

A REVIEW ON PERFORMANCE EVALUATION OF A KEYWORD SEARCH IN SOCIAL MEDIA.

PoojaHumbe*

Prof. M. M Ambekar**

Abstract

Data mining is the process of discovering patterns in data set according to the keyword. Keyword search is the most effective information discovery method in Databases. But somewhere, sometimes only searching the keyword is not enough, with searching restricting that keyword is became a necessity. Like in social media abuse of word is increasing. Previous systems have worked on only searching a keyword as per its application. But if the keyword is not appropriate in some applications then we have to restrict that word. So we are using keyword search for social media application in which it not just search the inappropriate words but also restrict that word from publishing and also give the class of inappropriate word used by person.

Keywords: Data mining; Keyword; Keyword search; Social media;Database.

* **P.G Student Department of Computer Science ,P.E.S. College of Engineering,Aurangabad.**

** **Associate Professor,Department Of Computer Science, P.E.S College of Engineering, Aurangabad.**

1. Introduction

The use of internet is increasing day by day there are so many people who use search engine daily and performs about 4 billion searches. As the searching information is increasing day by day the demand for the keyword search is also increasing rapidly. There are many existing relational keyword search systems. Keyword search systems should return whatever answers they can produce fast [1]. Data mining is nothing but extraction of data from a large data set. At the time of searching user inserts a keyword and gets a result related to that keyword after searching. [2]Keyword search changes for different applications and retrieval systems are different for that purposes.Applications change as per its use, also vary as per requirement [2].We are using the system in social media. As we know that these days Social Networking is becoming more and more popular. It is a platform where people can share their ideas and thoughts [3].Social media is used for communication purpose. But sometime message post by the users can contain some kinds of abusive or offensive contents [4]. As we know that as popularity of the websites increasing, their vulnerability to be attacked is also increasing[3].

So that our main aim is to restrict the inappropriate word used against any person or religion or community etc. For that purpose we are using inappropriate words database in which system not only searches the keyword but also restrict that word from publishing.Here each post is checked at run time not after publishing the message. Here keyword searching algorithm is used to search inappropriate content, keyword restriction algorithm is used to prevent that content from publishing and classification is used to classify the abusive word according to the category.

2. Literature Survey

2.1 BANKS: Browsing and Keyword Searching in Relational Databases

B. Aditya et.al [5] has proposed “*BANKS: Browsing and Keyword Searching in Relational Databases*”, in which system enables keyword based searched on relational databases. BANKS enables user to exact information without any knowledge of schema. Here user is able to get information by typing a few keywords then following the given hyperlinks and getting results.

2.2 SPARK: Top-k Keyword Query in Relational Databases”

Yi Luo et al [6] has proposed, “*SPARK: Top-k Keyword Query in Relational Databases*” in which they focus on effectiveness and efficiency of keyword query search. They have proposed a

ranking formula using previous information retrieval techniques. Main importance of this technique is it works on large scale real databases example of that is Customer Relationship Management.

2.3 Relational Keyword Search System

Pradeep M. Ghige and Ruhi R. Kabra [2] has proposed, “*Relational Keyword Search System*” in which they define the category of the search word first then after user will select the appropriate word meaning that is going to search. They statically add the database in their project. It also shows the keyword ranking and not requires the knowledge of database queries.

2.4 The How, When and Why of Sentiment Analysis.

Mrs. Vijyalaxmi M et.al [7] has proposed, “*The How, When and Why of Sentiment Analysis*” in which they applied Sentiment analysis wide domain to classifying and summarizing review and prediction.

2.5 Effectiveness of Social Media as a tool of communication and its potential for technology enabled connections: A micro-level study

Trisha Dowerah Baruah [8] has proposed a, “*Effectiveness of Social Media as a tool of communication and its potential for technology enabled connections: A micro-level study*” in which she has done analysis on the role, importance and the impact of social media as a tool of effective communication. To examine the effectiveness of social media, she has used the survey method for investigation in order to find out the growing importance.

3. Proposed Methodology

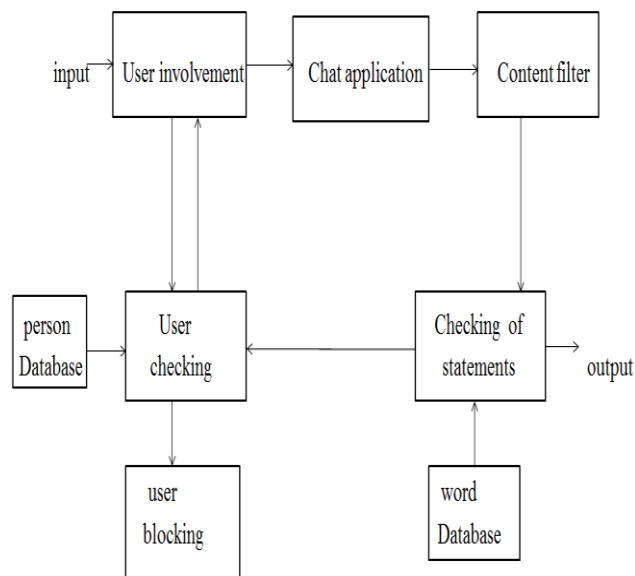
In this paper we are going to propose a system to perform an evaluation of a keyword search. We are using this system for social media application. That means system is based on inappropriate languages, words used while chatting on the net in social media. Here we are using two databases – one is database for the user and the second is database of inappropriate words.

User or person database contains all the information about the person like his name, mail id, password, DOB, gender, city etc. We are using person database to see how many times one particular person uses the inappropriate content. Here three chances will be given to the user for

his mischief with the warning and fourth time his account will be blocked permanently. His login credentials will be saved and he or she will not be allowed to be a member of chat group. Inappropriate words will be stored in database for checking each post before publishing.

When the user types some matter to be posted on chat application with or without inappropriate words to any person or religion or community. Then the post will be checked for inappropriate words if such word not found, the message is allowed to be posted. But if it contains some abusive words then that word or sss content will be restricted and also checks the frequency of that user/person to use abusive content and his or her chance to do mistake will be decrease by one.

3.1 Block Diagram



3.2 Modules

1. User involvement
2. User checking
3. Chat application
4. Content filtering
5. Checking of statement
6. User blocking

1. User involvement :-

In User involvement login of authorized person is done by using user name and password received after properly enrolling into the chat room.

2. User checking module:-

Before entering into chat application, user checking module checks the database of user. If the default variable count for that particular person is three then user is not allowed for chat, and for new user or the users who has a count less than three is allowed for chat.

3. Chat application:-

To use the chat application first user has to login with their user name and password. By checking the authentication and the database of that particular user, chat application will be open for them in which user is able to send data, receive data.

4. Content filtering :-

In content filter, it first takes input statement written by the user on chat room. Then removes the unwanted data like space (' '), comma(','), full stop('.') etc and send the string of word to the checking of statement module.

5. Checking of statement:-

In this module words are checked with database of inappropriate words if there is any inappropriate word found then fault count for that person will be incremented by one and chance is allotted for that person to use the application.

This chance will be incrementing his or her fault count by one and his or her chance to do mistake will be decrease by one.

6. User blocking:-

In this module person is blocked when the frequency of using inappropriate word by that person become three.

3.3 Algorithm

Step 1: user login

Step 2: before entering into chat application, frequency of using inappropriate word by that person is checked. If it is equals to 3 then don't allow that person for chat. And if count is less than 3 then

Step 3: allocating chat environment to the authorized user.

Step 4: User input into chat application.

Step 5: on clicking post button statement will be transfer to the filter.

Step 6: Filter will remove the unwanted data.

Step 7: transfer the output of filter to the statement checking module.

Step 8: in statement checking module if content doesn't matched with inappropriate word database, statement is allowed to move out. And if content matched then

a) Restrict that word.

b) User data is checked from user/person database.

c) For that user, chance to do the mistake will be decrease by one and allocates the remaining chances.

d) Classify the word according to the given category.

4. Conclusion

In existing systems as we have seen all systems have worked on searching keyword from databases but in addition we can see restriction of that keyword in this paper. Hence we have seen that how the inappropriate word is restricted from publishing on net in social media.

References

- [1] Kaveri A. Dighe and M. M. Naoghare, "Evaluating Performance of Keyword Search Systems", *International Journal of Science and Research (IJSR)*, Volume 4, pp. 741-744, October 2015.
- [2] Pradeep M. Ghige and Prof. Ruhi R. Kabra, "Relational Keyword Search System", *International Journal of Engineering Research and General Science* Volume 2, pp. 524-532, October-November, 2014.
- [3] Sandeep Kumar Rawat and Assistant Prof. Saurabh Sharma, "A Review on Spam Classification of Twitter Data Using Text Mining and Content Filtering" *International Journal of Advanced Research in Computer Science and Software Engineering*, Volume 5, pp. 485-488, June 2015

- [4] Krishna B. Kansara and Narendra M. Shekokar, "A Framework for Cyberbullying Detection in Social Network", *International Journal of Current Engineering and Technology*, Vol.5, pp. 494-498, Feb 2015.
- [5] B. Aditya et.al, "BANKS: Browsing and Keyword Searching in Relational Databases", *Proceedings of the 28th VLDB Conference*, Hong Kong, China, 2002.
- [6] Yi Luo, Xuemin Lin, "SPARK: Top-k Keyword Query in Relational Databases", *SIGMOD'07*, Chicago, China, 2007.
- [7] Mrs. Vijyalaxmi M et.al, "The How, When and Why of Sentiment Analysis", *Int.J.Computer Technology & Applications*, Vol 4 (4), pp. 660-666, July-August 2013.
- [8] Trisha Dowerah Baruah, "Effectiveness of Social Media as a tool of communication and its potential for technology enabled connections: A micro-level study", *International Journal of Scientific and Research Publications*, Volume 2, pp. 1-10, May 2012.