بسم الله الرحمن الرحیم

نظریه زبان‌ها و ماشین‌ها

جلسه ۱۹

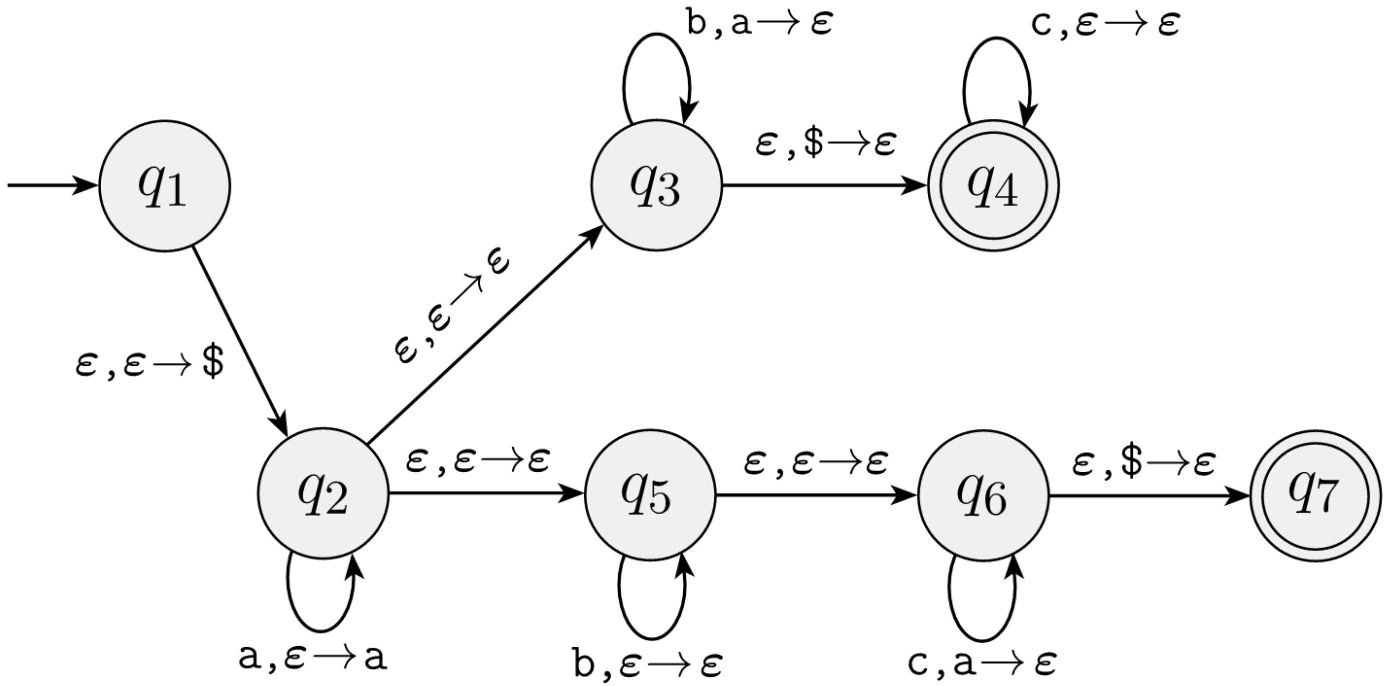
مجتبی خلیلی
دانشکده برق و کامپیوتر
دانشگاه صنعتی اصفهان

مثال

EXAMPLE 2.16

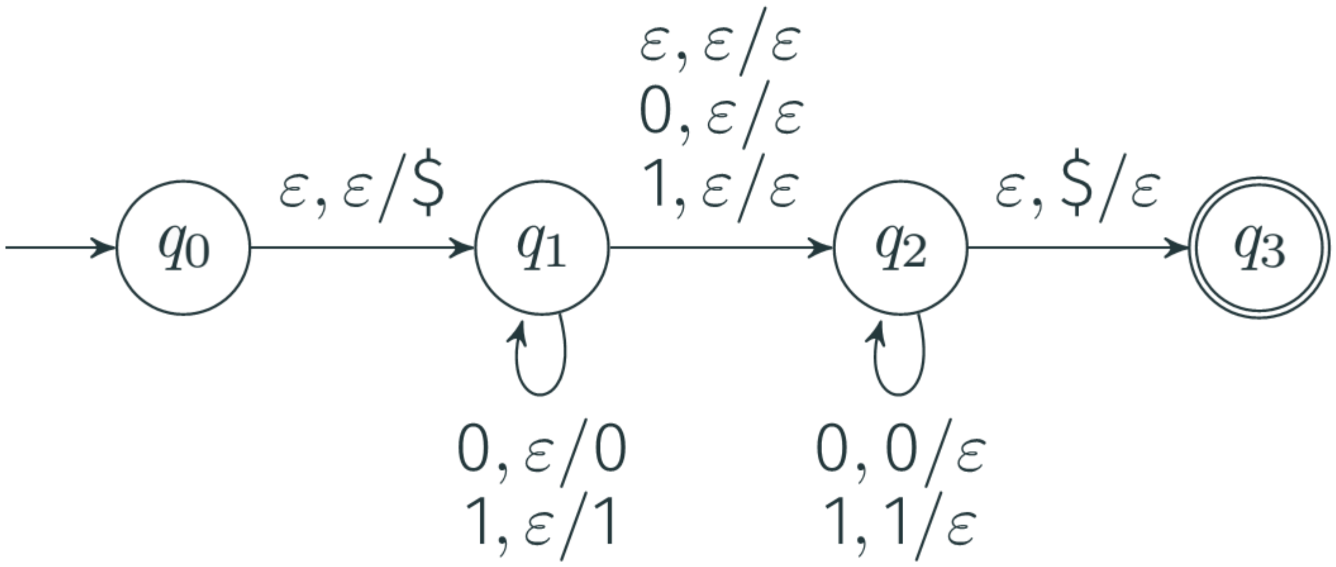
This example illustrates a pushdown automaton that recognizes the language

$$\{\mathtt{a}^i\mathtt{b}^j\mathtt{c}^k \mid i, j, k \geq 0 \text{ and } i = j \text{ or } i = k\}.$$



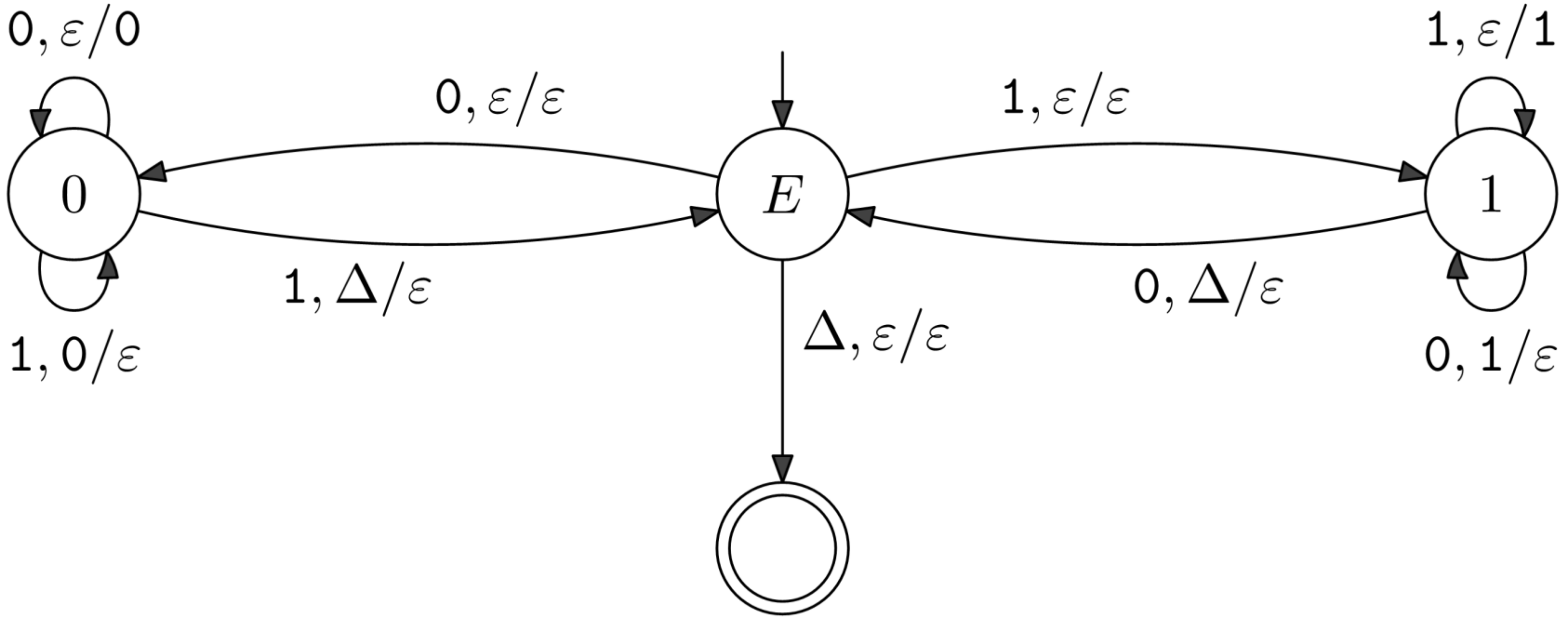
مثال

$L = \{w \mid w \in \{0,1\}^* \text{ and } w = w^R\}$



مثال

$$L = \{w \mid \ n_0(w) = n_1(w)\}$$



هم ارزی CFG و PDA

THEOREM 2.20

A language is context free if and only if some pushdown automaton recognizes it.

هم ارزی CFG و PDA

As usual for "if and only if" theorems, we have two directions to prove. In this theorem, both directions are interesting. First, we do the easier forward direction.

LEMMA 2.21

If a language is context free, then some pushdown automaton recognizes it.

از CFG به PDA

- ایده اثبات:

A = L(G) → A = L(P)

G is a CFG   P is a PDA

ساخت یک pda برای ورودی w تا تعیین کنیم آیا یک اشتقاق معتبر است.

از CFG به PDA

- ایده اثبات: ساخت یک pda برای ورودی w تا تعیین کنیم آیا یک اشتقاق معتبر (left-most) است.
  - استفاده از stack برای ذخیره کردن روند اشتقاق
  - ابتدا قرار دادن سمبل $ و متغیر آغازی در stack
  - تنها به top در stack دسترسی داریم
  - اگر در top یک ترمینال داریم، مقایسه آن با ورودی (pop کردن در صورت یکسانی- دغ رد کردن شاخه)
  - اگر در top یک متغیر داریم، جایگزینی آن از قوانین (به صورت نامعین)
  - اگر پس از اتمام رشته، $ در top بود وارد پذیرش شویم.

از CFG به PDA

- گرامر زیر را در نظر بگیرید:

$$A \rightarrow 0A1$$
$$A \rightarrow B$$
$$B \rightarrow \#$$

- مثلا برای رشته 00#11:

$$A \Rightarrow 0A1 \Rightarrow 00A11 \Rightarrow 00B11 \Rightarrow 00\#11$$

از CFG به PDA

- گرامر زیر را در نظر بگیرید:

$$A \rightarrow 0A1$$
$$A \rightarrow B$$
$$B \rightarrow \#$$

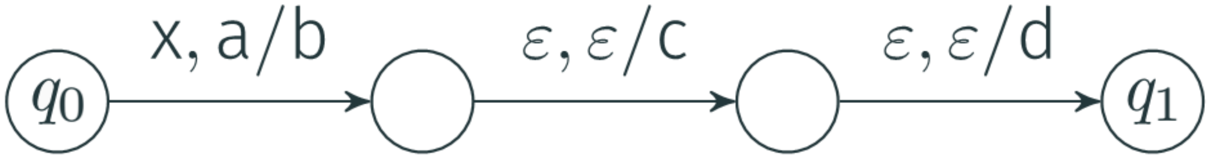
$$A \Rightarrow 0A1 \Rightarrow 00A11 \Rightarrow 00B11 \Rightarrow 00\#11$$

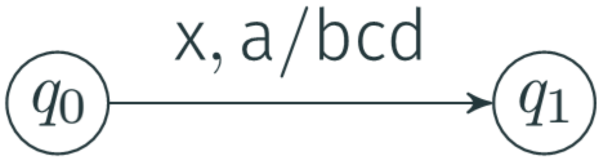
- مثلا برای رشته 00#11:

گذار	stack	ورودی

از CFG به PDA

- معادل هستند:





از CFG به PDA

IUT-ECE

- گرامر زیر را در نظر بگیرید:

$$A \to 0A1$$
$$A \to B$$
$$B \to \#$$

$$A \Rightarrow 0A1 \Rightarrow 00A11 \Rightarrow 00B11 \Rightarrow 00\#11$$

- مثلا برای رشته 00#11:

گذار	stack	ورودی
$\varepsilon, \varepsilon/\$A$	$\$A$	00#11
$\varepsilon, A/1A0$	$\$1A0$	00#11
$0, 0/\varepsilon$	$\$1A$	0#11
$\varepsilon, A/1A0$	$\$11A0$	0#11

از CFG به PDA

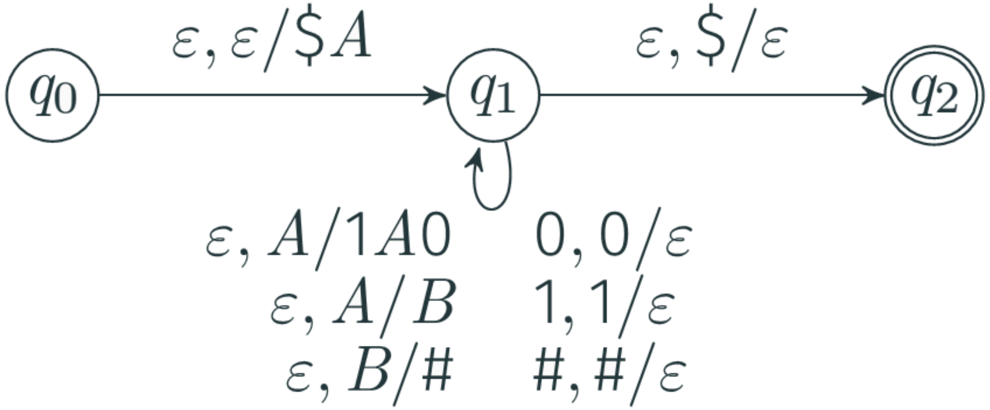
$$A \to 0A1$$
$$A \to B$$
$$B \to \#$$

$$A \Rightarrow 0A1 \Rightarrow 00A11 \Rightarrow 00B11 \Rightarrow 00\#11$$

ورودی	stack	گذار
$00\#11$	$\$A$	$\varepsilon, \varepsilon/\$A$
$00\#11$	$\$1A0$	$\varepsilon, A/1A0$
$0\#11$	$\$1A$	$0, 0/\varepsilon$
$0\#11$	$\$11A0$	$\varepsilon, A/1A0$

...

- داریم:



از CFG به PDA

- فرم کلی:

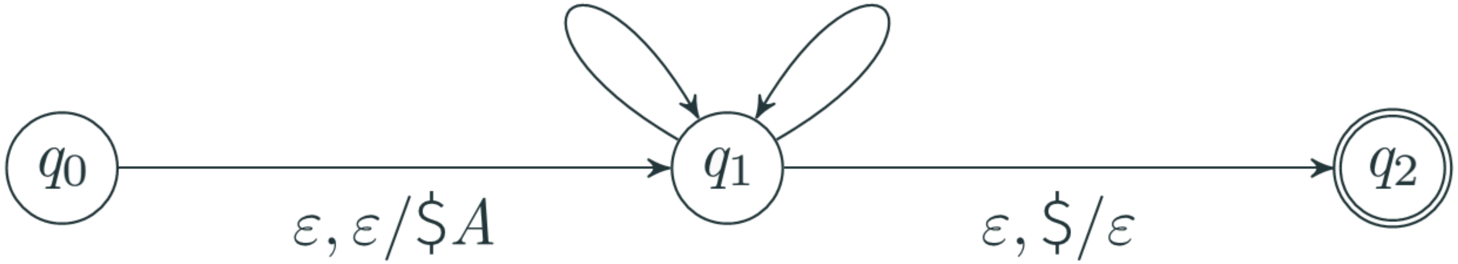
برای هر rule

$$A \rightarrow \alpha_1 \ldots \alpha_k$$

$$\varepsilon, A/\alpha_k \ldots \alpha_1$$

برای هر ترمینال

$$\mathrm{a}, \mathrm{a}/\varepsilon$$



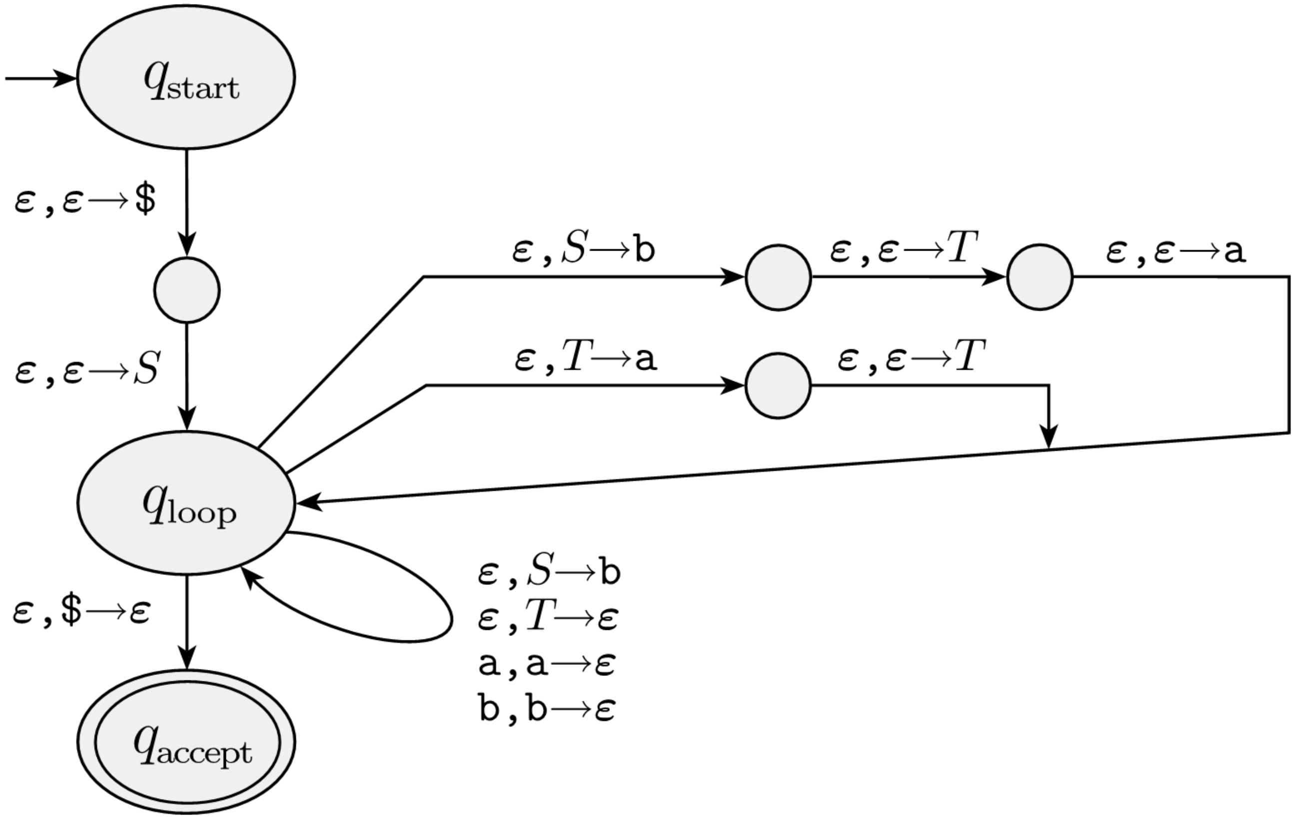
Loop

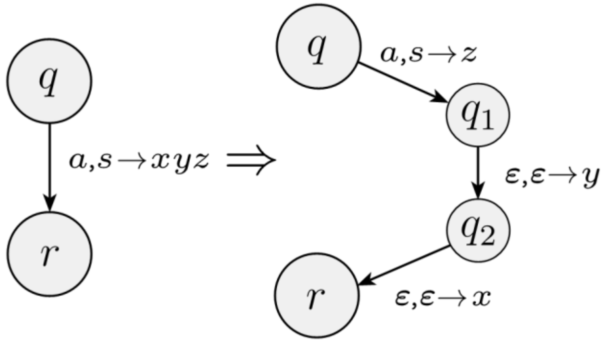
EXAMPLE 2.25

We use the procedure developed in Lemma 2.21 to construct a PDA P_1 from the following CFG G.

$$S \rightarrow \mathtt{a}T\mathtt{b} \mid \mathtt{b}$$
$$T \rightarrow T\mathtt{a} \mid \varepsilon$$

The transition function is shown in the following diagram.





IUT-ECE

از PDA به CFG

Now we prove the reverse direction of Theorem 2.20. For the forward direction, we gave a procedure for converting a CFG into a PDA. The main idea was to design the automaton so that it simulates the grammar. Now we want to give a procedure for going the other way: converting a PDA into a CFG. We design the grammar to simulate the automaton. This task is challenging because "programming" an automaton is easier than "programming" a grammar.

LEMMA 2.27

If a pushdown automaton recognizes some language, then it is context free.

از PDA به CFG

- ایده اثبات:

A = L(P) → A = L(G)

P is a PDA

G is a CFG

G باید همه رشته‌هایی که از حالت آغازی P به حالت پایانی آن میروند را تولید کند.
در نظر گرفتن متغیر B_{qp} برای دو حالت p و q و ....

خواص بستاری CFL

- زبانهای مستقل از متن نسبت به عملگرهای زیر بسته هستند:
  - اجتماع
  - الحاق
  - ستاره کلینی
  - معکوس
  - اشتراک با یک زبان منظم
  - ...

خواص بستاری CFL

- مثال(اجتماع):

$S_1 \rightarrow aS_1b \mid \varepsilon$ $\qquad \cup \qquad$ $S_2 \rightarrow aS_2a \mid bS_2b \mid cS_2c \mid \varepsilon$

$$S \rightarrow S_1 \mid S_2$$
$$S_1 \rightarrow aS_1b \mid \varepsilon$$
$$S_2 \rightarrow aS_2a \mid bS_2b \mid cS_2c \mid \varepsilon$$

خواص بستاری CFL

IUT-ECE

- مثال(ستاره):

$S_1 \rightarrow aS_1b \mid \varepsilon$

$S \rightarrow \varepsilon \mid S_1S$
$S_1 \rightarrow aS_1b \mid \varepsilon$

خواص بستاری CFL

- **قضیه**: اشتراک یک زبان مستقل از متن و یک زبان منظم، یک زبان مستقل از متن است.
- اثبات:

PDA $M_1 = (Q_1, \Sigma, \Gamma, \delta_1, q_1, F_1)$

$$L(M_1) = A$$

NFA $M_2 = (Q_2, \Sigma, \delta_2, q_2, F_2)$

$$L(M_2) = B$$

PDA $M = (Q, \Sigma, \Gamma, \delta, q_0, F)$

$$L(M) = A \cap B$$

خواص بستاری CFL

- **قضیه**: اشتراک یک زبان مستقل از متن و یک زبان منظم، یک زبان مستقل از متن است.

- اثبات:

$$\left.\begin{array}{l}\text{PDA } M_1 = (Q_1, \Sigma, \Gamma, \delta_1, q_1, F_1) \\ \text{NFA } M_2 = (Q_2, \Sigma, \delta_2, q_2, F_2)\end{array}\right\} \quad \text{PDA } M = (Q, \Sigma, \Gamma, \delta, q_0, F)$$

$$Q = Q_1 \times Q_2$$

$$q_0 = (q_1, q_2)$$

$$F = F_1 \times F_2$$

$$\delta\big((q,r),a,b\big) = \big\{\big((s,t),c\big) \bigm| (s,c) \in \delta_1(q,a,b) \text{ and } t \in \delta_2(r,a)\big\} \quad \text{for } a \in \Sigma_\varepsilon,\ b,c \in \Gamma_\varepsilon$$

- M رشته w را میپذیرد اگر M_1 و M_2 رشته w را بپذیرند.

مثال

- یک PDA یا CFG برای زبان زیر بنویسید:

$$B = \{\mathtt{a}^n\mathtt{b}^n\mathtt{c}^n \mid n \geq 0\}$$

لم تزریق برای CFL

- مشابه لم تزریق برای زبانهای منظم
- برخی زبانها توسط CFG تولید نمیشوند.
- برخی زبانها توسط PDA تشخیص داده نمیشوند.
- چگونه نشان دهیم زبانی CFL نیست؟

لم تزریق برای CFL

IUT-ECE

THEOREM 2.34

Pumping lemma for context-free languages If A is a context-free language, then there is a number p (the pumping length) where, if s is any string in A of length at least p, then s may be divided into five pieces $s = uvxyz$ satisfying the conditions

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

لم تزریق برای CFL

- ایده اثبات:
  - قبلا در اثبات لم تزریق برای زبانهای منظم دیدیم که اگر طول رشته بیشتر از تعداد حالتها بود آنگاه دست کم دو بار از برخی حالتها گذر کرده بودیم.
  - در اینجا همانند قبل است اما درباره متغیرهای گرامرها میتوانیم همین استدلال را کنیم که باید در روند اشتقاق دست کم دو بار از یک یا چند متغیر استفاده شود.

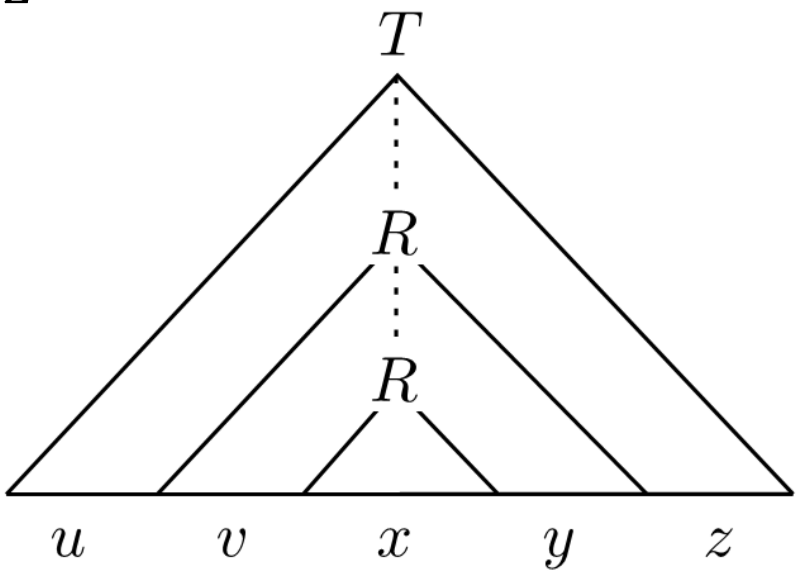
لم تزریق برای CFL

IUT-ECE

- بنابراین اگر طول رشته از حدی بیشتر باشد، برخی متغیرها باید دو بار در مسیر ریشه به برگ در درخت تجزیه وجود داشته باشند:

$$T \Rightarrow^* uRz \Rightarrow^* uvRyz \Rightarrow^* uvxyz$$

$$R \Rightarrow^* vxy \;, \qquad R \Rightarrow^* x$$

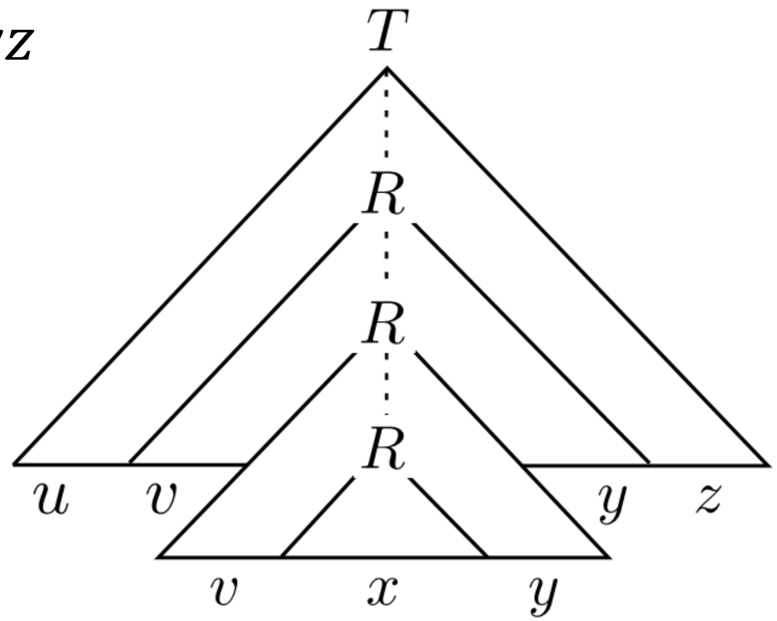


لم تزریق برای CFL

- جایگزینی زیردرخت دوم با زیردرخت اول (میتوان اینکار را تکرار کرد):

$$T \Rightarrow^* uRz \Rightarrow^* uvRyz \Rightarrow^* uvxyz$$

$$R \Rightarrow^* vxy \;, \qquad R \Rightarrow^* x$$



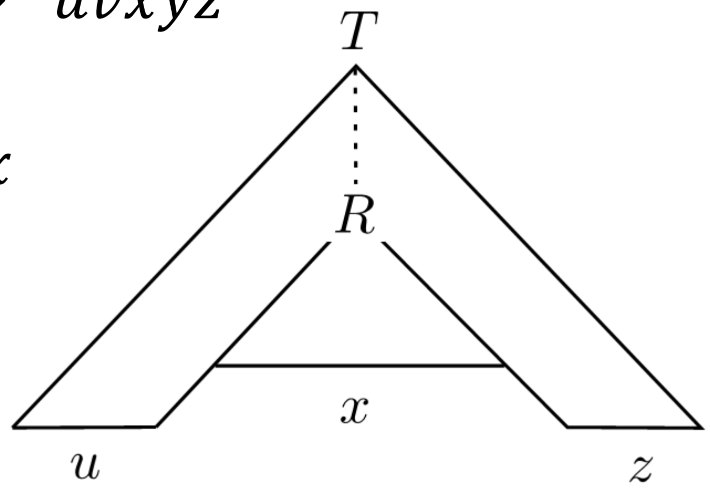
$$T \Rightarrow^* uRz \Rightarrow^* uvRyz \Rightarrow^* uvvxyyz$$

لم تزریق برای CFL

- میتوان زیردرخت دوم را حذف کرد و به نتیجه زیر رسید:

$$T \Rightarrow^* uRz \Rightarrow^* uvRyz \Rightarrow^* uvxyz$$

$$R \Rightarrow^* vxy \;, \qquad R \Rightarrow^* x$$



$$T \Rightarrow^* uRz \Rightarrow^* uxz$$

مثال

EXAMPLE 2.36

Use the pumping lemma to show that the language $B = \{\mathtt{a}^n\mathtt{b}^n\mathtt{c}^n | \, n \geq 0\}$ is not context free.

We assume that B is a CFL and obtain a contradiction. Let p be the pumping length for B that is guaranteed to exist by the pumping lemma. Select the string $s = \mathtt{a}^p\mathtt{b}^p\mathtt{c}^p$. Clearly s is a member of B and of length at least p. The pumping lemma states that s can be pumped, but we show that it cannot. In other words, we show that no matter how we divide s into $uvxyz$, one of the three conditions of the lemma is violated.

First, condition 2 stipulates that either v or y is nonempty. Then we consider one of two cases, depending on whether substrings v and y contain more than one type of alphabet symbol.

1. When both v and y contain only one type of alphabet symbol, v does not contain both a's and b's or both b's and c's, and the same holds for y. In this case, the string uv^2xy^2z cannot contain equal numbers of a's, b's, and c's. Therefore, it cannot be a member of B. That violates condition 1 of the lemma and is thus a contradiction.
2. When either v or y contains more than one type of symbol, uv^2xy^2z may contain equal numbers of the three alphabet symbols but not in the correct order. Hence it cannot be a member of B and a contradiction occurs.

One of these cases must occur. Because both cases result in a contradiction, a contradiction is unavoidable. So the assumption that B is a CFL must be false. Thus we have proved that B is not a CFL. ■

مثال

EXAMPLE 2.38

Let $D = \{ww|\ w \in \{0,1\}^*\}$. Use the pumping lemma to show that D is not a CFL. Assume that D is a CFL and obtain a contradiction. Let p be the pumping length given by the pumping lemma.

This time choosing string s is less obvious. One possibility is the string 0^p10^p1. It is a member of D and has length greater than p, so it appears to be a good candidate. But this string *can* be pumped by dividing it as follows, so it is not adequate for our purposes.

$$\overbrace{\underbrace{000\cdots000}_{u}\;\underbrace{0}_{v}\;\underbrace{1}_{x}}^{0^p1}\;\overbrace{\underbrace{0}_{y}\;\underbrace{000\cdots0001}_{z}}^{0^p1}$$

Let's try another candidate for s. Intuitively, the string $0^p1^p0^p1^p$ seems to capture more of the "essence" of the language D than the previous candidate did. In fact, we can show that this string does work, as follows.

We show that the string $s = 0^p1^p0^p1^p$ cannot be pumped. This time we use condition 3 of the pumping lemma to restrict the way that s can be divided. It says that we can pump s by dividing $s = uvxyz$, where $|vxy| \leq p$.

First, we show that the substring vxy must straddle the midpoint of s. Otherwise, if the substring occurs only in the first half of s, pumping s up to uv^2xy^2z moves a 1 into the first position of the second half, and so it cannot be of the form ww. Similarly, if vxy occurs in the second half of s, pumping s up to uv^2xy^2z moves a 0 into the last position of the first half, and so it cannot be of the form ww.

But if the substring vxy straddles the midpoint of s, when we try to pump s down to uxz it has the form $0^p1^i0^j1^p$, where i and j cannot both be p. This string is not of the form ww. Thus s cannot be pumped, and D is not a CFL. ■

مثال

IUT-ECE

EXAMPLE 2.37

Let $C = \{\texttt{a}^i\texttt{b}^j\texttt{c}^k \mid 0 \leq i \leq j \leq k\}$. We use the pumping lemma to show that C is not a CFL. This language is similar to language B in Example 2.36, but proving that it is not context free is a bit more complicated.

Assume that C is a CFL and obtain a contradiction. Let p be the pumping length given by the pumping lemma. We use the string $s = \texttt{a}^p\texttt{b}^p\texttt{c}^p$ that we used earlier, but this time we must "pump down" as well as "pump up." Let $s = uvxyz$ and again consider the two cases that occurred in Example 2.36.

مثال

1. When both v and y contain only one type of alphabet symbol, v does not contain both `a`'s and `b`'s or both `b`'s and `c`'s, and the same holds for y. Note that the reasoning used previously in case 1 no longer applies. The reason is that C contains strings with unequal numbers of `a`'s, `b`'s, and `c`'s as long as the numbers are not decreasing. We must analyze the situation more carefully to show that s cannot be pumped. Observe that because v and y contain only one type of alphabet symbol, one of the symbols `a`, `b`, or `c` doesn't appear in v or y. We further subdivide this case into three subcases according to which symbol does not appear.

مثال

a. *The* a*'s do not appear.* Then we try pumping down to obtain the string $uv^0xy^0z = uxz$. That contains the same number of a's as s does, but it contains fewer b's or fewer c's. Therefore, it is not a member of C, and a contradiction occurs.

b. *The* b*'s do not appear.* Then either a's or c's must appear in v or y because both can't be the empty string. If a's appear, the string uv^2xy^2z contains more a's than b's, so it is not in C. If c's appear, the string uv^0xy^0z contains more b's than c's, so it is not in C. Either way, a contradiction occurs.

c. *The* c*'s do not appear.* Then the string uv^2xy^2z contains more a's or more b's than c's, so it is not in C, and a contradiction occurs.

2. When either v or y contains more than one type of symbol, uv^2xy^2z will not contain the symbols in the correct order. Hence it cannot be a member of C, and a contradiction occurs.

Thus we have shown that s cannot be pumped in violation of the pumping lemma and that C is not context free. ■