

# **PCS 3216 – Sistemas de Programação**

## **Primeira Prova (2021)**

**Complemento ao enunciado**

**QUESTÃO 5**

# **COMPLEMENTO À QUESTÃO 5 DA PROVA**

# Preparação de textos e glossários

- **Escolha alguns arquivos .txt** contendo textos jornalísticos suficientemente longos, representativos de áreas de seu interesse (no mínimo, 3 áreas diferentes, e 3 textos por área), e outros, fora de área.
- Use o programa que você desenvolveu para coletar listas, alfabeticamente ordenadas, de todas as palavras neles encontradas. Ordene também essas listas segundo o **número de ocorrências das palavras** nos textos.
- Dentre as palavras mais encontradas nos diversos textos, selecione aquelas que, embora frequentes, são irrelevantes às áreas de interesse (como é o caso de pronomes, artigos, preposições, conjunções, etc.), compondo assim um **conjunto de palavras a descartar** nesta atividade.
- Use novamente o programa que você construiu para coletar as listas de **palavras**, agora **filtradas** de acordo com o conjunto de descartes construído no item anterior.
- A partir dessas listas, construa **glossários específicos, um para cada tema** de seu interesse.

# Atividades a executar

(Para efetuar estas tarefas, não utilize os mesmos textos a partir dos quais você preparou as tabelas de palavras do slide anterior.)

1. Para cada texto, execute o programa que você construiu, instruído pelo glossário do assunto desejado e pela lista de palavras a serem excluídas.
2. Marque nessa lista todas as palavras que tenham sido encontradas no glossário, e determine a percentagem de suas ocorrências no texto.
3. Proponha um critério de classificação de textos segundo os vários temas experimentados, usando para isso esses percentuais de ocorrência, e teste esse critério para classificar alguns textos, da área e fora dela.
4. Utilize textos adicionais e aplique o critério que você criou para estimar sua pertinência ou não a alguma das áreas de interesse.
5. Usando essa mesma técnica, amplie a capacidade de classificação do seu programa, agregando-lhe glossários associados a novas áreas.
6. Amplie o seu classificador para que contemple textos multidisciplinares, categorizando-os simultaneamente em mais de uma área, empregando para isso o seu critério de classificação, e aplicando-o aos textos disponíveis, e ao conjunto dos glossários montados anteriormente.