Exercise: gathering and formatting data

CSC230 Database Technologies for Analytics

23 October 2019

The Emacs text editor

Emacs is a popular and powerful text editor with a long history. In the 2013 movie The Internship, Nick Campbell (played by Owen Wilson) announces his opinion that Emacs should be the default editor in Linux. In this way, he impresses the engineer at Google with whom he has developed a romantic interest. The movie is a comedy and Wilson's line is funny, but he spoke the truth!

We will use the control (Ctrl) and escape (Esc) keys to enter commands in the Emacs text editor.

You will find the control key in the lower left corner of your keyboard and also to the right of the space bar between the alt key and the arrow keys.

You will find the escape key in the upper left corner of your keyboard.

When we press the control key, we will press another key at the same time.

- `c-x` means "press the Ctrl (control) and x keys simultaneously"

- `c-s` means "press the Ctrl (control) and s keys simultaneously"

- `c-c` means "press the Ctrl (control) and c keys simultaneously"

- `c-x c-s` means `c-x` and then `c-s` ("press Ctrl and x, and then press Ctrl and s")

- `c-x c-c` means `c-x` and then `c-c` ("press Ctrl and x, and then press Ctrl and c")

- . . . and so on

When we press the escape key, we will press just the escape key. Then, after releasing the escape key, we will press another key or type the name of a command.

- `esc-<` means "press the Esc (escape) key and then press the $<$ key"

- `esc->` means "press the Esc (escape) key and then press the $>$ key"

- `esc-x re-search-forward` means "press the Esc (escape) key, then press the x key, and then type 're-search-forward' after the prompt that will appear at the bottom of the window"

Here are some Emacs commands that will be especially useful in our exercise.

We need a way to open a file, save a file, and exit from Emacs.

- `c-x c-f` open a file (a prompt will appear at the bottom of the window— type in the name of a new or existing file)

- `c-x c-s` save a file

- `c-x c-c` exit from Emacs

With these commands, we can move about within a document and cut and paste text.

- `c-d` "delete" deletes a single character

- `c-k` "kills" deletes all text to the right of the cursor

- `c-y` "yanks" (pulls back from the dead) pastes previously killed text into the document at the position of the cursor

- `c-f` "forward" moves the cursor one character to the right

- `c-b` "backward" moves the cursor one character to the left

- `c-n` "next" moves the cursor one line down

- `c-p` "previous" moves the cursor one line up

- `c-a` moves the cursor to the beginning of the line

- `c-e` moves the cursor to the end of the line

- `c-.` sets a mark (to begin the definition of a region)

- `c-w` kill all text between the mark and the current position of the cursor

- `esc-<` moves the cursor to the top of the document

- `esc->` moves the cursor to the bottom of the document

- `esc-f` (forward) moves the cursor to the right by one word

- `esc-b` (backward) moves the cursor the left by one word

If you get into trouble, you can abort a command by typing c-g.

We need a way to search for text.

- `c-s` search for an exact match (a prompt will appear at the bottom of the window—type the text for which you want to search)

- `esc-x re-search-forward` execute a search for text that matches pattern that is described by a regular expression (again, a prompt will appear at the bottom of the window)

- `esc-x re-fo` is the same as `esc-x re-search-forward`

You can learn about regular expressions in Emacs here or here.

Here are some examples.

- To find a sequence of letters (a word), type this: `esc-x re-fo [[:alpha:]]+`

- To find a sequence of words in which each pair of consecutive words is separated by a single space, type this: `esc-x \([[:alpha:]]+ \)+`

- To find a sequence of digits that is followed by a period and then a space, type this (with a space after the period): `esc-x re-fo [[:digit:]]+\.`

Commands for searching and replacing might also help us.

- `esc-%` search for match and prompt for a "y" (yes) or "n" (no) at each match to replace or not replace

- `esc-x replace-string` search and replace all matches

We also need a way to record a sequence of commands and then execute the stored sequence.

- `c-x (` begin the definition of a macro

- `c-x )` end the definition of a macro

- `c-x e` execute a macro

Here is another PDF document (from another professor) that summarizes Emacs commands.

And here is another.

3

Practice

Log into our server. Copy brightest-stars.txt from the instructor's folder into your own.

```
cp ../leon/courses/cs230/brightest-stars.txt .
```

Open the file with Emacs.

```
emacs brightest-stars.txt
```

Move the cursor in the document left and right. Move a character at a time and a word at a time. Move to the beginning or end of a line with a single command.

Move up and down. Move a line at a time. Move to the top or bottom of the document with a single command.

Search for an exact match with a string that you specify. Search for a match with a pattern that you specify. For example, try specifying a pattern that describes a word or a pattern that describes a number.

Finally, follow the instructions within brightest-stars.txt to build a SQL INSERT command.

Challenge exercise

If you are enrolled in CSC314 Data Management Systems, learn how to use the AWK programming language to filter and reformat files.

Begin here.