Abstract

Web Usage Mining (WUM) and Web Page Recommendation are important in intelligent system nowadays. It is for the web system to direct users to the Web-Pages in their areas of interest. For that purpose Web Usage Mining plays an important role in finding the areas of interest of the user based on user’s previous usage. Recently Web 2.0 is in use, but it does not consider the semantic knowledge for recommendation of Web Pages to the user. Ontology can be used as background knowledge to improve patterns quality. A survey is done on the prediction method for the recommendation of the Web Pages to the user by using the History Knowledge, Ontology and semantic Network to increase the performance of the Web Usage Mining process.

1. Introduction

Web-Page Recommendation is becoming most useful and popular. Websites are growing in numbers. So it is challenging task of the webmasters to organize the contents of the particular websites to gather the needs of the users. This paper will be the better option to the Web Mining Process. Before getting deep into the main concept, let’s understand the supportive concepts.

The concepts we are going to use in the whole paper are as follows:

I. **Web Mining:** It is the application of data mining technique to search or discover patterns from the WEB. There are three techniques of Web Mining:

i. **Web Usage Mining**

ii. **Web Content Mining**

iii. **Web Structure Mining**

i. **Web Usage Mining:** In short WUM, is the process used to extract knowledge and useful information from web server or Server Logs. It is the process to find out what users want or is looking for on the Internet. It is totally a user’s interest whether to search for text or images or audio or video. Web Usage Data (WUD) captures the identity of the web users along with the browsing behavior on the web site.
ii. **Web Content Mining**: Web content mining is the mining, extraction and integration of useful data, information and knowledge from Web page content. The heterogeneity and the lack of structure that permits much of the ever-expanding information sources on the World Wide Web, such as hypertext documents, makes automated discovery, organization, and search and indexing tools of the Internet and the World Wide Web.

iii. **Web Structure Mining**: It is used for extracting patterns from hyperlinks in the web and for mining the document structure.

II. **Semantic Network**: The Semantic Network represents semantic relationships between concepts. It is often used as a form of Knowledge Representation. It can be directed or undirected graphs which consist of vertices which represent concepts and edges which represents the relations between two concepts.

III. **Ontology**: Ontology is the information about the concept. Ontology has been created to facilitate knowledge sharing and reuse in the decentralization and distributed context of the Web Page.

IV. **History Knowledge**: The user’s previous actions are recorded to predict the next web page in which user will have his interest. So the history knowledge is really very important.

V. **Domain Knowledge**: It is a valid knowledge which is used to refer an area of human endeavour or specialized discipline. It is the knowledge about the environment in which the target system usually operates.

VI. **Web Page Recommendation**: The process of recommending or suggesting the web pages to the users in their area of interest.

2. **Literature Survey**

Classical Techniques of web page recommendation includes the result which is generated in the form of web pages, so there is no semantic meaning of common Navigation Profile. There are many problems included in the classical Systems. One of these problems is NEW PAGE PROBLEM. The classical systems failed to recommends the newly added pages or products to the visitors since these pages or products are not in the current common navigation profiles.

So, to overcome the new page Problem the common navigation profile can be extracted in terms of semantic meaning or information. To do this ontology should be used. But in classical systems ontology was not used. Another problem is clustering in existing systems. In clustering number of recommended pages get increased. Unrelated pages or links are recommended which user never prefers.

Previous techniques used the RDF and LCS algorithms.

1. LCS: Longest Common Subsequence
2. RDF: Resource Description Framework

The comparison of two sequences to determine their similarity is one of the fundamental problems in pattern matching. Longest Common Subsequence generates a sequence or a list of recommended products to the user. It is useful for online shopping purpose but what about the other users who does not use such online shopping websites but they still want the personalization in their web page recommendation system.
3. Problem Statement

Problem in the previous systems is there was not any personalization based on ontology and domain or history knowledge. So a personalized-recommendation system which uses domain ontology and history knowledge to enhance the personalizations should be there to use for web page recommendation purpose.

4. Proposed Models

- ThiThanh Nguyen, Hai Yan Lu, Jie Lu proposed a model for personalized web page recommendation based on Ontology, Semantic Network and Domain Knowledge. There will be semantic web usage knowledge which an integration of both domain knowledge and web usage knowledge. The implementation of this model will demonstrate the results which produces significantly higher performances than the previous web usage mining and web page recommendation techniques.

- B. Liu, B. Mobasher and O. Nasraoui, B. Liu proposed a formal framework for integrating full domain ontologies and page clustering. It is a hybrid web recommendation system, which incorporates semantic knowledge and sequence information into pattern generation and clusters Web pages by using a constructed ontology for Web site. It is because the classical process was not preferable by the user for the personalization purpose. The Semantic Web Mining Process will provide an infrastructure that enables not just web pages but databases, programs, services, sensors, personal devices and even household appliances to both produce and consume data on the web.

- H. Dai and B. Mobasher proposed a system which used to represent the items and user-profiles based on ontology in order to provide semantic applications with personalized services. The proposed recommender system is Domain Independent system and it is implemented as a Web Service. This system uses both Implicit Feedback and Explicit Feedback Collection Method which is used to obtain the user’s interest in particular search area.

The first step of this proposed system is Feature Extraction. This approach extracts features from web documents and constructs relevant concepts. Then it builds the Ontology for the website usage by using the extracted features from web documents. After completion of this the semantic similarity of web documents is taken into consideration to cluster the web documents into different Semantic Themes the different themes may contain different preferences.

![Proposed Architecture](image-url)
5. System Overview

General overall framework of web personalization for Web Usage Mining consists of 3 phases:

1. Data Preparation
2. Transformation
3. Pattern Discovery and Recommendation

These phases are the raw phases of the Semantic Network which is used to make relation between concepts for the ontology. The Data Preparation phase is used to transform raw Web Log Files into transaction data that can be processed by data mining techniques. Ontology phase takes domain information and web pages as input and generates site ontology.
6. Conclusion

A survey is done to propose a new method to offer a better Web Page Recommendation System by using Semantic Enhancements by 3 models which are called as Knowledge Representation Models. One model is for Domain Knowledge, another is for ontology and the last is for Semantic Network. This is a conceptual prediction model which is used to integrate the web usage and domain knowledge. It is used to form a weighted semantic network of frequently viewed pages or terms. This system is proposed to predict next Web Page Requests of users through querying the knowledge.

References