

AED 2010/2011 – TP3

Descrição sumária dos ficheiros de input fornecidos

Input1: Hamlet, de Shakspeare

Texto real, com 32.858 palavras de um vocabulário de 4.583 palavras.

Input2: Hamlet, Ordenado

Idêntico ao input1, mas com as palavras ordenadas.

Input3

Texto sintético, com 100.000 palavras de um vocabulário de 10.000 palavras.
Distribuição aleatória e uniforme.

Input4

Texto sintético, com 200.000 palavras de um vocabulário de 10.000 palavras.
Este texto é uma duplicação directa do Input3.

Input5

Texto sintético, com 100.000 palavras
(vocabulário efectivo de aproximadamente 6.500 palavras).
As primeiras 10.000 palavras têm distribuição aleatória e uniforme.
As 90.000 palavras finais estão concentradas num vocabulário de 1.000 palavras
(15% do vocabulário, distribuição aleatória).

Input6

Texto sintético, com 100.000 palavras
(vocabulário de aproximadamente 6.500 palavras).
As primeiras 10.000 palavras têm distribuição aleatória e uniforme.
As 90.000 palavras finais estão concentradas nas 1.000 últimas palavras
(15% do vocabulário, as últimas palavras a ser inseridas).