CS5560 Knowledge Discovery and Management

In-Class-Exercise (ICE-1A)

June 5, 2017

Student Names: SreeLakshmi Nandanamudi, ClassID: 17

We are supposed to build a knowledge graph for the following input (text data).

1. Describe your knowledge about knowledge graph.
2. Why do we want to build such a knowledge graph?
3. What steps are required? Show your own workflow for this task.
4. What are the challenges?
5. Draw a knowledge graph for the given data.

> CHICAGO (AP) — Citing high fuel prices, United Airlines said Friday it has increased fares by $6 per round trip on flights to some cities also served by lower-cost carriers. American Airlines, a unit AMR, immediately matched the move, spokesman Tim Wagner said. United, a unit of UAL, said the increase took effect Thursday night and applies to most routes where it competes against discount carriers, such as Chicago to Dallas and Atlanta and Denver to San Francisco, Los Angeles and New York.

1. The knowledge Graph is a knowledge base used by Google to enhance its search engine's search results with semantic-search information gathered from a wide variety of sources.
The knowledge graph search API allows users to query Google's knowledge Graph database to obtain information on the entities contained therein.
According to Google, some typical use cases include:
→ Getting a ranked list of the most notable entities that match certain criteria.
→ Predictively completing entities in a search box.
→ Annotating/organizing content using the knowledge Graph entities.
Example: 1) If you search for a movie, you can directly book tickets to see it at your local cinema.
2) If you search for a local store, and you know exactly when the busiest times are.

2. A knowledge graph is a data structure that allows to contextualise entities, and organise correlations between entities or multiple types of entities.
The benefits of knowledge graph include:
1. Reputable open semantic database such as wikipedia or Freebase. They are easy to work with, are manageable in size and have pretty good quality, making them ideal as first steps. You will want to go beyond them eventually as relying on them exclusively might cause people to perceive your work as derivative-only.
2. General web content of higher quality such as newspaper and other article archives where content has been cleaned up from low quality stuff such as spam, porn and low quality ads.

3. open crawl database such as Common crawl comprising of unprocessed dumps of crawls of high quality links. They are a great source as selection of quality links saves you a tremendous amount of time and effort wasted in eliminating junk.

4. Crawling yourself is always an alternative where you have full control and with open source crawlers and cheap machines and bandwidth you can set up cheaply and quickly a rig crawling tens of millions of documents per day, which should be your target for serious crawling.

In addition to data acquisition, there are two main ways of knowledge processing approaches :

a) Simple heuristic approaches such as text processing based on regular expressions, simple parsing and NLP techniques.

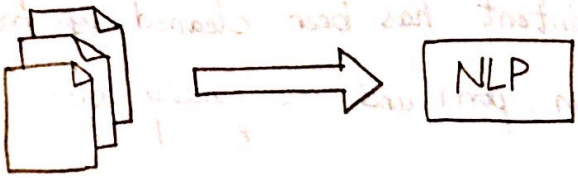
b) Deep Learning techniques where you employ novel methods to experiment with your own kinds of knowledge.

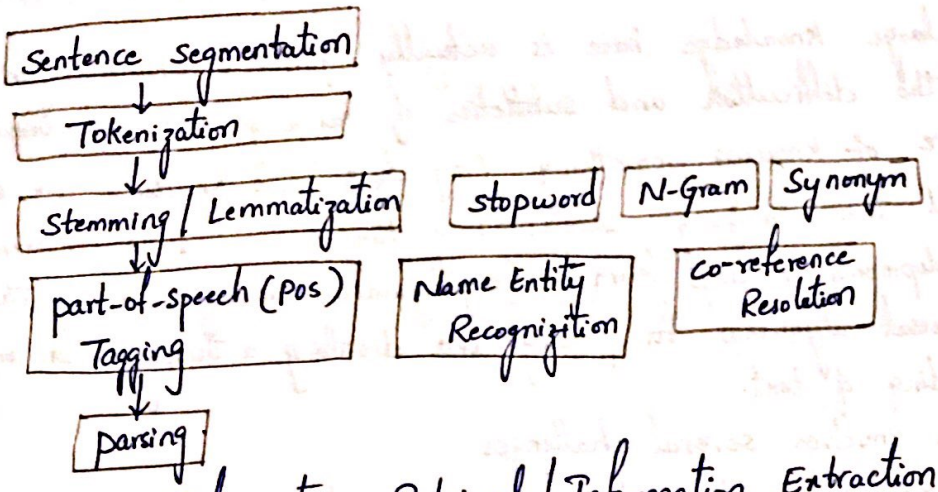
3. To build a knowledge graph, two basic steps are required :

1. What are the nodes ? In a knowledge graph, they will be related to semantic concepts such as persons, entities, events etc.
2. What are the edges ? They will be defined by relationships between nodes based on semantics.

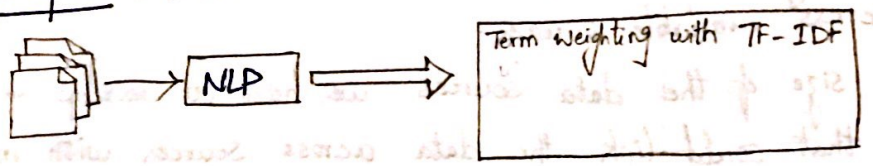
Workflow : We can build knowledge graph from text.

step1 : Natural Language processing (NLP)

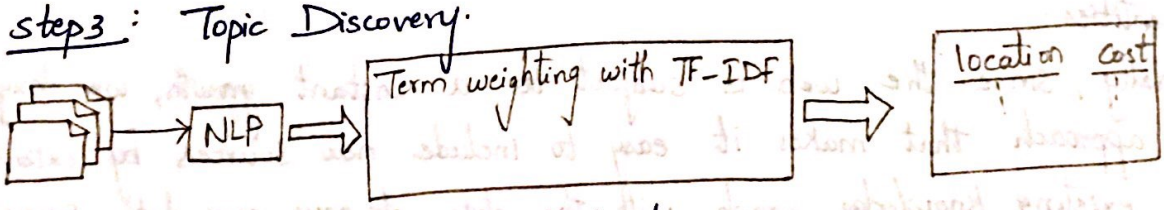




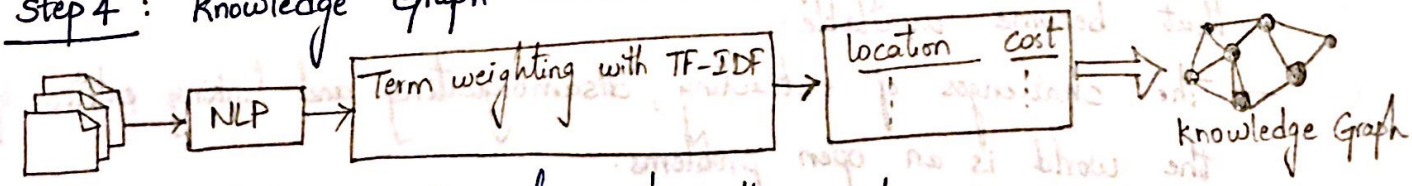
Step 2: Information Retrieval / Information Extraction



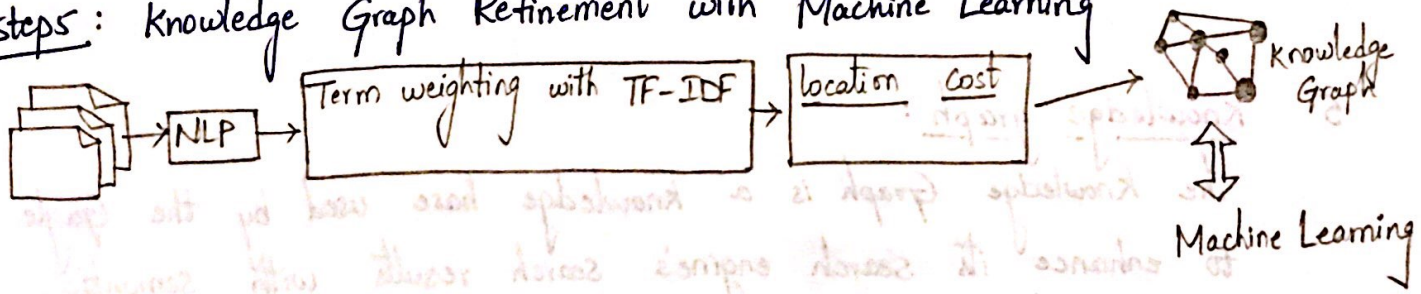
Step 3: Topic Discovery.



Step 4: Knowledge Graph Construction



Step 5: Knowledge Graph Refinement with Machine Learning



4. Challenges of Knowledge Graph:

1. The Knowledge Graph is not only for searching for pages that match your query terms but also for "entities" that the words describe.

→ Creating a large knowledge base is actually quite a challenge in part due to the difficulties and subtleties of language and the ethereal transient nature of knowing something. (e.g. facts and knowledge are contin evolving. For all the language understanding thrown around in conferences and recent developments in deep learning (e.g. dynamic memory networks) there is still no universal algorithm for parsing and distilling a thorough, non ambiguous, understanding of text.

Knowledge graph involves several challenges:

1. First, when consolidating the data from various data sources, we had to align the data across available sources.
2. Second, given the size of the data sources we had to develop a scalable approach that could link the data across sources with millions of entities.
3. Finally, since the web is subject to a constant growth, we designed an approach that makes it easy to include new sources by extending an existing knowledge graph with the data of any new data sources that become available.

→ The challenges of extracting, disambiguating and linking entities of the world is an open problems.

5. Knowledge graph:

The Knowledge Graph is a knowledge base used by the Google to enhance its search engine's search results with semantic-search information gathered from a wide variety of sources. The Google Knowledge Graph Search API is a tool that helps users find entities with in the Google Knowledge Graph.

5.