IJCTA, 9(36), 2016, pp. 281-289

Search Result Based Web Personalization Using XML Data with RPSO Clustering Algorithm

Ahmed Mudassar Ali* and M. Ramakrishnan**

Abstract : Internet is seen to have acquired the status of essential element in the lives of people. Applying the process of data mining for discovering knowledge out of data from web such as the web pages, hyperlinks, and web documents is called as web mining. Problem with the existing system is that execution time is more for the grouping process while the accuracy of retrieval of document from web is also not efficient. In addition to these two issues, the present system does not also focus on the role or domain-based search. Apart from these, search results retrieved are also not annotation result or according to users preference. The above stated are the primary issues with the present system which makes use of XML database. For overcoming the said issues, in our analysis, we have proposed a Partial Swarm Optimization (PSO) technique for grouping process. In addition to the process of clustering, this study also suggests three different types of processing that have been listed below:

Registration using user role, Extracting XML data by making use of tokenization, Clustering process using enhanced annotation label. Among these, in the first procedure, role will be defined prior to retrieval of search result out of the XML database. Secondly, this will help extract data out of the XML database by tokenization, and finally, group the (SRR) Search Result Records. In the process of clustering, focus is on two vital factors, the first one being attributes- that is, keywords, and the second one happens to be user role. Moreover, the process of rating-based ranking has been included for retrieving user preference results. In our proposed system, the time for the process of clustering is better when compared with the present system. In the end, we have also proposed one ranking procedure for the user preference.

Keywords : Role-PSO clustering, tokenization, data keywords, user Role, Rating based Rank, Annotation.

1. INTRODUCTION

It is well-known that a major portion in the vast web happens to be data-based [1], meaning that regarding several search engines, the data encrypted in returned resultant pages will come from underlying databases that are structured. Such kind of search engines is referred often to be web databases (WDB). One typical search result page that is returned through a WDB will have multiple search result records (SRRs). Every SRR will consist of manifold data units and every ingredient which will describe one particular aspect pertaining to some entity of real-world. The demand for gathering interested data out of the multiple WDBs is seen to be high [2]. It is getting difficult to run web search as it is tough for the users to build queries which need to be both sufficiently discriminating and sufficiently descriptive for finding only those particular web pages which prove relevant to users search objective. Unclear queries will result in search result groups that contain different page sets which meet distinct user search objectives. Each user will have a particular objective while searching for data by entering the keyword queries on any

* Research Scholar, Bharath University, Selaiyur, Chennai.

** Professor, School of Information Technology, Madurai Kamaraj University, Madurai, India.

search engine. The keyword queries happen to be intrinsically unclear but are often devised when user is engrossed in certain huge task [3]. For filtering out inappropriate results, users need to refine the search through modification of the query. Users need to understand result set for refining queries in effective manner, but that may prove to be time-consuming, in case result set turns out to be unorganized [4]. The process of Mining may be defined as extracting and integrating useful data, knowledge, and information out of the contents of web pages [5]. In the present system, data unit happens to be one text piece which will semantically represent one particular concept pertaining to an entity. This will correspond to value of some particular record under some attribute. It differs from the text node that will refer to some sequence of the text that is surrounded by an XML tag pair. It will describe relationships between data units and text nodes in detail [6]. Data annotation and extraction of web data falls are stored in a web database. Data annotation and data extraction problem, that is, assignment of labels that are meaningful to extracted information unit related to every SRR is one challenging task. Analyzing or annotating voluminous data in single website might reduce processing speed. We have proposed PSO Clustering strategy and user role-oriented process of query for retrieving the data in SRR in the early stages. Users perform the process of registration for identifying which particular role user will be accessing in XML database and will access that kind of information in SRR. This will be benefit by avoiding more result records information while time consumption is low. Also, in the process of extraction, user can run search the data by giving query. Our study proposes PSO clustering strategy to achieve enhanced speed of clustering and also accuracy in retrieving data. This process of clustering will make use of both user role and keyword. It will first cluster using the user role and then it will consider the keyword related to details of the users query content in the grouped user role records. Our paper suggests additional use of the process of annotation that will fetch quality data present in the clustering results. Such results produce sequence and label alignment. Here, the label having the alignment process makes use of rating-based ranking technique; it will display results according to the user preference while fulfilling expectation of the user in data retrieval.

2. RELEVANT WORK

Internet searchers have long since turned toward the most crucial entrance for the traditional individuals who search for vital information on web [7]. Custom web appearance is one common class of quest processes going for producing improved list items that are specifically custom-made to satisfy individual user needs. Since it proves costly, user data needs to be collected and analyzed for making sense of aim of the client behind the submitted inquiry. Results for PWS may be sorted generally into two types, namely, profile-oriented methods and snap [8]. The aim of our suggested approach is mining a search result that is effective and reduced time thus enhances the experience of searching. User's personal long-term dynamic profile is recorded and maintained on the basis of user search, and we make use of it for personalizing. For solving cold start issue, we make use of ontology in the client side. It is also useful in expanding a query and generating clusters with similar results. Users profile may be stored in the form of one weighted ontological tree. The web search inferences from present search engine may be taken and de-ranked in accordance with the client's profile [9]. The mining process of Web Content normally consists of mining process of the particular content that is held within particular web pages. Our papers aim is evaluating, suggesting, and enhancing use of the progressive web information clustering methods that are used highly with advent of need for mining huge content-oriented data sets which allow information analysts to carry out more effective performance of voluminous web information searches. In any search space, information will be available in some random style that may end up in trafficking when searching several times. Hence, in our paper, we have provided an enhanced algorithm that helps reducing search space within the search engines by making use of clustering methods [10]. By making use of this analysis, we wish to reduce various crises that exist in web pages. Fundamental point in our study is assessing, recommending, and developing the function of web information clustering methods mainly employed by specific arrival of mining voluminous content-oriented data sets allowing information analysts to carry out more effective implementation of voluminous web information search. Information prevails inside search space under

random style, thus causing trafficking at the time of being searched several times. Subsequently, in our proposed paper, we have made available an enhanced algorithm that might minimize search space by making use of clustering methods in the present search engines [11]. This study furnishes an automated annotating method that will first align data units found in any result page as different clusters similar to how information present in same cluster will reside in same semantic. Next, regarding this group, we will annotate the same from different factors and will aggregate the various annotations for predicting one final annotation label. Annotation wrapper pertaining to given search site will be constructed automatically and may be used for annotating fresh result pages out of the similar web databases. Here, our test results have proved very effective in the suggested system [12]. It will help the world in increasing useful information. The complete data has become web-based usage of HTML form-based interfaces of search.

3. PROPOSED METHODLOGY

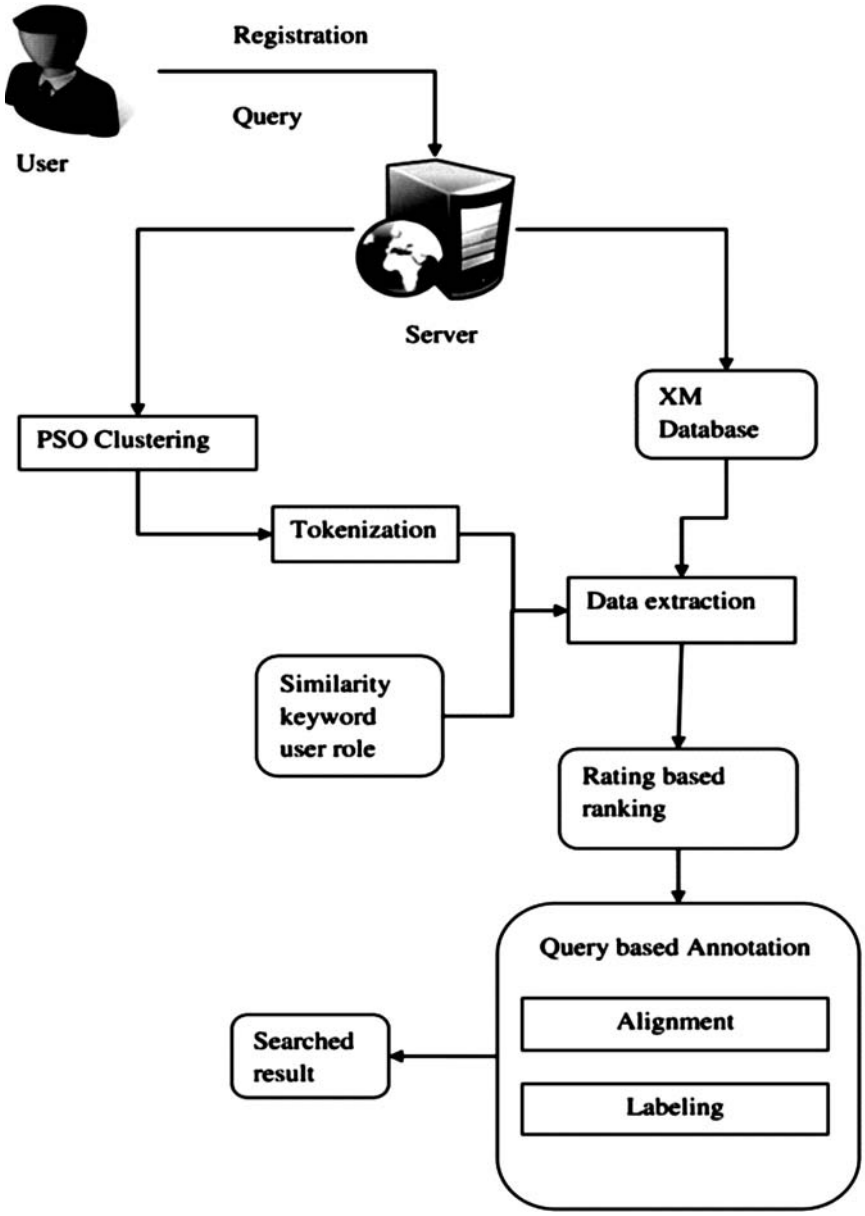

Figure 1: Overall Architecture

We have proposed in our paper keyword and a user role-based grouping approach for getting user-preferred SRR information in the early stage. User role group process is grouped individual user role; so, the given results will produce that particular kind of SRR results. The search result record information is grouped

based on similarity of keywords and label value. This kind of grouping strategy will require reduced time and will yield quality results, hence this study suggests PSO clustering program regarding faster grouping and low time consumption. SRR group results are being taken and are given to user preference by making use of the process of annotation. Such annotation alignment information is rating-based ranking having display and each with similar information items. These pages will function as containers for the grouped extracts related to results of search.

A. XML Query

Upon XQuery fetching the information, they get clustered together on the basis of their similarity, meaning, through pairing of keyword with the XML content and thereby binding together the contents that are similar. This is achieved by making use of XQuerys capability. The complete related records can be extracted out of the authentic XML document having the structure possessed.

B. PSO Clustering Process

Initially, we got web pages of the lists of the search result that are returned by any search engine on the web. We have derived the results from the Search Result Records (SRR). Therefore, the initial search happens to be a conventional Meta search that is based on keywords. Such web pages then get analyzed by one XML parser, and the relevant result items will be then extracted. Normally, there are query-based snippets and keywords that are available in every result item. Here, we presume that these contents prove informative enough as most of the present search engines have been designed for facilitating users related judgment by keywords only, hence it will be possible to provide contents that are the most relevant, in relation to any given query. Every extracted phrase will in fact be a candidate clusters name that will correspond to a group of documents which consists of the phrase. Our system will first identify user roles and the meaningful group labels and after that only assign the search results on to those labels for building the appropriate clusters.

C. Rating Based Ranking

Ranking can be defined as the relationship among an items set so that pertaining to any two given items, first one will be ranked lower than, ranked higher than, or else, ranked equal to the second one. It is called weak order in Math, or sometimes, complete preorder pertaining to objects. In essence, it may not be a complete order pertaining to objects, since any two disparate objects may have similar ranking. Rankings themselves get ordered totally. In our study, we have proposed to rating-based ranking by using higher rank preference data to be shown in SRR.

D. Query-Based Annotator (QA)

The fundamental concept behind this particular annotator happens to be that returned SRRs out of a WDB will always be related to a specified query. Particularly, query terms that are entered in search attributes in a local search link pertaining to a WDB will certainly be appearing in any of the retrieved SRRs. A query-oriented annotator works in the following sequence:

Considering any query having a group of query phrases that have been submitted on one attribute A in a local search link, initially identify the set that has got largest total instances of those query terms. Next, allot gn (A) as label for the cluster. As already mentioned, roles of any XML database will normally have all attributes pertaining to the latent database. And as a result of this, the query-oriented annotator will totally annotate the SRRs.

(*i*) Labeled

When one data has been aligned, each cluster will need to be paired with any explanatory keyword describing what it consists of. This can be accomplished by making use of labeling. Labels need to be relevant and perfectly appropriate to information items; they are selected according to the category that

denotes data items presented. In this manner, an XML page may be created and can be referenced according to the group naming related with the given label allotted. Therefore, in the end, output will be those groups which serve as XML references.

(*ii*) Alignment

One automatic annotation has been presented in our study. This initially aligns data units in any particular search page into various clusters in such a way that information inside the same cluster will be having similar semantic. Next, for every cluster, we annotate that from various aspects and then aggregate the various annotations for predicting one final label of annotation for the same.

E. Algorithm

Algorithm1: Role Based Clustering Framework

Input : A set of search results SRs returned from web

Output : Clustering of SRs into clusters C.

1. **Initializations :** C = 0; uK [*x*] = 0, uR [*y*] = 0 where *x* and *y* is an iterative variable; *x* = N and *y* = N where N is the initial assumptive of clusters C; index = 0.
2. uK = user Keyword; # user input
   uR = user Role; # user profile
3. *ds* = get Result Set (uK, uR)
4. Hash Map obj = result [N];
5. Repeat
6. For each *x*€uI and *ds* is null
7. Do
8. Query uK against the item *x*;
   Query uR against the role *y*;
9. Update *x*, *y* in Hash Map <obj>
10. End for
11. *x*++, *y*++;
12. Until *x* < = N && *y* < = N
13. Return *ds*

Algorithm 2: get Result Set (uI)

Input : search query in natural language

Output : Web search results:

1. **Initializations :** uK = null; uR = null ; builder Factory = instance; result [N] = null;
2. Repeat
3. X* = tokens of uK, uR;
4. For each *i*€uK & uR [N];
5. Query *x** *y** at every level;
6. Update the Result [*i*]
7. *i* ++;
8. End for;
9. Until *i* < element. length;
10. Return result [N];

Algorithm 3: Rating based ranking (Annotation)

1. Rat[*a*] = 0; a is iterative value; Ran = 0; *d* = 0; *d* is a data value;
2. *a* = 0;
3. if *d* < higher(*a*);
4. *a* ++;
5. until *d* = null;
6. until Rat [*a*]=null;
7. Ran = higher (*a*);
8. Return result [Ran];

4. RESULT AND DISCUSSION

Table 1 contains datasets of search result records of web database. It has some datasets with different search result records, no of terms in the record and no of cluster of search result records.

Table 1

Search Result Record datasets

Data	*Search Result Records*	*Number of terms*	*Number of clusters*
Dataset1	414	6429	9
Dataset2	313	5804	8
Dataset3	204	5832	6
Dataset4	878	7454	10

Table 2

Accuracy and Execution Time for Proposed Method

Methods	*Accuracy (%)*	*Execution Time (Seconds)*
XQuery Cluster	91.5	55
Role-PSO cluster	95.0	35

The Table 2 shows the accuracy and execution time for XQuery and Role-PSO. It is clear from the above table the proposed approach Role-PSO gives improved accuracy than existing approaches.

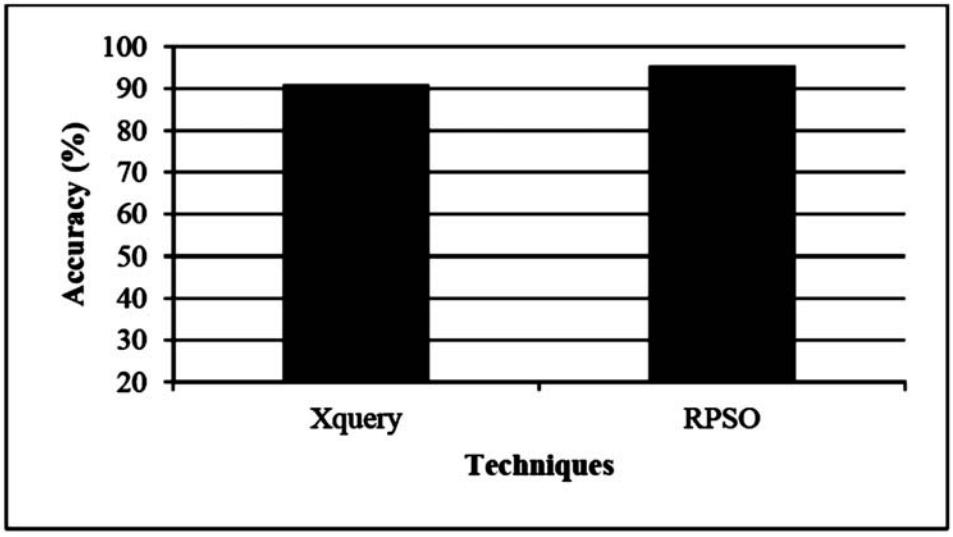

Figure 2: Accuracy For Proposed RPSO

Figure 2 shows the accuracy for XQuery and RPSO. The proposed methods of RPSO have high accuracy than other two methods.

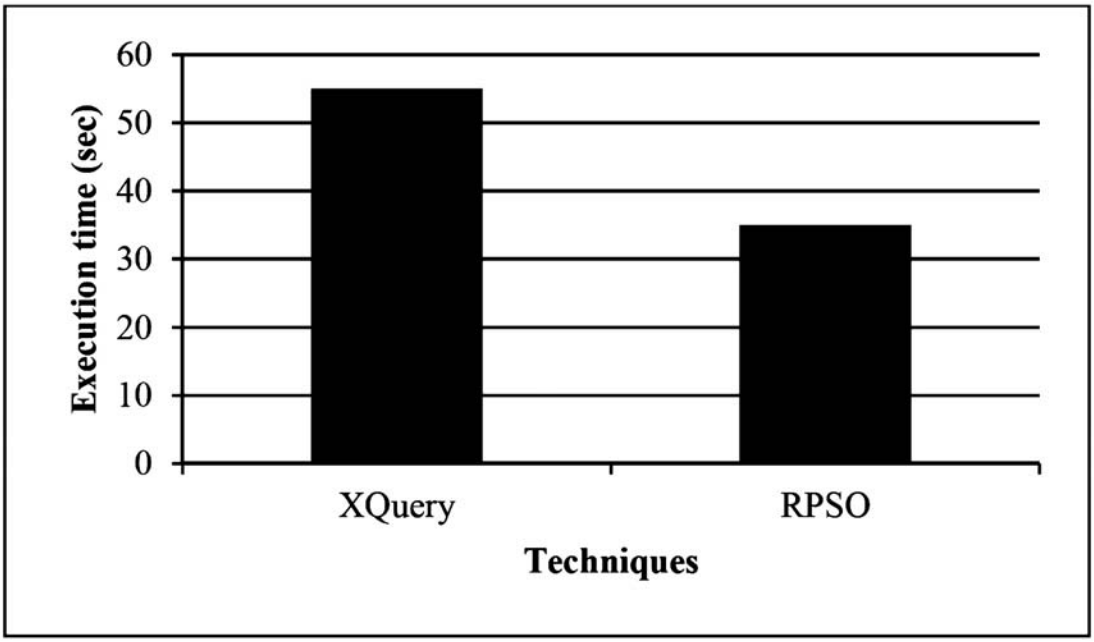

Figure 3: Execution time For Proposed RPSO

Figure 3 shows the execution time for proposed method Role-PSO. Proposed method has less execution time *i.e.*19 (sec) when compare with XQuery and Role-PSO.

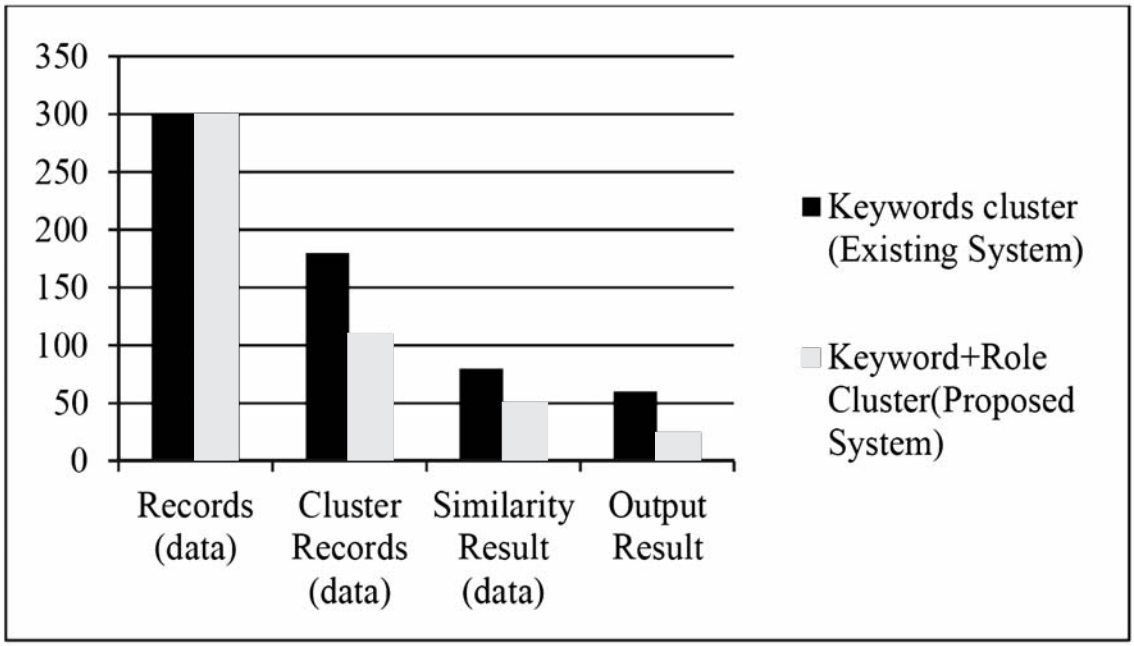

Figure 4: Overall comparison of Existing and proposed work

Figure 4 shows the comparison process of existing and proposed work. Here proposed system performs better than existing system in process of Cluster, Similarity Result.

Figure 5 explains annotation process in proposed system. It performs well in the area of Label and data alignment when compare with existing system and also it takes minimum time and provide accuracy result.

Figure 6 shows accuracy result of proposed techniques. This technique provides better result for xml database search engine.

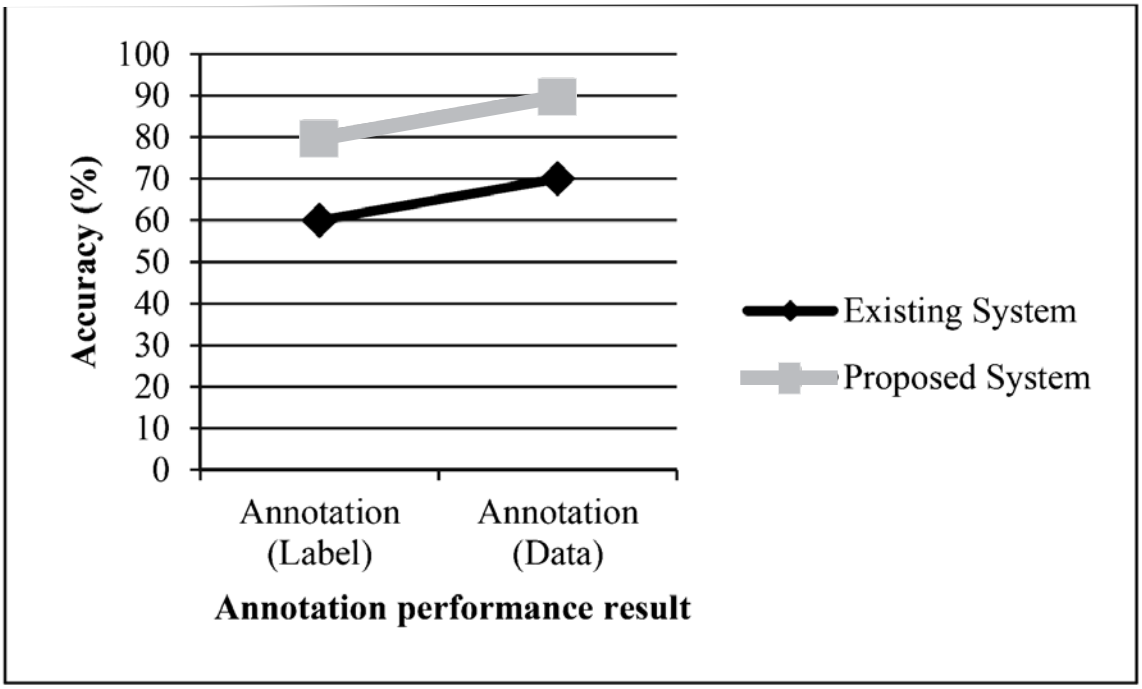

Figure 5: Annotation performance result

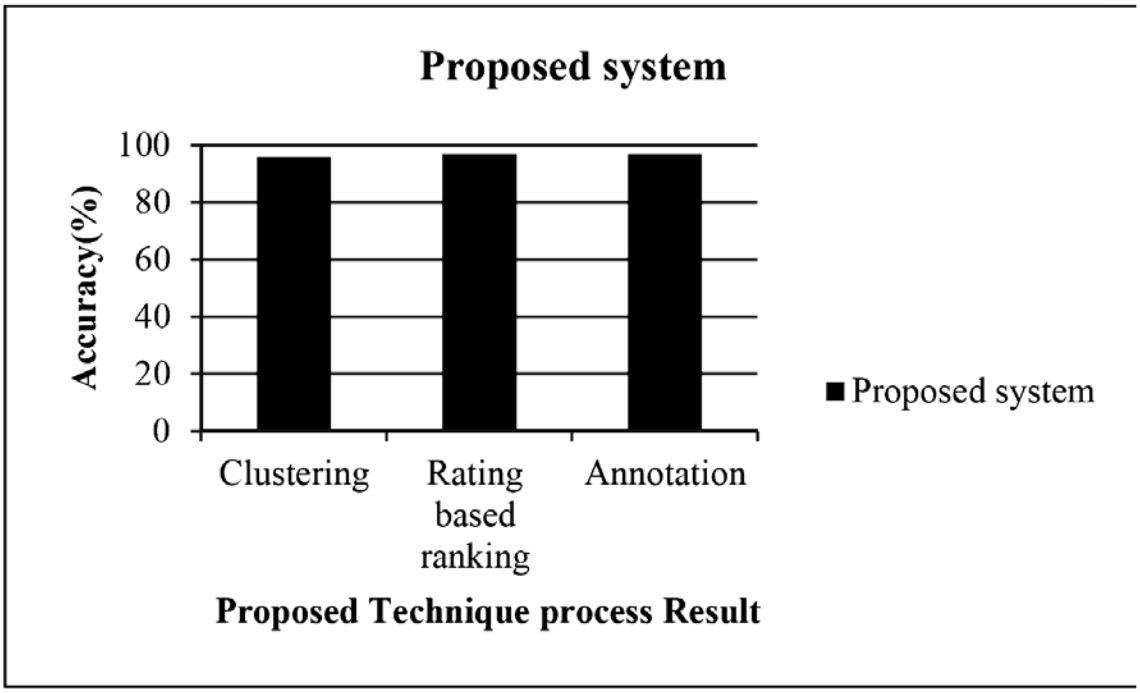

Figure 6: Result of Proposed Techniques

5. CONCLUSION

Our study proposes a Role-based grouping strategy for increasing the speed of clustering and the results have proven to be accurate SRR information in WDB. In the present techniques, query group process is done under keyword-based clustering; it takes longer time and grouping more than the number of record results. In order to solve this issue and improve the output and for getting efficiency toward user preference query, this study proposes an improved annotation alignment. Moreover, when the result from our result value assessment is compared with that of the existing system, our proposed system brings better results.

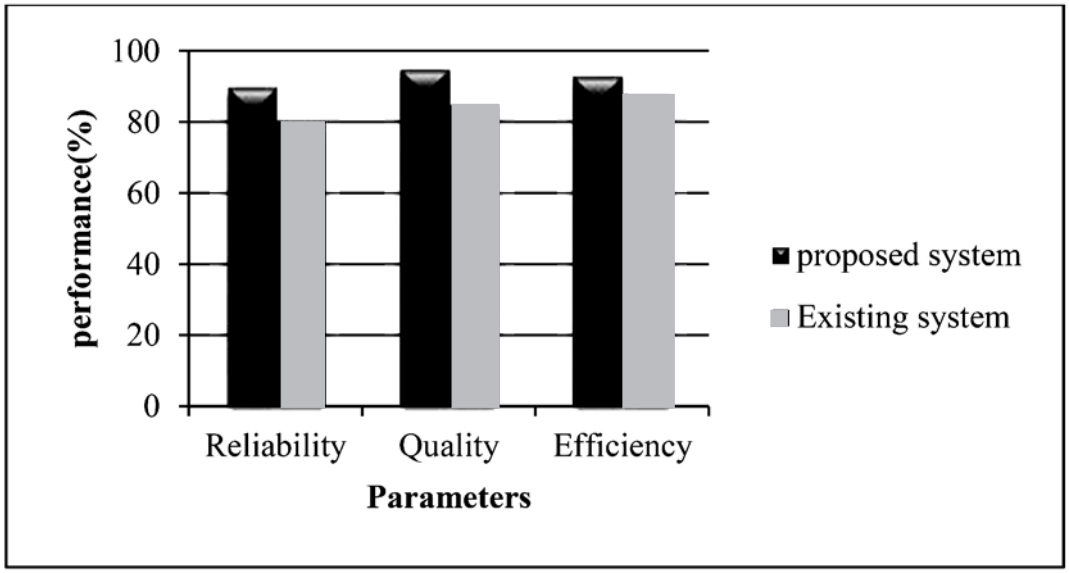

Figure 7: Parameter Results

Figure 7 shows the finally displayed output result which is better than existing technique.

6. REFERENCES

1. Saranya.J, SelvaKumar.M, Vigneshwaran.S, Danessh.M.S, "Annotating Search Results from Web Databases Using Clustering-Based Shifting", February 2014.
2. Kavita D. Satokar, A. R. Khare, "Search Result Clustering For Web Personalization", Sept 2011.
3. Ahu Sieg, Bamshad Mobasher, Robin Burke, "Feature Article: Learning Ontology-Based User Profiles: A Semantic Approach to Personalized Web Search" November 2007.
4. D. Satokar, A. R. Khare, Kavita, "Search result clustering for web personalization", International Journal of Advances in Engineering & Technology, Sept 2011.
5. Challenges and Surveys in Key Management and Authentication Scheme for Wireless Sensor Networks" in Abstract of Emerging Trends.
6. Boraste Jagruti Balkrishna Swati.K.Bhavsar, "Annotation Based Web Databases Search Technique", International Journal of Computer Science and Information Technologies, 2015.
7. Cluster based Key Management Authentication in Wireless Bio Sensor Network ", ,International Journal of pharma and bio sciences
8. Avinash A.P, Prashant M. Narayankar, Muniraj M, Jeevitha ,"Clustering method: search result based web personalization", 2012 .
9. Singh Bisht1, Prof. (Dr.) Sanjeev Bansal , "Optimization of Web Content Mining with an Improved Clustering Algorithm Solvers", November 2013.
10. Arti Sharma , Dr. Shashank Awasthi, "Improved Algorithm for Optimization of Web Content Mining", March 2015 .
11. Prof.J.Y.Kapadnis, Akshay.R.Gore, Hitesh.N.Sonawane, Gokul.S.Itnare, Ravindra.K. Balak, "Annotating Search Results from Web Databases using Alignment algorithm & different Annotators". Apr-2016.
12. Prof. P. S. Badgujar, Shagufta Pathan, Aayesha Sayyed, Salina Ansari, "Annotating Search Results from Web Database", 2015.