

# Semantic Network



Samridha M, Akshar Chawla, Shagnik Roy, Neharidha M

**Abstract:** Knowledge representation is an emerging field of research in Artificial Intelligence, Big data analytics, Semantic web, and Data Mining. Knowledge represented in an effective way helps in easy traversal, searching, reasoning, prediction, and inference. There are number of approaches, algorithms, techniques, and models that have been proposed for the same. Every approach has their own pros and cons. Hence our aim is to propose a simple and extremely effective way to represent knowledge, which has a greater expressiveness compared to logic and reverberates within the methods of people process data. In this paper we address the implementation of semantics network with simple yet powerful method, which despite the vague nature of English language, proves to give accurate results. The patterns discovered in sentences are listed in a defined order throughout the paper. The usage of Natural Language Toolkit and Posing tagging simplifies the most crucial task of tagging a word with its part of speech.

**Keywords:** Clauses, Natural Language Toolkit, NetworkX, Ontology, Pos tagging.

## I. INTRODUCTION

Human beings are good at understanding, reasoning, interpreting knowledge and are able to perform various actions in the real world using knowledge. For machine to do the same we have to represent knowledge in a form that helps it perform reasoning and interpretation using Artificial intelligence. Knowledge representation and reasoning represents information form real world for a computer to understand and then utilize this knowledge to complex real-life problems like communicating with human beings in natural language [1]. Knowledge representation is not just storing data in a database, it allows a machine to learn from that knowledge and behave intelligently like a human being. There are many approaches to knowledge representation, our paper focuses on representing knowledge using Sematic Networks. Semantic Networks are one of the oldest and most effective techniques for knowledge representation. A semantic net is a graph structure for representing knowledge in patterns of interconnected nodes and arcs, with nodes representing real world objects and arcs representing relationship between the nodes. Computer implementations of semantic networks were first developed for artificial intelligence and machine learning, but earlier versions have

been long used in philosophy, psychology, and linguistics [2]. They are graph structure that encode taxonomic knowledge of objects and their properties. Reasoning about property and set membership is easier and more efficient in a semantic network [3].

## II. TOOLS USED

### 1. Natural Language Toolkit

Natural Language Toolkit is a powerful suite of open source program modules, tutorials, and problem sets, providing ready to user computational linguistics courseware [4]. It contains libraries for symbolic and statistical Natural Language Processing for English written in python programming language. It contains text processing libraries for tokenization, parsing, classification, stemming, lemmatization, tagging and sematic reasoning.

### 2. Tokenization

Tokenization is process of breaking a stream of textual content up into words, terms, symbols, or some other meaningful elements called tokens [10]. Generally, tokenization occurs at the word level. Some of the tools for tokenization are Nlpdotnet Tokenizer, Mila Tokenizer, NLTK word Tokenizer.

### 3. Pos tagging

Part-Of-Speech (POS) tagging is the process of identifying nouns, verbs, adjectives, pronouns, conjunction, and other parts of speech in context. POS tagging can be used for Linguistic-text pre-processing before semantic analysis [5]. The state-of-art POS taggers achieve an accuracy of 83.6%-90.5% when applied on bug reports. The Stanford POS tagger performs the best, followed by TreeTagger[6]. Frequently used tags and their meaning are listed below and henceforth the given below abbreviation will be used that will be appearing throughout this paper are:

- NNnoun, singular
- NNS noun plural
- NNP proper noun, singular
- NNPS proper noun, plural
- VB verb, base form
- VBD verb, past tense
- VBG verb, gerund/present participle
- VBN verb, past participle
- VBP verb, sing. present, non-3d
- VBZ verb, 3rd person singular present
- RB adverb
- RBR adverb, comparative
- RBS adverb, superlative
- JJ adjective
- JJR adjective, comparative
- JJS adjective, superlative

Revised Manuscript Received on October 10, 2020.

\* Correspondence Author

**Akshar Chawla\***, B Tech in Computer Science and Engineering, Manipal Institute of Technology, Manipal, India.

**Shagnik Roy**, B Tech in Computer Science and Engineering, Manipal Institute of Technology, Manipal, India.

**Samridha M**, B Tech in Computer Science and Engineering, Manipal Institute of Technology, Manipal, India.

**Neharidha M**, B.E in Electronics and Communication, NMAM Institute of Technology, Nitte, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

- CC coordinating conjunction
- CD cardinal digit
- DT determiner
- EX existential there
- IN preposition/subordinating conjunction

For complete list refer [9].

## 4. NetworkX

NetworkX package provides classes for graph objects, generators to create standard graphs, IO routines for reading in existing datasets, algorithm to analyze the resulting networks and some basic drawing tools. It provides modularity of code and documentation. The graph internal data structures are based on an adjacency list representation and implemented as a python dictionaries; the outer dictionary is keyed by nodes to values that are themselves dictionaries keyed by neighboring node to the edge attributes associated with that edge. This “dict-of-dict” structure allows fast addition, deletion and lookups of nodes and neighbors in large graphs [7].

## III. UNDERSTANDING CLAUSES

Understanding clauses forms the basis for understanding the structure of sentences. Clause is a group of words, consisting of a subject and a finite form of a verb [8]. The two types of clauses are independent and dependent clause. Independent clauses are part of sentences that consist of a subject and a verb and requires no extra information in order to under. Dependent clauses are part of sentences that consist of a subject and a verb but do not express a complete thought. Independent clause starts subordinate conjunction such as “while”, “that”, “unless” etc. They give the background information but cannot stand on their own as sentences.

## IV. UNDERSTANDING SENTENCES

Understanding the types of sentences and their structures forms the basis for creating semantic network. Sentence is a group of words, usually containing a verb, that expresses a thought in the form of a statement, question, instruction or exclamation and starts with a capital letter [8]. Sentences can be broken down into clauses. Sentences can be broadly classified into simple, compound and complex sentences.

## V. METHODOLOGY

Text which is to be represented in the form of network is broken down into sentences using the fact that each sentence starts with capital letter and ends with a dot. Sentence which is in the form of string must be tokenized. NLTK provides tokenize function which converts string into list array of tokens. Each token should be tagged with its part-of-speech. NLTK provides pos-tag function which tags tokens with its part-of-speech and produces a 2D array as output. Further all processing uses this 2D array. From sentence which are in the form of 2D array clauses are extracted, depending on the number of clauses, types of clauses and how clauses are joined (when there are more than one clause) sentences are classified into simple, compound, and complex. Graph for the respective type of sentence is plotted using NetworkX tools.

### 1. Method to extract clause

All clauses follow subject-verb-object structure which is one of the simplest structures. Subject of the clause is part of

the sentence that contains the person or thing performing the action (or verb) in the sentence. clause may have compound subjects (two or more subjects) and they are separate with coordinating conjunction or comma. Inference can be drawn that tokens preceding the verb (VB, VBD, VBG, VBP, VBN, VBZ) are subject/object and are added to a set (from now onwards called subject set). From the subject set token with CC, CD, DT and EX tagged tokens are removed, since subject is person/object. Node is created for every element in subject set. Verb part of clause is all token with VB, VBD, VBG, VBP, VBN, VBZ tag. Clause can have compound verbs (two or more verbs), it is dealt in the same way as compound nouns were dealt. Clause may have optional noun at the terminal, so end of the verb part is marked by NN, NNS, NNP, NNPS or dot. If verb part of clause is not terminated with dot, then there is addition noun part whose end is marked with dot. The tokens in this part of clause are extracted and added to a set (from now onwards called object set). Node is created for every token in object set. For each clause a one-to-one mapping is created from subject set to object set with verb as edge label.

### 2. Method to determine the type of sentence

Simple sentences have only one independent clause and no dependent clause. Compound sentences have at least two independent clauses that have related ideas. The independent clauses are joined by a coordinating or by a semicolon. Coordinating conjunctions are “for”, “and”, “nor”, “but”, “yet”, “so” and “or”. Compound sentences will contain a comma before conjunction for correct punctuation. Any sentence which follows this pattern is a compound sentence

Complex sentences have one independent clause and at least one dependent clause. Subordinating conjunction link an independent clause to a dependent clause. The most common subordinating conjunctions are “although”, “because”, “before”, “even though”, “if”, “since”, “until” and “when”. Any sentence which follows this pattern is a compound sentence.

### 3. Miscellaneous minor details

After the type of the sentence is decided, parts of speech like pronoun, adjective and attributive nouns are dealt with

#### A. Pronoun

Pronouns are used in sentences to avoid repetition and make sentences easier to understand. A pronoun should refer clearly to one, clear, unmistakable noun coming before pronoun. Unfortunately, there is a good chance of faulty or vague pronoun reference. Some commonly made mistakes and how it can be dealt are given below

- Hidden antecedents

e.g.: The candy dish was empty, but we were tired of eating it anyways.

It appears as dish is being eaten, obviously dish cannot be eaten. Here dish functions as an adjective rather than a noun but NLTK considers it a noun germanely it considers most of the adjective as proper noun, singular (NNS) later sections elucidate how adjectives can be dealt.

- Too many antecedents:

e.g.: Take the radio out of the car and fix it.

Anyone who read this sentence would not know which item was to be fixed. This can be dealt by positing the pronoun's antecedent as the one coming before it.

**B. Adjective**

Adjectives are words that modify (or describe) nouns in a sentence, nevertheless sometimes they perform the function of the noun in the sentence/clause (Nominal adjective). They are preceded by the word "the" and can be found as the subject of a clause.

e.g.: The elderly are a great source of wisdom.

So, whenever adjective(JJ) preceded by the it should consider as noun (NN, NNS, NNP, NNPS).

**C. Attributive noun**

Attributive noun is a noun that modifies another noun and functions as an adjective. Also known as noun premodifier, a noun adjunct, and a converted adjective.

E.g.: We obtained the permit from a government official.

So, when there are two consecutive nouns and whenever the second noun does not start with a capital letter there is a good possibility of existence of attributive noun. According to Geoffrey Leech attributive noun of a sequence will be singular, yet there are examples of plural attributive noun, examples include "sports car," women leaders,". Considering Geoffrey Leech's statement as a posit we can consider NN and NNS as adjective when occurring with NNP or NNPS.

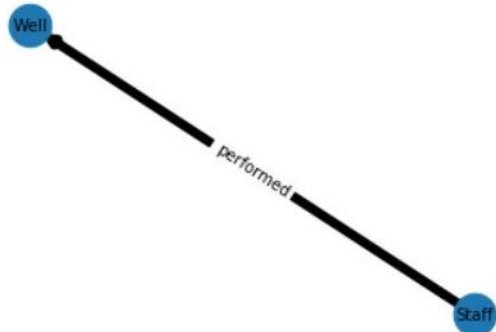
**VI. RESULT AND DISCUSSION**

**A. Simple Sentences**

- Simple sentence with single subject and single verb.

Input sentence: The staff performed well.

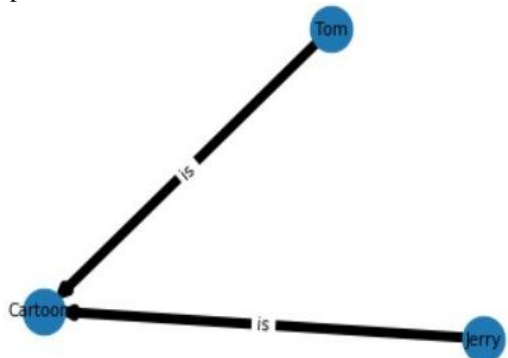
Output semantic Network:



- Simple sentence with compound noun

Input sentence: Tom and Jerry are cartoons

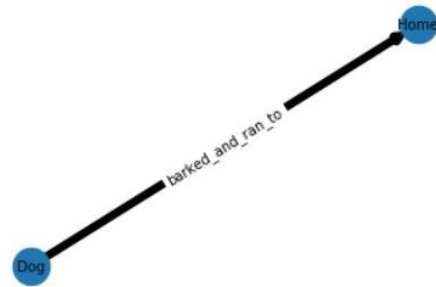
Output semantic network:



- Simple sentence with compound verb

Input sentence: The dog barked and ran home

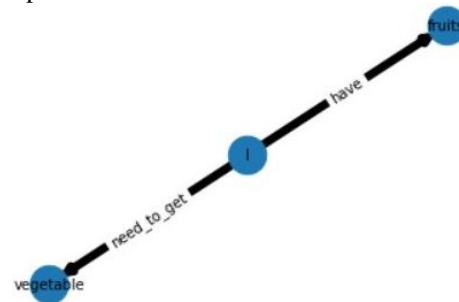
Output semantic network:



**B. Compound Sentences**

Input sentence: I need to get a vegetable and I have fruit.

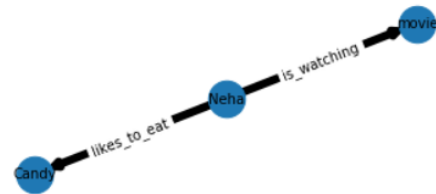
Output semantic network:



**C. Complex Sentences**

Input sentence: Neha likes to eat candy while she is watching Movie.

Output semantic network:



The methodology proposed works for best for proposed types of sentences.

**VII. CONCLUSION**

In the proposed methodology we dealt with three of the four types of sentence structures-based clauses. The methodology can be updated in future when required, to classify sentences into an additional type of complex-compound and build graph for the same. Since the usage of complex-compound sentences is not seen much in modern English this type has not been considered while proposing the methodology. Creating semantic network for any English sentence from the perspective of number and nature of Clause in sentence gives greater expressiveness, and accurate results. The proposed method is very simple and can be implemented with ease. The semantic network produced has enormous applications in various fields like Artificial Intelligence, Big data analytics, Semantic web etc. It can also be used for various other purposes like "Automatic question and answer development system".



## REFERENCES

1. Matthew Huntbach, *Artificial Intelligence I*, Queen Mary and Westfield College, London. Available: <https://www.coursehero.com/file/10092029/AINotes4/>
2. Atta ur Rahman, *Knowledge Representation: A Semantic Network Approach*, Handbook of Research on Computational Intelligence Applications in Bioinformatics, June 2016.
3. Nils J Nilsson, *Artificial Intelligence: A New Synthesis*, 1998.
4. Steve Bird, *NLTK: the Natural Language Toolkit, TMTNLP '02: Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics - Volume 1*.
5. Nyein PyaePyaeKhin, *Analyzing Tagging Accuracy of Part-of-speech Taggers*, International Conference on Genetic and Evolutionary Computing, September 2015.
6. Yuan Tian and David Lo, *A Comparative Study on the Effectiveness of Part-Of-Speech Tagging Techniques on Bug Report*, School of Information systems, Singapore Management University, Singapore.
7. Aric A. Hagberg, Daniel A. Schult and Pieter J. swart, "Exploring network structure, dynamics, and function using NetworkX", in Proceedings of the 7<sup>th</sup> Python in Science Conference (SciPy2008), GaelVaroquaux, Travis Vaught, and Jarrod Millman (Eds), (Pasadena, CA USA), pp. 11–15, Aug 2008.
8. Cambridge Dictionary, Cambridge university press, 2020.
9. Steve Bird, Ewan Klein and Edward Loper, "Natural Language Processing with Python", Chapter 5.
10. Dr. S.Vijayarani and Ms. R.Janani, "TextMining: Open Source Tokenization Tools-An Analysis", *Advanced Computational Intelligence: An International Journal (ACIJ)*, Vol3, No.1, January 2016.



**Neharidha M** is pursuing Bachelor of Engineering in Electronics and Communication from, NMAM Institute of Technology, Nitte, India. She is enthusiastic electronics Engineer with interest in developing her technical and management skills. She has a strong technical ability combined with an analytical approach with good communication skills. She is actively involved in various research activities and projects in her institution. She has worked on topics like advance Computer Communication, Digital Electronics and Antenna for medical application. Her research interests include the fields of Machine Learning, Big Data, Wireless networks. Robotics, IoT. Her hobbies include singing, reading novels, solving word puzzles, playing badminton.

## AUTHORS PROFILE



**Samridha M** is pursuing Bachelor of Technology in Computer Science and Engineering from Manipal Institute of Technology, Manipal, India. She is an enthusiastic Computer Science Engineer with interest in developing her technical and management skills. She has a strong technical ability combined with an analytical approach with good communication skills. She is actively involved in various research activities and projects in her institution. She aspires to become a cryptographer. Her research interests include the fields of Cryptography, Network security, Systems Security, Wireless networks, Block Chain, Edge Computing and IoT. Her hobbies include swimming, reading science novels, solving word puzzles, playing badminton.



**Shagnik Roy** is pursuing Bachelor of Technology in Computer Science and Engineering from Manipal Institute of Technology, Manipal, India. He is an enthusiastic Computer Science Engineer with interest in developing technical skills. He has a strong technical ability combined with Analytical approach with good smangement skills. He is actively involved in various projects in his institution. He has also worked as an Intern in Skybits Technologies & EBIW Info. Analytics Pvt. Ltd. In order to enhance his knowledge in the field of Web Scraping & Data Science. He aspires to be a Data Scientist. His research interests include the fields of Machine Learning, Artificial Intelligence and Robotics. His hobbies include solving varieties of puzzles, playing football, tennis.



**Akshar Chawla** is pursuing Bachelor of Technology in Computer Science and Engineering from Manipal Institute of Technology, Manipal, India. He is actively involved in various research activities and projects in his institution. He has shown keen interest in software development. He has also worked as an Intern in Verzeo. He worked on many projects which involves concepts of Computer Vision and Artificial intelligence. His research interests include the fields of Computer Vision, Big Data, Data Science and Artificial Intelligence and he aspires to work on these fields in future. His hobbies include gaming, solving math puzzles, playing basketball, gardening, and listening to music.

