دانشکده‌ی علوم ریاضی

نظریه‌ی زبان‌ها و اتوماتا	۱۴ آبان ۱۳۹۱

جلسه‌ی ۱۵ : گرامر مستقل از متن (استنتاج بازگشتی، اشتقاق، درخت تجزیه)

مدرّس: دکتر شهرام خزائی — *نگارنده: آرمیتا اردشیری*

۱ یادآوری

تعریف ۱ *یک گرامر مستقل از متن یک چهارتایی* $G = (V, T, P, S)$ *است، به طوری که:*

V: *مجموعه‌ای متناهی از متغیرهاست*
T: *مجموعه‌ای متناهی از پایانه‌هاست*
P: *مجموعه‌ای متناهی از قوانین تولید به فرم* $A \to \alpha$ *است که در آن* $\alpha \in (T \cup V)^*$ *و* $A \in V$
S: *متغیر شروع است که* $S \in V$

۲ روش های تعریف زبان یک گرامر مستقل از متن

سه روش کلی برای تعریف زبان یک گرامر مستقل از متن وجود دارد:
۱. استنتاج بازگشتی [۱]
۲. اشتقاق مکرر [۲]
۳. درخت تجزیه [۳]

۱.۲ استنتاج بازگشتی

در این روش به هر یک از متغیرها یک زبان نسبت داده می‌شود و با اعمال قوانین گرامر، رشته‌هایی که هر متغیر تولید می‌کند استنتاج می‌شود.

مثال ۱ *گرامر زیر را در نظر بگیرید:*

[۱] recursive inference
[۲] iterative derivation
[۳] parse tree

$$
\begin{array}{lrcl}
۱. & E & \rightarrow & I \\
۲. & E & \rightarrow & E+E \\
۳. & E & \rightarrow & E*E \\
۴. & E & \rightarrow & (E) \\
\\
۵. & I & \rightarrow & a \\
۶. & I & \rightarrow & b \\
۷. & I & \rightarrow & Ia \\
۸. & I & \rightarrow & Ib \\
۹. & I & \rightarrow & I۰ \\
۱۰. & I & \rightarrow & Ib
\end{array}
$$

می‌خواهیم با استنتاج بازگشتی نشان دهیم متغیر E رشته $a*(a+b۰۰)$ را تولید می‌کند.
جدول زیر به صورت خلاصه این استنتاج‌ها را نشان می‌دهد.
برای مثال خط (i) جدول بیان می‌کند که به وسیله قانون ۵ استنتاج می شود متغیر I رشته a را تولید می‌کند.
خطوط (ii) تا (iv) نشان می دهند چگونه رشته $(b۰۰)$ استنتاج می‌شود.
در نهایت خط آخر نشان می دهد متغیر E رشته $a*(a+b۰۰)$ را تولید می‌کند.

تعریف ۲ *زبان یک گرامر مستقل از متن، مجموعه تمام رشته‌هایی است که بتوان با استنتاج بازگشتی نشان داد متغیر شروع آنها را تولید می‌کند.*

۲.۲ اشتقاق

تعریف غیر رسمی: رشته α اشتقاقی از رشته β است که $(\alpha, \beta \in (T \cup V)^*)$ اگر متغیری مانند A در α وجود داشته باشد و قانون تولیدی مانند $A \rightarrow \gamma$ داشته باشیم که با جایگزینی A در α با γ، رشته β بدست آید.

تعریف ۳ *برای یک گرامر مستقل از متن $G=(V,T,P,S)$ که $\alpha, \beta \in (T \cup V)^*$ و $A \rightarrow \gamma \in P$ می‌گوییم $\alpha\gamma\beta$ اشتقاقی از $\alpha A\beta$ است و با $\alpha A\beta \Rightarrow \alpha\gamma\beta$ نشان می‌دهیم.*

تعریف ۴ *می‌گوییم رشته β اشتقاق مکرری از رشته α است و با $\alpha \overset{*}{\Rightarrow} \beta$ نشان می‌دهیم اگر در تعریف استقرایی زیر بگنجد:*
پایه: $\alpha \overset{*}{\Rightarrow} \alpha$
استقرا: اگر $\alpha \overset{*}{\Rightarrow} \beta, \beta \rightarrow \gamma$ آنگاه $\alpha \overset{*}{\Rightarrow} \gamma$

مثال ۲ *برای گرامر مثال قبل نشان دهید چگونه رشته $a*(a+b۰۰)$ از متغیر E اشتقاق می‌شود.*
در هر مرحله باید یکی از متغیرها را با یک بدنه قانون به طور صحیح جایگزین کنیم تا به رشته مورد نظر برسیم.

$$E \Rightarrow E*E \Rightarrow E*(E) \Rightarrow E*(E+E) \Rightarrow E*(I+E) \Rightarrow E*(a+E) \Rightarrow \ldots \Rightarrow a*(a+b۰۰)$$

تعریف ۵ *برای گرامر مستقل از متن $G=(V,T,P,S)$ زبان آن را با $L(G)$ نشان می‌دهیم و تعریف می‌کنیم:*

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

مثال ۳ نشان دهید گرامر $G=(\{S\},\{۰,۱\},P,S)$ با مجموعه قواعد تولید زیر زبان $L=\{w\in\{۰,۱\}^*|w=w^R\}$ را تولید می‌کند.

$$
\begin{aligned}
S &\rightarrow ۰\\
S &\rightarrow ۱\\
S &\rightarrow \epsilon\\
S &\rightarrow ۰S۰\\
S &\rightarrow ۱S۱
\end{aligned}
$$

گرامر فوق را می‌توان به صورت زیر به طور خلاصه نشان داد:

$$S\rightarrow \epsilon|۰|۱|۰S۰|۱S۱$$

برهان.
۱. ابتدا ثابت می‌کنیم تمام رشته‌های w که به فرم $w=w^R$ می‌باشند توسط این گرامر پذیرفته می‌شوند، یعنی:

$L\subseteq L(G)$
$w=w^R\rightarrow w\in L(G)$

برای این کار از استقرا روی طول w استفاده می‌کنیم.
پایه:حالت پایه زمانی است که طول رشته صفر یا یک باشد و در این دو حالت ثابت می‌کنیم اگر رشته با معکوسش برابر باشد حتما متعلق به زبان این گرامر است.
اگر طول رشته صفر باشد نتیجه می‌گیریم w برابر با ϵ است. و شرط $w=w^R$ برقرار است. با توجه به قانون $(S\rightarrow\epsilon)$ می‌دانیم ϵ اشتقاق مکرری از S است، پس متعلق به زبان این گرامر است.
اگر طول رشته یک باشد $w=۰$ یا $w=۱$ است، که برای هر دو حالت شرط $w=w^R$ برقرار است. با توجه به قوانین تولید $(S\rightarrow ۰)$ و $(S\rightarrow ۱)$ مشخص می‌شود این دو رشته متعلق به زبان این گرامر هستند.

استقرا:
اگر برای تمام رشته‌های به طول n فرض فوق برقرار باشد ثابت می‌کنیم برای رشته‌های به طول $n+۲$ هم فرض صادق است.
رشته $w=w^R$ به طول $n+۲$ را در نظر می‌گیریم، پس حرف ابتدایی و انتهایی در این رشته با هم برابر است.
بنابراین این رشته به یکی از دو فرم $۰x۰$ یا $۱x۱$ می‌باشد، که: $x=x^R$
چون طول رشته x برابر با n می‌باشد، طبق فرض استقرا می‌دانیم: $x\in L(G)$
چون x متعلق به $L(G)$ است، یعنی: $S\overset{*}{\Rightarrow}x$
پس با استفاده از یکی از دو قانون تولید زیر رشته w را می‌توان ساخت:

$S\rightarrow ۰S۰, S\overset{*}{\Rightarrow}x:$
$S\Rightarrow ۰S۰\overset{*}{\Rightarrow}۰x۰$

یا

$S\rightarrow ۱S۱, S\overset{*}{\Rightarrow}x:$
$S\Rightarrow ۱S۱\overset{*}{\Rightarrow}۱x۱$

۲.

$L(G) \subseteq L$
$w \in L(G) \rightarrow w = w^R$

حال باید اثبات کنیم هر رشته w که توسط این گرامر تولید می‌شود دارای شرط $w = w^R$ است. یعنی:

$(S \overset{*}{\Rightarrow} w) \rightarrow (w = w^R)$

به استقرا روی تعداد مراحل اشتقاق نشان می‌دهیم این فرض برقرار است.

پایه: حالت پایه زمانی است که طول اشتقاق برابر یک باشد. یعنی $S \Rightarrow w$. چون w یک اشتقاق یک مرحله‌ای از S است نتیجه می‌گیریم $S \rightarrow w$ متعلق به مجموعه قوانین تولید است. در مجموعه قوانین تولید سه قانون به این فرم هستند که عبارتند از:

$S \rightarrow \epsilon$
$S \rightarrow ۰$
$S \rightarrow ۱$

پس w برابر ۰ یا ۱ یا ϵ است. که در هر سه حالت شرط $w = w^R$ برقرار است. و حالت پایه اثبات می‌شود.
استقرا: اگر w یک اشتقاق n مرحله‌ای از S باشد فرض می‌کنیم شرط $w = w^R$ برقرار است (فرض استقرا). باید اثبات کنیم اگر w اشتقاق $n + ۱$ مرحله‌ای از S باشد هم شرط فوق برقرار است.
فرض کنید w یک اشتقاق $n + ۱$ مرحله‌ای از S باشد،مرحله اول اشتقاق نمی‌تواند از یکی از سه قانون اول تبعیت کند و باید به فرم یکی از دو قانون آخر باشد.
پس یکی از دو حالت زیر برقرار است:

$a) S \Rightarrow ۰S۰ \overset{*}{\Rightarrow} w$
$b) S \Rightarrow ۱S۱ \overset{*}{\Rightarrow} w$

در حالت اول $w = ۰x۰$ می‌باشد، که x یک اشتقاق n مرحله‌ای از S است وطبق فرض استقرا $x = x^R$ و در نتیجه:

$w = w^R$

برای حالت دوم $w = ۱x۱$ می‌باشد، که x یک اشتقاق n مرحله‌ای از S است وطبق فرض استقرا $x = x^R$ و در نتیجه:

$w = w^R$

■

تعریف ۶ اشتقاق چپ ترین:[۴]
اگر $G = (V, T, P, S)$ یک گرامر باشد و $w \in T^*$ و $(A \rightarrow \beta) \in P$, $\alpha \in (V \cup T)^*$ در این صورت می‌گوییم $w\beta\alpha$ یک اشتقاق چپ ترین از $wA\alpha$ است و با نماد $\underset{lm}{\Rightarrow}$ نشان می‌دهیم.

[۴] leftmost derivation

تعریف ۷ *اشتقاق راست ترین:*[5]
اگر $G = (V, T, P, S)$ *یک گرامر باشد و* $w \in T^*$ *و* $\alpha \in (V \cup T)^*, (A \to \beta) \in P$ *در این صورت می‌گوییم* $\alpha\beta w$ *یک اشتقاق راست ترین از* $\alpha A w$ *است و با نماد* $\underset{rm}{\Rightarrow}$ *نشان می‌دهیم.*

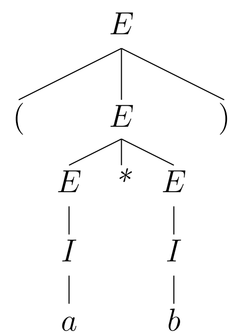
زبان یک گرامر را می‌توان با استفاده از اشتقاق‌های راست‌ترین و چپ‌ترین نیز تعریف کرد.

۳.۲ درخت تجزیه

فرض کنید $G = (V, T, P, S)$ یک گرامر مستقل از متن باشد، یک درخت مرتب یک درخت تجزیه برای گرامر G است اگر خصوصیات زیر را داشته باشد:

۱. ریشه درخت دارای برچسب S است.
۲. هر برگ دارای برچسبی از $T \cup \{\epsilon\}$ است.
۳. هر گره میانی دارای برچسبی از V است.
۴. اگر یک گره درونی برچسب $A \in V$ داشته باشد و فرزندانش از چپ به راست دارای برچسب های $X_1 X_2 \dots X_t$ باشند،
$A \to X_1 X_2 \dots X_t$ یکی از قوانین تولید باشد.
۵. گره ای که دارای برچسب ϵ است دارای هیچ همزادی نمی باشد.

مثال ۴ *درخت تجزیه زیر را در نظر بگیرید:*



این درخت تجزیه بیانگر اشتقاق مکرر $(a + b)$ *از متغیر* E *به صورت زیر است:*

$$E \Rightarrow (E) \Rightarrow (E * E) \Rightarrow (E * I) \Rightarrow (I + I) \Rightarrow (a * I) \Rightarrow (a + b)$$

در جلسه بعد تعریف دیگری برای زبان یک گرامر مستقل از متن با استفاده از تعریف درخت تجزیه ارائه می‌دهیم و ثابت می‌کنیم این تعریف با تعریف‌های ۲ و ۵ معادل است.

[5] rightmost derivation

	string inferred	for language of	production used	string(s) used
(i)	a	I	۵	$-$
(ii)	b	I	۶	$-$
(iii)	$b\circ$	I	۹	(ii)
(iv)	$b\circ\circ$	I	۹	(iii)
(v)	a	E	۱	(i)
(vi)	$b\circ\circ$	E	۱	(iv)
(vii)	$a+b\circ\circ$	E	۲	$(v), (vi)$
$(viii)$	$(a+b\circ\circ)$	E	۳	(vii)
(ix)	$a*(a+b\circ\circ)$	E	۴	$(v), (viii)$