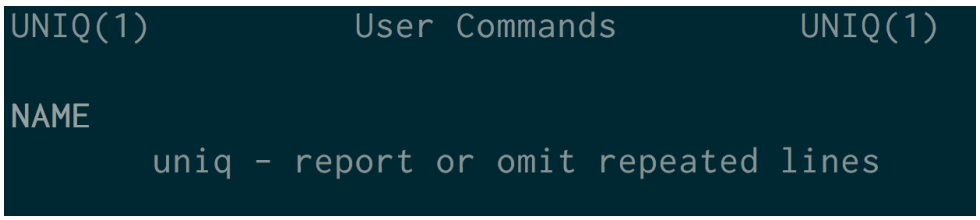
Last updated Saturday, Nov 16, 2019

Linux and Unix uniq command tutorial with examples

Tutorial on using uniq, a UNIX and Linux command for reporting or filtering repeated lines in a file. Examples of showing a count of occurrences, showing only repeated lines and ignoring characters and specific fields.

Estimated reading time: 4 minutes

Table of contents



What is the uniq command in UNIX?

The `uniq` command in UNIX is a command line utility for reporting or filtering repeated lines in a file. It can remove duplicates, show a count of occurrences, show only repeated lines, ignore certain characters and compare on specific fields. The command expects adjacent comparison lines so it is often combined with the `sort` command.

Uniq expects adjacent lines

The `uniq` commands expects adjacent lines in inputs. To find unique occurrences where the lines are not adjacent a file needs to be sorted before passing to `uniq`. `uniq` will operate as expected on the following file that is named `authors.txt`.

```
Chaucer
Chaucer
Orwell
Larkin
Larkin
```

As duplicates are adjacent `uniq` will return unique occurrences and send the result to standard output.

```
uniq authors.txt
Chaucer
```

```
Orwell
Larkin
```

Suppose that a file exists where the duplicates in the file are not adjacent.

```
Chaucer
Larkin
Orwell
Chaucer
Larkin
```

Passing this file to `uniq` will simply return the contents of the file. Where files are not already sorted the `sort` command can be used to sort the file first before piping to `uniq`. An article outlining the usage of `sort` is available here (/unix-sort/).

```
sort authors.txt | uniq
Chaucer
Orwell
Larkin
```

How to show a count of the number of times a line occurred

To output the number of occurrences of a line use the `-c` option in conjunction with `uniq`. This prepends a number value to the output of each line.

```
uniq -c authors.txt
2 Chaucer
2 Larkin
1 Orwell
```

How to only show repeated lines

To only show repeated lines pass the `-d` option to `uniq`. This will output only lines that occur more than once and write the result to standard output.

```
uniq -d authors.txt
Chaucer
Larkin
```

How to only show lines that are not repeated

To only show lines that are not repeated pass the `-u` option to `uniq`. This will output only lines that are not repeated and write the result to standard output.

```
uniq -u authors.txt
Orwell
```

How to ignore characters in comparison

To ignore characters in a comparison pass the `-s` option to `uniq`. This will ignore the characters specified in the comparison and output the result to standard output.

Suppose a list of authors exsits in a file that is saved as `authors.txt`. The file has some numbers in front of the names of the authors.

```
1Chaucer
2Chaucer
3Larkin
4Larkin
5Orwell
```

To return a list of the authors numbers can be ignored by using the `-s` option. This will skip the number of characters it is given before doing the comparison.

```
uniq -s 1 authors.txt
1Chaucer
3Larkin
5Orwell
```

Note that the first occurrence is taken and the line is printed out as is. If the output needs to cleaned this can be achieved by piping to something like `sed` .

```
uniq -s 1 compare.txt | sed s/^.//
Chaucer
Larkin
Orwell
```

How to ignore fields in comparison

To ignore fields in a comparison pass the `-f` option to `uniq` . This will run the comparison on the specified field and output the result to standard output.

Suppose a file exists with a list of cricketers and the clubs that they play for. This is saved as `cricketers.txt` .

```
Tom Westley Essex
Ravi Bopara Essex
Marcus Trescothick Somerset
Joe Root Yorkshire
Jonny Bairstow Yorkshire
```

A field is considered as a string of non-blank characters separated from adjacent fields by blanks. The `uniq` utility may be used to group by the county that these cricketers play for.

```
uniq -f 2 cricketers.txt
Tom Westley Essex
Marcus Trescothick Somerset
Joe Root Yorkshire
```

As with the `-s` option `uniq` outputs the first occurrence it finds. It is possible to combine with the `-c` option to output a count.

```
uniq -f -2 cricketers.txt
2 Tom Westley Essex
1 Marcus Trescothick Somerset
2 Joe Root Yorkshire
```

To just see the list of counties `sed` and `cut` may be used to clean this up.

```
uniq -f 2 -c cricketers.txt | sed 's/^\s*//' | cut -d ' ' -f 1,4
2 Essex
```