International Refereed Journal of Engineering and Science (IRJES)
ISSN (Online) 2319-183X, (Print) 2319-1821
Volume 4, Issue 10 (October 2015), PP.01-07

Semantic Representation for Natural Languages

Madhuri A. Tayal[1], M. M. Raghuwanshi[2], Latesh G. Malik[3]
[1](Research Scholar, G. H. Raisoni College of Engineering, Maharashtra, INDIA)
[2](Professor, YCCE, Nagpur, Maharashtra , INDIA)
[3](HOD, G.H. Raisoni College of Engineering, Nagpur, Maharashtra, INDIA)

Abstract:- For representing content of the text Semantic representation is a vital issue. Natural language processing is an arena of artificial intelligence and computational linguistics associated with the communications amongst processors and humanoid languages. The data is processed through word analysis, Grammar analysis, meaning analysis, Pragmatic analysis and Discourse processing. This paper verifies numerous semantic representation systems. The predicate logic algorithm in this paper separates the English sentences into shares of it. Then by in view of the categories of sentences (Simple, Questioning, Exclamatory etc.) sentences are converted to representation. Through NLP Chunker also the representation of sentences is accomplished. The algorithm has attained correctness of 75%. This illustration can be used in future for many NLP tasks.

Keywords:- POS tagger, Chunks, Chunker, FOL, Natural language

I. INTRODUCTION

Language is the main resource of communication utilized by the individuals. To show the better part of ideas and emotions it is the tool the whole world uses. It forms thought, has a arrangement, and conveys significance.

Semantic Illustration is an essential part of investigation in computational linguistics. Meaning of the sentences should be characterized in computer in such a method that can be recovered simply and powerfully. Many applications of representations are used in important areas of computational linguistics like machine translation, storytelling, question-answering,.[13], [14]. The elementary trouble of semantic representation is the development of an adequately specific notation for in place of semantic. Such symbolization is denoted to as a semantic representation technique [10], [11], [14].

Semantic is related with the implication of the language. The overall idea of semantic understanding is to take natural sentences and plot them onto certain demonstration of meaning. Checking whether there is a functional set of guidelines in the programming language or not it is verified by the semantic analysis. Any old noun phrase tracked by some verb phrase creates a syntactically accurate English sentence, a semantically accurate one has subject-verb organization, correct use of gender, and the constituents go together to direct an idea that creates logic. Semantic analysis is the forward-facing penultimate stage and the compiler's last turn to clear out inappropriate programs.

The real use of these representations has been verified by the creators of the illustration. Numerous researchers esteemed the meaning of semantic representation methods [4], [5], [6], [7], [8], [9]. The chief purpose of semantic representation is the formation of target language illustration of a sentence's there is a significant role that it shows.

This paper defines an illustration using predicate logic for English language reliant upon the dissimilar types of sentences. Section-2 delivers the facts for different ways of demonstrating semantic.Section-3 signifies algorithm for carrying out of First order logic. (FOL).Section-4 defines the investigational results for FOL algorithm.Section-5 defines how this illustration is useful for text summarization task. Section-6 advises conclusion and future work.

II. DIFFERENT WAYS OF REPRESENTING SEMANTIC

The succeeding is known models. They are used for semantic representation by many people.

A. Logic

For demonstrating actualities and properties of a world in a exact, unmistakable way is called as logic.[13].

The proposition is the important idea of propositional logic. A true or false value can be assigned to proposition. One often signifies propositions by sentences in a humanoid language.

First-order predicate calculus or FOL enhances predicates which can symbolize properties, e.g., mortal (person), or relationships, e.g., likes (Robert, Ice-cream),

- Existentially quantified variables, e.g., ∃ There exist X such that...

- Universally quantified variables, e.g., ∀ for all X such that...

Examples of predicate logic statements:

1. Rahul is a boy.: Boy (Rahul)
2. Machine made the products: made(machine, products)

Predicate logic uses the schemes of function, predicate, logical connectives, constant, variable, and quantifiers to indicate realities [13].In predicate logic, sentences can be separated into words e.g. nouns, pronouns, verbs and adjectives or even expressions. Such as there are frequently finite numbers of words or phrases in a language, therefore one can simply store words or phrases for demonstrating the semantic e.g. in the form of text. By means of this grouping of number of phrases, nouns, Amjad Ali, Mohammad Abid Khan [1][2][3] have suggested the semantic representation technique. In distinction to this method for representation of semantic, the planned method is founded upon the categories of sentences. The classification is done on the base of the kinds of the stuffs of that sentence. e.g. Inquiring, Simple sentence, Passive, facts, Exclamatory, etc. The complete process is stated in the next section.

The sentence can be retrieved efficiently using predicate logic. This is the advantage of predicate logic.

B. Rules

This is the powerful and simple method of semantic representation providing the flexibility of merging declarative and procedural illustration for using them in a combined form [8],[13]. Samples of production rules:

−IF complaint THEN accomplishment

−IF statement THEN decision

−IF proposition p1 and proposition p2 are true THEN proposition p3 is true.

A Rule-based system comprises a precise match the condition(s)/ premise(s) to predict the conclusion(s). This is very regulating, as real-world conditions are often undefined and do not contest precisely with rule requirements. The arrangement of a rule-based system displays at least two modules, the semantic base (Rule-base) and the inference engine. The Rule-base includes the domain semantic in the arrangement of rules. The inference engine signifies the reasoning scheme for examining the rules, in a semantic base, which allows finding an appropriate answer of a problem. [15].

The benefits of production rules are as follows:

1. Easiness.
2. Rules can be assumed by a new person also.
3. Rules can be modified.
4. Worthy for composite problems where persons have skilled knowledge.
5. Every single instruction/rule describes a minor and self-determining part of knowledge.
6. Rules are generally individually of extra rules.

This method is not well-organized representation arrangement for the knowledge representation. It is used merely for the representation of rule based knowledge. When definite rules are satisfied in a rule based system, then the system provides single answer and when certain other rules are fulfilled then it delivers a changed solution. Using the illustration of a sentence individual should be capable to recover the similar input text/sentences and from rule founded representation the retrieval of novel text/sentence is very tough job. Though Rules are having their individual applications similar, Medicinal Opinion, Instruction training, Prediction based schemes and Monitoring applications, for rural applications etc.

C. Frames

A frame is a type of a data-structure for signifying a categorized condition, like being in a sure kind of kitchen, or going to a wedding party. Involved to every single frame are a number of types of data. Specific of this fact is approximately how to use the frame.

We can reason of a frame as a system of nodes and associations. The "top levels" of a frame are secure, and signify things that are continuously correct about the invented condition. The minor stages have many terminals–"slots" that must be engaged by detailed instances or data. Every terminal can identify conditions its assignments must encounter. Easy circumstances are stated by markers that might need a terminal transfer to be a person, an object of adequate charge, or a pointer to a sub-frame of a definite type. Additional complex conditions can require relations amongst the things allocated to numerous terminals.

Groups of connected frames are associated together into frame-systems. The effects of significant actions are reflected by alterations amongst the frames of a system. These are used to create certain kinds of calculations inexpensive, to represent modifications of importance and consideration, and to account for the usefulness of "imagery.

Table 1: Frame for Newspaper

Slots	Fillers
Publisher	**Rohny**
Title	**Times of India**
Author	**H Kalen**
Edition	**Fourth**
Established Year	**1956**
Pages	**48**

A frame is not having capability of knowledge representation arrangement in linguistics. Because if someone requests to represent sentences of a language, it is problematic to piece the sentences into slots and their Fillers and vice versa is also very tough. Once sentences are represented in frames, then the supportive arguments of a sentence are not reserved and thus due to absenteeism of supportive arguments in representation the recovery of the similar input sentence is very thought-provoking.

D. Semantic Net

A semantic network or net is a diagram construction for representing knowledge in designs of interrelated nodes and arcs. Computer operations of semantic networks were mainly established for machine translation and artificial intelligence, but previous forms have long been used in psychology, philosophy, and linguistics.

A declarative graphic representation is shared to all semantic networks to represent knowledge and support automated systems for thinking about the knowledge. Some forms are extremely informal, but others are legally definite systems of logic. Subsequent are six of the maximum shared kinds of semantic networks. Accurately a semantic net is a well-defined as a labeled directed graph [13].

Recovering the real and accurate data from semantics networks is a very problematic task. Explicitly, if someone desires to stock and signify a text in processor in which there is no idea of objects and relations, then semantic network representation scheme is not suitable and capable to represent and make progress such text.

III. FOL Algorithm

The various stages for the FOL procedure are as follows.

1. Take POS tagger and associate it with Java. (Stanford Tagger)[12].
2. Sentence is required to be entered.
3. POS tagger (Tags yields (tags)).
4. Generating array of verbs. Allocate index to verb.
5. Generating array of subjects. Allocate index w.r.t verb.
6. Generating array of objects. Allocate index w.r.t verb.
7. Verify the type of sentence using end symbols and dissimilar determiners like (’?’, ’!’, ‘.’,’by’, CD, JJ etc.).
8. Inside sentences are organized into (Subject Verb Object), find Verb in the sentence, and keep it first. Find subject and Object, keep it therefore in FOL forms [Verb (Subject, Object)].
9. On behalf of complex sentences(S V O V O V O), catch central Subject and verb, then for the remaining of the sentence once more discover subject and verb.
10. If only one main subject found do step 8, else use conjunctions to link sentences.[Verb1 (S1, O1) ^ Verb2 (S2, O2)].

The phases for rebuilding the sentences after FOL are.

1. Recreate the sentence by means of Subject main S, subsequent Verb and then the object.
2. If it is passive sentences then rebuilding uses object main subsequent Verb and then the Subject.

A. *POS Tagger*

A Part-Of-Speech Tagger is a software routine that recites text in some language and allots parts of speech to every single word. It allots a part-of-speech like preposition, adverb, noun, verb, pronoun and adjective or other vocabulary class marker to every word in a sentence. This software is executed in Java. Amounts of Taggers are accessible Stanford Tagger; Apache Tagger, Berkerly Tagger, etc. are some of them. Out of which Stanford tagger has been used. Its simple download comprises two trained tagger models for English. Separately versions cover the similar source and other required files. The tagger can be reeducated on any language, given POS-annotated exercise text for the language. The contribution to a tagging algorithm is a arrangement of words of a natural language sentence and a enumerated tag set (a limited list of Part-of-speech tags). The various tags returned by a tagger is shown in table-2.

Table 2: POS Tagged output and respective meanings.

Tagger o/p	Meaning	Tagger o/p	Meaning	Tagger o/p	Meaning
CD	Cardinal Number	NNPS	Proper Noun, plural	TO	to
CC	Coordinating conjunction e.g. and, but, or...	NNS	Noun, plural	VBN	Verb, past participle
DT	Determiner	PDT	Predeterminer e.g. all , both ... when they precede an article	UH	Interjection e.g. uh, well, yes, my...
EX	Existential there	POS	Possessive Ending e.g. Nouns ending in 's	VB	Verb,baseform subsumes imperatives, infinitives and subjunctives
FW	Foreign Word	PRP	Personal Pronoun e.g. I, me, you, he...	VBD	Verb, past tense includes the conditional form of the verb to be
IN	Preposition or subordinating conjunction	PRP$	Possessive Pronoun e.g. my, your, mine, yours...	VBG	Verb, gerund or persent participle
JJ	Adjective	RB	Adverb Most words that end in -ly as well as degree words like quite, too and very	VBP	Verb, non-3rd person singular present
JJR	Adjective, comparative	RBR	Adverb,comparative Adverbs	VBZ	Verb, 3rd person singular present
JJS	Adjective, superlative	RBS	Adverb, superlative	WDT	Wh-determiner e.g. which, and that when it is used as a relative pronoun
LS	List Item Marker	RP	Particle	WP	Wh-pronoun e.g. what, who, whom...
MD	Modal e.g. can, could, might, may...	SYM	Symbol used for mathematical, scientific symbols	WP$	Possessive wh-pronoun
NN	Noun, singular or mass	TO	to	WRB	Wh-adverb e.g. how, where why
NNP	Proper Noun, singular				

B. Different Kinds of Sentences

Sentences comprises of Subject, Verb and Object agreeing to Wren and Martin. So, each sentence has to have a subject(S), Object (O) and a Verb (V). Particular sentences may have adjectives, adverbs and conjunctions. There are also sentences which are inquiring i.e. they ask a question and there are some which are exclamatory. So, sentences are considered in diverse type. It is important to sort sentences because the POS tagger gives the sentences as assembly of words. The meaning is not at all considered at that time. The foundation for the process of cataloguing is shown in the table 3. The cataloging is as follows:

1. Sentences having only subject, one verb and one object. (SVO)
2. Sentences having precisely one subject, one verb, one object and adjectives also.(SVO with ADJECTIVES)
3. Simple statement statements. (STATEMENT)
4. Sentences holding more than one noun and verbs. (COMPLEX)
5. Sentences which enquiry. (QUESTIONING)
6. Sentences which shout. (EXCLAMATORY)
7. Sentences which cover assisting verbs. (MORE THAN ONE VERB)
8. Sentences which comprises numeral value/statistics. (NUMBERS).
9. Sentences which are passive. (PASSIVE).

Table 3: Categorization of English sentences

Basis of categorizing	Category
With only single subject, one verb and one object.	**Subject, Verb, Object**
With only one subject, verb, and adjective monitored by a verb.	**SVO with adjective(JJ, JJR, JJS)**
Preliminary with This, That.	**STATEMENT**
Sentences having more than single subject or object and having "and"…"or" in it.	**Complex**
Sentences dismissing with a "?".	**Questioning**
Sentences dismissing with a "!".	**Exclamatory**
Sentences which comprises an assisting verb.	**More than one verb**
Sentences in which the subject keep an eye on "by".	**Passive**
Sentences in which the statistics are given.(Numbers-"MD"(POS tagger returns))	**Numbers**

Act out human language accepting on the computer is a big test. Natural language meanings in logical form to be represented are one of the approaches. The sentences and their equivalent reorientation results are revealed in next section.

C. NLP Chunker

The NLP Chunker tool will group the tokens of a sentence into larger chunks, each chunk corresponding to a syntactic unit such as a noun phrase or a verb phrase. It is useful in itself when looking for units of meaning in a sentence larger than the individual words. To perform the chunking task, a POS tagged set of tokens is required.

Ram had gone to Nagpur last summer.
Input sentence
[NP ram/NN] [VP had/VBD gone/VBN] [PP to/TO] [NP nagpur/NN] [NP last/JJ summer/NN] ./.
Corresponding Chunker output

Chunker Output	Associated Meaning
ADJP	**Adjective Phrase**
PP	**Prepositional Phrase**
ADVP	**adverb phrase**
PRT	**particle**
CONJP	**conjunction phrase**
SBAR	**Clause introduced by a subordinating conjunction**
INTJ	**interjection**
UCP	**unlike coordinated**
LST	**List marker**
vp	**Verb Phrase**
np	**Noun Phrase**

IV. FOL ALGORITHM USING NLP-CHUNKER

The steps for the FOL algorithm using NLP-chunker are as follows.

1. Download NLP Chunker.
2. Connect NLP Chunker with java.
3. Enter the sentence.
4. Convert the sentence into lowercase.
5. Use Chunker output and Split the sentences using ‘]’ symbol.
6. Remove all capital letters and symbols.
7. Create three arrays, containing Subject, Verb and Object.
8. Find first occurrence of Verb phrases and Noun phrases.
9. Arrange the remaining word phrases in the array named Object.
10. Rearrange the sentence in the form of :
    V[ ](S[ ],O[1]….O[n]).
    Sentences can be restructured using subsequent steps.
    The sentences can be rebuild using

1. Subject first ‘S’.
2. Then read the Verb phrase
3. Read the object phrase.

V. RESULTS

The sentences and their corresponding semantic representations are shown in the table 4.

Table 4: Results for Semantic Representation

Type of Sentence	Example	Output
SVO	The lion was lifting its paw again and again.	was_ lifting(the_ lion, its_paw, again_ and_ again)
SVO with Adjectives	He treated his slave badly.	treated(he, his_slave, badly)
STATEMENT	Computer helps drawing a picture.	helps_ drawing(computer, a_ picture)
Complex	A database is an organized collection of facts and information.	is(a_ database, an_ organized_ collection, of, facts_ and_ information)
Questioning	What is your job?	is(what, your_ job)
Exclamatory	What an idea!	(what, an_ idea)
More than one verbs	Computer helps drawing a picture.	helps drawing(Computer, picture)
Passive	Football is being played by Shyam.	is_ being_ played(football, by, shyam)
Numbers	The mountain is 800 meters high.	is(the_ mountain, 800_ meters, high)

VI. CONCLUSION

This paper stresses on semantic representation for natural language. It defines the Semantic Representation for English Language. Different semantic representation systems are obtainable and determined that predicate logic representation i.e. FOL is the modest and finest method for Semantic Representation. This algorithm calculates first order logic illustrations for sentences and also rebuilds the sentences from them using POS tagger as well as using NLP-chunker. The correctness of the system is further enlarged through NLP-Chunker. It creates the word phrases for the sentences. NLP-Chunker benefits while rebuilding the sentences more precisely. In future using some more tools the efficiency of the scheme can be enhanced.

REFERENCES

[1]. Amjad Ali, Mohammad Abid Khan, “Selecting Predicate Logic for Knowledge Representation by Comparative Study of Knowledge Representation Schemes” in 2009 International Conference on Emerging Technologies, IEEE 2009, pp 23-28.

[2]. Amjad Ali, Mohammad Abid Khan, Knowledge Representation of Urdu Text Using Predicate Logic, 2010 6th International Conference on Emerging Technologies (ICET), IEEE, pp. 293-298.

[3]. Dalia Fadl, Mostafa Aref, Rich Semantic Representation Based Approach for Text Generation, Ibrahim Fathy, The 8th International Conference on INFOrmatics and Systems (INFOS2012) pp- 20-28

[4]. Ralph M. Weischedel, Knowledge Representation and Natural Language Processing, Proceedings of the IEEE, Vol. 74, No **7,** July 1986, pp 905-920.
[5]. Mitsuru Ishizuka, Kazuya Tananashi, Hermut Prendinger, A Predicate Logic Version Method for Cost based Hypothetical Reasoning Employing an Efficient Propositional level Mechanism,
[6]. Aikaterini Mpagouli, Ioannis Hatzilygeroudis, Converting First Order Logic into Natural Language: A First Level Approach, University of Patras, Hellas.;
[7]. Toshiaki Fuziki, Hidetsugu Nana, Manau Okumura, Automatic acquisition of script knowledge from a text collection, EACL 2003 Budapest, Hungry, pp 91-94.
[8]. M. Minsky, A Framework for Representing Knowledge, The Psychology of Computers vision, McGraw Hill, 1975.
[9]. S. Sawai, H. Fukushima, M. Sugimoto, And N. Ukai Knowledge Representation and Machine Translation, COLING 1982, pp. 351-356.
[10]. P. J. Hayes, Some Problems and Non-Problems in Representation Theory, Proceedings AISB Summer Conference, July 1974, Essex University
[11]. STANFORD PARSER: http://nlp.stanford.edu/software/lex-parser.shtml
[12]. STANFORD POS TAGGER: http://nlp.stanford.edu/software/tagger.shtml
[13]. Books
[14]. Rich and Knight, "Artificial Intelligence", TATA Mc Graw Hill Second Edition.
[15]. Tanveer Siddiqui, U.S. Tiwari, Natural language Processing and Information Retrieval, (BOOK) Oxford University Press.
[16]. John Durkin, Expert Systems-Design and Development, Prentice Hall.