

## NLP Programming Assignment3 Strategy Write Up

Jingwen Zhong

For this programming assignment, the main comparison for simple pair I used Wu Palmer similarity along with the Trips ontology. However, there are a lot of complicated cases. I simply tried the example inputs files and outputs files to improve and adjust my codes.

My strategies are as follows:

1. Step one: Changing the format of input file to the desired format. For me this part is the hardest part, and also, I think this is the most important part, since I can decide how to compare and choose my result. However, I do think I can think of a better way to reformat the input file, so that the code will be easier.

Example: the format I changed for input example file 6 looks like this.

```
[[['THE-SET', 'THE ENGINEERS AND THE BOYS', 'PERSON', 'EB1'],  
  ['THE-SET', 'ENGINEERS', 'PERSON-DEFINED-BY-ACTIVITY', 'E11'],  
  ['THE-SET', 'BOYS', 'MALE-PERSON', 'B11'],  
  ['THE', 'COMPETITION', 'COMPETE', 'C1']],  
 [['THE-SET', 'ENGINEERS', 'ENGINEER', 'E1'],  
  ['VERB', 'HAD', 'HAVE', 'H1'],  
  ['A', 'BUDGET', 'BUDGET', 'B1']],  
 [['PRO-SET', 'THEY', 'PERSON', 'T1'],  
  ['VERB', 'BUILT', 'CREATE', 'B2'],  
  ['A', 'BOAT', 'VEHICLE', 'B3']],  
 [['THE-SET', 'BOYS', 'MALE-CHILD', 'B4'],  
  ['VERB', 'MADE', 'CREATE', 'M1'],  
  ['A', 'BOAT', 'VEHICLE', 'B5'],  
  ['SOME', 'PAPER', 'MATERIAL', 'P1']],  
 [['PRO-SET', 'THEY', 'PERSON', 'T2'],  
  ['VERB', 'LOST', 'LOSE-COMPETE', 'L1'],  
  ['THE', 'COMPETITION', 'COMPETE', 'C2']]]
```

2. Step two: Get the candidates

- a) If the <word> contains “SELF” or “SELVES”:

Refer it to the instance that is in the same sentence with it, pair them and assign similarity to 2.

- b) For sentence[i], compare it with sentence[i-1]. If two summaries have exactly <word> and <type> then pair them and assign similarity 2 to them. I hope the “probability” of two same summaries can be more than 1 so that it can be higher than the largest Wu Palmer similarity.
- c) If two summaries are not the same, pair them and assign their Wu Palmer similarity between their <type>(Ontology). Also, in this case I separate 2 cases, one is the entry with the <spec> that has “PRO” or “THE” in it. The other one is the one that has “PRO-SET” and also has “SET” shows up in previous sentences.
- d) Last, I compare sentence[i] to the one that before sentence[i-1]. And if they have same entry, assign them together. And I also assign similarity to 0.9, because of the input example file 6, where the two boat are bot the same boat and I want this similarity to be smaller than the other one.

3. Step three: Output:

I put all the pairs that are 100% matching to my output list, so that later I can

compare them to the other and solve the problem of reflexive constraints and then update the candidates. Then I compare the similarity, the higher one gets stay. Finally, I put all the updated candidates to the output file to get my final answer.

So far, I got all the examples out put file right, but obviously there are so many cases that I didn't think of.