



Laporan Praktikum Algoritma dan Pemrograman

Semester Genap 2023/2024

| | |
|---------------------------|-----------------|
| NIM | Samuel Natanael |
| Nama Lengkap | 71231050 |
| Minggu ke / Materi | 14 / RegEx |

SAYA MENYATAKAN BAHWA LAPORAN PRAKTIKUM INI SAYA BUAT DENGAN USAHA SENDIRI TANPA MENGGUNAKAN BANTUAN ORANG LAIN. SEMUA MATERI YANG SAYA AMBIL DARI SUMBER LAIN SUDAH SAYA CANTUMKAN SUMBERNYA DAN TELAH SAYA TULIS ULANG DENGAN BAHASA SAYA SENDIRI.

SAYA SANGGUP MENERIMA SANKSI JIKA MELAKUKAN KEGIATAN PLAGIASI, TERMASUK SANKSI TIDAK LULUS MATA KULIAH INI.

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
YOGYAKARTA
2024

BAGIAN 1: MATERI MINGGU INI (40%)

RegEx

Regular expression adalah sebuah teks (string) yang mendefinisikan sebuah pola pencarian sehingga dapat membantu kita untuk melakukan matching (pencocokan), locate (pencarian), dan manipulasi teks.

Regex sangat berguna untuk mencari pola dalam teks, seruit apapun pola tersebut. Python adalah salah satu bahasa yang mendukung library re.

Ada banyak kegunaan dari Regex. Salah satunya adalah mencari jumlah kalimat yang berawal dari kata "from:" yang berbeda pengirimnya di file mbox-short.txt

```
# Untuk menggunakan regex harus import re terlebih dahulu
import re

# Membuka file mbox-short.txt
handle = open('File/mbox-short.txt')
count = 0

List = list()
# Setiap kata yang dilooping akan di search
for line in handle:
    line = line.rstrip()
    # re.search untuk menentukan apakah kalimat tersebut terdapat "From:"
    if re.search('From:', line) and line not in List:
        List.append(line)
        count += 1

for i in List:
    print(i)

print("Count: ", count)
output:
From: stephen.marquard@uct.ac.za
From: louis@media.berkeley.edu
From: zqian@umich.edu
From: rjlowe@iupui.edu
From: cwen@iupui.edu
From: gsilver@umich.edu
From: wagnermr@iupui.edu
From: antranig@caret.cam.ac.uk
From: gopal.ramasammycook@gmail.com
From: david.horwitz@uct.ac.za
From: ray@media.berkeley.edu
Count: 11
```

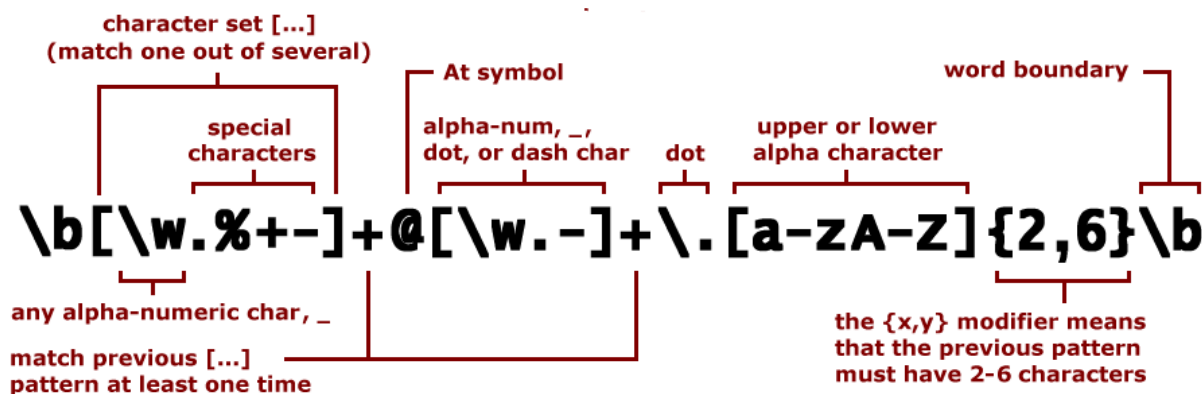
Pola Dalam RegEx

RegEx memiliki banyak cara untuk mengidentifikasi pola pada kalimat. Salah satu caranya adalah menggunakan beberapa qualifiers atau quantifiers untuk mencari pola yang diinginkan

Dibawah ini adalah beberapa command untuk menentukan pola yang bisa diinginkan dengan regex.

| QUALIFIERS | QUANTIFIERS | OTHERS |
|------------|-------------|--------|
| . | + | |
| \w | * | ^ |
| \d | qual? | \$ |
| \s | quant? | (...) |
| \W | {x} | \ |
| \D | {x,y} | |
| \S | {x,} | |
| [A-Z] | | |
| [a-z] | | |
| [0-9] | | |
| [abc] | | |
| [^abc] | | |

Gambar 14.1 Qualifiers, Quantifiers dll Dalam RegEx (diambil dari <https://www.thedataschool.com.au/mipadmin/regex-101-in-a-nutshell/>)



Gambar 14.2 Cara penggunaan RegEx (diambil dari <https://www.thedataschool.com.au/mipadmin/regex-101-in-a-nutshell/>)

Ada banyak website yang dapat digunakan untuk mempelajari atau memahami regex, seperti regexone.com dan regex101.com.

Exercise 2: Matching With Wildcards

| Task | Text | |
|-------|------|---|
| Match | cat. | ✓ |
| Match | 896. | ✓ |
| Match | ?=+. | ✓ |
| Skip | abc1 | |

...\.
Continue >

Gambar 14.3 Regexone.com

Dibawah ini adalah contoh penggunaan regex untuk mencari huruf kapital didalam kalimat

```
# Regex untuk mencari kata katipal
# String yang
indonesiaraya = """
Indonesia tanah airku
Tanah tumpah darahku
Di sanalah aku berdiri
Jadi pandu ibuku

Merdeka, merdeka
Tanahku, negeriku yang kucinta
Indonesia Raya
Merdeka, merdeka
Hiduplah Indonesia Raya

Indonesia Raya
Merdeka, merdeka
Tanahku, negeriku yang kucinta
Indonesia Raya
Merdeka, merdeka
Hiduplah Indonesia Raya
"""

kapital = re.findall("[A-Z][a-z]*\S", indonesiaraya)
print(kapital)

output :

['Indonesia', 'Tanah', 'Di', 'Jadi', 'Merdeka,', 'Tanahku,', 'Indonesia', 'Raya', 'Merdeka,', 'Hiduplah',
'Indonesia', 'Raya', 'Indonesia', 'Raya', 'Merdeka,', 'Tanahku,', 'Indonesia', 'Raya', 'Merdeka,', 'Hiduplah',
'Indonesia', 'Raya']
```

Regex diatas menggunakan

```
re.findall("[A-Z][a-z]*\S", indonesiaraya)
```

yang berarti mencari awalan A-Z yang diteruskan dengan a-z lalu berhenti sampai whitespace(spasi) menggunakan *\S

RegEx Functions

Re memiliki banyak fungsi untuk mencari string yang sesuai dengan pola yang dimasukan. Berikut adalah fungsi-fungsi yang disediakan dari regex.

| Function | Description |
|--------------------------------|--|
| <u>findall</u> | Returns a list containing all matches |
| <u>search</u> | Returns a <u>Match object</u> if there is a match anywhere in the string |
| <u>split</u> | Returns a list where the string has been split at each match |
| <u>sub</u> | Replaces one or many matches with a string |

Gambar 14.4 Fungsi RegEx (diambil dari https://www.w3schools.com/python/python_regex.asp)

1. re.findall()

```
# findall() function mengembalikan list yang cocok sesuai yang dicari
kalimat = "anton@mail.com dimiliki oleh antonius"
x = re.findall("anton", kalimat)
print(x)
# output : ['anton', 'anton']

# Jika tidak ada yang ditemukan, maka akan dikembalikan list kosong
x = re.findall("Portugal", kalimat)
print(x)
# output : []
```

2. re.search()

```
# search() function mencari string yang cocok, dan mengembalikan objek Match jika
ada yang cocok.
# Jika ada lebih dari 1 kali, maka hanya dikembalikan yang pertama
kalimat = "anton@mail.com dimiliki oleh antonius"
x = re.search("@", kalimat)
print(x)
# output : <re.Match object; span=(5, 6), match='@'>
```

```
# Jika tidak ada yang ditemukan maka dikembalikan None
kalimat = "anton@mail.com dimiliki oleh antonius"
x = re.search("gmail", kalimat)
print(x)
# output : None
```

(x.start() untuk mencari index kemunculan pertama dan x.end() untuk terakhir)

3. re.split()

```
# split() function mengembalikan list string telah dipisahkan pada setiap match
kalimat = "anton@mail.com dimiliki oleh antonius"
x = re.split("@", kalimat)
print(x)
# output : ['anton', 'mail.com dimiliki oleh antonius']

# Anda dapat mengontrol jumlah kemunculan dengan menentukan parameter maxsplit
kalimat = "anton@mail.com dimiliki oleh antonius, budi@gmail.co.id dimiliki oleh budi anwari"
x = re.split("@", kalimat, 1)
print(x)
# output : ['anton', 'mail.com dimiliki oleh antonius']
```

4. re.sub()

```
# fungsi sub() menggantikan kecocokan dengan teks pilihan
kalimat = "anton@mail.com dimiliki oleh antonius, budi@gmail.co.id dimiliki oleh budi anwari"
x = re.sub("anton", "budi", kalimat)
print(x)
# output : budi@mail.com dimiliki oleh budius, budi@gmail.co.id dimiliki oleh budi anwari
```

BAGIAN 2: LATIHAN MANDIRI (60%)

Pada bagian ini anda menuliskan jawaban dari soal-soal Latihan Mandiri yang ada di modul praktikum. Jawaban anda harus disertai dengan source code, penjelasan dan screenshot output.

SOAL 1

```
14,1.py > ...
1  # Import re dan datetime
2  import re
3  from datetime import datetime
4
5  # Mencari tanggal sekarang
6  x = datetime.now()
7  date = x.strftime("%x")
8
9  # String
10 string = """
11 Pada tanggal 1945-08-17 Indonesia merdeka. Indonesia memiliki beberapa pahlawan nasional,
12 seperti Pangeran Diponegoro (TL: 1785-11-11),
13 Pattimura (TL: 1783-06-08) dan KiHajar Dewantara (1889-05-02).
14 """
15 # Mencari tanggal tanggal di dalam string
16 tanggal = re.findall(r"\d*[-]\d*[-]\d*", string)
17
18 # def mencari selisih
19 def selisih(d1, d2):
20     # strptime untuk mengubah string tanggal menjadi format tanggal buat dihitung
21     d1 = datetime.strptime(d1, "%m/%d/%y")
22     d2 = datetime.strptime(d2, "%Y-%m-%d")
23     return abs((d2 - d1).days)
24
25 # def membalik tanggal
26 def balik(t):
27     tanggal_balik = ""
28     # Mengsplit tahun bulan dan hari menggunakan re split
29     t = re.split("-", t)
30
31     # Menambahkan tanggal kedalam string secara terbalik
32     for i in range(2,-1,-1):
33         tanggal_balik += t[i]
34         tanggal_balik += "-"
35     tanggal_balik = tanggal_balik[:-1]
36     return tanggal_balik
37
38 # Print setiap output
39 for i in range(len(tanggal)):
40     print(f"{balik(tanggal[i])} 00:00:00 berselisih {selisih(date, tanggal[i])} hari")
41
```

```
PS D:\Ngampus\Semester 2\PrakAlPro> & D:/Python/python.exe "d:/Ngampus/Semester 2/PrakAlPro/14,1.py"
17-08-1945 00:00:00 berselisih 28779 hari
11-11-1785 00:00:00 berselisih 87131 hari
08-06-1783 00:00:00 berselisih 88018 hari
02-05-1889 00:00:00 berselisih 49339 hari
```

SOAL 2

email.txt

```
1 anton@mail.com dimiliki oleh antonius
2 budi@gmail.co.id dimiliki oleh budi anwari
3 slamet@getnada.com dimiliki oleh slamet slumut
4 matahari@tokopedia.com dimiliki oleh toko matahari
```

14,2.py > ...

```
1 # import random, string dan re
2 import re
3 import random
4 import string
5
6 # Mengambil semua karakter dan angka
7 all_characters = string.ascii_letters + string.digits
8
9 # Membuka isi email
10 with open("email.txt", "r") as email:
11     string = email.read()
12
13 # Fungsi findall untuk mencari email dengan format a-zA-Z0-9 diawal, dipisah dengan @a-zA-Z0-9 dan .com atau .id
14 emails = re.findall(r"[a-zA-Z0-9.]+@[a-zA-Z0-9.]+\.[a-zA-Z]{2,}", string)
15
16 # def mencari pass
17 def cari_pass():
18     # Mengjoin huruf dan angka random sepanjang 8 karakter
19     password = ''.join(random.choices(all_characters, k=8))
20     return password
21
22 # def mencari username dengan split @ dan mengambil kata pertama
23 def cari_username(email):
24     username = re.split("@", email)[0]
25     return username
26
27 # print output
28 for i in range(len(emails)):
29     print(f"{emails[i]} username: {cari_username(emails[i])}, password: {cari_pass()} ")
```

```
PS D:\Ngampus\Semester 2\PrakAlPro> & D:/Python/python.exe "d:/Ngampus/Semester 2/PrakAlPro/14,2.py
anton@mail.com username: anton, password: 0U0vjSi4
budi@gmail.co.id username: budi, password: rOEQ1TZV
slamet@getnada.com username: slamet, password: f49lwPhz
matahari@tokopedia.com username: matahari, password: 0L1iEvVX
```

Github : <https://github.com/SamuelN1508/PrakAlPro14.git>