International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 6, No.2 (Mar-2017)

http://dx.doi.org/10.12785/ijcds/060203

Mining Social Media Text: Extracting Knowledge from Facebook

Said A. Salloum [1, 2], Mostafa Al-Emran [3], and Khaled Shaalan [1]

[1] *Faculty of Engineering & IT, The British University in Dubai, UAE.*
[2] *University of Fujairah, Fujairah, UAE.*
[3] *Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang, Pahang, Malaysia.*

Received 1 Nov. 2016, Revised 7 Dec. 2016, Accepted 23 Jan. 2017, Published 1 Mar. 2017

Abstract: Social media websites allow users to communicate with each other through several tools like chats, discussion forums, comments etc. This results in learning and sharing of important information among the users. The nature of information on such social networking websites can be straight forward categorized as unstructured and fuzzy. In regular day-to-day discussions, spellings, grammar and sentence structure are usually neglected. This may prompt various sorts of ambiguities, for example, lexical, syntactic, and semantic, which makes it difficult to analyse and extract data patterns from such datasets. This study aims at analyzing textual data from Facebook and attempts to find interesting knowledge from such data and represent it in different forms. 33815 posts from 16 news channels pages over Facebook were extracted and analyzed. Different text mining techniques were applied on the collected data. Findings indicated that Fox news is the most news channel that share posts on Facebook, followed by CNN and ABC News respectively. Results revealed that the most frequent linked words are focused on the USA elections. Moreover, results revealed that most of the people are highly interested in sharing the news of Mohammed Ali Clay through all the news channels. Other implications and future perspectives are presented within the study.

Keywords: Text mining, Social Media, Facebook, News channels.

1. INTRODUCTION

Social networking websites can help eliminate coordination issues among individuals that are at a considerable physical distance [1]. It can build the viability of social campaigns [2] by spreading the required data at any place and at any time. In any case, in social networking websites, individuals for the most part use unstructured or semi-structured language for correspondence. People tend to pay less attention to spellings and precise linguistic development of a sentence in day-to-day discussions. This may prompt diverse sorts of ambiguities, for example, lexical, syntactic, and semantic [3]. Hence, extracting logical patterns with exact data from such unstructured types of data is a complicated job. Social networking websites, for example, Facebook are opulent in writings that lets it users make different text pieces in the shape of comments, posts on timeline, online networking, and blogs. Since social networks have been widely used in the recent past, a lot of information is accessible by means of the Web. Utilization of text mining techniques on social networking websites can uncover important results about communication practices between people [4].

A study by [5] stated that text mining techniques extract data from various pieces of writing. A key component is the connection of the extracted data together to shape new knowledge or present new facts to be investigated further by more mainstream methods of investigation. Text mining is considered different as compared to what we are acquainted with as web search. In web search, the user is commonly searching for something that is known and has been composed by another person. The issue is to put aside all the data that at present is not relevant to your particular needs, keeping in mind that the final goal to find the relevant data.

This study is categorized as follows: section 2 provides a comprehensive background about the text mining field. Other related studies are addressed by section 3. Section 4 addresses the research methodology. The findings of this study are demonstrated in section 5.

E-mail: Salloum78@live.com, malemran@buc.edu.om, Khaled.shaalan@buid.ac.ae

Conclusion and future perspectives are illustrated in section 6.

2. BACKGROUND

A. Text Mining vs. Data mining

As compared to data mining, there is more convolutions in text mining because of its unstructured and obscure content [10]. Text mining [5] contains features of data mining, but the distinctive point between these processes is that data mining tools [11] are designed to cope with structured data from databases, while text mining is able to handle unstructured or semi-structured data sets which include full-text documents, emails, and HTML files etc. Consequently, text mining has more promising results for its users. Hitherto, data mining (using structured data) has attracted the attention of researchers. The problem affiliated with text mining is evident (i.e. the incapability of computers to comprehend natural language). The main goal of text mining is to explore the unknown information that is concealed from masses. The handling and mining process of data is significant, as it is the raw form of information, which eventually ends up by creating knowledge [12].

B. Text mining in social networks

In the digital arena, unstructured data is dominating over structured data with its volume up to 80% as compared to its counterpart (structured data) with merely 20% [13]. Social networks are gaining massive popularity as mediums of spreading information and additionally facilitators of social interactions. The activity of users gives a substantial understanding into individual conduct, experiences, point of views and interests. There is a tremendous option for adding new personality based qualities to user interfaces (UI). Customized frameworks utilized as in spaces like e-learning, information filtering, collaboration and e-commerce could incredibly profit by a UI that adjusts the communication (e.g. motivational techniques, presentation styles, interaction modalities and suggestions) as indicated by the personality of the user. Getting hold of past user interactions is just a beginning stage in understanding the client conduct from a personality perspective [14].

C. Text mining efforts in resolving various NLP issues

Text mining [15] is a blossoming new field that endeavors to gather significant data from text in human language. It might be generally described as the way toward examining content to extract data that is valuable for specific purposes. In contrast to the sort of information stored in databases, the text is unstructured, amorphous, and hard to manage algorithmically. However, in advanced society, text is the most widely recognized means of formal sharing of data. The field of text mining mostly manages texts whose purpose is the correspondence of factual data or opinions, and the inspiration for attempting to extract data from such text is naturally convincing—regardless of the possibility that success may be incomplete. Text mining seems to grasp the entire automatic natural language processing. For instance, investigation of linkage structures, references in academic writing and hyperlinks in the Web writing are important sources of data that lie outside the conventional area of NLP. NLP is one of the hot topics that concerns about the interrelation among huge amount of unstructured text on social media beside the analysis and interpretation of human-being languages [16], [17]. However, in actuality, most text mining endeavours intentionally disregard the more profound, subjective, parts of classic natural language processing for shallower methods more likened to those utilized as a part of practical information recovery.

D. Text Mining Methods and Techniques

1) Association Rule Extraction:

Association rule extraction proposed by [18] that keeps significance in text mining field. It involves finding association relationship between different feature words from the text collection. Such exploration of interesting association relationships among huge amounts of transaction records will facilitate in many decision making processes [19].

2) Information visualization:

Visual text mining or information visualization [20] puts large textual sources in a visual hierarchy or map and makes it facile to use browsing capabilities with simple searching. SaS visual analytics depicts a tool that shows mappings of large amounts of text, making visual analysis possible for the user. The user is liberated to interact with the document map by using features such as zooming, scaling, and creating sub-maps. Information visualization is useful when a user needs to narrow down a broad range of documents and explore related topics. The advantage identifies the hierarchy of events. It assists in providing a map of possible relationships between doubtful events so that they can conclude their investigation without presuming critical aspects of the cases.

3. RELATED WORK

Many researches have been performed to review the several methods to extract information. A large portion of the researches focused on the utilization of various procedures of text mining for unstructured datasets, are in the text document format. However, they do not particularly focus on the datasets in social networking websites. The present study endeavours to fully understand various text mining techniques and implement them in the social media.

Scholars of [21] used around 30,000 tweets before the World Cup started. An algorithm that joined the DBSCAN algorithm and a consensus matrix were used. Therefore, the study has left with the concerned tweets on those dominant topics. Cluster analysis was then put to look for the topics that the tweets discuss. The tweets were bunched via utilizing k-means (a popular clustering algorithm), and Non-Negative Matrix Factorization (NMF) and compared the outcomes. Both algorithms provided similar results, yet NMF turned out to be faster and the results could be interpreted easily. However, other text mining techniques were not applied on the data which makes this study incomplete.

An Algorithm became a topic of consideration for [22] which elucidated the factors to follow the appointed website or Web page as per user's request by employing a text mining technique in such process. This study also explained that how acquisition and expression of text characteristic can be made possible, how to carry out grouping and classification process for the data information with feedback judgment in combination with the Web page text contents applied in latter part. The study explained the distinguished features of the algorithm by using various experiments, and proved it as to be more promising than the conventional one. The traditional text classification technique only contains training and categorizing two processes. Its classification ability is fixedly constant, with no space for continuous study. The algorithm channels by using following chronology: "Training Categorizing → feedback judgment → feedback". It makes the algorithm possess cognitive capabilities. 500 pieces have been obtained from sport documents from (http:// sports.163. com) as training and testing documents. Use this paper improvement algorithm of text classification requires only 6 seconds and the accuracy of the classification acquires up to 91%.

The analysis on semantic polarity proposes the method of web text mining which is accentuated in the form of a novel by [22]. In the first stage of the novel addresses, a research made on semantic polarity and involves three main parts assets of data, feature sentences analysis and semantic polarity analysis, which are introduced in the web text mining procedure. In the second stage, various algorithms are analyzed which are associated with semantic polarity analysis in the web text mining method. The third stage is implemented to evaluate knowledge about certain valuable products which are consumed. These stages are compared in contemplation to represent the sensibility and efficacy of the method. Contrarily, there are various improvements that can still to be made to the technique, like generating more important dictionaries using other enhanced techniques.

With the consideration on the facts regarding association rule mining technique, a method was introduced by [19] for the purpose of classifying text from pre-classified text documents. The facts which are extracted for ultimate grouping were then availed by Naïve Bayes classifier. The process of text classification comprises of three categories with 115 papers (47 from Computer Science, 48 from Electrical and Electronic Engineering, and 20 from Mechanical Engineering). These are intended for the guidance of data. Furthermore, each transaction is preceded by the help of association rule mining method which assist after pre-processing the text data. Based on the associated word sets which are used in the training data, a new document is classified. Therefore, a document is classified depending upon the number of word sets which are utilized in the training data, so a new document is categorized by increasing the range of word sets used in training document. According to [24], these 115 abstracts are inadequate to enhance the process of learning when collated with 20,000 efficient abstracts present in Naive Bayes example of text classification which yield 89% efficiency.

With the intention to assimilate both (qualitative analysis and large-scale data mining methods), [25] initiated a workflow. The chief concern was the Twitter posts of engineering students to get well aware of their issues and problems in their educational experiences. A qualitative analysis of samples attained from around 25,000 tweets revolved around the college life of engineering students was conducted as well. The authors discovered the encounter problems of engineering students. For instance, a heavy burden of study, lack of social engagement, and sleep deprivation. In light of these results, the study executed a multi-label classification algorithm to classify tweets representing students' queries. To train a detector of student problems from approximately 35,000 tweets streamed at the geo-location of Purdue University, the authors utilized the algorithm. For the first time, the problems and experiences of the students have been addressed and exposed informally with the help of social media data through the study. A multi-label classifier to organize tweets founded on the groupings established within the content evaluation phase was constructed. In data mining and machine learning domain, there are various renowned classifiers extensively consumed. Results indicated that the Naïve Bayes classifier proved to be efficient on dataset as compared to another state-of-the-art multi-label classifiers.

The viability of demonstrating user's personality built on a proposed set of features extracted from the Facebook data was actually explored by [14]. In light of the motivating outcomes of the study, a performance of various classification techniques and the feasibility is

fully analyzed. The research comprises of a sample of 250 user instances from Facebook (activity and demographic data) from about 10,000 status updates, which was delivered by the My Personality project [26]. There are two interconnected objectives of the study, which are following: (1) to be aware of the pertinent personality-correlated indicators that are present in Facebook user data explicitly or implicitly; and (2) to discover the viability regarding prognostic character demonstration so as to support future intelligent systems. The authors stressed out on the elevation of the relevance of the features in a model, and in light of the features generated from a variety of sources can lead to improved performance of the classifiers under evaluation.

Social media played a significant role in altering the traditional media frontiers [29]. We can observe from the existing literature that there are a lot of issues in Facebook were not yet analyzed and explored; one of them is the news channels' pages. As a result, we are interested to concentrate on this issue and try to build a new knowledge from the text analysis results from these channels. Based on the existing literature, we are seeking to answer the following research questions:

Q1: What is the most news channel that shares posts on Facebook?

Q2: What are the most frequent linked words that are posted on news channels' pages provided by Facebook?

Q3: What are the most frequent posts that are shared by people among the news channels' pages provided by Facebook?

4. RESEARCH METHODOLOGY

The datasets have been collected via Facepager software which is usually used to extract public existing data from Twitter, Facebook, and other social media based API. It collects URLs from query setup. Then, the extracted data will be stored in a local database and could be exported to a CSV Excel form. In this study, around 37551 posts were collected from the sixteen most popular news sites worldwide on Facebook.

A typical issue with posts content information is the existence of linguistic noise. For our situation, it would be superfluous posts that are inconsequential on prevailing topics. This view includes the objects of our data collection. According to the study by [30], missing and garbage data have been removed from dataset and the cleaned data has been uploaded into RapidMiner tool.

During importing the dataset into RapidMiner tool, the irrelevant attributes have been excluded for enhancing the performance and data quality. Missing attributes have been removed from the analysis which contains missing data in order to increase the precision. The final number of the cleaned records that have been used for the investigation was 33815. Our collected data includes some special characters and empty cells. As per the study of [27], we deal with these data by removing the special characters and empty cells through the use of pre-processing techniques.

The primary steps include separation of documents into tokens with each word representation, usually known as Tokenization [28]. The next step carries out the transformation process of all the characters, creating a document in a lower case. The third step includes the stop words filtering, in which this operator filters English, stop words from a document eradicating equivalent stop word from the built-in stop word list. It is required that every token represents a sole English word. An operator applying eliminating all tokens identical to a stop word from file provided for the process. The file is required to possess one stop word per line. The last step associated with text processing is the filter tokens by length. This operator filters tokens based on their length; we stipulate the nominal number of characters that a token is 4, and the maximal number of characters that a token is 25.

5. RESULTS

Different text mining techniques have been applied on the collected data. Each of which has its own results and implications. These techniques were used in the literature. However, none of these techniques were applied on news channels pages provided by Facebook.

Q1: What is most news channel that shares posts on Facebook?

As per the collected data, one of the attributes contains the shared count for each post that was shared on Facebook through these news channels. As per the (Figure 1), it has been noticed that Fox news is the most news channel that share posts on Facebook, followed by CNN and ABC News respectively. On the other hand, Africa24News, Associated Press Alaska channel news were shown that they have the lowest share count of posts. As per the study [20], we used the cloud technique. As per (Figure 2), the word cloud (i.e. shared counts) of the sources depicted that the Fox news channels on the top among all news channels that shared posts on Facebook.

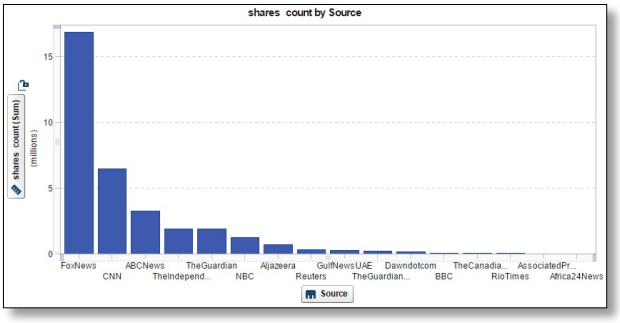

Figure 1. Share count by source.

Figure 2. Share count using word cloud by source.

Q2: What are the most frequent linked words that are posted on news channels' pages provided by Facebook?

In order to answer the above research questions, different techniques were applied on the collected data. First, we have applied the word frequency technique. As per (Figure 3), we can notice that the most frequent linked words among all the news channels is: "Trump" (i.e. represents Donald Trump, one of the candidates for the USA elections) followed by "President", "Clinton", "Republican", "Campaign", and "Nominee" respectively. These results indicate that the most frequent linked words are focused on the USA elections. Similarly to the results demonstrated by (**RQ1**), Fox News channel represents the most source that contains these words followed by CNN and ABC News respectively.

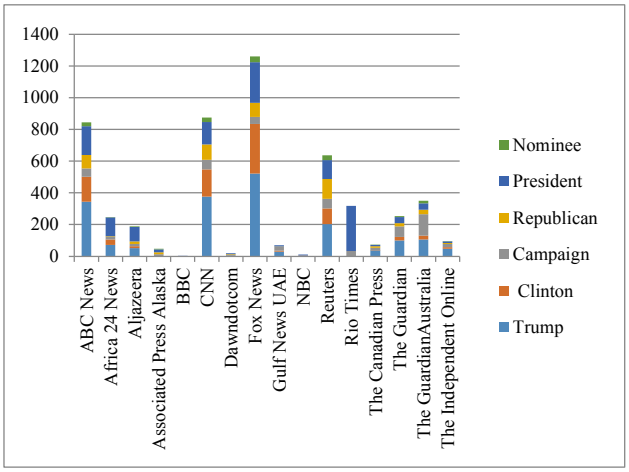

Figure 3. Word frequency distribution by sources.

For further investigation, the above most frequent words were analyzed and distributed among all the news channels in order to represent each word separately. As per (Figure 4), the word "Trump" was frequently mentioned by Fox News channel followed by CNN and ABC News respectively. According to (Figure 5), the word "President" was frequently used by Rio Times channel followed by Fox News and ABC News respectively. As per (Figure 6), the word "Clinton" was frequently utilized by Fox News channel followed by CNN and ABC News respectively. According to (Figure 7), the word "Republican" was frequently reported by Reuters channel followed by CNN and Fox News respectively. According to (Figure 8), the word "Campaign" was frequently reported by The Guardian Australia channel followed by The Guardian and (Reuters & CNN) respectively. Finally, as per (Figure 9), the word "Nominee" was frequently used by Fox News channel followed by Reuters and CNN respectively.

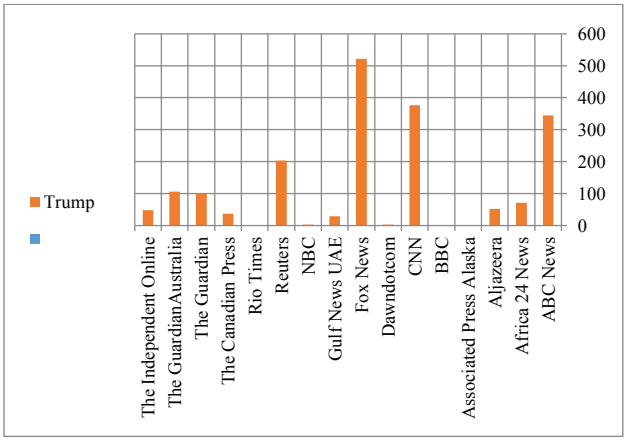

Figure 4. The distribution of the word "Trump" among all sources.

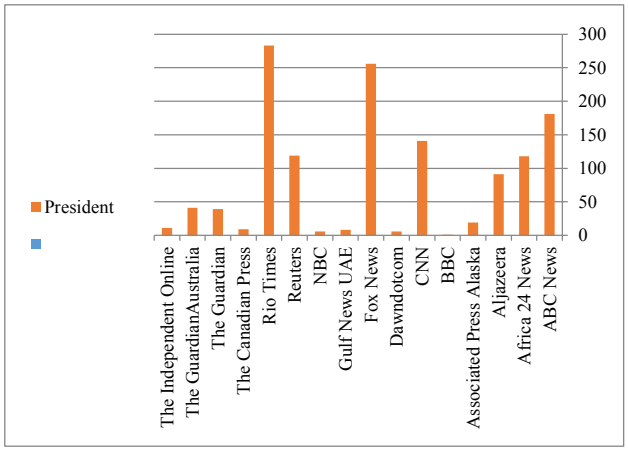

Figure 5. The distribution of the word "President" among all sources.

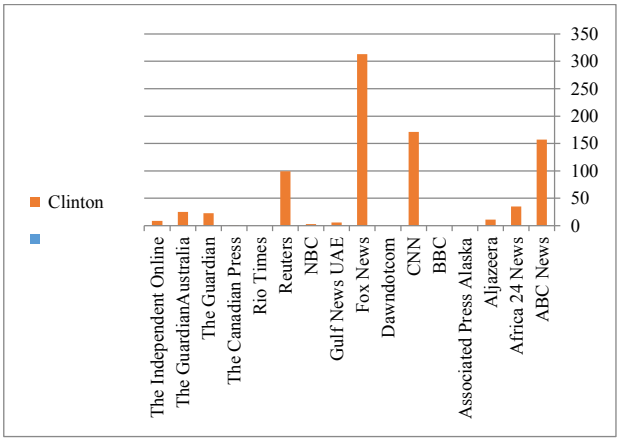

Figure 6. The distribution of the word "Clinton" among all sources.

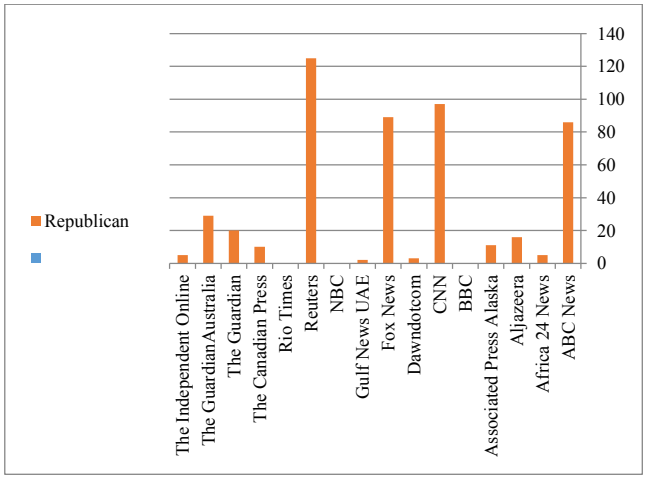

Figure 7. The distribution of the word "Republican" among all sources.

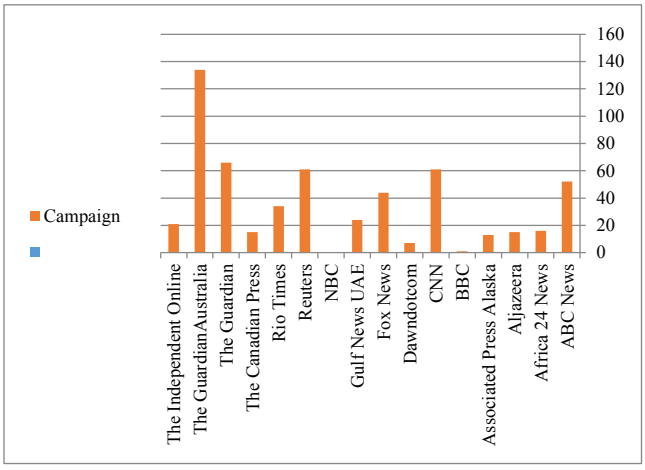

Figure 8. The distribution of the word "Campaign" among all sources.

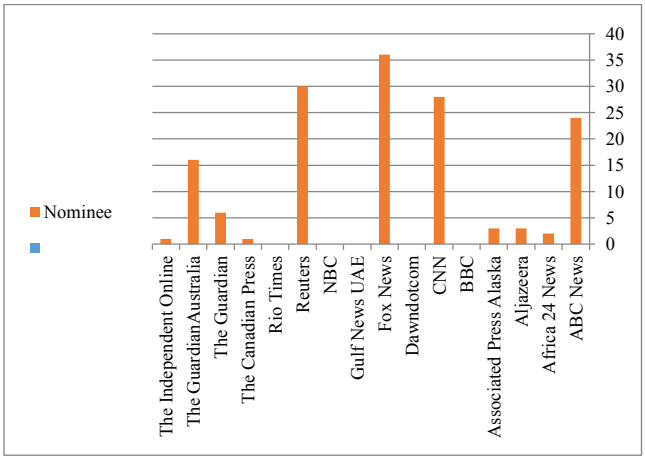

Figure 9. The distribution of the word "Nominee" among all sources.

Moreover, the concept link technique has been used to find and visualize the terms that are extremely connected with each other. The highly connected terms refer to the strongest relation among them. Figure 10 depict the term "Donald Trump" as the core of the tree structure along with all the relevant words that are linked to it. We can observe from the figure that most of the text posted by news channels is focused on the USA elections.

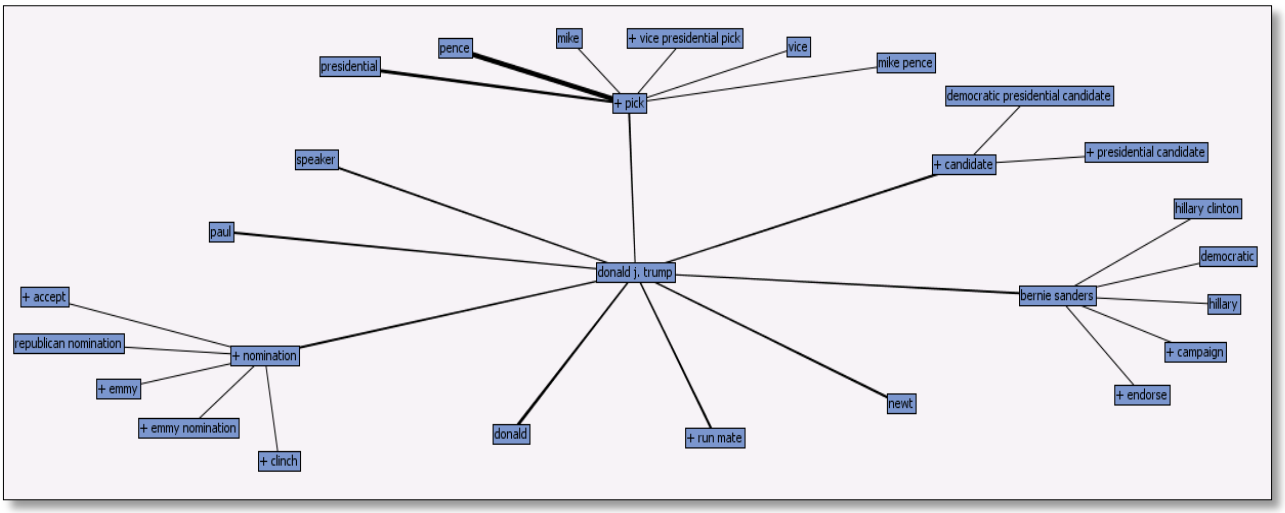

Figure 10. Concept link diagram.

Q3: What are the most frequent posts that are shared by people among the news channels' pages provided by Facebook?

As per the study of [19], we applied the association rule for the most frequently posts that were shared by people among all the news channels. Figure 11 revealed that there is a strong relationship between Mohammed (i.e. Mohammed Ali Clay, the American Olympic Boxer) and people. These results indirectly refer to the fact that most of the people are highly interested in sharing the news of Mohammed Ali Clay through all the news channels.

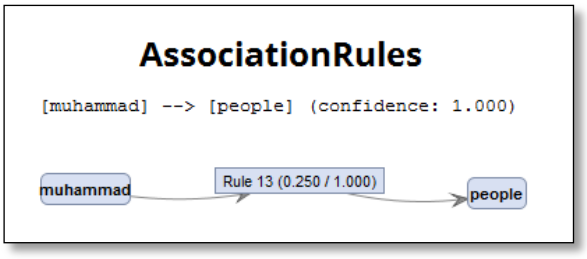

Figure 11. Association Rule diagram.

6. CONCLUSION

The evolving social media has radically transformed the communication method of people. This issue provides assistance to new agencies, companies affiliated with business arena. Specification of keywords and phrases may provide fruitful data to map out their future plans, easy decision making helps managing their business and communication becomes facile with current and potential customers. In this research, a comprehensive explanation about text mining and its research status is provided with reliable references. We found that news channels pages on social media were not got any attention from other scholars; the reason that makes us interested in investigating this particular issue. Through the use of Facepager tool, we collected 37551 posts from Facebook which were extracted from 16 international news channels. After pre-processing, the total number of posts were become 33815. Different text mining techniques have been applied on the collected data. Each of which has its own results and implications. These techniques were used in the literature. However, none of these techniques were applied on news channels pages provided by Facebook.

Findings indicated that Fox news is the most news channel that share posts on Facebook, followed by CNN and ABC News respectively. On the other hand, Africa24News, Associated Press Alaska channel news were shown that they have the lowest share count of posts. In accordance with the study [20], we used the cloud technique. Our results showed that the word cloud (i.e. shared counts) of the sources depicted that the Fox news channels is on the top among all news channels that shared posts on Facebook. Moreover, by applying the word frequency technique, we have noticed that the most frequent linked words among all the news channels is: "Trump" (i.e. represents Donald Trump, one of the candidates for the USA elections) followed by "President", "Clinton", "Republican", "Campaign", and "Nominee" respectively. These results indicate that the most frequent linked words are focused on the USA elections. Similarly, Fox News channel represents the higher source that contains these words followed by CNN and ABC News respectively. According to [19], we performed the association rule for the most frequently posts that were shared by people among all the news

channels. Our results revealed that there is a strong relationship between Mohammed (i.e. Mohammed Ali Clay, the American Olympic Boxer) and people. These results implicitly refer to the fact that most of the people are highly interested in sharing the news of Mohammed Ali Clay through all the news channels.

For future work, we are interested to collect and extract unstructured text from different news channels that provide news in Arabic language from Facebook and Twitter. Moreover, we are highly interested to do the same investigation on Twitter and do a comparison between Twitter textual data and Facebook. Furthermore, It is also possible to employ text mining tools in the form of intelligent agent that can obtain users personal profiles from social networks and pass on the appropriate information to users without the need to have an apparent request.

REFERENCES

[1] Evans, B. M., Kairam, S., & Pirolli, P. (2010). Do your friends make you smarter?: An analysis of social strategies in online information seeking. *Information Processing & Management*, *46*(6), 679-692.

[2] Li, J., & Khan, S. U. (2009, November). MobiSN: Semantics-based mobile ad hoc social network framework. In *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE* (pp. 1-6). IEEE.

[3] Sorensen, L. (2009, May). User managed trust in social networking-Comparing Facebook, MySpace and Linkedin. In *Wireless Communication, Vehicular Technology, Information Theory and Aerospace &Electronic Systems Technology, 2009. Wireless VITAE 2009. 1st International Conference on* (pp. 427-431). IEEE.

[4] Irfan, R., King, C. K., Grages, D., Ewen, S., Khan, S. U., Madani, S. A., ...& Tziritas, N. (2015). A survey on text mining in social networks. *The Knowledge Engineering Review*, *30*(02), 157-170.

[5] Berry Michael, W. (2004). Automatic Discovery of Similar Words. *Survey of Text Mining: Clustering, Classification and Retrieval", Springer Verlag, New York*, *200*, 24-43.

[6] Fan, W., Wallace, L., Rich, S., & Zhang, Z. (2005). Tapping into the power of text mining.

[7] SØRENSEN, H. T., Sabroe, S., & OLSEN, J. (1996). A framework for evaluation of secondary data sources for epidemiological research. *International journal of epidemiology*, *25*(2), 435-442.

[8] Zhang, J. Q., Craciun, G., & Shin, D. (2010). When does electronic word-of-mouth matter? A study of consumer product reviews. *Journal of Business Research*, *63*(12), 1336-1341.

[9] Sánchez, D., Martín-Bautista, M. J., Blanco, I., & de la Torre, C. J. (2008, December). Text knowledge mining: an alternative to text data mining. In*2008 IEEE International Conference on Data Mining Workshops* (pp. 664-672). IEEE.

[10] Akilan, A. (2015, February). Text mining: Challenges and future directions. In*Electronics and Communication Systems (ICECS), 2015 2nd International Conference on* (pp. 1679-1684). IEEE.

[11] Navathe, S. B., & Ramez, E. (2000). Data warehousing and data mining. *Fundamentals of Database Systems*, 841-872.

[12] Sukanya, M., & Biruntha, S. (2012, August). Techniques on text mining. In*Advanced Communication Control and Computing Technologies (ICACCCT), 2012 IEEE International Conference on* (pp. 269-271). IEEE.

[13] Chakraborty, G., & Krishna, M. (2014). Analysis of unstructured data: Applications of text analytics and sentiment mining. In *SAS global forum*(pp. 1288-2014).

[14] Markovikj, D., Gievska, S., Kosinski, M., & Stillwell, D. (2013, June). Mining facebook data for predictive personality modeling. In *Proceedings of the 7th international AAAI conference on Weblogs and Social Media (ICWSM 2013), Boston, MA, USA*.

[15] Witten, I. H. (2005). Text mining. *Practical handbook of Internet computing*, 14-1.

[16] Salloum, S. A., Al-Emran, M., & Shaalan, K. (2016). A Survey of Lexical Functional Grammar in the Arabic Context. Int. J. Com. Net. Tech, 4(3).

[17] Al Emran, M., & Shaalan, K. (2014, September). A Survey of Intelligent Language Tutoring Systems. In Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on (pp. 393-399). IEEE.

[18] Agrawal, R., Imieliński, T., & Swami, A. (1993, June). Mining association rules between sets of items in large databases. In *Acmsigmod record* (Vol. 22, No. 2, pp. 207-216). ACM.

[19] Rahman, C. M., Sohel, F. A., Naushad, P., & Kamruzzaman, S. M. (2010). Text classification using the concept of association rule of data mining. *arXiv preprint arXiv:1009.4582*.

[20] Qian, G., Sural, S., Gu, Y., &Pramanik, S. (2004, March). Similarity between Euclidean and cosine angle distance for nearest neighbor queries. In *Proceedings of the 2004 ACM symposium on Applied computing* (pp. 1232-1237). ACM.

[21] Godfrey, D., Johns, C., Meyer, C., Race, S., & Sadek, C. (2014). A case study in text mining: Interpreting twitter data from world cup tweets. *arXiv preprint arXiv:1408.5427*.

[22] Yin, S., Qiu, Y., & Ge, J. (2007, December). Research and realization of text mining algorithm on Web. In *Computational Intelligence and Security Workshops, 2007. CISW 2007. International Conference on* (pp. 413-416). IEEE.

[23] Yu, L., & Li, Q. (2009, September). A novel web text mining method based on semantic polarity analysis. In *2009 5th International Conference on Wireless Communications, Networking and Mobile Computing* (pp. 1-4). IEEE.

[24] Mitchell M. T., 1997. "Machine Learning", McGraw Hill, New York, 1997.

[25] Chen, X., Vorvoreanu, M., & Madhavan, K. (2014). Mining social media data for understanding students' learning experiences. *IEEE Transactions on Learning Technologies*, *7*(3), 246-259.

[26] Celli, F., Pianesi, F., Stillwell, D., & Kosinski, M. (2013, June). Workshop on computational personality recognition (shared task). In *Proceedings of the Workshop on Computational Personality Recognition*.

[27] Atia, S., & Shaalan, K. (2015, April). Increasing the Accuracy of Opinion Mining in Arabic. In *2015 First International Conference on Arabic Computational Linguistics (ACLing)* (pp. 106-113). IEEE.

[28] Verma, T., Renu, R., & Gaur, D. (2014). Tokenization and Filtering Process in Rapid Miner. *International Journal of Applied Information Systems*, *7*(2), 16-18.

[29] Mhamdi, C. (2016). Transgressing Media Boundaries: News Creation and Dissemination in a Globalized World. *Mediterranean Journal of Social Sciences*, *7*(5), 272.

[30] Zaza, S., & Al-Emran, M. (2015, October). Mining and Exploration of Credit Cards Data in UAE. In *2015 Fifth International Conference on e-Learning (econf)* (pp. 275-279). IEEE.

Said A. Salloum is currently a master student of Informatics (Knowledge and Data Management) at The British University in Dubai. He got his Bachelor degree in Computer Science from Yarmouk University. He is currently the Director of Computer Center at Al Fujairah University. Salloum is an Oracle expert since 2013 along with various recognized international certificates that are issued by Oracle.

Mostafa Al-Emran is a PhD student in Computer Science. He has graduated from The British University in Dubai with a distinction level along with the top Academic Excellence Award with MSc in Informatics (Knowledge and Data Management). He is currently the Head of Technical Support & Electronic Services Sections at Al Buraimi University College. Al-Emran got his Bachelor degree from Al Buraimi University College with the first honor degree in Computer Science. Currently, he is working on different research areas in Computer Science such as: M-Learning, Knowledge Management, Educational Technology and Data Analysis.

Khaled Shaalan is a full professor of Computer and Information Sciences at the British University in Dubai (BUiD). He is also a tenure professor at Cairo University. Prof Khaled is an Honorary Fellow at the School of Informatics, University of Edinburgh (UoE). He is currently the Head of PhD in Computer Science, MSc in Informatics, and MSc in IT Management programs. His main area of interest includes computational linguistics. He is an authority in the field of Arabic Natural Language Processing, and commands a great respect among the research community in the Arab world. He is the Head of Natural Language Research Group at BUiD. Prof Khaled has several research publications in his name in highly reputed journals such as IEEE Transactions on Knowledge and Data Engineering, Computational Linguistics, Journal of Natural Language Engineering, Journal of the American Society for Information Science and Technology, Expert Systems with Applications, Software-Practice & Experience, Journal of Information Science, and Computer Assisted Language Learning to name a few. He has guided several Doctoral and Master Students in the area of Arabic Natural Language Processing and Knowledge Management. He has done extensive research in the field of Arabic Named Entity Recognition and currently working on Arabic Question Answering.