1 Sistema de pesquisa numa base de textos

Construa um sistema de pesquisa para uma base de textos recorrendo às estruturas de dados estudadas nas aulas para representar dicionários.

1.1 Formato e conteúdo dos textos da base de textos

A sua base de textos é constituída por um conjunto de ficheiros com o texto de uma fábula ou de um conto infantil.

Além do texto da história o ficheiro tem o título da história, as personagens principais (nomes separados por vírgulas) e pode ter ou não uma conclusão.

Exemplo do ficheiro f1.txt:

título: A cegonha e o lobo

personagens: a cegonha, o lobo.

conclusões: É bom não contar com as promessas de quem está numa aflição pois elas podem não ser cumpridas.

texto: Um dia, estava um lobo a comer, quando se lhe atravessou um osso na garganta. Quase a sufocar e muito engasgado, foi pela floresta perguntando a todos os animais se o podiam ajudar. - Dou uma boa recompensa a quem me salvar - anunciava ele. Uma cegonha que pairava lá nos ares e o ouviu, veio de lá de cima e ofereceu-lhe a sua ajuda. - Abra a boca, senhor lobo - disse-lhe a cegonha, e meteu-lhe o comprido bico pela garganta abaixo. - Cá está o osso que tanto o incomodava - disse a cegonha assim que retirou a cabeça da boca do lobo. - Acho agora que tenho direito à minha recompensa, não é verdade? - Recompensa?, disse-lhe o lobo mostrando os seus grandes dentes e já a esquecer a aflição em que se encontrava quando tinha o osso atravessado. - Estás com muita sorte em não te ter cortado a cabeça quando a tinha dentro da minha boca. Queres recompensa maior, pássaro ingrato? E dizendo isto, voltou as costas à cegonha que não teve outro remédio que regressar aos céus.

Exemplo do ficheiro f2.txt

título: A camponesa e a bilha de leite

personagens: a Camponesa, a bilha de leite, o leite, a galinha.

conclusões: Não adianta chorar sobre leite derramado.

texto: Uma bela camponesa ia a caminho da feira levando à cabeça uma bilha cheia de leite acabado de ordenhar. E enquanto atravessava as ruas, começou a sonhar.

Depois de vender o leite, fico com dinheiro suficiente para comprar muitos ovos - pensava ela. - Ponho os ovos debaixo duma galinha choca e hei-de ter pelo menos uma dúzia de belos pintos que posso vender no mercado. Quando nascerem, será numa boa altura, precisamente quando a criação estiver mais cara, e assim poderei comprar um vestido novo.

De que cor será o meu vestido? - pensava ela. - Gosto do azul. Fica-me muito bem. Com uma saia muito rodada. Oh, hei-de ficar tão bonita que todos os rapazes da aldeia vão querer casar comigo.

Mas eu não lhes vou ligar nenhuma - disse ela de si para si mesma. - Viro-lhes a cara e sigo em frente como se os não visse.

E, juntando o gesto ao pensamento, virou bruscamente a cabeça.

Infelizmente, esqueceu-se de que levava a bilha à cabeça. A bilha desequilibrou-se e caiu com estrondo no chão derramando todo o leite ao quebrar-se.

Ai o meu rico vestido azul! - chorava ela. - Ai os meus pintainhos! Ai os meus ovos! Ai o meu leite!? Mas todos os seus ais foram inúteis. O leite já se sumira no chão. E sem ter nada para vender na feira, a rapariga voltou para casa com uma enorme tristeza.

1.2 Operações do sistema de Pesquisa

O seu sistema de pesquisa em bases de textos deve permitir as seguintes operações:

1. Listar todas as palavras que ocorrem nos textos e a sua frequência.
2. Dado o nome de uma história listar todas as palavras que ocorrem na história e a sua frequência.
3. Dado o nome de uma personagem, listar os títulos e o nome dos ficheiros de todas as histórias que tem a personagem.
4. Dada uma palavra listar todas conclusões das histórias que têm a palavra na conclusão.
5. Dada uma palavra listar os títulos e o nome dos ficheiros de todas as histórias que tem a palavra no texto.
6. Dado o nome dum ficheiro acrescentar a informação do ficheiro ao seu sistema de pesquisa. O ficheiro deve ter o formato descrito acima.

1.3 Entrega do trabalho

Junto com o programa deve entregar um relatório onde indica:

1. A estruturas de dados utilizadas pelo seu sistema de pesquisa.
2. A complexidade de cada operação implementada justificando a escolha das estruturas de dados usadas.
3. O espaço ocupado (em memória e no disco) pelo seu sistema, função do número de palavras e do número de ficheiros.
4. Instruções sobre o funcionamento do programa.

Regras:

- Os grupos deverão ter entre 1 e 3 elementos (grupos com mais de 3 elementos terão de ir obrigatoriamente à discussão do trabalho).
- A data para a entrega do trabalho é 07-06-2003 (Sábado) às 23.59h. Por cada dia de atraso serão descontados $2^{\text{dias}-1}$ valores na nota do trabalho.
- O trabalho deve ser entregue num ficheiro chamado t1.tar.gz, que descompacta para uma directoria chamada 'trabalho1'. Essa directoria deve conter uma subdirectoria 'doc', onde estará o relatório. O código e o Makefile (obrigatório, com *targets* 'build', 'clean' e 'run', pelo menos) devem ficar na directoria principal ('trabalho1').

Notas:

- Poderá haver lugar a discussão de alguns trabalhos, caso os docentes julguem necessário, para uma melhor classificação do trabalho. Os alunos em causa serão avisados antecipadamente da data, hora e local da discussão.

- Os trabalhos que não cumpram todos os requisitos terão penalizações, em valores, relativamente aos requisitos não cumpridos.
- A interpretação do enunciado, bem como a escolha das estruturas de dados a utilizar, fazem parte da avaliação.