COMP1204 - Coursework 1: Unix

Daniel Best (Student ID: 29777127)

March 5, 2019

1 Scripts

1.1 Basic File Processing (countreviews.sh)

The first task of this coursework was to perform some basic file processing on the TripAdvisor data. Specifically, this involved retrieving the number of reviews for each hotel file, and then outputting a sorted list of hotels, ordered by the most reviews to least.

The script file designed for this task, **countreviews.sh**, can be seen below:

```
#!/bin/bash

for file in $1/*; do
        echo "$(basename $file .dat) $(grep -E -c "<Author>" $file)"
done | sort -n -r -k 2
```

```
#!/bin/bash

for file in $1/*; do
    echo "$(basename $file .dat) $(grep -E -c "<Author>" $file)"
done | sort -n -r -k 2
```

This script loops through each **file** in given file path, which is represented as **$1/*** in the code. For each file, the script outputs the **basename**, i.e. the name of the file without the file path or extension, and the number of **<Author>** tags in the file, which represents the amount of reviews. This second value is found by using the **grep** command, with the **-c** argument specifying that the number of instances of the search query found should be returned, as opposed to the usual output of the found instances themselves.

Once the for loop has been completed, the output is piped into the **sort** command, which sorts the results in numerical order (**-n**), in reverse order (**-r**), and based on the second column (**-k 2**).

1.2 Data Analysis (averagereviews.sh)

2 Discussion