CS 164

- Please read all instructions (including these) carefully.
- Please print your name at the bottom of each page on the exam.
- There are seven questions on the exam, each worth between 10 and 20 points. You have 1 hour and 20 minutes to work on the exam, so you should plan to spend approximately 12 minutes on each question.

Midterm Examination

- The exam is closed book, but you may refer to your two sheets of prepared notes.
- Please write your answers in the space provided on the exam, and clearly mark your solutions. You may use the backs of the exam pages as scratch paper. Please do not use any additional scratch paper.
- Solutions will be graded on correctness and clarity. There are no 'tricky' problems on the exam-each problem has a relatively simple and straightforward solution. You may get as few as 0 points for a question if your solution is far more complicated than necessary.

NAME:

SID or SS#

Problem, Max points, Points

1 15
2 15
3 20
4 10
5 20
6 10
7 10

CS 164 Handout 11 1. Regular Expressions (15 points) Consider a language where real constants are defined as follows: A real constant contains a decimal point or E notation, or both. For instance, 0.01, 2.71821828, -1.2E12, and 7E-5 are real constants. The symbol '-' denotes unary minus and may appear before the number or on the exponent. There is no unary '+' operation. There must be at least one digit to left of the decimal
point, but there might be no digits to the right of the decimal point. The exponent following the "E" is a (possibly negative) integer. Write a regular expression for such real constants. Use the standard regular expression notation described by the following grammar: \( R \rightarrow (\epsilon | \text{char}\{R+R|R^*|RR|((R)) \) You may define names for regular expressions you want to use in more than one place (e.g., \( \text{foo} = R \)).

\[
\begin{align*}
digit &= 0+1+2+3+4+5+6+7+8+9 \\
\text{posint} &= digit \text{ digit}* \\
\text{int} &= ((\epsilon+\ -) \text{ posint} \ \text{exp} = E \text{ int frac} = .digit* \ \text{real} = (\text{int frac} (\text{exp} + \epsilon)) + (\text{int} (\text{frac} + \epsilon) \ \text{exp})
\end{align*}
\]

Fall 94 page 2 of 8

CS 164 handout 11

2. Finite Automata (15 poilats)

Consider a DFA with a start state so and a transition function \( \text{trans} \). For a state \( s \) and input character \( c \), \( \text{trans}(s,c) = s' \) if there is a transition from state \( s \) to state \( s' \) on character \( c \). If there is no transition from \( s \) on \( c \) then \( \text{trans}(s,C) = \text{none} \). The following algorithm simulates the behavior of such a DFA, accepting if the input string is accepted by the DFA and rejecting otherwise.

\[
\begin{align*}
\text{state} &\rightarrow \text{so} \\
\text{while there's input left do:} & \\
\text{char} &\rightarrow \text{next input character} \\
\text{if } \text{trans}(\text{state},\text{char}) &\rightarrow \text{none then stop and reject} \\
\text{state} &\leftarrow \text{trans}(\text{state},\text{char}) \\
\text{(end of loop)} & \\
\text{accept if state is a final state, otherwise reject}
\end{align*}
\]

Now consider an NFA with a start state so and a transition function \( \text{trans} \). In this case, for a state \( s \) and input character \( c \) (we now allow \( c = \epsilon \)), \( \text{trans}(s,c) \) is the set of states \( s' \) for which there is a transition from \( s \) to \( s' \) on \( c \). In the style of the algorithm above, give a (deterministic) algorithm that simulates the behavior of such an NFA. You may use the \( \epsilon \)-closure operation described in class and in the text.

Simulate subset construction.

Use state-set instead of state,

\[
\begin{align*}
\text{state-set} &\rightarrow c-\text{closure}(\text{so}) \\
\text{while there's still input left do:} & \\
\text{char} &\rightarrow \text{next input character} \\
\text{if } \epsilon-\text{closure}(\text{Union s subset state-set } \text{trans}(s, \text{char})) &\rightarrow 0 \text{ then stop and reject} \\
\text{state-set} &\leftarrow \epsilon-\text{closure}(\text{Union s subset state-set } \text{trans}(s, \text{char}))
\end{align*}
\]
3. Grammars (20 points)

Consider the following grammar. The nonterminals are E, T, and L. The terminals are +, id, (,), and ;. The start symbol is E.

E ---> i- E+T|T
T ---> id|id()|id(L)
L ---> E;L

Give an LL(1) grammar that generates the same language as this grammar. As part of your work please show that your grammar is LL(1).

(a) Eliminate left recursion:

E ---> TE
E' ---> +TE | epsilon
T ---> id | id() | id(L)
L ---> E; L

(b) Left factor:

E ---> TE'
E' ---> +TE | epsilon
T ---> id | id(L)
T' ---> epsilon | T''
T'' ---> ) | L)
L ---> EL'
L' ---> ; L epsilon

(c) Check that it's LL(l). For this part you just needed to give enough information to show that there would be no conflicts in the parsing table. The following is sufficient:

E --->+TE'epsilon First(+TE') = {+}

Follow(E) = {$,;}$

T'--->epsilon |(T" First((T")=()

Follow (T)={+,.,,;}

T"--->)(L) First()) =()

Follow (L)={id}

L--->;Lepsilon First(;L) = {;}

Follow(L)={}

Fall 94 page 4 of 8

CS 164 Handout 11

4. Parsing (10 points)

In both parts of this question, we are looking for clarity and brevity as well as the right idea. Suppose you are writing a parser for a programming language that includes the following syntax for looping constructs:

Loop --> do stmt while expr

| do stmt until expr

| do stmt forever

(a) (5 points) A predictive parser (i.e., a top-down parser with no backtracking) can't use this grammar. Give a brief (a couple of sentences) explanation of this fact.

When trying to expand a production for non-terminal "loop I I , the parser cannot decide which of the three productions to expand using only the next few input tokens.

(b) (5 points) Give a brief explanation of why a bottom-up parser does not have difficulty with this grammar.

When a bottom-up parser must decide which of the three productions to choose (reduce), the "while", "until", or "forever" has already been read and shifted onto the stack.
5. Parsing (20 points)

Consider the following grammar. The nonterminals are S' and S. The terminals are op and x. The start symbol is S'.

S' --> S
S --> S OP S|X

(a) (15 points) Draw the DFA built from sets of LR(O) items for this grammar. Show the contents of each state. (Note: Don't augment the grammar with a new start symbol.)

(b) (3 points) Is this grammar SLR(1)? Briefly explain why or why not.

(c) (2 points) Is this grammar LR(1)? Briefly explain why or why not.

6. Bison and Abstract Syntax 'JPrees (10 points)

Consider the following constructors for a tree language:

Expression app(Expression, Expression); Expression lambda(Expression, Expression); Expression ido;

Now consider the following Bison grammar:

%token ID LAMBDA

%type <Expression> Expr

%%

Expr : ID

{ $$ = id() ; }

| '(' LAMBDA ID '.' Expr ')'
Draw the abstract syntax tree that would be produced when parsing the sequence of tokens below. Label aB the nodes of your AST with the appropriate constructor.

\[
( ( \text{LAMBDA ID} . (\text{ID ID}) ) ( \text{LAMBDA ID} . \text{ID} ) )
\]

Self-application occurs when a function is called with itself as a parameter. This is, if \( f \) is a function taking a function as a parameter, the \( f(f) \) is an instance of self-application. Briefly explain why type checking of self-application must always fail using the type checking rule above.
Posted by HKN (Electrical Engineering and Computer Science Honor Society)
University of California at Berkeley
If you have any questions about these online exams
please contact examfile@hkn.eecs.berkeley.edu.