

# Generating UML Diagram Using Natural Language Processing and Use Case Diagram

Khan Abdul Shariff<sup>1</sup>, Faraz Alam<sup>2</sup>, Mohd. Asmar<sup>3</sup>, Amit Sharma<sup>4</sup>

<sup>1,2,3</sup>Student, Department of Computer Science & Engineering, B.B.D.I.T., Ghaziabad, India

<sup>4</sup>Professor, Department of Computer Science & Engineering, B.B.D.I.T., Ghaziabad, India

**Abstract:** This research paper presents a natural language processing based mechanized framework for producing UML graphs subsequent to dissecting the given Use case outline in the present situation. This model speak to investigating natural language processing and concentrate the relative data from the given UML chart which are given by the client. Client characterizes all the necessity in basic information dialect of machine language which are without setting dialect and punctuations. Subsequent to consolidating all the data we set up a relative model in the sentence structure dialect. Normal dialect preparing convert this in machine language comprehended to the machine. After this data, we can create different UML graph as activity diagram, succession diagram, use case diagram, class diagram and other many case instruments. These blueprint and standard CASE tools require a lot of extra time and attempts from the system analyst in the midst of the path toward making, planning, naming and finishing the UML diagrams. This is strong way to deal with create UML diagrams to save the time and spending plan of both the customer and structure analyst.

**Keywords:** Natural language preparing, UML outline, Use case diagram, Data flow graph, Software Requirement Specification.

## 1. Introduction

In the cutting edge age the looks and style of programming structuring has been totally changed. By the time each movement of programming structuring takes after the rules of Object Oriented arrangement plans. Starting late, there is no item which offers organizations to draw UML plots all the more profitably beside Rational Rose, Smart Draw etc and there is more likely that these are sensibly incredible programming. Generally, the structure master needs to finish a lot of work for inferring model justification and perceiving the customer necessities for UML charts by using widespread CASE instruments. Thus, there is extensive of wastage time in view of the dull thought of the available CASE gadgets for the required circumstance. These days everybody needs a quick and strong organization. So it was required that there should be a kind of watchful programming for delivering UML based documentation to save time and spending plan of both the customer and structure master. With a particular true objective to decide each and every such issue, we require programming, which energizes the two customers and programming engineers. The purpose of this paper is to demonstrate the usage

of NLP and territory cosmology techniques for the extraction of UML diagrams from easygoing basic language necessities by executing a modeling device that uses the predetermined frameworks.

## 2. Summary of issue

The issue especially tended to in this investigation is fundamentally related to the item examination and blueprint time of the item progression process. Hardly any years back data flow diagrams (DFD) were being used to symbolize the surge of data and address the customer's necessities. However, in current age, UML is used to model and guide the customer prerequisites, which is broader and genuine way to deal with depiction and it is significant for the later periods of programming enhancement. The item in the present market which give this office are just paint like contraptions as Visual UML, GD Pro, Shrewd Draw, Rational Rose et cetera. All of them have dull nature. To use the extensively over-trouble interface of these CASE instruments is a vexing issue. The system of making the UML charts through these item planning device is to a great degree troublesome and this is a smidgen more redundant and long methodology to perform. Thus, it was necessitated that any unmistakable individual included obligatory in programming enhancement may get his required yield with most extraordinary precision in minimum time

## 3. Related Solution

The value of the coordinated research was space specific yet it is very well done and may be overhauled adequately later on according to the requirements. Current created structure merge the limit of mapping customer essentials in the wake of examining the given requirements in plain substance and speak to in UML and Class Diagram, Action Diagram, Sequence Diagram, Use Case Diagram and Component Diagram.

## 4. Natural language processing

In the NLP we give any info any dialect and NLP convert that information which dialect are machine comprehend. The given info give the input in their very own machine language and NLP change that the yield dialect in our understanding dialect. The perception and multi-perspective getting ready of

the ordinary vernaculars that are furthermore called as "talk languages", is extremely one of the conflicts of more critical excitement for the artificial intellectual competence field. The symbol language are flighty and disproportionate. Usually, symbol language rely upon un-formal accentuations. There are the graphical and sociological components which affect the acts of basic languages. There are obscure game plan of words what's more, they in like manner change and contrast area to zone and time to time. Because of these assortments and anomalies, the symbol vernaculars have different flavors as English language has the larger section dozen esteemed flavors everywhere throughout the world. These flavors have unmistakable accents, set of vocabularies and phonological perspectives. These premonition moreover, debilitating mistakes and abnormalities in typical languages make it a troublesome task to process them when diverged from the formal dialects. Amid the time spent exploring and understanding the customary languages, diverse issues are for the most part looked by the experts. The issues related with the more vital nature of the ordinary language are verb's conjugation, accentuation, lexical abundance, issue of dubiousness. From this game plan the issue which ever causes more difficulties is issue of unclerness. Ambiguity could be easily disentangled at the sentence structure and semantic level by using a sound and generous run based system.

### 5. Component of NLP

The two noteworthy segments of Natural Language Processing are,

- Natural Language Understanding (NLU)
- Natural Language Generation (NLG)

#### A. Natural language understanding (NLU)

Here the discourse input gets changed into helpful portrayal with the end goal to break down the different parts of language. A natural dialect can be extremely equivocal (distinctive importance of same sentence)

#### B. Lexical ambiguity

The issue of words made or explained not adequately is disposed of in this circumstance. These sorts of goofs are only achievable through the examination with the explanations contained in a word reference. Lexical vulnerability is made when an equivalent word expect distinctive ramifications. For this circumstance that unclerness is created from the truth that which suggestions will be participated in which circumstance. For example we consider the "ice" modifier in the going with sentences: "That room is ice." "That individual is cool.

#### 1) Syntactical ambiguity

Language structure examination is performed on world level to perceive the word classification. The syntactic examination of the projects would need to be in a situation to segregate subject, verbs, objects and different supplements. It is minimal complex method. "Mario eats the apple." In this case, the genuine implications are that "Mario eats apple" yet vagueness

can be affirmed if this sentence is imagined as "the apple eats Mario"

#### C. Semantic ambiguity

To separate an articulation from the semantic viewpoint expects to give it a centrality. This should give you a chance to grasp we connected to a vital point. Semantic ambiguities are most customary due to the route that all things considered a PC isn't in a circumstance to perceive the authentic conditions. "The auto hit the post while it was moving." Every one of us would surely decode the articulation like "The auto, while moving, hit the post.¶, while nobody would be imagined to credit to the sentence the connoted "the auto hit the post while the pole was moving ".

#### D. Pragmatic ambiguity

Calm disapproved of ambiguities imagined when the correspondence happens between two individuals who don't have a similar setting. As following outline: "I will connect to the air terminal at 12 o'clock." In this representation, if the subject individual has a place with a different territory, the suggestions can be totally changed.

#### E. Disclosure ambiguity

Here the importance of the sentence is confirmed with sentence before it.

#### F. Natural language generation (NLG)

##### 1) Text planning

It incorporates the extricating learning from information base.

##### 2) Sentence planning

This incorporates choice of right words and shaping sentence which pursue the language structure.

##### 3) Text realization

Mapping the arranged sentence into the real world.

### 6. Architecture of designed system

The laid out UMLG system has ability to attract UML traces the wake of scrutinizing the substance circumstance gave by the customer. This structure pulls in diagrams four modules: Text input getting, content understanding, data extraction. Following figure demonstrate the design of UML is given beneath.

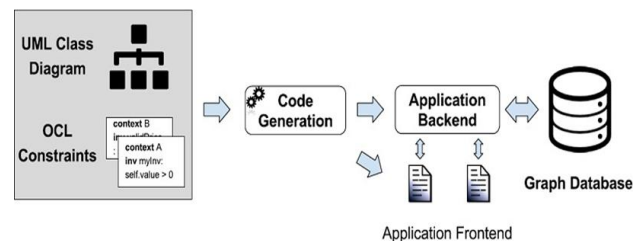


Fig. 1. Text input procurement

This module gets input content circumstance. Customer gives the business circumstance in from of segments of the substance. These modules scrutinize the data message in the

casing characters and make the words by connecting the information characters. This module is the execution of the lexical stage. Word references and tokens are created in this module.

#### A. Text understanding

This module examines the commitment from module as words. These words are requested into various classes as verbs, helping verbs, things, pronouns, modifiers, social words, conjunctions, et cetera.

#### B. Knowledge extraction

This module, isolates unmistakable dissents and classes and their specific properties on the bases of the data given by the previous module. Things are symbolized as classes and questions and their related properties are named as characteristics.

#### C. UML outline age

This is the last module, which finally utilizes UML pictures additionally, draws distinctive UML charts by solidifying available pictures according to the information isolated of the past module. As disengaged circumstance will be given for various diagrams as classes, progression, activity and use cases plots, so the distinctive limits are completed for specific.

#### D. Used methodology

Ordinary normal vernacular getting ready based systems customer run based structures. Authorities are another way to deal with location this issue. In the examination, a control based figuring has been used which has enthusiastic ability to scrutinize, fathom and separate the pined for information. As an issue of first significance fundamental parts of the vernacular sentence structure are evacuated as verbs, things, modifiers, et cetera then dependent on this removed information also getting ready is performed. In etymological terms, verbs much of the time demonstrate exercises, and thing phrases the articles that appreciate the action. Everything articulation's at that point part demonstrates how the dissent appreciates the movement.

As in the going with case:

"Robbie hit a ball with a racket."

A framework that sees such a sentence must discover, to the point that Role is the master since he plays out the movement of hitting, that the ball as the topical challenge since it is the dissent hit, and that the racket is an instrument since it is the mechanical assembly with which hitting is finished. In this way, sentence examination requires, mostly, the reactions to these exercises: The amount of topical parts gotten a handle on by various theories vacillates likely. A couple of individuals use about around six topical parts. Others use more conditions the same number of. The right number does not have any kind of effect much, as long as they will adequately wonderful to reveal ordinary restrictions on how verbs and topical precedents shape sentences.

- *Agent*: The master influences the action to occur as in

"Robbie hit the ball," Robbie is administrator who plays out the errand. In any case, in this delineation an inert sentence, the administrator moreover may appear in a prepositional "The ball was hit by Robbie."

- *Co-agent*: The word with may display a join articulation that serves an assistant in the key administrator. They two do the movement together "Robbie played tennis with Susie."
- *Beneficiary*: The beneficiary is the person for whom an action has bumble bee performed: "Robbie acquired the balls for Suzie." In this sentence Suzie is beneficiary.
- *Thematic object*: The topical inquiry is the issue the sentence is to a great degree about—commonly the inquiry, encountering a change. As often as possible the topical challenge is the equivalent a-s the syntactic direct dissent, as "Robbie hit the ball." Here the ball is topical inquiry.
- *Conveyance*: The development is something in which or on which voyages: "Robbie constantly passes through get ready."
- *Trajectory*: Motion from source to objective takes put over at course. ID contrast to the following part possible results, a couple of social words can serve to introduce bearing thing *phrases*: "Robbie and Suzie went in through the front gateway: he passed on her over the edge."
- *Location*: The territory is the place a movement occurs. As toward the path section, "a couple of social words round portion possible, which passes on infers despite filling in as a banner that a territory thing phrase is "Robbie and Suzie considered in the library, at a work zone, by the divider, a photograph, near the passage."
- *Time*: Time shows when an action occurs. Social words such as at, beforehand, and after present thing phrases filling in as time part fill "Robbie and Susie left before twelve."
- *Duration*: Duration decides to what degree a move makes. Social word, for instance, for show term. "Robbie and Susie kept running for a hour."

## 7. Conclusion and future scope

This examination is about the dynamic age of the UML diagrams by scrutinizing and separating the given circumstance in English vernacular given by the customer. The created structure can find the classes and questions and their properties and errors using a phony information framework, for instance, normal vernacular getting ready. By then the UML diagrams, for instance, Activity diagram. Sequence diagram. Component diagram. Use Case diagram. Would be drawn. The precision of the item is typical up to around 80% with the commitment of the item assemble gave that he has taken after the pre-necessities of the item to set up the data circumstance. The

given circumstance should be done and written in clear and cure English. Under the degree of our endeavor, programming will play out an aggregate examination of the circumstance to find the classes, their qualities and exercises.

It will likewise draw the accompanying graphs.

- Class Diagrams
- Activity graphs
- Use-Case Diagrams
- Sequence Diagrams
- Component Diagrams

The made system for delivering UML diagrams was started with the focuses that there should be an item which can peruse the circumstance given in English vernacular what's more, can draw the extensive variety of the UML diagrams, for instance, Class layout, activity diagram, progression diagram, use case plot, part diagram, association outline. Nevertheless, last two of them part chart, association layout are so far immaculate. There is similarly some edge of changes in the figuring for delivering starting four creates Class diagram, activity diagram, course of action layout, use case chart. Current accuracy of making diagrams is around 80% to 85%. It may be overhauled up to 95% by upgrading the figuring and inciting the limit of learning.

## References

- [1] Thomas C. Rindflesch (1999) Natural language processing.
- [2] Ruth Malan and Dana Bredemeyer, (2001) —Functional Requirements and Use Cases.
- [3] L. R. Tang and R. J. Mooney, (2001), —Using Multiple Clause Constructors in Inductive Logic Programming for Semantic Parsing, Proceedings of the 12th European Conference on Machine Learning (ECML- 2001), Freiburg, Germany, pages 466–477.
- [4] Woo, H., & Robinson, W. (2002). —A light-weight approach to the reuse of use-cases specifications, In Proceedings of the 5th annual conference of the Southern Association for Information Systems (pp. 330- 336) Savannah: SAIS.
- [5] Imran Sarwar Bajwa, M. Abbas Choudhary, (2006), “A Rule Based System for Speech Language Context Understanding,” Journal of Donghua University, (English Edition) 23 (6), pp. 39-42.
- [6] M. Ibrahim and R. Ahmad, "Class Diagram Extraction from Textual Requirements Using Natural Language Processing (NLP) Techniques," 2010 *Second International Conference on Computer Research and Development*, Kuala Lumpur, 2010, pp. 200-204.
- [7] D. K. Deeptimahanti and M. A. Babar, "An Automated Tool for Generating UML Models from Natural Language Requirements," 2009 *IEEE/ACM International Conference on Automated Software Engineering*, Auckland, 2009, pp. 680-682.
- [8] Farid Meziane, Nikos Athanasakis, Sophia Ananiadou, 2007, Generating Natural Language specifications from UML class diagrams, Springer-Verlag London Limited 2007.
- [9] Gobinda G. Chowdhury, 2001, Natural Language Processing.
- [10] L. Mich, NL-OOPs: —From Natural Language to Object Oriented Using the Natural Language Processing System LOLITA, Natural Language Engineering, 2(2), 1996.