

Implicit and Explicit Knowledge Mining of Crowdsourced Communities: Architectural and Technology Verdicts

Husnain Mushtaq, Babur Hayat Malik, Syed Azkar Shah, Umair Bin Siddique, Muhammad Shahzad, Imran Siddique
Department of CS & IT
The University of Lahore
Gujrat, Pakistan

Abstract—The use of social media especially community Q&A Sites by software development community has been increased significantly in past few years. The ever mounting data on these Q&A Sites has open up new horizons for research in multiple dimensions. Stackoverflow is repository of large amount of data related to software engineering. Software architecture and technology selection verdicts in SE have enormous and ultimate influence on overall properties and performance of software system, and pose risks to change if once implemented. Most of the risks in Software Engineering projects are directly or indirectly coupled with Architectural and Technology decisions (ATD). Advance Architectural knowledge availability and its utilization are crucial for decision making. Existing architecture and technology knowledge management approaches using software repositories give a rich insight to support architects by offering a wide spectrum of architecture and technology verdicts. However, they are mostly insourced and still depend on manual generation and maintenance of the architectural knowledge. This paper compares various software development approaches and suggests crowdsourcing as knowledge ripped approach and brings into use the most popular online software development community/Crowdsourced (StackOverflow) as a rich source of knowledge for technology decisions to support architecture knowledge management with a more reliable method of data mining for knowledge capturing. This is an exploratory study that follows a qualitative and qualitative e-content analysis approach. Our proposed framework finds relationships among technology and architecture related posts in this community to identify architecture-relevant and technology-related knowledge through explicit and implicit knowledge mining, and performs classification and clustering for the purpose of knowledge structuring for future work.

Keywords—StackOverflow; architecture and technology verdicts; crowdsourcing; data mining; explicit and implicit knowledge; software repositories; knowledge structuring

I. INTRODUCTION

Large sharing of data and information on social media and Community Q&A websites have become a potentially valuable and rich source of knowledge for a large spectrum of users across the world. It has become a norm of learning and professional communities to knock at web based social media communities when they feel a need to get insight about a new topic, subject or to solve a particular problem. The fact behind

is that with knowledge relevant web communities, such online forums, social media groups, Community Question Answering (CQA)/crowdsourcing, traditionally known as Question Answering (Q&A) websites gather contributions from a large pool of users with different levels of expertise and experiences. Recent time is witnessed the acute emergence and growing popularity of these Community Q&A sites among software professionals [1]. The use of social media as source of knowledge for software engineering process and practical learning has been on a steady rise in recent years.

Software architecture is a complex process of making technology and design decisions which impose potential risk on software development project success, directly or indirectly. Technology and Architectural Decisions not only affect a system's structural properties but also its quality attributes [15], [19]. It is dying hard to change an architectural solution and Technology decision after it has been implemented as such change become perilous risk for system and project [2]. Software architecting and implementation are different aspects of software projects because changing in technology is more complex and crucial as compared to bug fixing in a system. Architecture and Technology Knowledge plays a vital role regarding Architectural and Technology Decisions [2]. Existing architectural and Technology Knowledge (ATK) management approaches emphasis on fabrication of repositories of ATK to guide software architects and system analysts to make the desired decision [3]. These repositories of architecture and technology are built manually by architects and it is prolonged and mined-numbing process to populate and utilize such repositories. Manually populated architectural and technology knowledge repository can accumulate limited knowledge while, in the meantime, manual evolution and maintenance of these repositories is another laborious task.

Architectural and technology decisions (ATD) imprint their effects on all over software project. Any risk attached with such decisions may spoil the whole project due to dependency of SDLC on ATD. Conceptual ATDs effect the architecture configuration of the system which is not aligned implementation [12]. No doubt, risk is inevitable in large, even in medium and smaller; software projects and ATDs are about selection of concrete technology and architecture solutions, such as design patterns, frameworks and tools through which it

could be easy to implement risk avoidance and risk mitigation strategies [17]. The selection between technologies solutions architectural designs mostly depend upon architecture-technology relationship and its knowledge. About 30% of executive decisions constitute of a system ATDs, and most of them are technology decisions. Moreover, executives, technology experts, system analysts, architects and domain experienced personals are involved in ATDs and organizations spent enormous time in getting the people engaged and spent their dear cost to get valuable opinions. Major purpose of experts' opinion is to identify and avoid risks in start to make rightful ATDs. In addition, technology ATDs are the mostly documented for future use. Physical identification and involvement of experts across the globe and utilization of manual ATDs knowledge repositories is possible but not seems feasible where project time is less and cost is high.

Social software (e.g., forums) offers novel, advanced and mechanized methods to share, capture and utilize knowledge across the world. Now, Software engineers ask questions, and assimilate from the crowdsourced (people on social media, Q&A communities and forums) about architecture and technological related risks. Many architects mentioned social software as a rich source of knowledge for technology solutions where large number of experts and unlimited knowledge repositories and huge technology solutions are available [13]. Recent social media studies about software development gives an insight that developers utilizes the rich libraries of software developers' communities spread across the social media to discuss variety of concepts, such as architecture, design, technology and domain concepts and related software engineering risks [5], [6]. Hence it is revealed that architectural, technology, design and risk knowledge exists in software development communities. Ultimate approach of this paper is to complement architectural and technology management systems with efficient data mining methods for extraction, classification and clustering architecture, technology and risk related knowledge [9]. Clustering of architectural and technological risk with respect to software project domain and build relationship between risks and ATK clusters in developers and experts communities is purpose of this paper.

To achieve this, there is need to consider a well-structured online software development community as a rich knowledge repository of architectural, technology and related risks along with their solutions. StackOverflow (SO), as software development community websites, is large website with Question and Answer (Q&A) structure [5]. SO supports utilitarian knowledge management profile like incorporation of Q&A posts with context details and it ensures quality of post through categorization and tagging the posts. It is an emerging platform for data mining researchers to extract, classify and sort largely distributed useful knowledge through various data mining and machine learning techniques like clustering, k-means algorithm, association, etc. [9]. Architectural and Technology management approaches can be supported with progressive technology-related knowledge and software related risks, evaluated by a pool of technology experts across the globe via virtual means [10]. Fig. 1 shows the overall problem domain, crowdsourced knowledge and insourcing knowledge.

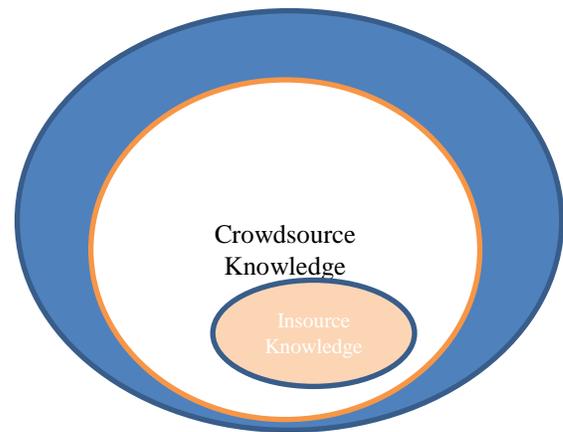


Fig. 1. Knowledge comparison of Crowd and insource w.r.t application knowledge.

The paper provides a novel framework with an objective to identify SO posts related to software architecture, technology and risks management Knowledge. Moreover, the study explores the key differences and relationships among Architecture, technology and programming posts. Based on this study, the paper focuses following Objectives:

- Extraction of raw data in form of SO Posts
- Classify SO posts as architectural, technology and risk through data mining techniques
- Find a relationship and association among SO posts using machine learning approaches
- Analyze triangular based relationships make best architectural and technological choice with minimum risk in software project.

II. LITERATURE REVIEW

The study organizes the literature review in related areas. This study is a multi-faced study that involves mining of Software Engineering Knowledge from unstructured and unorganized data from community Q&A platform using Data Mining and Machine Learning techniques [8]. This domain guided knowledge is then brought into use to make architectural and technology decisions. Right architectural decisions largely depend upon software architectural knowledge and its utilization in different kinds of applications. Software architects also rely and seek guidance from software architecture based knowledge repositories. These knowledge repositories are part of software repositories which are specifically being built to make their best use in future projects as guideline [2]. Both, architect and technology decisions are considered of highest importance to decide fate of software project [1]. Software project implementation depends upon right and suitable software architect and system components configuration decisions. Technological knowledge follows architectural knowledge in importance regarding software project and technology selection decision also rely on architectural design. Because there exist plus and minus of each technology for different design patterns [4]. Architecture and technology decisions are made by top management and properly documented due their importance in whole project

life. Q&A communities present themselves as rich candidate for conducting knowledge mining research.

The ever increasing volume of data and information on social media has become valuable and rich source of knowledge. Definitely, people look for fastest, largest and reliable platform when they need to learn about some new topics or to solve some specific problem [7]. Therefore, they give priority to consult with some web community like social media, online Question Answering (Q&A) Sites, which welcome data form large pool of users (students, Experienced, Experts, etc.). Recent years are witness of emergence and growing popularity of these sites among academicians and industrialists. Domain of computer sciences, Software engineering and Information Technology are major contributors and beneficiaries of these Q&A Sites Stackoverflow is good example of such community where millions of users from similar domain are registered for Quo pro quid [10]. Pearson's latest annual report on use of social media has revealed that there is an acute rise in use of social media by software developer community and graph of their dependency on these social Q&A sites has gone up tremendously[21]. Some researchers even say that online assistance has become indispensable for programmer and software engineers [25].

To date knowledge mining research on community Q&A has directed in line to predict answer quality, crowd participation rewards and users ranking and profiling expert finding, and success factors of community, subject related analysis [8]. Advances in Knowledge Discovery and Data Mining brings together the latest research in statistics, databases, knowledge discovery, machine learning, and artificial intelligence that are part of the exciting and rapidly growing field of Knowledge Discovery and Data Mining [13]. We believe, there is an apparent lack of knowledge mining of Community Q&A sites especially in context of decision making process of software engineering process: opinion mining. Mining community Q&A sites to extract knowledge (implicit and explicit) to refine decision making process [26][27]. This study describes a first level attempt to gather implicit and explicit knowledge form stackoverflow and make it useful in architectural and technology related verdicts.

III. SOFTWARE DEVELOPMENT APPROACHES V/S CROWDSOURCING APPROACH

A. Insourcing

In insourcing, organizations accomplish project goals by using internal expertise. Organization opt to hire new qualified human resource, shift staff from one project to another or train staff to complete a project, Instead of subcontracting to third parties [11]. Insourcing improves communication among staff members and internal IT resources are also innovated. Cost is the main challenges of, insourcing due to hiring new staffs and software licenses.

Comparing crowdsourcing with insourcing in software development, user participation, flexibility, openness, scalability and flexibility in insourcing are said to be lower than crowdsourcing. On the other hand, development time, development cost, trustworthiness, license requirement, business risk and operational control of insourcing are higher

than crowdsourcing [11]. Control over software development process is stronger in insourcing software development compared to outsourcing and crowdsourcing [11].

B. Outsourcing

In case, sufficient in-house expertise is not available then organizations choose outsourcing. It means, organizations contract with external companies for accomplishment of project. It not only reduces burden to find out human resource but reduces cost also. Finding the right service provider according to its expertise is the main task of software development via outsourcing [11].

Crowdsourcing and outsourcing are intermingled with each other [18]. Crowdsourcing utilizes open calls to large number of geographically distributed people to achieve tasks from volunteer workers (with all levels of expertise) while outsourcing makes contract with other companies or professional organizations. Besides, outsourcing performs is business relationships [20], while crowdsourcing is all about participation and motivation. Factors like development time and cost, confidentiality, software license issues, business related risks and level management control of outsourcing are said to be higher than crowdsourcing [11]. However, Transparency, ability to have tailored product, user participation, scalability of crowdsourcing factors are greater than outsourcing.

C. Opensourcing

Open source and crowd source are two different software development approaches. Projects developed by crowdsourcing are not distributed in public for free, while open source software development contains freely distributed software [16]. People may work independently or collaboratively in crowdsourcing tasks while people work in collaboration during open source software development [20]. Factors like user participation, openness, scalability and flexibility are mostly lower in open source software development as compared to crowdsourcing. On the hand, development time, development cost, confidentiality, license issues, business risks and management control are higher in opensourcing than crowdsourcing [11].

D. Nearshoring

It geographic proximity between client and sourcing locations [22], i.e. nearshoring is associated with outside of client country but proximate to sourcing countries. Client countries achieve their tasks at lower wages in sourcing countries by utilizing the proximity of geography, culture, language and economic characteristics between countries [22]. There are three major nearshoring clusters in the world: the USA and Canada, wealthy nations of Western Europe and Korea and Japan [23].

E. Offshoring

Project is performed between clients and supplier organizations located at different countries. The driving force behind offshoring is cost reduction. Communication limitation, language barriers, cultural differences and political issues are few drawbacks of offshore software development approach [24].

TABLE I. COMPARISON OF DIFFERENT SOFTWARE DEVELOPMENT PARAMETERS TO FIND THE BEST AND SUITABLE SOFTWARE DEVELOPMENT APPROACH

Software Development Parameters	Inourcing	Outsourcing	Open sourcing	v/s Crowdsourcing
User Participation	Lower -1	Lower -1	Lower -1	Higher +1
Flexibility	Lower -1	Lower -1	Lower -1	Higher +1
Openness	Lower -1	Lower -1	Lower +1	Higher +1
Scalability	Lower -1	Lower -1	Lower -1	Higher +1
Time	Higher -1	Higher -1	Higher +1	Lower +1
Cost	Higher -1	Higher -1	Higher +1	Lower +1
Trustworthiness	Higher +1	Higher +1	Higher +1	Lower -1
Licenses	Higher -1	Higher +1	Higher -1	Lower +1
Business related Risks	Higher -1	Higher +1	Higher -1	Lower +1
Management Control	Higher +1	Higher -1	Higher +1	Lower -1
Transparency	Higher +1	Higher -1	Higher +1	Lower -1
Risk Prediction	Lower -1	Lower -1	Lower -1	Higher +1
Participants	Lower -1	Lower -1	Lower -1	Higher +1
Points Gain	-7	-7	-1	+7

F. Crowdsourcing

Crowdsourcing has become an emerging platform for both academics and industrial world. It is applicable to software development approach in which open calls are utilized in order to have the tasks done by a large group of volunteer people connect through a platform like Stack overflow (SO). There are three main models of Crowdsourcing: peer production, competitions and micro-tasking [10]. Peer production crowdsourcing is the oldest model where people work collaboratively without any reward expectation. Competitions crowdsourcing approach constitutes that workers compete with each other for achieve projects' goals in order to gain monetary rewards. The requirements are submitted to the crowdsourced platforms. The copilot/platform manager splits the project into sub tasks which are competition tasks with different rewards. The large group of workers i.e. community propose diverse solutions for these tasks. The best solution is chosen among all solutions and winning solution is rewarded [10]. The micro-tasking crowdsourcing is last approach which divides works into several self-contained and small tasks to complete in a short time period by a large group of people via scalability feature of software works.

Table I shows that among all software development approaches, crowdsourcing seems better to adopt as most suitable and broad approach. Emergence of social media and bog data has changed the trend of software development. Crowdsourcing is a hidden potential that can be used in software development process especially in architectural and design decision that are considered more important and pivotal in SDLC. This study unveils the idea that how to mine big data and community Q&A platforms to utilize implicit and explicit knowledge scattered across the internet. The paper takes stackoverflow as targeted Q&A community and gives a framework to mine opinions and utilize that implicit and explicit knowledge in architectural and technology verdicts for software development.

IV. STACKEXCHANGE AND SO AS CROWDSOURCING COMMUNITY

StackExchange is a large network of Q&A community websites and each covers a particular topic at vast level like technology, science, business, etc. Currently, StackExchange is ranked at 170 in global traffic graph. It covers 119 topic websites and as well as 119 Meta websites. StackExchange is a platform where users create, vote and edit questions and answers and filter the answering posts using popularity voting mechanism. It also utilizes gamification to boost up user participation and game design elements.

SO was created in 2008 as part of StackExchange network and at now it is the most famous and collaborative website in the SE network. It is a free Q&A platform which facilitates exchange of knowledge among fresh, experts and experienced programmers. SO have over 3.5 million users registered with it. Since 2008, almost over 8 million questions and over 14 million answers have been posted on the SO website. All these posts have, collectively, turned the SO into a large repository of computer programming, software development and other related knowledge. This is an evidence about popular usage of SO for discussions and exchange of information about a particular technology and revealed that SO encompasses a wide spectrum of technologies. Almost 7,000 questions are posted on the SO website daily. Furthermore, SO maintains a complete record of each user including ones badges, points, and scores which may be utilized for various research purposes. Attributes of SO:

- Extensive coverage and comprehensiveness
- Up-to-dateness
- Rich Description.

V. CONCEPTUAL FRAMEWORK

A. Implicit Knowledge Mining/ Knowledge Discovery from Community Data

1) Tacit or implicit knowledge is the kind of knowledge that is not easy to transfer from person to person by means of writing it down or verbalizing it. Here in this paper, implicit knowledge refers to the useful information that is to be extracted from crowdsourcing Q&A platform and then to be used to make architectural and technology decisions for software projects. The knowledge explored from questions and answers posts posted on SO is implicit because data is posted for particular user in a specified sense but here we are using SO data to make different decisions of architecture and technology in software development process. Hence the knowledge extracted from such data can be called as implicit knowledge.

2) *Determine Knowledge Domain / Q&A Community*: The first and the foremost step involve in knowledge discovery is determination of knowledge domain. This study focuses the two main domains of software engineering, architecture and technology decisions and risk involved with these aspects of software engineering projects. As mentioned above, architecture and technology verdicts bear highest importance because any loop-hole in any of these decisions often poses risks in almost every phase of software development. Therefore, for right decision right knowledge is required to achieve right goal.

3) *Select Data Mart*: SO is among biggest Q&A networks online. One can use the data dump for research and mining MSR which contains all data since 2008 from inception of SO. Here the term data mart is used for specific kind of posts posted over SO. Data marts can be distinguish among IT, CS, software development, tools, technology, architect, programming, framework, etc. this paper selects architectural and technology data marts for data mining as shown in Fig. 2. Data marts can be import as XML data into database with ease for with further processing. The data dump includes all the questions and their answers including partially anonymized user data, user tags and actions logs and their rewarded reputation points.

4) *Data Preprocessing*: Creation of Target Data: It is selection of data of interest from data set.

Data cleaning: Data cleaning is process of making the selected or targeted data by removing the noisy, redundant and irrelevant data [14]. The ability to understand and to correct quality of data is imperative in getting to accurate final analysis. Data cleaning in data mining gives the user an insight to discover inaccurate or incomplete data before analysis phase.

Data reduction and transformation: This step involves the cleaned data set to be reduced by choosing dimension of interest as in this study, architecture and technology, and transforming to the format that can be easily and properly interpreted by data mining methods.

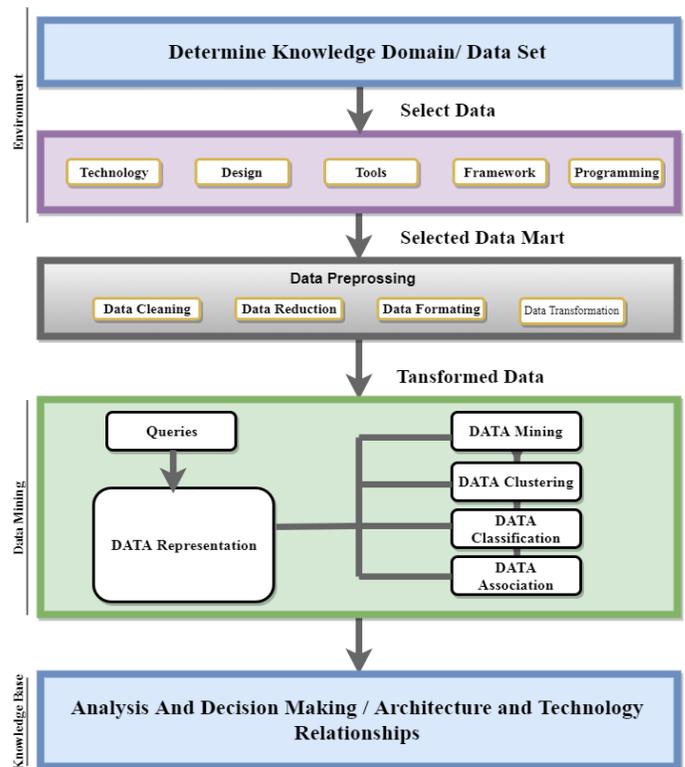


Fig. 2. Implicit knowledge mining framework.

5) *Queries*: The framework provides two different ways to extract usable data for knowledge presentation. First and short cut method is querying the CSV formatted data. CSV formatted data file is opened in excel and query is made to filter desired data. This is simple way just to get quick and to the point data. Limitation of query is that it does not give relationship and association among data that is the ultimate purpose of the study. Hence data mining is applied on formatted data to find knowledge patterns that are not possible in CSV query based method.

Application of data mining: According to knowledge determinants and goals of knowledge discovery, any of the mining intelligent methods can be applied according to data mining expertise and needs of knowledge discovery process. Data patterns are result of data mining process.

Association: Association technique is also named as relational technique. Data mining discovers patterns of relationship among data set in the transaction through if/then statements that help to unveil relationship between seemingly unrelated data in relational database or some other information repository this paper ponders to find association rules to find architectural and technology relationships.

Classification: This method classifies the data set into predefined classes or concepts. There are several mathematical techniques like decision-tree, neural network and other are utilized to classify the items in data set.

Clustering: Clustering is grouping technique of data mining that organizes the items of similar characteristics in clusters whose classes are known or unknown.

Regression: This method finds relationship between the data variables like dependent and independent variables in dataset.

Sequential patterns: This analysis discovers the regular and frequent patterns in data set.

6) *Evaluation of patterns:* Patterns discovered by data mining methods are then analyzed and evaluated according to the knowledge.

7) *Knowledge presentation:* Presentation of knowledge patterns to represent the decisions in form of metaphors like graphs. This is visualization stage of knowledge so that the discovered knowledge might be used in decision making of architectural and technology verdicts.

8) The knowledge discovered by using machine learning and data mining techniques using dump data of SO can be termed as Implicit knowledge because it is an inference from a large pool of data. The knowledge is not delivered in response of an explicit query rather it is extracted from dump data, formatted, cleaned and then summarized in form of patterns to make it useful in particular sense for which it might not be produced originally.

B. Explicit Knowledge Discovery / Knowledge Discovery from Experts

Contrary to implicit knowledge, explicit knowledge is directly discovered from experts in response of direct and to the point problem statement. This study proposes a three-tier or tri level mechanism to find experts from data set. Here, a question arises that to whom should put a problem statement to discover explicit knowledge. Each post in data set point a user. All these users are producer of knowledge repository by asking and answering question knowledge sharing communities like SO. Experts finding among Crowd or Expert Finding Mechanism is given as follows:

1) *Post acceptance level:* Each post on SO gets acceptance or rejections through ups and downs respectively from registered users across the community. Post quality, reliability and authenticity is measured from its ups and downs. A post with positive response reflects that user, who posted that post or comment, bears good knowledge about that domain.

2) *User badge:* SO also involve in gamification of community to attract more people to register themselves and to gear up activities of registered users across the community to share knowledge. Each badge reflects knowledge and experience level of SO users according to criteria of each badge set by community admin. Each user across the platform strive his best to achieve highest badge through raising quality questions and producing quality answers of difficult problems. This study utilizes user badge criteria to select user for gathering explicit knowledge about technology and architecture as shown in Fig. 3.

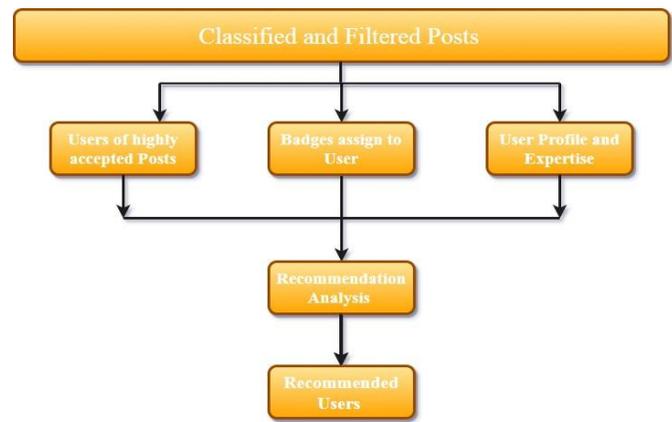


Fig. 3. Explicit knowledge mining framework.

3) *User profile and expertise:* Each registered user across the SO distinguishes himself through unique capabilities and expertise. User profile on the SO is a direct way to determine user expertise in required domain. This is the third criteria to select a user from data posts for explicit knowledge mining.

This study presents a triangular approach in selection of SO user for explicit knowledge gathering is done through three evaluation criterion. Only User posts in selected domain, user badge or user profile is not enough to select a user for inquiring about architectural or technology knowledge rather this technique gives weightage to the all three aspects of user selection framework to select right person.

VI. CONCLUSION

Goal of this study is to unveil the potential of crowdsourced experts and communities across the internet to support architectural and design management approaches and decision makings with a more efficient method for capturing architecture and technology related knowledge. Moreover, this paper distinguishes the crowdsourcing approach of software development as more suitable and rip with knowledge. The proposed framework of implicit and explicit knowledge checks if SO could be viable source for reusable architecture and design knowledge. The framework utilizes the data mining techniques to support software development team to get wide spectrum of opinions in form knowledge patterns discovered by the crowdsourcing knowledge mining framework. This study focuses both qualitative and quantitative aspects of SO knowledge repository. In our future work, we will implement this framework to find architectural and technology relationships and pattern through classification, clustering and association techniques. Moreover, we will extend our study to remaining core knowledge base area over SO and on other Q&A communities.

REFERENCES

- [1] (n.d.). Retrieved from https://en.wikipedia.org/wiki/Educational_data_mining
- [2] Achmad Arwan, S. R. (2015). Source Code Retrieval on StackOverflow Using LDA. 3rd International Conference on Information and Communication Technology (ICoICT), (pp. 295-299).
- [3] Arash Joorabchi, M. E. (2015, August). Text mining stackoverflow, An insight into challenges and subject-related difficulties faced by computer

- science learners. *Journal of Enterprise Information*, Vol. 29 No. 2, 2016, 255-275.
- [4] ASLI SARI, G. I. (2017). An Overview of Crowdsourcing Concepts in Software Engineering. *International Journal of Computers*, 106-114.
- [5] Bosch, A. J. (2005). Software architecture as a set of architectural design decisions. *WICSA*, (pp. 109-120). A. Jansen and J. Bosch, “,” in *WICSA*, 2005, pp. 109–120.
- [6] Fabio Calefato, F. L. (2015). Mining Successful Answers in Stack Overflow.
- [7] J. L. C. Ramos, R. E. (2016). A Comparative Study between Clustering Methods in Educational Data Mining. *IEEE LATIN AMERICA TRANSACTIONS*, (pp. 355-3361).
- [8] Juan Yang, S. P. (2016). Finding Experts in Community Question Answering Based on Topic-Sensitive Link Analysis. *IEEE First International Conference on Data Science in Cyberspace*, (pp. 55-60).
- [9] Katiyar, N. U. (2014). A Survey on the Classification. *International Journal of Computer Applications Technology and Research*, 725-728.
- [10] L. Bass, P. C. (2012). *Software Architecture in Practice* (3rd ed.). (A.-W. Professional, Ed.)
- [11] L. Bass, P. C. (2012). *Software Architecture in Practice*. Addison-Wesley Professional.
- [12] Latifa Guerrouj, O. B. (2016). Software Analytics: Challenges and Opportunities. *IEEE/ACM 38th IEEE International Conference on Software Engineering Companion*, (pp. 902-903). Austin, USA.
- [13] M Amala Jayanthi, R. L. (2016). Research Contemplate on Educational Data Mining. *IEEE International Conference on Advances in Computer Applications (ICACA)*, (pp. 110-114).
- [14] M Amala Jayanthi, R. L. (2016). Research Contemplate on Educational Data Mining. *IEEE International Conference on Advances in Computer Applications (ICACA)*, (pp. 110-114).
- [15] M. Soliman, M. R. (2015). Enriching architecture knowledge with technology design decisions. *WICSA*.
- [16] Maalej, D. P. (2013). How do open source communities blog. *Empirical Software Engineering*, vol. 18, no. 6, (pp. 1090–1124).
- [17] McConnell, S. (2004). *Code Complete* (2nd ed.). (Microsoft Press.
- [18] Meiyappan Nagappan, E. S. (n.d.). *Future Trends in Software Engineering Research* for.
- [19] Mohamed Soliman, M. G. (2016). Architectural Knowledge for Technology Decisions in Developer Communities. *IEEE/IFIP Conference on Software Architecture*, (pp. 128-133).
- [20] Muhammad Ahsanuzamman, M. A. (2016). Mining Duplicate Questions in Stack Overflow. *IEEE/ACM 13th Working Conference on Mining Software Repositories*, (pp. 402-412).
- [21] Neelamadhav Gantayat, P. D. (2015). The Synergy Between Voting and Acceptance of Answers on StackOverflow, or the Lack thereof. *12th Working Conference on Mining Software Repositories*, (pp. 406-409).
- [22] O. Zimmermann, J. K. (2009). Managing architectural decision models with dependency relations, integrity constraints, and production rules,”. *Journal of Systems and Software*, vol. 82, 1249–1267.
- [23] Philipp Berger, P. H.-P. (2016). A Journey of Bounty Hunters: Analyzing the Influence of Reward Systems on StackOverflow Question Response Times. *IEEE/WIC/ACM International Conference on Web Intelligence*, (pp. 644-649).
- [24] Piatetsky-Shapiro, G. (2017). *Advances in Knowledge Discovery and Data Mining*. (G. P.-S. Usama M. Fayyad, Ed.) America: American Association for Artificial Intelligence Press.
- [25] Questions, M.-c. M.-t. (2015). Jos´e R. Cede˜no Gonz´alez, Juan J. Flores Romeroy, Mario Graff Guerreroz and Felix Calder´onx. *IEEE / ROPEC 2015 - Computing*.
- [26] Thiago B. Procaci, B. P.-F. (2016). Finding Topical Experts in Question & Answer Communities. *IEEE 16th International Conference on Advanced Learning Technologies*, (pp. 407-411).
- [27] Yunxiang Xiongy, Z. M. (2017). Mining Developer Behavior Across GitHub and StackOverflow.