International Journal of Advanced Research and Development
ISSN: 2455-4030; Impact Factor: RJIF 5.24
Received: 01-06-2020; Accepted: 15-06-2020; Published: 25-07-2020
www.advancedjournal.com
Volume 5; Issue 4; 2020; Page No. 01-05

Natural language processing and its significance at higher education

T Kondala Rao[1], Alankrita Saxena[2]
[1, 2] Faculty of Education, Dayalbagh Educational Institute (Deemed University), Agra, Uttar Pradesh, India

Abstract
The paper aims to highlight the significance of natural language processing at higher education and how does it influence students' learning by making the technology accessible to all across the world. NLP is involved in each and every technology that students take into the consideration by studying, learning, discovering new information and exploring the different fields of knowledge. Natural language processing usually shortened as NLP and is a branch of Artificial Intelligence that deals with interaction between computers and humans. In the field of education NLP is getting upgraded and prominent day by day as students who are pursuing higher education use NLP for text analysis, checking up of plagiarism and summarization of the whole document. In this modern era where whole world is running business, learning deals and shopping online, E- learning applications and tools provide assistance to the learner to improve their education. Teachers are also getting friendly with technology and enhancing students' skills in the field of Artificial Intelligence. NLP is widely integrated with the large number of educational content such as science, linguistics, e- learning, evaluation system and contributes resulting positive outcomes in other educational settings such as schools, higher education system and universities.

Keywords: natural language processing, natural language generation, natural language understanding, deep learning, machine learning

Introduction
The theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages. Artificial intelligence (AI) is an area of computer science that emphasizes the creation of intelligent machines that work and reacts like humans. Some of the activities computers with artificial intelligence are designed for include Speech recognition, Learning, Planning, Problem solving. In computer science, artificial intelligence (AI), sometimes called machine intelligence, is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans. Natural Language Processing, usually shortened as NLP, is a branch of artificial intelligence that deals with the interaction between computers and humans using the natural language. The ultimate objective of NLP is to read, decipher, understand, and make sense of the human languages in a manner that is valuable. Most NLP techniques rely on machine learning to derive meaning from human languages.

Machine Learning
Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

Deep Learning
Deep learning is an artificial intelligence function that imitates the workings of the human brain in processing data and creating patterns for use in decision making.

Deep learning is a subset of machine learning in artificial intelligence (AI) that has networks capable of learning unsupervised from data that is unstructured or unlabeled. Also known as deep neural learning or deep neural network.

Meaning of Natural Language Processing
Natural Language Processing (NLP) is a tract of Artificial Intelligence and Linguistics, devoted to make computers understand the statements or words written in human languages. Natural language processing came into existence to ease the user's work and to satisfy the wish to communicate with the computer in natural language. Since all the users may not be well-versed in machine specific language, NLP caters those users who do not have enough time to learn new languages or get perfection in it. A language can be defined as a set of rules or set of symbol. Symbol are combined and used for conveying information or broadcasting the information. Symbols are tyrannized by the Rules. Natural Language Processing basically can be classified into two parts i.e. Natural Language Understanding and Natural Language Generation which evolves the task to understand and generate the text.

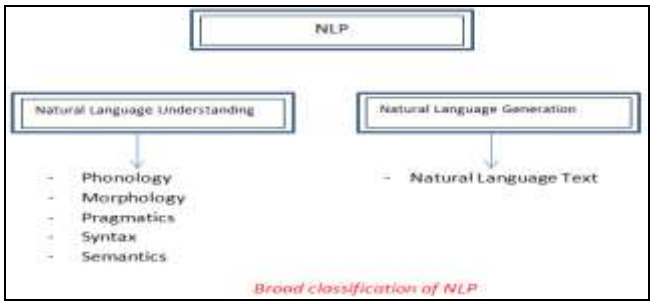

Fig 1

Brief History of NLP

1950- NLP started when Alan Turing published an article called "Machine and Intelligence."

1950- Attempts to automate translation between Russian and English

1960- The work of Chomsky and others on formal language theory and generative syntax

1990- Probabilistic and data-driven models had become quite standard

2000- A Large amount of spoken and textual data become available

Educational Implications of NLP

- NLP is very effective for providing knowledge and information to the students for the application of e-learning and NLP in understanding and also in dealing with the need of analyzing the text.
- In higher education understanding of text is based on the development of research- based analysis of the general and contextual learning.
- NLP in higher education is very useful in developing the understanding of students in natural settings and assessing the information available from various sources.
- NLP can help for academic writing, assessment purpose, writing test questions and for analysis of errors.
- NLP is useful in higher education especially in the field of research as it can help in finding plagiarism in the work.
- Finding appropriate documents on certain topics from a database of text.
- Extracting the information from messages or articles on certain topics.
- Translation of documents from one language to another.
- Helps students in summarizing texts for certain purpose (for example producing a 3 page summary of 1000 pages of government report)
- Grammar correction.
- Auto correct system that is frequently used by students
- Text analysis and sentiments analysis.

Levels of NLP

The 'levels of language' are one of the most explanatory method for representing the Natural Language processing which helps to generate the NLP text by realizing Content Planning, Sentence Planning and Surface Realization phases.

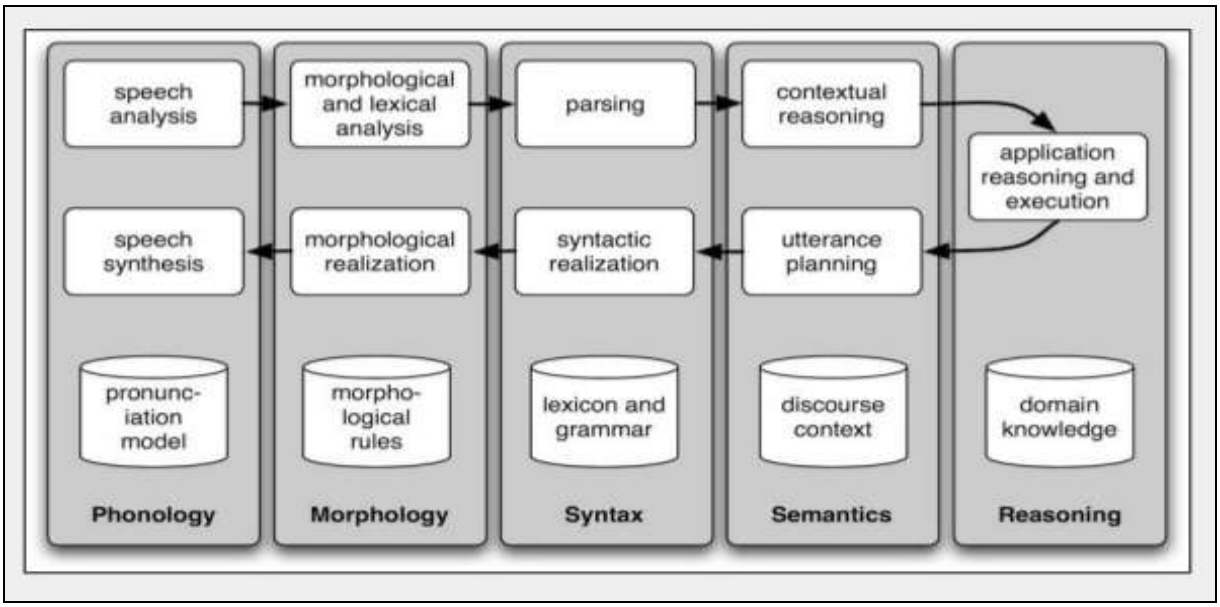

Fig 2

Morphological Analysis: The morphological level of linguistic processing deals with the study of word structures and word formation, focusing on the analysis of the individual components of words. The most important unit of morphology, defined as having the "minimal unit of meaning", is referred to as the *morpheme*.

- **Lexical Analysis**: The lexical analysis in NLP deals with the study at the level of words with respect to their lexical meaning and part-of-speech. This level of linguistic processing utilizes a language's *lexicon*, which is a collection of individual *lexemes.* A lexeme is a basic unit of lexical meaning; which is an abstract unit of morphological analysis that represents the set of forms or "senses" taken by a single morpheme.
- **Syntactic Realization**: The part-of-speech tagging output of the lexical analysis can be used at the syntactic level of linguistic processing to group words into phrase and clause brackets. Syntactic Analysis also referred to as *"parsing"*, allows the extraction of phrases which convey more meaning than just the individual words by themselves, such as in a noun phrase.
- **Semantic**: The semantic level of linguistic processing deals with the determination of what a sentence really means by relating syntactic features and disambiguating words with multiple definitions to the given context. This level entails the appropriate interpretation of the meaning of sentences, rather than the analysis at the level or individual words or phrases.
- **Discourse Integration**: The discourse level of linguistic processing deals with the analysis of structure and meaning of text beyond a single sentence, making connections between words and sentences.
- **Pragmatic Analysis**: The pragmatic level of linguistic processing deals with the use of real-world knowledge

and understanding how this impacts the meaning of what is being communicated. By analyzing the contextual dimension of the documents and queries, a more detailed representation is derived.

Components of Natural Language Processing

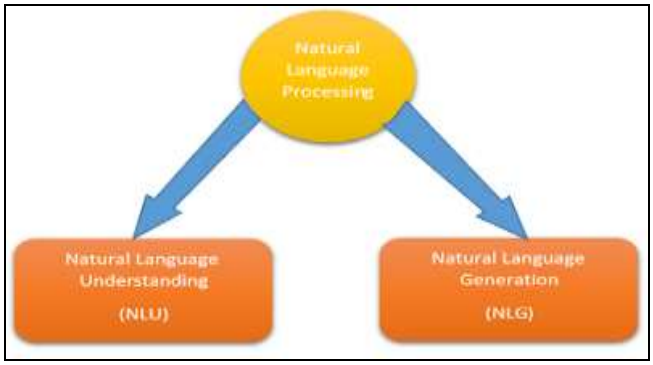

Fig 3

Natural Language Understanding

Natural-language understanding (NLU) or natural-language interpretation (NLI) is a subtopic of natural-language processing in artificial intelligence that deals with machine reading comprehension. NLU is the post-processing of text, after the use of NLP algorithms (identifying parts-of-speech, etc.), that utilizes context from recognition devices (automatic speech recognition) vision recognition, last conversation, misrecognized words from ASR, personalized profiles, microphone proximity etc.), in all of its forms, to discern meaning of fragmented and run-on sentences to execute an intent from typically voice commands. NLU has ontology around the particular product vertical that is used to figure out the probability of some intent. An NLU has a defined list of known intents that derives the message payload from designated contextual information recognition sources. The NLU will provide back multiple message outputs to separate services (software) or resources (hardware) from a single derived intent (response to voice command initiator with visual sentence (shown or spoken) and transformed voice command message to different output messages to be consumed.

- Mapping the given input in the natural language into a useful representation.
- Different level of analysis required:
- Morphological analysis
- Syntactic analysis
- Semantic analysis
- Discourse analysis

Natural Language Generation

Natural language generation (NLG) is the use of artificial intelligence (AI) programming to produce written or spoken narrative from a dataset. NLG is related to computational linguistics, natural language processing (NLP) and natural language understanding (NLU), the areas of AI concerned with human-to-machine and machine-to-human interaction. Natural language generation (NLG) is a particular AI-complete task that involves generating language from non-language inputs. Some experts might refer to a natural language generation application as a "translator" of text or other informational formats into spoken language.

- Producing output in the natural language from some internal representation.
- Different level of synthesis required:
- Deep planning (what to say),
- Syntactic generation

Components of NLG

Speaker and Generator

To generate a text we need to have a speaker or an application and a generator or a program that renders the application's intentions into fluent phrase relevant to the situation.

Components and Levels of Representation

The process of language generation involves the following interweaved tasks. Content selection: Information should be selected and included in the set. Depending on how this information is parsed into representational units, parts of the units may have to be removed while some others may be added by default.

Textual Organization

The information must be textually organized according the grammar, it must be ordered both sequentially and in terms of linguistic relations like modifications.

Linguistic Resources

To support the information's realization, linguistic resources must be chosen. In the end these resources will come down to choices of particular words, idioms, syntactic constructs etc. Realization: The selected and organized resources must be realized as an actual text or voice output.

Application or Speaker

This is only for maintaining the model of the situation. Here the speaker just initiates the process doesn't take part in the language generation. It stores the history, structures the content that is potentially relevant and deploys a representation of what it actually knows. All these form the situation, while selecting subset of propositions that speaker has. The only requirement is the speaker has to make sense of the situation.

Application of NLP

Machine Translation

As most of the world is online, the task of making data accessible and available to all is a challenge. Major challenge in making data accessible is the language barrier. There are multitude of languages with different sentence structure and grammar. Machine Translation is generally translating phrases from one language to another with the help of a statistical engine like Google Translate. The challenge with machine translation technologies is not directly translating words but keeping the meaning of sentences intact along with grammar and tenses. The statistical machine learning gathers as many data as they can find that seems to be parallel between two languages and they crunch their data to find the likelihood that something in Language A corresponds to something in Language.

Text Categorization

Categorization systems inputs a large flow of data like official documents, military casualty reports, market data, newswires etc. and assign them to predefined categories or

indices.

Spam Filtering

It takes the information of which words are used in a document irrespective of number of words and order. In second model, a document is generated by choosing a set of word occurrences and arranging them in any order. This model is called multi-nomial model, in addition to the Multi-variate Bernoulli model, it also captures information on how many times a word is used in a document.

Information Extraction

Information extraction is concerned with identifying phrases of interest of textual data. For many applications, extracting entities such as names, places, events, dates, times and prices is a powerful way of summarize the information relevant to a user's needs. In the case of a domain specific search engine, the automatic identification of important information can increase accuracy and efficiency of a directed search.

Summarization

Overload of information is the real thing in this digital age, and already our reach and access to knowledge and information exceeds our capacity to understand it. This trend is not slowing down, so an ability to summarize the data while keeping the meaning intact is highly required. This is important not just allowing us the ability to recognize the understand the important information for a large set of data, it is used to understand the deeper emotional meanings; For example, a company determine the general sentiment on social media and use it on their latest product offering. This application is useful as a valuable marketing asset.

Dialogue System

Perhaps the most desirable application of the future, in the systems envisioned by large providers of end user applications, Dialogue systems, which focuses on narrowly defined applications (like refrigerator or home theater systems) currently uses the phonetic and lexical levels of language. It is believed that these dialogue systems when utilizing all levels of language processing offer potential for fully automated dialog systems. (Elizabeth D. Liddy, 2001). Whether on text or via voice. This could lead to produce systems that can enable robots to interact with humans in natural languages.

Application of Natural Language Processing in Higher Education

Natural language process is an effective process to assist students in the process of scientific learning. Implementing NLP in the educational setting not only helps in developing effective language process, but it is also significant for enhancing the academic performance. The NLP techniques follow the approach of the natural process of language acquisition integrated with the scientific approach of using computer programs. Natural Language processing has various applications for educational purpose. NLP is also very effective for providing knowledge and information to the students for application of e-learning and NLP in understanding and dealing with the need of analyzing text.

Understanding of text is based on the development of research-based analysis of the general and contextual learning. Based on the research outcome, it is clear that students' output can be increased by implementing the NLP in the education. This approach is not only affective for its application in assessment, but it is also effective for writing purposes such as writing material for digital libraries, websites, and various other sources. Some of the applications of nlp are given below:

1. **Writing Mentor:** The writing mentor application is a google docs writing support add-on system. The app targets a wide range of post-secondary users, including struggling writers and English learners population enrolled in 2-4 year college. Students who are using Google Docs can install the app and use it to get feedback for text — specifically, actionable feedback about their writing related to claims and sources, topic development, coherence, and English conventions and word choice.
2. **The Language Muse Activity Palette:** It is a web-based application designed to support English Learners (ELs). Aligned with reading standards, the tool automatically generates customizable activities aimed to help ELs build the academic language skills needed for deeper reading comprehension in content areas. The language-based activities are intended to support content comprehension and language skills development through activities that afford practice with vocabulary, sentence structures, discourse and summary writing.
3. **Text Evaluator Capability:** Text evaluator is a fully-automated, web-based technology tool designed to help teachers, textbook publishers and test developers select texts for use in instruction and assessment. Text Evaluator incorporates a patented measurement approach that goes beyond traditional readability dimensions of syntactic complexity and vocabulary difficulty to address complexity variation due to cohesion, concreteness, academic orientation.
4. **Quillionz AI:** Artificial Intelligence (AI) is no longer just contained in science fiction films. It is increasingly becoming a significant part of our everyday lives. We already use tools like Siri and Amazon's Alexa are just beginning to see the possibilities of AI in education. Quillionz is the world's first intelligent tool for automatically generating questions from written content. This unique AI-driven platform has the potential to transform the way teachers create quizzes, tests, and assessments. Quillionz accomplishes this by leveraging the power and promise of artificial intelligence and machine learning algorithms. That's what makes it a game-changer. The platform creates a variety of questions based on your written content, including multiple-choice questions; fill in the blank questions, and short descriptive questions. Once your items are ready, Quillionz allows you to curate and revise them as much as you want. Choose the questions you like the most and edit them to make them perfect. From there, you can download the text file, convert it to a Word file and print your assessment to administer it to your students, or you can publish them as a QuilliQuiz, which is basically a self-assessment utility.

Steps for Quillionz

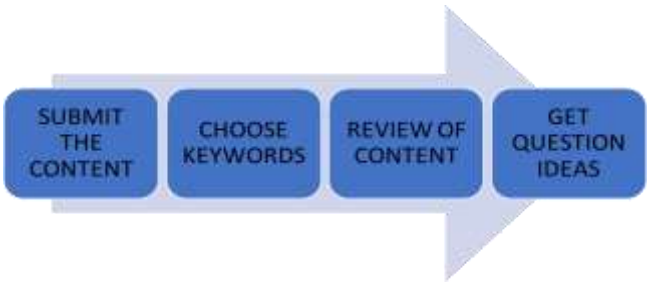

Fig 4

Ambiguities of NLP

- **Lexical Ambiguity:** Is the ambiguity of a single word. A word can be ambiguous with respect to its syntactic class. Eg: book, study. Lexical ambiguity can be resolved by Lexical category disambiguation i.e, parts-of-speech tagging. As many words may belong to more than one lexical category part-of-speech tagging is the process of assigning a part-of-speech or lexical category such as a noun, verb, pronoun, preposition, adverb, adjective etc. to each word in a sentence.
- **Syntactic Ambiguity:** The structural ambiguities were syntactic ambiguities. Structural ambiguity is of two kinds: Scope Ambiguity and Attachment Ambiguity.
- **Scope Ambiguity:** Scope ambiguity involves operators and quantifiers.
- **Attachment Ambiguity:** A sentence has attachment ambiguity if a constituent fits more than one position in a parse tree. Attachment ambiguity arises from uncertainty of attaching a phrase or clause to a part of a sentence.
- **Semantic Ambiguity:** This occurs when the meaning of the words themselves can be misinterpreted. Even after the syntax and the meanings of the individual words have been resolved, there are two ways of reading the sentence.
- **Discourse:** Discourse level processing needs a shared world or shared knowledge and the interpretation is carried out using this context. Anaphoric ambiguity comes under discourse level.

Conclusion

Natural Language Processing (NLP) is an effective approach for bringing improvement in educational setting. Implementing NLP involves initiating the process of learning through the natural acquisition in the educational systems. It is based on effective approaches for providing a solution for various problems and issues in education. Natural Language Processing provides solution in a variety of different fields associated with the social and cultural context of language learning. It is an effective approach for teachers, students, authors and educators for providing assistance for writing, analysis, and assessment procedures. Natural Language Processing is widely integrated with the large number of educational contexts such as research, science, linguistics, e-learning, evaluations system, and contributes resulting positive outcomes in other educational settings such as schools, higher education system, and universities.

References

1. Burstein J, Madnani N, Sabatini. Generating Language Activities in Real time for English Learners using language muse, 2017. Retrieved from https://www.ets.org/research/topics/as_nlp/educational_applications/
2. Khaled M. Natural language processing and its use in education, 2014. Retrieved from https://thesai.org/Downloads/Volume5No12/Paper_10-Natural_Language_Processing.pdf
3. Khurana D, Koli A, Khatter K. Natural language processing state of the art, current trends and challenges, 2016. Retrieved from https://www.researchgate.net/publication/319164243_Natural_Language_Processing_State_of_The_Art_Current_Trends_and_Challenges
4. Lynch M. Product review of Quillionz, the world's first AI- powered Question Generator, 2018. Retrieved from https://www.thetechedvocate.org/product-review-of-quillionz-the-worlds-first-ai-powered-question-generator/
5. Sethi G. Role of NLP in education, 2015. Retrieved from http://ijsae.in/ijsaeems/index.php/ijsae/article/viewFile/1143/797