School of Computing and Information Systems
COMP30026 Models of Computation
Week 8: Regular Expressions, Context-Free Grammars, and the Pumping Lemma for Regular Languages

Exercises

T8.1 Give regular expressions for the following languages over the alphabet $\Sigma = \{0, 1\}$.

(i) $\{w \mid w$ begins with a `1` and ends with a `0`$\}$

(ii) $\{w \mid w$ contains the substring `0101`$\}$

(iii) $\{w \mid$ the length of w is at most $5\}$

(iv) $\{w \mid w$ contains at least two `0`s and at most one `1`$\}$

(v) $\{w \mid w$ is any string except the empty string$\}$

Continued in P8.1

T8.2 Give context-free grammars for the following languages. Assume the alphabet is $\Sigma = \{\mathtt{0}, \mathtt{1}\}$.

(i) $\{w \mid$ the length of w is odd and its middle symbol is `0`$\}$

(ii) $\{w \mid w$ is a palindrome$\}$

Continued in P8.5 & P8.6

T8.3 A context-free grammar is called *ambiguous* if there is a string in its language which has two distinct parse-trees. A context-free language is called ambiguous (or *inherently ambiguous*), if all of its context-free grammars are ambiguous. Consider the context-free grammar $(\{A, B, T\}, \{\mathtt{a}, \mathtt{b}\}, R, T)$ with rules R:

$$\begin{array}{lcl} T & \rightarrow & A \mid B \\ A & \rightarrow & \mathtt{a}\ \mathtt{b} \mid \mathtt{a}\ A\ \mathtt{b} \\ B & \rightarrow & \epsilon \mid \mathtt{a}\ \mathtt{b}\ B \end{array}$$

Show that G is ambiguous, but $L(G)$ is not ambiguous.

Continued in P8.12

T8.4 The pumping lemma for regular languages is the following statement: if A is a regular language, then there exists an integer $p \geq 1$ (called the *pumping length*) such that, for any string $s \in A$ with $|s| \geq p$, there exist strings x, y and z such that $s = xyz$ and

- $y \neq \epsilon$,
- $|xy| \leq p$,
- $xy^iz \in A$ for all integers $i \geq 0$.

Use the pumping lemma for regular languages to prove that the following language is not regular:

$$A = \{\mathtt{0}^n\mathtt{1}^n\mathtt{2}^n \mid n \geq 0\}$$

Continued in P8.4

Homework problems

P8.1 Give regular expressions for the following languages over the alphabet $\Sigma = \{0, 1\}$.

(i) $\{w \mid w$ has length at least 3 and its third symbol is $\texttt{0}\}$

(ii) $\{w \mid$ every odd position of w is a $\texttt{1}\}$

(iii) $\{\epsilon, \texttt{0}\}$

(iv) The empty set

Bonus: Draw minimal NFAs for these languages, and those from T8.1

P8.2 String s is a *suffix* of string t iff there exists some string u (possibly empty) such that $t = us$. For any language L we can define the set of suffixes of strings in L:

$$\mathit{suffix}(L) = \{x \mid x \text{ is a suffix of some } y \in L\}$$

Let A be any regular language. Show that $\mathit{suffix}(A)$ is a regular language. *Hint:* think about how a DFA for A can be transformed to recognise $\mathit{suffix}(A)$.

P8.3 In general it is difficult, given a regular expression, to find a regular expression for its complement. However, it can be done, and you have been given all the necessary tricks and algorithms. This question asks you to go through the required steps for a particular example.

Consider the regular language $(\texttt{ba}^*\texttt{a})^*$. Assuming the alphabet is $\Sigma = \{\texttt{a}, \texttt{b}\}$, we want to find a regular expression for its complement, that is, for

$$L = \{w \in \{\texttt{a}, \texttt{b}\}^* \mid w \text{ is not in } (\texttt{ba}^*\texttt{a})^*\}$$

To complete this task, go through the following steps.

(a) Construct an NFA for $(\texttt{ba}^*\texttt{a})^*$. Two states suffice.

(b) Turn the NFA into a DFA using the subset construction method.

(c) Do the "complement trick" to get a DFA D for L.

(d) Reflect on the result: Wouldn't it have been better/easier to apply the "complement trick" directly to the NFA?

(e) Turn DFA D into a regular expression for L using the NFA-to-regular-expression conversion process shown in the lecture on regular expressions.

P8.4 A *palindrome* is a string that reads the same forwards and backwards. Use the pumping lemma for regular languages and/or closure results to prove that the following languages are not regular:

(a) $B = \{\texttt{a}^i\texttt{ba}^j \mid i > j \geq 0\}$

(b) $C = \{w \in \{\texttt{a}, \texttt{b}\}^* \mid w \text{ is not a palindrome}\}$

(c) $D = \{www \mid w \in \{\texttt{a}, \texttt{b}\}^*\}$

P8.5 Give context-free grammars for the following languages. Assume the alphabet is $\Sigma = \{\texttt{0}, \texttt{1}\}$.

(i) $\{w \mid w$ starts and ends with the same symbol$\}$

(ii) $\{w \mid$ the length of w is odd$\}$

P8.6 Construct a context-free grammar for the language $\{\texttt{a}^i\texttt{ba}^j \mid i > j \geq 0\}$.

P8.7 Show that the class of context-free languages is closed under the regular operations: union, concatenation, and Kleene star. *Hint:* show how context-free grammars for A and B can be manipulated to produce context-free grammars for $A \cup B$, AB, and A^*. Careful: The variables used in the grammars for A and in B could overlap.

P8.8 If we consider English words the "symbols" (or primitives) of English, we can use a context-free grammar to try to capture certain classes of sentences and phrases. For example, we can consider articles (A), nouns (N), adjectives (Q), intransitive verbs (IV), transitive verbs (TV), noun phrases (NP), verb phrases (VP), and sentences (S). List 5–10 sentences generated by this grammar:

$$
\begin{array}{lllllllll}
S & \to & NP\ VP & N & \to & \text{cat} & IV & \to & \text{hides} \\
A & \to & \text{a} & N & \to & \text{dog} & IV & \to & \text{runs} \\
A & \to & \text{the} & Q & \to & \text{lazy} & IV & \to & \text{sleeps} \\
NP & \to & A\ N & Q & \to & \text{quick} & TV & \to & \text{chases} \\
NP & \to & A\ Q\ N & VP & \to & IV & TV & \to & \text{hides} \\
N & \to & \text{bone} & VP & \to & TV\ NP & TV & \to & \text{likes}
\end{array}
$$

Are they all meaningful? Discuss "well-formed" versus "meaningful".

P8.9 How would you change the grammar from the previous question so that "adverbial modifiers" such as "angrily" or "happily" can be used? For example, we would like to be able to generate sentences like "the dog barks constantly" and "the black cat sleeps quietly".

P8.10 Consider this context-free grammar G with start symbol S:

$$
\begin{array}{lll}
S & \to & \mathtt{a\,b\,a} \mid \mathtt{a\,b}\,A \mid \mathtt{b}\,B \\
A & \to & \mathtt{b} \mid \mathtt{b}\,A \mid \mathtt{a}\,B \\
B & \to & \mathtt{a}\,A \mid \mathtt{a}\,B
\end{array}
$$

Draw an NFA which recognises $L(G)$. *Hint:* The grammar is a regular grammar; you may want to use the labels S, A, and B for three of the NFA's states.

P8.11 Consider the context-free grammar G with rules

$$S \to \mathtt{a\,b} \mid \mathtt{a}\,S\,\mathtt{b} \mid S\,S$$

Use structural induction to show that no string $w \in L(G)$ starts with `abb`.

P8.12 Consider the context-free grammar $(\{S\}, \{\mathtt{a}, \mathtt{b}\}, R, S)$ with rules R:

$$S \to \mathtt{a} \mid \mathtt{b} \mid S\,S$$

Show that the grammar is ambiguous; then find an equivalent unambiguous grammar.

P8.13 Give a context-free grammar for the language recognised by this DFA: