CS 352 – Compilers: Principles and Practice
Mid-term Examination, 3/6/08

Instructions: Read carefully through the whole exam first and plan your time. Note the relative weight of each question and part (as a percentage of the score for the whole exam). The total points is 100 (your grade will be the percentage of your answers that are correct).

This exam is **closed book, closed notes**. You may *not* refer to any book or other materials.

You have **75 minutes** to complete both (2) questions. Write your answers on this paper (use both sides if necessary).

Name:

Student Number:

Signature

1. [Regular expressions, scanning; 25%] C C integer literals are formed in the following way:

   - *Octal* literals consist of a leading `0` followed by zero or more octal digits (`0` through `7`)
   - *Hexadecimal* literals consist of a leading `0` followed by `x` or `X` followed by one or more hexadecimal digits ((`0` through `9`, `a` through `f`, or `A` through `F`)
   - *Decimal* literals consist of one non-zero decimal digit (`1` through `9`) followed by zero or more decimal digits (`0` through `9`).

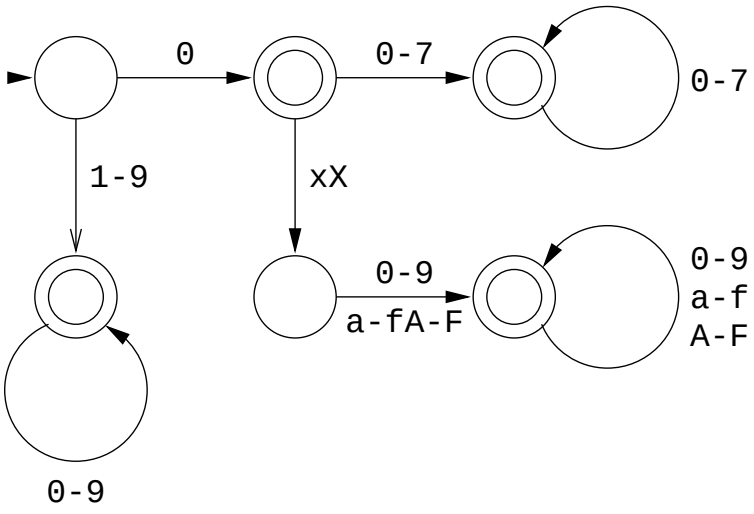
   (a) (10%) Give a regular expression for C integer literals.

   Answer:

   $$0(0-7)^* \mid 0(x \mid X)(0-9 \mid a-f \mid A-F)^+ \mid (1-9)(0-9)^*$$

   (b) (15%) By inspection (I don't expect you to go through any NFA-DFA construction to do this), draw the state diagram of a DFA (not an NFA!) for this literal form.

   Answer:



2. [Context Free Grammars, parsing; 75%] Consider the following simple context free grammars:

Grammar G_1	Grammar G_2	Grammar G_3
$S \rightarrow A$	$S \rightarrow A$	$S \rightarrow A$
$A \rightarrow \varepsilon$	$A \rightarrow \varepsilon$	$A \rightarrow \varepsilon$
$A \rightarrow bbA$	$A \rightarrow bAb$	$A \rightarrow Abb$

The start symbols are S, the non-terminals are S and A, and b is a terminal symbol. You may also treat \$ as a terminal to represent end-of-file. Note that these grammars all generate the same language $= L(G_1) = L(G_2) = L(G_3)$: strings consisting of even numbers of b symbols (including 0 of them).

(a) (20%) Construct the LL(1) parse table for G_1, showing FIRST and FOLLOW sets, and the productions predicted by each terminal symbol. Is G_1 LL(1)?

Answer:

	FIRST	FOLLOW	b	\$
S	$\{b,\$\}$	$\{\$\}$	$S \rightarrow A\$$	$S \rightarrow A\$$
A	$\{b,\$\}$	$\{\$\}$	$A \rightarrow bbA$	$A \rightarrow \varepsilon$

Since there are no multiply defined entries in the LL(1) parse table G_1 is LL(1).

(b) (5%) Is G_2 LL(1)? Why or why not?

Answer:

G_2 is not LL(1) since b predicts both $A \rightarrow \varepsilon$ and $A \rightarrow bAb$.

(c) (5%) Is G_3 LL(1)? Why or why not?

Answer:

G_3 is not LL(1) since it is left-recursive.

(d) (10%) Show the steps of a top-down LL parser as it parses the input *bbbb* using the grammar G_1, showing the input as it is consumed and the parse stack at each step of the parse. You can use your parse table from question 2a if that helps.

Answer:

Stack (top is left)	Input
S	$bbbb\$$
$A\$$	$bbbb\$$
$bbA\$$	$bbbb\$$
$bA\$$	$bbb\$$
$A\$$	$bb\$$
$bbA\$$	$bb\$$
$bA\$$	$b\$$
$A\$$	\$
\$	\$

(e) (20%) Construct the SLR(1) parse table for G_3, showing the LR(0) item sets, action entries, and goto entries. Is G_3 SLR(1)?

Answer:

Grammar G_3	
1	$S \rightarrow A$
2	$A \rightarrow \varepsilon$
3	$A \rightarrow Abb$

$0: \; S \rightarrow \bullet A$
$\quad A \rightarrow \bullet$
$\quad A \rightarrow \bullet Abb$
$1: \; S \rightarrow A\bullet$
$\quad A \rightarrow A \bullet bb$
$2: \; A \rightarrow Ab \bullet b$
$3: \; A \rightarrow Abb\bullet$

	b	\$	S	A
0	$r2$	$r2$		1
1	$s2$	a		
2	$s3$			
3	$r3$	$r3$		

(f) (10%) Show the steps of a bottom-up LR parser as it parses the input *bbbb* using the grammar G_3, showing the input as it is consumed, the parse stack at each step of the parse, and the action applied at each step. You can use your parse table from question 2e if that helps, in which case your stack elements will be parse table states. Alternatively, you can show a generic LR parser as it executes, with stack elements that are grammar symbols.

Answer:

Stack (top is right)	Input	Action
\$	*bbbb*\$	$r2$
\$$A$	*bbbb*\$	s
\$$Ab$	*bbb*\$	s
\$$Abb$	*bb*\$	$r3$
\$$A$	*bb*\$	s
\$$Ab$	*b*\$	s
\$$Abb$	\$	$r3$
\$$A$	\$	$r1$ or a
\$$S$	\$	a

or

Stack (top is right)	Input	Action
\$0	*bbbb*\$	$r2$
\$01	*bbbb*\$	$s2$
\$012	*bbb*\$	$s3$
\$0123	*bb*\$	$r3$
\$01	*bb*\$	$s2$
\$012	*b*\$	$s3$
\$0123	\$	$r3$
\$01	\$	a

(g) (5%) Of the *language* classes we have discussed in the course, what is the *smallest* category into which L fits? Justify your answer?

Answer:

The language is regular and can be defined by the regular expression $(bb)^*$.