CSCI3136
Assignment 1

Instructor: Alex Brodsky

Due: 3:00pm, Monday, January 20, 2014

1. Modern integrated development environments (IDEs) include syntax highlighting, which displays various parts of the source code in various fonts and colours depending on the language structure and terms being used. For example, variables may be highlighted in blue, keywords in green, function calls, in purple, etc.

   (a) **[7 marks]** Which parts of the compiler (compilation phases) would be needed to implement this functionality and why?

   The IDE would need to use the *lexical analysis*, *syntax analysis*, and possibly *semantic analysis* phases of the compiler depending on the language. [1]
   ***Lexical analysis* (a scanner) would be needed to tokenize the source code and identify the various types of tokens, e.g., keywords, literals, etc. [2]**
   ***Syntax analysis* (a parser) would be needed to identify the multi token language structures such as function calls, expressions, matching parentheses, braces, and quotes, etc. [2]**
   ***Semantic analysis* would be need to classify the various symbols in the source code, such as function names, variable names, class names, etc. [2]**

   (b) [3 marks] Suppose you only had time to implement one of these phases. Which one of these phases should be implemented? Be sure to justify your answer.

   The *lexical analysis* (a scanner) would be most beneficial [1] because the later stages rely on the scanner and cannot be implemented without it [1]. Furthermore, the scanner can be used to categorize most of the items in the source code that would be highlighted, such as comments and keywords.[1]

2. For each of the following languages give the regular expression that matches the strings in the language. Note: You can use . to denote "any" character in the alphabet.

   (a) **[5 marks]** The language, over the English alphabet, of strings that contain at least 2 vowels.

   $.*(a|e|i|o|u).*(a|e|i|o|u).*$

   (b) [5 marks] The language, over the alphabet of decimal digits, of strings that represent integers divisible by 25.

$.*(00|25|50|75)$

(c) [5 marks] The language of binary strings that do not contain three consecutive 1s.

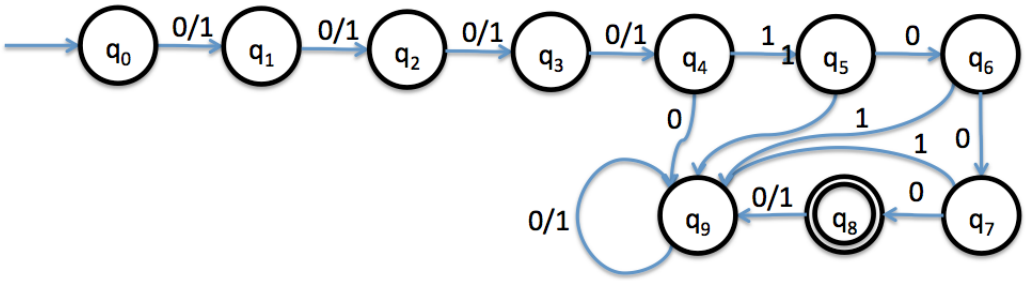
$0*(10|\epsilon)(10*100*)*(\epsilon|10*|10*10*)$

3. For each of the following langauges,

- State whether the language is finite or infinite.
- State whether the language is regular or nonregular.
- If you claim the language is regular: give a DFA (graphical representation) that recognizes the language.
- If you claim that the language is not regular, describe the intuition for why this is so.

Consider the following languages:

(a) **[8 marks]** The language of 8 bit binary strings that represent integers divisable by 8.
Finite and regular [1]

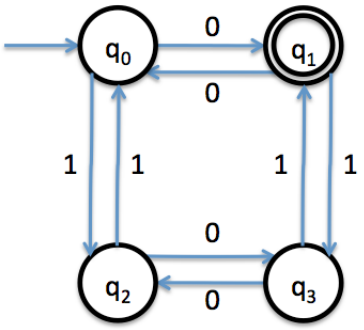


[Correct states: 3 marks, correct transitions 3 marks, Correc accepting state 1 mark]

(b) [8 marks] The language of binary strings such that the number of 0s is less than the number of 1s.
Infinite and not regular [2] mark] A DFA would need to keep track of the diffference betweeen the number of 0s read and the number of 1s read [2]. Since the difference can be arbitrarily large a any point during during a computation [2], a DFA would run out of states to keep track of the difference [2].

(c) [9 marks] The language, of binary strings with an odd number of 0s and an even number of 1s.

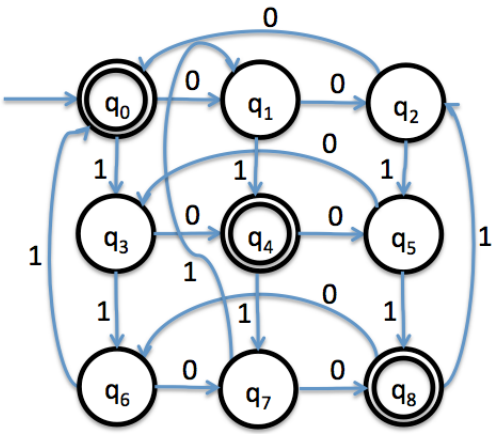
Infinite and regular [1 mark]



[Correct states: 3 marks, correct transitions 4 marks, Correct accepting state 1 mark]

(d) [Bonus 5 marks] The language of binary strings such that the number of 0s is congruent to the number of 1s modulo 3.

Infinite and regular [1 mark]



[Correct states: 2 marks, correct transitions 2 marks]

CSCI3136: Assignment 1

Winter 2014

Student Name	Login ID	Student Number	Student Signature

	Mark
Question 1	/10
Question 2	/15
Question 3	/25
Bonus	/5
Total	**/50**

Comments:

Assignments are due by 3:00pm on the due date before class and must include this cover page. Assignment *must* be submitted into the assignment boxes on the second floor of the Goldberg CS Building (by the elevators).

Plagiarism in assignment answers will not be tolerated. By submitting their answers to this assignment, the authors named above declare that its content is their original work and that they did not use any sources for its preparation other than the class notes, the textbook, and ones explicitly acknowledged in the answers. Any suspected act of plagiarism will be reported to the Facultys Academic Integrity Officer and possibly to the Senate Discipline Committee. The penalty for academic dishonesty may range from failing the course to expulsion from the university, in accordance with Dalhousie Universitys regulations regarding academic integrity.