

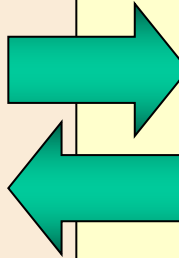
String Processing

ภาควิชาวิศวกรรมคอมพิวเตอร์
จุฬาลงกรณ์มหาวิทยาลัย

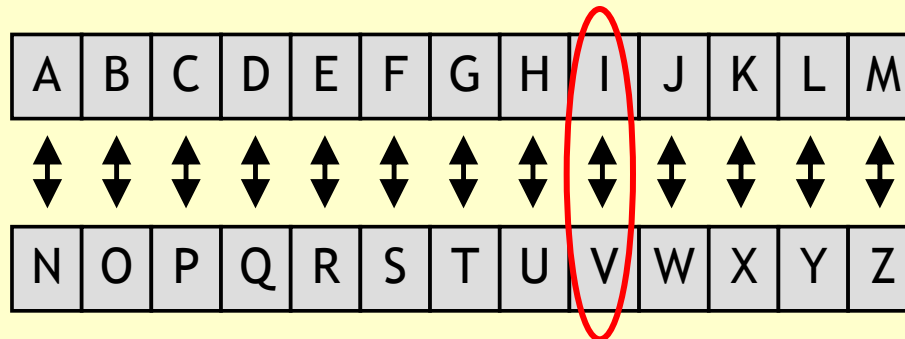
๒๕๕๙

Rot-13 Encoding

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



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



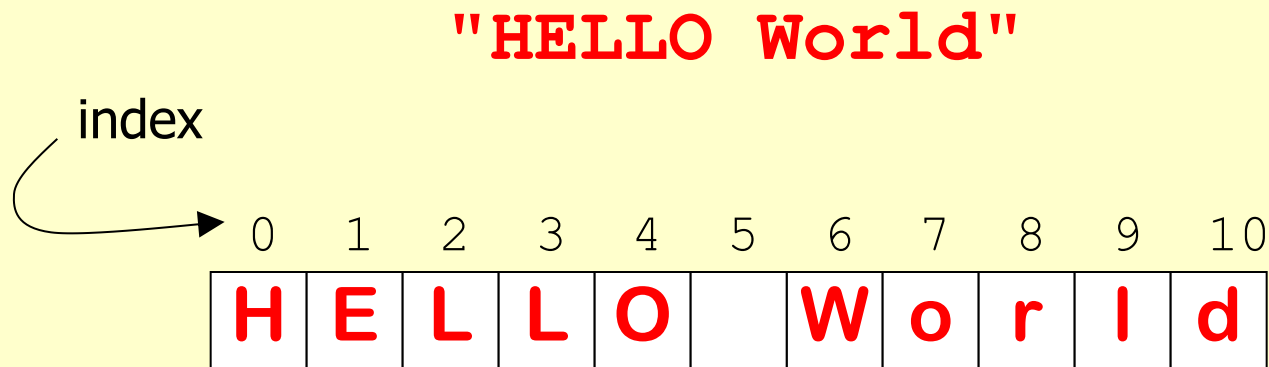
Topics

- บริการพื้นฐานของสตริง
 - len, lower, upper, strip, find, [], in
- Method chaining
 - input().strip().upper()
- การอ่านสตริงจากแฟ้มข้อความ
 - open, readline, for line in file, close
- รหัสพิเศษ escape characters
 - \n \\ \t
- การเขียนสตริงลงในแฟ้มข้อความ
 - open(..., "w"), write, close
- การอ่านสตริงจากเว็บ
 - urlopen, decode

ทบทวนเรื่องสตริง

- ได้
 - `"I'm a string."`
 - `'I said "This is a string".'`
 - `"""I said "I'm a string"."""`
 - `"abc" + "def"`
 - `2*("ab" * 3 + "x")`
- ไม่ได้
 - `"a" + 2`
 - `"a" * "b"`
- การแปลงประเภทข้อมูล
 - `str(2)` ได้ `'2'`, `str(2.3)` ได้ `'2.3'`,
 - `int("2")` ได้ `2`, `float("2.3")` ได้ `2.3`
- การอ่านจากแป้นพิมพ์ได้สตริงเสมอ
 - `line = input(">>")`

String = Sequence of characters



สตริงขนาด N ตัวอักษร มี index ตั้งแต่ 0 ถึง $N - 1$

```
>>> type("what am i")  
<class 'str'>
```

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
- `s[2]` ได้ "H", `s[-2]` ได้ "d"
- `s[2:7]` ได้ "Hello"
- `s[8:]` ได้ "World " `s[:7]` ได้ " Hello"
- `"lo" in s` ได้ True `"la" in s` ได้ False
- `if (c in "aeiou"):`
`print(c, "is a vowel")`

ตำแหน่งของตัวอักษรในสตริง

- ต้องเป็น int หรือเป็นช่วง int : int s[3] s[2:6]
- เป็นตัวแปรที่เก็บข้อมูลแบบ int s[k]
- เป็น expression ที่ให้ผลเป็น int s[k+j-3]

```
>>> s = "ABCDEFGHI"  
>>> i = 3  
>>> j = 5  
>>> print(s[i])
```

```
.....  
>>> print(s[5//3])
```

```
.....  
>>> print(s[i:j+1])
```

```
.....  
>>> print(s[6/3])  
.....
```

ข้อควรระวัง

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

ลองเขียนดู : email address อยู่ไหน ?

```
">"<a href="mailto:somchai@men.or.th">Send mail</a>
```

```
to <a href = "mailto:me@earth.com">me</a>
```

- หาตำแหน่งของ "mailto:"
- หาตำแหน่งของ ">"

```
txt = input(">> ").lower()
start = txt.find(_____)
stop  = txt.find(_____)
print(txt_____)
```

Method Chainings

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

- **input** ได้สตริง
- **strip** ได้สตริง
- **upper** ได้สตริง
- สามารถ ใช้ **strip** ต่อจาก **input**
ใช้ **upper** ต่อจาก **strip** ได้
และ ใช้ **find** ต่อจาก **upper** ได้

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

ลองเขียนดู : แปลงชื่อย่อเดือนเป็นเลขเดือน

```
> Oct
```

```
Oct --> 10
```

```
> jan
```

```
jan --> 1
```

```
month = _____.strip() # รับชื่อย่อจากผู้ใช้
#          012345678901234567890123456789012345
abbr = "JANFEBMARAPRMAYJUNJULAUGSEPNOVDEC"
k = abbr.find( month._____ )
if _____ :
    print("Invalid month abbreviation")
else :
    print(month, "-->", _____ )
```

ใช้ < <= == != >= > เปรียบเทียบสตริงได้

• "abc" == "ab" + "c"

• "a" != "A"

• "a" < "b"

• "abbbb" < "b"

• "A" < "a"

• "0" < "1"

• "z" < "ก"

• "ก" < "ฮ"

• "สระไก่" < "เขย"

• "12" < 23

	0001	0002	0003	0004	0005	0006	0007	0008	0009	000A	000B	000C	000D	000E	000F
0010	0011	0012	0013	0014	0015	0016	0017	0018	0019	001A	001B	001C	001D	001E	001F
0020	!	"	#	\$	%	&	'	()	*	+	,	-	.	/
0030	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>
0040	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N
0050	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^
0060	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n
0070	p	q	r	s	t	u	v	w	x	y	z	{		}	~
0E48	'	ก	ข	ฃ	ค	ฅ	ฆ	ง	จ	ฉ	ช	ฌ	ฎ	ฏ	ฐ
0E10	ฐ	ท	ฒ	ณ	ด	ต	ถ	ท	ธ	น	บ	ป	ผ	ฝ	พ
0E20	ภ	ม	ย	ร	ฤ	ล	ฌ	ว	ศ	ษ	ส	ห	ฬ	อ	ฮ
0E30	ะ	ั	า	ำ	ิ	ี	ึ	ุ	เ	อ	โ	ๅ	ๆ	็	่
0E40	เ	แ	โ	ใ	เ	า	ๅ	ๆ	็	่	้	๐	๑	๒	๓
0E50	๔	๕	๖	๗	๘	๙	๐	๑	๒	๓	๔	๕	๖	๗	๘

ไม่ต้องจำ

รูปแบบการประมวลผลแต่ละตัวอักษรในสตริง

```
k = 0
while k < len(s) :
    c = s[k]
    # นำ c ไปใช้

    k += 1
```

```
nvowels = 0
k = 0
while k < len(s) :
    c = s[k]
    if c in "aeiou" :
        nvowels += 1
    k += 1
```

`for c in s:` อ่านว่า สำหรับแต่ละตัวอักษร `c` ในสตริง `s`
`if c in s:` อ่านว่า ถ้ามีสตริง `c` ปรากฏอยู่ในสตริง `s`

```
for c in s :
    # นำ c ไปใช้
```

```
nvowels = 0
for c in s :
    if c in "aeiou" :
        nvowels += 1
```

แบบนี้ง่ายกว่า

ลองเขียนดู : เลขที่รับมามีเลขโดดค่าเพิ่มขึ้นเรื่อย ๆ

Enter digits : 024589

Yes, this is an increasing-digit number

Enter digits : 024189

No, this is not an increasing-digit number

```
n = input("Enter digits : ")
is_increasing = _____
prev_d = _____ # ตั้งแต่เริ่มต้นด้วยตัวอักษรที่ "น้อยๆ"
for d in _____ : # สำหรับแต่ละตัวอักษรใน n
    if _____ : # ถ้าตัวปัจจุบันน้อยกว่าตัวก่อนหน้า
        is_increasing = _____
        break
    prev_d = d # ตัวปัจจุบันจะเป็นตัวก่อนหน้าในรอบหน้า

if is_increasing :
    print("Yes, this is an increasing-digit number")
else :
    print("No, this is not an increasing-digit number")
```

ลองเขียนดู : แปลงเลขอารบิกเป็นเลขไทย

Enter digits : 24,589.00

๒๔,๕๘๙.๐๐

```
in_txt = input("Enter digits : ")
out_txt = ""
arabic_number = "0123456789"
thai_number    = "๐๑๒๓๔๕๖๗๘๙"

for _____ :

    k = _____

    if k < 0 :

        out_txt += _____

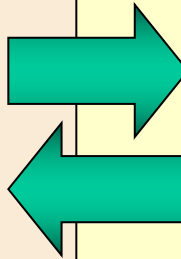
    else :

        out_txt += _____

print(out_txt)
```


ตัวอย่าง : rot-13

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



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

A	B	C	D	E	F	G	H	I	J	K	L	M
↕	↕	↕	↕	↕	↕	↕	↕	↕	↕	↕	↕	↕
N	O	P	Q	R	S	T	U	V	W	X	Y	Z

ตัวอย่าง : rot-13

```
line = input("Enter any text : ")
upper = "ABCDEFGHIJKLMNOPQRSTUVWXYZ"
lower = upper.lower() # lower สองที่ในบรรทัดนี้มีความหมายต่างกัน
upper2 = upper*2      # upper2 = upper + upper[:13] ก็ได้
lower2 = lower*2      # lower2 = lower + lower[:13] ก็ได้
rot13 = ""
for c in line:
    if c in upper:
        rot13 += upper2[upper.find(c)+13]
    elif c in lower:
        rot13 += lower2[lower.find(c)+13]
    else:
        rot13 += c
print("rot13 =", rot13)
```

1

2

3

4

5

01234567890123456789012345678901234567890123456789012345678901

"ABCDEFGHIJKLMNOPQRSTUVWXYZABCDEFGHIJKLMNOPQRSTUVWXYZ"

"ABCDEFGHIJKLMNOPQRSTUVWXYZABCDEFGHIJKLM"

ตัวอย่าง : rot-13

```
line = input("Enter any text : ")
upper = "ABCDEFGHIJKLMNOPQRSTUVWXYZ"
lower = upper.lower()
rot13 = ""
for c in line:
    if c in upper:
        k = upper.find(c)+13
        if k >= 26:
            k -= 26
        rot13 += upper[k]
    elif c in lower:
        k = lower.find(c)+13
        if k >= 26:
            k -= 26
        rot13 += lower[k]
    else:
        rot13 += c
print("rot13 =", rot13)
```

1	2
01234567890123456789012345	01234567890123456789012345
"ABCDEFGHIJKLMNOPQRSTUVWXYZ"	

ตัวอย่าง : rot-13

```
line = input("Enter any text : ")
upper = "ABCDEFGHIJKLMNOPQRSTUVWXYZ"
lower = upper.lower()
rot13 = ""
for c in line:
    if c in upper:
        rot13 += upper[(upper.find(c)+13) % 26]
    elif c in lower:
        rot13 += lower[(lower.find(c)+13) % 26]
    else:
        rot13 = rot13 + c
print("rot13 =", rot13)
```

```
k = upper.find(c)+13
if k >= 26:
    k -= 26
```

1 2

01234567890123456789012345

"ABCDEFGHIJKLMNOPQRSTUVWXYZ"

การอ่านสตริงจากเว็บ

The screenshot shows a web browser window with the address bar displaying `www.eng.chula.ac.th`. The page title is "ChulaEngineering". Below the title, there is a section titled "ข่าวสารล่าสุด" (Latest News). A right-click context menu is open over the page content, showing options such as "Back", "Forward", "Reload", "Save as...", "Print...", "Translate to English", "View page source" (highlighted), "View page info", and "Inspect element". A speech bubble points to the context menu with the text "click mouse ปุ่มขวา".

ChulaEngineering

ข่าวสารล่าสุด

"CLUE" เผยข่าวรายเดือนคณะวิศวกรรมศาสตร์ จุฬาฯ ฉบับที่ 12 (22-Jul-2015)

กรรมการคอมพิวเตอร์ ได้รับรางวัลจากการแข่งขันพัฒนาโปรแกรม

2015)

คณะวิศวกรรมฯ ให้การต้อนรับ

คณะวิศวกรรมฯ ให้การต้อนรับ

26 (16-Jul-2015)

คณะวิศวกรรมฯ ให้การต้อนรับ

คณะวิศวกรรมฯ ให้การต้อนรับ

คณะวิศวกรรมฯ ให้การต้อนรับ

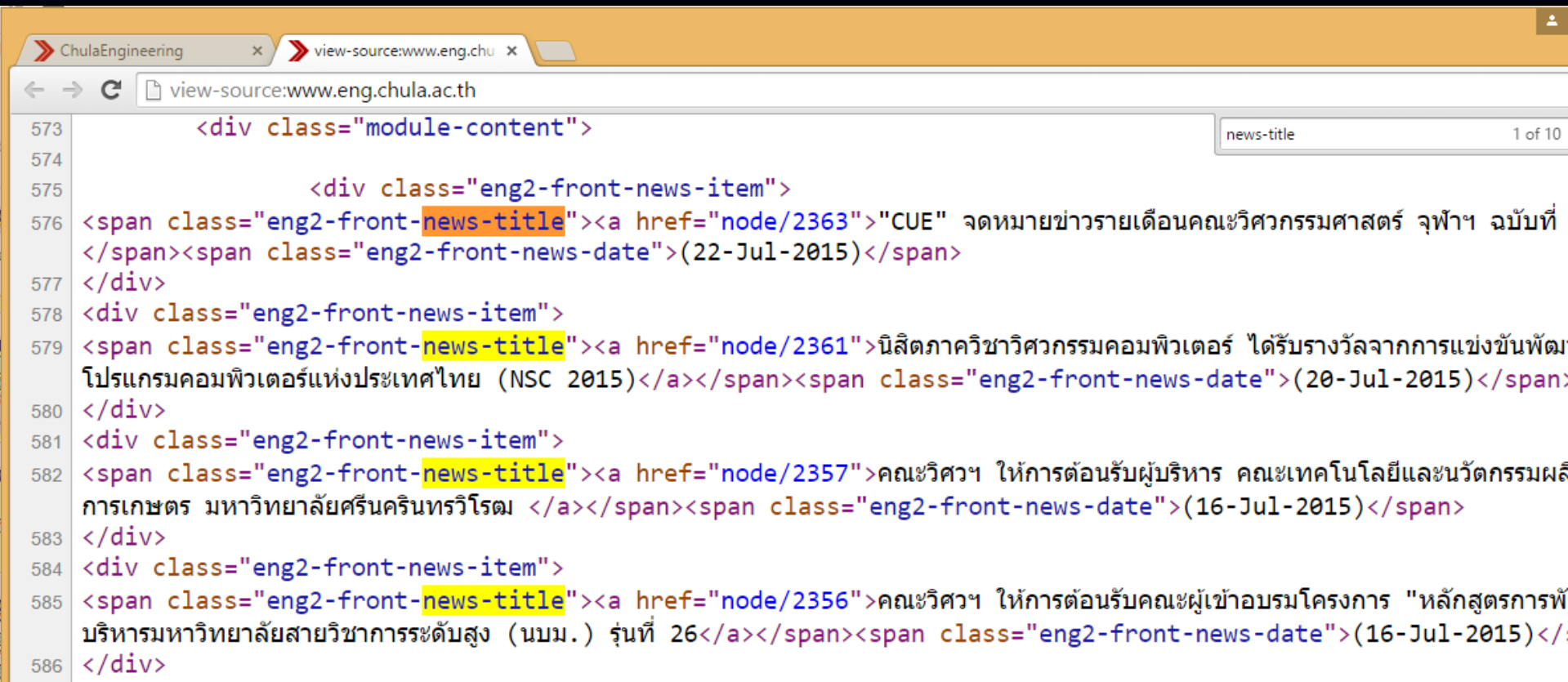
คณะวิศวกรรมฯ จัดกิจกรรม

คณะวิศวกรรมศาสตร์จะไม่ดำเนินการจัดสอบเข้าชั้นสำหรับรายวิชาของคณะวิศวกรรม

2558 เป็นต้นไป (15-Jul-2015)

click mouse ปุ่มขวา

View Page Source



```
573 <div class="module-content">
574
575     <div class="eng2-front-news-item">
576 <span class="eng2-front-news-title"><a href="node/2363">"CUE" จดหมายข่าวรายเดือนคณะวิศวกรรมศาสตร์ จุฬาฯ ฉบับที่
577 </span><span class="eng2-front-news-date">(22-Jul-2015)</span>
578 </div>
579 <div class="eng2-front-news-item">
580 <span class="eng2-front-news-title"><a href="node/2361">นิสิตภาควิชาวิศวกรรมคอมพิวเตอร์ ได้รับรางวัลจากการแข่งขันพัฒน
581 โปรแกรมคอมพิวเตอร์แห่งประเทศไทย (NSC 2015)</a></span><span class="eng2-front-news-date">(20-Jul-2015)</span>
582 </div>
583 <div class="eng2-front-news-item">
584 <span class="eng2-front-news-title"><a href="node/2357">คณะวิศวกรรมฯ ให้การต้อนรับผู้บริหาร คณะเทคโนโลยีและนวัตกรรมผลิ
585 การเกษตร มหาวิทยาลัยศรีนครินทรวิโรฒ </a></span><span class="eng2-front-news-date">(16-Jul-2015)</span>
586 </div>
587 <div class="eng2-front-news-item">
588 <span class="eng2-front-news-title"><a href="node/2356">คณะวิศวกรรมฯ ให้การต้อนรับคณะผู้เข้าอบรมโครงการ "หลักสูตรการพั
589 บริหารมหาวิทยาลัยสายวิชาการระดับสูง (นบม.) รุ่นที่ 26</a></span><span class="eng2-front-news-date">(16-Jul-2015)</span>
590 </div>
```

การอ่านสตริงจากเว็บ

```
import urllib.request

url = "http://www.eng.chula.ac.th"
web = urllib.request.urlopen(url)
for line in web:
    line = line.decode()
    print(line)
```

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0
Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-
transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" >
<head>
...
<div class="eng2-front-news-item">

<span class="eng2-front-news-title"><a href="node/2363">"CUE"
จดหมายข่าวรายเดือนคณะวิศวกรรมศาสตร์ จุฬาฯ ฉบับที่ 12</a></span><span
class="eng2-front-news-date">(22-Jul-2015)</span>

</div>
```

ตัวอย่าง : การสกัดหัวข้อข่าวเว็บคณะวิศวกรรมศาสตร์

ChulaEngineering x view-source:www.eng.chu x

view-source:www.eng.chula.ac.th

news-title 1 of 10

```
<div class="module-content">
  <div class="eng2-front-news-item">
    <span class="eng2-front-news-title"><a href="node/2363">"CUE" จดหมายข่าวรายเดือนคณะวิศวกรรมศาสตร์ จุฬาฯ ฉบับที่
    </span><span class="eng2-front-news-date">(22-Jul-2015)</span>
  </div>
  <div class="eng2-front-news-item">
    <span class="eng2-front-news-title"><a href="node/2361">นิสิตภาควิชาวิศวกรรมคอมพิวเตอร์ ได้รับรางวัลจากการแข่งขันพัฒนา
    โปรแกรมคอมพิวเตอร์แห่งประเทศไทย (NSC 2015)</a></span><span class="eng2-front-news-date">(20-Jul-2015)</span>
  </div>
```

ข่าวสารล่าสุด

"CUE" จดหมายข่าวรายเดือนคณะวิศวกรรมศาสตร์ จุฬาฯ ฉบับที่ 12 (22-Jul-2015)

นิสิตภาควิชาวิศวกรรมคอมพิวเตอร์ ได้รับรางวัลจากการแข่งขันพัฒนาโปรแกรมคอมพิวเตอร์แห่งประเทศไทย (NSC 2015)

คณะวิศวกรรมศาสตร์ ให้การต้อนรับผู้บริหาร คณะเทคโนโลยีและนวัตกรรมผลิตภัณฑ์การ

คณะวิศวกรรมศาสตร์ ให้การต้อนรับคณะผู้เข้าอบรมโครงการ "หลักสูตรการพัฒนานักบริหาร

26 (16-Jul-2015)

คณะวิศวกรรมศาสตร์ ให้การต้อนรับผู้แทนจาก Queensland University of Technology

คณะวิศวกรรมศาสตร์ ให้การต้อนรับผู้แทนจาก Opole University of Technology ประเทศ

คณะวิศวกรรมศาสตร์ ให้การต้อนรับอาจารย์จาก Massachusetts Institute of Technolog

คณะวิศวกรรมศาสตร์ จัดกิจกรรมต้อนรับน้องใหม่ "ค่ายวิศวกรรมมนูตร์ ครั้งที่ 13" (16-Jul-20

คณะวิศวกรรมศาสตร์จะไม่ดำเนินการจัดสอบเข้าชั้นสำหรับรายวิชาของคณะวิศวกรรม

2558 เป็นต้นไป (15-Jul-2015)

ตัวอย่าง : การสกัดหัวข้อข่าวเว็บคณะวิศวกรรมฯ

```
import urllib.request

url = "http://www.eng.chula.ac.th"
web = urllib.request.urlopen(url)
for line in web:
    line = line.decode()
    if "news-title" in line:
        k = line.find("<a href")
        k = line.find(">", k)
        j = line.find("</a>", k)
        print(line[k+1:j])
```

```
...
<span class="eng2-front-news-title"><a href="node/2363">"CUE"
จดหมายข่าวรายเดือนวิศวกรรมศาสตร์ จุฬาฯ ฉบับที่ 12</a></span><span
class="eng2-front-news-date">(22-Jul-2015)</span>
...
```

ลองเขียนดู : สภาพอากาศที่กรุงเทพฯ

<http://weather.yahooapis.com/forecastrss?w=12756350>

```
<rss xmlns:yweather="http://xml.weather.yahoo.com/ns...  
...  
<yweather:location city="Bangkok" region="" country="Thailand"/>  
<yweather:units temperature="F" ... />  
...  
<pubDate>Wed, 19 Aug 2015 8:30 am ICT</pubDate>  
<yweather:condition text="Fair" code="34" temp="88" date="Wed, 19  
Aug 2015 8:30 am ICT"/>  
...  
<yweather:forecast day="Wed" date="19 Aug 2015" low="80"  
high="97" text="PM Thunderstorms" code="38"/>  
<yweather:forecast day="Thu" date="20 Aug 2015" low="79"  
high="94" text="Thunderstorms" code="4"/>  
<yweather:forecast day="Fri" date="21 Aug 2015" low="80"  
high="94" text="PM Thunderstorms" code="38"/>  
<yweather:forecast day="Sat" date="22 Aug 2015" low="81"  
high="97" text="Partly Cloudy" code="30"/>  
<yweather:forecast day="Sun" date="23 Aug 2015" low="80"  
high="94" text="PM Thunderstorms" code="38"/>  
...
```