório de Introdução à Ciência da Computação I (SSC-102)

Trabalho 2

Professor: Dr. Jó Ueyama (*joueyama@icmc.usp.br*)
Monitor PAE: Fernando Alva Manchego (*falva@icmc.usp.br*)

Instituto de Ciências Matemáticas e Computação
Universidade de São Paulo, São Carlos

27 de maio de 2011

1 Objetivo

Praticar as técnicas de programação aprendidas em aula através da implementação de uma aplicação simples de correção ortográfica de textos.

2 Especificações

Nesta seção serão apresentadas as especificações funcionais do programa que será desenvolvido.

2.1 Descrição Geral

Um corretor ortográfico (como o do *Microsoft Word*), para poder verificar se um texto está bem escrito ou não, deve procurar por diferentes tipos de erros dentro das sentenças:

- Erros Lexicais: cada palavra é considerada de maneira isolada e é verificado se pertence ou não a um vocabulário de referência.
- Erros Gramaticais: avaliam-se as relações entre as palavras das sentenças, como: conjugações, correlação de número, gênero, etc.

No trabalho 1 foram implementadas algumas das funcionalidades do corretor ortográfico, principalmente aquelas referidas ao corretor léxico. Neste segundo trabalho, algumas dessas funcionalidades serão modificadas e algumas novas serão acrescentadas, especificamente, aquelas referidas com o corretor gramatical. A seguir, estas modificações e novas funcionalidades serão apropriadamente descritas.

2.2 Corretor Gramatical

O corretor gramatical que será implementado somente procurará por **erros nas correlações de gênero e número entre artigos e substantivos**. O programa, por exemplo, deverá identificar os seguintes erros:

Uns carro foi comprado na loja. (*Erro de número*)
O mãe do João foi no supermercado. (*Erro de gênero*)

A informação necessária para realizar este tipo de avaliação será ingressada ao programa através de dois **arquivos de texto**.

2.2.1 Arquivo “Artigos”

Este arquivo de texto terá a informação correspondente a os artigos. Para cada artigo será indicado:

- **nome:** o, os, a, as, um, uma, uns, umas, do, das, etc.
- **gênero:** M (*masculino*), F (*feminino*).
- **número:** S (*singular*), P (*plural*).

Por exemplo:

o M S
os M P
a F S
as F P
do M S
das F P

2.2.2 Arquivo “Palavras”

Este arquivo de texto terá a informação das outras palavras que poderiam aparecer no texto (similar ao léxico empregado no trabalho 1). Para cada palavra será indicado:

- **nome:** caneca, garrafas, computadores, livro, folha, sapato, etc.
- **gênero:** M (*masculino*), F (*feminino*), – (*não corresponde*).
- **número:** S (*singular*), P (*plural*), – (*não corresponde*).

Por exemplo:

caneca F S
computadores M P
em – –
sapato M S
garrafas F P
para – –

Com a informação fornecida pelos dois arquivos, o programa deverá realizar tanto a correção lexical como a gramatical.

2.3 Interface do Programa

O programa deverá apresentar uma interface que permita realizar as seguintes funcionalidades:

- Carregar os arquivos “artigos” e/ou “palavras”: O usuário deverá indicar o endereço no sistema de arquivos no qual estará cada arquivo.
- Avaliar um texto na tela ou em um arquivo e reportar os resultados na tela ou em e um arquivo: O usuário terá a possibilidade de escolher se o texto será ingressado diretamente por teclado (na tela) ou se prefere indicar o endereço no sistema de arquivos de um documento que precisa ser verificado. No primeiro caso, novamente será empregada a sinal ‘!’ para indicar o final do ingresso de dados. No segundo caso, o texto deve ser lido até o final do arquivo. O resultado da avaliação deve ser reportado na tela ou em um arquivo de texto (escolhido pelo usuário), seguindo as especificações da subseção 2.4.
- Terminar o programa: O usuário deverá ter a possibilidade de realizar as duas ações anteriores quantas vezes quiser sem que termine o programa, assim que precisa-se de uma opção para dar fim ao corretor ortográfico.

2.4 Saída do Programa

A saída do programa deve estar dividida em três seções: estatísticas gerais, correção lexical e correção gramatical.

2.4.1 Estatísticas Gerais

Aqui deve ser apresentada a seguinte informação:

- Quantidade de sentenças do texto: uma sentença se diferencia da outra por um ponto.
- Quantidade total de palavras no texto: uma palavra está separada de outra por um ou mais espaços (‘ ’), caracteres *tab* (\t) ou caracteres de pulo de linha (\n). No caso existam números no texto, eles não devem ser considerados como palavras.
- Quantidade de números: só serão considerados números inteiros positivos.
- Quantidade de palavras erradas por formato: consideram-se as palavras que contem números e letras (por exemplo, “er12as”, “12cas”).
- Lista de palavras erradas por formato.

2.4.2 Correção Lexical

Aqui deve ser apresentada a seguinte informação:

- Quantidade de palavras erradas por léxico: uma palavra é considerada errada se ela não pertence ao léxico.
- Lista de palavras erradas por léxico.

2.4.3 Correção Gramatical

Aqui deve ser apresentada a seguinte informação:

- Quantidade de pares de palavras erradas por gramática: considera-se um par como o conjunto *artigo + palavra*.
- Lista de palavras erradas por gramática, indicado se o erro foi por gênero ou número.

Um exemplo da saída do programa seria:

Estatísticas Gerais

Total de sentenças: 3
Total de palavras: 11
Total de números: 2
Total de palavras erradas por formato: 1
Palavras erradas por formato:
pai23

Correção Lexical

Total de palavras erradas por léxico: 3
sinol
computedor
caderne

Correção Gramatical

Total de pares de palavras erradas por gramática: 3
Uns carro (erro de número)
O mãe (erro de gênero)
Un computadoras (erro de gênero)/(erro de número)

No último caso do exemplo, que possui os dois tipos de erro, **somante é necessário indicar um deles**.

NOTA: *A memória necessária para armazenar as diferentes estruturas de dados que serão utilizadas no programa, assim como o texto, deve ser alocada dinamicamente somente no momento de sua utilização. O desperdício de espaço de memória deve ser evitado. A memória que não for mais utilizada deverá ser desalocada.*

3 Indicações Gerais

- O trabalho deve ser desenvolvido em grupos de **máximo três alunos**.
- Cada grupo deve projetar e desenvolver um *Corretor Ortográfico* seguindo **todas as especificações** dadas.
- Deve empregar-se a **linguagem de programação C**.
- Data de apresentação: **17/06**.

4 Critérios de Avaliação

- Cumprimento de todas as especificações funcionais indicadas.
- Ordem e clareza do código fonte: modularização, comentários e indentação apropriados.
- Apresentação do relatório: Deve ser entregue no mesmo dia da apresentação do trabalho. Não serão aceitos relatórios entregados fora do horário de aula.

5 Estrutura do Relatório

O relatório a ser apresentado deve cumprir com as especificações detalhadas nesta seção. Contudo, se o grupo considera necessário, pode acrescentar seções ou subseções para melhorar a clareza do documento.

- **Capa:** código e nome da disciplina, turma, nomes completos e números USP dos integrantes do grupo.
- **Conteúdo:**
 - Resumo
 - Introdução: apresentação geral do trabalho e organização do documento.
 - Especificação da Implementação: estruturas de dados utilizadas, pseudocódigo da função *main* mostrando a sua interação com as outras funções e pseudocódigo das funções principais do programa. Toda escolha deve ser apropriadamente justificada.
 - Dificuldades Apresentadas: relatar quais foram as principais dificuldades encontradas durante a implementação do programa e as decisões tomadas pelo grupo para superá-las.
 - Conclusões
 - Referências Bibliográficas

6 Considerações Finais

O **plágio** de programas não será tolerado. Quaisquer programas similares terão nota zero independente de qual for o original e qual for a cópia.