Survey Paper on Discovering of Tasks using Natural Language Processing

#1Prachi Rayate, #2D.M.Thakore

1prachirayate@gmail.com
2dmthakore@bvucoep.edu.in

#1M.Tech, Computer, B.V.D.U.C.O.E, Pune, India - 411043.
#2H.O.D, Computer, BV.D.U.C.O.E,Pune, India - 411043.

ABSTRACT

Product knowledge and software requirement plays an important role in software documentation. They are essential for software developers to build quality product. The developers are unable to read whole documentation of large corpus within maintenance. They need to receive a software documentation i.e. (development, designing and testing etc.) in a short time. An important documents are able to record in software documentation using customize portal. There live a space between information which developer wants and software documentation. To solve this issue, the main idea is to automatically extract task from large corpus task into different phases of documentation.   In this case, we design and implement extraction technique includes natural language processing, statistical analysis, text mining and some HTML files of text. This approach is an object oriented creates a new framework for text mining based on wed development application of open projects.

Keywords: Natural language processing, text mining, software documentation, HTML files, development tasks etc.

ARTICLE INFO

I. INTRODUCTION

"Natural language processing" refers to the use and ability of systems to process sentences in a natural language such as English. Software documentation has many forms which is captured and wanted by software developers [6]. There exists a space between information which a developer wants and software documentation. This problem attempt to solve by many organizations of software development and open source projects by creating web pages which generates very useful information. Sometimes developers won't be able to read the whole documentation of large systems during maintenance. Once read the whole documents, it does not tell enough information about its motivation and definition. By using statistical techniques, automatically of task structure is generally not able to done because the whole documentation consist of large information. The search engine is inadequate as per the users requirement that what they need and unable to express techniques. Most of the web search engines follow a software function that completes words to reduce the technique gap [4] and it has presented

users feedback with high fulfillment. Query logs are not available and to learn suitable model for customized search systems, the number of previous queries is so small [3]. N-grams are used to populate auto-complete field by researches; which is extracted from corpus.The extraction technique combines natural language processing (NLP) techniques, classification methods, text mining and the analysis of syntactical features of the text. There is no any need of machine learning techniques for task extraction. This approach requires text mining to summarizes text input and reduce redundancy with NLP. NLP will filter each word into a proper way and then whole documentation will analyze and then categorize into different software documentation phases (i.e. documentation, development and testing etc.).

II. EXISTING SYSTEM

This technique is used for discovering of task from customize system using different techniques. The tasks have been mentioned in software documentation for extraction purpose.Here task can be extracted from documentation and can also reduce a live space between information which developer wants and software documentation [1]. Syntactic analysis of natural language to sentences constructed from the terms which exist in program element identifiers to automatically extract ontology by mapping linguistic entities. And also shows parsing can be represented as a dependency tree [5]. The model information is used for selecting and ordering most relevant questions, selecting the most relevant answer and estimating the FAQs quality. In natural language documents, extraction mainly assigning specific form of data to discover relevant task. The identifier split into word forms and then tokenizes each name for automated analysis of task [2]. The product knowledge which is carried out in various ways of software documentation[6].The right information which developers want for making software documentations are not get easily [7]. To solve all these above problem, our main intension is to get easily information by developers whenever they requires.This approach is an object oriented creates a new framework for text mining based on the Knowledge Discovery from Databases (KDD).

III.SYSTEM DESIGN

After studying literature survey, these papers works modified and then trying to build a customize application that will give us best result like accuracy in terms of frequency, categorization of task. This approach can be useful to improve the precision of tasks of software documentation. So that it will become easy to reduce a live space between information which a developer wants and software documentation. Our main idea is to extract relevant task from software documentation by applying different techniques and algorithms.   Text Mining usually involves the process of structuring the input text deriving patterns within the structured data, and finally evaluation and interpretation of the data. After completion of this process, Natural Language Processing (NLP) will apply. NLP will apply in such a way that the task will be filtering and structuring with HTML files. HLML files will then further transfer into text files based on pre-processing. The extracted task will further then divide into different phases of software engineering (i.e. documentation, development and testing etc.). And finally we will calculate the accuracy of the extracted task with frequently asking questions to developers gives a proper feedback.

OUR CONTRIBUTION

Task extraction technique is mainly focus on software documentation of large corpus. In this case, our mainly idea is to discover relevant task from customize portal using NLP. This application is basically useful for online reviews. Important documents are able to record in software documentation. There live a space between information which developer wants and software documentation. To solve this issue, the main idea is to automatically extract task from large corpus of documentation and then categorize each task into different phases of documentation. This process is known as task extraction. We develop this technique for developer's documentation by which developers will get the required data from software documentation easily with customize portal. The task that have been described in the software documentation. This approach is mainly useful for standalone application develop using customize portal. Like google snippet, this would be beneficial for online reviews.

As we store the data into database we can use the data whenever we want it for getting task, for software documentation, for preparing KDD and other purposes as well. This work is mainly focus on software documentation to extract some task from large corpus through word net library for online reviews application. For this purpose, our approach is to take some data source file as a input and then apply text mining on that data. We are assigning NLP to each word describing in a document. That document will send us the extracted task and frequency of each word. Using this data received from the user we will calculate the real time results. We will be developing a documentation to categorize extracted task.

IV.CONCLUSION

This research is undertaken to extract relevant task from software documentation based on customize portal for online reviews. In this concept, NLP is used to get the required data and also used text mining, statistical techniques for data, sentiment analysis method and some more algorithms to calculate the results.

REFERENCES

- Christoph Treude, Martin P. Robillard, and Barth_el_emy Dagenais ,"Extracting Development Task To Navigate Software Documentation" in Proc, IEEE Soft,Vol.41 No.6,2015,pp,565-581
- S. Gupta, S. Malik, L. Pollock, and K. Vijay-Shanker, "Part-ofspeech tagging of program identifiers for improved text-based software engineering tools," in Proc. 21st IEEE Int. Conf. Program Comprehension, 2013, pp. 3–12.
- M. Barouni-Ebrahimi and A. A. Ghorbani, "On query completion in web search engines based on query stream mining," in Proc. IEEE/WIC/ACM Int. Conf. Web Intell., 2007, pp. 317–320.
- P. Mika, E. Meij, and H. Zaragoza, "Investigating the semantic gap through query log analysis," in Proc. 8th Int. Semantic Web Conf., 2009, pp. 441–455.
- S.L.Abebe and P.Tonella,"Natural language parsing of program element names for concept extraction," in Proc. 18th IEEE Int. Conf. Program Comprehension, 2010, pp. 156–159.
- C. Treude and M.-A. Storey, "Effective communication of software development knowledge through community portals," in Proc. 8th Joint Meet. Eur. Soft. Eng. Conf. ACM SIGSOFT Symp. Found. Soft. Eng., 2011, pp. 91–101.

- T. C. Lethbridge, J. Singer, and A. Forward, "How software engineers use documentation: The state of the practice," IEEE Soft., vol. 20, no. 6, pp. 35–39, Nov./Dec. 2003.