

“What can I help you with today?”

Exploring Opportunities of Learner Modeling for Online Educational Portals

Sebastian Wollny¹, Jan Schneider¹, Marc Rittberger¹ and Hendrik Drachsler^{1,2,3}

Abstract: Online educational Portals (OEPs) subsume a field of online repositories for a wide range of stakeholders. They are characterized by easily accessible structures that do not require visitor accounts. Most OEPs therefore provide content in one way to all visitors also known as “one size fits all” approach. With this study, we examine if Web Analytics data can be used to infer the modeling of learners for OEPs. This would be the basis for additional and more personalized ways of providing content to various stakeholders. In order to draw conclusions about opportunities and limitations of Web Analytics in this regard, the data structure of the Fachportal Pädagogik, as one of the largest educational OEPs in Germany, is compared with a common Learner Modeling Framework. The evaluation of the results finally leads to two major challenges that must be overcome in order to achieve personalized content and learning experiences on OEPs.

Keywords: Learner Modeling, Personalization, Adaptive Hypermedia, Learning Analytics, Recommender Systems, Open Educational Resources, Online Libraries, Online Educational Portals, Digital Assistants, Chatbots

1. Introduction

Online educational portals (OEP) are a container term for a variety of repositories containing online learning and teaching materials. The diversity of OEP repositories ranges from well-established online libraries over learning support services, such as edutags.de, until OER portals such as Merlot.org or the Education Portal and educational content servers such as Bildungserver.de in Germany. The Open Data movement is currently also creating new types of OEPs that will deliver content and data sets to a wide audience.

¹ DIPF | Leibniz Institute for Research and Information in Education, Educational Technologies,
Rostocker Straße 6, D-60323 Frankfurt am Main, {wollny | schneider.jan | rittberger | drachsler}@dipf.de

² Goethe-Universität Frankfurt am Main, Educational Technologies, Robert-Mayer-Str. 10, D-60054 Frankfurt
am Main [lastname@em.uni-frankfurt.de]

³ Open University of the Netherlands, Welten Institute, Valkenburgerweg 177, NL-6419 AT Heerlen,
[firstname.lastname@ou.nl]

Independent from the kind of data hosted in OEPs, they are all characterized by an easily accessible structure that can facilitate services for a variety of stakeholders at any moment. They are designed to fulfill the demands of various stakeholders (learners, teachers, curators, editors, etc.) [CEP14], [Ch18]. However, visitor accounts proved to be impracticable in the past as they were not used intensively. As a result, OEPs restricted themselves to a "one size fits all" approach that treats all visitors uniformly. Studies have shown a drop of visitors attributing it to the "one-size-fits-all" approach [CL07], [Jo09], [Wi07]. This raises the question of the "one size fits all" approach is still up-to-date, or whether personalized content is better suited to individual information needs.

In this paper we analyze the web analytics data structure obtained from the Fachportal Pädagogik, one of the largest scientific OEPs in Germany, to examine the opportunities and technical limitations of Web Analytics for modeling learners in the context of an OEP.

Based on the findings of the Web Analytics investigation, we argue that educational portals need to focus more on learner modeling in order to overcome the one-size-fits-all approach and increase their usability, educational effectiveness and general level of satisfaction [AB08]. Especially with the upcoming open data repositories⁴ joining the OEP family, it would be a great advantage, if OEPs would support the diversity of stakeholders with personalized content for individual information needs and develop beyond the one-size-fits-all approach.

The structure of this paper is as follows, first we shortly describe the state of the art of *Learner Modeling* and how it could be applied for OEPs (section 2), thereafter we review current information that can be taken out of Web Analytics in OEPs (section 3). Next, we introduce our Web Analytics study (section 4) and present the main findings (section 5). Finally, we discuss our results (section 6) and conclude our research with two challenges that OEPs have to overcome to fulfill the information needs of their stakeholders in the 21st century.

2. State of the Art on Learner Modeling

Looking at *Learner Modeling*, it is a technique that is strongly linked to the personalization of learning or adaptive learning [CV13]. Since most stakeholders of OEPs have specific information needs, we consider them as learners. The term *Learner Modeling* traditionally refers to the modeling of learners related to knowledge diagnosis and adaptive scaffolding. Through intensive research on User Modeling [BM07], the scope of this traditional *Learning Modeling* approach is extended and therefore used in

⁴ An overview for Germany is available on: <https://open-educational-resources.de>, 24.03.2019

this form. Especially in OEPs, adaptations or personalization of content could help learners to orient and structure their information needs [BR09]. Adaptations to Learners are in many OEP cases limited to recommendation systems, which rely on collaborative or semantic filtering techniques [Ve12], [Dr15] and are therefore often driven by a content-based approach.

Learner Modeling can be divided into three phases according to [Ga07]: 1. Data Collection, 2. Profile Construction and 3. Profile Application (Fig.1).

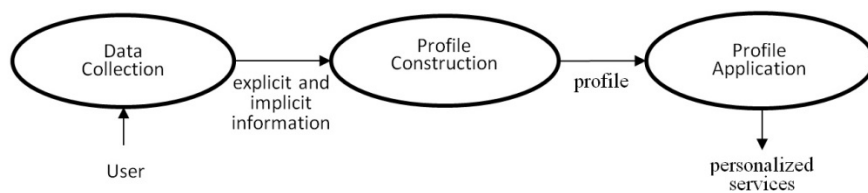


Fig. 1: Learner Modeling process [Ga07]

The data collected in the first phase can be distinguished between explicit and implicit. Among other behavioral data, the browser cache, search logs, and especially weblogs can be used to determine implicit learner profile data. In the next phase, collected data can be transferred to a learner profile, which is then applied in the third phase. For providing personalization based on *Learner Models*, it can be used for memorization, guidance, customization and task performance support of learners [Ca09].

As our study mainly focuses on creating learner models (profiles) out of collected Web Analytics data, we especially focus on phase one and two of the modeling process. A prerequisite for an extension of adaptations in the context of OEPs is the correct extraction of learner characteristics (Data Collection), which can be combined to a learner model (Profile Construction). There are several models of learners presented in previous work [RL08], [ND08], [Sp00]. For this paper, we selected a *Learner Modeling Framework* that is based on Brusilovsky and Millán [BM07]. This *Learner Modeling Framework* seems to be most appropriated for our purpose, as it is a common model in Learning Analytics [CMS17], Adaptive Learning [MDK11] and Recommender Systems for Learning [Ve12] to describe the characteristics of learners that are supported with personalized content. It consists of six characteristic categories about learners, which are put together to an overall model of a learner. These categories are 1. Knowledge, 2. Interests, 3. Goals and Tasks, 4. Background, 5. Individual Traits, 6. Context of Work. In the following, the categories are presented and explained by examples.

Knowledge

A learner's knowledge is technically often represented by directed graphs or ontologies. It is a changeable feature, as learners are able to learn and forget. Knowledge features are often considered important because they reflect the level of understanding in relation to the domain. After extraction, they can be used to adapt the learning process in the way, that it helps to expand knowledge and avoids unwanted repetitions.

Interests

Interest reflects the attention or focus of learners. It can be grouped into subject areas present in the respective domain and described by weighted keyword vectors or overlay models. Learner interests are an important reference point for generating learning paths. Especially for use cases in which certain subject areas are examined, current interests define directions for further exploration. This type of learner feature is often used as a core component of recommender systems to support learner exploration of repositories.

Goals and Tasks

Learning goals and tasks are related features. Goals are personally desired states of learners in the future. They are the cause of personal tasks that have to be mastered in order to reach the goal. Goals and tasks reflect the information needs of learners. They reflect the information needs of learners and can be selected or defined in adaptive learning systems. The nature of personal learning goals incorporates the relation to different time periods. While graduating from school can be a long-term goal, short-term goals are, for example, finding the solution of an equation.

Background

The learning background is a feature category that shares similarities with knowledge features. It defines past experiences that are not directly related to the respective domain. The background features of learners can be combined into personas or stereotypes. While a professor has a larger research background, a student has a more curricular background.

Individual Traits

Individual character traits are personal characteristics that cannot be classified in any of the previous categories. These are mainly preferences and habits that affect the learning process. These kinds of features can be extracted through psychological tests.

Learner Context

The learner context consists of all the features that describe the learner's current situation. In addition to the use of device types, the learner context also includes the learner's location, the affective state, social and personal context, and physical environment.

2.1. Example of applying a Learner Model to an OEP

The following table (Tab. 1) shows two examples of learners using an OEP. Both learners would like to find out something about self-regulated learning (SRL) but have different reasons and strategies. While Anna prefers to look for a variety of practical content on SRL, Bob would prefer a single meta-study. In addition, Anna would be happy about a paper in the context of biology teaching, while Bob prefers to look for

general insights. Both have different features in all feature categories, which could also lead to a differentiated treatment of these learners.

Visitor	Knowledge	Interests	Goals	Tasks	Back-ground	Indiv. Traits	Work Context
Anna	Didactics, Biology	Teaching Methods	Biology Lesson Preparation	Find SRL Research	Teacher	Wants to understand things in detail	Last year students wanted to work independently, Desktop, Evening
Bob	Psychometrics, Statistics	Mathematics, Psychology	Finding evidence for SRL Model	Analyze State-of-the-Art Papers	Researcher	Likes fast + precise results	Preparation of a Presentation, at work

Tab.1: Learner Model Examples

3. Web Analytics

Collecting data about learners is a requirement for modeling them. Web Analytics data can be seen as an indication for processes on websites. Similar to key performance indicators (KPI) from the business sector [CP12], these Web Analytics indicators are important for the interpretation of visitor actions on websites [Ke11]. Also for learning on OEPs Web Analytics data could be suitable to personalize learning. To support *Learner Modeling* on OEPs, Web Analytics measuring methods have the advantage of being universally applicable to a wide range of OEPs and common for data collection and analysis on these portals.

Web Analytics is defined as the collection, measurement, analysis, and reporting of web data with the main purposes to identify web traffic and usage patterns [BBC07]. The data usually comes from four sources: 1. Direct HTTP request data, 2. Network level and server-generated data associated with HTTP requests, 3. Application level data sent with HTTP requests, and 4. external data [ZP15].

In contrast to local device recordings, which provide a detailed learner model, website recordings have the advantage to collect learner data of large groups without the requirement to manually install software on each single device [Ga07].

In the following, we group the major Web Analytics techniques into the following four categories that we will also use of reviewing the Web Analytics results: A. Click-Stream Analysis, B. Session Analysis, C. Visitor Analysis, and D. Event Analysis.

A. Click Stream Analysis. The most often used methods can be subsumed by the Click-Stream Analysis. It contains the logging of click and HTTP requests that can be assembled to a personal usage log of a website. These clicks and requests are grouped into sessions or visits, which are defined as active time periods of visitors on the website. As indicators for the attractiveness of pages the click-through rate, bounce

rate and the time on page can be used. They express if the visitor stays on the website or rather browses to another website.

- B. Session Analysis.** For the analysis of sessions, four data types are important: 1. The landing page, which is the first visited page of a visitor, 2. The exit page, which is the last visited page of a visit, 3. The visit duration and 4. The referrer. Latter can be seen as the external page, which linked to the website and caused the landing page request. Nowadays a large amount of website traffic is referred by search engines.
- C. Visitor Analysis.** The visitor itself can be analyzed by the information they share with the website through their browsers. This could be browser settings or cookies, which are also used to identify the visitor. Whether a visitor is new or returning can be classified in this way too. An identification method, called Fingerprinting, is based on the measurement of subtle differences in communication between browsers and servers, which are caused by these individual configurations. This measurement and the associated analysis is carried out on the server side, which is why large visitor groups can be covered with it.
- D. Event-based Analysis.** These events are specific to the purpose of a website. For commercial websites, an event could be considered for example as a click on the “buy” button. Events can be defined flexibly. They enable website owners to protocol all necessary and interesting actions. By analyzing website events, the effectiveness of online advertisement campaigns or the effects of website adaptations can be measured.

With these four major categories in mind, we will examine in section 4, how data structures from Web Analytics can be matched to the Learner Modeling Framework.

4. Method

In this paper, we explore whether the use of Web Analytics techniques on the log data of educational portals can be used to create a learner model for an OEP without having to resort to visitor accounts. Our main Research Question is, therefore:

RQ1: Can we use Web Analytics log data of OEPs in order to create a learner model based on the Learner Modeling Framework?

In order to identify to what extent it is possible to create a model of learners, based on the *Learning Modeling Framework*, we guide our examination with the use of the following sub-questions:

RQ1a: Can we use Web Analytics to identify learner knowledge?

RQ1b: Can we use Web Analytics to identify learner interest?

RQ1c: Can we use Web Analytics to identify learner goals and tasks?

RQ1d: Can we use Web Analytics to identify learner backgrounds?

RQ1e: Can we use Web Analytics to identify learner individual traits?

RQ1f: Can we use Web Analytics to identify learner context?

To give answers to these questions we will analyze the Web Analytics log data structure of the Fachportal Pädagogik⁵. The findings to sub-questions are rated on two five-point scales, which we call the Applicability Score and Variety Score. While the Applicability Score describes the quality of data has been found, the Variety Score describes the number of deducible features has been found. The scales of the scores are described by the following table (Tab. 2):

Applicability Score		Variety Score	
1	No applicable data found (but indications may be possible)	1	No feature values deducible
2	Weak indications found	2	Small number of feature values deducible
3	Strong indications found	3	Medium number of feature values deducible
4	Noisy feature values found	4	Large number of feature values deducible
5	Applicable feature values found	5	Wide variety of feature values deducible

Tab. 2: Applicability Score scale and Variety Score scale

4.1. Data Source – Fachportal Pädagogik

The Fachportal Pädagogik is one of the most visited OEPs in the context of scientific education in Germany. With over 800.000 Online-Sessions per year, it offers its visitors access to an overview in digital databases, scientific full-texts and library registries. The data pool includes more than 1.000.000 entries on educational literature and obtains its content primarily from library networks.

We used the Matomo⁶ Web Analytics engine in order to track and analyze the data of the Fachportal Pädagogik. It enables state-of-the-art Web Analytics with the possibility to access raw data. Data obtained in this way is mainly stored in three tables. While the log action table contains all actions that have been detected so far, the log visit table is a

⁵ Reachable through <https://www.fachportal-paedagogik.de/>, 24.03.2019

⁶ Until December 2017 Matomo was called "Piwik".

record of session data. Actions can be mainly seen as visited URLs, which are assigned to a session by the “log link visit action” table.

By examining the elements of the Web Analytics log data tables we can identify the information types of the four Web Analytics technique categories (A to D) in Tab. 3. It indicates which of the three Matomo data tables are required to extract related information types. These must be considered in many cases in the context of the portal’s content and have to be combined with it to draw conclusions. In our case, this connection is made by the URL in the ‘log action’ table, since it determines the page content.

Information type	log action	log visit	log link visit action
A. Click-Stream Analysis			
<i>Click Paths</i>	x		x
<i>Time spent on pages</i>	x	x	x
<i>Classification of requested pages</i>	x	x	x
B. Session Analysis			
<i>Session indicators</i> (visit duration, referrer, returning visitor, ...)		x	
<i>Session behavior</i> (e.g. exploration / exploitation)	x		x
<i>Topic- and content type-wise classification of sessions</i>	x	x	x
C. Visitor Analysis			
<i>Visitor identification</i>		x	
<i>Time since the last visit</i>		x	
<i>Visitor preferred language</i>		x	
<i>IP address</i>		x	
<i>Visitor location (incl. time zone)</i>		x	
<i>Device information</i> (type / brand / OS version / screen resolution, ...)		x	

<i>Browser information</i> (type / version / plugins)	x		
D. Event-based Analysis			
<i>Bounce rate</i>	x	x	x
<i>Click-through rate</i>	x	x	x

Tab. 3: Overview of Web Analytics information types in Matomo log data.

5. Results

In the following, we present the findings we derived by comparing Web Analytics data and exemplary data table contents retrieved by a perennial record with feature categories of the *Learner Modeling Framework* in order to identify technical opportunities and limitations of modeling learners on OEPs. In addition, we rate the opportunity of Web Analytics to learner modeling with the Applicability Score (AS) and the Variety Score (VS) for each aspect of the *Learner Model Framework*.

RQ1a - Knowledge: Knowledge is usually inferred through assessments. We were not able to identify any type of knowledge assessment in the OEPs. However, we could identify some weak indicators about the learner's knowledge. For example, we could assume that after visiting a page, the learner has some knowledge about its content. **AS: 2, VS: 2**

RQ1b - Interests: The learner's click-path data informs us about the learner's attention, which in turn provides some indications about the learner's interest. The available data is limited to the interaction with the portal, therefore the identified interest of the learner is restricted to these interactions. **AS: 5, VS: 4**

RQ1c - Goals and Tasks: The OEP has no feature where it asks the learner directly for their current goals or tasks. Therefore, we were not able to identify any strong indicators about them. Nonetheless, we argue that for long sessions click-path can be used as an indicator of the current task. **AS: 2, VS: 3**

RQ1d - Background: The strongest indicator identified for inferring the background of the learner is the access to content that was created specifically for a visitor group (e.g. content for elementary school grammar teachers). However, not all the content stored in the portals is targeted only for specific groups. Prolonged tracking of the learner could strengthen this indicator by looking at more content accessed by the learner, as well as identifying their level of familiarity with the portal. **AS: 2, VS: 2**

RQ1e - Individual Traits: Web usage behavior is the only indicator we have been able to identify that is related to the individual characteristics of the learner. This provides us only a weak indication because the generalization of visitor behavior to individual

character traits is a complex process. Moreover, the possibilities to adapt the OEP in our case on the basis of specific individual traits are very limited. **AS: 3, VS: 2**

RQ1f - Learner Context: Regarding the learner context we could identify the location, local time, device being used by the learner. The practical personalization use of this type of contextual data is limited. Important aspects of the learner context such as affective state, social and personal context, and the physical environment were not identified. **AS: 2, VS: 2**

6. Discussion

The aim of this research is to assess the technical possibilities of *Learner Modeling* through Web Analytics. We retrieved results in all sub-questions and want to use them to answer the main Research Question RQ1 (Tab. 4):

	Knowledge RQ1a	Interests RQ1b	Goals and Tasks RQ1c	Background RQ1d	Indiv. Traits RQ1e	Work Context RQ1f	\emptyset
AS	2	5	2	2	3	2	2,7
VS	2	4	3	2	2	2	2,5

Tab. 4: Overview of Application Score and Variety Score ratings

The findings show limitations of the Web Analytics data structure in all feature categories. Besides RQ1b (Identification of learner interests), the data structure of Web Analytics engines does not generally allow to answer the sub-questions (RQ1a,c,d,e,f) positively. In terms of modeling knowledge (RQ1a), goals or tasks (RQ1c), the learner's background (RQ1d), individual traits (RQ1e) and learner's context (RQ1f) we found that there is the possibility that Web Analytics could give indications for some sessions or learners.

This indicates that data obtained through Web Analytics is already significantly limited in terms of applicability ($\emptyset = 2,7$) and variety of deducible learner features ($\emptyset = 2,5$) by its structure to model learners. Although Web Analytics seems to be suitable to model the interest of learners (RQ1b), many applications in this context are already covered by recommender systems. In addition, the interest feature can change dynamically over time and with context. A longer observation period is likely to improve data quality but not expected to lead to fundamental changes in the assessment.

In order to overcome the "one size fits all" approach of OEPs, alternative tools should be considered in addition to Web Analytics. In this context, interactive assistants in the form of Chatbots, new forms of search engines or content presentations are conceivable. Although learner profiles retrieved through Web Analytics foreseeable are too limited to represent the learner as a whole, they reflect the individual interaction of a resource and

must be considered in the context of data privacy. This raises two major challenges that need to be addressed in order to fulfill the OEPs stakeholder's information needs in the 21st century:

The first challenge can be named as “measurement problem”. It is concerned about the question, how we can obtain missing data that cannot be found in the current OEP website interaction record. The second challenge can be named as “focus problem”. It is concerned about the question, which learner features are needed for which OEPs in order to enable personalized learning experiences.

Further exploration of these challenges would help to understand personalized learning for OEPs, open education and educational online services.

Bibliography

- [AB08] Arh, T.; Blazic, B. J.: A case study of usability testing—the SUMI evaluation approach of the EducaNext portal. In *WSEAS transactions on information science and applications*, 2008, 5; pp. 175–181.
- [BBC07] Burby, J.; Brown, A.; Committee, W. S.A.: *Web analytics definitions*. In Washington DC: Web Analytics Association, 2007.
- [BM07] Brusilovsky, P.; Millán, E.: *User models for adaptive hypermedia and adaptive educational systems: The adaptive web*. Springer, 2007; pp. 3–53.
- [BR09] Brees, I.; Rittberger, M.: *Web 2.0 learning environment. Concept, implementation, evaluation*. PAU Education, 2009.
- [Ca09] Castellano, G. et al.: *Innovations in web personalization: Web Personalization in Intelligent Environments*. Springer, 2009; pp. 1–26.
- [CEP14] Camilleri, A. F.; Ehlers, U. D.; Pawlowski, J.: *State of the art review of quality issues related to open educational resources (OER)*. Luxembourg: Publications Office of the European Union, 2014.
- [Ch18] Chounta, I.-A.: *A review of the state-of-art of the use of Machine-Learning and Artificial Intelligence by educational portals and OER repositories (white paper)*, 2018.
- [CL07] Cuthrell, K.; Lyon, A.: *Instructional strategies. What do online students prefer*. In *Journal of Online Learning and Teaching*, 2007, 3; pp. 357–362.
- [CMS17] Chatti, M. A.; Muslim, A.; Schroeder, U.: *Toward an open learning analytics ecosystem: Big data and learning analytics in higher education*. Springer, 2017; pp. 195–219.
- [CP12] Chaffey, D.; Patron, M.: *From web analytics to digital marketing optimization. Increasing the commercial value of digital analytics*. In *Journal of Direct, Data and Digital Marketing Practice*, 2012, 14; pp. 30–45.
- [CV13] Chrysafiadi, K.; Virvou, M.: *Student modeling approaches. A literature review for the last decade*. In *Expert Systems with Applications*, 2013, 40; pp. 4715–4729.

- [Dr15] Drachsler, H. et al.: Panorama of recommender systems to support learning: Recommender systems handbook. Springer, 2015; pp. 421–451.
- [Ga07] Gauch, S. et al.: User profiles for personalized information access: The adaptive web. Springer, 2007; pp. 54–89.
- [Jo09] Jones, A.: Redisciplining generic attributes. The disciplinary context in focus. In *Studies in Higher Education*, 2009, 34; pp. 85–100.
- [Ke11] Kent, M. L. et al.: Learning web analytics. A tool for strategic communication. In *Public Relations Review*, 2011, 37; pp. 536–543.
- [MDK11] Magnisalis, I.; Demetriadis, S.; Karakostas, A.: Adaptive and intelligent systems for collaborative learning support. A review of the field. In *IEEE Transactions on Learning Technologies*, 2011, 4; pp. 5–20.
- [ND08] Nguyen, L.; Do, P.: Learner model in adaptive learning. In *World Academy of Science, Engineering and Technology*, 2008, 45; pp. 395–400.
- [RL08] Razmerita, L.; Lytras, M.D. Eds.: *Ontology-based user modelling personalization. Analyzing the requirements of a semantic learning portal.* Springer, 2008.
- [Sp00] Specht, M. Ed.: *ACE-adaptive courseware environment.* Springer, 2000.
- [Ve12] Verbert, K. et al.: Context-aware recommender systems for learning. A survey and future challenges. In *IEEE Transactions on Learning Technologies*, 2012, 5; pp. 318–335.
- [Wi07] Willems, J. Ed.: *Does style matter? Considering the impact of learning styles in e-learning.* Australasian Society for Computers in Learning in Tertiary Education, 2007.
- [ZP15] Zheng, G.; Peltsverger, S.: *Web analytics overview: Encyclopedia of Information Science and Technology, Third Edition.* IGI Global, 2015; pp. 7674–7683.