

Taxonomical Classification of Web Usage Mining Applications and its Ontological Representation

Satyaveer Singh¹ and Mahendra Singh Aswal²

¹Research Scholar, ²Assistant Professor, Department of Computer Science,
Gurukul Kangri Vishwavidyalaya, Haridwar, Uttarakhand, India

¹Assistant Professor, Department of Computer Science,
Maharishi Markandeshwar (Deemed To Be University), Mullana, Haryana India
E-Mail: satyaveer.sangwan@gmail.com

Abstract - Web usage mining is used to find out fascinating consumer navigation patterns which can be applied to a lot of real-world problems, such as enriching websites or pages, generating newly topic or product recommendations and consumer behavior studies, etc. In this paper, an attempt has been made to provide a taxonomical classification of web usage mining applications with two levels of hierarchy. Further, the ontology for various categories of the web usage mining applications has been developed and to prove the completeness of proposed taxonomy, a rigorous case study has been performed. The comparative study with other existing classifications of web usage mining applications has also been performed.

Keywords: Classification, Ontology, Web usage mining applications, Taxonomy

I. INTRODUCTION

Web usage mining is the appliance of data mining techniques to explore interesting usage patterns from web data which recognizes and better delivers the requests of web-based applications [1]. The overall working of web usage mining is accomplished into three steps, namely, data pre-processing, pattern detection and pattern analysis. During the data pre-processing, the non-responded web requests is removed from the raw weblog file. The cleaned weblog file is transferred into a user session file in order to achieve valuable pattern detection and analysis. Within pattern detection step, some web usage mining techniques is used to generate web access sequences from the user session file and entered into a dataset. Common web access pattern is extracted from these web access sequences. The extracted pattern is valuable web usage knowledge. In the last step, pattern is analysed and determined knowledge is used in a particular web applications, e.g. web page recommendation system. Major applications of web usage mining are to forecast forthcoming accesses. Web caching, web pre-fetching, intelligent online advertisements, web search engine, website restructuring, etc are the other web usage mining applications. The paper presents a taxonomical classification of web usage mining applications. The ontological representation of the taxonomy has also been presented.

The paper has been organized as follows. A brief survey of related work is contained in section-2. The taxonomical

classification of web usage mining applications has been introduced in the section-3. Section-4 defines the ontological representation of taxonomy and implementation. Section-5 discusses the case study of web usage mining applications. The comparison of our taxonomical classification to other existing classification of web usage mining applications has been presented in section-6. The paper is concluded in section-7.

II. RELATED WORK

In this section, the literature has been reviewed concerning the taxonomical classification of web usage mining applications. It has been found that not much work has been done in this direction. J. Srivastava, R. Cooley, and et al [1] describe the dimensions and application area that is used to categorize web usage mining projects. The five major elements- data sources used to gather input, input data types, the number of websites represented in each dataset and the application area concentrated on the project are applied to every web usage mining project. Various web usage mining applications are presented in [2] [3]. A short outline of the present research works in the field of web usage mining techniques, a number of prominent applications and the key area of future research are described in [4]. In [5], the authors have provided the general study on a large range of applications using web usage data that enables readers to evolve a significant understanding of the scope of clickstream data analysis. The paper [6] has presented work done in the web usage mining and the various applications of web usage mining. In [7], the author has presented the detailed procedure of web usage mining, distinct techniques for pattern detection and has also highlighted the various applications and tools used for web usage mining. The paper in [8] has reviewed the various web usage mining concepts, its techniques and applications. However, it has been observed that web usage mining applications are not taxonomically classified. The literature review also finds that no appreciable work has been done in the direction of ontological representation of web usage mining applications.

III. TAXONOMICAL CLASSIFICATION OF WEB USAGE MINING APPLICATIONS

Taxonomy is a particular classification arranged in a hierarchical structure [9]. Taxonomical classification of web usage mining applications has positioned all known web usage mining applications in a hierarchy or relationship map. The goal of using taxonomy is to improve our ability to find relevant content in a large scale collection of content and to learn from the patterns and trends that emerge from a large scale collection of content without manually reviewing each piece of content. With the help of taxonomy, web usage mining applications are arranged into categories and sub-categories according to the position in a hierarchy.

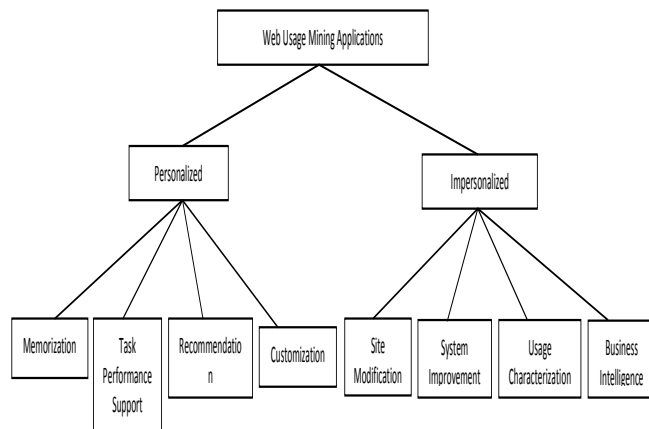


Fig. 1 Taxonomical classification of web usage mining applications

Taxonomical classification of web usage mining applications presents a novel classification approach giving primarily two levels of hierarchy in classification taxonomy as shown in Fig. 1. Web usage mining applications are classified into two main categories: personalized vs. impersonalized [10]. According to taxonomy the description of different categories are given below.

A. First Level Categories

1. Personalized

Personalized means discovering information preferences and needs of individual web users in order to provide personalized websites or web pages for certain types of users [10]. It could be characterized as a technique that grants personalized web action to an appropriate user or set of users. It is the arrangement of the personalization of customized products, services, information or information relating to products or service.

2. Impersonalized

Impersonalized means learning user navigation patterns in order to understand how users use the website [10]. It spotlights on the procedure to learn the user behaviour while navigating inside a website. With the help of user navigation patterns learning techniques, the information providers would be pleased to sight the enhancement of the

efficiency on their websites, which results in altering the website design or by biasing the user's behaviour towards fulfilling the objectives of the site. It is not just the method of modifying and preparing the site's interface for individuals, but also enhancing the site's static structure of the underlying hypertext system as well [10]. Excellent knowledge of the way of visitor's navigation in a website could avoid uncertainty and assist the provider to put the information appropriately.

B. Second Level Categories

1. Memorization

This is the straightforward style of web personalization. Basic information such as username, browsing time and history are stored in web server and used by memorization [11]. User salutation and bookmarks are two major functionalities performed by the memorization. For user salutation, the server uses this information when the user returns and shows the user's name jointly with a welcome message. The web pages visited by the user in the past stores are put into the system and presented to the user by way of a personalized bookmarking schema for that site.

2. Customization

Customization [12] is a more advanced web personalization strategy than memorization. It takes as input a user's preferences from registration forms and uses their knowledge and interest for customizing the content and structure of a page. However, user's knowledge, preferences and interests are taken into account by it to refer the rearrangement of the web page in terms of content, layout and structure etc.

3. Recommendation System

The recommendation system is a much more advanced web personalization strategy than customization and memorization [11]. It is fully based on hyperlinks; it analyzes the previous visit of a user and finds out his navigational path. Next time the system automatically recommends hyperlinks to the user according to his navigational path [13]. It also has an alternative set of hyperlinks for the user according to the interest and preferences.

4. Task Performance Support

Task performance support is a functionality that contains the implementation of an appropriate operation in the favour of a user. This is the most progressive personalization function, innate from personal assistants, which can be treated as client-side personalization systems [11].

5. Site Modification

Usability is the only one concern in the construction and realization of websites. The information provided by the web usage mining about the user behaviour helps in decisions about any rearrangement of the content and structure of the website [8].

6. Business Intelligence

The main aim of business intelligence is to assist people to make fine verdict to enhance company performance and to preserve competitive gain in the marketplace, i.e., it assists companies to make the top choices instantly and freely [7]. Web usage mining is the right approach for extracting information and building a valuable and well-informed catalogue about consumer behaviours.

7. Usage Characterization

The web usage used by the user is for various purposes. Data usage of users is characterized and classified according to their usage activities. The user’s behaviour can be observed by usage regularities on the website [8]. The navigational patterns are used for characterizing the users. Therefore, it contributes the accurate knowledge about the consumer’s communication with browser interface as well as the navigational policy used to browse an appropriate site. It also gives full statistics instance of a variety of client-side events such as the clicking back/forward buttons, saving a file, adding to bookmarks etc.

8. System Improvement

The web usage mining results can be used to enhance the performance of web servers and web-based applications. Web traffic behaviour is captured by the web usage mining and can be used to build up guidelines for web caching, network transmission, load balancing and data distribution [1]. The patterns identified by the web usage mining are helpful for discovering intrusion, fraud, attempted break-ins etc.

IV. IMPLEMENTATION AND ONTOLOGICAL REPRESENTATION OF TAXONOMY

This section presents the implementation and ontological representation of taxonomy. Ontological representation provides a graphically hierarchical representation of any taxonomy. Taxonomical classification of web usage mining application is implemented in a formal language such as RDF and OWL using Jena framework [14]. The output of the implementation generates OWL ontology for the taxonomical classification of web usage mining applications. The designed ontology is well formed and complete. The importance of the ontological implementation of taxonomy will not only keep the contents human understandable but also put in the aspect of machine interpretability and after a significant understanding of the related contents can be incorporated for a more optimal consumption. Graphviz- Graph Visualization Software [15] has been used for viewing the ontology on the graphical form. The graphical representation of the generated valid OWL ontology for the taxonomical classification of web usage mining applications is shown in Fig. 2.

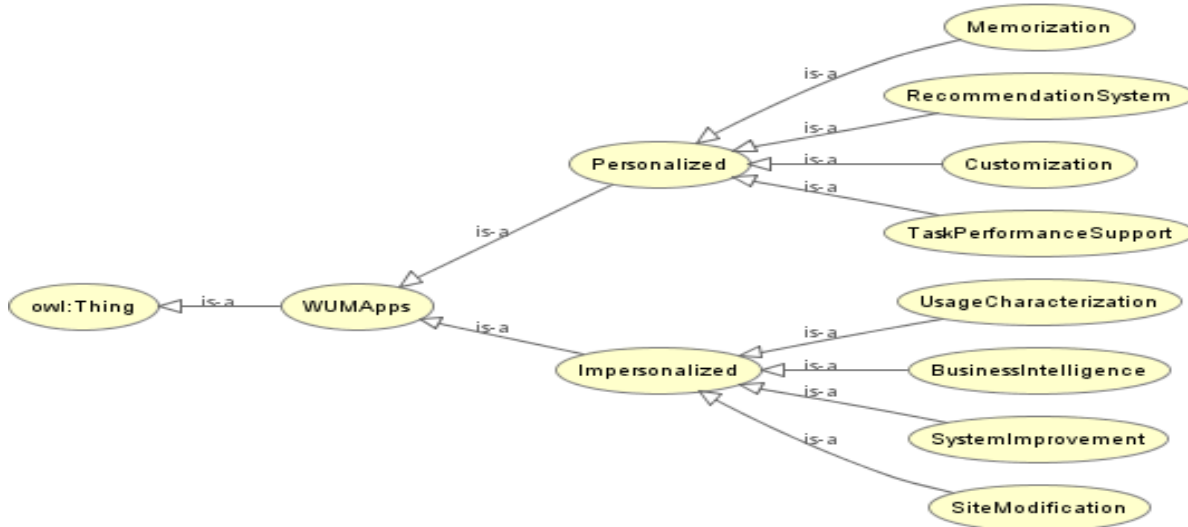


Fig. 2 The graphical representation of valid Ontology at RDF/OWL level

V. WEB USAGE MINING APPLICATIONS: CASE STUDY

To prove the completeness of proposed taxonomy, a rigorous case study has been performed. For this purpose,

we take the example of each web usage mining application and then define taxonomy class according to proposed taxonomical classification, as shown in Table I.

TABLE I TAXONOMY CLASS OF WEB USAGE MINING APPLICATION ACCORDING TO PROPOSED TAXONOMICAL CLASSIFICATION

S. No.	Name of the Web Usage Mining Application	Taxonomy Class According to Proposed Taxonomical Classification
1.	Oracle9iAS Personalization [16]	Web usage mining application → Personalized → Memorization
2.	My Yahoo[17] Google [18]	Web usage mining application → Personalized → Customization
3.	Movielens [19] Amazon [20] eBay [21]	Web usage mining application → Personalized → Recommendation system
4.	AdaptiveInfo.com[22]	Web usage mining application → Personalized → Task Performance Support
5.	Lynda [23] Creative Blog [24]	Web usage mining application → Impersonalized → Site modification
6.	Biz Intel [25] Webtrends [26]	Web usage mining application → Impersonalized → Business intelligence
7.	Google [18]	Web usage mining application → Impersonalized → Web usage characterization
8.	Schechter [27]	Web usage mining application → Impersonalized → System Improvement

VI. COMPARATIVE ANALYSIS

This section presents the comparison of our taxonomical classification of web usage mining applications to other existing classification of web usage mining applications.

The existing classification of web usage mining applications is presented in [1-8]. These are briefly compared with our taxonomical classification of web usage mining applications in Table II and each feature is briefly described below:

TABLE II COMPARISON OF PROPOSED TAXONOMICAL CLASSIFICATION TO OTHER EXISTING CLASSIFICATION OF WEB USAGE MINING APPLICATIONS

Features	Theoretical Complete Taxonomy	Implementation	Ontological Representation	Case Study
Srivastava J, Cooley R, <i>et al.</i> [1]	Only Classified	No	No	No
Li X, Zhang S [2]	Only Classified	No	No	No
Sisodia S D, S. Verma [3]	Only Classified	No	No	No
Suadaa H L [4]	Only Classified	No	No	No
Aggarwal S, Mangat V [5]	Only Classified	No	No	No
Malviya K B, J. Agrawal [6]	Only Classified	No	No	No
Aldekhail M [7]	Only Classified	No	No	No
Shrivastava N J, Singh P S [8]	Only Classified	No	No	No
Our Taxonomical Classification of Web Usage Mining Applications	Yes	Yes	Yes	Yes

A. Theoretical Complete Taxonomy

From the literature review, it is found that the existing classification of web usage mining applications [1-8] is only a categorization of the applications based on the function performed and not a complete taxonomy. Therefore, our taxonomical classification of web usage mining applications is a complete taxonomy.

B. Implementation

Our taxonomical classification of web usage mining applications is implemented in a formal language such as RDF and OWL using Jena framework [14]. The other existing classification of web usage mining applications [1-8] has not been implemented in any formal languages. They have only categorized web usage mining applications based on the function performed.

C. Ontological Representation

The output of the implementation has generated OWL ontology for the taxonomical classification of web usage mining applications. The graphical representation of OWL ontology is presented by using Graph viz- Graph Visualization Software [15] as shown in Fig. 2. The existing approaches of classification for web usage mining applications [1-8] have not provided any ontological representation because they are not implemented in a formal language such as RDF and OWL.

D. Case Study

A case study of our taxonomical classification has been presented to prove that our approach is theoretically complete. For this purpose, we cite the example of each web usage mining applications and define taxonomy class according to proposed taxonomical classification as shown in Table I.

VII. CONCLUSION

In this paper, the theoretical complete taxonomy for web usage mining applications has been provided. Taxonomical classification provided by this work positions all known web usage mining applications in a hierarchy or relationship map. The implementation and ontological representation of proposed taxonomy have also been presented. A case study on the different web usage mining applications has been performed and taxonomy class of web usage mining application has been defined according to proposed taxonomical classification. The work also presents the comparison of our taxonomical classification of web usage mining applications to other existing classification of web usage mining applications.

REFERENCES

- [1] J. Srivastava, R. Cooley, M. Deshpande and P. Tan, "Web usage mining: discovery and applications of usage patterns from web data", SIGKDD Explorations, ACM, Jan. 2000.
- [2] X. Li and S. Zhang, "Applications of web usage mining in e-learning platform", *International Conference on E-Business and E-Government, IEEE*, 2010.
- [3] S. Sisodia and S. Verma, "Web usage pattern analysis through web logs: a review", *Nine International Joint Conferences on Computer Science and Software Engineering, IEEE*, 2012.
- [4] L. H. Suadaa, "A survey on web usage mining techniques and applications", *International Conference on Information Technology Systems and Innovations (ICITSI), IEEE*, Bandung-Bali, Nov. 2014.
- [5] S. Aggarwal and V. Mangat, "Application areas of web usage mining", *Fifth International Conference on Advanced Computing and Communications Technologies, IEEE*, 2015.
- [6] B. K. Malviya and J. Agrawal, "A study on web usage mining: theory and applications", *Fifth International Conference on Communication System and Network Technologies, IEEE*, 2015.
- [7] M. Aldekhail, "Application and significance of web usage mining in the 21st century: a literature review", *International Journal of Computer Theory and Engineering*, Vol. 8, No. 1, Feb. 2016.
- [8] J. N. Shrivastava and S. P. Singh, "A survey of web usage mining: concepts with applications and its future scope", *International Journal of Computer Science Trends and Technology*, Vol. 4, No. 2, Apr. 2016.
- [9] Taxonomy - From Wikipedia, the free encyclopaedia. [Online] Available: <http://en.wikipedia.org/wiki/Taxonomy>, [accessed 12/07/2018].
- [10] S. Dalal, S. Kumar and V. Dixit, "Web mining an application of data mining", *International Journal of Computer Science and Information Technology Research*, Vol. 2, No. 3, 2014.
- [11] Prasanth, "Web personalization using web usage mining techniques", *International Journal of Current Engineering and Scientific Research*, Vol. 3, No. 3, 2016.
- [12] M. Bolin, M. Webber M, and *et al.* "Automation and customization of rendered web pages", *Proceedings of UIST, ACM Press*, 2005.
- [13] B. Mobasher, R. Cooley and J. Srivastava, "Automatic personalization based on web usage mining", *Communications of the ACM*, 2000.
- [14] Jena-a semantic web framework for Java. [Online] Available: <https://jena.apache.org/>, [accessed 12/07/2018].
- [15] Graphviz -Graph Visualization Software. [Online] Available: <https://graphviz.gitlab.io/download/>, [accessed 12/07/2018].
- [16] Oracle9iAS Personalization. Release Notes, Release 2 (v9.0.2) for UNIX.
- [17] Welcome to My Yahoo. [Online] Available: <https://my.yahoo.com/?fr=yfp-t-403>, [accessed 12/07/2018].
- [18] Google. [Online] Available: <https://www.google.co.in/>, [accessed 12/07/2018].
- [19] Movielens. [Online] Available: <https://movielens.org/>, [accessed 12/07/2018].
- [20] Amazon. [Online] Available: <http://www.amazon.in/>, [accessed 12/07/2018].
- [21] eBay. [Online] Available: <http://www.ebay.in/> 12/07/2018].
- [22] AdaptiveInfo.com, [Online] Available: http://www.hugedomains.com/domain_profile.cfm?d=adaptiveinfo&e=com, accessed 12/07/2018].
- [23] Lynda. [Online] Available: <https://www.lynda.com/>, [accessed 12/07/2018].
- [24] Creative Bloq. [Online] Available: <http://www.creativebloq.com/net-magazine>, [accessed 12/07/2018].
- [25] Biz Intel. [Online] Available: <http://biz-intel-inc.com/>, [accessed 12/07/2018].
- [26] Webtrends. [Online] Available: <http://www.webtrends.com/>, [accessed 12/07/2018].
- [27] S. Schechter, M. Krishnan and M. M. Smith, "Using path profiles to predict HTTP requests", *In Proceedings of the 7th International World Wide Web Conference*, Brisbane, Australia, April, 1998.