Problem Set 7

Fall 11

Due: Thursday, 1st December, 2011, 11:00 am before class begins

Please follow the homework format guidelines posted on the class web page:

`http://www.cs.illinois.edu/class/fa11/cs373/`

1. [**Category**: Aliens, **Points**: 5]

   Prove that the problem of deciding whether there are aliens is decidable or undecidable. More precisely, is there a TM that will take as input "Are there aliens?" and accepts it if there are aliens, and rejects it if there aren't any. (The TM rejects all other strings.)

2. [**Category**: Undecidability, **Points**: 20]

   Let L be the set of all encoding of Turing machines and words, $\langle M, w \rangle$ such that M when run on w at some point moves right for three consecutive steps. Prove that L is undecidable.

3. [**Category**: CFG design, **Points**: 20]

   Consider well-formed arithmetic expressions on numbers with four binary operators $\{+, -, *, /\}$ and one unary operator $\{-\}$ (negative sign). A number is any string over $\{0, 1, \ldots 9\}$ (starting with 0s is fine). To avoid ambiguity, consider expressions which are parenthesized every time an operation is used. Design a context-free grammar for arithmetic expressions. That is construct a grammar G such that $L(G)$ is the set of all valid arithmetic expressions.

   Here are three examples that should be in $L(G)$:

   $((((1335 + 21) * 3222) - 431)/565)$

   $(745 - (-((003 - (101 + 134545452)) + (345 - 4453))))$

   $(1/0)$

   Here are five examples that should NOT be in $L(G)$:

   $(1 + 2 - 3 * 4)$

   $(1 - -(2 + 3))$

   $(2 + 3($

   $1 + 2$

   $((1) + 2)$

   After your construction, show the following two strings are valid arithmetic expressions by explicitly showing every yield step of applying rules in G.

   (a) $(2 + (-(1 * 3)))$

   (b) $((4/5) + (5 * (6 + 7)))$

4. [**Category**: CFG Design, **Points**: 20]

We want to show that a subset of HTML documents is a context-free language. For our purposes, we will consider a subset of HTML restricted to the tags: `html, body, ul, li`. In particular, the document must have the open tags and close tags matched properly, and satisfy the following conditions:

- The document must start with an open `<html>` tag and close with `</html>` and there should be no other `html` tag and all text must be contained within these tags.
- There is only one open `body` tag (and its matching close tag)
- All `ul` tags occur within the `body` block. There can be any number of `ul` blocks, and all `li` must blocks occur within an immediate `ul` block. A `ul` block need not have any `li` blocks within it.
- There can be text anywhere within the `<html>` block, between any tags.
- Text is any sequence of a–z, A–Z, and the space character.

Hence such documents start with the `html` tag followed by some text followed by a `body` block. The `body` block consists of nested `ul` blocks that have sequences of `li` blocks, and text in between the tags.

For example, the following is a well-formed document:

```
<html> Heading <body> Blah Blah <ul><li> first item </li><li>second</li>
<ul><li>This is nested at second level</li></ul></ul></body> </html>
```

5. CNF Conversion [**Category**: Proof., **Points**: 7+7+6] Consider the grammar G:

$$\begin{aligned} S &\rightarrow 0A0 \mid 1B1 \mid BB \\ A &\rightarrow C \\ B &\rightarrow S \mid A \\ C &\rightarrow S \mid \epsilon \end{aligned}$$

(a) First, add a rule $S_0 \rightarrow S$ to G and eliminate ϵ-productions, obtaining G_1. Write down precisely the set of nullable variables, and the resulting grammar G_1.

(b) Eliminate any unit productions in G_1, obtaining G_2. Write down precisely the set of all transitive unit derivations, and the resulting grammar G_2.

(c) Put G_2 into Chomsky Normal Form G_3.

6. CYK [**Category**: Comprehension, **Points**: 20]

Use CYK algorithm to determine whether or not the given string belongs to the grammar. Your answer should include either "yes" or "no" and a chart that you built using CYK.

You are required to use the CYK algorithm; do not just give a derivation or an argument as to why the word does not belong to the language.

Determine whether the string (i) *aabbbb* and (ii) *aabaab* belong to the language.

$$S \longrightarrow AP \mid AB$$
$$E \longrightarrow AP \mid EB \mid b$$
$$P \longrightarrow EB$$
$$A \longrightarrow a$$
$$B \longrightarrow b$$