A New Framework for Construing Search Results

Togaru Vinay Kumar1, S Arvind2, M Janga Reddy3

1#Pursuing Mtech from CMRIT, Affiliated to JNTU, Hyderabad.
2#Working as HOD, Department of Computer Science & Engineering at CMRIT, Affiliated to JNTU, Hyderabad.
3#Working as a Principal at CMRIT, Affiliated to JNTU, Hyderabad.

Abstract— In current day almost everything is done online where the users can perform the tasks remotely. In addition to the online data distribution it has become very difficult to keep track of things together. This paper proposes a scheme where the data is segregated and annotated based upon the user intension. All the search history is gathered and segregated into categories based on the annotations given. Based on these the user can search and utilize the data very efficiently and can save time as the search performed purely on the user provided query. This technique is much better than compared with the existing system i.e. the results are shown or accessed through form-based HTML interfaces.

Keywords — Annotations, Segregated, Interfaces.

I. INTRODUCTION

This work resides on the platform “Data Mining”. From the name itself it is very clear that the proposed work mines the data from the large database to the user for further processing and better understanding. Data Mining is one such domain where when different mining algorithms employed results to kind of situation that it becomes very easy for every individual to do analysis. Data Mining is a subfield in Computer Science; it is the computational process of discovering patterns in large data-sets involving manipulations of machine learning, artificial intelligence, database systems and statistics. Apart from the data extraction it involves data management, database aspects, data pre-processing, inference and model considerations, complexity considerations, post-processing of discovered structures, online updating and visualization. The actual task is the automatic or semi-automatic analysis of huge quantities of data to extract previously anonymous interesting patterns such as groups of records, unusual sets and dependencies. This usually involves using database techniques such as spatial indices. These patterns can then be seen as a kind of summary of the input, and may be used in further analysis or, for example like, in machine learning and predictive analytics.

Fig. 1 Data Mining Process

Most of the websites available today are returning results to the users only based upon the search performed. The major drawback with the system is that the application is just doing the task of searching the content from the database and showing it to the user. In the proposed work, the application will categorize the data and based upon the user request it will fetch the results and annotate them for the user future purpose to save time based upon the fetching time. As it is understood that most of the applications in current day scenario are content based and not user specific based, so the proposed work must focus on the search mechanism. We can have a search mechanism which will annotate the search based upon the user intension. In addition to the annotation, if the results are sorted as the demand of the users then the application is going to have more significance than any other existing application. For having the above
dedicated output the application need to have two tasks to be performed one after the other i.e. search using annotation and once the search part is completed ranking the results using ranking algorithm.

II. RELATED WORK

The proposed work will have to do two mechanisms one after the other. First would be segregating the search results using the annotation technique and then ranking technique to display the results required by users for saving user search time. In case of searching the data from large set of database, we need to employ various techniques like, Data unit similarity, Data content similarity. Both of these techniques are for aligning data and showing the relevant results to the user.

Data Unit Similarity: The purpose for aligning data is to put data units of a similar group into one unit such that they can be annotated for showing the results in group wise manner. Similarity is checked for the search type and the database type indexes i.e.

$$Sim(d_1, d_2) = SimC(d_1, d_2) + SimU(d_1, d_2)$$

Where in the above equation we can see that the similarity is judged based upon two factors, similarity in common for the searched results and similarity to be searched in the database. All the filtered data will be shown to the user and will be purely an annotated result as the results are shown purely based on the user specific search request.

In this manner, each result fetched from the database will be compared with the other data values to make sure that the user is searching and getting only the relevant data that is useful to the user and thus saving time.

To have a search action in a sorted manner label should be added to the data sets like categorizing the search results based upon the label as shown in the below block.

<table>
<thead>
<tr>
<th>Label</th>
<th>Book</th>
<th>Author</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Book</td>
<td>Blackbook</td>
<td>Nageshwar</td>
<td>1.0</td>
</tr>
<tr>
<td>Book</td>
<td>Orient book</td>
<td>R Long</td>
<td>1.1</td>
</tr>
<tr>
<td>Book</td>
<td>Creative Sears mill</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Book</td>
<td>Black Book</td>
<td>Sathosh</td>
<td>1.1</td>
</tr>
</tbody>
</table>

From the above table we can understand that the search operation on this table will be done based on the label. As the label is book for the data mentioned above in the table, categorizing will be done as per the label specified. This way many such links can be provided to have a database with huge dataset in it. This type of annotating search result can be referred to as “Query Based Annotator”.

Coming with the second technique applied in the proposed work for the ranking search results. By doing so the user search time will be reduced and thereby increasing the search efficiency of the application and thus giving 90% success results to the users based on their query posted to the search engine. Ranking algorithm involves the technique of mining the user search keyword and then the user intension. Based on these two conditions the application will calculate the rank for specific search and mine it in the database.

Ranking will be done purely based on the user specific search. For example, consider a user wanted to know about Java and has type “java material”, for which the search engine gives out some links from where the user can gain the knowledge about that domain. So, user mandatorily clicks on the specific link and reads the information. In this manner every user with the search type “java material” will get similar links but every time a user getting added in the search operation the search results will be varying as the results will be shown purely based on the number of users clicking on a specific link. From this it is understood that users will click the link which will give them better information about the requested resource. The rank will be calculated purely based on the number of visits to a specific link. In this manner, the application will show the results to the user by
taking two algorithms into consideration. Both the algorithms have equal importance in their own manner. New approach added in this paper for the maximum benefit to user is done based on the Ranking algorithm.

Ranking algorithm works in the below manner,

```java
Return rank(int key, int[] a) {
    int le = 0;
    int ma = a.length - 1;
    while (le <= ma) {
        // Key is in a[lo..hi] or not present.
        int mid = le + (ma - le) / 2;
        if (key < a[mid]) ma = mid - 1;
        else if (key > a[mid]) le = mid + 1;
        else return mid;
    }
    return -1;
}
```

III. CONCLUSIONS

Proposed work gives out a clear cut idea about the search mechanism. Search optimization is very important task in any Data Mining project, any application or work that saves much of the user time in searching the data which user requested believes to be the best suited search engine. Work must involve different mechanisms to fetch out the data which is relevant to the user and must give maximum information from it. Likewise this paper composes two search mechanisms to give out the best output to the user i.e. Annotating search results and then applying the ranking algorithm. Both including improved the search efficiency when it is just compared with the annotation alone. Thus, here it concludes showing the best and suited search optimization engine for better and efficient results.

REFERENCES