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 pr several times

Regular Expressions

- Given:

    Σ an alphabet

    ( )

  Operators:

    + Union

    . Concatenation

    * star-closure

Expression: (a + b.c)* defined over Σ = {a, b, c}

Star-closure of {a} + {bc}

{Λ, a, bc, aa, abc, bca, bcbc, aaa, aabc, ….}

REGULAR EXPRESSIONS

Let Σ be an alphabet

1. Φ, Λ, and a ε Σ , regular expressions (primitive)
2. If r_1 and r_2 are regular expressions

    $r_1 + r_2$,

    $r_1 . r_2$,

    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. {Λ}
3. Any a ε Σ {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. $L(r_1 \, . \, r_2) = L(r_1) \, L(r_2)$
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_1r_2

EXAMPLE

- Describe as regular expressions over Σ={0, 1}
  - {101}
  - {01, 10}
  - {^, 0, 00, 000, ….}

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: $(a+b)\cdot a^*$

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

Example

- Regular expression $$r = (aa)^*(bb)^*b$$

- What is the regular languages that r describes?

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

Equivalent Regular Expressions

- Definition:

Regular expressions r_1 and r_2

are **equivalent** if

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

Example

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

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

$$r_2 = (1^* 011^*)^* (0 + \lambda) + 1^* (0 + \lambda)$$

$L(r_1) = L(r_2) = L$ $\Rightarrow$ r_1 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

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

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

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

- Give examples for words in L

{a ab abb abbb …..}

RE-2

- Example

$\sum$ ={a, b}

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

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

- Give examples for words in L

{ab abb abbb …..}

RE-3

- Example

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

- 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 abbc abbbc …..}

RE-4

- Example

$\sum$ ={a, b}

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

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

- Give examples for words in L

{Λ ab abb abababb ….. b bbb…}

RE-5

- Example

$\sum$ ={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

{Λ a b aa ab bb aaa abb abbb bbb…..}

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.

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

- Give examples for words in L

{a aaa aaaaa …..}

RE-7.1

- Example

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

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

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

- Give examples for words in L

{ab cb abbb cbbb …..}

RE-7.2

- Example

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

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

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

- Give examples for words in L

{ab c abbb cbb abbbbb …..}

RE-8

- Example

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

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

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

- Give examples for words in L

{ab cb aabb cbbb …..}

RE-9.1

- Example

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

- Formally describe all words with length three.

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

- List all words in L

  {aaa aab aba abb baa bab bba 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

$\sum$ ={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 … ddd}

- What is the count of words?

$4^3 = 64$

RE-10.1

- Example

Σ ={a, b}, What does L describe?

- $L = \underbrace{(a+b)^*}_{\text{Any string of a's and b's}}\underbrace{a}_{\text{Single a}}\underbrace{(a+b)^*}_{\text{Any string of a's and b's}}$

- Give examples for words in L

{a ab aab bab abb …..}

RE-10.2

Ambiguity: ***abbaab can be parsed in 3 ways***

abbaab

Λ a bbaab

abbaab

abb a ab

abbaab

abba a b

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

RE-11

- Example

$\sum$ ={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…..}

RE-12

- Example

$\sum$ ={a, b}

- Formally describe all words with <u>*exactly*</u> *two a's*.

  L = b*ab*ab*

- Give examples for words in L

  {aa, aab, baba, and bbbabbbab …..}

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

RE-13.1

- Example

$\sum$ ={a, b}

- Formally describe all words with at least one <u>a</u> and at least one <u>b</u>.

1) L = **(a + b)*a(a + b)* b(a + b)***

= (anything) a (anything) b (anything)

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***

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***

- 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?

RE = **(a+b)*(ab+ba)(a+b)***