Theoretical Computer Science (M21276)

Part A/6: Regular grammars & beyond

(Oct 9-13, 2023)

Question 1.

- What is the difference between regular and context-free grammar?
- Can you give an example of a language which is context-free but not regular?
- How can we prove that a given language is not regular?
- What is the pumping lemma?
- What is the pigeonhole principle? Why within ANY group of 13 people, there must be at least two who have their birthdays in the same months.

Question 2. Consider the following grammar with A, B nonterminals, S the initial nonterminal; and 0, 1 terminals

$$\begin{aligned} S &\to 0 \mid 0A \\ A &\to 1B \\ B &\to 0 \mid 0A \end{aligned}$$

(i) What type is this grammar? Give the name of the smallest Chomsky class to which the language belongs to.

Answer: Regular grammar, regular language.

(ii) Which strings belong to the language generated by this grammar? Find three strings from the language and describe the structure of all strings belonging to the language.

Answer: $L = \{0(10)^n \mid n \geq 0\}$

(iii) If it is possible, find a regular expression for the language.

Answer: $0(10)^*$

(iv) If it is possible, construct a DFA which recognises the language. Does you DFA has a minimum number of states?

Answer: The language is regular, a DFA can be constructed.

Question 3. Consider the following grammar:

$$\begin{aligned} S &\to 0A \mid 1A \\ A &\to 1BB \\ B &\to 11 \mid 01 \end{aligned}$$

with the nonterminals S (initial), A, B and a set of terminals $\{0, 1\}$.

(i) What type is this grammar? Is the language finite? Give the name of the smallest Chomsky class to which the language belongs to.

Answer: Context free, but for a regular language, we can change the productions:

$$
\begin{aligned}
S &\to 0A \mid 1A \\
A &\to 1B \\
B &\to 1111 \mid 1101 \mid 0111 \mid 0101
\end{aligned}
$$

(ii) What is the structure of the strings from the language?

(iii) If the language is regular, can you find a regular expression for the language? *Answer:* $(01 + 11)(01 + 11)(01 + 11)$

Question 4. For each of the following languages over the alphabet $\{a, b\}$, state whether it is regular or not. If not, give a reason why is not regular. If the language is regular, can you construct an NFA which recognise it?

(i) $\{(ab)^n \mid n > 100\}$

Answer: Regular

(ii) $\{(ab)^n \mid n < 100\}$

Answer: Finite $\implies$ regular

(iii) $\{a^n b^n \mid n < 100\}$

Answer: Finite $\implies$ regular

(iv) $\{a^n b^n a^n \mid n > 1\}$

Answer: Not regular, can be proved using the pumping lemma.

Question 5. Find a context-free grammar for the following languages over $\Sigma = \{a, b\}$

(i) $L = \{ab(bbaa)^n bba(ba)^n \mid n \geq 0\}$

Answer: Non-terminals: S (initial), A, Terminals: $T = \{a, b\}$
the productions: $S \to abA$, $A \to bbaaAba$, $A \to bba$

(ii) $L = \{a^n b^m \mid 0 \leq n \leq m + 3\}$

Answer: Non-terminals: S (initial), A, B, Terminals: $T = \{a, b\}$
the productions: $S \to A| \ aA \mid aaA \mid aaaA$, $A \to aAb \mid B$, $B \to Bb \ |\Lambda$

(iii) $L = \{a^n b^m \mid 0 \leq 2n \leq m \leq 3n\}$

Answer: Non-terminals: S (initial), Terminals: $T = \{a, b\}$
the productions: $S \to \Lambda \mid aSbb \mid aSbbb$,

Question 6. Consider the following grammar:

$$\begin{aligned}
S &\to SABC \mid ABC \\
AB &\to BA \\
BA &\to AB \\
BC &\to CB \\
CB &\to BC \\
AC &\to CA \\
CA &\to AC \\
A &\to a \\
B &\to b \\
C &\to c
\end{aligned}$$

with the nonterminals S (initial), A, B, C which generates the language L over the alphabet $\{a, b, c\}$.

(i) Find three strings from L.

(ii) Is *abac* in the language? What is the structure of the strings from the language?

Answer: No, same number of a's, b's, and c's.

(iii) Can you find a regular expression for L?

Answer: No, L is not a regular language.

(iv) What type is the grammar? Give the name of the smallest Chomsky class to which the language L belongs to.

Answer: Following our definition, the grammar is unrestricted, more precisely monotomic. Therefore the language is context-sensitive, it can be generated by a monotonic grammar (see the question above). The language can't be a context-free because we need to be able to keep in memory, how many a's, b's and c's we have seen. So the language is context-sensitive.

Question 7. Consider the following grammar over the alphabet $\Sigma = \{0, 1\}$, and the set of non-terminal $N = \{S\ (initial), A, B\}$:

$$\begin{aligned}
S &\to 0AB \\
B &\to SA \mid 01 \\
A1 &\to SB1 \\
1B &\to 0 \\
A0 &\to S0B
\end{aligned}$$

(i) What type is this grammar?

Answer: Phrase structure, or unrestricted

(ii) Is the language empty?

Answer: Yes, every A creates and S and vice versa.

Question 8. A monotonic grammar with the restriction "any production rule is allowed as long as there are no rules for making strings shorter, except one $S \to \Lambda$" also describes a context-sensitive language. Can you rewrite the rule of the type $XY \to YX$ using only context-sensitive production rules?

Answer: e.g. $AB \to AX$, $AX \to BX$, $BX \to BA$.

Question 9. Let L and M be context-free languages. Can you decide whether the languages $L \cup M$, $L \cdot M$, L^* are context-free?

Answer: All three languages are context-free. If L and M are context-free grammars with the start symbols A and B, then the grammar for $L \cup M$ is context-free and starts with the two productions $S \to A \mid B$; The grammar for $L \cdot M$ starts with the production $S \to AB$; and the grammar for L^* starts with the production $S \to \Lambda \mid AS$.