_____

Bespoke Image Search Engine Based On User Sensitivity

Dr. A.Suresh

Professor & Head,
Department of Computer Science and Engineering,
Asan Memorial College of Engineering and Technology.
Cghengalpet, Chennai, Tamil Nadu, India.
prisu6esh@yahoo.com

Abstract - Every user has a specific goal when searching for information. Based on user interest the goal has to be achieved. A system for re-ordering the search and to provide the users results of more relevance we go for Personalization. Goal of Personalized Web Search is to search a particular user based interests which allows them for more efficient and relevant information access. In order to make the information more reliable we go for Personalized Image Search Engine based on user feedback. Based on user feedback and clickthrough, our Search Engine categories the information using images and gives the desired results. This makes our search with a specific limit. Hence my proposed approach considers user feedback to provide accurate search results and to increase the performance of the search engine. This helps user to identify topics from different senses of the term, and then personalize results of image search by displaying to the user only those images.

Keywords: Uses Search Goal, Feedback Session, Personalization, and Clickthrough.
_____***** _____

I. INTRODUCTION

As the amount of information on the Web increases rapidly, it creates many new challenges for Web search. When the same query is submitted by different users, a typical search engine returns the same result, regardless of who submitted the query. This may not be suitable for users with different information needs. For example, for the query "apple", some users may be interested in documents dealing with "apple" as "fruit", while some other users may want documents related to Apple computers. One way to disambiguate the words in a query is to associate a small set of categories with the query. For example, if the category "cooking" or the category "fruit" is associated with the query "apple", then the user's intention becomes clear. Current search engines such as Google or Yahoo! have hierarchies of categories to help users to specify their intentions. The use of hierarchical categories such as the Library of Congress Classification is also common among librarians. A user may associate one or more categories to his/her query manually. For example, a user may first browse a hierarchy of categories and select one or more categories in the hierarchy before submitting his/her query. By utilizing the selected categories as a context for the query, a search engine is likely to return documents that are more suitable to the user. Unfortunately, a category hierarchy shown to a user is usually very large, and as a result, an ordinary user may have difficulty in finding the proper paths leading to the suitable categories. Furthermore, users are often too impatient to identify the proper categories before submitting his/her queries. An alternative to browsing is to obtain a set of categories for a user query directly by a search engine. However, categories returned from a typical search engine are still independent of a

particular user and many of the returned categories do not reflect the intention of the searcher.

II. BACKGROUND

The main data source in the web usage mining and personalization process is the information residing on the web site's logs. Web logs record every visit to a page of the web server hosting it. The entries of a web log consist of several fields which represents the date and time of the request, the IP number of the visitor's computer (client), the URI requested the HTTP status code returned to the client, and so on. The web logs' file format is based on the so called extended log format. Prior to processing the usage data using web mining or personalization algorithms, the information residing in the web logs should be pre-processed. . The web log data pre-processing is an essential phase in the web usage mining and personalization process. An extensive description of this process can be found. In the sequel, we provide a brief overview of the most important pre-processing techniques, providing in parallel the related terminology. The first issue in the *pre-processing* phase is *data preparation*. Depending on the application, the web log data may need to be cleaned from entries involving page accesses that returned, for example, an error or graphics file accesses. Furthermore, crawler activity usually should be filtered out, because such entries do not provide useful information about the site's usability.

A. *Web usage mining and personalization*

Web usage mining is the process of identifying representative trends and browsing patterns describing the

_____

activity in the web site, by analyzing the users' behavior. Web site administrators can then use this information to redesign or customize the web site according to the interests and behavior of its visitors, or improve the performance of their systems. Moreover, the managers of e-commerce sites can acquire valuable business intelligence, creating consumer profiles and achieving market segmentation. There exist various methods for analyzing the web log data. Some research studies use well known data mining techniques such as association rules discovery, sequential pattern analysis, clustering, probabilistic models, or a combination of them. Since web usage mining analysis was initially strongly correlated to data warehousing, there also exist some research studies based on OLAP cube models. Finally some proposed web usage mining approaches that require registered user profiles, or combine the usage data.

B. *Semantic web based search*

Semantic search seeks to improve search accuracy by understanding searcher intent and the contextual meaning of terms as they appear in every search engine in order to generate one or more relevant results. Systems consider various pointing including context of search, location, variation of words, generalized and specialized queries, concept matching and natural language queries to provide relevant search results. Major web search engines like Google and Bing incorporate some elements of semantic search. User provides the search engine with a phrase which is intended to denote an object about which object is trying to gather the information. There is no particular document which the user knows about it. The user trying to locate the number of documents which together will provide the desired information Semantic search lends itself well with this approach that is closely related with exploratory search.

III. PRELIMINARIES

Generally, a session for web search is a series of successive queries to satisfy a single information need and some clicked search results. It mainly focuses on inferring user search goals for a particular query. Therefore, the single session containing only one query is introduced, which distinguishes from the conventional session. Meanwhile, the feedback session in this paper is based on a single session, although it can be extended to the whole session. The proposed feedback session consists of both clicked and unclicked URLs and ends with the last URL that was clicked in a single session. It is motivated that before the last click, all the URLs have been scanned and evaluated by users. Therefore, besides the clicked URLs, the unclicked ones before the last click should be a part of the user feedbacks. Each feedback session can tell what a user requires and what he/she does not care about. Moreover, there are plenty of diverse feedback sessions in user click-through logs.

Therefore, for inferring user search goals, it is more efficient to analyze the feedback sessions than to analyze the search results or clicked URL's directly. Due to usefulness of user search goals, it is categorized into three classes: query classification, search result reorganization, and session boundary detection

A. *Latent Semantic Indexing*

It is an indexing and retrieval method that uses a mathematical technique called singular value decomposition (SVD) to identify patterns in the relationships between the terms and concepts contained in an unstructured collection of text. LSI is based on the principle that words that are used in the same contexts tend to have similar meanings. A key feature of LSI is its ability to extract the conceptual content of a body of text by establishing associations between those terms that occur in similar contexts. It uncovers the underlying latent semantic structure in the usage of words in a body of text and how it can be used to extract the meaning of the text in response to user queries, commonly referred to as concept searches. Queries, or concept searches, against a set of documents that have undergone LSI will return results that are conceptually similar in meaning to the search criteria even if the results don't share a specific word or words with the search criteria.

IV. PROPOSED SYSTEM

The proposed feedback session consists of both clicked and unclicked URLs and ends with the last URL that was clicked in a single session. It is motivated that before the last click, all the URLs have been scanned and evaluated by users. Therefore, beside the clicked URLs, the unclicked ones before the last click should be a part of the user feedbacks. It discovers user search goals for some popular queries offline at first. Then, when users submit one of the queries, the search engine can return the results that are categorized into different groups according to user search goals online. Thus, users can find what they want conveniently. A novel way to map feedback sessions to pseudo-documents, includes two steps. They are the following,

- Representing the URLs in the feedback session.
- Forming pseudo document based on URL representation.

A. *Representing URL in feedback session*

A session for a web search is a series of successive queries to satisfy a single information need and some clicked search results. The single session containing only one query is introduced which differs from conventional session. The Feedback session is based on single session, although it can be

extended to the whole session. The feedback session is defined as the series of both clicked and unclicked URL's and ends with the last URL that was clicked in a session from user click through logs. Then map the feedback session to pseudo documents which can efficiently reflect user information needs

B. Forming pseudo document based on URL representation

The feedback session vary lot for different click through and queries, it is unsuitable to use feedback session directly for inferring user search goals. Different feedback sessions have different numbers of URLs, they may have different dimensions. For a certain query, users will usually have same keyword representing their interests in their mind. The keyword is to determine whether a document can satisfy their needs. Some representation method is needed to describe feedback sessions in more efficient and coherent way.
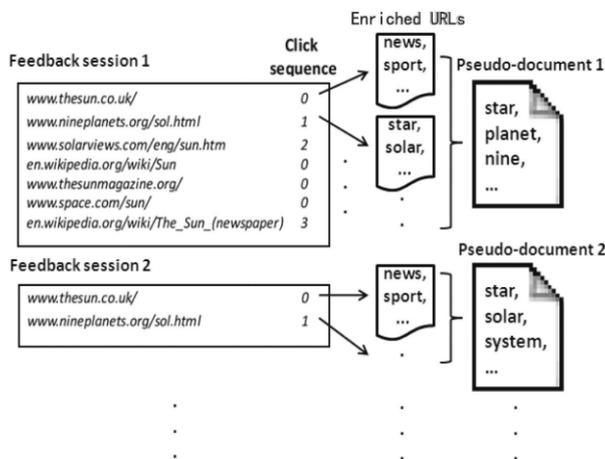

Fig. 1 forming pseudo document

Feedback sessions can be considered as a process of resampling. If we view the original URLs in the search results as original samples, then feedback sessions can be viewed as the "processed" or "resample" samples which differ from the original samples and reflect user information needs. Feedback session is also a meaningful combination of several URLs. Therefore, it can reflect user required information more precisely and there are plenty of feedback sessions to be analyzed. In our proposed work, the images are submitted to search engines to represent the required information of users where it is the quick way for the users to visualize rather than text based search.

V.   SYSTEM DESIGN

The system design is the process of defining the architecture for a system to satisfy specified requirements. System design could be seen as the application of systems theory to product development.
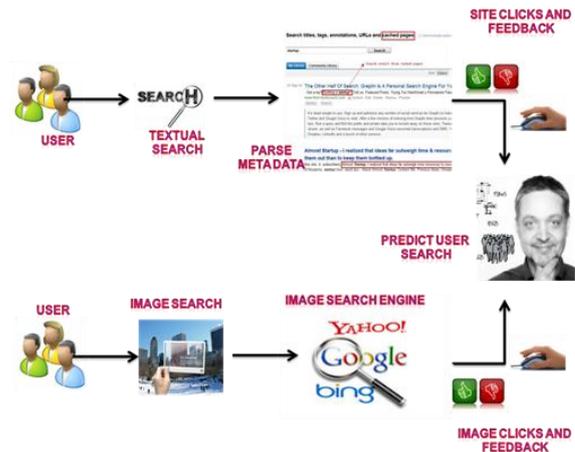

Fig. 2 System Design

VI.   CONCLUSION

The data that are in the form of tags are created mainly to help users organize and manage their own content. This type data can also be used to target relevant content to the user through personalization. The Personalized Image Search Engine is used to retrieve the information using images on the search engine rather than text search. Depending upon number of clicks on images, the feedback is generated. The feedback sessions is analyzed to infer user both the clicked URLs and the unclicked ones before the last click are considered as user implicit feedbacks and taken into account to construct feedback sessions. Therefore, feedback sessions can reflect user information needs more efficiently. After generating the feedback it creates a pseudo document which gives the searching information in the form of images.

REFERENCES

[1] Chaitali Jambhulkar, Ankita Borana, Chaitali Deshmukh, Priyanka Nanaware &Chandrama G. Thorat , "Personalized Image Search using Hybrid Re-ranking Based on Ontology", "Department of Computer, Rajarshi Shahu College of Engineering, India", 2012

[2] Dongyuan Lu, Qiudan Li, "Personalized Search on Flickr based on Searcher's Preference Prediction", "Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China", 2011.

[3] Hao Chen, Susan Dumais, "Bringing Order to the Web: Automatically Categorizing Search Results","School of Information Management & Systems University of California Berkeley, CA 94720 USA,"2010.

[4] Huanhuan Cao, Daxin Jiang, Jian Pei, "Context-Aware Query Suggestion by Mining Click-Through and Session Data", "University of Science and Technology of China, Microsoft Research Asia, Simon Fraser University",2008

[5] Xiaohui Tao, Yuefeng Li and Ning Zhong, I"A Personalized Ontology Model for Web Information" IEEE Transaction on Knowledge and Data Engineering, Vol 23, No.4, April 2011.

[6] Bhaganagare Ravishankar C, Dharmadhikari Dipa D Web Personalization Using Ontology: A Survey ISSN: 2278-0661 Volume 1, Issue 3(May-June 2012), PP 37-45.

[7] P. Devisree and P. Revathi," Web Search Personalization with Ontological User Profiles and World Knowledge Base" International Conference on Computing and Control Engineering (ICCCE 2012), ISBN 978-1-4675-2248-9.

___

[8] Chawla, I "An overview of personalization in web search. Electronics Computer Technology (ICECT), 2011 3rd International Conference Volume: 3 published on 2011.

[9] "Jayanthi, J.; Dept. Comput. Sci., Sona Coll. of Technol., Salem, India; Jayakumar, K.S.; Surendran, S."Published Electronics Computer Technology (ICECT), 2011 3rd International Conference on (Volume:6) April 2011.

BIBLIOGRAPHY

Dr. A. Suresh., B.E., M.Tech., Ph.D works as a Professor & Head of the Computer Science and Engineering Department in ASAN Memorial College of Engineering & Technology, Chengalpet, Chennai, Tamilnadu, India. He has more than 17 years of experience in teaching and his areas of specializations are Data Mining, Artificial Intelligence, Image Processing, Neural Networks and System Software. He has published many of his research work in national and international journals & conferences and he has published one book in the name of Data structures & Algorithms in DD Publications.

___