[redacted]

MIDTERM TEST

CST8152 – COMPILERS – Section 10

STUDENT NAME: [redacted]

STUDENT ID:

TEST INSTRUCTIONS:

The Midterm Test consists of 30 questions. At the end of test there are three bonus questions. You may replace with the bonus question any one of the preceding compulsory questions.

Every properly answered question is worth **1.0**. The test counts for 30% of your final grade.

Read each question carefully before you answer. You have 55 minutes to complete the test and 5 minutes to submit on Brightspace. Late tests will not be accepted.

Work at a steady pace and you will have ample time to finish. When answering the true-false and multiple-choice questions you must click on the Answer box under the question and select the appropriate letter. If the question have a text field, must type your answer in the text filed. If the question is indicated with an **M,** it has more than one correct answer. In the Multiple Answers text field you have to enter the letter which are in front of the appropriate options (for example: c, d, e).

Save your work time to time.

When you are finished save your document.

Next step is to sign the finished test. If you do not know how to do that,
the SigningMidterm.pdf document (posted on Brightspace) contains instruction how to do that.

Once you sign the Midterm you have to upload the signed pdf on Brightspace.

Good luck, and do not forget that:

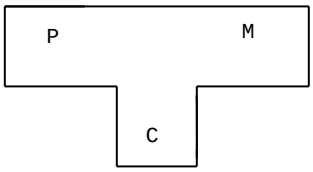
“ THE PROMISE OF THE FUTURE LIES NOT IN TECHNOLOGY BUT IN YOU!”

Professor: Svillen Ranev,
March 2020

1. Which one of the following is a list of typical compiler components?
   a. lexical analyzer, syntax analyzer, semantic analyzer, code optimizer
   b. preprocessor, scanner, semantic analyzer, code generator
   c. lexical analyzer, parser, assembler, symbol table
   d. lexical analyzer, syntax analyzer, loader/linker, code optimizer
   e. all of the above are components

Answer: e

2. Here is a T-shaped tombstone diagram, where P, M and C are different languages.



   The diagram above represents
   a. an interpreter written in language P.
   b. a compiler written in language C.
   c. a compiler written in language P.
   d. an interpreter written in language C.

Answer: b

3. The compiler front end includes these parts (phases) of a compiler that depend on the source language, and are in general independent of the target machine.
   a. True b. False

Answer: a

4. Several *phases* of compilation are usually implemented in a single *pass* consisting of reading an input file and writing an output file.
   a. True b. False

Answer: a

5. An interpreter is a compiler that does not produce a target program as a result of the translation.
   a. True b. False

Answer: a

6. A context free grammar has four components

   1. A set of terminals, known as token symbols.
   2. A set of non-terminals.
   3. A set of production.
   4. A designation of one of the non-terminals as a start (goal) symbol.

7. Which of following is not a property of a parse tree?
   a. Each leaf represents one production.
   b. Each leaf is labeled by a token.
   c. Each interior node is labeled by a nonterminal.
   e. The root is labeled by the start symbol.
   f. all of the above are properties.

Answer: a

8. What is an ambiguous grammar?
   A grammar that can generate
   different parse trees using the same parsing method from and input string

9. Given the grammar
   <exp> **->** <exp> +< exp> | <exp> * <exp >|< factor>
   <factor> **->** a | b | c | d | 0 | 1 | 2 | 3

   Write at least 5 strings that belong to the language defined by the grammar.
   a0, abb, ddddaaa, ddd22312, b2asda

10. When a compiler is written in the language it compiles, the process of writing (building) the compiler is called bootstrapping.

11. How do the parse trees for left-associative and right-associative operators differ?

    The parse tree for the right-associative operators grows towards the left,

    whereas the parse tree for the left-associative operators grows toward the right.

12. The word terminal is a synonym for token when we are talking about syntactic (syntax) grammars.

13. There are problems with the grammar described in question 9. What are the problems?
wrong order of precedence, Ambiguous

14. Every construct (set of strings) that can be described by regular expressions can also be described by context-free grammar.
a. True b. False

Answer: a

15. From scanner implementation point of view what is the best way to describe the lexical part of a language?
a. Regular grammar
b. Regular expressions
c. BNF grammar
d. Lexemes

Answer: b

16. A sequence of input characters that matches a pattern for a token is called a lexeme.

17. Given the following language construct
s = 2 * d + 1 / r
Define the language tokens and their attributes using the format: **{** *token description* **;** *attribute description* **;** *lexeme(s) found in the construct* **}**. For example: {assignment operator; no attribute; =}
assignment operator; s,r; no attributel 1.0, 2.0; floating point literal; No lexeme found}

18.**M** What is usually eliminated by the lexical analyzer and never sent to the parser?
a. White space
b. Comments
c. Lexemes representing numbers
d. Variable names
e. Lexemes representing string literals

Multiple Answers: A,B,C

19. Many lexemes can match a pattern for a token.
    a. True b. False

Answer: a

20. Which of the following is not an operation used in regular expressions?
    a. production
    b. closure
    c. concatenation
    d. union | alternation

Answer: a

21. Given the regular expressions defining a pattern for a token floating-point literal

    FPL **=** (+|-|ε)D^{*}. D^{+} E(+|-)?DD
    D **=** 0 | 1 | 2 | 3 | 4 | 5 |6 | 7 | 8 | 9
    write at least 5 structurally different strings that match the **FPL** token definition.

    DE1231, DE124

    ED321, ED3423

    01231D

22. When converting an NFA to a DFA, the resulting DFA will theoretically have more states than the original NFA.
    a. True b. False

Answer: a

23. A DFA is an NFA with two restrictions:

    1. Transitions that do not contain E(epsilon)
    2. Transitions from states must be unique

24. The entry for row ***s*** and symbol ***c*** in the transition table is always a single state for
    a. NFA
    b. DFA
    c. both NFA and DFA

Answer: a

25. Given the following few strings samples which belong to a set of strings:

    d db da dabd dbaabad daaabd dbbbdd dababababddd dddd

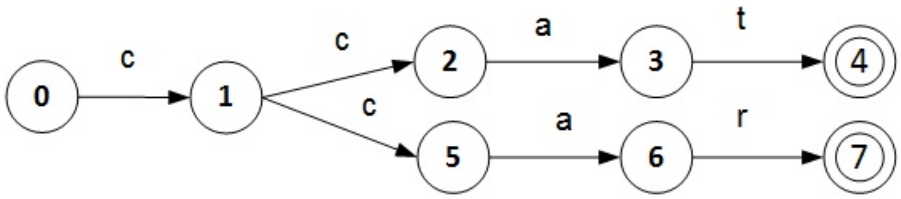
    write a regular expression that defines the whole set of strings.

    d+(a|b)*d+

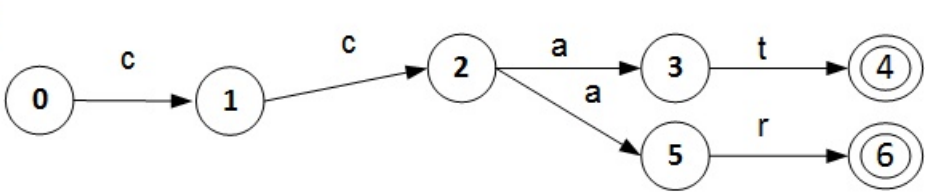
26. Given the regular expression **c(cat|car)**

    Which one of the following three transition diagrams that will recognize the strings defined by the regular expression is not a NFA?

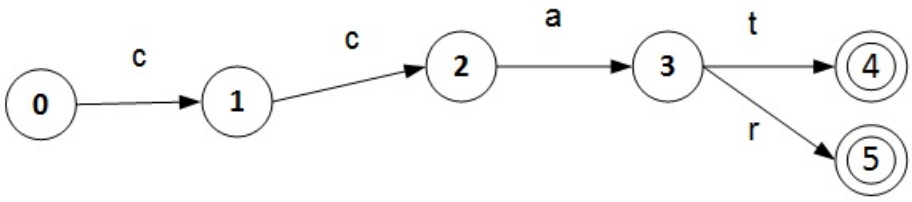
1



2



3



Answer: 3

The questions Q27 to Q30 are related and they will be evaluated and marked in pairs.
For example, even your regular expression have some minor errors, but the transition table reflects your regular expressions correctly, you shall still get full marks for the transition table. If the answers to some questions contain major mistakes (more than 50% wrong), the related questions cannot receive partial marks greater than 50%.

27. The following grammar defines an octal integer literal (**OIL**) :

<octal integer literal> **->** 00 **|** 0 <octal non zero digit><opt_octal digits>
<opt_octal digits> **->** <octal digits> **|** ε
<octal digits> **->** <octal digit> **|** <octal digits><octal digit>
<octal digit> **->** 0 **|** <octal non zero digit>
<octal non zero digit> **->** 1 **|** 2 **|** 3 **|** 4 **|** 5 **|** 6 **|** 7

Write at least 10 strings (lexemes) that belong to the language defined by the grammar.

0, 00, 01, 010, 04, 056, 0777, 000, 0123, 0900

Write at least 5 strings that cannot be generated by the grammar (illegal literals).

1, 10, 25, 09, 028

28. Write regular expressions that define the same set of strings (the same language) for the octal integer literal (OIL) as the set defined by the BNF grammar given in question 27. You should use regular definitions in the regular expressions. For example, the following regular definition
OnZ = 1 | …| 7 or OnZ = [1-7] , OD = [0-7] could be useful.

OIL = Z+ | O ONZ OOD* ONZ = [1-7] OD = [0-7]

29. Using the regular expression defining a pattern for octal integer literals (OIL) written in question 28, write at least 5 lexemes (strings) that can be matched by the OIL.

OIL = Z+10 ONZ OOD*

30. Create a transition table for OIL. You should first draw on paper the transition diagram (DFA) for OIL before populating the table.

State	Input Symbols			Accepting state type
	0	[1-7]	other	
0	1	ES	ES	NOAS
1	2	4	3	NOAS
2	2	ES/AS3	3	NOAS
3	IS	IS	IS	NOAS-OIL
4	4	IS	3	NOAS
5	IS	IS	IS	ANSR

BONUS QUESTIONS:

#1. Given the regular expression

S = (yada)$^+$ blah

write a grammar that describes the same language.

S = (yada)(blah

(yada)-><yada>|<yada><yada>yada->yada|E

[172]

#2. Write a regular expression describing the pattern for the comments as defined in the PLATYPUS language.

!! [^ CR]*

#3. Do you think that learning about compilers makes you a better programmer?
a. Yes
b. No
c. Cα&¶~C~β^-☺P - Encripted answer

Answer: a

FEEDBACK QUESTIONS (optional):

1. **The Midterm Test is** HARD
2. **The Final Test should be** EASIER
3. **Comments on the course** :)

CHECK YOUR ANSWERS AGAIN,
and
remember
that

"In examinations the foolish ask questions that the wise can not answer."
Oscar Wilde