CS 536

Midterm Exam

Wednesday, April 2, 2008

12:30 — 2:20 PM

1325 CSST

Instructions

Answer any four questions. (If you answer more, only the first four will count.) Point values are as indicated. Please try to make your answers neat and coherent. Remember, if we can't read it, it's wrong. Partial credit will be given, so try to put something down for each question (a blank answer always gets 0 points!).

1. (25 points)
   Let AllButFirst be the operator that systematically removes the first character from a set of non-null strings. For example, AllButFirst ({abc,xy,a,b,bb}) = {bc,y, λ,b}. Let R be any regular expression that does not generate λ. Show that AllButFirst(R) is a regular set.

2. (25 points)
   Let Double be the set of strings defined as {s|s=ww}. Double contains only strings composed of two identical repeated pieces. For example, if we use a vocabulary consisting of the letters a and b, the following strings (and many more!) are in Double: aa, bb abab abbabb, aaaabaaaab.

   Is Double a regular set? Why?

3. (a) (10 points)
   Let FA be any finite automaton (my choice). Explain how to decide whether **all** the strings FA accepts are exactly two characters in length.

   (b) (15 points)
   Let S_1 and S_2 be sets of strings. Define S_1 - S_2 to be the set of all strings in S_1 that aren't also in S_2. Assume R_1 and R_2 are regular sets. Show that R_1 - R_2 is also a regular set.

4. (25 points)
   Recall that in building parsers, the **first** relation is key in making parsing decisions. Define InFirst(A,a), where A is a nonterminal and a is a terminal, to be true if it is the case that
   A $\Rightarrow^+$ a…. (and false otherwise) That is, InFirst(A,a) tells us if a string of symbols beginning with the terminal symbol a may be derived from A.

   Give an algorithm that computes InFirst(A,a) given a grammar G.

5. (25 points)
Assume we add a simple parameterless macro facility to CSX. The command

```
#define ident   string
```

(in the file being scanned) will define identifier `ident` to be an abbreviation for `string`, an ordinary CSX quoted string. Whenever `ident` is scanned (as an identifier) after the `#define`, it will be replaced with the text of `string`. This text (with enclosing quotes removed and escaped characters expanded), is immediately scanned. `string` **may** contain macro calls. For example, we might use the following to abbreviate calls to `println` in Java:

```
#define pr "system.out.println"
```

Assume your scanner calls a subroutine `getNextChar()` whenever it wants the next character to be scanned. You may modify `getNextChar()` in any way you wish. Outline how you'd add this macro facility to your CSX scanner.

6. Consider the following three context-free grammars. Which are LL(1)? Why? Which are LALR(1)? Why?

(i) (8 points)

S → A S
S → b
A → a
A → λ

(ii) (8 points)

S → (S)
S → [S]
S → [S)
S → a

(iii) (9 points)

S → D P a
D → b
D → λ
P → c
P → λ