Project

→ Loops & self reference in Construction of Dictionaries

- Human lexicon is not a simple one-to-one mapping of concepts onto words but rather a complex web of semantically related parts.
- Thus a lexicon can be represented as a graph with words as nodes with edges drawn based on a variety of possible relationships such as word co-occurrence in texts, thesauri or word association experiments on humans
- Dictionary graphs, in which directed links are drawn from a word to the words in its definition, allow one to identify sets of words with equivalent meanings simply by selectively iterating through the descendants of a given node.
- Loops can be found in dictionary networks. These loops are particularly intriguing as they represent a form of self-reference, a condition that has been used in classical logic
- It has been found that the ~~introduction of new concept~~ insertion of a loop in a dictionary network can only lead to the introduction of a new concept in the lexicon of the dictionary.
- Introduction of a new concept in a lexicon can also be associated with the introduction of atleast one word that was not definable at earlier times
- The aforementioned relationship between concepts and loops reflects the basic intuition that new concepts must be self contained and as such the collection of words used to represent them must be self-referential.
- Loops are strictly not isolated but are often linked to form larger, yet still semantically coherent strongly connected components, which represent distinct semantic ideas.
- Possible combinations of Dictionary networks and CBR
  - A Synonym retrieval system - cases can be represented as sub-graphs from the dictionary network. Each subgraph corresponds to a concept in language lexicon.
  - Meaning retrieval system - A system to retrieve the definition of a new word. Somewhat similar to the synonym extracting system.
  - Translation system based on a dictionary network of words and their respective translation.

- Challenge - How to create the case base?
  - Should sub-graphs from dictionary networks be used for cases?
  - Should the case base be created from an abstraction obtained from the dictionary network.

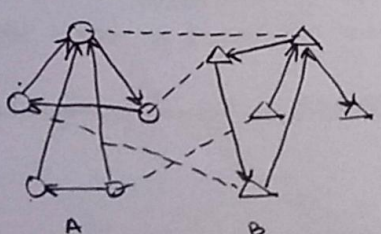
- Using the semantic information from dictionary networks, the retrieval of cases from a case base can be made more efficient and relevant to the test case.

- Footprint algorithm to come up with minimum words/net that explains other words.

Project

- The dictionary network will encode some relationship between words that are present in it. The different options that are there are -
  - semantic relationship
  - Morphological relationship
  - Phonological relationship
- Semantic relationships encode the meaning of the word, how synsets are dependent with one another.
- Other information like Part of speech information etc can also be encoded in the network.
- Since the objective is to understand ~~what kind of~~ the relationships between word meanings in two different resources, we need to define first how to define the meaning of a word from the resources that are being used.
  - For example, wordnet uses synsets to define words present in the resource
  - wikipedia on the other hand define words using article, which contain a large amount of unstructured information.
- Dictionary networks can be used for two tasks
  - language production
  - language learning
  - One of the two needs to be chosen to fix what problem we will work on.
- Things to decide
  - Type of problem (language production / language learning)
  - ~~Resource~~ Multi-lingual / Monolingual network
  - Relationships to use to create the network.
- A possible network structure for a bilingual system.

  2 types of edges in the graph.
  - Same language edges
  - Multi language edges.



The (0, 0) pair are the corresponding article titles in the two languages. Graph A, encodes the word meanings as derived from language A, wikipedia article. Graph B does the same for ~~a~~ language B.

This created network can be used for getting direct translations of words from one ~~sentence to~~ language to another

- If two resources of the same language are used, then the same graph structure can be used to generate ~~the~~ a definition of the word that both the resources agree with.

- The above two are language generation tasks. To be able to generate a viable sentence using the words ~~out~~ of similar meaning, extracted from the graph, a grammar engine will be required for both the languages to be able to generate meaningful ~~se~~ definitions / translation.

- The same graph can also used for language learning tasks. Since a composite resource ~~of multi~~ spanning multiple resources / languages can be created, using the graph mentioned before, a an word-meaning-association mapper can be created for a chatbot or an artificial language generation models..

- This network can be used to extract grammatical ~~rule~~ rules of a language as well.

Project

Text the ~~data~~ graph for the following properties and relationship.

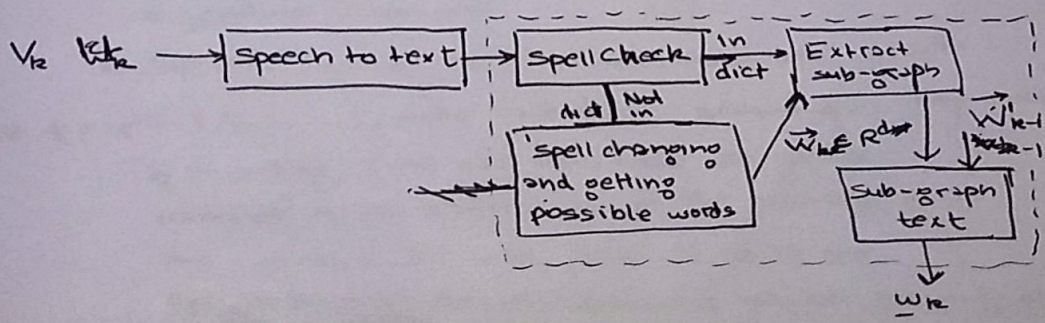
→ 1. This can be applied with the speech to text system, ~~the~~ ~~analysis~~ for DIC. The neural network for the DIC speechToText takes in a voice and converts it to text. There will be ~~issues~~ errors in the output of ~~the~~ speechToText.

A Each word in the output can be checked with the dictionary network, if present, → ~~Assu The A possible word is used~~ (A possible word is the output), then ~~the~~ ~~we~~ a sub-graph will be extracted, with all the words within a distance n from the word, will be extracted and used to test if the word has been correctly deduced).

If the word is not present in the dictionary - dictionary either does not have the word or, the word is not possible. If it is the the latter then the word is input to a spell-corrector that can provide ~~the~~ a spelling that is a word, which is then tested ~~of~~ to see if the word is correctly deduced.., ~~I~~ else we can not do anything and the output will be wrong.

Hypothesis:- This system can help us tell if the dictionary sub-graph have any relation to word use in speech and also help understand what sense of a word ~~can be us~~ has to be used for the particular word.

The following flowchart depicts the ~~epoch~~ system.



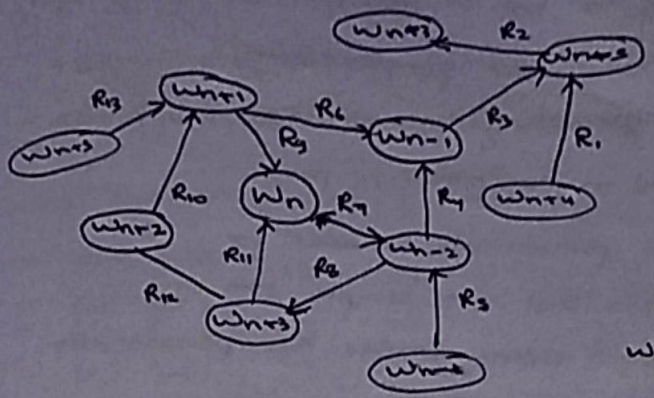
2. Find out ~~whatre~~ what are the different types of graphical properties that tell about the structure of the graph and the relationships between ~~wor~~ nodes and edges. [As ~~one~~ There was a table in ~~some~~ a paper find it]

3. Refer to PPT for more hypothesis and tests.

4. Differentiable edges by semantic relationships like Part of speech. [A pre-trained model for POS tagger can be used.] This when combined with the idea of ~~getting the~~ cores of ~~different~~ can reveal a lot about possible applications of dictionary networks

5. All of these experiments with modifications ~~can be~~ need ~~used~~ to be arranged in the order of complexity to make programming them easy.

6. ~~There~~ The objective could be changed to performing experiments on the graph and ~~on~~ demonstrate possible ~~reason~~ applications of dictionary networks.
If demonstration is not possible, then a detailed enough reason from all the experiments performed, ~~what~~ has to be provided as to why dictionary network would be suitable for the application. Throughout the experiments, more information will be added to compare performance, by adding more data to create more concrete relationships.



where z = number of words in the dictionary

$W_R \in$ Words in Dictionary $R \in \{1, 2, \ldots z\}$

$R_1, R_2 \ldots R_n$ are relationships from one node to another.

This is a directed graph.

Let E be the set of all edges.

Each edge has the value of relationship R_i.

The Relationships may be like

1. R can be the part of speech in the definition of w that was R the word in the preceding node.
2. R can be nothing. The word in the preceding node is just a word in the definition of w. This is the standard dictionary networkes.
3. The R can be a wordNet relationship between the two words.. (Mentioned in PPT).
4. collocational relationships or other relationships possible.

what → Decide what tests to perform all the hypothesis on the last two pages.

5. The value of R will change the structure of the graph. so only one of the above mentioned 3 cases will be a standard dictionary.
6. Capturing loops in the structure. Make sets of with tuples of size 2, 3, 4, 5 ... L, with each filled with loop words of in a loop of the said size, these sets can be examined to see relationships in the meaning of all the words in the loop. Have to decide the test to analyze the relationship between meaning of words in a loop.

- Agendo of tomorrow's meeting

  - Progress of the last decided tasks
  - ~~Be what~~ Decide the pre-processing required for each of the hypothesis described on the previous pages.
  - Write the pseudocode ~~between~~ for all the tests that are simple ~~and~~ and can be found online. Divide the pseudocodes between the two ~~in~~ team members.
  - Discuss further plans
  - Need to do all the things mentioned in the previous pages.
  - ~~Discuss~~
  - Plan next ~~weekend visit~~ meeting with a hard-deadline.
  - Git - create repository on github for the project.:
  -