International Journal of Engineering and Management Research, Vol.-2, Issue-6, December 2012
ISSN No.: 2250-0758
Pages: 57-62
www.ijemr.net

Web Search Information Retrial Using PFusion Architecture

A.Srinivas[1], M.V.Bramhananda Reddy [2], G.Sreenivasula Reddy[3], A.Rama Mohan Reddy[4]
[1]Dept. of CSE, Holy Mary Institute of Technology & Science, Hyderabad, Andhra Pradesh, INDIA
[2, 3] Dept. of CSE, VIGNANA BHARATHI INSTITUTE OF TECHNOLOGY, Proddatur, A.P., INDIA
[4]Dept. of CSE, S.V.U. College of Engineering, Sri Venkateswara University, Tirupati, INDIA

ABSTRACT

The emerging Peer-to-Peer model has become a very powerful and attractive paradigm for developing Internet-scale systems for sharing resources, including files and documents. The distributed nature of these systems, where nodes are typically located across different networks and domains, inherently hinders the efficient retrieval of information. In this paper, we consider the effects of topologically aware overlay construction techniques on efficient P2P keyword search algorithms. We present the Peer Fusion architecture that aims to efficiently integrate heterogeneous information that is geographically scattered on peers of different networks. The aim of PFusion Architecture is to integrate the heterogeneous information i.e. text, graph, sound, files, mails, Message etc. by means of an efficient scarch mechanism. It is powerful and attractive paradigm of Internet Systems for Searching and Passing of Message on different domains. There is an overall construction technique to get the good results for obtaining the shortest path of routing for the required Search and Retrieval of information. There is an improvement of latency time i.e. the time taken for the Retrieval of information from the Server. The basic idea is that an overlay Networks of nodes is constructed on top of heterogeneous Operating System and Network overlays which are flexible and deployable approaches that allow users to perform distributed operation without modifying the underlying the Physical Networks.

Keywords— Fusion, Internet System, Operating System, Peer to Peer, paradigm, Physical Networks.

I. INTRODUCTION

Web sites such as Youtube.com and Yahoo Video allow users to upload, search, browse, and view on demand the video clips of other users through a keyword-based search interface. Such systems typically exploit a centralized storage and retrieval infrastructure that has a number of disadvantages and limitations: 1) the service can easily become a bottleneck during periods of high demand and is also a single point of failure, 2) the infrastructure is expensive and requires extensive administration, and 3) the content can be censored.

We are all witnessing the rise in a new form of journalism, where non journalists have an active role in collecting, analyzing, and generating newswire based on their personal rules of fairness and objectivity, often also referred to as Citizen Journalism. The Web site voiceofsandiego.com establishes half of its content from contributing authors, and this new way of performing journalism has profound implications on the future of news media. Content in such applications is currently communicated to users through centralized Web sites, which suffer from the same disadvantages as the P2P video-sharing application.

Demerits of existing system: The service can easily become a bottleneck during periods of high demand and is also a single point of failure. The infrastructure is expensive and requires extensive administration. The content can be censored.

II. P2P SYSTEM

We model such a service on the premise of an unstructured P2P system, where each user stores locally its own video clips and performs the search and retrieval functions. An important point in such Internet-scale applications is that the large-scale data transfer and retrieval can be expensive when the network connections among the clients are arbitrary, due to unpredictable communication latencies, excessive resource consumption, and changing resource availability in inter domain routing. Thus, we seek to optimize the overlay by establishing connections between peers based on the criterion of network proximity. In particular, peers minimize the network distance from their neighboring nodes by establishing connections to nodes that belong to the same domain. Using a topologically aware P2P system, besides that of overcoming the problems of centralization, would also enable users to more easily retrieve local or regional content. We present the architecture of an Internet scale middleware that can be used for efficient content based search and retrieval in a variety of contexts.

III. FRONT END USED

Microsoft Visual Studio. Net used as front end tool. The reason for selecting Visual Studio dot Net as front end tool as follows: Visual Studio .Net has flexibility, allowing one or more language to interoperate to provide the solution. This Cross Language Compatibility allows to do project at faster rate. Visual Studio. Net has Common Language Runtime, which allows the entire component to converge into one intermediate format and then can interact. Visual Studio. Net has provided excellent security when your application is executed in the system. Visual Studio.Net has flexibility, allowing us to configure the working environment to best suit our individual style. We can choose between a single and multiple document interfaces, and we can adjust the size and positioning of the various IDE elements. Visual Studio. Net has Intelligence feature that make the coding easy and also dynamic help provides very less coding time. The working environment in Visual Studio.Net is often referred to as Integrated Development Environment because it integrates many different functions such as design, editing, compiling and debugging within a common environment. In most traditional development tools, each of separate program, each with its own interface. The Visual Studio.Net language is quite powerful – if we can imagine a programming task and accomplished using Visual Basic .Net. After creating a Visual Studio. Net application, if we want to distribute it to others we can freely distribute any application to anyone who uses Microsoft windows. We can distribute our applications on disk, on CDs, across networks, or over an intranet or the internet. Toolbars provide quick access to commonly used commands in the programming environment. We click a button on the toolbar once to carry out the action represented by that button. By default, the standard toolbar is displayed when we start Visual Basic. Additional toolbars for editing, form design, and debugging can be toggled on or off from the toolbars command on the view menu. Many parts of Visual Studio are context sensitive. Context sensitive means we can get help on these parts directly without having to go through the help menu. For example, to get help on any keyword in the Visual Basic language, place the insertion point on that keyword in the code window and press F1. Visual Studio interprets our code as we enter it, catching and highlighting most syntax or spelling errors on the fly. It's almost like having an expert watching over our shoulder as we enter our code.

IV. BACK END USED

Microsoft SQL SERVER 2000 used as back end tool. The reason for selecting SQL SERVER 2000 as back end tool as follows: SQL SERVER, DATABASE. A database management, or DBMS, gives the user access to their data and helps them transform the data into information. Such database management systems include dBase, paradox, IMS, Sql Server and SQL Server. These systems allow users to create, update and extract information from their database.

I. FEATURES OF SQL SERVER (RDBMS)

SQL SERVER is one of the leading database management systems (DBMS) because it is the only Database that meets the uncompromising requirements of today's most demanding information systems. From complex decision support systems (DSS) to the most rigorous online transaction processing (OLTP) application, even application that require simultaneous DSS and OLTP access to the same critical data, SQL Server leads the industry in both performance and capability. SQL SERVER is a truly portable, distributed, and open DBMS that delivers unmatched performance, continuous operation and support for every database.

II. OUTPUT DESIGN

Outputs from computer systems are required primarily to communicate the results of processing to users. They are also used to provide a permanent copy of the results for later consultation. The various types of outputs in general are: External Outputs, whose destination is outside the organization, Internal Outputs whose destination, is with in organization and they are the User's main interface with the computer. Operational outputs whose use is purely with in the computer department. It is not always desirable to print or display data as it is held on a computer. It should be decided as which form of the output is the most suitable. Output Media: In the next stage it is to be decided that which medium is the most appropriate for the output. The main considerations when deciding about the output media are: The suitability for the device to the particular application. Keeping in view the above description the project is to have outputs mainly coming under the category of internal outputs. The main outputs desired according to the requirement specification are: The outputs were needed to be generated as a hot copy and as well as queries to be viewed on the screen. Keeping in view these outputs, the format for the output is taken from the outputs, which are currently being obtained after manual processing. The standard printer is to be used as output media for hard copies.

INPUT DESIGN

Input design is a part of overall system design. The main objective during the input designs is as given below: To produce a cost-effective method of input. To achieve the highest possible level of accuracy. To ensure that the input is acceptable and understood by the user.

The main input stages can be listed as below: Data recording, Data transcription, Data conversion, Data verification, Data control, Data transmission, Data validation and Data correction.

INPUT TYPES

It is necessary to determine the various types of inputs. Inputs can be categorized as follows: External inputs, which are prime inputs for the system. Internal inputs, which are user communications with the system. Operational, which are computer department's communications to the system, Interactive, which are inputs entered during a dialogue.

INPUT MEDIA

At this stage choice has to be made about the input media. To conclude about the input media consideration has to be given to; Type of input, Flexibility of format , Speed, Accuracy, Verification methods, Rejection rates, Ease of correction, Storage and handling requirements , Security, Easy to use and Portability. Keeping in view the above description of the input types and input media, it can be said that most of the inputs are of the form of internal and interactive. As Input data is to be the directly keyed in by the user, the keyboard can be considered to be the most suitable input device. Error avoidance, error detection, data validation: user interface design. User_initiated intergfaces User initiated interfaces fall into tow approximate classes: Command driven interfaces: In this type of interface the user inputs commands or queries which are interpreted by the computer. Forms oriented interface: The user calls up an image of the form to his/her screen and fills in the form. The forms oriented interface is chosen because it is the best choice.

III. COMPUTER-INITIATED INTERFACES

The following computer – initiated interfaces were used: The menu system for the user is presented with a list of alternatives and the user chooses one of alternatives. Questions – answer type dialog system where the computer asks question and takes action based on the basis of the users reply. Right from the start the system is going to be menu driven, the opening menu displays the available options. Choosing one option gives another popup menu with more options. In this way every option leads the users to data entry form where the user can key in the data.

ERROR MESSAGE DESIGN

The design of error messages is an important part of the user interface design. As user is bound to commit some errors or other while designing a system the system should be designed to be helpful by providing the user with information regarding the error he/she has committed. This application must be able to produce output at different modules for different inputs.

PERFORMANCE REQUIREMENTS

Performance is measured in terms of the output provided by the application. Requirement specification plays an important part in the analysis of a system. Only when the requirement specifications are properly given, it is possible to design a system, which will fit into required environment. It rests largely in the part of the users of the existing system to give the requirement specifications because they are the people who finally use the system. This is because the requirements have to be known during the initial stages so that the system can be designed according to those requirements. It is very difficult to change the system once it has been designed and on the other hand designing a system, which does not cater to the requirements of the user, is of no use. The requirement specification for any system can be broadly stated as given below: The system should be able to interface with the existing system, The system should be accurate. The system should be better than the existing system and The existing system is completely dependent on the user to perform all the duties.

SYSTEM ANALYSIS

The DDNO Module is a distributed overlay construction module utilized to cluster topologically lose-by nodes together. This is achieved by having each node connect to d=2 random neighbors and d=2 other nodes in the same domain (siblings), where d denotes the number of neighbors. Sibling nodes are efficiently discovered by the deployment of distributed lookup messages and local Zone Caches, which contain information on which domains are reachable in an r-hop radius. The ISM is a keyword search mechanism used by each pFusion node. ISM consists of the following two subcomponents: 1) a Profile Mechanism, which a peer uses to build a profile for each of its neighboring peers (that is, the query/query hit pairs), and 2) Relevance Rank (RR), which is a peer-ranking mechanism that uses the local profiles to select the neighbors that will lead a query to the most relevant answers. The Local Information Retrieval Engine (LIRE) is a local index utilized by each node in order to efficiently access its local data repository. Specifically, LIRE organize local information into disk based indexes, which allow the efficient execution of a wide range of queries. Note that the indexes in our setting are incrementally updated as new information arrives in the local repository. The merging of query results is performed at the querying node, which ranks results on their local score returned by the source. LIRE can also use an external data fetcher for retrieving and storing content pertinent to the interests of the node owner.

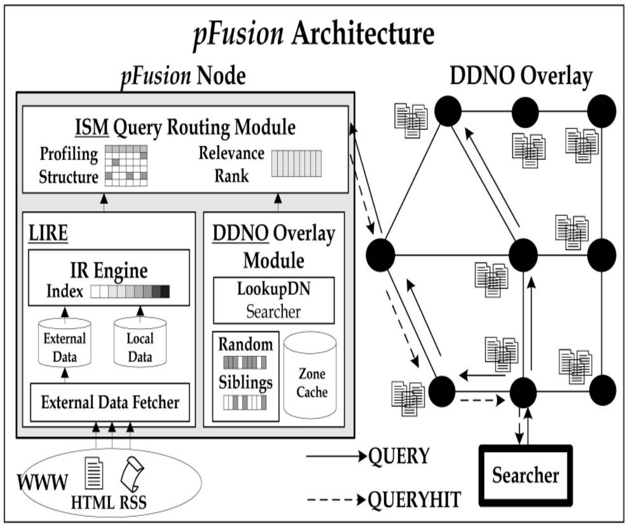


Figure: System Architecture

SYSTEM DESIGN

Design is concerned with identifying software components specifying relationships among components. Specifying software structure and providing blue print for the document phase. Modularity is one of the desirable properties of large systems. It implies that the system is divided into several parts. In such a manner, the interaction between parts is minimal clearly specified. Design will explain software components in detail. This will help the implementation of the system. Moreover, this will guide the further changes in the system to satisfy the future requirements.

IV. DEVELOPMENT OF SYSTEM AND TESTING

SYSTEM MAINTENANCE

The objectives of this maintenance work are to make sure that the system gets into work all time without any bug. Provision must be for environmental changes which may affect the computer or software system. This is called the maintenance of the system. Nowadays there is the rapid change in the software world. Due to this rapid change, the system should be capable of adapting these changes. In our project the process can be added without affecting other parts of the system. Maintenance plays a vital role. The system liable to accept any modification after its implementation. This system has been designed to favor all new changes. Doing this will not affect the system's performance or its accuracy.

IMPLEMENTATION

Implementation is the most crucial stage in achieving a successful system and giving the user's confidence that the new system is workable and effective. Implementation of a modified application to replace an existing one. This type of conversation is relatively easy to handle, provide there are no major changes in the system. Each program is tested individually at the time of development using the data and has verified that this program linked together in the way specified in the programs specification, the computer system and its environment is tested to the satisfaction of the user. The system that has been developed is accepted and proved to be satisfactory for the user. And so the system is going to be implemented very soon. A simple operating procedure is included so that the user can understand the different functions clearly and quickly. Initially as a first step the executable form of the application is to be created and loaded in the common server machine which is accessible to the entire user and the server is to be connected to a network. The final stage is to document the entire system which provides components and the operating procedures of the system. Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective. The implementation stage involves careful planning, investigation of the existing system and it's constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods. Implementation is the process of converting a new system design into operation. It is the phase that focuses on user training, site preparation and file conversion for installing a candidate system. The important factor that should be considered here is that the conversion should not disrupt the functioning of the organization.

V. RESULTS

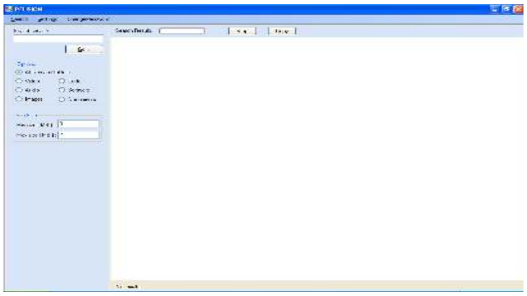
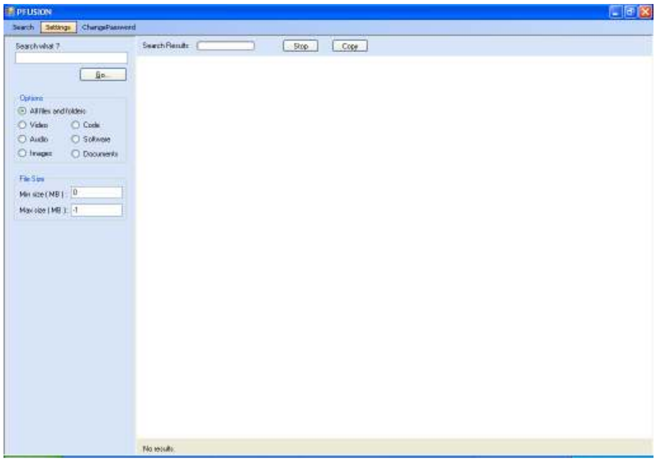


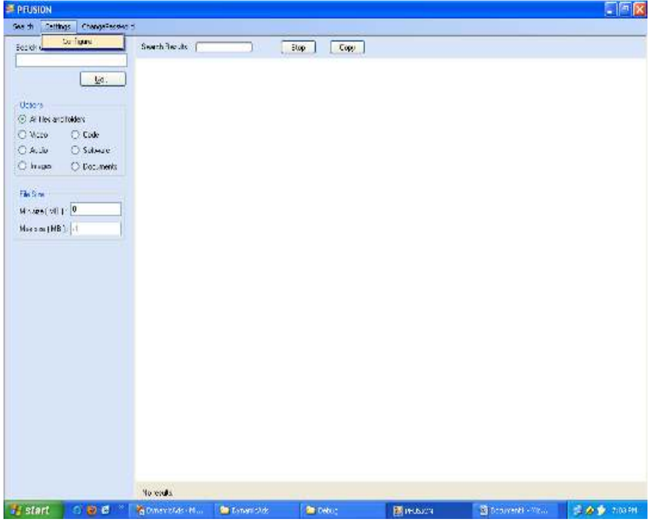
Figure 9.1: File Storage Screen



● All files and folders
○ Video ○ Code
○ Audio ○ Software
○ Images ○ Documents

Min size (MB) : 0
Max size (MB) : -1

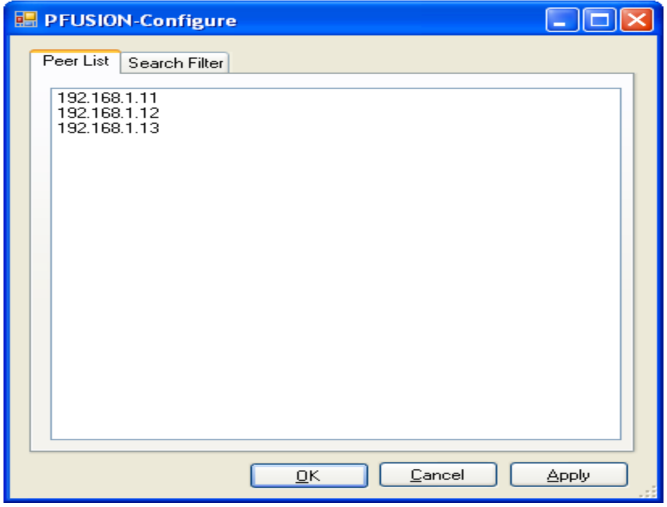
Figure 9.2: Search Screen



● All files and folders
○ Video ○ Code
○ Audio ○ Software
○ Images ○ Documents

Min size (MB) : 0
Max size (MB) : -1

Figure 9.3: Option Screen



192.168.1.11
192.168.1.12
192.168.1.13

Figure 9.4: Peer List Screen

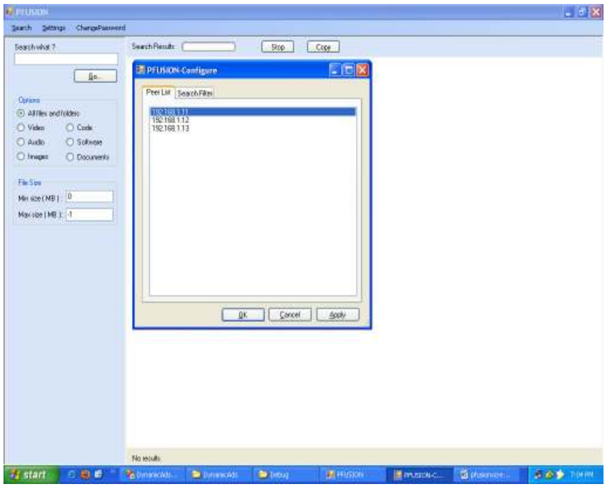
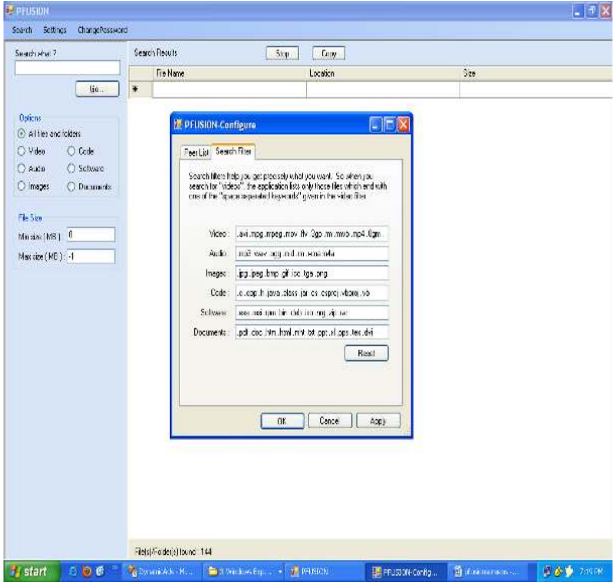


Figure 9.5: Search File Screen Selected I.P. Address



Search filters help you get precisely what you want. So when you search for "videos", the application lists only those files which end with one of the "space separated keywords" given in the video filter

Video : .avi .mpg .mpeg .mov .flv .3gp .rm .rmvb .mp4 .ogm .

Audio : .mp3 .wav .ogg .mid .ra .wma .m4a

Images : .jpg .jpeg .bmp .gif .ico .tga .png

Code : .c .cpp .h .java .class .jar .cs .csproj .vbproj .vb

Software : .exe .msi .rpm .bin .deb .iso .msg .zip .rar

Documents : .pdf .doc .htm .html .mht .txt .ppt .xl .pps .tex .dvi

File(s)/Folder(s) found : 144

Figure 9.6: Search File/Folder Screen (Page Extinction)

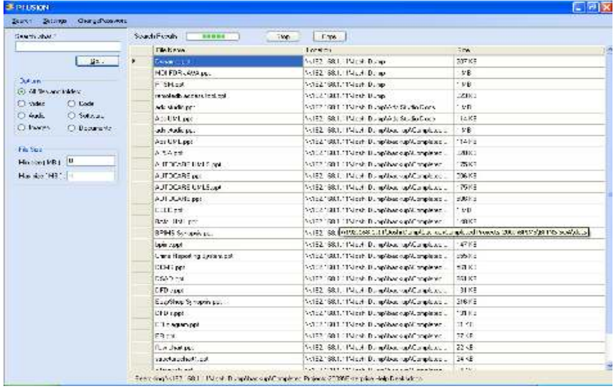


Figure 9.7: Search Results

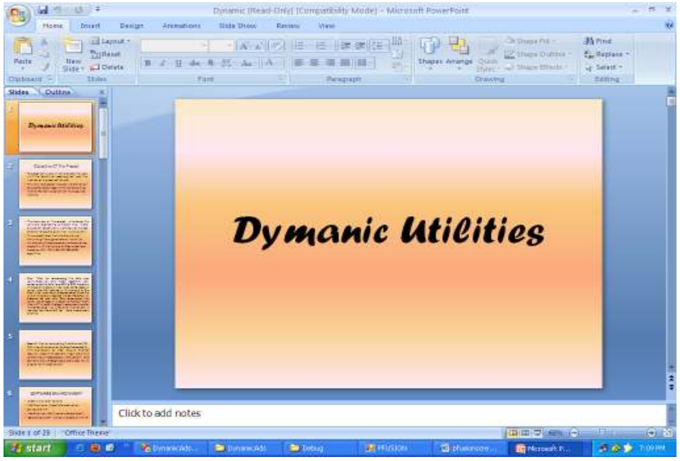


Figure 9.8: Successful Login Open File Screen

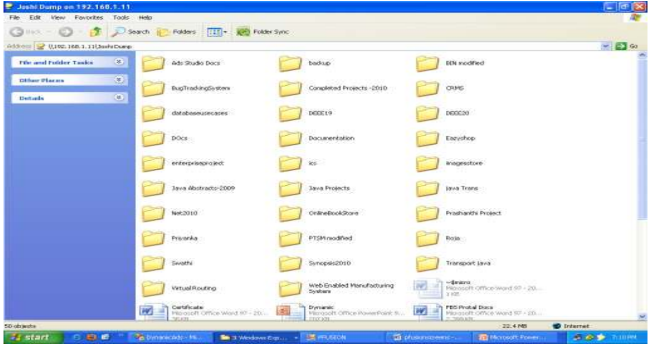
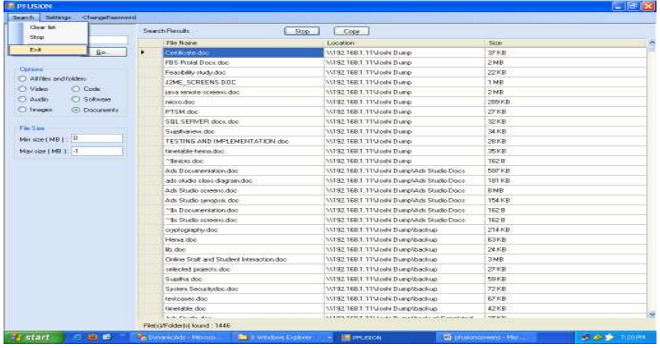


Figure 9.9: Display All Folder's Screen



File Name	Location	Size
Certificate.doc	\\192.168.1.11\Joshi Dump	37 KB
FBS Protal Docs.doc	\\192.168.1.11\Joshi Dump	2 MB
Feasibility study.doc	\\192.168.1.11\Joshi Dump	22 KB
J2ME_SCREENS.DOC	\\192.168.1.11\Joshi Dump	1 MB
java remote screens.doc	\\192.168.1.11\Joshi Dump	2 MB
micro.doc	\\192.168.1.11\Joshi Dump	289 KB
PTSM.doc	\\192.168.1.11\Joshi Dump	27 KB
SQL SERVER docs.doc	\\192.168.1.11\Joshi Dump	32 KB
Sujathanew.doc	\\192.168.1.11\Joshi Dump	34 KB
TESTING AND IMPLEMENTATION.doc	\\192.168.1.11\Joshi Dump	28 KB
timetable-hema.doc	\\192.168.1.11\Joshi Dump	35 KB
~$micro.doc	\\192.168.1.11\Joshi Dump	162 B
Ads Documentation.doc	\\192.168.1.11\Joshi Dump\Ads Studio Docs	587 KB
ads studio class diagram.doc	\\192.168.1.11\Joshi Dump\Ads Studio Docs	181 KB
Ads Studio screens.doc	\\192.168.1.11\Joshi Dump\Ads Studio Docs	8 MB
Ads Studio synopsis.doc	\\192.168.1.11\Joshi Dump\Ads Studio Docs	154 KB
~$s Documentation.doc	\\192.168.1.11\Joshi Dump\Ads Studio Docs	162 B
~$s Studio screens.doc	\\192.168.1.11\Joshi Dump\Ads Studio Docs	162 B
cryptography.doc	\\192.168.1.11\Joshi Dump\backup	214 KB
Hema.doc	\\192.168.1.11\Joshi Dump\backup	63 KB
lib.doc	\\192.168.1.11\Joshi Dump\backup	24 KB
Online Staff and Student Interaction.doc	\\192.168.1.11\Joshi Dump\backup	3 MB
selected projects.doc	\\192.168.1.11\Joshi Dump\backup	27 KB
Sujatha.doc	\\192.168.1.11\Joshi Dump\backup	59 KB
System Securitydoc.doc	\\192.168.1.11\Joshi Dump\backup	72 KB
testcases.doc	\\192.168.1.11\Joshi Dump\backup	67 KB
timetable.doc	\\192.168.1.11\Joshi Dump\backup	42 KB

File(s)/Folder(s) found : 1446

Figure 9.10: Select Exit Option Screen

VII. CONCLUSION

We considered and evaluated the impact of the use of topologically aware overlay networks on the performance of fully distributed P2P information retrieval techniques. Specifically, we show that it is possible to efficiently organize the overlay network using only local information in order to significantly improve the query latency. We also show how to take advantage of this organization when routing the queries in the network. Our experimental results demonstrate that our approach optimizes many desirable properties such as aggregate delays, recall rates, and the number of messages. We believe that our techniques are simple to enable seamless integration into existing overlay systems, with minimal changes.

REFERENCES

[2] R.A. Baeza-Yates and B.A. Ribeiro-Neto, Modern Information Retrieval. ACM Press Series/Addison-Wesley, May 1999.

[3] M. Bawa, R.J. Bayardo, S. Rajagopalan, and E. Shekita, “Make It Fresh, Make It Quick—Searching a Network of Personal Webservers,” Proc. 12th Int’l World Wide Web Conf., pp. 577- 586, 2003.

[4] B. Bolloba´ s, Modern Graph Theory, Graduate Texts in Mathematics, vol. 184. Springer, 1998.

[5] B.H. Bloom, “Space/Time Trade-Offs in Hash Coding with Allowable Errors,” Comm. ACM, vol. 13, no. 7, pp. 422-426, 1970.

[6] H. Cai and J. Wang, “Foreseer: A Novel, Locality-Aware Peer-to- Peer System Architecture for Keyword Searches,” Proc. Fifth ACM/ IFIP/Usenix Int’l Conf. Middleware, pp. 38-58, 2004.

[7] P. Cao and Z. Wang, “Efficient Top-K Query Calculation in Distributed Networks,” Proc. ACM Symp. Principles of Distributed Computing, pp. 206-215, 2004.

[8] M. Castro, P. Druschel, Y.C. Hu, and A. Rowstron, “Topology- Aware Routing in Structured Peer-to-Peer Overlay Networks,” Proc. IFIP/ACM Int’l Conf. Distributed Systems Platforms, 2001.

[9] Y. Chawathe, S. Ratnasamy, L. Breslau, N. Lanham, and S. Shenker, “Making Gnutella-Like P2P Systems Scalable,” Proc. Conf. Applications, Technologies, Architectures, and Protocols for Computer Comm., pp. 407-418, 2003.

[10] B.F. Cooper, “Guiding Queries to Information Sources with InfoBeacons,” Proc. Fifth ACM/IFIP/USENIX Int’l Conf. Middleware, pp. 59-78, 2004.