Personalized content search and retrieval in an interest group environment

Nishant Rayan, Sridhar.G, Rajesh.A, Dr.S.Swamynathan
{nishantrayan,sridhar87,shift2pappu}@gmail.com, swamyns@cs.annauniv.edu
Department Of Computer Science & Engineering, College Of Engineering, Guindy,
Anna University, Chennai, India

Abstract--***Personalization is a much required feature in today's world wrecked by information explosion. A typical software professional finds himself searching for a needle in a haystack of information. There is a way to redeem ourselves from such a plight. Formation and management of interest groups is becoming a widely accepted phenomenon in today's virtual world of internet. Our approach to content search and retrieval (CSR) is based on the interest groups that make use of existing architecture known as pFusion[1]. pFusion has gifted us with an efficient query routing mechanism within a single interest group. We have built a global CSR architecture that offers personalization to individuals in an interest group.***

Index terms **-- Content search and retrieval, personalization, interest group.**

I. INTRODUCTION

Many systems have been developed to cater to the search requirements of the people. Google is a well known search engine that is popular among people, especially software professionals. While search techniques and algorithms are highly mature and have been used by people for many years, personalized retrieval of content and presentation techniques are still in their infancy. By personalization we mean

1. Presentation of user-expected results
2. High relevancy of the retrieved results
3. Use of user search behavior in determining relevancy of search results.

Interest groups help us in partitioning the huge information base into several smaller domains so that search and retrieval become efficient, pFusion is a relatively new architecture that has efficient mechanism for query routing and overlay management. It has been studied closely and experiments have suggested that due to the enormity of content in any reasonably large scale network, pFusion fails to perform well as expected. This can be attributed to the ineffective peer network management in a large scale network.

Our approach combines the interest group architecture to the pFusion to deliver the much need personalization of search.

II. RELATED WORKS

A. Search Engines with personalization:

There are many popular search engines that promise personalization. Google has recently launched its own product that offers personalized searching. The user's search history is taken into account while presenting the search results. Users feel good to see results related to their previous searches. Studies have shown that users are much satisfied to see results based on the search behavior of other users who share their interests.

B. Other Search Engines

Eurekster is an interesting spin on search. It tries to leverage the search patterns of your

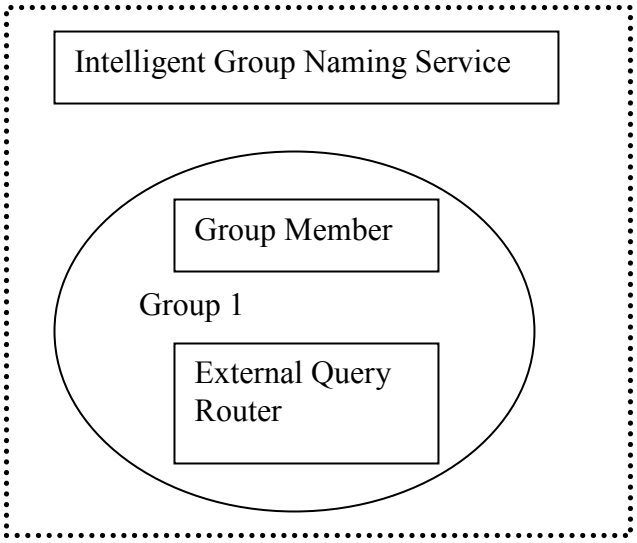
respected friends to help you find stuff. Eurekster could identify strong colleague networks

Feedster is a blog search engine. While it doesn't claim to specifically create social networks, one could see how it could be used to identify communities of cross posters in blogs.

III. PROPOSED SOLUTION

Our solution combines the best out of interest group management strategies and p2p architecture known as pFusion[1]. The system that we propose is reminiscent of a much required personalized content search and retrieval system.

A. Group Architecture



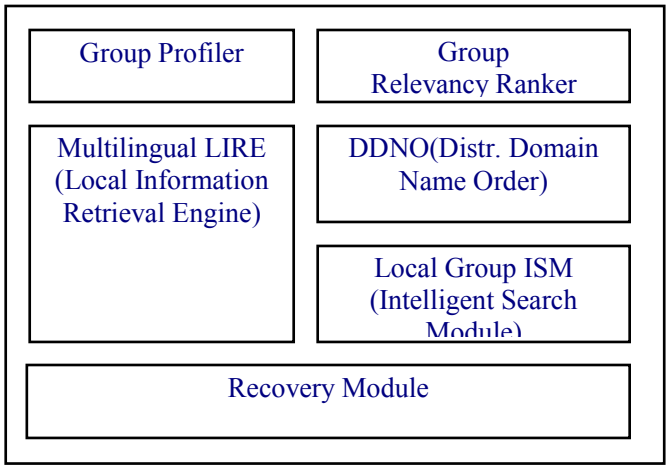
a. Intelligent Group Naming Service (IGNS):
Group creation is done by registering the group with the IGNS which contains a hashed table data structure to store group information. The groups are searched based on the user queries. The IGNS has the capability to extract query terms from the user queries and list the most relevant groups. It returns the group list to the user.
b. Interest Group
A group typically has several hundred users who share similar interest. The similarity in interest is apparent in the commonality of query terms in their search

The interest groups can be visualized as a logical partitioning of the huge information base. Each interest group can be uniquely identified by a name or a semantic title.

c. Group member
A user who performs frequent searches on information that is available in the interest group in which he has joined, is represented as a group member.

d. External Query Router(EQR)
Despite having several hundreds of group members, an interest group may not fully satisfy the requirement of its group member. This situation occurs if a user searches for content outside the interest group that he is a member of. This anomaly is handled by a privileged group member called External query router. The role of an EQR is to route the query to external groups and also to enable group members to join external groups.

B. Group Member



a. Group profiler: Manages the profiles of groups in which the users have joined.
b. Group Relevancy ranker: Gives a list of groups that are relevant to the search terms in the user queries

c. DDNO(Distributed Domain Name Order): It is a pFusion module that takes care of overlay management functions
d. Local Group Intelligent Search module: Performs the task of query routing. This is an important pFusion module in the system.

C. External Query Router

The EQR maintains a global picture of the interest group. When a user sends a query that has to be sent to an external group the EQR selects the group from the union list of groups in which the members in the interest group have joined. This global picture is maintained in a similar manner to that of the group member except that query is rerouted to the appropriate group member. This function is performed by the ***Intelligent Group selection module (IGSM).***

IV. WORK IN PROGRESS

Our system's query routing mechanism is more efficient compared that of the pFusion which handles local queries only. The information contents that are distributed across many peers in a group are retrieved and presented to the user. The system has certain inherent features that are highlighted below.

1) High Relevancy:
The modules, as shown in the architectural diagrams use specialized algorithms for selecting and returning the best possible results and hence they are qualified as intelligent. There is no doubt that the results returned are highly relevant to the user.

2) Personalization:
Instead of using the user history for achieving the personalization, the system monitors the groups that the user has joined to achieve the same. This can be illustrated by the same old APPLE example that is often used to demonstrate the effect of personalization

Let's say the user requires content for the search term apple (apple fruit). The user is already a member of the group which is semantically identified as mango fruit. The query is directed to the EQR of the mango fruit group. The EQR finds that one of the member of the group is also a member of the group named apple fruit. The query is immediately routed to the apple group and user finds the result on apple fruit top on the result list. Thus personalization is achieved.

3) Presentation of results:
The results of the search are presented to the user in the most appropriate form. Though the system focuses on text results some of the contents in multimedia formats require a different type of presentation. This is achieved by using a separate presentation module at the user end

V. FUTURE WORK

A. Multimedia content search and retrieval:
Our system primarily targets text content, however some systems require multimedia content search and retrieval. The techniques used in our system can be easily extended to cater to such requirements.

B. Extension to Social networking:
Social networking is a blooming concept that is more sophisticated than our original interest group based system. Much greater levels of performance can be achieved using social networking.

C. Support for Smart devices:
The compact smart devices such as PDA's and Mobile phones that are personal to the user, can be integrated into the system and will add a whole new dimension to the entire system.

D. Extension to Large Networks:
The WWW is a challenging environment for any system to function properly due to its shear size. The system can be adapted easily to such large scale network environments.

E. Distributed Web Crawlers:
With more coordination among peers the system can be extended to a distributed crawler system where each peer crawls the web independently and brings the results to the group which is shared by all the members.

VI. CONCLUSION

Our system holds much promise. The system in itself is distributed thereby reducing the overload on any particular node in the network. It doesn't exert any high performance demand on the peers nodes except the EQR which requires certain performance level. We have designed it to be extensible so that future improvements can be done with minimal effort. The content search and retrieval system can be integrated into a much larger system that requires an efficient query routing mechanism in a huge information base. The system fairs well in an environment where large number of people with similar interests are present. Based on their search activities they fall into the same group and benefit from each other's information content.

VII. REFERENCES

[1] Demetrios Zeinalipour-Yazti,Vana Kalogeraki, Dimitrios Gunopulos, **"pFusion: A P2P Architecture for Internet-Scale Content-Based Search and Retrieval"**
IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 18, NO. 6, JUNE 2007

[2] Tao Gu ,Daqing Zhang, **"A Peer-to-Peer Overlay for Context Information Search"** 2005 IEEE.

[3] Haitao Chen, Zunguo Huang, Zhenghu Gong,
"Efficient Content Locating in Peer-to-Peer Systems"
Proceedings of the 2005 IEEE International Conference on e-Business Engineering (ICEBE'05)

[4] Fr´ed´eric Dang Ngoc, Joaqu´ın Keller_, Gwendal Simon,
"MAAY: A Decentralized Personalized Search "
Proceedings of the 2005 Symposium on Applications and the Internet (SAINT'06) ,2006 IEEE

[5] Francisco Matias Cuenca-Acuna, Christopher Peery, Richard P. Martin, Thu D. Nguyen,
PlanetP: Using Gossiping to Build Content Addressable Peer-to-PeerInformation Sharing Communities Appears in the 12th IEEE International Symposium on High Performance Distributed Computing, 2003

[6] Matthias Bender, Tom Crecelius, Mouna Kacimi, Sebastian Michel, Josiane Xavier Parreira, Gerhard Weikum,
Peer-to-Peer Information Search: Semantic, Social, or Spiritual, appears in Bulletin of the IEEE Computer Society Technical Committee on Data Engineering,2007

[7] Haibo Yu, Tsunenori Mine and Makoto Amamiya "**An Architecture for Personal Semantic Web Information Retrieval System – Integrating Web services and Web contents"** Proceedings of the IEEE International Conference on Web Services (ICWS'05)

[8] Carla Carvalho, Al´ıpio M. Jorge, Carlos Soares
Personalization of e-newsletters based on web log analysis and clustering, Proceedings of the 2006 IEEE/WIC/ACM International Conference
on Web Intelligence (WI 2006 Main Conference Proceedings)(WI'06),IEEE 2006

[9] Souptik Datta, Kanishka Bhaduri, Chris Giannella,
and Hillol Kargupta **"Distributed Data Mining in Peer-to-Peer Networks",** Published by the IEEE Computer Society,2006

[10] Rohit Gupta, Varun Sekhri, and Arun K. Somani,
"**CompuP2P: An Architecture for Internet Computing Using Peer-to-Peer Networks",** IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 17, NO. 11, NOVEMBER 2006