iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** IV **Month of publication:** April 2018

DOI: http://doi.org/10.22214/ijraset.2018.4279

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

Social Tagging System for Community Detecting using NLP Technique

Mrs. C. Gomathi[1], S. Jameena Banu[2]
[1]*Assisstant Professor, Dept of CSE/IT, UCE-BIT CAMPUS*
[2]*PG Student, Dept of CSE/IT, UCE-BIT CAMPUS*

Abstract: ***Social tagging systems for community detecting in social networks, users usually have different intentions when tagging. Therefore, social tags may describe quite a few different aspects of the community. Social tag network is a cross-linked social graph, or a bipartite graphs (a network with two classes of vertices). These cross-linked social graphs model the associations between co-occurring tags, tags and community detection. In order to model network of social tag at an abstract level, we will represent such system as bipartite graphs with edges. Automatically clustering social tags into semantic communities would greatly boost the ability of communities to retrieve the most relevant ones at the same time improve the accuracy of tag-based service recommendation. An increasing number of users interact, collaborate, and share information through social networks. Unprecedented growth in social networks is generating a significant amount of unstructured social data. From such data, distilling communities where users have common interests and tracking domain wise comments over time are important research tracks in fields such as opinion mining. The existing community detection methods are time consuming, making it difficult to process data in real time. In this paper, dynamic unstructured data is modeled as a stream. Tag assignments stream clustering (TASC), an incremental scalable community detection method, is proposed based on natural language processing. The tags and latent interactions among users are incorporated in the method In our experiments, the social dynamic behaviors of users are first analyzed in face book and twitter datasets.***

Index Terms: Social Tagging system, Tag assignment, Community detection, Natural Language Processing, Tag based services

I. INTRODUCTION

Social tagging applications allow users to provide, create, and share more information online. Users create and share information they are interested in the form of textural content, multimedia content and social relationship information. Sharing the textural content online is widely popular and this can be in various forms such as tags, blogs, reviews, micro-blogs, comments, posts, documents and others. Due to rapid growth in popularity and a high degree of activity by their users, social tagging system has become the rich source of user information which shows their interest and ideas to explore today. Social tagging systems which users share their interested resources with other users and express their interest online by using del.icio.us, Citeu Like, Bisonomy, last.fm, flickr.com. So,the increasing popularity of these systems demonstrates that users of these systems are requiring recommender systems in order that they can access the resources they are interested in easily. Accordingly, this paper considers how to detect the community in social networks by using NLP technique can be employed to this social tagging environment and proposes an performance of community detecting that use fundamental tagging data available implicitly in these systems. The fundamental data available is tagging data and resource information (descriptions, content, etc). The proposed system focuses on top-N recommendation task which suggests a list of items. The reason is that it will be difficult to evaluate rating predictions in such systems that have no explicit user rating data. The social tagging systems for community detecting, users do not provide explicit rating on resource items that they interest. Instead of rating on items, users annotate the resources using keywords called tags. Therefore, tags indicate user's interest and preferences. The proposed system will analyze user's tagging behavior and try to estimate user's interest and preferences. The proposed system of community detecting to generate the implied topics (latent topics) on the given resources and based on these resulted topics, user's interests on these latent topics are estimated to create a user interest profile. The approach of topic modeling, LDA (Latent Dirichlet Allocation) is used to derive latent topics from the collection of resources. In topic modeling, a document is transformed into a bag of words, in which all of the words of a document are collected and the frequency of the occurrence is recorded. In LDA, documents are represented as a mixture of implied (or latent) topics, where each topic can be described as a distribution of words. In social tagging environment, instead of documents (D), users annotate each resource using keywords called tags. Therefore, in order to create topic models using LDA, resources are taken as

documents and all of the words in a document (resource) are a set of tags used to describe it by the users. Therefore, each document in social tagging system is a bag of tags used to annotate a resource.

II. RELATED WORK

Alhamid, Mohammed F et. al., [1] proposed an traditional machine learning methods and algorithms can focus their attention on understanding the user behavior on the social networks. There is a need for models and algorithms that can be employed to perform analysis on a wide variety of application and resources available on social networks. The proposed algorithm aims to achieve an automated process to mine social data automatically, discover point of interests, determines the most optimal parameters systematically, and tunes the those parameters to give the best accuracy and performance based on the user needs. In addition, this study provides a model that can be employed to perform analysis on a wide variety of application and problem domains related to utilizing social networks in a smart city environment. The use of the social analysis creates tremendous research opportunities for data scientists, machine learning researchers and application domain experts for a number of reasons. First, the social network can be used to automates many of the crucial tasks/processes involved in performing social network mining and machine learning analytics. Second, the social networks greatly accelerate the development of novel and new distributed algorithms and models that can handle social network analytics volume and/or high-dimensional. Hence, practitioners can mine and analyze social datasets to gain more useful insights into the users behavior on the social network and what their activities means and deliver the results significantly faster. Third, the output of the proposed model can also directly be used for two specific application domains: (1) social community's identifications and

(2) users interest data analysis. The contributions of this paper come to address the following: First, investigate content discovery in social networks according to the user type of interest. Second, analyze and discover the relations of trending contents, and third, identify social community using topics of interests. The proposed methodology accomplishes the objective contributions in this paper as follows: in order to fulfill the first objective of enabling content discovery, to employ social tagging to build relations between posts and sentiment analysis on textual contents to build emotional background. In order to achieve this objective, It bring two contributions. First, to design an algorithm to recognize the user's basis interaction activities on social networks as well as extracting more complex contextual situations in the online environment such as locations and emotional related wording. Second, to design an algorithm to provide the user with personal suggestions type of information, based on the recognized interest and context. In [2], they proposed social networking websites such as Facebook, Delicious, and Flickr introduce a new concept of social tagging recommender systems. These websites allow the user to associate resources with customized tags, which in turn help in carrying out the recommendation of resources. Over the last few years, a large collection of data has been generated, but the issue of privacy has been overlooked. An adversary with or without malicious intentions may reidentify a particular user in a social tagging dataset and extract the user's past tagging records. Since the tags directly expose user preferences, there exists an issue of privacy threats to the user profile. Various privacy-preserving approaches have been proposed over last few years. However, it is observed that incorporating privacy enrichment strategies come at the cost of deterioration in the quality of recommendation being offered to the user. Thus, the question that needs to be addressed is that design an efficient tagging recommender system which incorporates the user privacy with a minimal deficiency in the quality of recommendation. To propose a mechanism to address the problem stated above and implement an effective privacy based social tagging recommender system. Social tagging systems involve the use of customized tags. Tags refer to the labels given by the user to the pages or websites that they find interesting. These tags help the user to identify which pages he liked previously, and it is used for providing recommendations. Social tagging [3] is becoming widely as an important tool to classify and organize resources for sharing and finding them. Users create or/and upload content, annotate it with chosen keywords (tags), and share it with other users. In this way, the user is becoming active and involved in information production where he can enrich the content of these resources. Moreover, it can be assumed that collected tags by the user are part of this user preferences or interest, and the more a tag is used by a certain user, the more important that tag is for him. Researches have shown that social tagging can be used to improve recommender systems. To further understand the recommendation concept, we quote different types that LinkedIn offers: (1) Skills recommendation, the user is led to quote a number of skills, all of his friends have the opportunity to confirm. Therefore, they recommend him for one or more skills. (2) Free recommendation, user friends have the possibility to write a free text to recognize and testify some of his skills. (3) System recommendation, in this case, the system recommends or suggests for the user people who have similar profiles, skills and interests. In our context, we are interested in the third type, in which the system recommends handicrafts women for customers, recommends suppliers for handicrafts women and different users for each other. Social tagging enables actors such as end users, handicraft women, customers and suppliers to tag content. This action may enrich the user profile, indeed, the set of tags handled by a given

user is considered as an important indicator on user preferences. To exploit the resulting users interest to recommend mainly artisans to customers. Tang et al.,[4] is proposed by solving the low-rank approximation problem of the image-tag-user associated tensor. It illustrates the tri-clustering method is proposed to divide the tensor into several sub-tensors, in order to overcome the challenges of large-scale tensor factorization. Two variants of TTC are proposed respectively, by considering the two assumptions whether or not the sub-tensors are independent of each other. It construct three similarity matrices (i.e., image-image, user-user and tag-tag similarity matrices) from the data. The goal is to divide the large-scale and super-sparse tensor into some sub-tensors and select the denser sub-tensors. To expect that the tensor partition process can group the heterogeneous data, which is similar in some aspects, into one common sub-tensor. To this end, implement the tensor partition by utilizing the proposed tri-clustering. After tri-clustering, to select the denser sub-tensors and keep the ones with a relatively larger number of observed entries. The purpose is to complete the selected sub-tensors to refine social image tags. The tensor Tucker model and low-rank approximation to implement the sub tensor completion. Here, to investigate the dependence or independence among all the selected subtensors, and propose two variants of TTC (i.e., TTC1 and TTC2). TTC1 assumes that these sub-tensors are independent, while TTC2 assumes that these sub tensors are dependent. It integrate these reconstructed sub-tensors into the expected tensor. Finally, re-rank the tags of images based on the values of the entries. Wang, Xiaofang, et al. [5] proposed an social tagging systems, e.g., Flickr, YouTube, Amazon, allow users to annotate items (images, videos, products, etc.) with tags in order to facilitate later retrieval. The tags and items information are sharable among users, which allows users to easily find users with a common interest. Groups in social tagging systems are self-organized user communities revolving around a common interest. Along with the fast development of social tagging systems, the numbers of groups are increasing rapidly. In such a scenario, group recommendations have become an increasingly important application of social tagging systems in recent years, it can inform users about groups that they might be interested in. focus on recommending groups to users in social tagging systems. The experiments shown that quaternary relations can obtain more accurate recommendation results. We propose a 4-order tensor model , which integrates the information of users, tags, items and groups information in a unified framework To reduce the dimensionality of the 4-order tensor and mine the potential semantic relations between users and groups. Experiments on the dataset crawled from Flickr.com and comparisons with the binary and ternary relations model demonstrate the validity of our approaches. In order to acquire the users-tags-items-groups quaternary relations, we integrate these four types of entities (users, tags, items, and groups) into a 4-order tensor recommendation framework, and then, we use the Higher-Order Singular Value Decomposition (HOSVD) and Higher Order Orthogonal Iteration (HOOI) methods to reveal the latent semantic association among these entities and recommend groups to users. In this experiments on a Flickr dataset demonstrate the effectiveness of approach. In [6] an illustrate the social bookmarking systems. It attracted the interest of the research community and they provide a vast amount of user-generated annotations (tags) and reflect the interest of millions of users. The social aspect of these services derives from the fact that resources (usually web pages) are tagged by the community as a whole and not only by the creator of content alone as it is the case for services like Flickr or YouTube. Collaborative tagging, as considered it provide relevant meta-data and is expected to boost the semantic quality of labels. The on-line social detecting communities creates new opportunities in areas such as product tracking or marketing. So, bookmarking systems, too, have recently attracted interest in this regard. A model of social bookmarking systems based on a set of bipartite graphs. Then, social bookmarking systems can be detected by means of a probabilistic generative model with corresponding smoothing priors. In our experimental evaluation, we consider a corpus of 105 million delicious bookmarks, discuss the trend detection capabilities of approach and conclude by comparing our findings to the trends detected by other measures. This, however, does not take into account that most collaborative tagging systems are vulnerable to spam. After an initial analysis of our dataset, we found some users of delicious to be bots. The behavior of these automatic “users” varies but is generally characterized by a very high rate of participation and anomalous tag assignments. In order to limit the influence of spam, to apply the diffusion-of-attention concept presented. Wu, Boya, et al., [7] discussed about the problem is non-trivial and presents us with several challenges. First, although a few works demonstrate the existence of a correlation between the demographics and emotions of users, it is still unclear whether the correlation exists on image social networks. Second, model user demographics and other information (visual features, social correlations, and so forth) in a joint framework. To consider the model emotions of social images and unveil nine major emotion categories, namely, awe, amusement, contentment, excitement, anger, disgust, sadness, fear and boredom. Then, to investigate whether user demographics such as gender, marital status and occupation are related to the emotional tags of social images. To uncover several patterns, and a partially labeled factor graph model named the demographics factor graph model (D-FGM) is proposed to leverage user demographics in the modeling as different factors. The experimental results confirm the accuracy of the proposed model, achieving improvement compared with naive Bayesian and improvement compared with SVM (support vector

machine). The effectiveness of the user demographics factors is also demonstrated by the factor contribution analysis, which reveals some interesting behavioral phenomena.

III. PROPOSED METHOD

In social tagging systems, tag assignments continuously arrive because of the ongoing activities of users. Thus, the community detection method should be constantly updated to handle incoming tag assignments, which show the latest user interests. After processing an incoming tag assignment, the algorithms should output the communities in time. Clustering methods are useful tools for extracting patterns from data. However, it is difficult for traditional clustering methods to handle data streams. Moreover, the data stream requires clustering algorithms to detect emerging clusters, merge old clusters, and efficiently discard expired clusters. These requirements make data stream clustering a significant challenge for single-pass algorithms. To address this challenge, new clustering methods have been studied to efficiently discover valuable information in the data stream. In this project implement the community based detection system based on Natural language processing (NLP) is a field of computer science, artificial intelligence concerned with the interactions between computers and human (natural) languages, and, in particular, concerned with programming computers to process large natural language corpora. The challenges in natural language processing frequently involve natural language understanding, natural language generation, connecting language and machine perception, dialog systems, or some combination thereof. We can extract the keywords from face book and twitter comments and match the keywords with trained database. Finally label the domains with improved accuracy rate. The proposed work is shown in fig 1.

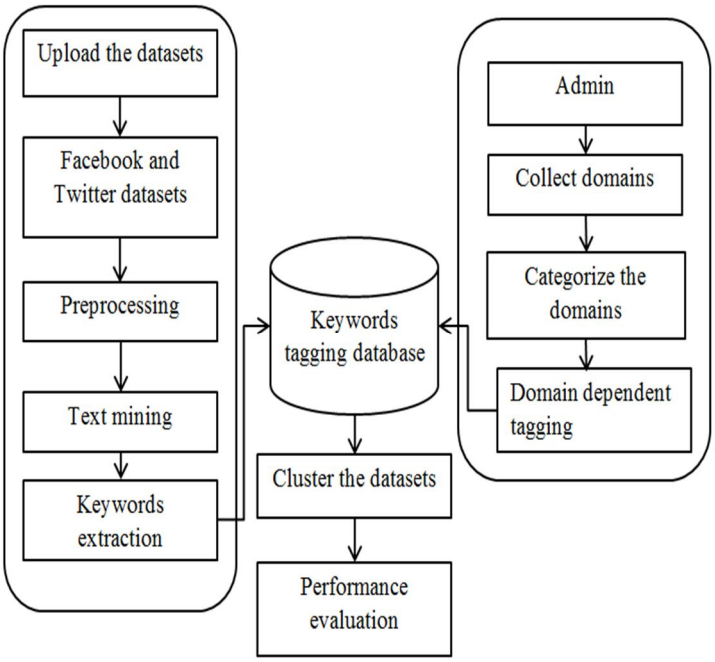


Fig 1. Social Tagging System Architecture Design

1) *Step 1:* First step of dataset acquisition to upload the datasets from twitter and face book datasets. The datasets contains various attributes such as user name and comments.
2) *Step 2:* Second step contains preprocessing such as missing value estimation and irrelevant data removal. And also analyze annual reports, finally normalize the data.
3) *Step 3:* To train the keywords related to politics, entertainment, studies and medicine domains. Domain dependent keywords are trained. Keywords are analyzed as centroid values for clustering.
4) *Step 4:* To read the comments from uploaded datasets. Apply stop words removal, stemming words analysis are implemented to detect keywords. Identify the similarity of keywords with trained database. This module eliminates the noisy words from text datasets.
5) *Step 5:* The final step to categorize the each domain comments for separate datasets and visualize the graph to easily survey the datasets in effective manner. The graphical design is used to calculate the comments counts to analyze overall reports.

IV. ANALYSIS AND IMPLEMENTATION

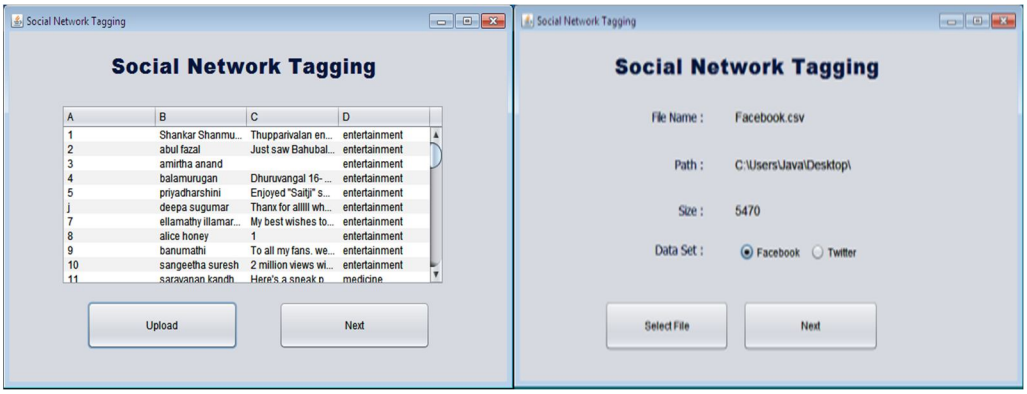


Fig 2.Upload Dataset

It describes to upload the datasets from twitter and face book datasets. The datasets contains various attributes such as user name and comments and to upload the datasets are in the form of CSV format in training phase.

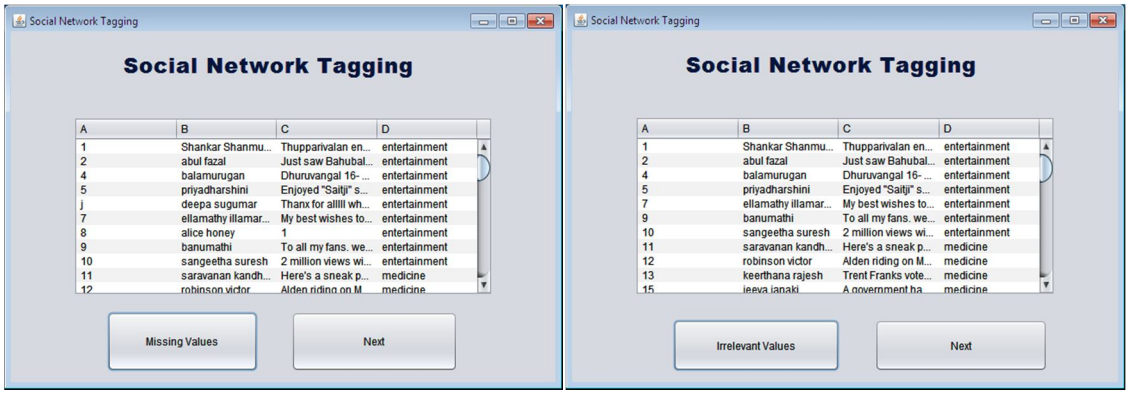


Fig 3.Finding missing values and irrelevant values in dataset

It analyze annual reports, finally normalize the data and to perform data cleaning process. Find irrelevant datasets and also missing datasets. Structured datasets are forward to further modules.

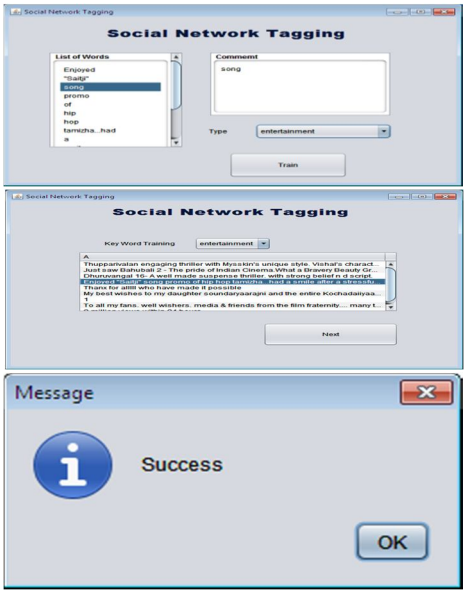


Fig 4. Keyword Training

Administrator can train the keywords related to politics, entertainment, studies and medicine domains. Domain dependent keywords are trained. Keywords are analyzed as centroid values for clustering. Finally database contains various different domain keywords.

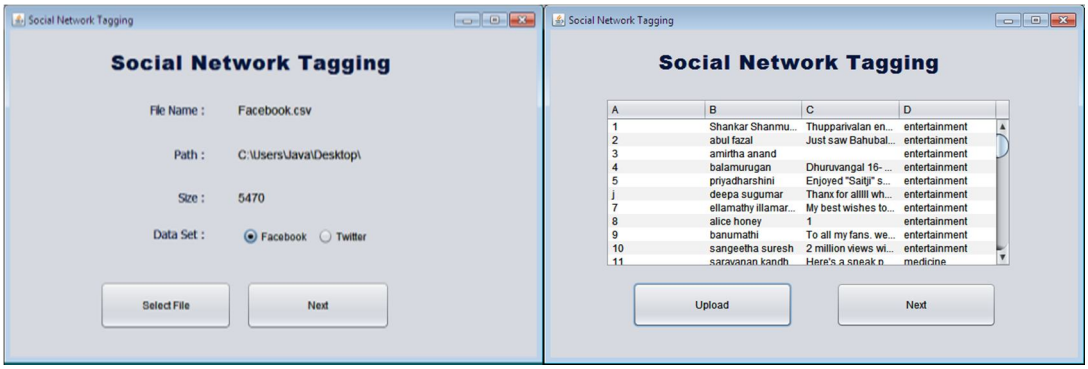


Fig 5 .Testing Dataset

It describes to module upload the datasets from twitter and face book datasets. The datasets contains various attributes such as user name and comments and to upload the datasets are in the form of CSV format .

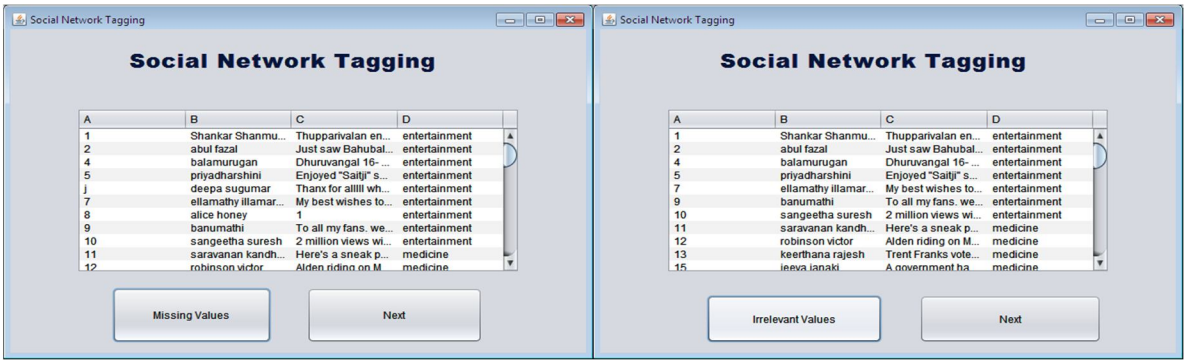


Fig 6.Remove missing values and irrelevant values

It analyze annual reports, finally normalize the data and to perform data cleaning process. Eliminate irrelevant datasets and also missing datasets. Structured datasets are forward to further modules.

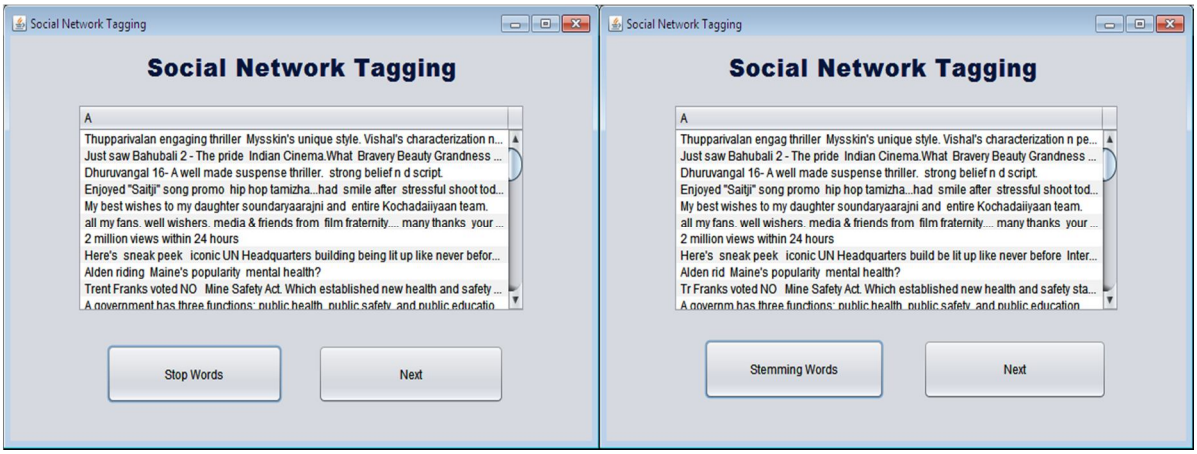


Fig 7.Remove stop words and stemming words

It illustrates to read the comments from uploaded datasets. Apply stop words removal, stemming words analysis are implemented to detect keywords. Identify the similarity of keywords with trained database. This module eliminates the noisy words from text datasets.

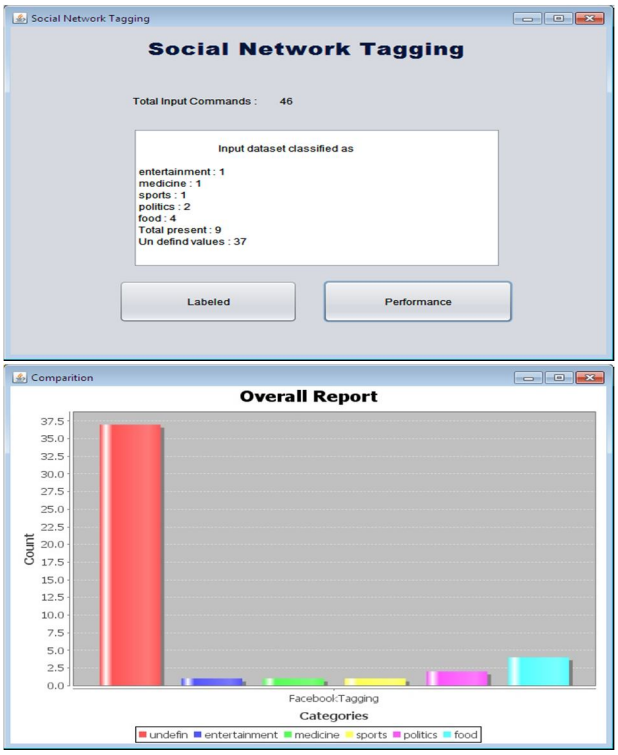


Fig 8.Performance of Facebook Tagging

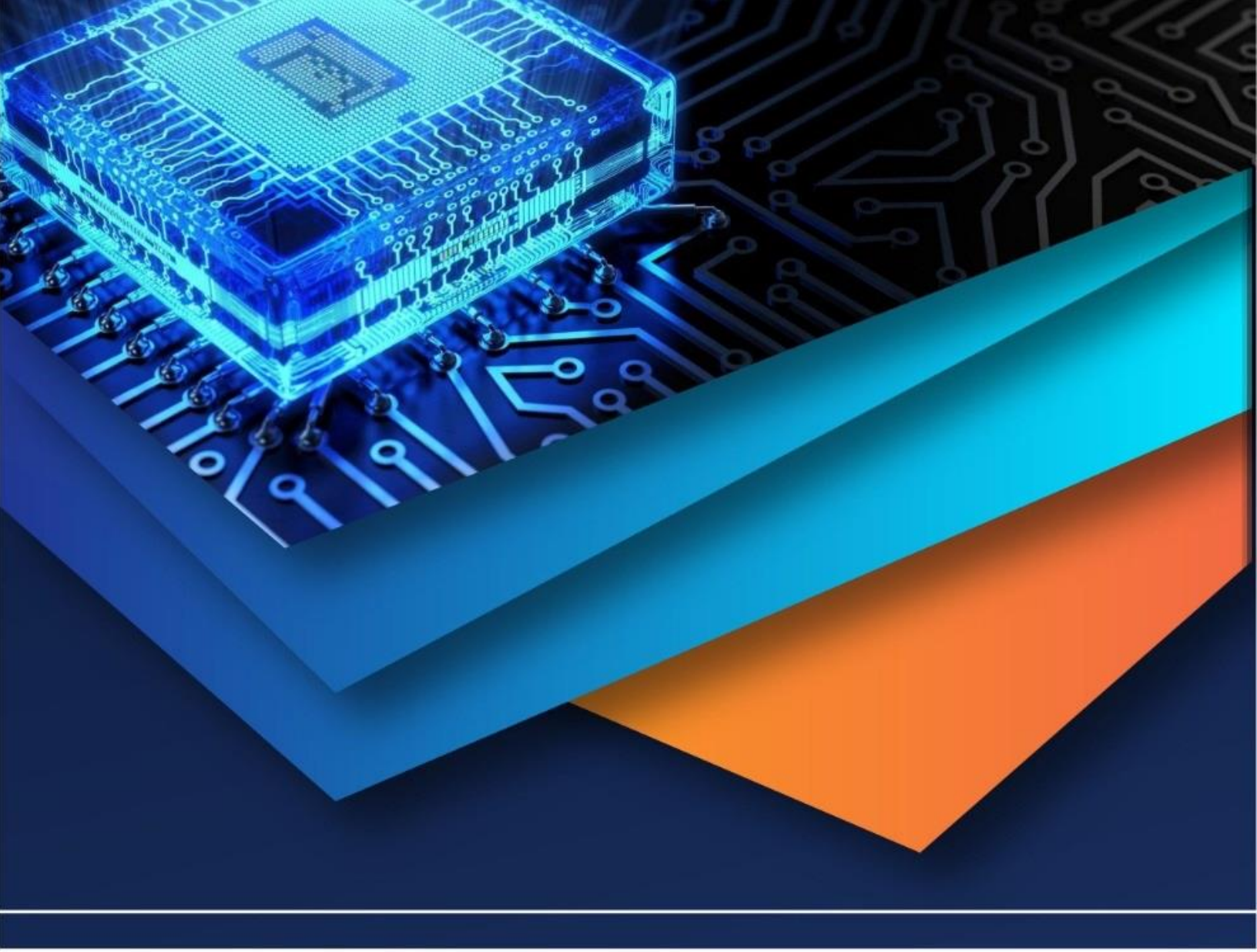
Finally, administrator categorize the each domain comments for separate datasets and visualize the graph to easily survey the datasets in effective manner. The graphical design is used to calculate the comments counts to analyze overall reports.

V. CONCLUSION

Frequent user activity in online social networks generates a large amount of unstructured social data. Analysis of big social data in real time can enable improved decision making, timely execution of meaningful actions, and accurate prediction of topic trends. Moreover, real-time analysis is useful for discovering the opinions of users over time and recommending to users their information of interest. The work presented in this paper considered the above in the context of big social data processing. Big social data has highly dynamic characteristics; consequently, it is a challenging to efficiently and effectively retrieve useful information from it. In fact, crowd intelligence is embedded in the behaviors of online users. Therefore, two datasets collected from face book and twitter and Comments were analyzed to elucidate dynamic user behaviors. The results showed that user behaviors among the two social tagging systems were different. In social networks, community detection is important for other studies, such as opinion mining and personalized information retrieval. To efficiently and effectively detect communities, an incremental method was applied. The results showed that the communities more effectively.

REFERENCES

[1] Alhamid, Mohammed F. "Social tagging recommendation system for smart city environments." Multimedia & Expo Workshops (ICMEW), 2016 IEEE International Conference on. IEEE, 2016.
[2] Katarya, Rahul, and Om PrakashVerma. "Privacy-Preserving and Secure Recommender System Enhance with K-NN and Social Tagging." Cyber Security and Cloud Computing (CSCloud), 2017 IEEE 4th International Conference on. IEEE, 2017.
[3] Kichou, Saida, et al. "Handicraft women Recommendation Approach based on User's Social Tagging Operations." Web Intelligence (WI), 2016 IEEE/WIC/ACM International Conference on. IEEE, 2016.
[4] Tang, Jinhui, et al. "Tri-clustered tensor completion for social-aware image tag refinement." IEEE transactions on pattern analysis and machine intelligence 39.8 (2017): 1662-1674.
[5] Wang, Xiaofang, et al. "Group recommendation in social tagging systems by consistent utilization of items and tags information." Informative and Cybernetics for Computational Social Systems (ICCSS), 2016 3rd International Conference on. IEEE, 2016.
[6] Wetzker, Robert, et al. "Detecting trends in social bookmarking systems using a probabilistic generative model and smoothing." Pattern Recognition, 2008. ICPR 2008. 19th International Conference on. IEEE, 2008.
[7] Wu, Boya, et al. "Inferring Emotional Tags From Social Images With User Demographics." IEEE Transactions on Multimedia (2017).



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY