SURVEY OF NLP APPLICATIONS AND TOOLS

[1]Shashank Girepunje,[2]Awantika Singh,[3]Koena Ghosh

[1]Assistant Professor, [2]Assistant Professor, [3]Assistant Professor
[1]Computer Science Department,
[1]Kalinga University, Raipur, India

,Computer Science Department,
Kalinga University,
New Raipur, India.

Abstract— Natural language processing (NLP) has as of late increased a lot of consideration for speaking to and investigating human language computationally in the field of Artificial Intelligence. It has spread its applications in different fields, for example, machine interpretation, email spam detection, data extraction, summarization, clinical, and question noting and so on. The paper explains all the aspects of NLP trailed by introducing the history and the different utilizations of NLP, current research trends and introduction of tools required for processing.

Index Terms: **- Natural Language processing, Artificial Intelligence.**

I. INTRODUCTION

Artificial Intelligence (AI) is a part of software engineering that manages how machines can discover answers for complex issues or problem. By and large, 'intelligence' might be viewed as the capacity to procure information and abilities and a few analysts characterize AI as "the examination and plan of savvy operators." John McCarthy, who authored the term AI in 1955, characterized it as "the science and building of making astute machines." AI is as of now utilized in a wide scope of fields, including clinical diagnosis, stock exchanging, robotics & many others.

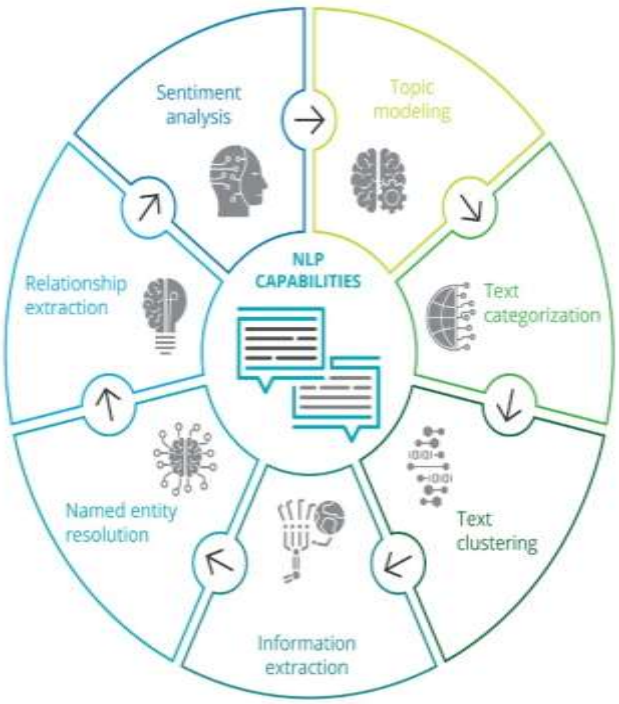

Fig 1.1: Natural language Processing & its applications

Natural Language Processing (NLP) is a tract of Artificial Intelligence and Linguistics, gave to cause system to comprehend the sentences or words written in human languages. Natural Language Processing (NLP) appeared to facilitate the user's work and to fulfill the desire to speak with the computer in regular language[1]. Since all the clients may not be knowledgeable in machine explicit language, NLP provides food those clients who need more time to learn new dialects or get flawlessness in it. NLP is a concept in which computers analyze, understand, and derive meaning from human language in a meaningful manner. By utilizing NLP, developers can organize the data and prepare structure for knowledge or information to perform tasks such as auto summarization, machine translation, sentiment analysis, speech recognition, and topic segmentation. These types of activities' are widely used in industries for many purposes.

Natural Language Processing (NLP) deals with actual text data. The text data is organized into machine format by text processing [2]. Artificial Intelligence (AI) specific tools use this information provided by the text processing and applies a lot of math's to determine whether something is positive or negative.

A few strategies exist to decide a creator's view on a subject from natural language data. Some type of AI approach is utilized and which has fluctuating level of adequacy. One of the kinds of normal language handling is assessment mining which manages following the temperament of the individuals with respect to a specific item or point. This product gives programmed extraction of assessments, feelings and conclusions in content and furthermore tracks mentalities and emotions on the web. Individuals express their perspectives by composing blog entries, remarks, audits and tweets pretty much a wide range of various points. Following items and brands and afterward deciding if they are seen decidedly or contrarily should be possible utilizing web. The supposition mining has somewhat various undertakings and numerous names, for example slant investigation, supposition extraction, estimation mining, subjectivity examination, influence investigation, feeling investigation, survey mining, and so on. Regular language preparing (NLP) is a field of software engineering, man-made consciousness, and computational phonetics worried about the cooperation's among PCs and human (common) dialects. Accordingly, NLP is identified with the region of human–PC cooperation. Numerous difficulties in NLP include: characteristic language understanding, empowering PCs to get importance from human or natural language info; and others include natural language generation.

In this paper we explore the various aspects of natural language processing. Some usage of NLP in various fields, research trends and description of NLP tools are also given.

II. NATURAL LANGUAGE PROCESSING

The study of natural language preparing by and large began during the 1950s, albeit some work can be found from before periods. In 1950, Alan Turing distributed an article titled "Computing Machinery and Intelligence" which proposed what is currently called the Turing test as a basis of insight.

Moreover, up to the 1980s, most NLP frameworks depended on complex arrangements of manually written standards. Beginning in the late 1980s, be that as it may, there was a revolution in NLP with the presentation of AI (ML) calculations for language processing. This was because of both the consistent increment in computational force (see Moore's law) and the slow decreasing of the predominance of Chomsky a speculations of phonetics (for example transformational syntax), whose hypothetical underpinnings debilitated the kind of corpus phonetics that underlies the ML way to deal with language handling. Probably the soonest utilized ML algorithms, for example, choice trees, delivered frameworks of hard in the event that rules like existing hand written rules.

Since the alleged "statistical learning" in the mid-1990s, much common language handling research has depended vigorously on ML and as of now depending considerably more on ML in view of the huge advancement the now-celebrated subfield of ML called Deep Learning (DL). During the 2010s, deep learning (DL) dominated and profound neural system style ML strategies got across the board in natural language processing, because of results indicating that such methods can accomplish state-of-the-art results brings about more NLP tasks, for example, language modeling, parsing and numerous others.

Natural Language Processing includes following phases of handling to be specific, lexical (structure) investigation, parsing, semantic analysis, disclosure integration, and program analysis. Some notable application territories of NLP are Speech Recognition, Optical Character Recognition (OCR), Machine Translation, and Chatbots. As of late, Machine Learning algorithms are utilized to process Natural Language input by considering a huge number of instances of content — words, sentences, and passages — composed by people. By examining these examples, preparing calculations increase a comprehension of the "unique situation" of human discourse, composing, and different methods of correspondence. The AI and profound learning algorithms are generally is utilized to create structures for NLP and proficiently perform normal NLP talks.

Liddy (1998) and Feldman (1999) propose that [1] so as to comprehend natural language, it is essential to have the option to recognize among the seven related degrees of examination that individuals use to extracts importance from content or natural languages, as given underneath:

- Phonological level
- Discourse level
- Lexical level
- Semantic level
- Syntactic level
- Pragmatic level
- Morphological level

III. APPLICATIONS OF NLP

3.1. Sentiment Analysis

Major application of NLP is in sentiment analysis or opinion mining. As innovation builds step by step, tremendous volumes of information additionally increment [6]. Longer than 10 years back, a Gigabyte information produced every day, except starting now and into the foreseeable future it were in a moment or two. Time makes a huge difference; the world is traveling through

creating side to an ever increasing extent. Various divisions create information inconceivably, for example, Institutions, Companies, Social media destinations, Hospitals, Companies, Governments, and so on., among these 99 out of 100 uses Social media locales everywhere throughout the world. Different online life destinations like Twitter, YouTube, Facebook, Whatsapp, Instagram, Snap Chat, and so on., Sentiment Analysis is otherwise called Opinion mining. It is the strategy for discovering the tone behind words. It assists with increasing a comprehension of the mentalities, suppositions and feelings communicated inside an online notice. Assessments, input and evaluates gave by web clients show perspectives and notions towards explicit points, items, or services. It is hard to peruse and comprehend the huge volume of information.

In this way, web based life slant examination assumes a fundamental job in tackling and settling on better choices. Different internet based life destinations that are managing are Twitter, YouTube, and Online news. From numerous locales, gathering various informational indexes and doing changes and getting the outcomes characterizing the passionate status. Assumption examination goes under Natural Language Processing. The notion is a disposition or a feeling or an inclination. It predominantly learns about clients feelings towards specific items or people or services.

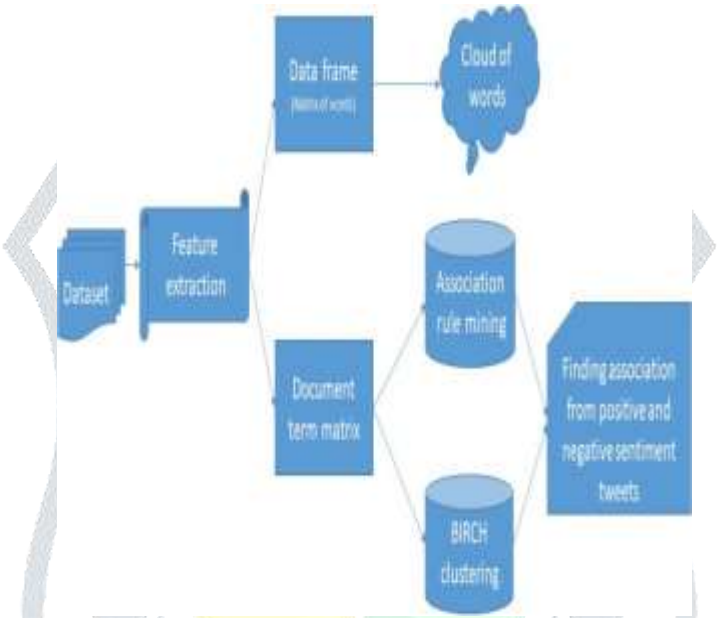

Fig 3.1.1: Working of Sentiment Analysis

With the assistance of the web, individuals ready to post their perspectives, comments and their contend through social media sites, for example, gatherings, small scale web journals, web based systems administration locales. The internet based life destinations that are managing here are Twitter, YouTube, and Online news. Each site has its upsides and downsides over 1.5 million tweets created every day by clients. Each acclaimed individual like political pioneers, entertainers, on-screen characters, organization holders, Institutions, utilizes Twitter a ton to post their ongoing exercises and perspectives. To know the enthusiastic status of tweets, Twitter information removed and do a few changes and dissect them and classify them into the positive, negative, nonpartisan, effectively positive, solid negative, feebly positive, powerless picture. Any twitter client can remove the open information from Twitter utilizing Application Programming Interface (API). Twitter at present has principally three distinct API's, for example, the REST API, the Search API and the Streaming API. Engineers can ready to accumulate status information and client data by utilizing REST API. While the Search API permits designers to inquiry explicit Twitter content. Though, the Streaming API ready to gather the real-time information on Twitter.

To achieve better results, API's can be mixed. Textual data just as accounts can be isolated and researched. YouTube chronicles are notable video goals for reviews. Social affair video comments from different customers and can be rotted them into positive, negative and fair-minded. By this, it is definitely not hard to portray that is video is defended, notwithstanding all the difficulty or not. Anything in this overall population is news. Before the nineteenth century, typescripts, papers which were used to consider the news. Regardless, directly, advancement grows, news can be scrutinized, knew through the Internet, which is in any case called Online-News. Distinctive reports related to different regions can be exhorted, inspected and requested them into positive, negative and neutral reliant on the furthest point regard.

3.2. Machine Translation

Automatic or machine translation is maybe one of the most challenging AI problmes given the smoothness of natural language. Traditionally, rule-based frameworks were utilized for this assignment, which were supplanted during the 1990s with statistic techniques. All the more as of late, profound neural system models accomplish best in class brings about a field that is relevantly named neural machine translation. Machine translation (MT) alludes to completely computerized programming that can

make an interpretation of source content into target dialects. People may utilize MT to assist them with rendering content and discourse into another language, or the MT programming may work without human intercession.

As the greater part of the world is on the web, the undertaking of making information open and accessible to everything is a test. Significant test in making information available is the language hindrance. There are large number of dialects with various sentence structure and language structure. Machine Translation is by and large deciphering phrases starting with one language then onto the next with the assistance of a measurable motor like Google Translate. The test with machine translation innovations isn't legitimately deciphering words yet keeping the importance of sentences unblemished alongside language and tenses. The factual AI accumulates the same number of information as they can see that appears as equal between two dialects and they crunch their information to discover the probability that something in Language A compares to something in Language B. With respect to Google, in September 2016, reported another machine translation framework dependent on Artificial neural systems and Deep learning . As of late, different techniques have been proposed to consequently assess machine translation quality by contrasting theory interpretations and reference interpretations.

MT tools are regularly used to decipher immense measures of data including a great many words that couldn't in any way, shape or form be interpreted the conventional way. The nature of MT yield can differ extensively; MT frameworks require "preparing" in the ideal space and language pair to build quality. There are more than 100 suppliers of MT advancements. Some of them are carefully MT engineers, others are interpretation firms and IT monsters. A portion of the names are demonstrated as follows:

- Google Translate
- Yandex Translate
- IBM - Watson Language Translator
- Language Engineering Company

There are three main approaches to machine translation:

- First-age rule-based (RbMT) frameworks depend on endless calculations dependent on the sentence structure, linguistic structure, and style of a language.
- Statistical systems (SMT) showed up with search and enormous information. With loads of equal writings opening up, SMT designers figured out how to design coordinate reference writings to discover interpretations that are measurably destined to be reasonable. These frameworks train quicker than RbMT, if there is sufficient existing language material to reference.
- Neural MT (NMT) utilizes AI innovation to show programming how to deliver the best outcome. This procedure devours a lot of preparing power, and that is the reason it's regularly run on illustrations units of CPUs. NMT began picking up perceivability in 2016. Numerous MT suppliers are currently changing to this innovation.

3.3. Text Categorization

Categorization systems inputs a huge progression of information like authority records, military setback reports, advertise information, newswires and so on and allot them to predefined classes or files. A few organizations have been utilizing order frameworks to classify inconvenience tickets or protest demands and directing to the fitting work areas. Another use of content classification is email spam filters. Spam filters are getting significant as the main line of resistance against the undesirable messages. Bogus negative and bogus positive issues of spam filters are at the core of NLP innovation, its brought down to the test of removing importance from strings of content. A sifting arrangement that is applied to an email framework utilizes a lot of conventions to figure out which of the approaching messages spam is and which are most certainly not. There are a few kinds of spam filters accessible.

Content filters: Review the content within the message to determine whether it is a spam or not.

Header filters: Review the email header looking for fake information.

General Blacklist filters: Stopes all emails from blacklisted recipients.

Rules Based Filters: It uses user-defined criteria. Such as stopping mails from specific person or stopping mail including a specific word.

Permission Filters: Require anyone sending a message to be pre-approved by the recipient.

Challenge Response Filters: Requires anyone sending a message to enter a code in order to gain permission to send email.

3.4. Information Extraction

Information Extraction (IE) alludes to the utilization of computational strategies to distinguish pertinent snippets of data in archive created for human use and convert this data into a portrayal appropriate for PC based capacity, handling, and recovery. The contribution to IE framework is an assortment of records (email, pages, news gatherings, news stories, business reports, investigate papers, web journals, resumes, recommendations, etc) and yield is a portrayal of the important data from the source archive as indicated by some particular measures. The capacity of individuals to viably utilize this immense measure of data is

low as this errand is very exhausting, dull and devours part of time. This blast of data and requirement for increasingly complex and productive data dealing with apparatuses featured the need of data extraction and recovery innovation (Neil et al., 1998). Information Extraction advancements serve to proficiently and successfully dissect free content and to find important and pertinent information from it as organized data. Henceforth, the objective of IE is to extricate striking realities about pre-indicated sorts of occasions, elements, or connections, so as to fabricate progressively important, rich portrayals of their semantic substance, which can be utilized to populate databases that give increasingly organized information.

IV. Tools for NLP

Previously, no one but specialists could be a piece of natural language processing ventures that necessary unrivaled information on arithmetic, AI, and linguistics. Presently, engineers can utilize instant apparatuses that streamline content preprocessing with the goal that they can focus on building AI models. There are numerous instruments and libraries made to take care of NLP issues. Peruse on to learn progressively 8 astonishing Python Natural Language Processing libraries that have throughout the years helped us convey quality tasks to our customers.

Natural Language Toolkit (NLTK)

Link: https://www.nltk.org/

NLTK is a basic library underpins errands, for example, classification, stemming, labeling, parsing, semantic thinking, and tokenization in Python. It's essentially your primary apparatus for regular language handling and AI. Today it fills in as an instructive establishment for Python designers who are plunging their toes in this field (and AI). The library was created by Steven Bird and Edward Loper at the University of Pennsylvania and assumed a key job in advancement NLP explore. Numerous colleges around the world presently use NLTK, Python libraries, and different devices in their courses. This library is really flexible, however we should concede that it's likewise very hard to use for Natural Language Processing with Python. NLTK can be somewhat moderate and doesn't coordinate the requests of snappy paced creation use.

TextBlob

Link: https://textblob.readthedocs.io/en/dev/

TextBlob is an unquestionable requirement for designers who are beginning their excursion with NLP in Python and need to take advantage of their first experience with NLTK. It fundamentally gives tenderfoots a simple interface to assist them with learning most essential NLP undertakings like sentiment analysis, pos-labeling, or thing phrase extraction.

CoreNLP

Link: https://stanfordnlp.github.io/CoreNLP/

This library was created at Stanford University and it's written in Java. All things considered, it's furnished with wrappers for some, various dialects, including Python. That is the reason it tends to be helpful for engineers keen on taking a stab at normal language preparing in Python. What is the best favorable position of CoreNLP? The library is truly quick and functions admirably in item improvement conditions,. Besides, some of CoreNLP parts can be incorporated with NLTK which will undoubtedly support the productivity of the last mentioned.

Gensim

Link: https://github.com/RaRe-Technologies/gensim

Gensim is a Python library that spends significant time in recognizing semantic comparability between two records through vector space demonstrating and subject displaying toolbox. It can deal with huge content corpora with the assistance of proficiency information gushing and gradual calculations, which is beyond what we can say about different bundles that lone objective bunch and in-memory handling. What we love about it is its mind blowing memory use streamlining and handling speed. These were accomplished with the assistance of another Python library, NumPy. The instrument's vector space demonstrating abilities are likewise first class.

spaCy

Link: https://spacy.io/

spaCy is a generally youthful library was intended for creation use. That is the reason it's a great deal more open than other Python NLP libraries like NLTK. spaCy offers the quickest syntactic parser accessible available today. In addition, since the toolbox is written in Cython, it's likewise extremely rapid and productive.

Rapid Miner

Link: https://rapidminer.com/

RapidMiner is an information science programming stage created by the organization of a similar name that gives an incorporated domain to information arrangement, AI, profound learning, content mining, and prescient investigation. It is utilized for business and business applications just as for explore, instruction, preparing, quick prototyping, and application advancement and supports all means of the AI procedure including information arrangement, results perception, model approval and optimization. RapidMiner is created on an open center model. The RapidMiner Studio Free Edition, which is constrained to 1 legitimate processor and 10,000 information columns, is accessible under the AGPL license, by relying upon different non-open source segments. Business evaluating begins at $5,000 and is accessible from the engineer.

V. CONCLUSION

In this paper, we have explored all the aspects of Natural language processing. The introduction part gave the decent introduction of NLP and its history. Then various applications of NLP has also discussed with proper description. NLP is very vast field to explore but we have shown essential applications in software market. There are some tools which are useful for implementations and research oriented work.

REFERENCES

[1] Diksha Khurana," Natural Language Processing: State of The Art, Current Trends and Challenges",2017.

[2] R. Kibble," Introduction of Natural language", subject guide by university of london,2013.

[3] Sudhir K Mishra,"Artificial intelligence and Natural language processing", book, 2018.

[4] Yasir Ali Solangi,"Review on Natural language processing and its toolkits for opinion mining & sentiment Analysis", International conference on technologies & applied science, 2018.

[5] S. Sun, C. Luo, and J. Chen, "A review of natural language processing techniques for opinion mining systems," *Inf. Fusion*, vol. 36, pp. 10–25, 2017.

[6] Y. Kai, Y. Cai, H. Dongping, J. Li, Z. Zhou, and X. Lei, "An effective hybrid model for opinion mining and sentiment analysis," *2017 IEEE Int. Conf. Big Data Smart Comput. BigComp 2017*, pp. 465–466, 2017.

[7] Feldman, S. (1999). NLP Meets the Jabberwocky: Natural Language Processing in Information Retrieval. ONLINE-WESTON THEN WILTON-, 23, 62-73.

[8] "Natural Language Processing." Natural Language Processing RSS. N.p., n.d. Web. 25 Mar. 2017