Prof. Javier Esparza
Philipp Czerner

Technical University of Munich
Chair for Foundations of Software Reliability

Automata and Formal Languages

Winter Term 2023/24 – Exercise Sheet 6

Exercise 6.1.

(a) Build the automata B_p and C_p for the word pattern $p = mammamia$.

(b) How many transitions are taken when reading $t = mami$ in B_p and C_p?

(c) Let $n > 0$. Find a text $t \in \{a, b\}^*$ and a word pattern $p \in \{a, b\}^n$ such that testing whether p occurs in t takes n transitions in B_p and $2n - 1$ transitions in C_p.

Exercise 6.2.

In order to make pattern-matching robust to typos we want to include also "similar" words in our results. For this we consider words with a small Levenshtein-distance (edit-distance) "similar".

We transform a word w to a new word w' using the following operations (with $a_i, b \in \Sigma$):

- *replace* (R): $a_1 \dots a_{i-1}a_ia_{i+1} \dots a_l \to a_1 \dots a_{i-1}ba_{i+1} \dots a_l$
- *delete* (D): $a_1 \dots a_{i-1}a_ia_{i+1} \dots a_l \to a_1 \dots a_{i-1}\varepsilon a_{i+1} \dots a_l$
- *insert* (I): $a_1 \dots a_{i-1}a_ia_{i+1} \dots a_l \to a_1 \dots a_{i-1}a_iba_{i+1} \dots a_l$

The Levenshtein-distance (denoted $\Delta(w, w')$) of w and w' is the minimal number of operations (R,D,I) needed to transform w into w'. We denote with $\Delta_{L,i} = \{w \in \Sigma^* \mid \exists w' \in L.\, \Delta(w', w) \leq i\}$ the language of all words with edit-distance at most i to some word of L.

(a) Compute $\Delta(become, bekommen)$ and $\Delta(become, werden)$.

(b) Let p be the pattern $ABBA$. Construct an NFA-ϵ locating the pattern or variations of it with edit-distance 1.

(c) Prove the following statement: If L is a regular language, then $\Delta_{L,n}$ is a regular language.

Exercise 6.3.

Consider transducers whose transitions are labeled by elements of $(\Sigma \cup \{\varepsilon\}) \times (\Sigma^* \cup \{\varepsilon\})$. Intuitively, each transition reads one or zero letter and writes a word of arbitrary length. Such a transducer can be used to perform operations on strings, e.g. upon reading `"singing in the rain"` it could write `Singing In The Rain`.

Sketch such ε-transducers for the following operations, each of which is informally defined by means of three examples. For each example, when the transducer reads the string on the left, it should write the string on the right. You may assume that the alphabet Σ consists of $\{a, b, \dots, z, A, B, \dots, Z\}$, a whitespace symbol, and an end-of-line symbol. Moreover, you may assume that every string ends with an end-of-line symbol and contains no other occurrence of the end-of-line symbol.

(a)

Input	Output
`Automata and Formal Languages`	`AFL`
`Technical University of Munich`	`TUM`
`Max Planck Institute`	`MPI`

(b) For this exercise, Σ is extended with $\{,,\cdot\}$.

Input	Output
`Ada Lovelace`	`Lovelace, A.`
`Alan Turing`	`Turing, A.`
`Donald Knuth`	`Knuth, D.`

(c) For this exercise, Σ is extended with $\{0, 1, \ldots, 9, (,), +\}$. We want to transform phone-numbers into a normal form, where they are prefixed with a country code.

Input	Output
`004989273452`	`+49 89 273452`
`(00)4989273452`	`+49 89 273452`
`273452`	`+49 89 273452`
`2 7 3 4 5 2`	`+49 89 273452`
`498949`	`+49 89 498949`
`+49 89 498949`	`+49 89 498949`