

Filtering Repeated Lines Out

Review different ways to search for repeated lines in a text or a file.

uniq

Definition:

`uniq` command in bash is a command line utility to filter and view multiple repeated lines.

This command works on *adjacent comparison lines* so it is often combined with the **sort** command.

It can be used to:

- remove duplicates.
- show only repeated lines.
- show a count of repeated occurrences.
- comparing particular fields and ignoring certain inputs.

Syntax:

```
uniq [option] [input[output]]
```

Options:

| Option | Description |
|--------|--|
| -c | Prefix lines with a number showing how many times they occurred. |





| | |
|----|---|
| -d | Only print duplicated lines. |
| -u | Only print unique lines. |
| -Z | End lines with 0 byte (NULL), instead of a newline. |
| -w | Compare no more than N characters in lines. |
| -i | To perform case-insensitive comparisons. |
| -f | To avoid comparing first N fields of a line before determining uniqueness. (Field is a set of characters delimited by a white space.) |
| -s | To avoid comparing first N characters before determining uniqueness. |

Examples:





- To display repeated lines:





- To ignore first field, containing line numbers from each line while doing comparison, use `-f` option and execute the following snippet. In this way, both lines will be considered alike:

| | |
|---|---|
| <div>main.sh</div> <div>file.txt</div> | <div>uniq -f 2 file.txt</div> <div></div> |
| <div></div> <div></div> | |

- Now using **-s** option, we can ignore characters instead of fields while comparing:

| | |
|---|---|
| <div>main.sh</div> <div>file.txt</div> | <div>uniq -s 3 file.txt</div> <div></div> |
| <div></div> <div></div> | |

- **uniq** command works on the input that is already sorted. Hence, it is usually combined with the **sort** command to find repeated lines:

| | |
|---|---|
| <div>main.sh</div> <div>file.txt</div> | <div>sort file.txt uniq</div> <div></div> |
| <div></div> <div></div> | |

- To show a count of the number of times a line occurred, we use the option **-c**:

| | |
|---|--|
| <div>main.sh</div> <div>file.txt</div> | <div>sort file.txt uniq -c</div> <div></div> |
| <div></div> <div></div> | |

- To show lines that are not repeated, we use the option **-u**:

| | |
|--------------------|--|
| <div>main.sh</div> | <div>sort file.txt uniq -u</div> <div></div> |
|--------------------|--|

file.txt

