

การทำงาน

โปรแกรมค้นหาสิ่งกีดขวาง

รายละเอียดโปรแกรม

โปรแกรมนี้เกิดจากความขี้เกียจคลิกลิงก์ทีละลิงก์ในหน้าตารางเรียน
ที่มีลิงก์รวมเกือบ 50 ลิงก์ จึงเกิดเป็นไอเดียในการหาตัวช่วยค้นหาลิงก์ทั้งหมด
เพื่อนำไปใช้งานในรูปแบบอื่นที่ต้องการอีกที เช่น ทำเว็บรวมลิงก์

อนึ่ง ลิงก์ที่รวบรวมมา เป็นลิงก์จากหน้าเว็บรวมวิชาเลือกที่ใช้ฟิลเตอร์ดังนี้:

- ปีการศึกษา 2559
- ชั้นปีที่ 1

ภาษาที่ใช้: Python (3.5)

ลักษณะของ URI

URL เต็ม

[http://www.reg.kmitl.ac.th/teachtable_v20/teachtable_show.php?
&faculty_id=01&dept_id=x&curr_id=0&curr2_id=x&year=2559&semester=2&class=21](http://www.reg.kmitl.ac.th/teachtable_v20/teachtable_show.php?&faculty_id=01&dept_id=x&curr_id=0&curr2_id=x&year=2559&semester=2&class=21)

พารามิเตอร์ใน URL ที่มีการเปลี่ยนแปลง

`http://www.reg.kmitl.ac.th/teachtable_v20/teachtable_show.php?
&faculty_id=01&dept_id=x&curr_id=0&curr2_id=x&year=2559&semester=2&class=21`

รหัสคณะ

ปีการศึกษา

ภาคเรียน

รหัสกลุ่มวิชา

แทน URI ด้วย Regular Expression

faculty_id=\d\d&dept_id=x&curr_id=\d&curr2_id=.&year=\d{4}&semester=\d&class=\d{1,}

\d

ตัวเลข

.

ตัวอะไรก็ได้ที่ไม่ใช่ line break

\d{4}

ตัวเลข 4 ตัว

\d{1,}

ตัวเลข 1 ตัวขึ้นไป

อัลกอริทึม

```
from re import findall
```

```
from codecs import open
```

อิมพอร์ตฟังก์ชัน findall จากโมดูล re และอิมพอร์ตฟังก์ชัน open จากโมดูล codecs

```
webpage = open('2559_y1_s2.html', 'r', 'utf-8').read()
```

เปิดและอ่านข้อมูลจากไฟล์ แล้วเก็บไว้ในตัวแปร webpage

```
regex = 'faculty_id=\d\d&dept_id=x&curr_id=\d&curr2_id=.\&year=\d{4}&semester=\d&class=\d{1,}'
```

แพทเทิร์น RegEx สำหรับค้นหา เก็บไว้ในตัวแปร regex

```
results = findall(regex, webpage)
```

เก็บผลลัพธ์ที่ได้ไว้ในตัวแปร results (ได้ออกมาเป็นอาร์เรย์) สำหรับไปใช้งานต่อ

โค้ดเริ่มต้น


```
from re import findall
```

```
webpage = open('2559_y1_s2.html', 'r').read()
```

```
regex = 'faculty_id=\d\d&dept_id=x&curr_id=\d&curr2_id=.\&year=\d{4}&semester=\d&class=\d{1,}'
```

```
results = findall(regex, webpage)
```

```
for i in results:
```

```
    print(i)
```

ตัวอย่าง:

```
print("Found", len(results), "results.")
```

แสดงผลลัพท์เป็นบรรทัดๆ

[illegible]

ผลลัพธ์ในเทอร์มินัล


```
[...shortened...]

raw_results = findall(regex, webpage)

final_results = []

for i in raw_results:
    if i not in final_results:
        final_results.append(i)

create_text_file = open("results.txt", "w")

for i in final_results:
    create_text_file.write(i + '\n')

create_text_file.close()

print("Found", len(final_results), "results.")

print("Results saved to results.txt")
```

ตัวอย่าง:

เอ็กซ์พอร์ตผลลัพธ์เป็นบรรทัดๆ
โดยลิงก์ไม่ซ้ำกัน

```
< > results.txt
1 faculty_id=01&dept_id=x&curr_id=0&curr2_id=x&year=2559&semester=2&class=10
2 faculty_id=01&dept_id=x&curr_id=0&curr2_id=x&year=2559&semester=2&class=9
3 faculty_id=01&dept_id=x&curr_id=0&curr2_id=x&year=2559&semester=2&class=21
4 faculty_id=03&dept_id=x&curr_id=0&curr2_id=x&year=2559&semester=2&class=9
5 faculty_id=03&dept_id=x&curr_id=0&curr2_id=x&year=2559&semester=2&class=10
6 faculty_id=04&dept_id=x&curr_id=0&curr2_id=x&year=2559&semester=2&class=9
7 faculty_id=04&dept_id=x&curr_id=0&curr2_id=x&year=2559&semester=2&class=10
8 faculty_id=05&dept_id=x&curr_id=0&curr2_id=x&year=2559&semester=2&class=21
9 faculty_id=05&dept_id=x&curr_id=0&curr2_id=x&year=2559&semester=2&class=9
10 faculty_id=05&dept_id=x&curr_id=0&curr2_id=x&year=2559&semester=2&class=10
11 faculty_id=06&dept_id=x&curr_id=0&curr2_id=x&year=2559&semester=2&class=10
12 faculty_id=06&dept_id=x&curr_id=0&curr2_id=x&year=2559&semester=2&class=9
13 faculty_id=11&dept_id=x&curr_id=0&curr2_id=x&year=2559&semester=2&class=10
14 faculty_id=15&dept_id=x&curr_id=0&curr2_id=x&year=2559&semester=2&class=9
15 faculty_id=15&dept_id=x&curr_id=0&curr2_id=x&year=2559&semester=2&class=10
16
17
```

ผลลัพธ์ในไฟล์เท็กซ์

พัฒนาโดยนักศึกษาชั้นปีที่หนึ่ง

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง