Exercise 1: Regular Expressions

Consider the regular expression \( r = (aa^* + ba)^* \cdot (bbb + ab^*) + (ab)^* \). State for each of the following words whether they are contained in \( L(r) \)

\( \epsilon, baabbb, bbbb, abab, ababa, ababaa, a. \)

In case that a word is contained in \( L(r) \) show how to obtain the word from \( r \), e.g., by marking the corresponding parts in \( r \).

Sample Solution

- \( \epsilon \): \( (aa^* + ba)^* \cdot (bbb + ab^*) + (ab)^* \)
- \( baabbb \): \( (aa^* + ba)^* \cdot (bbb + ab^*) + (ab)^* \)
- \( bbbb \): The word is not contained in the language as using \( (aa^* + ba)^* \) or \( (ab)^* \) implies that the word is empty or contains an \( a \) and \( (bbb + ab^*) \) cannot produce four consecutive bs.
- \( abab \): \( (aa^* + ba)^* \cdot (bbb + ab^*) + (ab)^* \)
- \( ababa \): The string is not contained in the language. The word cannot be created with \( (ab)^* \) as all words created with \( (ab)^* \) end with a \( b \).

The word can also not created with \( (aa^* + ba)^* \cdot (bbb + ab^*) \): All words created with that regular expression have to use the part \( (bbb + ab^*) \) for a single time at the end of the word, which for \( ababa \) implies that \( (bbb + ab^*) \) would have to generate an \( a \) (the expression cannot generate \( ba \) or any longer substring of the word ending at the end).

This means that we would have to generate \( abab \) with \( (aa^* + ba)^* \) but all strings generated by that expression cannot end with a \( b \).

- \( ababaa \): \( (aa^* + ba)^* \cdot (bbb + ab^*) + (ab)^* \)
- \( a \): \( (aa^* + ba)^* \cdot (bbb + ab^*) + (ab)^* \).

Exercise 2: Regular Expressions 2

1. Let \( L_1 \) be the language consisting of words of the form \( w_1w_2w_3 \) with \( w_1, w_2, w_3 \in \{a, b, c\}^* \) and \( w_1 \) contains no \( a \)'s and \( w_2 \) contains no \( b \)'s and \( w_3 \) contains no \( c \)'s. Give a regular expression that generates \( L_1 \).

2. Let \( L_2 \subseteq \{a, b\}^* \) be the language of all words that do not have any of the words \( \{aaa, aaaa, \ldots\} \) as a consecutive substring. Give a regular expression that generates \( L_2 \).
Sample Solution

1. \((b + c)^* \cdot (a + c)^* \cdot (a + b)^*\).

2. \(b^* + (b^*(a + aa)b^+)^*(a + aa)b^*\) or alternatively \((\epsilon + a + aa)(b + ba + baa)^*\).

Exercise 3: Pumping Lemma

Use the pumping lemma to show that \(L = \{a^i b^j \mid i \neq j\}\) is not regular.

Sample Solution

We prove that \(L\) is not regular by showing that \(L\) violates the pumping condition, i.e. always contains a string that cannot be pumped. Let \(p\) be the pumping length. We investigate the string \(s = a^p b^p + p \notin L\). Furthermore \(s\) is longer than \(p\).

Consider a partition \(xyz = s\) with \(|y| \geq 1\) and \(|xy| \leq p\). It follows that \(y = a^q\) with \(1 \leq q \leq p\). Set \(k = p/q + 1 \in \mathbb{N}\). Then we have that \(xy^k z = a^{p+q(k-1)} b^p + p \notin L\). Regardless of the partition we gave a pumped string that is not contained in the language.

Remark: This solution was pretty tough. You should also look at exercises from previous years to find easier pumping lemma questions.

Exercise 4: Context Free Grammar

Give a context free grammar for the language \(L = \{a^i b^j \mid i \neq j\}\).

Sample Solution

\(S\) is the start symbol.

\[
S \rightarrow AaTb \mid aTbB \mid A \mid B \\
T \rightarrow \epsilon \mid aTb \\
A \rightarrow a \mid aA \\
B \rightarrow b \mid bB.
\]