

# ASSESSING AND MINING RELEVANT DATA ON ONLINE JAVA FORUM USING SEMANTIC ALGORITHM

Rajathi V<sup>1</sup>, Sunanthini J<sup>2</sup>, and Dhanalakshmi B<sup>3</sup>

<sup>1</sup>Department of Computer Science and Engineering, JEPPIAAR S.R.R Engineering College, Padur, Chennai 603103

<sup>2</sup>Department of Computer Science and Engineering, JEPPIAAR S.R.R Engineering College, Padur, Chennai 603103

<sup>3</sup>Department of Computer Science and Engineering, JEPPIAAR S.R.R Engineering College, Padur, Chennai 603103

Email: srisankar64@gmail.com, sunanthjohn@gmail.com, dhana.baskaran@gmail.com

## ABSTRACT:

Online forum mainly helps the user to collect and share their experiences and gain knowledge by posting their content in the forum. The user can post the data which is relevant or irrelevant to the content in the forum. It does not give exact solution or content to the user post. Forum rankings are based on the user's activity and the number of postings so it does not give best solution to the problem. The aim of the present work have the forum admin for uploading only the relevant data in the forum database. Using this forum database the user can upload and getting only the relevant data in the java forum using semantic algorithm. The semantic algorithm is used to represent interactions between the users in the forum based on the semantic content of the data which gives only the relevant content for the user posting. In this semantic algorithm similarity computation to be done for identifying the group of similar content of users' posting. Using homogeneous function, classification of relevant and irrelevant data can be computed by using string comparison in the forum database. Using cognitive cue-competition technique, the irrelevant content can be blocked by the admin in forum. This approach characterizes user interaction,

helps to gain exact relevant knowledge from the users' postings in the java forum.

**Keywords:** Data mining, exploratory analysis, online forum, cognitive technique, network modeling.

## *I. INTRODUCTION*

Forum is an online discussion site where the group of people can exchange their ideas, views and experience. It is a platform to build the relationship among people to gather better knowledge. Java forum have the user interface login with valid email and password to upload relevant data. Using these users' have created the online discussion group to share their relevant ideas to the users' posting. Social networks are used to tie the relationship with friends and share their information but in this forum the discussion is based on the particular topic so the users' should aware of the topic. Forums allow the users to share their experiences which is relevant to the user posting. It gives opportunities to hospital management system, Information technology sectors, healthcare systems, industry, educational institutions to give the better solution. Using a network-based approach [1] for posting java related forum posts. Initial exploratory analysis [3] is an approach to analyzing datasets to summarize their main characteristics to make decisions based on the

data set. After this analysis semantic algorithm is used to compute the similarity of user postings. Semantic search seeks to improve search accuracy by understanding the searcher's intent and the contextual meaning of terms as they appear in the searchable data space, to generate more relevant results. Using cluster analysis [3] to identify the group of similar content. Homogeneous function is used classify the relevant and the irrelevant data. Relevant content can posted to the forum discussion page .Irrelevant content can blocked by the forum admin using cognitive technique [4]. In this technique there are three stages for blocking the content.

## ***ii. RELATED WORK***

M. Chau and H. Chen [2] describes traditional search engines are used for searching the relevant content, it is difficult to search the content in the web. The approach that overcome these difficulties to retrieve the more relevant information by using Topic-specific-search-engine. In this Search engine there are two issues, what is the way to find the document which is relevant to the web and what is the way to filter the documents which is irrelevant from the collection of documents present in the web. The second issue can be solved by using machine learning approach. It consists of two types of analysis web content and web structure. Using machine learning algorithms the input is based on the webpage, content link based properties. The implementation can be done by using forward and backward propagation neural network and vector machine.

Altug Akay [1] describes content matching algorithm is used with TF-IDF technique to discover the most frequent words. Using TF-IDF most frequent words are given as output that may be relevant or even irrelevant to the user's post. It does not give exact solution. Text containing the highest TF-IDF scores represent the data with a vector set in the words list. Each post was converted into a numerical vector as non-zero variables representing the TF-IDF scores in the wordlist term that were present in the respective post. The rest of variables had values assigned to 0 and stored in the database. This database was used for term

enrichment analysis to characterize the content retrieved from users' postings. It does not give exact solution or content to the user posts. User can post unwanted content so it leads to many problems. It takes more time to get the content based on the positive and negative terms by using TF-IDF scores within each post. Forum rankings are based on the user's activity and the number of postings in the forum so it does not give best solution to the problem

## ***iii. PROPOSED SYSTEM***

In this proposed system semantic algorithm [10], is used to accurately represent user interactions by relying on the data's semantic content which gives only the relevant content for the user posting. Semantic search seeks to improve search accuracy by understanding the searcher's intent and the contextual meaning of terms as they appear in the searchable data space, to generate more relevant results. In this semantic algorithm, similarity computation to be done for identifying the group of similar content of users' posting. Using the homogeneous function, classification of relevant and irrelevant data can be computed by string comparison in the forum database. Using cognitive cue-computation technique the irrelevant content can be blocked by the admin in forum. This approach characterizes user interaction, helps to gain exact relevant knowledge from the users' postings in the java forum. Block the unwanted content posted by user in the java forum based on cognitive technique. It takes less amount of time and gives the relevant content for the user posting because of semantic algorithm. Computation of content similarity is used to find relevant and irrelevant content which was done by the admin in the forum that emerged Users effectiveness. Provides more security and viewing only the relevant content in the forum.

## ***iv. SYSTEM DESIGN***

In this system the admin and user registration is done by valid E-mail and password. After login admin upload the relevant content based on the forum and the contents are stored in the database. User uploaded content are analyzed by the exploratory analysis. It gives the processed clean data for the computation of similarity of user postings using semantic algorithm. Using

homogeneous function the algorithm separates the relevant and irrelevant content by string comparison using array list. In this array list the data's are compared to the forum database to give only the relevant data to the users' in the forum discussion page. The irrelevant contents are blocked by using cognitive technique.

*System architecture*

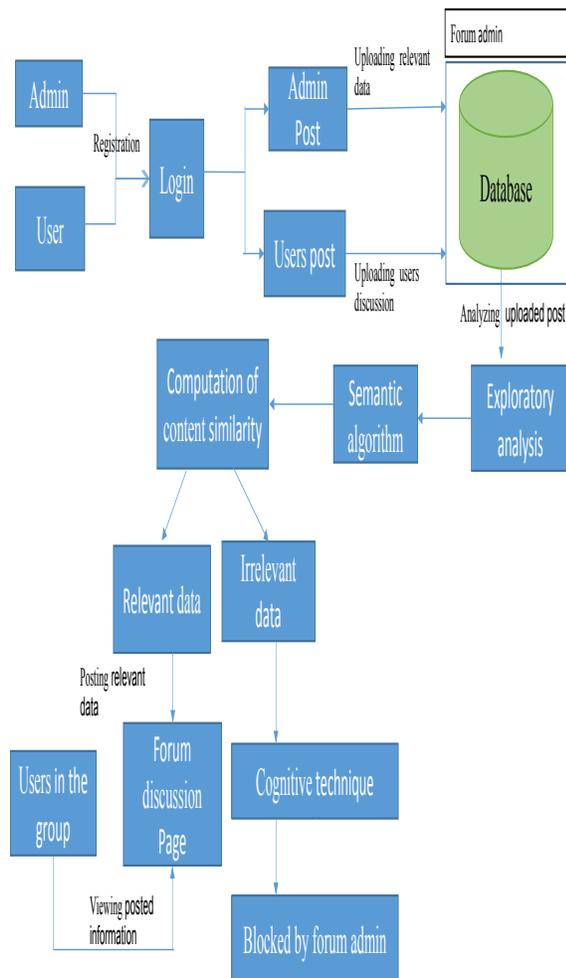


Fig. 1. Architecture of online java forum

**V. METHODS**

*A. DATA COLLECTION*

The admin collect [9] the java related data from www.Herbschildt.com and stored in the forum database. Using this website the admin gather relevant questions and answers for java forum for posting only the relevant content. Content stored in the database are regularly updated by the admin in forum. Forum database contents are only updated by the admin in the forum.

**JAVA FORUM DATABASE**

Question:What are buzzwords in java?

Answer : simple,secure,portable,object oriented,Robust,multithreaded,high performance,dynamic.

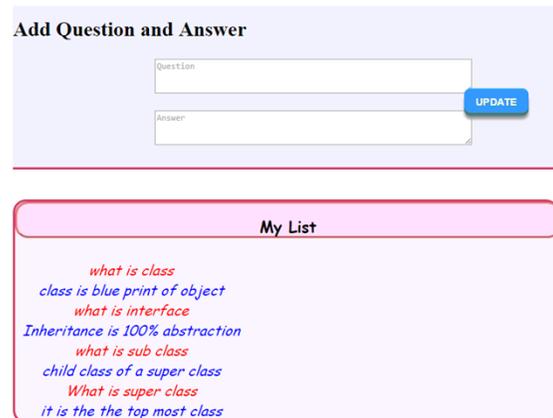


Fig. 2. Admin uploading post

*B. EXPLORATORY ANALYSIS*

In this analysis [6]the raw datasets are collected data preprocessing is done to make clear datasets to make decisions. Using semantic algorithm the user and admin uploaded post are analysed by this technique.Initial exploratory analysis was performed to determine structures based on user opinions among the posts. The results were a compilation of user's clusters and their correlated opinion of the posts.

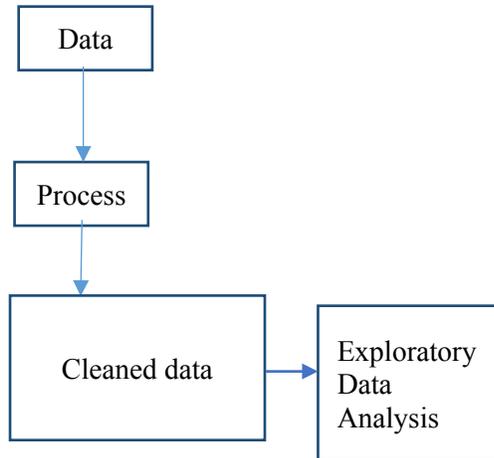


Fig. 3. Exploratory analysis framework

### C. COMPUTATION OF CONTENT SIMILARITY

The semantic algorithm is used for finding computation of content similarity of the data sets. Content similarity is done by using homogeneous function [7] in the forum by string comparison using array list. It gives the data which is irrelevant and relevant to the content. The forum admin posting only the relevant content to the forum discussion page. In the forum database is in the form of array index to check whether the user posting relevant questions and answers which is related to the java content.

### D. POSTING RELEVANT DATA IN FORUM

Semantic algorithm is used computing similarity of user postings in the forum. Using homogeneous function compare the analyzed dataset to the forum database dataset. Both the datasets have the same type of data means the user can upload their posts. This function denotes the datasets which are of same kind and conceptual meaning. So that only the user can upload only the relevant questions and answers in the forum discussion page. Forum posting can be done by the network modeling technique. In this technique [8] separating a high side with sensitive data and low side with internet connection. Benefit of the technique is to avoid the intruder to modify the data which is uploaded by the forum admin in the database

and the user uploaded post to the forum discussion page to provide secure information.



Fig.4. Uploading users' post

### E. BLOCKING IRRELEVANT DATA IN FORUM

Homogeneous function classification gives the relevant and irrelevant data from user posting. Using cognitive technique [4] the irrelevant content can be blocked by forum admin. In this technique there are three stages to perform blocking, the first stage is to be the encoding stage the irrelevant content can be encoded. The second stage is the retention stage the action of absorbing and holding the substance of encoded data. The final stage is the performance stage the irrelevant content are blocked in the forum discussion page. This technique is also called as cue-competition technique for blocking the irrelevant data.

## VI. RESULTS

In this java forum the user can post any number of relevant questions and answers. The accuracy [10] can be measured by using likes and dislikes for the user posting in the forum the user can upload their posts by using semantic algorithm. Using this algorithm the user can post and view only the relevant information in the forum discussion page. Users gain exact relevant knowledge in the particular content. Other users in the forum discussion page can view the other user posting information. Using this forum the user have exact knowledge about the particular topics. It is the place where the user can gather and share knowledge with other users. User viewing only relevant

posting so the user is not deviate from the content they want and also the user don't waste the time for any irrelevant posting.



Fig. 5 Forum discussion page

EXPERIMENTAL EVALUATION

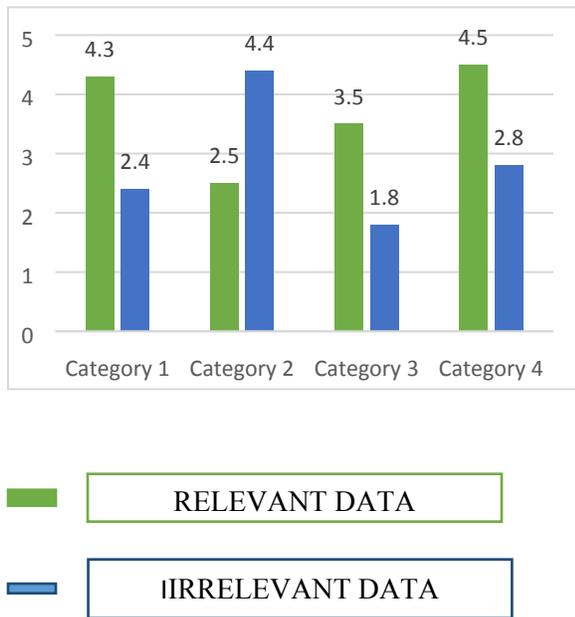


Fig. 6. Graphical analysis

In this experiment analysing the user uploaded post based on the relevant and irrelevant in the forum database and classify by the catagories.

**vii. CONCLUSION**

This approach characterizes user interaction, helps to gain exact relevant knowledge from the user's posting in the java forum. In this semantic algorithm, similarity computation is to be done

for identifying the group of similar content of user's posting.Using cognitive cue-computation technique the irrelevant content can be blocked by the admin in forum. Semantic algorithm represent user posting information based on the data's semantic content which gives only the relevant content for the user posting in the forum.

**Reference:**

[1] Altug Akay, Andrei Dragomir and Bjorn –Erik Erlandsson, "Assessing Antidepressants Using Intelligent Data Monitoring and Mining of Online Fora", IEEE Journal of Biomedical and Health Informatics, Vol 20, No.4, July 2016.

[2] A. J. Ferrari et al., "Burden of depressive disorders by country, sex, age, and year: Findings from the global burden of disease study 2010," PLOS Med., Nov. 2013.

[3] D. Liben-Nowell and J. M. Kleinberg, "The link prediction problem for social networks," J. Amer. Soc. Inf. Sci. Technol., vol. 57, pp. 556–559, May 2007.

[4] K.Babu ,P.Charles, "A System to Filter Unwanted Words Using Blacklists In Social Networks" International Journal Of Computer Science and Information Technologies, vol.5(2), 2014.

[5] L. Getoor and C. Diehl, "Link mining: A survey," SIGKDD Explor. Newslett., vol. 7, pp. 3–12, Dec. 2005.

[6] A. Ng, A. Zheng, and M. Jordan, "Stable algorithms for link analysis," in Proc. SIGIR Conf. Inf. Retrieval, New Orleans, LA, USA, 2001, pp. 258–266

[7] B. Taskar, M. Wong, P. Abbeel, and D. Koller, "Link prediction in relational data," presented at the Advances Neural Information Processing Systems., Vancouver, BC, USA, 2003 .

[8] P. J. Carrington, J. Scott, and S. Wasserman. "Models and Methods in Social Network Analysis".. Cambridge University Press, Cambridge, 2005.

[9] S. Chakrabarti. "Mining the Web". Morgan Kaufman, 2002.

[10] D. J. Cook and L. B. Holder. " Graph-based data mining". IEEE Intelligent Systems, 15(2):32–41, 2000.