ISSN: 0976-3031

Available Online at http://www.recentscientific.com

CODEN: IJRSFP (USA)

International Journal of Recent Scientific Research
Vol. 8, Issue, 8, pp. 19585-19589, August, 2017

International Journal of Recent Scientific Research

DOI: 10.24327/IJRSR

Research Article

MEDICAL APPLICATION USING NLP OVER CLOUD

Anshul Garg*., Vaibhav Singhal and Rachna Jain

Completed Bachelors of Technology (B.Tech) from Bharati Vidyapeeth's College of Engineering, IPU, India

DOI: http://dx.doi.org/10.24327/ijrsr.2017.0808.0725

ARTICLE INFO

Article History:

Received 05th May, 2017
Received in revised form 21st June, 2017
Accepted 06th July, 2017
Published online 28th August, 2017

Key Words:

Artificial Intelligence, Clinical decision support system, Cloud computing, Natural language processing, Parser, POS Tagger, Semantic analyzer, Syntax Analyzer, Tokenizer.

ABSTRACT

Cloud computing is a gushing technology where all the computing resources are needed on the cloud rather than having local servers or personal gimmicks to handle applications. Cloud storage means "*the storage of data online in the* cloud," where in a company's data is hoarded in and accessible from multiple distributed and connected resources that engross a cloud.
Cloud storage can provide the welfare of greater accessibility and reliability; fleet deployment; rugged protection for data backup and disaster recovery purposes. Natural Language Processing (NLP) in the medical field describes potential to influence the way in which every day clinical dismay and medical research is supervised. NLP systems cater access to structured content embedded in coarse medical texts, therefore enabling automated processing. This paper describes an advent and implementation which leverages cloud-based deployment of medical related NLP application which can pinpoint the disease on the basis medical condition or symptoms provided by user. This NLP interface extracts, process, synthesize, compare/contrast, scrutinizes, and manage medical text data in a flexibly immune and scalable architecture.

INTRODUCTION

CLOUD Computing means being able to access and modify content online. It offers online data storage, infrastructure and application. Cloud computing is an internet-based computing in which a number of secluded servers are networked such that they allow sharing of data-processing tasks, centralized data storage, online access to computer services. Cloud computing stakes users and enterprises with various capabilities to hoard and process their data in third-party data centers and it relies on sharing of resources to achieve coherence and of scale similar to a unity over a network. Clouds can be classified accede to the deployment models as public, private or hybrid.

Public cloud: The public cloud allows systems and services to be easily accessible to the general public. Public cloud may be less secure because of its openness.

Private Cloud: The private cloud allows systems and services to be accessible within an organization. It offers increased security because of its private nature.

Hybrid Cloud: The hybrid cloud is a mixture of public and private cloud. However, the critical activities are performed using a private cloud while the non-critical activities are performed using public cloud.

Natural languages are languages naturally evolved and used by human beings. Natural Language Processing (NLP) also called computational linguistics is a scientific study of language from computational perspective. In NLP, the aim is to fabricate computer understand the commands given in a particular language and perform it. The goal of NLP is the construction of a specific language comprehension and a production theory to such a level of detail that a person is able to write a computer program that would understand or produce natural language.

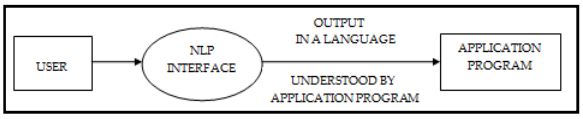

Fig 1 Block diagram of NLP system

Natural language processing system takes a sentence that compromises of string of words as its input and produces structured representations capturing the meaning of those words as their output. The nature of this output depends heavily on the efficiency of the processing mechanism. A natural language understanding system serving as an interface to a database might accept questions in English which relate to the kind of data held by the database. In this case the meaningof

Corresponding author:* **Anshul Garg
Completed Bachelors of Technology (B.Tech) from Bharati Vidyapeeth's College of Engineering, IPU, India

the input (the output of the system) might be expressed in terms of structured SQL queries which can be directly submitted to the database.

In normal NLP system, the input is given in natural language. This input is given to a parser which generates the syntactical structure of a sentence. Then the semantic analyzer captures its semantic details and generate a more deep structure with association between different words. The conversion rules accept it and make it compatible for database point of view. The database handler works on it to generate processed form from storage point of view.

The different stages of NLP are as follows:

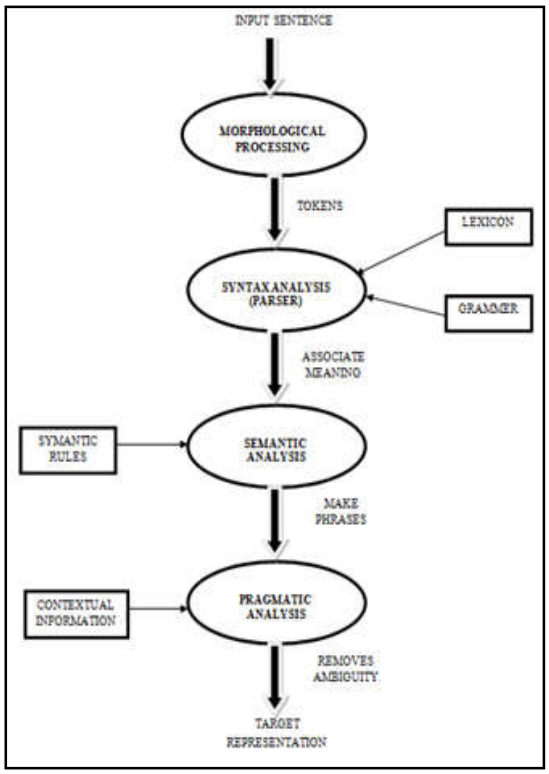

Fig 2 Stages of NLP system

Morphological Processing

It is the preliminary stage of NLP. It deals with the word formation which takes place before syntax analysis. The purpose of this stage of language processing is to break strings of language input into sets of tokens corresponding to discrete words, sub-words and punctuation forms. Each individual word is analyzed according to their components called morphemes. Whereas, non-token words such a punctuations are separated. Morphology is concerned primarily with the errand of braking words into morphemes. A morpheme is a minimal meaningful unit of language that can't be broken down further. Modification typically occurs by the addition or removal of prefixes and/or postfixes but other textual changes can also take place. For example, word "happy" consist of a single morpheme as it can't be broken down further. Whereas word "cats" consist of 2 morphemes i.e. morpheme "cat" and morpheme "s" which indicate plurality.

Syntax Analysis

This stage also comprises of lexical analysis, which checks the validity of words according to the lexicon. Lexicon stands for dictionary. It makes the collection of all possible valid words of the language along with its meaning.

Syntactical Analysis is performed after that. It study the formal relationship between the words of sentences according to grammar rules. It requires the knowledge of grammar and parsing technique. The grammar is formal specification of rules allowable in a language and parsing is a method of analyzing a sentence to determine its structure according to grammar.

Semantic Analysis

A language processor must carry out a number of different functions primarily based around syntax analysis and semantic analysis. The purpose of syntax analysis is two-fold: to check that a string of words (a sentence) is well-formed and to break it up into a structure that shows the syntactic relationships between the different words. Semantics further deals with the meaning of Natural Language Sentences. In this phase the meaning of the sentence is understood. NLP semantics consist of two major features:

1. representation of the gist in a way that allows all the necessary manipulation
2. relating these representation to the portion of linguistic model which will define its structure

Pragmatics Analysis

After semantic analysis the next stage of processing deals with pragmatics. Unfortunately there is no universally agreed distinction between semantics and pragmatics. Semantic analysis associates meaning with isolated utterances/ sentences, whereas Pragmatic analysis interprets the results of semantic analysis from the perspective of a specific context.

Pragmatics is the study of the relation between the language and content of use. Content of use includes thing which are the replication of identity of people. In this phase, the main intention of the speaker is understood. Pragmatic analysis simply fits actual objects/events which exist in a given context with object references obtained during semantic analysis. In some situations pragmatic analysis can disambiguate sentences which cannot be fully disambiguated during the syntax and semantic analysis phases.

This paper integrates all these elements to develop a CDS over cloud. Computerized Clinical Decision Support (CDS) aims to aid decision making relating to the health care providers and the public. It provides a mechanism involving easy accessibility of health-related information at any point and time, when needed. Natural Language Processing (NLP) is instrumental in using free-text information stored in database over cloud to drive CDS. Thus, representing clinical knowledge and CDS interventions in standardized formats, which is widely acceptable and understood by everyone. The early innovative NLP research of clinical narrative was followed by a period of stable research conducted at the major clinical centers and a shift of mainstream interest to biomedical NLP. This review focuses on the recently renewed interest in development of fundamental NLP methods and widely available cloud environment, which will provide intuitive features to its user.

Moreover, the omnipresent nature of cloud make this application widely available to multiple user from anywhere.

Need for NLP in Clinical Decision Support System

NLP is used in the CDS system to dissect all the symptoms that are input to match them with the database symptoms. No client just states the symptoms one by one. They use proper sentences to tell their symptoms. Of these full sentences the symptoms have to be taken out and other parts of the speech are to be rejected. This makes the identification of the symptoms very easy.

Working

Terms Used

NLP (Natural language processing): It is a field of artificial intelligence concerned with the interaction between computers and humans (natural languages). As the name suggests the natural or human language of the human is taken as input and processed to get the desired output.

1. Clinical decision support system (CDS): It is a health information technology which aims to aid he decision making of healthcare providers and the public by providing easily accessible health-related information. It uses artificial intelligence and NLP for its use.
2. Tokenizer: It is the first step in NLP where the input language is broken up in tokens or lexemes. Each word is the smallest unit called a token.
3. POS Tagger (Parts of speech tagger): It is used to put a parts of speech tag to each token and make some sense of the input. Example: The girl is sick. 'The' is tagged as an article. Girl as a noun. 'Is' is a verb. 'Sick' as an adjective.
4. Parser: In a parser a tagged tokens are parsed using some algorithm and parsing tree is created.

Proposed Work

This algorithm uses weights ranging from 0 to 5. Each symptom of the disease is assigned a weight depending on the rarity or the intensity of the symptom. An intense symptom is frequently the rare symptom which leads to a rare yet a malign disease. A summation function is used and all the weighted symptoms of the disease are added to give the final value of each disease.

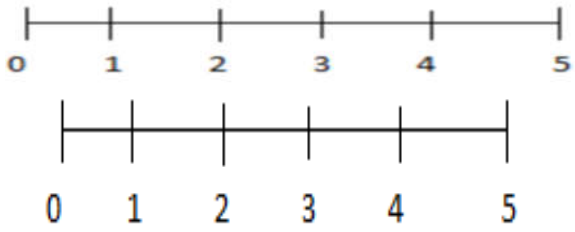

Fig 3 The weight-scale used for each symptom.

These final values are commuted for each of the diseases in the database. Next, when the client inputs the symptoms he is ailing from, each of the symptoms is assigned the weight that is already stored in the database. Then the summation function is applied and the value of the disease is calculated (adding all the weighed symptoms). This value of the disease is matched with the value of the diseases already in the database. Now the problem is that a disease may have many symptoms that are not already registered and also many diseases may have overlapping symptoms. Thus there is a error rate in the matching of the diseases. For this we take an error rate of ± 10 to accommodate all the diseases within the range. After these diseases are shortlisted the symptoms input are matched with the diseases and the diseases with the highest match are chosen.

Let S_i be the ith symptom of the input disease and D_j the final value of the jth disease. n is the total number of symptoms that are input for a disease. Hence the summation function is:

$$D_j = \sum s1 + s2 + s3.....si + ...sn$$

Modules

Voice to text: In the voice to text module the speech signals that are input to the hardware are recognised and then converted to digital signals. This is done by the PC sound card and is achieved through the sampling of the speech signal at regular intervals. Next the software acoustic model breaks the word into plosive constant sounds (constant stops produced by obstructing the air flow in the vocal tract) and matches them to the known phenomes. The software language model matches phenomes are joined to make the textual representation of the speech signals.

POS tagger: also called the parts of speech tagger is used to tag each word of the sentence input with the corresponding parts of speech. Tagging of each part-of-speech is more difficult than just having a list of words and their parts of speech, because many words can represent more than one part of speech at different times, and because some parts of speech are complex or unspoken. This is very common in natural languages. For example, even "economics", which is usually thought of as a plural noun, can also be a verb:

"Economics is said to be difficult": Here the word economics is being used as a plural noun.

"The economics of the situation demand that": Now here the word economics is being used as a verb.

Correct grammatical tagging will reflect that "economics" is used here as a verb, not as plural noun.

Clinical decision maker: The clinical decision support system (CDS) is used to make the decision regarding the diagnosis. It is a computerized decision maker. It is fast and accurate with a timely and instant diagnosis. It uses the POS tagger and the algorithm stated to calculate the final value of the disease. The benefit of deploying it on a cloud is that this information is accessible anywhere and can be shared with any number of doctors. It is helpful as the patients' medical papers can even be uploaded and kept handy whenever needed.

Database searcher: In this module the medical database is searched using the fastest algorithm*** for the corresponding disease value taking into account the ± 10 error rate. This result (disease) is displayed back. This benefits both the patient and the doctors.

Fig 4 GUI of the login page

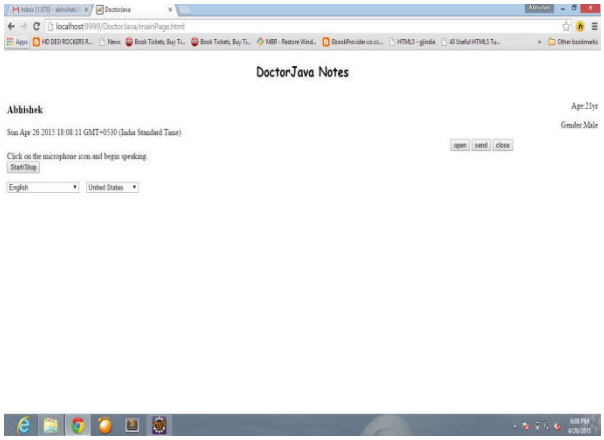

Fig 5 GUI of the voice to text input of the symptoms from the patient.

Geolocation module: In a case where there is no diagnose of a disease or more than one disease is output a geolocation application is called which states the nearest hospital for the convenience of the patient. A google API is used to implement this module.

CONCLUSION

Through this paper we can conclude that the clinical decision support system can be effectively used using the above proposed algorithm. Further using NLP to identify the symptoms and different parts of speech make the system very efficient and accurate. The system being on cloud makes it available to all, making it convenient for the patients and the doctors likewise to access the system.

Future Scope

This project can further be expanded by removing the ambiguity of the symptoms and increasing the tagging and parsing process. A better function than just a summation function can be used to reduce the error rate below 10. Database can be increased by adding more diseases. And each patient and doctor can even manage their medical accounts and information. Further, a payment gateway can be associated with this application that will help the user in the easy online purchase of the medicines or any other form of services.

References

1. Kyle Chard, Michael Russell, Yves A. Lussier, Jonathan C. Silverstein. Scalability and Cost of a Cloud-based Approach to Medical NLP.
2. A. Aronson. A. effective mapping of biomedical text to the umlsmetathesaurus: the metamap program. In Proceedings of the AIMA Symposium, 2001.
3. A. Aronson and F. Lang. An overview of metamap: historical perspective and recent advances. *Journal of the American Medical Informatics Association,* 17:229-236, 2010.
4. D. Carrell. A strategy for deploying secure cloud-based natural language processing systems for applied research involving clinical text. In Proccedings of the 44th Hawaii International Conference on System Sciences (HICSS), 2011.
5. K. Chard, M. Russell, E. Mendonc¸a, Y. A. Lussier, and J. Silverstein. A cloud-based approach to medical nlp. In Submitted to: AIMIA, 2011.
6. R. S. Crowley, M. Castine, K. Mitchell, G. Chavan, T. McSherry, and M. Feldman. Caties: a grid based system for coding and retrieval of surgical pathology reports and tissue specimens in support of translational research. *Journal of the American Medical Informatics Association*, 17(3):253- 264, 2010.
7. H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan. Gate: an architecture for development of robust hlt applications. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, 2002.
8. J. T. Dudley, Y. Pouliot, R. Chen, A. A. Morgan, and A. J. Butte. Translational bioinformatics in the cloud:an affordable alternative. *Genome Medicine*, 2(8):51-57, 2010.
9. C. Friedman, G. Hripcsaka, W. DuMouchela, Johnsona, and Claytona. Natural language processing in an operational clinical information system. *Journal of Natural Language Engineering*, 1(1):83-108, 1995.
10. A. Rosenthal, P. Mork, M. Li, J. Stanford, D. Koester, and P. Reynolds. Cloud computing: A new business paradigm for biomedical information sharing. *Journal of Biomedical Informatics*, 43:342-353, 2010.
11. G. Savova, J. Masanz, P. Ogren, J. Zheng, S.Sohn, K. Kipper-Schuler, and C.G.Chute. Mayo clinical text analysis and knowledge extraction. *Journal of the American Medical Informatics Association,* 17(5):507-513, 2010.
12. D. P. Wall, P. Kudtarkar, V. A. Fusaro, R. Pivovarov, P. Patil, and P.J. Tonellato. Cloud computing for comparative genomics. BMC Bioinformatics, 11(1):259, 2010.
13. Anil A. Thomas *et al*. Extracting data from electronic medical records; validation of natural language processing program to assess prostate biopsy results. *World Journal of Urology*, February 2014, Volume 32, Issue 1, pp 99-103.
14. K. Vijaykumar, C. Arun. Automated risk identification using NLP in cloud based development environments. *Journal of Ambient Intelligence and Humanized Computing,* pp1-13, 2017.

15. David R Kaufman *et al*. Natural Language Processing-Enabled and Conventional data Capture Methods for Input to Electronic Health Records: A Comparative Usability Study. JMR Medical Informatics, 206 Oct-Dec; 4(4); e35.
16. Friedman C, Elhadad N. Natural language processing in health care and biomedicine. In: Shortliffe EH, Cimino JJ, editors. Biomedical Informatics. 4th edition. London: Springer; 2014. pp. 255-284.
17. Healthit. 2012. [2015-08-23]. EHR Adoption & Meaningful Use Progress is Assessed in New Data Briefshttp://www.healthit.gov/buzz-blog/meaningful-use/ehr-adoption-meaningful-use-progress-assessed-data-briefs/ *webcite*.

How to cite this article:

Anshul Garg., Vaibhav Singhal and Rachna Jain.2017, Medical Application Using NLP Over Cloud. *Int J Recent Sci Res.* 8(8), pp. 19585-19589. DOI: http://dx.doi.org/10.24327/ijrsr.2017.0808.0725

****