

ASSIGNMENT 2: SED

CS 3424 - Systems Programming

For this assignment, you will use **sed**, **bash**, and the other utilities you have used in class to create a program for use in an academic setting for which redacting sensitive information for extra-departmental communication is important. Your program should take the names of one or more files that are to be redacted as *command line arguments*.

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

Hint: make use of the `\W` character class (equivalent to `[^a-zA-Z0-9_]`).

Redaction and Substitution Rules

For all files so specified, the following changes shall be made *in place*. No other changes should be made to these files.

Note, that substitutions should only be made for *whole* matches for all of the criteria that follow. For example, if they are embedded within a larger sequence, such as “*The password is canterbury987*”, you should *not* redact anything within this sequence.

- **Student myUTSA ID's** consist of *abc123*, where *abc123* represents an actual myUTSA ID. These ID's should be redacted by simply replacing them with the sequence XYZ000.
- **Student Banner ID numbers** must be redacted, and are identifiable as consisting of a sequence beginning with an “at-sign” (@) followed by precisely eight (8) digits, as in: @83098247. These ID numbers should be replaced simply with the sequence @XXXXXXXX.
- **Student Grades** should be similarly obliterated. Grade letters will consist of a single letter (case-insensitive) from the set $\{A, B, C, D, F\}$, *optionally* followed by a symbol from the set $\{+, -\}$. Grades should be redacted through simple substitution with the character X.
- **Student Grade Point Averages** will consist of a single digit from the set $\{0, 1, 2, 3, 4\}$, followed by a period / dot (.) and ending with one to three digit decimal places. GPA's should simply be replaced with the sequence “X.X”.
- **Lines slated for redaction** will consist of a line that begins with one or more “X” characters. When these lines are encountered, they should be deleted in their entirety.

Example

Original redactme.txt:

```
1 Student Transcript for John Doe, jkl567 , @99029832
2
3 Semester Fall 2020
4   CS 3424 Systems Programming          B+
5   CS 3733 Operating Systems            C-
6   CS 4853 Adv. Systems Programming     D
7   MAT 1013 Calculus I                  B-
8 Semester grade point average:         2.18
9
10 X Academic probation active
11 X Academic holds: 3
```

Redacted Version of redactme.txt:

```
1 Student Transcript for John Doe, XYZ000, @XXXXXXXXX
2
3 Semester Fall 2020
4   CS 3424 Systems Programming          X
5   CS 3733 Operating Systems            X
6   CS 4853 Adv. Systems Programming     X
7   MAT 1013 Calculus I                  X
8 Semester grade point average:         X.X
```

Script Execution

Your program should be invoked through a single bash file (see below) with the filename(s) containing the sensitive data as argument(s).

Example: \$ assign2.bash redactme.txt

Assignment Data

A sample input file can be found in:

/usr/local/courses/ssilvestro/cs3424/Spring21/assign2

When using this data, remember that your 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 `assign2.sed` file which is used for a `sed` executed to be invoked via `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.txt` using `diff` (check the man-pages on how to use it). You should create your own test cases to test for the recursion feature.

Submission

Turn your assignment in via Blackboard. Your zip file, named `a2-abc123.zip` (where *abc123* represents your myUTSA ID) should contain only your singular `bash` and `sed` files.