Gebze Technical University
Department of Computer Engineering
CSE 654 / 484
Fall 2019

Homework 01
Due date: Oct 16th 2019

In this homework we will write a simple FST that handles one Turkish morphological rule:

The Word Root mutates:
Words ending in -P -Ç -T -K change to -B -C -D -Ğ when suffixed with a vowel:
dolap → dolabı
kağıt → kağıdı

Before, we start doing this homework, we need to familiarize ourselves with the FSA tools. Download and install the Turkish morphological parser at https://github.com/coltekin/TRmorph
You may need to install additional tools such as foma to make the tool work on Ubuntu.

Experiment with the tools to parse some Turkish words and generate words from the parsed sentences. For example, produce similar parses such as below
echo "yüklendim" | flookup trmorph.fst
yüklendim yükle<v><pass><t_past><1s>
yüklendim yük<n><D_lAn><v><t_past><1s>

Do the same experiment with word generation.

Read the documentation for program options and other uses. Also read
http://www.let.rug.nl/~coltekin/trmorph/ for the TRmorph Analysis Symbols and become familiar with them.

In this homework, as we mentioned before, we will write a simple FST that handles the word root mutation rule.

Example

Surface: kağıdı, Lexical: kâğıt<N><acc>
Surface: kalemi, Lexical: kalem<N><acc>

Find at least 20 words that goes thru this rule. Find 20 other words that this rule does not apply. Show that your transducer produces correct morphological analysis for the 40 words.

Submit your homework to the moodle page, you will demo your homework after the submission.