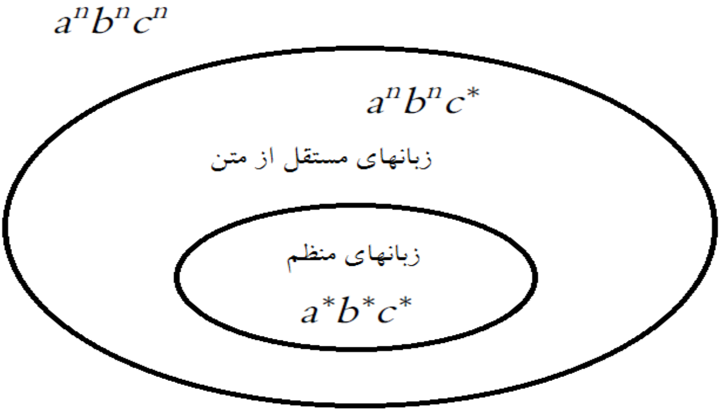
درس مبانی نظریه محاسبه

جلسه پانزدهم

لم تزریق برای زبانهای مستقل از متن

The pumping lemma for context-free languages

زبانهایی با الگوهای ساده وجود دارد که مستقل از متن نیستند.



چگونه می‌توان اثبات کرد که زبان $a^n b^n c^n$ مستقل از متن نیست؟

لم تزریق برای زبانهای مستقل از متن

لم: اگر A یک زبان مستقل از متن باشد عدد p وجود دارد بطوریکه هر رشته s در زبان A که طولش حداقل p باشد را می‌توان بصورت $s = uvxyz$ نوشت درحالیکه شرایط سه‌گانه زیر برقرار است.

1. $\forall i \geq 0, \quad uv^ixy^iz \in A$
2. $|vy| > 0$
3. $|vxy| \leq p$

بررسی صدق لم تزریق برای یک زبان مستقل از متن

$$A = \{a^n b^n \mid n \geq 0\}$$

انتخاب $p = 2$ را در نظر می‌گیریم. رشته $a^k b^k$ را برای $k \geq 1$ در نظر بگیرید.
تقسیم بندی زیر را در نظر بگیرید.

$$u = a^{k-1}, \quad v = a, \quad x = \epsilon, \quad y = b, \quad z = b^{k-1}$$

$$s = \underbrace{\overbrace{a \ldots a}^{u}}_{k-1}\underbrace{\overbrace{a}^{v}}_{1}\overbrace{}^{x}\underbrace{\overbrace{b}^{y}}_{1}\underbrace{\overbrace{b \ldots b}^{z}}_{k-1}$$

روشن است با تزریق (یا حذف) هماهنگ y و v رشته حاصل در زبان A باقی می‌ماند.

$$uv^i xy^i z \in A$$

بررسی صدق لم تزریق برای یک زبان مستقل از متن

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

انتخاب $p = 2$ را در نظر می‌گیریم. رشته $s \in B$ با طول حداقل 2 را در نظر بگیرید. فرض کنید s شامل زیررشته ab باشد (اگر نباشد حتما شامل ba است و ما بجای ab آن را در نظر می‌گیریم.)

- قسمت u را از ابتدای s تا اولین رخداد ab قرار می‌دهیم.
- $y = b$ $\qquad$ $x = \epsilon$ $\qquad$ $v = a$
- قسمت z را بقیه رشته (بعد از اولین رخداد ab) قرار می‌دهیم.

$$s = \underbrace{\ldots}_{u}\overbrace{\underbrace{a}_{v}\underbrace{\quad}_{x}\underbrace{b}_{y}}^{\text{اولین رخداد}}\underbrace{\ldots}_{z}$$

روشن است با تزریق (یا حذف) هماهنگ y و v رشته حاصل در زبان A باقی می‌ماند.

بررسی صدق لم تزریق برای یک زبان مستقل از متن

$$C = \{ww^R \mid w \in \{a,b\}^*\}$$

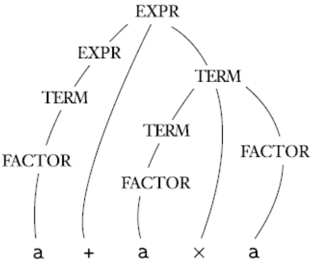
تمرین: برای رشته‌های این زبان قسمتهای u, v, x, y, z را چگونه انتخاب می‌کنید.

اثبات لم تزریق برای زبانهای مستقل از متن

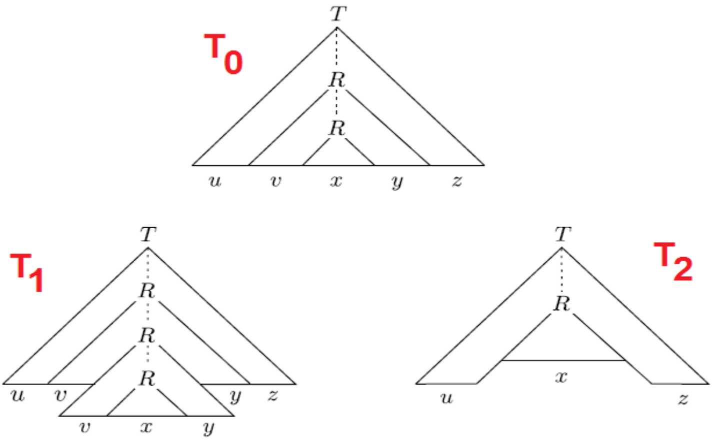
فرض کنید A یک زبان مستقل از متن باشد. پس A را می‌توان با گرامر مستقل از متن $G = (V, \Sigma, R, S)$ بیان کرد.

رشته w در زبان A را در نظر بگیرید. w توسط گرامر G تولید می‌شود و لذا یک درخت تجزیه بر اساس این گرامر دارد. برای مثال:

$$\begin{aligned}
\mathrm{EXPR} &\rightarrow \mathrm{EXPR} + \mathrm{TERM} \mid \mathrm{TERM} \\
\mathrm{TERM} &\rightarrow \mathrm{TERM} \times \mathrm{FACTOR} \mid \mathrm{FACTOR} \\
\mathrm{FACTOR} &\rightarrow (\ \mathrm{EXPR}\ ) \mid \mathtt{a}
\end{aligned}$$



ایده اثبات: چون مجموعه متغیرهای V متناهی است و طول عبارات داخل قوانین R نیز متناهی هستند، اگر طول رشته w به اندازه کافی بزرگ باشد، در درخت تجزیه w حتما یک متغیر تکرار می‌شود بنا به اصل لانه کبوتر.



جزئیات اثبات: گرامر G را در نظر بگیرید. فرض کنید b حداکثر طول یک عبارت در سمت راست قوانین گرامر G باشد. b همان ماکزیمم تعداد فرزندان در درخت تجزیه یک رشته است.

$$
\begin{array}{lcl}
 & & \overbrace{\text{EXPR} + \text{TERM}}^{b = 3} \\
\text{EXPR} & \rightarrow & \text{EXPR} + \text{TERM} \mid \text{TERM} \\
\text{TERM} & \rightarrow & \text{TERM} \times \text{FACTOR} \mid \text{FACTOR} \\
\text{FACTOR} & \rightarrow & (\ \text{EXPR}\ ) \mid \text{a}
\end{array}
$$

گزاره: اگر در درخت ریشه‌دار T هر راس حداکثر b فرزند داشته باشد و ارتفاع درخت h باشد آنگاه T حداکثر b^h برگ دارد. توجه کنید ارتفاع فاصله دورترین برگ تا ریشه است.

با توجه به گزاره بالا، اگر طول رشته w را حداقل $p = b^{|V|+1}$ بگیریم آنگاه ارتفاع درخت تجزیه w حداقل $|V| + 1$ خواهد بود. در نتیجه طول مسیر از ریشه تا یکی از برگها دست کم $|V| + 1$ خواهد بود. در این مسیر $|V| + 2$ راس وجود دارد. پس تعداد متغیرها در مسیر بیشتر از $|V|$ است و بنا به اصل لانه کبوتر یکی از متغیرها باید تکرار شود.

اگر y و v هر دو تهی باشند چه؟

در این صورت درخت تجزیه $T2$ هم که اتفاقا تعداد رئوس کمتری نسبت به $T0$ دارد رشته w را تولید خواهد کرد. پس اگر از اول فرض کنیم در بین همه درختهای تجزیه‌ای که w را تولید می‌کنند درختی با کمترین تعداد رئوس را انتخاب کرده‌ایم، این مشکل پیش نمی‌آید. پس در شروع اثبات فرض می‌گیریم که درخت تجزیه با کمترین تعداد رئوس را انتخاب کرده‌ایم.

حال چرا $|vxy| \le p$ ؟

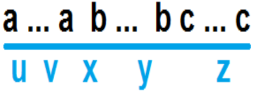
متغیر R بالایی را طوری انتخاب می‌کنیم که طول مسیر از R به پایین حداکثر $|V| + 1$ باشد. زیردرخت مربوطه حداکثر $p = b^{|V|+1}$ برگ خواهد داشت.
پس طول vxy حداکثر برابر با p است.

قضیه: زبان $A = \{a^n b^n c^n \mid n \geq 0\}$ مستقل از متن نیست.

اثبات: با برهان خلف. فرض کنید A مستقل از متن باشد. پس شرایط لم تزریق در مورد A صادق است. فرض کنید p انتخاب شده است. باید نشان دهیم یک رشته با طول حداقل p که در زبان A وجود دارد که نمی‌توانیم آن را به پنج قسمت $uvxyz$ تقسیم کنیم بطوریکه شرایط سه‌گانه لم تزریق برقرار باشند.

رشته $a^p b^p c^p$ را انتخاب می‌کنیم.

چون باید داشته باشیم $|vxy| \leq p$ پس زیررشته vxy نمی‌تواند شامل هر سه کاراکتر a و b و c باشد. در نتیجه هر طور که محل v و y را انتخاب کنیم، تزریق باعث می‌شود که رشته حاصل الگوی مطلوب را بهم زند و از زبان خارج شود.



قضیه: زبان $B = \{ww \mid w \in \{a,b\}^*\}$ مستقل از متن نیست.

اثبات: با برهان خلف. فرض کنید A مستقل از متن باشد. پس شرایط لم تزریق در مورد A صادق است. فرض کنید p انتخاب شده است. باید نشان دهیم یک رشته با طول حداقل p که در زبان A وجود دارد که نمی‌توانیم آن را به پنج قسمت $uvxyz$ تقسیم کنیم بطوریکه شرایط سه‌گانه لم تزریق برقرار باشند.

یک انتخاب بد: $s = a^p b a^p b$

یک انتخاب بهتر: $s = a^p b^p a^p b^p$

چون باید داشته باشیم $|vxy| \leq p$
زیررشته vxy هر کجای رشته که انتخاب شود تزریق باعث می‌شود که رشته حاصل الگوی مطلوب را بهم زند و از زبان خارج شود.

a ... a b ... b a ...a b ... b

u v x y z

قضیه: زبان $A = \{a^i b^j c^k \mid i \leq j \leq k\}$ مستقل از متن نیست.

اثبات: با برهان خلف. فرض کنید A مستقل از متن باشد. پس شرایط لم تزریق در مورد A صادق است. فرض کنید p انتخاب شده است. باید نشان دهیم یک رشته با طول حداقل p که در زبان A وجود دارد که نمی‌توانیم آن را به پنج قسمت $uvxyz$ تقسیم کنیم بطوریکه شرایط سه‌گانه لم تزریق برقرار باشند.

رشته $a^p b^p c^p$ را انتخاب می‌کنیم.

چون باید داشته باشیم $|vxy| \leq p$ پس زیررشته vxy نمی‌تواند شامل هر سه کاراکتر a و b و c باشد. البته اینجا بر خلاف زبان $a^n b^n c^n$ تنها تزریق کافی نیست و باید از حذف نیز برای ایجاد تناقض استفاده کنیم.