International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Content and Usage Based Ranking for Enhancing Search Result Delivery

Shital C. Patil[1], R. R. Keole[2]

[1]H.V.P.M's College of Engineering & Technology, Amravati University, India

[2]Department of Computer Science and Engineering, H.V.P.M's college of Engineering & Technology, Amravati University, India

Abstract: ***In today's e-world search engines play a vital role in retrieving and organizing relevant data for various purposes. However, in the real ground relevance of results produced by search engines are still debatable because it returns enormous amount of irrelevant and redundant results. Providing relevant information to user is the primary goal of the website owner. Web mining is ample and powerful research area in which retrieval of relevant information from the web resources in a faster and better manner. Web Mining consists of three different categories, namely Web Content Mining , Web Structure Mining and Web Usage Mining . Web content mining improves the searching process and provides relevant information by eliminating the redundant and irrelevant contents. However for a broad-topic and ambiguous query, different users may have different search goals when they submit it to a search engine. Web usage mining plays an important role in inferring user search goals as they can be very useful in improving search engine relevance and user experience. Thus the project focuses on combine approach of web usage mining and web content mining to improve the search engine results by inferring user search goals. In the proposed work a new approach is introduced to re-order the search results based on the contents and user interest rather than keyword and page ranking provided by search engines. Based on the user query, search engine results are retrieved. Every result is individually analyzed based on the user query and page contents and particular score is awarded to each result. Finally, the relevancy of the particular link against user request is computed by summarizing all the scores and the reordered list is displayed to the user. When the user visits the web page out of this reordered list, the query, url and the contents extracted from the web page are stored in the server log. When next time user enters a query the scores are awarded to each result link based on the data in the server log which indirectly incurs the user interest.***

Keywords: Search Engine Result, Information Retrieval, Web Usage Mining, Web Content Mining, Re-ranking.

1. Introduction

It is not exaggerated to say the Web World Web is the most excited impacts to the human society in the last 10 years. It changes the ways of doing business, providing and receiving education, managing the organization etc. The most direct effect is the completed change of information collection, conveying, and exchange. Today, Web has turned to be the largest information source available in this planet. The Web is a huge, explosive, diverse, dynamic and mostly unstructured data repository, which supplies incredible amount of information, and also raises the complexity of how to deal with the information from the different perspectives of view – users, Web service providers, business analysts. With the exponential growth of WWW, it has become difficult to access desired information that matches with user needs and interest [1]. The users want to have the effective search tools to find relevant information easily and precisely. Therefore, Web mining becomes an active and popular research field.

Web mining is the process of discovering knowledge, such as patterns and relations, from Web data. Web mining generally has been divided into three main areas [1] [2]: content mining, structure mining and usage mining. Each one of these areas are associated mostly, but not exclusively, to these three predominant types of data found in the Web:

- Content: The real data that the document was designed to give to its users. In general this data consists mainly of text and multimedia.
- Structure: This data describes the organization of the content within the Web. This includes the organization inside a Web page, internal and external links and the website hierarchy.
- Usage: This data describes the use of a website or search engine, reflected in the Web server's access logs, as well as in logs for specific applications.

There is not a clear-cut distinction among these categories, and all three mining tasks can be combined [3].

1.1 Overview

In order to retrieve user requested information, search engine plays a major role for crawling web content on different node and organizing them into result pages so that user can easily select the required information by navigating through the result pages link. This strategy worked well in earlier because, number of resources available for user request is limited. Also, it is feasible to identify the relevant information directly by the user from the search engine results. When the Internet era increases, sharing of resource also increases and this leads to develop an automated technique to rank each web content resource. Different search engine uses different techniques to rank search results for the user query.

Web content mining improves the searching process and provides relevant information by eliminating the redundant and irrelevant contents according to user queries [4]**.** However, sometimes queries may not exactly represent users' specific information needs since many ambiguous queries may cover a broad topic and different users may want to get information on different aspects when they

Licensed Under Creative Commons Attribution CC BY

submit the same query. 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 and potential to capture different user search goals in information retrieval. The inference and analysis of user search goals can have a lot of advantages in improving search engine relevance and user experience:

- 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.
- Restructure web search results [5] according to user search goals by grouping the search results with the same search goal.

1.2 Web Mining Taxonomy

Web Mining can be broadly divided into three distinct categories, according to the kinds of data to be mined. A figure [6] depicting the taxonomy is shown in Figure 1.

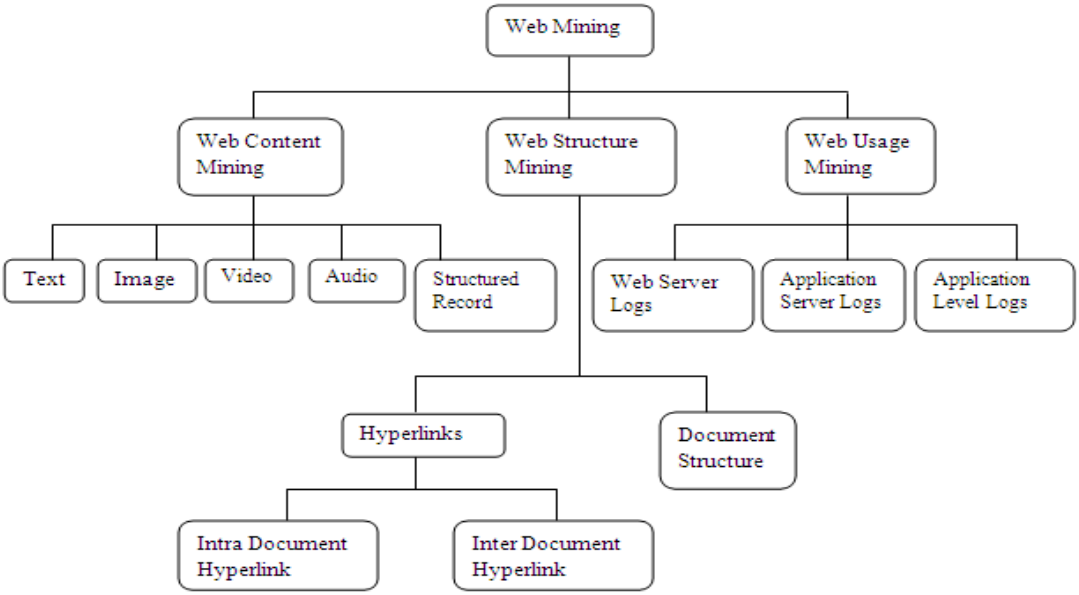

Figure1: Web mining Taxonomy

1.2.1 Web Content Mining

Web content mining is used to examine the content of Web pages as well as results of Web searching. The content may include text as well as graphics data. Web content mining is further divided into Web page content mining and search results mining. Web page content mining is traditional searching of Web pages with the help of content while search results mining is a further search of pages found from a previous search.

1.2.2 Web Structure Mining

Web structure mining is done at the hyper link level. This kind of mining tries to discover the model underlying the link structure of the web.

1.2.3 Web Usage Mining

Web usage mining is the process of extracting useful information from server logs e.g. use Web usage mining is the process of finding out what users are looking for on the internet. Some users might be looking at only textual data, whereas some others might be interested in multimedia data. Web Usage Mining is the application of data mining techniques to discover interesting usage patterns from Web data in order to understand and better serve the needs of Web-based applications. Usage data captures the identity or origin of Web users along with their browsing behavior at a Web site.

1.3 Motivation

Currently, a major challenge is to build communication between search engines and web users. However most search engines can only use queries rather than Web user profiles due to the difficulty of automatically acquiring web user profiles. The first reason for this is that web user may not know how to represent their topic of interest. The second reason is that web users may not wish to invest a great deal of effort to dig out few relevant pages from hundreds of thousands of candidates provided by search engines. The real motivation behind this work in this dissertation is to help in the resolution of this problem by taking one step further toward a satisfactory solution. The intention is to create a system that is able to recognize the user interest by observing the user's search history and gives relevant results to the user. So the motivation behind this work is as follows,

- The major source of information is from Web.
- Traditional information search approaches are hardly appropriate due to enormous size.
- Typical queries retrieve hundreds of documents; most of them have no relation with what the user is looking for.
- So the main problem is regarding the interest of the user and to display required documents from the Web at the top position.

Licensed Under Creative Commons Attribution CC BY

- One of the solution to this is combine approach of the web usage mining and web content mining. Web content mining removes the persistent problem while Web usage mining plays an important role in inferring user search goals.

1.4 Objectives

The dissertation work will focus on the problem of displaying search results from the retrieving Web documents using combination of content mining and usage mining. And the aim of the dissertation work is to enhance the search results delivery by inferring the user search interest. This dissertation work will try to achieve some or all of the following objectives,

- Retrieve top n search results.
- Re-order the results according to user query.
- Extract the contents from user click through data and maintain the log at server containing query, url and extracted contents.
- Analyze the user search interest from the server log.
- Provide the results to the user according to usage made by user.

1.5 Brief description of the Dissertation work

The proposed system is based on combine approach of the content mining and usage mining. In this dissertation work focus is on the problem of displaying the useful search results from the web on the top position according to the query entered by the user and the user interest, from the stored log information. Following are the steps performed for the system

Step 1: Processing user request to obtain the results from search engine.
Step 2: Extraction of top n results from search engine.
Step 3: Analysis of results for giving score to each link based on the user query.
Step 4: Re-ordering the results according to the scores.
Step 5: Providing re-ordered list of results to the user.
Step 6: Contents extraction from the page clicked by the user.
Step 7: Maintaining the server log containing url, query, number of clicks and the extracted contents.

The rest of the paper is organized as follows. The Literature Review is covered in Section 2. Section 3 covers the analysis of problems with the existing search engines. Section 4 discusses the description about the proposed concept with architecture and components. In chapter 5 the proposed work is discussed with specification and techniques used. In chapter 6 experimental results are discussed with comparison with the existing system. Finally chapter 7 concludes by summarizing the various systems proposed for mining the web information, as described in the literature review and the selection of the combine approach of content mining and usage mining.

2. Literature Review

Due to the heterogeneity of network resources and the lack of structure of web data, automated discovery of targeted knowledge retrieval mechanism is still facing many research challenges. Moreover, the semi structured and unstructured nature of web data creates the need for web content mining. In paper [7] the author differentiates web content mining from two different points of view. Information retrieval view and database view. In paper [8] research area of web mining and different categories of web mining are discussed briefly. They also summarized the research works done for unstructured data and semi structured data from information retrieval view.

Effective organization of search results is critical for improving the utility of any search engine. The utility of a search engine is affected by multiple factors. While the primary factor is the soundness of the underlying retrieval model and ranking function, how to organize and present search results is also a very important factor that can affect the utility of a search engine significantly. Compared with the vast amount of literature on retrieval models, however, there is relatively little research on how to improve the effectiveness of search result organization. The most common strategy of presenting search results is a simple ranked list [4]. Intuitively, such a presentation strategy is reasonable for non-ambiguous, homogeneous search results; in general, it would work well when the search results are good and a user can easily and many relevant documents in the top ranked results. However, when the search results are diverse (e.g., due to ambiguity or multiple aspects of a topic) as is often the case in Web search, the ranked list presentation would not be effective; in such a case, it would be better to group the search results into clusters so that a user can easily navigate into a particular interesting group.

People attempt to infer user goals and intents by predefining some specific classes and performing query classification accordingly. Lee et al. [9] consider user goals as “Navigational” and “Informational” and categorize queries into these two classes. Other works focus on tagging queries with some predefined concepts to improve feature representation of queries. However, since what users care about varies a lot for different queries, finding suitable predefined search goal classes is very difficult and impractical.

Methods of organizing search results based on text categorization are studied in [3]. In this work, a text classifier is trained using a Web directory and search results are then classified into the predefined categories. The authors designed and studied different category interfaces and they found that category interfaces are more effective than list interfaces. However predefined categories are often too general to reflect the finer granularity aspects of a query.

Clustering search results [10] is an effective way to organize search results, which allows a user to navigate into relevant documents quickly. As a primary alternative strategy for presenting search results, clustering search results has been studied relatively extensively. The general idea in virtually

Licensed Under Creative Commons Attribution CC BY

all the existing work is to perform clustering on a set of top-ranked search results to partition the results into natural clusters, which often correspond to different subtopics of the general query topic. A label will be generated to indicate what each cluster is about. A user can then view the labels to decide which cluster to look into. Such a Such a strategy has been shown to be more useful than the simple ranked list presentation in several studies. However, this clustering strategy has two deficiencies which make it not always work well:

1) The clusters discovered in this way do not necessarily correspond to the interesting aspects of a topic from the user's perspective. For example, users are often interested in finding either “phone codes" or “zip codes" when entering the query “area codes." But the clusters discovered by the current methods may partition the results into “local codes" and “international codes." Such clusters would not be very useful for users; even the best cluster would still have a low precision.
2) The cluster labels generated are not informative enough to allow a user to identify the right cluster.
3) Since feedback is not considered, many noisy search results that are not clicked by the users may be analysed as well.
4) Wang and Zhai clustered queries and learned aspects of these similar queries , which solves the problem in part. However, their method does not work if we try to discover user search goals of one single query in the query cluster rather than a cluster of similar queries. For example, in [11], the query “car” is clustered with some other queries, such as “car rental,” “used car,” “car crash,” and “car audio.” Thus, the different aspects of the query “car” are able to be learned through their method. However, the query “used car” in the cluster can also have different aspects, which are difficult to be learned by their method.

Some works take user feedback into account and analyze the different clicked URLs of a query in user click-through logs directly. However the number of different clicked URLs of a query may be not big enough to get ideal results.

Web usage mining aims to capture, model, and analyze the behavioral patterns and profiles of users interacting with the Web. Data stored in usage logs can be used for solving navigational problem [12], improving web search [5], recommending queries [13], suggesting authoritative web sites [14], and enhancing performance of search engines [15]. A good survey of web usage mining can be found in [16].

3. Analysis of Problem

Nowadays, most of the people rely on web search engines to find and retrieve information. When a user uses a search engine such as Yahoo or Google or bing to seek specific information, an enormous quantity of results are returned containing both the relevant document as well as outlier document which is mostly irrelevant to the user. Therefore discovering essential information from the web data sources becomes very important for web mining research community. As the Web’s contents grow, it becomes increasingly difficult to manage and classify its information. The high level of competition in the Web makes it necessary for websites to improve their organization in a way that is both automatic and effective, so users can reach effortlessly what they are looking for. The problems are:

- **Incomplete or Limited Information Problem:** A number of heuristic assumptions are typically made before applying any data mining algorithm; as a result some patterns generated may not be proper or even correct.
- **Incorrect Information problem:** When a web site visitor is lost, the clicks made by this visitor are recorded in the log, and many mislead future recommendations. This becomes more problematic when a website is badly designed and more people end up visiting unsolicited pages, making them seem popular.
- **Persistence Problem:** When a new pages are added to a web site, because they are not visited yet, the recommender system may not recommend them, even though they could be relevant Moreover, the more a page is recommended, the more it may be visited, thus making it look popular and boost its candidacy for future recommendation.
- **Incorrect recommendation**: Since what user cares about varies a lot for different queries, finding suitable predefined search goal classes is very difficult and impractical.

4. Proposed System

The inference and analysis of user search goals can have a lot of advantages in improving search engine relevance and user experience which can be achieved by web usage mining while web content mining removes persistence problem. The proposed system focuses on combine approach of web usage mining and web content mining. It presents weighted technique to mine the web content catering to the user needs. In the proposed work a new approach is introduced to rank the relevant pages based on the content and keywords rather than keyword and page ranking provided by search engines.

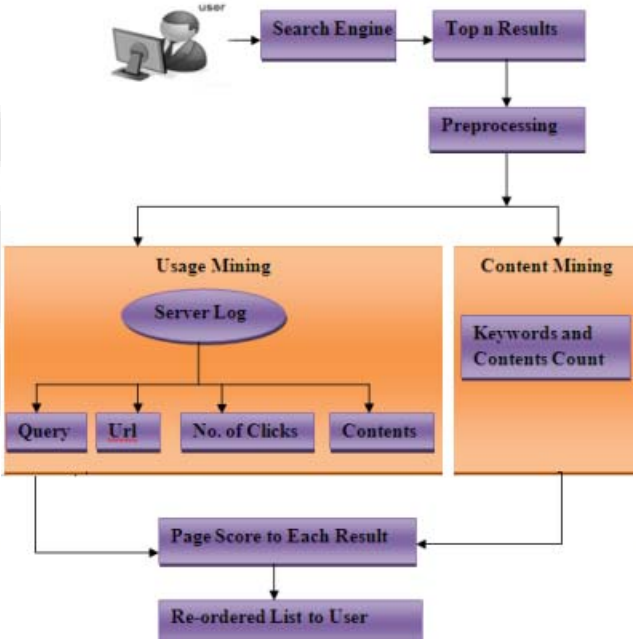

Figure 2: Architecture of Proposed System

4.1 Content Mining

Licensed Under Creative Commons Attribution CC BY

Based on the user query, search engine results are retrieved. Every result is individually analyzed based on keywords and content. The particular score is awarded to every result on the basis of keyword present in the title, count of keyword in web page and total number of words present in the web page. The total relevancy of the particular link against user request is computed by summarizing all the scores of the keyword and content words. Finally, all these results are re-ordered by arranging them with their decreasing order of scores. Thus the result now present on the top is the most relevant according to the user query.

4.2 Usage Mining

User clicks any one of the links presenting to him/her after re-ordering. The query user enters, url, no. of clicks and the contents retrieved from that web page are stored on the server log, which represents user's interest for that particular query. When next time user will enter any other or same query as entered before, the scores will be awarded on the basis of data stored in server log. And again the re-ordered list of results will be presented to the user. Thus the result then present on the top is the most relevant according to user's interest.

5. Specification

Implementation is the stage of project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective. The implementation stage involves careful planning, investigation of the existing system and its constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.
The proposed system is divided into four parts.

5.1. Data Extraction

Data extraction is the first stage in this project. Since we are concerned only about the content on the websites, we use web content mining to extract the data. Web content mining is very similar to text mining and data mining. In web content mining the data is either semi-structured or unstructured whereas in data mining the data is more structured. Due to the exponential growth in web content and its usage, many applications are built to use this data and present the data in a user-friendly manner. There are various problems involved in extracting data from the Internet due to lack of a standard structure used by websites to store the data. When user enters the query in the system API, the query is forwarded to the search engine and the top n results are extracted from the search engine.

5.2. Preprocessing

Pre-Processing is an important step in text based mining. Real world data tend to be dirty, incomplete and inconsistent. Data pre-processing techniques can improve the quality of the data, thereby helping to improve the accuracy and efficiency of the subsequent mining process.

5.3. Page Score

(a) by content mining
For re-ordering the results, every result link is given the specific page score. The page score is depend on the number of times the keyword in the user query appears in the web content and the number of total contents on the web page. Some additional weight is given if the keyword in the user query appears in the title. The value of score of each result is sorted in descending order to get the most relevant content for the user query. Re-ordered results are sent back to the user so that the top most page is more relevant for the user query. When user clicks any of the result, the url, query and contents extracted from that web page are stored in the web page. The contents are extracted by NLP using POS tagger.

- **Natural Language Processing**
  NLP provides means of analyzing the text. The goal of NLP is to make computers analyze and understand the languages that humans use naturally. As the reader has probably already deduced, the complexity associated with natural language is especially key when retrieving textual information to satisfy a user's information needs. This is why in Textual Information Retrieval, NLP techniques are often used both for facilitating descriptions of document content and for presenting the user's query, all with the aim of comparing both descriptions and presenting the user the documents that best satisfy their information needs

- **Part of Speech Tagging**
  When the user clicks the result link he is interested in, the web page gets open in the browser. The contents present on the web page are tokenized and tagged using POS tagging. A Part-Of-Speech Tagger (POS Tagger) is a piece of software that reads text in some language and assigns part of speech to each word (and other token), such as noun, verb, adjective, etc., Following is the list of POS. Out of all these only noun and adjectives are stored into server log.

b) by usage mining

When user enters the query it is first check for the server log entries. If the log contains some entries for previously accessed data then the page score is given on the basis of data present in the server log. Presence of same url and more number of clicks for any url gives additional score to that particular result.

5.4. Server Log Maintenance

For storing the history of user's access behaviour, the server log is maintained. The log contains query entered by user, the URL of the selected result, the contents retrieved from the selected page and number of times he made the click on that particular URL. The log is updated automatically whenever the user selects the web page. The WAMP server has been used to maintain the server log.

Licensed Under Creative Commons Attribution CC BY

6. Experimental Results

Experiment is conducted with a user query "sun". Top 10 results are retrieved from the search engine. Following table represents results retrieved from search engine.

Table 1: Top 10 results retrieved for query "sun

S.No	ID	URL
1	SR1	http://www.thesun.co.uk/
2	SR2	http://en.wikipedia.org/wiki/su
3	SR3	http://www.oracle.com/us/sun
4	SR4	http://www.torontosun.com/
5	SR5	http://nineplanets.org/sol.html
6	SR6	http://www.calgarysun.com/
7	SR7	http://www.vancouvarsun.com/
8	SR8	http://www.baltimoresun.com/
9	SR9	http://edmotionsun.com/
10	SR10	http://www.sun-sentinel.com/

Keyword and content based ranking approach is applied and results are reordered. At the present stage there is no data in the server log.

These re-ordered result list was presented to the user and asked to select any one of the results. The user was totally unaware about the proposed and objectives of the project. User entered his choice. The required data is now stored in the server log. Depend on user's choice the results are now re-ordered by combine approach of usage and content mining. The order of the results provides by search engine and proposed system is compared by using manual ranking from user. Table 2 shows this comparison.

ID	*Search Engine Ranking*	*Content Based Ranking*	*Content and Usage Based Ranking*	*Manual Ranking*
SR1	1	9	6	6
SR2	2	1	2	2
SR3	3	10	10	10
SR4	4	7	7	7
SR5	5	2	3	3
SR6	6	8	9	8
SR7	7	3	4	4
SR8	8	4	1	1
SR9	9	6	8	9
SR10	10	5	5	5

6.1 Performance Evaluation

Performance evaluation of the proposed approach and the search engine is done based on Precision measure.
Precision measure is calculated based on the following formula.

$$\text{Precision} = \frac{tp}{tp + fp}$$

Where,
tp – True Positive (Correct result)
fp – False Positive (Unexpected Result)

Table 2 represents the matching of manual ranking against proposed approach ranking and search engine ranking. Document SR1, 2, 5, 7, 8, 9 represents the mismatching of manual ranking against content based ranking and documents SR6, 9 represents the mismatching of manual ranking against content and usage based ranking.

Table 3: Precision Measure

Different Methods	tp	fp	Precision
Search Engine Ranking	2	8	0.2
Content Based Ranking	4	6	0.4
Content & Usage Based Ranking	8	2	0.8

From the table 3, it is understood that precision of the search-engine is 0.2, precision of content mining is 0.4 and precision of combine approach of content and usage mining is 0.8 out of 1. The results of the performance measure are plotted in Figure 3.

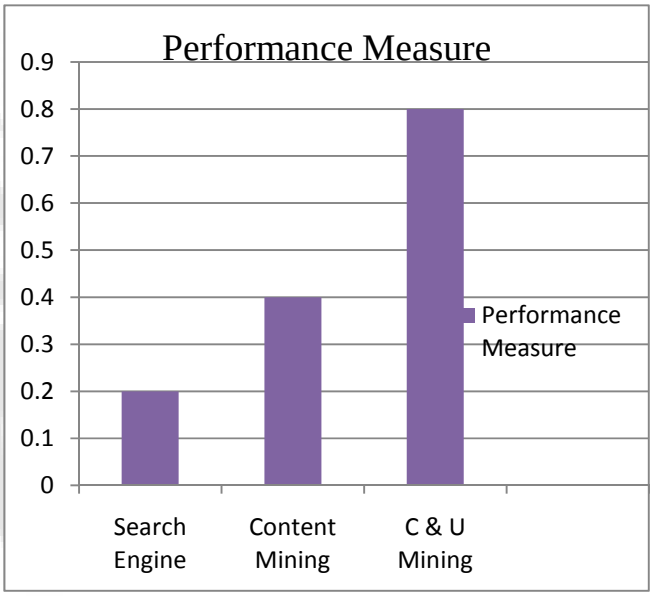

Figure 3: Performance Measure

7. Conclusion

Among the various techniques proposed for organizing and re-ordering of search results from the web, as described in the literature review, combination of content and usage mining to re-ordering the search results was selected for the implementation of proposed work to provide useful information to the user.

To gather the useful information from the web page, the user interested in, the log is maintain on the database of WAMP server. From this server log, interest of particular user can be found easily.The inference and analysis of user search goals can have a lot of advantages in improving search engine relevance and user experience which can be achieved by web usage mining while web content mining removes persistence problem. This work is the combination of content and usage mining which works better than using any one of them. The experimental results demonstrate that proposed model is promising. The proposed approach provides a solution to minimize the persistence and incorrect information problem in a single model. The findings can be applied to the design of Crawling Systems and for Discrimination Prevention. The

Licensed Under Creative Commons Attribution CC BY

system also has contribution to the field of Information Retrieval.

7.1 Future Work

The proposed work is focused only on re-ordering the search results, retrieved from search engine, according to the user's interest and can be further extended to work for searching the results according to the user interest. Proposed work is based on text based mining. As the information on web is present in the images, audio and video formats also, the future work can be focused on all these different formats.

References

[1] R. Cooley, B. Mobasher and J. Shrivastava, "Web Mining: Information and Pattern Discovery on the World Wide Web", Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence, pp. (ICTAI), 1997.
[2] R. Kosala, H. Blockeel, "Web Mining Research: A Survey", SIGKDD Explorations, Newsletter of the ACM Special Interest Group on Knowledge Discovery and Data Mining Vol. 2, No. 1 pp 1-15, 2000.
[3] I. Mele, " Web Usage Mining for Enhancing Search – Result Delivery and Helping Users to Find Interesting Web Content," ACM SIGIR Conf. Research and Development in Information Retrival (SIGIR '13), pp. 765-769, 2013.
[4] P. Sudhakar, G. Poonkuzhali, R. Kishor Kumar, "Content Based Ranking for Search Engines", Proc. International Multi Conference of Engineers and Computer Scientists (IMECS 12), 2012.
[5] Z. Lu, H. Zha, X. Yang, W. Lin, Z. Zheng, "A New Algorithm for Inferring User Search Goals with Feedback Sessions," Proc. IEEE Transactions on Knowledge and Data Engineering, pp. 502-513, 2013
[6] Jaideep Srivastava, Prasanna Desikan, Vipin Kumar, "Web Mining -Concepts, Applications & Research Directions", University of Minnesota, USA, Chapter 3. Pg 52-71.
[7] Gibson, J., Wellner, B., Lubar, S, "Adaptive web-page content identification", In WIDM '07: Proceedings of the 9th annual ACM international workshop on Web information and data management. New York, USA,2007.
[8] Georgies Lappas, An overview of web mining in societal benefit areas, The 9th IEEE International Conference on E-Commerce Technology, IEEE 2007.
[9] U. Lee, Z. Liu, and J. Cho, "Automatic Identification of User Goals in Web Search," Proc. 14th Int'l Conf. World Wide Web (WWW '05), pp. 391-400, 2005.
[10] O. Zamir and O. Etzioni, " Web Document Clustering: A Feasibility demonstration," ACM (SIGIR, 99) , pp. 46-54.
[11] X. Wang and C.-X Zhai, "Learn from Web Search Logs to Organize Search Results," Proc. 30th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '07), pp. 87-94, 2007.
[12] J. Zhu, J. Hong, and J. G. Hughes. Pagecluster: "Mining conceptual link hierarchies from web log files for adaptive web site navigation," ACM Trans. Internet Technol., 4(2), 2004.
[13] P. Boldi, F. Bonchi, C. Castillo, D. Donato, A. Gionis, and S. Vigna. The query-ow graph: model and applications. In CIKM, 2008.
[14] R. W. White, M. Bilenko, and S. Cucerzan. "Studying the use of popular destinations to enhance web search interaction,". In SIGIR, 2007.
[15] Y. Xie and D. O'Hallaron, "Locality in search engine queries and its implications for caching," In IEEE Infocom 2002, pages 1238{1247, 2002.
[16] J. Srivastava, R. Cooley, M. Deshpande, and P.-N. Tan. Web usage mining: discovery and applications of usage patterns from Web data. SIGKDD Explor. Newsl., 1(2):12, 23, 2000.

Author Profile

Ms. Shital C. Patil received B.E. degree in Computer Science and Engineering from Sant Gadge Baba Amravati University in 2012. Presently in pursuing M.E second year Computer Science & Engineering.

Licensed Under Creative Commons Attribution CC BY