

String and File Processing

ภาควิชาวิศวกรรมคอมพิวเตอร์

จุฬาลงกรณ์มหาวิทยาลัย

๒๕๖๒

ทบทวนเรื่องสตริง

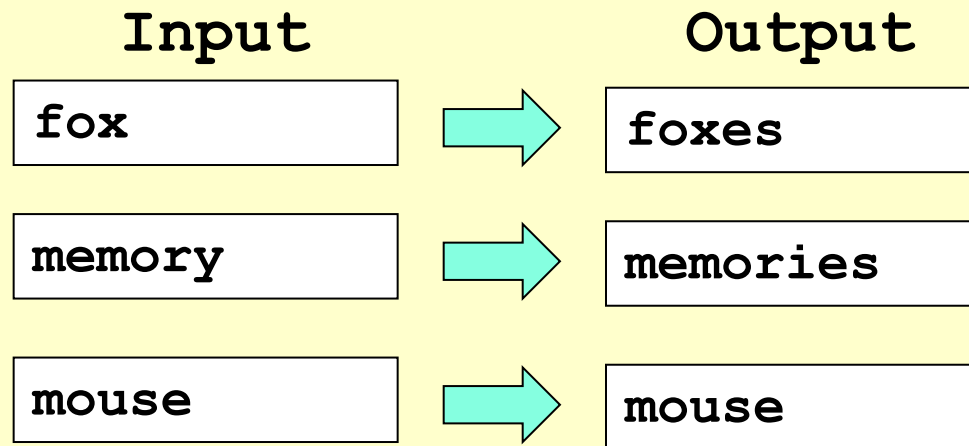
```
s = "I'm a string"
t = 'I said "This is a string".'
chars = input()
c = 0
for ch in chars:
    if ch in s:
        c += 1
for i in range(len(t)):
    if t[i] in s:
        c += 1

r = ""
for k in range(2,10,2):
    r += str(k)      # 2468
r = 2*r             # 24682468
```

แบบฝึกหัด: เอกพจน์ → พหูพจน์

แบบง่าย ๆ

- ลงท้าย s, x, หรือ ch → เติม es ต่อท้าย
- ลงท้าย y แต่ตัวก่อน y ไม่ใช่สระ → เปลี่ยน y เป็น ies
- ถ้าไม่ตรงกับกฎสองข้อข้างบนนี้ → เติม s ต่อท้าย



Espresso, Macchiato, Ristretto, Americano, Latte, Cappuccino, Mocha, Affogato

```
x = input().split(", ")
out = []
k = 0
for e in x:
    out.append((s+" "*12)[:12])
    k += 1
    if k % 3 == 0:
        print("".join(out))
        out = []
print("".join(out))
```

Affogato

Americano

Cappuccino

Espresso

Latte

Macchiato

Mocha

Ristretto



```
import urllib.request

def find(s, start, c):
    for i in range(start, len(s)):
        if s[i] == c: return i
    return -1

url = "http://air4thai.pcd.go.th/services/getNewAQI_XML.php?stationID=52t"
web = urllib.request.urlopen(url)
for line in web:
    line = line.decode()
    if "<PM25 value=" in line:
        i = find(line, 0, '"')
        j = find(line, i+1, '"')
        print("PM 2.5 =", line[i:j])
        break
```

อักขระพิเศษ (Escape Characters)

```
s = "Hello"
print(s)                # Hello
s = "\"Hello\""
print(s)                # "Hello"
s = "'Hello'"
print(s)                # 'Hello'
s = "\"'\Hello\""
print(s)                # "'\Hello"
s = "Hello\nPython"
print(s)                # Hello
                        # Python
```

`\"` คือ `"`
`\'` คือ `'`
`\\` คือ `\`

`\n` คือ รหัสให้ขึ้นบรรทัดใหม่

ตัวอย่าง: แทน " ' / \ () , . : ; ด้วยช่องว่าง

```
def blank(t):  
    result = ""  
    for c in t:  
        if c in "\"' / \\ , . : ;":  
            result += "  
        else:  
            result += c  
    return result
```

บริการของสตริง : String Methods

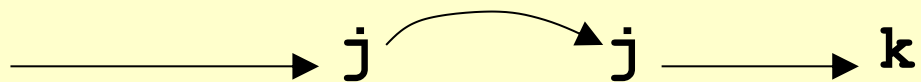
01234567890123

s = " **Hello World** "

- **len(s)** ได้ 14, **len("")** ได้ 0
- **s.lower()** ได้ " **hello world** "
- **s.upper()** ได้ " **HELLO WORLD** "
- **s.strip()** ได้ "**Hello World**"
- **s.find("o")** ได้ 6 **s.find("ex")** ได้ -1
- **s.find("o",7)** ได้ 9

ตัวอย่าง: การใช้งาน

```
t = input().strip()      # ป้องกันกรณีผู้ใช้เผลอกด space
if t.upper() != "YES":  # ตรวจสอบได้ทั้ง Yes, yes, ...
    j = t.find("mailto:")
    if j >= 0:
        j += len("mailto:")
        k = t.find("@", j)
        if k >= 0:
            username = t[j:k]
            ...
```



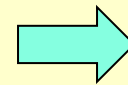
```
<a href="mailto:somchai@gmail.com">...
```

แบบฝึกหัด: camelCase

camelCase เป็นรูปแบบการเขียนวลีที่ไม่มีเครื่องหมายวรรคตอน โดยนำคำต่าง ๆ ในวลีมาเขียนติดกันหมด แต่เขียนให้แต่ละคำขึ้นต้นด้วยตัวใหญ่ ยกเว้นเฉพาะคำแรกสุดเป็นตัวเล็กหมด (ตัวเลขคงไว้เหมือนเดิม)

Input

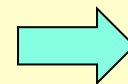
An example of "camel case".



Output

anExampleOfCamelCase

Emergency call 911



emergencyCall911



```
import urllib.request

url = "http://air4thai.pcd.go.th/services/getNewAQI_XML.php?stationID=52t"
web = urllib.request.urlopen(url)
for line in web:
    line = line.decode()
    pattern = '<PM25 value="'
    i = line.find(pattern)
    if i >= 0:
        i += len(pattern)
        j = line.find('"', i)
        print("PM 2.5 =", line[i:j])
        break
```

```
def split(t, s):  
    x = []  
    k0 = 0  
    k = t.find(s, 0)  
    while k >= 0:  
        x.append(t[k0:k])  
        k0 = k+1  
        k = t.find(s, k0)  
    x.append(t[k0:])  
    return x
```

Method Chaining vs Function Composition

```
x = math.radians(d)
s = math.sin(x)
y = abs(s)
r = round(y, 2)
```

```
r = round(abs(math.sin(math.radians(d))), 2)
```

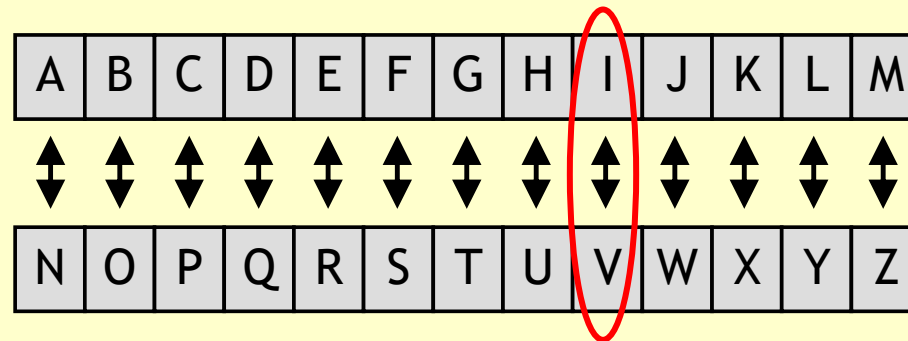
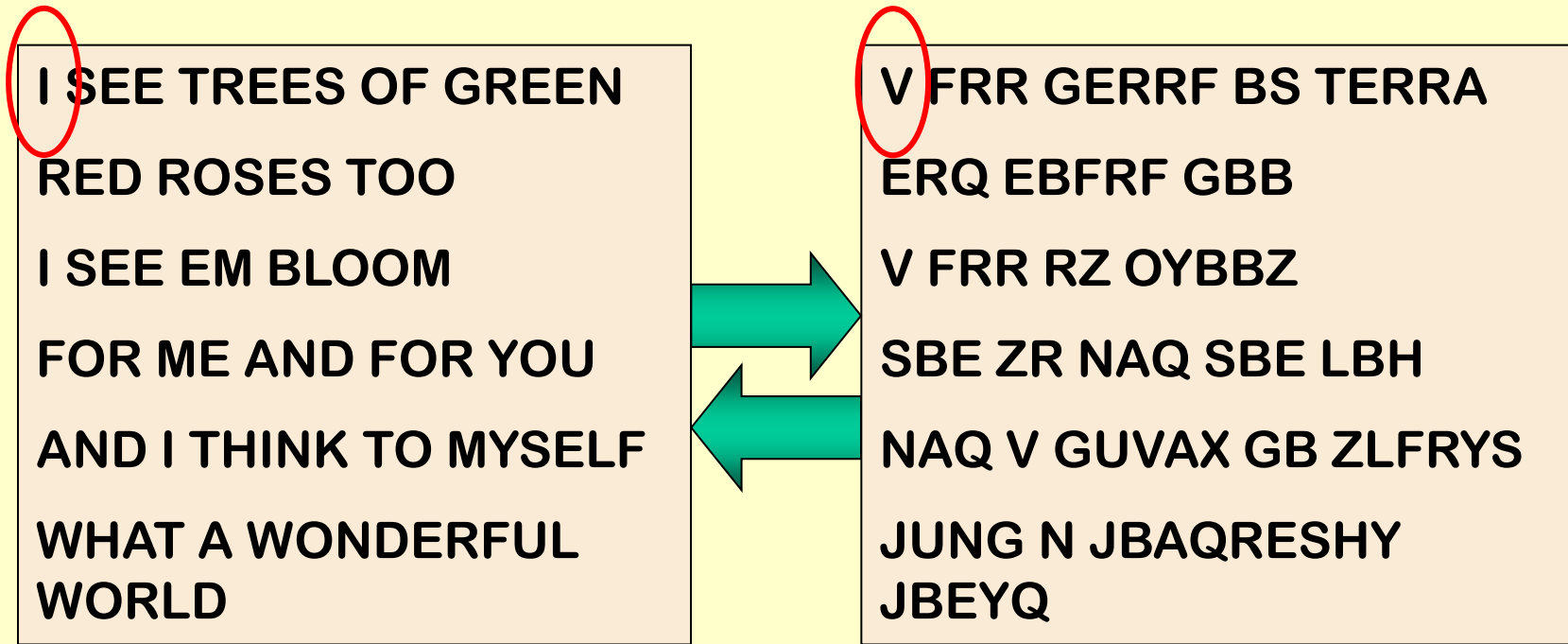
```
line1 = input()
line2 = line1.strip()
line3 = line2.upper()
i      = line3.find("OK")
```

```
i = input().strip().upper().find("OK")
```

ข้อควรระวัง


- ห้ามเปลี่ยนค่าภายในสตริง
 - `s[2] = "a"`
 - `s[3:7] = "-^o^-"`
- string methods ไม่เปลี่ยนตัวสตริง
 - `s = "Hello"`
 - `s.lower()` ได้ 'hello' แต่ s เก็บค่าเดิม 'Hello'
- แต่เปลี่ยนค่าซึ่งเก็บในตัวแปรได้
 - `s = "Hello"`
 - `s = s.lower()` แบบนี้ s เก็บค่าใหม่ "hello"

ตัวอย่าง : rot-13



ตัวอย่าง : rot-13

```
def rot13(txt_in):  
    upper = "ABCDEFGHIJKLMNOPQRSTUVWXYZ"  
    alphabets = 2*upper  
    txt_out = ""  
    for ch in line:  
        j = alphabets.find(ch)  
        if j != -1:  
            txt_out += alphabets[j + 13]  
        else:  
            txt_out += ch  
    return txt_out  
line = input()  
while line.strip().upper() != "END":  
    print(rot13(line))  
    line = input()
```

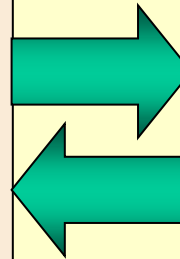


ABCDEFGHIJKLMNOPQRSTUVWXYZABCDEFGHIJKLMNOPQRSTUVWXYZ

แบบฝึกหัด: rot-13

ปรับโปรแกรมให้ เข้า/ถอดรหัสได้ทั้งตัวอักษรเล็กและใหญ่

I see trees of green
red roses too
I see em bloom
for me and for you
And I think to myself
what a wonderful world
end



V frr gerrf bs terra
erq ebfrf gbb
V frr rz oybbz
sbe zr naq sbe lbh
Naq V guvax gb zlfrys
jung n jbaqreshy jbeyq

การอ่านข้อมูลจากแฟ้มข้อความ

```
fin = open( filename, "r" )  
  
line = fin.readline()  
...  
for line in fin:  
    ...  
  
fin.close()
```

"r" บอกว่าเปิด
แฟ้มเพื่อ **อ่าน**

readline อ่านหนึ่งบรรทัด
ถัดไปเข้ามาเป็นสตริง

for แบบนี้ อ่านทีละบรรทัด
ถัดไปเป็นสตริงใส่ตัวแปร line
รอบละหนึ่งบรรทัด จนหมดแฟ้ม

เลิกอ่านแฟ้ม
แล้วก็ปิดแฟ้ม

ตัวอย่าง: หาคะแนนเฉลี่ยนิสิต 3 คนแรกในแฟ้ม

```
sum_points = 0; n = 3

infile = open("data.txt", "r")
for k in range(n):
    line = infile.readline() # อ่านบรรทัดถัดไป
    x = line.split()
    sum_points += float(x[1])
infile.close()

print("Average =", sum_points/n)
```

	x[0]	x[1]
data.txt	6230012121	90
	6230351221	80
	6231027921	79
	6230548121	70

Average = 83.0

ตัวอย่าง: อ่านแฟ้มคะแนนมาแสดงเรียงจากมากมาน้อย

```
students = []

infile = open("data.txt", "r")
for line in infile: # อ่านรอบละบรรทัดจนหมดแฟ้ม
    sid,point = line.strip().split()
    point = float(point)
    students.append([point, sid])
infile.close()

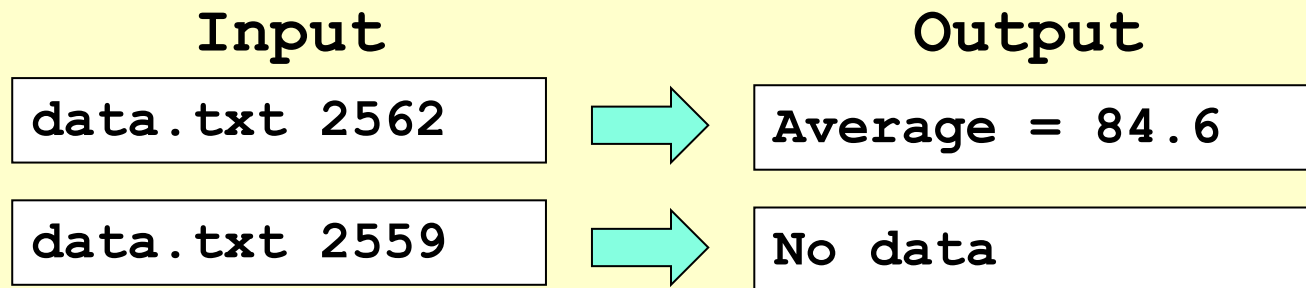
students.sort() # เรียงจากน้อยไปมากตามคะแนน

for [point,sid] in students[::-1]:
    print(sid, point)
```

data.txt

6230012121	90
6230351221	80
6231027921	79
6230548121	70

แบบฝึกหัด: หาคะแนนเฉลี่ยของนิสิตบางกลุ่มในแฟ้ม



data.txt

6230012121	90
6130351221	80
6231027921	79
5830548121	65
6031087221	70
6230550321	72
6230432721	87
6230215221	95
6130518321	72

การบันทึกข้อมูลลงแฟ้มข้อความ

```
fout = open( filename, "w" )
```

"w" บอกว่าเปิด
แฟ้มเพื่อ **เขียน**

```
fout.write("First line")
```

write บรรทัดสตริง
ต่อท้ายในแฟ้ม

```
fout.write("Text\n")
```

ต้องการขึ้นบรรทัด
ใหม่ต้องใส่ \n เอง

```
fout.close()
```

บันทึกเสร็จแล้ว
ก็ปิดแฟ้ม

ตัวอย่าง: บันทึก Triangular Number ร้อยตัวลงแฟ้ม

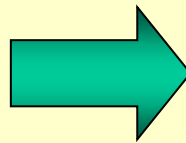
```
fout = open("tri_numbers.txt", "w")
for k in range(1,101):
    n = k*(k+1)//2    # triangular number
    fout.write(str(n)+" ")
    if k%10 == 0:    # ครบ 10 ตัว ขึ้นบรรทัดใหม่
        fout.write("\n")
fout.close()
```

```
1 3 6 10 15 21 28 36 45 55
66 78 91 105 120 136 153 171 190 210
231 253 276 300 325 351 378 406 435 465
496 528 561 595 630 666 703 741 780 820
861 903 946 990 1035 1081 1128 1176 1225 1275
1326 1378 1431 1485 1540 1596 1653 1711 1770 1830
1891 1953 2016 2080 2145 2211 2278 2346 2415 2485
2556 2628 2701 2775 2850 2926 3003 3081 3160 3240
3321 3403 3486 3570 3655 3741 3828 3916 4005 4095
4186 4278 4371 4465 4560 4656 4753 4851 4950 5050
```


แบบฝึกหัด: สร้างแฟ้มที่เป็น rot-13 ของอีกแฟ้ม

I see trees of green
red roses too
I see em bloom
for me and for you
And I think to myself
what a wonderful world

data.txt



V frr gerrf bs terra
erq ebfrf gbb
V frr rz oybbz
sbe zr naq sbe lbh
Naq V guvax gb zlfrys
jung n jbaqreshy jbeyq

rot13.txt