We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Knowledge Extraction from Open Data Repository

Vijayalakshmi Kakulapati

Abstract

The explosion of affluent social networks, online communities, and jointly generated information resources has accelerated the convergence of technological and social networks producing environments that reveal both the framework of the underlying information arrangements and the collective formation of their members. In studying the consequences of these developments, we face the opportunity to analyze the POD repository at unprecedented scale levels and extract useful information from query log data. This chapter aim is to improve the performance of a POD repository from a different point of view. Firstly, we propose a novel query recommender system to help users shorten their query sessions. The idea is to find shortcuts to speed up the user interaction with the open data repository and decrease the number of queries submitted. The proposed model, based on pseudo-relevance feedback, formalizes exploiting the knowledge mined from query logs to help users rapidly satisfy their information need.

Keywords: Data Mining, Query, Public Open Data, Social Network, Knowledge Extraction

1. Introduction

SNS (Social networking services) is online services, platforms, or sites designed to support the development of Internet-based communities or community links between, for instance, individuals who often regularly interact with hobbies, experiences, and emotional interactions. SNS includes an account of every member and its community ties and a range of additional capabilities (typically a biography). A significant number of SNS are social media and allow consumers to communicate Online like e-mail and automatic messages. While SNS is often an individual-centered service in a broad context, social media facilities are a team. Social media platforms may be regarding constitute SNS. Online Communities enable individuals inside individual unique systems to exchange opinions, tasks, experiences, and goals.

Social networks communicate to interact in many innovative approaches, such as shows, hashtags, perform and engages electronically, revealing further cooperation and projected benefit that could scarcely imagine only a short time before. Online communities can play a significant role in the organizational processes as well as helping to develop company concepts and emotional responses and give up different prospects for the examination of social interactions and social behavior.

Presently, people rely on social media and its vast and diverse wealth and have progressively penetrated each human living area. Increasingly individuals prefer to engage valuable time on social media to develop a significant social entertaining community and again try to communicate with each other so often that the interaction around them is robust. POD repository analytics is perhaps a commonly used scientific and commercial approach for investigating the social media of interpersonal, organizational, and corporate links. The necessity for solid knowledge in DPO analysis has lately increased with ready availability to computational power and the rise of social popular social networking platforms such as LinkedIn, Twitter, Netlog, and more.

Twitter social network by study the contents of the tweets and the links between the tweets to extract knowledge from log data. By selecting buzzwords, began the 'Twitter review and then collecting all Twitter posts (Tweets) correlated to the keywords. It is a social-economical problem in India. Mining the query log based on social networks like Facebook, Twitter, etc. Study and address the discovery, access, and citation of POD repositories like Twitter data sets; and strengthening educational programs of academics of current and future generations specializing in such areas. This is an auspicious time for extracting useful information from social media query log data. Substantial efforts to decipher large amounts of data are steps towards complete search log records integrating POD repository analysis. From these data sets, we extract valuable knowledge.

The search log obtained by user actions with the Public Open Data (POD) database is an excellent data collection for improving its efficiency and the effectiveness of the online community. The data in the user input logs are gathering from individuals who communicate on online platforms. The search log assessment is complicating due to the variety of customers and diverse resources. As a result, numerous scientific articles written about query log analysis.

The word "data set" can also describe the data in a set of specifically relevant tables that correlate to a specific investigation or occurrence. Records generated by satellites testing hypotheses using devices aboard communications satellites are one instance of such a category. A data source is the standard measurement for data provided in a POD repository in the open data domain. Over a quarter a million datasets are gathering on the European Open Data platform. Alternative interpretations were presented in this area, although there is presently no accurate statement. Various difficulties (relevant data resources, non-relational datasets, etc.) make reaching a compromise more challenging. The utilization of query logs for knowledge discovery improves the speed of the POD repositories and improves the use of open data source capabilities.

POD repository analysis and mining for valuable extract knowledge from query log data. We perform on knowledge discovery, ML or similar approaches, challenges connected to pre-processing and model assessment, for data sets (web usage log files, query logs, collection of documents), and collaborative data (images, videos, and their explanations multi-channel handling data). We summarize the fundamental results concerning query logs: analyses, procedures used to retrieve knowledge, the outstanding results, most practical applications, and open issues and possibilities that remain to be studied. We discuss how the retrieved knowledge can be utilized to progress different social media class features, mainly its effectiveness and efficiency.

In addition, several concurrent inquiries of multiple distinct users are addressing by business social networks. The query stream has simultaneously been defining by a stop-time rate, making it impossible for the POD repository to generate massive query load times without over-sizing [1]. Web Search engines Query Logs Social

Network Analysis [2], Web search engine quality approaches developed for query logs, November 2013.

1.1 Motivation

To observe and comment on Twitter users, Twitter is used to share channels for personal information (intimate and confidential life). Since this proportion of Twitter users' Tweets increases significantly and the clickthrough rate, the number of RSs (recommender systems) changes its methodological approaches for the same query-based RSs system. Hence, user perception and feelings include effectively user tweets connected to user's decision making. It is hard to complete and comprehend the retrieval of the needed words from the customer. ML classification approaches from the content evaluation represent the finest and most helpful approach for trend analytics and sentiment modeling "learn the user's query patterns and generate the query functions with good predictions." ML-based RSs are used to classify user tweets and make suggestions. The feature selection and classification responsibilities in the ML method of customer tweets are significant in designing effective RS. But there is no assessment interpretation for ML-based RSs. The primary goal of the work for RS is to give awareness based on recommendations, choices, and choices by enhancing the effectiveness of their suggestions through a suitable system model.

A Recommender system for dynamic Tweets offers content based on user preferences and desires by evaluating the chronological, current, and user tweets posting pertinent data. Interactive tweets that are most suited to specific users and tweet logs data are proposing. The excess issues of information for data analysis and collection are reducing, and the search publishes user tweets in a dynamically and personalized fashion and makes precise and accurate demands.

DTSRS (Dynamic Tweets Status Recommender system) retrieves relevant data from public tweets that allow an adequate content-aware comprehension, accessible to most appropriate tweets for redundant user tweets. The advantage of the developed system DTSRS is that the user does not have to consume enough attention-seeking tweets, helping to make an effort too minimal and testing the tweets' dynamics to the user, reducing and increasing the user's happiness.

User-Query Centered Recommender System (UQCRS) is applied to exploit different measures to demonstrate the efficiency of recommendations delivered. The proposed algorithm exhibits an effective result to the search shortcuts issues.

1.2 Challenges in twitter content query-based recommendation

1. Retrieving the collection of Twitter tweets that match with one or more content keywords of user-query.

2. Ranking the query within the text.

3. Develop a method for user-query-centered knowledgebase integration.

4. Predict the outcome of the automatic query-analyzer of Twitter tweets concerning the recommendations

These challenges can be solved using query expansion and semantic models. In query expansion, the reformulation of the query is made based on the vocabulary mismatch among the query and content retrieved. Through semantic models, similar words of the user query are extracting.

1.3 Objectives

1. Scalability and Real-Time Performance Analysis.

2. Discovering the inherent variability in mining the POD repository using query log data.

3. Comprehensive quality measures analysis.

4. Algorithmic consistency across domains.

5. Valuable identification of information and domain services using POD repository data analysis.

CF (Collaborative Filtering) and CBR (Content-based recommender) systems are the predominant forms of recommendation. The content-based recommender system, which is our subject of interest? In the Twitter recommender system, Twitter tweets' essential nature is noisy and with less content for understanding because of the exact use of posting user [3]. Creating the strictly relevant content of the recommendation system with the account authority is considered in [4], with the learning to rank algorithm considered. The relevant content is a similar type of information retrieval [5] using the tweet contents posted by the user and user friends, which provides recommended set of tweets to the user. The Recommendation System [6, 7] is using to construct the query in the Twitter content query-base recommender system by identifying relevant Recommended tweets.

2. Literature survey

Social media connects next to each other individuals in various methods, such as web-based gaming, marking, earning, and socializing, showing effective technologies to collaborate and communicate that were unthinkable only recently. In addition, online communities contribute to the corporate strategy and assist in altering economic models, sentiments and introduce various opportunities for studying direct intervention and collaborative actions.

Several previous researchers suggested using the Internet search query records to derive linguistic relations between queries or terms. The idea that the web search query logs provide knowledge via clicks confirms the relation among searches and records selected by individuals. The writers relate questions and words in the information gathered based on these data. This technique has also been using to group requests from log files. The cross-reference text is linked to similarities depending on query information, proximity editing, and hierarchical resources to identify better clusters. Such clusters are utilizing to discover identical queries for querying systems.

Twitter is a massive amount of information social network, to perform an analysis on Twitter, a keyword-based search for possible and relevant posts [8], where such search keywords cover all the possible tweets of the user [9], which is a lengthy and time-consuming process. Typically, to reduce the complexity of searching the posts from a Twitter data source, a user search keyword identification is made [10] to reduce the manual effort. User search keywords extraction is developing on the target keywords instead of the general word phrase of the keyword selected. This keyword extraction process is iterative because of the user's regular interaction in the social network through a web search and advertising. So there is a need for a query

recommendation and query expansion system to improve the keyword extraction process from Twitter tweets and provide recommendations for Twitter users. The keyword's frequency statistics and machine learning models recommend [11, 12] classifying the keywords and extracting them from the Twitter tweets. The search-based dataset is given in [13] to find the keyword topics and search the keywords in the dataset, but for the enormous tweets, this method results in relevant tweets or empty tweets because of a vast Twitter dataset. Therefore, keyword recommendation in search based through query suggestion is recommended [14, 15]. In keyword recommendation, the query system is designed based on the relevant keywords in the Twitter tweets through query log mining and search query suggestions [16]. The query expansion [17] from the original query is complete with expanding and improving query ranking for the searched tweets through query suggestions.

In recent years, POD repository analyses have received considerable interest mainly because of the increase in online microblogging and media disseminating websites and the generating substantial of an extensive POD repository. Furthermore, despite increasing attention, the significant financial uses of online communities in extraction are not very well understood. While there is a great deal of study on various issues and methodologies of POD repository extraction, there is a difference between approaches produced and used in practical situations by the researchers. Thus, such approaches are yet relatively unknown to their corporate development implications.

Though there is a significant difference between the recently established POD deposition extraction technologies and their application. Some sectors, such as the telecommunications business, whereas charges are appropriately structured to indicate the names and the persons as operators), POD results indicate POD results substantial market orientation synergy. Moreover, most POD repository analytics studies have focused on generic issues previously mentioned rather than considering particular commercial uses. As a result, the possible use of the POD repository evaluation and analysis in the industry is hardly known.

1. The future generation information processing must make it much easier to routinely analyze diverse data sets such as extensive reading, videos, and knowledge created by users like blogging.

2. Innovative approaches are essential for the data analysis of numerous contributing inputs, for example, contextual performance, clinical development research findings, previous warnings, incidence history.

3. Novel approaches require decentralized systems where the related items are searching, and the similarity's reliability needs to be measured.

4. In observational predictive analytics, topic specialists require sophisticated higher transmission. That applies to advanced methods such as cartoons and non-traditional ones such as poetic analytics.

5. A recurrent topic seems to be the necessity to integrate consumer demands into any novel computational technology, system, or technique as a type of data with the involvement of domain knowledge.

Mining methods workflow contains a group of mining data and models, with an utmost data operator work to set the parameters of the mining model used. In mining, the data is not expressing indirect form, but it is unseen in the model connectors. The user provides the indirect form of data and applies a model on the

indirect form of data, generating the direct form of data. During this process, mining techniques should distinguish between components, are data model, operators, and parameters. Enable the user in designing such mining for the web. There is a need for the development of online data workflow through concepts and categories. Online data refers to a frequent visitor of a group or several web pages in social networks to cover and gather the complete user required information through locating the web page and fetching the desired user valid information.

Web pages are complete application-specific to fetch the user's desired target information through the user-defined keywords using a constrained specific web application to provide up-to-date information through the Online Social Networks (OSN). OSN is protecting billions of active and passive web user's knowledges. The rapid change in social networking sites has proven an exponential growth in user information and knowledge exchange rate. According to [18], two-thirds of the online users browse a social network or an eCommerce website, with an average of 10% of all internet utilization time. By covering such a large amount of helpful information exchange, OSNs through social media become an excellent platform for mining techniques and research in data analysis.

The method allows data on social media user tweets for goods and commodities. In [19], a RES approach is recommending to give a level of precision compared to the previous approaches used in the tweet assessment of the consumer. Fasahte et al. [20] has presented the method to anticipate tweets by utilizing the Online Reviews dataset sorting procedures. It investigates the search engine extraction and training algorithm to collect data from the unstructured text in the available online content. In addition to the keyword-based evaluation, the data model on the Internet is connecting with complex searches. They utilized to locate tweets on various tweets while maintaining the surfing data operational inside the account location. The data collection, processing of data, and data sampling are all three aspects of tweets availability. Verma et al. [21] developed a dynamic analysis classification technique by implementing ML and evaluated the different variables in these learning approaches. A public repository response assessment technique [22] discussed huge data volumes on Twitter to create the emotional state of every message. Rosenthal et al. [23] describes the user opinion mining system used to extract similar users' views from the person's view using a moderate data analysis method. Ibrahim et al. [24] established an online emotional assessment that supplied many functional tweets out of interest to identify comparable personal data. The decision relates to extracting features, extensive conversion, and different recognition using machine learning techniques in many tweet solutions for the clustering techniques, correlating the query response pattern, relationship regulations for Twitter tweet extraction, and visualization in the Tweet API application.

The phrase retrieval from several texts is provided by [25–27] since the words should be user-specific, and the searching procedure should be preserving. Due to the powerful conventional method to all these, it is possible to investigate a method that relies on the recommendation, utilizes iteratively in the search engine and advertising searching.

Optimization techniques provide AI (Artificial Intelligence) and NLP (Natural Learning Processing) capabilities in order to deliver necessary assessed user suggestions interpretations in different networks/services of social network applications [28, 29]. Interface design such as mobile web apps permits various movies, cuisine, literature, YouTube, healthcare and more information related material. Films, culture, and entertainment are communal societies. Depending on the user's awareness of the material, the recommender system [30] has problems with confidentiality and protection. Thus, classic recommenders are becoming inevitable for current user ratings and Twitter posts to evaluate user-generated content [31, 32].

A social media micro-blogging process recommendation enables 140-character consumers to utilize tweets and retweets, known as individual tweet statuses [33]. Such tweets are related unidirectionally, as it posts the tweet, and others who tweet reciprocate follow a tweet. The predefined list solely concerns the retweeting user. Tweets contain a specific user interest context and content, for example, movies or music [34]. Tweets are connecting to the abovementioned subjects. The person who posts is the origin of news, and the customer who retweets is the follower data source, try to evaluate tweet content. Microblogging is a form of communication via the consumer Twitter medium of complete information. Twitter members are classifying into three categories: the first relates to posts, where people organize into a significant number of supporters. The latter concerns the tweet requesters, where the users submit a rarely post to comprehend and observe tweet material continuously. The last concerns user twitch [35] relationships, including friends/relatives in which all the posts are personalized.

These classes classify according to the post and retweet consumers, followers, and comparable results [36]. The rating is for multiple users, supporters, and retweets with the suggested providers and searchers of knowledge. Depending on these ordering, users and followers get the reputation of relatively large text containing signs of impact in the production of tweets [37]. Posts and repost individuals are classified first, social media posts and tweets are the subsequent most important for tweet users, and the third reciprocate relation to the discussion forums of these ratings [38]. Subsequently, the relevance rating and connectivity are estimated to identify the Twitter post system by giving the prominent tweet the correct weight, which affects the linked user's position for various stuff exchanging relations [39, 40].

3. Recommender system

In this paper, UQCRS proposes a concept of the recommendation system at the user-query level that mainly aims to find the correct tweet's information extraction through content to user requirements.

3.1 Proposed recommender system

Unlike the previous works of the recommendation system, the proposed UQCRS is a system capable of performing Twitter mining from a vast Twitter database through query logs and user tweets for understanding the user interaction in the Twitter tweets. UQCRS provides the search-based tweets content recommendation through the found user-query centered content in the short and long tweets to depict the intention of the user tweets. In the proposed system, the workflow is: firstly, the Twitter background knowledge is extracting for user-query-centered knowledgebase integration. Secondly, implementing the strategy of UQCRS on the Twitter knowledge repositories. And finally, evaluation and illustration through the discussion of the proposed URCRS system are made.

3.1.1 Content-based twitter tweets detection

Twitter users communicate with the recommender systems through a user interface like a web portal or a mobile app. Depending on the user availability, the user interacts with the social network to extract the information in the tweets on user interest, which is predicting by the tweets ranking method to provide the list of proposed tweets based on the content keywords of the user query. The UQCRS data system depends on the database it stores and updates the tweets based on the

content and ratings of tweets through the query search. UQC recommendation system tweets content architecture depends on the profile and database of user Twitter profile that store query information and update the entities continuously through Twitter user customer recommendation, as represented in **Figure 1**.

With the content feedback and query-centered analysis, recommender systems are implementing in **Figure 1**, used within the e-commerce websites, to guide the Twitter customers through retrieving log data and Twitter mining originate themselves.
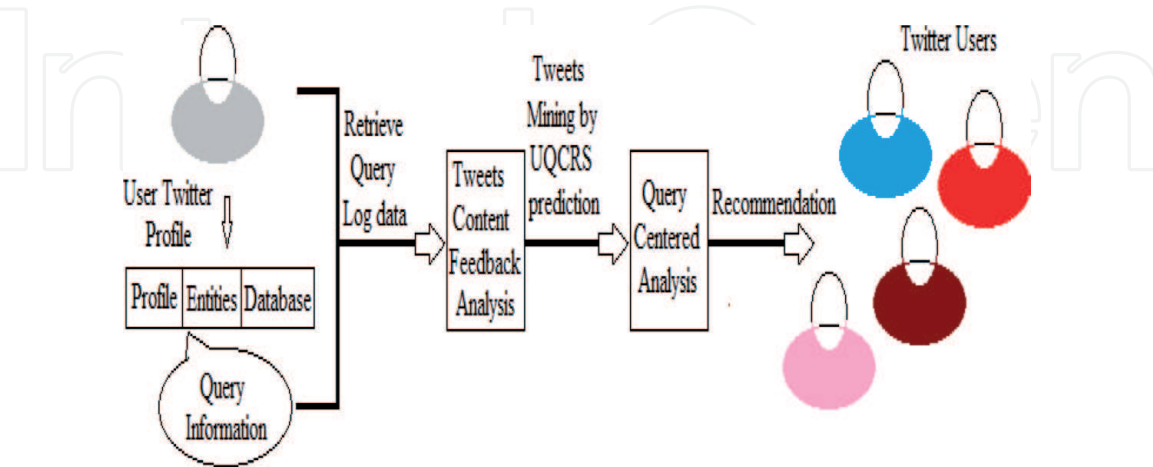

Figure 1.
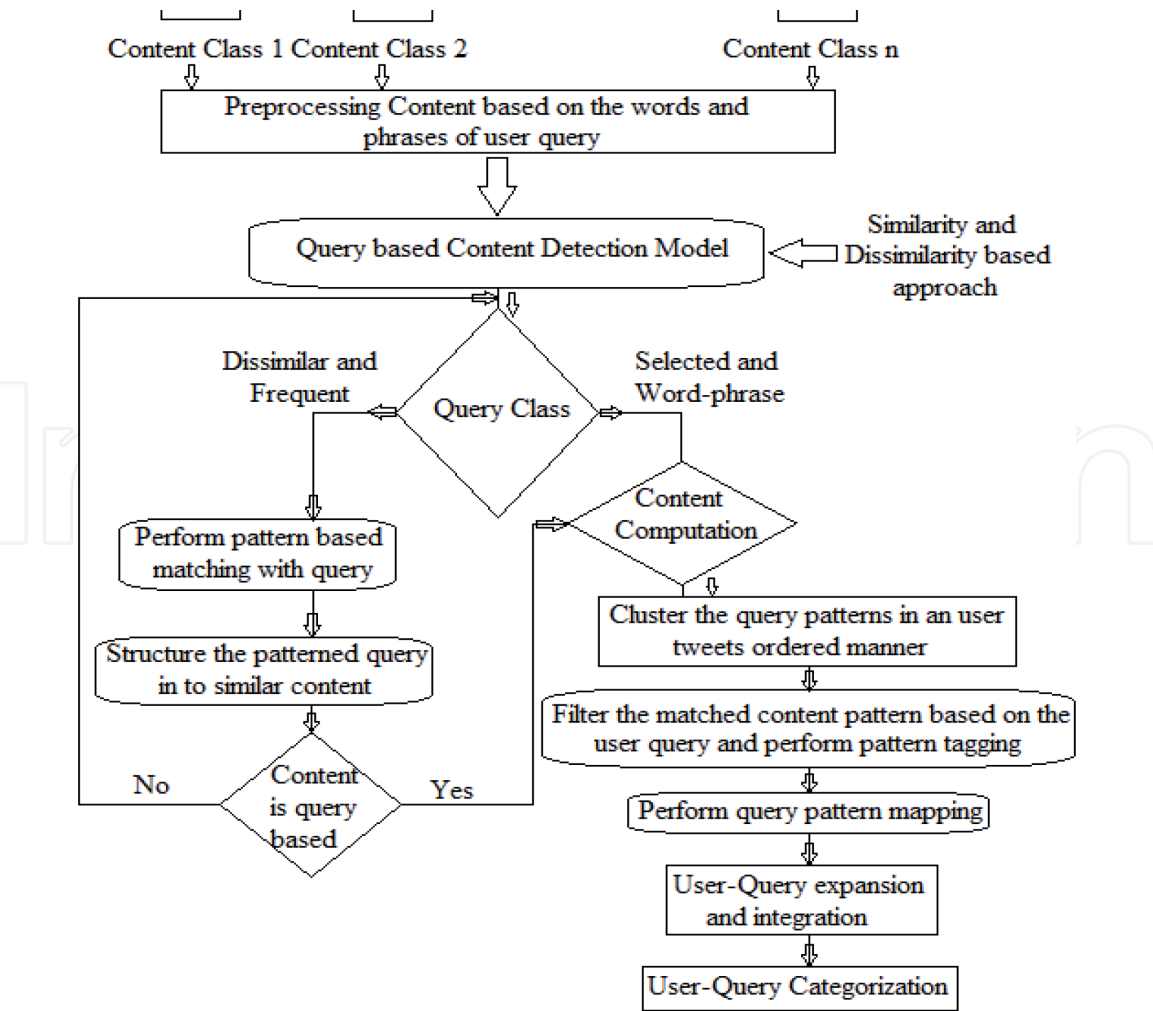*Architecture of the UQCRS implementation.*

Figure 2.
User-query pattern categorization.

The tweets clustering process and the tweets' filtering are performing, shown in **Figure 2**, of the tweets as given by the user recommendation, which analyses at every instant of user interaction.

In the proposed system, each analysis is a service, and the Content Detection Model algorithm's operation is explaining in **Figure 3**, which is describing in two parts. The first content phrase built as a final set of query phrases per query log

---

Algorithm 1: Content Detection Model

input: *set of content class $C[t_i]$ for query log q_i*

output: *set of similar tweets T_s*

T_s = Null

index.pattern($C[t_i]$, p), p is the content phrase

for all non-empty class $C_{x,y} \in$ index do

   $S_r = C_{x,y}$ $C_s=[C_{x-1,\ y-1}\ ,..., C_{x+1,\ y+1}]$

 if $|C_s| \geq \tau$, τ is the ordered pattern

for each $s_r \in S_r$ do

      f=Filter (s_r, p), $|f| \geq \tau$, map (s_r, c_s) $\leq$ p, $c_s \in C_s$

for each $c_j \in$ f do

  if $\{s_r, c_j\}$ is matched then

         compute query mapping $\{q_1, q_2,..q_n\}$

         given by q_i & number of p

  for each query mapping

         $q_k \in \{q_1,q_2\}$ do

           $q = q_k \cap$ f

 if $|q| \geq \tau$ then T_s.Add(q)
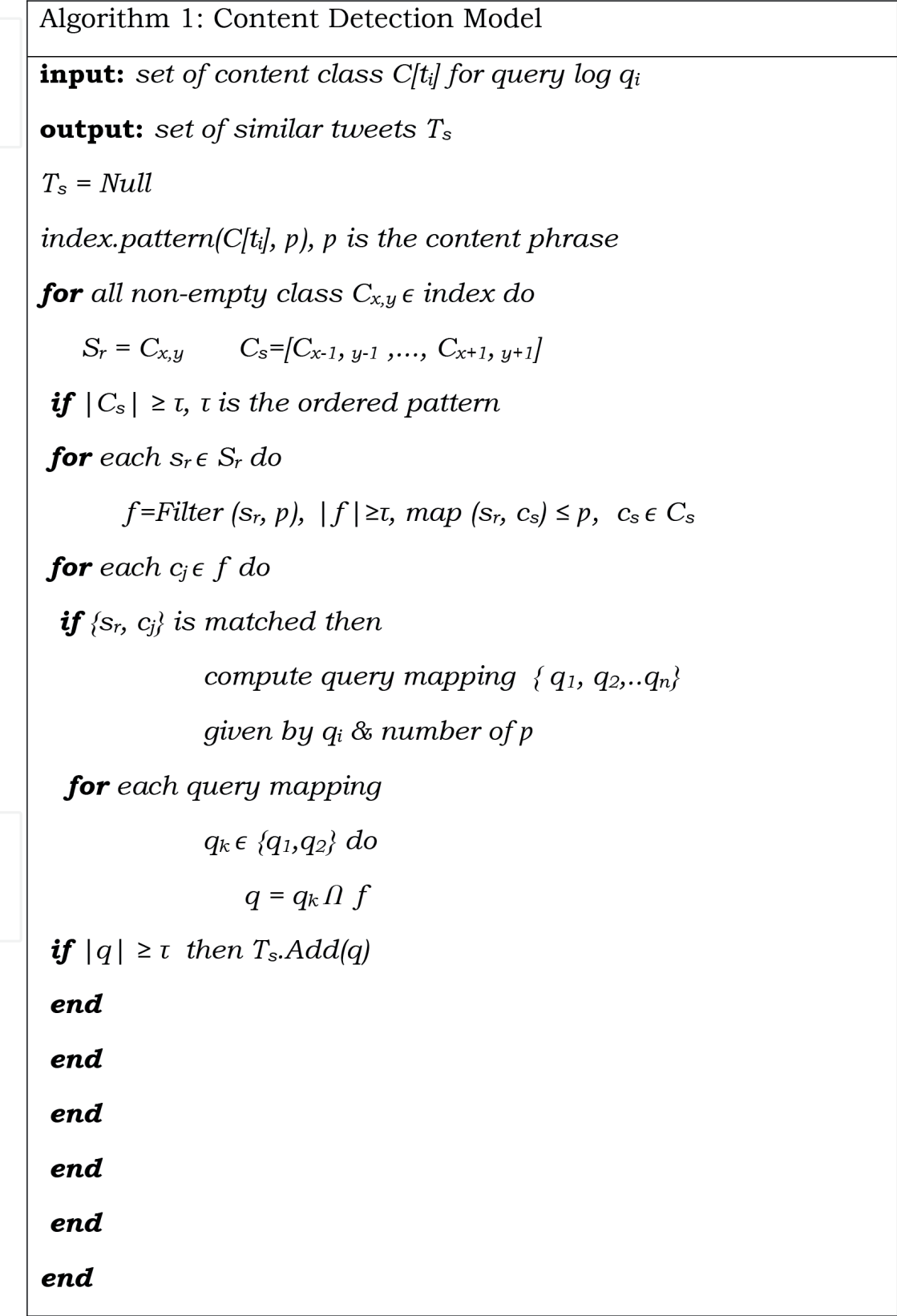
end

end

end

end

end

end

Figure 3.
User twitter tweets detection.

brings together a maximum number of ordered patterns that make each filtered design generate enough matched tweets. In the second, query mapping is computing that shares similar tweets during consecutive query logs, exceeding the maximum ordered pattern.

3.1.2 User-query centered knowledgebase integration

User query analysis is complete by extracting knowledge from query log data, shown in **Figures 3** and **4**. Here, the scenario is that successive user tweets are removed based on the query log data, and the matched query is accessed
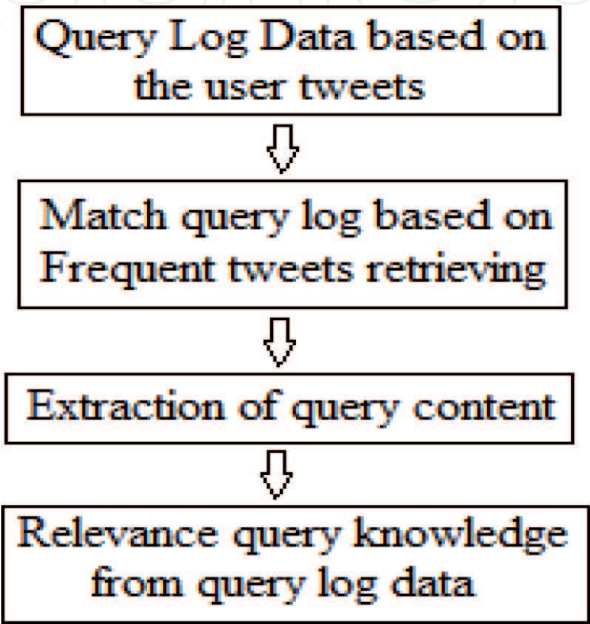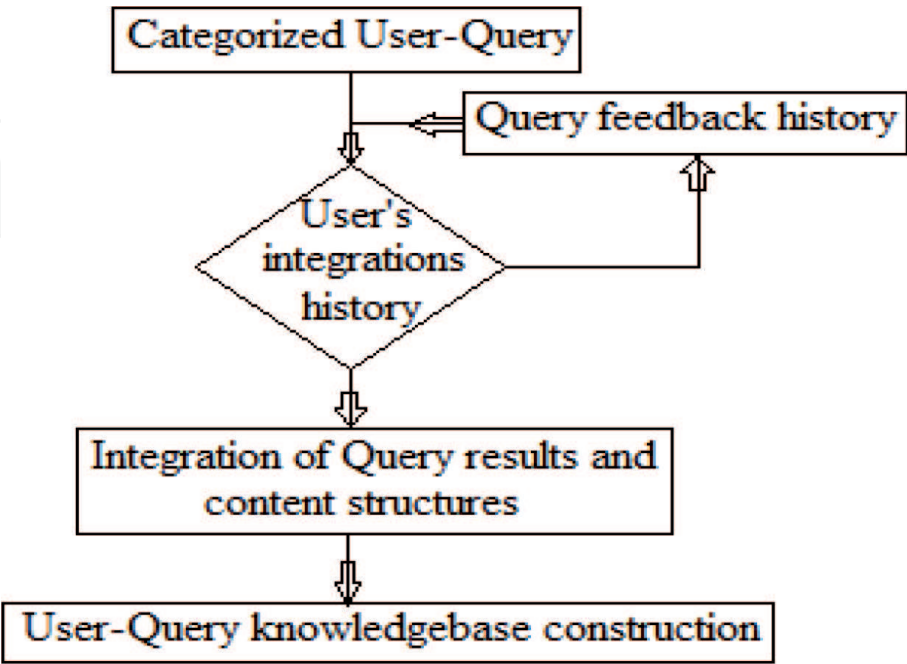
Figure 4.
Extracting the knowledge from query log data.

Figure 5.
User-query centered knowledgebase integration.

correspondingly through subsequent extraction. The accessed query content related to a Twitter account focuses on extracting the respective accounts retrieved from the tweets account. The latter tweet's accounts focus on the generic tweets with similar content, with identical query phrases. But during relevance query knowledge, similar user query search with less effective and provide the presence of user tweets with high similarity with different text types.

The scenario of user-query knowledgebase construction is showing in **Figure 5**. Here, from the previous methods, with the selected user-query procedure used as a tweet query category, user integration history is proposed, with knowledge extraction at every query feedback history loop-back. For query integration, the projected tweets from one day to many days are structured for continuous knowledge construction through the constant tweet's attribution and trace the future knowledge analysis, which evolves in the future based on the tweeted user query.

The model of the proposed recommendation system is shown in **Figure 6**, consists of:

a. a set T_U of N users, $T_U = \{T_U1, T_U2, T_U3,...., T_UN\}$

b. a set C of M items, $C = \{c_1, c_2, c_3,....., c_M\}$

c. a query cluster matrix Q_C, $Q_C = [qc_{mn}]$ where $m \in T_U$ and $n \in C$

d. a set f of N feature query sets, $f = [f_{mn}]$

e. a tweet knowledge weights $K_\omega = \{\omega_1, \omega_2, ... \omega_N\}$

User item set is associated with the number of feature vectors representing the tweet customers with different tweet phrases assigned to the user-query content model. In the recommender content model, the decision ranking prediction compares the users and Twitter queries in the categorized user-query item set and tweet's weights.

Figure 6.
The proposed user-query centered recommender system.

4. Results and discussions

For experiments, a random public user Twitter dataset and real-time data using the API of Twitter is complete. The Twitter tweets containing the keywords "basket," "pencil," "work," "enter," and "formal" from the public domain are taking as the standard bag-of-words approach. Used this dataset for classification and collected 300 documents in each of the public domains.

For the classification of tweets, the true +ves, true -ves, false +ves, and false -ves constraints are utilized to equate the consequences of the classifier under the test with investigation techniques, which is illustrating in **Figure 7**.

The relations between TP, FP, FN, and TN are:

a. The relation $TP/[TP+FP]$ is describing as precision, which is the correctly classified metric.

b. The relation $TP/[TP+FN]$ is describing as Recall, which is the actual classified metric.

	Actual (Expectation)	Class
Predicted Class (Observations)	True +ve (TP) Correct Result	False +ve (FP) Unexpected Result
	False -ve (FN) Missing Result	True -ve (TN) Correct absence of result

Figure 7.
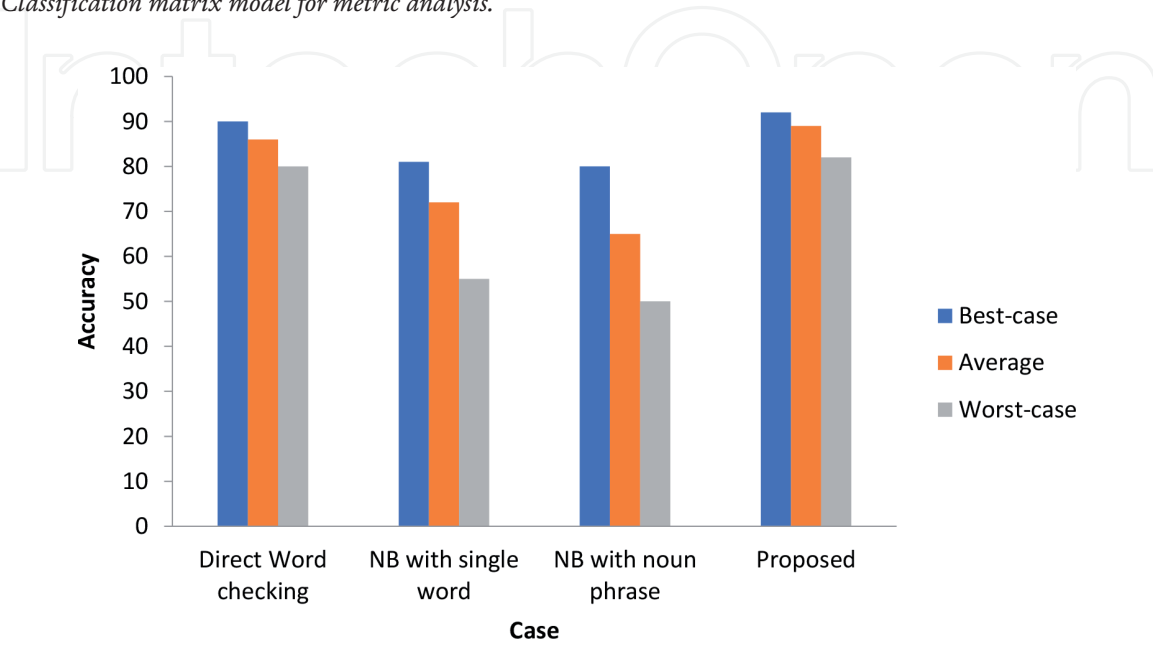*Classification matrix model for metric analysis.*

Figure 8.
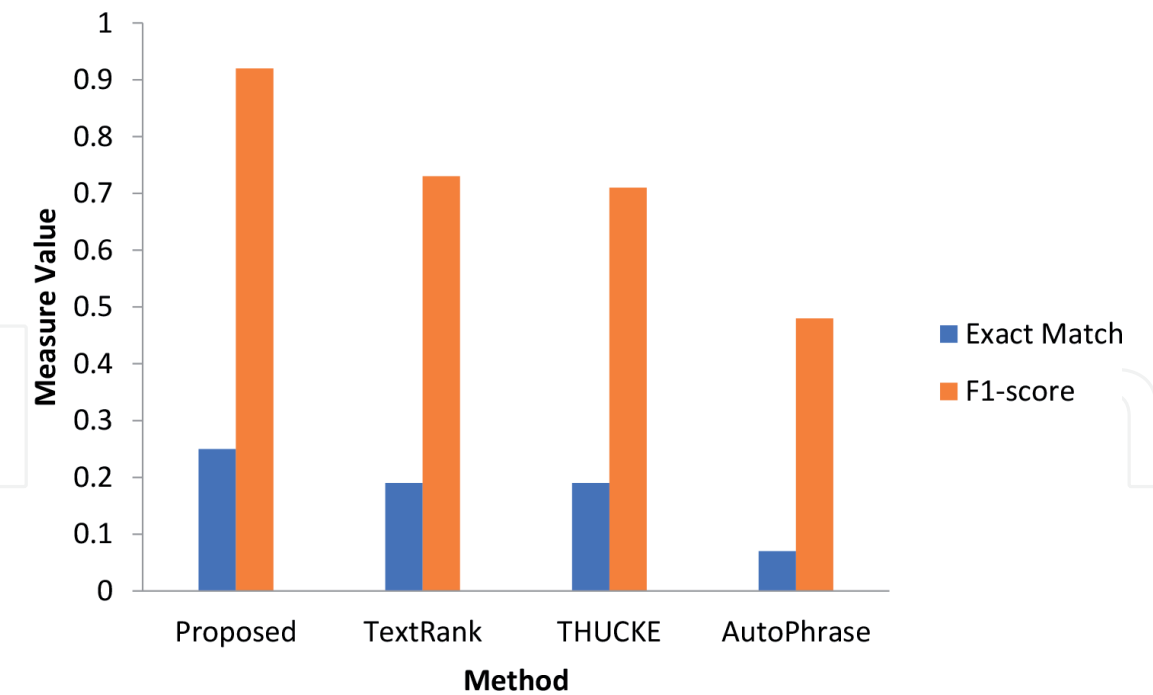*Accuracy comparison of a classified tweet.*

Figure 9.
Comparison of different algorithms for measured values.

c. Relation $2 * \left[\dfrac{precision * recall}{precision + recall} \right]$ is describing as the F-measure, which is a measure of precision and recall.

Figure 8 shows the accuracy of classifying a user query tweet in the users defined the recommended system. The highest accuracy is achieved through the proposed work, with an incredible number of word phrases, depending on the content of the user query and compared with Naive Bayes classifier (NB) [41].

Figure 9 compares the proposed system with different dataset approaches in terms of F1-score and the exact match. Because of the phrase-based content mining is made and tweets analysis is made accurately on two other datasets of different methods [42, 43] and proposed.

5. Conclusions

Social media connects connected individuals in numerous ways. For instance, online people can interact, communicate, collaborate, and socialize, showing new types of integration and interaction that were difficult to imagine only a short period before. Online communities also play a significant role in entrepreneurship, influence company feelings, and concepts and create many opportunities for unique investigation of person and team performance.   In this chapter, a user query-centered recommendation system is deliberate to improve user-query analysis and tweets analysis for Twitter tweets. The proposed and implemented query catego-rization gives a satisfactory exact match performance in tweets categorizing. The content model is the most important model for the majority of tweets. Apart from finding the tweets, the phrases are identified and located in accuracy performance metric. As discussed, content integration is helpful for tweets match retrieval. For the given user tweet query, if the aim is to retrieve similar tweets from the public repository, retrieving the keywords is improved with the use of words and phrases.

Also, a novel algorithm based on content detection is using to extract the tweets using the bags-of-word method. Using the tweet's knowledge weights the proposed recommendation system avoids the dissimilar tweet's pattern identification problem. The above said three parameters are complete, which indicates that the proposed approach produces better accuracy results than the other methods.

6. Future scope

Future work focuses on The DTSR System may be extended along with additional user profiles such as film playlists, community groups, social media tweets, user emotion, user posts, and feature tweets to better the method recommended.

Author details

Vijayalakshmi Kakulapati
Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India

*Address all correspondence to: vldms@yahoo.com

IntechOpen

References

[1] References: Wayne Xin Zhao, et al. "Incorporating Social Role Theory into Topic Models for Social Media Content Analysis," IEEE Transactions On Knowledge And Data Engineering, Vol. 27, No. 4, April 2015.

[2] Metin Turan, et al. "Automatize Document Topic and Subtopic Detection with Support of a Corpus," Social and Behavioral Sciences, Published by Elsevier, DOI: 10.1016/j.sbspro.2015.02.373, 2015.

[3] K. D. Rosa, et al. "Topical clustering of tweets," Proceedings of the ACM SIGIR: SWSM, 2011.

[4] Y. Duan, L. et al. "An empirical study on learning to the rank of tweets," in Proceedings of the 23rd COLING, 2010, pp. 295-303.

[5] M. Pennacchiotti, et al. "Making your interests follow you on Twitter," in Proceedings of the 21st CIKM, 2012, pp. 165-174.

[6] A. Pal et al., "Identifying topical authorities in microblogs," in Proceedings of the 4th ACM WSMINING. ACM, 2011, pp. 45-54.

[7] J. Weng, et al. "Twitter rank: finding topic sensitive influential Twitterers," in Proceedings of the 3rd ACM WSMINING, 2010, pp. 261-270.

[8] Turney, P. D. 2000. Learning algorithms for keyphrase extraction. Information Retrieval 2(4):303-336.

[9] Zhao, W. X et al. 2011. Topical keyphrase extraction from Twitter. In ACL, 379-388.

[10] El-Kishky, A.; et al. 2014. Scalable topical phrase mining from text corpora. VLDB 8(3):305-316.

[11] Danilevsky, M.; et al. 2014. Automatic construction and ranking of topical keyphrases on collections of short documents. In Proceedings of MINING.

[12] Zhao, W. X.; et al. 2011. Topical keyphrase extraction from Twitter. In ACL, 379-388.

[13] King, G.; et al. 2014. Computer-assisted keyword and document set discovery from the unstructured text—copy at http://j.mp/1qdVqhx 456.

[14] Luke, T.; et al. 2013. A framework for specific term recommendation systems. In SIGIR, 1093– 1094.

[15] Bhatia, S; et al., 2011. Query suggestions in the absence of query logs. In SIGIR, 795-804.

[16] Zhang, Y.; et al. 2014. Bid keyword suggestion in sponsored search based on competitiveness and relevance. Information Processing & Management 50(4):508-523.

[17] Hahm, G. J.; et al. 2014. A personalized query expansion approach for engineering document retrieval. Advanced Engineering Informatics 28(4):344-359.

[18] Global Faces and Networked Places, A Neilsen report on Social Networking's New Global Footprint, March 2009. Neilsen company.

[19] Z. Tan et al. "An efficient similarity measure for user-based collaborative filtering recommender systems inspired by the physical resonance principle," IEEE Access, vol. 5, pp. 27211-27228, 2017.

[20] U. Fasahte, et al. "Hotel recommendation system," Imperial Journal of Interdisciplinary Research, vol. 3, no. 11, 2017.

[21] A. Verma et al. "A hybrid genre-based recommender system for movies

using genetic algorithm and kNN approach," International Journal of Innovations in Engineering and Technology, vol. 5, no. 4, pp. 48-55, 2015.

[22] H. Jazayeriy et al. "A fast recommender system for the cold user using categorized items," Mathematical and Computational Applications, vol. 23, no. 1, p. 1, 2018.

[23] Rosenthal, S., et al. (2017). SemEval-2017 task 4: Sentiment analysis in Twitter. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017) (pp. 502-518).

[24] M. Ibrahim et al. "Design and application of a multivariant expert system using Apache Hadoop framework," Sustainability, vol. 10, no. 11, p. 4280, 2018

[25] Turney, P. D. 2000. Learning algorithms for keyphrase extraction. Information Retrieval 2(4):303-336.

[26] Zhao, W. X.; et al. 2011. Topical keyphrase extraction from Twitter. In ACL, 379-388.

[27] El-Kishky, A.; Song, Y.; Wang, C.; Voss, C. R.; and Han, J. 2014. Scalable topical phrase mining from text corpora. VLDB 8(3):305-316.

[28] H. Li, D. et al. "Disaster response aided by tweet classification with a domain adaptation approach," Journal of Contingencies and Crisis Management, vol. 26, no. 1, pp. 16-27, 2018.

[29] H.-T. Zheng, et al. "A deep temporal neural music recommendation model is utilizing music and user metadata," Applied Science, vol. 9, no. 4, p. 703, 2019.

[30] Java, A., et al.: Why we twitter: understanding microblogging usage and communities. In: Proc. of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis. pp. 56-65 2007.

[31] Joachims, T., et al.: Accurately interpreting clickthrough data as implicit feedback. In: Proc. of the 28th Int. ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR'05). pp. 154-161 2005.

[32] Kwak, H., et al.: What is Twitter, a social network or a news media? In: Proc. of the 19th Int. Conf. on World Wide Web (WWW'10). pp. 591-600 2010.

[33] Liben-Nowell, D., et al.: The link prediction problem for social networks. In: Proc. of the 12th Int. Conf. on information and knowledge management. pp. 556-559. CIKM '03, ACM, New York, NY, USA 2003.

[34] M. Gao et al. A movie recommender system from tweets data, January 2019, http://cs229.stanford.edu/proj2015/299_report.pdf.

[35] M. Ibrahim et al. "Design and application of a multivariant expert system using Apache Hadoop framework," Sustainability, vol. 10, no. 11, p. 4280, 2018.

[36] M.-Y. Hsieh, et al. "Building a mobile movie recommendation service by user rating and APP usage with linked data on Hadoop," Multimedia Tools and Applications, vol. 76, no. 3, pp

[37] Muhammad Ibrahim, et al. " A Neural Network-Inspired Approach for Improved and True Movie Recommendations," Hindawi, Computational Intelligence and Neuroscience, Volume 2019, pp-1-19.

[38] O. Loyola-Gonzalez et al. "Fusing pattern discovery and visual analytics approach in tweet propagation,"

Information Fusion, vol. 46, pp.
91-101, 2018.

[39] Phelan, O. et al..: Using Twitter to
recommend real-time topical news. In:
Proc. of the 3rd ACM Conf. on
Recommender Systems (RecSys'09).
pp. 385-388 2009.

[40] S. Kumar, et al. "Movie
recommendation system using
sentiment analysis from microblogging
data," 2018, https://arxiv.org/
abs/1811.10804.

[41] Rada Mihalcea et al. 2004. Textrank:
Bringing order into text. In EMNLP,

[42] Zhiyuan Liu, et al. 2011. Automatic
keyphrase extraction by bridging the
vocabulary gap. In Proceedings of the
Fifteenth Conference on Computational
Natural Language Learning. ACL,
135-144.

[43] Jingbo Shang, et al. 2018. Automated
phrase mining from massive text
corpora. IEEE Transactions on
Knowledge and Data Engineering 30, 10
(2018), 1825-1837.