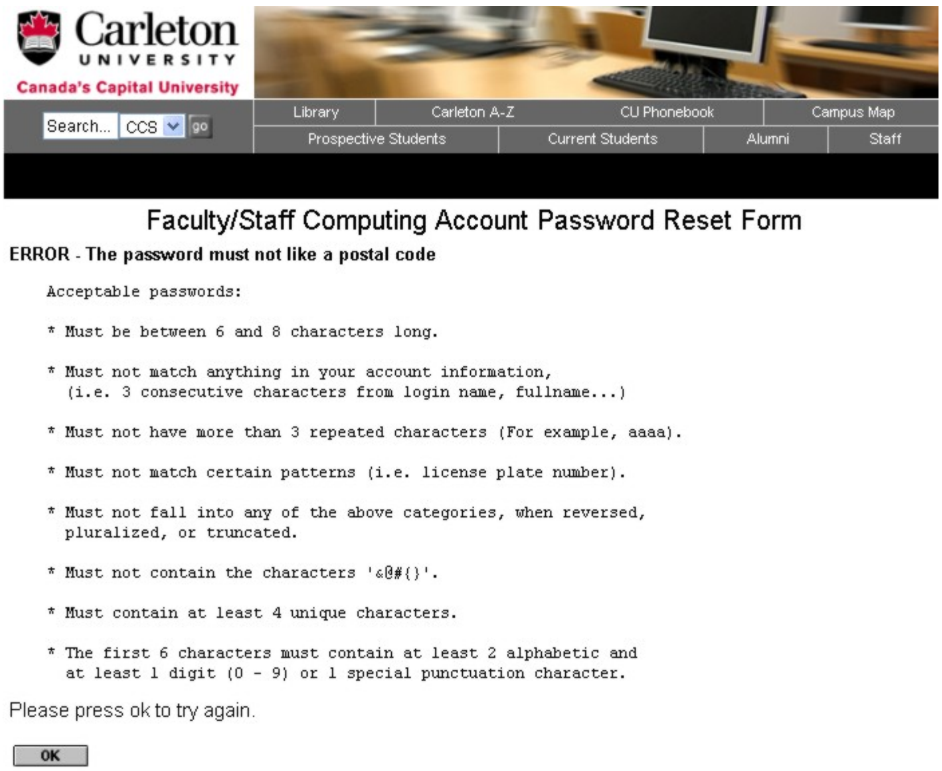
Homework 1

MATH4805/COMP4805/MATH5605 Theory of Automata

Fall 2011 – Due on October 27th

(1) (10 marks) To obtain a CONNECT account at Carleton University, members of the faculty and staff must choose a *connect password* that satisfies a number of basic rules. The bulleted list in Figure 1 states a number of the rules explicitly, and the error message at the top of the figure shows that additional rules need to be satisfied.



Carleton UNIVERSITY
Canada's Capital University

Search... CCS go | Library | Carleton A-Z | CU Phonebook | Campus Map | Prospective Students | Current Students | Alumni | Staff

Faculty/Staff Computing Account Password Reset Form

ERROR - The password must not like a postal code

```
Acceptable passwords:

* Must be between 6 and 8 characters long.

* Must not match anything in your account information,
  (i.e. 3 consecutive characters from login name, fullname...)

* Must not have more than 3 repeated characters (For example, aaaa).

* Must not match certain patterns (i.e. license plate number).

* Must not fall into any of the above categories, when reversed,
  pluralized, or truncated.

* Must not contain the characters '&@#{}'.

* Must contain at least 4 unique characters.

* The first 6 characters must contain at least 2 alphabetic and
  at least 1 digit (0 - 9) or 1 special punctuation character.
```

Please press ok to try again.

OK

Figure 1: Error message from Carleton's *Faculty/Staff Computing Account Password Reset Form* `https://luminis.carleton.ca/secure-cgi/passwd.cgi`

(a) (2 marks) Why are the connect passwords a recognizable language?

Consider an alternate definition of a password based on the connect password. Our alphabet will be $A = L \cup U \cup D \cup P \cup Q$ where $L = \{a, b, \ldots, z\}$ (lowercase letters), $U = \{A, B, \ldots, Z\}$ (uppercase letters), $D = \{0, 1, \ldots, 9\}$ (digits), $P = \{!, \hat{}, ., ?\}$ (valid punctuation) and $Q = \{\&, @, \#, \{, \}\}$ (invalid punctuation). A *password* is any string $w \in A^*$ that satisfies all of the following conditions

- $|w| \geq 6$

- must not have a substring of length three that is also a substring of length three in your login name
- must not have a substring of length three of the form aaa for $a \in A$
- must not have a substring that is a postal code
- must not have a substring that is a license plate
- the reverse of w and the Kleene closure of w satisfy the above substring conditions
- must not contain any symbol in Q
- must contain at least four unique symbols
- the first six symbols in w must contain at least one symbol from $L \cup U$, one symbol from D, and one symbol from P.

In this description assume that your *login name* is your first name, a *postal code* is any string in $U \times D \times U \times D \times U \times D$, and a *license plate* is any string in $U \times U \times U \times D \times D \times D$.

(b) (8 marks) Justify that $Ł = \{w \in A^* \mid w \text{ is a password}\}$ is recognizable. In your answer you must use the closure properties of recognizable languages. Apply these closure properties using at least one finite automata, at least one non-deterministic or ϵ finite automata, and at least one regular expression. Do not justify the correctness of the individual automata and regular expressions.

(2) (10 marks) You do not need to justify your answers to the following two parts.

(a) (5 marks) Give a finite automaton that accepts $Ł_s$ and a regular expression that generates $Ł_s$ for $Ł_s = \{w \in \{0,1\}^* \mid w \text{ has suffix } 010\}$.

(b) (5 marks) Give a finite automaton that accepts $Ł_o$ and a regular expression that generates $Ł_o$ for $Ł_o = \{w \in \{0,1\}^* \mid w \text{ has an even \# of 0s and an odd \# of 1s}\}$.

(3) (10 marks) Prove your answers to the following two parts.

(a) (5 marks) Is $Ł_p = \{w \in \{0,1\}^* \mid w \text{ is a palindrome}\}$ recognizable?

(b) (5 marks) Is $Ł_e = \{w \in \{0,1\}^* \mid w \text{ has an equal \# of 01 and 10 substrings}\}$ recognizable?

(4) (10 marks) Construct a NFA recognizing

$$\{a^n | n \leq 4 \text{ or } n \equiv 3, 4 \bmod 5\}.$$

<u>Using the *subset construction*</u>, give a DFA that recognizes L.

(5) (10 marks) let $n \geq 1$. Show that the language $(0+1)^*1(0+1)^{n-1}$ can be recognized by an NFA with at most $n+1$ states. Give and justify a lower bound on the number of states a DFA must have to recognize this language. <u>Find the best lower bound you can. Not all lower bounds will get full marks.</u>

(6) (10 marks) Construct an ϵ-automata to recognize $(a^2b^* + b^2a^*)(ab + ba)$.

(7) (10 marks) Prove that the following equalities hold for regular expressions:

(i) $r^* = (rr) * + r(rr)^*$

(ii) $(r+s)^* = (r^*s^*)^*$

(iii) $(rs)^*r = r(sr)^*$

(8) (10 marks) Recall that a language $Ł \subseteq A^*$ satisfies the *Pumping Property* if the following holds: $\exists$ number n such that $\forall w \in Ł$ with $|w| \geq n$ $\exists x, y, z \in A^*$ such that

- $w = xyz$,
- $|xy| \leq n$,
- $|y| > 0$, and
- $\forall i \geq 0$ $xy^iz \in Ł$.

(a) (3 marks) Prove that the following language $Ł_a$ satisfies the Pumping Property

$$Ł_a = \{w \in \{0,1,2\}^* \mid w = 0^a1^b2^b \text{ or } w = 1^a2^b \text{ for some } a, b \geq 0\}.$$

A language $Ł \subseteq A^*$ satisfies the *Generalized Pumping Property* if the following holds: $\exists$ number n such that $\forall pws$ Ł with $|w| \geq n$ $\exists x, y, z \in A^*$ such that

- $w = xyz$,
- $|xy| \leq n$,
- $|y| > 0$, and
- $\forall i \geq 0$ $pxy^izs \in Ł$.

(b) (1 mark) Informally, what is the difference between the Pumping Property and the Generalized Pumping Property?

(c) (3 marks) Prove that $Ł_a$ does not satisfy the Generalized Pumping Property.

(d) (3 marks) Prove that the Generalized Pumping Property holds for any recognizable language.

(9) (10 marks) [MATH 5605 only] Prove or disprove that recognizable languages are closed under shuffles. Given two languages $Ł_1$ and $Ł_2$ over alphabet A, the *shuffle* of $Ł_1$ and $Ł_2$ is $\{a_1b_1a_2b_2 \cdots a_kb_k \mid a_1a_2 \cdots a_k \in Ł_1 \text{ and } b_1b_2 \cdots b_k \in Ł_2\}$.