Linguagens e Gramáticas

Fundamentos de Matemática Discreta para a Computação

Linguagens e Gramáticas

- Para definir uma linguagem:
  - Alfabeto
  - Palavras ou cadeias de caracteres
- A especificação da construção apropriada de sentenças é chamada **sintaxe da linguagem**.
- Estudaremos a sintaxe da uma classe de linguagens chamadas **gramática com estrutura de frase**.

Modelos Abstratos de Máquina

- Linguagens
  - Gramática com estrutura de frase
    - Definição 1: Um vocabulário (ou alfabeto) V é um conjunto finito e não-vazio de elementos chamados símbolos. Exemplo: letras e dígitos.
    - Uma palavra (ou sentença) sobre V é uma cadeia de tamanho finito sobre V. A cadeia vazia ou cadeia nula, denotado por λ, é a cadeia que não contém símbolos.
    - Se V = {a, b} então:
      - V* = {λ, a, b, aa, ab, ba, bb, aaa, ...}
    - O conjunto de todas as palavras sobre V é denotado por V*.

Linguagens e Gramáticas

- Uma linguagem sobre V é um subconjunto de V*.
- Linguagens podem ser especificadas de várias formas:
  a) Listando todas as palavras da linguagem.
  b) Fornecendo um certo critério que deve ser satisfeito pelas palavras que pertençam à linguagem.
  c) Através do uso de uma **gramática**.
- Uma **gramática** provê um conjunto de vários tipos de símbolos e um conjunto de regras para produção de palavras.

Linguagens e Gramáticas

- Uma **gramática** possui um vocabulário V, que é o conjunto de símbolos utilizados para derivar membros da linguagem.
- Alguns elementos do vocabulário não podem ser substituídos por outros símbolos. Eles são chamados de **terminais** e os outros símbolos do vocabulário que podem ser substituídos por outros símbolos são chamados de **não-terminais**. Os conjuntos de terminais e não-terminais são denotados por T e N respectivamente.
- Existe um membro especial do vocabulário chamado **símbolo inicial**, denotado por S.

Linguagens e Gramáticas

- As regras que especificam quando podemos substituir uma cadeia de V* por outra string são chamadas de produções da gramática. Denotamos por $Z_0 \rightarrow Z_1$ a produção que especifica que Z_0 pode ser substituído por Z_1 em uma cadeia.
- Definição 2: Uma gramática com estrutura de frase G = (V, T, S, P) consiste de um vocabulário V, um subconjunto T de V de elementos terminais. Um símbolo inicial S de V e de um conjunto de produções P. Toda produção em P deve conter somente símbolos não terminais do seu lado esquerdo.

Linguagens e Gramáticas

- Exemplo: Seja G = (V, T, S, P) onde V = {a, b, A, B, S}, T = {a, b}, S é o símbolo inicial e P = {S → ABa, A → BB, B → ab, AB → b}.

$$
\begin{tikzcd}
 & \text{S} \arrow[d] & \\
\text{ba} & \text{ABa} \arrow[l] \arrow[r] \arrow[d] & \text{BBBa} \arrow[dd] \\
 & \text{Aaba} \arrow[d] & \\
 & \text{BBaba} \arrow[d] & \text{abababa} \\
 & \text{abababa} &
\end{tikzcd}
$$

Ambigüidade – 2 caminhos diferentes que levam a uma mesma palavra.

Linguagens e Gramáticas

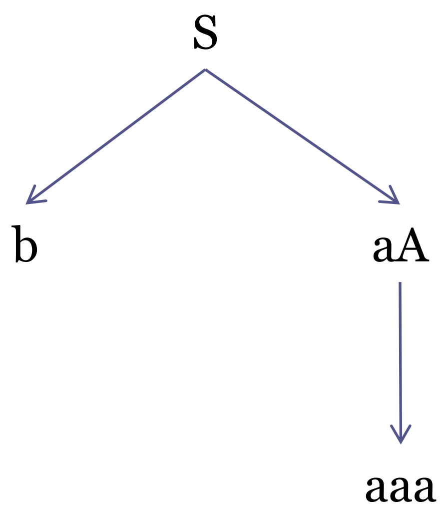
- Definição 3: Seja G = (V, T, S, P). Sejam w_0 = Iz_0r (ou seja, a concatenação de I, z_0 e r) e w_1 = Iz_1r, cadeias sobre V. Se $z_0 \rightarrow z_1$ é uma produção de G, dizemos que w_1 é diretamente derivável de w_0 e escrevemos $w_0 => w_1$. Se $w_0, w_1, \ldots, w_n$ são cadeias sobre V, tal que $w_0 => w_1 => \ldots => w_n$, então dizemos que w_n é derivável de w_0 e escrevemos $w_0 \doteq> w_n$. A seqüência de passos utilizados para obter w_n a partir de w_0 é chamada de **derivação**.

Linguagens e Gramáticas

- Definição 4: Seja $G = (V, T, S, P)$. A linguagem gerada por G (ou a linguagem de G), denotado por $L(G)$, é o conjunto de todas as cadeias de terminais que são deriváveis a partir do estado inicial S, ou seja, $L(G) = \{w \in T^* \mid S \overset{*}{=}> w\}$

Linguagens e Gramáticas

- Exemplo: Seja G a gramática com vocabulário V = {S, A, a, b}, T = {a, b}, S = S e P = {S → aA, S → b, A → aa}. Qual a linguagem L(G) desta gramática?



L(G) = {b, aaa}

Linguagens e Gramáticas

- Exemplos de concatenação:
  - a^5 = aaaaa
  - $(bc)^4$ = bcbcbcbc
- Exemplo: Seja G a gramática com vocabulário V = {S, 0 , 1}, T = {0, 1}, S = S e P = {S → 11S, S → 0}. Qual a linguagem L(G) gerada por esta gramática?
- Resposta: A linguagem de todas as cadeias que possuem um número par de 1's e que terminam com 0. L(G) = $\{1^n0 \mid n \text{ é par}\}$

Linguagens e Gramáticas

- Exemplo: Ache a gramática com estrutura de frase que gere o conjunto $\{0^m 1^n\}$ m e n $\in Z^+$.
  - $G1 = (V, T, S, P)$
    - $V = \{S, 0, 1\}$
    - $T = \{0, 1\}$
    - $P = \{S \rightarrow 0S, S \rightarrow S1, S \rightarrow \lambda\}$
  - $G2 = (V, T, S, P)$
    - $V = \{S, A, 0, 1\}$
    - $T = \{0, 1\}$
    - $P = \{S \rightarrow 0S, S \rightarrow 1A, S \rightarrow 1, A \rightarrow 1A, A \rightarrow 1, S \rightarrow \lambda\}$

Tipos de Gramáticas com Estrutura de Frase

- Uma gramática do tipo 0 não tem restrições em suas produções.
- Uma gramática do tipo 1 pode possuir produções de forma $w_1 \rightarrow w_2$ onde o tamanho de w_2 é maior que o tamanho de w_1 ou da forma $w_1 \rightarrow \lambda$.
- Uma gramática do tipo 2 pode possuir produções apenas de forma $w_1 \rightarrow w_2$, onde w_1 é um único símbolo e é um símbolo não-terminal.

Tipos de Gramáticas com Estrutura de Frase

- Uma gramática do tipo 3 pode possuir produções apenas da forma $w_1 \rightarrow w_2$ com $w_1 = A$ e tanto $w_2 = aB$ ou $w_2 = a$ onde A e B são símbolos não-terminais e a é um símbolo terminal, ou ainda com $w_1 = S$ e $w_2 = \lambda$.
- Gramáticas do tipo 3 são também chamadas de **gramáticas regulares**. Uma linguagem gerada por uma gramática regular é chamada **regular**.
- Exemplo: 0^n1^n é uma linguagem tipo 2, pois as produções da gramática são $S \rightarrow 0S1$ e $S \rightarrow \lambda$.

Linguagens e Gramáticas

- Em estudos anteriores, consideramos o conjunto S^* consistindo de todas as cadeias finitas (strings) de elementos de um conjunto (alfabeto) S.
- Existem várias interpretações para os elementos de S^*, dependendo da natureza de S.
  - se pensarmos em S como um conjunto de "palavras", então S^* pode ser considerado como o conjunto de todas as possíveis "sentenças" formadas a partir de palavras de S. Claro que tais sentenças não necessitam ter um significado.

Linguagens e Gramáticas

- Devemos pensar em uma linguagem como uma especificação completa, pelo menos em princípio, de 3 elementos.
  1. Deve existir um conjunto S de palavras que fazem parte da linguagem.
  2. Um subconjunto de S* deve ser designado como o conjunto de "sentenças corretamente construídas" da linguagem.
  3. Deve ser determinado quais das sentenças corretamente construídas têm significado e qual seu significado.

Comprimento de cadeias

- Seja w_n uma palavra em $\{a,b\}^*$ $n = 1, 2, 3...$
  - Exemplos:
    - w_1 = aab
    - w_2 = abbba
    - w_3 = ababababaaa
  - O comprimento de w é representado por $|w|$, assim:
    - $|w_1| = 3$
    - $|w_2| = 5$
    - $|w_3| = 9$

Prefixo, sufixo, subpalavra e expressão regular

- Seja w uma palavra de {a, b}* e w = aabab
  - Conjunto dos prefixos de w:
    - P = {a, aa, aab, aaba aabab}
  - Conjunto dos sufixos de w:
    - S = {b, ab, bab, abab, aabab}
  - Conjunto das subpalavras de w:
    - $P \cup S$
- Expressão regular
  - $a^*bc^+b^2$
    - * → qualquer quantidade de vezes (inclusive zero vezes).
    - + → uma ou mais vezes.
    - x (algum número natural) → x vezes.

Linguagens e Gramáticas

- Exemplo:
  - Seja S o conjunto dos inteiros, dos símbolos +, -, × e ÷ e os parênteses ( ).
  - Nós obteremos uma linguagem se considerarmos apropriada aquelas cadeias em S* que representam de maneira não ambígua expressões algébricas.
    - $((3-2)+(4\times 7))\div 9$
    - $(7-(8-(9-10)))$
    - São sentenças corretamente construídas nesta linguagem.

Linguagens e Gramáticas

- Exemplo: Qual é uma das gramáticas que produzem as sentenças abaixo?
  - $((3-2)+(4\times 7))\div 9$
  - $(7-(8-(9-10)))$

S	S
ABC	A
(ABA) ÷9	(ABA)
(A+A) ÷9	(C-(ABA))
((ABA)+(ABA)) ÷9	(7-(C-(ABA)))
((C-C)+(CxC)) ÷9	(7-(8-(C-C)))
((3-2)+(4×7)) ÷9	(7-(8-(9-10)))

V = {S, A, B, C, +, -, ×, ÷, (,), 2, 3, 4, 7, 8, 9, 10}
T = {+, -, ×, ÷, (,), 2, 3, 4, 7, 8, 9, 10}
S = S
P = { S → A,
S → A B C,
A → (A B A),
A → C,
B → + | - | × | ÷,
C → 2|3|4|7|8|9|10
}

Linguagens e Gramáticas

- Exercício
  - Para a gramática G = (V, T, S, P) com:
    - V = {E, I, a, b, +, *}
    - T = {a, b, +, *}
    - S = E
    - P = { E → I | E + E | E * E
      I → a | b | Ia | Ib }
  - Encontre a árvore de derivação da palavra **a+b*a**.

E
E+E
I+E
a+E
a+E*E
a+I*E
a+b*E
a+b*I
a+b*a

E
E*E
E+E*E
I+E*E
a+E*E
a+I*E
a+b*E
A+b*I
a+b*a