1st mini-project report
Natural Language

Authors: Jascha Jestel, Piotr Migdałek

Project objective:

The primary objective of this project was to develop a date reading system using finite-state transducers (FSTs). The aim was to design a system capable of converting dates in various formats, potentially mixed, to a standarized formats that mimic human transcription. To achieve this, we employed OpenFST library which we used for constructing, combining and optimizing finite-state transducers.

Transducer creation:

To obtain required transducers outlined in the instructional paper, we performed several operations on them, all implemented within *run.sh* script. *mix2numerical* was created by concatenation of *mmm2mm* and auxiliary *h_identity* transducer. *en2pt* was received by inversing the *pt2en* FST. Moreover, *datenum2text* was obtained by concatenation of month, day, year and auxiliary h_comma and h_slash FSTs; transitional results were stored in the temporary files (*temp0*, *temp1* and *temp2*) and deleted afterwards. Additionally, *mix2text* FST came from composing mix2numerical and *datenum2text*. Lastly, union of FSTs *mix2text* and *datenum2text* gave the final date2text transducer.

Script Modifications:

Additionally, we made important changes to *word2fst.py* parsing script. In order for the names of the month to be represented as tokens, not a collection of letters, we modified the script using Python library *re*. The change was crucial to properly assess the conceptual soundness of our system by conducting adequate tests using the 3. variant of tests where the test transducer is created using the python script. When the run.sh script is executed, all the mandatory tests are executed using the given 1. and 3. method. We prefer the 3. version because it's the quickest and more intuitive because it just prints the output tape using the output symbols txt. Therefore, to execute the run.sh, syms.txt and syms-out.txt need to present like in the tiny example.

Project Work Distribution:

The workload was evenly distributed between the two group members, with each member contributing 50% to the project's completion. The architecture of the transducers was collaboratively designed and implemented by Jascha who also modified the Python script. Piotr was responsible for conducting tests and preparing this report.

Conclusion:

In conclusion, our project successfully achieved its goal of creating a date conversion system using finite-state transducers. We effectively used the OpenFST library to construct and optimize the required transducers and made significant effort to understand underlying mechanics of conducting FST-based project.