UTD Summer Workshop on NLP with Python

Assignment 2

Upload your completed Python program file to the MS Teams "Assignments" → Assignment 2" folder. Be sure to include your name in the comments at the top of the file in the following format:

```
###############################################################################
#
#          FILE:
#            filename.py
#        AUTHOR:
#            Your Name
#   DESCRIPTION:
#            Assignment 2
#            Description of your program, what it does
#  DEPENDENCIES:
#            Created with Python 3.10.11 (Python version)
#            Any dependencies, i.e. extra libraries required to run (like
#            datetime, NLTK, or re)
#
###############################################################################
```

Extra help references:

- Python re reference – https://docs.python.org/3/library/re.html
- NLTK references
  - Tokenize – https://www.nltk.org/api/nltk.tokenize.html
  - WordNet –
    - Sample usage

Submit a single file Python program that does the following:

1. Process a natural language text file (text_web.txt) and send the output in parallel to both (a) the screen, using the Python print() function, and (b) a separate file named "output.txt".

2. Provided a plain text file (text_web.txt), your Python program should first tokenize the text—remove any html tags using regualar expressions, then split on sentences using newlines "\n" as a delimiter.

3. Then ***for each sentence***:

   3.1. Display a visual separator, such as: `print("=" * 80)`

   3.2. Display the text of the sentence followed by the POS-tagged full sentence. For example:
   ```
   John came from the store.
   ('John', 'NNP'), ('came', 'VBD'), ('from', 'IN'), ('the', 'DT'), ('store',
   'NN')]
   ```

   3.3. Identify the first noun in the sentence and display it followed by its synsets, e.g.
   ```
   actor [Synset('actor.n.01'), Synset('actor.n.02')]
   ```

3.4. For each synset in §3.3, display its synset followed by the fist 30 characters of its definition. Then locate the nearest common hypernym synset among the remaining nouns in the sentence. Under each synset, display an indented line that displays the most closely related synset for each of the other nouns in the sentence with their 30 character truncated definition