

Understanding language using Dictionary Networks

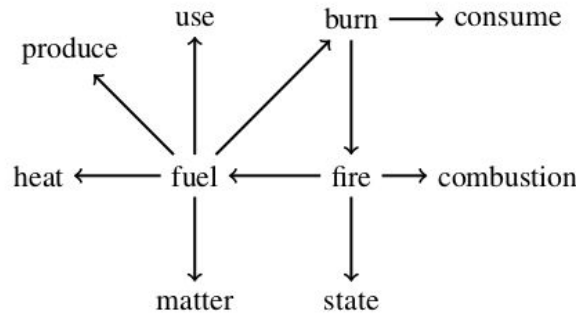
Shreyas Shandilya
Krishna Raj

What are Dictionary Networks?

- A Dictionary is a collection of definitions of words used in a language
- A Network defined by the relationship of words as used in a dictionary exhibits properties that are related to the use and meaning of different concepts as defined by a language.

Properties of a dictionary network and sub-graphs in the network can be studied to understand the semantic, phonological or morphological relationship between words.

Dictionary networks may be called concept networks



FIRE: Fuel in a state of combustion.

FUEL: Any matter used to produce heat by burning.

BURN: To consume with fire.

Figure : 1 : A Toy dictionary network from the definition of three words

Prior work on Dictionary Networks

- BabelNet: Building a Very Large Multilingual Semantic Network; Roberto Navigli and Simone Paolo Ponzetto
- Loops and Self-Reference in the Construction of Dictionaries; David Levary, Jean-Pierre Eckmann, Elisha Moses, and Tsvi Tlusty
- Wiktionary and NLP: Improving synonymy networks; Emmanuel Navarro, Franck Sajous, Bruno Gaume, Laurent Prévot, Hsieh ShuKai and others
- Dictionaries as Networks: Identifying the graph structure of Ogden's Basic English; Camilo Garrido and Claudio Gutierrez
- Collocational networks and their application to an E-Advanced Learner's Dictionary of Verbs in Science (DicSci) : Araceli Alonso, Chrystel Millon, Geoffrey Williams
- Language Resources for a Network-based Dictionary : Veit Reuer

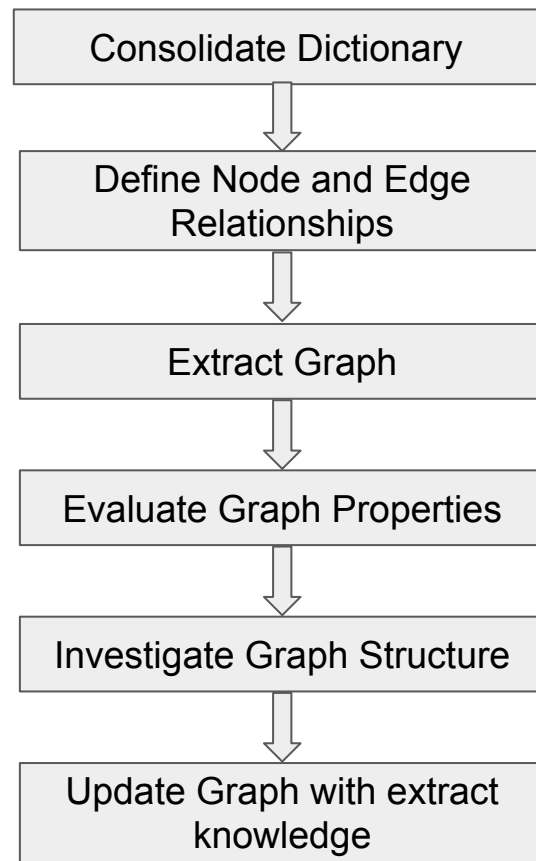


Figure 2 : Methodology involved

Application of Dictionary Networks

- Sense Disambiguation
- Language Etymology Research
- Study of synonymy networks
- As Lexical Resource for the following category of language related tasks -
 - Language Generation
 - Language Learning
- Research of Cognitive aspects of Language

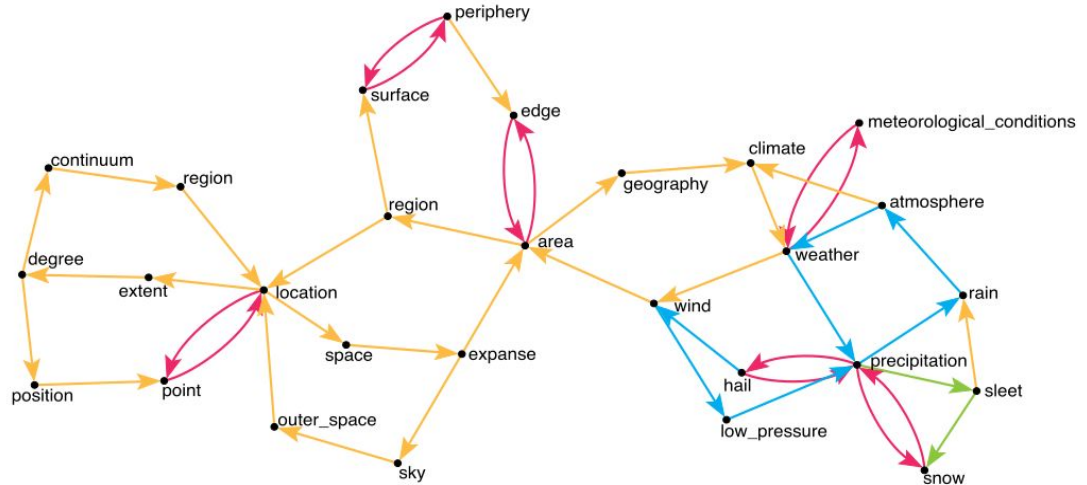


Figure 3 : An example of a large, strongly connected component in the decomposition.

Problem Formulation

- Define a graph $G = (V, E)$, defined by a set V of n vertices and a set $E \subset V^2$ of m edges.
- V is a set of words
- E is defined by a relation R such that $E \rightarrow E : (w_1, w_2) \in E$ if and only if $w_1 \rightarrow w_2$
- Evaluation of the graph structure and comparison with graph of other resources, reveal hidden relationships characterizing the meaning of words and word use

Current Hypothesis

- Graph created from a language resource will have a core: a strongly connected subgraph, with all a very high degree for each node.
- The core consists of words that largely define each other. Loops with just 2 nodes are observed to be very common, thus the dense connections.
- The words in the core are the words that form the smallest units of concepts that then upon combination formed more complex concepts in the language.
- The current hypothesis states that there can be multiple cores, each representing a concept independent of the others in the language. Moreover, these independent concepts are the basic building blocks of the language that give meaning to all words and their relationships

Plan Ahead

- Implementation of the graph from resources in Python
- Experiments with graphs from Wiktionary to investigate the modifications observed by addition of sense knowledge from WordNets.
- Experiments to investigate the existence of independent concepts.
- Experiments with different relationships to be used to define the edges of the graph

Thank You