International Journal of Advance Foundation and Research in Computer (IJAFRC)
Volume 1, Issue 3, March 2014. ISSN 2348 - 4853

Domain Personalized Multimedia Encyclopedia Search by Web Mining

Archana Y. Panpatil (ME-II, Computer)

Department of Computer Engineering,M.I.T. College of Engineering, Pune.University of Pune
Email:archana.panpatil@gmail.com

A B S T R A C T

Most present day search engines have a different way in the sense that same search results are returned for all users who submit the same query at certain times. They don't take the users interests and preferences in the retrieval process. Integrating user context related search in the retrieval process helps to deliver more customizable search results thus provides a personalized search experience for the user. Domain personalizing web search involves the process of identifying user interest, according to the domain during interaction with user, and using that information to deliver results that are more relevant to the user. The proposed system consists of Support vector machine algorithm using latent semantic analysis for searching domain wise semantic and customizable result. We will propose a system that will describe a compressive framework, giving support to a wide range of personalization facilities in a multimedia content search, along with available content search to provide users with personalized content search, ranking, retrieval, user friendly, interesting and interactive system.

Index Terms: Multimedia mining, support vector machine (SVM), semantic clustering, recommended links, domain based search engine.

I. INTRODUCTION

Today, internet search engines have become an indispensable part of our life. They have enabled mass participation and collaboration by hundreds of millions of people around the world. Now-a-days people are able to find all sorts of information instantly from anywhere. Search engines are also included within large web sites such as social networking sites, e-commerce sites, and corporate sites [1]. The exceedingly difficult nature of the problem of understanding user intent and matching it with the world's accumulated knowledge stored on the World Wide Web has attracted large scale research and development efforts from the academia as well as the industry. Also search engines perform a larger role in commercial applications; the desire to increase their effectiveness grows [2]. However, search engines order their results based on the small amount of information available in the user's queries and by web site popularity, rather than individual user interests. Thus, every user gathers the same results for the same query, even if they have widely different interests and backgrounds. To address this issue, interest in personalized search had grown in the last few years, and user profile creation is an important component of any personalization system. Explicit customization has been widely used to personalize the look and content of many websites.

Personalization is the process of presenting the right information to the right user at the right moment. In order to learn about a user, systems must collect personal information, analyze it, and store the results of the analysis in a user profile. Information can be collected from users in two ways: explicitly, for example asking for feedback such as preferences or ratings; or implicitly, for example observing user behaviors

such as the time spent reading an on-line document. Commercial systems tend to focus on personalized search using an explicitly defined profile. Explicit construction of user profiles has some drawbacks. The users may provide inconsistent or incorrect information, the profile is static whereas the user's interests may change over time, and the construction of the profile places a burden on the user that she may not wish to accept. On the other hand, implicitly created user profiles do not place any burden on the user and they provide an unbiased way to collect information. Thus, many research efforts are underway to implicitly create accurate user profiles. User profiles can also be divided in other two groups: the ones representing user's preferences (e.g., search engines preferred, types of documents) and the ones representing user's interests (e.g., sports, photography.)We focus our attention on maintaining user's interests. To achieve effective domain personalization, these profiles should be able to distinguish between long-term and short-term interests. Several systems have attempted to provide personalized search based upon user profiles that capture one or more of these aspects.

Personalization techniques that incorporate user interests and preferences into the search may address some of these issues. Profile of user involvement is a process of learning which is involved broadly in personalization. It is then used to deliver personalized content to the user. Personalization of web search usually involves filtering or re-ordering the results returned from a standard search engine, or directly incorporates user interests into retrieval process itself to present personalized results. Caved in a query, a personalized search can provide different answers for different users or even different results for the same user in different contexts. A user profile that represents the interests of a specific user can be used for supplementary information about the search that, currently, is represented only by the query itself. According to the user's profile web will provide a number of links priority wise. The user will check among those links which link is useful or accurate for his search. If it is correct, then user will further proceed, if not, then the user will type keyword on search bar. After the keyword search user will get results and semantic clustering will be done. In semantic clustering keyword will be matched with already displayed text, the result will be displayed according to search of links with highest priority [1] [2] [3].

II. RELATED WORK

In 2012, the researchers Richang Hong, Zhen-Jun, Yue Gao, Tat-Seng Chua, and Xindong Wu "Multimedia Encyclopedia Construction by Mining Web Knowledge" briefly introduced concept of Mediapedia which aims to construct Multimedia encyclopedia by mining web knowledge. The Mediapedia distinguishes itself from traditional encyclopedia in its multimedia presentation, fully automated, production, dynamic update and the flexible framework where each module is extensible to potential application. The drawback of paper was the result not as good as expected. Assuming that the distribution of images from Flicker can automatically make a tradeoff between ''typical'' and ''different''. However, this is not the truth for all the concepts [1] [6].

In 2012, the researchers Xiaoou Tang, Ke Liu, Jingyu Cui, Fang Wen, and Xiaogang Wang "Intent Search: Capturing User Intention for One-Click Internet Image Search" briefly introduced the concept of a novel Internet image search approach which only requires one-click user feedback. The drawback of paper was sometimes duplicate images show up as similar images for the query, the quality of reranked images [2] [8].

In 2013, the researchers Ting Yao, Chong-Wah Ngo, and Tao Mei," Circular Reranking for Visual Search", briefly introduced the concept of a circular reranking which explores information exchange and reinforcement for visual search ranking. The drawback of paper was The degree of improvement, though,

is limited by how accurate the modality importance and fusion weights can be estimated, which could be notice from empirical results when compared to the oracle setting of circular reranking [3] [7].

In 2012, the researchers Bo Gong, Linjun Yang, Chao Xu, Xian-Sheng Hua" Ranking Model Adaptation for Domain-Specific Search", vertical search engines emerges and increases dramatically, a global ranking model, which is trained over a dataset sourced from multiple domains, does not give a sound performance for each specific domain with special topics, document formats and domain-specific features. The drawback of paper was time-consuming for learning [4] [5].

III. PROPOSED SYSTEM

The plan is to build a structure that always provides domain wise personalized results.

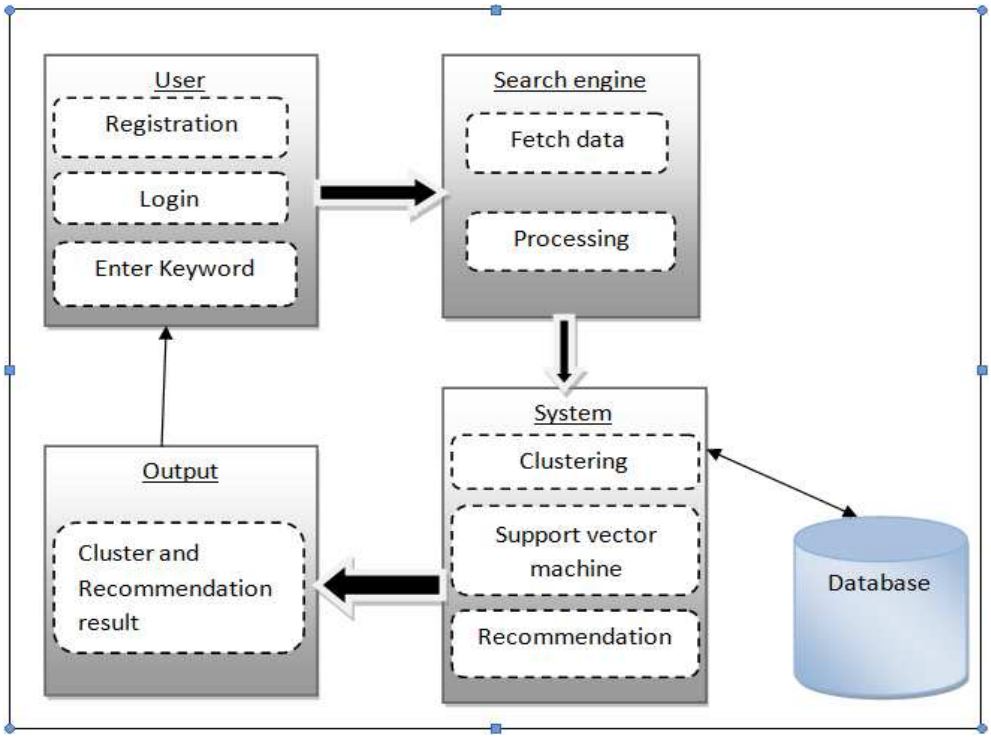

Fig 1. Proposed System Architecture

The explanation of internal work of the system is as follows:

1. User login to maintain user history profile
2. Then users enter the keyword and fetch relevant data from google. And then perform processing using Support vector machine algorithm using LSA.
3. Link analysis of a domain is performing using LSA with the help of tags.
4. Semantic Clustering of domains are performed when similar domains cluster occurs.
5. Query related links recommended domain wisely
6. If not, then the data is fetched from Google and n number of links visited or hit by the number of users will be displayed
7. The profile will be updated.
8. According to domains displayed to user, user will search data or images in an easy and fast way.
9. Output will be displayed in the form of multimedia including domain cluster and recommended links for customizable keyword search.

The key practical challenges are 1) Display customizable search result domain wisely, 2) Update User history profile dynamically, 3) Perform personalization, learning and re-ranking on search result, 4) Display relevant result to the user and improve the performance of the system. 5) Improved the quality of re-ranked images, 6) Display result in multimedia format.

IV. Mathematical Model

Let S be a system which provides domain personalized multimedia encyclopedia search by using web mining .Which provides recommended links, multimedia content search facilities and integrating user context in the retrieval process can help deliver more customizable search results, thereby providing a domain personalized search experience to the user. The available content search provide users with domain wise personalized content search, recommended link, ranking, retrieval, user friendly, interesting and interactive system. Suppose vector X_i and Y_i is the element of a training data set, where vector X_i is the feature vector (with information about features) and Y_i is the label (which classifies the category of vector X_i). A typical SVM classifier for such data set can be defined as the solution of the following optimization problem. The solution of the above optimization problem can be represented as a linear combination of the feature vectors X_i.

S= {I, F, R, Er, Q, P, D, H}

Where,

I= set of input query.
Er=set of error state
Q=set of query analysis for features extracted.
P=set of domain personalization search
H=set of profile history.
D=set of domain name.
R=set of recommended links.
RK=set of ranking search.
F=set of final state results observed in terms for recommended links, text and visual document search.

A. Set Theory

I= {I_1, I2, I3..... ....In| In □ □}
Q= {Q_1, Q2, Q3..... ....Qn| Qn □ □}
P= {P_1, P2, P3..... ....Pn| Pn □ □}
R= {R_1, R2, R3..... ....Rn| Rn □ □}
RK= {RK_1, RK2, RK3..... ....RKn| RKn □ □}
H= {H1, H2, H3..... ....Hn| Hn □ □}
Er = {E1, E2, E3... Ek}
D= {D_1, D2, D3..... ....Dn| Dn □ □}
F= {}

B. Functions

a) f1 = Users login the system for maintain profile history.
b) f2=If users are not already register then firstly register.
c) f3= User enter query keyword.

d) f4=that query pass to domain personalized search engine.
e) f5= Match that query keyword with existing keyword in domain. And display that domains name by applying semantic clustering using SVM and AP algorithm.
f) f6=User select that domain as per their self-preferences.
g) f7=User get recommended links according to user history.
h) f8=Display the searched customizable personalized visual document result and recommended links simultaneously on single panel.

C. Mapping Diagram

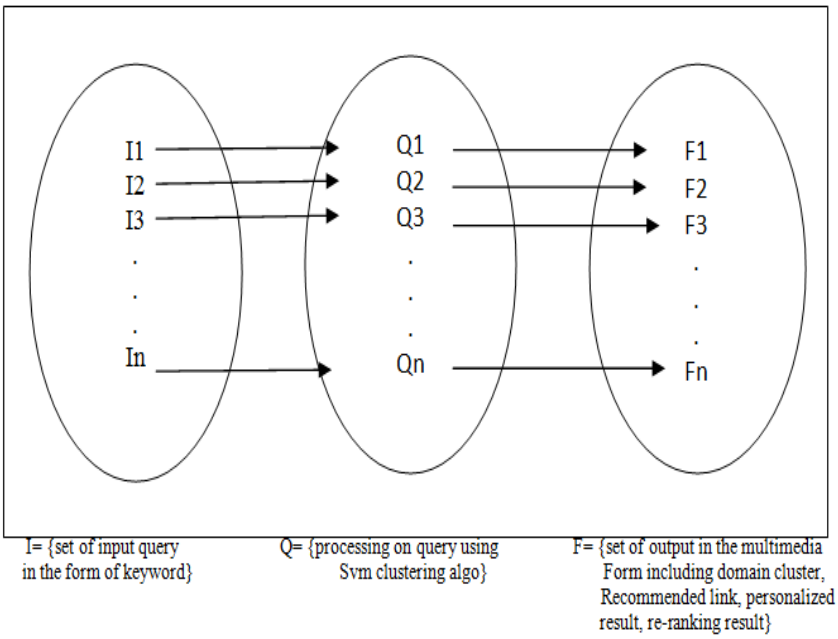

Fig 2. Mathematical module

V. Implementation Details

Efforts which are considered have been devoted to the implementation of efficient optimization method for solving the support vector machine dual problem.

The support vector machine (SVM) algorithm (Cortes and Vapnik, 1995) probably uses kernel learning algorithm. Relative robust pattern recognition performance using well established concept in optimization theory is achieved.

Despite of mathematical class, the implementation of efficient SVM solvers has diverged from the classical methods of numerical optimization. This divergence is common to virtually all learning algorithms. The numerical optimization literature has focused on the asymptotical performance: how fast the accuracy of solution increases with computing time. In case of learning algorithms, two other factors reduce the impact of optimization accuracy.

To compute a linear SVM, only a single weight vector w needs to be stored, rather than all of the training examples that correspond to non-zero Lagrange multipliers. If the joint optimization succeeds, the stored weight vector needs to be updated to reflect the new Lagrange multiplier values. The weight vector update is easy, due to the linearity of the SVM.

A. Serialization Algorithm steps

1. Start.
2. User login to maintain a user's history profile
3. If users is not already register then firstly register, then proceed to next step.
4. After login user enter the keyword query.
5. If keyword matches with existing query those stored in profile history according to their domain wise.
6. Semantic Clustering of domains are performed using support vector machine, when similar domain keyword occur.
7. Than it displays recommended link according to domain.
8. According to domains displayed to user, user will search data or images in easy and faster way.
9. Output will be displayed in the form of multimedia including domain cluster and recommended links.
10. End.

B. Platform Details

The hardware and software requirements of the system are as follows:
Desktop Computer with following specifications
Operating System : XP/ Higher version of Window
Database : Oracle
User Designer Tool : Net beans/Eclipses
Software Component : JAVA Version 1.6 or Adv. Java
Hard Disk : 1 GB Minimum or onwards
RAM : 256 MB or Higher
Processor : Intel P Family or Equivalent
Display Options : VGA (104, 768)
Input Device : Keyboard, Mouse.

VI. Results

User enters the query. According to the user's profile web will provide a number of links priority wise. The user will check among those links which link is useful or accurate for his search. If it is correct, then user will further proceed, if not, then the user will type keyword on search bar. After the keyword search user will get results and semantic clustering will be done. In semantic clustering, keyword will be matched with already displayed text; the result will be displayed according to search of links with highest priority.

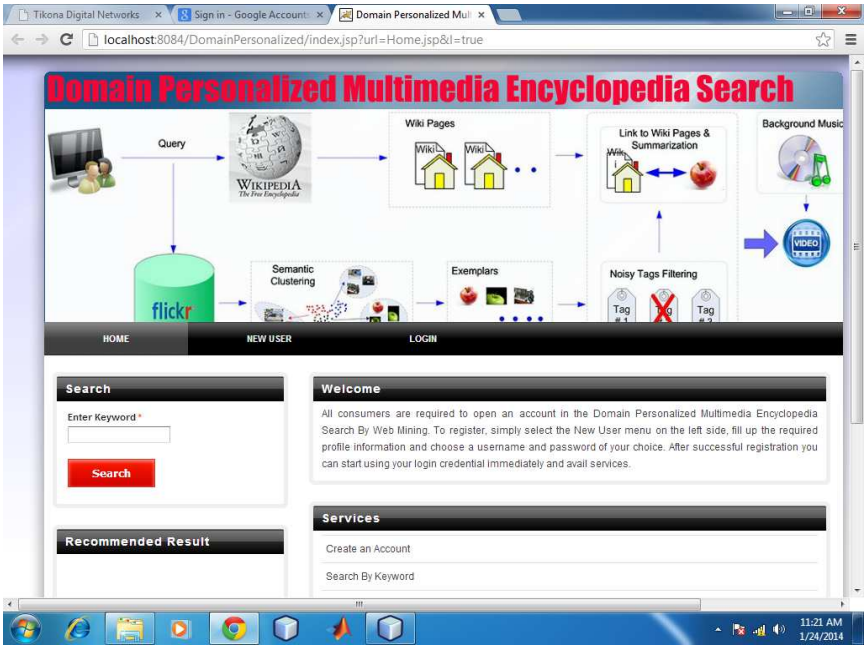

Fig 3. Domain Personalized Multimedia Encyclopaedia Search

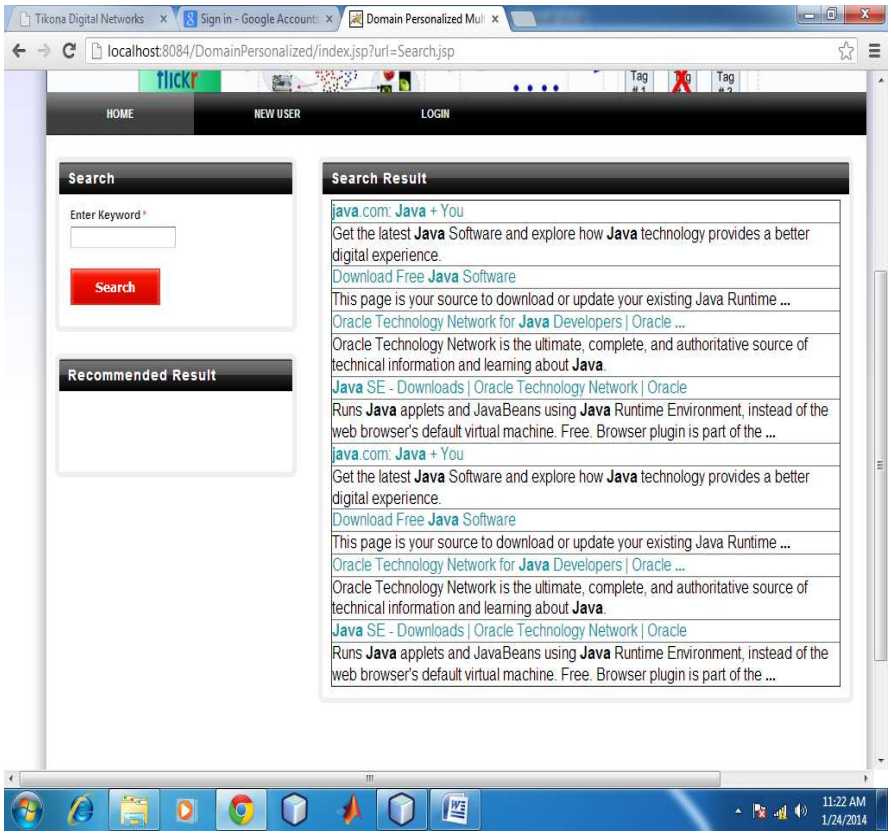

Fig 4. Recommended Pages

VII. Conclusion

This research is about personalized web search on multimedia content. Our approach involved building an interest profile based on his interaction with web search results and his browsing behavior according to domain wise search. Domain based Personalization of search results of multimedia content is achieved by ranking search results based on proximity to the user's interest profile. The strength to recognize user interests in a completely non-invasive way and the accuracy of the personalized results are some of the major advantages of our approach.

VIII. Acknowledgment

The authors thank Prof. Rekha Sugandhi (Guide) and Dr. Prasanna Joeg, Head of Computer Engineering Department, MIT College of Engineering, Pune (MH); for his kind support in providing laboratory infrastructure facility required for this research work.

IX. References

[1] Richang Hong, Zhen-Jun, Yue Gao, Tat-Seng Chua, Xindong Wu,"Multimedia Encyclopedia Construction by Mining Web Knowledge", 2012 Elsevier on Signal processing, July 2012, pp 2361-2368.

[2] Xiaoou Tang, Fellow, Ke Liu, Jingyu Cui, Fang Wen and Xiaogang Wang, " Intent Search: Capturing User Intention for One-Click Internet Image Search", 2012 IEEE Transaction on pattern analysis and machine intelligence, Vol. 34, no. 7, July 2012, pp 1342-1353.

[3] Ting Yao, Chong-Wah Ngo, and Tao Mei, " Circular Reranking for Visual Search", 2013 IEEE Transaction on image processing, vol. 22, no. 4, April 2013, pp 1644-1655.

[4] Bo Geng, Linjun Yang, Chao Xu And Xian-Sheng Hua," Ranking Model Adaptation For Domain-Specific Search", 2010 IEEE Transactions On Knowledge And Data Engineering, Vol. Xx, No. X, March 2010, pp 1-15

[5] D. Cai, X. He, Z. Li, W.Y. Ma, J.R. Wen, "Hierarchical clustering of WWW image search results using visual, textual and link information", in: Proceedings of ACM Multimedia, New York, USA, October 2004, pp 10-16.

[6] T.S. Chua, R. Hong, G. Li, J. Tang, "From text question–answering to multimedia QA", in: ACM Multimedia Workshop on Large-Scale Multimedia Retrieval and Mining (LS-MMRM), Beijing, China, October 2009, pp 19-23.

[7] R. Hong, J. Tang, H.K. Tan, S. Yan, C.W. Ngo, T.S. Chua, "Beyond search: event driven summarization for web videos", ACM Transactions on Multimedia Computing, Communications and Applications 7 (4) (2011) 35.

[8] T. Gonzalez, "Clustering to minimize the maximum inter cluster distance", Theoretical Computer Science 38 (2) (1985) 293–306.

