

DESAFIO 01 – TEORIA DA INFORMAÇÃO

Desenvolva um programa para determinar o **modelo sem memória de ordem-1** para uma determinada linguagem, cuja amostra é fornecida como o arquivo “Desafio.txt”. Uma vez obtido o modelo da linguagem, desenvolva outro programa, para “gerar texto” segundo o modelo obtido.

Seus programas devem operar em modo **terminal**, ou seja, com uma interface de linha de comando, não-gráfica, e deve ser desenvolvido utilizando a linguagem de programação **C (ANSI) ou Free Pascal**, compiláveis com “gcc” ou “fpc”, conforme o caso.

Seu programa deve atender às seguintes funcionalidades:

- seleção do alfabeto pelo usuário: Z_{26} , Z_{27} , ou Z_{256} (no caso de Z_{26} ou Z_{27} , caracteres minúsculos, na amostra, devem ser convertidos para caracteres maiúsculos, antes de serem processados);
- qualquer caracter (símbolo) não pertencente ao alfabeto selecionado e presente no arquivo amostra deve ser ignorado para a construção do modelo;
- a especificação do “arquivo amostra” deve ser feita via entrada pelo teclado, dando liberdade de se escolher qualquer arquivo;
- o arquivo para armazenamento do modelo também deve ser escolhido via teclado. Deve ser visualizável, de alguma forma, o modelo obtido, onde sejam apresentados os percentuais de ocorrência de cada caracter.
- além da distribuição da frequência dos símbolos para a amostra e alfabeto selecionados, o programa de construção do modelo deve também apresentar o valor da entropia da linguagem segundo o modelo desenvolvido.
- O programa de geração de texto deve permitir, via teclado, a escolha do modelo a ser utilizado, o número de caracteres a serem gerados, e um arquivo de destino.

Uma vez desenvolvidos os programas, faça os seguintes experimentos:

- Determine o modelo R1 (ordem-1) para a amostra fornecida, considerando o alfabeto como Z_{26} ;
- Repita (1), agora considerando o alfabeto como Z_{27} ;
- Repita (1), agora considerando o alfabeto como Z_{256} ;
- Compare os três resultados anteriores e justifique eventuais diferenças;
- Sabendo-se que a entropia de uma linguagem pode ser interpretada como a quantidade média de bits necessária para representar (em binário) cada símbolo do alfabeto da linguagem, proponha uma codificação binária eficiente para cada símbolo do alfabeto da linguagem em questão (observe que a amostra fornecida utiliza uma codificação de 8 bits por símbolo - ASCII). Utilize para tal o modelo em Z_{27} .

Observações:

- ✓ o trabalho poderá ser feito em duplas ou individualmente;
- ✓ o código do programa deve ser do próprio aluno(a)/alunos(as), contendo um cabeçalho no arquivo fonte que identifique o programa, o aluno(a) e seu propósito;
- ✓ é livre a consulta a qualquer referência bibliográfica ou da Internet, as quais, se utilizadas, devem ser mencionadas;
- ✓ o código deve ser suficientemente documentado para permitir sua fácil compreensão e uso por terceiros.
- ✓ O programa deve ser acompanhado de um pequeno texto descritivo dos procedimentos computacionais adotados para a solução do problema.
- ✓ Este trabalho contará como um “Trabalho Avulso” para a avaliação T1, do primeiro

bimestre, como consta Plano de Ensino da disciplina.

O trabalho deve ser encaminhado por e-mail ao professor até o dia 20/03/2017, contendo o código fonte e eventuais documentos complementares, tais como resultados de testes efetuados e um fluxograma ou algoritmo para a solução do problema. Sob nenhuma circunstância serão aceitos trabalhos entregues após esta data.

Nota:

Para o cômputo da nota será considerado o seguinte:

- | | | |
|------|------------------------------------------------|------------|
| i. | Relatório de desenvolvimento e resultados | 1,0 pontos |
| ii. | Criatividade | 2,0 pontos |
| iii. | Documentação do projeto | 3,0 pontos |
| iv. | Resultado Final (operacionalidade do programa) | 4,0 pontos |

A identificação de cópia do programa, de qualquer origem, acarretará a atribuição da nota Zero no trabalho como um todo. Programas idênticos receberão também a nota Zero.