Volume 4, Issue 1, January 2014 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Design and Architecture for Web Graph Mining Base Recommender System for Query, Image and Social Network using Query Suggestion Algorithm and Heat Diffusion Method

Asst. Prof. Sonal Patil
Information Technology
G.H. Raisoni Institute of Engineering and Management
Jalgaon, India

Ankush Mahajan
Computer Science and Engineering
G.H. Raisoni Institute of Engineering and Management
Jalgaon, India

Abstract— Recommendation techniques is now a day's very important. various kinds of recommendations are done on the Web, example movies, music, images, books recommendations, query suggestions and tags recommendations, etc. it is not important that what kind of data sources are used for the recommendations, these data sources can be modeled or taken in the various types of graphs. In this paper, We are discussing a general framework on Web graphs mining based recommender system for Query, Image-tags and Social Network using Query Suggestion Algorithm and Heat Diffusion Method 1) we first propose a novel diffusion method which shows similarities between different nodes and generates recommendations; 2) secondly we illustrate Architecture of web graph mining base recommender system which function for query, image and social network using three different data sets respectively 3) And at the end we illustrate the design using UML diagrams as a Visual models for the same; Hao Ma, Irwin King et al in their paper "Mining Web Graphs for Recommendations" have proposed a system for query suggestion and image recommendation using heat diffusion by taking reference of that we are adding a social recommendation. Aim of this paper is to review heat diffusion and depict the architecture and design for the proposed system.

Keywords— Recommendation, diffusion, query suggestion, image recommendation, social recommendation

I. INTRODUCTION

Organize and use information effectively and efficiently is a very difficult task now a days. Mining useful information on web from different sources is also difficult. To satisfy the need of information of web user, recommender system has been well studied in academic and in industries. There are various recommender systems available on web. For example Movielens, Which recommends movies to user based on the already collected and well organized data which was taken through a feedback from the previous users who rate the movies on that web site. Another beautiful example can be a online shopping web site i.e. Amezon.com. Generally, recommender systems are based on Collaborative Filtering [1], which is a technique that automatically predicts or infer the interest of an active user by collecting rating information from other similar users or items. This consideration of collaborative filtering is that the active user or a current user will prefer those items which other similar users prefer. Based on this simple concept, collaborative filtering has been widely employed in some large, well-known commercial systems, including product recommendation at Amazon, movie recommendation at Netflix, etc. In this paper we will see the architecture and the visual models of the web graph mining base recommender system which works for the query, image-tags and social network using query suggestion and heat diffusion method. Visual Models are represented through a UML diagrams.

II. HEAT DIFFUSION

Heat diffusion is a physical phenomenon. In a medium, heat always flows from a position with high temperature to a position with low temperature. Recently, heat diffusion-based approaches have been successfully applied in various domains such as classification and dimensionality reduction problems [2]. Actually, the process of people influencing others is very similar to the heat diffusion phenomenon. In a social network, the innovators and early adopters of a product or innovation act as heat sources, and have a very high amount of heat. These peoples start to influence others, and diffuse their influence to the early majority, then the late majority. Finally, at a certain time point,

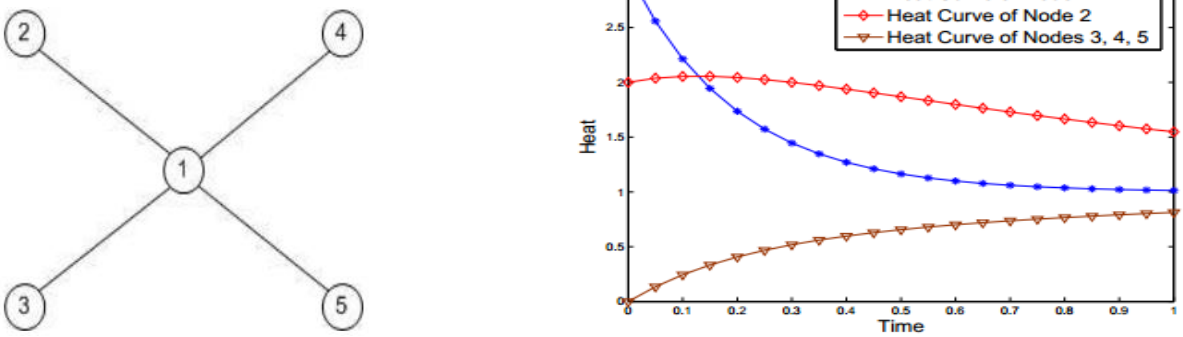
heat is diffused to the margin of this social network, and the laggards adopt this product or innovation.

The heat flows throughout a geometric manifold with initial conditions can be described by the following second order differential equation:

$$\frac{\partial f(x, t)}{\partial t} - \Delta f(x, t) = 0,$$
$$f(x, 0) = f0(x) \qquad (1)$$

Where f (x, t) is the temperature at location x at time t,beginning with an initial distribution f0(x) at time zero, and

Δf is the Laplace-Beltrami operator on a function f [2].



(a) A small undirected social network (b) Curve of heat change with time

Fig.1. Heat diffusion examples on an undirected graph [2]

In the light of the several successful existing applications of the heat kernel, it is natural to investigate the heat equation whose special solution is the heat kernel Kt(x, y)[2]. The heat kernel Kt(x, y) describes the heat distribution at time t diffusing from the initial unit heat source at position y, and thus describes the connectivity (which is considered as a kind of similarity) between x and y[2]. However, it is very difficult to represent the social network as a regular geometry with a known dimension. This motivates us to investigate the heat flow on a graph. The graph is considered as an approximation to the underlying manifold, and so the heat flow on the graph is considered as an approximation to the heat flow on the manifold. In this paper, we model a social network as a graph, and each user in the social network is defined as a node on this graph. The relationships between peoples are represented by edges that connect nodes. Hao Ma, Haixuan Yang et al in their paper Mining Social Networks Using Heat Diffusion Processes for Marketing Candidates Selection, have explained in detail about heat diffusion on directed and undirected graph as well as on directed social network with prior knowledge [2].

III. ARCHITECTURE OF WEB GRAPH MINING BASE RECOMMENDER SYSTEM

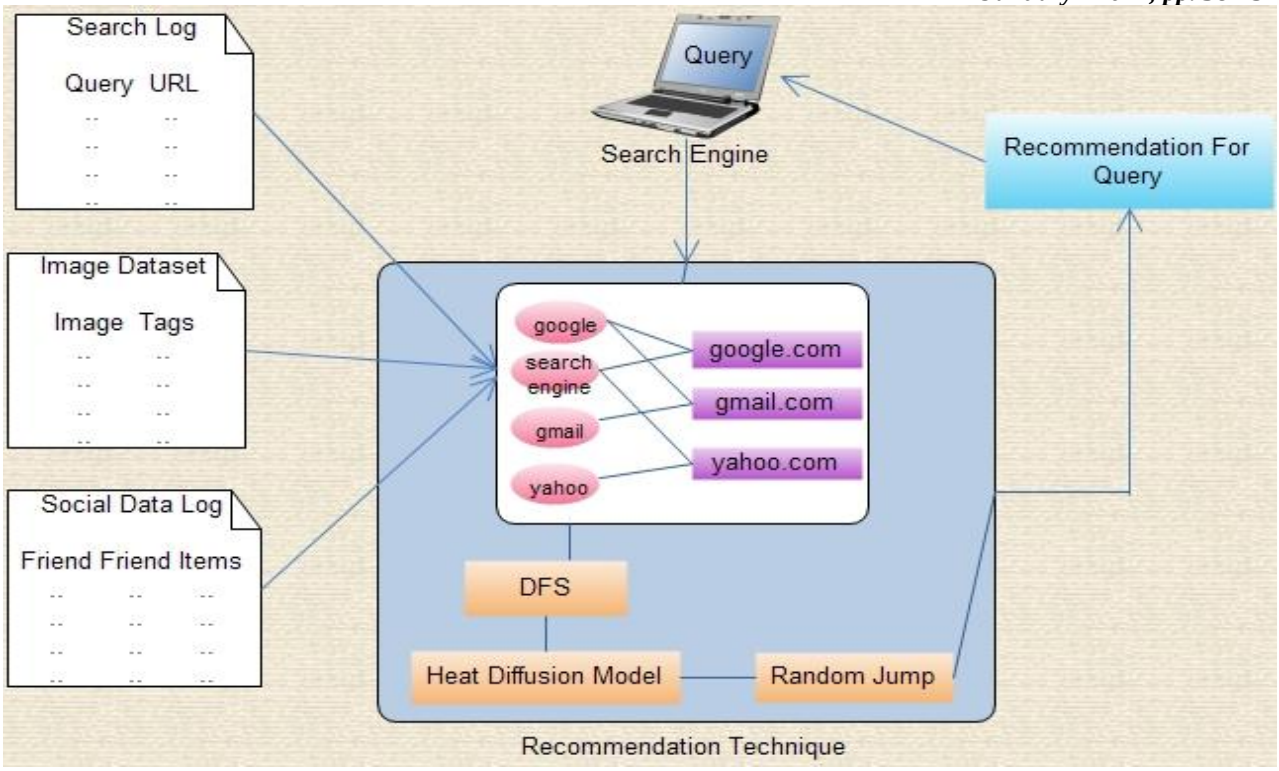


Fig.2 Web graph mining base recommender system Architecture

System architecture is the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures of the system. System architecture can comprise system components, the externally visible properties of those components, the relationships (e.g. the behavior) between them. It can provide a plan from which products can be procured, and systems developed, that will work together to implement the overall system. In our web graph mining base recommender system we are using three different data sets, so that recommender system will fetch a appropriate data from the data set whenever required.

Above fig.2 shows three different data sets Query-url, Image-tag and Friend-Item for Query suggestion, Image recommendation and social recommendation respectively. Other box which shows a recommender Technique which comprises a graph construction, heat diffusion model and random jump. In graph construction we propose to construct a graph for different data sets. These graphs will be bipartite graphs. In query recommendation query and url will be node and there will be connecting edges between them, and the graph will be constructed on the basis of the available data at hand in the data set which will be a historical data.

IV. DESIGN USING UML DIAGRAMS AS VISUAL MODELS

A. Use case diagram

A use case illustrates a unit of functionality provided by the system. The main purpose of the use-case diagram is to help development teams visualize the functional requirements of a system, including the relationship of "actors" to essential processes, as well as the relationships among different use cases. Use-case diagrams generally show groups of use cases, either all use cases for the complete system, or a breakout of a particular group of use cases with related functionality to Show a use case on a use-case diagram, you draw an oval in the middle of the diagram and put the name of the use case in the center of, or below, the oval. To draw an actor (indicating a system user) on a use-case diagram, you draw a stick person to the left or right of your diagram. Following diagram shows the relationships of the user or actors with the use cases which are shown in a oval shape.

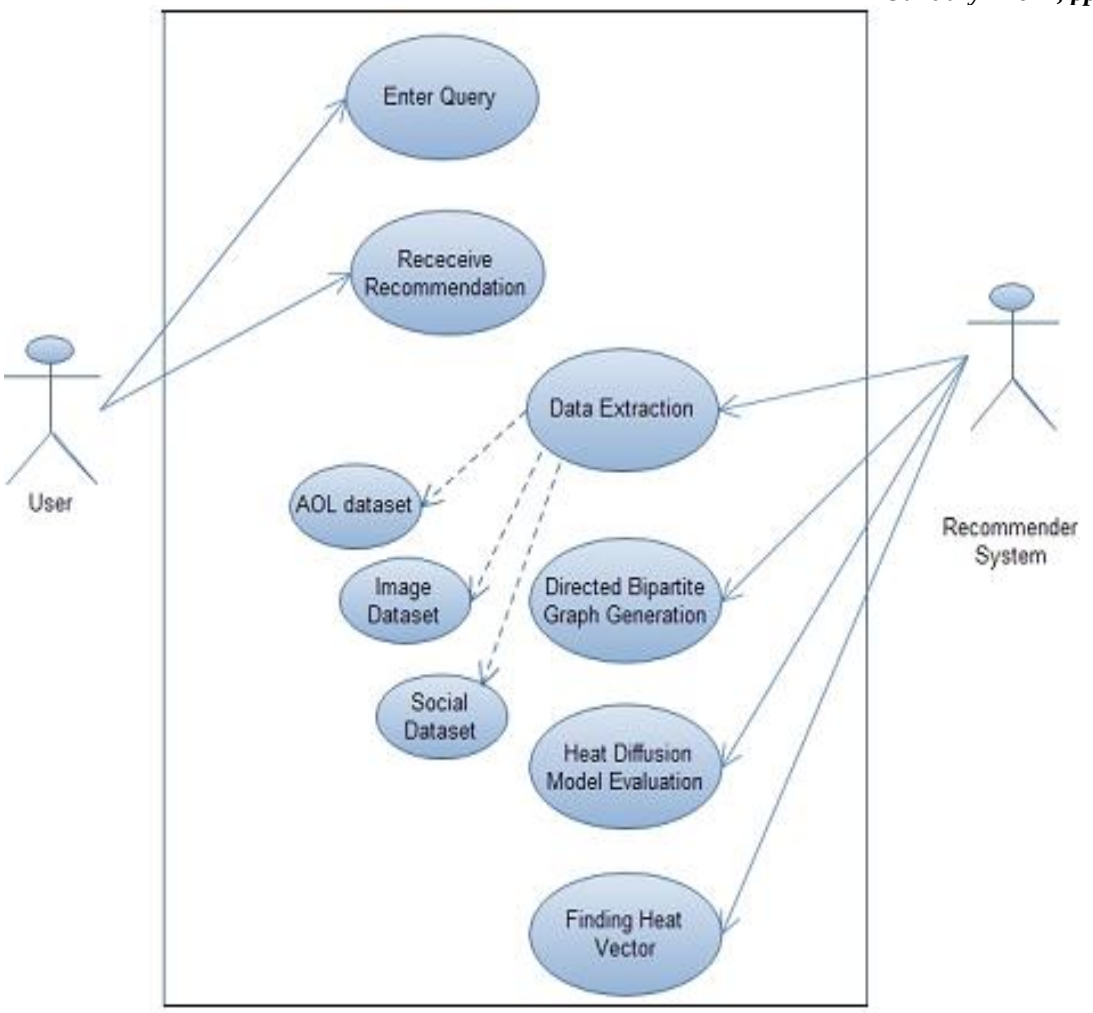


Fig.3. Use case Diagram

B. *Activity diagram*

Activity diagram is typically used for business process modeling, for modeling the logic captured by a single use case, or for visualizing the detailed logic of a business rule. Complicated process flows in the system are captured in the activity diagram. Similar to a state diagram, an activity diagram also consists of activities, actions, transitions, initial and final states, and guard conditions. However, difference is state diagrams are in context of simulation while activity gives detail view of business logic. Activity diagrams are "less technical" in appearance, compared to sequence diagrams, and business-minded people tend to understand them more quickly.

Following activity diagram shows how a business process flows from activity to activity. For three different operation activity diagram takes three different routes from three different datasets i.e. query-url dataset, image-tag dataset and social dataset. At each dataset once a query is submitted the bipartite graph is generated, heat diffusion method is applied and then depending upon the heat values the results are sorted out and final query is recommended.

Fig.4 Activity diagram

C. State diagram

A state diagram, as the name suggests, represents the different states that objects in the system undergo during their life cycle. Object change in response to certain simulation so this simulation effect is captured in state diagram. Therefore, it has a initial state and final state and events that happen in between them. Whenever you think that some simulations are complicated, you can go for this diagram. Following state diagram depict all the necessary states of the system and how a control transfer from one state to another state. The state diagram models the different states that a class can be in and how that class transitions from state to state. It can be argued that every class has a state, but that every class shouldn't have a state diagram. Only classes with "interesting" states, that is, classes with three or more potential states during system activity should be modeled.

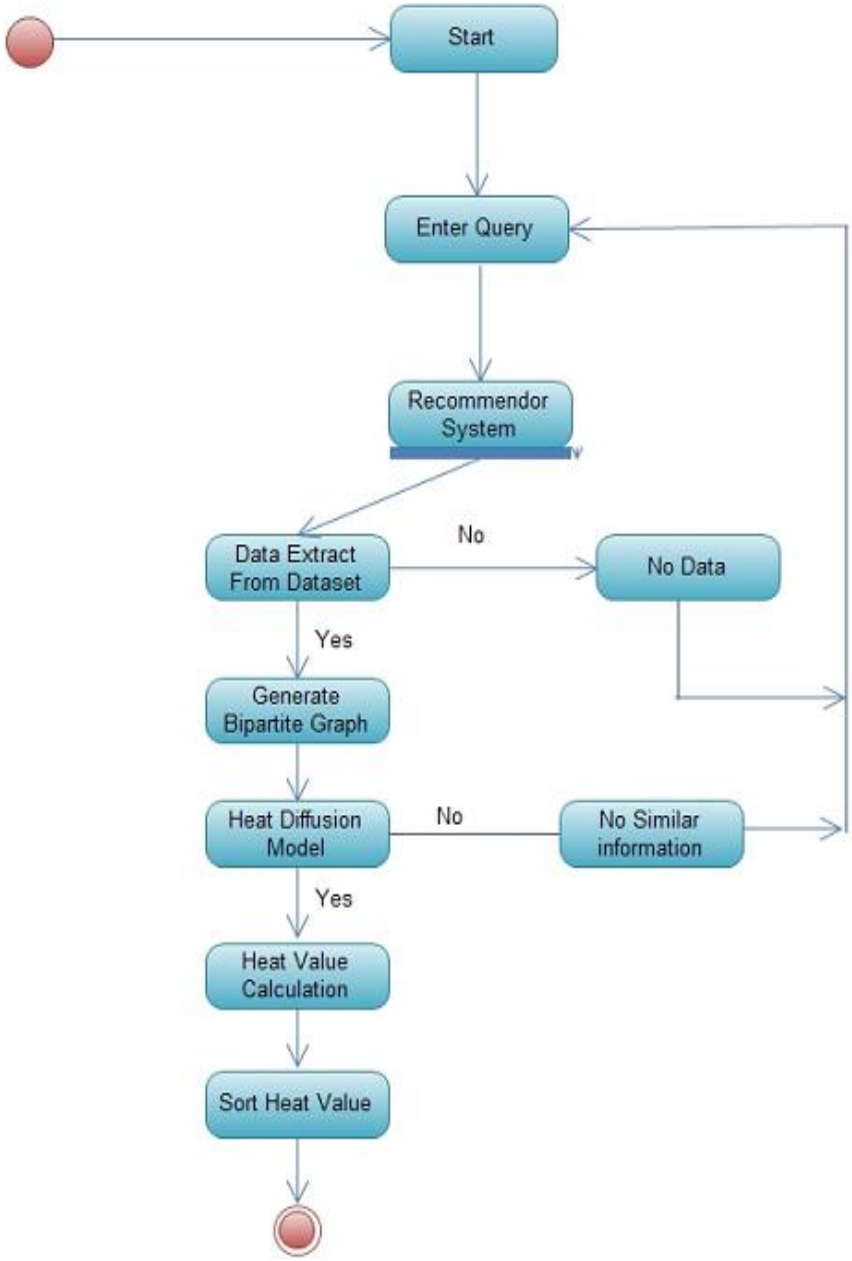


Fig.5 State diagram

D. Data flow diagram

A data flow diagram (DFD) is a graphical representation of the "flow" of data through an information system, modeling its process aspects. DFDs can also be used for the visualization of data processing (structured design). DFD shows what kinds of information will be input to and output from the system, where the data will come from and go to, and where the data will be stored.

1) Level 0 Diagram

Level 0 diagram shows all the processes that comprise the overall system, it also shows how information moves from and to each process and finally Adds data stores. Following figure shows all the processes which are in our recommender system, namely user, recommender system in middle and recommended query and finally a click-through data, which is a historical data stored in a database.

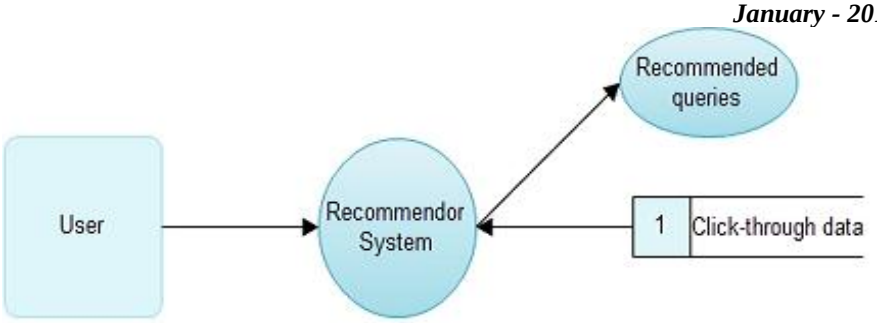


Fig.6 Level 0 DFD Diagram

2) *Level 1 Diagram*

Level 1 diagram shows all the processes that comprise a single process on the level 0 diagram. It also shows how information moves from and to each of these processes and more detail the content of higher level process. Level 1 diagram may not be needed for all level 0 processes.

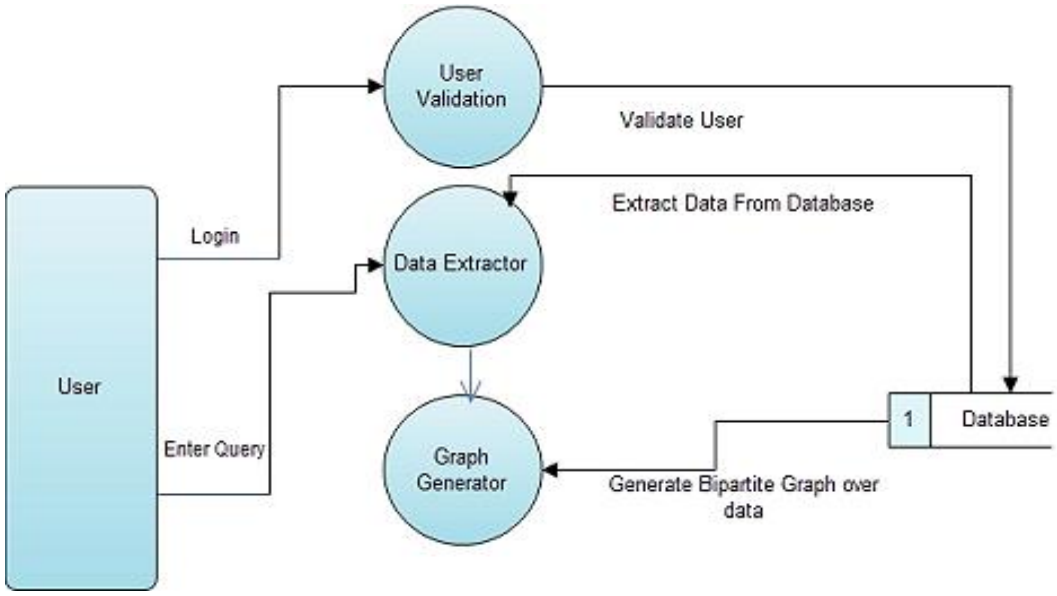


Fig.7 Level 1 DFD Diagram

3) *Level 2 Diagram*

There can be one level 2 DFD for each process of the Level 1 DFD. Level 2 shows a process broken down into greater detail. Level 2 Diagrams are only necessary where the Level 1 process is more complex, and where the particular process is relevant to the analysis. Following figure shows a level 2 diagram of our system.

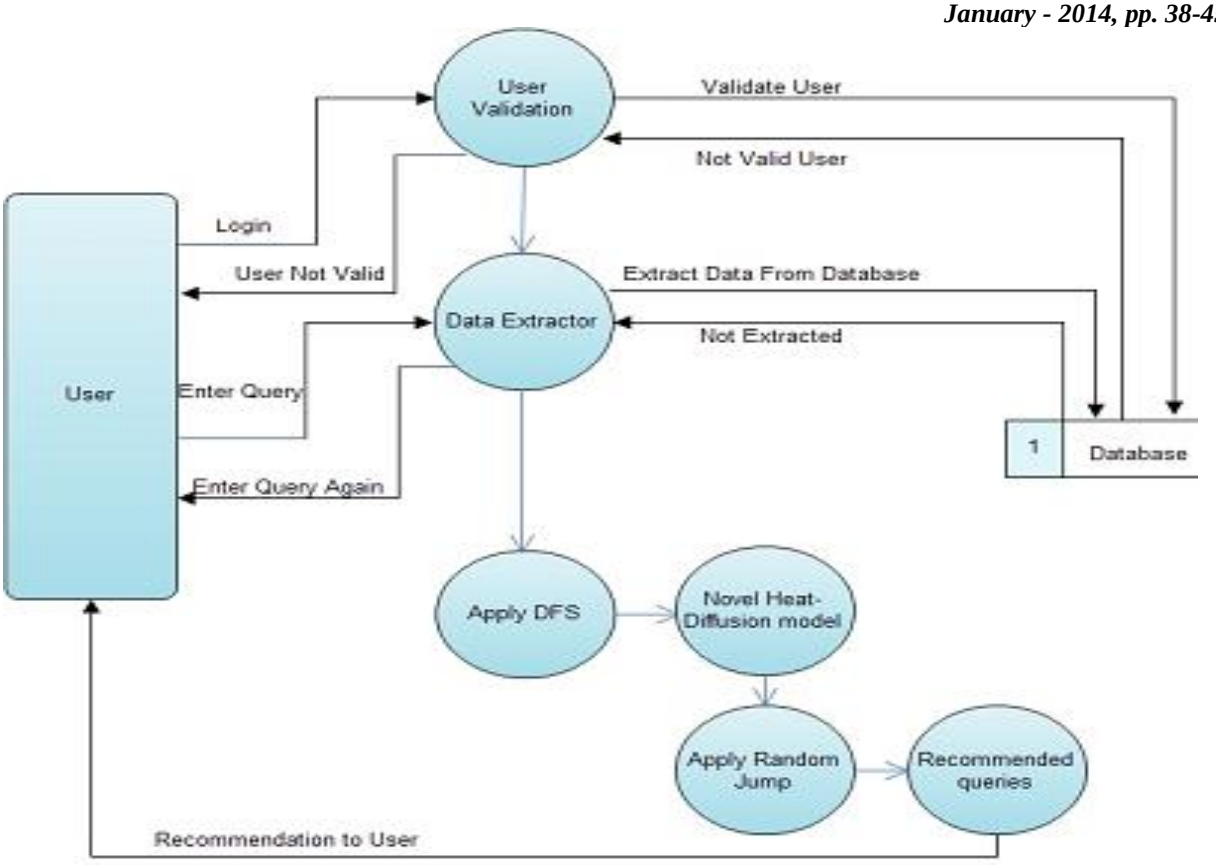


Fig.8 Level 2 DFD Diagram

E. ***Deployment diagram***

Deployment diagrams show the hardware for your system, the software that is installed on that hardware, and the middleware used to connect the disparate machines to one another. It shows how the hardware and software work together to run a system. In one, line its shows the deployment view of the system.

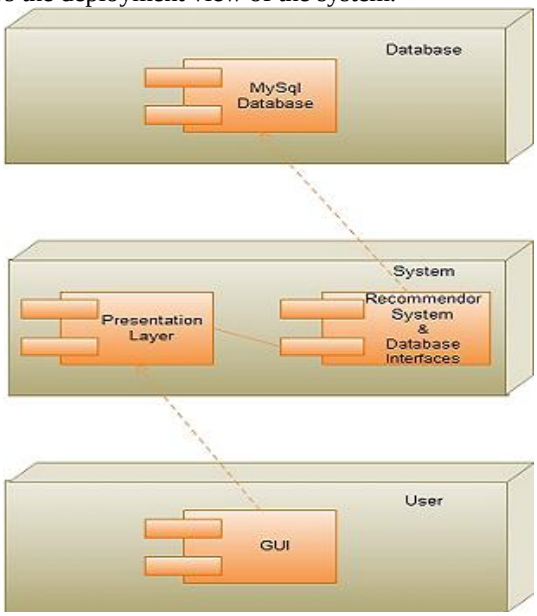


Fig.9 Deployment Diagram

F. **Sequence diagram**

Sequence diagrams can be used to explore the logic of a complex operation, function, or procedure. They are called sequence diagrams because sequential nature is shown via ordering of messages. First message starts at the top and the last message ends at bottom. The important aspect of a sequence diagram is that it is time-ordered. This means that the exact sequence of the interactions between the objects is represented step by step. Different objects in the sequence diagram interact with each other by passing "messages". Above Sequence diagram shows that how the interaction between object is represented step by step.

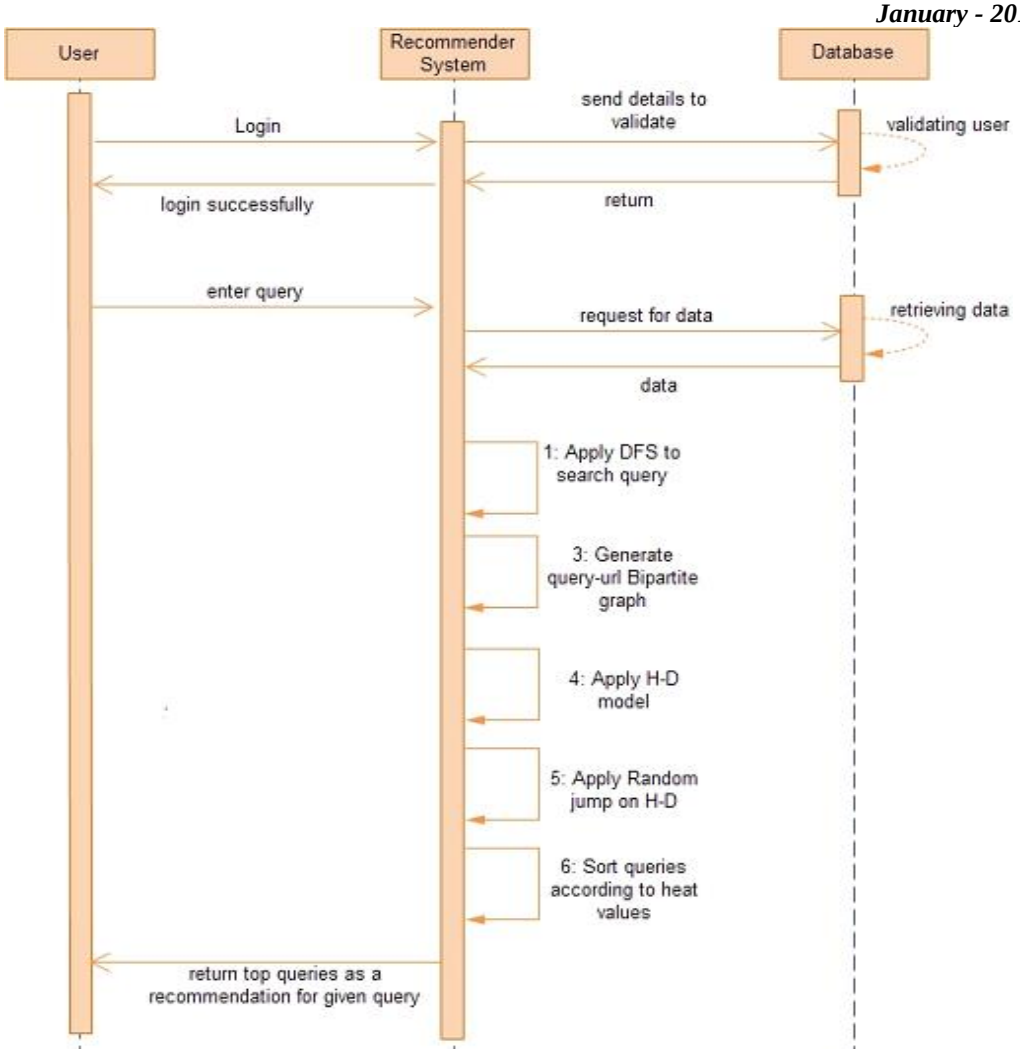


Fig.10. Sequence Diagram

V. CONCLUSION

In this paper we have seen the architecture of the web graph mining base recommender system for query, image, and social network. We also saw the visual model of the same, which was depicted in UML. Various UML's are shown i.e. use case diagram, activity diagram, state diagram, data flow diagram (level 0, level 1, level 2), deployment diagram and ultimately a sequence diagram. Compendiously we studied about Design and architecture for the proposed system.

REFERENCES

[1] Hao Ma, Irwin King and Michael Rung-Tsong Lyu, "Mining Web Graphs for Recommendations ", IEEE Transaction on knowledge and data engineering, Vol.24, No.6, June 2012.

[2] Hao Ma, Haixuan Yang, Michael R. Lyu and Irwin King, "Mining Social Networks Using Heat Diffusion Processes for Marketing Candidates Selection", published in international conference on information and knowledge management – CIKM, 2008.

[3] R.A. Baeza-Yates, C.A. Hurtado, and M. Mendoza," Query Recommendation Using Query Logs in Search Engines," Proc. Current Trends in Database Technology (EDBT) Workshops ,pp. 588-596, 2004

[4] D. Beeferman and A. Berger, "Agglomerative Clustering of a Search Engine Query Log ,"KDD '00: Proc. Sixth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 407-416, 2000.

[5] J.S. Breese, D. Heckerman, and C. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering,"Proc. 14th Conf.Uncertainty in Artificial Intelligence (UAI),1998.

[6] J. Canny, "Collaborative Filtering with Privacy via Factor Analysis," SIGIR '07: Proc. 25th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval ,pp. 238-245, 2002.

[7] N. Craswell and M. Szummer, "Random Walks on the Click Graph,"SIGIR '07: Proc. 30th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval,pp. 239-246, 2007

[8] H. Cui, J.-R. Wen, J.-Y. Nie, and W.-Y. Ma, "Query Expansion by Mining User Logs," IEEE Trans. Knowledge Data Eng.,vol. 15, no. 4, pp. 829-839, July/Aug. 2003.

[9] A.S. Das, M. Datar, A. Garg, and S. Rajaram, " Google News Personalization: Scalable Online Collaborative Filtering," WWW '07: Proc. 16th Int'l Conf. World Wide Web, pp. 271-280, 2007.

[10] G. Dupret and M. Mendoza, "Automatic Query Recommendation Using Click-Through Data,"Proc. Int'l Federation for Information Processing, Professional Practice in Artificial Intelligence (IFIP PPAI), pp. 303-312, 2006.

[11] N. Eiron, K.S. McCurley, and J.A. Tomlin, "Ranking the Web Frontier,"WWW '04: Proc. 13th Int'l Conf. World Wide Web,pp. 309-318, 2004.

[12] J.L. Herlocker, J.A. Konstan, L.G. Terveen, and J.T. Riedl, "Evaluating Collaborative Filtering Recommender Systems," ACM Trans. Information Systems ,vol. 22, no. 1, pp. 5-53, 2004.

[13] G. Jeh and J. Widom, "Simrank: A Measure of Structural-Context Similarity," KDD '02: Proc. Eighth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining,pp. 538-543, 2002.

[14] W. Gao, C. Niu, J.-Y. Nie, M. Zhou, J. Hu, K.-F. Wong, and H.-W.Hon, "Cross-Lingual Query Suggestion Using Query Logs of Different Languages,"SIGIR '07: Proc. 30th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval,pp. 463-470, 2007.

