# CS476: Automata Theory and Formal Languages

## Homework 1

Due: 01/11/2011 17.00

---

### Questions

1. **True-False:** (20pts) State whether the following statements are true or not. You must give a BRIEF explanation or show a counter example to receive full credit.

   (a) (5pts) All finite languages are necessarily regular.

   (b) (5pts) $(RS + R)^* RS = (RR^* S)^* + RS$.

   (c) (5pts) $(01 + 10)^*$ is the regular expression for the set of all strings with equal number of 0s and 1s.

   (d) (5pts) With pumping lemma, we can prove that $L = \{w : \text{w does not contain 01 as a substring}\}$ is regular.

2. **Finite Automata:** (20pts) Give DFA's for the following languages.

   (a) (10pts) $L = \{w : n_0(w) \mod 3 > n_1(w) \mod 3\}$

   (b) (10pts) $L = \{w \in \{0,1\}^* : \text{the first two symbols and the last two symbols of } w \text{ are identical }\}$.

3. **Regular Expressions:** (20pts) Give a regular expression for each of the following languages.

   (a) (10pts) $L = \{w \in \{a,b\}^* : w \text{ does not contain the substring } aba\}$.

   (b) (10pts) $L = \{w \in \{0,1\}^* : w \geq 72\}$.

4. **Regular-or-Not:** (20pts) Prove or disprove that the following languages are regular.

   (a) (10pts) $L = \{a^n b^l a^k : n + l + k > 5\}$.

   (b) (10pts) $L = \{a^n b^l a^k : n > 5, l > 3, k \leq l\}$.

5. **RL Closure:** (20pts) Prove that the family of regular languages are closed under the following operations where $L$ denotes a regular language.

   (a) (10pts) *even* removes the odd-ordered symbols from all the strings in $L$ such that $even(a_1 a_2 a_3 a_4 a_5 a_6 \cdots) = a_2 a_4 a_6 \cdots$

   (b) (10pts) *minus*4 removes the 4th symbol (from the left) from all the strings in $L$ (strings of length 0, 1, 2 and 3 are not changed).

6. **Perl:** (20pts) *Perl* is a language with a lot of scripting capabilities. It provides powerful text processing facilities. In this exercise, you will use the regular expression capabilities of *Perl*.

   (a) (10pts) In this part, you will write a script such that given a file the script displays some information about the strings in the file such that

      i. The number of strings that does not contain 101.
      ii. The number of strings that contains at least one pair of consecutive 1s.
      iii. The number of strings that starts with 11 and ends with 01.
      iv. The number of strings that does not contain more than one occurence of the string 00. (The string 000 should be viewed as containing two occurences of 00.)

   The alphabet is $\Sigma = \{0, 1\}$ hence the strings are binary strings. The strings can be separated by any kind of whitespaces, i.e., tab, space, newline etc.

   (b) (10pts) Assume you are a TA of a course and each student submits his/her homework by email as a PDF file with the following name format:

$$\{ID\}\,\{Delimiter\}\,\{Name\}\,\{Delimiter\}\,\{Surname\}\,.pdf$$

   Write a *Perl* script to recognize the Id, Name, and Surname of each student and output them in a tab-delimited format

$$\{ID\}\,\backslash tab\,\{Name\}\,\backslash tab\,\{Surname\}$$

   such that

      i. The delimiter between Id, Name, and Surname can be a space, a hypen, an underscore, or a dot.
      ii. The student Id starts with 19, 20, or 21 and consists of 8 digits.
      iii. Name may contain an optional middle name. Regardless of the delimiter between the first and middle names of the student, there should be a single space between them in the output format.
      iv. Allow Turkish characters in names. You should be able to recognize both lowercase and uppercase letters.
      v. If you encounter a homework submission with a filename in a wrong format, output

$$\{Filename\}\,\backslash tab\,WRONG\ FORMAT\ !$$

      where Filename is the whole name of the file including the extension.

   **SAMPLE INPUT:**

```
19901234-Kuddusi Müftüoğlu.pdf
21001234_Mustafa_Kamil_ABİTOĞLU.pdf
homework.pdf
```

   **SAMPLE OUTPUT:**

```
19901234      Kuddusi         Müftüoğlu
21001234      Mustafa Kamil   ABİTOĞLU
homework.pdf  WRONG FORMAT !
```

Answers for questions 1-5 should be returned in hard copy. The answer (perl scripts) for question 6 should be returned in e-mail to your teaching assistant (emre.yilmaz@cs.bilkent.edu.tr), with the subject line cs476hw1, as an attachment zip file named *NameSurname.zip* including q6part1.pl and q6part2.pl. Good luck to all.