

***Licenciatura em Engenharia Informática
Departamento de Engenharia Informática
Instituto Superior de Engenharia do Porto***

Manipulação de Ficheiros de Texto

Sandra Manuela Lemos Machado

Versão 2 – Março 2002

Este documento pretende constituir um guia de trabalho para as aulas de Sistemas Operativos 1, introduzindo utilitários que permitem a manipulação de ficheiros de texto ASCII no sistema operativo Unix. Não pretende ser um manual dos comandos apresentados, por isso se quiser realmente saber como os utilitários funcionam use o manual do Unix (man comando).

1 CONTAGEM DE LINHAS NUM FICHEIRO DE TEXTO

WC (Word Count) é o utilitário Unix que permite a contagem de linhas, palavras e caracteres num ficheiro de texto.

1.1 *Sintaxe*

```
wc [ -cwl ] nome_dos_ficheiros
```

1.2 *Exercícios*

- 1) Experimente as três opções do comando `wc` do seguinte modo:
 - a) Conte apenas as palavras do ficheiro `passwd` (-w);
 - b) Conte apenas os caracteres do ficheiro `passwd` (-c);
 - c) Conte apenas as linhas do ficheiro `passwd` (-l).
- 2) Crie uma sequência de comandos que mostre apenas o número de utilizadores presentes no sistema.
- 3) Crie uma sequência de comandos que mostre o número de ficheiros e directórios da sua directoria HOME.

2 SUBSTITUIR CARACTERES NUM FICHEIRO DE TEXTO

O comando `tr` (translate) na sua forma mais simples permite substituir os caracteres especificados na *string_input* pelos especificados na *string_output*.

O comando `tr` e um filtro, lê o *standard input* (teclado) e devolve a saída para o *standard output* (monitor). O *standard input* pode ser redireccionado para um ficheiro (<) ou fornecido por um *pipe* (|).

O *standard output* também pode ser redireccionado (>), embora no Unix não seja permitido usar o mesmo ficheiro como entrada e saída de um comando, pode-se contornar esta limitação recorrendo a um ficheiro temporário.

Com a opção `-d`, os caracteres não são substituídos, mas removidos.

A opção `-s` remove os caracteres que se repetem, torna-se muitas vezes útil para remover os duplos espaços, triplos, etc., mantendo apenas um espaço.

2.1 *Sintaxe*

```
tr [-cds] [string_input] [string_output]
```

2.2 Exercícios

- 1) Substitua todas as letras minúsculas de um ficheiro de texto por maiúsculas. Verifique novamente o seu conteúdo.
- 2) Use a linha de comandos que da alínea anterior, mas altere-a de modo a que as alterações provocadas pelo comando `tr` fiquem guardadas no ficheiro.
- 3) Verifique as opções `-d` e `-s` do seguinte modo:
 - a) Apague todas as vogais do ficheiro usado acima.
 - b) Crie um ficheiro chamado `listagem` com a listagem longa da sua HOME. Retire ao ficheiro todos os espaços que não sejam simples (duplos espaços, triplos, etc.)
- 4) Crie uma linha de comandos que substitua todos os caracteres `"\"` para `"*` e os `","` para `"-` no ficheiro `/etc/group`.
- 5) Crie uma sequência de comandos que mostre quantos níveis abaixo da directoria `root` está localizado o seu directório corrente.

3 EXTRAIR COLUNAS DE UM FICHEIRO DE TEXTO

O comando `cut` permite visualizar ficheiros como matrizes com linhas e colunas e permite extrair as colunas especificadas.

O comando `cut` pode analisar o ficheiro de duas formas: os ficheiros podem ser constituídos por colunas separadas por determinado carácter (por defeito espaço, mas é configurável através da opção `-d`). Para extrair colunas a este tipo de ficheiros tabulares usa-se `-f` seguido de uma lista de colunas a extrair (por exemplo `-f1,4,3`).

A opção `-s` obriga o comando `cut` a suprimir linhas do ficheiro que não contenham o carácter delimitador.

A opção `-c` permite cortar um conjunto de caracteres especificando, para isso, a sua posição na linha, por exemplo `cut -c1`, corta apenas o primeiro carácter de cada linha.

3.1 Sintaxe

```
cut -flista [-dchar] [-s] [ficheiro]
```

```
cut -clista [ficheiro]
```

3.2 Exercícios:

- 1) Explique como extrair as colunas 21 a 43 do ficheiro `/users/1/bach/so1/fortunes.txt`;
- 2) Explique como extrair as colunas 1 a 3 do ficheiro `listagem` criado anteriormente;

- 3) Crie um ficheiro **nomes** com todos os nomes dos ficheiros presentes no ficheiro **listagem**;
- 4) Crie uma sequência de comandos que liste todos os ficheiros do directório **/usr/bin** juntamente com o seu ultimo tempo de acesso. Guarde essa lista num ficheiro com o mesmo nome na sua **HOME**, num formato de 3 colunas (nome do ficheiro, dia e mês).
- 5) Construa uma lista com todos os utilizadores do sistema, a sua directoria e a sua **shell**, separe cada um dos itens com um espaço e mostre o resultado num ecrã de cada vez.
- 6) Conte o número de linhas de um ficheiro de texto que comecem por maiúsculas.

4 COLAR FICHEIROS

Muitas vezes existe a necessidade de cortar várias colunas a um ficheiro para depois as voltar a colar noutro com numa ordem diferente. O utilitário Unix **paste** lê um ou mais ficheiros e cola as colunas correspondentes pela ordem pretendida.

Por defeito as colunas são separadas por espaço, mas podem ser separadas por qualquer caracter especificado com a opção **-d**, esta opção permite a utilização de uma lista de caracteres, sendo que eles são usados ciclicamente (1º, 2º, 3º, 1º, 2º, ...).

A opção **-s** permite juntar informação de vários ficheiros numa linha e não em colunas.

O comando **paste** não é um filtro, é necessário indicar na linha de comandos os ficheiros a usar, porque o utilitário não lê o *standard input*. Se em alguma situação houver a necessidade de usar o **paste** como filtro, deve escrever-se um hífen (-) na linha de comandos em vez do nome dos ficheiros (`who | cut -f1 -d" " | paste -s -`).

4.1 Sintaxe

```
paste [-s] [-dlista] [ficheiros]
```

4.2 Exercícios

- 1) Crie um ficheiro, chamado **nomes**, com apenas o nome dos ficheiros retirado do ficheiro **listagem**. Crie um ficheiro, chamado **permissões**, com as permissões associadas a cada ficheiro listado. Crie um ficheiro chamado **nomes+perm** que contenha apenas duas colunas, a primeira com os nomes dos ficheiros e a segunda com as permissões;

5 JUNTAR FICHEIROS DE TEXTO COM CAMPOS EM COMUM

O comando Unix **join** permite juntar dois ficheiros de texto com pelo menos um campo em comum. A coluna comum chama-se referência. Para juntar os dois ficheiros com o utilitário **join** a referência tem de estar ordenada por ordem alfabética. Este utilitário não é um filtro, mas pode ser usado numa *pipeline*, substituindo o primeiro ficheiro por um hífen, como no comando **paste**.

Quando usado sem opções o **join** junta os ficheiros pela primeira coluna, sempre que o campo referência não fôr o primeiro, deve usar-se a opção **-j**. Para exemplificar veja-se a seguinte linha de comandos:

```
join -j1 4 -j2 3 exemplo1 exemplo2.
```

Neste caso junta os ficheiros exemplo1 e exemplo2 pelo campo 4 do exemplo1 e pelo campo 3 do exemplo2.

Por defeito os dois ficheiros são juntos pelo campo referência, se alguma linha do ficheiro não contiver o respectivo campo, não é exibido no resultado. Sempre que isto consista numa limitação para o utilizador, pode usar a opção **-a**, seguida de **1** (-a1) para o caso de querer na saída os campos que não juntou no primeiro ficheiro e analogamente para o ficheiro dois a opção **-a2**.

O caracter delimitador também pode ser configurado usando o **-t** seguido de um caracter (-t"*").

Nem sempre é necessário apresentar todos os campos dos ficheiros a juntar, a opção **-o** permite escolher os campos a juntar, mas não mostra o campo referência, a não ser que seja explicitamente especificado na linha. Por exemplo:

join -o 1.3 1.4 2.2 2.3 2.5 exemplo1 exemplo2

Neste exemplo o resultado conterà as colunas 3 e 4 do exemplo1 e as colunas 2,3,5 do exemplo2, o campo referência é o primeiro, visto que nada foi especificado.

Outra opção que pode ser útil quando se usa o comando **join** é **-e**, seguida de uma *string* que será exibida na saída sempre que existam campos em branco resultantes, por exemplo se na opção **-o** especificarmos uma coluna que não exista.

5.1 Sintaxe

join [-a1 -a2] [-o lista] [-j1 no -j2 no] [-tchar] [-e string] file1 file2

5.2 Exercícios

1) Considere os ficheiros abaixo:

turmas		horario		
Pedro	A	3 ^a	A	Pedro
Nuno	B	4 ^a	B	Nuno
Patricia	B	4 ^a	B	Patricia
Ricardo	F	5 ^a	F	Ricardo
Luis	F	5 ^a	F	Luis

- c) Escreva uma linha de comandos que junte as colunas 1,2 do ficheiro turmas com as colunas 1 e 2 do ficheiro horário, separe as colunas com o caracter ":";
- d) Os ficheiros acima podem ser juntos por outra coluna que não a primeira, qual? Que alteração seria preciso fazer nos ficheiros?