

بسم الله الرحمن الرحيم

کلاس نکته و تست
نظریه زبان ها و ماشین ها

ابوالفضل گیلک

جلسه دوازدهم – فصل ششم

زبان های مستقل از متن

حل تست های 15 سال اخیر

ارشد – دکتری

مهندسی کامپیوتر – علوم کامپیوتر

فصل 6

ساده سازی گرامرهای مستقل از متن

تست های 1 الی 17
(کل تست های فصل 6)

1- در گرامر $G = (\Sigma, V, S, P)$ تمام قواعد آن به صورت $A \rightarrow a\alpha$ است که $A \in V, a \in \Sigma$ و $\alpha \in \{\Sigma \cup V\}^*$ کدام گزینه صحیح است؟
(علوم کامپیوتر - 95)

(1) زبان‌های این نوع گرامرها زیرمجموعه اکیدی از زبان‌های مستقل از متن هستند.

(2) اگر $|\alpha| > 1$ آن‌گاه زبان گرامر G منظم نیست.

(3) زبان $L = \{wa : w \in \Sigma^*, a \in \Sigma\}$ دارای گرامری به شکل G نیست.

(4) اگر $|\alpha| \leq 2$ و α شامل حداکثر یک متغیر باشد، آن‌گاه زبان G منظم است.

2- زبان ناتهی و مستقل از متن L با گرامر $G = (V, \Sigma, S, P)$ به فرم نرمال چامسکی توصیف می‌شود که در آن مجموعه متغیرها (سمبل‌های غیرپایانی) و P مجموعه قوانین است. آن‌گاه بهترین کران بالا برای طول کوتاه‌ترین عضو L کدام است؟
(علوم کامپیوتر - 93)

$$2^{|P|+|V|} \quad (4)$$

$$2^{|V|+1} \quad (3)$$

$$2^{|P|+1} \quad (2)$$

$$2^{|P|-1} \quad (1)$$

3- گرامر زیر را در نظر بگیرید:

$$G : \begin{cases} S \rightarrow \lambda \\ S \rightarrow 0S110 \\ S \rightarrow SX \\ X \rightarrow 011X \\ X \rightarrow \lambda \end{cases}$$

(علوم کامپیوتر - 93)

کدام گزینه صحیح است؟

(1) در هر گرامر G' معادل G متغیر Y که متغیر شروع نیست وجود دارد طوری که $Y \rightarrow \lambda$.

(2) در هر گرامر G' معادل G با متغیر شروع S' باید یک قانون به صورت $S' \rightarrow \lambda$ موجود باشد.

(3) گرامر $G' = (V, \{0,1\}, S', P)$ معادل G وجود دارد که در آن برای هیچ متغیری مثل X ، نمی‌توان داشت $X \Rightarrow X\alpha$ که $\alpha \in (V \cup \{0,1\})^*$.

(4) در هر گرامر G' معادل G با متغیر شروع S' حتماً می‌توان داشت $S' \Rightarrow S'\alpha$ شامل سمبل‌های پایانی و غیرپایانی است.

4- گرامر $G = \langle S, V = \{S, X, Y\}, R, \{0,1\} \rangle$ را با قوانین تولید زیر در نظر بگیرید:

$$R : \begin{cases} S \rightarrow 0X1 \mid \lambda \\ 0X \rightarrow Y \\ Y \rightarrow \lambda \mid Y1 \end{cases}$$

(علوم کامپیوتر - 91)

کدام گزینه درست است؟

(1) G یک گرامر مستقل از متن است.

(2) G به فرم نرمال چامسکی است.

(3) یک λ -NFA وجود دارد که زبان آن با زبان G برابر است.

(4) هر گرامر دلخواه برای بیان زبان گرامر G ، دارای حداقل 2 متغیر و 2 قانون است.

5- گرامر زیر را در نظر بگیرید:

$$\begin{aligned} E &\rightarrow T \mid E + T \\ I &\rightarrow a \mid b \mid Ia \mid Ib \mid IO \mid II \\ F &\rightarrow I \mid (E) \\ T &\rightarrow F \mid T * F \end{aligned}$$

(علوم کامپیوتر - 89)

کدام گزینه صحیح است؟

(1) این گرامر به فرم نرمال چامسکی است.

(2) عبارت $a + a * a$ در زبان این گرامر نیست.

(3) عبارت $a + a * a$ فقط به یک صورت توسط این گرامر تولید می‌شود.

(4) کلماتی وجود دارند که به بیش از یک صورت توسط این گرامر تولید می‌شوند.

6- کدام گزینه در مورد گرامر زیر درست است؟ (علوم کامپیوتر - 88)

$S \rightarrow AB$
 $A \rightarrow 0 \mid AS$
 $B \rightarrow 1$

(1) یک گرامر به فرم نرمال گریباخ است. (2) یک گرامر به فرم نرمال چامسکی است.

(3) یک گرامر به فرم نرمال چامسکی و فرم نرمال گریباخ است. (4) هیچ کدام

7- ماشینی که با دریافت یک گرامر دلخواه به فرم نرمال چامسکی و یک رشته دلخواه w از واژه‌های زبان گرامر، تعیین می‌کند که آیا w به زبان گرامر تعلق دارد یا خیر، مفروض است. بهترین عملکرد زمانی ممکن برای این ماشین برحسب $|w|$ (طول رشته w) کدام است؟ (مهندسی کامپیوتر - 83)

(1) $O(|w|^2)$ (2) $O(|w|^3)$ (3) $O(2^{|w|})$ (4) $O(\log |w|)$

8- می‌خواهیم قواعد تولید λ (Production) λ ، بی‌فایده (useless) و واحد (unit) را از یک گرامر مستقل از متن که زبان آن فاقد λ است حذف کنیم. کدام ترتیب برای حذف درست است؟ (مهندسی کامپیوتر - 81)

(1) واحد، λ ، و بی‌فایده (2) بی‌فایده، λ ، و واحد (3) λ ، بی‌فایده، و واحد (4) λ ، واحد، و بی‌فایده

9- فرض کنید G یک گرامر مستقل از متن به فرم نرمال چامسکی باشد که رشته مبهم w را تولید می‌کند، در این صورت کدام گزینه درست است؟ (مهندسی کامپیوتر - 81)

(1) حداکثر تعداد مراحل در اشتقاق‌های چپ برای w برابر $|w| + 2$ است.

(2) تعداد مراحل در اشتقاق‌های چپ برای w ثابت است.

(3) حداقل تعداد مراحل در اشتقاق‌های چپ برای w برابر $|w| + 2$ است.

(4) تعداد مراحل در دو اشتقاق چپ متفاوت برای w ممکن است متفاوت باشد.

10- فرض کنید G گرامری به فرم نرمال چامسکی، L زبان تولید شده توسط آن و w رشته‌ای در L به طول n باشند. اگر طول طولانی‌ترین مسیر جهت‌دار در درخت اشتقاق متناظر با w (ارتفاع درخت اشتقاق)، l باشد آن‌گاه مجموعه تمامی مقادیر ممکن n کدام یک از مجموعه‌های زیر است؟ (علوم کامپیوتر - 81)

(1) $\{n \in \mathbb{N} \mid 1 - 1 \leq n \leq 2^l\}$ (2) $\{n \in \mathbb{N} \mid 1 \leq n \leq 2^{l-1}\}$ (3) $\{n \in \mathbb{N} \mid 1 \leq n \leq 2(l-1)\}$ (4) $\{n \in \mathbb{N} \mid 1 \leq n \leq 2l\}$

11- گرامر $G = (\{S, X, Y\}, \{a, b\}, S, R)$ را با مجموعه قوانین زیر در نظر بگیرید:

$R : S \rightarrow XY \mid a$
 $X \rightarrow YS \mid b$
 $Y \rightarrow XS \mid b$

کدام گزاره نادرست است؟ (علوم کامپیوتر - 87)

(1) گرامر G یک گرامر مستقل از متن است.

(2) گرامر G به فرم نرمال چامسکی است.

(3) $baba$ توسط گرامر G تولید می‌شود.

(4) $baba$ فقط به یک روش از روی قوانین گرامر G تولید می‌شود.

12- اگر G یک گرامر مستقل از متن و رشته w با طول k متعلق به $L(G)$ باشد، کدام یک از گزینه‌های زیر نادرست است؟ (مهندسی کامپیوتر - 79)

(1) اگر گرامر G به شکل نرمال چامسکی باشد، برای اشتقاق w به $2k-1$ مرحله نیاز است.

(2) با حذف قواعدی که با فرم $A \rightarrow B$ هستند از گرامر G ، ممکن است تعداد مراحل اشتقاق رشته w کمتر شود.

(3) اگر گرامر G به شکل نرمال گریباخ باشد، برای اشتقاق w ، حداقل به k مرحله نیاز است.

(4) تعداد مراحل اشتقاق، بستگی به فرم گرامر ندارد و قابل پیش‌بینی نمی‌باشد.

13- کدام گزاره صحیح است؟ (علوم کامپیوتر - 84)

- (1) هر زبان مستقل از متن حتماً می‌تواند توسط یک گرامر مبهم تولید شود.
- (2) هر زبان مستقل از متن می‌تواند توسط یک گرامر مستقل از متن تولید شود که برای هر قانون آن مثل $X \rightarrow w$ داشته باشیم $|w| \leq 2^{|w|}$ ($w \in (V \cup \Sigma)^*$).
- (3) اگر L یک زبان متناهی باشد حتماً L^* یک زبان نامتناهی است.
- (4) تعداد حالات برای اتوماتون‌های مینیمال زبان‌های منظم L و L^R همواره برابر است.
- 14- حداقل پیچیدگی زمانی الگوریتم تجزیه‌ای که بتواند هر رشته متعلق به یک گرامر مستقل از متن مبهم دلخواه به فرم نرمال چامسکی را تجزیه (پارس) کند کدام است؟ (دقت کنید که الگوریتم تجزیه، گرامر را به هیچ وجه تغییر نمی‌دهد).

(مهندسی کامپیوتر - 89)

$$O(n^4) \quad (4) \quad O(n^3) \quad (3) \quad O(2^n) \quad (2) \quad O(n^3 \log n) \quad (1)$$

15- گرامر

$$\begin{aligned} G : S &\rightarrow AB \\ A &\rightarrow BB \mid a \\ B &\rightarrow BA \mid b \end{aligned}$$

(مهندسی کامپیوتر - 77)

را در نظر بگیرید. فرض کنید $n = |w|, w \in \{a, b\}^*$.

(1) اگر w توسط G تولید شود، در حالت کلی الگوریتمی با پیچیدگی زمانی $O(n)$ وجود دارد که درخت اشتقاقی برای w را بدست دهد.

- (2) در حالت کلی الگوریتمی وجود ندارد که بتوان گفت آیا w توسط G تولید می‌شود یا خیر.
- (3) در حالت کلی الگوریتمی با پیچیدگی زمانی $O(n^3)$ وجود دارد که بتوان گفت آیا w توسط G تولید می‌شود یا خیر.
- (4) هیچ‌کدام

(مهندسی کامپیوتر - 96)

16- کدام عبارت صحیح است؟

- (1) زبان‌های منظم ذاتاً مبهم نیستند.
- (2) زبان‌هایی که بتوان برای آن‌ها گرامر به فرم چامسکی ارائه نمود ذاتاً مبهم نیستند.
- (3) هر زبانی که برای آن یک گرامر مبهم وجود داشته باشد ذاتاً مبهم است.
- (4) هیچ‌کدام

17- گرامر G با متغیر شروع S بر $\Sigma = \{a, b, c\}$ با قوانین تولید زیر را در نظر می‌گیریم:

$$\begin{aligned} S &\rightarrow XY \mid W \\ X &\rightarrow aXb \mid \varepsilon \\ Y &\rightarrow cY \mid \varepsilon \\ W &\rightarrow aWc \mid z \\ Z &\rightarrow bZ \mid \varepsilon \end{aligned}$$

با قرار گرفتن کدام یک از دو مقدار b یا c به جای a در $a^{1400}b^{1401}c^{1402}$ ، رشته w حاصل در $L(G)$ قرار می‌گیرد؟

(علوم کامپیوتر - 1403)

- (1) فقط c (2) فقط b (3) هم b و هم c (4) نه b و نه c

پاسخ‌ها

فصل 6: ساده‌سازی گرامرهای مستقل از متن

تست‌های 1 الی 17

(کل تست‌های فصل 6)

1- گزینه (1) صحیح است.

در گرامر نرمال گریباخ قواعد به فرم $A \rightarrow a\alpha$ هستند که $A \in V, a \in \Sigma$ و $\alpha \in V^*$ در گرامر های صورت سوال قواعد به فرم $A \rightarrow a\alpha$ هستند که $A \in V, a \in \Sigma$ و $\alpha \in (\Sigma \cup V)^*$. پس هر گرامر نرمال گریباخ یک گرامر از نوع صورت سوال است. هر زبان مستقل از متن فاقد λ یک گرامر نرمال گریباخ دارد پس هر زبان مستقل از متن فاقد λ یک گرامر از نوع خواسته شده دارد. گرامرهای مورد نظر قادر به تولید λ نیستند. پس زبان های تولید شده توسط این گرامر ها مجموعه زبان های مستقل از متن فاقد λ است.

2- گزینه (1) صحیح است.

مثال برای $|P|=1$: در گرامر $G: S \rightarrow a$ فقط رشته $w = a$ تولید می شود و $|w|=1$.
 مثال برای $|P|=2$: در گرامر $G: S \rightarrow AA, A \rightarrow a$ رشته $w = aa$ کوچکترین رشته است و $|w|=2^{2-1}=2$.
 مثال برای $|P|=3$: در گرامر $G: S \rightarrow AA, A \rightarrow BB, B \rightarrow a$ رشته $w = aaaa$ کوچکترین رشته است و $|w|=2^{3-1}=4$.

3- گزینه (3) صحیح است.

رشته λ عضو $L(G)$ است پس در هر گرامر G' معادل با G این رشته تولید می شود. می توان تولید λ را مستقیم با متغیر شروع انجام داد یا آن را به شکل غیر مستقیم با واسطه تولید کرد. پس هردو گزینه 1 و 2 نادرستند. گرامر G دارای بازگشتی چپ است می توان با حذف بازگشتی چپ از G به گرامر معادل G' رسید که فاقد بازگشتی چپ است و در عوض قاعده بازگشتی راست دارد.

4- گزینه (3) صحیح است.

قاعده $0X \rightarrow Y$ نشان می دهد گرامر G مستقل از متن نیست پس به فرم چامسکی هم نیست. باجانشانی $0X \rightarrow Y$ در قاعده $S \rightarrow 0X1$ متوجه می شویم که $L(G) = \{1\}^*$ زبانی منظم است. هر زبان منظم پذیرنده λ -NFA دارد. این زبان با گرامر منظم $G': S \rightarrow 1S1$ نیز تولید می شود.

5- گزینه (3) صحیح است.

گرامر داده شده فرم چامسکی ندارد. با توجه به آنکه در این گرامر عملگر جمع (+) در همان سطر اول و عملگر ضرب (*) در قواعد اشتقاق بعدی ظاهر می شود پس رشته هایی را تولید می کند که اولویت عملگر (*) در آنها بیشتر از جمع (+) است. (اولویت عملگرها در رشته های تولیدی، عکس ترتیبی است که در گرامر دیده می شوند) اگر هر دو عملگر را در سطر اول در قواعد اشتقاق E داشته باشیم، گرامر مبهم می شود اما هم اکنون غیر مبهم است. در این گرامر هر دو قاعده بازگشتی $E \rightarrow E+T$ و $T \rightarrow T * F$ بازگشتی چپ هستند پس هر دو عملگر جمع و ضرب در رشته های تولید شده دارای شرکت پذیری چپ هستند بنابراین برای هر $w \in L(G)$ مانند $a + a * a$ فقط یک اشتقاق آن هم از نوع اشتقاق چپ وجود دارد.

6- گزینه (2) صحیح است.

همه قواعد به فرم $A \rightarrow BC$ یا $A \rightarrow a$ هستند که $A, B, C \in V, a \in T$. پس گرامر داده شده گرامر نرمال چامسکی است.

7- گزینه (2) صحیح است.

ماشین مورد نظر یک پیاده سازی از الگوریتم CYK است. پیچیدگی زمانی این الگوریتم $O(|w|^3)$ می باشد.

8- گزینه (4) صحیح است.

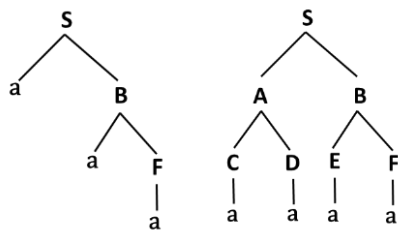
در متن درس ترتیب صحیح و علت آن گفته شده است. به طور خلاصه، حذف قواعد واحد، ممکن است باعث تولید قواعد زائد شود پس نمی تواند آخرین گام باشد. حذف قواعد λ ممکن است باعث تولید قواعد واحد شود پس نمی تواند آخرین گام باشد.

9- گزینه (2) صحیح است.

مطابق متن درس در گرامر نرمال چامسکی تعداد مراحل اشتقاق برای رشته به طول n برابر $2n-1$ است و مبهم بودن گرامر روی تعداد مراحل اشتقاق تأثیری ندارد. مثلاً در گرامر چامسکی مبهم G برای تولید رشته aa با هر روش، 3 مرحله اشتقاق لازم است:

$$\begin{array}{ll} G : S \rightarrow AB \mid AA & S \rightarrow AB \rightarrow aB \rightarrow aa \\ B \rightarrow a & S \rightarrow AA \rightarrow Aa \rightarrow aa \\ A \rightarrow a & S \rightarrow AA \rightarrow aA \rightarrow aa \end{array}$$

10- گزینه (2) صحیح است.



همه قواعد به فرم $A \rightarrow a$ یا $A \rightarrow BC$ هستند. طولانی‌ترین مسیر از ریشه تا برگ در درخت اشتقاق همان ارتفاع درخت است. برای مثال اگر $\ell = 3$ باشد، کمترین و بیشترین

طول رشته در دو درخت رسم شده به دست می‌آید: $3 \leq |w| \leq 2^2$

در حالت کلی نیز داریم: $\ell \leq |w| \leq 2^{\ell-1}$

11- گزینه صحیح وجود ندارد.

کلید اولیه گزینه 4 است اما این گزینه نیز دارای ابهام است. گزینه‌های 1 و 2 به وضوح درستند. گزینه 3 نیز درست است:

$$S \rightarrow XY \rightarrow YSY \rightarrow YSXS \xrightarrow{*} baba$$

در مورد گزینه 4، برای تولید رشته $baba$ باید از قاعده $S \rightarrow XY$ شروع کنیم سپس به ترتیب دلخواه از قواعد $Y \rightarrow XS$ و $X \rightarrow YS$ استفاده کنیم. در پایان، جانشانی‌های $S \rightarrow a, X \rightarrow b, Y \rightarrow b$ نیز با ترتیب‌های مختلف می‌تواند انجام شود. پس $baba$ با روش‌های مختلفی قابل تولید است اما درخت اشتقاق آن یکتا است. منظور طراح از (فقط یک روش) نامشخص است. اگر منظور یکتایی درخت اشتقاق باشد درست است در غیر اینصورت نادرست است.

12- گزینه (4) صحیح است.

فرض کنیم رشته w با طول k متعلق به $L(G)$ باشد، برای تولید آن در گرامر چامسکی دقیقاً به $2k-1$ مرحله اشتقاق و در گرامر گریباخ حداقل به k مرحله اشتقاق احتیاج داریم و تعداد مراحل اشتقاق در یک گرامر ساده $O(k)$ است. حذف قواعد یکانی ممکن است تعداد مراحل اشتقاق را کاهش دهد. مثلاً در گرامر مستقل از متن زیر با حذف قاعده یکانی $A \rightarrow B$ تعداد مراحل اشتقاق برای تولید ab کاهش می‌یابد:

$$\begin{array}{l} G : S \rightarrow aA \\ A \rightarrow B \\ B \rightarrow b \end{array}$$

با توجه به موارد فوق روشن است که تعداد مراحل اشتقاق، به فرم گرامر وابسته است. پس گزینه 4 نادرست است.

13- گزینه (2) صحیح است.

گزینه ۱ برای زبان‌های مستقل از متن ناپای درست است اما زبان مستقل از متن تهی اصلاً رشته‌ای ندارد که بتوان برای آن دو اشتقاق چپ متفاوت یافت؛ پس گزینه ۱ نادرست است. گزینه ۳ نیز برای زبان‌های $L = \{\}$ و $L = \{\lambda\}$ نادرست است زیرا برای این زبان‌ها $L^* = \{\lambda\}$ است.

گزینه ۴ نیز نادرست است. تعداد حالات DFA کمینه L و $\bar{L}$ برابر است اما در مورد L و L^R لزوماً چنین نیست. وقتی در DFA کمینه L جهت یال‌ها را معکوس می‌کنیم ممکن است به NFA تبدیل شود. اما گزینه ۲ صحیح است. برای زبان مستقل از متن $L - \{\lambda\}$ یک گرامر نرمال چامسکی وجود دارد. در این گرامر طول سمت راست هر قاعده تولید یک یا دو است. می‌توانیم در صورت نیاز قاعده $S \rightarrow \lambda$ را نیز به این گرامر اضافه کنیم تا گرامری با شرایط گزینه 2 برای L داشته باشیم.

14- گزینه (3) صحیح است.

از الگوریتم CYK برای تجزیه رشته‌ها می‌توان استفاده کرد و می‌دانیم این الگوریتم از مرتبه زمانی $O(n^3)$ است. گرامر ورودی باید به شکل نرمال چامسکی باشد و مبهم بودن گرامر تأثیری بر مرتبه زمانی CYK ندارد.

15- گزینه (3) صحیح است.

گرامر داده شده به شکل نرمال چامسکی است، از الگوریتم CYK می‌توان برای عضویت w به $L(G)$ استفاده کرد. پیچیدگی زمانی این الگوریتم $O(n^3)$ است. $n = |w|$

16- گزینه (1) صحیح است.

زبان‌های منظم، غیرمبهم هستند. هر زبان مستقل از متن (فاقد λ) گرامر نرمال چامسکی دارد اما برخی از این زبان‌ها ذاتاً مبهم هستند پس گرامر غیر مبهم ندارند. در واقع گرامر چامسکی ممکن است مبهم باشد پس گزینه ۲ صحیح نیست. برای تمامی زبان‌های ناپای گرامر مبهم وجود دارد. برخی از آنها گرامر غیرمبهم نیز دارند اما برخی از آنها ذاتاً مبهم هستند.

17- گزینه (2) صحیح است.

بهتر است ابتدا زبان $L(G)$ را به دست آوریم. با توجه به قواعد $Z \rightarrow bZ \mid \varepsilon$ متغیر Z به b^k تبدیل می‌شود. بنابراین با توجه به قواعد $W \rightarrow aWc \mid Z$ در نهایت به $a^n b^k c^n$ تبدیل می‌شود. با توجه به قواعد $Y \rightarrow cY \mid \varepsilon$ متغیر Y به c^m تبدیل می‌شود. با توجه به قواعد $X \rightarrow aXb \mid \varepsilon$ متغیر X به $a^p b^p$ تبدیل می‌شود. با جانشانی این نتایج در قاعده $S \rightarrow XY \mid W$ معلوم می‌شود که $L(G)$ دارای دو نوع رشته است: $a^p b^p c^m$ و $a^n b^k c^n$ به عبارتی یا تعداد a ها برابر است یا تعداد a ها برابر c ها است:

$$L(G) = \{a^n b^m c^k \mid m = n \vee k = n\}$$

در رشته $w = a^{1402} b^{1401} c^{1400}$ اگر $\alpha = a$ قرار دهیم: $\alpha \notin L(G)$ و اگر $\alpha = b$ قرار دهیم:

$$w = a^{1402} b^{1402} c^{1400} \in L(G)$$

پیروز و سربلند باشید.
پایان جلسه دوازدهم نکته و تست