

## MAC 323 – Algoritmos e Estruturas de Dados II

Primeiro semestre de 2020

### Grafos de textos – Entrega: 19 de maio

No EP1 vocês implementaram uma tabela de símbolos e construíram um dicionário de palavras que ocorrem em um texto. Neste EP nós vamos fazer alguns experimentos sobre dicionários de palavras. Para isso, vamos criar um grafo que liga as palavras do dicionário, que deverá ser a primeira tarefa deste EP.

#### Construção do Grafo

Considere um grafo cujos vértices são as palavras com pelo menos  $k$  letras no seu texto. As arestas do grafo vão ligar palavras parecidas, ou seja, palavras que satisfazem uma das operações abaixo:

- remoção (inserção) de uma letra. Duas palavras são vizinhas se uma é igual à outra com uma de suas letras removidas. Assim, por exemplo, **alta** e **ata** são vizinhas. Da mesma forma, **casa** e **casca**.
- troca de letras de uma mesma palavra. Duas palavras são vizinhas se uma pode ser obtida a partir da troca de posição de duas letras da outra. Por exemplo, **nato** e **nota** são vizinhas, assim como **nota** e **tona**.
- substituição de uma letra. Duas palavras são vizinhas se uma é obtida a partir da outra substituindo uma das letras. Por exemplo, **gato** e **rato** são vizinhas, assim como **prata** e **preta**.

Com estas operações construímos um grafo para cada texto que for lido. Os experimentos serão feitos sobre este grafo.

## Experimentos sobre grafos de textos

Vamos analisar algumas propriedades do grafo construído acima. São elas, entre outras que vocês podem propor:

- **medidas:** quantos vértices e arestas o grafo tem? Qual o grau médio dos vértices? O grafo é denso (o número de arestas é da ordem do número de vértices ao quadrado)?
- **conexidade:** o grafo construído é conexo? Quantas componentes ele tem? Qual o tamanho médio das componentes? E da maior (menor)?
- **caminhos e ciclos:** dadas duas palavras  $a$  e  $b$  do dicionário, existe um caminho que liga  $a$  a  $b$ ? As palavras aparecem em algum ciclo do grafo? Qual a distância média (número de palavras em um menor caminho do grafo entre duas palavras) entre duas palavras no grafo?

Como vocês podem ver, diversas perguntas interessantes podem ser feitas sobre estes grafos. Será que textos de tipos diferentes têm comportamentos semelhantes? Por exemplo, se o texto for um manual técnico de um equipamento ou um romance de Camões, os grafos são parecidos?

Para os interessados em visualizar os grafos obtidos, existem ferramentas que desenhavam grafos, como o Pajek, por exemplo.

Mais uma vez os monitores vão disponibilizar uma interface para fazer as consultas ao grafo. Além do código, esperamos que você envie um relatório com seus testes, como nos EPs anteriores.