

# CS 536

## Midterm Exam

Tuesday, April 17, 2007

5:00 — 7:00 PM

1240 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. Write regular expression definitions for the following token classes:

(a) (10 points)

Unsigned integer literals that represent integers evenly divisible by 4. That is, 0, 8, 120, and 1000000 are allowed, but 1, 007 and 123 are not allowed.

(b) (15 points)

A C/Java-style multi-line comment that begins with `/*` and ends with `*/`. Within the body of the comment neither `*/` nor `/*` may appear. Thus

```
/* Compute a = b/c*d */
```

is OK, but

```
/* /* Incorrectly nested comment?? */
```

and

```
/* Incorrectly terminated comment?? */ */
```

are not allowed.

2. (25 points)

A **palindrome** is a word that reads the same forwards or backwards. Examples include `otto`, `abccba` and `bbbbbb`.

Let  $P$  be the set of all palindromes built using only lowercase letters. Create a regular expression or finite automaton for  $P$  or show that no such regular expression or finite automaton is possible.

3. (a) (15 points)  
Let FA be any finite automaton (my choice). Explain how to decide whether **all** the strings FA accepts always have at least two a's somewhere within them.

(b) (10 points)  
Show that there exists some deterministic finite automaton that **must have** two or more final states.

4. Write JLex regular expression definitions that match the following strings

(a) (7 points)  
The four characters: "\n"

(b) (7 points)  
Any odd number of backslash characters (e.g., \ or \\ or \\\ or \\\ \, etc.).

(c) (11 points)  
A CSX multi-line comment, delimited by { and }, that is allowed to contain *no more* than two new-line characters. That is, the comment may appear entirely on one line, or it may span two or three lines, but no more than three lines.

5. (25 points)  
Assume a reduced grammar, G. Define CanGen(A), where A is a nonterminal, to be the set of terminals, b, such that  $A \Rightarrow^+ \dots b \dots$ . That is, CanGen(A) contains all the terminal symbols that may be derived starting with A.

Give an algorithm that computes CanGen(A) given grammar G.

6. (a) (8 points)  
Consider the following context free grammar:

$$\begin{aligned} S &\rightarrow L a b \\ S &\rightarrow L a c \\ L &\rightarrow \lambda \end{aligned}$$

Is this grammar LL(1)? Why? Is this grammar LALR(1)? Why?

(b) (8 points)  
Consider the following context free grammar:

$$\begin{aligned} S &\rightarrow L S a \\ S &\rightarrow b \\ L &\rightarrow \lambda \end{aligned}$$

Is this grammar LL(1)? Why? Is this grammar LALR(1)? Why?

(c) (9 points)  
Consider the following context free grammar:

$$\begin{aligned} S &\rightarrow L a S \\ S &\rightarrow b \\ L &\rightarrow \lambda \end{aligned}$$

Is this grammar LL(1)? Why? Is this grammar LALR(1)? Why?