Web page and Image Recommendation based on web usage and domain knowledge

Ms.J.SathiaParkavi[1], R.Pavithra[2], M.Suriyaprabha[3], C.Vimala[4]

[1]Assistant Professor, Saranathan College of Engineering
[2,3,4] Student, Saranathan College of Engineering

Abstract— Generally Recommendation system recommend the user to satisfy their needs. Web page and Image recommendation System helps the user to select the correct Web pages and Images related to query given by them. This System can take the advantage of Semantic Reasoning capabilities to overcome the limitations of the current systems. (i.e)This system use the High Standard than the current system. And the main aim is to improve the Quality of Recommendation. When the user gives the search query, the system will collect all the web pages and images immediately. There are two types of Search queries. For Product based queries the system will provide the web pages and Images based on the Rating which is already provided by the users. And for Non product based search queries the web page and Image will be displayed based on the frequent use of the previous user. The Web pages and Images which is frequently used by other users previously, will take place at the top of the websites. And an Unwanted Advertisements will be avoided in this System .So that users can avoid the unnecessary access. Images are displayed on the structured manner and Distance, Skew, Entropy, Mean values are displayed for each images. Domain based inference method is used for model the User's Interests and Stemmer algorithm is used to find Root words.

Key terms: Recommendation, Rating, Web pages, Web images, Frequent usage.

1. Introduction

Generally in a web page recommendation system, queries are submitted through search engines to retrieve the information needed by the users. However, sometimes queries may not exactly retrieve users' specified information. Different users may want to get information on different aspects when they submit the same query. In the existing system, user may get an unwanted Advertisements during their search. For example, when the query "the sun" is submitted to a search engine, some users want to locate the homepage of a United Kingdom newspaper, while some others want to learn the natural knowledge of the sun. Therefore, it is necessary to capture different user search goals in information retrieval system. Resultant information from the webpages would satisfy his/her need. User search goals can be considered as the clusters of information needs for a query. The inference and analysis of user search goals can have a lot of advantages in improving search engine relevance and user experience. Some advantages are summarized as follows. First, we can restructure web search results according to user search goals by grouping the search results with the same search goal; thus, users with different search goals can easily find what they want. Second, user search goals represented by some keywords can be utilized in query recommendation thus, the suggested queries can help users to form their queries more precisely and it provides the Semantic Information. Third, the distributions of user search goals can also be useful in applications such as re-ranking web search results that contain different user search goals. i.e The search results which is needed by the users will be reordered based on the users' rating which is already provided by the other users.

In this Recommendation system, many works of user's search goal analysis have been investigated. They can be summarized into two classes: Query clustering and Query classification.

In Query Clustering, when the user gives the query in the search engine it will collect all the informations about the searched queries. The related informations also displayed in the web page. Using Stemming algorithm, the root words are fetched from the search queries. In Query classification, Queries are checked whether it is 'Product based' or 'Non product based'. If the Query is Product based means it will be consider the 'Ratings' which is already provided by the users for the Web pages. For Non product based, the web pages are displayed based on the Frequent use of the previous users.

In Image pre-processing, when the user gives the query in the search engine, first it checks the root word and it provide

the Images. Here the Resulted images are displayed based on the frequent use of the previous users. And in this system images are in the structured manner. And there is some advantages that each images having the Entropy, Skew, Mean values and Distance. (i.e) time taken to display each images.

2. Literature survey

Bringing Order to the Web: Automatically Categorizing Search Results

Hao Chen

School of Information Management & Systems University of California

This model was then used to classify new web pages returned from search engines on-the-fly. This approach has the advantage of leveraging known and consistent category information to assist the user in quickly focusing in on task-relevant information. The interface allows users to browse and manipulate categories, and to view documents in the context of the category structure.

Automatic Identification of User Goals in Web Search

Uichin Lee University of California

In this paper we study whether and how we can automate this goal-identification process. We first present our results from a human subject study that strongly indicates the feasibility of automatic query-goal identification.

Query Recommendation using Query Logs in Search Engines

Ricardo Baeza-Yates1, Carlos Hurtado1

In this paper we propose a method that, given a query submitted to a search engine, suggests a list of related queries. The related queries are based in previously issued queries, and can be issued by the user to the search engine to tune or redirect the search process.

Varying Approaches to Topical Web Query Classification

Steven M. Beitzel

Telcordia Technologies, Inc. One Telcordia Drive

We have evaluated three differing approaches to topical web query classification. We find that training explicitly from classified queries outperforms bridging a document taxonomy for training by as much as 48% in F1.

Context-Aware Query Suggestion by Mining Click-Through and Session Data¤

Huanhuan Cao1 Daxin Jiang2

In this paper, we propose a novel context-aware query suggestion approach which is in two steps. In the *o²ine model-learning step*, to address data sparseness, queries are summarized into concepts by clustering a click-through bipartite.

1) Picture Collage

[28]In this paper, we address a novel problem of automatically creating a picture collage from a group of images. Picture collage is a kind of visual image summary — to arrange all input images on a given canvas, allowing overlay,to maximize visible visual information. We formulate the picture collage creation problem in a Bayesian framework.

2) Inferring Users' Image-Search Goals with Pseudo-images

[13]In this paper, we propose a novel approach to capture user search goals in image search by exploring pseudo-images which are extracted by mining single sessions in user click-through logs to reflect user information needs. Moreover, we also propose a novel evaluation criterion to determine the number of user search goals for a query. Experimental results demonstrate the effectiveness of the proposed method .

3) Personalized Image Search through Tag-based User Profile on Social Websites

[22]In this paper we exploit the Social annotations and Novel Framework for considering the user query relevance and user specific-topic to learn personalized image search. The proposed framework contains two techniques:1) Utility and Prediction model for social annotations. We introduce a Hit Matrix technique for user query relevance and preference into the specific topic space. Performance evaluation shows that our proposed method outperforms the existing method and also shows that the developed model demonstrates the effectives of the Personalized Image Search

4) Towards a comprehensive survey of the semantic gap In visual image retrieval

[15]This paper adopts the premise that the 'semantic gap' is an incompletely surveyed feature in the landscape of visual image retrieval, and proposes a framework within which this deficiency might be made good. Simple classifications of types of image and of types of user are proposed. Consideration is then given in outline to how semantic content is realised by each class of user within each class of image. The argument is advanced that this realisation finds expression in perceptual, generic interpretive and specific interpretive content.

5) The effect of specialized multimedia collections on WEB SEARCHING

[2]In this paper, we report the results of a research study evaluating the effect of separate multimedia Web collections on individual searching behavior. The AltaVista search engine has an extensive multimedia collection and uses tabs to search specific collections. The motivating questions for this research are: (1) What are

the characteristics of multimedia searching on AltaVista? and (2) What are the effects on Web searching of separate multimedia collections? The results of our research show that multimedia searching is complex relative to general Web searching.

6) Searching for multimedia: analysis of audio, video and image Web queries

In this case, the problem seems to be in representing audio and image information needs with textual queries, or with representing retrieved multimedia documents as short textual abstracts. In order to express a non-textual information need in only textual terms, the user takes on an additional cognitive load. In order to make relevance judgments, the user must visually inspect the full record in order to know if the retrieved document contains the requested multimedia information.

7) Hierarchical Clustering of WWW Image Search Results Using Visual, Textual and Link Information

In this paper, we described a method to organize WWW image search results. Based on the web context, we proposed three representations for web image, i.e. representation based on visual feature, representation based on textual feature and representation induced from image link graph. Spectral techniques were applied to cluster the search results into different semantic categories. For each category, several images were selected as representative images according to their Image Ranks, which enables the user to quick understanding the main topics of the search results.

8) Learn from Web Search Logs to Organize Search Results

In this paper, we propose to address these two de_-ciencies by (1) learning \interesting aspects" of a topic from Web search logs and organizing search results accordingly; and (2) generating more meaningful cluster labels using past query words entered by users. We evaluate our proposed method on a commercial search engine log data.
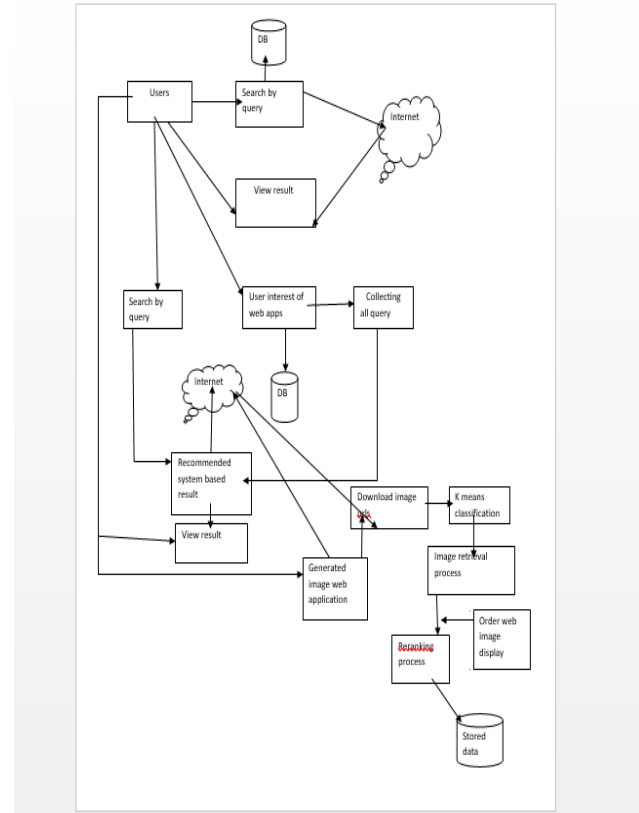
3.SYSTEM ARCHITECTURE

Fig1: Proposed System Architecture

In our System, first we have Login page which is used for an authentication purpose. For product based companies, these type of authentication will be used for security purpose. For successful login, the user have to Register themselves with the system. If any of the user try to login themselves without registration means those people are advised to make a previous registration.

When the user clicks the Register they are redirected to the Registration page. New user will give their 'Username', 'Password' and 'Email ID'. These details are stored in the SQL table which is created already. These Registration details are used to help any Organizations to maintain their Employees informations.

Search Engine for Text Preprocessing:

Once the user make register themselves, they are intended to the page of Search Engine. There are two types of Search queries. One is product based and another one is Non product based. When the user search their queries, the unwanted word like 'is, was, the and so on' are removed by the use of

'Stemming Algorithm'. A simple stemmer looks up the inflected form in a lookup table. The advantages of this approach are that it is simple, fast, and easily handles exceptions. The disadvantages are that all inflected forms must be explicitly listed in the table: new or unfamiliar words are not handled, even if they are perfectly regular (e.g. iPads ~ iPad), and the table may be large. For languages with simple morphology, like English, table sizes are modest, but highly inflected languages like Turkish may have hundreds of potential inflected forms for each root.

Queries:

For Product based queries, Web pages are displayed based on the 'Ratings' which are provided already by the users. And for the Non product based queries, the Frequent usage of the previous users are considered.

Unwanted advertisements are removed in our proposed system. So that, users can be avoided to access the unnecessary Web pages.

Click through:

The Resultant Queries are stored in the SQL table using the Click through button. Using this method Employees or users are noticed by their company to maintain the search history.

Search Engine for Image Preprocessing:

For Image Preprocessing, Images are displayed based on the Frequent use of the Previous users. And Images are displayed in Structured manner. **SVM** algorithm is used to find the Frequent usage of the previous users. The Skew, Mean, Entropy and Distance values are displayed for each of the Images. To find these values we can use 'Affine Transformations'. Search Images are stored in the Datasets.

Affine Transformation:

Affine transformation is a linear mapping method that preserves points, straight lines, and planes. Sets of parallel lines remain parallel after an affine transformation. The affine transformation technique is typically used to correct for geometric distortions or deformations that occur with non-ideal camera angles.

4. Experimental Outputs:

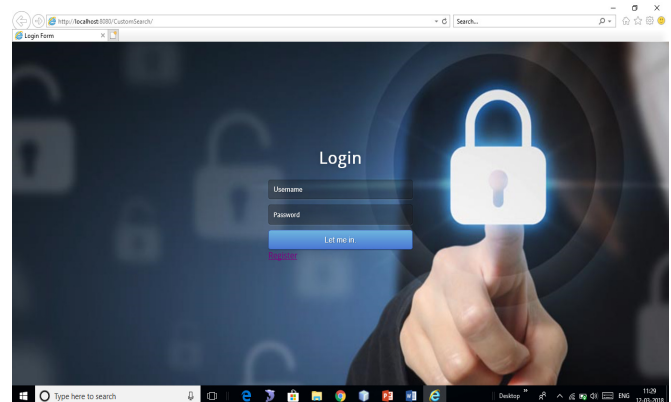The outputs taken in various stages during the implementation are presented below.

Fig 2. Login Page of the Proposed System
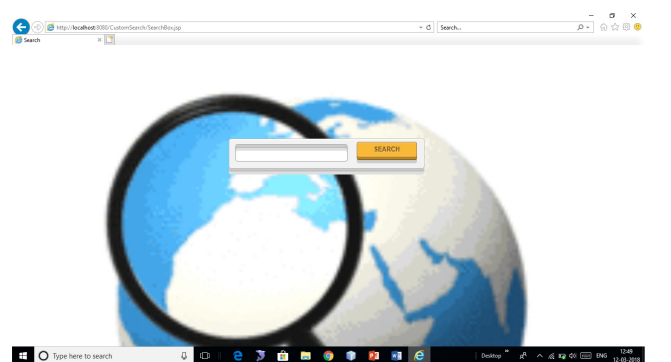

Fig 3. New user Registration page
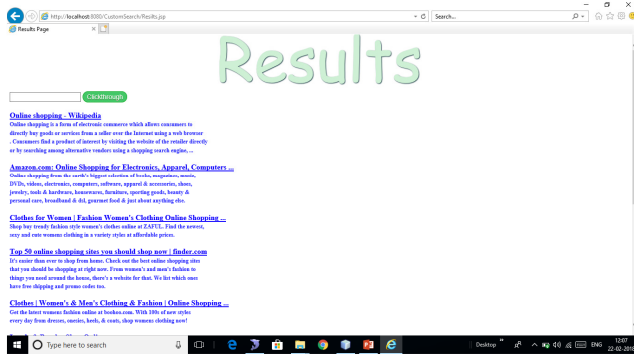

Fig 4. Search Engine for Text in proposed system
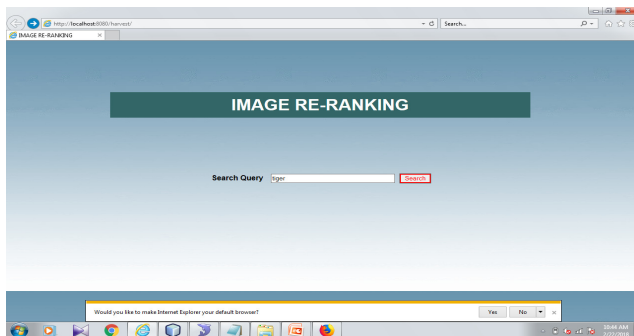
Fig 5. Result page for Text

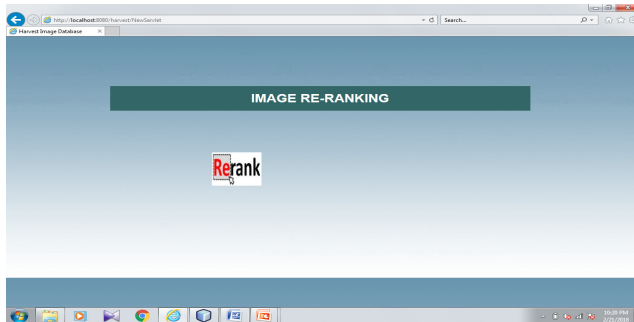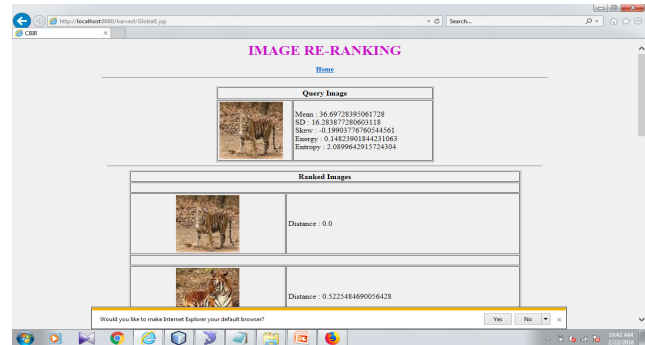Fig 6. Search Engine for Image in Proposed System

Fig 7. Re-ranking of Images

Fig 8. Result page for Image

5. Conclusion

The very big issue that everyone of us are misleading by the Existing system. Rating and Frequent use of the previous user for Web pages and Images are considered in Proposed System. This system is mainly developed to provide accurate and efficient Web pages and Images according to the user queries.

6. References

[1]B. Liu, B. Mobasher, and O. Nasraoui, "Web Usage Mining," in *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*, B. Liu, Ed.: Springer-Verlag Berlin Heidelberg, 2011, pp.527-603.

[2] B. Mobasher, "Data Mining for Web Personalization," in *The Adaptive Web*. vol. 4321, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds.: Springer-Verlag Berlin, Heidelberg, 2007, pp. 90-135.

[3] G. Stumme, A. Hotho, and B. Berendt, "Usage Mining for and on the Semantic Web," AAAI/MIT Press, 2004, pp. 461-480.

[4] H. Dai and B. Mobasher, "Integrating Semantic Knowledge with Web Usage Mining for Personalization," in *Web Mining:Applications and Techniques*, A. Scime, Ed. Hershey, PA, USA:IGI Global, 2005, pp. 276 - 306.

[5] S. A. Rios and J. D. Velasquez, "Semantic Web Usage Mining by a Concept-Based Approach for Off-line Web Site Enhancements," in *Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT '08. IEEE/WIC/ACM International Conference on*, 2008, pp. 234- 241.

[6] S. Salin and P. Senkul, "Using Semantic Information for Web Usage Mining based Recommendation," in *24th*

International Symposium on Computer and Information Sciences, 2009., 2009, pp. 236-241.

[7] A. Bose, K. Beemanapalli, J. Srivastava, and S. Sahar, "Incorporating Concept Hierarchies into Usage Mining Based Recommendations," in *Proceedings of the 8th Knowledge discovery on the web international conference on Advances in web mining and web usage analysis* Philadelphia, PA, USA: Springer-Verlag, 2007, pp. 110-126.

[8] N. R. Mabroukeh and C. I. Ezeife, "Semantic-Rich Markov Models for Web Prefetching," in *2009 IEEE International Conference on Data Mining Workshops* Miami, Florida, USA, 2009, pp. 465-470.

[9] M. O'Mahony, N. Hurley, N. Kushmerick, and G. Silvestre,"Collaborative recommendation: A robustness analysis," *ACM Transactions on Internet Technology,* vol. 4, pp. 344-377, 2004.

[10] G. Stumme, A. Hotho, and B. Berendt, "Semantic Web Mining: State of the art and future directions," *Web Semantics: Science, Services and Agents on the World Wide Web,* vol. 4, pp. 124-143, 2006.

[11] L. A. Alvares Aldaco, "Comportamiento de la deserción y reprobación en el colegio de bachilleres del estado de baja california: Caso plantel ensenada," in *Proc. 10th Congr. Nat. Invest. Educ.*, 2009, pp. 1–12.

[12] F. Araque, C. Roldán, and A. Salguero, "Factors influencing university drop out rates," *Comput.Educ.*, vol. 53, no. 3, pp. 563–574, 2009.

[13] M. N. Quadril and N. V. Kalyankar, "Drop out feature of student data for academic performance using decision tree techniques," *Global J.Comput. Sci. Technol.*, vol. 10, pp. 2–5, Feb. 2010.

[14] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," *Expert Syst. Appl.*, vol. 33, no. 1, pp. 135–146, 2007.

[15] C. Romero and S. Ventura, "Educational data mining: A review of the state of the art," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 40, no. 6, pp. 601–618, Nov. 2010.