



EECS6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Tuesdays, Thursdays 10:00-11:20 – LAS 3033
Winter Semester, 2015

THE SMALL ASSIGNMENT

1. In this question we ask whether an element x is in a set S , by giving conditions on x , S , and other sets. For each of the parts below, answer A, B, or C, with the following meanings:

- A: x is surely in the set S .
B: x cannot possibly be in the set S .
C: It cannot be determined from what is given whether x is or is not in S .
- a) x is in $S \cup T$ and x is NOT in T .
b) x is NOT in $S \cap T$ and x is NOT in T .
c) x is NOT in $T - S$ and x is in T .
d) x is in $R - (R - T)$ and $S + R \cap T$.
e) x is NOT in $R - (R - T)$ and $S = R \cap T$.

2. Write a regular expression for the language L , over alphabet $\{0, 1, 2\}$, such that every 0 that is not the last (rightmost) symbol is immediately followed by a 1, and every 1 that is not the last symbol is immediately followed by a 0. By "immediately followed" we mean "with no intervening symbols. Thus, 010 is in L , but 001 is not, because the first 0 is not *immediately* followed by 1. 021 is not in L for the same reason.
3. Let L be the language consisting of all strings of zero or more 0's followed by one or more 1's followed by two or more 2's. For example 001122, 122, and 0111122 are in L : 012 (too few 2's), and 0112122 (a 2 precedes a 1) are not.

Write a regular expression whose language is L

Are the following regular expressions and, if so, what strings do they produce?

$L(01)$

$L(01+0)$

$L(0(1+0))$

Note order of precedence of operators.

$L(0^*)$

$L((0+10)^*(\epsilon+1))$

4. Give a context free grammar generating the same language, It is sufficient just to give the productions, assuming that S is the start symbol

5. Arithmetic expressions with operator + and parentheses can be generated by the grammar

$$E \rightarrow E+E \mid (E) \mid a$$

where a stands for any number. This grammar is ambiguous.

(a) Give an example of a string that has two or more leftmost derivations or parse trees.

(b) Design an unambiguous grammar for the same language.

6. Here is a context free grammar $G = (\{S, A, B\}, \{0, 1\}, P, S)$, where P is the set of productions_

$$S \rightarrow 0A \mid 1B \mid \epsilon$$

$$A \rightarrow 1S \mid 0AA$$

$$B \rightarrow 0S \mid 1BB$$

Intuitively, A generates strings with one more 1 than 0, B generates strings with one more 0 than 1, and S generates the strings with equal numbers of 0's and 1's

Give a leftmost derivation of the string 0011

Give a rightmost derivation of the string 001011

Draw a parse tree for the string 1001

How does this grammar work?

$$\begin{array}{l} S \rightarrow AB \mid CD \\ A \rightarrow aA \mid \epsilon \\ B \rightarrow bBc \mid E \mid cD \\ C \rightarrow aCb \mid E \mid aA \\ D \rightarrow cD \mid \epsilon \\ E \rightarrow bE \mid b \end{array}$$

7. Show how one would represent the following natural language statements in predicate calculus form.

- (a) Sidney Crosby or Evgeni Malkin is a centre.
- (b) If Fred is a dog and Bruce is a cat, then Fred chases Bruce.
- (c) No one likes Bill.
- (d) Bill gave the dirty old green book to his mother.
- (e) Computing Science is part of the Faculty of Applied Science.
- (f) All dogs chase some cat (or other).
- (g) All women are mortal.
- (h) Someone is loved by everybody.

8. What about these sentences make them difficult to interpret?

- a) The man saw the boy with the binoculars.
- b) They are hunting dogs.
- c) Free whales.
- d) Police help dog bite victim.
- e) He saw that gas can explode.
- f) We saw her duck.
- g) The kiwi eats roots and leaves.
- h) The old man the boats.
- i) I once shot an elephant in my pajamas.
- j) Every farmer who owns a donkey beats it.
- k) The rat the cat the dog bit chased escaped.
- l) John wants to marry the prettiest blonde.