More Applications

of

the Pumping Lemma

class 7

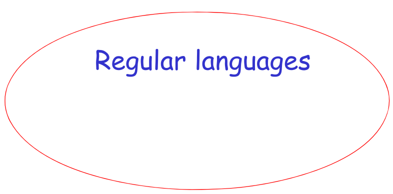
1

The Pumping Lemma:

- 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: $x\,y^i\,z \in L \qquad i = 0, 1, 2, \ldots$

2

Non-regular languages $\qquad L = \{vv^R : v \in \Sigma^*\}$



3

Theorem: The language

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

is not regular

Proof: Use the Pumping Lemma

4

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

Assume for contradiction
that L is a regular language

Since L is infinite
we can apply the Pumping Lemma

5

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

Let m be the integer in the Pumping Lemma

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

length $|w| \geq m$

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

6

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

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

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

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

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

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

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

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

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

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

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

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

$$xy^2z = \underbrace{a...a}_{x}\underbrace{a...a}_{y}\underbrace{a...a}_{y}\underbrace{a...ab...bb...ba...a}_{z} \in L$$

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

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

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

$\Downarrow$

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

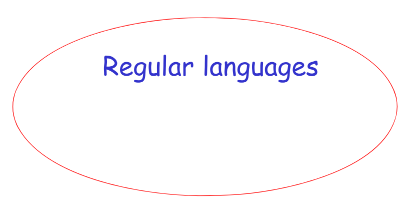
CONTRADICTION!!!

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

Conclusion: L is not a regular language

Non-regular languages

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



Theorem: The language

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

is not regular

Proof: Use the Pumping Lemma

13

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

Assume for contradiction that L is a regular language

Since L is infinite we can apply the Pumping Lemma

14

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

Let m be the integer in the Pumping Lemma

Pick a string w such that: $w \in L$ and length $|w| \geq m$

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

15

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

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

$$xyz = \overbrace{\underbrace{a...a}_{x}\underbrace{a...a}_{y}\underbrace{a...a}_{}}^{m}\overbrace{b...b}^{m}\overbrace{c...c...c}^{2m}$$

z

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

16

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

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

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

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

17

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

From the Pumping Lemma: $xz \in L$

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

z

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

18

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

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

$\Downarrow$

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

CONTRADICTION!!!

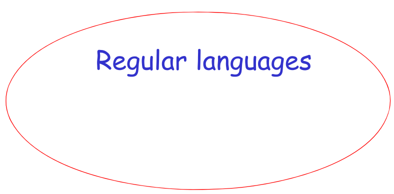
19

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

Conclusion: L is not a regular language

20

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



21

Theorem: The language $L = \{a^{n!} : n \geq 0\}$
is not regular

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

Proof: Use the Pumping Lemma

22

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

Assume for contradiction
that L is a regular language

Since L is infinite
we can apply the Pumping Lemma

23

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

Let m be the integer in the Pumping Lemma

Pick a string w such that: $w \in L$
length $|w| \geq m$

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

24

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

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

$$xyz = a^{m!} = \overbrace{\underbrace{a...a}_{x}\underbrace{a...a}_{y}a...a}^{m}\underbrace{\overbrace{a...aa...a}^{m!-m}}_{z}$$

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

25

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

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

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

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

26

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

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

$$xy^2z = \overbrace{\underbrace{a...a}_{x}\underbrace{a...a}_{y}\underbrace{a...a}_{y}a...a}^{m+k}\underbrace{\overbrace{a...aa...a}^{m!-m}}_{z} \in L$$

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

27

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

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

$\Downarrow$

There must exist p such that:

$m!+k = p!$

28

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

$\le m!+m!$

$< m!m + m!$

$= m!(m+1)$

$= (m+1)!$

$\Downarrow$

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

$\Downarrow$

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

29

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

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

$\Downarrow$

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

CONTRADICTION!!!

30

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

Conclusion: L is not a regular language

31

Lex

32

Lex: a lexical analyzer

- *A Lex program recognizes strings*
- For each kind of string found the lex program takes an action

33

Input

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

→ Lex program →

Output

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

34

In Lex strings are described with regular expressions

Lex program

Regular expressions

"+" "-" "="	/* operators */
"if" "then"	/* keywords */

35

Lex program

Regular expressions

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

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

36

integers

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

37

identifiers

(a|b|..|z|A|B|...|Z)+ [a-zA-Z]+

38

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

Examples:

Regular expression	Action
\n	linenum++;
[0-9]+	prinf("integer");
[a-zA-Z]+	printf("identifier");

39

Default action: ECHO;

Prints the string identified
to the output

40

A small lex program

```
%%
[ \t\n]         ;  /*skip spaces*/

[0-9]+          printf("Integer\n");

[a-zA-Z]+       printf("Identifier\n");
```

41

Input

```
1234   test
var  566    78
9800
```

Output

```
Integer
Identifier
Identifier
Integer
Integer
Integer
```

42

Another program

```
%{
int linenum = 1;
%}
%%
[ \t]          ;   /*skip spaces*/
\n             linenum++;
[0-9]+         prinf("Integer\n");
[a-zA-Z]+      printf("Identifier\n");
.              printf("Error in line: %d\n",
                   linenum);
```

43

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

44

Lex matches the longest input string

Example: Regular Expressions "if"
"ifend"

Input:	ifend	if
Matches:	"ifend"	"if"

45

Internal Structure of Lex

Lex

→ Regular expressions → NFA → DFA → Minimal DFA →

The final states of the DFA are associated with actions

46