

AN EFFECTIVE SEARCH ON WEB LOG FROM MOST POPULAR DOWNLOADED CONTENT

Brindha.S¹ and Sabarinathan.P²

¹PG Scholar, Department of Computer Science and Engineering, PABCET, Trichy

²Assistant Professor, Department of Computer Science and Engineering, PABCET, Trichy

ABSTRACT

A Web page recommender system effectively predicts the best related web page to search. While searching a word from search engine it may display some unnecessary links and unrelated data's to user so to avoid this problem, the conceptual prediction model combines both the web usage and domain knowledge. The proposed conceptual prediction model automatically generates a semantic network of the semantic Web usage knowledge, which is the integration of domain knowledge and web usage information. Web usage mining aims to discover interesting and frequent user access patterns from web browsing data. The discovered knowledge can then be used for many practical web applications such as web recommendations, adaptive web sites, and personalized web search and surfing.

KEYWORDS

Web Usage Mining, Ranking, Histories, Domain Knowledge, page recommendations.

1. INTRODUCTION

The main goal of this mining is used to find best link for user's searching. Web usage mining is the process of extracting knowledge from web user's access by using data mining technologies. This web usage mining application is called as recommender system. This recommender system is to improve Web site usability. web usage mining prediction process is structured according to web server activity and analyzing historical data such as server access log file or web logs which are captured from the server then these web logs are used capturing the intuition list of the user so as to recommend page views to the user whenever he/she comes online for the next time.

Our paper, we present architecture for capturing recommendations in the form of intuition list of user. Intuition list consist of list of pages visited by user as well as the list of pages visited by other user of having similar usage profile.

The results represent that improved accuracy of recommendations. The Web usage mining process [6] consist of following three inter-dependent stages: collection of data, pre-processing, pattern discovery and analysis. In the pre-processing stage, the click stream data is cleaned and divided into a set of user transactions represents the behavior of each user during different sessions. In the pattern discovery stage, statistical, database, and machine learning operations are executed to get hidden patterns revealing the usual behavior of users, summary statistics on Web resources, sessions, and users. In the final stage of the process, the extracted patterns and statistics

are further analyzed, filtered, which result in aggregate user models that is used as input to applications such as recommendation engines, visualization tools, and Web analytics and report generation tools. The overall process is depicted in Fig. 2. There is different types of models are available.

1.1 Ontology

Ontology is describing the detailed information [1,5,7] from the domain data mining and knowledge discovery it includes definition of basic data mining entities (e.g., data type, dataset, data mining task, data mining algorithm etc.) and allows extensions with more complex data mining entities (e.g. constraints, data mining scenarios and data mining experiments).

1.2 Semantic Network

The term denotes a network which represents semantic relations [2,3,4] between concepts. This is often used as a form of knowledge representation. Semantic data mining is a data mining approach where domain ontology's are used as background knowledge. Such approach is motivated by large amounts of data.

1.3 Conceptual Prediction Model

It is necessary first to present the current status of the field and to identify the associated difficulties. Potential solutions can then be sought. The process of identifying valid, novel, potentially useful, and ultimately understandable patterns from data and also combines the ontology and semantic network model for getting the perfect result by filtering those models result.

2. EXISTING SYSTEM

In an Existing System either ontology or semantic network model was used. The performance of existing approaches depends on the sizes of training datasets. The bigger the training dataset size is, the higher the prediction accuracy is. However, these approaches make Web-page recommendations solely based on the Web access sequences [3] learnt from the Web usage data. Therefore, the predicted pages are limited within the discovered Web access sequences. Integrating semantic information with Web usage mining achieved higher performance than classic Web usage mining algorithms. However, one of the big challenges that these approaches are facing is the semantic domain knowledge acquisition and representation. Manually building ontology of a website is challenging given the very large size [1] of Web data in today's websites. So the performance of the system will be degraded.

3. PROPOSED SYSTEM

In this system using conceptual prediction model which combines the ontology model and semantic network model Proposed system presents a new method to provide better Webpage recommendation based on Web usage and domain knowledge, which is supported by three new knowledge representation models and a set of Web-page recommendation strategies. The first model is an ontology-based model [1] that represents the domain knowledge of a website. The

construction of this model is semi-automated so that the development efforts from developers can be reduced. The second model is a semantic network [2] that represents domain knowledge, whose construction can be fully automated. This model can be easily incorporated into a Web-page recommendation process because of this fully automated feature. The third model is a conceptual prediction model, which is a navigation network of domain terms based on the frequently viewed Web-pages and represents the integrated Web usage[2] and domain knowledge for supporting Web-page prediction. The construction of this model can be fully automated.

The recommendation strategies make use of the domain knowledge and the prediction model through two of the three models to predict the next pages with probabilities for a given Web user based on the current Web-page navigation state.

4. SYSTEM ARCHITECTURE

Architecture describes about the process while searching a word in search engine. User gives the query to the query processor, that query processor is to searching is based on 3 models. Ontology model, Semantic network & Conceptual prediction model, Ontology contains user queries and elaborated content. Semantic contains the relation between the data and corresponding result. By combining these 2 models it has been proposed a conceptual prediction model based upon filtering used to find the result set and also download ratio scheme is used to find the ranking results based on content downloading. These 3 models based on following techniques

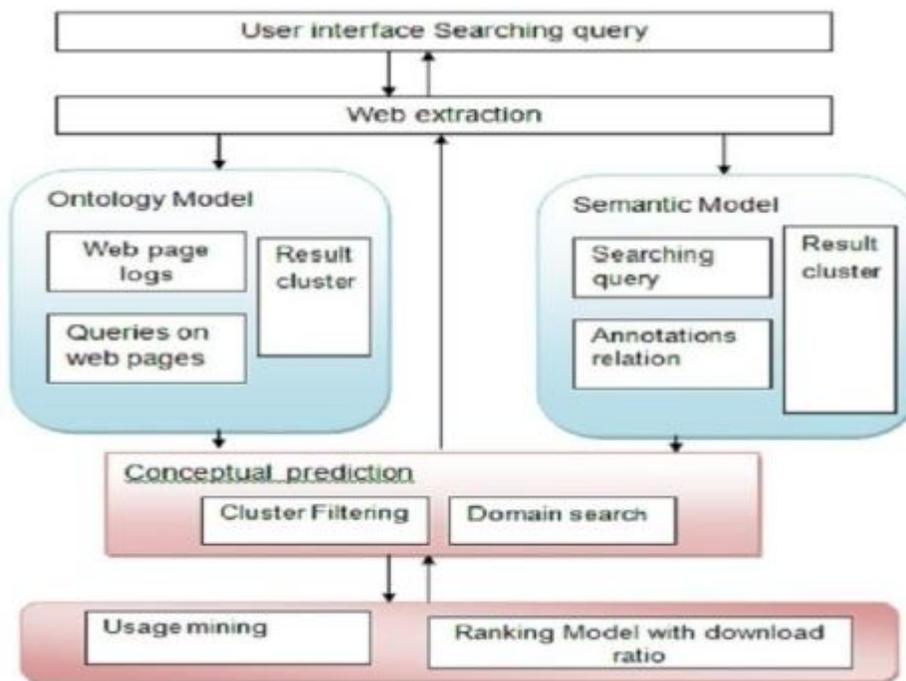


Figure 1. Overall System architecture

5. TECHNIQUES

In our Proposed Work Illustrates following techniques,

5.1. Sequential Pattern Construction

Sequential pattern mining is an important data mining problem with broad applications. It is challenging since one may need to examine a combinatorial Explosive number of possible subsequence patterns.

5.2. Hybrid Clustering

Clustering algorithms often require that the entire dataset be kept in the computer memory. When the dataset is large and does not fit into available memory, one has to compress the dataset to make the application of clustering algorithms possible.

5.3. Apriori Algorithm

The Apriori Algorithms an influential algorithm for mining frequent item sets for Boolean association rules.

Key Concepts: Frequent Item sets: The sets of item which has minimum support (denoted by Item set) Apriori Property: Any subset of frequent item set must be frequent.

6. IMPLEMENTATION

Types to be describes are as follows,

- 6.1 Data Creation and Manipulations
- 6.2 User interface
- 6.3 Query Processing
- 6.4 Usage and Relationship mining
- 6.5 Ranking Model

Figure 2. Usage based Result



Table 1. Ranking Result

No.	Authors	Websites	Users	Rank
1.	Loizou	www.uuu.com	9	1
2.	Selva saa	www.sa.com	7	2
3.	Selva roy	www.zx.com	4	3

6.1 Data Creation and Manipulation

In our type, we chose to create the many website for the specific search. Here the data are posted one by one by admin. The data are created by article posting. All WebPages are manages by admin.

6.2 User Interface

Based on the user’s application logic, User gives the different inputs of query to the query processor .It may be a keyword or content then searching results are retrieved by clusters and that results are filtered by usage.

6.3 Query Processing

This type initiates the data search at server side. Query processing is checking the user query these results are retrieved from the database. Query processing results are combination of WebPages and relationships. And all these queries are checked by the processor for log creation and comparison. This gives the related data’s.

6.4 Usage and Relationship Mining

In This Type Describes About Usage Mining [6]. Web Page Usage Classifications Are Identified And The Matching Results Are Obtained Based On Semantic Relation [8] And Content Relation. Ranking Is Detected By Using Clustering Data And Will Get The Final Results, And These Results Are Updated By Server.

6.5 Ranking Model

In this type the results are produced based on ranking is used to generate the following results and analyze the following functions,

Reports: Article reports User queries report

Analysis: Relations Cluster formation

7. SUMMARY

This paper illustrates, the related works on web usage mining process including web usage data, preprocessing links, and the Sequence pattern construction techniques. Usage based data is the main source for web usage mining; it mainly includes web server logs, proxy server logs and client browser logs. they are the most widely used source in research on web usage mining. Web search access patterns from websites. However, it also includes data's from user profiles, registration details, cookies, user queries and bookmarks from the interactions of users while surfing on the Web. Web usage data are mainly divided into three types, namely web server logs, proxy server logs and client browser logs.

These paper techniques are generally used for extracting statistical knowledge from weblogs. Such knowledge is most useful for analyzing web traffic of a website. Apriori technique can be used for finding related pages that are most often referred together in an access session. Clustering technique can be used to discover user clusters from web logs. Sequential patterns are sequences of web pages accessed frequently by users. Such patterns are useful for discovering user behavior and predicting future pages to be visited by the user.

8. CONCLUSIONS

A new web usage mining process for finding sequential patterns in web usage data which can be used for predicting the possible next move in browsing session's three new models has been proposed. One is an ontology based model which defines domain knowledge. Second is semantic network model which defines relationship and histories. A conceptual prediction model is also proposed to integrate the Web usage and domain knowledge to form a weighted semantic network. Results are filtered in this conception prediction model. That links are displayed in the web page. These frequently used links only updated as a first link and also while downloading a file and that link will be recommended in the web log as a first link and that is the best web page.

ACKNOWLEDGEMENTS

I don't have enough words to describe the profound gratitude and sense of indebtedness which I feel to express towards my supervisor Mr.P.SABARINATHAN, Assistant Professor, Department

of Computer Science & Engineering, PAVENDAR BHARATHIDASAN COLLEGE OF ENGINEERING AND TECHNOLOGY for his invaluable guidance, persistent and useful suggestions, moral support and for making an environment conducive during the course of investigation reported in the present dissertation. Without his constant help and keen interest, it would have been difficult for me to sustain efforts for its completion. I am also grateful to my guide and my respected parents for all possible encouragement and inspiration from time to time given in this submission.

REFERENCES

- [1] Boyce S. and Pahl C.(2007) 'Developing Domain Ontologies for Course Content', Educational Technology & Society, vol.10, pp.275-288.
- [2] Dai M. and Mobasher B.(2005) 'Integrating Semantic Knowledge With Web Usage Mining for Personalization', in Web Mining: Application And Techniques, Global, pp.276-306.
- [3] Ezeife C.I. and Lu Y.(2005) 'Mining Web Log Sequential Patterns with Position Coded Pre-Order Linked WAP Tree', Data Mining and Knowledge Discovery, vol.10, pp.5-38.
- [4] Ezeife C.I. and Lu Y.(2009) 'Fast Incremental Mining of Web Sequential Patterns with PLWAP Tree', Data Mining and Knowledge Discovery, vol.19, pp.376-416.
- [5] Eirinaki M ., Mavroiedi D ., Tsatsaronis G. and Vazirgiannis M.(2006) 'Introducing in Web Personalization :The Role of Ontologies', Mining, pp.147-162.
- [6] Liu B ., Mobashar B. and Nasraoui O.(2011) ' Web Usage Mining ' , in Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, pp.527-603.
- [7] Oberle D . ,Grimm S.and Staab S.(2009) ' An Ontology for Software ', in Handbook on Ontologies vol.2 pp.383- 402.
- [8] Rios S.A. and Velasquez J .D. (2008) 'Semantic Web Usage Mining by Concept - Based Approach for Off-line Web Site Enhancements ' , in Web Intelligence and Intelligent Agent Technology, pp. 234-241
- [9] Stumme G.,Hoth A.And Berendt B.(2004) 'Usage Mining for and on the Semantic Web', pp.461-480.
- [10] Zhou B. (2004) 'Intelligent Web Usage Mining ' , Nanyang Technological University.

Authors

Brindha.S received her B.Tech degree in Information Technology from M.I.E.T Engineering College, Tiruchirappalli in 2012. She is currently doing her ME-Computer Science in Pavendar Bharathidasan College of Engineering and Technology, Tiruchirappalli.



Sabarinathan.P received his BE degree in Computer Science from Annai Mathammal Sheela Engineering College, Namakkal in 2007 and received his ME degree in the same stream in 2010 from Dhanalakshmi Srinivasan Engineering College, Perambalur. He is currently working as an Assistant Professor in Pavendar Bharathidasan College of Engineering and Technology, Tiruchirappalli and his area of interest includes MANET and Data mining.

