International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 24 (2015) pp 44263-44270

An Effective Collaborative Querying Using Distributed Clustering Algorithm on Multiple Domain

D. Madhusubram

Research Scholar, Bharathiar University, Coimbatore, India dms_ram@yahoo.com

S.P. Shantharajah,

Professor, Sona College of Technology, Salem spshantharaj@gmail.com

Abstract

Web search engines help user's to identify the useful information from multiple domain websites. Traditionally, search engines have focused on verification technique using k-nearest neighbor from single domain. However, multiple domain based query verification process requires a larger area to be covered. In designing user driven queries, user query collaboration to deal user friendly format are important and often crucial. User driven queries often involves intuitive interaction with datum of multiple domains and thus obtaining user friendly format is often difficult. To extend the work for effective collaboration of user queries, Distributed Multiple Domain Information Clustering (DMDIC) method is proposed in this paper. DMDIC method provides effective collaborative querying by initially measuring the similarity between user queries. The similarity features of two types are evaluated using DMDIC method to improve the collaboration between the queried results. These two types are referred to as the Prior Query Content based similarity measure and Prior Answer Content based similarity measure on multiple domains. These two similarity features are integrated in the second part of DMDIC method, and then an algorithm called as the Distributed clustering algorithm is employed. The Distributed clustering in DMDIC method chooses the titles as the cluster head. Distributed clustering algorithm clusters the similar user queries and develop a user friendly format for effective collaborative querying on multiple domains. Experiment is performed on factors such as collaborative querying efficiency, cluster processing time and precision ratio.

Keywords: Distributed Clustering Algorithm, User Query, content based similarity measure, Collaborative Querying, Multiple Web Domain.

Introduction

The recent advancements in user driven queries on multiple web domains have created a greater attention. Users now require the information that is available in web domains ranging from single web domains to multiple web domains. Meanwhile, we have seen an expansion in the use of collaborative queries such as widely equipped information for genes, proteins, drugs in a more user friendly format.

An efficient K-nearest-neighbour query verification technique with improved distance (k-NN using DIST) [1] applied Voronoi diagram to prove the collaboration between query results. However, in practice, multiple domain based query verification process is not carried out effectively. Another search platform for user driven queries named as ReVeaLD (a Real-time Visual Explorer and Aggregator of Linked Data) [2] framework increase intuitive interaction with datum of multiple domains. But, ReVeaLD was not extended to handle user query collaboration to yield user friendly format satisfying only the administrator.

Role-based collaborative search process was introduced in [3] satisfied the role of both gatherer and surveyor for efficient partitioning of data. However, there were certain limitations in collaborative information processing environment and therefore could not be applied in web database. To provide solutions to perform efficient matching of records using web database, Unsupervised Duplicate Detection (UDD) [4] was introduced. The method proved to be efficient with higher precision reducing the average execution time.

With higher fluidity and increasing unpredictability of arrivals of data streams, one of the solutions to this problem is adaptive query processing. An adaptive model called Double Index Nested-loops Reactive join (DINER) [5] was introduced a two-way join algorithm for increasing the rate of results produced. DINER combined intuitive flushing and re-entrant join technique to maximize the productivity rate and provided a flexible switch between processing in-memory and disk-resident tuples. Though the method was proved to be robust, but at the cost of computation. To minimize the cost of computation, three approximate oracles were introduced in [6]. However, measures were not included for traffic congestion.

Concept-based user profiles were introduced in [7] that included both positive and negative preferences and address the problems related to traffic congestion using web search engine log files. Though traffic congestion was addressed, ranking user interests whenever traffic was overload remains unaddressed. An efficient ranking mechanism was introduced in [8] based on concept hierarchies. The method proved to be efficient in terms of processing time.

In this work, focus on effective collaboration of the user queries Prior Query Content based similarity measure and Prior Answer Content based similarity measure on multiple domains. The Distributed Clustering algorithm is introduced for developing a user friendly format for effective collaborative querying on multiple domains. The analyzed result is used for collaborative query efficiency based on similar user queries. Efficient querying is performed to assess

larger upon multiple domains with differing domain size and hence to evaluate the precision ratio. As a result, query similar measure and answer similar measure easily evaluates the combined similarity measure of multiple domains in an efficient manner.

The structure of paper is as follows. In Section 1, collaborative querying with respect to multiple domains with existing works is described. In Section 2, literatures related to query result efficiency and multiple domains is elaborated by comparing the current history. Section 3 explains about the proposed work Distributed Multiple Domain Information Clustering (DMDIC) method with neat architecture diagram and algorithmic steps to increase the precision ratio. Section 4 analyzes the experimental results and Section 5 provides the result analysis with the help of table values and graph. Finally, concluding remarks is included in Section 6.

Related Works

An effective collaboration mechanism is highly required to meet the demands of multiple heterogeneous sensor networks. A framework for collaborative heterogeneous sensor networks called as COSE [9] was designed where several users communicate with each other for successful processing of queries. Though average running time was reduced multiple pipelines remained unaddressed. To solve the issue, an efficient indexing mechanism was introduced in [10] using a storage scheme based on cluster. With this, the number of accessed records got increased but dominant relationship model remained unaddressed. Saturn [11] using distributed hash tables provided solution for dominant relationship that included load balance and fault tolerance.

With the increased browsing of books, stories, web page recommendation has received greater attention in recent years. A novel method was introduced in [12] using semantic enhancement that combined both the knowledge of domain and web usage for a specific website. Though precision was improved, semantic web page recommendation have to be provided.

A personalized search engine model was introduced in [13] to address semantic nature of data that increases accuracy, reliability and quality of data. However, with the increasing amount of data, the rate at which query processing was handled have to be solved. An approach based on collaborative search [14] using web log data addressed similar user community and improved collaborative ranking. But multiple domains were not included. Data integration [15] offered query processing in terms of local, global and local global mechanism. Query Redefining [16] initial user query is redefined using condition redefine query.

Based on the aforementioned methods, in our work we design an efficient distributed multiple domain information clustering method to improve the collaborative querying efficiency.

Distributed Multiple Domain Information Clustering Method

In this section, a detailed description of the proposed distributed multiple domain information clustering (DMDIC) method is presented. First, an illustrative diagram of collaborative querying system is briefly explained, followed by an elaborate discussion of the key elements collaborative query processing. Next, an architecture diagram of DMDIC method is presented with detailed description about prior query content based similarity measure and prior answer content based similarity measure is presented. Finally, the distributed clustering algorithm is discussed in details. Certain assumptions have been made during the modelling of the problem as a collaborative query system framework.

The main objective of distributed multiple domain information clustering is to provide effective collaborative querying to the users. The collaborative querying is performed using multiple domain information. The web information plays a vital role in retrieving the information based on the needs of the user. The similarity between the previous user request and current user request helps to easily collaborates user query in DMDIC method. The collaborative system is depicted clearly in Figure 1.

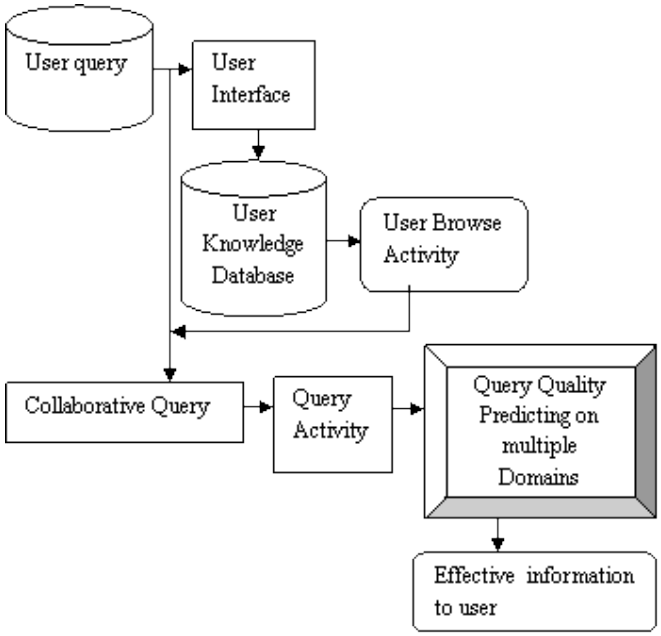

Figure 1 Collaborative Querying System

Whenever the user needs certain information, the intended user places the query through a designed DMDIC method. A newly asked query enters into the interface and evaluate all (i.e.,) multiple domains to attain effective collaborative system. The collaborative querying system extracts the activity of web users and analyses the profiles with the aid of user knowledge database.

The most appropriate query attains all the responses from multiple domains for quality prediction. Through the activity browsed by the users, the knowledge from usage database is continuously gathered. The gathered knowledge is then shared with questions of different users. In DMDIC method, the entire user activity is stored in usage knowledge database.

The user knowledge database uses the similarity measure of prior queries and answer for the effective collaborative result. The Prior Query content based similarity measure uses the previous user query format (i.e.,) terms to identify the

similarity. Prior Answer content based similarity measure is also used on identifying the related answer providing URL from the multiple domains. DMDIC method uses the prior query and answer content to collaborate the system using the distributed clustering algorithm.

Distributed algorithm sets up a clustering that satisfies the properties listed in the element of DMDIC method. The algorithm is executed on each user query in such a way that a query decides its own role depending solely on the decision of its answer which has comparatively the higher weight value. Depending on the function, a specific procedure is executed using distributed clustering algorithm for each user query. The distributed clustering chooses the titles as the cluster head.

For each user query UQ_i the element unit vectors for distributed clustering is represented as,

$$UQ_i = (D_i, T_i, RURL_i)$$

Equation (1)

In Equation (1), D_i is the document related to user queries and T_i is the related title whereas $RURL_i$ is the element unit vector representing the results of the URLs. The URL returns response of UQ_i in DMDIC method. The overall architecture Diagram of DMDIC method is depicted in Figure 2.

As illustrated in Figure 2, user request are submitted on the web search engine. The DMDIC method computes the similarity between the user requests. As depicted in the figure, the user request similarity depends on two factors. They are prior user query content based similarity measure and prior user answer content based similarity measure. The Distributed Clustering algorithm uses the title as the cluster head and improves the collaborative querying efficiency and cluster rate.

To enhance teaching and learning of software engineering.

A reflective process model has been used by adapting Boehm' spiral model [1] of software.

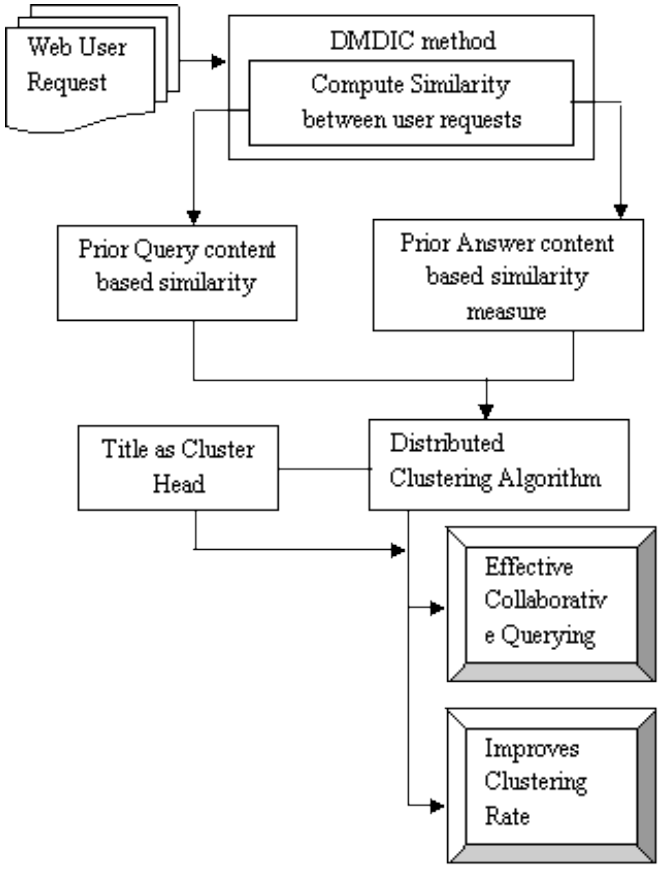

Figure 2 Architecture Diagram of DMDIC Method

Prior Query Content based Similarity measure

The first step in the formation of distributed multiple domain information clustering is to define the prior query content based similarity measure. Let us consider a user query followed by which the query terms are examined with previous query terms vectors using DMDIC method. All key terms are extracted after which the processing operation takes place for measuring similarity. The extracted user key terms are used to represent the user query UQ_i on Document D_i, loop of the spiral has been assigned the following four activities:

$$Weight\ (W) of\ PQCS = \sum_{i=1}^{M} Prior\ Frequency_{\ i} * N$$

Equation (2)

The weight (W) of Prior Query Content based Similarity (PQCS) is measured in (2) which is the product of frequency level of previous query with 'N' number of user queries related to specific topic. The count of similarity measure in DMDIC method is computed as,

$$N = \log\left(\frac{W_{PQCS}}{Prior\ Frequency_{\ i}}\right)$$

Equation (3)

The count of similarity measure 'N' is obtained using the logarithmic form. The query prior frequency level is identified to match the key terms and recognize the similarity measure. The prior query content based similarity measure is expressed as,

$$Query\ Similarity\ Measure = \frac{\sum_{i=1}^{n}(W_{current\ query}) * \sum_{i=1}^{n}(W_{PQCS})}{\sqrt{\sum_{i=1}^{n}(W_{Current\ Query})^2} * \sum_{i=1}^{n}(W_{PQCS})^2}$$

Equation (4)

In DMDIC method, similarity is measured based on the current user query and prior user query correspondingly to attain higher result on collaborative querying.

Prior Answer Content based Similarity measure

The feasibility of similarity measure largely depends upon the prior answer content in determining the precision ratio of the model. Essentially, the formulation of prior answer content similarity measure follows a two-step process, where during the first step the answer content based similarity measure in DMDIC method is computed based on the answer URL pages. The second step performs the answer terms along with titles where the titles are used to easily represent answer for each user query by verifying multiple web domains.

$$A_i = \{URL\left(UQ_{current\ query} \cap UQ_{PQCS}\right)\}$$

Equation (5)

The URL of the previous query result UQ_{PQCS} is measured with the aid of current query URL that efficiently identifies the similarity. With the help of the intersection point, the similarity value is identified. The answer content A_i based similarity rate is measured as,

$$Answer\ Similar\ Measure = \frac{A_i}{Max\ (UQ_{current\ query}, UQ_{PQCS})}$$

Equation (6)

In order to fetch effective query answer from multiple domains, (6) is used to identify the similarity measure of prior URL and current URL. The maximum user query result provides satisfied result to the users with lesser false positive rate.

Distributed Clustering Method

With the obtained prior query content based and prior answer content based, the final part combines them for collaborative querying on multiple domains. The integration of these two similarity measure is described as,

Combined Similarity Measure = Query Similar Measure + Answer Similar Measure

Equation (7)

The combined similarity measure obtained from (7) are grouped (i.e.,) clustered using the distributed clustering algorithm in DMDIC method. The distributed clustering algorithm using DMDIC method clusters all types of user queries on varying domains to attain higher precision rate. The distributed clustering with cluster head is shown in Figure 3.

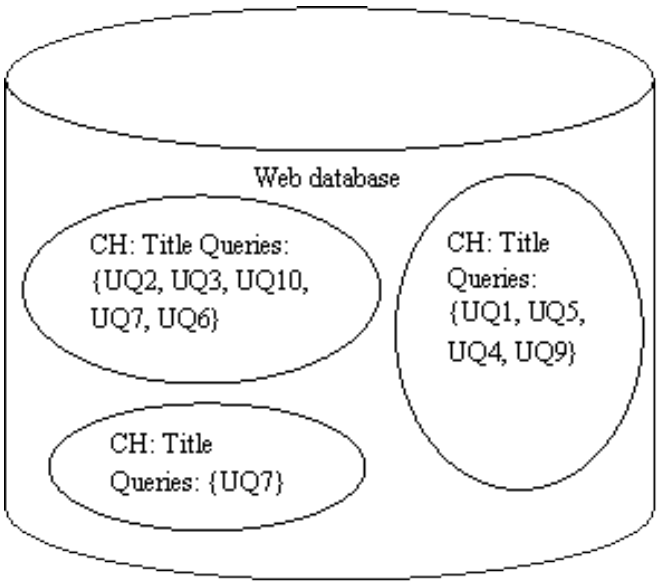

Figure 3 Distributed Clustering Representation

Figure 3 illustrates the representation of distributed clustering that selects the title as Cluster Head (CH) with user queries of '10' different users taken to demonstrate distributed clustering. The combined (i.e., query and answer) similarity measure result from different users are grouped together using distributed clustering procedure. The {UQ2, UQ3, UQ10, UQ7, UQ6} user queries are grouped together using the similarity measure value obtained (4) and (6). In a similar manner, the other similar information is grouped. The distributed clustering algorithm is described below,

//Distributed Clustering Algorithm
Begin
Step 1:
Step 1: If (W_PQCS=Combined Similarity Measure) Then
Step 1.1: Cluster similar weight user queries
Step 1.2: Set "Title" as Cluster Head
Step 1.3: Cluster with similar user query points
Step 2: Else form another cluster with CH
Step 3: Clustered based on multiple domain content analysis
Step 4: Repeat step from 1 to 3 for all user collaborative querying
End

The above algorithm describes the steps involved in the design of distributed clustering based on the 'title' as cluster head. Here, multiple domain contents are analyzed to provide effective collaborative clustering. Clusters are easily defined as the most likely belonging to the same distribution in

DMDIC method, thereby reducing cluster processing time. Content based similarity measure in query and answer form remains unchanged even on combined equation of DMDIC method.

Experimental Evaluation

Distributed Multiple Domain Information Clustering (DMDIC) method uses JAVA platform. The Java platform with Weka tool uses the MSNBC.com Anonymous Web Dataset for experimental work which is based on multiple web domain information. The DMDIC method uses the data obtained from the pages visited by the users at msnbc.com which are in order of appearance. MSNBC.com Anonymous Web contains the 989818 users with visits made at an average of 5.7 per user whereas the URL category ranges from 10 to 5000 different domains.

Distributed Multiple Domain Information Clustering (DMDIC) method compares the result with the existing K-Nearest-Neighbor using distance (k-NN using DIST) [1] query verification technique and ReVeaLD (a Real-time Visual Explorer and Aggregator of Linked Data) [2] framework. The experiment is conducted on factors such as collaborative querying efficiency, cluster processing time, precision ratio and false positive rate.

The measure of collaborative querying efficiency is obtained using the entropy factor. The entropy value for DMDIC method is obtained using the ratio factor of similarity measure, the ratio between prior query content based PQC and prior answer content based similarity measure PAC respectively.

$$Entropy_{PQC,PAC} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} P(PQC_i, PAC_j) * \log\ P\ (PAC_j/PQC_i)}{\sum_{j=1}^{n} P(PAC_j) * \log(PAC_j)}$$

Equation (8)

The cluster processing time using DMDIC method is the time taken to perform combined similarity measure, i.e., the summation of time taken to process the query similarity measure $Time_{QSM}$ the time to perform answer similarity measure respectively$Time_{ASM}$.

$$CPT = Time(Time_{QSM} + Time_{ASM})$$

Equation (9)

The ratio of the number of unique answer returned $Unique\ answer\ returned_{mwd}$ divided by the number of queries $Number\ of\ queries_{mwd}$ implies precision ratio of a user's queries from multiple web domains is given as below:

$$Precision = Unique\ answer\ returned_{mwd} / Number\ of\ queries_{mwd}$$

Equation (10)

The false positive rate using DMDIC method is the ratio at which the answer is returned from multiple web domains but actually does not has the unique answer.

Results Analysis of DMDIC Method

The Distributed Multiple Domain Information Clustering (DMDIC) method is analyzed against K-Nearest-Neighbor using distance (k-NN using DIST) [1] query verification technique and ReVeaLD (a Real-time Visual Explorer and Aggregator of Linked Data) [2] framework. Each method has its own collaborative query efficiency. The existing and proposed result is analyzed with the help of table values and graph points. Table 1 tabulates the collaborative query efficiency with respect to the number of users considered for the purpose of experiment being conducted using JAVA.

Table 1 Tabulation for collaborative query efficiency comparing with existing methods k-NN using DIST and ReVeaLD

Number of users	Collaborative Query Efficiency (%)		
	DMDIC	k-NN using DIST	ReVeaLD
20	55.35	45.22	40.11
40	58.45	48.32	43.21
60	71.25	61.12	56.01
80	74.35	64.22	59.11
100	80.25	70.12	65.01
120	82.25	72.12	67.01

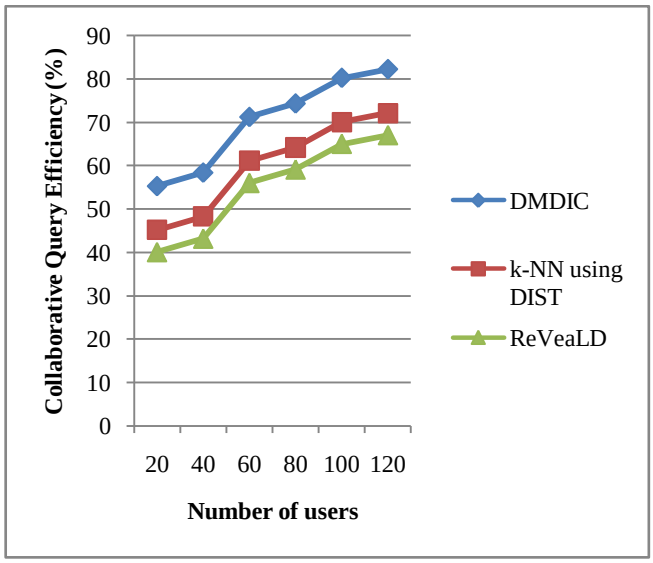

Figure 4 Measure of collaborative query efficiency

Figure 4 show that the proposed Distributed Multiple Domain Information Clustering (DMDIC) method provides higher collaborative query efficiency when compared to k-NN using DIST [1] and ReVeaLD [2]. This is because of the application of two types of similarity measure called the Prior Query content based similarity measure and Prior Answer content based similarity measure with the aid of user knowledge database that eventually separate the similar and dissimilar

queries and improves the collaborative query efficiency by 12 – 18 % when compared to k-NN using DIST [1]. In addition to that with the use of the appropriate similarity measures from multiple domains helps in improving the collaborative query efficiency by 18 – 27 % than the ReVeaLD [2] technique.

Table 2 Tabulation for Cluster Processing Time

Domain sets	Cluster Processing Time (ms)		
	DMDIC	k-NN using DIST	ReVeaLD
2	0.135	0.215	0.235
4	0.155	0.235	0.245
6	0.185	0.265	0.355
8	2.015	3.185	3.295
10	2.005	2.395	2.505
12	2.250	3.250	3.550

The cluster processing time of our DMDIC method is presented in table 2. It is easy to find that the cluster processing time is improved using DMDIC method than the state-of-art methods.

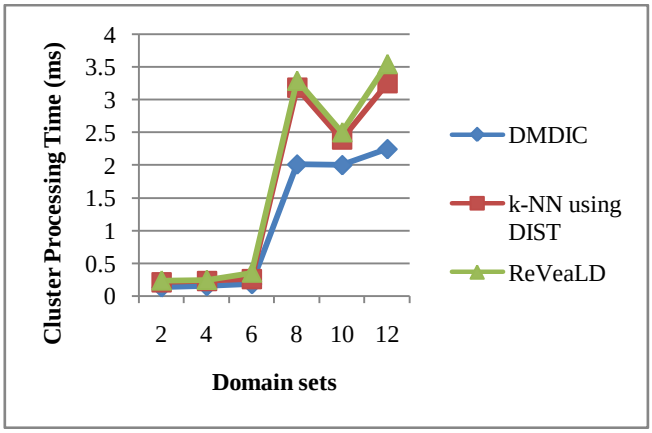

Figure 5 Measure of cluster processing time

Figure 5 describes the cluster processing time based on the differing domain sets. The number of images used is exactly pre-processed using the interface and improve the cluster processing time by 19 – 59 % when compared to the k-NN using DIST [1] technique. As the number of domain sets is increased, thought he clustering processing time also gets increased but comparatively DMDIC method is better than the two other methods [1] [2]. In addition, the combined similarity measure obtained are grouped (i.e.,) clustered using distributed clustering algorithm thereby reducing cluster processing time by 24 – 91 % compared to ReVeaLD [2].

Table 3 Tabulation for Precision

Number of users	Precision (%)		
	DMDIC	k-NN using DIST	ReVeaLD
20	48.12	20	48.12
40	51.05	40	51.05
60	68.45	60	68.45
80	72.32	80	72.32
100	85.75	100	85.75
120	88.55	120	88.55

Table 3 describes the precision efficiency on differing set of users based on the several numbers of queries given as input. As the number of users count improves, the efficiency rate of precision is also improved.

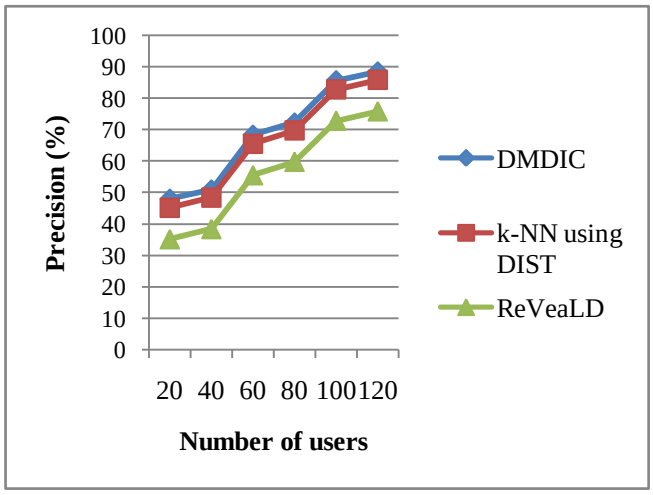

Figure 6 Measure of precision

To ascertain the performance of the precision, comparison is made with two other existing techniques, K-Nearest-Neighbor using distance (k-NN using DIST) [1] query verification technique ReVeaLD (a Real-time Visual Explorer and Aggregator of Linked Data) [2] framework. In Figure 6, the number of users to be ascertained is varied between 20 and 120. From the figure it is illustrative that the precision that measures the unique answer returned from multiple web domains is higher using the proposed Distributed Multiple Domain Information Clustering (DMDIC) method when compared to the two other existing works. This is because with the formulation of prior answer content similarity measure identified based on the answer URL pages, the rate of precision gets increased from 3 – 6 % compared to k-NN using DIST [1]. Furthermore, by verifying multiple web domains using distributed clustering algorithm, the DMDIC method provides satisfied result to the users with lesser false positive rate increasing the precision ratio by 14 – 26 % compared to ReVeaLD [2].

Table 4 Tabulation for false positive rate.

Number of users	False Positive Rate (%)		
	DMDIC	k-NN using DIST	ReVeaLD
20	18.20	20	18.20
40	21.41	40	21.41
60	28.52	60	28.52

80	42.79	80	42.79
100	55.85	100	55.85
120	58.82	120	58.82

The comparison of false positive rate is presented in table 4 with respect to the varying number of users in the range of 20 – 120 with varying range of queries to be measured from multiple domain sets. With increase in the number of users who provide queries, the false positive rate to perform accurate query retrieval is also increased.

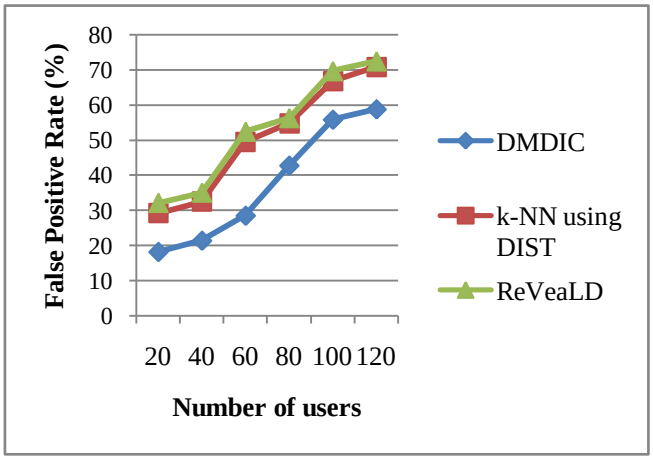

Figure 7 Measure of False positive rate

In figure 7, we depict the false positive rate attained using varying the sizes of the users from 120 different users for experimental purpose and applied in JAVA. From the figure, the value of false positive rate achieved using the proposed DMDIC method is lower when compared to two other existing techniques namely, k-NN using DIST [1] and ReVeaLD [2]. Besides we can also observe that by increasing the number of users who provide the queries, the false positive rate is increased using all the methods. But comparatively, it is lower in DMDIC method because the two similarity measures are separated and then integrated using the distributed clustering algorithm on the basis of titles that significantly minimizes the false positive rate by 19 – 73 % and 23 – 83 % compared to k-NN using DIST [1] and ReVeaLD [2] respectively.

Conclusion

An effective collaborative querying using distributed clustering algorithm is applied on multiple domains has been designed to achieve effective collaboration in terms of precision on MSNBC.COM anonymous web data set. We adopt two similarity measures called as the prior query content and prior answer content between web user requests that produces effective collaborative query efficiency with minimal cluster processing time. The proposed Distributed Multiple Domain Information Clustering (DMDIC) method is user friendly because of the introduction of distributed clustering algorithm that efficiently integrates the two similarity measures that results in reducing the false positive rate. Experimental evaluation is conducted with the MSNBC.com Anonymous Web Dataset to analyze the robustness with respect to collaborative query efficiency, cluster processing time, precision and false positive rate. Performances results reveal that the proposed DMDIC method provides higher level of collaborative query efficiency and precision by reducing the false positive rate. Compared to the existing query processing methods on multiple web domains, the proposed DMDIC method is comparatively better than to the state-of-art works.

References

[1] Yinan Jing., Ling Hu., Wei-Shinn Ku., and Cyrus Shahabi., "Authentication of k Nearest Neighbor Query on Road Networks," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 6, JUNE 2014.

[2] Maulik R. Kamdar., Dimitris Zeginis., Ali Hasnain., Stefan Decker., Helena F. Deus., "ReVeaLD: A user-driven domain-specific interactive search platform for biomedical research," Journal of Biomedical Informatics., Elsevier Journal., 2013.

[3] Chirag Shah, Jeremy Pickens, Gene Golovchinsky," Role-based results redistribution for collaborative information retrieval", Information Processing and Management, Elsevier, Feb 2010.

[4] Weifeng Su, Jiying Wang, and Frederick H. Lochovsky, Member," Record Matching over Query Results from Multiple Web Databases", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 22, NO. 4, APRIL 2010.

[5] Mihael A. Bornea, Vasilis Vassalos, Yannis Kotidis, and Antonios Deligiannakis," Adaptive Join Operators for Result Rate Optimization on Streaming Inputs", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 22, NO. 8, AUGUST 2010.

[6] Jagan Sankaranarayanan, Member, IEEE, and Hanan Samet,," Query Processing Using Distance Oracles for Spatial Networks", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 22, NO. 8, AUGUST 2010.

[7] Kenneth Wai-Ting Leung and Dik Lun Lee," Deriving Concept-Based User Profiles from Search Engine Logs", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 22, NO. 7, JULY 2010.

[8] Abhijith Kashyap, Vagelis Hristidis, Michalis Petropoulos, and Sotiria Tavoulari," Effective Navigation of Query Results Based on Concept Hierarchies",IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 4, APRIL 2011.

[9] Yuan He, Member, IEEE, and Mo Li," COSE: A Query-Cent Tien Tuan Dao ?, Tuan Nha Hoang, Xuan Hien Ta, Marie Christine Ho Ba Thoric Framework of Collaborative Heterogeneous Sensor

Networks", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 23, NO. 9, SEPTEMBER 2012.

[10] Lei Zou, Member, IEEE, and Lei Chen," Pareto-Based Dominant Graph: An Efficient Indexing Structure to Answer Top-K Queries", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 5, MAY 2011.

[11] Theoni Pitoura, Member, IEEE, Nikos Ntarmos, and Peter Triantafillou," Saturn: Range Queries, Load Balancing and Fault Tolerance in DHT Data Systems", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 7, JULY 2012.

[12] Thi Thanh Sang Nguyen, Hai Yan Lu, Jie Lu," Web-page Recommendation based on Web Usage and Domain Knowledge", IEEE Transactions on Knowledge and Data Engineering, Vol. 26, Issue. 10, Sep 2014.

[13] Tien Tuan Dao, Tuan Nha Hoang, Xuan Hien Ta, Marie Christine Ho Ba Tho," Knowledge-based personalized search engine for the Web-based Human Musculoskeletal System Resources (HMSR) in biomechanics", Journal of Biomedical Informatics, Elsevier, Nov 2012.

[14] Haoran Xie, Xiaodong Li, JiantaoWang, Qing Li Yi Cai," The Collaborative Search by Tag-Based User Profile in Social Media", Hindawi Publishing Corporation, The Scientific World Journal, Volume 2014.

[15] Niharika Pujari, Debahuti Mishra, Kaberi Das," Query Reformulation: Data Integration Approach to Multi Domain Query Answering System", Int. J. of Computer and Communication Technology, Vol. 2, No. 1, 2010.

[16] Madhusubram D, Shantharajah S.P," A PREDICTIVE USERS PREFERENCE IDENTIFICATION TECHNIQUE TO IMPROVE WEB USERS QUERY RELATIONAL PROCESSING", Journal of Theoretical & Applied Information Technology, Vol. 67, Issue 2, 2014.