Proof of Pumping Lemma using Regular Expression

Junfu Chen

January 22, 2019

Definition. A regular expression R is a **reduced regular expression** if $R = A_1B_1^*A_2B_2^*...A_{n-1}B_{n-1}^*A_n$ for which

1. $A_i \in \Sigma^*$ for all i (i.e. A_i contains no union or star operation).

2. B_i is non-empty regular expression for all i.

Lemma (Pumping Lemma for Reduced Regular Expression)**.** *Let R be a reduced regular expression. Then there exists an integer $p \geq 1$ depending only on R such that any string $w \in L(R)$ of length at least p can be written as $w = xyz$, satisfying the following conditions:*

1. $|y| \geq= 1$

2. $|xy| \leq p$

3. $(\forall n \geq 0)(xy^n z \in L(R))$

Proof. Let p be the length of R plus 1. Take any $w \in L(R)$ with $|w| \geq p$. Since R is a reduced regular expression, $R = A_1B_1^*A_2B_2^*...A_{n-1}B_{n-1}^*A_n$ with $A_i \in \Sigma^*$ and B_i nonempty regular expression. Since $w \in L(R)$, there exists $a_1, a_2, ..., a_n, b_1, b_2, ..., b_{n-1}$ such that $w = a_1b_1...a_{n-1}b_{n-1}a_n$ with $a_i \in L(A_i)$ and $b_i \in L(B_i^*)$ for all i.

Since $a_i \in L(A_i)$ and $A_i \in \Sigma^*$, $|a_i| \leq |A_i|$. Then

$$|w| = \sum_{i=1}^{n} |a_i| + \sum_{i=1}^{n-1} |b_i|$$
$$|w| \leq \sum_{i=1}^{n} |A_i| + \sum_{i=1}^{n-1} |b_i|$$

Since $|w| \geq p = |R| + 1$, $\sum_{i=1}^{n} |A_i| \leq |R|$, we must have $\sum_{i=1}^{n-1} |b_i| \geq 1$. Hence there exists at least one k such that $|b_k| \geq 1|$. Pick the b_k with the smallest index, such that $|b_i| = 0$ for all $i < k$.

Since $b_k \in L(B_k^*)$, there exists $y \in B_k$ such that $b_k = y^t$ for some $t \geq 1$ and $y^n \in B_k^*$ for all n.

Let $x = a_1b_1a_2...b_{k-1}a_k$, $z = y^{t-1}a_{k+1}b_{k+1}...b_{n-1}a_n$, then we have

- $|y| \geq= 1$ since $|y^t| \geq 1$.

- $|xy| \leq p$ since $|xy| = \sum_{i=1}^{k} |a_i| + \sum_{i=1}^{k-1} |b_i| + |y| \leq \sum_{i=1}^{k} |A_i| + 0 + |B_k| \leq p$

- $(\forall n \geq 0)(xy^n z \in L(R))$ since

$$\begin{aligned} xy^n z &= (a_1 b_1 a_2 ... b_{k-1} a_k)(y^n)(y^{t-1} a_{k+1} b_{k+1} ... b_{n-1} a_n) \\ &= a_1 b_1 a_2 ... b_{k-1} a_k y^{n+t-1} a_{k+1} b_{k+1} ... b_{n-1} a_n \in L(R) \end{aligned}$$

  because $a_i \in A_i$ for all i, $b_i \in B_i^*$ for all $i \neq k$ and $y^{n+t-1} \in B_k$.

Thus proving the pumping lemma. $\square$