Abstract—Web mining is the application of the data mining which is useful to extract the knowledge. Web mining has been explored to different techniques have been proposed for the variety of the application. Most research on Web mining has been from a ‘data-centric’ or information based point of view. Web usage mining, Web structure mining and Web content mining are the types of Web mining. Web usage mining is used to mining the data from the web server log files. Web Personalization is one of the areas of the Web usage mining that can be defined as delivery of content tailored to a particular user or as personalization requires implicitly or explicitly collecting visitor information and leveraging that knowledge in your content delivery framework to manipulate what information you present to your users and how you present it. In this paper, we have focused on various Web personalization categories and their research issues.

Index Terms—Web mining, web usage mining, web personalization.

I. INTRODUCTION

Data Mining (the analysis step of the Knowledge Discovery in Databases process, or KDD), a relatively young and interdisciplinary field of computer science, is the process of discovering new patterns from large data sets involving methods from statistics and artificial intelligence but also database management. Web mining is the application of data mining techniques to extract knowledge from Web data including web documents, hyperlinks between documents, usage logs of web sites, etc [1].

Two different approaches were taken in initially defining Web mining. First was a ‘process-centric view’, which defined Web mining as a sequence of tasks [2]. Second was a ‘data-centric view’, which defined Web mining in terms of the types of Web data that was being used in the mining process [3].

We decompose the web mining to following sub tasks [4]:

- Resource Finding: the task of retrieving indented Web documents.
- Information Selection and Pre-processing: automatically selecting and pre-processing specific information from retrieved web resources.
- Generalization: automatically discovers general patterns at individual web sites as across multiple sites.
- Analysis: validation and /or interpretation of the mined patterns.

II. OVERVIEW OF THE VARIOUS PERSONALIZATION CATEGORIES

Web Personalization is a strategy, a marketing tool, and
an art. The objective of a Web personalization system is to “provide users with the information they want or need, without expecting from them to ask for it explicitly” [8]. Personalization requires implicitly or explicitly collecting visitor information and leveraging that knowledge in your content delivery framework to manipulate what information you present to your users and how you present it. A personalization mechanism is based on explicit preference declarations by the user and on an iterative process of monitoring the user navigation, collecting its requests of ontological objects and storing them in its profile in order to deliver personalized content [9].

A. Phases of Web Personalization

The Web Personalization process divides in to four distinct phases [5].

- **Collection of Web data**—In this, implicit data includes past activities/click streams as recorded in Web server logs and/or via cookies or session tracking modules. Explicit data usually comes from registration forms and rating questionnaires. In some cases, Web content, structure, and application data can be added as additional sources of data, to shed more light on the next stages.

- **Preprocessing of Web data**—In this, Data is frequently pre-processed to put it into a format that is compatible with the analysis technique to be used in the next step.

Preprocessing may include cleaning data of inconsistencies, filtering out irrelevant information according to the goal of analysis. Most importantly, unique sessions need to be identified from the different requests, based on a heuristic, such as requests originating from an identical IP address within a given time period.

- **Analysis of Web Data**—Also known as Web Usage Mining, this step applies machine learning or data mining techniques to discover interesting usage pattern and statistical correlation between web pages and user groups. This step frequently results in automatic user profiling, and is typically applied offline, so that it does not add a burden on the web server.

- **Decision making/Final Recommendation**—It makes use of the results of the previous analysis step to deliver recommendations to the user. It involves generating dynamic Web content on the fly, such as adding hyperlinks to the last web page requested by the user. This can be accomplished using a variety of Web technology options such as CGI programming.

B. Personalization Categories

Here the table shows the details of personalization categories and its description.

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Strategies</th>
<th>Description</th>
</tr>
</thead>
</table>
• Most widespread form of personalization, user information such as name and browsing history is stored (e.g. using cookies), to be later used to recognize and greet the returning user.  
• Implemented on the Web server.  
• Can also jeopardize user privacy. |
| 2.      | Customization[5] | • Takes as input a user’s preferences from registration forms in order to customize the content and structure of a web page.  
• Process tends to be static and manual or at best semi-automatic  
• Implemented on the Web server.  
• E.g. My Yahoo and Google |
| 3.      | Guidance or Recommender Systems[5] | • Tries to automatically recommend hyperlinks that are deemed to be relevant to the user’s interests, in order to facilitate access to the needed information on a large website.  
• It is implemented on the Web server.  
• Relies on data that reflects the user’s interest implicitly (browsing history as recorded in Web server logs) or explicitly (user profile as entered through a registration form or questionnaire). |
| 4.      | Task Performance Support[5] | • A personal assistant executes actions on behalf of the user, in order to facilitate access to relevant information.  
• This approach requires heavy involvement on the part of the user, including access, installation, and maintenance of the personal assistant software.  
• It also has very limited scope in the sense that it cannot use information about other users with similar interests. |
| 5.      | Web usage data mining personalization[9] | • The customer preference and the product association are automatically learned from click stream.  
• In order to avoid the poor recommendations that will lead to disappoint customers, customers who are likely to buy recommended products are selected using decision tree induction. |
The evolution of the algorithms to take into account additional implicit user feedback of the final products chosen and not only the e-shops and services [2]. Researchers must perform more empirical studies that cover different types of online service providers (for example, game malls). Such work will help identify the generic architecture and common components (for example, profiling and matching) reusable in Web personalization [8].

It will be interesting to compare personalized recommender system with a standard collaborative filtering based methodology in the aspect of recommendation performance. And it will also be an interesting research area to conduct a real marketing campaign to customers using this methodology and to evaluate the performance [12].

Extracting useful patterns and rules using data mining techniques in order to understand the users’ navigational behavior, so that decisions concerning site restructuring or modification can then be made by humans [5].

Computing similarity between domain objects and aggregate domain-level patterns, as well as learning techniques to automatically determine appropriate combination functions used in the aggregation process. [13]

Can exploit and enable a more effective integration and mining of content, usage, and structure data from different sources promise to lead to the next generation of intelligent Web applications [1].

### REFERENCES

7. Q. Han, X. Gao, and W. Wu, Study on Web Mining Algorithm Based on Usage Mining, 2010.