"Pattern Matching"

Day 4

Introduction

Udeshya Dhungana
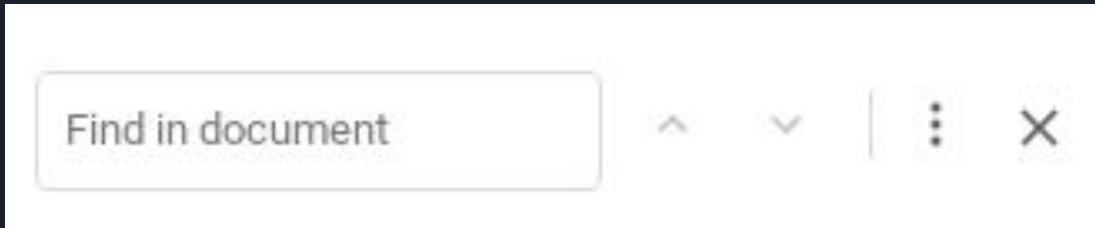
BCT 3rd Year Student

Pulchowk Campus

Objectives

- To learn about Regular Expressions (RegEx)
- To learn to use `re` module in python
- Construct regex for phone number
- Construct regex for email addresses
- To use regex to filter data

RegEx?

Find in document

- You might be familiar with 'Ctrl + F' feature. It's almost everywhere
- Ctrl + F is good, but do you know what's better?
  - RegEx

- RegEx matches patterns instead of searching for literal contents

If you have a problem, and you try solving it using RegEx, congratulations!

Now you have two problems!

But, will I even use this thing?

- Yes, and you'll need this much more than you think

- Web Development
- Data Wrangling
- UNIX systems
- …

Let's try a problem without regex

Let's write a python script that decides whether or not the supplied input is a phone number

And this time with RegEx

RegEx workflow in python

- Create the matching RegEx string
- Compile
- Search / Match
  - `re.search()` to search for regex pattern in a string
  - `re.fullmatch()` to check whether a string matches a regex

`search` or `fullmatch`

- `search` if you want to SEARCH the string matching the regular expression inside another string

- `fullmatch` if you want to identify if a string fully agrees with the regular expression

Symbols and their meaning

Round 1: Simple Ones

Symbol	Meaning
abc	Letters
123	Digits
\d	Any digit (From 0 to 9)
\D	Any non-digit
\w	Any alphanumeric character
\W	Any non-alphanumeric character

Round 2

Symbol	Meaning
.	Any character
\.	Period
*	Zero or more repetitions preceding item
*	Asterisk
+	One or more repetitions of preceding item
\+	Plus sign
?	Optional character
\?	Question mark

Round 3

Symbol	Meaning
[abc]	Only a or b or c
[^abc]	Not a, b nor c
[a-z]	From a to z
\s	Any whitespace
{m}	M repetitions
(abc|def)	abc or def
(...)	Capture Group
^...$	Starting with and ending with

Where can I use this?

- Web Forms

- Web Scraping
  - BeautifulSoup