

Gebze Technical University
Department of Computer Engineering
CSE 654 / 484
Fall 2021

Homework 03
Due date: Jan 24th 2022
No Late Submissions

In this homework, we will use the Turkish language model that we developed for HW2 to correct Turkish sentences written using English letters only. For example, if the input sentence is “beklenen olumlu tepkiyi alamadi”, the output sentence will be “beklenen olumlu tepkiyi alamadı”. Note that “olumlu” could have been “ölümlü” but we did not choose it.

Your task is to use your language model to assign probability values for each possible sentence and make your decision based on the result. Here are the steps that you may use to complete this task

- Input the sentence written in English letters only, convert it to small case letters
- Mark all the letters that may change on a Turkish alphabet version. These letters are u, i, g, c, o, s.
- Ask your language model the probability of each possible sentence in Turkish character set. If there are N such letter in our sentence, then there are 2^N possible sentences.
- Choose the sentence with the highest probability.

Write a report of your findings on the performance of your language model. You should test all models that use different N values for N-Grams. In order to test your system, take some text from a source that uses Turkish alphabet. Convert this text to English alphabet version using the FST below, which will give you both versions of the same sentence.

karakterler	ı	i	o	ö	u	ü	s	ş	c	ç	g	ğ
ı	1	1	0	0	0	0	0	0	0	0	0	0
o	0	0	1	1	0	0	0	0	0	0	0	0
u	0	0	0	0	1	1	0	0	0	0	0	0
s	0	0	0	0	0	0	1	1	0	0	0	0
c	0	0	0	0	0	0	0	0	1	1	0	0
g	0	0	0	0	0	0	0	0	0	0	1	1

What to submit to the Teams page

- Your source code
- Your report that includes your method details and your performance tables
- test run results for at least 20 sentences

