## 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.

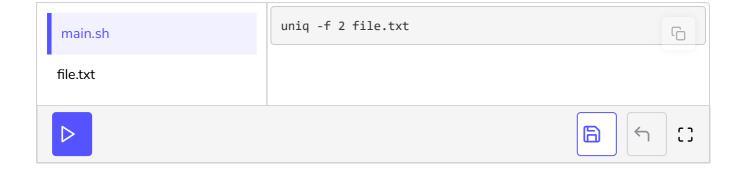
Only print duplicated lines. -d Only print unique lines. -u End lines with 0 byte (NULL), -Zinstead of a newline. Compare no more than N -W characters in lines. To perform case-insensitive -i comparisons. To avoid comparing first N fields of a line before determining uniqueness. (*Field* is a set of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

### **Examples:**

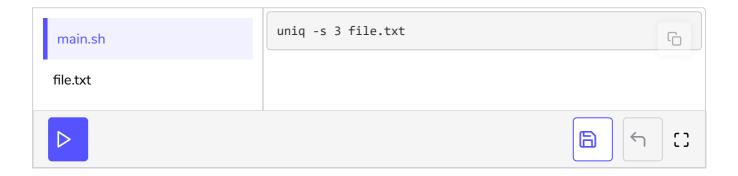
• To display repeated lines:



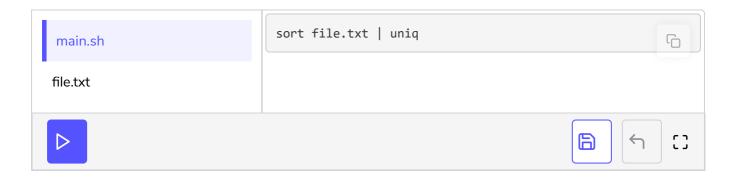
• To ignore first field, containing line numbers from each line while doing comparison, use -f option and execute the following snippet. In this way, both lines will be considered alike:



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



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



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



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



file.txt	
	ر ر