LN Report - MP2

Group 23

Daniela Mendes 87646

Viviana Bernardo 87709

1 Options taken

In this project we implemented various non-deterministic Finite-State Transducers (FST) that convert a Portuguese word's lemma and a given morphological classification to the corresponding word, depending on the classification we attributed.

We also implemented an FST that would do the exact opposite, which means that, given a Portuguese word, the FST returns its lemma and all the possible morphological classifications.

Then, while processing the input, we decided to overwrite the symbols we wanted to delete, or replace, with epsilon (eps).

Afterwards, we used the fstunion operation, available within the OpenFST tool, to join the lemma2verbip.fst, lemma2verbis.fst and lemma2verbif.fst into one FST only (lemma2verb.fst). Then, we chose the fstinvert operation to invert lemma2word.fst into word2lemma.fst

Lastly, we used the fstcompose operation to test our FSTs with the inputs "porteiro+N+fp" (test1.txt), "lentamente+ADV" (test2.txt), "matas" (test3.txt) and "plantas" (test4.txt).

2 Comments on developed solution

We opted to use non-deterministic finite state machines, so that the transducer would be able to process the word and decide when to move forward to analyse the morphological syntax.

When we tested test1.fst with the three FSTs, both lemma2verb.fst and word2lemma.fst failed (i.e. the corresponding PDFs were empty), which was expected since test1 contained a lemma to be translated to a name, and these two transducers that failed translate a lemma to a verb or a word to a lemma, respectively. lemma2word.fst performed successfully, returning "porteiras".

A similar thing happened with test2.txt and test3.txt:

test2.txt failed with lemma2verb.fst and word2lemma.fst, since test2.txt contained a lemma to be translated to an adverb and these two transducers translated a lemma to a verb and a word to a lemma, respectively. lemma2word.fst performed successfully again, returning "lentamente".

test3.txt failed with lemma2verb.fst and lemma2word.fst, since test3.txt contained a word to be translated to a lemma, and these two transducers translated a lemma to a verb or to a word, respectively. word2lemma.fst performed successfully, returning "matar+V+ip+2s" and "mato+N+fp", which was expected since this word works both as a verb and as a noun.

Our group tested a forth word, contained in test4.txt, and this test showed us that the transducer word2lemma.fst does not always return a correct output: When tested with the word "plantas", the mentioned transducer returned 2 lemmas, one of them being "planto+N+fp", which is wrong since "planto" is not a word. However, since "plantas" ended with "as", the transducer interpreted this word as the feminine and plural form of the noun "planto" instead of "planta". This happened because, while developing the transducer, we only considered the masculine and singular form of a noun as a lemma.