A SURVEY PAPER ON NEW QA SYSTEM THAT PREDICTS A USER'S FUTURE INTEREST AS NEXT QUESTION

Karthik S[1], Hamsaveni.M[2]
[1]M.Tech student, [2]Assistant Professor
Dept. of Computer Science & Engineering, Vidyavaradhaka College of Engineering, Mysuru.

Abstract: This paper proposes, a new form of QA system, The Question answering is a new form of data retrieval which gives the knowledge. It allows the users to enter the question and waits for the answer as response, then the system will predict the next question based on the user's past searched questions in the QA system. In this system, the users current interactions is maintained in a question log, from the log the user sessions get extracted ,from the user sessions the questions are stored in databases. Based on the user sessions the system gives the next question on the user's future interest. For the prediction the system makes use of a datamining technique called "Association rule''.

Keywords: Question lo; datamining; association rule; user sessions;

I. INTRODUCTION

The QA system provides answers to the users question rather than retrieval the whole document. The extraction of information from the question log, this procedure is called as "Log analysis"[1]. During the interactions the user search a set of queries that specify the information about what they are interested in. this interactions is maintained in a specified log. The main aim of doing log analysis is to extract the information for the prediction .the data mining can be applied to extract the information from the log for the prediction of users future interest. Mining processes can be applied to large search query logs in order to extract knowledge about user interests[4]. The data mining technique called association rule mining is concerned with "maintaining the data bases, it consists of queries, relations, transactions, associations, co-relations among the item sets on the basis of previous transactions". This information provides how item sets are related to each other and how they tend to group together. This is in particular a necessary step for the design of true usercentric applications in which user search behaviors are identified and taken into account. Providing related queries for search engine users can help them quickly find the desired content. Recently, some search engines started showing related search keywords in the bottom of the result page. Their main purpose is to give search engine users a comprehensive recommendation when they search using a specific query. Recommending the most relevant search keywords set to users not only enhances the search engine's hit rate, but also helps the user to find the desired information more quickly [2]. Question Classification is technique used to extract useful information from the question by identifying its class. Then, to provide user with relevant set of answers, the appropriate answer type needs to be identified on the basis of user's expectation. If the user asks "Who is the First Prime Minister of India?", the user expects "Pandit Jawaharlal Nehru" as the answer which is the name of a person[6]. For this, the question class "Who" is mapped to the expected answer type i.e. "Person".Fig.1 shows the architecture of a general Question answering system.

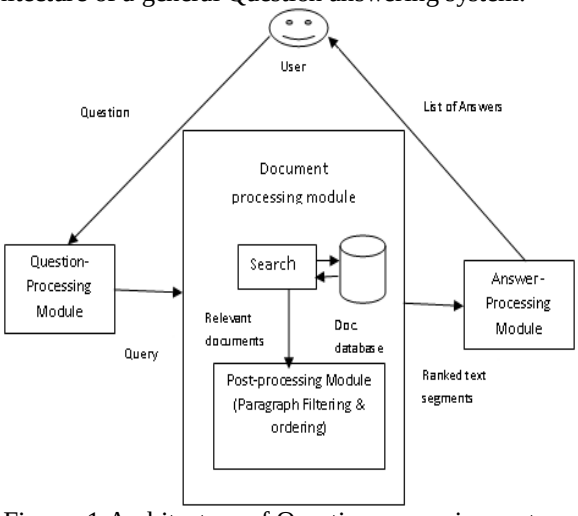

Figure. 1:Architecture of Question answering system.

II. RELATED WORKS

A. Question log analysis.

The user submits his queries on the interface of the QA system. The user asks question from the QA system by entering a question on the QA interface. This question is classified by the question classifier module. The module classifies the questions according to their Question type and then converts the rest of the question into query.[3] For the query a search is done by the searcher module.

B. User session extraction:

After storing thedate, time, question in the question log ,The next step is to extract the user sessions, By considering t as the session length ,the user sessions are extracted from the question log for a time unit t. The main aim ofextracting the user sessions is to extract the questions that have been fired within a specific time of t units.

C. Query Extraction:

After the user sessions are extracted, from the user sessions questions have been extracted. For this purpose, this module uses the question log.

D. Query Filter module:
This module takes the input as question which is extracted by the question extraction and then seperates the questions on the basis of their question type. A separate database is maintained for the each question type.

E. Query preprocessor:
The main aim of these step is to retrieve the query terms that form a set of elements .there are two levels of query preprocessing system: The query level and terms level.
The query processing is nothing but a filtering process; first some queries resulting from the format of the logs are removed. Next one is te "bad queries" are filtered out, that means by a bad query a non-interpretable. In the term level, can apply a lexica queries, each query is divided into a set of tokenizer. The separated questions in the query filter are stored in a database and then these questions are sent to the query preprocessor. This module divides the question type from question and converts the remaining part of the question into query. A Query is a set of terms contained in the question.

F. Association rules generation:
The main aim of using Association rules mining is to identify all the rules in a market basket data analysis. This type of data also referred to as a transaction data. This approach is basically used to analyze the purchase of items by customers in a shop or a supermarket and how these are related to one another. The set of items a customer buys is referred to as an item set, and market basket analysis seeks to find relationship between purchases. For each customer transaction, one data record is generated. The association rule generation achieved from a set F of frequent item sets in an extraction context D[1]. For example the probability that a customer will by beer without a bar meal is referred to as the support for the rule. The conditional probability that a customer will purchase crisps is referred to as the confidence.

E. Next question predictor:
This module takes as the input from the association rules that have been generated by Association Rule Generation module. The Next question predictor which generates the next question on the basis of the association rules that have been generated by the current module and predict the next Question by the association rules generated [7]. A predicted questions database is maintained PQDi for each QuDi. Each PQDi contains QTi and a set of predicted questions. For example, if the user is looking for "what is testing", then the module predicts that he or she may be interested in "what is testing in software". Similarly, if the user is looking for "how to play cricket" then he or she may be interested in "how to play basketball and baseball". This module gives the predicted questions in PQDi as input to the Question classifier which performs its task and then provides the input to the searcher that searches for each predicted query in the Question classification based index and maintains the search results as answers for later reference[1]. In future, if the user enters a question that matches with any of the predicted questions in PQuDi, then the system provides the stored answers to the user. This increases the efficiency of the QA system and provides a speedy response to the user.

III. CONCLUSION

This paper proposes a new form of Question answering system. This approach is used to retrieve the answers based on the user's future interest as a next question. This system uses a concept called association rule mining for prediction of next question based on the current interaction with the system. The next question prediction system that predicts the users' next requests based on their current interactions with the system from the search query logs. The technique of Association rule discovery is found as one of the most important techniques in the field of data mining.

REFERENCES

[1] Rosy Madan, A.K Sharma Ashutosh dixit, "A Data mining approach to predict user's next question in QA system" ,2015.
[2] L. Limam, D. Coquil, H. Kosch, and L. Brunie, "Extractinguser interests from search query logs: A clustering approach," DEXA '10 Proceedings of the 2010 Workshops on Databaseand Expert Systems Applications, 2010.
[3] Z. Zhang, and O. Nasraoui, "Mining search engine query logsfor query recommendation," in Proceedings of the 15thinternational conference on World Wide Web, pp. 1039–1040, 2006
[4] Tan, Steinbach, Kumar, "Introduction to Data Mining,"LectureNotes,2004
[5] G. Dupret, and M. Mendoza, "Recommending Better Queries Based on Click-Through Data". LNCS, Springer, 2005.
[6] R. Mudgal, R. Madaan, A.K. Sharma, and A. Dixit, "A Novelarchitecture for question classification based indexing schemefor efficient question answering," International Journal ofComputer Engineering & Applications (IJCEA), ISSN: 2321-3469, Volume-2, Issue-2, June 2013.
[7] K.H Lin, "Predicting Next Search Actions with Search Engine Query Logs," Web Intelligence and Intelligent Agent Technology (WI-IAT), IEEE/WIC/ACM International Conference on (Volume: 1), 2011.

