

درس مبانی نظریه محاسبه

جلسه دهم

گرامر و زبان، گرامرهای منظم

Regular Grammars

گرامرها ابزاری برای توصیف زبان

- ◀ در فصل قبل دیدیم که از عبارات منظم برای توصیف زبانهای منظم استفاده می‌شود.
- ◀ در این فصل از ابزار گرامرها برای توصیف زبانها استفاده می‌کنیم.
- ◀ ایده استفاده از گرامرها (بویژه گرامرهای مستقل از متن) به مطالعات روی زبانهای انسانی برمی‌گردد. گرامر مثل یک دسته قوانین کلی است که ساختار جملات یک زبان را مشخص می‌کند. گرامرها رابطه بین عبارات اسمی و فعلی را بصورت یک رابطه بازگشتی بیان می‌کنند (یک عبارت اسمی ممکن است داخل یک عبارت فعلی ظاهر شود و بالعکس).
- ◀ گرامرهای مستقل از متن در طراحی و تولید زبانهای برنامه‌نویسی نیز نقش مهمی را ایفا می‌کنند.

نمونه‌ای از یک گرامر

$$\begin{aligned} G_1 = \quad & A \rightarrow 0A1 \\ & A \rightarrow B \\ & B \rightarrow \# \end{aligned}$$

- ▶ یک گرامر مجموعه‌ای متناهی از یک سری قوانین جایگذاری substitution rules است.
- ▶ هر قانون جایگذاری دو طرف که با یک فلش از هم جدا شده است. طرف چپ (که معمولاً) شامل یک متغیر variable است. در طرف راست رشته‌ای که ترکیبی از متغیرها و ترمینالها terminals می‌باشد قرار می‌گیرد.
- ▶ بطور قراردادی، متغیرها را با حروف بزرگ نشان می‌دهیم. یکی از متغیرها به عنوان متغیر شروع start variable مشخص می‌شود.
- ▶ ترمینالها یا حروف نهایی با حروف کوچک نشان داده می‌شود. اینها همان حروف الفبای زبان هستند.

تولید رشته‌ها از قوانین گرامر

با استفاده از قوانین گرامر می‌توان رشته‌هایی را تولید کرد که زاده آن گرامر هستند (عضوی از زبان آن گرامر هستند). برای اینکار به صورت زیر عمل می‌کنیم.

◀ متغیر شروع را می‌نویسیم.

◀ یکی از قوانین که طرف چپ آن متغیر شروع باشد را انتخاب می‌کنیم و متغیر شروع را با طرف راست قانون عوض می‌کنیم.

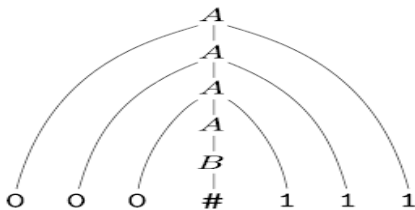
◀ در عبارت حاصل یک متغیر را انتخاب کرده (می‌تواند متغیر انتهای سمت چپ باشد). و قانونی را پیدا می‌کنیم که متغیر مربوطه طرف چپ آن باشد. متغیر را با عبارت طرف راست جایگزین می‌کنیم

◀ این کار را آنقدر ادامه می‌دهیم تا اینکه در عبارت حاصل متغیری باقی نماند.

$$\begin{aligned}
 G_1 = \quad & A \rightarrow 0A1 \\
 & A \rightarrow B \\
 & B \rightarrow \#
 \end{aligned}$$

$$A \Rightarrow 0A1 \Rightarrow 00A11 \Rightarrow 000A111 \Rightarrow 000B111 \Rightarrow 000\#111.$$

به عمل تولید یک رشته توسط قوانین گرامر یک اشتقاق derivation گفته می‌شود. اصطلاحاً می‌گوییم که رشته مورد نظر از آن گرامر مشتق شده است. متناظر با یک اشتقاق یک درخت تجزیه parse tree وجود دارد که در زیر یک نمونه از آن را مشاهده می‌کنید.



زبان یک گرامر

$\langle \text{SENTENCE} \rangle \rightarrow \langle \text{NOUN-PHRASE} \rangle \langle \text{VERB-PHRASE} \rangle$
 $\langle \text{NOUN-PHRASE} \rangle \rightarrow \langle \text{CMPLX-NOUN} \rangle \mid \langle \text{CMPLX-NOUN} \rangle \langle \text{PREP-PHRASE} \rangle$
 $\langle \text{VERB-PHRASE} \rangle \rightarrow \langle \text{CMPLX-VERB} \rangle \mid \langle \text{CMPLX-VERB} \rangle \langle \text{PREP-PHRASE} \rangle$
 $\langle \text{PREP-PHRASE} \rangle \rightarrow \langle \text{PREP} \rangle \langle \text{CMPLX-NOUN} \rangle$
 $\langle \text{CMPLX-NOUN} \rangle \rightarrow \langle \text{ARTICLE} \rangle \langle \text{NOUN} \rangle$
 $\langle \text{CMPLX-VERB} \rangle \rightarrow \langle \text{VERB} \rangle \mid \langle \text{VERB} \rangle \langle \text{NOUN-PHRASE} \rangle$
 $\langle \text{ARTICLE} \rangle \rightarrow \text{a} \mid \text{the}$
 $\langle \text{NOUN} \rangle \rightarrow \text{boy} \mid \text{girl} \mid \text{flower}$
 $\langle \text{VERB} \rangle \rightarrow \text{touches} \mid \text{likes} \mid \text{sees}$
 $\langle \text{PREP} \rangle \rightarrow \text{with}$

به مجموعه همه رشته‌هایی که از گرامر G مشتق می‌شوند زبان گرامر G گفته می‌شود. برای مثال چند رشته که از گرامر بالا مشتق شده است در زیر نشان داده شده است.

a boy sees

the boy sees a flower

a girl with a flower likes the boy

گرامر منظم

در گرامرهای منظم راست هر قانون (قاعده) به یکی از سه شکل

$$\begin{cases} A \rightarrow a \\ A \rightarrow aB \\ A \rightarrow \epsilon \end{cases}$$

می باشد بطوریکه A و B متغیر بوده و a یک ترمینال است.

در زیر یک نمونه از گرامر منظم راست را می بینید.

$$\begin{cases} S \rightarrow aA \mid bS \\ A \rightarrow bS \mid aA \\ A \rightarrow \epsilon \end{cases}$$

$$S \Rightarrow aA \Rightarrow abS \Rightarrow abaA \Rightarrow aba$$

$$S \Rightarrow aA \Rightarrow aaA \Rightarrow aabS \Rightarrow aabaA \Rightarrow aaba$$

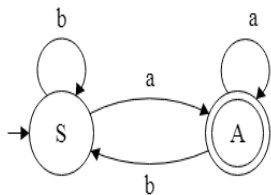
بطور مشابه در گرامرهای منظم چپ هر قانون به یکی از سه شکل زیر می باشد.

$$\begin{cases} A \rightarrow a \\ A \rightarrow Ba \\ A \rightarrow \epsilon \end{cases}$$

در اینجا در عبارات تولید شده، متغیر همواره در انتهای سمت چپ قرار می گیرد.

گرامرهای منظم راست و گرامرهای منظم چپ از لحاظ قدرت توصیف یکسان هستند. زبانهای منظم معادل با گرامرهای منظم راست (چپ) هستند.

قضیه: ماشین متناهی معین M را می توان با گرامر منظم راست (چپ) G بیان کرد بطوریکه $L(G) = L(M)$.



$$\begin{cases} S \rightarrow aA \mid bS \\ A \rightarrow bS \mid aA \\ A \rightarrow \epsilon \end{cases}$$

برای هر وضعیت یک متغیر تعریف می کنیم. متغیر شروع S معادل با وضعیت شروع q_0 است. قانون $A \rightarrow aB$ را قرار می دهیم اگر و فقط اگر از وضعیت متناظر با A یک فلش با برچسب a به وضعیت متناظر با B داشته باشیم. قانون $A \rightarrow \epsilon$ را اضافه می کنیم اگر و فقط اگر A متناظر با یک وضعیت پذیرش باشد.

به طریق مشابهی می توان گرامرهای منظم راست (چپ) را تبدیل به ماشینهای متناهی معین کرد.

زبانهای منظم



ترکیب قوانین چپ و راست

دقت کنید از ترکیب قوانین گرامرهای منظم راست و چپ، گرامرهایی حاصل می‌شود که قدرت بیشتری از زبانهای منظم دارند. برای مثال گرامر زیر معادل با زبان $\{a^n b^n \mid n \geq 0\}$ است که می‌دانیم نامنظم است.

$$\begin{cases} S \rightarrow aA \\ A \rightarrow Sb \\ S \rightarrow \epsilon \end{cases}$$

$$S \Rightarrow aA \Rightarrow aSb \Rightarrow ab$$

$$S \Rightarrow aA \Rightarrow aSb \Rightarrow aaAb \Rightarrow aaSbb \Rightarrow aabb$$

تعریف رسمی یک گرامر مستقل از متن

یک گرامر مستقل از متن با ۴ عنصر مشخص می‌شود.

$$G = (V, \Sigma, R, S)$$

- ◀ V یک مجموعه متناهی از متغیرهاست.
- ◀ Σ مجموعه‌ای متناهی از ترمینالهاست.
- ◀ R یک رابطه از V به $(V \cup \Sigma)^*$ است. در واقع هر عضو R یک قانون (قاعده) است بطوریکه در سمت چپ یک متغیر و در سمت راست قانون دنباله‌ای متناهی است ترمینالها (به اضافه رشته تهی) و متغیرهاست.
- ◀ S عضوی از V است که به آن متغیر شروع (ابتدایی) گفته می‌شود.

یک مثال از گرامر مستقل از متن

$$G_2 = (\{S\}, \Sigma = \{a, b\}, R, S)$$

$$R = \begin{cases} S \rightarrow aS \\ S \rightarrow bS \\ S \rightarrow \epsilon \end{cases}$$

سوال: بنظر شما زبان این گرامر چیست؟ توصیفی ساده برای آن ذکر کنید.

چند نمونه رشته مشتق از گرامر G_3

$a, b, aba, abaabb$

$$L(G_2) = (a + b)^* = \Sigma^*$$

یک مثال از گرامری که مستقل از متن نیست!

$$G = (\{S, B\}, \Sigma = \{a, b, c\}, R, S)$$

$$R = \begin{cases} S \rightarrow aBSc \\ S \rightarrow abc \\ Ba \rightarrow aB \\ Bb \rightarrow bb \end{cases}$$

نام گرامرهای مستقل از متن از این حقیقت آمده که جایگذاری متغیرهای سمت چپ قانون، هر زمان که یکی از آن‌ها در یک عبارت ظاهر شود امکان‌پذیر است. بدین معنی که این جانشینی به دیگر نشانه‌های عبارت (متن) وابسته نیست. این ویژگی از این ناشی می‌شود که تنها یک متغیر در سمت چپ قانون‌های تولید قرار می‌گیرد.

ثابت شده است که بعضی از زبانهای طبیعی مثل (آلمانی، فارسی و ترکی) مستقل از متن نیستند.

یک مثال از گرامر مستقل از متن

$$G_3 = (\{S, A\}, \Sigma = \{a, b\}, R, S)$$

$$S \rightarrow aA$$

$$A \rightarrow aA \mid bA \mid \epsilon$$

سوال: بنظر شما زبان این گرامر چیست؟ توصیفی ساده برای آن ذکر کنید.

چند نمونه رشته مشتق از گرامر G_3

$a, aa, abbbbbb, abaabb$

$$L(G_2) = a\Sigma^*$$

$$G_4 = (\{S\}, \Sigma = \{a, b\}, R, S)$$

$$S \rightarrow aS \mid bS \mid bb$$

سوال: بنظر شما زبان این گرامر چیست؟ توصیفی ساده برای آن ذکر کنید.

چند نمونه رشته مشتق از گرامر G_3

$$bb, aabb, abbbbb, abaabb$$

$$L(G_2) = (a + b)^*bb = \Sigma^*bb$$

$$G_5 = (\{S\}, \Sigma = \{a, b\}, R, S)$$

$$S \rightarrow aS \mid bS \mid Sa \mid Sb \mid bb$$

سوال: بنظر شما زبان این گرامر چیست؟ توصیفی ساده برای آن ذکر کنید.

چند نمونه رشته مشتق از گرامر G_3

$bb, aabb, abbbbb, abaabb$

$$L(G_2) = \Sigma^* bb \Sigma^*$$

$$G_6 = (\{S\}, \Sigma = \{a, b\}, R, S)$$

$$S \rightarrow aSa \mid bSb \mid \epsilon$$

سوال: بنظر شما زبان این گرامر چیست؟ توصیفی ساده برای آن ذکر کنید.

چند نمونه رشته مشتق از گرامر G_3

$bb, bbbb, aa, abbbba, abaabb$

$$L(G_5) = \{ww^R \mid w \in \Sigma^*\}$$

$$G_7 = (\{S\}, \{a, b\}, R, S).$$

The set of rules, R , is

$$S \rightarrow aSb \mid SS \mid \varepsilon.$$

سوال: بنظر شما زبان این گرامر چیست؟ می‌توانید توصیفی ساده برای آن ذکر کنید؟

چند نمونه رشته مشتق از گرامر G_3

$$ab, abab, aabb, abaabb$$

اگر a را پرانتز باز (در نظر بگیریم و b را پرانتز بسته) آنگاه زبان گرامر بالا همه عبارات شامل پرانتز باز و بسته است که پرانتزبندی آن درست است.

تمرین: یک گرامر مستقل از متن برای زبان زیر بنویسید.

$$L = \{x\#y \mid x, y \in \{a\}^*, |x| = |y|\}$$

$$L = \{x\#y \mid x, y \in \{a\}^*, |x| \neq |y|\}$$

تمرین: یک گرامر مستقل از متن برای زبان زیر بنویسید.

$$L = \{w \in \{a, b\}^* \mid n_a(w) = n_b(w)\}$$