

Analysis of Question and Answering Recommendation System

Jayesh Kale¹, Hetansh Bharad², Sourav Ghatage³, Soham Meher⁴, Dr. Shwetambari Chiwhane⁵

^{1,2,3,4}B.E Student, Dept of computer Engineering, NBN Sinhgad school of Engineering, Ambegaon, Pune-411041, Maharashtra, India

⁵Prof. Dr., Dept of computer Engineering, NBN Sinhgad school of Engineering, Ambegaon, Pune-411041, Maharashtra, India

Abstract - Question and answer (Q&A) forums are getting bigger day by day as number of internet user increases every user will have a specific and new problem which is either already solved or yet to be solved. So for creating a platform for college student where they can find all the relatable questions and also ask questions about the studies whose answer can be given by the community. In this paper we are focusing on the literature survey on various features and technology which are involved in creating a forum with a recommendation engine. Explained steps are Recommendation engine, Leading user detecting, FAQ detection and Open Information extraction system. Based on the analysis, the ideal process and procedure is acknowledged, which will make the forum effective and used in all devices.

similar question is found that question may have many answers to it. So, now to need to get the best answer. For that, we create a pair of question and answer i.e. if two question has five answers to it then we create ten pairs of question and answer. After creating this pairs we need to rank the best answer for which we are using like count and comment polarity. Every answer has an option for adding a comment and like to it. Users can like the answer if they like and comment on the nature of the answer. While evaluating question-answer pair we use a combined score of like count and comment polarity which is classified into positive, negative and neutral to rank the best pair of the question and answer.

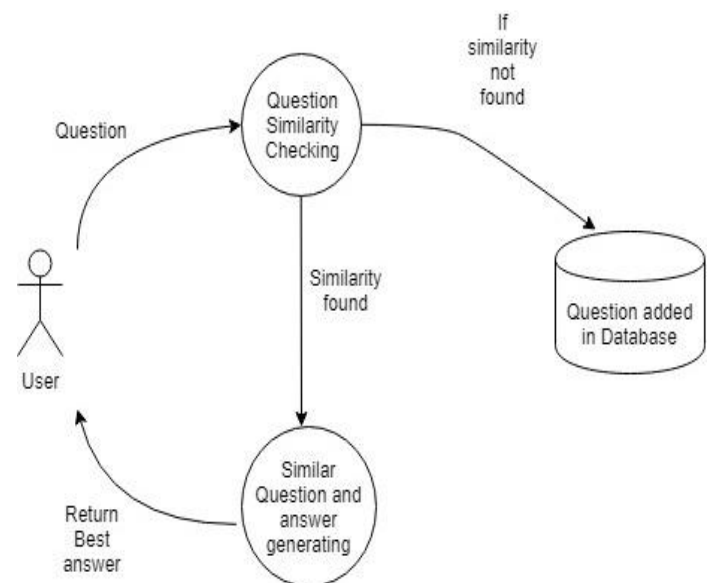
1. INTRODUCTION

Analysis and review being done on various techniques applied for question and answering system. A brief analysis of each method was carried out to identify its own merits and limitation. The literature survey has different approaches/tools to develop an efficient question and answering system. The survey helps to gather ideas from the existing methods such as accuracy of recommendation engine and leading users and open information extraction system. Technologies we are using to accomplish this are namely flutter, machine learning, flask, dart.

2. RELATED WORK

2.1. Answer Ranking in Community Question Answer (QA) System and Questions Recommendation

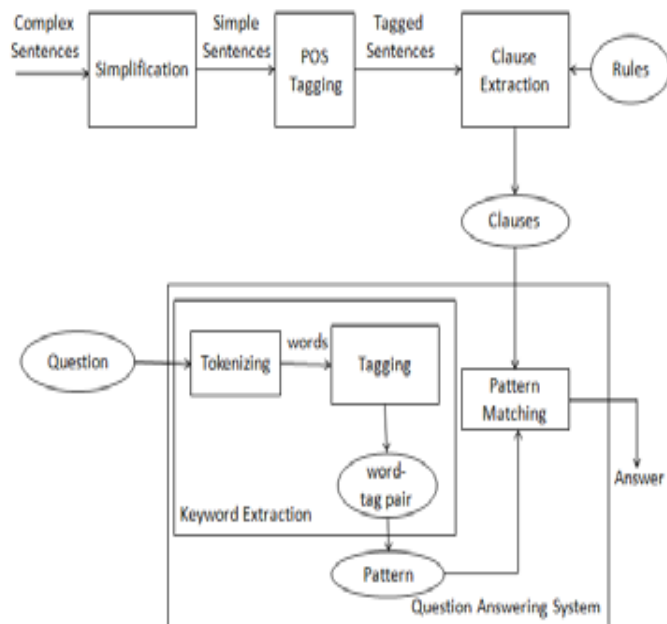
When a user asks a question that question may contain unnecessary words that may not affect the outcome of the question and so those stop words are removed first. After removing the stop words we get a string of tokens that only contains words that are important. Now, we already have a database of questions and answers and we may have similar meaning question already present in the system so instead of storing the question to be answered by another user in the system we first compare the given question with the question already present in the system using Levenshtein distance algorithm which gives us the score of similarity between two string. If the similar question is not found the system prompt to add a question to the database. If the



2.2. Open information extraction for QAS

In Open Information Extraction (OIE) systems like TextRunner extracts information from a huge corpus irrespective of the domain. The TextRunner consists of three major modules namely self-supervised, single pass extractor, redundancy based assessor. The first module inputs a small corpus to the learner and outputs a classifier that used to classify the text as trustworthy or not. The second module extracts tuples for all possible relations by making a single pass over the entire corpus. The extractor generates one or more candidate tuples from each sentence. It sends the tuples to the classifier and retains only the ones that are

labeled trustworthy. The probability to each retained tuple is assigned using the third module. System consists of modules namely sentence simplification, POS tagging, clause extraction and question answering. First, the document that consists of large number of sentences is taken as input. The sentences present in the document may be complex or compound. A sentence simplification process is necessary to simplify the complex and compound sentences into simple ones. These sentences are later tagged with the POS tags. The tagged sentences are then taken as input and extended set of rules are applied on it to produce the necessary clauses. These sentences in document can directly be split into clauses from which relations and other sub-clauses can be produced. Then, the extracted relations and clauses are used to operate the question answering (QA) system. Our QA system consists of a set of questions from which the key words are detected and compared with the extracted clauses and relations to determine the answer to those questions.



2.3. Leading User Detecting model

The model mainly consists of two parts: leading capacity estimation and collective decision. Making use of Q&A data and social data of a professional CQA service, users' leading capacity is analyzed from three aspects, and then the generated standalone rankings are merged into a synthetic ranking by using an optimized Borda count decision making method. Measuring leading ability, which are authority, activity and influence. First, an important feature of leading users is the ability to generate knowledge of high quality in the community. Hence authority in one or several fields is necessary. Second, leading users should spend much spare time in participating in kinds of normal routine in the community, to make contributions constantly. Third, great social influence is also a common requirement. It will help to

improve others' enthusiasm, and let more and more people benefit from the knowledge base. Thus, the leading capacity is modeled with a triple: LEADING = (ATH, ACT, INF). The Borda count makes a collective decision by giving each candidate a certain number of points corresponding to the position in which it is ranked by each criterion.

Q(u): The questions that user u has asked

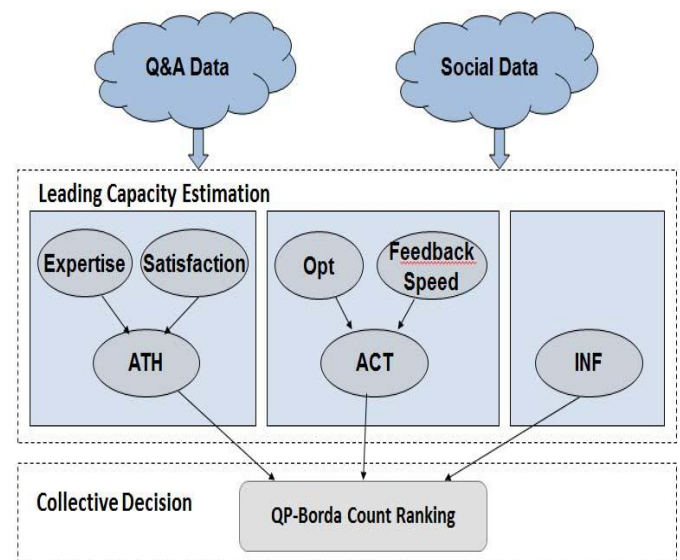
A(u): The questions that user u has answered

V(u): The votes that user u has given to others

S(u): The votes that user u has received from others

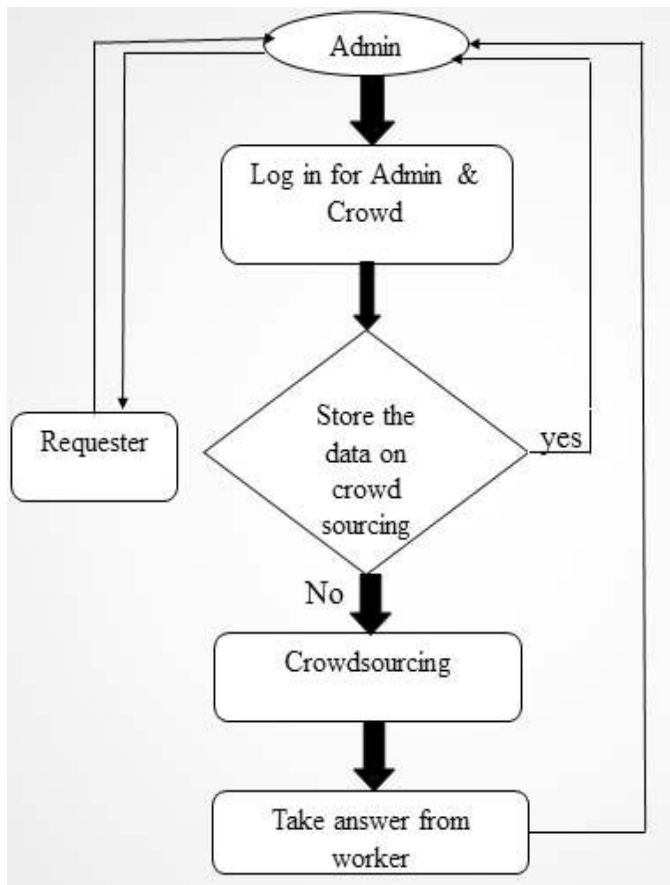
Fl(u): The users who follow user u

QAFI(u): The users who follow the questions asked



2.4. Analysis of QAS with crowd sourcing

The main objective of analysis is to supply a platform for open communication to social media and community for finding their queries, To design of crowd sourcing application which will use for public required data, to make QA system that produces an appropriate declare the received question and gift it to the verbalize for this purpose. The work includes the info for personal Crowd sourcing data and sort of question answer it's solely utilized by administrator. The principle of crowdsourcing in education- Crowdsourcing is large in different industries, however, has however to have an effect on education. It does not be that academics area unit recreating lesson over and another time. With the common core dealing, it is sensible that academics would be exchanging best practices and methodology that job. Academics area unit probing for quality materials that are vetted.



2.5. An Answer Recommendation for a Community Question Answering Systems

The question descriptions are represented with the query likelihood framework and word embedding methods. The similarity of two cases is measured by the distance of two vectors. The most similar historical cases could be selected according to the similarity ranking. Afterwards, in order to select the high quality answers, the answers belong to the cases are chose to estimate the quality. Another key operation in cQA is answers quality estimation. For each historical question, there are many answers from different doctors. The quality of the answers is varied because the doctors' experiments and knowledge are uneven. It's essential to select the high quality answers of the question. There are two kinds of popular approaches to assess the quality of answers, crowding sourcing methods [16, 17] and machine learning methods. To recommend answer to a particular question on medical QA systems. There are three steps for the answer recommendation system, including similar cases retrieval, answers quality estimation and answer recommendation. The data is collected from the QA system. The algorithm is tested on the dataset and the results show an increase in performance of answer recommendation.

2.6. Exploiting Response Patterns for Identifying Topical Experts

Identifying DQs: The classification of DQ shows an accuracy of above 95% with F1 measure of 0.9. This validates the authenticity of our classification method. It supports view count as a basis of measuring the difficulty of the questions. Upon checking manually we find that each question labeled DQ possesses one or more of the features FDQ. We observed that some of difficult questions are inter topical questions i.e. a question in a topic becomes a bit difficult in the topic when it involves the knowledge of other topics too. Better results are obtained when we use view count along with favorite count together as an assumption of DQ. The bar graphs in figures 1a and 1b, respectively, shows the accuracy and F1 measure of the features of DQ and their combination. Almost all accuracy values are above 50% indicating that the features are correctly selected for the DQ identification, however combining all the features provides the best accuracy. 2) Identifying experts: Table IV represents the result of the comparison of the List Ex 1 & List Ex 2. We present the results of four different aspects of experts, the first basis being the more conventional approach of expert identification. We observe that the 3rd & 4th assumption of an expert out weights the conventional one. We observe that basis 3, where we combine 'Best Answer' count with answer count on questions with relatively low answers, provides a higher accuracy and f1 measure with a reduced set of features as compared to the best answer basis (basis 1). However including the Difficult Questions feature (FEX 11) combined with the users 'Best Answers' as criteria of expertise, provides the best results in terms of accuracy and F1 score. Thus this basis indicates a more reliable classifying model with much reduced set.

Topic	Expert Basis	Instances	Approx % correctly matched instances	Accuracy	F1 measure
C	Best Answers	31681	99	94.772	0.926
	Difficult Questions		49	54.791	0.547
	Best Answers & Ans on Ques. Low Ans count		99	99.369	0.991
	Best Answers & Difficult Questions		99	99.598	0.995
Python	Best Answers	33747	99	94.724	0.927
	Difficult Questions		66	61.151	0.581
	Best Answers & Ans on Ques. Low Ans count		99	97.536	0.967
	Best Answers & Difficult Questions		99	99.321	0.99
Java	Best Answers	40140	97	82.75	0.773
	Difficult Questions		64	54.981	0.544
	Best Answers & Ans on Ques. Low Ans count		99	97.262	0.963
	Best Answers & Difficult Questions		99	98.125	0.974

3. PROPOSED SYSTEM

“Ask A Question ” which is one of the components of the system option will lead a user to a screen where a user will be able to see multiple questions that are currently been live on the system and also some question that the user previously asked and some question related to the user’s activity. This feature will allow the user to ask a question in the community. Now, when want to ask a question on the system we have given the option for the type of question that the user wants to post of that community like in the system we have implemented we are working on computer science community and in that java based question. We have also provided a similarity question provision for the question that is already asked by the other user can be used for reducing the redundancy. “All Question”, an option that is what makes this system worth its time. The named Question answer system needs those questions asked by the user to be answered and in this option, it is what is been done. All questions of the community are that have been posted will be visible here; also a filtering option is available to get to some specific top n that community. A user will be able to answer the question and will be stored in the database with an answer id connecting question id to be used later. “Recommend Question” option which will help users to get a recommendation of the question for himself based on the activity of the user and will be also able to suggest questions to other user based on the interest. This system also has some major functionality that makes this system different and interesting for the user to work with. Question and their answer that user likes can be sent to his email in just a few steps and can also be bookmarked for the future reference. This both the future can ease the process of searching a question which a user already asked, answered or want to just refer again. Comments and like feature is also been provided for the user and are been used extensively. Comments are been used to classify answer in three categories: positive, neutral, and negative. This will be done on bases of a dictionary where all almost around 5000 words with is used in a positive, negative and neutral sentence is supplied to the system. After this classification, we are going to use other users likeness towards that answer to project it for another user to be viewed on the top of many more. This is wanting answer ranking in community question answer is all about to flash those answer on top of the searched question so that user will be getting rid of that manual search labor.

4. CONCLUSION

Based on this literature survey, presents an analysis of different techniques used in the process of making question and answer forum. This study paper provides the fundamental techniques, recommendation procedure, and ranking methods required in order to build Q&A forum or platform. Exploring the proposed work in field of Q&A platform principle, literature’s and commercial utilization experiences, different feature and techniques to achieve

performance, robustness, security and accuracy. A careful analysis of requirement and system performance will be helpful for developing new efficient algorithm with better performance in the research work. For future research work we intend to consider the other option and metrics to evaluate the performance.

REFERENCES

- [1] “Answer Ranking in Community Question Answer(QA) System and Questions Recommendation” by “Sunny S. Shah”, “Tejas S. Bavaskar”, “Sourabh S. Ukhale”, “Akash S. Kalyankar”.
- [2] “Robust Web Service Recommendation via Quantile Matrix Factorization” by “Rui Zhu”, “Di Niu” and “Zongpeng Li”.
- [3] “Question Answering System Using Web Snippets” by “Menaha R”, “Udhaya Surya A”, “Nandhni K” and “Ishwarya M.”
- [4] “Exploting response patterns for identifying topical experts in stackoverflow” by “Manish Bhanu”, “Joydeep Chandra.”
- [5] “An Open Information Extraction For Question Answering System” by “D.Thenmozhi”, “G.Ravi Kumar”.
- [6] “Leading Users Detecting Model in Professional Community Question Answering Services” by “Siqi Song”, “Ye Tian”, “Wenwen Han”, “Xirong Que”, “Wendong Wang”.