CSC488S/CSC2107S Compilers and Interpreters

Mid Term Test Solution

1. [10 marks] The main issue here is that the lexical analyzer has to process on order of magnitude more text than in the traditional case. Would need some very efficient character processing, probably hand written.

Some of the classification of token types is done in the XML, but the lexical analyzer would still have to classify the various kinds of <reserved> words.

If the lexical analyzer can't (shouldn't) assume perfectly formed XML input, then it will have to validate the XML, an extra level of processing.

A possible solution might be to just tokenize everything, and let the parser sort it out. e.g. with syntax rules like:

if ::= RESERVED IF UNRESERVED

But a careful analysis of processing efficiency would be required.

2. [20 marks] a) Yes the grammar is LL(1)

b)

			Predict Sets
1	$V \rightarrow$	WX'c'	first(W) + first(X'c')
1	V	W X C	
	\rightarrow		$\{b\} \cup first(\mathbf{X})$
	\rightarrow		$\set{b} \cup \set{a,c,d}$
	\rightarrow		{ a , b, c, d }
2	\rightarrow	'e' Y W	{ e }
3	$W \rightarrow$	'b' X Y Z	{b}
4	\rightarrow	λ	follow(W)
	\rightarrow		$first(X) \cup follow(V) \cup follow(X)$
			Rule: 1 2 5
	\rightarrow		$\{a,c,d\} \cup \{f,\$\} \cup \{a,c,d,e,f,\$\}$
	\rightarrow		{ a, c, d, e, f, \$ }
5	$X \rightarrow $	Y'a'W	$first(Y) \cup \{a\}$
	\rightarrow		$\{d\} \cup \{a\}$
	\rightarrow		{ a, d }
6	\rightarrow	'c' 'a'	{ c }
7	$Y\rightarrow$	'd' Y	d j
8	\rightarrow	λ	follow(Y)
	\rightarrow		$first(W) \cup follow(V) \cup first(Z) \cup \{a\}$
			Rule 2 2,4 3 5
	\rightarrow		$\{b\} \cup \{f, \$\} \cup \{e, c\} \cup \{a\}$
	\rightarrow		{ a, b, c, e, f, \$ }
9	$Z\rightarrow$	'e' V 'f'	{ e }
10	\rightarrow	'c'	{c}

Many solutions missed including \$ in the predictor sets

Some solutions wrongly assumed that the Predict sets had to be disjoint across all rules, not just the rules for a particular non-terminal.

3. [20 marks]



Most common problem was incompleteness. Missing entries, missing links.

4. [20 marks]

a) Storing identifier string in the symbol table would make each symbol table entry large which is undesirable. Need separate table for identifier names. Expect most identifiers to be short, saving space for largest possible identifier is wasteful

b) Makes processing the language more difficult. Lexical analysis can only identify *potential* reserved words. Syntax analysis has to have enough power to distinguish reserved words from keywords.

c) Need extra mechanisms for handling long list of initial values. Perhaps store them in a separate table. Need to do semantic analysis on the initial values to ensure type correctness.

d) Requires a more powerful parsing technique, perhaps larger parse tables. Recursive descent parser might be a good choice, only use extended lookahead when it is necessary.

e) Requires multiple pass semantic analysis. First pass captures all the declarations, second pass checks usage.

Note: many answers assumed that the variable was never declared, but that's not what the question said.

5. [20 marks]

First layout the inner structure marks

assignments	w midterm w f final					
80	8 8 32 8 24 32					
align 0 mod 32	length 192 bits 24 bytes					

Then layout the outer structure

name		number	g r d	marks		rawMark	m r	course mark
160	32	64	8 24	192	32	64	8 24	32
align 0 mod 64	length 64	40 bits	80 bytes					

One common problem was forgetting to include the inner structure marks in the outer structure.