Assignment 1

Dr. Athanasios N. Karasimos
ΓΕ77 Υπολογιστική Γλωσσολογία
ΒA in Linguistics | School of English Language and Literature | NKUA

Student’s Name:

**Regular Expressions & Minimum Edit Distance**

Assignment 1

Regular Expressions and Minimum Edit Distance

Deadline: 11/05/2018 Filename: ΓΕ77\_Assignment1\_Surname\_Name

# Task 1: Regular Expressions

## Write regular expressions for the following examples:

(a.) the set of all alphabetic strings.

(b.) the set of all lower case alphabetic strings ending in a *b*.

(c.) the set of all strings from the alphabet a, b such that each *a* is immediately preceded by and immediately followed by a *b*;

(d.) all strings that have both the word *hobbit* and the word *orc* in them (but not, e.g., words like *hobbits* that merely *contain* the word *hobbit* – in any order!);

(e.) the set of all the names and surnames of your classmates (do not forget the whitespace). You do not need to know all the names.

## 1.2. Give two or three examples from the given regexp OR describe what kind of matches you will get (i.e. the regexp /d[a-k]{2,4}/ will give daa, dbak, dfelk etc. OR ‘words’ that start with *d* followed by two, three or four characters ranged from *a* to *k*)

(a.)/[Tt]han(ks|x)/

(b.)/\d\s\w\d\s\w\sAlphanumeric\sSequenses?/

(c.) /(b+eah)\*/

(d.) /(Jan|Feb|Mar|Apr|Mai|Jun|Jul|Aug|Sep|Oct|Noe|Dec)\s\d{4}/

(e.) /^[Tt]he\s[a-z]+\./

# Task 2: Minimum Edit Distance

## 2.1. Figure out whether *drive* is closer to *brief* or to *divers* and what the edit distance is to each. You may use any version of distance cost that you like.

## 2.2. Use the [Levenshtein Demo](http://www.let.rug.nl/kleiweg/lev/) to test 2 pairs of examples. Do not use any of the embedded examples (i.e. industry/ interest). Include here the minimum edit tables and explain step-by-step the path (insertion, deletion or substitution).