IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-6, June 2018 E-ISSN: 2347-2693

Social Networking using Semantic Web with Social Tagging System

R. Indra[1*], M. Thangaraj[2]

[1] Department of Computer science,Government Arts College for Women, Tamil Nadu, India
[2] Department of Computer Science, Madurai Kamaraj University, Madurai, Tamil Nadu , India

Corresponding Author: indrakannan2000@yahoo.co.in, Tel.: +91 9442722566

Available online at: www.ijcseonline.org

Accepted: 16/Jun/2018, Published: 30/Jun/2018

Abstract— Social networks are simple graphs where nodes represent either the people or the groups and links represent their relationships. Social networks explicitly exhibit relationships among individuals and groups. It is used for trust calculation, information sharing and recommendation, ontology construction and relation and relevance detection. Semantic Web is not a separate web but it is an extension of the current web in which information has well-defined meaning enabling computers and people to work better with cooperation. Semantic web (SW) has proven to be a useful data integration tool, facilitating the meaningful exchange of heterogeneous data. Ontology is a tool for data integration. Social Tagging Systems (STSs) allow collaborative users to share and annotate many types of resources (webpages, songs, etc.) with descriptive and semantically meaningful information called tags. Our proposed system constructs social network among individuals based on users' interest predicted from their tag usage in the Social Tagging System using semantic web.

Keywords—Social Tagging System (STS), Social Network (SN), Semantic Web (SW), Randomized Singular Value Decomposition (RSVD), Recommendation, User interest.

I. INTRODUCTION

Social Networking services are under the control of the database owner who has an interest in keeping the information bound to site. Virtual or online communities are groups of people connected through the Internet and other information technologies. The communication technologies used to support virtual communities are evolved with the Internet and include electronic mailing lists, bulletin boards, Usenet, IRC, Wikis, FOAF and blogs. Virtual communities built on social network structures began appearing in 2002 and have become most popular among web-based applications. Such sites allow individuals to publish personal information in a semi-structured form and to define links to other members with whom they have relationships of various kinds. Current examples include Friendster, LinkedIn, Tribe.net and Orkut.

STSs are web-based systems that allow users to upload their resources, and to label them with arbitrary words, so-called tags. Various STSs are available, each supporting different kind of resources. For example, Flickr for sharing photos, Last.fm for sharing music, BibSonomy for sharing bookmarks and scientific publication references etc. STS helps users to find the resources they are interested in, to discover other users by finding people who tagged the same resource or used the same tag and support the discovery of new information using the set of all tags made by all users. As the tagging activities increase, a lightweight collaborative classification system, known as Folksonomy emerges. Folksonomy is an aggregation of tags for multiple resources, shared by multiple users.

The SW is specifically a web of machine-readable information whose meaning is well-defined by standards. The basic layer of data representation is standardized as the Resource Description Framework (RDF). RDF is based on existing standards of eXtensible Markup Language (XML), Uniform Resource Indicators (URIs) and Unicode. Built on top of RDF, the Ontology layer exists in the form of Web Ontology Language (OWL). The concept of SW is quite different from the artificial Intelligence. In SW, machines will have semantics that permit them to solve well-defined problems through the sequential processing of operations. In SW, information will have well-defined meaning as a result of the use of Semantic mark-up languages. These languages will be included in the existing Web pages in the layered cake architecture. Ultimate benefit would be a collateral outcome of creating genuinely global knowledge representation.

The rest of the paper is organized as follows. Related work is discussed in Section II. The proposed architecture and its components are explained in section III. Section IV discusses the Implementation details. The merits and demerits of the proposed system and its future scope are discussed in section V.

II. RELATED WORK

In this section we present some of the research literature related to finding user interest. Discovering social interest shared by groups of users is very important because it helps to connect people with common interest and encourages people to contribute and share more contents. The main difficulty to solve this problem is in detecting and representing the interest of the users. As the existing approaches are all based on the online connections of users they are unable to identify the common interest of users who have no online connections.

Social interest discovery approach is based on user–generated tags [1]. Patterns of frequent co-occurrences of user tags, is used to characterize and capture topics of users' interest. Internet Social Interest Discovery System (ISID) discover the common users' interest and cluster users and their saved URLs by different interest topics. ISID uses association rules to identify the frequent tag patterns for the posts by considering user and resource as 'transaction id' and 'tags' as item set. Here the disadvantage is the ambiguity in tags due to syntactic variation of tags.

The collaborative nature of social network systems and their flexibility in tagging produces multiple variations of same tag frequently. Syntactic variations of tags are grouped together using fuzzy similarity and cosine measures. That improves the ISID algorithm [2]. However, this uses only the tags used by the users to find the user interest without predicting the tags that can be used by the users.

The 3-order tensor is used to model the three types of entities (user, item and tag) that exist in social sites [3]. In 3-order tensors, High Order singular Value Decomposition (HOSVD) is applied to reveal the latent semantic associations between users, items and tags. Smoothing technique called Kernel-SVD is also applied to address the sparseness of data. Here, the new users are recommended for a tag, according to their total weight, which results by aggregating all items, which are labelled with the same tag by the target user. Thus the interest group is formed based on the tags used by the users as well as future tags predicted by the system. However, this system lacks SW.

An interest group based on tagging data is extracted from tagging behaviours using Formal Concept Analysis (FCA) [4]. Formal Concept Analysis (FCA) is a mathematical theory used for conceptual data analysis and unsupervised machine learning. Based on FCA, given a set of users 'U' and considering the tags that they have in common, the interest group of 'U' is the set of users who are using these tags. The intent of a set of users 'U' is the set of tags which are used by every user in 'U'. The extent of a set of tags 'T' is the set of users using every tag in 'T'. Thus an interest group would be a set of users that use a significantly similar collection of tags to identify their resources. Here the SW is used to represent tagging activities after finding user interest group using FCA. Yet, latent associations that exist among user, tag and items are not revealed.

Existing methods extract social network from homepages [5], web pages, emails, publication archives and FOAF (Friend-Of-A-Friend) profiles [6], [7], Traditional web [8], Traditional web, email and sensor [9], E-mail archives, FOAF documents and DBLP [10], E-mail and the traditional web [11] but not from STSs.

Markov Model predicts user interest and behaviour using web usage mining [12] and another method available for finding user interest is sentiment analysis [13]. But both of these methods are not applicable for STSs.

Hence, to alleviate all the above mentioned problems a SNSTS framework (Social Networking in Social Tagging Systems) that constructs a social network based on user interest is proposed. This is accomplished by building SW after applying Randomized Singular Value Decomposition (RSVD) to reveal latent semantic associations between users, items and tags. Before applying RSVD stemming using porter stemmer algorithm is used to remove syntactic variation of tags.

III. PROPOSED SNSTS ARCHITECTURE

The architecture of the proposed framework is shown in Figure 1. This framework extracts STS data in the form of triples from the web, stores in the database and performs stemming on the tags to find root words. The triples are represented using Tensor and matrix unfolding of the Tensor is performed. The system reveals latent association in the tensor by applying Latent Semantic Indexing (LSI) using RSVD on the resultant matrices of matrix unfolding. Friends are identified based on the tag usage for each item and Jaccard coefficient. Then the triples along with the weight are stored on the database. From the database, Ontology is constructed representing tagging activities and social network. Social Network shows the relationships between friends. The various components involved in the framework are explained below.

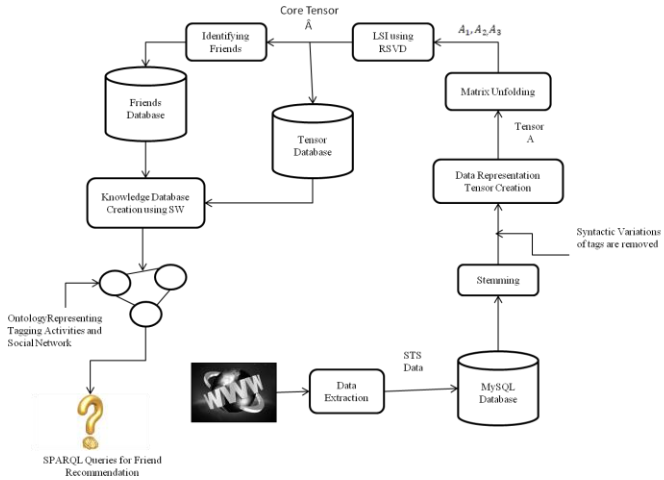


Figure 1. *SNSTS Architecture*

III.I. Data Extraction

In STS, a user 'u' tags an item 'i' with a tag 't' resulting in a collection of triples {u, i, t}. These triples are retrieved from the web and stored in the database.

III.II. Stemming

From the tags all characters which are neither numbers nor letters are removed. Then the tags are stemmed using porter stemmer algorithm to convert them into base form word in order to remove syntactic variations.

III.III. Data Representation (Tensor Creation)

A tensor is a multidimensional matrix. An N-order tensor $\mathcal{A}$ is denoted as $\mathcal{A} \in R^{I_1 \ldots I_N}$ with elements $a_{i_1 \ldots i_N}$. In this paper we only use 3-order tensors. All distinct users, tags and items are found from the database and stored in the 3-order tensor with user in the first dimension, tag in the second dimension and item in the third dimension. The size of the three order Tensor is no. of distinct users, tags and items respectively. Then each record containing user, tag and item of the user's posts are read from the database until all the records are exhausted. For each record the occurrence count is stored in the corresponding place of the tensor. Thus the tensor $\mathcal{A}$ is constructed.

III.IV. Matrix unfolding

Given a 3-order Tensor $\mathcal{A}$, three matrix unfolding operations [3] [14] are defined as in (1):

$$A_1 \in R^{I_1 x I_2 I_3} \quad A_2 \in R^{I_2 x I_1 I_3} \quad A_3 \in R^{I_3 x I_1 I_2} \tag{1}$$

where A_1, A_2 and A_3 are called the 1-mode, 2-mode and 3-mode matrix unfolding of $\mathcal{A}$, respectively. Each A_n, $1 \le n \le 3$, is called the n-mode matrix unfolding of $\mathcal{A}$ and is computed by arranging the corresponding fibers of $\mathcal{A}$ as columns of A_n [14]. Matrix unfolding of Tensor $\mathcal{A}$ (constructed in section 3.3) is performed for all the three modes, so that Tensor is matricized as A_1, A_2 and A_3 .

III.V. LSI using RSVD

Algorithm RSVD is applied on the resultant matrices to perform LSI [14] [15]. It reveals latent semantic associations that exist among user, item and tag. Resultant core Tensor $\hat{\mathcal{A}}$ is the approximation of the original Tensor $\mathcal{A}$. Output of RSVD is interesting because it reveals the new associations among user, item and tag. These new associations are the predictions of future tagging instances of users.

III.VI. Identifying Friends

Each user is compared with every other user item wise to find out the number of common tags. Then these tags are accumulated for all items. Similarity between two users P & Q is calculated using Jaccard coefficient (2), which is the ratio of number of common tags between P and Q with the total number of tags of P and Q.

$$\text{Similarity (P, Q)} = \frac{|P \cap Q|}{|P \cup Q|} \tag{2}$$

where $|P \cap Q|$ is the number of common tags used by P & Q and $|P \cup Q|$ is the total number of tags used by both P & Q. In this way for each user, similar users (friends) are found out and stored in a friends-database with one record for each user and his friend. If a user has more than one friend separate records are created for each friend. Similar users are based on their tag usage for each item. As tags reveal users' interests, this method can be called as finding similar users based on their interests. Existing methods use only two dimensions (user, tag) for finding similar users and do not predict future tag usages[16][17]. But in this method third dimension item is included to find out similar users and also future tagging activities of users are predicted using LSI. As this system uses three dimensions, user's different interest for an item is taken into consideration.

Then the whole core Tensor $\hat{\mathcal{A}}$ is stored in tensor-database with one record for each (user, tag, resource) triple along with the weight. If a user uses more than one tag for a resource then a separate record will be created for each tag.

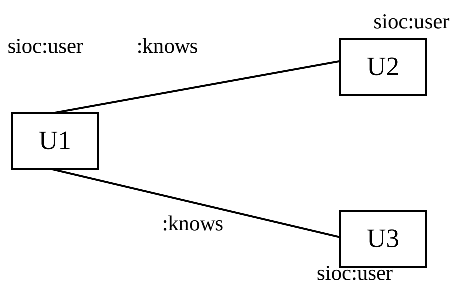
III.VII. Knowledge Database Creation using SW

III.VII.I.Ontology Representing Tagging Activities

The databases tensor and friends, created above is converted into knowledge base using ontology. The classes and properties from various ontologies like SCOT (http://rdfs.org/scot/ns#) ontology, SIOC (http://rdfs.org/sioc) ontology and tags (http://www.holygoat.co.uk/owl/redwood/0.1/tags/) ontology are used in our ontology [15]. Each record (user, tag, item and weight) of a tensor-database representing a tagging activity is represented using tags:Tagging class. The entities of a tagging activity such as user is represented using class sioc:user, tag using scot:tag, item using sioc:item and weight is a float connected using dataproperty with tags:tagging. The class tags:Tagging has object properties tags:taggedBy, tags:associatedTag and tags:taggedResource to link to sioc:user, scot:tag and sioc:item respectively. All tagging activities of a user are collectively represented using scot:TagCloud class. This class has two object properties viz. scot:taggingActivity and scot:contains to establish a relationship with tags:tagging and scot:tag respectively. The scot:tag is connected to sioc:user using scot:usedBy and to sioc:item using scot:tagOf. The property scot:has_tag connects sioc:item with scot:tag. If the names of the multiple tags in tagging events coincide then those tags are aggregated to one unique scot:Tag class. The classes sioc:user and sioc:item have other properties to describe the user and item respectively.

III.VII.II. Ontology Representing Social Network

In the ontology constructed above each entity sioc:user is connected with his friends using the object property :knows. The object property :knows has both domain and range as sioc:user. For instance if user U1 has two friends U2 and U3, it is established as follows.



If user U2 has friend U3 then U2 and U3 are connected using :knows object property. In this way social network is constructed. Friends database is used to establish this network. Existing methods [8], [9] use FOAF ontology with FOAF profile to represent social network. But we are creating our own object property :knows to constitute social network. The object property :knows is a bidirectional property. That is if U1 knows U2 then U2 knows U1.

III.VIII. Friend Recommendation

As the social network is represented using semantic web, Simple Protocol And RDF Query Language (SPARQL) is used to query and retrieve the required information.

IV. PERFORMANCE ANALYSIS

This system is implemented using Java in NET BEANS 8.0. MySQL is used in the backend. Weka is used for filtering. The data set extracted from STS BibSonomy [18] is used to assess the proposed system. It contains all public bookmarks and publication posts of BibSonomy until (but not including) 01-01-2009. To get more dense data p-core at level 10 is used. Since it is too large, a sample of 1000 users is selected at random from this dataset. To assess how accurate or correct the system is, we use the metrics of precision, recall and F-measure. For a test user that receives a list of N identified friends (top-N list), precision, recall and F-measure are defined as follows.

Precision is the percentage of retrieved friends that are in fact relevant the query user (3). (i.e., Correct responses)

$$\text{Precision} = \frac{|\{Relevant\} \cap \{Retrieved\}|}{|\{Retrieved\}|}$$

$$\text{Precision} = \frac{\text{Number of relevant friends in the top-}N\text{ list}}{N} \quad (3)$$

Recall is the percentage of friends that are relevant to the query user and in fact retrieved (4).

$$\text{Recall} = \frac{|\{Relevant\} \cap \{Retrieved\}|}{|\{Relevant\}|}$$

$$\text{Recall} = \frac{\text{Number of relevant friends in the top-}N\text{ list}}{\text{Total no.of relevant friends}} \quad (4)$$

To trade off Recall for Precision or vice versa F-score is used (5). It is defined as the harmonic mean of Recall and Precision.

$$F = \frac{2 \times Precision \times Recall}{(Precision + Recall)} \quad (5)$$

The proposed system is compared with the recommendation system 'User Recommendation with Tensor Factorization in Social Networks' (URTFSN) [16]. Figures 2- 4 show the results of performance comparison with the said method in the metrics of Precision, Recall and F-measure respectively.

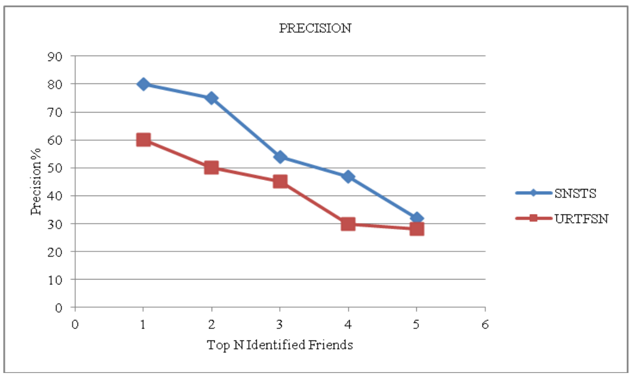


Figure 2. *Plot of Precision*

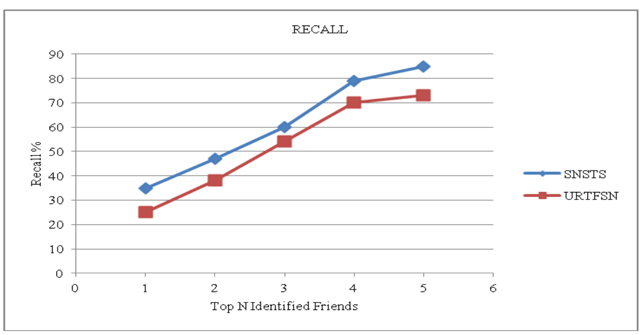


Figure 3. *Plot of Recall*

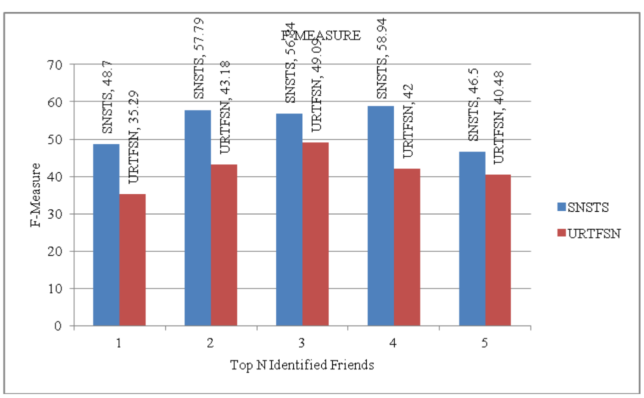


Figure 4. *Plot of F-Measure*

From Figure 2 Precision is found to decrease as N increases where as in Figure 3, Recall increases as N increases. SNSTS accomplishes 80 percent of Precision when we identify top-1 friend, whereas URTFSN gets a Precision of only 60 percent. Furthermore SNSTS is more powerful than URTFSN getting a maximum recall of 85 percent while the latter's 73 percent. From these figures it is understandable that STSTS attains 10 to 20 percent increase in both Precision and Recall. This confers that our proposed system constantly outperforms other approach in all metrics.

V. CONCLUSION AND FUTURE DIRECTIONS

The main purpose of this research is to present a framework that constructs social network among individual users based on their interests. User interest is calculated using Jaccard Coefficient based on the tags used by the users for each item in the STS. This framework outperforms over the existing methods by revealing latent associations among user, item and tag using RSVD. That is future tag usages of users are predicted and included in the interest calculation. Another advantage of this system is using SW for the representation of social network & tagging activities and three recommendations such as friend recommendation, item recommendation and tag recommendation that can be performed from the same. One drawback of this system is, the type of relationships between users is not included.

This research can be further extended by Social Network Analysis (SNA). SNA can be applied on a network of known criminals and their relationships. This analysis identifies various groups and subgroups, key individuals and links between groups. Centrality can be detected using graph properties including degree (the number of direct links), betweenness (geodesics passing through) and closeness (sum of geodesics). Each of these indices is evidence for different individual roles: a high degree suggests leadership and high betweenness indicates a "gatekeeper".

REFERENCES

[1] X. Li, L. Guo, and Y. Zhao, "*Tag-based social interest discovery*", Proceedings of the 17th international conference on World Wide Web," page 675--684. New York, NY, USA, ACM, 2008.

[2] Jose Javier Astrain, Alberto Cordoba, Francisco Echarte, Jesus Villadangos, "*An Algorithm for the Improvement of Tag-based Social Interest Discovery*," SEMAPRO, The Fourth International Conference on Advances in Semantic Processing, 2010.

[3] Panagiotis Symeonidis, Alexandros Nanopoulos and Yannis Manolopoulos, "*A Unified Framework for Providing Recommendation in Social Tagging systems Based on Terenary Semantic Analysis*," IEEE Transactions on Knowledge and Data Engineering, Vol. 22, No. 2, February 2010.

[4] Hak-Lae Kim, John G. Breslin, Stefan Decker, and Hong-Gee Kim, "*Mining and Representing User Interests: The Case of Tagging*

Practice," IEEE Transactions On Systems, Man and Cybernetics-Part A: Systems and Humans, Vol-41, No.4, July 2011.

[5] L. A. Adamic, E. Adar, "*Friends and neighbors on the Web*," Social Networks, Vol. 25, No. 3., pp. 211-230, July 2003.

[6] Peter Mika, "*Flink: Semantic web technology for the extraction and analysis of social networks*," Web Semantics: Science, Services and Agents on the World Wide Web, Elsevier, 2005.

[7] Yutaka Matsuo, Junichiro Mori, Masahiro Hamasaki, TakuichiNishimura, HideakiTakeda, KoitiHasida, MitsuruIshizuka, "*POLYPHONET: An advanced social network extraction system from the Web, Web Semantics: Science, Services and Agents on the World Wide Web*," Volume 5, Issue 4, Pages 262-278, December 2007.

[8] J Mori, Y Matsuo, K Hashida, "*Web Mining Approach for a User-centered Semantic Web*" … - Proc. Int'l Workshop on …, 2005.

[9] Junichiro Mori, Tatsuhiko Sugiyama, and Yutaka Matsuo, "*Real-world oriented information sharing using social networks*," In Proceedings of the international ACM SIGGROUP conference on Supporting group work (GROUP '05), ACM, New York, NY, USA, 81-84, 2005.

[10] Jin Y., Matsuo Y., Ishizuka M., "*Extracting Social Networks Among Various Entities on the Web*," In: Franconi E., Kifer M., May W. (eds) The Semantic Web: Research and Applications, ESWC 2007, Lecture Notes in Computer Science, vol. 4519. Springer, Berlin, Heidelberg.

[11] Culotta, Aron, Bekkerman, Ron and McCallum, Andrew, "*Extracting social networks and contact information from email and the Web,*" Computer Science Department Faculty Publication Series, 33, 2004.

[12] Neetu Anand, Tapas Kumar, "*Prediction of User Interest and Behaviour using Markov Model*", International Journal of Scientific Research in Computer Science and Engineering, Vol.5, Issue.3, pp.119-123, 2017.

[13] C. Nanda, M. Dua, "*A Survey on Sentiment Analysis*", International Journal of Scientific Research in Computer Science and Engineering, Vol.5, Issue.2, pp.67-70, April 2017.

[14] P Drineas, MW Mahoney, "*A randomized algorithm for a tensor-based generalization of the singular value decomposition,"* - Linear algebra and its applications, Elsevier, 2007.

[15] R. Indra, M. Thangaraj, "*An Integrated Recommender System using Semantic Web with Social Tagging System*", International Journal on Semantic Web and Information Systems (IJSWIS), Vol. 15, Issue 2, No. 3, 2019. [In Press]

[16] Zhenlei Yan, Jie Zhou, "*User Recommendation with Tensor Factorization in Social Networks,*" ICASSP, 2012.

[17] Shuhui Jiang, Xueming Qian, Jialie Shen, Yun Fu, Tao Mei, "*Author Topic Model based Collaborative Filtering for Personalized POI Recommendation,*" In Multimedia, IEEE Transactions on, IEEE, volume 17, 2015.

[18] Knowledge & Data Engineering Group, University of Kassel: Benchmark Folksonomy Data from BibSonomy, version of December 31st, 2006.

Authors Profile

Radhakrishnan Indra received her post-graduate degree in Computer Science from Bharadhidasan University, Trichy in 1994 and M.Phil. degree in Computer Science from the same university in 2000. She is currently pursuing part-time Ph.D. in M.S.University, Thirunelveli and working as the Associate Professor in Computer Science Department, Government Arts College for Women, Sivagangai, Tamilnadu, India. She has 24 years of teaching experience and 5 years of Research Experience. Her main research work focuses on Data Mining, Social Network and Semantic Web.

Muthuram Thangaraj received his post-graduate degree in Computer Science from Alagappa University, karaikudi, M.Tech degree in Computer Science from Pondicherry University and Ph.D. degree in Computer Science from Madurai Kamaraj University, Madurai, TN, South India in 2006. He is now working as the Associate Professor of Computer Science Department at M.K.University. He is an active researcher in Semantic Web, Information Retrieval, Big Data and Data Science and has published more than 100 papers in Journals and Conference Proceedings. He is a member in the Society of Digital Information and Wireless Communication (SDIWC). He has served as program chair and program committee member of many international conferences held in India for about one decade.

