MA/CSSE 474 Exam 3 Winter 2013-14 Name________________________ Section: 02(3rd) 03(4th)

Closed book and notes, except for three 8.5 x 11 sheets of paper (can be 2-sided). Also the description of the TM macro language from session 29.

No electronic devices, especially ones with headphones.

Scores:

Problem	Possible	Score
1	28	
2	30	
3	20	
4	10	
5	10	
6	15	
7	15	
8	8	
9	7	
10	20	
11	7	
Total	170	

1. (28 points) For each of the following statements, circle T or F to indicate whether it is *True* or *False*. If it is sometimes False, you should choose *False*. You do not have to give proofs or counterexamples. For each part, you get 2 points for circling *IDK* (I don't know), 4 for circling the correct answer, 0 for leaving it blank, and -1for circling the incorrect answer.

a) T F IDK Some languages are not decidable, but every language is semi-decidable.
b) T F IDK If R is regular and $R \cap L$ is not context-free, then L is not context-free.
c) T F IDK The complement of every context-free language is non-context-free.
d) T F IDK Every context-free language is decidable.
e) T F IDK The set of decidable languages is closed under complement.
f) T F IDK The set of semidecidable languages is closed under complement.
g) T F IDK The set of non-semidecidable languages is closed under complement.

2. (30 points) For each of the following languages, circle
R if the language is Regular,
CF if it is Context-free but not Regular,
D if it is Decidable but not Context-free
SD if it is Semidecidable but not Decidable
¬SD if it is *not* Semidecidable
IDK if you don't know.

Scoring: Correct answer 3, IDK 1, incorrect answer 0.

a) R CF D SD ¬SD IDK $WW^R = \{ww^R : w \in \{a,b\}^*\}$.

b) R CF D SD ¬SD IDK { <M> : where M is a TM with 5 states, and tape alphabet {☐, a}}.

c) R CF D SD ¬SD IDK $\{a^n : n \geq 0\}$.

d) R CF D SD ¬SD IDK $\{a^nb^n : n \geq 0\}$.

e) R CF D SD ¬SD IDK $\{a^nb^nb^n : n \geq 0\}$.

f) R CF D SD ¬SD IDK {<M, w> : TM halts when started with input w}.

g) R CF D SD ¬SD IDK {<M, w> : TM does not halt when started with input w}.

h) R CF D SD ¬SD IDK $\{wxw^R : w,x \in \{a,b\}^+\}$.

i) R CF D SD ¬SD IDK $L(G)$, where $G = \{S \rightarrow TSb \mid S \rightarrow Tb, T \rightarrow Ta, T \rightarrow \varepsilon\}$.

j) R CF D SD ¬SD IDK $\{x\#y\#z : x, y, z \in \{a, b\}^+$ and $(|x| - |y| = 2$ or $|y| - |z| = 2\}$.

3. (20 points) Let G be the grammar $S \rightarrow SS \mid (S) \mid \varepsilon$.

L(G) is the language *Bal* of all strings of balanced parentheses; that is, those strings that could appear in a well-formed arithmetic expression. We may want to show that L(G) = *Bal*, which requires two inductive proofs:

If w is in L(G), then w is in *Bal*.

If w is in *Bal*, then w is in L(G).

We prove only the first part. You will see below the sequence of steps in the proof, each with the reason for its validity omitted. These reasons belong to one of three classes. You do not have to write out reasons; just **correctly choose and write A, B, or C in each blank**. There are ten such blanks.

A) Use of the induction hypothesis.

B) Reasoning about properties of grammars, e.g., that every derivation has at least one step.

C) Reasoning about properties of strings, e.g., that every string is longer than any of its proper substrings.

The proof is by induction on the number of steps in the derivation of w.

Base: One step. The only 1-step derivation of a terminal string is $S \Rightarrow \varepsilon$ because ________

ε is in *Bal* because ________

Induction: An n-step derivation for some n>1.

The derivation $S \Rightarrow^{n} w$ is either of the form

(a) $S \Rightarrow SS \Rightarrow^{n-1} w$ or of the form

(b) $S \Rightarrow (S) \Rightarrow^{n-1} w$

because ________

Case (a):

$w = xy$, for some strings x and y such that $S \Rightarrow^{p} x$ and $S \Rightarrow^{q} y$,

where $p<n$ and $q<n$ because ________

x is in *Bal* because ________

y is in *Bal* because ________

w is in *Bal* because ________

Case (b):

$w = (z)$ for some string z such that $S \Rightarrow^{n-1} z$ because ________

z is in *Bal* because ________

w is in *Bal* because ________

4. (10 points) What are the two ways that a nonterminal symbol in a CFG can be *useless*?

a)

b)

5. (10 points) Here is a context-free grammar G:

$S \rightarrow AB \mid CD$
$A \rightarrow BG \mid 0$
$B \rightarrow AD \mid \varepsilon$
$C \rightarrow CD \mid 1$
$D \rightarrow BB \mid E$
$E \rightarrow AF \mid B1$
$F \rightarrow EG \mid 0C$
$G \rightarrow AG \mid BD$

In the list below, Circle each nullable symbol of G and do not circle non-nullable symbols.

S A B C D E F G 0 1

6. (15 points) Here are eight simple grammars, each of which generates an infinite language of strings. These strings tend to look like alternating *a*'s and *b*'s, although there are some exceptions, and not all of the grammars generate all such strings.

a) S → abS | ab
b) S → SS | ab
c) S → aB; B → bS | a
d) S → aB; B → bS | b
e) S → aB; B → bS | ab
f) S → aB | b; B → bS
g) S → aB | a; B → bS
h) S → aB | ab; B → bS

For each of the following grammar pairs, circle Y if those two grammars generate the same language, and N if they do not.

Y N a and f

Y N a and b

Y N c and f

Y N d and g

Y N e and h

7. (15 points) The language L = {*ss* | *s* is a string of *a*'s and *b*'s} is not a context-free language. In order to prove this, we need to show that for every integer *k*, there is some string *w* in L, of length at least *k*, so that no matter how we break *w* up as *w*=*uvxyz*, subject to the constraints $|vxy| \leq k$ and $|vy| > 0$, there is some $q \geq 0$ such that uv^qxy^qz is not in L. Let us focus on a particular *k*=7 and *w*=*aabaaaba*. It turns out that this is not a good choice of *w* for *k*=7, since there are some ways to break *w* up for which we can find the desired *q*, and other ways of breaking it up for which we cannot. Identify from the list below the choices of *u*,*v*,*x*,*y*,*z* for which there is an *q* that makes uv^qxy^qz **not** be in L.

We show the breakup of *aabaaaba* by placing four |'s among the *a*'s and *b*'s. The resulting five pieces (some of which may be empty) are the five strings. For instance, aa|b||aaaba| means *u*=aa, *v*=b, *x*=ε, *y*=aaaba, and *z*=ε.

For each way of breaking up the string, circle Y if this way of breaking up the string has a way of pumping that takes us outside of L, and N if every choice of q keeps the pumped string in L.

Y N aab|a|aab|a|

Y N aa|ba|aa|ba|

Y N |aa|b|aa|aba

Y N aab|a|a|a|ba

Y N |a|abaa|a|ba

8. (8 points) Consider the pushdown automaton with the following transition rules:

1. $\delta(q, \varepsilon, \varepsilon) = \{(q, Z_0)\}$
2. $\delta(q,0,Z_0) = \{(q,XZ_0)\}$
3. $\delta(q,0,X) = \{(q,XX)\}$
4. $\delta(q,1,X) = \{(q,X)\}$
5. $\delta(q,\varepsilon,X) = \{(p,\varepsilon)\}$
6. $\delta(p,\varepsilon,X) = \{(p,\varepsilon)\}$
7. $\delta(p,1,X) = \{(p,XX)\}$
8. $\delta(p,1,Z_0) = \{(p,\varepsilon)\}$

The start state is *q*. For which of the following inputs can the PDA enter state *p* for the first time with the input empty and the stack containing XXZ_0 [i.e., the configuration (p,ε,XXZ_0)]? Mark the correct such input. Only one is correct.

___ 0101010

___ 011011011

___ 011001101

___ 1001101

9. (7 points) Start with the grammar G:

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

If we use the nondeterministic top-down parsing algorithm from class (and the textbook) to convert this grammar to a PDA, which of the following transitions are part of the PDA? Circle Y if the given transition is part of the PDS, and N if it is not. p is the start sate, q is the accepting state (don't forget that in order to accept a string w, the stack must also be empty after reading w).
Notation: ((r, a, b), (t, d) means "in state a, read b from input, pop c off the stack, push d, go to state d"

Y N ((p, ε, ε),(q, S))

Y N ((p, ε, S) , (q, ε))

Y N ((q, ε, S) , (q, AS))

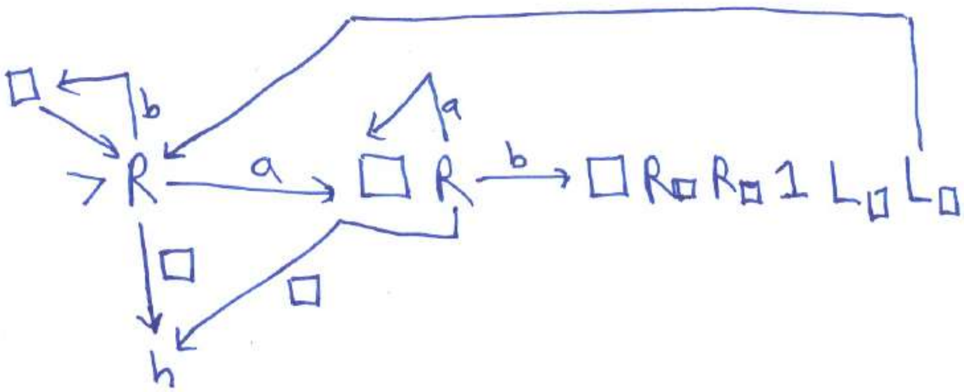
Y N ((q, ε, A) , (q, A0))

Y N ((q, ε, A) , (q, 1))

Y N ((q, 1, 1) , (q, ε))

Y N ((q, 1, ε) , (q, 1))

10. (20 points) Consider the following TM M. Try to figure out what M does when started with a string of a's and b's (you do not have to *describe* what it does in general, but understanding that will help you answer the questions. For each of the two inputs below, if M is started with that input on its tape, show the tape contents after M halts.



Recall that a TM always starts with its read/write head at the last blank square before the input string.

Input: baabaaabaaaab Output: ___________________________

Input: abbabbbabbbba Output: ___________________________

11. (7 points) Where does the "k" in the statement of the Pumping Theorem for context-free languages come from? [Hint: for a regular language, k is the number of states in a DFSM that recognizes the language.]

A brief answer (two or three sentences) is probably sufficient to let me know you have the idea.