

Automated Answering System for a Product Using Opinion Mining

¹Varun Mohan, ²Dheeraj Varma, ³Yogesh Tathe, ⁴Rahul Sinnarkar
^{1,2,3,4} Students

^{1,2,3,4} Department of Computer Engineering,
^{1,2,3,4} Maharashtra Institute of Technology, Pune

Abstract—Though computers are considered as machines, they are expected to give answers to questions in normal English language, just the way human beings can do. To train a computer to answer English language questions is an interesting and challenging problem. These problems are handled under two categories: open Domain problems and close domain problems. In this paper a question answering scheme is proposed for close domain systems. The question answering system for closed domain deals with questions under specific domain. Answers to questions from close domain cannot be searched using a search engine. During a close domain, answers to questions are not alive in the public area and so they cannot be search using a search engine. Hence, the answers to such questions are maintained in a database by the domain expert. During retrieval of answers, the best matched answer searched from database is returned to the user. To perform this matching a template matching technique is used. Thus the main challenge of the proposed scheme is to understand the English language question to get the correct answer from the database.

Index Terms—Answering Machine, Closed Domain Problem, Natural Language Processing, Template Matching.

I. INTRODUCTION

The interaction between computer and human being is always interesting and challenging. Due to increased use of computers, question answering systems for computers are developed to answer the questions asked in natural English language. Such mechanisms allow users to ask questions in a natural language and give a concise and accurate answer. Understanding user questions in natural language requires some pre-processing on questions. This pre-processing is carried out using Natural Language Processing (NLP). Natural language processing is the computerized approach to analyzing text based on both a set of theories and a set of technologies. The goal of the Natural Language Processing is to design and build methods that will analyze, understand, and generate languages that humans use naturally. With the use of NLP, it is possible to address the computer in the simpler way which the computer can understand. Natural language processing is the computerized approach to analyzing text based on both a set of theories and a set of technologies. In this paper, an automated answering system is proposed for closed domain system which accepts question in normal English language and returns appropriate answer in form of output.

Developing mechanisms for using computers to answer user questions is becoming an interesting problem with the increased use of computers. Such mechanisms allow users to ask questions in a natural language and give a concise and accurate answer. Understanding user questions in natural languages requires Natural Language Processing (NLP). Being an active area of research, NLP plays a big role in the ICT and Question Answering (QA) systems. Natural language processing is the computerized approach to analyzing text based on both a set of theories and a set of technologies. It will become important to be able to ask queries and obtain answers, using natural language (NL) expressions, rather than the keyword based retrieval mechanisms. The QA system can better satisfy the needs of users as they will provide an accurate, quicker, convenient and effective way of giving answers to user questions.

The approach we have adopted in this project is an automated FAQ (Frequently Asked Question) answering system that replies with pre-stored answers to user questions asked in ordinary English, rather than keyword or syntax based retrieval mechanisms. This is achieved using a template matching technique with some other mechanisms like disemvoweling, matching synonyms, etc. The natural language processing technique developed for FAQ retrieval does not analyze user queries. Instead analysis is applied to FAQs in the database. Thus, the work of FAQ retrieval is reduced to keyword matching creating an illusion of intelligence. The system is both evolving and portable. Evolving because its question answering ability improves as more questions are asked and new FAQ entries are created. It is portable because the system could be used for any problem domain (closed) by changing the knowledge base.

II. PREVIOUS WORK DONE

In [1] contains drawbacks at different sentiment level. This paper contents description about the challenges and techniques of opinion mining. Following table show some points related to this.

Document level classification	It does not give details of what people likes or dislikes because writer comments only the specific aspects of product.
	It is not applicable on forums and blogs as they contain only few opinionated sentences on features of object.
	Document level opinion mining defines the polarity of document. But a positive phrase dose not indicates that the user likes everything and similarly a negative phrase does not indicate that the opinion holder dislikes everything. Just imagine for a moment, if a document containing review in which holder has likes the story of movie but dislikes its sound and print. The overall sentiment of the document is negative but the holder still likes the movie. So such kind of reviews shows the wrong classification. Similarly if some user dislikes the movie but likes everything else, again the review will be classified as positive due to the average orientation of the positive phrases
Sentence level classification	A user can express different views in a single sentence. If a user expresses his likeness of picture quality and dislike the sound of the movie, the review will be ranked as neutral at sentence level. As such kind of the sentences the average orientation of positive and negative phrases will be equal and one cannot find out what user wants to convey.
	Similarly if user expresses his opinion about the likeness of the movie in just one sentence and the rest of the sentences are expressing the user dislike some feature then the classification of the document could be wrong and will create negative impact.
NLP	A drawback of the NLP approach is that it could really cut very badly if they are used grammatically incorrect text. Currently a large part of the web based sentiment data fall into this category, methods to detect and correct bad English, if any, would be necessary before using them on a larger scale.

Table .1 drawbacks at different sentiment level

In [2] an important part of their information-gathering behaviour has always been to find out what other people think. The sudden eruption of activity in the area of opinion mining and sentiment analysis, which deals with the computational treatment of opinion, sentiment, and subjectivity in text, has thus occurred at least in part as a direct response to the surge of interest in new systems that deal directly with opinions as a first-class object. This survey covers techniques and approaches that promise to directly enable opinion-oriented information seeking systems. Their focus is on methods that seek to address the new challenges raised by sentiment aware applications, as compared to those that are already present in more traditional fact-based analysis. They include material on summarization of evaluative text and on broader issues regarding privacy, manipulation, and economic impact that the development of opinion-oriented information-access services gives rise to. To facilitate future work, a discussion of available resources, benchmark datasets, and evaluation campaigns is also provided.

In [3] the development of the Web 2.0 led to the birth of new textual genres such as blogs, reviews or forum entries. The increasing number of such texts and the highly diverse topics they discuss make blogs a rich source for analysis. This paper presents a comparative study on open domain and opinion QA systems. A collection of opinion and mixed fact-opinion questions in English is defined and two Question Answering systems are employed to retrieve the answers to these queries. The first one is generic, while the second is specific for emotions. We comparatively evaluate and analyze the systems' results, concluding that opinion Question Answering requires the use of specific resources and methods.

III. PROPOSED SYSTEM ARCHITECTURE

In this section we depict the working of our system. The system can be subdivided into three main modules: (1) preprocessing (2) question template matching, and (3) answering. Figure shows the system architecture of the question and answering system. Each module is described in detail in the following subsections.

In this project the following modules are described:

- 1) Preprocessing.
- 2) Question template matching
- 3) Answering

1. Pre-Processing Module

Pre-processing module mainly consists of three operations:

- (1) perform stemming.
- (2) pos tagging.
- (3) stopword removal and
- (4) query intention.

First step here is to perform stemming which reduces the word to its stem or root. Example by performing stemming the words like "fishing", "fished" and "fisher" are reduced to their root word i.e fish. Next step in this module is pos tagging here the words

are divided according to their tags which we have in tagging. Stop words are the words that add no effect to the meaning of a sentence even if they are removed. Removing stop words is done to increase the effectiveness of the system by saving time and disk space. Examples of stop words are the, a, and, etc. In the final step here we take into account the query intention if it is like a positive intent or a negative one.

2. Question-Template Matching Module

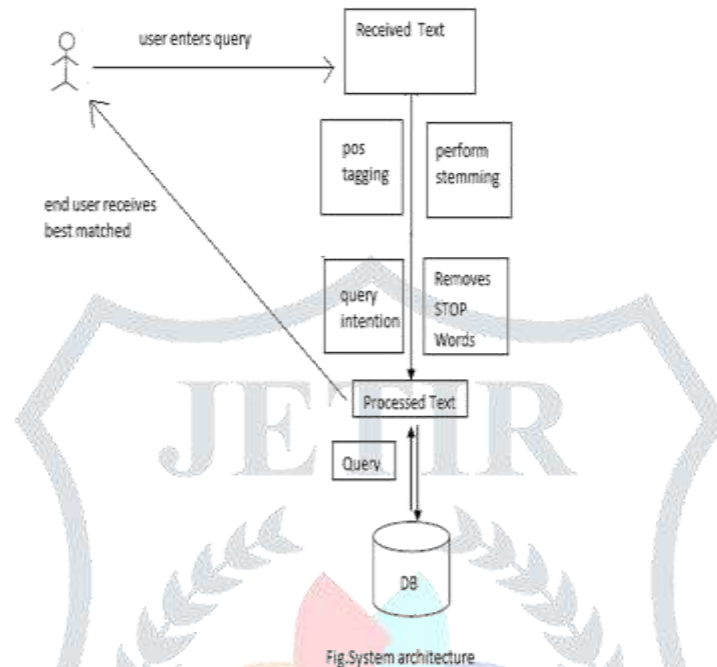


Fig .1 System Architecture

The pre-processed text is matched against each and every pre stored template until it finds the best matched template with the received text. In order to do this, templates are created according to a specific syntax. Further in this module, words that are considered to have synonyms are referred in a synonym file. This synonym file can be modified according to the relevant domain and are updated from a standard database such as WordNet. It is worth noting that the templates here are for questions and not for answers. The main target of this system is to identify the closest template that matches the question we have received from the user.

3. Answering Module

Since each and every template representing a question are pre stored in a database with its answer, just when the best matched template for the question is found, the corresponding answer will be returned to the end user. If the question is not found in the pre stored database then it is stored in the database for further use.

IV. CONCLUSION

We conclude that our system is user friendly automatic answering system with the ability of detecting and answering questions asked in English. This system allows the user to ask their respective queries about a specific product using their computers. These systems also have the ability to overcome the mismatches we might encounter due to spelling mistakes. This system will do processing on the queries asked by the user and after searching the answer it will provide the appropriate answer to the user. We propose to explore the shortcomings as future work.

REFERENCES

- [1] A Survey Paper: Areas, Techniques and Challenges of Opinion Mining, IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 6, No 2, November 2013 ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784.[2] Opinion mining and Sentiment analysis, Foundations and Trends in Information Retrieval Vol. 2, No 1-2 (2008) 1–135c 2008 Bo Pang and Lillian Lee. This is a pre-publication version; there are formatting and potentially small wording differences from the final version.
- [3] B. Pang, and L. Lee, Opinion mining and sentiment analysis. Foundations and Trends R_In Information Retrieval Vol. 2, Nos. 1–2 (2008) 1–135, 2008.
- [4] Q. Su, X. Xu1, H. Guo, Z. Guo, X. Wu, X. Zhang, B. Swen1and Z. Su “Hidden Sentiment Association in Chinese Web” Proceedings of the 17th international conference on World Wide Web Pages 959-968 in 2008.

- [5] T. Yao and L. Li “A Kernel-based Sentiment Classification Approach for Chinese Sentences” World Congress on Computer Science and Information Engineering in March 31 2009-April 2 2009.
- [6] D Wang, S Zhu, and T Li. “SumView: A Web-based engine for summarizing product reviews and customer opinions” Expert Syst. Appl. 40(1):27-33 (2013).
- [7] T. Khushboo “Mining of Sentence Level Opinion Using Supervised Term Weighted Approach of Naïve Bayesian Algorithm” ,Int. Journal. Computer Technology & Applications, Vol 3 IJCTA | MAY-JUNE 2012.
- [8] J. Han and M. Kamber, “Data Mining: Concepts and Techniques”, third edition Publication Date: July 6, 2011.[9] N, Anwer and A, Rashid “Feature Based Opinion Mining of Online Free Format Customer Reviews Using Frequency Distribution and Bayesian Statistics” Networked Computing and Advanced Information Management (NCM), 2010 Sixth International Conference on 16-18 Aug. 2010.
- [10] B. Seerat and F.Azam “Opinion Mining: Issues and Challenges (A survey)” International Journal of Computer Applications (0975 – 8887) Volume 49– No.9, July 2012.
- [11] D. S. Deshpande “ A Survey on Web Data Mining Applications” Emerging Trends in Computer Science and Information Technology - 012(ETCSIT2012) Proceedings published in International Journal of Computer Applications® (IJCA).
- [12] N. M. Shelke, S. Deshpande and V. Thakre “Survey of Techniques for Opinion Mining” International Journal of Computer Applications (0975 – 8887) Volume 57– No.13, November 2012.

