IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-5, Issue-6 E-ISSN: 2347-2693

Category Based Search for Collaborative Environment

T. V. Salokhe[1*], P.M.Pawar[2]

[1*]Information Technology, Smt.Kashibai Navale College of Engineering, Savatribai Phule pune University, Pune, India
[2]Information Technology, Smt.Kashibai Navale College of Engineering, Savatribai Phule pune University, Pune, India

[*]*Corresponding Author:Salokhe.tushar30@gmail.com Tel.: +91 9595414949*

Available online at: www.ijcseonline.org

Received: 25/May/2017, Revised: 02/Jun/2017, Accepted: 20/Jun/2017, Published: 30/Jun/2017

***Abstract*—** The Foremost goal of CBSCE (Category based search for Collaborative Environment) is to minimize the time required to obtain particular information, also increase user satisfaction with the result that they are going to get for the specific search. CBSCE provides enhanced search results based on previous user interplays with the systems by tracking each user's performance every time user logs in the system. Existing system used HMM(Hidden markow Model) model which is intensely complicated and hard to extend further, as the primary goal is session clustering, session clustering along with HMM model is what very hard to link such results which take time and less efficient operation. It provides user interaction related search based on existing interactions, for that HMM, is used, Instead of HMM SVM(Support Vector Machine) is a technique you can customize as per the need and which is very flexible with session clustering.

***Keywords*—** Hidden markow model, category based search for knowledge sharing system, Support vector machine, Knowledge sharing

I. INTRODUCTION

Knowledge sharing (KS) method is actually about making the knowledge or good understanding sources accessible to the right people at the right time. Knowledge sharing is accordingly perhaps the single most significant features of this process since the huge majority of KS leads are determined. Knowledge sharing defines as unless push or pull. At the end knowledge worker actively tries out knowledge sources such as library search, managing with a co-worker, etc.

Knowledge sharing dependent on the bias and readiness of the knowledge worker try out and be responsive to those knowledge sources. The right culture, influences, including present [1]. The knowledge sharing task is most useful in IT sector, and it utilized for content management and data and text mining.

Content managed to update, distribute, tag, and otherwise handle the content. Including the range of functions, Consisting of web content management and manage the document systems. They may be used to send and design documents and multimedia material. It also identifies key users and their roles. Specify tasks and duties also numerous cases of content sections or types with defining its workflow tasks [2].Content managers change when variations in content made. Tracking and managing variants of content is done. Supporting way for preserving the issued material is necessary. The tabular model determines the problem about how to decide the grid presentation in a continuous space [3]. Document management systems use different new indexing, searching, and retrieval technique to provide advanced knowledge sharing.

The study of knowledge sharing involves technology transfer and feature literature. The research participating in this area has aim on explaining successes or failures in economic growth by professional involvement. While some researcher stated that rates in the physical and human environment are close enough. Central to each technique is a perception of the exchanging of ideas. The knowledge sharing describes manufacturing processes and organizational designs that are important to them. Continually communicating with users and sources to create helpful KS, which will improve the imported knowledge and equipment.

The original goal of the system is to explain a fast, efficient search result which is relevant to the user requirements.

The other section of the paper is described as follows:
Section II Related work. Section III Implementation detail. Section IV Results and discussions and Section V conclusion and future scope. Section and Section.

II. RELATED WORK

A. 1. Spectral Clustering:-

Spectral clustering method makes use of the spectrum of using the similarity matrix of the data to perform dimensionality reduction before clustering in few dimensions [4, 5]. Primarily, spectral clustering searches a partition of the similar graph like the edges between different clusters have very small weights, and the edges of clusters have very high loads. Spectral clustering also has the advantage that enables the clustering based on click session information.

B. Pairwise Similarity:-

Pairwise similarity discovered from a variety of different signals, they are lexicon similarity and query semantic category similarity, for measuring the similarity belonging to a couple of questions. The inner product of the foretold topic is used to calculate the semantic similarity that belongs to the queries. Various features that are comparable to each other are normalized them into the range [5].

C. Support Vector Machine

In many supervised learning, labelling instances to create a training set is time-consuming and costly; thus, finding ways to minimize the number of labelled examples is useful. Usually, the training set is chosen to be a random sampling of cases [6].However, in number of cases, active learning can be employed. Here, the learners can case actively keep the training data. It hoped that allowing the learner this extra versatility will reduce the learner's need for large quantities of labelled data. Pool-based active learning for classification was proposed by Lewis and Gale (1994).The learner has access to provisions of unlabelled data and can request the right class label for a specific number of instances in the pool [6, 7, 8].

III. METHODOLOGY

The goal of the proposed system is to present a fast, efficient search result which is most related to the user requirements.

In the proposed procedure admin trained the dataset using the CSVM (Customized support vector machine) algorithm. In CSVM algorithm, it will calculate the Link and Domain relevance.as per the relevancy and session clustering overall search result is going to calculate.

A. System Overview

In proposing a method decomposed our problem statements into different modules as admin, user, and query searching using rules created by a CSVM algorithm with a Session clustering algorithm [9]. Figure. 1 shows overall system architecture of CBSKSS.

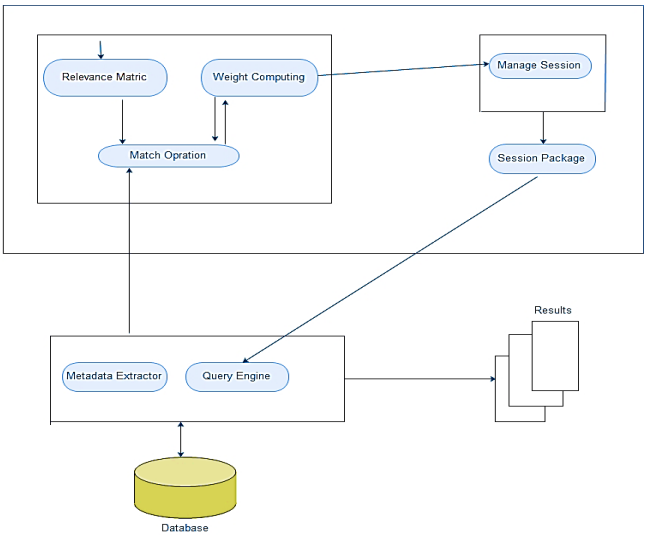

Figure. 1. System Architecture

Following modules are explained for Knowledge sharing data in the collaborative environment.

a. Admin login**:** In This Module admin maintained various Ebook, and user files details using a relational database with a respective attribute. The databases contain attributes like Location, LinkR, DomainR, quantity, Status of products.

b. Query-similarity: The main purpose of our system is reducing a searching time required to search accurate user desired results. To solve this, it uses the index concept. Indexing used to reduce execution time.

When customer login and search the Ebooks and User files details with a particular query, then product details will retrive from the database using rule Session clustering and CSVM [9]. Details are displayed on different types of databases, using join query. Also, the user can give feedback to that Search.

c. Database Training: In this system it allows you to train your database by your own.whenever user searches for any keyword and clicks on any searched link, the relevancy of that particular link It going to increase.so as the number of searches and clicks goes on increasing the more database is going to train itself and that will result in better satisfactory search results for users.

d. Searching Criteria: For categories searching, it uses CSVM algorithm. From which user get correct results from complex data [10]. For creating patterns and rules Session Clustering algorithm is used.

In this system, it provides following search criteria

1. LinkR Method: System will provide search result based on LinkR calculation.
2. DomainR Method: System will provide search result based on DomainR calculation.

e. Display result**:** Results presented Which are most recently searched.

B. Constraints

1. The Database should load.

2. Provide login and password for each user.

3. Only registered user able to view the search.

4. User search by inserting Keywords.

C. Mathematical Model

- Let member in a collaborative environment as a "User". In our problem, it have a group of h users $\{e_1, \ldots, e_h\}$ where each user e_i generates a sequence of sessions
- $S_{(i)} = \{S_{(i)1}, \ldots, S_{(i)} N_i\}$.
- $S = \{S_1, \ldots, S_N\}$, with totally N sessions.
- Sub-Task:
- Partition W into a set of clusters $C = \{C_1, \ldots, C_t\}$ where each cluster represents a task;
- Partition sessions in each C_i into a set of micro aspects
- $S_i = \{s_{i1} \ldots s_{iti}\}$, where each micro aspect
- Compute the association weight between e_i and s_{jk} as $|W_{(i)} T_{sjk}|$. Given a query q, produce a ranking of {ei} according to their relevance it can also consider it as weights.
- Let $\mathbf{L_r}$ be the Link_Relevancy, $\mathbf{D_r}$ be the Domain_Relevancy.
- Given training data $\{x_1 \ldots x_n\}$ that are vectors in some space **X**.
- SVMs here used is technique that separates the training data based on support vector (L_r, D_r)
- Function used to calculate the closest data among space **X** using SVM
- $$f(x) \left(\sum_{i=1}^{n} \alpha(X_i, X) \right)$$

Match result (M_r) =MinDistance $(X_i (X_{ij}, Xik)) => L_r, Dr$

IV. RESULTS AND DISCUSSION

The primary aim of CBSKSS is to provide user number of links which are related to the search input as well as most recently links should show first. This is managed by LinkR, Domain R and session clustering. Results analyzed based on the Domain R, Link R, and a number of clustered sessions, a number of links without relevancy and number of links with relevancy. The proposed search mechanism here will enable the system to classify all searches in categories and apply SVM technique with the help of click based expert search method to mine the result set. Proposed implementation provides better search results and reduces extra time to retrieve related information. It will be easy to expand proposed functionality as SVM can be easily extensible in nature. In future user based properties can be used to further enhance search result.

Proposed Modules:

- Search: Registered users will be able to search the files. The search will be based on relevancy, links with higher relevancy will always be on the top.
- Session Clustering: Every time a new request hits the server and user logins to the system a session is created and maintained by this module. Clustering, Search results between sessions are deep analysed and matched, after matching various clusters has been created

Relevancy Calculation: Relevancy calculation is the main module of this system, every time user searches any file and then clicks on particular file then relevancy of that particular file as well as related domain relevancy increases by one.

To check the effectiveness of proposed system, Time required to search any keyword is calculated, both proposed system and existing system follows session clustering and relevancy metric, but proposed algorithm is designed to optimize time with the same result set. Following is the performance snapshot of proposed system and existing system. Performance result shows every keyword searched and time required for it. In below Figure 2 and Figure 3 shows time elapsed for search keywords, Fig 2 shows data for existing system and Figure 3 shows data for implemented system, by analyzing both data it is clear that time required for search is implemented system is faster than existing system, time require to search keyword 'java' in existing system is more than time require search same keyword in implemented system.

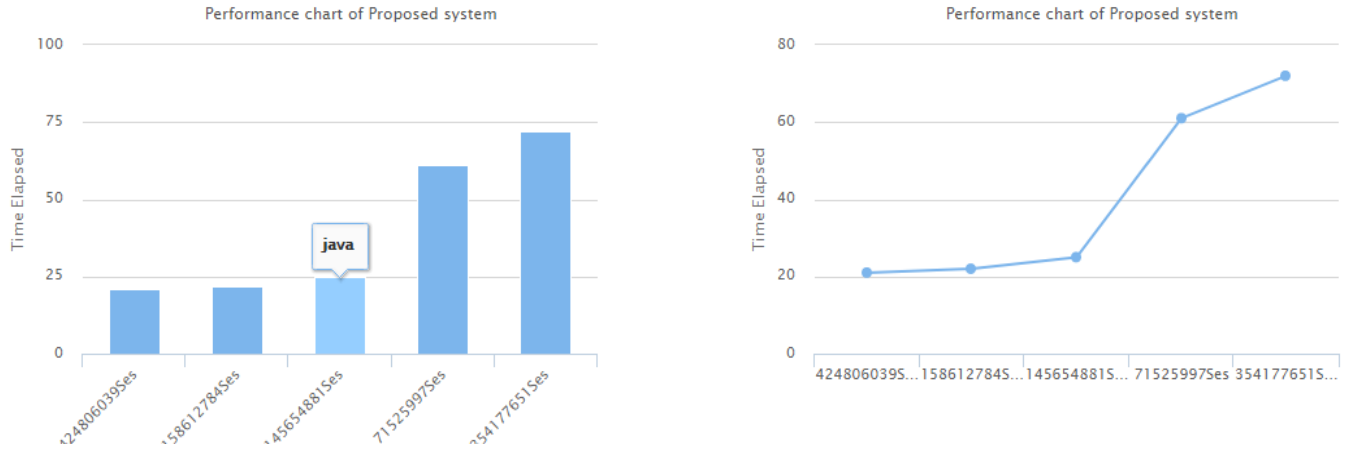

Figure. 2.Performace of Proposed System

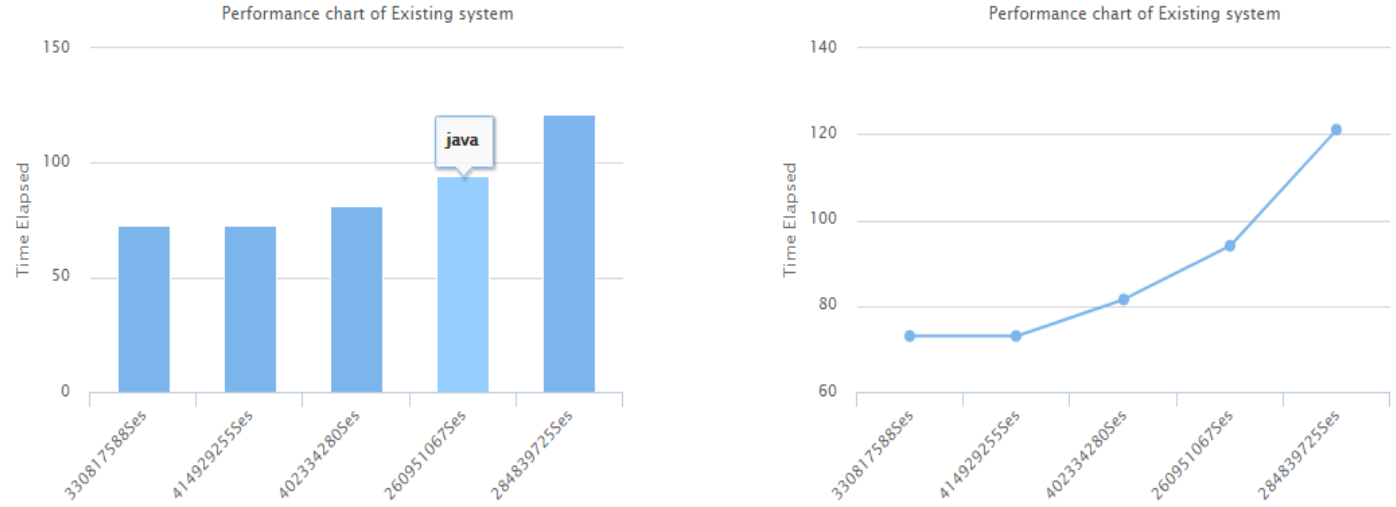

Figure. 3.Performace of Existing System

V. CONCLUSION & FUTURE SCOPE

Existing system is based on single keyword matching, this is resolved in implemented system, as well as per implemented results, it is clearly visible that implemented system takes less time to search and search data is most relevant to the search keywords. In Figure 2 time required to search Keyword "java" is more than time for search "java" in Figure 3. In category based search for collaborative environment, the focus is totally on the most frequent used data first. The primary goal is to refine results based on input keyword and relevancy of data.

The new system can extend towards more user level parameters to improve search results more precise to the distinct domain. In the future, can work on a thorough understanding of relevancy metric, more refined search results and custom GUI (Graphical user interface). Also, Relevancy calculation and session clustering can be further modified by adding more related criteria as per the domain in which application is going to execute.

REFERENCES

[1] S.H. Beigy, H, "*Expertise Finding, in Bibliographic Network: Topic Dominance Learning Approach*", Cybernetics, Volume-44, Issue-12, IEEE transaction, Iran, 2646 – 2657, 2014.

[2] H. Shen; Z. Li; J. Liu; Grant, J.E., "*Knowledge Sharing in the Online Social Network of Yahoo! Answers and Its Implications*", Volume-64, Issue-6, IEEE Transactions on Computers, USA, 1715 - 1728, 2014.

[3] K. Hwang, "*Model Learning and Knowledge Sharing for a Multiagent System With Dyna-Q Learning*", Volume-44, Issue-5 IEEE TRANSACTIONS ON CYBERNETICS, Taiwan, 978-990, 2014.

[4] Z. Guan, S. Yang, H. Sun, M. Srivatsa, X. Yan," *Fine-Grained Knowledge Sharing in Collaborative Environments*", Volume-27, Issue-8, IEEE Transactions on Knowledge and Data Engineering, Northwest Univ. of China, Xi'an, China, 2163 – 2174,2015.

[5] A. Appice,D. Malerba, "*A Co-training Strategy for Multiple View Clustering in Process Mining*", Volume-9, Issue-6, IEEE TRANSACTIONS ON SERVICES COMPUTING, Italy, 43–50,2015

[6] Z. Lu, X. Yang," *Inferring User Image-Search Goals under the Implicit Guidance of Users*", Volume-24, Issue-3, IEEE Circuits and Systems Society, China,394 - 406, 2014

[7] S. Tong, D. Koller*," Support Vector Machine Active Learning with Applications to Text Classification*", Journal of Machine Learning Research, Stanford CA, 45-66, 2001.

[8] Y. Song, Ming-Wei Chang, "*Learning to Extract Cross-Session Search Tasks*", the University of Illinois at Urbana-Champaign Urbana IL, USA,1-11,2013

[9] B. Li, X. Zhu, "*Rating Knowledge Sharing in Cross-Domain Collaborative Filtering*", Volume-45, Issue-5, Cybernetics, IEEE Transaction, Sydney, 1068 - 1082,2015

[10] N. Kolhe,H. Kulkarni,I. Kedia,S. Gaikwad, "*Comparative Analysis of Cluster based Boosting*", Volume-3, Issue-10, International Journal of Computer Sciences and Engineering, pune india, 60-70, 2015

