

ASSIGNMENT 2: SED

CS3423 - Systems Programming

Rocky Slavin and Steven O'Hara - UTSA

For this assignment, you will use **sed**, **bash**, and the other utilities you have used in class to create a program for redacting personally-identifiable information. Your program should take a path to a directory and an integer representing the number of days old a file within that directory must be to be redacted (use the date it was last modified). The program should take an *optional* third parameter, **-r**, which will cause the program to operate recursively within the directory. Without this option, the program should not traverse into subdirectories.

This assignment requires only sed, bash, and the other utilities used so far in class. **Do not** use awk, Python, or any other languages/utilities.

Redaction Rules

For all files fitting the criteria, the following changes should be made *in place*. No other changes should be made to the file.

- **Social Security numbers** should be replaced with `%%-%%-%%%`
- **Phone numbers** should have their digits replaced with `'#'`. Phone numbers can take the following forms where `d` is a digit and `s` is any number of whitespace characters, including zero.
 - `ddd-dddd`
 - `ddd-ddd-dddd`
 - `(ddd)sddd-dddd`
- **Any strings between `'%'` characters** should be completely removed, including the `'%'` character. You may assume such instances will not span multiple lines.
- **Any line beginning with `'*' followed by one or more spaces`** should have the line `"* ATTENTION *` (note the single spaces around "ATTENTION") preceded by a blank line and followed by a blank line.
- **Dates of the form `"DD/MM/YYYY"`** should be converted to `"YYYY-MM-DD"`. Days and months will not be zero-padded before or after they are updated.

Example

Original redactme.txt:

```
1 *Cast and characters
2 * Bob Denver is Gilligan , the inept , accident-prone First Mate
3 (affectionately known as "Little Buddy" by "the Skipper") of the SS Minnow.
4 Denver was not the first choice to play Gilligan; actor Jerry Van Dyke, phone
5 210-222-3333, was offered the role on 2/11/1963, but he turned it down,
6 believing that the show would never be successful. He chose instead to play the
7 lead in My Mother the Car, which premiered the following year and was cancelled
8 after one season. The producers looked to Bob Denver, the actor who had played
9 Maynard G. Krebs, ss #111-22-3333, the goofy but lovable beatnik in The Many
10 Loves of Dobie Gillis. None of the show's episodes ever specified Gilligan's
11 full name or clearly indicated whether "Gilligan" was the character's first name
12 or his last. In the DVD collection , Sherwood Schwartz states that he preferred
13 the full name of "%Willy Gilligan%" for the character.
```

Redacted Version of redactme.txt:

```
1 *Cast and characters
2
3 * ATTENTION *
4
5 * Bob Denver is Gilligan , the inept , accident-prone First Mate
6 (affectionately known as "Little Buddy" by "the Skipper") of the SS Minnow.
7 Denver was not the first choice to play Gilligan; actor Jerry Van Dyke, phone
8 ###-###-#### was offered the role on 1963-11-2, but he turned it down,
9 believing that the show would never be successful. He chose instead to play the
10 lead in My Mother the Car, which premiered the following year and was cancelled
11 after one season. The producers looked to Bob Denver, the actor who had played
12 Maynard G. Krebs, ss #%%-%-%%% the goofy but lovable beatnik in The Many
13 Loves of Dobie Gillis. None of the show's episodes ever specified Gilligan's
14 full name or clearly indicated whether "Gilligan" was the character's first name
15 or his last. In the DVD collection , Sherwood Schwartz states that he preferred
16 the full name of "" for the character.
```

Script Execution

Your program should be invoked through a single bash file (see below) with the path to the directory containing files as the argument, an integer representing how old the files should be before they are redacted, and an optional `-r` option to make the program behave recursively (this will always be the third argument).

Example: `$ assign2.bash /path/to/dir 2 -r`

Assignment Data

A sample input file can be found in:

`/usr/local/courses/rslavin/cs3423/Spring19/assign2.`

When using this data, remember that you will be made to overwrite the files. Be sure to make a backup of the files and restore them every time you run the script.

Script Files

Your program should consist of *exactly* two files:

- `assign2.bash` - the main file which is initially invoked
- *Exactly* one `.sed` file which is used for a sed invocation run in `assign2.bash`.

Verifying Your Program

Your program must work for *arbitrary* files by applying the rules above. You can test your program with the input provided in `redactme.txt` and compare the output with `redacted.out` using `diff` (check the man-pages on how to use it). You should create your own test cases to test for the recursion feature. Use the `touch -t` command to change a file's date.

Submission

Turn your assignment in via Blackboard. Your zip file, named `abc123.zip` (with your abc123) should contain only your bash and sed files.