DATA RETRIEVAL IN DISTRIBUTED DATABASES USING MAS

[1] Mr. Vaibhav Parjane, [2]Mr. Manish Rai, [3]Dr. MohitGangwar
[1]Student, [2]Head of Department, [3]Principal
[1]Department of Computer Science and Engineering,
[1]RKDF College of Engineering, Bhopal, India

Abstract: In distributed systems, to search out information is costly task because they have to be compel led to transfer information from node containing information to the node wherever query is generated, this wil l consume latency, network traffic etc.For reducing these parameters mobile agents are accustomed fetch information from nodes wherever information resides. Along side mobile agents directory containing information concerning database kept on completely different nodes is employed to focus retrieval method solely to those nodes that are containing answers to the query. 3 kinds of agents area unit accustomed fetch data namely coordinator, search and local agent.

Key words
Mobile agents, data retrieval, distributed data, Natural Language Processing

I. INTRODUCTION

Traditional Client-Server architecture for distributed data retrieval do not offer much flexibility. There are several steps in client server architecture ; set up a connection between client and server, sending request to database servers and receive results from all database servers. But client has to send requests to each database server containing distributed data. With increase in no. of servers no. of requests increase which increase bandwidth consumption. Also heterogeneity of databases affect retrieval. Another approach is to transfer database to a node where request is generated. But this approach also consumes huge amount of bandwidth. Also unnecessary data is transferred over the network. The challenge is retrieving data from distributed databases ,usually heterogeneous, with minimum consumption of network bandwidth.

In mobile computing atmosphere, users will access information freelance of their location . But accessing this info shouldn't prohibit quality of application. From information management purpose of reading data, mobile users will handle solely fraction of information since mobile devices area unit having restricted resources.The invention of low price and nonetheless moveable mobile devices have enabled mobile users to figure from anyplace, at anytime. Along with growing technology scores of folks area unit exploitation these devices and through that these area unit accessing distributed information residing on distributed nodes. So there is ought to develop a system that ought to give required information to those mobile devices with a minimum use of resources. these days mobile devices area unit developed to use web. These devices area unit GPRS enabled that provide the simplest way to attach alternative devices to transfer information or to request information. These devices are often wont to access distributed access through GPRS therefore on preserve limited resources of those mobile devices. The devices area unit preserved by exploitation mobile agents. By this technology all the computations are performed at the nodes themselves.

Mobile agent technology is used as a useful and efficient tool for searching and retrieving data in distributed environment where the data is stored at a various nodes of the system.A mobile agent is an executing program that can migrate during execution from machine to machine in a heterogeneous network.[18] . The advantage of mobile agent is that it searches for info rather than users. Mobile agents carry code of execution or query to be applied on information. They run the query on the information and returns to the node that who had created that agent. This agent carries answer to the query .It is an economical alternative to transfer information and
additionally it reduces execution time. therefore by exploitation of mobile agents we are able to cut back bandwidth needed for operation. during this paper we've got urged a query retrieval approach by exploitation directories and multi agent system.

The Distributed Information Retrieval task deals with the collection of information from multiple and usually heterogeneous information sources that exist in a distributed environment. One way to address these issuesis to use information agents. These Distributed Information Retrieval agents should be able to:

- ready to serve a requst for information from user ,
- Convert the request into a code that can be understood by resources,
- determine the knowledge sources that contain information relevant to the request,
- send the request to those sources,
- collect the corresponding results,and
- methods for returning results.

2. Related Work

With development of distributed applications need for retrieving data from scattered data arised.Many approaches are made for retrieval of distributed data.In [1] the process involving distributed data access from a mobile device, using mobile agents is described. To answer any query in the distributed environment the search is conducted to answer the query

only in the databases, which are present in the systems.This approach is useful in situations where required data is scattered on most of the nodes.But in case where data is concentrated to very few nodes. Because unnecessary search agents are sent to the nodes where required data is not present.In [3] the planning issues and implementation details of an entire MAP(Multi Agent Platform) model addresses numerous problems like security mechanisms, fault olerance, methods for building network-aware mobile agents etc.The MAP system has been has been implemented in Java and optimized for network and systems management applications. In [4],a system supported mobile crawlers which uses mobile agent is planned. The approach implements mobile agents to crawl the pages.Aglet platform [19] is used for implementation of mobile agents. These mobile crawlers or mobile agents that crawl the pages determine the changed pages at the remote web site while not downloading them.But it downloads those pages solely,which have truly been changed since last crawl. therefore it'll cut back the net traffic and load on the remote web site. Distributed info framework supported mobile agent is planned in [5],which represents agent distribution and also the advancement of every agent.The distributed info system is meant and enforced by the employment of agent technology in developing information gathering management application platform of Down Hole Operation Company. under the IBM Aglet development platform the distributed question system is accomplished supported Mobile Agents,the running results of system shows high pertinence and capability of the model. In [6], Papastavrou et. Al. have suggested the framework for accessing Web-based distributed data . For that they have used mobile agents.The suggested system also supports for light-weight, portable, and autonomous clients.Also it is able to operate on slower networks. The implementation of the mobile agents is done by using the aglet platform.The system performs well in wireless and dial-up environments and also for average size transactions, As compared to a client-server platform the proposed system provides a performance improvement of roughly multiples of ten. For the fixed network, the gains are near about 20% and 30 %, respectively. In [7] an experimental mobile-agent system searching technical reports distributed across multiple machines is proposed. The application was implemented on the DAgents system[20].It uses statistical information retrieval system called Smart.Smart uses the vector-space model to measure the textual similarity between documents and is wrapped inside a stationary agent on each node. Kawamura et al. [8] have proposed three types of agents to process accessing of data from distributed databases namely direct access, stationery agent access, and mobile agent access. Also, Ismail et al. [9] have analyzed the comparisons between a Java applet-based approach and a mobile agent technology in accessing distributed databases from the Web.It presents a performance evaluation of the mobile agent paradigm in comparison to the client/server paradigm.It is implemented on top of the Java environment,using respectively RMI, the Aglets mobile agents platform and a mobile agents prototype.Menczer [10] designed and implemented Myspiders, a multi-agent system for information discovery in the Internet.Myspiders is a threaded multiagent system designed for information discovery.It uses an adaptive population of intelligent agents mining the Web online at query time. The usage of mobile agents for information filtering is studied by Theilmann et al.[11].It implements mobile agents for filtering distributed information resources and coordinating mobile agent dissemination that minimizes communication costs. Nguyen et.al. [12] has developed an agent system that helps information retrieval from the Internet. Their system differs from Myspiders because it uses consensus methods for resolving differences in response sets i.e. answers from various nodes and it uses multiple agents like managing agents and search agents for the retrieval task.

Tool used for implementation of mobile agents is JADE(Java Agent DEvelopment Framework)[13].JADE is a Framework which is implemented in Java language. It works as a middle-ware for developing mobile managing agents and search agents for the retrieval task. Tool used for implementation of mobile agents is JADE(Java Agent DEvelopment Framework)[13].JADE is a Framework which is implemented in Java language. It works as a middle-ware for developing mobile agents. JADE platform follows all guidelines suggested by FIPA(Foundation for Intelligent Physical Agents)[14] specifications. JADE also allows debugging and deployment of mobile agents.FIPA is an IEEE Computer Society standards organization. It works for agent technology and the interoperability with other technologies. In JADE, this platform can be distributed across nodes of the distributed system. Also JADE facilitates mobility of agents from one node to another one when it is necessary.JADE is used as add-on to JAVA. When it is used with netbeans it should be added to libraries of our project in order to use it.

3. Programmer's design

3.1. System Architecture

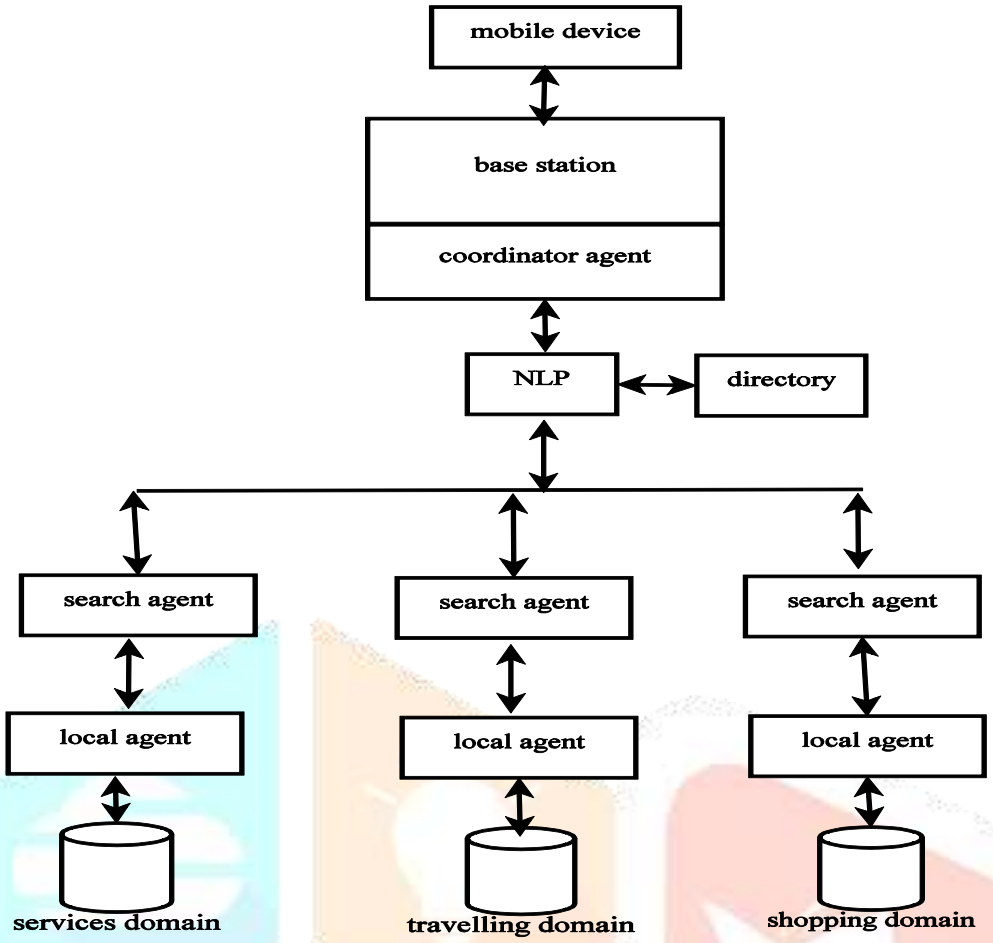

Figure 1: Proposed system

It classifies each such domain as referred by query and extracts results from each domain For example, it will Search all the Employee Information from Employee Domain like Name, Designation, Gender and sends the result back to NLP.The NLP Convert in-
formation from computer databases into readable human language. It uses Natural Language Generation (NLG) for generating natural language from a machine representation system such as a knowledge base or a logical form. The converted output can be viewed from mobile unit.

Finally, these important features are passed to the co-ordinator agent. The co-ordinator agent handles the query answering process. It acts as an information gateway to the records sources it manages.In contrast to the above, the user agent is the one that the end user interacts with. It formulates the user's query, entered via a application, translates into an appropriate query message format and displays the answers.

The user agent makes use of the services of a corresponding Co-ordinator agent. This agent accepts requests from user agents. It has the role to identify which database the user is actually referring. Concluding with the overall agent architecture, there are three agents: the co-ordinator agent, the search agent and local agent.

3.2. Mathematical Model

Problem description

Let s be MAS which will process a query such that

$s = \{M,B,C,A,L,D\}$

where

M is set of mobile phones.

B is set of base stations

C is set of coordinator agents

A is set of search agents

L is set of local agents

D is set of database servers

$A = \{a0, a1, a2, a3\}$

$M = \{m0,m1, ...,mn\}$

B = *b*0
C = *c*0
L = {*l*0, *l*1, *l*2, *l*3}
D = {*d*0, *d*1, *d*2, *d*3}
note:-Since only four distributed databases are considered for implementation so only four elements for a set are considered.

Vein Diagrams

Following figures are showing vein diagrams of the proposed system

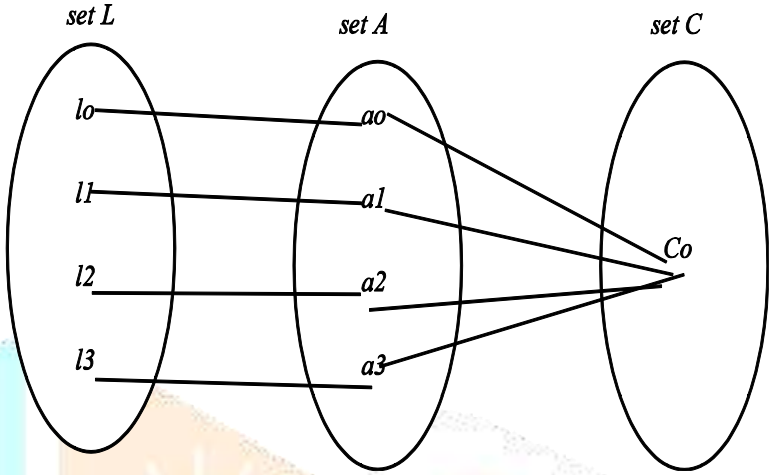

Figure 2: vein diagram 1

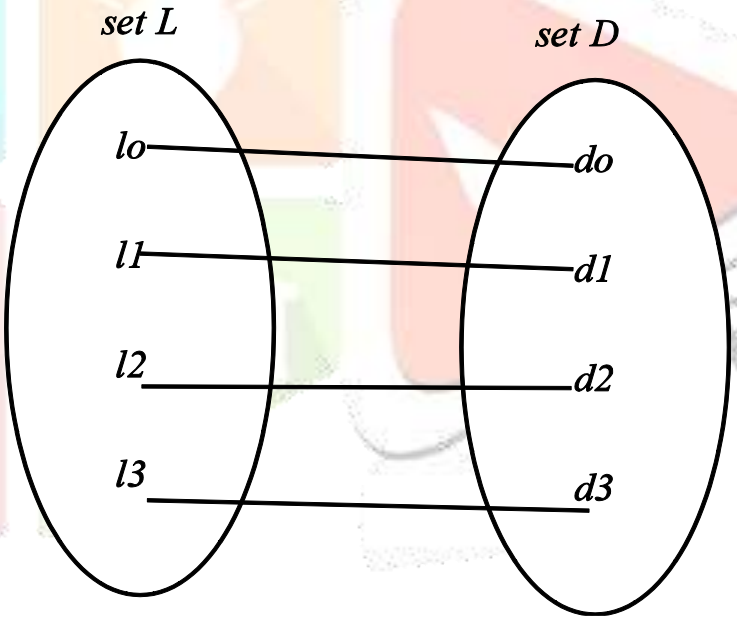

Figure 3: vein diagram 2

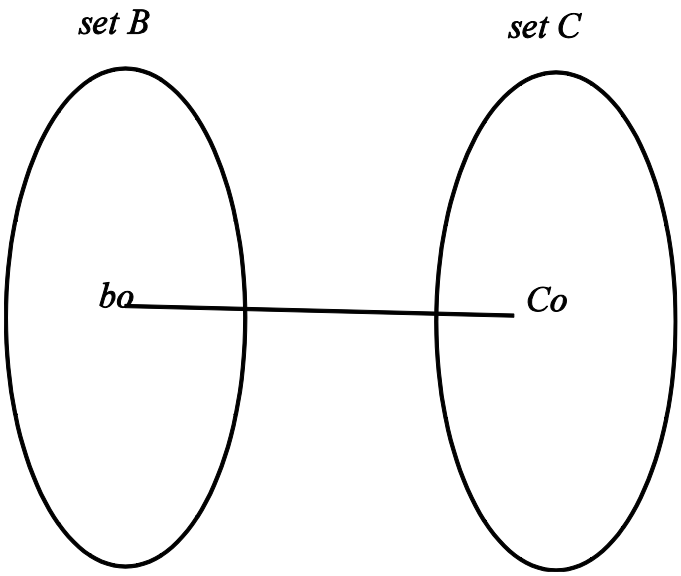

Figure 4: vein diagram 3

DFA theory

Definition:

A deterministic finite automaton (DFA)

1. A finite set of states (often denoted Q)
2. A finite set S of symbols (alphabet)
3. A transition function that takes as argument a state and symbol and returns a state (often denoted d)

The transition function d is a function in d: $Q \times S = Q$

4. A start state often denoted *s*0
5. A set of final or accepting states (often denoted F)

So a DFA is mathematically represented as a 5-tuple (Q, S, d, *s*0, F)

Mathematical Model for the system

Finite set of states (Q) = { *s*0, *s*1, *s*2, *s*3, *s*4 } where

1. *s*0 denotes initial state and final state and *s*1, *s*2, *s*3,*s*4 are internal states.
2. A set of final or accepting states (F) = { *s*0 }
3. A Finite set S of symbols = { English query,SQL query,result,directory result}

	F0	F1	F2	F3	F4
F0	1	1	0	0	0
F1	1	0	1	0	0
F2	0	1	1	1	0
F3	0	1	0	0	1
F4	0	0	0	1	1

- Function *f*0 will accept a query from a user and it will pass the query to coordinator agent for further processing,and when results will return it will display those results to user.
- Function *f*1 will process query given by base station.
- Function *f*2 will convert the English query to SQL query by using NLP approach.
- Function *f*3 will send search agent to the particular node for retrieving data.Search agent contains query to be fired.
- In function *f*4 local agent will accept query and fire on database,will retrieve results and pass to search agent.

3.3. Dynamic Programming and Serialization

Proposed system can be divided into two main parts;mobile and stationary.Mobile part includes all mobile agents used in the system.In proposed system search agents are mobile,travelling across network.These agents carry query which is to be fired on database.Stationary part includes local and coordinator agents.Local agents are present at nodes where database resides.These agents accept query from search agent,fire it on the database,collect results and forward these results to search agent.Local agents are well aware of DBMS which is maintaining distributed database.Coordinator agent coordinates retrieval process. Coordinator agent carry out retrieval process by creating search agents,sending them to appropriate nodes and collecting results from those search agents.

Necessary mapping functions

f(*M*) ! *B* ,such that
e.g.*f*(*m*0) | *b*0*€M*
f(*C*) | *A* ,such that
e.g.*f*(*c*0) | {*a*0, *a*1, *a*2, *a*3}*€A*
f(*B*) | *C* ,such that
e.g.*f*(*b*0) | *c*0*€C*
f(*A*) | *L* ,such that
e.g.*f*(*a*0) | *l*0*€L*
f(*L*) | *D* ,such that
e.g.*f*(*l*0) | *d*0*€D*

3.4. Data independence and Data Flow architecture

As shown in figures,query for data is accepted through mobile device.This query is accepted through simple GUI.The communication between mobile and base station (coordinator machine) is done through GPRS.At base station coordinator agent is present which accepts the Natural language query and converts it to SQL query by using NLP.Then coordinator agent identifies node(s) which are having answers to given query by using a directory .

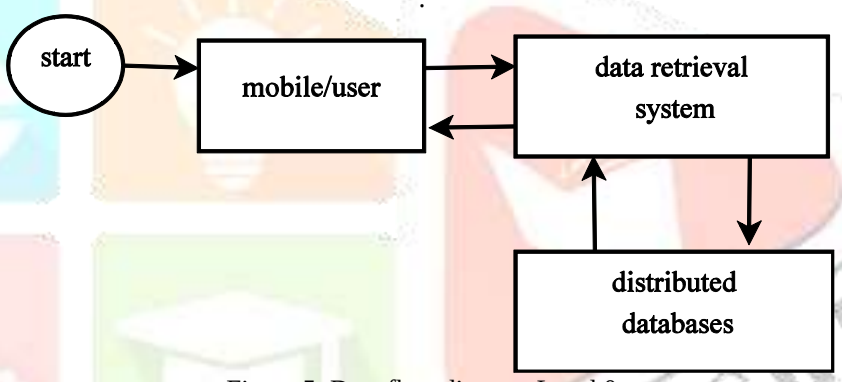

Figure 5: Data flow diagram.Level 0

This directory is maintained at base station which contains information about data present in distributed databases.Suppose N no. of nodes are containing answers to given query then coordinator agent creates N no. of search agents.Each search agent contains query.Search agents travel through network and reaches to their respective nodes.These nodes are containing database servers which are containing distributed data.Also local agents are present at these nodes.Search agents forward query to these local agents.As local agents are having information about how to retrieve data from their databases they do so after accepting a query from search agent.Local agent searches for answers to given query,retrieves them and forwards them to search agent.Search agents returns back to its originator i.e. at base station.Same data is accessed by mobile user through his mobile phone as a web page.

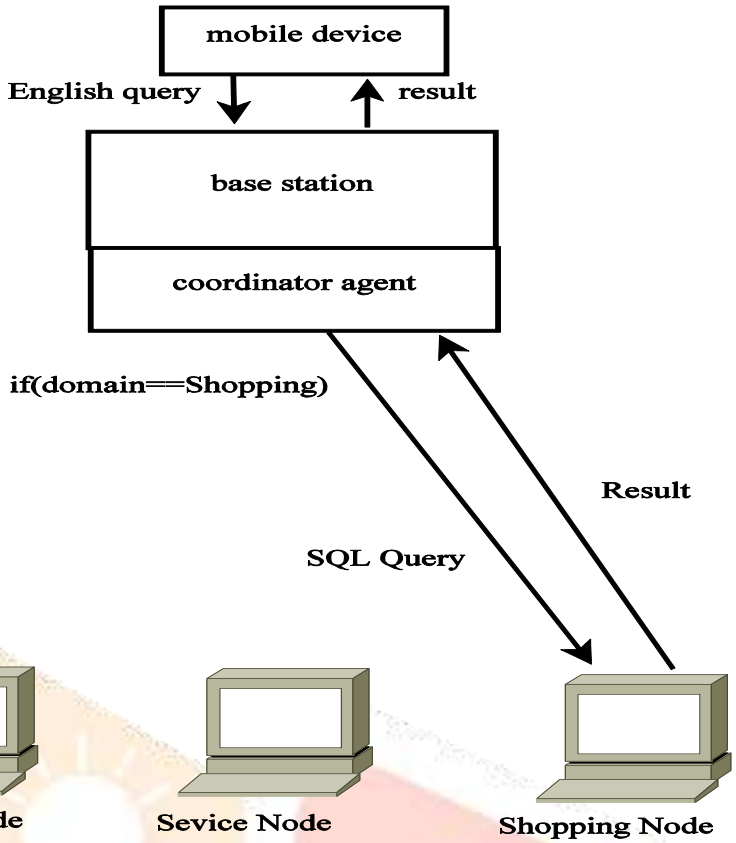

Figure 6: Data flow diagram Level 1

4. .**Implementation Details**

Admin details

Homepage for administrator of the system is shown. Administrator has all privileges for accessing all databases. He can add , edit or remove any of the record from database.

Figure 7: Admin details

Admin can also view some queries which are forwarded by users for further processing, accordingly admin can insert stop words in stop word dictionary.

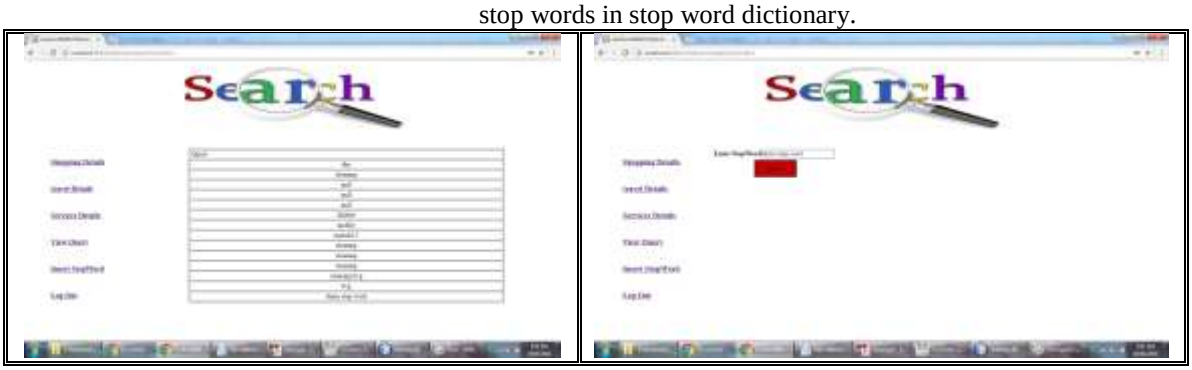

Figure 8: Insert stop word

User Implementation

The homepage for user is shown in form.This window appears when user successfully login to system.It provides user facility to search data.Also it provides logout button. The search window for a user is shown in next snapshot. It provides a text field to provide a query for searching data. The message is displayed when we are providing a query for data which is not available is shown in next snapshot. In such case this query is forwarded to administrator for further processing.

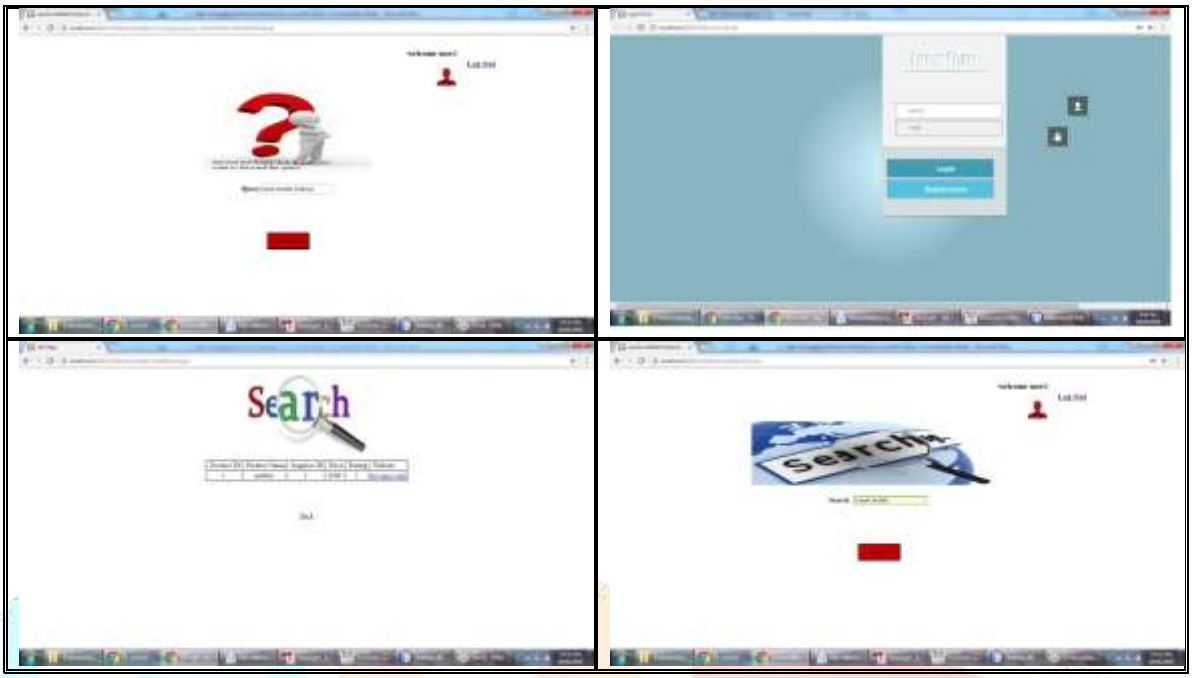

Figure 9: User implementation

5. Results

As graph is showing results of base paper are compared with proposed system.The time required to retrieve data is plotted against number of database servers.It shows that time remains constant with the increasing number of servers.Because only single agent is created by proposed system.

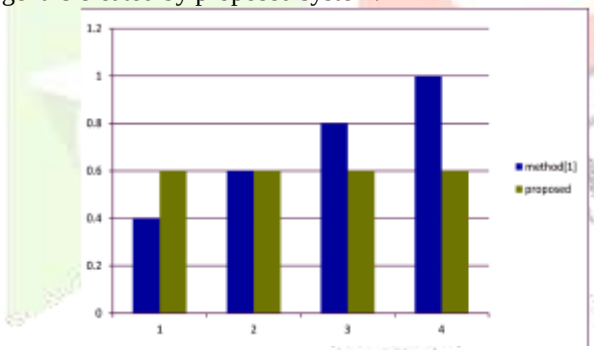

Figure 10 Graph 1

Following graph is showing comparisons of time taken vs. no. of agents.In this graph time taken by agents remain constant because there is no extra processing is required when no. of agents are increased,because all agents are working independent of each other.As compared to base paper methodology explained in [1] results are improved as shown in graph 2.

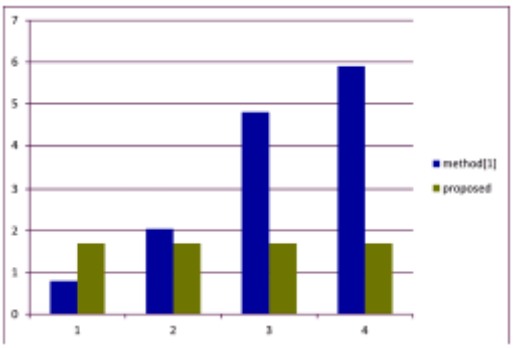

Figure 11: Graph 2

6. Future Scope:

This system will work best in the databases where we are well aware of what type of data is stored in distributed databases. In future this system can be improved by working about security of agents, implementing the same concept for homogeneous type of distributed database. Also to work on maintenance of directory about data stored in databases which specifies domain of the databases.

7. Conclusion

Hence, we have implemented retrieval system by using mobile agent & natural language processing. This approach allows user to access data remotely. Also this system is developed as server pages, so it can be accessed by any device either a computer or smart mobile device which is having a web browser. This feature is giving freedom to user or even to administrator of the system to access the system from any place at any time. Also one more feature of this system is that for a user it is not necessary to know the syntax of writing the query. User just needs to provide request of data by using simple English phrase. This system also improves retrieval process by focusing on the nodes which has answer to the query. In this way we can improve system performance by using mobile agent system and Natural Language Processing.NLP approach will improve system efficiency, reduce system overhead, data congestion and network overhead.

8. References

[1] M. Murali Dr. R. Srinivasan, *Multi-Agent System for Distributed Data Retrieval using PQR Approach* International Journal of Computer Applications (0975−8887) Volume 29 No.3, September 2011.

[2] Ekta Agrawal et. al, *Natural Query Generation By Single Token Matching* International Journal of Emerging Technology and Advanced Engineering (ISSN 2250-2459, Volume 2, Issue 8, August 2012).

[3] D Gavalas, G. Tsekouras, C. Anagnostopoulos , *A mobile agent platform for distributed network and systems management* Journal of Systems and Software 82 (2) (2009) pp.355 − 371.

[4] Anbukodi.S,Muthu Manickam.K, *Reducing Web Crawler Overhead using Mobile Crawler* PROCEEDINGS OF ICETECT 2011.

[5] Li Yang et.al., *Application Research on distributed database access of Down-Hole Operation producingsystem based on Mobile-Agent Technology* 3rd International Conference on Advanced Computer Theory and Engineering(ICACTE) 2010.

[6] Stavros Papastavrou et.al., *Mobile Agents for World Wide Web Distributed Database Access* IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 12, NO. 5, SEPTEMBER/OCTOBER 2000.

[7] Brewington, B. et al., *Mobile agents in distributed information retrieval* Klusch, M.(ed.) Intelligent Information Agents, 355 − 395. Springer, Heidelberg (1999).

[8] T. Kawamura, S. Joseph, A. Ohsuga, S. Honiden,*Designing multi-agent systems based on pairwise agent interactions* IEICE Transactions on Information Systems E 84-D (8) (2001) 968 − 980.

[9] L. Ismail, D. Hagimont, pp. 306 − 313., *A performance evaluation of the mobile agent paradigm* Fourteenth Conference on Object-Oriented Programming, Systems, Languages and Applications.
(OOPSLA), Denver, November 1999.

[10] Menczer F., *Complementing Search Engines with Online Web Mining Agents* Decision Support Systems 35, 195 − 212 (2003).

[11] W. Theilmann, K. Rothermel, *Optimizing the dissemination of mobile agents for distributed information filtering* IEEE Concurrency 8 (2) (2000), 53 − 61.

[12] Nguyen, N.T., Ganzha, M., Paprzycki, M. *A Consensus-based Multi-agent Approach for Information Retrieval in Internet* Alexandrov, V.N., van Albada,G.D., Sloot, P.M.A.,Dongarra, J.J. (eds.) ICCS 2006.LNCS, vol. 3993, pp. 208−215. Springer, Heidelberg (2006).

[13] JADE homepage *http://jade.tilab.com/.*

[14] FIPA website *http://www.fipa.org.*

[15] R. Alonso, H.F. Korth, *Database system issues in nomadic computing* Proceedings of ACM SIGMOD Conference on Management of Data, 1993, pp.388 − 392.

[16] Y. Fu, S. Madria,*Multi-layered databases for intelligent query answering in mobile environments* International Workshop on Reliable and Secure Applications in Mobile Environments, New Orleans (also invited paper in NSF workshop), October, 2001.

[17] Tainchi Lu, Chinghao Hsu , *Mobile agents for information retrieval in hybrid simulation environment*,Journal of Network and Computer Applications 30 (2007) 244 − 264.

[18] Katia P. Sycara , *Multiagent Systems* American Association for Artificial Intelligence SUMMER 1998.

[19] Aglet web site h*ttp://aglets.sourceforge.net*

[20] Robart Gray et.al., *DAgents:Security in a multiple-language multi agent system* from book Mobile Agent and Security, 154 − 187,Springer-Verlog,1998