

CGS600A: Computational Tools in Cognitive Science

Assignment #1: Pattern matching, regular expressions

Max marks: 80

Due on/before: 08.00, 13-Aug-2018.

5-Aug-2018

Note: Some questions may require the use of different options in (e)grep and may also require multiple (e)grep commands - since (e)grep is line oriented. Read the (e)grep manual carefully. Use `info grep` in Linux for the full manual or you can find it at GNU's `grep` webpage.

1. Write (e)grep commands for the following:

- (a) An integer (without leading 0s, can be negative) that is divisible by 5.
- (b) Prints lines where after a punctuation character (e.g. '.', ',', etc.) there is no white space and the next character immediately follows the punctuation character. For example:

`This is,not correct.`

`This is, correct.`

If the punctuation character is at the end of a line or a file it is legal.

- (c) You have a file that contains the following pattern at different places in the file.

`FIRST`

`SECOND`

`THIRD`

Note that these lines contain only the text shown. Your command has to count the number of times such a pattern occurs in the file.

[5,10,20=35]

2. Currency amounts can be written in many ways. For example:

`Rs.23`, `Rs.23.00`, `Rs.23.0` or `Rs.0.50`, `Rs..50`, `Rs.0.5` etc.

Your command should print all correctly matching currency amounts. For testing assume the file contains one currency amount per line in Rupees.

[20]

3. Given text marked up with HTML tags you want to identify lines that contain a valid anchor tag (i.e. the `<a ...>` pattern) that opens in a new tab or window. Typically, this will look like:

```
<a href="link-spec" target="_blank">
```

However, many variations are legal. Here are the ones you should account for:

- a) One or more spaces can be present after `<` or before `>` or on either side of `=` or between attribute-value pair `href` and `target`.

- b) The attribute-value pairs href and target can occur in any order and other attribute-value pairs may also be present.
- c) The tag and attribute names (like href, target) can be in any case.

Assume `link-spec` can be any non-empty string and both attribute-value pairs href and target must be present.

[25]