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

THE SMALL ASSIGNMENT (answers)

---

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*. Ans.: A
   b) *x* is NOT in $S \cap T$ and *x* is NOT in *T*. Ans.: C.
   c) *x* is NOT in $T - S$ and *x* is in *T*. Ans.: A
   d) *x* is in $R - (R - T)$ and $S + R \cap T$. Ans.: A
   e) *x* is NOT in $R - (R - T)$ and $S = R \cap T$. Ans.: B. Note that (d) and (e) exploit the law $(R - (R - T)) \equiv (R \cap T)$.

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.

   As per the given conditions, the 2's can only occur in the beginning of the sentence. So one of the possible representations is 2*(0(10)* | (1(01)*)

   Grading - There are many equivalent regular expressions, all were considered. In case the expression was incorrect, we tried to mention either the illegal expressions generated or some legal expression missed.

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

   0*11*222*

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

   L(**01**) = {01}.

   L(**01**+**0**) = {01, 0}.

   L(**0**(**1**+**0**)) = {01, 00}. -note order of precedence of operators.

   L(**0***) = {ε, 0, 00, 000,… }.

   L ((**0**+**10**)*(ε+**1**)) = all strings of 0's and 1's without two consecutive 1's.

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

   There were many acceptable grammars. One is the following

   S → A B C
   A → 0A | ε
   B → 1B | 1

   C → 2C | 22

   Common errors with a included being careless with the regular expression (forgetting a 1, for example, and just having 01*222*).

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

   E -> E+E | (E) | 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.

   Part a) Consider the string a + a + a, we have following two left-most derivations for the same:

   Derivation 1 : E ⇒ E + E ⇒ a + E ⇒ a + E + E ⇒ a + a + E ⇒ a + a + a
   Derivation 2 : E ⇒ E + E ⇒ E + E + E ⇒ a + E + E ⇒ a + a + E ⇒ a + a + a

   Hence this grammar is ambiguous.

   Part b) The following is an equivalent un-ambiguous grammar.

   E → E + T|T
   T → (E)|a

   Points were deducted for:
   - Not showing(or providing some reasoning) why a particular string is ambiguous.
   - The modified grammar is ambiguous.
   - The modified grammar though not ambiguous is not equivalent to the original grammar.

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

   S → 0A j | 1B | ε
   A → 1S j 0AA
   B → 0S j 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

   S→0A→00AA→001SA→001A→0011S→0011

   Give a rightmost derivation of the string 001011

   S→0A→00AA→00A0AA→00A0A1S→00A0A1→00A01S1→00A011→001S011→001011

   Draw a parse tree for the string 1001

How does this grammar work?

```
S -> AB | CD
A -> aA | ε
B -> bBc | E | cD
C -> aCb | E | aA
D -> cD | ε
E -> bE | b
```

A generates zero or more *a*'s.
D generates zero or more *c*'s.
E generates one or more *b*'s.
B first generates an equal number of *b*'s and *c*'s, then produces either one or more *b*'s (via *E*) or one or more *c*'s (via *cD*). That is, *B* generates strings in *b*c** with an unequal number of *b*'s and *c*'s.
Similarly, *C* generates unequal numbers of *a*'s then *b*'s.
Thus, *AB* generates strings in *a*b*c** with an unequal numbers of *b*'s and *c*'s, while *CD* generates strings in *a*b*c** with an unequal number of *a*'s and *b*'s.

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

   (a) Sidney Crosby or Evgeni Malkin is a centre.

   [or (centre(Sidney-Crosby)) (centre (Evgeni-Malkin))]

   (b) If Fred is a dog and Bruce is a cat, then Fred chases Bruce.

   [dog(Fred) & cat(Bruce)] → chases(Fred, Bruce)

   (c) No one likes Bil.

   ~∃x [person(x) & likes(x, Bill)]

   (d) Bill gave the dirty old green book to his mother.

   gave(Bill, mother1, d-o-g-book1)

   (e) Computing Science is part of the Faculty of Applied Science.

   part-of(CS, Applied Science)

   (f) All dogs chase some cat (or other).

   (x) ∃y [dog(x) → cat(y) & chases(x,y)]

   (g) All women are mortal.

   (x) [woman(x) → mortal(x)]

   (h) Someone is loved by everybody.

   y (x) [loves(x,y)]

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

   a) The man saw the boy with the binoculars. (PP atachment ambiguity)
   b) They are hunting dogs. (ambiguous “hunting dogs” vs. hunting ”dogs”)
   c) Free whales. (ambiguous “free whales” vs.”free” whales)
   d) Police help dog bite victim. (ambiguous “Police help dog” vs. “Polics help dog-bite”)
   e) He saw that gas can explode. (ambiguous – “can” as noun or auxiliary)
   f) We saw her duck. (ambiguous “duck” as noun or verb)
   g) The kiwi eats roots and leaves. (ambiguous – 3 verbs vs. eats “rootsand leaves”)
   h) The old man the boats. (awkward but okay)
   i) I once shot an elephant in my pajamas. From G.Marks – “What he was doing in my pajamas I do not know” - ambiguous
   j) Every farmer who owns a donkey beats it. (ambiguous – anaphoric reference)
   k) The rat the cat the dog bit chased escaped. (awkward but okay)
   l) John wants to marry the prettiest blonde. (referentially opaque – many ways ambiguous – does he know the prettiest blonde or whoever he marries he wants to be the prettiest blonde – at the time of wanting or of marrying????)