

CSE 484 HOMEWORK 1

MYBASICFST

FINITE STATE TRANSDUCER FOR

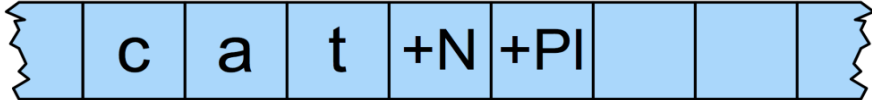
PLURAL AND POSSESSIVE SUFFIXES

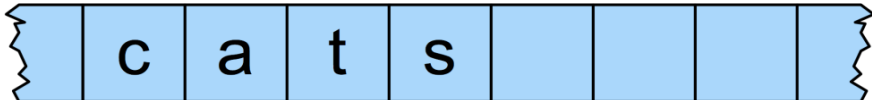
ÇAĞRI ÇAYCI

1901042629

OVERVIEW

The goal of this assignment is to develop a finite state transducer for the Turkish language, specifically focusing on plural and possessive suffixes. To achieve this objective, we examined the following examples.

Lexical 

Surface 

Our main goal is to accurately convert the input we receive on the surface level to the lexical level. For instance, when the Finite State Transducer (FST) is given "arabalar" as input, it should generate the output "araba+N+PL." This ensures a straightforward and precise transfer, aligning the input representation with the appropriate lexical structure. This pattern extends to various instances:

"araba" -> "araba+N"

"arabalar" -> "araba+N+PL"

"arabam" -> "araba+N+PSS"

"arabalarım" -> "araba+N+PL+PSS"

"araban" -> "araba+N+PSS"

"arabaların" -> "araba+N+PL+PSS"

"arabamız" -> "araba+N+PSS"

"arabalarımız" -> "araba+N+PL+PSS"

"arabanız" -> "araba+N+PSS"

"arabalarınız" -> "araba+N+PL+PSS"

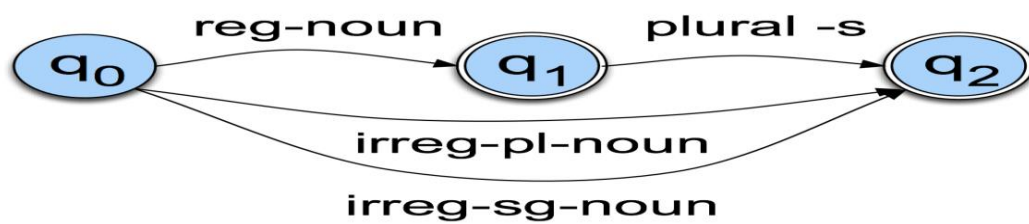
"arabası" -> "araba+N+PSS"

"arabaları" -> "araba+N+PL+PSS"

In this way, FST can transform the provided words from a surface level to a lexical level. It also provides information about whether the string is accepted or not.

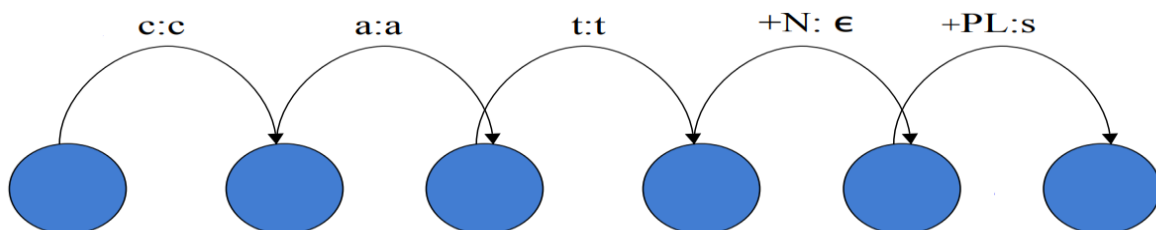
DESIGN

The book examples are examined when designing this finite state transducer. Firstly, the following approach (Figure 1) was going to be applied by dividing the words into two as those that take the suffix 'ler' for pluralization and those that take 'lar'. This method effectively facilitates the transition from surface level to lexical level pluralization. However, when incorporating possessive suffixes, we encounter a need for the last letter of the word. The selection of possessive suffixes is influenced by this last letter. For instance, "araba" takes "-m" as its possessive suffix, while "taşıt" takes "-ım" instead, showcasing the variation in possessive suffixes based on the last letter of the word. For this reason this approach could not be applied.

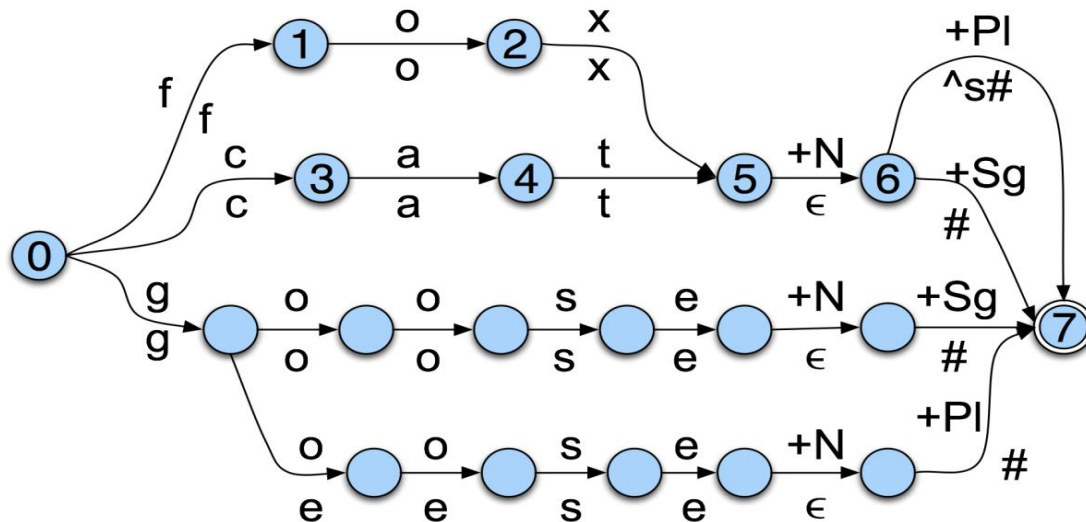


(Figure 1)

In the second approach, illustrated in Figure 2, the finite state transducer will be systematically applied letter by letter, progressing towards the final letter. Consequently, each word in my dataset is symbolically represented by its corresponding FST (Figure 3).



(Figure 2)



(Figure 3)

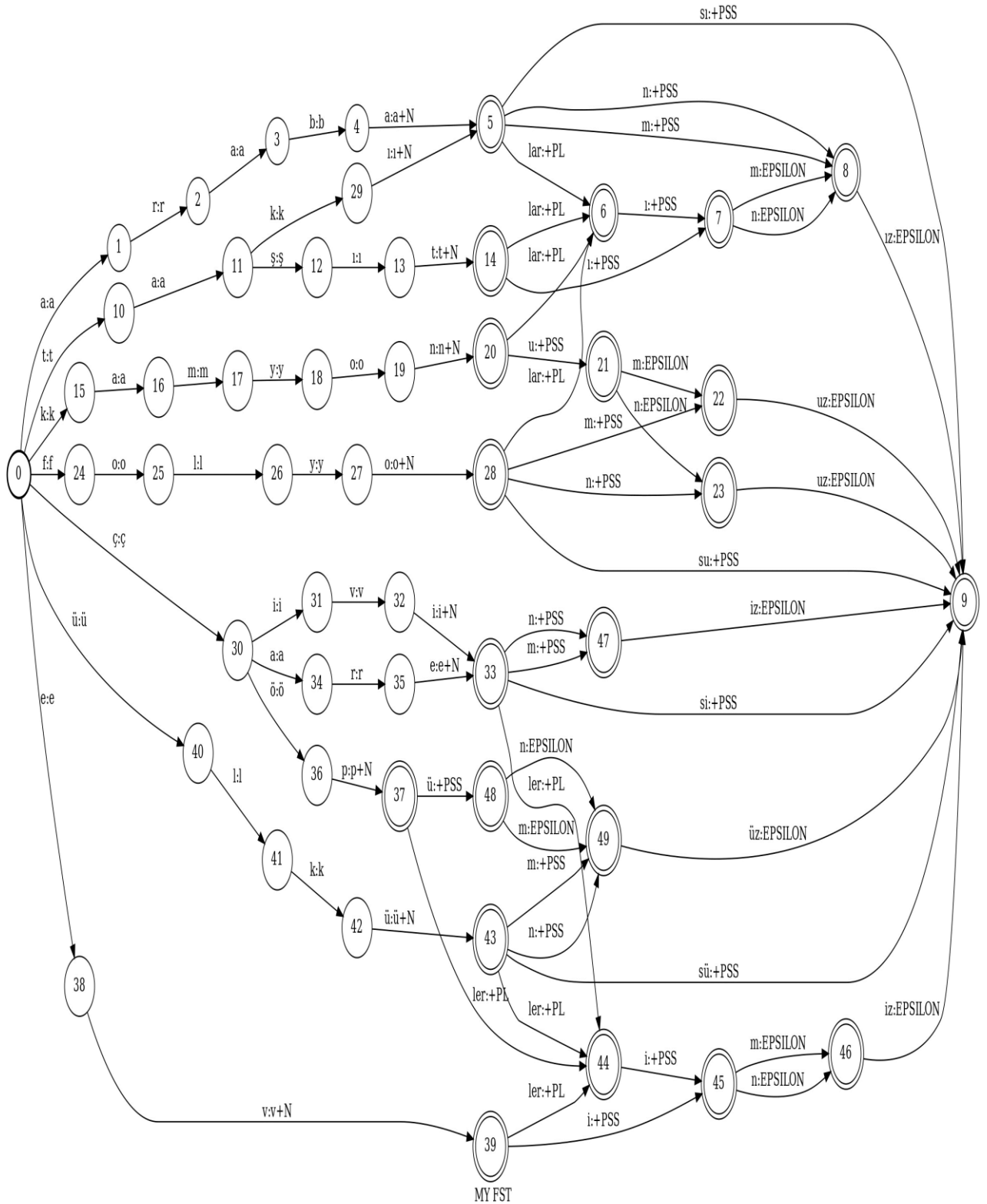
Finally, we can access the last letter of the word. So, we can write rules for that.

Possessive Suffix Rules

1. The last syllable contains 'a' or 'i' and consonant ending:
Add "-i" or "-in" or "-im" for possessive suffix.
2. The last syllable contains 'o', or 'u' and consonant ending:
Add "-u" or "-un" or "-um" for possessive suffix.
3. "a" or "i" ending: Add "-m" or "-n" "-si".
4. "o" or "u" ending: Add "-m" or "-n" "-su".
5. "e", "i" ending: Add "-m" or "-n" "-si"
6. "ö", "ü" ending: Add "-m" or "-n" "-sü"

Plural Suffix Rules

1. The last syllable contains a back vowel: Add "-lar".
2. The last syllable contains a front vowel: Add "-ler".



(Figure 4)

In Figure 4, you can see the final version of our finite state transducers. With this, we can handle all the words in our datasets, including different versions like plurals and possessive forms.

REQUIREMENTS

This assignment was successfully executed using the pywrapfst library in a Python environment on Linux. Furthermore, the graphviz library was employed for visualizing the Finite State Transducer (FST). The following commands are recommended for usage:

1. To test with built-in test cases:

```
$ python3 pluralFST.py < test.txt
```

2. To test with new test cases:

```
$ python3 pluralFST.py
```

3. To visualize the FST:

```
$ dot -Tpng -Gdpi=500 pluralFST.dot -o pluralFST.png
```

Note: I executed my Python code using python3. The specific command may vary depending on your installation method.

Note: As per the assignment instructions, I have selected 10 words to create a Finite State Transducer (FST). The chosen words are: 'araba,' 'taşıt,' 'kamyon,' 'folyo,' 'takı,' 'çivi,' 'çare,' 'ülkü,' 'çöp,' and 'ev.' Please use those words and note that all letters in the words must be in lowercase for testing purposes.

