A Lexicon-Based Approach to Build Reputation from Social Media

Haifa Al-Hussaini
Information Systems Department
King Saud University
Riyadh, Saudi Arabia
Haifa.alhussaini@imamu.edu.sa

Hmood Al-Dossari
Information Systems Department
King Saud University
Riyadh, Saudi Arabia
hzaldossari@ksu.edu.sa

***Abstract*— Nowadays, many social media platforms are widely used to express people's opinions about their daily experiences and interests. These platforms encourage people to exchange and share information about a particular brand, company or even a political point of view. Consequently, huge amount of data which can be extracted and analyzed to obtain some useful knowledge are available. In this paper, we propose to build a reputation of a given service provider (i.e. brand, product or service) from the collected social media data. To do so, we have developed a lexicon as a basic component for sentiment polarity in Arabic idioms. That is, the lexicon is used to classify words extracted from "Tweets" into either a positive or negative word. We use beta probability density functions to combine feedback from the lexicon to derive reputation scores. The experimental results show that our proposed approach is consistent with sentiment analysis approach results.**

***Keywords*— Lexicon, Reputation, Sentiment Analysis, Social Network.**

I. INTRODUCTION

Today social media platforms such as Facebook, Twitter, LinkedIn and MySpace are widely used. These platforms provide easy tracking and monitoring tools to share and exchange information about a particular brand, a company or a political point of view [20]. Such huge amount of data can be analyzed and exploited to extract some useful knowledge. Thus, there is a need for computation theories and tools to extract knowledge from the rapidly growing online contents in the social media [8].

Sentiment analysis or sometimes called opinion mining is a field of study that examines people's opinions towards entities such as products, organizations, and services [5]. The authors in [4] found that Twitter sentiment was linked with public opinion polls and Twitter sentiment was applied to predict election results.

A Reputation system gathers evidence about the attributes of individual objects, combines the evidence, and broadcasts the results. The objects can be products, users, or even digital content and the results are called reputation scores [26]. Identifying the reputation of social network entities can support developing better strategies for new product advertisements or election campaigns [17].

Although sentiment analysis approaches and reputation systems are correlated, it is important to highlight their similarities and differences. Basically, the reputation is a subjective notion that relies on what people think about something or someone. It is clear that the sentiment notion is a single opinion while the notion of reputation is an aggregate of opinions. That is, the reputation of a given product, service or company is based on a set of opinions which have accumulated over time. It is also important to emphasize that reputation is different from brand although they sometimes overlap between them. Generally, the two concepts are based on the historical information of a given entity (e.g. product, service or company).

In this paper, we developed a lexicon-based approach called ARST (Arabic Reputation System based on Twitter) to build reputation score from social media (i.e. Saudi dialect in Twitter). To do so, the lexicon is used to classify words extracted from "Tweets" into a positive or negative words. Then, we use beta probability density functions to combine feedback from the lexicon to derive reputation score. The experimental results show that our proposed approach is consistent with sentiment analysis approach results.

The rest of this paper is organized as follows. Section II is a literature review. Section III contains the framework of the proposed approach. Section IV displays detailed results of the chosen approach and outlines the future work. Section V is the conclusion of this paper.

II. LITRUTURE REVIEW

Most of the reputation models are based on service consumer ranking. However, one main obstacle for any reputation system is scare of ratings (i.e. there is no sufficient data to build the reputation). Several researches have been proposed to address this problem in the past. In [21], authors showed that they rely on using existing information in different contexts to be able to solve the problems of scarcity and cold starts for new user, service or product. They proposed graph-based representation model attempting to solve these problems by looking for other contextual data when there is no information for prediction. To provide a flexible sparsity, a framework is proposed to use similarity as credibility weight and an inference scheme will based on that weight [18]. In our study, we propose to handle the scarcity problem by using the data available in social media (e.g. Twitter, Facebook).

Another obstacle in building the reputation of a given service is that users have no motivation to provide their ratings, so they abstain to leave their rating after using the services [1].

IRJECE

To address this problem, an extensible reputation model which uses measurable QoS constrains based on objective and automatic feedback is proposed [11]. The proposed model uses a sociological theory that relies on two main characteristics: service consumer count and service invocation count.

Sociologically and in vice versa manner, higher reputations are obtained by the services with more consumers and service invocations. These are objective measurements that overcome the problem of subjectivity. However, invocation fraud that is based on fake invocation is still problematic. The solution is in using invocation per IP and invocation frequency. The invocation per IP is the counting of invocation from every single IP. Invocation frequency means observing the time between two invocations from the same IP. Limited invocation frequency indicates fraud invocation [11]. This model has traffic barriers and a prototype is made to measure the reputation measures among web services.

SocialTrust is a proposed reputation system that has three chief components. First, social networks which is a network of friends and partners. Consequently, each node preserves two lists of mutual-trusted nodes. Each node needs to keep its friendship with other by producing no harm to them and like to have more partners to gain more benefits. Second, a trivial reliable server selection is made to choose a client of the highest local ranking offered by other services. Third, reputation evaluation is based on the amount of credits that each node has collected through rating. Reputation evaluation is used to adjust reputation reward or punishment after a transaction [15]. This reputation system is implemented for peer to peer networks.

A reputation system based on consumer ratings and expectations to produce robust reputation calculation was made in [12]. The ratings were observed over time to define the changing behavior of service providers. It declared that this field of research is in its early stage, so extended studies are needed. Also, the proposed algorithm is not tested among real world data.

A probabilistic graphical reputation model is created to embody the relationship between social brands and users [17]. It collects the network information as well as the feedback of the users. This model reduces unfair outcome from a single user and a single comment. It jointly concludes the brand reputation. The implementation of this model is based on a parallel block-based Markov Chain Monte Carlo (MCMC) sampling method. The model is assessed by using a large amount of Facebook data. In the future, the study will focus on gathering and assisting data from other social platforms like Twitter, Google+, and LinkedIn to obtain reliable reputation assessment. However, this model is very complicated and impractical.

Another reputation model is based on mining textual feedback of services to build an expectations model for each consumer is proposed in [3]. This model is created for e-commerce settings and the research is based on real data gained from eBay. It is based on textual feedback not rating; however, it is not applied on social media.

In [9], authors described in details the tough task of tracking on Twitter. They proposed that Galois lattices solve efficiently the problems of this task. Galois lattices is used for conceptual classification. It groups exhaustively objects in classes as concepts, using their shared properties. It is based on a Boolean context matrix. The authors emphasized that it is not enough to gather the most frequently used words, the results must be displayed in an informative and readable way. They proposed a visual tool called Topigraphic Network of Tags to represent a tag cloud in a network of concepts. This tool is able to visualize the more found significant concepts about a given search on Twitter [9].

To easily monitor the reputation of a company in the Twitters, a strategy that arranges a set of tweets into diverse clusters based on the tweets topics is developed in [19]. The acquired clusters are prioritized into different priority ranks. A cluster with high priority represents a topic which may affect the reputation of a company, and deserves instant attention.

The testing results showed that this strategy is viable even though it does not use any external knowledge resource. The algorithm is evaluated against reliability and sensitivity. Reliability is the precision of binary relationships predicted by the system. Sensitivity is the similarly that is represented as a recall of relationships.

Online Reputation Monitoring Assistant (ORMA) is a semi-automatic tool [14]. It supports experts in their daily work of monitoring the reputation of entities companies or public concepts in Twitter. The tool automatically interprets tweets to see if it is about the specified entity. Then, reputational polarity is calculated to see if the tweet's inference is positive or negative as regards the specified entity reputation. The tool displays groups of relevance in decreasing order of relevance from a reputational perception. ORMA was tested by thirteen experts using half a million tweets for 61 various entities and proved efficiency. The input data in this model were tweets about entities in both English and Spanish.

Arabic language is a Semitic language which causes problems for natural language processing. It has numerous challenges such as: diacritics which are rarely used in writing. In Arabic two words which look alike may be indistinguishable for machines without the use of diacritics. As a result, the efficiency of word-based syntactic features is reduced. In addition, Arabic language has high inflectional and derivational regional languages. As such there are problems in stemming Arabic text as well as in lexical extraction. Moreover, Arabic nouns do not begin with capital letter as in English. This causes a problem in extracting proper nouns from Arabic texts [23].

Another problem is that a given root can take several forms depending on the context such as (أحبُ,يحب , يحبون, أحبو). Also, negation words that negate past or present tense verbs. They change the meaning of the verb to exactly the opposite

Users' opinions on social networks are critical because they have a significant influence toward marketing, decision making, purchasing decisions, political elections, travel plans, school applications, and so on [17]. In addition, many people in social media use the dialect of their country instead of using MSA (Modern Standard Arabic) [24].

REFERE

Presenting a feature-based sentence level approach is one way for Arabic sentiment analysis. Authors in [13] described in detail how they use Arabic lexicon and Egyptian dialect to improve the detection of the sentiment polarity in Arabic sentences. In our study, we introduce a Saudi-Lexicon to be used to build a reputation score for a given entity.

III. A LEXICON-BASED APPROACH TO BUILD REPUTATION

The main goal of this research is to build a reputation of an entity based on Saudi tweets in Twitter. The aim is to use the collected opinions of Twitter users in calculating the reputation of a given entity such as iPhone 6 or Macdonald's. The proposed system consists of four main components: Data collection, Data preprocessing, Lexicon development and reputation calculation.

A. Data Collection

The collected data in our system are tweets. They are collected automatically using Apache Nutch tool. Apache Nutch is a complete open source Web crawler written in Java, depending on Apache Hadoop data structures. It is used in distributed architecture and it can grab web page hyperlinks automatically. In our proposed system, Nutch was modified to crawl only Arabic tweets. Apache Nutch is so highly modular that most of its functions can be changed via plugins.

B. Data Preprocessing

In ARST, Apache Solr is used for preprocessing. Apache Solr is a fast open-source Java search server. Solr allows the easy creation of search engines including searching websites, databases and files [16]. The basic component of Solr that makes it able to perform fast search responses, is the index. It creates and uses to search text. Index is similar to retrieving pages in a book based on a keyword by scanning the index at the back of a book. This is called an inverted index. In Solr, a document is the element of search and index. An index contains one or more documents, and a document contains one or more Fields. In database terms, a Document is a table row, and a Field is a table column [16].

When working on Solr, the first thing is to specify the schema: the kinds of fields, which field is the unique key, which fields are obligatory, and how to index and search each field. Solr expands the diversity of field types available in Apache Lucene [16].When data is added to Solr, it passes through a series of transformations before being added to the index. This is considered as "analysis phase". Examples of these transformations are lower-casing, removing word stems, removing duplicates, removing stop-words, splitting on whitespace, splitting words into sub-words that help to match delimiters with words. By using an external synonyms file that we edit manually from an Arabic dictionary, we can match strings of tokens, replaces them with other tokens and get them from Solr. This work is done within query time. The end result of this phase are a series of tokens which are then added to the index. The same series of transformations happen to the query string during the search process to perform matching [16].

Solr has three main concepts that are used to process the text. These concepts are known as analyzers, tokenizers, and filters. Each field has analyzer and tokenizer. Field analyzers are used when a document is "indexed" within query time. It examines the field text and generates a "token stream".

Analyzers may be a single class or they may be composed of a series of tokenizer and filter classes. Tokenizers breakdown the field data into lexical units, or tokens. Filters inspect a stream of tokens and keep them, alter or remove them, or create new ones. Tokenizers and filters may be combined to form chains. Such a sequence of tokenizers and filters is called an analyzer [7].

For indexing Arabic content, Solr offers support for the Light-10 (PDF) stemming algorithm, and Lucene includes an example of a stop word list. This algorithm describes both character normalization and stemming. It is divided into two filters to provide flexibility as follows: - solr.ArabicStemFilterFactory and solr.ArabicNormalizationFilterFactory [25].

C. Lexicon Development

The problem of extracting sentiments from text is a very complex task, in particular due to the significant amount of Natural Language Processing (NLP) required. This task becomes even more difficult when dealing with morphologically rich languages such as Modern Standard Arabic (MSA) and when processing brief, noisy texts such as "tweets" or "Facebook statuses". People in social media use the dialects of their country instead of using MSA [24].

In this study, we have proposed to develop a lexicon for Saudi dialects. The proposed Lexicon is made out of Saudi tweets as it is specialized in the Saudi dialect. The lexicon is made by collecting 1000 tweets written in Saudi dialect using NodeXL. After collecting the sentiment words in each tweet, which reached 1400 word, the indication of it whether positive or negative is identified. Fig. 1 below shows part of the collected data and the sentiment of the words in each tweet.



Fig 1: Sentiment Analysis of the Collected Data from Tweets (Positive Sentiment)

In this research, a Dictionary-based approach is used to extract the polarity of each sentence (in our work: tweets). After that, the logic of the opinion words is analyzed and the sentiment of the text is classified (positive or negative).

In general, the techniques that follow Dictionary-based approach are based on lexicons. They map the words to their semantic value by using the dictionary. In this study, we allot 90% of the data that we collected for developing the lexicon and used 10% of the data for testing. The importance of having

IRJECE

such a lexicon in our system, motivate us to create the first Saudi dialect lexicon. Below are the steps followed to build the lexicon:

- Collecting 1000 tweets written in Saudi dialect from Twitter from various fields such as restaurants, companies etc.
- Extracting the Saudi sentiment words, idioms.
- Assigning each word with a sentiment score; positive (+1) or negative (-1).
- Removing duplicate words.
- Ignoring question tweets, retweet and spams.
- Adding idioms or a sentence that express positive or negative opinion.

Different ways of writing entities are anticipated to be used as synonymous so as to gather the tweets associated with a specific entity efficiently. The names of products, organizations, companies, and other are gathered to create an Entity Lexicon as shown in Fig. 2.

	اسم المنتج الرئيسي	المرادفات	المرادفات	المرادفات
2	stc	الاتصالات	الاتصالات السعودية	اس تي سي
3	ماكدونالدز	مالك		
4	كنتاكي	كي اف سي	kfc	
5	شانيل	شنيل		

Fig 2: Entity Lexicon

D. Reputation Calculation

In this research, the reputation function is based on the Beta Probability Density Function. It is used to show the probability distributions of binary events. In this research the binary events are positive or negative sentiment. It is a mathematical base that is used to represent the reputation ratings based on feedbacks, which are tweets in this research. It is based on posteriori probabilities of binary events to convey beta distributions [1]. It is a continuous function which uses two parameters which are α and β.

In our proposed approach, we view extracted words from the collected tweets as a set of Bernoulli trials: positive word (with the entity under hand) or negative word (against the entity under hand), and then model it as Beta distributions:

$$R = \frac{\alpha}{\alpha + \beta} \qquad (1)$$

Where $\alpha = r + 1$ and $\beta = s + 1$, and r is the observed number of positive words and s the negative ones. The ratio of α and β will determine where in the interval [0,1] the reputation score peaks, and a high α will cause the score to occur close to 1.

The idea of adding 1 each to r and s (and thus 2 to r + s) follows Laplace's rule of succession for applying probability to inductive reasoning [10]. This rule in essence reflects the assumption of an equi-probable prior, which is commonly adopted in probabilistic reasoning. That is, when we have no information about an entity initially, (i.e. r=0 and s=0), we have R= 1/(1+1) =0.5, suggesting that a positive and negative word about the entity is equally likely. It is noticed in this study rating nearer to "1" donates good reputation score and the vice versa. Algorithm 1 describes how to calculate reputation score in detail.

Algorithm 1: Pseudo-Code for a lexicon-based

Input: W, L
Output: Reputation score (R)
Process:

1. $\alpha = 0$;
2. $\beta = 0$;
3. For each wi do
4. IF wi $\in \Omega$
5. $\alpha = \alpha + 1$;
6. Else
7. IF wi $\in \pi$
8. $\beta = \beta + 1$;
9. Else
10. ¥ ← wi;
11. End IF
12. End IF
13. End For
14. $R = \alpha / (\alpha + \beta)$;
15. Return R;

In this research, we use a lexicon-based approach to compute reputation score of a given entity. The developed lexicon L:=(Ω, π, λ, ¥) where Ω is a set of positive words in Saudi dialect, π is a set of negative words, λ is a set of natural words (i.e. neither positive nor negative), and ¥ is a set of non-known words that need to be checked by the lexicon moderator to be added later on to one of the three groups.

The input for the proposed algorithm is represented by W and L. W is a set of words about a given entity that has been extracted from Twitter and L is the developed lexicon. The algorithm uses two variables: α and β, the first for positive ratings (Line1) and the second is for calculating the number of negative ratings (Line2). After that, the algorithm starts to scan the words in the database (from Line 3 to Line 13). For each word W_i, it checks If the word belongs to the positive words of lexicon Ω; then α will increased by 1 (line 5). If it is not, the second condition will be checked and if the word belongs to the negative words π then β will be increased by 1 (Line 8). If the word is not considered as positive or negative words, it will then added to the non-known list ¥ (line 10). The lines from 4 to 12 are repeated for each word from the collected tweets.

Once the algorithm has scanned all the collected words, the reputation score will be calculated by dividing the total number of positive ratings (α) by the sum of positive and negative ratings (Line 14) and the result will be returned to the requester (Line 15).

IV. RESULTS AND DISCUSSION

To evaluate our proposed approach, we collected tweets from Twitter about 5 different restaurants. On average, 62 tweets were collected about each restaurant. For validity

IRJECE

purposes, the proposed approach has been compared to sentiment analysis approach. In the sentiment analysis approach, each collected tweet was classified manually as positive, negative or natural. The result is shown in Figure 4.

Here we introduced a specific example to explain how to calculate reputation score using our proposed approach (word-level) and sentiment analysis approach (sentence-level). In this example: "مطعم لوسين يجنن وأفضل مطعم" means "Lusin restaurant is amazing and the best restaurant", to calculate the reputation score we do two sets of calculations for each restaurant:

- Sentence-level: labeling sentiment of the sentence as a whole. Depending on the previous example, we give +1 for this tweet.
- Word level: set each word that has positive or negative sentiment. By referring to our example, we give +1 for "amazing" and +1 for "best". So the total will be +2 for this tweet. We can find these words in the positive words of the lexicon.



Fig 4: Reputation Score Of Lexicon-Based Vs. Sentiment Analysis

Figure 4 shows how results of our proposed approach consists with sentiment analysis approach. That is, the calculated reputation scores for the five restaurants are so closed to those scores calculated by the sentiment analysis approach. For example, restaurant 1 has the highest scores in both approaches (0.962 using our proposed approach and 0.964 using the other approach). More important, the order of the five restaurants are equivalent by the two approaches. We also investigated how FourSquare Application rates these restaurants and found that the restaurant 1 is the best one which supported the results of our proposed approach.

The precision of the lexicon in our system is 100% since it detect all the words that have sentiments in each tweet. To calculate recall of our proposed lexicon, we test about 300 tweets and extract 225 distinct words that have sentiment (both positive and negative). Our lexicon detects 120 words from the total which means that the recall of our lexicon is 53%.

It was approved that users tend to repeat the letter more than once to emphasize the meaning or feeling which is one of the challenges in social media [24]. One of the reasons that justify lower percentage of the recall is that Saudi users often duplicate letters to express overstatement of their opinion. For example: "لذيييذ" is like "لذيذ" which means "delicious" with more weights of pronunciation. Furthermore , the language used in social media is highly unstructured and full of slang words, misspellings, contractions and abbreviations, such as "lol" (Laughing Out Loud) or "soooo cooool" [2],[22].

Dealing with the hash-tag "#" is another problem. It is used to spot, organize and filter tweets according to topics or categories. People use this symbol before related keywords in their tweets to help classify and find them more easily in Twitter Search. Emoticons is like ":-)" which means a humorous comment and interpreting them is a challenge task.

Saudi tweets are full of spelling mistakes and the structured build is loose. The words are used with no regard to grammar or spelling. Take for example "أبغى" meaning "I want = أريد" pronounced in a wrong way "أبغا" and written as such. The language used has nothing to do with MSA (Modern Standard Arabic). An example of this is " يرضيهم شركه الاتصالات اني اطارد وراهم في الدرج عشان الشبكه". The tweet-writer just wonders that STC Company is pleased at his running after them on the stairs for the net service. However, some Saudi tweet-writers shape the words according to mood. Some use symbols, pictures and numbers to stand for words while others use proverbs and verses from poetry. A reverse round about misleading way of writing is gaining popularity among some tweet-writers.

V. CONCLUSION

In this paper, we proposed a lexicon-based approach to calculate reputation score from Twitter. We build a Saudi dialect lexicon from Saudi tweets to improve addressing the sentiment of the Arabic tweets. The achieved results indicate that our ARST system is very promising. To this end, we collected about 1300 tweets (for developing and testing the proposed the lexicon). The experimental results show that our proposed approach is consistent with sentiment analysis approach results.

To improve our proposed approach, as a future work we will cover question tweets besides handling negation in the sentimental analysis model. In our lexicon "أحب" is a positive word but negation changes it to a negative one. Also, we intend to focus on the usage of intensifiers (quantitative) words, repeating letters of words to produce certain effects. Finally, we plan to make the proposed lexicon available online for the community.

REFERENCES

[1] A. Jøsang and R. Ismail, "The Beta Reputation System," in In Proceedings of the 15th Bled Electronic Commerce Conference, Bled, Slovenia, 2002.

[2] A. Joshi , A. R. Balamurali , P. Bhattacharyya and R. Mohanty, "C-Feel-It: a sentiment analyzer for micro-blogs," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Systems Demonstrations, Stroudsburg, 2011.

[3] A. M. ElMessiry, X. Gao and M. P. Singh, "Incorporating Expectations as a Basis for Business Service Selection," in Service-Oriented Computing, Springer Berlin Heidelberg, 2010, pp. 486-500.

IRJECE

[4] A. Tumasjan, T. O. Sprenger, P. G. Sandner and I. M. Welpe, "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment," in the Fourth International AAAI Conference on Weblogs and Social Media, 2010.
[5] B. Liu, Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies, Chicago: Morgan & Claypool, 2012.
[6] B. O'Connor, R. Balasubramanyan, B. R. Routledge and N. A. Smith, "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series," in the Fourth International AAAI Conference on Weblogs and Social Media, 2010.
[7] C. Targett, "Understanding Analyzers, Tokenizers, and Filters," 2014. [Online]. Available: https://cwiki.apache.org/confluence/display/solr/Understanding+Analyzers%2C+Tokenizers%2C+and+Filters.
[8] D. K. Simonton, "Creative thought as blind-variation and selective-retention: Combinatorial models of exceptional creativity," Physics of Life Reviews, vol. 7, no. 2, p. 156–179, June 2010.
[9] E. Cuvelier and M.-A. Aufaure, "A Buzz and E-Reputation Monitoring Tool for Twitter Based on Galois Lattices," in Conceptual Structures for Discovering Knowledge, Springer Berlin Heidelberg, 2011, pp. 91-103.
[10] E.S. Ristad. A natural law of succession. Arxiv preprint cmp-lg/9508012, 1995
[11] G. Yang and H.-p. Chen, "An Extensible Computing Model for Reputation Evaluation Based on Objective and Automatic Feedbacks," in Advanced Language Processing and Web Information Technology, 2008. ALPIT '08. International Conference on, Dalian Liaoning, 2008.
[12] H. Al-Dossari, "A reputation-based approach for consumer protection in Saudi Arabia," in Computational Aspects of Social Networks (CASoN), 2012 Fourth International Conference on, Sao Carlos, 2012.
[13] H. S. Ibrahim, S. M. Abdou and M. Gheith, "SENTIMENT ANALYSIS FOR MODERN STANDARD ARABIC AND COLLOQUIAL," International Journal on Natural Language Computing (IJNLC), vol. 4, April 2015.
[14] J. Carrillo-de-Albornoz, E. Amigó, D. Spina and J. Gonzalo, "ORMA: A Semi-automatic Tool for Online Reputation Monitoring in Twitter," in Advances in Information Retrieval, Springer International Publishing, 2014, pp. 742-745.
[15] K. Chen , H. Shen, K. Sapra and G. Liu, "A Social Network Integrated Reputation System for Cooperative P2P File Sharing," in Computer Communications and Networks (ICCCN), 2013 22nd International Conference on, Nassau, 2013.
[16] K. Tan, "Basic Solr Concepts," 2015. [Online]. Available: http://www.solrtutorial.com/basic-solr-concepts.html.
[17] K. Zhang, D. Downey, . Z. Chen , Y. Xie , Y. Cheng , A. Agrawal , . W.-k. Liao and A. Choudhary , "A probabilistic graphical model for brand reputation assessment in social networks," in IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, New York, 2013.
[18] L. Xiong, L. Liu and M. Ahamad , "Countering Feedback Sparsity and Manipulation in Reputation Systems," in Collaborative Computing: Networking, Applications and Worksharing, 2007. CollaborateCom 2007. International Conference on, New York, 2007.
[19] M. A. Qureshi, C. O'Riordan and G. Pasi, "Concept Term Expansion Approach for Monitoring Reputation of Companies on Twitter," in CLEF (Online Working Notes/Labs/Workshop) , Rome, 2012.
[20] M. Matešić , K. Vučković and Z. Dovedan, "Should academia care about online reputation management and monitoring?," in MIPRO, 2010 Proceedings of the 33rd International Convention, Opatija, Croatia, 2010.
[21] M. Tavakolifard and K. C. Almeroth, "Social Computing: An Intersection of Recommender Systems, Trust/ReputationSystems, and Social Networks," IEEE, 2012.
[22] P. H. Calais Guerra, A. Veloso, W. M. Meira and V. Almeida , "From bias to opinion: a transfer-learning approach to real-time sentiment analysis," in the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, 2011.
[23] R. Al-Eid , R. Al-Khalif and A. S. Al-Salman , "Measuring the credibility of Arabic text content in Twitter," in Digital Information Management (ICDIM), 2010 Fifth International Conference on, Thunder Bay, 2010.
[24] R. M. Duwairi, R. Marji, N. Sha'ban and S. Rushaidat, "Sentiment Analysis in Arabic tweets," in Information and Communication Systems (ICICS), 2014 5th International Conference on, Irbid , 2014.
[25] Solr wIKI, "Language Analysis," 2014. [Online]. Available: https://wiki.apache.org/solr/LanguageAnalysis#Arabic.
[26] Y. Sun and Y. Liu, "Security of online reputation systems: The evolution of attacks and defenses," IEEE Signal Processing Magazine, vol. 29, no. 2, pp. 87 - 97, 2012.

IRJECE