

Natural Language Processing Homework 7

Katie Chang + Willis Wang

April 28, 2016

README

1 Q2

1.1 first

First accepts a language that starts with at least one 0 and ends with at least three 1s. There can be any combination of 0s and 1s in between the leading 0 and the trailing 111. Since the ? denotes optionality, the fourth 1 is not necessary for the string to be accepted in this language.

ii. This language is a subset of the language Bit^* .

The 0 and 1 are quoted in the .grm file because these two FSMs are defining specific strings that this language will use. Essentially, the quotes indicate the initializing of the letters of the alphabet that this language will use.

iii. There are 5 states and 9 arcs in this FST.

1.2 b. second

i. `export Second = Optimize[Zero+ Bit* One One One];`

ii. If First and Second are equivalent, then Disagreements as a set should be empty.

Running `fstinfo` on `Disagreements.fst`, we can see that this FST has 0 states and 0 arcs, and in fact has a 0 count of everything. Therefore, it should be unable to accept any sort of string.

1.3 c

- i. First.fst now has 20 states and 25 arcs. Second.fst now has 13 states and 17 arcs. Disagreements.fst has 68 states and 88 arcs (!!!).
- ii. fstview does not work remotely, so I will do it later.
- iii. The results are the same..;???

1.4 d

After creating Disagreements.fst, the number of states and arcs is the same as before.

2 Q3

See binary.grm.

3 Q4

Used overleaf.com to generate LaTeX document.