uniq with flags for manipulating duplicate lines in Linux

The uniq command in Linux helps you identify and manage duplicate lines within text files. Here's a breakdown of its usage and common flags:

Basic Functionality:

- By default, uniq compares entire lines and removes consecutive duplicate lines, keeping only the first unique occurrence.

Flags:

- -c (count): Prefixes each unique line with the number of times it appears (useful for identifying frequencies).
- -d (repeated): Prints only lines that appear more than once (opposite of default behavior).
- -i (ignore case): Makes the comparison case-insensitive.
- -w (skip characters): Compares lines based on a specific number of characters at the beginning (useful for skipping leading identifiers).
- -f (skip fields): Compares lines based on all characters except for those in specified fields (useful for ignoring specific data points).
- -s (skip characters): Ignores a specific number of characters at the beginning of each line (similar to -w but can skip any position).
- -o (output separator): Specifies a delimiter to use between the count (with -c) and the line (defaults to a space).

Examples:

1. **Remove duplicate lines (default):**
   Bash
   ```
   uniq lines.txt
   ```

2. **Count occurrences of each line:**
   Bash
   ```
   uniq -c data.txt
   ```

3. **Print only duplicate lines (more than once):**
   Bash
   ```
   uniq -d errors.log
   ```

4. **Compare lines ignoring case:**
   Bash
   ```
   uniq -i mixed_case.txt
   ```

5. **Compare lines based on the first 10 characters:**
   Bash
   ```
   uniq -w 10 user_ids.txt
   ```

6. **Compare lines ignoring the second field (separated by spaces):**
   Bash
   ```
   uniq -f 2 transactions.txt
   ```

Tips:

- Use sort before uniq to ensure the lines are sorted for accurate duplicate identification.
- Combine flags like -c and -d for specific analyses (count duplicates and print them only).
- Experiment with -w, -f, and -s to compare lines based on specific parts of the data.

By effectively using these flags, you can customize uniq to manage duplicate lines in your text files, making it a valuable tool for data cleaning and analysis.