Leverage Web Analytics for Real Time Website Browsing Recommendations

Claudio Sapateiro[1], João Gomes[2]

[1] School of Technology, Polytechnic of Setúbal, Setúbal,Portugal
claudio.sapateiro@estsetubal.ips.pt
[2] Business School, Polytechnic of Setúbal, Setúbal,Portugal
joao.gomes@estudantes.esce.ips.pt

Abstract. As a websites' structure grow it is paramount to accommodate the alignment of user needs and experience with the overall websites' purposes. Toward this requirement, the proposed website navigation recommendation system suggests to users, pages that might be of her interest based on past successful navigation patterns of overall site's usage. Most of existing recommendation systems adopts traditionally one of the web mining branches. We take a different stance, on web mining usage, and alternatively considered the real time enactment of web analytic tools supported analysis given their current maturity and affordances. On this basis we provide a model, its implementation and evaluation for navigation based recommendations generation and delivery. The developed prototype adopted a SaaS orientation to promote the underlying functionalities integration within any website. Preliminary evaluation's results seem to favor the validation of the present contribution rational.

Keywords: Web analytics, Recommendation systems, Web usage recommendation, Navigation recommendations, User experience.

1   Introduction

Extracting knowledge about web users browsing activity have been a field of research for many application domains, ranging from pursuing user experience (UX) improvement (e.g. recommendation and personalization) to digital marketing endeavours efficacy (e.g. considering segments definition and targeting). Whatever the purpose, it had been approached in different ways and had delve with several challenges. For instance, the term Web mining became mainly popular from the application of datamining (DM) techniques to extract knowledge from Web data including web documents, hyperlinks between documents, usage logs, etc. [1]. Currently, Web mining arena comprehends three major categories: 1) web content mining, 2) web structure mining and 3) web usage mining. Web content mining deals with discovery and categorization of semantic information of website's contents [2]. On the other hand, Web structure mining focus on the web's hyperlink structure, usually involving the analysis of in- and out-links of webpages, as used in search

engines results ranking [3]. Particularly, Web usage mining, had encompassed the application of data mining techniques to aid navigation needs of webpages' end-users [4]. Typical such analysis pursues usage patterns and fed some application domains as: personalization, recommendation (navigation oriented not content based) or adaptive interfaces e.g. [5][6].

This work focus on the web usage domain, in line with navigation oriented recommendations generation but departing from the more traditional approaches that rooted web mining. Typical web mining endeavours follows the conventional data mining workflow: collect data, pre-process it, apply the elected analysis method and produce and validate a model aiming for prediction and/or classification toward end goals such as personalization or recommendation. The current work may be seen more in line with the evolution of the earlier OLAP (On Line Analytical Processing) systems, a branch of Businesses Intelligence that also started to tackle, as an alternative perspective, the analysis of web related data e.g. [7]. Although integrated in some approaches e.g. [8], one do acknowledge and must put forward the divide of the two perspectives. While web mining had mainly aligned with DM methods, whereas OLAP mainly roots on a more descriptive statistics grounding. We argue that producing DM models rely on representative data featuring which typically requires an explicit offline process to accommodate the stages of their development and validation before coming actionable; further they are prone to regular revisions to accommodate website restructure endeavours, dynamic trends of usage and/or growth of website usage data. Nowadays, the OLAP analysis flavour had experience an evolution given the implantation and maturity of web analytics tools (WAT). Such tools allows to monitor website traffic data at a level that affords to readily identify design issues regarding structure, and usage behaviours. This work departs from the typical website develop-monitor-evaluate-redesign (DMER) cycle fed by current web analytics (WA) based analysis, by using WAT provided APIs to actually enact in real time customized on demand analysis' results. The current proposal claim that we may provide some assistance to website browsing through recommendations based on other users successful navigation paths supported on data collected and aggregated by WAT, therefore proactively and up-to-date address possible navigation faults while redesign, if needed, would be reconsidered.

The paper is organized as follows: In the next section we discuss related work and fundamental associated concepts, on section 3 we present the proposed model for web analytics based navigation recommendations generation, section 4 presents the consequent functional prototype developed under the Software as a Service (SaaS) paradigm in order to readily promote provided functionality integration within any website; an evaluation of current work is provided and discussed on section 5 and section 6 concludes the paper presenting the concluding remarks and pointing future work directions.

2 Background

2.1 Related Work

Traditionally, common applications of recommendation systems are toward e-commerce (product) recommendations or content based recommendation; supported on techniques such as association rules elicitation, clustering or collaborative filtering to list a few e.g. [9] . Considering our current aims we had focused particularly on the body of work in web usage analysis, which had produced contributions according five major orientations: 1) Sessions' analysis, 2) Cluster analysis and visitor segmentation, 3) Association analysis and 4) Temporal oriented analysis and 5) Classification and prediction based on user transactions [8]. Session analysis is a common form of generating statistically aggregated measures on sessions. Standard statistic techniques are used to gain knowledge about users' behaviour, considering e.g. time of the day, most frequent: landing, exit or bounce pages; average time on page, etc. Clustering, toward web usage analysis domain, commonly aims to group users either through Web usage mining methods or WA, each within their respective approach variants, focuses on the elicitation of similar browsing patterns. Some works show that behaviour clustering is challenging and often does not presents significant impact on Web surfer behaviour prediction [10]. A possible interpretation of such outcomes may be the lack of sensitivity to user's context. The other classical web usage analysis technique is the discovery of association rules which targets to find groups of pages (sets) that are commonly accessed together. This enables websites to organize the site content and navigation structure more efficiently. Most common approaches to association discovery are based on one of the myriad of variants of the original *Apriori* algorithm e.g. [11]. Although often used with satisfactory results, recommendation based on association rules inherently face degradation if dataset is sparse. This is often the case in Web usage domain given that users' visits only cover a small fraction of the available items (pages). Moreover association rules algorithms' order of complexity becomes combinatorial, so unsuited when targeting ordered sets (as e.g. in navigation flows, considering pages' precedence). The temporal analysis of navigational patterns constitutes the more implicitly time oriented web usage analysis technique. Specific methods to unveil trends attempt to find from inter-session data, sequential patterns on a time-ordered scale [12]. The last considered web usage analysis focus concerns classification and prediction based on Web user transactions. Typically, the focus of classification and prediction is to develop and make use of models produced either by some of the previous techniques discussed (namely clustering) or additional ones (e.g. decision trees or naive Bayesian classifiers) in order to classify a website user/visit accordingly the selected independent variables space that bounds the model. This often requires to develop users' (and new users) profiles through features collection and extraction, to position the user within a model. As earlier stated, in this work we are not focused on generating and updating offline models but replace traditional *classification/predication* by navigation recommendations real time forged through framing the user current website visit within the ones similar from the existing record of website's navigation data.

2.2 Web Analytics

Web Analytics Association defines web analytics as the "measurement, collection, analysis and reporting of Internet data for the purposes of understanding and optimizing Web usage". Despite similar goals had been pursued in the past, the dynamics, volume and complexity of the websites ecosystem had fueled the current state of the art of WAT [13]. Such tools as for instance, Piwik, Adobe Analytics, KISSMetrics, Clicky and Google Analytics (to mention some of the most present nowadays) became a hardly neglectable aid regarding website usage understanding and optimization to further inform user experience design and goals conversion (given the very purpose(s) of the website). To further discuss WA, we must put forward fundamental concepts underlying, WA, WAT (and their concretization in google analytics (GA) in particular - which given its representativeness has supported our developments). WAT collect website traffic data based on users' sessions' clickstream. Such data is organized as *dimensions* (e.g. regional origin, user browser, operating system and device) and *metrics* (e.g. number of sessions, page views, time on pages, bounce rate). GA also address the WA concepts of *goal* and *goal conversion*. Goals can be established quantitatively as for instance, number of users reaching a predefined web page, number of users that visit a given number of pages or number of users that trigger a specific event (e.g. submitted an opt-in form). Accordingly, a user reaching the predefined goals constitutes a conversion. The overall analytical flow concludes, based on the concept of *report* which may be regarded as the output of an inquiry query to the dataset for selected metrics aggregated for the also selected set of dimensions. Reports are further furnished by the possibility of *filtering* the query output data for particular subset of values of interest. Therefore, WAT provides an integrated framework that affords multi-dimensional operations (as e.g. slicing and dicing, rolling up and drilling down) over the dataset with different granularity levels on metrics. Therefore, WA paradigm, grounded on the current generation of tools, provided a number of new functional affordances that eased the burden of conducting websites related data analysis and monitoring, and furthermore, due their often present API provided a way to generate integrated, up to date reports on the fly. Moreover, such API, had allowed in fact to induce consequent actions given some analytical indicators in close to real time on an automated manner as explored in the present work. As with any WA based endeavor, the challenge is to bound the almost unlimited set of dimensions and metrics available through WAT to the proper level of coverage and granularity for the given goals.

3 Web Analytics Based Recommendations

The present line of work adopts website navigation recommendation based on results of analytical queries submitted to the WAT monitoring the website, and the consequent enactment of respective results, as recommendations, in real time. The model tackles two complementary issues: 1) the rational underlying recommendations generations and 2) how and when the recommendations are conveyed to the end user.

3.1 Recommendation Model

One should make a clear distinction on the rational that underlies the proposed model. We are interested on providing browsing suggestions considering similar navigation paths (given the overall website traffic data) and not based or requiring to identify the (returning) user; neither classify her regarding subjects, contents of interest or market segment. So we are not addressing any form of personalization, which its value to recommendations generation in some scenarios is still debatable [14]. Our assumption is that on a given context, given momentarily needs, a user may benefit from being scoped into a behavioral cluster regarding a navigation flow already defined by a set of past users that shared and fulfilled the same needs satisfactorily meeting website designated goals.

The distinctive nature of this work is that such clusters are not inferred by datamining algorithms based on data featuring and required associated processes, but readily and dynamically formed reflecting actual usage measures provided by WAT. Clearly, in order to generate effective recommendations, website usage data must be framed into what constitutes a successful navigation flow. If such care wasn't accommodated in the model one may end up to suggest meaningless navigational paths fueled up by massive usage data of wander browsing sessions, a distinct form of directed browsing [15]. We relied on the WA concept of goal to support what constitutes the criteria for a successful navigation, i.e. a navigation path that worth to recommend, matching user current browsing behavior with navigation paths of successful (goal conversion) sessions. The definition of goals is delegated for website stakeholders. As so, we particularly focus the scope on websites that include purpose and goals definition policies on their development process.

The current model posits that recommendations consist in a set of suggestions that are delivered to a website user proposing the next page she may navigate to, considering her current navigational flow, routing her into a navigation path that had already lead to most goal conversions (of past users that also tread her up to the moment path). Produced recommendations are ranked according the number of users that had choose the current recommended next page and had goal converted within their session. Accordingly, in each page the user navigates into (whether or not a recommended one) a new set of recommendations is produced aiming to direct the user on track to a goal conversion path (acknowledging that her current flow was updated). Both the model and its implementation, further detailed on section 4, were thought and developed as highly customizable. This means, that it is possible, to some extent, to select which WA dimensions are included in recommendations generation. For instance, a website may experience different usage given user's country, device and/or time frame (e.g. if it is under a particular promotional/marketing campaign, redefining momentarily the trends of its usage). Such constrains may be included on recommendations production, e.g. on considering successful navigation paths restricted to those subsets that share the same device that the user is actually using. Country and language dimensions may help to distinguish, and effectively recommend accordingly, differences regarding regional or cultural trends in website usage.
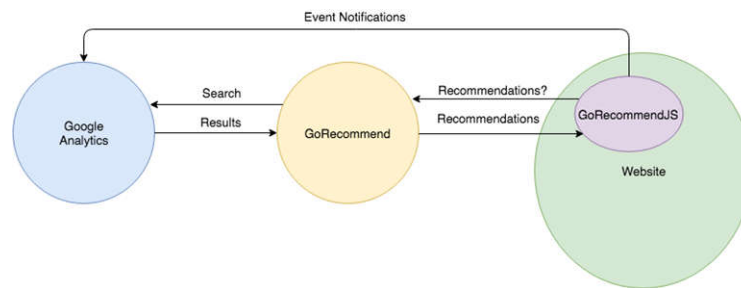
3.2 Recommendations Notification Model

As stated the proposed model also encompasses another concern within the recommendation arena: how and when are recommendations delivered to the end user? Careful must be taken in order to not be too intrusive or disturb the user experience. While the ultimate form of recommendations as a list may be (re)defined (build upon current provided implementation) by websites developers and/or designers, there is an underlying model for recommendations deliver/notifications that we also grounded on WA data. We specified implementation requirements for recommendations delivery on two aspects: number of recommendations and notification timing. Such concerns were grounded on the basis that if users optionally hold some degree of control, within a non-intrusive process, toward the customization of the interface providing the recommendations they may be more prone to adopt (adapt) it [16]. As so, the number of recommendations to deliver may be defined either by the developer (as he configures the provided model implementation integration within the website) and by the end-user in real time (at a given time user may require more or less suggestions). Additionally, the model foresees the possibility to redefine the timing at which recommendations are provided (delaying/advancing). This customization is available to the website developer and may be made as a function of the average and standard deviation of time on page for the current page extracted from the WAT. This allows that users aren't immediately notified with recommendations as they enter in a webpage, but also allows to deliver such recommendations before the average time of such page abandon. Nevertheless, as stated, the model also envisioned the ability of end-users easily fine tune such timing, if desired.
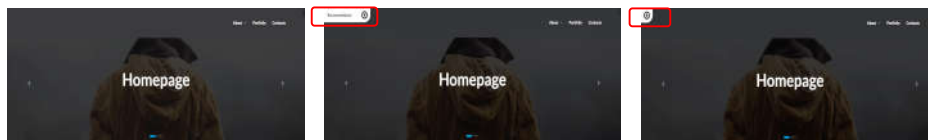
4 Functional Prototype

Two main factors guided the functional prototype development: 1) naturally it should strictly follow the requirements specification outlined by the envisioned model for recommendations generation and delivery, and 2) it should be developed within the SaaS perspective and easily integrated within any website. The overall system architecture is depicted in figure 1. The "GoRecommend" is a server service, developed in PHP through the *Symfony* framework that provides the implementation of the proposed model rational described on section 3. As posited, recommendations are produced based on the results of GA API requests given the respective data provided by the "GoRecommend.js" a customizable client side javascript library that should be integrated in the website. The *GoRecommend* service is REST compliant and communication with the GoRecommend.js is based on json. The *GoRecommend* service maintains a cache and tracks current user navigation for coherence on recommendations proposals. The *GoRecommend.js* client library along with the GA property ID, for site monitoring, constitute the integration requirements to include and start using the recommendation system, although with all configurable parameters reverted to default values. Customization through *GoRecommend.js* by the website developer is possible along several aspects introduced by the proposed model,

namely: 1) time frame covered by GA queries, 2) Activate optional dimensions that may be include on GA queries (e.g. language, country or device), 3) real time on page accuracy improvement, 4) recommendation delivery widget appearance (e.g. positioning, background and text colors), 5) Threshold for the number of recommendations to convey and 6) The timing for delivering recommendations notification to end-user. As posited by the model, such time may be a function of average time on page collected by GA. Such time measure often suffers from several distortions; actually getting the accurate time on page and/or delimiting users' web sessions has posited challenges continuously addressed for some years to the present e.g. [17][18]. We had tackled this problem by implementing a combination of techniques that allows to achieve more accurate measures regarding the GA time on page metric.

 Fig. 1. Overall functional prototype architecture

Figure 2 exhibits a time ordered sequence regarding the recommendations notification. In Fig. 2a (left most) the website visitor just entered *the home page*, after the notifications' configured timing has been achieved an animation on the upper left corner (default positioning) is showed (highlighted in Fig. 2b) and retracted although maintaining an iconic presence (highlighted in Fig. 2c).

Fig. 2. Time ordered sequence of recommendation widget introduction/appearance in user's visit to a webpage

Finally, one concludes the developed functional prototype presentation addressing the recommendation client-side widget affordances to end-user level customizations. Accordingly to the requirements put forward by our model, when widget is activated (mouse over) we may notice that the bottom of the widget provides 4 buttons. Two buttons are related with adding or removing recommendations and therefore controlling the number of suggestions showed/present in the widget (Fig. 3); the other two buttons are related with the ability to delay or quicken the recommendations' notification (widget) appearance timing.
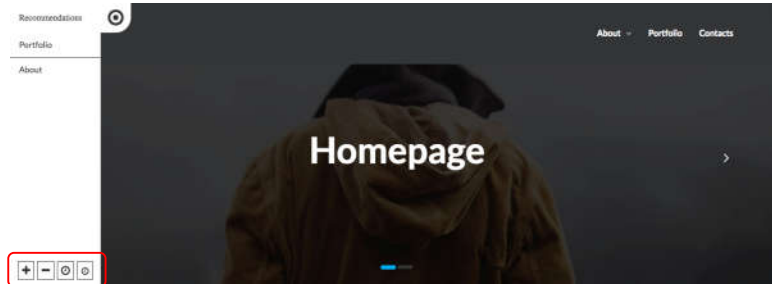
Fig. 3. Recommendation widget customization buttons

5 Evaluation

The evaluation of current proposal was conducted by relying on the GA traffic data from two organizations' websites. One acknowledge that the evaluation of the current work may encompass several dimensions: 1) Does the implementation is able to provide the envisioned rational of the model, 2) Are the generated recommendations effective for the end user and 3) Is the user experience with the recommendation widget satisfactory. Engaging users in evaluation trials (evaluation dimensions 2 and 3) is often a very demanding endeavor if statistically significant results are pursued. Moreover introduce such widget in an organization website (e.g. produce A/B tests) must at least had some forms of earlier validations support. Clearly, it is not possible to jump directly to evaluation dimensions 2 and 3 without a body of results in dimension 1. This paper presents and discusses the dimension 1 evaluation's results based on two distinct test cases, provided in the following two subsections.

5.1 First Test Case

In the first test case we settled a goal (achieve a predefined webpage on the website) collect (through GA) website traffic data and produce sets of recommendations for each page according the underlying logic of the model. After, we redefined the goal and re-produced the recommendations in order to verify if recommendations start to reflect different browsing paths aligned with the current updated goal. Such was indeed the case, reflecting that the implementation copes with the model. Table 1 exhibits the recommendations (order by respective rank, with a threshold of 60) for a period of 60 days while the goal was first reaching "*contacts*" webpage and after the goal redefinition, placing the goal on the "*portfolio*" webpage. One may notice that given the website small structure many pages may lead directly to a recommendation toward the goal page.

Table 1. Partial outcome of recommendations generated on the first test case

Goal: reach "*contacts*" webpage			Goal change: reach "*portfolio*" webpage		
(current) page	(next page) Recommendation	Rank from conversions	(current) page	(next page) Recommendation	Rank from conversions
/ (home)	/contacts	1030	/ (home)	/portfolio	1920
	/portfolio	60		/contacts	450
	/about	60		/about	170
/portfolio	/contacts	170	/contacts	/portfolio	80
	/about	110		/	70
/about	/contacts	120		/about	60
	/portfolio	60	/about	/portfolio	100
				/	60
…	…	…	…	…	…

5.2 Second Test Case

The second test case addresses the natural issue of accumulating simultaneously several goals within the website. For this purpose we have relied on other organization's website (with a larger structure) traffic data.

Table 2. Partial outcome of recommendations generated on the second test case

Goal: reach "*contacts*" webpage			Simultaneous Goals: "contacts" & "*clients*"		
(current) page	(next page) Recommendation	Rank from conversions	(current) page	(next page) Recommendation	Rank from conversions
/ (home)	/contacts	560	/ (home)	/contacts	470
	/where	100		/clients	240
	/who	60		/where	120
/clients	/where	60		/services	60
/who	/contacts	230	/clients	/contacts	60
/where	/contacts	110	/contacts	/clients	140
/services	/contacts	170		/who	110
	/logistics	70		/	90
				/services	60
			/team	/who	70
			/projects	/team	60
			/who	/contacts	70
				/clients	60
				/team	60
				/who	60
			/where	/contacts	140
				/clients	90
			/services	/logistics	100
				/contacts	80
				/clients	60
…	…	…	…	…	…

After an initial definition of a goal (reach *"contacts"* webpage) a second goal was added (reach *"clients"* webpage), and indeed the recommendations had reflected goals coexistence by starting to include recommendations of entry pages in any of the navigation flows that had led to one of the defined goals conversion accordingly (Table 2). One may notice that some recommendations lean toward one of the goals, while others may contemplate both, in consideration for the given navigation path. Additionally in some cases only indirect paths are suggested following the assumed conversions' threshold.

6   Conclusions

This work proposes an alternative approach to websites' navigation recommendations generation. Rather than incurring through the common web mining based approaches we redirect the focus to the real time enactment of web analytical tools supported analysis, given their current maturity and functional affordances. We do not claim that traditional web mining approaches have no place, given the assumptions that we have discussed on related work, but rather both approaches may complementary co-exist, leveraging each other strengths. We presented the underlying rational of our model for the recommendations production and delivering. Both aspects were reflected on the consequent functional prototype implementation as highly customizable to readily accommodate its functionality integration within any website; further promoted by the adopted SaaS orientation.

Our overall recommendation logic relies on the assumption that if a website usage and mission policy bounds website's development process, goals, in the web analytics sense, may be defined and guide navigation recommendations production. Moreover, we focus on website usage patterns, rather than on user's identification (or any form of personalization) under the assumption that in a given context a user may have navigation requirements closer to those that have pursued and effectively benefited from the same goals' underlying criteria; rather than those within her classified personalized *"segment"* or *"cluster"*.

We discussed three different evaluation dimensions and respective challenges that bound the current proposal validation; and presented the conducted evaluation process and results for the first (root) dimension. This refers to the extent that the implementation consubstantiated the model; and despite acknowledging that further evaluation should be conducted, the achieved results are favorable. We must make notice that a biased thinking may emerge from the examples that supported the evaluation, which is that if a goal is to reach a webpage within the website than the recommendation should readily be toward the goal page(s). This is a reductive conclusion since (notably larger sites) navigation may inherently hold an increased value besides meeting the end goal and; goals' and their logic can entail other purposes rather than just reach the end-page(s) (as presented and discussed). As an example, it will be naïve to expect that a visitor will readily choose follow a recommendation to fill and submit an opt-in form (as goal conversion) without some browsing activity that will motivate her to do so, possible based on intermediaries goals (although not to be seen has hard coded suggestions), or informed through

suggestions matching browsing path records of successful and proficient navigations, as envisioned. Moreover, also for website's stakeholder, several reasons may dictate that it may be desirable that visitors navigate around (guided by best/rich navigation sessions) before end-goals conversions.

All in all we perform a little step toward to the real time enactment of website usage data analysis' results given its underlying policies based on a paradigm that may fuel a new perspective on developing adaptive interfaces, as we are currently exploring along with user level validation dimensions.

References

[1]   P. Mehtaa, B. Parekh, K. Modi, and P. Solanki, "Web Personalization Using Web Mining: Concept and Research Issue," *Int. J. Inf. Educ. Technol.*, vol. 2, no. 5, p. 510, 2012.

[2]   P. Fragkou, "Information extraction versus text segmentation for web content mining," *Int. J. Softw. Eng. Knowl. Eng.*, vol. 23, no. 8, pp. 1109–1137, 2013.

[3]   C. E. Dinucâ, "Web structure mining," *Ann. Univ. Petrosani Econ.*, vol. 11, no. 4, pp. 73–84, 2011.

[4]   Q. Han, X. Gao, and W. Wu, "Study on Web Mining Algorithm based on usage mining," in *9th CAID/CD*, 2008, pp. 1121–1124.

[5]   P. Ackermann, C. A. Velasco, and C. Power, "Developing a Semantic User and Device Modeling Framework That Supports UI Adaptability of Web 2.0 Applications for People with Special Needs," in *International Cross-Disciplinary Conference on Web Accessibility*, 2012.

[6]   D. Parra, "Beyond lists: studying the effect of different recommendation visualizations," *Proc. sixth ACM Conf. Recomm. Syst.*, pp. 333–336, 2012.

[7]   R. Alves, O. Belo, and F. Costa, "Mining Clickstream-Based Data Cubes.," in *11th Database Engineering and Applications Symposium*, 2007, pp. 120–128.

[8]   B. Liu, B. Mobasher, and O. Nasraoui, "Web Data Mining," *Web Data Min. Explor. Hyperlinks, Contents, Usage Data*, pp. 449–483, 2007.

[9]   R. Suguna and D. Sharmila, "An Efficient Web Recommendation System using Collaborative Filtering and Pattern Discovery Algorithms," *Int. J. Comput. Appl.*, vol. 70, no. 3, pp. 37–44, 2013.

[10]   M. T. Hassan and A. Karim, "Impact of behavior clustering on web surfer behavior prediction," *J. Inf. Sci. Eng.*, vol. 27, no. 6, pp. 1855–1870, 2011.

[11]   T. Slimani and A. Lazzez, "Efficient Analysis of Pattern and Association Rule Mining Approaches," *Int. J. Inf. Technol. Comput. Sci.*, vol. 6, no. 3, pp. 70–81, 2014.

[12]   M. Abramson, "Learning Temporal User Profiles of Web Browsing Behavior," *Nav. Res. Lab.*, pp. 1–9, 2014.

[13]   K. Nakatani and T. Chuang, "A web analytics tool selection method: An analytical hierarchy process approach," *Internet Res.*, vol. 21, no. 2, pp. 171–186, 2011.

[14]   C. Li, "When does web-based personalization really work? The distinction between actual personalization and perceived personalization," *Comput. Human Behav.*, vol. 54, pp. 25–33, 2016.

[15]   L. Chiarandini, "Characterizing and Modeling Web Sessions with Applications," Universitat Pompeu Fabra, 2014.

[16]   J. E. Pérez, X. Valencia, M. Arrue, and J. Abascal, "Elaborating a web interface personalization process," *ACM Int. Conf. Proceeding Ser.*, vol. 07–09–Sept, pp. 1–4, 2015.

[17]   B. Berendt, B. Mobasher, and M. Spiliopoulou, "Measuring the accurary of sessionizers for web usage analysis," *Proc. Work. Web Min. 1st SIAM Int. Conf. Data Min.*, pp. 7–14, 2001.

[18]   A. Halfaker, O. Keyes, D. Kluver, T. Nguyen, K. Shores, and A. Uduwage, "User Session Identification Based on Strong Regularities in Inter-activity Time," in *24th International Conference on WWW*, 2015, pp. 410–418.