# Enhanced Approach To Find Topic Experts In Forum

**Mayur Jakate[1], Lata Ragha[2], Jaydip Damle[3]**
[1] Dept of Computer Engineering
[2]Professor and HoD, Dept of Computer Engineering
[3]Founder and MD
[1] Terna Engineering College, Nerul, Navi-Mumbai
[2]Fr. C. Rodrigues Institute of Technology, Vashi, Navi-Mumbai
[3]PollenGrain Technology and Soluction Private LTD, Navi-Mumbai

***Abstract-*** *Social media Networks and micro blogging are now used all over the world. It has become very difficult to identify which among this information are important for us. So, Expert finding has become a hot topic on social network. Information from Experts is considered to be trustworthy and relevant to satisfy our need but several attempts use the relations among users and list of label query which extracted from given query. In literature the authors used to find the expert from twitter corpus that is specific to twitter data. We propose a method that uses stack overflow API data set to find an expert of different topics which are at the pic of micro blogging. We use different traditional method to explore the topics and user among their relationship such as semi supervised algorithm; page rank algorithm, HITS Algorithm and LDA Algorithm Our work is divided into three stages such as data set formation, extract complete information form dataset and apply suitable algorithm to get desired output.*

***Keywords***- Expert search, micro-blogging, graph-based ranking, PageRank computation, Hyperlink-Induced Topic Search (HITS) algorithm and Social Network Analysis (SNA),Community Expertise Network (CEN). content analysis .

## I. INTRODUCTION

Expert finding is one of the exchanging points in internet based life with result to get the best profile of the individual having pertinent learning on the specific subject. As of late master discovering issue pull in the consideration of via web-based networking media, for example, forum like Stack overflow, it highlights questions and replies on a wide scope of subjects given by various individuals [1]. As we probably aware, learning sharing is a standout amongst the most critical utilizations of online networks in virtual space of the Internet. For instance in [2] and [3] number of factor that effect of information partaking in online networks get recognized.

In online networks we don't realize information dimension of User so estimation of answer and remark are hazy. This is the greatest test in smaller scale blogging networks. By distinguishing information dimension of every User we can get the most significant response for the appropriate response posted by the User.

Information dimensions of Users are misty and along these lines estimation of answers and remarks are obscure. Also this is one of the greatest difficulties in online networks. By deciding learning dimension of an individual User and discovering specialists in an online network, we can decide the appropriate responses which are increasingly solid.
In online networks, substantial volume of data identified with inquiries posted by Users is another essential test that makes addresses concealed by specialists who can react to them. In this way, the reaction time for reacting addresses takes longer. By utilizing Expert finding techniques and making recommender frameworks dependent on these strategies, questions can be presented to people who have sufficient information to react them. Also, it is conceivable to make basic inquiries concealed to specialists; in this manner covering them to not squander their very own occasions for responding to straightforward inquiries.

As previously mentioned, in online networks the strategies for Expert finding and deciding skill dimension of Users are very critical since the significant volume of data can be used. As a rule, there are two fundamental methodologies for discovering specialists.

- The main methodology concentrates on Social Network Analysis (SNA)
- The second methodology accentuations on content analysis (CI)

Both of the previously mentioned methodologies have a few deformities. For example in the approaches based on social network analysis, content of users' messages are not considered and users may send many irrelevant or empty messages. Thusly this expands Users' correspondences and may cause a few slip-ups in discovering specialists. Furthermore, in the methodologies dependent on substance examination, interchanges between people are not considered and there is no qualification among traded reactions. Be that as it may, ability of a person who reacts to a specialist might be able to easily compare to one who reacts to an ordinary or learner User. Therefore, for higher precision the mixture strategies ought to be connected for to find Expert

## II. RELATED WORK

Prior to now, most investigates in the field of Expert finding had been done in associations. Anyway at present, more interests are appeared for discovering specialists in virtual condition, particularly in informal organizations and online networks. A few investigations depicted in this segment are identified with the association and a few others are identified with the Internet.

Master Finder Systems (MFS's) are considered as a feature of the CSCW frameworks (Computer Supported Cooperative Work). For instance, they are intended to discover individuals with extraordinary skill or learning in online networks, who have capacity to react to a specific inquiry. Also, MFS's set up a critical class of the recommender frameworks [12].

As referenced in past area, there are two principle approaches for discovering specialists. The principal approach centers around Social Network Analysis (SNA) and the second methodology accentuations on substance examination. Considering the main methodology so far system based positioning techniques and calculations have been utilized to distinguish specialists, for example PageRank and HITS.

In any network-based ranking algorithm individuals are considered as nodes and relationship between them are considered as links of a network. At the point when data is traded between two hubs a connection between them is molded. For instance,

On the off chance that individual A reacts to individual B a connection from A is drown to B. In the wake of making every single imaginable connection between all people a system which is known as the Expertise Network (EN) is built up [12]. At the present time arrange based positioning calculations can discover critical hubs and show the specialists. For instance in [12], they planned to discover distinctive strategies to distinguish and rank specialists by molding EN, and after that the execution of these techniques has been thought about. SNA was utilized for discovering specialists. In [13], the point was to discover specialists in Meta Filter online network utilizing SNA approach. In [17] specialists were found by methods for SNA in Friend feed online network. In [18], Thiago Baesso and his partners have broken down some chart measurements and calculations so as to discovering specialists in various gathering.

The second methodology for discovering specialists in online networks is focused on content analysis. In this methodology, content mining systems are used. For this reason the substance of messages sent by Users are investigated and dependent on data separated from instant messages, User's learning model or a likelihood model of the connection between the User and the messages are created. Learning model and likelihood model can be used to recognize master Users. For instance in [1], User's information demonstrating has been utilized to recognize specialists.

There are not many examinations that have hybrid methods for discovering master. For instance, in [13], a hybrid method has been utilized for discovering specialists in an interpersonal organization of scientists. In [14], traded messages have been utilized to set up both previously mentioned methodologies known as informal organization examination and content analysis. In [15], Bozzon and his associates have presented a technique for discovering master in informal organizations dependent on content investigation and interpersonal organization setting. What's more in [16] and [17], by methods for consolidating highlights of the two methodologies, specialists have been recognized.

It is mentionable that so far a portion of the works in the field of Expert finding are tied in with using Expert finding to help different applications. Planning recommender frameworks is a standout amongst the most vital applications that use master discovering calculations. For instance, in [12], a recommender framework for interface personalization of Stack Overflow is given. In [17] a recommender framework system is accommodated improving knowledge sharing in online forums. Another case of using master discovering frameworks is to take care of complex issues in associations

In our examination a hybrid method is displayed for Expert finding in online networks. By methods for this technique specialists can be distinguished with high exactness. In our technique content investigation is performed by dissecting idea guide and interpersonal organization examination depends on PageRank calculation we additionally use HITS calculation. Subtleties of the proposed strategy will be portrayed beneath.

## III. PROPOSED METHODOLOGY

In this area, structure of the proposed strategy is depicted. Our proposed system has four primary advances.

- Data Extraction
- Topic Identification
- Social network analysis
- User Ranking and Experts Finding

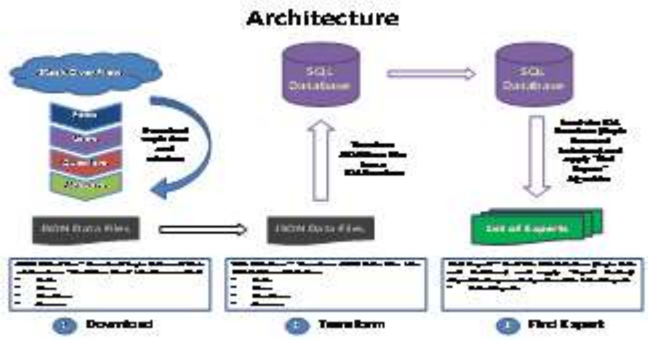


Figure 1.Framework of proposed strategy demonstrates ventures of the system and subtleties for each progression.

Figure 1, demonstrates the proposed architecture, in the accompanying we quickly clarify the proposed technique by a precedent and after that subtleties for each progression will be depicted.

### A. Information extraction

In this progression, fundamental data pertinent to the User's profile and the User's posts are separated. Subtleties of data extraction from every one of the referenced sources are tended to as pursues.

At first, the structure of web in Stack Overflow online network is considered. This is important to discover delivers identified with Users' data in JSON forma, User's messages and other attractive data. For this reason, the URL addresses were inspected and rationale next to URLs was found. At that point data identified with the Users was removed. The most critical data identified with every User profile incorporates the accompanying things:

User ID that is a novel identifier

- User handle that is a novel name
- User's posts
- Number of User's posts
- Number of User questions

Since examination of the a large number of messages assembled in the site is unimaginable and the greater part of the Users are not at present dynamic, our investigation was centered around the individuals who were generally dynamic. For this reason, first, the data document was changed over to an organization which could be entered as contribution to the ETL (Extract-Transform-Load). By utilizing the ETL, data of Users were gone into SQL Server to effectively run question on it and concentrate wanted Users.

In this examination, we have considered Users as dynamic Users on the off chance that they have in excess of 100 posts. The complete number of dynamic Users in the Stack Overflow discussion till December 2018.

### B. Extraction of User's posts

In the wake of extricating User data, strings of inquiry/answer can be removed. It was important to extricate as much as strings of inquiry/answer since Stack Overflow discussion does not enable access to a particular User's posts. In this way, first, all strings of inquiry/answer must be cleared and afterward those presents related on explicit User can be extricated from these strings. In posts extraction, source code and different statements ought to be expelled.

At long last an information structure is made for every User that contains the accompanying data:

- Thread ID that user submits a message to it
- Subject of thread
- If the post is kind of response, then who has received a response?
- The content of messages that have been sent by each user

### C. Topic Identification

In this progression Topics of the users'' posts are removed and after that committed score to every users is determined dependent on Topics map. Subtleties of this progression are portrayed straightaway.

**Extraction of Topics**

At first, Topics in traded questions and replies of every user are separated. Since these Topics ought to be removed and contrasted and Stack Overflow subject rundown for this venture we think about just five hotly debated issues recorded as Java, Python, JavaScript, C#, .NET. For this it is important to extricate all hubs of Stack Overflow ahead of time. For every hub, different catchphrases with same importance are considered too.

Subsequent to making an information structure for the Stack Overflow, Topics of traded presents are extricated concurring on the Topics map. Toward the finish of this stage, every user has an information structure which incorporates

Topics of each inquiry and watchwords important to reaction presented on that question.

**Ascertaining separation between Topics in the Topics map**

To figure separate between the Topics found from reactions and the Topics derived from inquiries ought to be removed. In such manner, the most limited separation from one Topic to every topic in the Topics map is extricated. For this reason it is important to draw a chart of relations between Topics. The quantity of Topics molding the Topics map was 5. In the wake of illustration a diagram, by utilizing Dijkstra's calculation, the briefest way between any two hubs in an undirected chart was determined. The yield of this stage is a two-dimensional lattice that holds remove between Topics.

**Computing loads of the Topics in users' reactions**

At this stage, the normal separation between every theme accordingly and all Topics in the inquiry is determined by condition (2).

$$AvgDist(R) = \frac{\sum_{Q \in Questions}^{N} (Dist(R,Q))}{N} \quad (2)$$

Where R is Topic of the reaction, Q is Topics of the inquiry, Questions are on the whole Topics in the inquiry, Dist( R,Q ) is remove between Topics R in the reaction, and Topics Q in the inquiry and N is the quantity of Topics in the inquiry.

In the numerator of condition (2), whole of separations identified with the Topics R in the reaction from the majority of the Topics in the inquiry has been determined.

AvgDist (R), has been determined for all Topics of the reaction. At last every subject in the reaction has been supplanted with normal separation of the Topics from all Topics of the inquiry.

**Figuring positioning scores**

At this stage, the last scores for users has been determined dependent on condition (3).

$$Score(I) = \sum M \in Messages \sum R \in Responses = \left(\alpha . \mathrm{Rep}(\mathrm{R}) \frac{\beta}{\beta AvgDist(R)}\right) \quad (3)$$

Where Score ( I ) is score of users I, Messages will be messages of users I, Responses are Topics in the reaction of the message M, Rep (R) is the quantity of emphases of Topics R in the reaction of the message M and Weight (R) is weight of Topics R in the reaction of the message M.

In view of condition (3), clearly score of every user has been determined dependent on two measures. One is the quantity of cycles of Topics which is utilized by users accordingly. Another measure is likeness between Topics accordingly and Topics being referred to. Since the normal separation of Topics accordingly from the Topics being referred to be conversely relative to closeness, consequently AvgDist are utilized contrarily in ascertaining score for users. $\alpha$ and $\beta$ are coefficients with qualities somewhere in the range of 0 and 1. $\alpha$ demonstrates the effect of the quantity of Topics which are in the users reaction, and $\beta$ shows the effect of separation between Topics in users reaction and Topics being referred to. In this examination, the ideal qualities for these coefficients are determined. To accomplish the ideal qualities state space is hunt by changing 0.01 interims down $\alpha$ and $\beta$. The ideal qualities got for these coefficients are similarly 0.5. By utilizing these coefficients, the best connection between the scores acquired from the proposed strategy and the scores given by Stack Overflow, is determined.

**D. Social network analysis**

In this progression, we portray how the system is made among users and after that how the PageRank calculation works on the system of users and scores are given to them.

**Create network of users**

Online people group typically have a talk string structure. A user posts a topic or question, and afterward some different users present an answer on an inquiry or take an interest in an exchange. By utilizing these posting/answering strings in a network, we can make a post-answer system of users, as appeared in Figure 2.

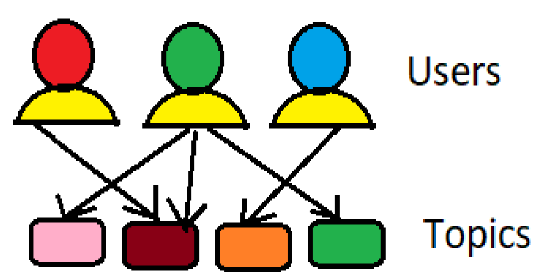


Figure 2.Post-answer arranges. In this figure dashed bolts demonstrate subjects or questions and reactions are appeared strong dark bolts.

At the point when a user answering to an inquiry or a theme, typically this shows the respondent user has a larger amount of ability regarding the matter than the individual who makes the inquiry. Interfacing examiners to respondents by directional bolts from examiners to respondents makes a system which is called Community Expertise Network (CEN). The CEN made from Figure 2, is appeared in Figure 3.

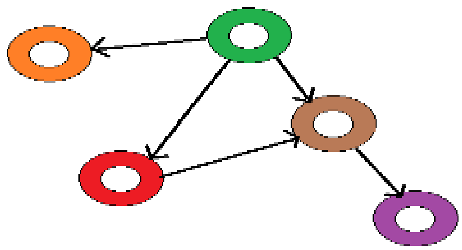

Figure 3.Expertise networks A Community Expertise Network (CEN) made from the Post-Reply Network.

Informal organization investigation for positioning users in online networks depends on CEN which is made among users. Inbound connect to a hub in CEN shows that the user connected to this hub answers to the user who is on opposite side of the connection. Whatever the quantity of inbound connects to a hub is more, demonstrates that connected user to that hub has higher skill.

**2.PageRank computation**

We describe the difference between the community expertise network and the network which is created in the web. We also describe how the PageRank algorithm is used for giving scores to users in a community expertise network.

In a community expertise network (CEN) it may be possible to have more than one link between two nodes, however in a network which is created in the web, only one link can be existed in each direction.

Suppose a CEN has been formed between three users A, B and C as shown in Figure 4. Transition probability matrix for Figure 4, is shown in Figure 5. Such as PageRank Algorithms row of the matrix replaced with 1/n if all values are zero, n is the number of nodes in the network.

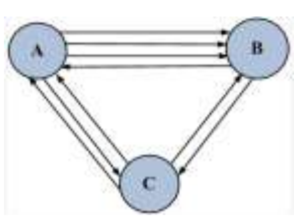


Figure 4.Sample of expertise network. A CEN has been formed between three users as nodes, questions and answers between users have been showed as arrows.

$$\begin{array}{cc} & \begin{array}{ccc} A & B & C \end{array} \\ \begin{array}{c} A \\ B \\ C \end{array} & \begin{bmatrix} 0 & \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{2}{3} & \frac{1}{3} & 0 \end{bmatrix} \end{array}$$

Figure 5.Weights for sample CEN. This matrix shows weights of arrows between users as nodes in the Sample CEN.

Weight for x and y calculated as equation (4). In equation (4), n is the total number of node or users and R xy is the out links from x to y and R xi is the outbound links from x to i.

$$\text{Weight}(xy) = \frac{Rxy}{\Sigma = \Sigma_{i=1}^{N} Rxi} \quad (4)$$

## IV. IMPLEMENTATION

**1. Data extraction Phase**

We use Stack overflow website as a base for this project, we analysis and discover expert list from the data available on this site. Stack overflow provide their API (Application programming interface) so that we can extract their data.

Initially data available as a link so we convert data into JSON format with respect to their file.JSON stands for JavaScript Object Notation. JSON is a lightweight format for storing and transporting data. JSON is often used when data is sent from a server to a web page. JSON is "self-describing" and easy to understand.

For this project we use four JSON file as a input, below are the list of file name which we are used in this project.

- Post.json
- Question.json
- Answer.json
- User.json

**2. Data Load Phase**

When data in store into respective json file, next step is to load the data into database for further processing, we use SQL Server as a backend database for this project which is mention as below figure


Figure 7.Data load on home page

**3. Final Output Phase**

We extract data from source then we load the data into our database, after data analyses we applying algorithms we get the desire output in the form of list of people having expertise in their respective area.


Figure 8. List of expert with the rank in their subject area

## IV. RESULT ANALYSIS

**Figure out scores**

After making of progress likelihood network dependent on CEN we different damping factor by all the lattice components. Damping factor typically set to 0.8, we likewise have utilized this esteem.

In the PageRank calculation 1-d is the likely of transport activity. In transport, surfer can bounce into every hub in the system. The goal of transport task is chosen haphazardly. In the event that a hub doesn't have any yield connect, at that point the transport activity will be finished with the likelihood of 1/n in which n is the quantity of hubs/users in the network aptitude organize, else it will be finished by likelihood of 1-d. At long last the measure of (1-d)/n will be added to all components of progress likelihood grid. Subsequent to running PageRank calculation scores will be resolved.

**Positioning users and discover specialists**

In this progression, scores of substance examination and interpersonal organization investigation are joined and the last scores are acquired. Subtleties of this progression are depicted in the accompanying.

**Institutionalized scores**

For consolidating join examination and substance investigation approaches, scores that are given to users must be institutionalized. Scores of users in system investigation dependent on PageRank calculation are somewhere in the range of 0 and 1. Anyway scores in substance examination not restricted to a specific range. Scores of substance investigation will likewise being the scope of 0 and 1, utilizing condition (5). In this condition, Max and Min, separately, are the most noteworthy and least scores.

$$Stand\ (value) = (value - Min) / (Max - Min) \quad (5)$$

**Figure out last scores**

After scores got with the two methodologies, were a similar range, scores were consolidated to get a last score for every client, utilizing condition (6).

$$Score\ (p) = T.Score(T) + N.Score(N) \quad (6)$$

In condition (6), Score (T) is score of substance investigation and Score (N) is score of connection examination. T and N, individually are considered as weight for substance investigation and connection examination. At conclusive, users are positioned with Score (p) qualities and top users are resolved as master users.

To assess the proposed technique, all the subsection of Stack Overflow is utilized. To begin with, number of reactions for every subsection was determined and topics which the quantity of reactions for them is under 100 have been avoided. At long last, 5 topics have remained.

Spearman connection between our outcomes and scores arranged by Stack Overflow was determined independently for the 5 topics and the whole Stack Overflow, the general relationship was determined by taking the normal of these relationships. The general relationship was determined with various qualities for weight of substance investigation and interpersonal organization examination in condition (6). Table 1, demonstrates these relationships.

Table 2.Bold numbers indicate that when "T 0.2-N 0.8" Spearman correlation between our results and scores prepared by Stack Overflow online community reaches the max. value

| Weights of content analysis (t) and social network analysis (n) | Average spearman correlation |
|---|---|
| T 0.0 - N 1.0 | 0.877142549 |
| T 0.1 - N 0.9 | 0.901276748 |
| T 0.2 - N 0.8 | **0.904018451** |
| T 0.3 - N 0.7 | 0.903982375 |
| T 0.4 - N 0.6 | 0.901998248 |

The outcomes demonstrate that the hybrid method is superior to the strategies for substance investigation or informal organization examination alone. As you find in first column of Table 2, when just informal organization examination is utilized normal spearman connection is 0.877 and when just substance investigation is utilized as you find in last line of Table 2, normal spearman relationship is 0.829. Be that as it may, when weight of informal organization examination is 0.8 and content investigation is 0.2, spearman relationship between our outcomes and scores arranged by Stack Overflow achieves most extreme esteem.

Table 3(A):

| **Category** | **NQ** | **NR** | **NU** | **A(Q)** |
|---|---|---|---|---|
| *ALL (Entire StackOverflow)* | 6465 | 345206 | 614 | 10.52 |
| *JAVA* | 367 | 23456 | 254 | 1.44 |
| *C#* | 308 | 4869 | 152 | 2.02 |
| *Python* | 337 | 31096 | 105 | 3.20 |
| *JavaScript* | 415 | 200 | 28 | 14.82 |
| *.NET* | 1657 | 54839 | 150 | 11.04 |

Table 3(B):

| **Category** | **A(R)** | **P(80Q)** | **P(80R)** | ***SpCo*** |
|---|---|---|---|---|
| *ALL* | 562.22 | 9.60 | 3.74 | ***0.96*** |
| *JAVA* | 92.34 | 33.85 | 3.14 | ***0.89*** |
| *C#* | 32.03 | 18.42 | 6.57 | ***1*** |
| *Python* | 296.15 | 12.38 | 3.80 | ***0.91*** |
| *JavaScript* | 148.03 | 10.71 | 7.14 | - |
| *.NET* | 365.59 | 4.66 | 5.33 | ***0.88*** |

Table 3(A) and 3(B) shows detailed information for each subsection of Stack Overflow and entire Stack Overflow separately. Spearman correlation between our results and scores prepared by Stack Overflow has been presented for each subsection; these correlations are for best weights of social network analysis and content analysis.

In Table 3 the abbreviations are defined as:

- NQ: Number of query asked by user
- NR: Number of query response by user
- NU: Number of actively participate users
- A(R): Average number of responses get from per user
- A(Q): Average number of questions get for per user
- P(80Q): Percentage of users who submit 80Percentage of questions
- P(80R): Percentage of users who submit 80Percentageof responses
- SpCo: Spearman correlation taken by our results and scores prepared by Stack Overflow

For some of the topics the number of responses is less than 100, the correlation is not valid for them and is not calculated.In Table 3, the value of NR and A(R) are more important for our study, because the proposed method is based on user's responses and if these numbers are much higher, accuracy will be higher as well.

**Now we will compare our hybrid method with other methods.**

There are some essential strategies which are utilized for examination; these techniques have been portrayed in [12]. We quickly present these essential strategies and correlation our half and half strategy with these techniques in the accompanying.

**AnswerNum:** in this strategy specialists have been related to tallying of answers of one client.

**Indegree:** in this strategy specialists have been related to tallying of clients that one client has sent solutions for their inquiries.

**Z-score:** in the event that one client makes n = a + q posts, q is the quantity of inquiries and the quantity of answers, at that point Z-score has been determined with condition (7).

$$z = \frac{a-q}{\sqrt{a}+b} \qquad (7)$$

On the off chance that Z-score has been considered for the quantity of inquiries one client asked and replied, the strategy called Z-number, and if Z-score has been considered for the quantity of clients one client answered to and got answers from, the technique called Z-degree.

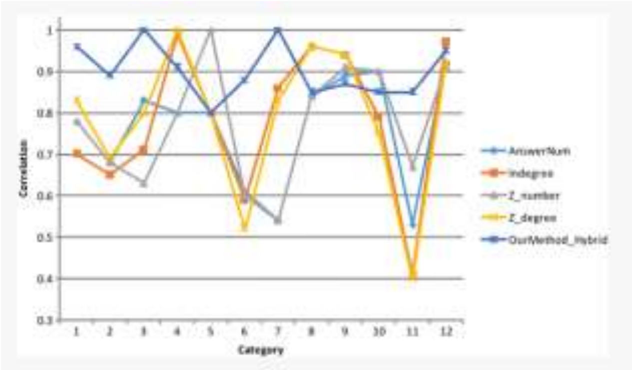

Figure 9.Comparison with different techniques in various classes. This figure demonstrates the consequences of examination between our hybrid technique and fundamental strategies in various classes of StackOverflow.

Figure 9 demonstrates correlation between our hybrid technique and essential strategies in various classifications of Stack Overflowand Figure 10 shows examination between our hybrid strategy and fundamental techniques in normal of all classes of Stack Overflow.

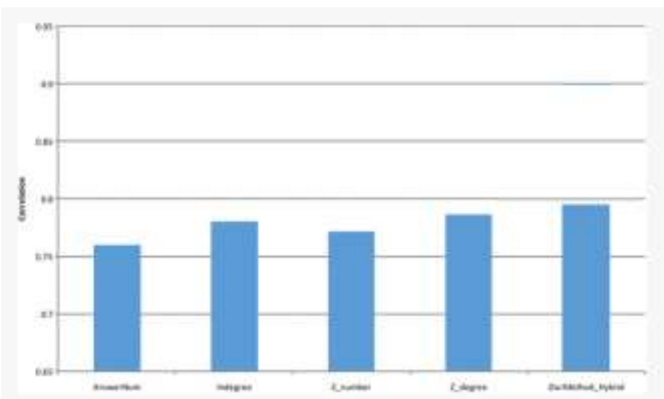

Figure 10.Correlation with different strategies in normal all things considered. This figure demonstrates the aftereffects of correlation between our proposed technique and fundamental strategies in normal of all classifications of Stack Overflow.

As you find in Figure 9 our mixture strategy is superior to different techniques in the greater part of classes of java online network, furthermore as you find in Figure 10 our strategy is superior to different strategies in normal of all classifications of Stack Overflow.

## V. CONCLUSION

In this paper we addressed a problem of topic specific expert finding in forum. We integrated different algorithms and methods to identify topics and users from dataset.Our method aimed to assign similar ranking scores to the similar users, and meanwhile the ranking scores are subjected to the supervised information from the input data which is provided. Based on the computed ranking scores, we selected the top-N relevant users for any given topic.

## REFERENCES

[1] Kardan A, Garakani M, Bahrani B (2010), “A method to automatically construct a user knowledge model in a forum environment”, Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval, pp 717–718

[2] Chen I (2007), “The factors influencing members”, continuance intentions in professional virtual communities - A longitudinal study”, Journal of Information Science, 33(4), pp 451–467

[3] Zhang U, Ackerman M, Adamic L, Nam K (2007) “QuME: a mechanism to support expertise finding in online help-seeking communities”, Proceedings of the 20th annual ACM symposium on User interface software and technology, pp 111–114

[4] J. Weng, E.-P.Lim, J. Jiang, and Q. He (2010) “Twitterrank: Finding topic-sensitive influential Twitterers”, Proceedings of ACM International Conf. on Web Search Data Mining, pp. 261–270.

[5] Banu D. Gunel, Pinar Senkul (2013), “Integrating Semantic Tagging with Popularity-Based Page Rank for Next Page Prediction”, Computer and Information Sciences III.

[6] Weiming Yang (November 2016), “An Improved HITS Algorithm Based on Analysis of Web Page Links and Web Content Similarity”, International Conference on Cyber worlds.

[7] FabricioAparecidoBreve, Daniel Carlos GuimarãesPedronette (2016), “Combined unsupervised and semi-supervised learning for data classification”, IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP).

[8] Sanjeev Patel, KritiKhanna, Vishnu Sharma (2016), “Documents ranking using new learning approach”, International Conference on Computing, Communication and Automation (ICCCA)

[9] Link for dataset:https://api.stackexchange.com/docs

[10] http://www.aclweb.org/anthology/J97-1003

[11] https://isi.edu/integration/papers/bowu12-iiweb.pdf

[12] Zhang J, Ackerman M, Adamic L (2007),“Expertise networks in online communities: structure and algorithms”, Proceedings of the 16th international conference on World Wide Web (WWW '07).,pp 221–230

[13] Zhang J, Tang J, Li J (2007),“Expert Finding in a Social Network, Advances in Databases: Concepts, Systems and

Applications”, Springer, Berlin Heidelberg, vol. 4443, pp. 1066-1069

[14] Campbell C, Maglio P, Cozzi A, Dom B (2003),“Expertise identification using email communications”, Proceedings of the twelfth international conference on Information and knowledge management (CIKM '03)., pp 528–531

[15] Bozzon A, Brambilla, Ceri S, Silvestri M, Vesci G (2013),“Choosing the right crowd: expert finding in social networks”, Proceedings of the 16th International Conference on Extending Database Technology (EDBT '13). ACM, New York, NY, USA, pp 637–648

[16] Serdyukov P, Rode H, Hiemstra D (2008),“Modeling multi-step relevance propagation for expert finding”, Proceedings of the 17th ACM conference on Information and knowledge management (CIKM '08)., pp 1133–1142

[17] Li Y, Liao T, Lai C(2012),“A social recommender mechanism for improving knowledge sharing in online forums”, Information Process Management, 48(5), pp.978–994