

# ASSIGNMENT 2: SED

---

For this assignment, you will use `sed` and `bash` **only**, to the exclusion of all other utilities we have discussed in class, to create a program for use by a municipality in both redacting sensitive information from internal communications as well as simplifying and standardizing the format of these documents prior to their release and circulation. Your program should take the names of one or more input files that are to be redacted as command line arguments. Redaction should be performed **in-place** (i.e., the original files contents should be changed subsequent to your script's invocation; however, while developing/troubleshooting, perhaps turning off in-place editing in order to avoid permanently modifying the input file(s) would be helpful to you).

This assignment requires only `sed` and `bash`. **Do not** use `grep`, `find`, `awk`, Python, or any other languages/utilities.

## Redaction and Substitution Rules

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

- **Driver's License numbers** begin with `xxDL`, where `xx` is a two-letter state code identifying the origin of the issuing state. Following this code is a space character, then a license number composed of exactly 6 to 8 digits in length. For example, the following are all valid driver's license numbers, as well as their corresponding redacted form (see details below):

- TXDL 12345678 → DL <redacted>

- WADL 1234567 → DL <redacted>

- VA DL 123456 → DL <redacted>

The state code should be redacted by simply removing the two-character state code (and optional space separator) from the string, leaving only the letters "DL". Further, the license number should be redacted by simply replacing the entire number sequence with the string <redacted>.

- **Credit Card Numbers** Credit card numbers must be redacted, but it is desirable that the censored text still retain information suitable for the identification of both the type of card (i.e., the issuer), as well as its last four digits. Each card network specifies a unique first digit to the cards they issue, and typically contain 16 digits in total. American Express is an exception to this, in which the second number can only be a 4 or a 7, and the total number of digits is only 15. In summary:

- Visa cards: Begin with a 4 and have 16 digits

- Master Cards: Begin with a 5 and have 16 digits

- American Express cards: Begin with a 3, followed by a 4 or a 7, and have 15 digits
- Discover cards: Begin with a 6 and have 16 digits

Card number data should be redacted as in the following examples. Note that 16 digit numbers may or may not be separated into groups of four using hyphens – this is optional. Hyphenation of American Express cards into sections of 4, 6, and 5 digits is similarly common, and also optional. Examples of the expected substitutions for each card type are as follows:

- 5098-7811-1980-3129 → MC-3129
- 3770-718209-88900 → AMEX-8900
- 6133-0012-6612-5389 → DISC-5389
- 4329911234331302 → VISA-1302

- **Texas Vehicle License Plate numbers** should be similarly obliterated. Texas vehicle plates appear in one of two formats, but both will be written with the letters TX and an optional space preceding them. The first type is six alphanumeric characters, optionally separated by a hyphen in the middle. The second type begins with three alphabetic characters, followed by four digits, again optionally separated by a hyphen. Examples of valid type one Texas license plate numbers are:

- TX 64Q7JL
- TX 64Q-7JL
- TX64Q7JL
- TX64Q-7JL

Examples of valid type two Texas license plate numbers are:

- TX RYD3792
- TX RYD-3792
- TXRYD3792
- TXRYD-3792

These numbers should be redacted by simply replacing the license plate number with the string <redacted>.

- **Current Date Placeholder** The document authors may use the shorthand symbol <date> in order to insert the current date (i.e., today's date). Regardless of the date on which your script is run, this placeholder should be updated with the correct current date.

- **Municipality Name Placeholder** The authors of these documents may use the shorthand symbol <mname> in order to designate the full name of their municipality. Any such references should be replaced with the full name: City of Brook Haven, Connecticut.

## Example

*Note: Obtain original copies from course directory; do not attempt to copy/paste from this PDF, as certain character renderings/substitutions have occurred (e.g., en-dash vs hyphen, whitespace, etc.).*

### Original redactme.txt:

```
1 <mname>
2 Date: <date>
3 Title: [INTERNAL] Memo #762
4
5 Dear staff:
6
7 Memorandum #762 has been amended as follows , in accordance with the
8 updated employee operations and purchasing policy*:
9
10 The only employees authorized to operate vehicle #89 (license plate
11 TX RYN9001), vehicle #1467 (license plate TX 89D-7VP), and vehicle #1468
12 (license plate TX DPD-4981) are those employees who possess the following
13 driver 's licenses :
14
15     TXDL 09182745
16     CADL 00173389
17     PADL 451903
18     AKDL 6770912
19
20 Further , usage of city credit cards will be strictly limited to the
21 following departmental cards , until further notice :
22
23     5090-4839-9284-2931
24     3445-717892-72600
25     6093-2033-0662-8088
26     4373130058122033
27
28 Thank you ,
29     Mgmt
30
31 * Policy revision date 2/1/13 (originally passed 7/13/92).
```

### Redacted Version of redactme.txt:

```
1 City of Brook Haven , Connecticut
2 Date: 7/09/2024
3 Title: [INTERNAL] Memo #762
4
5 Dear staff:
6
```

```
7 Memorandum #762 has been amended as follows , in accordance with the
8 updated employee operations and purchasing policy *:
9
10 The only employees authorized to operate vehicle #89 (license plate
11 TX <redacted>), vehicle #1467 (license plate TX <redacted>), and vehicle #1468
12 (license plate TX <redacted>) are those employees who possess the following
13 driver 's licenses :
14
15     DL <redacted>
16     DL <redacted>
17     DL <redacted>
18     DL <redacted>
19
20 Further , usage of city credit cards will be strictly limited to the
21 following departmental cards , until further notice :
22
23     MC-2931
24     AMEX-2600
25     DISC-8088
26     VISA-2033
27
28 Thank you ,
29     Mgmt
30
31 * Policy revision date 2/1/13 (originally passed 7/13/92) .
```

## Script Execution

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

**Example invocation:** `$ assign2.bash redactme.txt redactme2.txt redactme3.txt`

## Script Files

Your program should consist of *exactly* two files:

- `assign2.bash` - the main file which is initially invoked
- *Exactly* one sed script named `assign2.sed` to be used within a sed invocation in `assign2.bash`. You should not have (nor need) more than two sed invocations from `assign2.bash`.

## Verifying Your Program

Your program must work for *arbitrary* files by applying the rules above. You can test your program with input of the type provided in above and compare the output with your modified output using `diff` (check the man-pages on how to properly use this command).

## Submission

Turn your assignment in via Blackboard. Your submission, named `a2-abc123.zip` (with your myUTSA ID) should contain only your two bash and sed files. Your submission must adhere to the above naming/format requirement, as well as only contain a one or two files at most, as mentioned above. No extra directories or files should be included.