

Query Processing Using Open NLP Tool

S. Karthick*, C. Arun, S. Selvakumarasamy and Ganesh Arora

Department of Software Engineering, SRM University, kattankulathur, Chennai, Tamil Nadu, India;
karthik.sa@ktr.srmuniv.ac.in, arun.c@ktr.srmuniv.ac.in, selvakumarasamy.s@ktr.srmuniv.ac.in,
ganeshmodi555@gmail.com

Abstract

Background/Objectives: PROSOL is a community-driven android based application that allows users to interact with the different Educational Leads from different universities, Institutions and others linked with the Application. **Methods/Statistical analysis:** PROSOL implements the query processing using Suffix Array Algorithm, which places the searches based on the prefix value of the query, if the query doesn't matches with the existing options then new query handler will be generated to handle the request. **Findings:** Although many similar concepts exists, what separates PROSOL from other such existing interfaces is that, PROSOL in itself is a separate world of queries under which raised queries by one user will be resolved by team of experts synchronized with the application interface .This application gives members chance to earn points based on their response to the queries, upon which rewards will be given .This makes the application interactive and also encourages the participation of more members. **Applications /Improvement:** Use to assess the skill levels of peoples and, also acts as forum to solve problems of users.

Keywords: Android, Answers, PROSOL, Questions

1. Introduction

Questions and Answers Forums are kind of online discussion sites where people can interact with others through posting messages¹. These forums are far different form chat rooms where messages are more verbose than one line of text. Apart from these all the posted messages are first reviewed by the moderator to classify the content, after that it made visible to interested participants². The forums are usually and questions and answer³ based community which encourage the people to post their queries which could replied peoples who were part of the community.

Forums have a well-known terminology associated with them; e.g., a single conversation is called a "thread", or topic. A discussion forum is categorized or tree-like in structure: a forum can contain a number of sub-forums,

each of which may have several topics. In every topic, each debate started as a new thread. The thread is made visible to the user who has shown interest related to the specific topic, so that they can simply view the conversation or they can participate in the debate. Forums also provides the feature that a user can be of being anonymous while asking queries and giving answers^{4,5}. For answering or asking any question users require to login whereas just for reading asked questions it is not mandatory to login⁶.

The most common structure of most of the forums is a tree like directory structure. At very first page option of "Categories" is provided. Each category is a forum where users can ask relevant questions. Thus each forums may nested sub forums. The last levels of this hierarchy constitute topics (commonly called threads). These threads allow members to start their debates or posts. Forums are organized into a finite set of generic topics (usually with

*Author for correspondence

one main topic) which are managed by a group known as members, and governed by a group of people classified as moderators. It can also have a directed graph structure⁷. Where the nodes represent the members and edges are the interactions among them. There are three formats for classifying the threads based on their context of information: Non-Classified/Semi-Classified /Fully Classified. If the Post are not related to one another, which is entirely new is categorized as Non-classified Thread. If a user post has already registered with some response or the debate in progress are categorized to semi-classified thread format is best. If a user has a query⁸ and it has been already replied by the participant then it is categorized as fully classified thread format is best⁹.

Forum is usually a medium of communication between the people community for various reasons such as organizing debates, FAQ, comparisons, polls of opinion, etc. Nonsense or unsocial behavior is quite common on the forums⁶. The usual reasons for such behaviors are discussion on controversial topics, improper understanding, in-conclusive, in-complete, or confused nature of the problem or due to ability of user in understanding the topic. Other problems which is most frequent on forum is the deviation of the users from the topic and the reason behind is there replies to a topic are often worded aimed at someone's point of view, which usually change the direction and motivation of the discussion. For avoiding such kind of situations some users are provided with the privilege of blocking the thread if found controversial to them.

The other problem with the contemporary forums is the limited set of experts. This affects the quality of answers and latency of the replies. There are number of well-known forums available e.g., stack overflow, wiki-how, Yahoo forums, Google Forums. These all forums are having same problem that the set of users who answered the questions is limited¹⁰. Many of these sites have automated deletion of questions if they are not answered in stipulated time. And even the users who have asked the removed questions are not notified about this³.

In The next section of this paper advantages and disadvantages of existing forums is discussed. In the 3rd section new idea is proposed for overcoming the above mentioned disadvantages. In 4th section conclusion is made about the advantages of using the proposed methods over the other methods.

2. Materials and Methods

2.1 Existing System

PROSOL features make it unique. There is no such kind of systems which can redirect the asked queries to experts. For example if a user ask questions related to virtual machine then the system redirect, it to relevant people who have shown interest or tagged with virtual machine¹⁰. This feature not only joins the two kinds of its users but also reduce the replies latency of the asked questions. What makes PROSOL different from other applications is, its architecture. Its Architecture is optimized and scalable thus can work efficiently. The following section explains PROSOL's architecture in more detail.

2.1.1 Proposed System

The features supported by the PROSOL are:

1. Provides option for selecting types of query mode (custom or default).
 - » Under the custom query mode users can add the members manually as supported in "whatsapp" and send their queries to them. In this mode users are forced to add tags relevant to queries. Where as in default mode users are supposed to add tags and based on this queries' domain is determined. After this queries are sent to domain related experts or teachers or other verified users.
2. Notify the user about the status of asked question.
3. History of the asked questions is organized on the basis of category of questions.
 - » It allows the users to organize their queries history on the basis of queries type. The structure of history is tree like directory structure in that internal node represents categories and leaf nodes queries.
4. A search box that lets users searching for any kind of question (as provided by the Quora).
 - » Is Implemented using Suffix Array Algorithm. Suffix Array provides searching based on prefix. If the asked question is not found then this module will report to New-query handler. This will create the query and pass it to Notification handler.
5. Question-feed provides domain wise top questions (based on number of up votes) of the day.
6. Answers are displayed along with following information

- » Name and status of the user who has asked the question.
 - » Name and status of the user who has given the answer.
 - » Number of up-votes and down-votes.
7. Use new bucket based methodology for storing the user's queries. This methodology is explained in the following sub-section.

2.1.2 Bucket Based Methodology

1. ER diagram for Bucket Based Methodology.

Explanation:

Whenever users ask questions it is passed to server. Before storing into database it is passed from three steps which are:

- a. Asked question passed to text classifier. For text classification OpenNLP tool POS tagger is used. POS tagger classify query into nouns, pronouns, and other parts of speech.
- b. Obtained POS list is processed in this step. Each part is stored in its type table suppose noun (NN) is to be stored then it will be stored in table named "NN". Each stored POS is called tag. Once it's stored a unique tag id will be generated against its name.
- c. Once tag Id is obtained from the previous step. Each tag id will be linked to the question using

many-to-many relation. Linked tag id and question stored in "NNTaggedQues" and "AskedQuesInfo" tables respectively

In PROSOL following POS are used as tables. These are linked with "AskedQuesInfo" table as shown in the diagram.

1. Coord conj
2. Card num
3. Determiner
4. Existent
5. Foreign word
6. Preposition
7. Adj
8. comp
9. Adj super
10. Lis item
11. Modal
12. Noun, singular, Adverb, Plural
13. Comparative, superlative
14. Wh-determiner
15. Wh-pronoun
16. WRB Wh-adv erb

Whenever users search for a query same steps are used just instead of storing the tags it looks for the tags in the respective tables and fetch the linked question to it. After this it prioritizes the questions based on number of times each question is repeating, for prioritization prior-

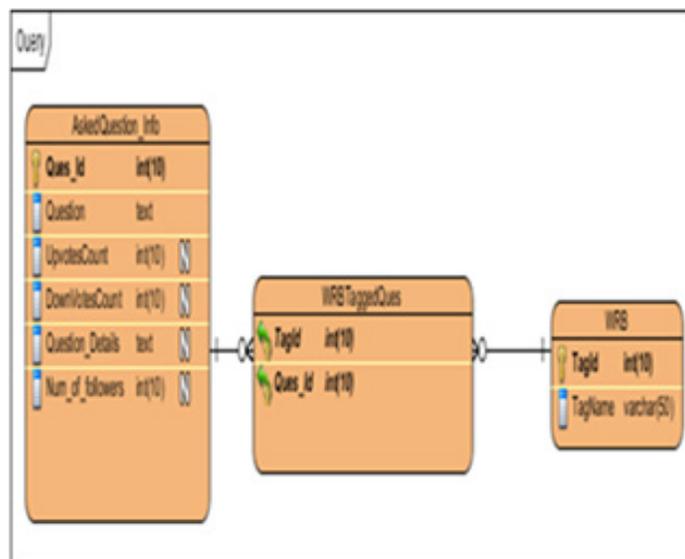


Figure 1. Query Processing.

ity queue is used. The static guise of a Query processing of the PROSOL is depicted (Figure 1).

2.2 Architecture of PROSOL

Various aspects of the PROSOL have been expressed using different Scenario based architecture. The first

scenario reveals how a user places a query to PROSOL System (Figure 2). The Second Scenario demonstrated how the queries placed by users will receive the response from various people from the audience in the PROSOL (Figure 3). The third scenario demonstrates the process of submitting a new query to the PROSOL system (Figure

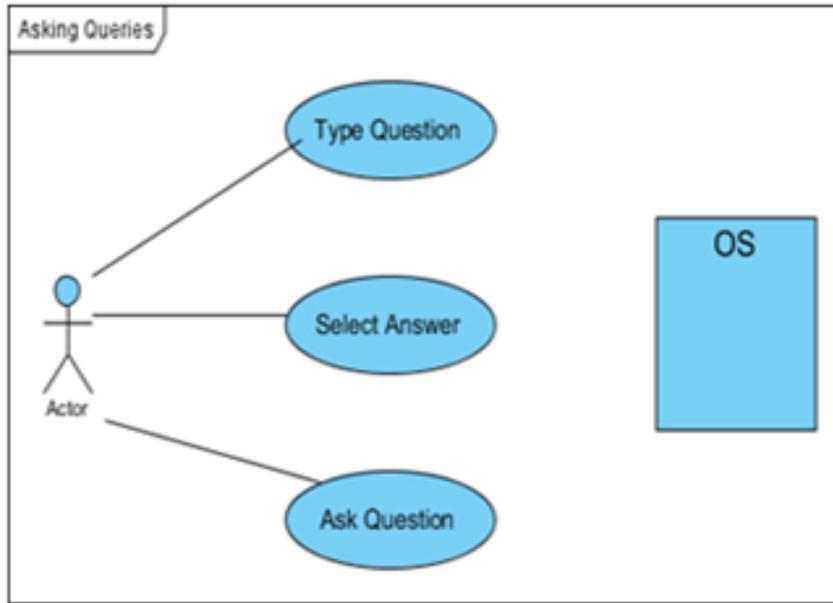


Figure 2. User placing a query to the PROSOL.

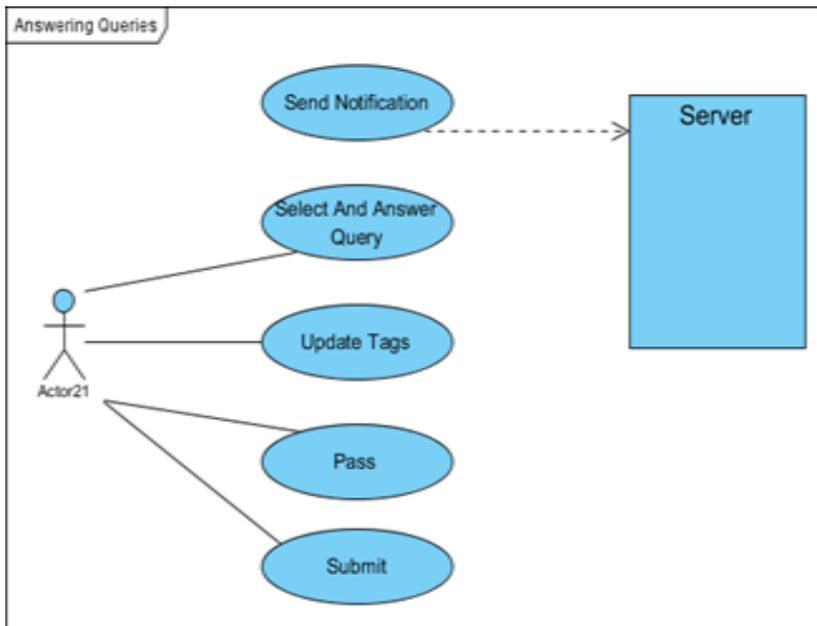


Figure 3. User submitting a response to the query.

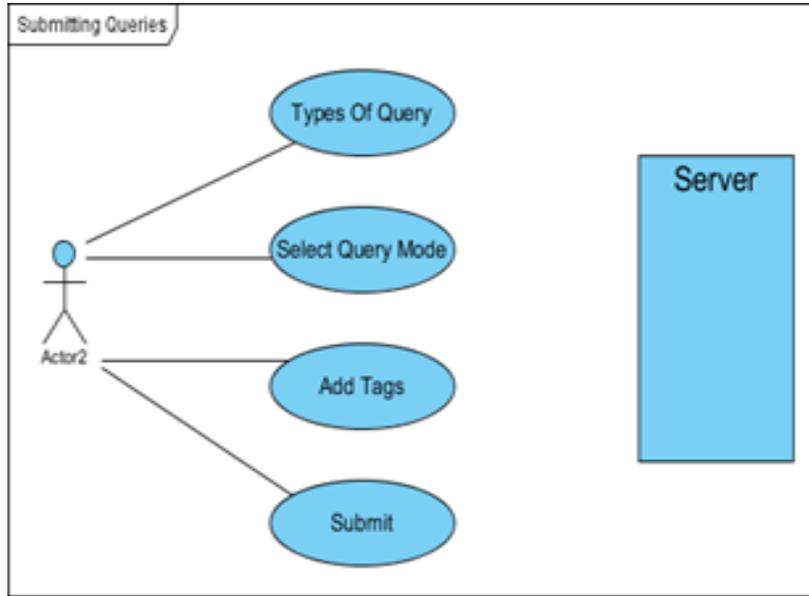


Figure 4. Scenario of submitting the query.

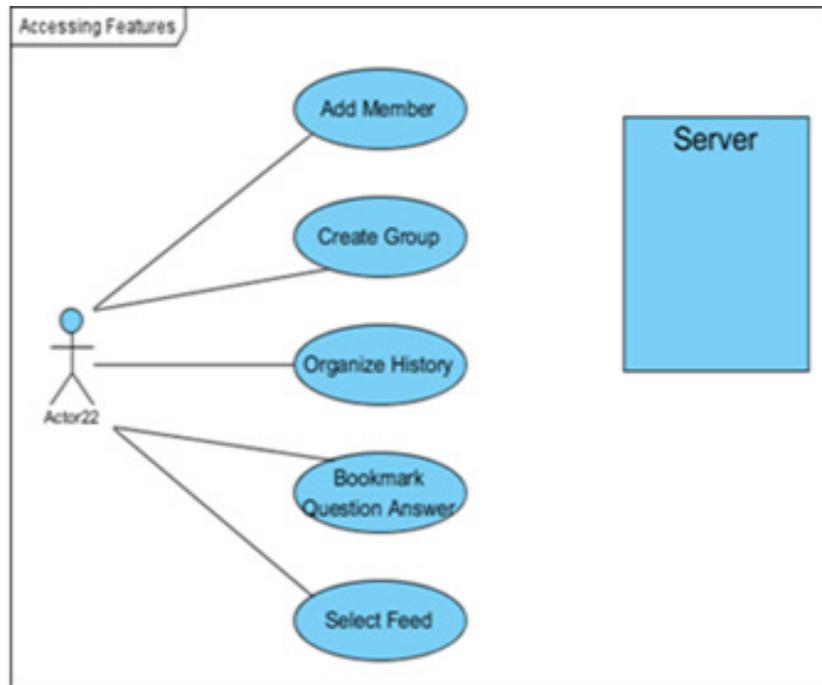


Figure 5. Features of the system

4). The Last scenario demonstrates the various features that exhibit in the PROSOL (Figure 5).

The (Figure 6) Architecture of the system reveals the functionality of the system that could be used by the users

of different community of people associated with the PROSOL community, it's also expressed in the form of conceptual diagram (Figure 7).

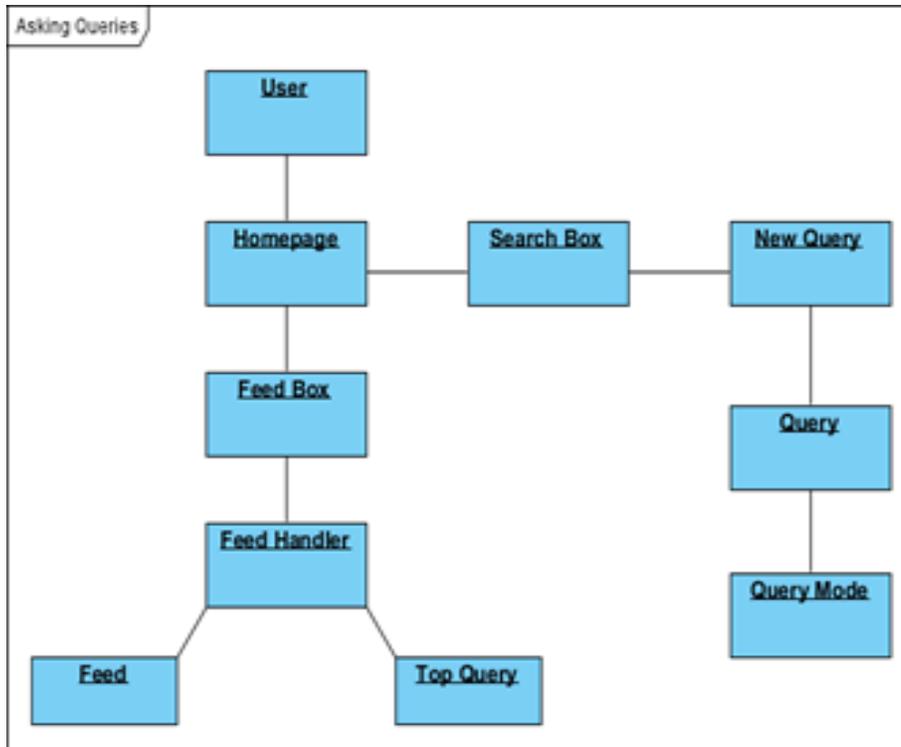


Figure 6. Architecture of the system.

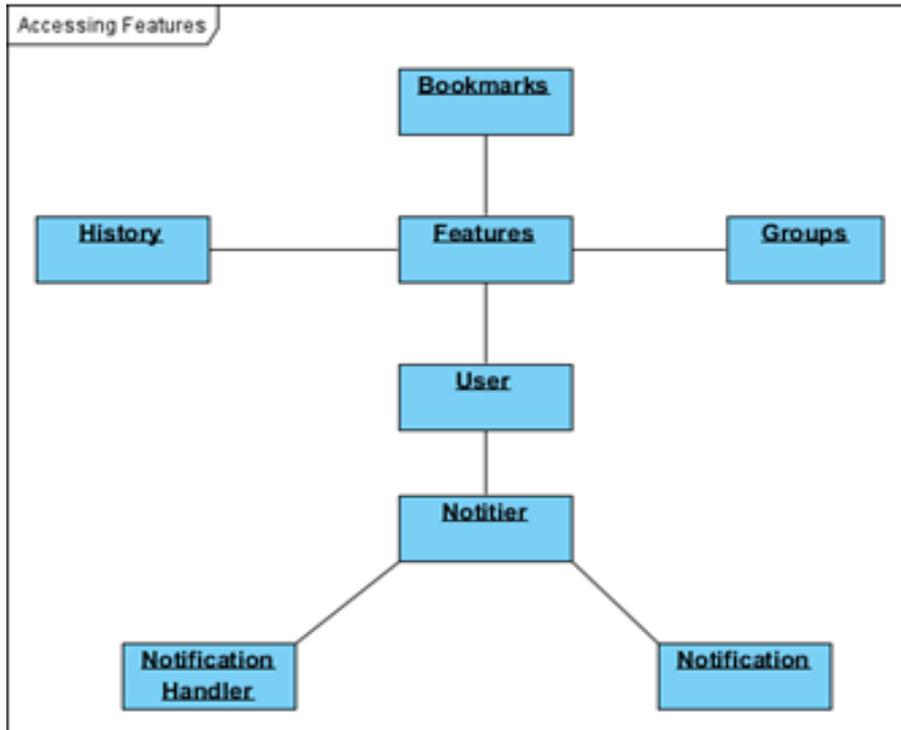


Figure 7. Conceptual diagram.

3. Conclusion and Future Enhancements

PROSOL address the problem solving by connecting the teachers or professors from various universities, we can not only improve the quality of answers but also can reduce the replies latency. This arrangement brings the win-win situation for both the side of users, those who asked the doubts having the better chances of getting quality replies where as those who answers queries will be awarded with reputation points and prizes. Apart from this the proposed algorithm for opinion mining is having optimal complexity thus can be used in the devices having low processing capabilities such as mobile phones.

4. References

1. Bulletin Community Forum - FAQ: What is a bulletin board, www.vbulletin.com/forum/help. Data accessed :9-1-2008
2. Anonymity-Quora-featureK. <https://www.quora>.Data accessed: 4-2014
3. Debate tools - LessWrong Wiki, https://wiki.lesswrong.com/wiki/Debate_tools#Explore. Data accessed: 26 - 2016.
4. Non-Threaded vs Semi-Threaded vs Threaded Message Board, <http://www.bulletinboards.com/ThreadHelp.cfm>, Data cessed:2-4-2016.
5. Sewak M. Finding a Growth Business Model at Stack Overflow, Inc. Stanford Case Publisher, 2010, p.1-35.
6. Moore M, Casal C. Users of Yahoo Answers seek advice, opinion, expertise. University of Michigan News Service, Ann Arbor, Michigan, 2012.
7. Modine M, Austin A. Wolfram Alpha to venture beyond boron nucleus. The Register. Southport, England: Situation Publishing, Retrieved 2011.
8. Liu L, Yandong Y, Agichtein E. On the Evolution of the Yahoo! Answers QA Community. Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, 2008, 737-738.
9. Murugan K, Ravichandran T. Intelligent query processing in temporal database using efficient context free grammar. Indian Journal of Science and Technology. 2012; 5(6):1-6.
10. Sharma L, Dhir V, Kaur K. A New Model for Question-Answer based Dialogue System for Indian Railways in Hindi Language. Indian Journal of Science and Technology. 2015; 8(32):1-4.