

NLP HW4 - DXP170020

Writeup Question 2.1:

We can find parts of speech only on the original sentence and not on lemmatized version of the sentence. Also, lemmatization is the process of converting a word to its base form. Hence lemmatization considers the context and converts the words to its meaningful base form.

Writeup Question 3.1:

The other hearst pattern that we can come up with in the wikipedia_corpus are as follows :

Hearst Pattern1 : ((NP_\w+ ?(,)?) + (and |or)?) as NP_\w+)

Example sentence : In this genre , the mandolin is the main instrument , simultaneously that the cuatro and the guitar serve as companions . in this genre , the mandolin be the main instrument , simultaneously that the cuatro and the guitar serve as companion .

Chunked Sentence Form of above Example Sentence : in NP_this_genre , NP_the_mandolin be NP_the_main_instrument , simultaneously that NP_the_cuatro and NP_the_guitar_serve as NP_companion .

Example Sentence : By the mid 80s , KMGK suddenly found itself fighting for market share as KRNQ and KLYF started to hit on KMGK 's previously untouched land of hit music on FM radio .

Chunked Sentence Form of above Example Sentence : by NP_the_mid 80s , NP_KMGK suddenly find itself fight for NP_market_share as NP_KRNQ and NP_KLYF start to hit on NP_KMGK 's previously NP_untouched_land of NP_hit_music on NP_FM_radio .

Hearst Pattern2 : (NP_\w+ (,)?) like (NP_\w+ ? (,)?(and |or)?) +)

Example Sentence : The West Series was first formed in 1954 as a proving ground for drivers from the West who could not travel to race in the more traditional stock car racing regions like North Carolina and the rest of the South .

Chunked Sentence Form of above Example Sentence : NP_the_West_Series be first form in @card@ as NP_a_prove_ground for NP_driver from NP_the_West who could not travel to NP_race in the more NP_traditional_stock_car race NP_region like NP_North_Carolina and NP_the_rest of NP_the_South .

Example Sentence : A kora is built from a large calabash cut in half and covered with cow skin to make a resonator , and has a notched bridge like a lute or guitar .

Chunked Sentence Form of above Example Sentence : NP_a_kora be build from NP_a_large_calabash_cut in NP_half and cover with NP_cow_skin to make NP_a_resonator , and have NP_a_notch_bridge like NP_a_lute or NP_guitar .

Hearst Pattern3 : (NP_\w+ (,)?) particularly (NP_\w+ ? (,)?(and |or)?) +)

Example Sentence : The restaurant is known for many menu items , particularly Oysters Rockefeller , of which more than a million orders have been served since the business opened in the nineteenth century .

Chunked Sentence Form of above Example Sentence : NP_the_restaurant be know for NP_many_menu_item , particularly NP_Oysters_Rockefeller , of which more than a million NP_order have be served since NP_the_business open in NP_the_nineteenth_century .

Example Sentence : Kurilsk is a center for fishing and fish farming , particularly salmon .

Chunked Sentence Form of above Example Sentence : NP_Kurilsk be NP_a_center for NP_fishing and NP_fish_farming , particularly NP_salmon .

Writeup Question 5.1:

Precision: 1.000000

Recall:0.126984

F-measure: 0.225352

Writeup Question 5.2:

In the dataset, the (hyponym,hypernym) pair consists single, usually lemmatized, tokens. On the other hand, the extractions from our code extract a lot of multi-word noun phrases. This could negatively affect the performance on the provided dataset. We can drastically improve if we post process our extractions to output only single token extractions. We can use our knowledge about noun phrases (that the head of the noun phrase is usually the last token) and lemmatization to figure out post processing methodology that might improve performance.

Writeup Question 6.1: 4 days

Writeup Question 6.2: No

Writeup Question 7.1:

The new hearst patterns are as follows :

1. Hearst Pattern and Sentence
 - a. **Pattern :** (NP_\w+ (,)?other than (NP_\w+ ? (,)?(and |or)?)+), before
 - b. **Original Sentence :** In the first four volumes he treats (a) of the prices of corn , and the circumstances affecting prices ; (b) the prices of produce other than corn ; and (c) the state of the circulation .
2. Hearst Pattern and Sentence
 - a. **Pattern:** (NP_\w+ (,)?particularly (NP_\w+ ? (,)?(and |or)?)+), before
 - b. **Original Sentence:** The restaurant is known for many menu items , particularly Oysters Rockefeller , of which more than a million orders have been served since the business opened in the nineteenth century .
3. Hearst Pattern and Sentence
 - a. **Pattern:** (NP_\w+ (,)?like (NP_\w+ ? (,)?(and |or)?)+), before
 - b. **Original Sentence:** A kora is built from a large calabash cut in half and covered with cow skin to make a resonator , and has a notched bridge like a lute or guitar .
4. Hearst Pattern and Sentence

- a. **Pattern:** (NP_\w+ (,)?compare to (NP_\w+ ? (,)?(and |or)?)+), before
- b. **Original Sentence:** Further , the steady beeping of the radio beacon aboard " Sputnik 1 " as it passed overhead every 96 minutes was widely viewed on both sides as effective propaganda to Third World countries demonstrating the technological superiority of the Soviet political system compared to the American one .

Writeup Question 7.2:

Precision: 0.966102

Recall:0.150794

F-measure: 0.260870