

# Introduction to Computational Linguistics (Exercise 2)

Solve the given problems and send the answers to [pavlina@pu.acad.bg](mailto:pavlina@pu.acad.bg), please mention your name and student id.

## **Part 1**

1. The function D-RECOGNIZE only solves a subpart of the important problem of finding a string in some text. Extend the algorithm to solve the following two deficiencies: (1) D-RECOGNIZE currently assumes that it is already pointing at the string to be checked, and (2) D-RECOGNIZE fails if the string it is pointing includes as a proper substring a legal string for the FSA. That is, D-RECOGNIZE fails if there is an extra character at the end of the string.
2. Write a FST for generation/analysis of the following strings: speak (VB), speaks (VBZ), spoke (VBD), spoken (VBN), speaking (VBG), peak (NN), peaks (NNS), and show the analysis of “speaking” and the generation of “peaks”.

## **Part 2**

Visit the website <http://www.xrce.xerox.com/competencies/content-analysis/demos/english> and write down your experience with the tokenizer and the morphological analyzer there.