Expert Finding System using Latent Effort Ranking in Academic Social Networks

Sobha K. Rani
Department of Information Technology, MVGR College of Engineering, Vizianagaram, 535005, India
Email: sobharani.t@gmail.com

KVSVN Raju, V. Valli Kumari
Professor & Director (R&D), ANITS, Visakhapatnam, 531162, India
Professor, Department of CSSE, AU College of Engineering, Andhra University, Visakhapatnam, 530003, India
Email: {kvsvnraju, vallikumari}@gmail.com

Abstract- The dynamic nature of social network and the influence it has on the provision of immediate solutions to a simple task made their usage prominent and dependable. Whether it is a task of getting a solution to a trivial problem or buying a gadget online or any other task that involves collaborative effort, interacting with people across the globe, the immediate elucidation that comes into anyone's mind is the social network. Question Answer systems, Feedback systems, Recommender systems, Reviewer Systems are some of the frequently needed applications that are used by people for taking a decision on performing a day to day task. Experts are needed to maintain such systems which will be helpful for the overall development of the web communities. Finding an expert who can do justice for a question involving multiple domain knowledge is a difficult task. This paper deal with an expert finding approach that involves extraction of expertise that is hidden in the profile documents and publications of a researcher who is a member of academic social network. Keywords extracted from an expert's profile are correlated against index terms of the domain of expertise and the experts are ranked in the respective domains. This approach emphasizes on text mining to retrieve prominent keywords from publications of a researcher to identify his expertise and visualizes the result after statistical analysis.

Index Terms- Expertise Matching, Latent Semantic Analysis, Text Mining, Social Networks, Information Retrieval Systems.

I. INTRODUCTION

Knowledge database is required to be maintained by different organizations in their R&D department or Academic Committee, where they continuously need to interact with experts from different engineering disciplines. As some topics need inter-disciplinary domain knowledge, experts who can be referred for multiple fields of interest are to be identified. Expertise mining from social network helps us identify potential members who would offer their knowledge upon request in their fields of interest. In the context of research publications, it is necessary to identify reviewers who can review scientific publications that deal with topics from multiple but interrelated domains. Hence, there is always a need to identify an expert with prominent and versatile knowledge in order to enhance the collaborative effort in the research front.

Academic Social networks like Arnetminer[14], Microsoft Research[15] etc. are playing major role in simulating the collaborative effort offered by experts across the globe. They offer services like topic search, journal search, person search and represent the relation amongst them using graphs. A social network is represented as a graph $G(V,E)$, where the vertices/nodes are the members and the edge represents the relation of co-authorship as depicted by Academic Research of Microsoft [15], is shown in Fig.1. The edge weight represents the frequency of co-authorship.
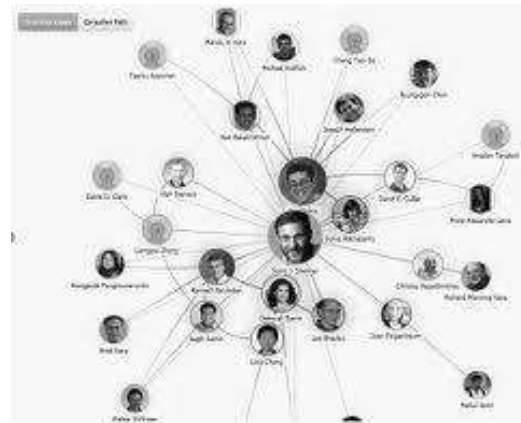

Fig.1. Co-authorship of an Academic Social Network

The authors are sorted by indices like h-index, g-index based on their publications and their citations. However, relevance of publications with a specific domain is not exactly identified, but will be categorized based on their keywords. Different research works [3], [4], [5] were made to identify multi-domain experts based on their profile pages, and their specified research interests. With a thorough analysis made on the concept of publications and versatile research domains, it is found necessary to clearly define the correlation of expertise of each individual and the concept hierarchy of abstract research domains.

Information Retrieval Systems and Search Engines have their own strategies for identifying relevancy of objects i.e. documents and people with respect to the given query. The basic statistical principles of frequency

and co-occurrence will be used to measure the similarity of web documents. Text documents are represented by Vector Space Model [16] as an array of valid tokens or index terms. The model is best suited for information filtering, retrieval, indexing and relevancy ranking. The dimension of the vector is the number of distinct words on the vocabulary.

$$\sum_{l=1}^{M} sim(p_l, CI) \qquad (1)$$

Where d_j represents the document vector and $w_{i,j}$ represents the weight of each distinct term in the document containing t terms. The weight can simply be replaced by the word's frequency or the weightage based on the importance to the current context like inverse document frequency.

The idea of Latent Semantic Analysis is that text has a higher order structure which is obscured by word usage due to polysemy[12]. In order to identify the relevancy of publications with the concept hierarchy of domains, our proposed method uses Term Frequency and Inverse Document Frequency [17] and [19]. These factors reflect the importance of given word with respect to a document in the entire corpus. Term frequency $tf(t,d)$ is the number of times the term t occurs in document d. To prevent bias towards longer documents, the value is normalized and the augmented frequency is used to measure the term frequency. Inverse Document Frequency $idf(t,D)$ is a measure whether the term is common or rare across the document corpus D. This is calculated by dividing the total number of documents by the number of documents containing the term. The product of TF and IDF will tend to filter out common terms in the document dataset.

Computing Document Similarity is another important phase in the proposed architecture as publications and their authors can be clustered based on their specific research domains. The publications are computed for similarity with respect to the concept hierarchy. Expertise Ranking is derived by computing the similarity of an expert's publications with concept hierarchy of the research domains. The similarity of document vectors is measured by using Cosine Similarity [8] and [18].

The preprocessing and graphical analysis of results of this approach is performed by using the environment of R[12][13], an integrated suite of software facilities for data manipulation, calculation and graphical display. Text Mining and Social Network Analysis modules provided by R gave us a flexibility to perform our experimentation and visualize results in an effective manner. Building of Term-Document Matrix from a given corpus of text documents and creating a word cloud from the frequent words is made so simple with the command line interface provided by R.

Rest of the paper is organized as follows: Collaboration Analysis in Social Networks is discussed in Section III, architecture of proposed Expert Finding System is described in Section IV. Latent Effort Ranking is discussed in Section V, the relevant experimentation along with results is presented in Section VI and the final conclusions are given in Section VII.

II. RELATED WORK

Language modeling by [1] uses various textual data sources to build a framework for expert finding given a particular topic. Propagation based expert scoring mechanism proposed by [2] has shown a method to analyze dense social network and use of personal profile document to initialize the process of ranking. Expert scoring started by probability based information retrieval, later uses personal relation of co-authorship to propagate the scores- across the academic social network.

Team formation problem addressed by [3] has started with the belief that effective communication among team members is an important factor for the successful completion of any task. Pool of individuals that are members of a social network and their skill set is considered for building an expertise team. The communication cost incurred towards graph analysis is also considered.

Wikipedia based Expert Finding System built by [4] extracts expert-expertise information from their publications, ranks them and provides multi-disciplinary search using link pattern analysis from Wikipedia. Author-topic model devised by [5] have shown the methods of identifying a reviewer for peer review of research publications. A methodology to create an author Persona per topic which enhances the search process is also shown. Multi-dimensional social recommender systems developed by [6] have projected different layers of social networks in a photo sharing system based on objects they access and their social connections.

Collaborative recommender systems developed by [7] have shown the use of similarity measures like Cosine Similarity and Pearson Correlation Coefficient to analyze product reviews by people on internet. Word similarity computed by [8] has dealt with vector space model of words based on semantic similarity used for document clustering. They have shown the importance of choosing an appropriate co-occurrence measure to compute the similarity of words. Social search systems developed by [9][10] have projected the importance of information seeking from peer group during search process. Answering systems and feedback systems are gaining focus due to the large attention from different categories of people. The importance of understanding the relation between data, information and representation in data storage and retrieval systems is analyzed in [26].

III. COLLABORATION ANALYSIS IN SOCIAL NETWORKS

Academic social networks, an offshoot of social network create a platform for researchers to publish their work and share knowledge with their peer group. Any technical reference can be had through them by referring to the current research trends, existing publications, people collaboration etc. They also rank researchers based on their effort, but is only quantitative. Different measures like h-index etc. are computed based on publications and citations [20]. However, the computation purely depends on count of publications and

coauthors. There is an immense need of analysis of research trends and other technical details, which can only be acquired from the qualitative analysis of the contents of publications.

Current research of expert finding is trying to build the gap between the need of the user and the availability of huge volumes of data. When a user is supplied with the most suitable answer within a short of span or a single click of a mouse, the satisfaction levels will be much better. When a novice wants to get some information, more belief would be on the statements given by an expert who is proven so in the domain. A common user can be transformed as an expert by spending effective time in acquiring knowledge and sharing with peer group. In applications like social search and social feedback systems, always a suggestion from an expert is given more weight. Collaborative effort analysis on the academic social network results in a team of experts who can be productive for solving issues dealing with multiple domains. Whether it is identifying a reviewer for a research publication, contacting a domain expert from industry or University is a task that will be made easy by this approach of collaborative expert matching system. This system tries to extract the latent expertise of a member of network that is hidden in his research publications. Standard Information Retrieval techniques make this approach superior to existing techniques of expert finding. It is predictable that the analysis of existing association between experts of different domains will result in a faster and stronger collaborative effort.

IV. ARCHITECTURE OF EXPERT FINDING SYSTEM (EFS)

A. Keyword Extraction from Publications

A research publication contain title, abstract, introduction, architecture, experimentation, results and concluding remarks followed by citations. Out of all these, the prominent zone that emphasizes the concept of the publication is the "Abstract". Existing research methods like [4] perform the concept matching using title and Wikipedia, but the abstract contain more information about the concept. Also, for almost all publications, the abstract part will be made accessible to a common user and even a web crawler. Hence, this approach tries to perform information extraction and visualization of expertise in a different perspective.

Fig.2 shows a sample abstract along with the tokens extracted from the abstract. Only the frequent terms are shown here after eliminating stop words like "*an*", "*the*", "*is*" etc.

B. Referencing Concept Hierarchy

Information retrieval systems have an important phase of indexing which makes the entire system effective [21]. Indexing is the oldest technique that helps in identifying contents of an item to assist retrieval. Due to the evolution of digital libraries, storage capability and retrieval efficiency are increased to cater to the needs of current digital environment. Different automatic indexing algorithms exist that create index terms for every domain

and provide access points for users through search engines. Relevance ranking strategies that involve feedback mechanism enhance the quality of retrieval which is measured using Precision and Recall.

This paper studies the problem of latent community topic analysis in text-associated graphs. With the development of social media, a lot of user-generated content is available with user networks. Along with rich information in networks, user graphs can be extended with text information associated with nodes. Topic modeling is a classic problem in text mining and it is interesting to discover the latent topics in text-associated graphs. Different from traditional topic modeling methods considering links, we incorporate community discovery into topic analysis in text-associated graphs to guarantee the topical coherence in the communities so that users in the same community are closely linked to each other and share common latent topics. We handle topic modeling and community discovery in the same framework. In our model we separate the concepts of community and topic, so one community can correspond to multiple topics and multiple communities can share the same topic. We compare different methods and perform extensive experiments on two real datasets. The results confirm our hypothesis that topics could help understand community structure, while community structure could help model topics.

[1]	"analys"	"can"	"communit"	"differ"	"discover"
[6]	"graph"	"help"	"inform"	"latent"	"method"
[11]	"model"	"multipl"	"problem"	"share"	"text"
[16]	"text-assoc"	"topic"	"user"		

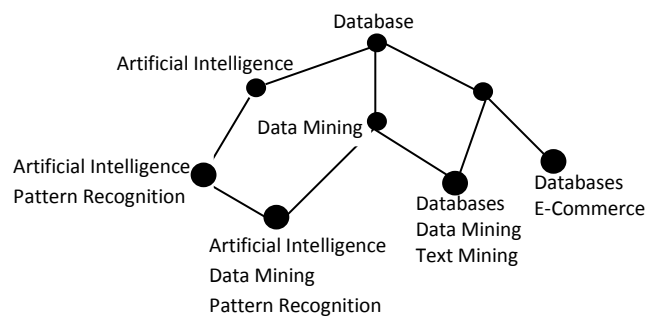Fig. 2. Frequent Terms extracted from abstract of given publication (After Stemming and Stopword Removal)

Fig. 3. Reference Concept Hierarchy of Research Domains

Fig. 3 depicts a sample hierarchy of concepts that are branches and siblings of computer science domain. Some terms are related to a specific domain and some terms cannot be decided to which they belong to. Hence, not only keyword extraction, but also contiguous phrase extraction has to be performed to identify domain is some scenarios. Each domain will be having a set of index terms and the relevancy of the terms w.r.t. the domain is measured quantitatively. Words and then documents are clustered based on the semantic similarity to refine the search process. Terms extracted in previous phase of 4A are matched against this concept hierarchy and its respective index terms to identify the expertise and rank the expert based his expertise in every domain.

As explained in the above phases, the complete architecture is shown in Fig. 4 which demonstrates the flow of information extraction and ranking followed by expertise derivation.
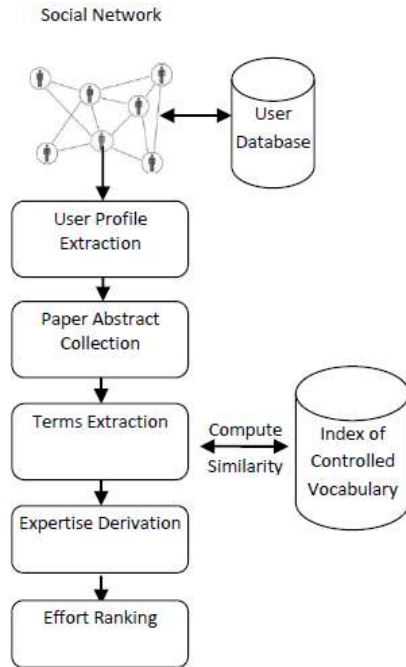
Fig. 4. Architecture of Expertise Finding System

V. LATENT EFFORT RANKING

A. Algorithm 1

To Compute Expertise Ranking

Input: Author Profile and Publications, P Output: Author Expertise Index/Domain

Step 1: From a given set of Researchers, R, for each Researcher, $R_i \subseteq R$, extract profile document

Step 2: Extract set of Publications P for R_i

Step 3: For each Publication, P_j, create TF-IDF vector, using Algorithm2.

Step 4: Compute Cosine Similarity between TF-IDF weight Vector and the Concept Hierarchy of index terms i.e. controlled vocabulary, using the equation

$$sim\left(p_j, CI\right) = \frac{p_j . CI}{\|p_j\| \|CI\|} \Rightarrow \frac{\sum_{i=1}^{N} W_{i,j} W_{i,CI}}{\sqrt{\sum_{i=1}^{N} W_{i,j}^2} \sqrt{\sum_{i=1}^{N} W_{CI}^2}} \quad (2)$$

Step 5: Compute cumulative similarity for R_i in each Domain D_l, build a feature vector and rank them in descending order per each research domain.

$$\sum_{l=1}^{M} sim(p_l, CI) \quad (1)$$

Step 6: Compute Correlation measures between documents in the corpus to identify their semantic similarity and to refine the controlled vocabulary.

B. Algorithm 2

To Compute TF-IDF weight vector for a given document

Input: Title and Abstract of Publication

Output: TF-IDF weight vector

Step 1: For each publications, $P_j \subseteq P$, perform tokenization i.e. extract tokens from Title and Abstract.

Step 2: Remove Stop Words, the set of insignificant words

Step 3: Perform stemming i.e. conflate tokens into their root form

Step 4: Compute Term Frequency (TF(t,d)) where t is the term/token, d is the document and w is the set of words/tokens in the document.

$$TF\left(t,d\right) = 0.5 + \frac{0.5 * f(t,d)}{\max\{f(w,d) : w \in D\}} \quad (3)$$

Step 5: Compute Inverse Document Frequency (IDF), where numerator, |D| is the total # documents in the document set and the denominator is the # documents containing term t.

$$IDF = \log \frac{|D|}{|\{d' \in D \mid t \in d'\}|} \quad (4)$$

Step 6: Compute TF-IDF Weight i.e. product of TF-IDF

VI. EXPERIMENTATION AND RESULT ANALYSIS

A. KeyWords /Controlled Vocabulary Identification

Several researchers are selected per each domain of an academic social network, and their publications are collected from Google Scholar, CiteSeer and other open access journals for this analysis. 20 researchers from 8 research domains of "Data Mining", "Information Retrieval", "Social Network Analysis", "Text Mining" etc. are selected for analysis and for each researcher 30 publications in the last two years are analyzed.

Some of the sample keywords that can be used as glossary terms for refereeing concept hierarchy in the domains of "Data Mining" and "Artificial Intelligence" [23],[24] are given in Table 1a and 1b. Such a set of finite index terms is referred to as the controlled vocabulary [21]. This index simplifies the search process at the cost of time. The format of index even supports the process of ranking to decide relevance of the word with reference to the specific domain.

Table 1a. Concept Keywords in "Data Mining"

Accuracy	Association Rule	Back Propagation
Binning	Brute Forcing	Cardinality
CART	Clustering	Cross Validation
Decision Trees	Entropy	Error Rate
Expert System	Data Analysis	Field
Fuzzy Logic	Genetic algorithm	Gini Index
Hypothesis	Intelligent Agent	KDD
Knowledge Discovery	Lift	Machine Learning
Model	Nearest Neighbor	Neural Network
OLAP	Outlier Analysis	Overfitting
Prediction	PCA	Regression
Reinforcement learning	Sampling	Segmentation
Supervised learning	Support	Visualization

Table 1b. Concept Keywords in "Artificial Intelligence"

ALICE	artificial intelligence	autonomic computing
backward chaining	Bot	CAPTCHA
chat bot	Cobot	computational linguistics
expert systems	forward chaining	4GL
character recognition	KQML	natural language
neural network	NLP	pervasive computing
robotics	rule-based system	Turing

B. Significance of "R" in the pre-process

Word cloud represents most frequent words occurring in a given text and those frequent terms can be used to refer to the concept hierachy for expertise matching.

Fig. 5b. Wordcloud with stopwords

Fig. 5a. Wordcloud without stopwords

Sample wordcloud[12] created from 15 abstracts of research publications collected in the domain of "Expertise Search" are shown in Fig.5. Stop word Elimination, punctuation eliminations, graph plotting are some functionalities in the package of R.

Sample syntax to perform the preprocessing is given below.

```
library (tm)
library (wordcloud)
corpus ← Corpus(DataframeSource(data.frame(texts)))
tdm ←
        TermDocumentMatrix(corpus,control=list(word
        Lengths=c(2,Inf)) )
findFreqTerms(tdm,lowfreq=4)
        newcorpus ← tm_map(corpus,
                removePunctuation)
        newcorpus ← tm_map(corpus, removeWords,
        stopwords("english"))
        wordcloud(words=names(wrdfrq),freq=wrdfrq,
        min.freq=4,random.order=F,
        colors=grayLevels)
```

Properties of Term-document Matrix and its equivalent bar plot is shown in Fig.6
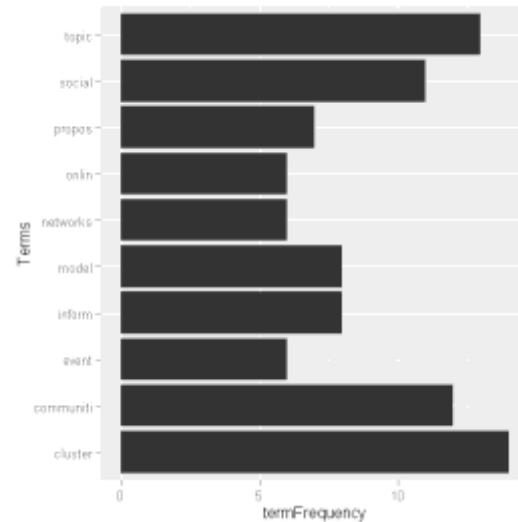

Fig. 6 Bar Plot of Term Frequency

➤ inspect(tdm[35:40,])
 A term-document matrix (1275 terms, 14 documents)
 Non-/sparse entries: 338/762
 Sparsity: 69%
 Maximal term length: 16

C. Computation of Latent Semantic Similarity:

Semantic analysis is performed on the text documents to identify their domains and computation of similarity with the concept index further. The correlation measures computed between some documents of the corpora shows that the correlation between some set of documents is negative, implying that those are not relevant to a single domain.

Table 2 Correlation measures b/w documents in corpus

Kendall Corr. Coeff	Text1	Text2	Text3	Text4
Text1	1	0.66	-0.44	-0.23
Text2	0.66	1	-0.11	0.11
Text3	-0.44	-0.11	1	0.79
Text4	-0.23	0.11	0.79	1

Kendall Corr. Coeff	Text1	Text2	Text3	Text4
Text1	1	0.66	-0.44	-0.23
Text2	0.66	1	-0.11	0.11
Text3	-0.44	-0.11	1	0.79
Text4	-0.23	0.11	0.79	1

Pearson Corr. Coeff	Text1	Text2	Text3	Text4
Text1	1	0.78	0.11	0.42
Text2	0.78	1	0.71	0.90
Text3	0.11	0.71	1	0.95
Text4	0.42	0.90	0.95	1

Spearman Corr.Coeff.	Text1	Text2	Text3	Text4
Text1	1	0.74	-0.50	-0.21
Text2	0.74	1	-0.01	0.29
Text3	-0.50	-0.01	1	0.88
Text4	-0.21	0.29	0.88	1

Relative comparison of the correlation measures is shown in Table 2 and Fig.7. Different correlation measures are considered due their significance in normalization of the resultant values. These are measures of how well two datasets fit on a straight line, and its value lies between -1 to 1. Pearson Correlation Coefficient is influenced by outliers and non-linearity, which can be overcome by Spearman's correlation coefficient. It is calculated to the ranks of the data rather than to the actual data values. Kendall's Tau is a measure of correlation between two ordinal-level variables of m*n tables. From the statistics it is evident that Text1 & Text 2, Text3 & Text4 are similar and Text1 & Text4 are dissimilar to one another.
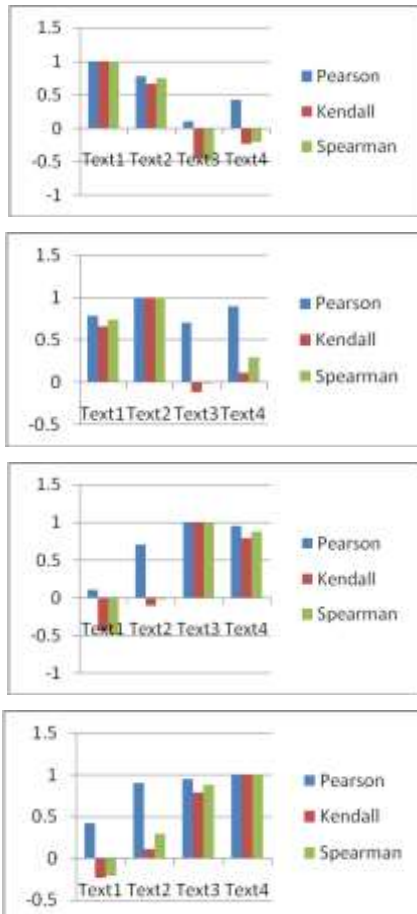

Fig. 7 Correlation Measures of Text Documents w.r.t. Text 1 to Text4

Table 3. TF-IDF values in corpus

Document/ TF IDF	Cluster	commun	heterogeneous	Purpose
Text1	0.000	0.007	0.007	0.000
Text2	0.000	0.013	0.000	0.000
Text3	0.012	0.000	0.006	0.012
Text4	0.012	0.000	0.000	0.006

The computed values of Inverse Document Frequency for a set of sample text documents, is shown in Table 3, which depicts the overall impact of the terms in the document corpus. The importance of IDF[18] lies in the fact that the influence of keyword w.r.t. the entire document corpus related to a specific research domain that comprises the controlled vocabulary. It can be observed that the non-occurrence of a word in a document is represented by 0.

D. Ranking of Author Association in a specific domain

Based on the process described in the above sections, an author's cumulative relevance factor w.r.t. a specific research domain is computed using the cosine similarity with reference to the bag of words from the controlled vocabulary.

$$\sum_{l=1}^{M} sim(p_l, CI)$$

Where M is the #publications and *sim()* computes the Cosine Similarity of each publication w.r.t. the controlled vocabulary of research domain. A feature vector is built for each researcher with the ranking in each research. When our interface queries for a researcher in a set of domains, the aggregate similarity of each domain for the particular researcher is computed and the resultant list is displayed. Each vector represents one publication (P) of author and column represent research domains(R) and $sim_{i,j}$ gives the Cosine Similarity of the publication w.r.t. the specific domain, where i varies from 1 to # domains and j varies from 1 to #publications. Final aggregation of all publications is the summation of the similarity values per each domain.

VII CONCLUSIONS

Expertise Matching is an important task that satisfies the need of reviewer systems, collaborative recommender systems or even question answer systems. This approach of publication analysis and retrieval of hidden expertise is proven to be the better approach in identifying the exact expertise of an individual in different research domains. Further analysis is required to simplify the information retrieval process that suits to any social networks, not only academic social networks. Text mining of information retrieval strategy chosen for this analysis has made this approach faster to process large corpus of text documents.

REFERENCES

[1] Balog, Krisztian, Leif Azzopardi, and Maarten de Rijke. "A language modeling framework for expert finding." *Information Processing & Management* 45.1 (2009): 1-19.

[2] Zhang, Jing, Jie Tang, and Juanzi Li. "Expert finding in a social network." *Advances in Databases: Concepts, Systems and Applications*. Springer Berlin Heidelberg, 2007. 1066-1069.

[3] Lappas, Theodoros, Kun Liu, and Evimaria Terzi. "Finding a team of experts in social networks." *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009.

[4] Yang, Kai-Hsiang, et al. "EFS: Expert finding system based on Wikipedia link pattern analysis." *Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on*. IEEE, 08.

[5] Mimno, David, and Andrew McCallum. "Expertise modeling for matching papers with reviewers." *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2007.

[6] Kazienko, Przemyslaw, Katarzyna Musial, and Tomasz Kajdanowicz. "Multidimensional social network in the social recommender system." *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on* 41.4 (2011): 746-759.

[7] Munoz-Organero, Mario, et al. "A collaborative recommender system based on space-time similarities." *Pervasive Computing, IEEE* 9.3 (2010): 81-87.

[8] Bordag, Stefan. "A comparison of co-occurrence and similarity measures as simulations of context." *Computational Linguistics and Intelligent Text Processing*. Springer Berlin Heidelberg, 2008. 52-63.

[9] Chi, Ed H. "Information seeking can be social." *Computer* 42.3 (2009): 42-46.

[10] Trier, Matthias, and Annette Bobrik. "Social search: exploring and searching social architectures in digital networks." *Internet Computing, IEEE* 13.2 (09): 51-59.

[11] Maybury, Mark T. "Expert finding systems." *MITRE Cente_for_Integrated_Intelligence_Systems, MIT,* (2006).

[12] http://cran.r-project.org/

[13] http://www.rdatamining.com/

[14] www.aminer.org

[15] www.academic.research.microsoft.com

[16] Santos, Igor, et al. "Enhanced Topic-based Vector Space Model for semantics-aware spam filtering." *Expert Systems with applications* 39.1(12): 437-444.

[17] Chow, Tommy WS, Haijun Zhang, and M. K. M. Rahman. "A new document representation using term frequency and vectorized graph connectionists with application to document retrieval." *Expert Systems with Applications* 36.10 (2009): 12023-12035.

[18] http://en.wikipedia.org/wiki/Tf/idf

[19] http://nlp.stanford.edu/IR-book/html/htmledition/inverse-document-frequy-1.html

[20] Jie Tang, Jing Zhang , Limin Yao, Juanzi Li , Li Zhang and Zhong Su, *"ArnetMiner: Extraction and Mining of Academic Social Networks"*, *Knowledge Discovery & Datamining'08,* August 24–27, 2008, Las Vegas, Nevada, USA.

[21] Gerald Kowalski, "Information Retrieval Systems", Kluwer Academic Press

[22] Zhao, Ying, and Falk Scholer, "Predicting query performance for user-based search tasks." *Australian Database Conference,* Vol.7. 2007.

[23] http://www.thearling.com/glossary.htm

[24] http://www.webopedia.com/ ComputerScience/AI

[25] Becker, J., & Kuropka, D. (2003). "Topic-based vector space model", in Proc. of the 6th Intnl Conf. on Business Information Systems (p 7–12).

[26] Lin Liu,Junkang Feng,"A Semantic Approach to the Notion of Representation and Its Application to Information Systems", IJITCS, vol.3, no.5, pp.39-50, 2011.

Authors' Profiles

Sobha K. Rani is currently working as an Associate Professor in the Department of Information Technology, MVGR College of Engineering, Vizianagaram, Andhra Pradesh. She has a teaching experience of 9 years and has practical exposure to work on different platforms since 1996. She is pursuing Ph.D. in the areas of data mining, information retrieval and social networks.

KVSVN Raju is one of the first generation educationists of CSE in India and is a Professor & Director (R&D) at Anil Neerukonda Institute of Technology and Sciences(ANITS), Bheemunipatnam, Andhra Pradesh. Earlier he worked for 32 years at Dept of CSSE, College of Engineering, Andhra University, as Assistant Professor to Professor. His research areas include Data Engineering, Security Engineering, Software Engineering and Web Engineering and he supervised so many Ph.Ds in these areas. His professional body memberships include IEEE, IETE, CSI, ISTE and Inst. of Engineers.

Valli V. Kumari is a Professor at Dept. of CSSE, College of Engineering, Andhra University, Visakhapatnam. She has a total of 25 years teaching experience. She won "Gold Medal" for *Best Research* in 2008. She is a fellow of IETE and life member of CSI, ISTE, CRSI etc. Her research areas include Network Security, Privacy Preservation, Image Processing and Web Mining. She is certified by Microsoft in VC++ and IBM in DB2 and is well versed with so many other latest technologies.