International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 04 | Apr42018 www.irjet.net p-ISSN: 2395-0072

Highlighting Document Streams Using Multi-Regional Input Output (MRIO)

Muthamil Selvi N[1], Muthubhavani K[1], Nandhinidevi S[1], Dhivya S[2]

[1,2] *Department of Computer Science and Engineering, K.L.N College of Information Technology*

Abstract-*The processing of document streams in emerging applications, such as news update filtering and social network notifications, demand presenting end-users with the most relevant content to their preferences. The main objective of this project is to support large number of people and to consume memory by highlighting the keywords, which are preferred by user.A stream of documents flows into a central processing server, which hosts a set of Continuous Top Query Documents.User preferences are indicated by a set of keywords .A central server monitors the document stream and continuously reports to each user, the documents that are most relevant to keywords. The user downloaded documents are get ranked using ranking algorithm. Based on that ranking the documents are get listed to user. The user view the interested documents and then in a fraction of seconds the keywords presented in a documents are get highlighted .The user can download the documents by sending request to the admin. The OTP get generated using RSA algorithm .The admin response to the user request by sending OTP to the Users mail-id. Using OTP the user can download the document.*

Keywords—Highlighting, Document stream, Ranking, Downloading.

I. INTRODUCTION

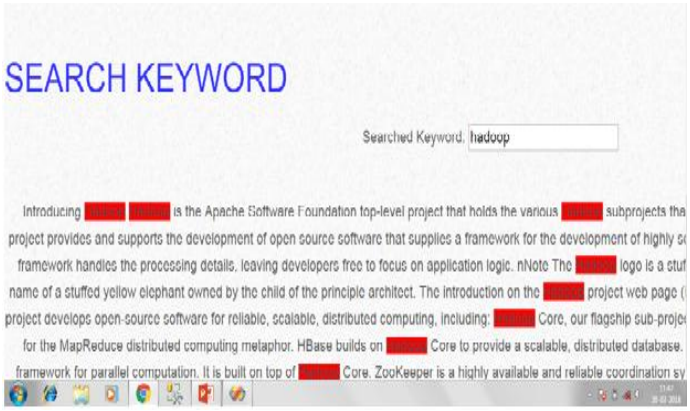
Highlighting document streams using MULTI REGIONAL INPUT OUTPUT (MRIO) has been originated from continuous top-k monitoring document streams. Our objective is to support large number of users and to consume memory by highlighting the keywords, which are preferred by user. In a database the admin upload the documents. On user side, the user have to register with mail-id. Admin send password to registered users mail-id. Using the password the user login to the account. User preferences are indicated by a set of keywords. The keywords are get stored in the cookies for highlighting purpose. Using matching algorithm the user preferred keywords are matched with file name in the database. A central server monitors the document stream and continuously reports to each user, the documents that are most relevant to keywords. Using Multi Regional Input Output technique the documents are get bounded from multiple region and list to the user. The documents list based on the ranking of user downloaded documents. From the list of documents the user can view the interested document. When user view the documents, then in a fraction of seconds the keywords presented in a documents are get highlighted. After viewing the documents, if the content of the documents are needed to the user, they can download the documents by sending request to the admin. The admin response to the user request by sending OTP to the Users mail-id. Using RSA algorithm the OTP get generated. Using OTP the user can download the document. The downloaded documents are get ranked using ranking algorithm. User can view the statistics of top five documents that are downloaded by user.

II. INDEX & QUERY PROCESSING

Upload the document files in MySQL database .The database holds the list of an entry for every document that includes keywords. Then all the document lists are sorted in ascending order of document ID.The task of the admin is to update all query results when new documents arrive. Admin maintains the registered user's information.

III. REVERSE ID-ORDERING FOR CTQD

When a new document arrives, we need to update the index and re-evaluate each and every query. The user given keywords are matched with file name. The keyword related documents are bounded from multiple regions. Listed files are reverse id ordered based on ranking of user downloads.



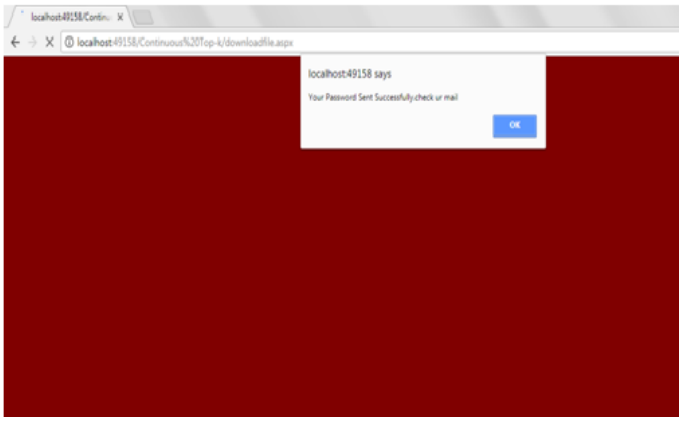
IV. MINIMAL REVERSE ID-ORDERING

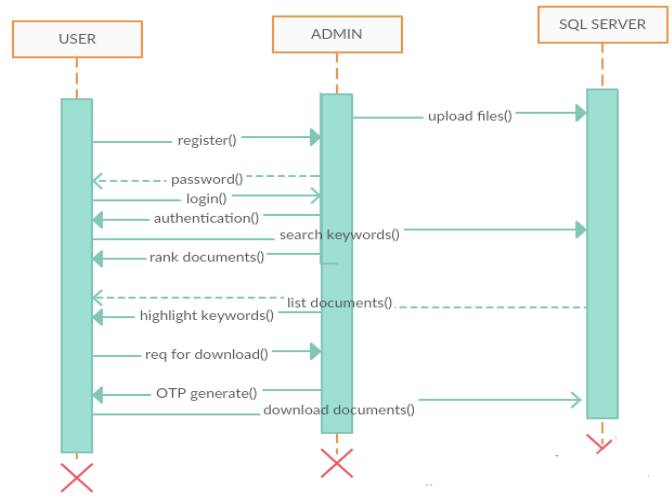
Minimal Reverse ID-Ordering is used reduce iteration in Reverse ID Ordering process. The user queries are stored in cookies. The queries which are presented in the documents are get highlighted by using jQuery. User can identify the documents which are topmost downloaded documents by then registered users.

V.FILE ACCESS PERMISSION

If the user interested on the content of the file they can make request for the file to the admin. The admin send one time password (OTP) to the user mail ID. Using that OTP

they make download on the interested file. The OTP was generated using RSA algorithm.



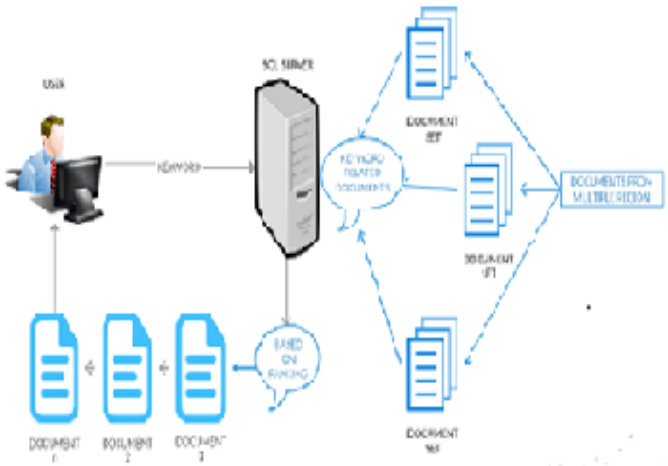


VI. EXISTING SYSTEM

In traditional text search the documents are listed based on user view. It comprises a list for every term in the dictionary; the list for a term holds an entry for each document that contains the term. The documents are in the form of snapshots for only 1 to 3 pages. User can able to download only the PDF files and they can only copy and paste the contents present in the document. The content of the document exceeds user capacity to discover and understand the information. The user can't able to download the notepad and word files.

VII. PROPOSED SYSTEM

Multi regional input output (MRIO) technique are used to search the documents from various region. Keywords related documents are listed based on ranking of downloaded files. The search keyword are stored in cookies to highlight the document streams by using pattern matching algorithm. Ranking algorithm are used to rank the downloaded files. RSA algorithm are used to generate the OTP to download the documents. The efficient filtering and monitoring of rapid streams is key to many emerging applications. By sending the request to admin the user can able to download the interested documents.



VIII. RELATED WORK

In information retrieval, the objective is to retrieve the information based on user search keyword. The keyword matched information content are screen to the user. Document clustering techniques face an important problem that is the threshold value determination During document clustering, threshold value determination based on content of the documents in the cluster. Threshold value can be calculated using cosine similarity between the document in the cluster and its representation. Publish-subscribe is a message pattern where the publisher of messages categorize their message into classes, and the subscribers receive only those messages that fall in their classes of interest.

IX. CONCLUSION

. An analysis on RIO reveals that the key factor that determines its performance is the number of iterations it executes. This motivates our approach MRIO, which not only reduces the number of iterations, but is proven to minimize it. Our Future enhancement is to implement the file database in cloud to increase the size of the dataset. To include images, video in the documents.

X. REFERENCES

[1]Leong Hou U and Junjie Zhang," Continuous Top-k Monitoring on Document Streams" IEEE Transactions on knowledge and data engineering, vol. 29, no. 5, may 2017.

[2]K. Mouratidis and H. Pang, "Efficient evaluation of continuous text search queries," IEEE Trans. Knowl. Data Eng., vol. 23, no. 10,pp. 1469–1482, Oct. 2011.

[3] P. Haghani, S. Michel, and K. Aberer, "The gist of everything new: Personalized top-k processing over web 2.0 streams," in Proc. 19thACM Int. Conf. Inf. Knowl. Manage., 2010, pp. 489–498.

[4] K. Mouratidis and H. Pang, "Efficient evaluation of continuous text search queries," IEEE Trans. Knowl. Data Eng., vol. 23, no. 10, pp. 1469–1482, Oct. 2011.