Day 4: 'Uniq' Command #2

Objective

In this challenge, we practice using the *uniq* command to eliminate consecutive repetitions of a line when a text file is piped through it.

Resources

Plain Uniq

If this is the file *test.txt*:

```
00
00
01
01
00
00
02
02
```

This is the output on passing it through the *uniq* command, either via pipes or as input via STDIN:

```
Command:  uniq < test.txt

00
01
00
02
```

The first two lines of the original file are the same, (00). The next two lines are (01) which are followed by two repetitions of (00) again and two repetitions of (02). The *uniq* command replaces the consecutive repetitions with only one line in each case.

Uniq with counts

```
uniq -c < test.txt
```

This example indicates the count of repetitions for each of the lines it collapses.

If this is the test file, *testCounts.txt*:

```
00
00
01
01
00
00
02
02
03
aa
aa
aa

Command: uniq -c < input00.txt

    2 00
```

```
2 01
2 00
2 02
1 03
3 aa
```

The first number is the *count* of the repeated occurrences in the original file.

Printing only duplicate lines

The *-d* option only prints those lines that are followed by one or more repetitions immediately after them:

```
uniq -d < testCounts.txt
```

OR

```
cat testCounts.txt | uniq -d
```

OR

```
uniq -d testCounts.txt
```

Printing only unique lines

The *-u* option only prints those lines that are succeeded and preceded by different lines:

```
uniq -u < testCounts.txt
```

OR

```
cat testCounts.txt | uniq -u
```

OR

```
uniq -u testCounts.txt
```

Other possible options:

- Limit comparison only to the first N characters (using the *-w* option).
- Avoid comparing the first N characters (using the *-s* option).
- Ignore variations in case between lines (using the *-i* option).
- Avoid comparing the first N fields (using the *-f* option).
  (This may be useful while processing TSV files when you'd like to ignore the first column if it has serial numbers.)

You might find these examples interesting and useful.

Task

Given a text file, count the number of times each line repeats itself. Only consider consecutive repetitions. Display the space separated count and line, respectively. There shouldn't be any leading or trailing spaces. Please note that the *uniq -c* command by itself will generate the output in a different format than the one expected here.

Sample Input

```
00
00
01
01
00
00
02
02
03
aa
aa
aa
```

Sample Output

```
2 00
2 01
2 00
2 02
1 03
3 aa
```

Explanation

```
00 is repeated twice
01 is repeated twice
00 is repeated twice
02 is repeated twice
03 occurs once
aa is repeated thrice
```