

University of North Texas  
CSCE5290/CSCE4930, Fall 2017  
Natural Language Processing  
Instructor: Dr. Eduardo Blanco

Homework#Opt. Optional, SRL  
Due date: December 14, 2017  
**Name:**

### Instructions

Submit your code, a README file and the output of your code on Blackboard. You may be asked to demonstrate your code to the TA. **This is an optional assignment. If you complete it, your homework grade will be the average of your 4 best homework submissions. No partial credit will be given. Your implementation must match the sample output provided or you will not receive any credit.**

### Question 1 [100pt]

Given a file with parse trees, implement code that does the following for each parse tree:

1. Loops through all the leaves and selects the verbs.
2. For each verb, selects potential arguments following the heuristic we saw in class (Xue and Palmer)
3. For each verb and potential argument, extracts at least the following features:
  - the verb form (the actual predicate),
  - the POS of the verb,
  - the number of leaves in the potential argument, and
  - whether the argument occurs before or after the verb (i.e., the direction).

Sample input and output files are provided in Blackboard.

Follow these steps:

1. Make sure you have Python 2.7 installed.
2. Install the nltk software library, version 3.2.4 (<http://www.nltk.org/>). This command will install the right version if you use Ubuntu:

```
$ sudo pip install -U nltk==3.2.4}.
```

The sample code may work with other versions, and you are welcome to modify it however you see fit (as long as you get the same output).

3. Download the nltk data (<http://www.nltk.org/data.html>). This command should work from the Python interpreter:

```
>>> import nltk
>>> nltk.download()
```

It will take a while to download all the data. What you really want is a sample of the Treebank.

4. Modify the file `srl.py` to complete the assignment. To run the code, run the following command:

```
$ python srl_sol.py > my_output.txt
```

Your implementation works when `my_output.txt` is exactly the same than `output.txt`. You can check the differences between both files running this command:

```
$ diff output.txt my_output.txt
```