

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

A extração dos itens 1,2,3,4,5,6,11 foram realizadas.

Item 1

Utilizou-se “cut” para selecionar a coluna 2 em “title.all.tsv”, após esse comando foi realizado um pipe para um “sort” afim de ordenar a seleção, em seguida outro pipe para “uniq” e com isso listar as entradas únicas dessa seleção como saída para o arquivo.

| ENTRADA | CUT -F 2 | SORT | UNIQ |
|---------|----------|------|------|
| F A X | A | A | A |
| G B Y | B | A | B |
| H A Z | A | B | C |
| K C U | C | C | |

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 | itens 2,3,4,5,6,11 |
| 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

