Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 132 (2018) 1375–1384

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Computational Intelligence and Data Science (ICCIDS 2018)

Fuzzy Logic in Natural Language Processing – A Closer View

Charu Gupta[a], Amita Jain[b,*], Nisheeth Joshi[c]

[a]*Assistant Professor, Bhagwan Parshuram Institute of Technology, Delhi-110089, India*
[b]*Assistant Professor, Ambedkar Institute of Advanced Communication Technologies and Research, Delhi, India*
[c]*Associate Professor, Banasthali Vidyapith, Rajasthan, India*

Abstract

The natural instinct of human beings is very complex. The understanding of this instinct requires a clear dimensional analysis of the knowledge of discourse. The computer systems are now trained to understand how things work in real-world domain for intelligent analysis. This effort although very progressive has a limitation. There is an intelligence gap which makes human one step above the machine. Fuzzy logic can be used to make a machine understand this intelligence gap in a better way. Fuzzy logic is the science which makes a computer understand and think the way humans do. The aim of this study is two folds: first, to understand fuzzy logic, a computational Intelligence technique, for effective decision making and second, to illustrate through real world examples the existence of this intelligence gap using well known natural language processing applications like Google Search Engine, Google Translator and MIT Start. To the best of our knowledge, there is no work available in the literature which exemplifies this intelligence gap in such a simple manner. The examples are chosen carefully to illustrate and demonstrate the applications of fuzzy logic in natural language processing environment for every reader.

Keywords:Fuzzy logic, Natural language processing, Google Search Engine, MIT start, Google Translator.

1. Introduction

Computer Science is an interesting field of study as it inherits its methods from both Mathematics and Information Technology & Engineering. It includes approaches, methods consisting of mathematical axioms, postulates, and proofs, quantification, comparison, measurement, simulation, implementation and experimentation. It

Corresponding author: charugupta0202@gmail.com

10.1016/j.procs.2018.05.052

has been rightly said that computer science investigates the contemporary phenomenon within its real life context; when the boundaries between phenomenon and context are not clearly evident[1,2]. A very special area with emerging importance within Computer Science is *Computational Intelligence* that is often referred to as Soft Computing, or Bio-inspired Computing, as its methods include various techniques and technologies borrowed from soft sciences, and are often motivated by phenomena observed in biological organisms, or systems, including the partial imitation of human perception, cognition, and reasoning; and the behaviour of groups of human beings, especially social systems[3,4,5].

Nowadays, Computational Intelligence (CI) and, within this broader area, Fuzzy Systems has become one of the most successful technologies for sophisticated control and reasoning/inference systems [6, 7]. The reason for its success lies in the fact that Fuzzy Systems closely resembles human decision making with the ability to solve problems and generate solutions which are precise. It is also important that fuzzy models are transparent and easily adaptable. Further, the fact that fuzzy approaches use the sub-symbolic information available in the universe of discourse providing the context to the given model and solution [8, 9]. Its advantage is that it infers acceptably good solutions from approximate information, by reducing the computational complexity tremendously, compared to classic approaches. The gap in engineering design which is left vacant by purely mathematical methodologies is often filled by fuzzy logic, and other CI techniques. Other approaches require hard core mathematical equations to model real life complex problems. But fuzzy systems can accommodate the ambiguities of the real world more aptly due to the formation of soft or vague boundaries, instead of hard (crisp) boundaries [10]. The problem specifications are converted here into automatic systems by intuitive design methodologies. The applications of fuzzy theory were mainly industrial, such as the very first one: the process control of cement kilns. However, as the technology was further embraced, fuzzy systems were used in many more applications, some of them rather sophisticated. The applications of Fuzzy technology have virtually exploded, affecting things we use every day, including domotic systems[11], vehicular technology[12], decision support in management[13], economy[14], diagnostics[15] and many other areas. The importance of fuzzy systems can be understood with the fuzzy washing machine, fuzzy microwave oven, fuzzy functions in cars, like driver adaptive automatic gear and fuel consumption optimisation, and many more. Practically, the most exciting thing about it is the simplicity involved in operating it, which is provided by the transparency of fuzzy models in contrary to many other mathematical (approximation) models. The growth of fuzzy systems applications over the years has seen growth in various segments, as it is shown in Figure 1. The growth has been analysed with the amount of research done in fuzzy theory, both at conceptual level and in the application field. The information for analysis has been collected from the well-known database of Google Scholar [16]. The results have been shown graphically indicating the volume of growth in fuzzy systems and allied techniques.

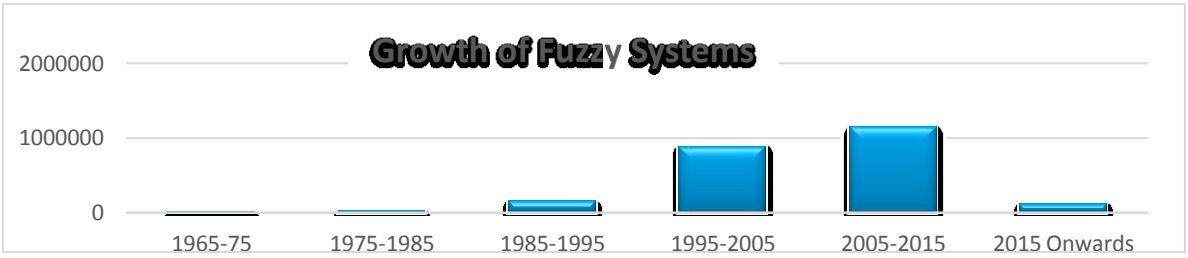

Figure 1: Fuzzy Systems Growth

In problems where fuzzy logic can be applied, there is a multitude of interesting and challenging problems. Such are the size of the large data sets (long time series, spatial data with high resolution, etc.), the heterogeneity of different data sources; such as sources of quantitative information, like measurement and simulation, and sources of subjective (often only qualitative) information, like expert knowledge and subjective evaluations; the coexistence of different data structures and data formats (e.g. time series, spatial data), and the presence of different types of data (e.g. quantitative and qualitative data), and the uncertainty of those data. An inherent uncertainty of this information is there because of the use of random variables, of inaccurate data, and approximate estimations, the incomparability

of the data, the partially vague expert knowledge, and the subjectivity of the information. In view of these factors fuzzy approaches find a wide field of applications and Natural Language Processing(NLP) is one of them.

The proper selection and the efficient use of fuzzy logic provide good solutions to various data analytics problems. Descriptive analysis and predictive modelling provide useful insight into complex computational problems. These insights eventually help in finding inter-relations among various phenomena. Fuzzy logic often provides action oriented insights, which may lead to new discoveries and impacts. It is, thus, the need of the hour to carefully consider and deploy the qualitative amalgamation of fuzzy logic in data analytics. Also, in order to address the issues of data science and data analytics, the modelling of solutions in view of simplicity and flexibility makes them suitable for being solved by fuzzy models. These methods are able to deal with vagueness, uncertainty and imprecision, which usually occur in human reasoning.

The aim of this paper is to understand the existence of uncertainty in natural language. The vagueness, imprecision can be exemplified with some well-known applications of *Google Search Engine* [17], *Google Translator*[18] and *MIT Start*[19]. The authors have carefully chosen examples which clearly illustrate the intelligence gap of the machines in understanding what is the desired information required by the user. This study will prove to be an effective tool in using fuzzy logic to natural language processing and will arose interest of academicians, researchers to focus more on domain specific attributes of natural language processing where fuzzy logic can be applied.

The rest of this paper is organized as follows: section 2 discusses the objective of this study. Section 3 presents the motivation which led to the analysis of fuzzy logic in natural language processing. Section 4 illustrates the observational analysis of well-known applications of *Google Search Engine, Google Translator and MIT Start* with a view to understand the intelligence gap. Section 5 presents a conclusion of the study.

2. Objective

Natural language has semantics, pragmatics for the people communicating in a particular language to express their sentiments, feelings, behaviour. Human brain is a marvellous machine which can translate and understand the text/audio in natural language.

It is also to be observed that natural language has characteristics like- ambiguity, uncertainty, vagueness, and un-descriptiveness. Some of the natural languages are free order as well, for example Hindi. These inherent characteristics make the outcome of a natural language processing system inaccurate.

To design an application to make a computer understand the natural language and interpret the way humans do, is the primary task of any natural language processing system. Natural language processing is a wide area with multi-facets applications in various domains. Natural language processing is ubiquitous and is likely to see a more powerful and useful future. Some of the unknowing use of natural language processing tools in our day-to-day life are- predictive typing, auto-correct, spell checker, grammar checker, duplicate detection, spam detection and so on.

3. Motivation

Uncertainties can be handled by fuzzy logic. Fuzzy logic is a generalization of bivalent logic with more flexibility for handling vagueness and ambiguity. In fuzzy logic, degree of membership is given to the elements of a set which lies between [0, 1]. Unlike crisp set, a discrete value is given to each element of a set. With the understanding of the presence of ambiguity and uncertainty, fuzzy logic can provide useful insights to system which processes natural language. In fuzzy set theory, the task of evaluating text/audio in natural language involves three basic steps:

a. Categorization
b. Prioritization
c. Detection of Specific theme

These basic steps form the backbone of any fuzzy logic based natural language processing system. Natural language processing systems which can handle fuzzy logic based statements should be encouraged. This gives more scalability, flexibility and efficiency to the natural language processing system.

It is also important that developed fuzzy models are transparent and easily adaptable. Its advantage is that it infers acceptably good solutions from approximate information, by reducing the computational complexity tremendously, compared to classical approaches. Further, the fact that fuzzy approaches use the sub-symbolic information available in the universe of discourse, providing the context to a given model and solutions, adds additional advantage.

Nowadays, Fuzzy Systems has become one of the most successful technologies for sophisticated control and reasoning/inference systems. The reason for its success lies in the fact that Fuzzy Systems closely resembles human decision making with the ability to solve problems and generate solutions which are precise.

4. Observational Analysis

In this section, some examples are used as queries to the applications to understand how ambiguity, uncertainty, vagueness in natural language input text affects poorly on the performance of the state-of-art NLP systems. The application of fuzzy logic to improve these systems is also discussed. The performance is analysed for most important natural language processing applications including Machine Translation, Question Answering System and information retrieval. The most powerful systems are considered. They are the "START developed at MIT for Question Answering", "Google search engine" and "Google Machine Translation".

a. ***Start, MIT's natural language question answering system***.
Let us begin by understanding the poor reasoning of the system using ***Start, MIT's natural language question answering system***. Figure 2 and 3 show the results for two queries (1 and 2) semantically related:
Query 1: Show me the picture of a tiger (showed the correct answer)
Query 2: Show me how a tiger *looks like* (showed the message "I don't know the answer")

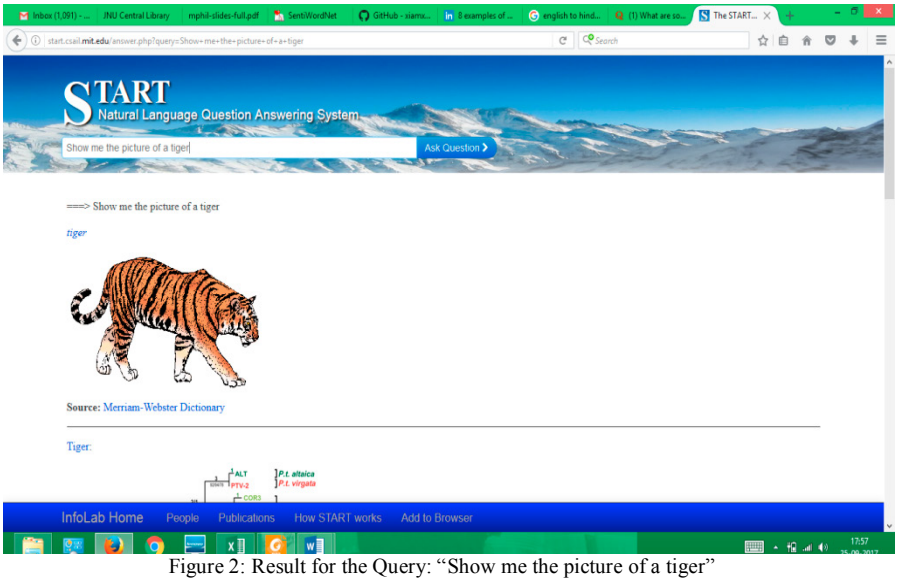

Figure 2: Result for the Query: "Show me the picture of a tiger"
(MIT Natural Language Question Answering System)

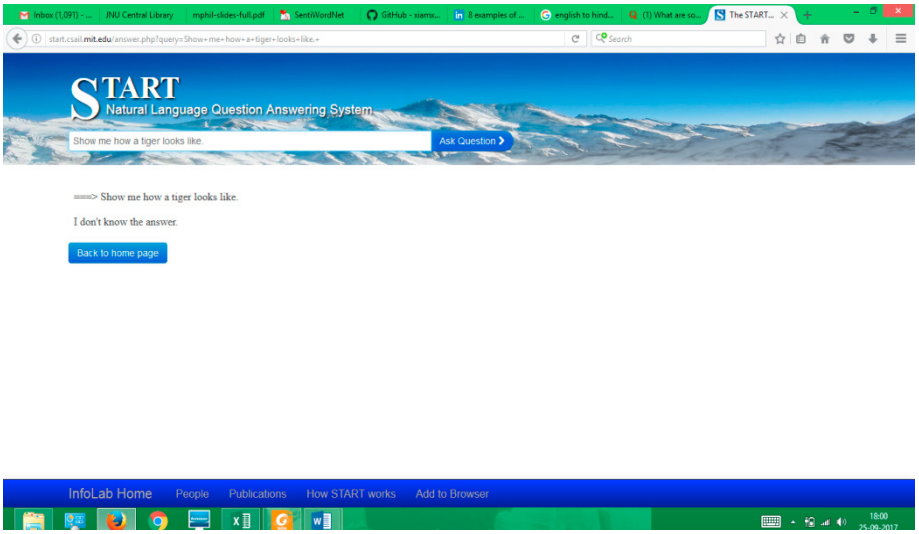

Figure 3: Result for the Query: "Show me how a tiger looks like"
(MIT Natural Language Question Answering System)

Here, *"looks like"* has not been correctly reasoned by the system and thus shows incorrect result ("I don't know the answer").

b. ***Google search engine***

Another evident tool which is synonymous to searching "almost anything" is Google Search Engine. In this time, ***Google search engine*** offers the most powerful search capability. It has been observed even the most powerful search engine is unable to fully understand the uncertainty in the search query (input from the user in natural language). The following search examples were carried out on Google Search Engine to illustrate the poor reasoning of the system (refer to Figures 4-7).

Query 3: what is the age of narendramodi (proper result)
Query 4: for how many years has narendramodi been a live (inappropriate result)
Query 5: tallest man in the world (proper result)
Query 6: a man who has the most height in the world (inappropriate result)

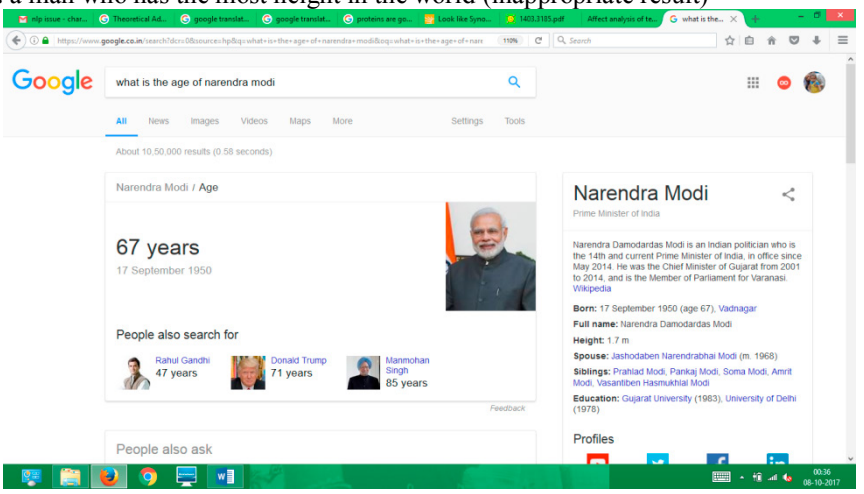

Figure 4: Correct Result retrieved from Google Search Engine for Query 3

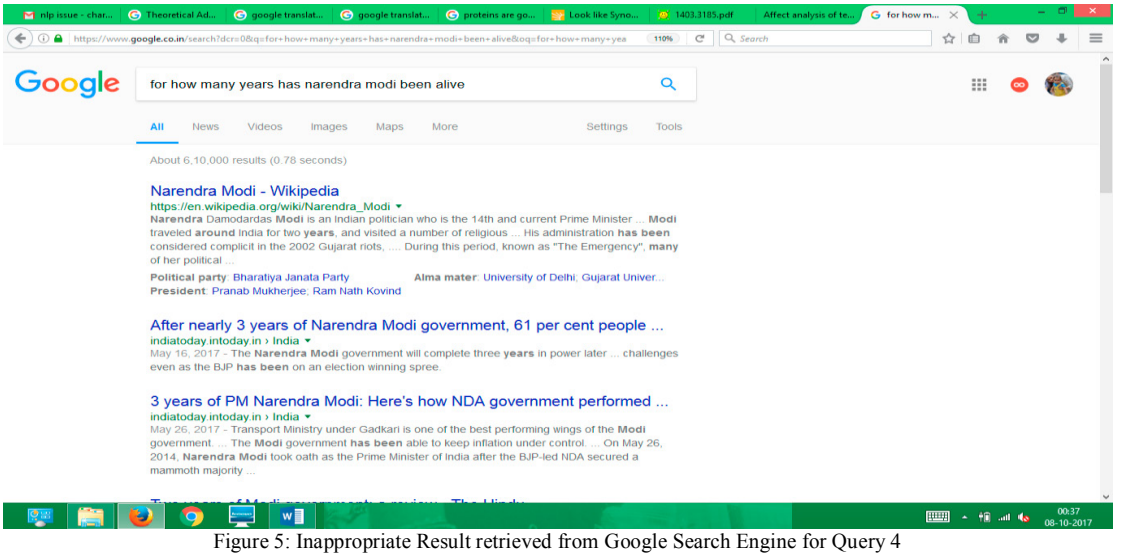

Figure 5: Inappropriate Result retrieved from Google Search Engine for Query 4

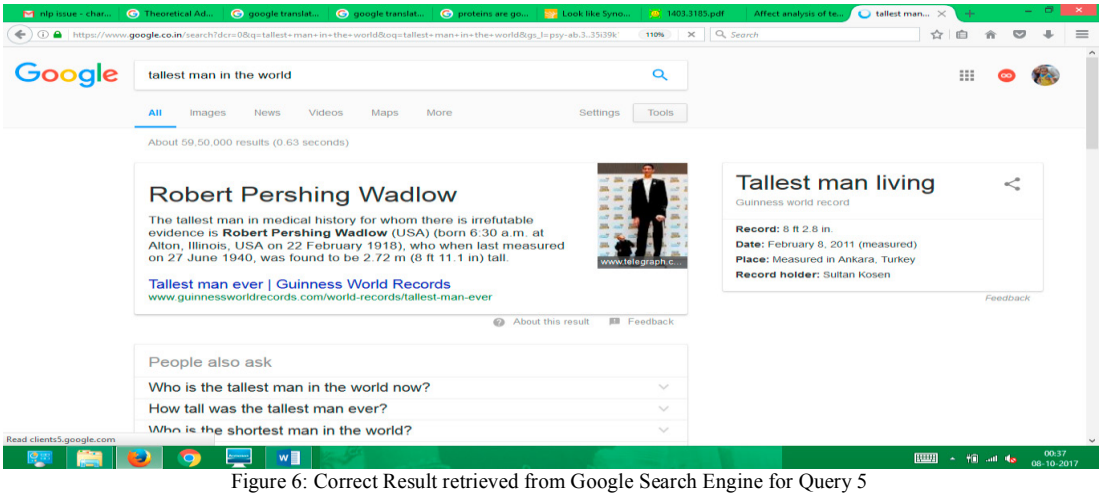

Figure 6: Correct Result retrieved from Google Search Engine for Query 5

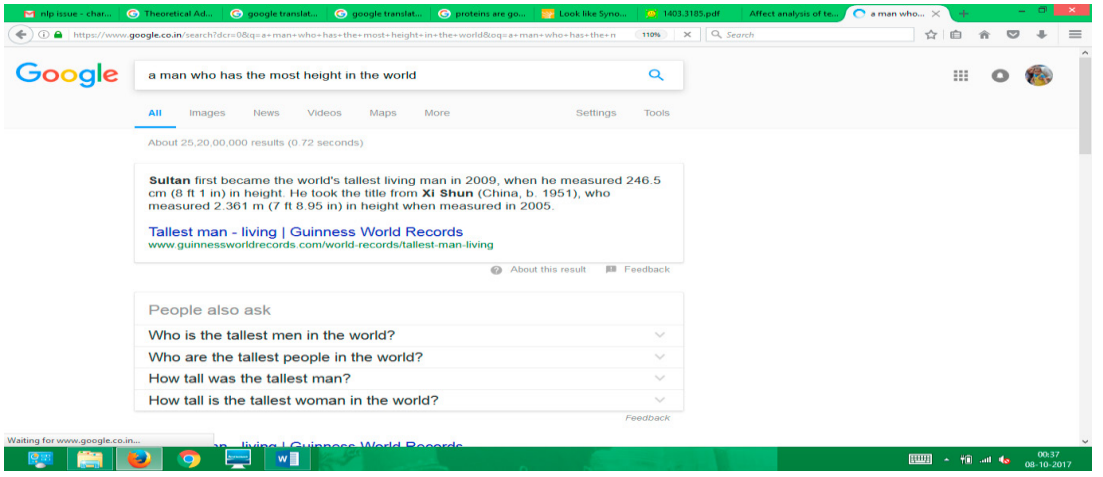

Figure 7: Inappropriate Result retrieved from Google Search Engine for Query 6

Among Query 3-6, the only difference is in the presentation of the sense of the query. The uncertain nature of the queries in 4 and 6, confuse the system and thus it results in inappropriate results. Another illustrative example is of **Sentiment Analysis** with **Google Search Engine** (refer to Figure 8 and 9). Consider the following set of Queries.

Query 7: good budget restaurants
Query 8: suggest restaurants which are not heavy on pocket

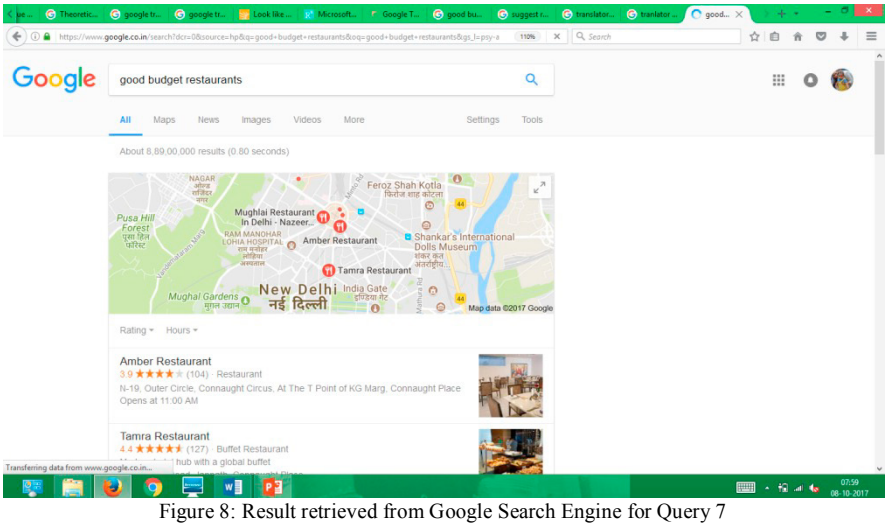

Figure 8: Result retrieved from Google Search Engine for Query 7

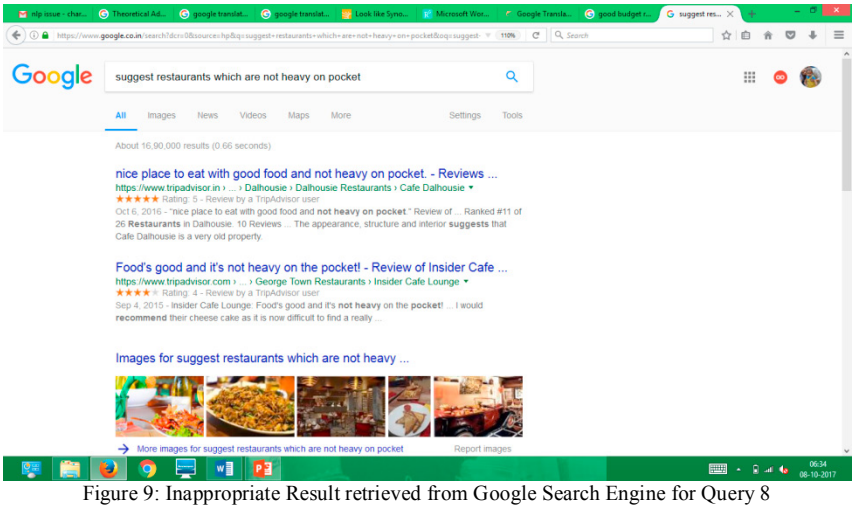

Figure 9: Inappropriate Result retrieved from Google Search Engine for Query 8

c. ***Google Machine Translation system***

Let us further understand the poor effect caused by ***Google Machine Translation system for English to Spanish***. Figures 10-11 show the translation inadequacies of the Queries (Query 9-10)

Query 9: (In English) there is a slight difference between warm and hot
Query 10: (In Spanish) esunaligeradiferencia entre caliente y caliente (is a slight difference between hot and hot)

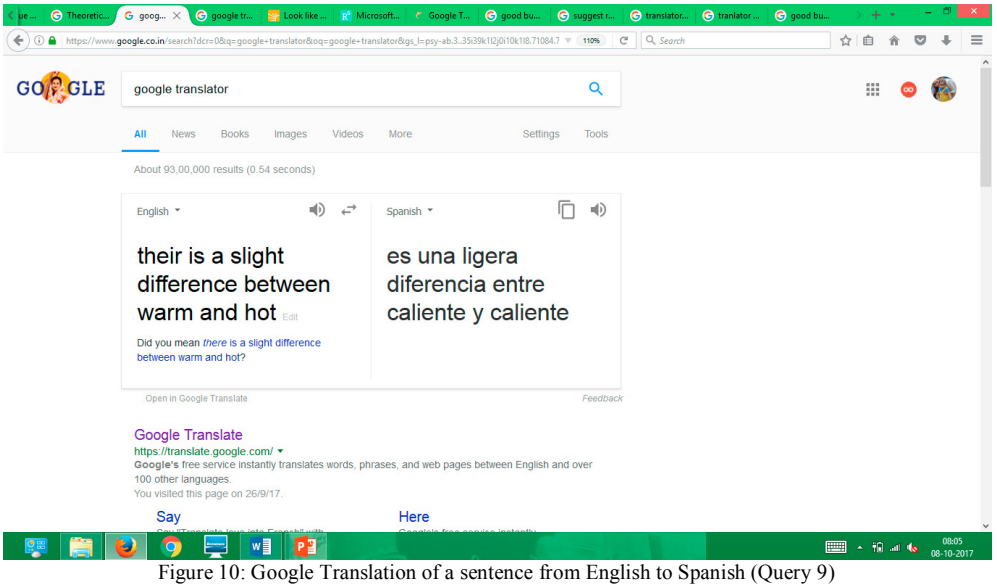

Figure 10: Google Translation of a sentence from English to Spanish (Query 9)

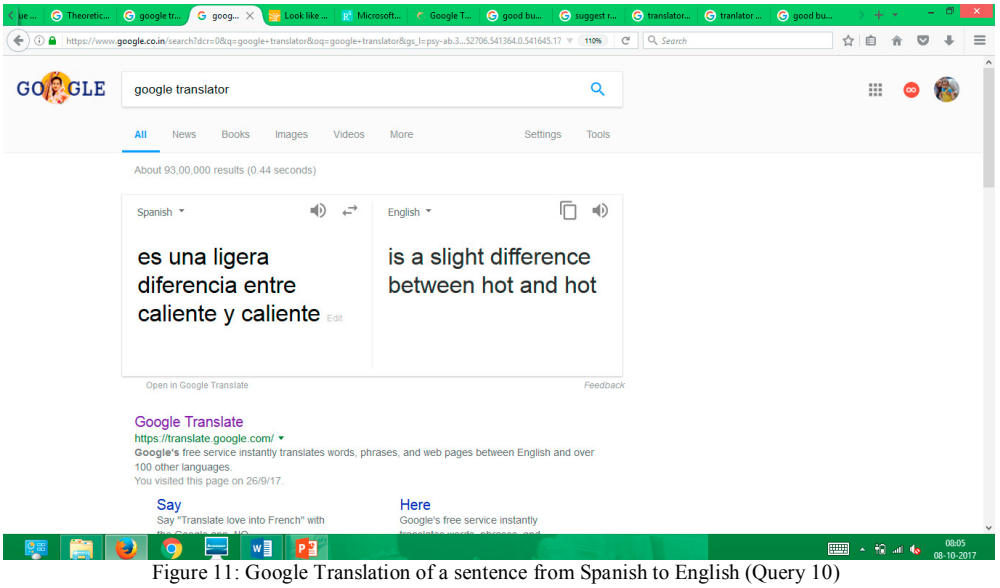

Figure 11: Google Translation of a sentence from Spanish to English (Query 10)

Further to illustrate the inability of Google Translator reasoning we consider query sentences involving **word sense disambiguation and named entity recognition.** They are also poorly translated by the Google Translator. Figure 12 and 13shows the results for Queries 11 and 12.

Query 11: (In English) Washington voted Washington to power
Query 12: (In Spanish) Washington votó a Washington al poder (Washington voted to power)

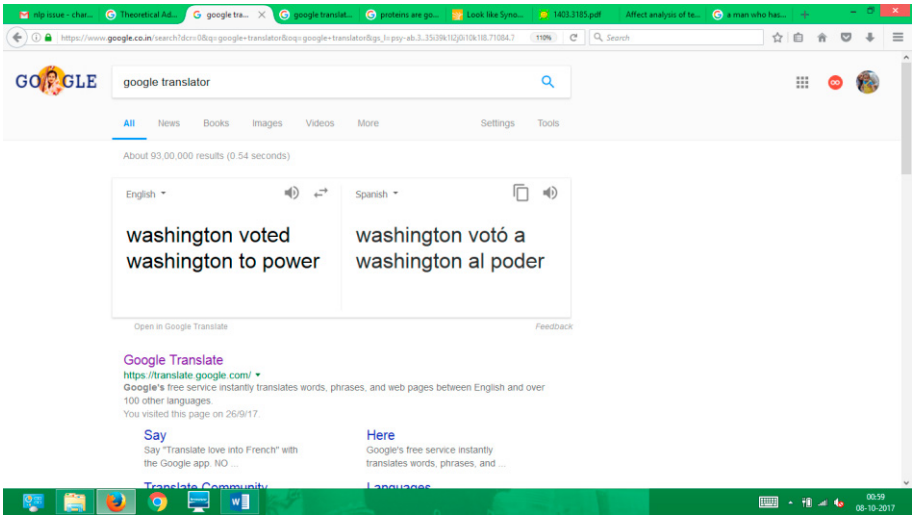

Figure 12: Google Translation of a sentence from English to Spanish (Query 13)

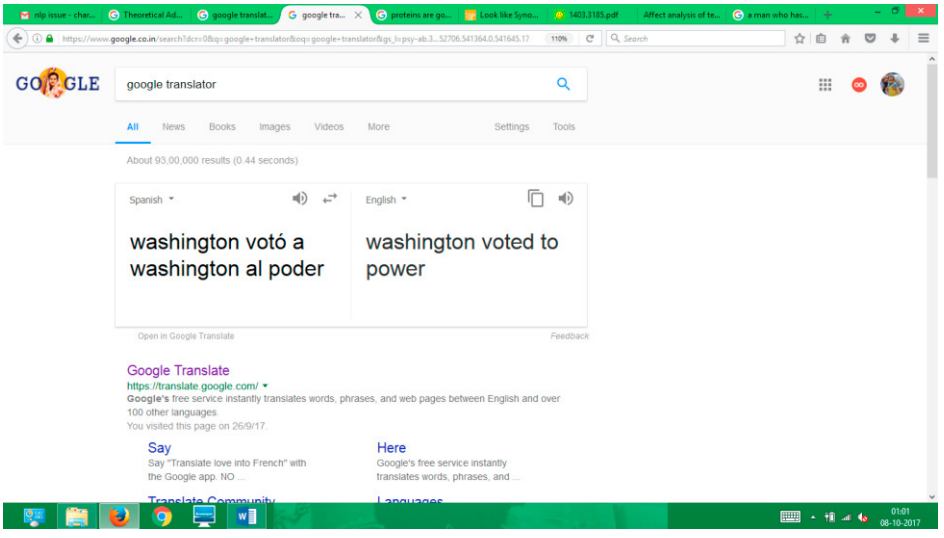

Figure 13: Google Translation of a sentence from Spanish to English (Query 14)

The incorrect translation results from the basic inadequacy of a computer system to process words which are best interpreted by common sense reasoning (involves world knowledge, domain knowledge, constraints). **All the examples stated above raise a need to understand the (human) contextual capability and uncertainty in the language by a system.**

These examples **inspire us to explore** the use of fuzzy logic in understanding the text in a better way. In the above examples, words such as *little, some, few, most, how many looks like* have not been interpreted correctly. This creates a gap between the **expected processing** of natural language and the **output of the processed text.**

5. Conclusion

Today a computer can translate a sentence but not proofread it. The idea is to understand that the out of the three things (computation, memory and common sense) that a human uses efficiently; a computer can inherently work on only two i.e. computation and memory. Fuzzy logic gives reasonable, sensible answers to questions which involve

common sense reasoning of a system. Nowadays, Fuzzy Systems has become one of the most successful technologies for sophisticated control and reasoning/inference systems. The reason for its success lies in the fact that Fuzzy Systems closely resembles human decision making with the ability to solve problems and generate solutions which are precise. The current state-of-the art methods which available in the literature for translation and retrieval focus more on designing a rule basedof the language. So there is a strong need to build a natural language system equipped with common sense (functionality).

It is also important that developed fuzzy models are transparent and easily adaptable. It is so because it infers acceptably good solutions from approximate information, by reducing the computational complexity tremendously, as compared to classical approaches. Further, fuzzy approaches use the sub-symbolic information available in the universe of discourse providing the context to the given model and solution adds additional advantage.

The idea of conjoining natural language processing with Fuzzy Logic can become a good platform to bring quantifiable change to the scientific output of natural language processing system. Systems thus developed using fuzzy logic can be used for the welfare of society and mankind.

References

[1] U.S. Tiwari and T. Siddiqui, Natural Language Processing and Information Retrieval, Oxford University Press, 2008.

[2] C.D. Manning and H. Schutze, Foundations of Statistical Natural Language Processing, The MIT Press.

[3]P. Bhattacharya, Notes on Natural Language Processing, “http://nptel.ac.in/courses/106101007/2” [last accessed on: 1-2-17]

[4]The University of Sheffield, Natural Language Processing Research Group, “http://nlp.shef.ac.uk/”[last accessed on: 23-4-17]

[5] New Jersey institute of Technology, Web Mining and its Latest Trends, “http://graduatedegrees.online.njit.edu/mba-resources/mba-articles/web-mining-and-its-latest-developments/”[last accessed on: 27-9-17]

[6] Zadeh L.A. (1975) The concept of linguistic variable and its application to an approximate reasoning. Information Sciences, 8, 199-249.

[7] Zadeh L.A. (1965). Probability measures of fuzzy events. Journal of

mathematical Analyses and Applications, vol. 23, no. 2, pp. 421-427.

[8] Zadeh L.A. (1965). Fuzzy sets. Information and Control, 8, 338–353.

[9] Klir G.J., Folger T.A.: Fuzzy Sets, Uncertainty, and Information: Prentice-Hall, NJ, 1988.

[10] Yager, R. R., &Zadeh, L. A. (Eds.). (2012). An introduction to fuzzy logic applications in intelligent systems (Vol. 165). Springer Science & Business Media.

[11]Acampora, G., &Loia, V. (2005). Fuzzy control interoperability and scalability for adaptive domotic framework. *IEEE Transactions on Industrial Informatics*, *1*(2), 97-111.

[12]Erdinc, O., Vural, B., Uzunoglu, M., &Ates, Y. (2009). Modeling and analysis of an FC/UC hybrid vehicular power system using a wavelet-fuzzy logic based load sharing and control algorithm. *international journal of hydrogen energy*, *34*(12), 5223-5233.

[13]Lin, C., & Hsieh, P. J. (2004). A fuzzy decision support system for strategic portfolio management. *Decision support systems*, *38*(3), 383-398.

[14]Draeseke, R., & Giles, D. E. (2002). A fuzzy logic approach to modelling the New Zealand underground economy. *Mathematics and Computers in Simulation*, *59*(1-3), 115-123.

[15]Phuong, N. H., &Kreinovich, V. (2001). Fuzzy logic and its applications in medicine. *International journal of medical informatics*, *62*(2-3), 165-173.

[16]https://scholar.google.co.in/ [Last accessed on: 5-1-18]

[17]https://www.google.co.in/?gfe_rd=cr&dcr=0&ei=d8Z4WpvkLK6cX8eVglg [Last accessed on: 5-1-18]

[18] https://translate.google.com/[Last accessed on: 5-1-18]

[19] http://start.csail.mit.edu/index.php[Last accessed on: 5-1-18]