

Regular Expression (RE)

Lecturer: Manar Elkady, Ph.D

REGULAR EXPRESSIONS

- Regular expression over a set of terminals Σ is defined recursively as:
 1. any terminal symbol, member in Σ , is a regular expression.
 2. The union of two regular expressions, r_1 , r_2 , written as $r_1 + r_2$, is a regular expression
 3. The concatenation of two regular expressions r_1 , r_2 , written as r_1r_2 is also a regular expression.



REGULAR EXPRESSIONS

4. The iteration (or closure) of a regular expression, written as r^* is also a regular expression.
5. If r is a regular expression then (r) is also a regular expression
6. The regular expressions over Σ are those obtained recursively by applying rules 1-5 once or several times



REGULAR EXPRESSIONS

- Given:

Σ an alphabet

()

Operators:

+ Union

. Concatenation

* star-closure

Expression: $(a + b.c)^*$ defined over $\Sigma = \{a, b, c\}$

Star-closure of $\{a\} + \{bc\}$

$\{\Lambda, a, bc, aa, abc, bca, bcbc, aaa, aabc, \dots\}$



REGULAR EXPRESSIONS

Let Σ be an alphabet

1. Φ , Λ , and $a \in \Sigma$, regular expressions (primitive)

2. If r_1 and r_2 are regular expressions

$r_1 + r_2$,

$r_1 \cdot r_2$,

r_1^* , and

(r_1) are regular expressions

3. A string is a regular expression iff it can be derived from primitive regular expressions by *finite* number of applications of (2)



REGULAR LANGUAGES

Language $L(r)$ denoted by regular expression r is:

1. Φ empty set
2. $\{\Lambda\}$
3. Any $a \in \Sigma$ $\{a\}$



REGULAR LANGUAGES

if r_1 and r_2 are regular expressions

4. $L(r_1 + r_2) = L(r_1) \cup L(r_2)$

5. $L(r_1 \cdot r_2) = L(r_1) L(r_2)$

6. $L((r_1)) = L(r_1)$

7. $L(r_1^*) = (L(r_1))^*$



REGULAR LANGUAGES

- Recursive definition
- Last 4 rules reduces a language to simpler components
- First 3 are termination conditions
- Precedence: $*$ $.$ $+$
- We may omit “.” : $r_1 . r_2$ written $r_1 r_2$



EXAMPLE

- Describe as regular expressions over $\Sigma=\{0, 1\}$
 - $\{101\}$
 - $\{01, 10\}$
 - $\{\epsilon, 0, 00, 000, \dots\}$



EXAMPLE

- Describe the set of all strings of 0's and 1's ending with 00.



EXAMPLE

- Give a regular expression representing a language of strings L where every 0 is immediately followed by at two 1's.



Example

■ Regular expression: $(a + b) \cdot a^*$

$$\begin{aligned} L((a + b) \cdot a^*) &= L((a + b)) L(a^*) \\ &= L(a + b) L(a^*) \\ &= (L(a) \cup L(b)) (L(a))^* \\ &= (\{a\} \cup \{b\}) (\{a\})^* \\ &= \{a, b\} \{\lambda, a, aa, aaa, \dots\} \\ &= \{a, aa, aaa, \dots, b, ba, baa, \dots\} \end{aligned}$$

Example

- Regular expression $r = (aa)^*(bb)^*b$
- What is the regular languages that r describes?

$$L(r) = \{a^{2n}b^{2m}b : n, m \geq 0\}$$

Equivalent Regular Expressions

- Definition:

Regular expressions and r_1 r_2

are **equivalent** if

$$L(r_1) = L(r_2)$$

Example

$L = \{\text{all strings with no two consecutive 0's}\}$

$$r_1 = (1 + 01)^* (0 + \lambda)$$

$$r_2 = (1^* 0 1 1^*)^* (0 + \lambda) + 1^* (0 + \lambda)$$

$L(r_1) = L(r_2) = L \implies r_1 \text{ and } r_2$
are equivalent
regular expr.

Example

- Regular expression

$$r = (0 + 1)^* 00 (0 + 1)^*$$

$L(r)$ = { all strings with at least
two consecutive 0's }

More RE Examples

RE-1

➤ Example 1

$$\Sigma = \{a, b\}$$

- Formally describe all words with a followed by any number of b's

$$L = a b^* = ab^*$$

- Give examples for words in L

$$\{a \text{ ab abb abbb } \dots\}$$

RE-2

➤ Example

$$\Sigma = \{a, b\}$$

- Formally describe all words with a followed by one or more b's

$$L = \text{a} \text{b}^+ = \text{abb}^*$$

- Give examples for words in L

$$\{\text{ab abb abbb} \dots\}$$

RE-3

➤ Example

$$\Sigma = \{a, b, c\}$$

- Formally describe all words that start with an a followed by any number of b's and then end with c.

$$L = a b^* c$$

- Give examples for words in L

$$\{ac \ abc \ abbcb \ abbbcb \ \dots\}$$

RE-4

➤ Example

$$\Sigma = \{a, b\}$$

- Formally describe the language that contains nothing and contains words where any a must be followed by one or more b's

$$L = b^*(abb^*)^* \text{ OR } (b+ab)^*$$

- Give examples for words in L

$$\{\Lambda \text{ ab abb abababb } \dots \text{ b bbb} \dots\}$$

RE-5

➤ Example

$$\Sigma = \{a, b\}$$

- Formally describe all words where a's if any come before b's if any.

$$L = a^* b^*$$

- Give examples for words in L

$$\{\Lambda a b aa ab bb aaa abb abbb bbb.....\}$$

$$\text{NOTE: } a^* b^* \neq (ab)^*$$

because first language does not contain abab but second language has.
Once single b is detected then no a's can be added

RE-6

➤ Example

$$\Sigma = \{a\}$$

- Formally describe all words where count of a is odd.

$$L = a(aa)^* \text{ OR } (aa)^*a$$

- Give examples for words in L

$$\{a \text{ aaa aaaaa } \dots\}$$

RE-7.1

➤ Example

$$\Sigma = \{a, b, c\}$$

- Formally describe all words where single a or c comes in the start then odd number of b's.

$$L = (a+c)b(bb)^*$$

- Give examples for words in L

$$\{ab \ cb \ abbb \ cbbb \ \dots\}$$

RE-7.2

➤ Example

$$\Sigma = \{a, b, c\}$$

- Formally describe all words where single a or c comes in the start then odd number of b's in case of a and zero or even number of b's in case of c.

$$L = ab(bb)^* + c(bb)^*$$

- Give examples for words in L

$$\{ab \ c \ abbb \ cbb \ abbbbb \ \dots\}$$

RE-8

➤ Example

$$\Sigma = \{a, b, c\}$$

- **Formally** describe all words where one or more a **or** one or more c comes in the start then one or more b's.

$$L = (\cancel{a^+} + \cancel{c^+}) \cancel{b^+} = (aa^* + cc^*) bb^*$$

- Give examples for words in L
 $\{ab \ cb \ aabb \ cbbb \ \dots\}$

RE-9.1

➤ Example

$$\Sigma = \{a, b\}$$

- Formally describe all words with length three.

$$L = \overbrace{(a+b)^3}^{\text{red}} = (a+b) (a+b) (a+b)$$

- List all words in L

$$\{aaa \text{ aab } aba \text{ abb } baa \text{ bab } bba \text{ bbb}\}$$

- What is the count of words of length 4?

$$16 = 2^4$$

- What is the count of words of length 44?

$$2^{44}$$

RE-9.2

➤ Example

$$\Sigma = \{a, b, c, d\}$$

- Formally describe all words with length three.

$$L = \cancel{(a+b+c+d)}^3 = (a+b+c+d) (a+b+c+d) (a+b+c+d)$$

- First and last words in L:

$$\{aaa \dots ddd\}$$

- What is the count of words?

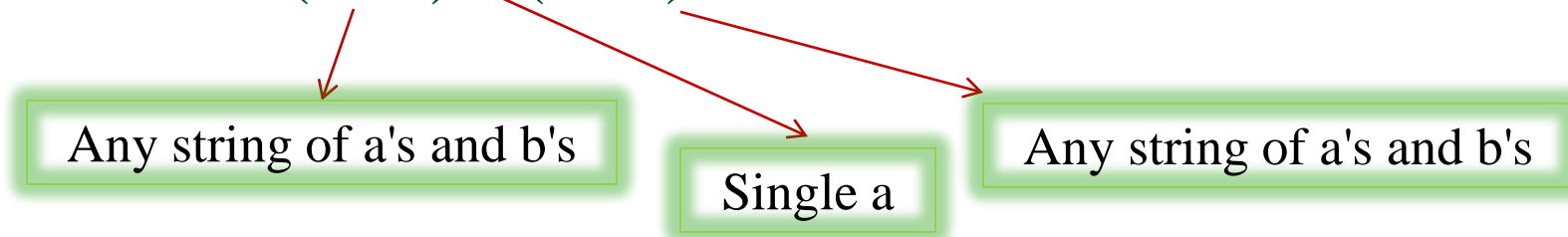
$$4^3 = 64$$

RE-10.1

➤ Example

$\Sigma = \{a, b\}$, What does L describe?

- $L = (a+b)^* a (a+b)^*$

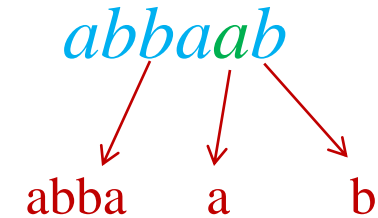
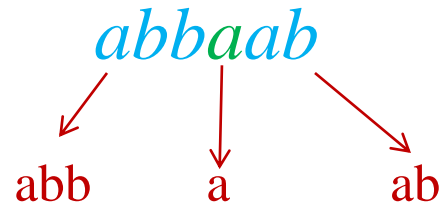
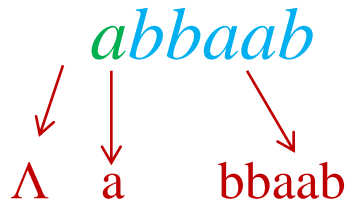


- Give examples for words in L

$\{a \text{ } ab \text{ } aab \text{ } bab \text{ } abb \text{ } \dots\}$

RE-10.2

Ambiguity: abbaab can be parsed in 3 ways



- $L = (a+b)^* a (a+b)^*$

RE-11

➤ Example

$$\Sigma = \{a, b\}$$

- Formally describe all words with at least two a's.

$$1) L = b^*ab^*a(a + b)^*$$

- Start with a jungle of b's (or no b's) until we find the first a, then more b's (or no b's), then the second a, then we finish up with anything.

- Give examples for words in L

$$\{abbbabb \ aaaaa \ bbbabbbbabab \dots\}$$

RE-12

➤ Example

$$\Sigma = \{a, b\}$$

- Formally describe all words with exactly two a's.

$$L = b^*ab^*ab^*$$

- Give examples for words in L

$$\{aa, aab, baba, \text{ and } bbbabbbab \dots\}$$

- To make the word *aab*, we let the first and second b^* become Λ and the last becomes b

RE-13.1

➤ Example

$$\Sigma = \{a, b\}$$

- Formally describe all words with at least one a and at least one b.

$$\begin{aligned} 1) L &= (a + b)^* a (a + b)^* b (a + b)^* \\ &= (\text{anything}) a (\text{anything}) b (\text{anything}) \end{aligned}$$

But $(a+b)^* a (a+b)^* b (a+b)^*$ expresses all words **except** words of the form some b's (at least one) followed by some a's (at least one). **bb^*aa^***

RE-13.2

$$2) L = (a+b)^*a(a+b)^*b(a+b)^* + bb^*aa^*$$

$$\begin{aligned}\text{Thus: } & (a+b)^*a(a+b)^*b(a+b)^* + (a+b)^*b(a+b)^*a(a+b)^* \\ & = (a+b)^*a(a+b)^*b(a+b)^* + bb^*aa^*\end{aligned}$$

- Notice that it is necessary to write bb^*aa^* because b^*a^* will admit words we do not want, such as aaa .

Does this imply that

$$(a+b)^*b(a+b)^*a(a+b)^* = bb^*aa^*??$$

NO!

Left side includes the word aba , which the expression on the right side does not.

RE-13.3

What about this RE? Does it Formally describe all words with at least one a and at least one b?

$$\text{RE} = (\mathbf{a+b})^*(\mathbf{ab+ba})(\mathbf{a+b})^*$$