

**MAC 122 - Princípios de Desenvolvimento de Algoritmos****Segundo semestre de 2014****Artista Aleatório – Entrega: 10 de novembro**

O objetivo deste exercício-programa é fazer um gerador de textos aleatórios “semelhantes” a uma obra literária.

Você deverá implementar vários níveis para seu programa. No nível 0, o mais simples, seu programa lerá um texto e calculará o número de ocorrências de cada caractere. Por exemplo, para a “Ilíada” (em inglês) de Homero (<http://www.gutenberg.org/cache/epub/6130/pg6130.txt>) que tem 1.201.894 caracteres os cinco caracteres mais frequentes são em ordem: “ ” (branco 226.304 ocorrências), e (108.789), t (69.284), s (63.348) e o (61.113). Já “Os Lusíadas” (em português do século XVI) tem 360.008 caracteres, e os 5 mais frequentes são “ ” (49.803), e (31.839), a (31.818), o (29.293) e r (17.608). Com esta contagem você consegue gerar um texto “que se parece” com sua obra predileta, gerando cada caractere com probabilidade proporcional à frequência com que aquele caractere ocorre no texto.

No nível 1 faremos algo um pouco mais sofisticado, determinando, para cada caractere, a probabilidade de que seja seguido por um outro caractere. Assim, em português, a probabilidade de que o caractere “p” seja seguido por um “a” é bem maior do que por um “q”. De forma geral, no nível  $k$  o próximo caractere será gerado com base nos últimos  $k$  caracteres gerados, de acordo com as probabilidades de seu texto predileto. Ou seja, você analisará todas as ocorrências deste trecho de  $k$  caracteres no texto (nessa ordem) e verificará quais caracteres ocorrem após este trecho e com qual probabilidade. Então, seu programa sorteará um destes caracteres. Você vai observar que com um nível moderado (5 a 7) o seu texto já vai ter algumas características de seu texto base.

Dessa forma seu programa deverá ler alguns parâmetros:

- um inteiro não negativo  $k$ ;
- um inteiro não negativo  $t$  (o tamanho do texto a ser produzido) ;
- um arquivo com o texto base com pelo menos  $k$  caracteres (“Os Lusíadas”, “Ilíada”, seu programa, etc);

Com base nestes parâmetros seu programa deverá:

- gerar um trecho de  $k$  caracteres aleatoriamente com base nas probabilidades do texto base;
- a partir daí, o próximo caractere deve ser escolhido com base na frequência com que cada letra segue os últimos  $k$  caracteres no seu texto base nesta ordem, até que  $t$  caracteres são gerados;

No site do Projeto Gutenberg (<http://www.gutenberg.org/>) há vários livros disponíveis (formato UTF-8) que podem servir de base para sua “produção literária aleatória”. Junto com o fonte de seu EP você deve entregar mais uma vez um relatório com os principais testes que você fez, textos obtidos, etc. Neste exercício será analisado este relatório e a qualidade de sua implementação (corretude, eficiência, elegância, etc).

**Bônus:** Você poderia pensar em geradores aleatórios de pinturas ou de músicas com base nas obras de seus pintores ou músicos prediletos. A ideia seria a mesma: o próximo pixel (nota) seria sorteado entre as ocorrências em sua(s) obra(s) favorita(s) de acordo com a probabilidade de ocorrência. Neste caso você deverá estudar o formato de armazenamento de imagens (músicas).

Este EP está baseado no experimento descrito na coluna “Computer Recreation” de A.K. Dewdney da revista *Scientific American*’ de junho de 1989: “A potpourri of programmed prose and prosody”.