

FILTRO DE CORREIO ELETRÔNICO

Y. KOHAYAKAWA

Data de entrega. Você deve entregar este EP até o dia **27/2/2004**, pelo Panda.

1. INTRODUÇÃO

Neste exercício-programa de recuperação, você deve construir um *filtro de correio eletrônico*.

2. O PROBLEMA A SER RESOLVIDO

O programa que você escreverá neste exercício deve resolver o problema que descrevemos a seguir.

O usuário terá um arquivo de *padrões*, onde um *padrão* é uma expressão comum em *spam*, mensagens eletrônicas indesejadas, com conteúdo de interesse dúbio, como por exemplo “Ganhe muito dinheiro!” ou “Visite nosso site!”. O usuário também tem um arquivo com suas mensagens eletrônicas, e ele deseja produzir dois arquivos a partir dele: um contendo as mensagens genuínas, e o outro contendo as mensagens que contêm um ou mais dos padrões especificados. Naturalmente, a idéia é que este processo tenta separar as mensagens interessantes dos *spam*.

Observações. Vamos definir um *padrão* como sendo um *string* do C que não contém o caracter `'\n'`. O arquivo de padrões, digamos **padroes**, contém um padrão por linha. Você pode supor que há no máximo 1000 padrões neste arquivo, e que cada padrão tem no máximo 256 caracteres.

As mensagens eletrônicas do usuário devem vir em um arquivo, digamos **correio**, com as mensagens separadas por uma linha em branco. Você deve supor as mensagens começam com o texto **"From "** (note o espaço em branco depois de **"From"**).

Execução. O usuário executará o seu programa com o seguinte comando:

```
meu_prompt $ eprec padroes arq_spam < correio > correio_limpo
```

No caso dos seguintes arquivos:

- **padroes**
- **correio**,

o comando acima deverá gerar as seguintes saídas:

- **arq_spam**
- **correio_limpo**.

3. EFICIÊNCIA DE SEU PROGRAMA

Procure implementar uma estrutura de dados que torne seu programa eficiente, tanto em termos de tempo como em termos de espaço. Você deve imaginar que o arquivo `correio` conterá um grande número de mensagens (digamos, da ordem de 2000 ou 3000 mensagens). Ademais, você deve supor que `padroes` contém, digamos, algo como 500 padrões.

A sua nota dependerá da correção e da eficiência de seu programa!

4. OBSERVAÇÕES

1. *Este EP é individual.*
2. Seja cuidadoso com sua programação (correção, documentação, apresentação, clareza do código, etc), dando especial atenção a suas estruturas de dados. A correção será feita levando isso em conta.
3. *Não imponha restrições arbitrárias sobre a entrada*, além das explicitamente dadas neste enunciado. Quando usado de forma inesperada, o seu programa deve parar de forma ‘graciosa’.
4. Organizem-se na turma para fazer bons testes. Comparem entre vocês o desempenho de seus programas.
5. Seja criativo. Se seu programa fizer algo a mais do que foi pedido você poderá ganhar algum bônus na nota. Por outro lado, você *deve* respeitar as especificações dadas neste enunciado.
6. Entregue seu EP seguindo os moldes usuais: cabeçalho claro, instruções claras para o monitor poder compilar e executar seu programa, e eventuais arquivos de teste (não muito grandes).

Observação final. Enviem eventuais dúvidas para mim ou para o monitor, por correio eletrônico. Não estamos acompanhando o fórum do Panda.