

فرادرس

فرا تراژیک کلاس درس
www.faradars.org

اصول طراحی کامپایلرها

مدرس:

منوچهر بابایی

کارشناس ارشد مهندسی کامپیوتر (نرم افزار)

دانشگاه کاشان

درس سوم:

گرامرها

FaraDars.org

گرامر

- ویژگی نحوی ساختارهای زبان های برنامه نویسی با استفاده از یک گرامر مستقل از متن توصیف می شود.
- گرامر یک نمایش دقیق و قابل فهم از زبان برنامه نویسی را ارائه می کند.
- یک گرامر به زبان این امکان را می دهد که به صورت تکرارپذیر تکامل یافته یا توسعه پیدا کند. این عمل با افزودن ساختارهای جدید برای انجام اعمال جدید انجام می شود.

فرم کلی گرامرهای مستقل از متن

$$S \Rightarrow x \quad S \in V \quad x \in (V \cup T)^*$$

$$E \rightarrow E + T \mid T$$

$$T \rightarrow T * F \mid F$$

$$F \rightarrow (E) \mid id$$

مثال: گرامری مستقل از متن برای عبارات محاسباتی اولیه

$\text{expr} \rightarrow \text{expr} + \text{term}$

$\text{expr} \rightarrow \text{expr} - \text{term}$

$\text{expr} \rightarrow \text{term}$

$\text{term} \rightarrow \text{term} * \text{factor}$

$\text{term} \rightarrow \text{term} / \text{factor}$

$\text{term} \rightarrow \text{factor}$

$\text{factor} \rightarrow (\text{expr})$

$\text{factor} \rightarrow \text{id}$

$\text{id} + * \text{id}$

$1 + * 2$ ✗

$1 + 2 * 3$ ✓

○ برای سادگی گرامرها معمولاً از قواعد زیر استفاده می شود:

۱- نمادهای زیر پایانه هستند:

-حروف کوچک مانند a, b, c, \dots

-نمادهای عملگر مانند جمع و تفریق....

-نمادهای پرانتز، کاما...

-ارقام

۲- نمادهای زیرغیرپایانه هستند:

-حروف بزرگ (در سمت چپ قوانین)

-اسامی حروف کوچک

$$A \rightarrow \alpha_1$$

*درگرامر قواعد $A \rightarrow \alpha_p$ را می توان به صورت $A \rightarrow \alpha_1 | \alpha_p | \alpha_{p+1} | \dots$ نوشت

$$A \rightarrow \dots$$

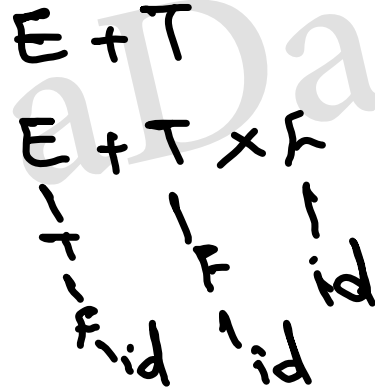
مثال: ساده شده گرامر مثال قبل

$$E \rightarrow E + T \mid E - T \mid T$$

$$T \rightarrow T * F \mid T / F \mid F$$

$$F \rightarrow (E) \mid id$$

$id + id * id$



اشتقاق (Derivation)

- فرایند تولید یک رشته از نماد شروع را اشتقاق گوییم.
- تولید یک رشته به وسیله گرامر مستقل از متن از نماد شروع آغاز می شود.
- در هر مرحله یک غیر پایانه به وسیله سمت راست قاعده تولید آن گسترش داده می شود.
- روند تولید تا زمانی ادامه می یابد، که هیچ غیرپایانه ای باقی نماند.

دو نوع اشتقاق وجود دارد:

۱- اشتقاق راست (RMD)

۲- اشتقاق چپ (LMD)

اشتقاق راست: در این اشتقاق همواره سمت راست ترین غیر پایانه، با سمت راست، قاعد تولید آن جایگزین می شود.

اشتقاق چپ: در این اشتقاق همواره سمت چپ ترین غیر پایانه، با سمت راست، قاعده تولید آن جایگزین می شود.

اشتقاق چپ

id+id

E+T

T+T

F+T

id+T

id+F

Id+id

اشتقاق راست

id+id

E+T

E+F

E+id

T+id

F+id

Id+id

$$E \rightarrow E + T \mid E - T \mid T$$

$$T \rightarrow T * F \mid T / F \mid F$$

$$F \rightarrow (E) \mid id$$

درخت تجزیه: یک نمایش گرافیکی برای اشتقاق می باشد. در واقع نحوه جایگزینی غیرپایانه ها با سمت راست قاعده تولید را نشان می دهد.



نحوه ساختن درخت تجزیه بصورت زیر است:

(الف) در ریشه درخت تجزیه، نماد شروع گرامر می آید.

(ب) در برگ های درخت تجزیه، پایانه ها یا لاندا می آید.

(ج) هر گره میانی درخت تجزیه یک غیرپایانه است.

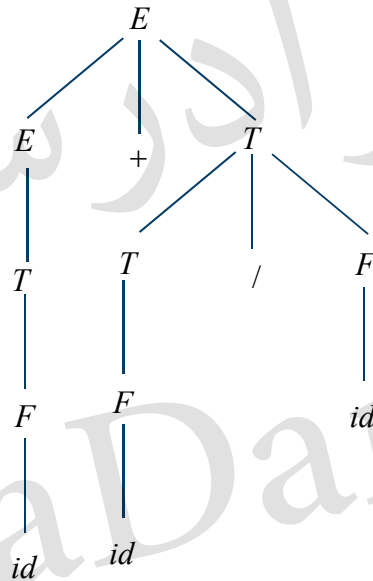
(د) اگر A یک غیرپایانه و گره میانی درخت تجزیه باشد که x_1, x_2, \dots, x_n فرزندان این گره هستند، آنگاه در گرامر باید قاعده تولیدی به صورت $A \rightarrow x_1 x_2 x_3 \dots x_n$ وجود داشته باشد.

$$E \rightarrow E + T \mid E - T \mid T$$

$$T \rightarrow T * F \mid T / F \mid F$$

$$F \rightarrow (E) \mid id$$

id+id/id



*در اشتقاق ترتیب به کارگیری قواعد مشخص است اما در درخت تجزیه این موضوع مشخص نیست.

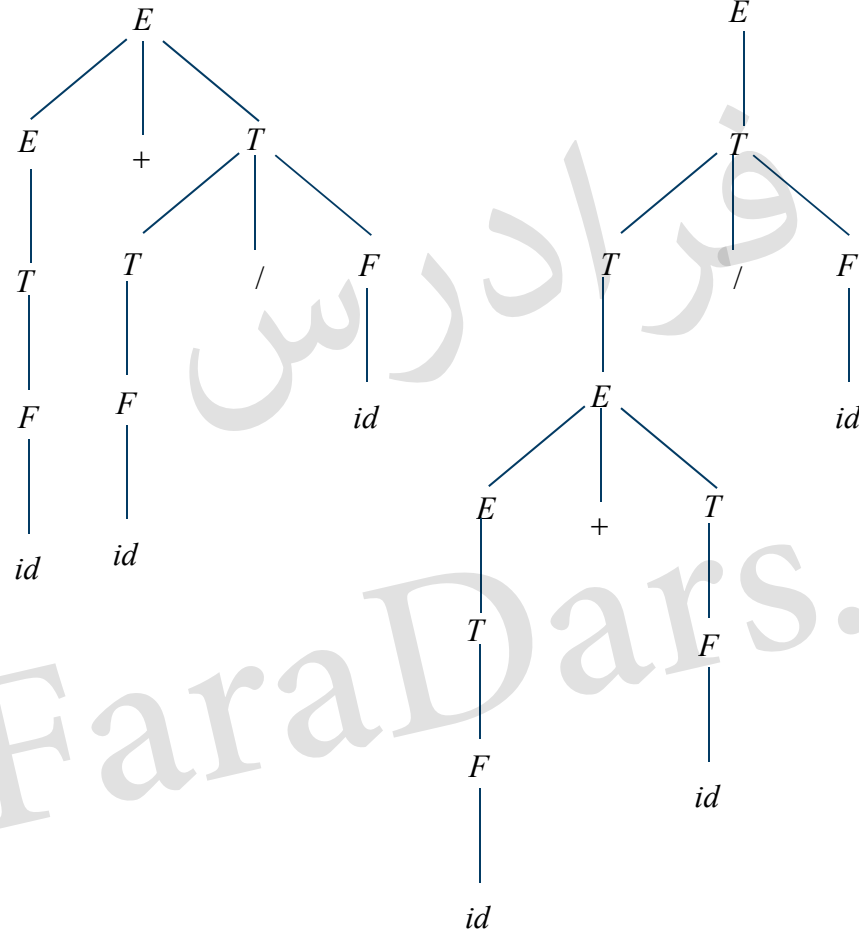
گرامر مبهم: در صورتی که در یک گرامر برای یک رشته بتوان دو اشتقاق چپ یا دو اشتقاق راست و یا دو درخت تجزیه متفاوت ایجاد نمود، آنگاه گرامر مبهم است.

$$E \rightarrow E + T \mid E - T \mid T$$

$$T \rightarrow T * F \mid T / F \mid F \mid E$$

$$F \rightarrow (E) \mid id$$

id+id/id



$$\begin{aligned} E &\rightarrow E + T \mid E - T \mid T \\ T &\rightarrow T * F \mid T / F \mid F \mid E \\ F &\rightarrow (E) \mid id \end{aligned}$$

تمرین کتاب AHO

مثال: گرامر مستقل از متن زیر و رشته $aa+a^*$ را در نظر بگیرید.

$$S \rightarrow SS + | SS^* | a$$

الف) یک سمت چپ ترین و سمت راست ترین اشتقاق برای رشته بنویسید

ب) یک درخت تجزیه برای رشته رسم کنید.

ج) آیا گرامر مبهم است؟
مبهم نیست

د) زبان توصیف شده توسط گرامر چیست؟

(الف)

$$S \rightarrow SS + | SS * | a$$

$$aa + a*$$

استقراء راسه

$$SS *$$

$$Sa *$$

$$SS + a *$$

$$Sa + a *$$

$$aa + a *$$

استقراء جیب

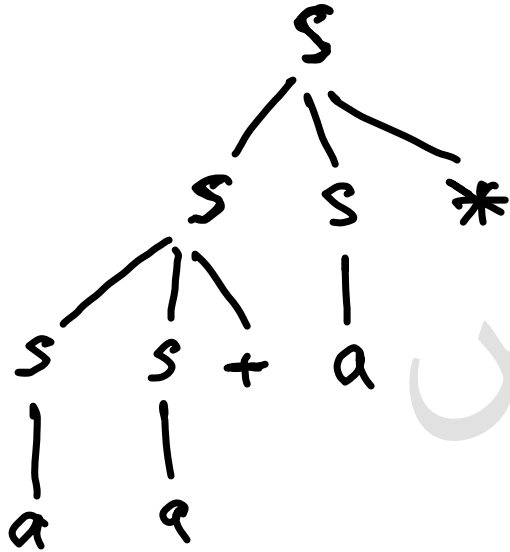
$$SS *$$

$$SS + S *$$

$$aS + S *$$

$$aa + S *$$

$$aa + a *$$



FaraDars.org

(> چهار دسته مختلف تولید شده توسط گرامر :

$$S \rightarrow SS + \mid SS * \mid a$$

a

SS+

aa+

SSS++

aa*

SSS*S++

aaaa*a++

→ a+a*a+a

این ترم‌های عبارات ریاضی بوم پیوندی با هم ترکیبی +، * را تولید می‌کند.

تمرین کتاب AHO

مثال: گرامر مستقل از متن زیر و رشته 000111 را در نظر بگیرید.

$$S \rightarrow \bullet S 1 \mid \bullet 1$$

الف) یک سمت چپ ترین و سمت راست ترین اشتقاق برای رشته بنویسید

ب) یک درخت تجزیه برای رشته رسم کنید.

ج) آیا گرامر مبهم است؟ **بهم نیست**

$$n \geq 1 \mid 0^n 1^n$$

د) زبان توصیف شده توسط گرامر چیست؟

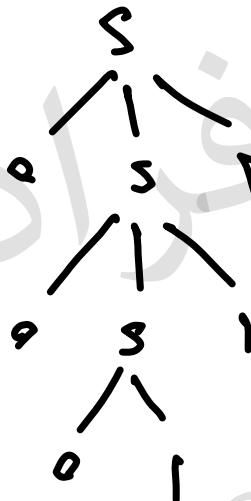
(الف)

031

4.511

000 111

(ب)



تمرین کتاب AHO

مثال: گرامر مستقل از متن زیر و رشته $a(a+a)^*a$ را در نظر بگیرید.

$$S \rightarrow S + S \mid (S) \mid S^* \mid a$$

الف) یک سمت چپ ترین و سمت راست ترین اشتقاق برای رشته بنویسید

ب) یک درخت تجزیه برای رشته رسم کنید.

ج) آیا گرامر مبهم است؟

د) زبان توصیف شده توسط گرامر چیست؟

$$S \rightarrow S + S \mid SS \mid (S) \mid S^* \mid a$$

$$(a + a) * a$$

الف)

SS استعاره راسه

SS

Sa

S*S

S*a

(S)*S

(S)*a

(S+S)*S

(S+S)*a

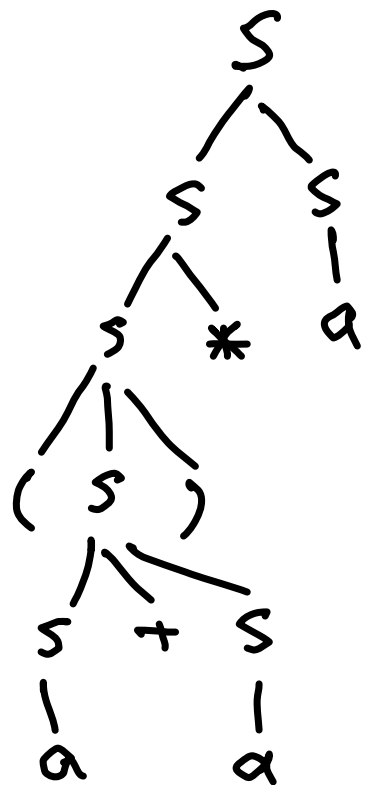
(a+S)*S

(S+a)*a

(a+a)*S

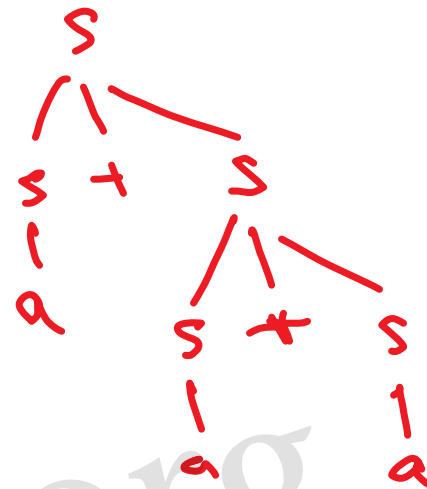
(a+a)*a

(a+a)*a



(ب)

$a + a + a$



گرامر مبهم است.

مثال: برای هریک از زبان های زیر یک گرامر بنویسید

الف) مجموعه تمام رشته ها از صفر و یک که تعداد یک ها و صفرها یکسان نیست.

ب) مجموعه تمام رشته ها که تعداد a ها دو برابر تعداد b ها است.

FaraDars.org

الف)

$$S \rightarrow p \mid q$$

$$p \rightarrow 0p1 \mid 1p0 \mid pp \mid Ap \mid 0$$

$$A \rightarrow 0A \mid \lambda$$

$$q \rightarrow 0q1 \mid 1q0 \mid qq \mid Bq \mid 1$$

$$B \rightarrow 1B \mid \lambda$$

$$P: n_0 > n_1$$

$$q: n_0 < n_1$$

(ب)

a a b
a b a
b a a

$S \rightarrow Saab | asab | aasb | aabs$

$S \rightarrow Saba | asba | absa | abas$

$S \rightarrow Sbaa | bsaa | basa | baas$

$S \rightarrow ss | \lambda$

FaraDars.org

مثال: آیا گرامرهای زیر مبهم هستند؟

$$S \rightarrow AB \mid aaB$$

$$A \rightarrow a \mid Aa$$

$$B \rightarrow b$$

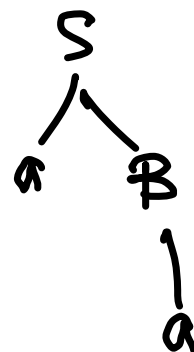
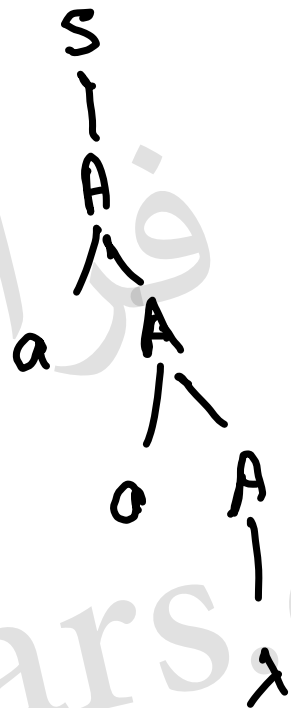
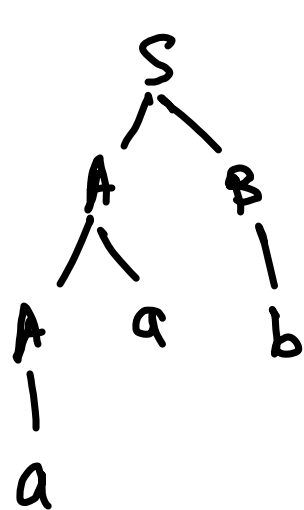
گرامر مبهم است: aab

$$S \rightarrow A \mid aB$$

$$A \rightarrow aA \mid \lambda$$

$$B \rightarrow bB \mid a$$

گرامر مبهم است: aa



دو درخت تجزیه برای aab

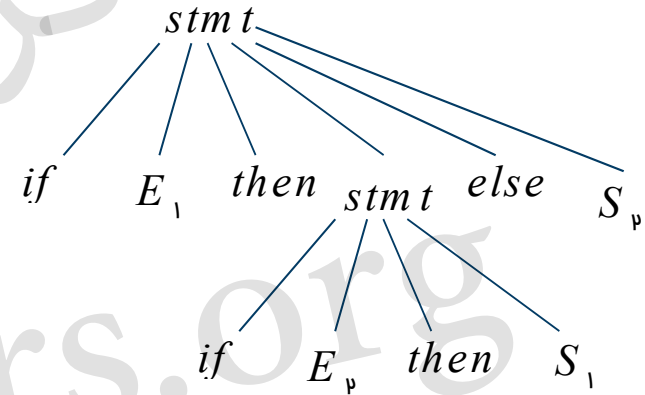
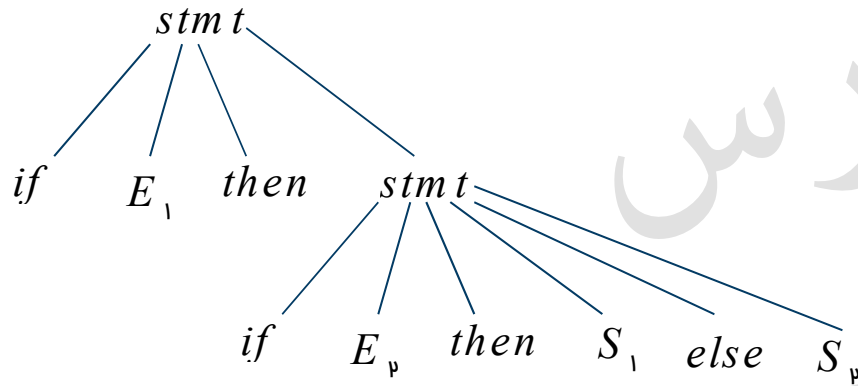
دو درخت تجزیه برای aa

روش های رفع ابهام:

۱- تغییر گرامر: هر گرامر یک زبان را توصیف می کند. اما برای هر زبان گرامرهای متعددی می توان نوشت، لذا با تغییر گرامر می توان ابهام آن را رفع نمود.

Stmt → **if** *expr* **then** *Stmt*
| **if** *expr* **then** *Stmt* **else** *Stmt* | **other**

if **E1** **then** **if** **E2** **then** **S1** **else** **S2**



گرامر غیر مبهم معادل:

Stmt \rightarrow matched|unmatched

Matched \rightarrow **if** expr **then** matched **else** matched|other

Unmatched \rightarrow **if** expr **then** stmt

| **if** expr **then** matched **else** matched|other

روش های رفع ابهام...

۲- استفاده از قوانین جانبی: بدون تغییر گرامر، قوانینی برای انتخاب یک درخت تجزیه از بین درخت های تجزیه استفاده می کنیم. در این صورت گرامر برای هر رشته مربوط زبان فقط یک درخت تجزیه خواهد داشت و مثل این که مبهم نیست.

الف) استفاده از اولویت عملگرها

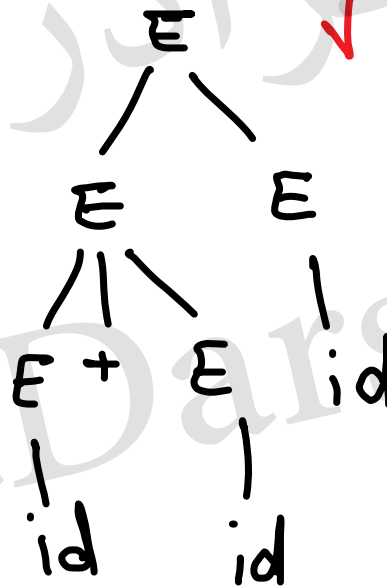
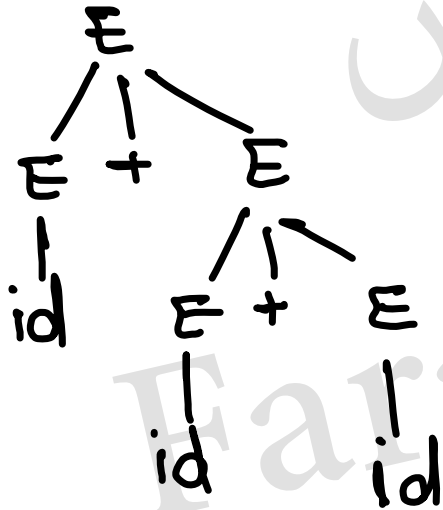
ب) استفاده از شرکت پذیری

یا مثلاً در گرامر مبهم مربوط به if، می توان گفت که هر else به نزدیک ترین if مربوط می شود.

$E \rightarrow E + E$

$E \rightarrow id$


$id + id + id$



استفاده از شرط نبوی
حیث برای همگرایی

بازگشتی چپ:

اگر سمت راست قاعده تولید با غیرپایانه سمت چپ قاعده تولید شروع شود، در این صورت گرامر دارای بازگشتی چپ است.

$$A \rightarrow A\alpha \quad \alpha \in (V \cup T)^*$$


انواع بازگشتی چپ: ۱- بازگشتی چپ مستقیم

$$A \xrightarrow{*} A\alpha \quad \alpha \in (V \cup T)^*$$

۲- بازگشتی چپ غیرمستقیم

$$S \rightarrow AB \rightarrow S \rightarrow SB$$

$$A \rightarrow S$$

$$\alpha_i \in (V \cup T)^*$$

با A شروع می‌شوند: $\beta_i \in (V \cup T)^*$

حذف بازگشتی چپ مستقیم:

$$A \rightarrow A\alpha_1 | A\alpha_p | \dots | A\alpha_m | \beta_1 | \beta_p | \dots | \beta_n$$

قواعد به دو دسته تقسیم می‌شوند:

$$1- \quad A \rightarrow A\alpha_i \quad i = 1 \dots m$$

$$2- \quad A \rightarrow \beta_i \quad i = 1 \dots m, \text{ که } \beta_i \text{ با } A \text{ شروع نمی‌شوند}$$

$$A \rightarrow \beta_1 A' | \beta_p A' | \dots | \beta_n A'$$

$$A' \rightarrow \alpha_1 A' | \alpha_p A' | \dots | \alpha_m A' | \lambda$$

$$\begin{array}{l} A \rightarrow A\overset{\beta_1}{b} | a | \overset{\beta_2}{b} \\ \downarrow \\ A \rightarrow aA' | bA' \\ A' \rightarrow bA' | \lambda \end{array}$$

حذف بازگشتی چپ غیرمستقیم:

for (each i from 1 to n) {

for (each j from 1 to $i - 1$) {

به جای هر قاعده $A_i \rightarrow A_j \alpha$ قواعد $A_i \rightarrow \delta_1 \alpha \mid \delta_p \alpha \mid \delta_s \alpha \mid \dots$

$A_j = \delta_1 \mid \delta_p \mid \delta_s \mid \dots$

}

بازگشتی چپ مستقیم را حذف می کنیم.

}

غیر پایانه S, A
1 2

$$S \rightarrow Ab \mid b$$

$$A \rightarrow Sb \mid Aa$$

$$\Rightarrow A \rightarrow Abb \mid bb \mid Aa$$

$$A \rightarrow A\alpha_1 \mid A\alpha_2 \mid \beta$$



$$S \rightarrow Ab \mid b$$

$$A \rightarrow bbA'$$

$$A' \rightarrow aA' \mid bbA' \mid \lambda$$

$$\begin{aligned} S &\rightarrow Ab \mid b \\ A &\rightarrow Sb \mid Aa \end{aligned} \Rightarrow$$

$$\begin{aligned} S &\rightarrow Ab \mid b \\ A &\rightarrow SbA' \\ A' &\rightarrow aA' \mid \lambda \end{aligned} \quad (1)$$

$$\begin{aligned} S &\rightarrow SbA'b \mid b \\ A' &\rightarrow aA' \mid \lambda \end{aligned} \quad (2)$$

↓↓↓ ↓↓↓

$$\begin{aligned} S &\rightarrow bS' \\ S' &\rightarrow bA'bS' \mid \lambda \\ A' &\rightarrow aA' \mid \lambda \end{aligned}$$

۱- ابتدا بازگشتی جدیدی را حذف کردیم
۲- پس بازگشتی جدیدی را حذف کردیم

$$\begin{array}{l}
 \xrightarrow{\alpha} \\
 E \rightarrow E + T \mid T \\
 \xrightarrow{\beta} \\
 T \rightarrow T * F \mid F
 \end{array}$$

$$F \rightarrow (E) \mid id$$

↓↓↓↓

$$E \rightarrow TE'$$

$$E' \rightarrow +TE' \mid \lambda$$

$$T \rightarrow FT'$$

$$T' \rightarrow *FT' \mid \lambda$$

$$F \rightarrow (E) \mid id$$

$$N \rightarrow AT$$

$$A \rightarrow Na \mid a$$

$$T \rightarrow Ta \mid b$$

$$\Downarrow \quad \Downarrow \quad \Downarrow \quad \underline{\alpha} \quad \underline{\beta}$$

$$N \rightarrow NaT \mid AT$$

$$A \rightarrow a$$

$$T \rightarrow Ta \mid b$$

$$\underline{\alpha} \quad \underline{\beta}$$

$$\Rightarrow N \rightarrow A TN'$$

$$N' \rightarrow aTN' \mid \lambda$$

$$A \rightarrow a$$

$$T \rightarrow bT'$$

$$T' \rightarrow aT' \mid \lambda$$

$$N \rightarrow \beta N'$$

$$N' \rightarrow \alpha N' \mid \lambda$$

FaraDars.org

$$N \rightarrow BNd \mid a$$

$$B \rightarrow b \mid \lambda$$

$$\Downarrow \Downarrow \Downarrow \alpha \beta_1 \beta_2$$

$$N \rightarrow Nd \mid BNd \mid a$$

$$B \rightarrow b$$

$$\Rightarrow N \rightarrow BNdN' \mid aN'$$

$$N' \rightarrow dN' \mid \lambda$$

$$B \rightarrow b$$

FaraDars.org

مثال: بازگشتی چپ گرامرهای زیر را رفع کنید؟

$$A \rightarrow Aa \mid aBc \mid \lambda$$

$\alpha_1 \quad \beta_1 \quad \beta_2$

$$B \rightarrow Bb \mid bc$$

$\alpha_2 \quad \beta_3$

$$S \rightarrow Aa \mid b$$

$$A \rightarrow Ac \mid Aad \mid bd \mid \lambda$$

$\alpha_1 \quad \alpha_2 \quad \beta_1 \quad \beta_2$

$$A \rightarrow aBcA' \mid A'$$

$$A' \rightarrow aA' \mid \lambda$$

$$B \rightarrow bcB'$$

$$B' \rightarrow bB' \mid \lambda$$

$$S \rightarrow Aa \mid b$$

$$A \rightarrow bA' \mid A'$$

$$A' \rightarrow cA' \mid adA' \mid \lambda$$

فاکتورگیری چپ:

برای یک غیرپایانه ممکن است، انتخاب های مختلفی از یک قاعده تولید وجود داشته باشد. لذا برای راحت کردن تجزیه پیشگو، از فاکتورگیری چپ استفاده می کنیم.

$$A \rightarrow \alpha\beta_1 \mid \alpha\beta_p$$

$$A \rightarrow aA \mid a$$

فاکتورگیری چپ:

$$A \rightarrow \alpha\beta_1 \mid \alpha\beta_p \mid \dots \mid \alpha\beta_n \mid \delta$$

$\Downarrow \Downarrow \Downarrow$

$$A \rightarrow \alpha A' \mid \delta$$

$$A' \rightarrow \beta_1 \mid \beta_p \mid \dots \mid \beta_n$$

$$A \rightarrow \alpha A \mid \alpha \implies \begin{array}{l} A \rightarrow \alpha A' \\ A' \rightarrow A \mid \lambda \end{array}$$

مثال: فاکتورگیری چپ را برای گرامر زیر انجام دهید؟

$$S \rightarrow aABb \mid aBC$$

$$A \rightarrow aBB \mid aB$$

$$B \rightarrow bB \mid b$$

$$C \rightarrow cC \mid c$$

$$S \rightarrow aS'$$

$$C \rightarrow cC'$$

$$S' \rightarrow ABb \mid BC$$

$$C' \rightarrow C \mid \lambda$$

$$A \rightarrow aBA'$$

$$A' \rightarrow B \mid \lambda$$

$$B \rightarrow bB'$$

$$B' \rightarrow B \mid \lambda$$

این اسلاید ها بر مبنای نکات مطرح شده در فرادرس
«آموزش طراحی کامپایلر»
تهیه شده است.

برای کسب اطلاعات بیشتر در مورد این آموزش به لینک زیر مراجعه نمایید
faradars.org/fvsft104