Automatic Concept Map Generation from Text-based Learning Material

07.04.2018
— **Guide : Prof. P. K. Bhowmick**

Akshay Patil

Ashwin Bhoyar

Suyog Chadawar

Overview

Concept map is an effective tool to present summary of a learning material. Concept map is represented by a set of concepts and their relationships. They are different from formal ontology in the sense that concepts maps do not consider standard vocabulary to represent the relations. The concepts and relations are formed using free text.

Packages Used

1. **Pysolr:** To index the Glove Word Vectors
   https://pypi.python.org/pypi/pysolr/3.6.0
   http://lucene.apache.org/solr/

2. **Dbpedia Spotlight:** To find Important Concepts
   https://pypi.python.org/pypi/pyspotlight/0.7.1

3. **Stanford CoreNLP OpenIE Tool:** To find the relation among Concepts
   https://pypi.python.org/pypi/pycorenlp/
   DATA **:** https://stanfordnlp.github.io/CoreNLP/

4. **NLTK:** To tokenize the free text into sentences
   https://pypi.python.org/pypi/nltk

5. **Networkx:** To generate the Graph (Concept map).
   https://pypi.python.org/pypi/networkx/2.1

6. **XML2Dict:** To convert annotated XML file into cleaned text file.
   https://pypi.python.org/pypi/XML2Dict/0.2.1

Tools Required

Graphviz
sudo apt-get install graphviz

Data Requirements

- **GloVe Pre trained word vectors (Wikipedia2014+Gigaword 5)**
  http://nlp.stanford.edu/data/glove.6B.zip

Note: Before following below steps Make sure you have downloaded all the Python Packages, Tools and Data Requirements mentioned above

How to Run

1. Run **ResolveCoreferences.py**: This will generate file having coreferences resolved.
   python ResolveCoreferences.py text_File

2. Start Solr Server:
   In Solr Directory put command
   ./solr start

3. Open **GloveIndexing.py** and Update GloVe 100d txt file path and then Run it. This will index words with their vectors in Solr.

   python GloveIndexing.py

4. Download Stanford CoreNLP 3.9.1 , Extract it and Run CoreNLP Server by using below command :
   java -mx4g -cp "*" edu.stanford.nlp.pipeline.StanfordCoreNLPServer

   This will provide OpenIE service.

5. Run **ConceptGenerator.py:**

    ```
    python  ConceptGenerator.py  path_to_source_file
    ```

    This will Generate the following files:

    - Graphviz Graph - concept_map.png
    - Matplotlib Graph