IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Volume-4, Issue-7 E-ISSN: 2347-2693

A Systematic Survey of Natural Language Processing (NLP) Approaches in Different Systems

N.Vasunthira Devi[1*], Dr. R. Ponnusamy[2]

[1]*Ph.D Research Scholar, Department of Computer Science, AVC College (Autonomous) Mannampandal, Mayiladuthurai*
[2]*Professor, Department of Computer Science & Engineering, Sri Lakshmi Ammal Engineering College, Chennai*

www.ijcseonline.org

Received:12/Jun/2016 Revised:21/Jun/2016 Accepted:14/Jul/2016 Published:31/Jul/2016

***Abstract*—** Natural Language Processing and Programming Languages are both established areas in the field of Computer Science, each of them with a long research tradition. Natural Language processing Frameworks can make conceivable the use of natural language to express ideas. In education system, Natural Language Processing gives solution in a assortment of differing fields related with the social and cultural context of language learning. Natural Language Processing is broadly integrated with the expansive number of education connections such as research, science, linguistics, e-learning, evaluations system, and contributes resulting positive outcomes in other education settings such as schools, higher education system, and universities. In healthcare, NLP Frameworks decrease fetched as well as improve the quality of electronic healthcare systems. It is therefore against this foundation that this paper reviews the NLP techniques, uses in education, healthcare and their applications as well as their limitations. The reason of this relook is to overview and report the current state and the future bearings of the use of NLP innovations in different frameworks in the corporate world.

***Keywords*—** *NLP, Education, Application, Education System, Electronic Healthcare Systems, Healthcare, NLP Techniques.*

I. INTRODUCTION

Natural Language Processing (NLP) is an area of relook and application that explores how PCs can be used to understand and control natural language text or speech to do helpful things. NLP scientists aim to gather learning on how human beings understand and use language so that proper apparatuses and Frameworks can be created to make PC frameworks understand and control natural languages to perform the desired tasks. In later years, the natural language text understanding and processing innovations have also gained an increasing level of sophistication. NLP innovations are becoming extremely essential in the creation of user-friendly decision-support frameworks for everyday non-expert users, especially in the areas of learning acquisition, data recovery and language translation. The reason of this relook is to overview and report the current state and the future bearings of the use of NLP innovations and frameworks in the corporate world.

II. GOAL

The objective of natural language processing (NLP) is to assemble computational models of natural language for its investigation and generation. First, there is mechanical inspiration of building intelligent PC frameworks such as machine translation systems, natural language interfaces to databases, man-machine interfaces to PCs in general, speech understanding systems, text investigation and understanding frameworks etc. Second, there is a cognitive and etymology inspiration to gain a better insight into how individuals communicate using natural language (NL).

The objective of the Natural Language Processing (NLP) is to accomplish human-like language processing, plan and assemble programming that will analyse, understand, and create languages that individuals use naturally, so that eventually you will be able to address your PC as though you were addressing another person.

A. NLP and Education Setting

There are a number of differing powerful approaches in the NLP, which help in education settings such as part of empirical data, corpora, and other such linguistic aspects, which are essential and powerful for the process of language learning. Corpora are very effective, which gives a expansive number of computational data for talkeded and written language. For example, in British English, BNC (the British National Corpus) gives a expansive data about the vocabulary usage. The expansive gathering of data gives adequate data regarding the usage of words, which help enhancing the data and academic skills of the students.

There are different powerful approaches, which are powerful for managing designs of grammar and other linguistic approaches. NLP is also an powerful technique for appraisal process to enhance the ability of understudies to recognize the connections of differing words and the use of such words in the look engine for generating treasure. Therefore; it is an

powerful approach, which permits learners and teachers to use these words more efficiently. The appraisal procedure requires entering correct data in the text in request to enter in the next level. NLP appraisal permits investigation of the students' data by coordinating it with the requirement of the content.

B. NLP in Healthcare Systems

The ubiquitous nature of Data and Communication Innovation (DCI) in later times has offered differing apparatuses such as Electronic Medical Records (EMR) and Electronic Health Records (EHR) which are profoundly beneficial to the healthcare system. These apparatuses optimize healthcare forms by giving convenient access to healthcare information, reducing healthcare fetched and errors, ensuring security and confidentiality of healthcare data and also giving an powerful strategy of storing expansive volumes of health-related data relating to diagnosis, medication, laboratory test results, pathologists, radiology as well as other imaging data which are profoundly unorganized and narrative in nature. However, it is troublesome for electronic healthcare frameworks to understand the data contents of the unorganized and narrative texts simply because they are composed of heterogeneous linguistic structures, varied expressions communicated in differing natural languages as well as the use of differing terms to denote a single concept. Consequently, the healthcare space is portrayed by ambiguity. Thus, the accessibility to valuable and meaningful healthcare data for finding and treatment in a convenient way becomes a challenge. Hence, the healthcare framework is portrayed by high fetched and high error rates. Nevertheless, Natural Language Processing (NLP) Frameworks have been used to structure data in healthcare frameworks by removing applicable data from narrative texts so as to provide data for decision making. Hence, NLP Frameworks decrease healthcare fetched and they are also noteworthy for the improvement of healthcare processes. It is therefore against this foundation that this paper examines NLP Frameworks used in healthcare, their importance to the healthcare space as well as their impediments in healthcare.

III. APPROACHES TO NATURAL LANGUAGE PROCESSING

Natural language processing approaches fall roughly into four categories: representative, geometric, connectionist, and hybrid.

A. Representative Approach

Typical approaches accomplish profound investigation of linguistic marvels and are based on explicit representation of facts about language through well-understood learning representation schemes and related algorithms.

B. Geometric Approach

Geometric approaches describe different mathematical Frameworks and often use expansive text corpora to create rough generalized models of linguistic marvels based on actual cases of these marvels given by the text corpora without adding noteworthy linguistic or world knowledge. In contrast to typical approaches, Geometric approaches use noticeable data as the primary source of evidence. Factual approaches have normally been used in undertakings such as speech recognition, lexical acquisition, parsing, part-of-speech tagging, collocations, Factual machine translation, and Factual grammar learning, and so on.

C. Connectionist Approach

Generally speaking, a connectionist model is a Framework of interconnected straightforward processing units with learning stored in the weights of the connections among units (32). Local interactions among units can result in dynamic global behavior, which, in turn, leads to computation. Some connectionist models are called localist models, assuming that each unit represents a particular concept. Comparable to the Geometric approaches, connectionist approaches also create generalized models from cases of linguistic phenomena. What separates connectionism from other Factual methods is that connectionist models consolidate Factual learning with different theories of representation - thus the connectionist representations permit transformation, inference, and manipulation of rationale formulae. In addition, in connectionist systems, linguistic models are harder to observe due to the truth that connectionist architectures are less constrained than Factual ones.

D. Assessment among Approaches

We have seen that similarities and contrasts exist among approaches in terms of their assumptions, philosophical foundations, and source of evidence. In expansion to that, the similarities and contrasts can also be reflected in the forms each approach follows, as well as in framework aspects, robustness, flexibility, and suitable tasks.

Process: Relook using these differing approaches follows a general set of steps, namely, data collection, data examination/model building, rule/data advancement and application of rules/data in system. The data gathering stage is basic to all three approaches although Factual and connectionist approaches normally require much more data than typical approaches. In the data examination/model building stage, typical approaches rely on human

investigation of the data in request to structure a hypothesis while Factual approaches manually characterize a Factual model that is a rough generalization of the collected data. Connectionist approaches assemble a connectionist model from the data. In the standard / data advancement stage, manual efforts are typical for typical approaches and the hypothesis shaped in the previous step may evolve when new cases are encountered. In contrast, Factual and connectionist approaches use the Factual or connectionist model as guidance and assemble rules or data things automatically, usually in relatively expansive quantity. After building rules or data items, all approaches then naturally apply them to particular undertakings in the system. For instance, connectionist approaches may apply the rules to train the weights of joins among units.

Framework aspects: By framework aspects, we mean source of data, hypothesis or model shaped from data analysis, rules, and premise for evaluation.

- *Data:* As mentioned earlier, typical approaches use human introspective data, which are usually not directly observable. Factual and connectionist approaches are assembled on the premise of machine noticeable facets of data, usually from text corpora.

- *Hypothesis or model based on data analysis:* As the outcome of data analysis, a hypothesis is shaped for typical approaches whereas a parametric model is shaped for Geometric approaches and a connectionist model is shaped for connectionist approaches.

- *Rules:* For typical approaches, the standard advancement stage usually results in rules with detailed criteria of standard application. For Geometric approaches, the criteria of standard application are usually at the surface level or underspecified. For connectionist approaches, individual rules normally can't be recognized.

- *Premise for Evaluation:* Assessment of typical frameworks is normally based on intuitive judgments of unaffiliated subjects and may use system-internal measures of growth such as the number of new rules. In contrast, the premise for assessment of Geometric and connectionist frameworks are usually in the structure of scores computed from some assessment function. However, if all approaches are utilized for the same task, then the results of the assignment can be evaluated both quantitatively and qualitatively and compared.

- Robustness: Typical frameworks may be fragile when exhibited with unusual or boisterous input. To bargain with anomalies, they can anticipate them by making the grammar more general to accommodate them. Compared to typical systems, Factual frameworks may be more strong in the face of unexpected input given that training data is sufficient, which may be troublesome to be assured of. Connectionist frameworks may also be strong and fault tolerant because learning in such frameworks is stored across the network. When exhibited with boisterous input, they degrade gradually.

- Flexibility: Since typical models are assembled by human investigation of well-formulated examples, typical frameworks may lack the adaptability to adapt powerfully to experience. In contrast, Factual frameworks permit broad coverage, and may be better able to bargain with unrestricted text (21) for more powerful handling of the assignment at hand. Connectionist frameworks display adaptability by powerfully acquiring proper conduct based on the given input. For example, the weights of a connectionist Framework can be adapted in real-time to improve performance. However, such frameworks may have difficulty with the representation of structures needed to handle complex calculated relationships, thus limiting their abilities to handle high-level NLP (36).

- *Suitable tasks:* Typical approaches seem to be suited for marvels that display identifiable linguistic behavior. They can be used to model marvels at all the different linguistic levels portrayed in prior sections. Geometric approaches have proven to be powerful in modeling language marvels based on incessant use of language as reflected in text corpora. Linguistic marvels that are not well understood or do not display clear regularity are candidates for Geometric approaches. Comparable to Geometric approaches, connectionist approaches can also bargain with linguistic marvels that are not well understood. They are helpful for low-level NLP undertakings that are usually sub undertakings in a bigger problem. To summarize, symbolic, statistical, and connectionist approaches have exhibited differing characteristics, thus some issues may be better tackled with one approach while other issues by another. In some cases, for some particular tasks, one approach may prove adequate, while in other cases, the undertakings can get so complex that it might not be conceivable to choose a single best approach. In addition, as Klavans and Resnik pointed out, there is no such thing as a "purely statistical" method. Every use of measurements is based upon a typical model and measurements alone is not adequate for NLP. Toward this end, Factual approaches are not at odds with typical approaches. In fact, they are rather complementary. As a result, scientists have begun developing hybrid Frameworks that utilize the strengths of each approach in an attempt to address NLP issues more effectively and in a more flexible manner.

IV. APPLICATIONS

Natural language processing gives both hypothesis and executions for a range of applications. In fact, any application that utilizes text is a candidate for NLP. The most incessant applications utilizing NLP incorporate the following.

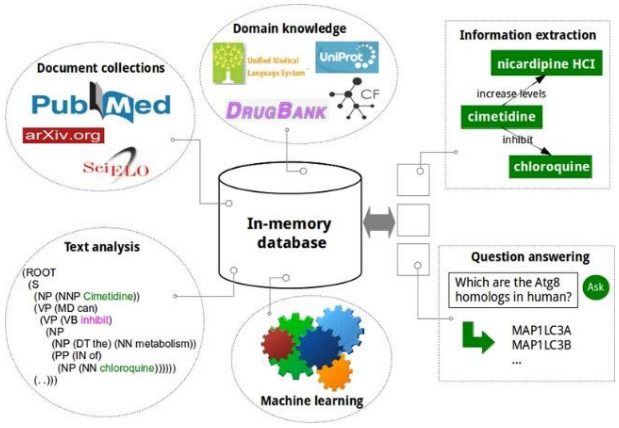

Fig: 4.1 NLP Applications

A. Machine Translation

Machine translation alludes to programmed translation of text from one human language to another. It is necessary to have an understanding of words and phrases, grammars of the two languages involved, semantics of the languages, and world knowledge. Different levels of NLP have been utilized in MT systems, ranging from the "word-based" approach to applications that incorporate higher levels of analysis.

B. Speech Recognition

Speech recognition is the process of mapping acoustic speech signals to a set of words. The difficulties arise due to wide variations in the pronunciation of words, homonym and acoustic ambiguities.

C. Speech Combination

Speech combination alludes to programmed production of speech (expression of natural language sentences). In request to create utterances, text has to be processed. So, NLP remains an essential component of any speech combination system.

D. Natural Language Interfaces to Databases

Natural language interfaces permit querying a organized database using natural language sentences. PCs have been broadly used to store and manage expansive amounts of data. The data might pertain to railway reservation, library, banking, management information, and so on.

E. Data Recovery

Data recovery is concerned with identifying reports applicable to a user‟s request. It is given the noteworthy presence of text. Factual approaches for accomplishing NLP have seen more utilization, few executions utilize NLP. NLP Frameworks have found helpful applications in data recovery such as indexing, word sense disambiguation, inquiry modification and learning bases have also been used in IR system.

F. Data Extraction

Framework captures an data extraction and factual data output contained in a document. An data recovery framework reacts to a user‟s data need. The data need is not communicated as a catchphrase inquiry in an data recovery system. Instead it is specified as pre-characterized database schemas or templates. An IR framework identifies a subset of reports in a expansive repository of text database. IE focuses on the recognition, tagging, and extraction into a organized representation, certain key elements of information, e.g. persons, companies, locations, organizations, from expansive collections of text. These extractions can then be utilized for a range of applications including question-answering, visualization, and data mining.

G. Question Answering

Question Answering gives a list of potentially applicable reports in response to a user‟s query, questionanswering gives the client with either just the text of the answer itself or answer-giving passages. Given a question and a set of documents, a question answering framework attempts to find the exact answer, or at slightest the exact portion of text in which the answer appears. A question answering framework is differing from an data extraction framework in that the content that is to be extracted is unknown. In general, a question answering framework benefits from having an data extraction framework to recognize entities in the text. A question answering framework requires more NLP than an data recovery framework or an data extraction system.

H. Text Summarization

The higher levels of NLP, especially the speech level, can empower an implementation that reduces a bigger text into a shorter, yet richly constituted abbreviated narrative representation of the original document. Text summarization

deals with the creation of summaries of reports and includes syntactic, semantic, and speech level processing of text.

I. Exchange Frameworks

Frameworks envisioned by expansive providers of end-client applications. Exchange systems, which usually focus on a narrowly characterized application (e.g. your refrigerator or home sound system), currently utilize the phonetic and lexical levels of language. It is believed that utilization of all the levels of language processing explained above offer the potential for truly habitable Exchange systems.

V. USES OF NLP

There are many applications of natural language processing created over the years. They can be mainly divided into two parts as follows.

A. Text-based applications

This includes applications such as searching for a certain topic or a catchphrase in a data base, removing data from a expansive document, translating one language to another or summarizing text for differing purposes.

B. Exchange based applications

Some of the typical cases of this are answering frameworks that can answer questions, administrations that can be given over a telephone without an operator, teaching systems, voice controlled machines (that take instructions by speech) and general issue solving systems.

VI. FRAMEWORKS USED IN ANALYZING NLP

There are several main Frameworks used in analyzing natural language processing. Some of them can be briefly portrayed as follows.

A. Plan coordinating

The thought here is an approach to natural language processing is to decipher input expressions as a entirety father than building up their understanding by combining the structure and meaning of words or other lower level constituents. That implies the interpretations are gotten by coordinating designs of words against the input utterance. For a profound level of investigation in plan coordinating a expansive number of designs are required indeed for a restricted domain. This issue can be ameliorated by hierarchical plan coordinating in which the input is gradually canonicalized through plan coordinating against sub phrases. Another way to decrease the number of designs is by coordinating with semantic primitives instead of words.

B. Linguistically driven Parsing

Syntax implies ways that words can fit together to structure higher level units such as phrases, clauses and sentences. Therefore linguistically driven parsing implies understanding of bigger groups of words are assembled up out of the understanding of their syntactic constituent words or phrases. In a way this is the opposite of plan coordinating as here the understanding of the input is done as a whole. Syntactic analyses are gotten by application of a grammar that determines what sentences are legal in the language that is being parsed.

C. Semantic Grammars

Natural language investigation based on semantic grammar is bit comparable to linguistically driven parsing except that in semantic grammar the classes used are characterized semantically and syntactically. There here semantic grammar is also involved.

D. Case outline instantiation

Case outline instantiation is one of the major parsing Frameworks under active relook today. The has some very helpful computational properties such as its recursive nature and its ability to consolidate bottom-up recognition of key constituents with top-down instantiation of less organized constituents.

VII. FUTURE OF NLP

NLP's future will be recharacterized as it faces new mechanical challenges to make more user-friendly systems. It is also forceing NLP more towards Open Source Development. If the NLP group embraces Open Source Development, it will make NLP frameworks less proprietary and therefore less expensive. The frameworks will also be assembled as easily replaceable components, which take less time to assemble and more user-friendly. Web entryway administrations interface are becoming progressively user-friendly. NLP will progressively play a basic part in the plan and advancement of successful Web portals. Searching must not require an education in SQL, Boolean logic, lexical analysis, or the underlying structures of data repositories. Clients overwhelmingly accept look usefulness that is natural language-based. Searches of all sorts of data are expected to decipher and expand queries lexically, while simultaneously conveying exact results focused on the essence of the search. These results should be ranked by perceived relevancy to the query. Queries, whether of organized data records or documents, should deliver answers

– not database records or gathering of documents. A look apparatus may also support a portal's presentation and personalization features, giving clients control over the level of detail and presentation of the answer set. The look apparatus should capacity against both organized and unorganized sorts of data stores with a single query, conveying a single, combined answer set that is data neutral – be able to return streaming video resources as well as database fields or applicable segments of text documents. A new market for look innovation is emerging, one in which established vendors are seeking to broaden their usefulness and new innovation is coming to market with innovative approaches against new Web-based engines.

Several other future applications of NLP, most of them currently under development, are as follows:

- Conversational systems. The first challenge for a speech recognition framework used in these frameworks still remains to be proper recognition of what is being talkeded by a wide assortment of individuals with differing vocabularies and accents. Frameworks where a PC would be able to read a book, store the data about the book, and then answer questions about the book. These sorts of framework would be dealing with advanced type of auto indexing.

- Artificial Neural Networks. One of the interesting products now being introduced on the market is DolphinLook technology. Dolphins learn by recognizing the attributes of objects off of which they bounce sonar waves. They learn by categorizing and remembering the different reflections that come back from the objects. In a comparable manner, this approach relates words to one another so that, in ambiguous situations, their linguistic part becomes evident.

- Microsoft MindNet – combination of an extensive database and calculations that can characterize relationships. The project is attempting to use dictionaries in sindeed languages and a assortment of encyclopedias to make a framework that recognizes connections among straightforward words (from the dictionaries) and phrases or sentences (from the encyclopedias). The connections are assembled and identified by straightforward questions directed at the system. MindNet also promises to be a powerful apparatus for machine translation. The thought is to have MindNet make separate calculated webs for English and another language, Spanish. MindNet then annotates these matched logical forms with data from the English-Spanish translator memory, so that translation can proceed smoothly in either direction.

- Medication Assistant – a therapeutic DSS, which models the effects of therapy on patients with cardiovascular and other therapeutic conditions. Prolog programming language, used in this DSS to control NLP joins hierarchically linked data and grammatically corrects text.

- Chatterbots– although they exist already, new generations of them are being constantly developed. Chatterbots use natural language processing to simulate conversations with users. Web sites are beginning to install chatterbots as Web guides and customer administration agents.

VIII. CONCLUSION

With over sixty years of NLP relook and development, the natural language frameworks are still very complicated to design. Most of all, NLP frameworks are still not perfect because natural human language is complex, and it is troublesome to capture the entirety linguistic learning for hundred percent accuracy in processing. Indeed though hundreds of organizations are replacing some administration reps with voice software, emergency administrations like 911 will continue to be handled by individuals for at slightest another decade or so because of their basic nature. The current voice frameworks still need adjustments -- some can't understand heavy accents, speech impediments or quiet voices. If the data frameworks group reacts to the challenge by building NLP frameworks with reusable parts via Open Source programming, the future of NLP will start looking indeed brighter. There are still unresolved challenges for programming programs to represent the entirety knowledge, the differing connections and cultures of the world.

References:

[1] Chetan Arora, Mehrdad Sabetzadeh, Lionel Briand & Frank Zimmer, "Automated Checking of Conformance to Requirements Templates Using Natural Language Processing", in IEEE Transactions on Software Engineering, Vol-41, Issue: 10, PP. 944 – 968, 2015.

[2] Henrik Leopold, Jan Mendling and Artem Polyvyanyy, "Supporting Process Model Validation through Natural Language Generation", in IEEE Transactions on Software Engineering, Vol- 40, Issue-8, PP.818 – 840, 2015.

[3] Julia Hirschberg, Bruce W. Ballard and Donald Hindle, "Natural language processing", in AT&T Technical Journal, Vol- 67, Issue-1, PP. 41 – 57, 1988.

[4] Erik Cambria and Bebo White, "Jumping NLP Curves: A Review of Natural Language Processing Research", in IEEE Computational Intelligence Magazine, Vol-9, Issue-2, PP.48 – 57, 2014.

[5] Michael T. Mills and Nikolaos G. Bourbakis, "Graph-Based Methods for Natural Language Processing and

Understanding-A Survey and Analysis", in IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol-44, Issue-1, PP.59 – 71, 2014.

[6] T. Patten and P. Jacobs, "Natural-language processing", in IEEE Expert, Vol- 9, Issue- 1, PP. 35, 1994.

[7] Minhwa Chung and D. Moldevan, "Applying parallel processing to natural-language processing", in IEEE Expert, Vol-9, Issue-1, PP. 36 – 44, 1994.

[8] B. W. Wah and T. S. Huang, "Report on workshop on high performance computing and communications for grand challenge applications: computer vision, speech and natural language processing, and artificial intelligence", in IEEE Transactions on Knowledge and Data Engineering, Vol- 5, Issue-1, PP. 138 – 154, 1993.

