apl. Prof. Dr. Thomas Noll

Daniel Zilken, Roy Hermanns

Compiler Construction 2025

— Exercise 0 —

General Remarks

- Exercises are *optional*, i.e., not required for admission to exams. However, corrections to students' solutions are provided as annotations to the submissions.
- Please use the corresponding task in the Moodle room to submit your solution. Paper submissions are not accepted.
- Please hand in your solutions in *groups of four* and hand in only one solution per group. You can use the forum in this Moodle room to find group members.

Exercise 1 (0 Points)

Which of the following statements hold?

(a) Deterministic finite automata (DFA) are strictly less expressive than regular expressions.

(b) Non-deterministic finite automata (NFA) are strictly more expressive than DFA.

(c) The regular languages are closed under:

  (i) union,

  (ii) intersection,

  (iii) complement,

  (iv) concatenation,

  (v) Kleene closure.

(d) Context Free Languages (CFL) are closed under:

  (i) union,

  (ii) intersection,

  (iii) complement,

  (iv) concatenation,

  (v) Kleene closure.

(e) DCFL is the set of context free languages that are accepted by deterministic push down automata. Is DCFL = CFL?

Solution:

(a) No. One can be effectively converted to another.

(b) No. NFA can be converted to a DFA using the powerset construction (Rabin-Scott construction).

(c) Yes for all. Each operation can be implemented by manipulating a DFA or NFA.

(d) Context Free Languages (CFL) are closed under the following:

  (i) Yes. S_1 and S_2 are starting Non-terminal for two Grammar. $S \to S_1 \mid S_2$.

  (ii) No. $L_1 = \{a^i b^i c^j \mid i, j \in \mathbb{N}\}$ and $L_2 = \{a^i b^j c^j \mid i, j \in \mathbb{N}\}$

  (iii) No. CFLs are closed under union and if they were closed under complement, they would also be closed under intersection.

(iv) Yes. $S \to S_1 S_2$.

(v) Yes. Let S be the starting non-terminal. New Grammar $S' \to SS' \mid \varepsilon$.

(e) No. Just an intuitive example: Palindromes are in CFL but not in DCFL.

Exercise 2 (0 Points)

(a) Describe the language of the following regular expression in words:

$$r = (0+1)^*0(0+1)^*0(0+1)^*.$$

(b) Construct the regular expression for...

(i) the set of all strings with at most one pair of consecutive 0's and at most one pair of consecutive 1's,

(ii) the set of all strings with equal number of 0's and 1's such that no prefix has two more 0's than 1's nor two more 1's than 0's.

(c) Construct a context free grammar (CFG) for a set of strings of $\{(,)\}^*$ such that every string of the set has equal number of left and right parenthesis, and every prefix has at least as many left parenthesis as right parenthesis.

Solution:

(a) The set of all strings containing at least two 0's.

(b) (i) The regular expression is:

$$\begin{aligned} (0+\varepsilon)(10)^*(01)^*(10)^*(1+\varepsilon) \\ +(0+\varepsilon)(10)^*(01)^*0 \\ +(1+\varepsilon)(01)^*(10)^*(01)^*(0+\varepsilon) \\ +(1+\varepsilon)(01)^*(10)^*1 \end{aligned}$$

(ii) The regular expression is:

$$(01+10)^*$$

(c) The CFG is:

$$S \to (\ S\ )\ S \mid \varepsilon$$

Exercise 3 (0 Points)

(a) Let r and s be regular expressions. Consider the set X such that $X = r.X + s$. Under the assumption that the language of r does not contain ε (i.e., $\varepsilon \notin L(r)$), find X.

(b) (i) Show that the language $L = \{0^{i^2} \mid i \in \mathbb{N}\}$ is not regular.

(ii) Show that the language $L = \{a^i b^i c^i \mid i \in \mathbb{N}\}$ is not a CFL.

Solution:

(a) Guess the solution to be r^*s (because it feels right).

$$X = r.X + s \qquad (\star)$$

By substituting r^*s for X in the equation $(\star)$, we observe that r^*s satisfies this property. Thus, $L(r^*s) \subseteq L(X)$. Translating the equation to the topology of strings,

$$X = R.X \cup S$$

where $R = L(r)$ and $S = L(s)$. Let $\varepsilon \notin L(r)$. Suppose the solution of the above equation is $X = R^*S \cup C$, where $R^*S \cap C = \emptyset$. This new set must also satisfy the equation. Thus,

$$R^*S \cup C = R(R^*S \cup C) \cup S = RR^*S \cup RC \cup S = (RR^*S \cup S) \cup RC = R^*S \cup RC$$

Since $R^*S \cap C = \emptyset$, $C \subseteq RC$. Now we can see why the condition $\varepsilon \notin R$ is important, mainly $C \subseteq RC$ has no solution for C when $\varepsilon \notin R$.

The proof also shows what happens if $\varepsilon \in R$, $r^*s + c$ is a solution of equation $(\star)$ for any c.

(i) Let L be regular and n be the integer in the pumping lemma. Let $w = 0^{n^2}$. By pumping lemma $w = xyz$ with $|y| \geq 1$, $|xy| \leq n$ and xy^iz is in L for all i. For $i = 2$, $n^2 < |xy^2z| \leq n^2 + n$. Or $|zy^2z|$ is strictly between n^2 and $(n+1)^2$.

(ii) Assume L is defined by a CFG. Let n be the pumping integer. We choose a string $z = a^n b^n c^n$. We write $z = uvwxy$. By the premises of the pumping lemma $|vwx| \leq n$, thus it cannot contain both a and c. W.l.o.g. assume vwx does not contain c. If $v = a^k b^l$ or $x = a^k b^l$ then pumping directly leads to a word not in L. Therefore assume $v = a^k, w = a^l b^p, x = b^q$. Then for $i := 2$: $uv^2wx^2y = a^{n+k} b^{n+q} c^n$. Now either $n + k > n$ or $n + q > n$ and therefore $uv^2wx^2y \notin L$.