Assignment 4

CS 6320.002: Natural Language Processing
Name: Manas Bundele
MMB180005

Part 2:

Writeup Question 2.1)

Lemmatization converts a token to its base word based on its POS tag. The form of the word carries information related to its POS tag. The same word can be a noun or a verb in different forms or in contexts which they appear. For e.g. if we say, 'He is caring' then POS tag will be [('He', 'PRP'), ('is', 'VBZ'), ('caring', 'VBG')] but if we lemmatize it then it will be [('He', 'PRP'), ('be', 'VB'), ('care', 'NN')] for lemmatized sentence 'He be care'. Thus, as word forms carry contextual information, we POS actual sentence instead of lemmatized ones.

Part 3:

Writeup Question 3.1)

1. X (e. g. Y) or X e. g. Y

   Here X is hypernym whereas Y is a hyponym and can be comma separated words.

   Example from wiki corpus: Other common themes include **place names (e. g. Washington , DC** ; Lower East Side ; North Carolina ; Paris ; Venice) , **animals (e. g. bear , goldfish , jellyfish , rabbit , bat , dog , boa constrictor , cockroach)** , as well as **themes common throughout Merritt 's work (e. g. the moon , dancing , rain , and eyes)**

2. X which is called Y

   Here, X - hypernym, Y – hyponym. Y can be in quotes or round brackets.

   Example from wiki corpus: Works of art depicted Ceres conventionally with a scepter , a basket of flowers and fruit , and a garland made of corn ears (note that " corn " in this instance refers to wheat , barley , or some other old world food grain , not to **the new world food grain maize , which is called " corn "** in the United States and some other areas of the Western hemisphere)

Part 5:

Writeup Question 5.1) Precision: 1.000000
                      Recall:0.148148
                      F-measure: 0.258065

       I made a change in the hearst patterns, making all the words lemmatized instead of their original form because we are looking for the pattern in lemmatized NP tagged and chunked text. Thus, 'including' changes to 'include' in pattern 4. This gives a count of that pattern as 10382 after using lemmatized form of 'including' instead 0 for its original form.

Writeup Question 5.2) For the NP tagging, we are considering phrases that have determiners, adjectives and nouns and thus this NP becomes a hypernym or hyponym text in a predicted tuple. Our test set only has nouns as hypernyms or hyponyms. So, since we match the tuple (hyponym, hypernym) directly, we might not get many matches, as our prediction also has DT and JJ in a NP. To fix this, we can split the hypernym or hyponym string and keep the nouns and then try matching the tuples to test data. That way, we will get better results.

Part 6:

Writeup Question 6.1)

It took maybe a day or a bit more than that to complete the assignment (except writeup and extra credit).

Writeup Question 6.2)

Yes, I got stuck in evaluate predictions to calculate false negatives so discussed it with Shobhana. Apart from that, didn't have any other doubts to discuss with anyone but discussed some parts in the assignment with few other people whoever asked me about it.

Part 7:

Writeup Question 7.1)

1. X (e. g. Y) or X e. g. Y

   Here X is hypernym whereas Y is a hyponym and can be comma separated words.

   Example from wiki corpus: Other common themes include **place names (e. g. Washington , DC** ; Lower East Side ; North Carolina ; Paris ; Venice) , **animals (e. g.**

bear , goldfish , jellyfish , rabbit , bat , dog , boa constrictor , cockroach) , as well as **themes common throughout Merritt 's work (e. g. the moon , dancing , rain , and eyes)**

2. X which is called Y
   After lemmatization: X which be call Y

   Here, X - hypernym, Y – hyponym. Y can be in quotes or round brackets.

   Example from wiki corpus: Works of art depicted Ceres conventionally with a scepter , a basket of flowers and fruit , and a garland made of corn ears (note that " corn " in this instance refers to wheat , barley , or some other old world food grain , not to **the new world food grain maize , which is called " corn "** in the United States and some other areas of the Western hemisphere)

3. X is a Y
   After lemmatization: X be a Y

   Here, X-hyponym, Y-hypernym.

   Example from wiki corpus: The **Frog cake is a dessert** in the shape of a frog 's head , composed of sponge cake and cream covered with fondant .

4. X like Y

   Here, X-hypernym, Y-hyponym. Y can be multiple words separated by comma or 'and' or 'or'. In this case, there can be multiple exceptions, so we might need to take care of that as well.

   Example from wiki corpus: They fast from solid food and take only **fluids like coconut water or milk**.

Writeup Question 7.2)

Precision: 0.983871
Recall:0.161376
F-measure: 0.277273

After adding the extra patterns, the results have improved a bit. Also, I have used lemmatized versions of patterns as we match them with lemmatized text and without using lemmatized version of patterns, there was no match.