1 Language Theory

Let Σ be a finite set of letters in the alphablet.

We define $\Sigma^* = 2^{\Sigma}$.

A **language** is a subset of Σ^* over Σ.

We define the following operations:

Concatenation:

If $x, y \in \Sigma^*$, then $x : y \in \Sigma^*$.

For $L, L' \in \Sigma^*$, we have $L \cdot L' = \{x : y \,|\, x \in L, y \in L'\}$.

We call $x \in \Sigma^*$ a **proper prefix** of $y \in \Sigma^*$ if there exists a string $x' \in \Sigma^*$ such that $y = xx'$ and both x and x' are not λ.

We call $x \in \Sigma^*$ a **proper suffix** of $y \in \Sigma^*$ if there exists a string $x' \in \Sigma^*$ such that $y = x'x$ and both x and x' are not λ.

We call $x \in \Sigma^*$ a **substring** of $y \in \Sigma^*$ if there exist $x' \in \Sigma^*$ and $x'' \in \Sigma^*$ such that $y = x'xx''$.

2 Regular Expressions

Let Σ be a finite alphabet. The **set of regular expressions** R over Σ is the inductively defined set of strings R such that $\emptyset, \lambda \in R$ and $\Sigma \in R$. If $r, r' \in R$, then $r + r'$, $r \cdot r'$ and r^* are in R.

A **generalised regular expression** allows complementation and intersection, so that $r \cap r' \in R$, $r - r'$ and $\overline{r} \in R$.

The language denoted by a regular expression is $d(r)$, where $\mathfrak{L} : R \to \{L \,|\, L \subseteq \Sigma^*\}$ is defined inductively as follows:

- $\mathfrak{L}(\emptyset) = \emptyset$
- $\mathfrak{L}(\lambda) = \{\lambda\}$
- $\mathfrak{L}(a) = \{a\}$ for each $a \in \Sigma$
- $\mathfrak{L}(r + r') = \mathfrak{L}(r) \cup \mathfrak{L}(r')$
- $\mathfrak{L}(r - r') = \mathfrak{L}(r) \cdot \mathfrak{L}(r')$
- $\mathfrak{L}(r^*) = (\mathfrak{L}(r))^*$
- $\mathfrak{L}(\overline{r}) = \Sigma^* - \mathfrak{L}(r)$
- $\mathfrak{L}(r - r') = \mathfrak{L}(r) - \mathfrak{L}(r')$

e.g. $\mathfrak{L}(0^*(10^*10^*)^*)$ is the set of all strings over $\{0, 1\}$ with an even number of 1's.

Claim. Let $L = \mathfrak{L}(0(00)^*(11)^* + (00)^*1(11)^*) = \{0^m1^n \,|\, m + n \text{ is odd}\}$.

Proof.

Let $x \in L$ be arbitrary.

Then $x = 0^m1^n$, where $m + n$ is odd.

First, suppose that m is odd. Say $m = 2k + 1$. Then n is even, so $n = 2l$ for some $k, l \in \mathbb{N}$.

Then $x = 0^{2k+1}1^{2l} = 0(00)^k(11)^l \in \mathfrak{L}(0(00)^*) \cdot \mathfrak{L}((11)^*) = \mathfrak{L}(0(00)^* \cdot (11)^*) = \mathfrak{L}(r)$.

Suppose now that m is even. The proof is similar.

Hence $L \subseteq \mathfrak{L}(r)$.

Exercise: Continue the proof. □

3 Finite Automata

A **finite automata** is a determenistic finite state automation (abbreviated as DFA or DFSA) which has a finite set of states Q, a finite alphabet Σ, an initial state $q_0 \in Q$, a set of final (also known as accepting) states and a state transition function $\delta : Q \times \Sigma \to Q$.

An **extended transition function** $\delta^* : Q \times \Sigma^* \to Q$ is defined by $\delta^*(q, \lambda) = q$ for $q \in Q$ for all $a \in \Sigma$ and all $x \in \Sigma^*$ we have that $\delta^*(q, ax) = \delta^*(\delta(q, a), x)$.

Equivalently, for all $x \in \Sigma^*$ and $a \in \Sigma$ we have $\delta^*(q, ax) = \delta(\delta^*(q, a), x)$.

Let $A = (Q, \Sigma, \delta, q_0, f$, and denote $\{x \in \Sigma^* \mid \delta^*(q_0, x) \in f\}$ as $\mathfrak{L}(A)$.