Chapter I. Regular expression

First. concept

Regular Expression as a form of expression, powerful and convenient.

Like a "wild card" (通配符, 无特定含义)

Second. Steps

1. Before you can use regular expression, you must import the library using "import re"

- New function "find" as a searcher, which can find the characters from the complete sentences.

eg. text = 'Hello World, I am Python'

Extraction = text.find('Python')

print(Extraction)

⟹ 14 (letter P's order)

- New function "rstrip" means remove all the character at the end of the sentence.

eg. text = 'Hello World, I am Python !!!'

modify_text = text.rstrip('!')

print(modify_text)

⟹ "Hello World, I am Python"

Third. "^X.*:" as a regular expression

- Match the start of the line (Match the capital X)
- Match any characters (match any times)
- The line should have the colon

result:

X-Sieve:

X-DSPAM-Result:

X-Plane is behind schedule

Extracting Data

- if we want the matching strings to be extracted, we use re.findall() (important!)

square bracket

regular expression: [0-9]+

Greedy matching will match every situation if it's match the requirements.

non-Greedy matching: just match the first situation if it's match the requirements.

eg. import re

lin = 'From stephen.marquard@uct.ac.za'

y = re.findall('@([^]*)', lin)

- match many of them
- match a non-blank character
- look through the string until you find this sign.

W+ one or more

String.split()