Linguagens Formais

EDUARDO FREIRE NAKAMURA

Instituto de Computação
Universidade Federal do Amazonas
nakamura@icomp.ufam.edu.br

[1]Este material utiliza conteúdo das aulas fornecidas pelo Prof. Vilar da Câmara Neto (disponível em http://http://prof.vilarneto.com).
[2]Permissão de uso fornecida pelos autores.
[3]As figuras utilizadas neste material são de domínio público, disponíveis na Internet sem informações de direitos autorais.

Conceitos Básicos

OBJETIVO

CONHECER OS CONCEITOS BÁSICOS USADOS NA DEFINIÇÃO DE LINGUAGENS FORMAIS

Linguagens formais

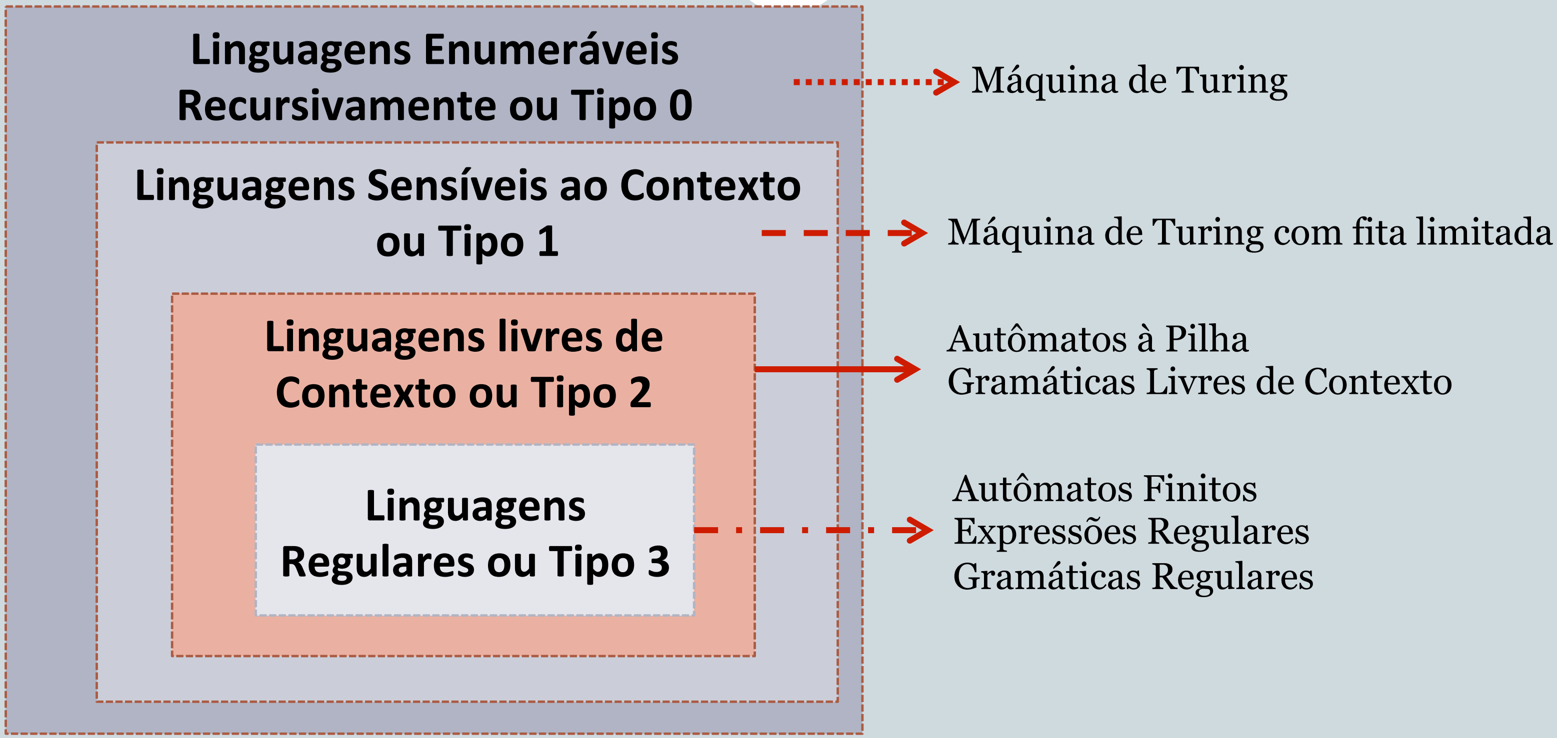
- O conteúdo de Linguagens Formais envolve três principais termos:

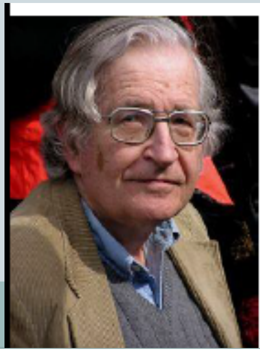
Linguagens	Gramáticas	Autômatos

- Esses termos possuem conceitos próprios, porém, estão relacionados entre si.

- É importante compreender o significado de cada termo.

Hierarquia de Chomsky





Avram Noam Chomsky

- Nasceu em dezembro de 1928 na Philadelphia
- Linguísta, filósofo, cientista cognitivo → considerado o pai da linguística moderna
- Professor Emérito no MIT

https://chomsky.info/

Linguagens formais

- Linguagens construídas respeitando **duas restrições**
  1. **Sintaxe bem definida**: sempre possível verificar se uma sentença pertence ou não à linguagem
  2. **Semântica precisa**: toda sentença da linguagem possui um significado único

www.opendocs.org

Alfabetos

Definição

- Toda linguagem formal possui um alfabeto
  - Conjunto finito e não vazio de símbolos

- O alfabeto de uma linguagem será representado por $\sum$ (sigma)

Exemplo 01

- $\sum = \{0,1,2,3,4,5,6,7,8,9\}$
- Permite representar **números naturais** na base decimal
  - 1
  - 20
  - 131
  - 1235562

Alfabetos

Definição

- Toda linguagem formal possui um alfabeto
  - Conjunto finito e não vazio de símbolos

- O alfabeto de uma linguagem será representado por ∑ (sigma)

Exemplo 02

- ∑ = {0,1,2,3,4,5,6,7,8,9,",",-}

- Permite representar todos os **números reais** na base decimal
  - 9,80665
  - -0,05
  - 1,13
  - 902

*não é necessário usar todos os símbolos

Alfabetos

Definição

- Toda linguagem formal possui um alfabeto
  - Conjunto finito e não vazio de símbolos
- O alfabeto de uma linguagem será representado por $\sum$ (sigma)

Exemplo 03

- $\sum = \{0,1\}$
- Permite representar sequências de 0s e 1s
  - 0
  - 10
  - 101
  - 110010

Alfabetos

Definição

- Toda linguagem formal possui um alfabeto
  - Conjunto finito e não vazio de símbolos
- O alfabeto de uma linguagem será representado por $\sum$ (sigma)

Exemplo 04

- $\sum = \{0\}$
- Permite representar sequências de 0s
  - 0
  - 000
  - 0000
  - 0000000000000

Alfabetos

Definição

- Toda linguagem formal possui um alfabeto
  - Conjunto finito e não vazio de símbolos

- O alfabeto de uma linguagem será representado por $\sum$ (sigma)

Exemplo 05

- $\sum = \{a,b,c,s\}$
- Permite representar algumas palavras
  - *casa*
  - *assa*
  - *babaca*
  - *abacaba*
  - *aaaaaaaaaaaaaaaa*

Palavra

Definição	**Exemplo**
• Uma **palavra** sobre o alfabeto $\sum$ é uma sequência finita de símbolos de $\sum$	• $\sum = \{0, 1\}$ ○ 011 ○ 10010 ○ 01001 ○ 101010 Palavras sobre $\sum$

Palavra

- Tamanho de uma palavra *p*, representado por $|p|$
  - Número de símbolos usados na palavra
- Exemplo
  - $p = 0$, então $|p| = 1$
  - $p = 100$, então $|p| = 3$
  - $p = 1011$, então $|p| = 4$
- Existe uma **palavra vazia**, ou seja, sem símbolos
  - Representada por λ
  - $|\lambda| = 0$

Linguagem

Linguagem

Uma **linguagem sobre um alfabeto** $\sum$ é um subconjunto (possivelmente infinito) de todas as possíveis palavras sobre $\sum$

Uma **linguagem** define quais **palavras** sobre $\sum$ são "**válidas**" segundo algum critério preestabelecido

Tamanho da Linguagem

O **tamanho de uma linguagem** L, representado por $|L|$, é o número de palavras da linguagem (possivelmente infinito)

Linguagem

- Exemplo

Dada a linguagem L_1 definida como "todas as palavras sobre o alfabeto $\sum$ = {0,1} com no máximo 3 símbolos"

$$L_1 = \{\lambda, 0, 1, 00, 01, 10, 11, 000, 001, 010, 011, 100, 101, 110, 111\}$$

Tamanho de L_1 é dado por $|L_1| = 15$

Linguagem

- Exemplo

Dada a linguagem L_2 definida como "todas as palavras sobre o alfabeto $\sum = \{a,b,c\}$ com dois símbolos e que não começam com *a*"

$$L_2 = \{\ ba,\ bb,\ bc,\ ca,\ cb,\ cc\}$$

Tamanho de L_2 é dado por $|L_2| = 6$

Linguagem

- Exemplo

Dada a linguagem L_3 definida como "todos os números binários com um dígito que começam com 0 e terminam com 1"

$$L_3 = \varnothing$$

Tamanho de L_3 é dado por $|L_3| = 0$

Linguagem

- O alfabeto $\sum = \{0,1,2,3,4,5,6,7,8,9\}$ pode representar números naturais (na base decimal)

Todas as palavras sobre $\sum$ representam um número?

Todas as palavras sobre $\sum$ são válidas?

λ é uma palavra sobre $\sum$, mas não é um número natural!

Linguagem

- "A linguagem dos números naturais é constituída por todas as palavras sobre o alfabeto $\sum = \{0,1,2,3,4,5,6,7,8,9\}$ que possuem pelo menos um dígito"

- "Uma palavra *p* sobre o alfabeto $\sum = \{0,1,2,3,4,5,6,7,8,9\}$ representa um número natural se, e somente se, $|p| > 0$"

Linguagem

- A linguagem dos números naturais é definida por

$$L_{\mathrm{N}} = \{p \text{ é uma palavra sobre } \Sigma \mid |p| > 0\}$$

Linguagem

- A linguagem dos números naturais é definida por

$$L_{\mathrm{N}} = \{p \text{ é uma palavra sobre } \Sigma \mid |p| > 0\}$$

"Conjunto das palavras p sobre o alfabeto ∑ tal que o tamanho de p é maior do que zero"

Linguagem

- A linguagem dos números naturais é definida por

$$L_N = \{p \text{ é uma palavra sobre } \Sigma \mid |p| > 0\}$$

"Conjunto das palavras p sobre o alfabeto Σ *tal que o tamanho de p é maior do que zero"*

Linguagem

- A linguagem dos números naturais é definida por

$$L_{\mathrm{N}} = \{p \text{ é uma palavra sobre } \sum \mid |p| > 0\}$$

"Conjunto das palavras p sobre o alfabeto ∑ tal que o tamanho de p é maior do que zero"

Linguagem

- A linguagem dos números naturais é definida por

$$L_{\mathrm{N}} = \{p \text{ é uma palavra sobre } \Sigma \mid |p| > 0\}$$

“Conjunto das palavras p sobre o alfabeto ∑ tal que o tamanho de p é maior do que zero”

Qual o valor de $|L_{\mathrm{N}}|$?

Linguagem

- O alfabeto $\sum$={0,1,2,3,4,5,6,7,8,9,,,-} pode representar números reais
- Algumas palavras sobre $\sum$ são "válidas", outras não

- Válidas
  - 1958
  - -55
  - 2,1
  - -9,44
- Inválidas
  - -
  - 1,2,3
  - λ
  - 20-7
  - ,-2

Formalização de linguagens

- A linguagem dos números reais é constituída por todas as palavras sobre o alfabeto $\sum$ = {0,1,2,3,4,5,6,7,8,9,,,-} que obedeçam às seguintes regras
  1. Possuem pelo menos um dígito (0 a 9)
  2. Podem possuir um único hífen (-), desde que seja o primeiro símbolo
  3. Possuem no máximo uma vírgula (,), desde que seja precedida e seguida por pelo menos um dígito

★ Difícil de entender

★ Difícil de verificar se está correta

★ Difícil de escrever um algoritmo para verificar se uma palavra representa ou não um número válido

Operações

OBJETIVO

COMPREENDER E SABER APLICAR OPERAÇÕES BÁSICAS DEFINIDAS PARA LINGUAGENS FORMAIS

Operações básicas

- Certas operações permitem representação conveniente de linguagens
  - Repetição
  - Fecho de Kleene e Fecho Positivo de Kleene
  - Agrupamento
  - Concatenação
  - União
  - Interseção
  - Subtração
  - Complemento

Repetição

- O operador de repetição aparece como um "expoente" após um símbolo e indica o número de repetições desse símbolo
  - 0^2 é uma sequência de dois "0"s: 00
  - 1^5 é uma sequência de cinco "1"s: 11111
  - z^3 é uma sequência de três "*z*"s: zzz
  - k^0 é uma sequência de zero "*k*"s: λ

*Qualquer símbolo repetido zero vezes gera λ

Repetição

- O operador de repetição é muito útil na definição de linguagens

$$L = \{a^k \mid k < 5\}$$

Repetição

- O operador de repetição é muito útil na definição de linguagens

$$L = \{a^k \mid k < 5\}$$

"*Conjunto de sequências de a's dado que tenham menos de 5 letras*"

Repetição

- O operador de repetição é muito útil na definição de linguagens

$$L = \{a^k \mid k < 5\}$$

"Conjunto de sequências de a's dado que tenham menos de 5 letras"

Eduardo Freire Nakamura (nakamura@icomp.ufam.edu.br)

Repetição

- O operador de repetição é muito útil na definição de linguagens

$$L = \{a^k \mid k < 5\}$$

"Conjunto de sequências de a's dado que tenham menos de 5 letras"

Repetição

- O operador de repetição é muito útil na definição de linguagens

$$L = \{a^k \mid k < 5\}$$

"Conjunto de sequências de a's dado que tenham menos de 5 letras"

Repetição

- O operador de repetição é muito útil na definição de linguagens

$$L = \{a^k \mid k < 5\}$$

"Conjunto de sequências de a's dado que tenham menos de 5 letras"

Ou ainda

"Conjunto de sequências de a's com menos de 5 letras"

$$L = \{\lambda, a, aa, aaa, aaaa\}$$

Repetição

- O operador de repetição também pode ser aplicado sobre conjuntos (neste caso, gera-se conjuntos de palavras)
  - $L = \{0,1\}^2$ é o conjunto de todas as palavras sobre {0,1} que possuem tamanho 2
    - $L = \{00, 01, 10, 11\}$
  - $L = \{a,b\}^3$ é o conjunto de todas as palavras sobre $\{a,b\}$ que possuem tamanho 3
    - $L = \{aaa,aab,aba,abb,baa,bab,bba,bbb\}$
  - $L = \{0,1\}^0$ é o conjunto de todas as palavras sobre {0,1} que possuem tamanho 0
    - $L = \{\lambda\}$

Só existe uma palavra de tamanho 0 (λ)*, independentemente do alfabeto!*

Repetição

- Exemplo 1

$$L = \{p \in \{0,1\}^n \mid n \leq 2\}$$

Repetição

- Exemplo 1

$$L = \{p \in \{0,1\}^n \mid n \leq 2\}$$

"*Conjunto das palavras sobre o alfabeto* {0,1} *dado que o número de símbolos de* {0,1} *é no máximo 2*"

Repetição

- Exemplo 1

$$L = \{p \in \{0,1\}^n \mid n \leq 2\}$$

"*Conjunto das palavras sobre o alfabeto* {0,1} *dado que o número de símbolos de* {0,1} *é no máximo* 2"

Repetição

- Exemplo 1

$$L = \{p \in \{0,1\}^n \mid n \leq 2\}$$

"Conjunto das palavras sobre o alfabeto {0,1} *dado que o número de símbolos de* {0,1} *é no máximo* 2"

Repetição

- Exemplo 1

$$L = \{p \in \{0,1\}^n \mid n \leq 2\}$$

"*Conjunto das palavras sobre o alfabeto* {0,1} *dado que o número de símbolos de* {0,1} *é no máximo* 2"

Ou

"*Conjunto de palavras sobre* {0,1} *com no máximo dois dígitos*"

$$L = \{\lambda, 0, 1, 00, 01, 10, 11\}$$

Repetição

- Exemplo 2

$$L = \{p \in \{0,1\}^n \mid n > 0\}$$

“Conjunto das palavras sobre o alfabeto {0,1} *com pelo menos um símbolo”*

- Pode ser lido como “Conjunto de números binários”
- Note que esta linguagem possui infinitas palavras, i.e., $|L| = \infty$

Fecho de Kleene

- As repetições “zero ou mais vezes” e “uma ou mais vezes” são tão comuns que há operadores específicos para elas:
  - O Fecho de Kleene, representado por “elevado a asterisco”, aplicado a um conjunto e significa “repetido zero ou mais vezes”
  - Exemplo 1
    - $L = \{s\}^*$
    - Ou seja, $L = \{\lambda, s, ss, sss, ssss, sssss, ...\}$

Fecho de Kleene

- Exemplo 2

$$L = \{a,b,c\}^*$$

Ou seja

$$L = \{p \in \{a,b,c\}^n \mid n \geq 0\}$$

Ou seja

$$L = \{\lambda, a, b, c, aa, ab, ac, ba, bb, bc, ca, cb, cc, aaa, aab, \ldots\}$$

Eduardo Freire Nakamura (nakamura@icomp.ufam.edu.br)

Fecho de Kleene

- Exemplo 3

$$L = \{0,11\}^*$$

- As seguintes palavras fazem parte de *L*

$$L = \{\lambda, 0, 11, 00, 011, 110, 1111, 000, 0011, ...\}$$

- Ou seja, sequências arbitrárias de 0 e 11
- Neste caso, 1 nunca aparece sozinho!

Fecho de Kleene

- É comum o Fecho de Kleene aplicado diretamente ao alfabeto
  - $L = \sum^*$
- Significa "a linguagem contendo todas as palavras sobre o alfabeto $\sum$"

Fecho Positivo de Kleene

- Fecho Positivo de Kleene
  - "elevado a mais"
  - Se aplica a um conjunto
  - Repetido uma ou mais vezes

- Exemplo 1
  - $L = \{0\}^+$
  - $L = \{0, 00, 000, 0000, 00000, ...\}$

Fecho Positivo de Kleene

- Fecho Positivo de Kleene
  - “elevado a mais”
  - Se aplica a um conjunto
  - Repetido uma ou mais vezes

- Exemplo 2
  - $L = \{0,1\}^+$
  - $L = \{p \in \{0,1\}^n \mid n \geq 1\}$
  - $L = \{p \in \{0,1\}^n \mid n > 0\}$

  - $L = \{0,1\}^+$ é uma representação concisa do conjunto de números binários!

Eduardo Freire Nakamura (nakamura@icomp.ufam.edu.br)

Fecho Positivo de Kleene

- Fecho Positivo de Kleene
  - “elevado a mais”
  - Se aplica a um conjunto
  - Repetido uma ou mais vezes

- Exemplo 3
  - $L = \{0,11\}^+$
  - $L = \{0, 11, 00, 011, 110, 1111, 000, 0011, \ldots\}$
  - λ não faz parte de L

Fecho Positivo de Kleene

- Fecho Positivo de Kleene
  - “elevado a mais”
  - Se aplica a um conjunto
  - Repetido uma ou mais vezes

- Exemplo 4 (e se?)
  - $L = \{\lambda, 0, 11\}^+$
  - $L = \{\lambda, 0, 11, 00, 011, 110, 1111, 000, 0011, \ldots\}$

Eduardo Freire Nakamura (nakamura@icomp.ufam.edu.br)

Fecho Positivo de Kleene

- Casos interessantes
  - $L = \emptyset$, então $L^* = \{\lambda\}$
  - $L = \emptyset$, então $L^+ = \emptyset$
  - $L = \{\lambda\}$, então $L^* = \{\lambda\}$
  - $L = \{\lambda\}$, então $L^+ = \{\lambda\}$
- Em geral
  - $\lambda \in L$, então $\lambda \in L^+$, caso contrário λ não pertence a L^+

Concatenação

- A concatenação de duas palavras é a palavra formada pelos símbolos da primeira palavra seguidos dos símbolos da segunda palavra

- A concatenação de duas palavras p e q é escrita pq

- Exemplos
  - p = 1 e q = 0
    - pq = 10
  - r = 1101 e s = 01
    - rs =110101
  - m = **`pala`** e n = **`vra`**
    - mn = **`palavra`**
  - a = **`aaba`** e b = λ
    - ab = **`aaba`**
  - c = λ e d = λ
    - cd = λ

Concatenação

- Concatenação de duas linguagens L_1 e L_2 é a linguagem contendo todas as concatenações de cada palavra de L_1 com cada palavra de L_2

- $L_1L_2 = \{p_1p_2\}$ para cada $p_1 \in L_1$ e $p_2 \in L_2$

- Exemplos
  - $L_1 = \{a,b,c\}$ e $L_2 = \{x,y\}$
    - $L_1L_2 = \{ax, ay, bx, by, cx, cy\}$
  - $L_1 = \{0,00\}$ e $L_2 = \{\lambda,1\}$
    - $L_1L_2 = \{0, 01, 00, 001\}$
  - $L_1 = \{\lambda,0\}$ e $L_2 = \{\lambda,0\}$
    - $L_1L_2 = \{\lambda, 0, 00\}$
  - $L_1 = \{0\}^*$ e $L_2 = \{\lambda,1\}$
    - $L_1L_2 = \{\lambda, 1, 0, 01, 00, 001, \ldots\}$
  - $L_1 = \{0\}^*$ e $L_2 = \{1\}^*$
    - $L_1L_2 = ?$

Concatenação

- Se uma das linguagens concatenadas for vazia
- O resultado também é um conjunto vazio!

- Exemplos
  - $L_1 = \{a,b,c\}$ e $L_2 = \varnothing$, $L_1L_2 = \varnothing$
  - $L_1 = \varnothing$ e $L_2 = \{\lambda\}$, $L_1L_2 = \varnothing$
  - $L_1 = \varnothing$ e $L_2 = \{0,1\}$, $L_1L_2 = \varnothing$
  - $L_1 = \varnothing$ e $L_2 = \varnothing$, $L_1L_2 = \varnothing$

Cuidado! A linguagem $\varnothing$ é diferente da linguagem $\{\lambda\}$!

Agrupamento

- O agrupamento é uma forma conveniente de especificar que uma operação de repetição se aplica sobre uma sequência de outras operações

- O agrupamento é representado por um par de parênteses em torno da sequência de símbolos: (sequência)

Eduardo Freire Nakamura (nakamura@icomp.ufam.edu.br)

Agrupamento

- Exemplos
  - $L_1 = (\{a\}\{a,b\})^2$
    - Linguagem formada por duas vezes a concatenação $\{a\}\{a,b\}$
    - $L_2 = \{aaaa, aaab, abaa, abab\}$
  - $L_2 = (\{a\}\{b\})^0$
    - Linguagem formada por zero vezes a concatenação $\{a\}\{b\}$
    - $L_2 = \{\lambda\}$

Agrupamento

- Exemplos
  - $L_4 = (\{a,b\}\{c,d\})^*$
    - Linguagem formada pelo fecho de Kleene sobre a concatenação $\{a,b\}\{c,d\}$
    - $L_4 = \{\lambda, ac, ad, bc, bd, acac, acad, acbc, acbd, adac, ...\}$
    - $L_4 = \{ac, ad, bc, bd\}^*$
  - $L_5 = (\{a,b\}\{c,d\}^+)^*$
    - $\mathrm{L}_5 = \{$

      $\lambda, ac, ad, acc, acd, ...,$
      $bc, bd, bcc, bcd, ...,$
      $acac, acacc, acacd, acaccc, ...,$
      $accac, accad, accacc, ...$

      $\}$
  - Sempre há uma sequência de *c*'s e/ou *d*'s após cada *a* e cada *b*!

Eduardo Freire Nakamura (nakamura@icomp.ufam.edu.br)

União

- União de duas linguagens L_1 e L_2
- Linguagem que contém todas as palavras de L_1 e L_2
- A união é representada pelo operador $\cup$

- Exemplos
  - $L_1 = \{0,1,01\}$ e $L_2 = \{\lambda,0,11\}$
    - $L_1 \cup L_2 = \{\lambda, 0, 1, 01, 11\}$
  - $L_1 = \{a\}^+$ e $L_2 = \{\lambda\}$
    - $L_1 \cup L_2 = \{a\}^*$
  - $L_1 = \{0,1\}^*\{0\}$ e $L_2 = \{0,1\}^*\{1\}$
    - $L_1 \cup L_2 = \{0, 1\}^+$

Interseção

- Interseção de duas linguagens L_1 e L_2
- Linguagem que contém todas as palavras comuns a L_1 e L_2
- A interseção é representada pelo operador $\cap$

- Exemplos
  - $L_1 = \{\lambda, 0, 1, 01\}$ e $L_2 = \{\lambda, 1, 111\}$
    - $L_1 \cap L_2 = \{\lambda, 1\}$
  - $L_1 = \{aa\}^+$ e $L_2 = \{a^n \mid n \leq 5\}$
    - $L_1 \cap L_2 = \{aa, aaaa\}$
  - $L_1 = \{01\}^*$ e $L_2 = \{0\}^*\{1\}^*$
    - $L_1 \cap L_2 = \{\lambda, 01\}$

Diferença

- Diferença entre duas linguagens L_1 e L_2
- Linguagem que contém as palavras de L_1 que não pertencem a L_2
- A diferença é representada pelo operador −

- Exemplos
  - $L_1 = \{\lambda,0,1,01\}$ e $L_2 = \{\lambda,1,111\}$
    - $L_1 - L_2 = \{0, 01\}$
  - $L_1 = \{a\}^+$ e $L_2 = \{a^n \mid n \leq 5\}$
    - $L_1 - L_2 = \{a^n \mid n > 5\}$
  - $L_1 = \{10\}^*$ e $L_2 = \{01\}^*$
    - $L_1 - L_2 = \{10\}^+$

Complemento

- Complemento de uma linguagem *L*
- Linguagem que contém as palavras que não pertencem a *L*
- $\sum^* - L$
- O complemento é representado por uma barra sobre a linguagem
- $\overline{L} = \sum^* - L$

- Exemplos

$$\sum = \{a,b\} \text{ e } L = \{a\}^* \rightarrow \overline{L} = \{a,b\}^* \{b\} \{a,b\}^*$$

$$\sum = \{a,b\} \text{ e } L = \{a\}^+ \rightarrow \overline{L} = \{a,b\}^* \{b\} \{a,b\}^* \cup \{\lambda\}$$

$$\sum = \{0,1\} \text{ e } L = \{0,1\}^* \{1\} \rightarrow \overline{L} = \{0,1\}^* \{0\} \cup \{\lambda\}$$

Exercícios

- Formalize as linguagens abaixo
  1. Conjunto de todos os números binários com tamanho par e cujos dígitos nas posições pares (2o dígito, 4o dígito, etc.) são obrigatoriamente 0.
  2. Conjunto de todos os números binários que contenham a sequência 0000.
  3. Conjunto de todos os números binários que contenham a sequência 0000 pelo menos três vezes.