

Trabalho Final de ICC (2017/02)

Pablo Cecilio, Marco Antônio, Lucas Souza, Arthur Rocha

1 Introdução

O trabalho consiste no processamento e manuseio entre bases de dados usando a programação em bash. Sendo esse processado através do uso de um script contendo os comandos necessários para executar diversas ações específicas entre as bases de dados originais, gerando a saída requisitada pelos itens abaixo descritos.

2 Pré-processamento

Afim de concatenar os dois arquivos foi utilizado o comando “join”, uma linha de comando que concatena linhas de dois arquivos através de uma coluna em comum.

```
join -t $'\t' -o 2.1,2.2,2.3,2.4,2.5,2.6,2.7,2.8,2.9,1.2,1.3  
↪ title.ratings.tsv title.basics.tsv
```

De modo a conferir se essa saída foi gerada corretamente, foi utilizado um “wc -l” comparando “titles.ratings.tsv” e o arquivo gerado “titles.tsv”.

```
wc -l title.ratings.tsv  
wc -l titles.tsv
```

Para finalizar o pre-processamento, o comando “sed” foi utilizado afim de remover a primeira linha do arquivo “titles.tsv” para gerar um novo arquivo, “titles.all.tsv”. Novamente a saída foi verificada com a contagem das respectivas linhas através do comando “wc -l”.

```
sed '1d' titles.tsv > titles.all.tsv  
wc -l titles.all.tsv
```

3 Extrações

Descreva aqui quais itens foram feitos, quais não foram feitos e quais sua equipe não tem certeza da resposta. Explique os motivos.

Item 1

Discuta aqui cada comando utilizado para resolver o problema. O exemplo hipotético da seção de “Entrada e saída” da especificação poderia ser explicado da seguinte forma:

Para contar o número de itens do arquivo “title.basics.tsv” nós utilizamos o comando cat para imprimir o arquivo na saída padrão e utilizamos o comando wc (parâmetro -l) em cascata (pipe) para contar quantas linhas foram impressas, resultando no total de linhas do arquivo “title.basics.tsv”.

Item 2

...

Item 18

4 Organização do trabalho

Explique nessa seção como o trabalho foi distribuído para o grupo. Por exemplo, toda a equipe fez todos os itens junto ou as tarefas foram distribuídos para os membros da equipe? Crie uma tabela onde a linha é o nome do membro da equipe e a coluna são as atividades desempenhadas. Essa seção será utilizada para a avaliação e arguição. Portanto, sejam honestos.

MEMBRO	ATIVIDADES
Pablo Cecilio	pre-processamento; script.sh; item 1; documentação
Marco Antônio	—
Lucas Souza	—
Arthur Rocha	—

5 Conclusão

Discuta aqui os principais desafios enfrentados no trabalho e como cada desafio foi superado. Discuta também o aprendizado da equipe durante o trabalho.

6 Extra

