

CMD

Jesús Manuel Solís Durán

October 2, 2024

La Tarea consiste en ver las diferencias entre encontrar las 10 mejores y peores películas en que se encuentran en el archivo comprimido aclImdb, En este PDF, solo estaremos estudiando la parte en linea de comandos. Para ello usaremos 3 comando awk, sed, paste.

1 CMD

Primero utilizaremos el comando awk para tomar el segundo elemento del titulo de los archivos en train y guardándolos en puntaje_negativo y puntaje_positivo.

```
jesolis14@shadow:~/Escritorio/PCD/A-02/aclImdb/train$ ls neg | awk -F'[_ .]' '{print $2}' > puntaje_negativo.txt
jesolis14@shadow:~/Escritorio/PCD/A-02/aclImdb/train$ ls pos | awk -F'[_ .]' '{print $2}' > puntaje_positivo.txt
```

A continuación buscaremos las urls correspondientes en los archivos urls_pos y urls_neg, lo cual lo guardaremos en los archivos urls_negativo y urls_positivo.

```
jesolis14@shadow:~/Escritorio/PCD/A-02/aclImdb/train$ sed 's/\\usercomments$//' urls_neg.txt > urls_negativo.txt
jesolis14@shadow:~/Escritorio/PCD/A-02/aclImdb/train$ sed 's/\\usercomments$//' urls_pos.txt > urls_positivo.txt
```

Lo que haremos enseguida es usar el comando paste para crear 2 archivos que contenga las urls y los correspondientes puntajes.

```
jesolis14@shadow:~/Escritorio/PCD/A-02/aclImdb/train$ paste urls_negativo.txt puntaje_negativo.txt > critica_negativo.txt
jesolis14@shadow:~/Escritorio/PCD/A-02/aclImdb/train$ paste urls_positivo.txt puntaje_positivo.txt > critica_positiva.txt
```

Para finalizar crearemos usando el comando awk encontraremos el promedio de los urls iguales, para así poder obtener un promedio de los puntajes dados en el directorio train. Después usamos sort para ordenarlos de mayor a menor y tomando head para los de mayor puntaje y tail para los de menor.

```
jesolis14@shadow:~/Escritorio/PCD/A-02/aclImdb/train$ awk '{critica_negativo[$1]++$2; count[$1]++} END {for (url in critica_negativo) print url, critica_negativo[url]/count[url]}' critica_negativo.txt | sort -k2 -n | head -n 10 > peores_peliculas.txt
jesolis14@shadow:~/Escritorio/PCD/A-02/aclImdb/train$ awk '{critica_positiva[$1]++$2; count[$1]++} END {for (url in critica_positiva) print url, critica_positiva[url]/count[url]}' critica_positiva.txt | sort -k2 -n | tail -n 10 > mejores_peliculas.txt
```

Por último dejo una imagen con todas los comandos realizados.

```
jesolis14@shadow:~/Escritorio/PCD/A-02/aclImdb/train$ ls neg | awk -F'[_ .]' '{print $2}' > puntaje_negativo.txt
jesolis14@shadow:~/Escritorio/PCD/A-02/aclImdb/train$ ls pos | awk -F'[_ .]' '{print $2}' > puntaje_positivo.txt
jesolis14@shadow:~/Escritorio/PCD/A-02/aclImdb/train$ sed 's/\\usercomments$//' urls_neg.txt > urls_negativo.txt
jesolis14@shadow:~/Escritorio/PCD/A-02/aclImdb/train$ sed 's/\\usercomments$//' urls_pos.txt > urls_positivo.txt
jesolis14@shadow:~/Escritorio/PCD/A-02/aclImdb/train$ paste urls_negativo.txt puntaje_negativo.txt > critica_negativo.txt
jesolis14@shadow:~/Escritorio/PCD/A-02/aclImdb/train$ paste urls_positivo.txt puntaje_positivo.txt > critica_positiva.txt
jesolis14@shadow:~/Escritorio/PCD/A-02/aclImdb/train$ awk '{critica_negativo[$1]++$2; count[$1]++} END {for (url in critica_negativo) print url, critica_negativo[url]/count[url]}' critica_negativo.txt | sort -k2 -n | head -n 10 > peores_peliculas.txt
jesolis14@shadow:~/Escritorio/PCD/A-02/aclImdb/train$ awk '{critica_positiva[$1]++$2; count[$1]++} END {for (url in critica_positiva) print url, critica_positiva[url]/count[url]}' critica_positiva.txt | sort -k2 -n | tail -n 10 > mejores_peliculas.txt
jesolis14@shadow:~/Escritorio/PCD/A-02/aclImdb/train$ █
```