**CS1005 Logic & Computation  
Lab sheet 5: Language and regular expressions**

This lab will

* Reinforce and deepen your understanding of **sets**, and **regular expressions.**

1. Given that the set V contains the vowels in the English language
   1. Write out set V as lower case letters
   2. Is i ∈V ?
   3. Is j ∈V ?
   4. Is ‘ae’ ∈V ?
2. Given that the set P contains the prime numbers from 1 to 10 (***is 1 a prime number?****)*, and the set Q contains the integers in the year that Alan Turing invented the concept of the Turing Machine
   1. Write out set P
   2. Write out set Q
   3. Write out the intersection P ∩ Q. How many numbers are in the intersection?
   4. Write out the union P ∪Q. How many numbers are in the union?
3. The set P contains the integers from 1 to 10, and the set Q contains the integers in the year 1953
   1. Give the set P
   2. Give the set Q
   3. Give the intersection of P and Q
   4. Give the union of P and Q
   5. Give as a set those elements in P which are not in Q
   6. Give as a set those elements of Q which are not in P
4. Given the sets of vowels V and the set N comprising the letters in the name ‘peter’
   1. Give the set V
   2. Give the set N
   3. Give the union of V and N
   4. Give the intersection of V and N
   5. Give the cardinality of the union of V and N
   6. Give the cardinality of the intersection of V and N
5. Construct a regular expression defining each of the following languages over the alphabet Σ = {a, b}
   1. All strings that contain only one a
   2. All strings that contain exactly three b’s in total, and any number of a’s anywhere
   3. All strings that end in a double letter
   4. All strings which have a double letter somewhere in them
   5. All strings which contain one a followed by any number of b’s, or one b followed by any number of a’s
6. Describe in English the languages associated with the following regular expressions
   1. (a|b)\* a(ε | bbbb)
   2. a(a | bb)\*
   3. (a(a | bb)\*)\*
   4. a(aa)\*b(bb)\*
   5. (a(aa)\*b(bb)\*)\*
7. Show by example that the following pairs of regular expressions define the same language over the language Σ = {a, b}
   1. (ab)\*a and a(ba)\*
   2. (a\* | b)\* and (a | b)\*
   3. (a\* | b\*)\* and (a | b)\*
8. Given that the genetic code comprises 4 letters: Σ = {A,C,G,T}, and biologists refer to sequences rather than strings
   1. Give a regular expression which encodes the sequence TATA
   2. Give a regular expression for GC followed by any DNA letter followed by GC (i.e. there will be 5 letters in the sequence)
   3. Given the following 2 DNA sequences, derive one regular expression to describe both of them:
      1. TATAATGCGCCC
      2. TATAATCCTAGC
9. 1. Download grepwin from

<https://github.com/stefankueng/grepWin/releases/tag/1.9.0>

The exe file should be able to run directly from a local directory. Some documentation can be found at <https://github.com/stefankueng/grepWin>

Use this program to search through some text files for specific patterns.

2. You can use the log file shutdown.log in the materials subfolder of this tutorial folder.

E.g. search for all pids

(i) over 99 [i.e. with 3 or more digits]. Hint: search for a pattern starting with ‘pid: ‘ and followed by the definition of your numbers.

(ii) between 100 and 999 [i.e. having only 3 digits]. Hint – how will you ensure that the pid has only 3 digits and not 4 digits?

(ii) between 10 and 99 [i.e. having only 2 digits]

(ii) between 0 and 9 [i.e. having only 1 digit]

1. Explore regular expressions on the site
   1. <https://regexr.com>

Note that you can put the two sequences from 6.3 in the text area, and search with the expression that you have made for the answer to 6.3