Gramática regular

IBM1088 Linguagens Formais e Teoria da Computação

Evandro Eduardo Seron Ruiz

evandro@usp.br

Universidade de São Paulo

Frase do dia

Através de três métodos podemos encontrar a sabedoria: primeiro pela reflexão que é o mais nobre deles; segundo, pela imitação, que é o mais fácil e; terceiro, pela experiência que é o mais amargo.

Confúcio[a]

[a]Filósofo chinês (551 A.C. — 479 A.C.).

Objetivos

- Conceituar gramáticas regulares como outro método de formalizar linguagens regulares
- Capacitar o estudante no entendimento de regras de produção de gramáticas simples
- Envolver GR, AF e LR como um único grande tópico de estudo

Conteúdo

Expressões regulares

- Uma linguagem = conjunto de palavras
- Linguagem de programação conjunto de todos os programas
- portanto, LP = conjunto infinito
- Conjunto infinito não é um formalismo adequado para um computador
- Conjunto finito de regras é mais viável
- Gramática = conjunto finito de regras que geram palavras

Expressões regulares

Uma gramática irrestrita é uma quádrupla ordenada
$G = (V, T, P, S)$ em que:

- V é um conjunto finito de símbolos variáveis ou não-terminais;
- T é um conjunto finito de símbolos terminais disjunto de V;
- P é um conjunto finito de pares denominados regras de produção tal que a primeira componente é a palavra $(V \cup T)^+$ e, a segunda palavra é palavra de $(V \cup T)^*$; e
- S é o elemento de V denominado variável inicial.

Em outras palavras

Gramática

Uma gramática é um formalismo para geracão de todas as palavras de uma linguagem.
Uma gramática permite derivar ("gerar") todas as palavras de uma linguagem.

Regras de produção

- Regra de produção: (α, β)
- Representação: $\alpha \rightarrow \beta$

Regras de produção

- Regra de produção: (α, β)
- Representação: $\alpha \rightarrow \beta$
- Seqüência de regras de produção: $\alpha \rightarrow \beta_1, \alpha \rightarrow \beta_2, \ldots, \alpha \rightarrow \beta_n$
- Forma abreviada: $\alpha \rightarrow \beta_1 | \beta_2 | \ldots | \beta_n$

Regras de produção

- Regra de produção: (α, β)
- Representação: $\alpha \rightarrow \beta$
- Seqüência de regras de produção: $\alpha \rightarrow \beta_1, \alpha \rightarrow \beta_2, \ldots, \alpha \rightarrow \beta_n$
- Forma abreviada: $\alpha \rightarrow \beta_1 | \beta_2 | \ldots | \beta_n$
- Aplicação sucessiva de uma RP: derivação de uma palavra

Derivação

Seja $G = (V, T, P, S)$ uma gramática:

Derivação

Uma derivação é um par da relação

$$(V \cup T)^{+} \Rightarrow (V \cup T)^{*}$$

que pode ser representada na forma infixa como $\alpha \Rightarrow \beta$.

Definição da relação derivação

A relação $\Rightarrow$, de $\alpha \Rightarrow \beta$, é indutivamente definida como:

- Para toda produção na forma $S \rightarrow \beta$ (S símbolo inicial) temos que:

  $$S \Rightarrow \beta$$

  e;
- Para todo par $\alpha \Rightarrow \beta$, em que $\beta = \beta_u\beta_v\beta_w$, se $\beta_v \rightarrow \beta_t$ é regra de P, então

  $$\beta \Rightarrow \beta_u\beta_t\beta_w$$

Explicando...

- A derivação é a substituição de uma palavra de acordo com uma regra de produção
- Quando desejamos explicitar a regra de produção $p \in P$ que define a derivação $\alpha \Rightarrow \beta$, usamos a notação

$$\alpha \Rightarrow^{p} \beta$$

Quais são os passos possíveis de uma derivação?

- $\Rightarrow^*$ Fecho transitivo e reflexivo da relação $\Rightarrow$, ou seja, zero ou mais passos;
- $\Rightarrow^+$ Fecho transitivo da relação $\Rightarrow$, ou seja, um ou mais passos;
- $\Rightarrow^i$ Exatos i-passos da relação $\Rightarrow$, $i \in \mathbb{N}$.

Exemplo

Suponha $G = (V, T, P, N)$ uma gramática capaz de gerar qualquer número natural válido. Assiml G, na qual,:

- $V = \{N, D\}$, símbolos não terminais;
- $T = \{0, 1, 2, \ldots, 9\}$, símbolos terminais; e
- $P = \{N \rightarrow D, N \rightarrow DN, D \rightarrow 0 \mid 1 \mid \ldots \mid 9\}$,

gera, sintaticamente, $\mathbb{N}$.

Exemplo

Suponha $G = (V, T, P, N)$ uma gramática capaz de gerar qualquer número natural válido. Assiml G, na qual,:

- $V = \{N, D\}$, símbolos não terminais;
- $T = \{0, 1, 2, \ldots, 9\}$, símbolos terminais; e
- $P = \{N \to D, N \to DN, D \to 0 \mid 1 \mid \ldots \mid 9\}$,

gera, sintaticamente, $\mathbb{N}$.
Perceba que G gera distintamente 0123 e 123

Regra × Passo da derivação

Passo da derivação	Regra usada
$N \Rightarrow$	$N \rightarrow DN$
$DN \Rightarrow$	$D \rightarrow 2$
$2N \Rightarrow$	$N \rightarrow DN$
$2DN \Rightarrow$	$D \rightarrow 4$
$24N \Rightarrow$	$N \rightarrow D$
$24D \Rightarrow$	$D \rightarrow 3$
243	

Existe alguma outra derivação para 243? Sim

Regra $\times$ Passo da derivação

Passo da derivação	Regra usada
$N \Rightarrow$	$N \rightarrow DN$
$DN \Rightarrow$	$D \rightarrow 2$
$2N \Rightarrow$	$N \rightarrow DN$
$2DN \Rightarrow$	$D \rightarrow 4$
$24N \Rightarrow$	$N \rightarrow D$
$24D \Rightarrow$	$D \rightarrow 3$
243	

Existe alguma outra derivação para 243? Sim

Portanto, indica-se que

- $\Rightarrow^* 243$
- $\Rightarrow^+ 243$
- $\Rightarrow^6 243$

Exercício

Suponha $G = (V, T, P, N)$ uma gramática definida por:

- $V = \{A, B\}$, símbolos não terminais;
- $T = \{a, b\}$, símbolos terminais; e
- $P = \{S \rightarrow A, S \rightarrow B, B \rightarrow bB,\ A \rightarrow aA,\ A \rightarrow \epsilon, B \rightarrow \epsilon\}$,

a qual gera palavras em $\Sigma = \{a, b\}$.
Descreva as regras $p \in P$ que produza $w = aaaa$.
Três minutos para a solução...

Passo da derivação × Regras

Passo da derivação	Regra usada
	S
$A \Rightarrow$	$S \rightarrow A$
$aA \Rightarrow$	$A \rightarrow aA$
$aaA \Rightarrow$	$A \rightarrow aA$
$aaaA \Rightarrow$	$A \rightarrow aA$
$aaaaA \Rightarrow$	$A \rightarrow aA$
aaaa	$A \rightarrow \epsilon$

Gramáticas e restrições

- A definição de gramática é bem ampla
- Pela definição podemos definir qualquer linguagem

Gramáticas e restrições

- A definição de gramática é bem ampla
- Pela definição podemos definir qualquer linguagem
- Linguagens regulares são linguagens restritas
- Para definir LR através de gramáticas precisamos restringí-la

Gramáticas e restrições

- A definição de gramática é bem ampla
- Pela definição podemos definir qualquer linguagem
- Linguagens regulares são linguagens restritas
- Para definir LR através de gramáticas precisamos restringí-la
- Restringir uma gramática = restringir regras de produção

Gramáticas lineares

Existem 4 formas de gramáticas lineares, a saber:

1. **GLD** Gramática linear à direita
2. **GLE** Gramática linear à esquerda
3. **GLUD** Gramática linear unitária à direita
4. **GLUD** Gramática linear unitária à esquerda

Vejamos como elas se comportam...

GLD

Seja $G = (V, T, P, S)$ uma gramática. Sejam $A, B \in V$ e $w \in T$.

Gramática linear à direita

Na GLD todas as regras de produção são da forma:

$$A \rightarrow wB, \text{ ou, } A \rightarrow w$$

Ou seja, o lado esquerdo só possui uma variável e o lado direito da produção também só possui, se existir, uma variável que sucede uma subpalavra de terminais.

GLE

Seja $G = (V, T, P, S)$ uma gramática. Sejam $A, B \in V$ e $w \in T$.

Gramática linear à esquerda

Na GLE todas as regras de produção são da forma:

$$A \rightarrow Bw, \text{ ou, } A \rightarrow w$$

Ou seja, o lado esquerdo só possui uma variável e o lado direito da produção também só possui, se existir, uma variável que antecede uma subpalavra de terminais.

GLUD

Seja $G = (V, T, P, S)$ uma gramática. Sejam $A, B \in V$ e $w \in T$.

Gramática linear unitária à direita

Na GLD todas as regras de produção são da forma:

$$A \rightarrow wB, \text{ ou, } A \rightarrow w$$

e, adicionalmente,

$$|w| \leq 1$$

GLUE

Seja $G = (V, T, P, S)$ uma gramática. Sejam $A, B \in V$ e $w \in T$.

Gramática linear unitária à esquerda

Na GLD todas as regras de produção são da forma:

$$A \rightarrow Bw, \text{ ou}, A \rightarrow w$$

e, adicionalmente,

$$|w| \leq 1$$

Unificação da Gramáticas Lineares

Seja L uma linguagem. Então:

Teorema

L é gerada por uma GLD see, L é gerada por uma GLE see, L é gerada por uma GLUD see, L é gerada por uma GLUE.

Ou seja, as diversas formas de gramáticas são formalismos equivalentes.

Gramática linear $\times$ Gramática regular

Definição

Uma gramática G é dita uma gramática regular (GR) se G é uma gramática linear.

Sobre a linguagem gerada por uma GR

Definição

Seja $G = (V, T, P, S)$ uma gramática.
A linguagem gerada por G, denotada por

$$L(G)$$

é tal que:

$$L(G) = \{w \in T^* \mid S \Rightarrow^+ w\}$$

Lembrem-se $\Rightarrow^+$ é o fecho transitivo da relação, ou seja, corresponde a um ou mais passos sucessivos de derivação.

Teste!

Descreva L

Descreva a linguagem gerada pela expressão regular $E = a(ba)^*$

Teste!

Descreva L

Descreva a linguagem gerada pela expressão regular $E = a(ba)^*$

$E = a(ba)^*$ corresponde linguagem formada por todas as palavras que começam com a e que pode ser seguida por uma seqüência de '*ba*'s.
Podem ser considerados quaisquer dos quatro tipos anteriores de gramáticas regulares.
Cinco minutos para a resposta...

Exemplo: GLD

Seja $L = a(ba)^*$ uma linguagem. L pode ser gerada pelas 4 formas de gramáticas lineares, como veremos:

GLD

Seja $G = (\{S, A\}, \{a, b\}, P, S\}$, para P as seguintes regras de produção:

$$S \rightarrow aA$$
$$A \rightarrow baA \,|\, \epsilon$$

e ainda ...

continuação: GLE

Para $L = a(ba)^*$:

GLE

Seja $G = (\{S, A\}, \{a, b\}, P, S\}$, para P as seguintes regras de produção:

$$S \rightarrow a$$
$$S \rightarrow Sba$$

e ainda ...

continuação: GLUD

Para $L = a(ba)^*$:

GLUD

Seja $G = (\{S, A, B\}, \{a, b\}, P, S\}$, para P as seguintes regras de produção:

$$\begin{aligned} S &\rightarrow aA \\ A &\rightarrow bB \mid \epsilon \\ B &\rightarrow aA \end{aligned}$$

Reparem que $|w| \leq 1$.
Ainda temos mais uma, a GLUE...

continuação: GLUE

Para $L = a(ba)^*$:

GLUE

Seja $G = (\{S, A\}, \{a, b\}, P, S\}$, para P as seguintes regras de produção:

$$S \rightarrow a \mid Aa$$
$$A \rightarrow Sb$$

Finalizando assim as gramáticas lineares.

Mais um teste!

Descreva L

Descreva a linguagem gerada pela expressão regular
$E = (a + b)^*(aa + bb)$

Mais um teste!

Descreva L

Descreva a linguagem gerada pela expressão regular
$E = (a + b)^*(aa + bb)$

$E = (a + b)^*(aa + bb)$ corresponde linguagem formada por todas as palavras que terminam com *aa* ou com *bb*.

Outra linguagem gerada por uma GR

A linguagem $L = (a+b)^*(aa+bb)$ pode ser gerada por duas gramáticas lineares.
Vejamos a primeira:

GLD

Seja $G = (\{S, A\}, \{a, b\}, P, S\}$, para P as seguintes regras de produção:

$$S \to aS \mid bS \mid A$$
$$A \to aa \mid bb$$

Vejamos agora a segunda gramática geracional...

outra forma...

Ainda considerando $L = (a+b)^*(aa+bb)$: Vejamos a primeira:

GLE

Seja $G = (\{S, A\}, \{a, b\}, P, S\}$, para P as seguintes regras de produção:

$$S \to Aaa \,|\, Abb$$
$$A \to Aa \,|\, Ab \,|\, \epsilon$$

Finalizando assim as duas formas.

GLD e GLE

Suponha $|w| \geq 1$ e $A, B \in V$, ou seja, não terminais.

Se uma gramática tiver ambas as produções, linear à direita ($A \rightarrow wB$) e linear à esquerda ($A \rightarrow Bw$) então a correspondente linguagem gerada poderá não ser regular, ou seja, esta não é uma gramática regular.

GR e LR

Teorema

Se L é uma linguagem gerada por uma gramática regular, então L é uma linguagem regular

GR e AFNϵ

GLE

Seja $G = (\{S, A, B\}, \{a, b\}, P, S\}$, para P as seguintes regras de produção:

$$
\begin{aligned}
S &\rightarrow aA \\
A &\rightarrow bB \mid \epsilon \\
B &\rightarrow aA
\end{aligned}
$$

GR e AFNϵ

GLE

Seja $G = (\{S, A, B\}, \{a, b\}, P, S\}$, para P as seguintes regras de produção:

$$S \rightarrow aA$$
$$A \rightarrow bB \mid \epsilon$$
$$B \rightarrow aA$$

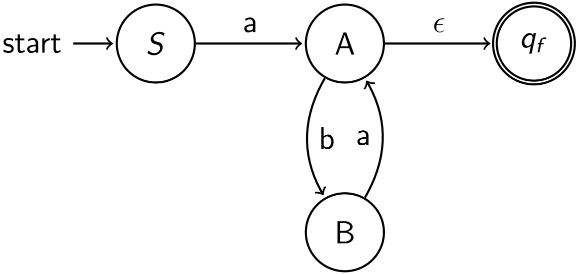
Primeira pergunta: Descreva G como uma ER

GR e AFNϵ, continuação

O AFNϵ que reconhece a linguagem gerada por G é

$$N = \{\{S, A, B, q_f\}, \{a, b\}, \delta, S, \{q_f\}\}$$

ilustrado abaixo:



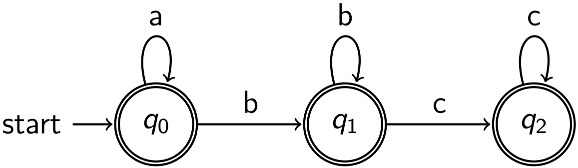
Exercício em sala

Vamos tentar agora resolver um problema inverso ao anterior.
Vamos construir uma gramática a partir de um AFD.
Veja o AFD abaixo:

Considere o autômato finito determinístico

$$M = (\{q_0, q_2, q_2\}, \{a, b, c\}, \delta, q_0, \{q_0, q_1, q_2\}),$$

como ilustrado a seguir:



Recupere a gramática a partir do AFD M.

Dica:

Estas são as regras de produção da gramática:

$$S \to q_0$$
$$q_0 \to aq_0$$
$$q_0 \to bq_1$$

Resposta

Estas são as regras de produção da gramática:

$$
\begin{aligned}
S &\rightarrow q_0 \\
q_0 &\rightarrow aq_0 \\
q_0 &\rightarrow bq_1 \\
q_1 &\rightarrow bq_0 \\
q_1 &\rightarrow cq_2 \\
q_2 &\rightarrow cq_2 \\
q_0 &\rightarrow \epsilon \\
q_1 &\rightarrow \epsilon \\
q_2 &\rightarrow \epsilon
\end{aligned}
$$

Caros,

Ainda veremos gramáticas mais complexas. Procurem fixar o conteúdo enquanto ainda estamos no início deste tópico de gramáticas.