

Counting hotel reviews

Sandor Kovacs id: 31773478

March 2021

1 Introduction

In this coursework the problem was to sort hotels according to their review number. This way the popularity and effectiveness of a hotel could be measured.

2 Overall solution

I tried doing this problem by first finding the directory where the files are. After that, finding a unique word which appears only once in every review was a key aspect of this project so we could count the number of reviews. Then we needed to sort the hotels by their number of reviews.

3 The code

```
#!/bin/bash
cd $1
for FILE in *;
do
echo -n $FILE | sed -e "s/.dat//"
echo -n " "
grep -i -c "<Author>" $FILE
done | sort -n -k2 -r
```

3.1

```
cd $1
```

This line is used to change the directory (cd) so that we are in the same directory where the "hotel number.dat" files are. Now the \$1 is to take the argument from the command line which represents the path where the files are.

3.2

```
for FILE in *;
```

This loops through every single file in the directory which we are currently in represented by *.

```
do
```

```
done
```

The "do" and "done" parts represent the beginning and the ending of the for cycle. The code written within these two command will be executed for each file in the directory.

3.3

```
echo -n $FILE | sed -e "s/.dat/"
echo -n " "
grep -i -c "<Author>" $FILE
```

The first line of the code eliminates the ".dat" part from the file so only the name of the hotel remains. The "-n" part allows us to write more argument in the same line. After this we put a space as a separator. The third line counts the number of times "Author" appears in the document and then writes that number.

3.4

```
| sort -n -k2 -r
```

This line sorts the output of the for cycle according to the second column "-k2". After that it reverses the order as the original sorted it in increasing order but we need it in decreasing order.