More Applications

of

the Pumping Lemma

The Pumping Lemma:

เป็นการพิสูจน์แบบ contra

- Given a infinite regular language L
- there exists an integer m
- for any string $w \in L$ with length $|w| \geq m$
- we can write $w = x\ y\ z$
- with $|x\ y| \leq m$ and $|y| \geq 1$
- such that: $x\ y^i\ z \in L \qquad i = 0, 1, 2, \ldots$

Non-regular languages

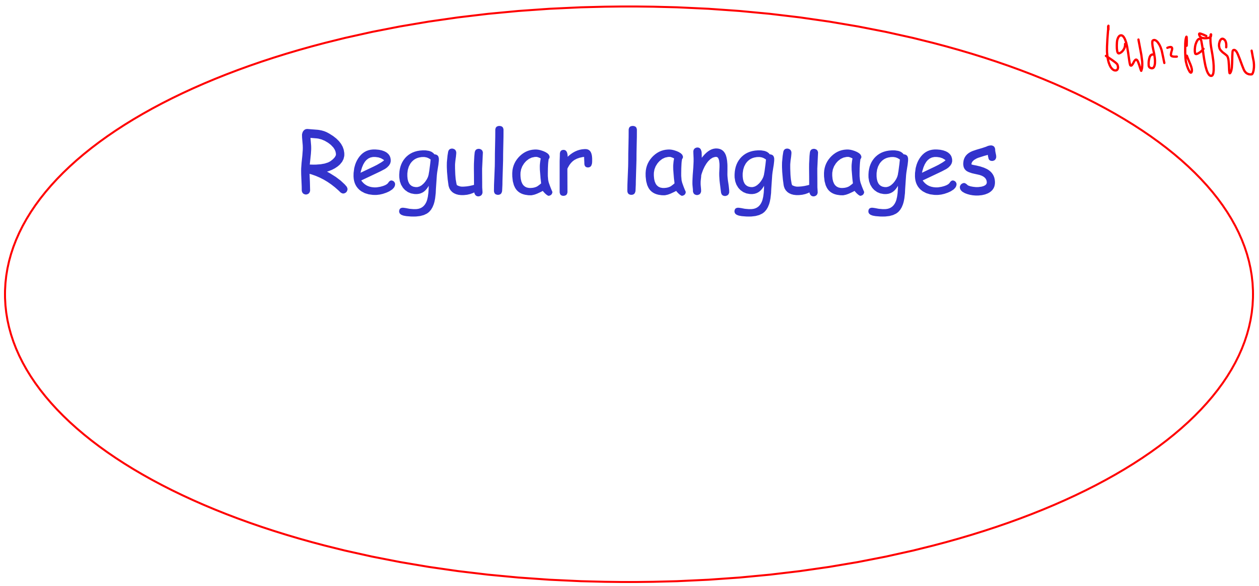
$$L = \{vv^R : v \in \Sigma^*\}$$

ยาก

เพราะเขียน DFA ไม่ได้

มันไม่รู้ต้อง

ทวนตอนไหน



Theorem: The language

$$L = \{vv^R : v \in \Sigma^*\} \qquad \Sigma = \{a, b\}$$

is not regular

Proof: Use the Pumping Lemma

$$L = \{vv^R : v \in \Sigma^*\}$$

Assume for contradiction
that L is a regular language

Since L is infinite
we can apply the Pumping Lemma

$$L = \{vv^R : v \in \Sigma^*\}$$

Let m be the integer in the Pumping Lemma

1

Pick a string w such that: $w \in L$ and

length $|w| \geq m$

พยายามอย่าสร้างสตริงใหม่

We pick $w = a^m b^m b^m a^m$

ถ้าเขียนเป็น

$a^m a^m$

ไม่ผ่านนะ

(2) Write $a^m b^m b^m a^m = x\,y\,z$

เสี่ยงบอก

From the Pumping Lemma

it must be that length $|x\,y| \le m, \quad |y| \ge 1$

$$xyz = \overbrace{\underbrace{a\ldots a}_{x}\underbrace{a\ldots a}_{y}\ldots a}^{m}\,\overbrace{b\ldots b}^{m}\,\overbrace{b\ldots b}^{m}\,\overbrace{a\ldots a}^{m}$$

$$z = \ldots a\,b\ldots b\,b\ldots b\,a\ldots a$$

Thus: $y = a^k, \quad k \ge 1$

$$x\,y\,z = a^m b^m b^m a^m \qquad y = a^k, \quad k \geq 1$$

3

From the Pumping Lemma: $x\,y^i\,z \in L$

$$i = 0, 1, 2, \ldots$$

Thus: $x\,y^2\,z \in L$

$$x\ y\ z = a^m b^m b^m a^m \qquad\qquad y = a^k, \quad k \geq 1$$

From the Pumping Lemma: $x\ y^2\ z \ \in\ L$

$$xy^2z = \overbrace{a...aa...aa...a...a}^{m+k}\overbrace{b...b}^{m}\overbrace{b...b}^{m}\overbrace{a...a}^{m} \in L$$

$$\underbrace{a...a}_{x}\underbrace{a...a}_{y}\underbrace{a...a}_{y}\underbrace{a...ab...bb...ba...a}_{z}$$

$(x-y)(y)^2$

$(x)+(y)^2$

Thus: $a^{m+k} b^m b^m a^m \quad \in L$

$$a^{m+k} b^m b^m a^m \quad \in L \qquad\qquad k \geq 1$$

ไม่อยู่ ∴ L เป็น non-regula language

BUT: $L = \{vv^R : v \in \Sigma^*\}$

⬇

$$a^{m+k} b^m b^m a^m \quad \notin L$$

CONTRADICTION!!!

Therefore: Our assumption that L is a regular language is not true

Conclusion: L is not a regular language

กลับมา

$\overbrace{aa\ldots aa}^{m} \mid \overbrace{aa\ldots aa}^{m}$

$|x$ — $||-y|$

$y = (aa)^k ; k \geq 1$

เลือก $i=0$ $a^{m-2}a^m \Rightarrow a^{m-1}a^{m-1}$

เลือก $i=2$ $a^{m+2}a^m \Rightarrow a^{m+1}a^{m+1} \rightarrow$ ไม่หลุด

Note: Can we use a string $w=a^m a^m$?

Non-regular languages

$$L = \{a^n b^l c^{n+l} :\ n, l \geq 0\}$$

แบบรนเขียนไม่ได้

Regular languages

Theorem: The language

$$L = \{a^n b^l c^{n+l} : \; n,l \geq 0\}$$

is not regular

Proof: Use the Pumping Lemma

$$L = \{a^n b^l c^{n+l} : n, l \geq 0\}$$

Assume for contradiction
that L is a regular language

Since L is infinite
we can apply the Pumping Lemma

$$L = \{a^n b^l c^{n+l} : \ n, l \geq 0\}$$

Let m be the integer in the Pumping Lemma

Pick a string w such that: $w \in L$ and

length $|w| \geq m$

①

We pick $w = a^m b^m c^{2m}$

Write $a^m b^m c^{2m} = x\ y\ z$

2

$y = a^u ; u \geq 1$

From the Pumping Lemma

it must be that length $|x\ y| \leq m, \quad |y| \geq 1$

$$xyz = \overbrace{a...aa...aa...a}^{m}\overbrace{b...b}^{m}\overbrace{c...cc...c}^{2m}$$

$$xyz = \underbrace{a...a}_{x}\underbrace{a...a}_{y}\underbrace{a...ab...bc...cc...c}_{z}$$

Thus: $y = a^k, \quad k \geq 1$

$$x\ y\ z = a^m b^m c^{2m} \qquad\qquad y = a^k, \quad k \geq 1$$

From the Pumping Lemma: $\quad x\ y^i\ z \ \in\ L$

$$i = 0, 1, 2, \ldots$$

ถ้า=0

Thus: $\quad x\ y^0\ z \ = xz \ \in\ L$

$$x\ y\ z = a^m b^m c^{2m} \qquad\qquad y = a^k, \quad k \geq 1$$

From the Pumping Lemma: $\quad xz \in L$

$$xz = \underbrace{\overbrace{a \ldots a a \ldots a}^{m-k}}_{\ldots} \overbrace{b \ldots b}^{m} \overbrace{c \ldots c c \ldots c}^{2m} \in L$$

$$xz = \underbrace{a \ldots a}_{x}\underbrace{a \ldots a b \ldots b c \ldots c c \ldots c}_{z} \in L$$

ตัวปั๊มออกมา

Thus: $\quad a^{m-k} b^m c^{2m} \in L$

$$a^{m-k} b^m c^{2m} \quad \in L \qquad\qquad k \geq 1$$

BUT: $$L = \{a^n b^l c^{n+l} : \; n,l \geq 0\}$$

$$\Downarrow$$

หลุด → l is not res

$$a^{m-k} b^m c^{2m} \quad \notin L$$

CONTRADICTION!!!

Therefore: Our assumption that L is a regular language is not true

Conclusion: L is not a regular language

Non-regular languages $L = \{a^{\cancel{n!}}: \ n \geq 0\}$

$1! \quad 2! \quad 3!$

Regular languages

Theorem: The language $L = \{a^{n!} : \ n \geq 0\}$

is not regular

นิยามของ! $n! \ = \ 1 \cdot 2 \cdots (n-1) \cdot n$

Proof: Use the Pumping Lemma

$$L = \{a^{n!} : n \geq 0\}$$

Assume for contradiction
that L is a regular language

Since L is infinite
we can apply the Pumping Lemma

$$L = \{a^{n!} : \; n \geq 0\}$$

Let m be the integer in the Pumping Lemma

①

Pick a string w such that: $w \in L$

length $|w| \geq m$

We pick $w = a^{m!}$

Write $\quad a^{m!} = x\,y\,z$

2

From the Pumping Lemma

it must be that length $\quad |x\,y| \le m, \quad |y| \ge 1$

ทิ้งหมด - ก้อนหน้า

$$xyz = a^{m!} = \overbrace{a\ldots aa\ldots aa\ldots a}^{m}\,\overbrace{a\ldots aa\ldots a}^{m!-m}$$

$$= \underbrace{a\ldots a}_{x}\,\underbrace{a\ldots a}_{y}\,\underbrace{a\ldots aa\ldots aa\ldots a}_{z}$$

Thus: $\quad y = a^k, \quad 1 \le k \le m$

$$x\,y\,z = a^{m!} \qquad\qquad y = a^k, \quad 1 \le k \le m$$

From the Pumping Lemma: $x\,y^i\,z \in L$

$$i = 0, 1, 2, \ldots$$

เลือก i=2

Thus: $x\,y^2\,z \in L$

$$x\,y\,z = a^{m!} \qquad\qquad y = a^{k}, \quad 1 \le k \le m$$

From the Pumping Lemma: $x\,y^2\,z \;\in\; L$

$$xy^2z = \overbrace{a\ldots a\,a\ldots a\,a\ldots a\,a\ldots a}^{m+k}\,\overbrace{a\ldots a\,a\ldots a}^{m!-m} \;\in\; L$$

$$xy^2z = \underbrace{a\ldots a}_{x}\,\underbrace{a\ldots a}_{y}\,\underbrace{a\ldots a}_{y}\,\underbrace{a\ldots a\,a\ldots a\,a\ldots a}_{z} \;\in\; L$$

Thus: $a^{m!+k} \in L$

ได้ออกมา

ก็ต้องมาดูแล้วว่า $a^{m!+k} \rightarrow a^{n!}$ ได้หรือไม่

$$a^{m!+k} \in L \qquad\qquad 1 \le k \le m$$

Since: $L = \{a^{n!} : \; n \ge 0\}$

$\Downarrow$

There must exist p such that:

$$m!+k = p!$$

However: $m!+k \le m!+m$ **for** $m > 1$

$\le m!+m!$ ใส่เครื่องหมายเพิ่มไป ตามกรณีไม่เปลี่ยน

$1 \le k \le m$

ตอนเลือก $|x+y| \le m$

คูณเข้าไป

$< m!m + m!$

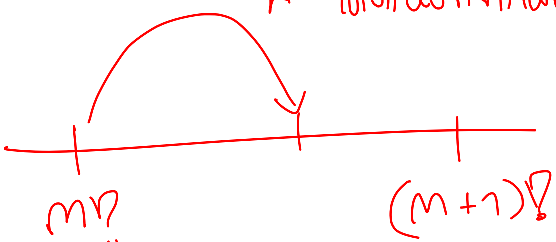
affect

$= m!(m+1)$ $\because (m+1)!$ → $(m+1)(m!)$

$= (m+1)!$

k ได้น้อยกว่าไม่ถึง m

จำนวนที่ครอบได้ไปแพร่ไกลกว่ามาก m (ตา)



$m!+k < (m+1)!$

ห่างแค่ +k

$m!+k \ne p!$ **for any** p

∴ ไม่มีสมการเปรียบอยู่ในจำนวน ! ใดๆ ได้เลย

L is not regular

$$a^{m!+k} \in L \qquad 1 \le k \le m$$

BUT: $L = \{a^{n!}: n \ge 0\}$

$$\Downarrow$$

$$a^{m!+k} \notin L$$

CONTRADICTION!!!

Therefore: Our assumption that L is a regular language is not true

Conclusion: L is not a regular language

Lex

ออกข้อสอบไม่ลึกมาก

ที่เรียนไปทั้งหมดนั้นคือ ส่วนประกอบหนึ่งของ compiler จริงๆเขาเรียกกันส่วนนี้ว่า Lex

Lex: a lexical analyzer

- A Lex program recognizes strings

free source

- For each kind of string found
  the lex program takes an action

ตัวอักษร

Input

```
Var = 12 + 9;
if (test > 20)
   temp = 0;
else
   while (a < 20)
      temp++;
```

แปลงเป็น token (list)

Lex program

Output

ชื่อ token

Identifier: Var
Operand: =
Integer: 12
Operand: +
Integer: 9
Semicolon: ;
Keyword: if
Parenthesis: (
Identifier: test
....

In Lex strings are described with regular expressions

การที่เราจะให้ได้ output ของ lex ออกมาแบบนี้ได้ เราต้องป้อน input เป็น regular expression

Lex program

Regular expressions

"+"
"-"
"="

/* operators */

"if"
"then"

/* keywords */

Lex program

Regular expressions

```
(0|1|2|3|4|5|6|7|8|9)+    /* integers */

(a|b|..|z|A|B|...|Z)+     /* identifiers */
```

integers

(0|1|2|3|4|5|6|7|8|9)+

[0-9]+

identifiers

(a|b|..|z|A|B|...|Z)+

[a-zA-Z]+

Each regular expression
has an associated action (in C code)

Examples:

เมื่อdefine regular exp เสร็จแล้วต้องมา define action ให้มันด้วย

Regular expression	Action
`\n`	`linenum++;` เพิ่มindex ขึ้น 1
`[0-9]+`	`printf("integer");` token
`[a-zA-Z]+`	`printf("identifier");` token

ถ้าที่รับเข้ามาไม่ match กับ res ใดเลย → เข้าสู่กรณี default

Default action: ECHO;

อ่านไรเข้ามา print ออก

Prints the string identified to the output

A small lex program

```
%%
                                no action
[ \t\n]        ;                /*skip spaces*/
[0-9]+         printf("Integer\n");
[a-zA-Z]+      printf("Identifier\n");
```

Input

1234 test

var 566 78

9800

Output

Integer

Identifier

Identifier

Integer

Integer

Integer

Another program

```
%{
int linenum = 1;
%}
%%
```

สัญลักษณ์การแบ่งส่วน อย่าไปสนใจ

```
[ \t]           ;    /*skip spaces*/
\n              linenum++;
[0-9]+          prinf("Integer\n");
[a-zA-Z]+       printf("Identifier\n");
.               printf("Error in line: %d\n",
                       linenum);
```

แสดงให้ดูว่า error เกิดบรรทัดไหน

Input	Output
1234 test var 566 78 9800 + temp	Integer Identifier Identifier Integer Integer Integer Error in line: 3 Identifier

Lex matches the longest input string

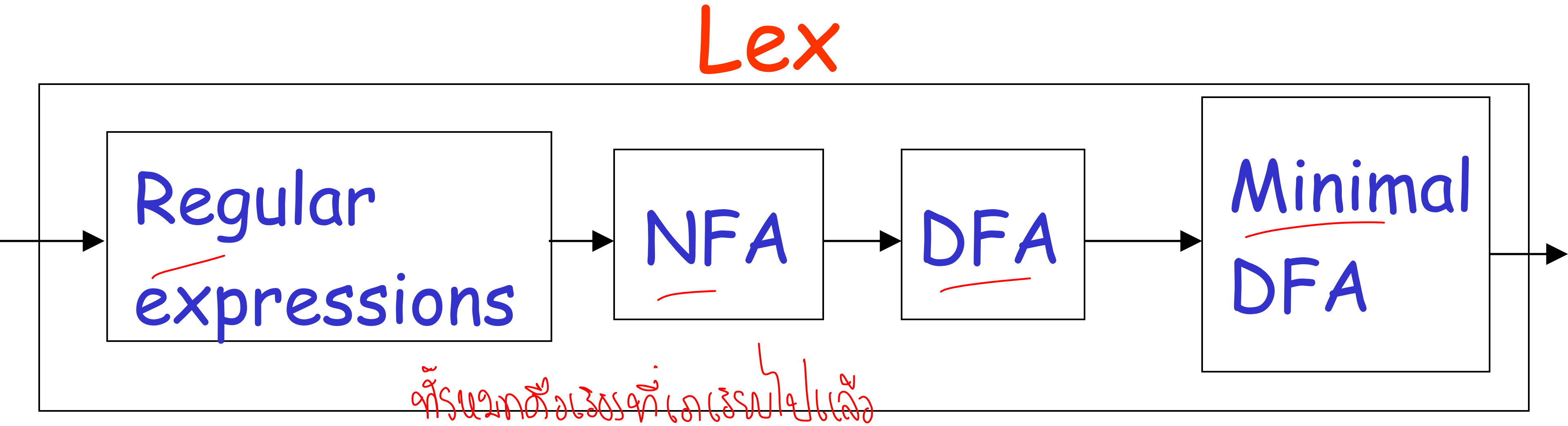
ผลการ match string ของ lex → เลือกยาวสุด

Example: Regular Expressions "if"

"ifend"

Input:	ifend	if
	อันยาวสุด	ยาวสุด
Matches:	"ifend"	"if"

Internal Structure of Lex



The final states of the DFA are associated with actions