ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

Content Based Filtering In P2P Lookup Service

Lavanya.M
PG Student,
Priyadarshini Engineering College, Vaniyambadi, Vellore-635751
Lavimani11@gmail.com.
Vanathi.A
Assistant Professor,
Priyadarshini Engineering College, Vaniyambadi, Vellore-635751
Vausha.2006@gmail.com

ABSTRACT:

This paper study about keyword search function using peer-to-peer network system. The key of MTAF (mandatory traffic adversary frequency) used to carefully select bunches of terms without incurring negatives and to forward the content items towards selected terms of lower content. Users to create subscription filters based on the keyword search. In similarity based replication of filter used to mitigate the effect of hotspot s these application that arises some fields are more popular than others. Unfortunately these application suffer from lot of website in our system we efficiently improved performance and searching technique of hotspot terms

I. INTRODUCTION

The potential of Peer-to-Peer (P2P) technologies for building distributed applications at a very large scale has been commonly recognized. Existing P2P systems such as Vuze, Bit torrent and eMule connect millions of machines to provide Internet-scale content sharing and keyword-based content searching services. This is due to the desirable properties of scalability, fault tolerance, short routing paths and anonymity protection by distributed hash tables (DHTs)and P2P networks A significant body of research work has been dedicated to studying keyword search and dissemination in DHT.

We leverage the publish/subscribe (pub/sub) fashion to design a scalable keyword-based content alert mechanism, called MTAF. Similar to Vuze subscription, MTAF offers the functions of filter subscription and content alert. Nevertheless, when fresh content is available, MTAF forwards the associated metadata information (consisting of a set of keywords to describe the raw content) and match it with filters. If matched filters are found, MTAF then timely notifies subscribers of the fresh content

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

a) Distributed Hash Table

A number of DHT-based architectures have been proposed in the literature that match the content with queries (and filters) based on keywords. The main element of these architectures is that the implementation utilizes the key-to-node mapping of the DHT to designate one of the peer nodes as a home node for each content term.

b) Information filtering

Information filtering is a name used to describe a variety of processes involving the delivery of information to people who need it. Although this term is appearing quite often in popular and technical articles describing applications such as electronic mail, multimedia distributed systems, and electronic office documents, the distinction between filtering and related processes such as retrieval, routing, categorization, and extraction is often not clear.

c) Information Retrieval

Probabilistic retrieval models compute , which is the probability that a user's information need is satisfied given a particular .Objects are usually We consider an information need as a complex proposition about the content of an object, with possible values true and false. Queries are regarded as representations of the information need

II.RELATED WORK

The area of information filtering and dissemination (IFD), and could be treated as the plugin of the information retrieval (IR) model into the classic publish/subscribe (pub/sub) paradigm [6]. Therefore, we investigate the literature in the four areas of the keyword-based IR, pub/sub, IFD, and related works of SSBF. First, from the first generation of P2P Napster, the keyword-based IR has been a fundamental functionality, and an important research topic

.

Second, MTAF shares commonalities with the publish/ subscribe (pub/sub) paradigm [6], and our work can be treated a subclass of the pub/sub paradigm. The works [7], [13], [14], [20], [27], [29] could be regarded as instances of a pub/sub application. However, most of these works employ topic-based or content-based subscription semantics.

Third, the centralized IDF has been widely studied in the literature, including the classic centralized work SIFT [32] and In Route [4]. Feed Tree [27] proposed to use a P2P environment instead of a centralized web server to disseminate RSS to subscribers. [13] devised a dynamic clustering method for reducing the maintenance overhead given a large number of RSS URLs. Corona [15] used smart distributed polling and client push messages to inform users of new items.

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

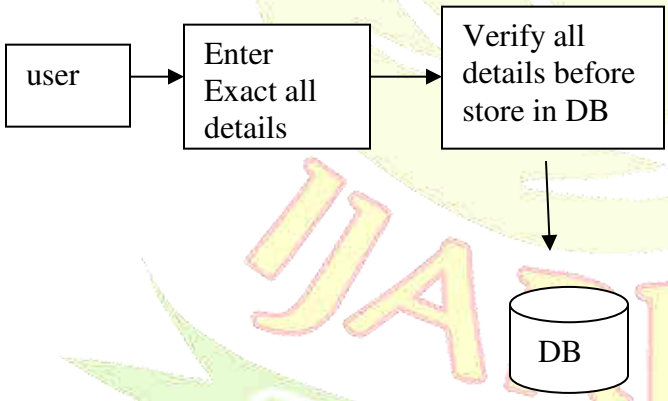
In terms of the related works of SSBF, there are two kinds of bloom filters in the literature to encode groups of members (i.e., sets of sets) [10]. The first approach [5] built a bloom filter for each group of members, and all element of the group are then encoded to the associated bloom filter. Another approach [5], [12] assigned each group with an Id and represented the group Id with some bits. A bloom filter was then assigned to each bit of the group Id, such that log k bloom filters were used to represent k groups.

II.PROPOSED FRAMEWORK

1) User

a) Registration

The User need to enter exacts all the details which are used to enter login page. If registration page successful completion, it will take up to login page else it will remain in the registration page itself. If it is an existing member then it will move to the login page.



b) Login and Authentication

The user need to enter exact Email Id and password which is given in the registration, if login success means it will take up to main page else it will remain in the login page itself. If it is a new user then it will move to the registration page.

c) Choose Search Type

User can select one search type from two type search (Web and Window) engine. If user selects one search type from two type of search, user can enter the one of search module.

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

d) Search and View Selected File

User search select the file based on keyword. The file must be in the following image or video. User can search and view all related files and selected one particular file from related files.

e) Online Watch and Download file

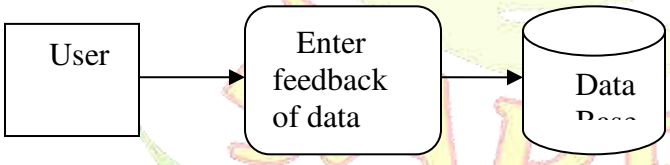
User can see the list of related files. If user select particular file, user can online watch and download selected file.

c) User Upload Video

User can upload favorite video for user compatible search and forward to all searchable member views and download that files.

d) Feedback

User can enter details for feedback. If user can post details with feedback, admin will get user details.



ALGORITHM USED FOR CENTRALIZED PRELIMINARY:

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

Algorithm 1: CENTRALIZED_MTAF (filters $\mathcal{F}$, doc d)

```
1 create a sorted heap H;
2 for each term t_i that appears both in F and d do
3 |   add pair (t_i, |F_i|) to heap H;
4 while H is not empty do
5 |   pick the term t_i in the pair (having the currently largest
  |   |F_i|) popped from H;
6 |   match doc d with all filters in F_i;
7 |   for (each term t_j(≠ t_i) appearing in F_i) do
8 |   |   F_j ⇐ F_j − F_i ⊓ F_j;
9 |   |   update the pair with term t_j in H with new |F_j|;
```

Alg. 1 follows the greedy algorithm of the set cover problem to select a subset of terms ti 2 d. Given such terms ti, the physical node then retrieves the associated positing lists and matches d with the filters Fi. First, lines 1-3 initialize a heap H to maintain the pair of hti; jFiji, where ti is the term in d and jFij is the number of filters in Fi, and the pair popped from H is the one with the largest jFij.

Inside the while loop of lines 4-9, line 5 selects the term ti in H associated with the largest jFij, and line 6 matches d with all filters in Fi. Consider that a filter f 2 Fi might also contain other terms tj, and thus f also appears in Fj. For each such term tj, line 8 removes the filters f 2 Fi from Fj, and line 9 updates the pair having term tj in H by new jFjj. If Fj is empty, i.e., jFjj ¼ 0, the pair htj; jFjji is removed from H. The selection of terms is finished when H is empty.

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

Algorithm 2: DIIT_MTAF(k sets of replicated terms $S_1...S_k$, doc d)

```
1  create k flags m[1...k] with each element equal to 0;
2  for each term t_i that appears in d do
3      for 1 ≤ j ≤ k do
4          if (t_i appears in the set S_j) and (m[j] == 0) then
5              among all terms in S_j, choose a term t_i w.p. 1/|S_j|;
6              forward d to the node n_i and to R_i;
7              set m[j] = 1;
8              break;
```

For only two posting lists Fi and Fj, a positive benefit Bij caused by the merging of Fi and Fj can obviously lead to an improvement in performance (the detail to compute Bij refers to the Appendix A.1). However, the general problem of the merging all posting lists so as to minimize the total cost of matching is NP-hard when the total number of terms (and accordingly, lists) is greater than 2. This claim can be easily proved by reducing the known graph partition problem [9] to the problem of optimal merging. We propose a heuristic solution in an iterative manner as follows.

First, given the initial set of T inverted lists, we compute the pairwise Bij for any two posting lists Fi and Fj (1 ~ i ¼6 j ~ T and jFij 9 jFjj). Among all such pairs, we find the one with the largest positive benefit and merge the two associated inverted lists into a single one. After the first round, we have ðT ~ 1Þ posting lists. We continue this process if there exists a pair of lists whose merging would bring a positive benefit.

2) ADMIN

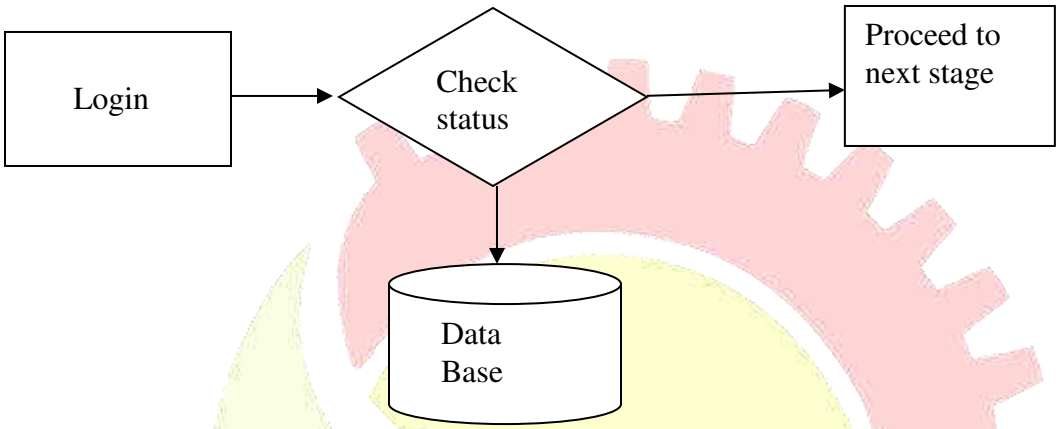
a) Authentication

Admin has to provide exact username and password here. If login success means it will take up to main page else it will remain in the login page itself.

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

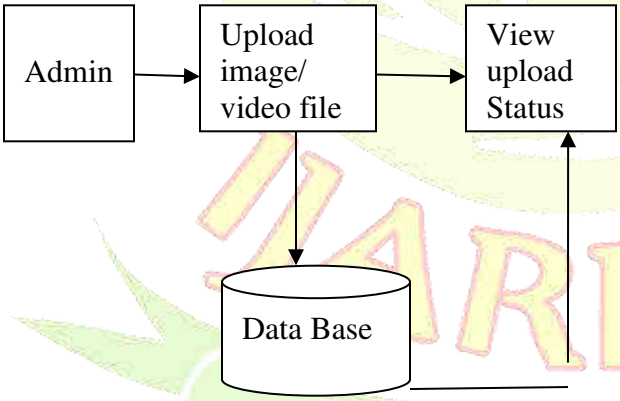
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016



b) Uploading File

Admin can upload image, video for new search of files or events. If user can search new files, admin posted new files will be displayed.



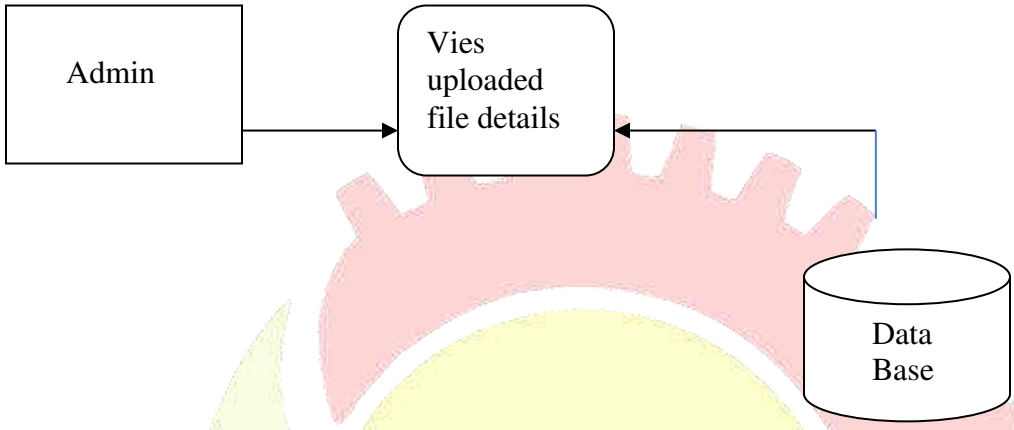
a) View Uploaded File Details

Admin can view uploaded file details. If admin can view uploaded new files status, admin can also delete particular uploaded file.

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

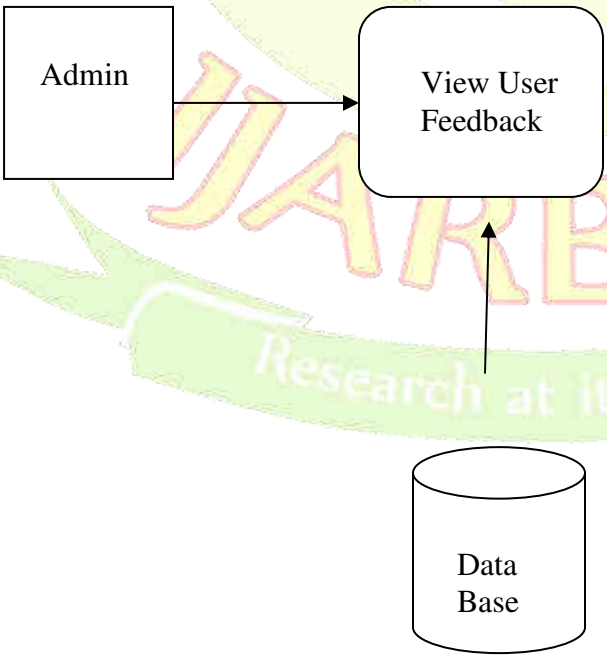
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016



b) View Feedback Status

Admin can views user feedback of status. If user can post user details with feedback, admin will get and view user details.



ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

TABLE 1
Summary of Short Names Label Meaning

Centralized Alg	**Baseline**: select only common terms **Greedy:** select terms by Alg1. **Merging:** select items by Alg 1 plus cost model **All:** select all terms in adhoc
DHT Alg **STAIRS**	**Baseline:** forward does to home nodes of common terms. **MIAI:** Forward does by Alg 1 Forward does by Alg 1 **STAIRS:** forward does by Alg2 improved replication and SSRP.

IV: CONCLUSION

Finally, the periodical retrieval scheme out performs the versions when the number of filters is small, but otherwise when the number of filters is higher, the throughput of the retrieval scheme is only. This result verifies that the retrieval schemes incur high overhead and large processing time when given a very large number of filters

REFERENCES

[1] G. Ausiello, P. Crescenzi, G. Gambosi, and V. Kann, Complexity and Approximation: Combinatorial Optimization Problems and Their Approximability Properties. Berlin, Germany: Springer-Verlag, 1999.

[2] F. Bonomi, B. Prabhakar, andY. Lu, ‘‘Bloom Filters: Design Innovations and Novel Applications,’’ in Proc. 43rd Annu. Allerton Conf., 2005, pp. 1006-1015.

[3] J.P. Callan, ‘‘Document Filtering with Inference Networks,’’ in Proc. SIGIR, 1996, pp. 262-269.Fig. 10.Evaluation on STAIRSþ. (a) Load balancing. (b) Throughput. RAO ET AL.: MTAF: ADAPTIVE DESIGN FOR KEYWORD-BASED CONTENT DISSEMINATION ON DHT NETWORKS 1083

[4] F. Chang, W.-C.Feng, and K. Li, ‘‘Approximate Caches for Packet Classification,’’ in Proc. IEEE INFOCOM, 2004, pp. 2196-2207.

[5] P.T. Eugster, P. Felber, R.Guerraoui, and A.-M.Kermarrec, ‘‘The Many Faces of Publish/Subscribe,’’ ACM Comput. Surveys, vol. 35, no. 2, pp. 114-131, June 2003.

[6] F. Fabret, H.-A. Jacobsen, F. Llirbat, J. Pereira, K.A. Ross, and D. Shasha, ‘‘Filtering Algorithms and Implementation for Very Fast Publish/Subscribe,’’ in Proc. SIGMOD Conf., 2001, pp. 115-126.

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

[7] M.J. Freedman and R. Morris, ‘‘Tarzan: A Peer-to-Peer Anonymizing Network Layer,’’ in Proc. ACM Conf. Comput. Commun. Security, 2002, pp. 193-206.

[8] M.R. Garey and D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness. San Francisco, CA, USA: Freeman,

[9] F. Hao, M.S. Kodialam, T.V. Lakshman, and H. Song, ‘‘Fast Multiset Membership Testing Using Combinatorial Bloom Filters,’’ in Proc. IEEE INFOCOM, 2009, pp. 513-521.

[10] J.M. Hellerstein J. Li, , M.F. Kaashoek, D.R. Karger,B.T. Loo, and R. Morris, ‘‘On the Feasibility of Peer-to-Peer Web Indexing and Search,’’ in Proc. IPTPS, 2003, pp. 207-215.