Instituto Superior Técnico

Natural Language

Mini Project 1

2023/2024

Group and relative contribution (%)

Tiago Miranda, no. 93416 .... 50%
Nuno Machado, no. 94021 .... 50%

We chose to divide the project in the following way: Nuno started solving questions a and c and Tiago questions b and d. All transducers were tested independently by both. The work wasn't performed in parallel however, both students helped each other in every question of this project.

a

Since the years that will be used as inputs will always belong to the interval [2001-2099] we opted to only allow the numbers 2 and 0 for the first two digits of the given year. Also considering that the day and year must remain unchanged, if the day is inputted with a single digit, it'll be outputted with a single digit.

b

We opted to create an auxiliary transducer called **transpt2en.fst** that translates the month abbreviation from Portuguese to English. Then, just as in exercise a) we concatenated it with the transducer **dd_aaaa.fst**. In order to convert a date in a mixed condensed format from English to Portuguese we inverted the **pt2en.fst** with the command *fstinvert*.

c

Firstly we defined the transducers **day.fst**, **month.fst** and **year.fst**. Then to make the transducer **datenum2text.fst** we created two extra transducers: **comma.fst** (to add a comma) and **slash.fst** (to consider a slash and remove it). All of these were concatenated in pairs following the order: **month.fst**, **slash.fst**, **day.fst**, **slash.fst**, **comma.fst** and **year.fst**. Since the concatenations were done in pairs several auxiliary transducers were needed. After the final one was obtained all the auxiliary ones were deleted.

d

To obtain **mix2text.fst** we performed an union (*fstunion*) of 2 transducers: the composition (*fstcompose*) of **pt2en.fst**, **mix2numerical.fst** and **datenum2text.fst**; and the composition of **mix2numerical.fst** and **datenum2text.fst**. To obtain **date2text.fst** we performed an union (*fstunion*) of **datenum2text.fst** and **mix2text.fst**. Again, several auxiliary transducers were generated that were deleted after obtaining the desired results.