Mining Web Analytics Data for Information Wikis to Evaluate Informal Learning

Heba M. Ismail [✉], Boumediene Balkhouche, Saad Harous
UAE University, Abu Dhabi, UAE
`Hebaismail20@gmail.com`

Abstract—Information wikis and especially Wikipedia have become one of the most attractive environments for informal learning. The nature of wikis enables learners to freely navigate the learning environment and independently construct knowledge without being required to follow a predefined learning path in line with the constructivist learning theory. Link-based navigation and keyword-based search methods used on Wikipedia and similar information wikis suffer from many limitations. In our paper, we present an effective recommendation system that provides easier and faster access to relevant content on Wikipedia to support informal learning. In addition, we evaluate the impact of personalized content recommendations on informal learning from Wikipedia and show how web analytics data can be used to get an insight on informal learning in similar environments.

1 Introduction

Personalization has proved to achieve better learning outcomes by adapting to specific learners' needs, interests, and/or preferences [1]. Traditionally, most personalized learning software systems focused on formal learning [2], [3], [4]. Formal learning software systems attempt to model formal education normally delivered at schools or colleges by defining specific learning content aligned with a curriculum, learning outcomes, and assessments. However, learning personalization is not only desirable for formal learning, it is also required for informal learning which is self-directed, does not follow a specified curriculum, and does not lead to formal qualifications [5].

Studies of informal learning reveal that up to 90% of adults are engaged in hundreds of hours of informal learning [6]. It has also been estimated that up to 70% of learning in the workplace is informal [7]. Several recent studies investigated how online text issuing platforms such as wikis and blogs can contribute to informal learning. Wikis among other informal learning platforms are found to attract an increasing attention for informal learning, especially Wikipedia [8], [9], [10], [11]. A study that targeted high school students at six campuses in the U.S. between April and May 2009, showed that up to 82% of students in higher education turn to Wikipedia to give their research a jump start, and 76% of students use Wikipedia to find the meaning of terms in certain

topics [12]. As of September 2015, Wikipedia reported around 374 million unique visitors per month, and 71,000 active contributors working on more than 47,000,000 articles in 299 languages [13]. This makes Wikipedia one of the greatest sources of knowledge on the web.

To support informal learning on Wikipedia and similar environments, it is important to provide easy and fast access to relevant content. However, navigation on information wikis suffer from several limitations. One method to navigate articles is using keyword-based search. However, in many cases, users may fail to identify representative keywords. Another method to navigate articles is following hyperlinks. This method is powerful but may divert the user away from the main topic of interest. In addition, links mentioned in an article cannot fully cover all related articles in the whole corpus. One of the reasons is because there is no term describing related articles within the current article, or simply because some links might not be working.

Recommendation systems (RSs) have long been used to effectively provide useful recommendations in different technology enhanced learning (TEL) contexts [14], [15]. Nevertheless, to our best knowledge, no effective recommendation system has yet been designed to support informal learning on Wikipedia and similar information wikis.

On the other hand, the evaluation of recommender systems in general is a complicated task, because of:

- The diversity of different measures that need to be considered, e.g. accuracy, novelty, scalability, serendipity [16]
- The availability/unavailability and adequacy/inadequacy of benchmark datasets
- The number of users that such evaluations may require.

In addition to these factors, evaluation of TEL recommender systems for informal learning is rather a challenging activity due to the inherent difficulty in measuring the impact of recommendations on informal learning with the absence of formal assessment and commonly used learning analytics.

To this end, in this paper, we present a personalized content recommendation framework for information wikis in addition to an evaluation framework that can be used to evaluate the impact of personalized content recommendations on informal learning from wikis. Our framework and preliminary evaluation were presented in EDUCON19 [17]. However, this paper mainly extends our evaluation approach and presents a primary design of an evaluation framework based on web data analytics.

The introduced recommendation framework models learners' interests by continuously extrapolating topical navigation graphs from learners' free navigation and applying graph structural analysis algorithms to extract interesting topics for individual users. Then, it integrates learners' interest models with fuzzy thesauri for personalized content recommendations. Our evaluation approach encompasses two main activities. First, we evaluate the impact of personalized recommendations on informal learning by assessing conceptual knowledge in users' feedback. Second, we analyze web analytics data to get an insight into users' progress and focus throughout the test session.

We introduce an overview of related literature in Section 2. In Section 3, we explain the major components of the recommendation framework. In Section 4, we describe

our evaluation process, the experimental set up, and the results of our experiment. Conclusion is presented in section 5.

2 Background and Literature Review

2.1 Recommender systems for technology enhanced learning

Recommender systems (RSs), sometimes called recommender engines, typically make recommendations through two main approaches:

- *Collaborative filtering (CF)*, sometimes called association-based filtering.
- *Content-based filtering (CB)*, also called personality-based approach.

CF approaches primarily generate a user model based on a user's past interaction such as products previously purchased, books read/downloaded, articles navigated, courses completed/viewed, and/or ratings/likes/dislikes/reviews of those items as well as similar decisions made by other users. This model is then used to predict items, or ratings for items, that the user may have an interest in. On the other hand, CB approaches utilize attributes or features of an item to recommend additional items with similar properties. These approaches are often combined in hybrid recommenders [18].

Many technology-enhanced learning (TEL) systems utilize different types of recommender engines to support learning [14]. As classified by Drachsler et al. [15], TEL recommender systems reported in the literature support various tasks such as finding good learning content [19], [20], suggesting the most effective paths through a plethora of learning resources to achieve a certain competence [21], [22], or suggesting peers learners, which is very central recommendation task for distance education settings where learners usually feel isolated and sometimes demotivated [23].

Even though, the reported research studies in TEL RSs show interesting results especially in online learning environments with focused learning objectives and well-defined learning content and learners' base, there remain some challenges inherent in delivering recommendations for massively diverse unstructured content with massive user base as seen in Wikipedia. CF approaches have long been singled out for being less effective in recommending content to new users with no or minimum interaction data, a case that is called the cold start problem. In addition, CF approaches are less effective when items are massively diverse, hence, fewer user groups will exhibit similar interaction history. Moreover, CB approaches are less effective with unstructured text such as Wikipedia content, especially that converting unstructured text into bag-of-words representation eliminates essential semantic relationships in the text.

Therefore, different variations of recommendation models have been used to address the challenges associated with designing recommendations for Wikipedia.

2.2 Wikipedia recommender systems

A few recommendation models have been proposed to provide article recommendations in Wikipedia. For example, Sriurai et al [24] used the Latent Dirichlet Allocation

(LDA) algorithm to generate topic-based recommendations. The proposed topic-based model is used to generate topic features used to classify articles against topics using LDA. The model was evaluated with an unspecified number of articles by 5 assessors. Each assessor was given a number of recommended articles and linked articles, i.e., linked through hyperlinks within articles, and asked to give a relevance score from 1 to 5. The average relevance score for recommended articles surpasses the relevance score by 1.2. The approach is neither designed to generate personalized recommendations, nor accounts for changing interests. Rather, fixed recommendations are presented to all readers following a pre-built topic distribution. In addition, those recommendations were not used in any learning task to evaluate their impact on readers learning.

On the other hand, Adline & Mahalakshmi [25] proposed a more sophisticated article quality framework to classify and recommend Wikipedia articles into usage categories. Users are assumed to be looking for some article's quality features according to their usage purpose. Thus, article's quality measures such as characters count, sections length, organization, readability, and structure are used to categorize and recommend articles under three usage categories:

- Readable
- Learnable
- Referable

For evaluation, 50 users were asked to categorize 150 articles into best, average or worst for each usage category. Users' ratings were then compared to system's ratings using error measures and system ratings turned to be very accurate. In addition to the fact that the proposed framework treats all users equally and does not account for personalization, it treats Wikipedia as a definitive and comprehensive source of knowledge and categorize the articles based on usage purposes accordingly. However, Wikipedia articles are not meant to be an ultimate source of knowledge, rather they serve the purpose of giving a quick introduction to a topic, a jump start to a new research topic, and a list of some good references. Moreover, evaluating the quality of articles is a valuable contribution, but, assuming that users of different types would like to have articles of certain quality measures is a very strong assumption that is not verified in the paper, or in other cited papers.

In addition to the new variations of content-based recommendations, researchers started to utilize new variations of search algorithms to deliver structural recommendations [26]. In structural recommendation techniques, content or/and users are represented using graphs. Graph search and ranking algorithms are then used to recommend nodes, links, or different combinations of both. A recent research study by Schwarzer et al [27] proposed a structural recommendation framework for Wikipedia articles based on a modified form of Co-Citation Proximity Analysis (CPA) utilizing page links rather than citations. The proposed recommendation framework is not personalized to individual users. Moreover, the accuracy of the proposed framework was evaluated using Wikipedia's "See also" sections which account for 17% of the corpus only, and a Wikipedia clickstream dataset which are not fully user generated. Even though, results show high performance of the proposed framework, it lacks reliability. Furthermore, the study did not evaluate the impact of recommendations on learning.

Therefore, since we are addressing personalized informal learning, there is a need to model a personalized content recommendation framework for Wikipedia as well as evaluate the impact of recommendations on informal learning.

3 Personalized Content Recommendation Framework

The proposed approach first captures raw learning interests for every individual learner in a topical navigation graph, TNG, by tracking individual learning sessions. We model the learner navigation as a directed graph, TNG (V, E). Every vertex, V, in TNG corresponds to a topic or a wiki page, and every edge, E, in TNG corresponds to a navigational action. Then, structural topical graph analysis algorithms, adapted from Leak et al. [28], are used to rank the raw topics captured in the navigation graph in the previous step. Topics that receive high ranking in the structural analysis are considered the user interest model, UIM. UIMs are then associated with semantically relevant topics found in inverted indices of topics, IIT, generated based on concepts from fuzzy set information retrieval model [29], to deliver personalized content recommendations.

Our framework is composed of four main modules:

- Session tracking
- TNG analyzer
- Personalization
- Semantic analysis modules

Figure 1 illustrates our conceptualization of the proposed framework. The semantic analysis module is designed (to be used offline) to build and process custom corpora and generate inverted indices of topics used online by the personalization module to generate personalized content recommendations based on the learner models generated by the TNG Analyzer module. We briefly describe each module in the following sections.
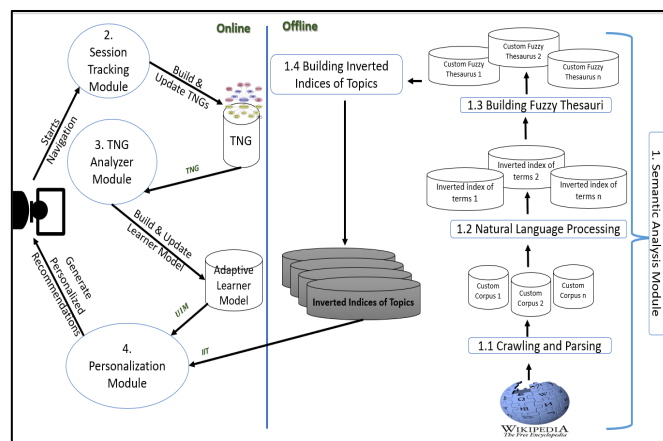

Fig. 1. Personalized content recommendation framework

3.1 Semantic analysis module

The objective of this module is to generate inverted indices of topics, IIT. These indices associate each topic with a set of semantically relevant learning documents, web pages. First, custom corpora are extracted from Wikipedia for each main topic category[1] using a web crawler. From this step we get a custom corpus for each main topic category such as science, art, culture, etc. Second, natural language processing tasks such as stemming, tokenization, and stop words removal of the custom corpora is performed to generate inverted indices of unique terms. An inverted term index indicates, for each unique word in the corpus, the documents in which it appears, and its positions or occurrences in that document. Third, the inverted indices of terms are used to generate custom fuzzy thesauri that define the semantic similarity, Cf, between every two distinct terms in each custom corpus as explained in [30], [31]. Finally, custom fuzzy thesauri are used to calculate the semantic similarity between distinct topics and all documents, web pages, in each custom corpus. At this stage topics are extracted at the page level. To calculate the semantic similarity between a topic and a web page, every term, Ti, in every topic, Topic, is compared with every word, wj, in a document, d, to retrieve the corresponding semantic similarity factor, Cfij, from the corresponding custom fuzzy thesaurus which indicates the word-word semantic similarity. Once a term, Ti, is compared to each word, wj, in each document, d, the semantic similarity between the term and the whole document is calculated as follows:

$$\mu_{T,d} = 1 - \Pi\left(1 - Cf_{ij}\right) \tag{1}$$

which indicates the Term-Document semantic similarity. The average of all μ-values of the terms, Ti, composing a given topic, Topic, and a given document, d, is calculated to yield the overall similarity between the topic and the document as follows:

$$Sim\left(\text{Topic}, \text{d}\right) = (\mu\_(T_1, d) + \mu\_(T_2, d) + \cdots + \mu\_(T_i, d))/i \tag{2}$$

Table 1 shows a snapshot from an inverted topic index. The greater the Topic_Document_Similarity the more semantically similar the document, wiki page, to the topic of interest. Analysis of semantic similarity is done at the whole document level covering all sections and sub sections of corresponding wiki pages.

Table 1. Sample entry in the inverted topic index

Topic	Document	Topic_Document_Similarity
Amazon River	Celtic mythology	0.087005
	Chew Valley	0.055887
	Chew Valley Lake	0.091634
	Colorado River	0.333333

---

[1] https://en.wikipedia.org/wiki/Category:Main_topic_classifications

3.2 Session tracking module

The session tracking module first captures raw learning interests for every individual learner in a topical navigation graph, TNG, by tracking individual learning sessions. A learning session starts when the learner first accesses the wiki and ends when the learner leaves the wiki domain. We model the learner navigation as a directed graph, TNG (V, E). Every vertex, V, in TNG corresponds to a learning topic in the wiki environment. A learning topic corresponds to the overall subject of the article. Pages that do not have learning content are filtered out and not captured in the graph. Every edge, E, in TNG corresponds to a navigation action performed by the user to access an article or to move from one article to another. Navigation actions occur through clicking on hyperlinks within the page, browsing back and forward, or clicking on topics' indices provided in the wiki. The process of capturing navigation into TNG is dynamic and continuous throughout the learning session. Figure 2 illustrates changes in TNG throughout a typical navigation session.

3.3 TNG analyzer module

We adapt The Hub-Authority and Root-Distance Model (HARD), and The Connectivity Root-Distance Model (CRD) concept maps' topological analysis models from Leak et. al. [32], [28], to calculate topics' structural weights relevant to individual learners' navigation behavior. The analysis of the structural weights goes through two steps.

- First, the structural characteristics of each topical node in TNG need to be defined as per the selected model. For the CRD model, each topical node, V, needs to be characterized for its connectivity, outgoing connections, $o(v)$, and incoming connections, $i(v)$, and direct steps from the first topical node, $d(v)$. For the HARD model, each topical node, v, needs to be characterized as being a hub, $h(v)$, with mostly outgoing connections, authority, $a(v)$, with mostly incoming connections, or upper node, $u(v)$, that is closer to the starting node in TNG.
- Second, using the structural characteristics, the relative node's weight $W(v)$ is calculated as follows:

For the CRD Model:

$$W(v) = \left(\alpha \cdot o(v) + \beta \cdot i(v) \right) \cdot \left(\frac{1}{d(v)+1} \right)^{\frac{1}{\delta}} \tag{3}$$

And for the HARD model:

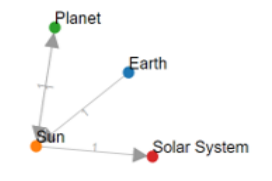$$W(v) = \alpha \cdot h(v) + \beta \cdot a(v) + \gamma \cdot u(v) \tag{4}$$

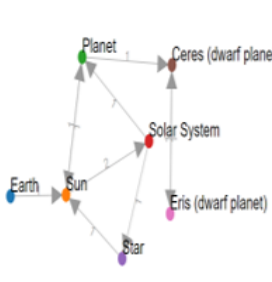| User TNG	Graph Structure
	V = { Earth, Sun} **E** = { (Earth, Sun)}
	V = { Earth, Sun, Planet} **E** = { (Earth, Sun), (Sun, Planet)}
	V = { Earth, Sun, Planet, Solar System} **E** = { (Earth, Sun), (Sun, Planet), (Planet, Sun), (Sun, Solar System)}
	V = { Earth, Sun, Planet, Solar System, Star, Ceres (Dwarf Planet), Eris (Dwarf Planet)} **E** = { (Earth, Sun), (Sun, Planet), (Planet, Sun), (Sun, Solar System), (Solar System, Star), (Star, Sun), (Sun, Solar System), (Solar System, Planets), (Planets, Ceres (Dwarf Planet)), (Ceres (Dwarf Planet), Eris (Dwarf Planet), (Eris (Dwarf Planet)) }

Fig. 2. Illustration of session tracking and TNG building processes

3.4 Personalization module

Based on the structural weights calculated earlier, a weighted or ranked topical navigation graph can be used to extract the most interesting topics that form a learner interest model, UIM. UMI is then used to associate semantically similar articles from topics' inverted indices, IIT, to generate personalized content recommendations such that:

Personalized Content Recommendations for user i =

$$W(umi_i) * Sim(\{Topic \in \text{UIM}_i \}, \{ d_n : d_n \in IIT \}) \tag{5}$$

Therefore, learning documents with higher semantic similarities to topics that have higher structural weights in the learner model are retrieved and recommended to the learner. Adaptation is accomplished through continuous update of TNG and, accordingly, the structural weights as well as the personalized recommendations. Figure 3 illustrates how structural weights adapt to changes in user's TNG, presented earlier in

Figure 2, and accordingly changes the topics of interests in the learner model as well as the personalized list of recommended articles.

HARD Topological Weights	Recommended Topics
W (Earth) = 0.83 W (Sun) = 1.00	1. Earth's Atmosphere 2. Flat Earth
W (Earth)= 0.53 W (Sun)= 0.75 W (Planet) = 1	1. Ceres (Dwarf Planet) 2. Definition of Planet 3. Earth's Atmosphere
W (Earth)= 0.38 W (Sun) = 0.66 W (Planet) = 0.92 W (Solar System) = 1	1. Solar Eclipse. 2. Ceres (Dwarf Planet). 3. Timeline of Discovery of Solar System planets. 4. Earth's Atmosphere.
W (Earth)= 0.25 W (Sun)= 0.58 W (Planet) = 0.69 W (Solar System) = 0.80 W (Star)= 0.87 W (Ceres (dwarf planet)) = 0.94 W (Eris (dwarf planet)) = 1	1. Ceres (Dwarf Planet). 2. Eris (Dwarf Planet). 3. Binary Star 4. Solar Eclipse 5. Ceres

Fig. 3. Illustration of how TNG structural weights change the learner interest model and accordingly the personalized list of recommended articles.

4 Evaluation

The proposed approach is aimed at achieving effective and adaptive personalization of unstructured learning content in the form of personalized recommendations to support informal learning in wikis. Consequently, our evaluation encompasses two main objectives:

- Evaluating the effectiveness of personalized content recommendations.
- Evaluating the impact of personalized recommendations on informal learning.

Traditionally, the quality of a recommender system is defined in terms of objective statistical metrics calculated by comparing system's behavior against some historical data [33] which is commonly referred to as offline evaluation. However, it is believed that evaluations of systems involving user models cannot and should not be separated from actual users [34]. As a result, recommendation systems research is exploring user-centric directions for measuring and improving the subjective quality of RSs from the point of view of the user [35]. A major advantage of user studies is that they allow for collecting information about user interaction and as well as testing different scenarios. Since we need to evaluate the impact of recommendations on informal learning, we carried out a user-centric evaluation. Therefore, we designed user studies to evaluate the effectiveness of the proposed approach.

4.1   Evaluation metrics

In our evaluation we use two types of metrics: user-centric quality metrics to evaluate the effectiveness of the personalized recommendations; and objective educational metrics and web analytics data to evaluate the impact of recommendations on learning.

For the user-centric metrics, we evaluate two user-centric quality metrics that have been commonly used in the literature [36]:

- *Perceived accuracy or relevance*: How much the recommendations match the users' interests, preferences, and tastes.
- *Overall users' satisfaction*: The global users' feeling of the experience with the RS.

For educational metrics, we focus on knowledge assessment as we are evaluating informal learning and we are not following a curriculum or predefined learning objectives upon which we can evaluate learners. Knowledge assessment allows measuring the outcomes of learning and determines the effectiveness of the learning process. As knowledge structure cannot be observed directly, various indirect methods are used instead. Concept maps (CM) are one of such methods [37]. Therefore, to evaluate informal learning, we design a conceptual knowledge assessment rubric adapted from concept map-based rubrics[2]. Our conceptual knowledge rubric was presented in EDUCON19 [17]. The proposed rubric is a simplified rubric aimed at assessing conceptual knowledge in essays for primary students. Essays can be assessed against five criteria: structure, relationships, exploratory, communication, and writing quality. Essays can be assessed on a scale of 1 to 4 against each criterion based on some characteristics that are explained in the rubric.

4.2   Technological framework

To run our user studies, we developed three web-based encyclopedias equipped with user navigation tracking and analysis algorithms as well as the proposed personalized content recommendation engine. Our online test encyclopedias are listed in Table 2.

---

[2] https://teach.its.uiowa.edu/sites/teach.its.uiowa.edu/files/docs/docs/Concept_Map_Rubrics_ed.pdf

The three websites are XHTML based. The tracking and analysis scripts are developed using PHP 5.5 and JavaScript ES5. All user navigation data is kept in MySql 5.6.32.

Table 2. Test websites

Website URL	Type
www.theknowledge.site	No personalized support
www.hardtest.site	Personalized content recommendations using HARD model
www.crdtest.site	Personalized content recommendations using CRD model

4.3 Learning content

We use content from the 2007 Wikipedia DVD Selection[3] which is a free, hand-checked, and non-commercial selection from Wikipedia, targeted around the UK National Curriculum. It is about the size of a fifteen-volume encyclopedia including all topics in Wikipedia rated "Good" or higher by Wikipedia itself at date of production.

4.4 Data collection techniques

We use questionnaires to collect users' feedback about some aspects of the system during the experiments. Questionnaires collect both users' demographic attributes and their opinions about perceived accuracy and overall satisfaction. In addition, we asked the participants to submit essays related to their topics of interest. Moreover, we run tracking scripts to collect navigation-related data.

4.5 Participants

Experiments were carried out at a local private school teaching the UK National Curriculum. All year-5 students were invited to participate in the experiments. Consent forms were sent to interested students' parents to allow their children to participate in the experiments. A total of eighty students from year-5 participated in the experiments.

4.6 Procedure

A writing challenge was announced among year-5 students. In the announcement we invited the students to use an online encyclopedia during their break hours at the school to learn about any topic related to the space and then submit an essay about their topic of interest. The question in the announcement states the following: "If you could go to space at some point in your life, what would you most like to see or experience? Choose anything in the universe and write about it." The experiments were carried out during term three of the school year by then the participants had covered enough material related to space as part of their science subject. We needed to confirm this information from teachers to ensure participants' familiarity with the topic of the experiments as

---

[3] https://en.wikipedia.org/wiki/Wikipedia:Wikipedia_CD_Selection

well as to ensure that participants are capable of learning and writing about the space. Hence, we control the factors of previous experiences and minimum required skill levels that commonly impact any learning process. Furthermore, we forced a fixed design for all the test sessions in terms of time, location, class setup, and duration to eliminate the impact of these factors on the experimental results. For example, some students might be very tired at the end of school day compared to their agility level in the early morning and thus may be less capable to learn. Moreover, some classrooms might have more comfortable setups, lighting, or conditioning system which may have impact on their attention or engagement in the experiment. So, we carried all the test experiments in the same computer lab. The variable factors were limited to website setups in terms of recommendations' logic as explained earlier.

Forty students used the online encyclopedia with personalized recommendations, and forty students used the website without any recommendations. Each group has all levels of students. Students could use the website in informal settings during break time for one hour during which they could read about any topic related to space, take notes, save some pictures, and ask questions to the study moderator whenever they needed help. At the end of the session, students were asked to complete a questionnaire to rate their experience on a scale of "1" to "4", where "1", e.g. "not useful" or "not relevant", represents the worst impression, and "4", e.g. "very useful" or "very relevant", represents the best impression. We used expressive responses rather than points as we found it to be more suitable for the selected age group. Afterwards, the students could use the information they collected from the encyclopedia to write an essay and email it to the study moderator. All students completed the questionnaires and rated their experience, but, only 32 students out of the 80 participants submitted written essays. Nevertheless, we selected only 22 essays (11 from the personalized support group and 11 from the control group) for the assessment of informal learning and excluded 10 submissions that were entirely copied from the online encyclopedia. Prizes were given to the best three essays.

4.7   Results and discussion

User-centric quality metrics: As highlighted in previous sections, link-based navigation suffers from many limitations. To verify those findings, we asked the students whether it was easy for them to find the information they were looking for by just using the navigational tools supported in the online encyclopedias such as subject index and hyperlinks. We found that 43.59% of the students in the control group took long time to find the information compared to 29.73% of the students in the group with personalized support as shown in Figure 4 (A). Interestingly, the percentage of students who faced difficulty in navigation on the encyclopedias with personalized support is relatively smaller than the percentage of students who faced difficulty in navigation on the encyclopedias without personalized support (control groups).

Moreover, results show that the proposed personalized content recommendation framework generates highly relevant recommendations as shown in Figure 4 (C). In addition, considering the overall user satisfaction criteria, results show that more than 90% of the 40 users who used the encyclopedia with personalized recommendations

found the recommendations to be useful, and more than 80% thought that it would be helpful to have similar recommendations on other websites that they commonly used for information search as shown in Figure 4 (B) and (D) respectively.
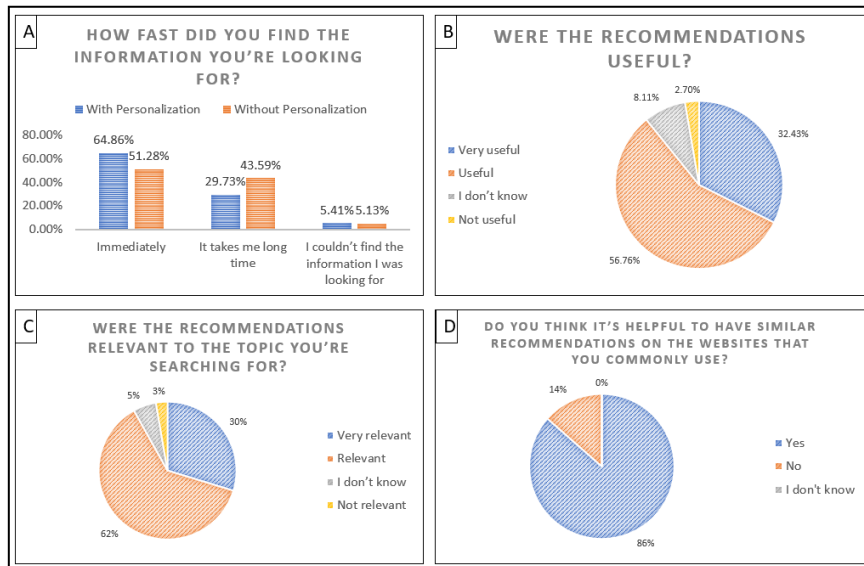

Fig. 4. Results of user experience questionnaires

Conceptual knowledge assessment: Two assessors evaluated the students' essays using the conceptual knowledge rubric explained earlier. Evaluation of conceptual knowledge reveals that users, who used the online encyclopedia with personalized recommendations, achieved higher scores on conceptual knowledge assessment compared to those who used Wikipedia without recommendations. The average score for students who used the encyclopedia with personalized recommendations was 14.9 compared to 10.0 for the students who used the encyclopedia without recommendations as shown in Table 3. The results are statistically significant at alpha level 5%, $\alpha = 0.05$, using t-Test for small independent samples with P-Value = 0.0 < 0.05.

Moreover, the assessors found that participants who used the encyclopedia with personalized recommendations were able to make use of a larger number of concepts, make comparisons, and state relations between concepts.

Table 3. Conceptual knowledge assessment results

With Personalization		Without Personalization	
Topic	*Result*	*Topic*	*Result*
A Trip to Mars	18	Sun	12
Mars	14	Black Holes	12
Black Holes	16	Black Hole	11
Jupiter	14	Neptune	12
The Cat's Eye Nebula	12	Black Hole	11
Pluto	15	Mars	9
Milky Way	13	Black Hole	8
Lunar Eclipse	15	The Universe	12
Venus And Mercury and Earth	16	Lunar Eclipse	9
The Hubble Telescope	16	Neptune	8
Black Holes	16	Moon	8

Web analytics-based evaluation: Web analytics is the measurement, collection, analysis and reporting of web data for purposes of understanding and optimizing web usage [38]. With the inapplicability of formal assessment of learning in informal learning settings it is difficult to collect commonly used learning analytics for evaluation purposes. Therefore, we decide to examine the possibility of using web analytics data, which can be generated from any typical web navigation session, to induce some helpful insights about learners' performance. We propose an initial design of an evaluation framework based on web analytics data (Fig. 5) that can be used to evaluate informal learning in similar environments.
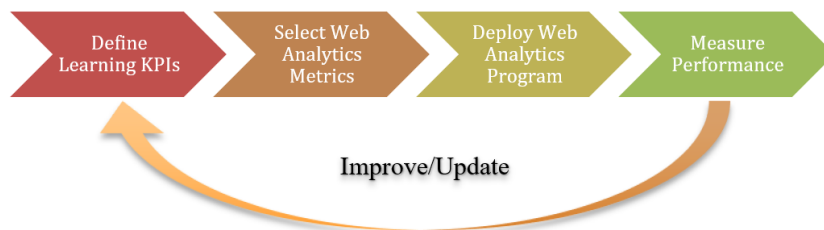

Fig. 5. Web analytics-based evaluation framework

In the following paragraphs we explain different activities involved in our web analytics-based evaluation.

Defining Key Performance Indicators (KPIs): KPIs are defined as "the critical (key) indicators of progress toward an intended result. KPIs provide a focus for strategic and operational improvement, create an analytical basis for decision making and help focus attention on what matters most [39]".

Considering the context of informal learning on information-oriented websites such as Wikipedia, users typically visit the website to learn about diverse topics of interest for various purposes. Additionally, users may have a new learning objective for every new visit to the website. Thus, our objective here is to maximize the value of each visit

by providing faster and easier access to relevant content. Therefore, the required KPIs in this context should help us measure and quantify whether users of the website succeed to gain adequate access to relevant content in every visit.

Accordingly, we consider initially the following three KPIs for each user every time he/she visits the website:

- ***The frequency of relevant topics visited by the user***: We quantify this KPI at the document level, i.e. we consider the main topic of each document/webpage which can be indicated by the page title in the context of information wikis.
- ***The frequency of relevant keywords in the visited pages***: We extract the main keywords from the collection of visited pages for each user. We use Term Frequency Inverse Document Frequency, TF-IDF, to measure the importance of individual keywords in the collection. At a high level, a TF-IDF weight finds the words that have the highest ratio of occurring in the current document vs the frequency of occurring in the larger set of documents. As a result, terms that have very high frequency in all the documents in a certain collection will end up having very low TF-IDF, hence, they do not represent important keywords. Whereas, terms that receive high frequency at the document level compared to low frequency at the collection level will have very high TF-IDF scores and as such are considered important keywords. Afterwards, keywords undergo semantic relevance test to select relevant keywords which can be used to quantify the frequency of relevant keywords.
- ***The frequency of relevant phrases in the visited pages***: We apply similar TF-IDF approach explained in KPI number two at the phrase level. We consider the phrase to be composed of two terms.

These KPIs quantify at the document, phrase, and keyword levels how much relevant content the user was able to access during his/her visit.

Selecting Web Analytics Metrics: Web analytics metrics aim at counting different events or things related to users' navigation on a website. For example, among the commonly used metrics are:

- *Hits*: Represent the total number of requests made to the server during a given time period, e.g. month, day, hour.
- *Files*: Represent the total number of hits (requests) that actually resulted in something being sent back to the user. That is, not all hits will send data, such as 404-Not Found requests and requests for pages that are already in the browsers cache. So, by looking at the difference between hits and files, we can get a rough indication of repeat visitors, as the greater the difference between the two, the more people are requesting pages they already have cached, i.e. have viewed already.
- *Pages (Views)*: Are those URLs that would be considered the actual page being requested, and not all the individual items that make it up such as graphics and audio clips. This metric is sometimes called impressions, and defaults to any URL that has an extension of .htm, .html or .cgi.
- *Visits*: Occur when some remote site makes a request for a page on a server for the first time. If the same site keeps making requests within a given timeout period, they will all be considered part of the same Visit. If the site makes a request to a server,

and the length of time since the last request is greater than the specified timeout period, common default is 30 minutes, a new Visit is started and counted, and the sequence repeats. Since only pages will trigger a visit, remote sites that link to graphic and other non-page URLs will not be counted in the visit totals, reducing the number of false visits.

- ***Sites***: Is the number of unique IP addresses/hostnames that make requests to a server.
- ***Kbytes (KB)***: Is 1024 bytes (1 Kilobyte). It is used to show the amount of data that is transferred between the server and the remote machine, based on the data found in the server log.

In our evaluation, the metric that can help us calculate all the desired KPIs is the page view metric.

Choosing and Deploy Web Analytics Program: We evaluated three web analytics programs, namely, Webalizer4, AWStats5, and Google Analytics6. Google Analytics is a client-side analytics tool for which data is collected by a JavaScript code added to the website's HTML pages. Whereas, the first two are server-side. That is, they use the data contained in the server logs. We excluded Google Analytics since we are already running number of JavaScript on our test environments for tracking navigation graphs and for personalized recommendations. We choose AWStats as it gives full list of visited URLs that can be easily used for scrapping and further processing required to quantify the KPIs mentioned earlier.

Using the page metric, we identify for each user group viewed pages during the test session by applying time and date filters to AWStats setups. Then, we run a web scrapper application to extract viewed pages found in the AWStats' web analytics log files of both groups. During scrapping we allow repeated extraction of pages. We count repeated page views as they give an indication of the amount attention a user gives to a specific topic. Table 4 illustrates an example of AWStats page view analytics which we use in our evaluation.

Table 4. Snapshot from page view analytics using AWStats

136 different pages-url	Viewed	Average size	Entry	Exit
/wp/a/Acetic_acid.htm	115	73.73 KB	110	42
/wp/p/Prehistoric_man.htm	55	41.94 KB	48	22

Performance Evaluation based on Web Analytics Data: Analysis of web analytics data revealed that users, who used the encyclopedia with personalized support, navigated more articles related to their topics of interest compared to participants who used the encyclopedia without any personalized support. Users in the control group navigated a total of 226 articles compared to 644 articles navigated by the users in the personalized support group. These numbers include repeated views to the same articles. Manual analysis of the visited articles by both groups revealed that users in the control group were generally focused but visited less diverse topics related to "space" and some

---

[4] http://www.webalizer.org/
[5] https://awstats.sourceforge.io/
[6] https://analytics.google.com/analytics/web/#/

of them visited a few irrelevant topics such as "art" and "children charity". However, the other group of users visited more diverse pages related to "space". This might have resulted in helping the students who used the online encyclopedia with personalized support to use a larger number of related concepts and state relations among concepts. We can see as well in Table 3 that the students in the personalized support group submitted essays of more various topics compared to the control group students who submitted limited number of topics, mainly focused on "Black Hole" and "Neptune".

Moreover, by performing keyword extraction and phrase extraction on the collection of visited pages of both groups we are able to get further validation on the observations highlighted by the manual analysis. Table 5 shows statistics on viewed pages, frequency of extracted keywords, and frequency of extracted phrases.

Table 5. Statistics of visited pages extracted from users' web analytics logs

Visited Pages Analytics	Control Group	Personalized Support Group
Visited Pages	226	644
keywords Extracted	840,346	2,449,305
Phrases Extracted	447	1000

By considering the twenty highest frequency keywords and phrases of both groups, we can see that, for both groups, the top 50 keywords are mostly relevant to the topic of space. This gives a good indication that users were focused on the topic of space. However, the frequency of top keywords viewed by the personalized-support group significantly surpasses the frequency in control group as illustrated in Figure 6 and Figure 7. For example, "Earth" keyword's frequency is 9,441 in the personalized support group compared to 3,600 in the control group. This in turn, indicates that for the personalized support group more relevant articles related to "earth", which is an important topic in the space, were visited by the personalized support group. These results reinforce the manual analysis carried earlier.
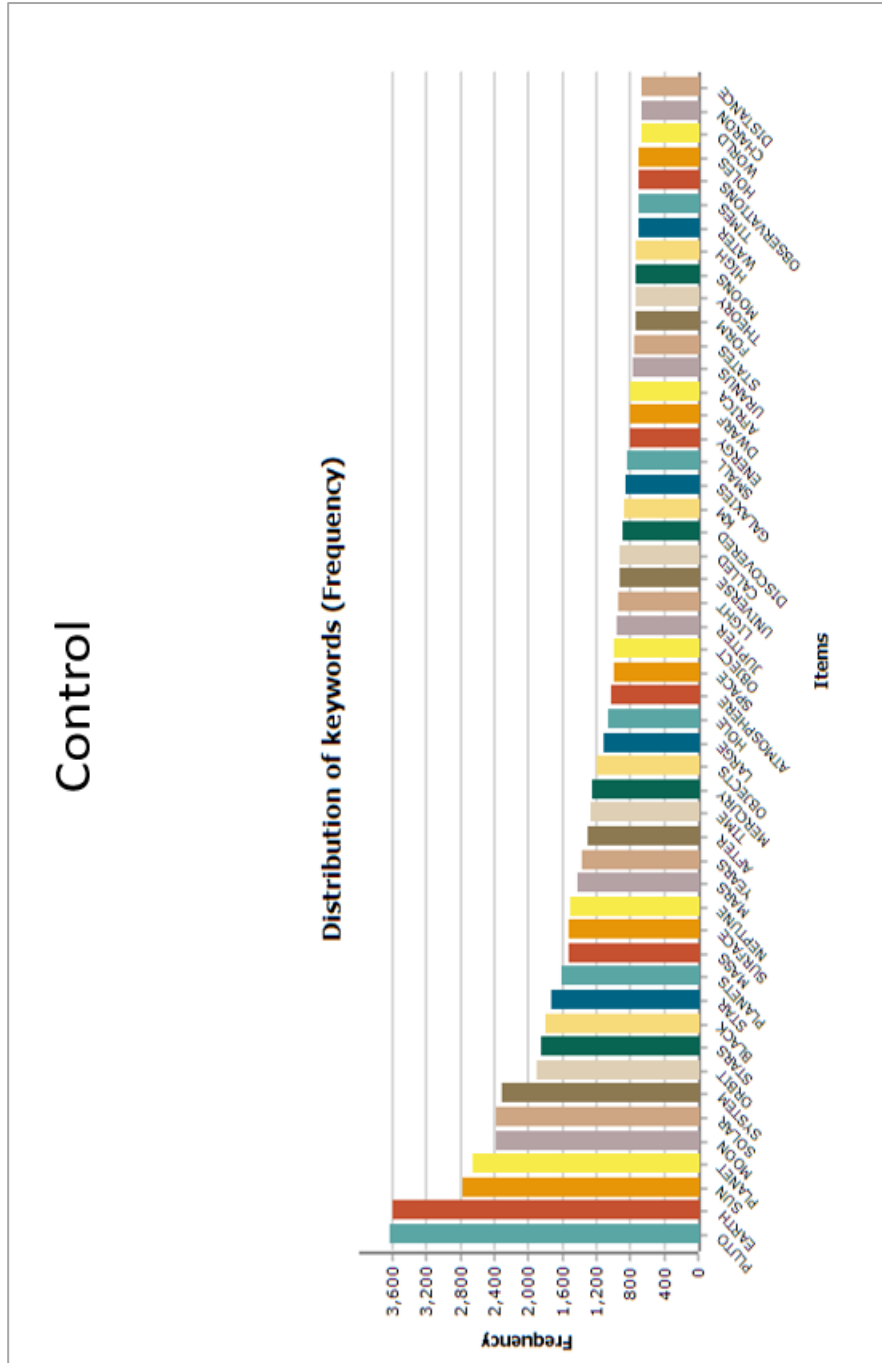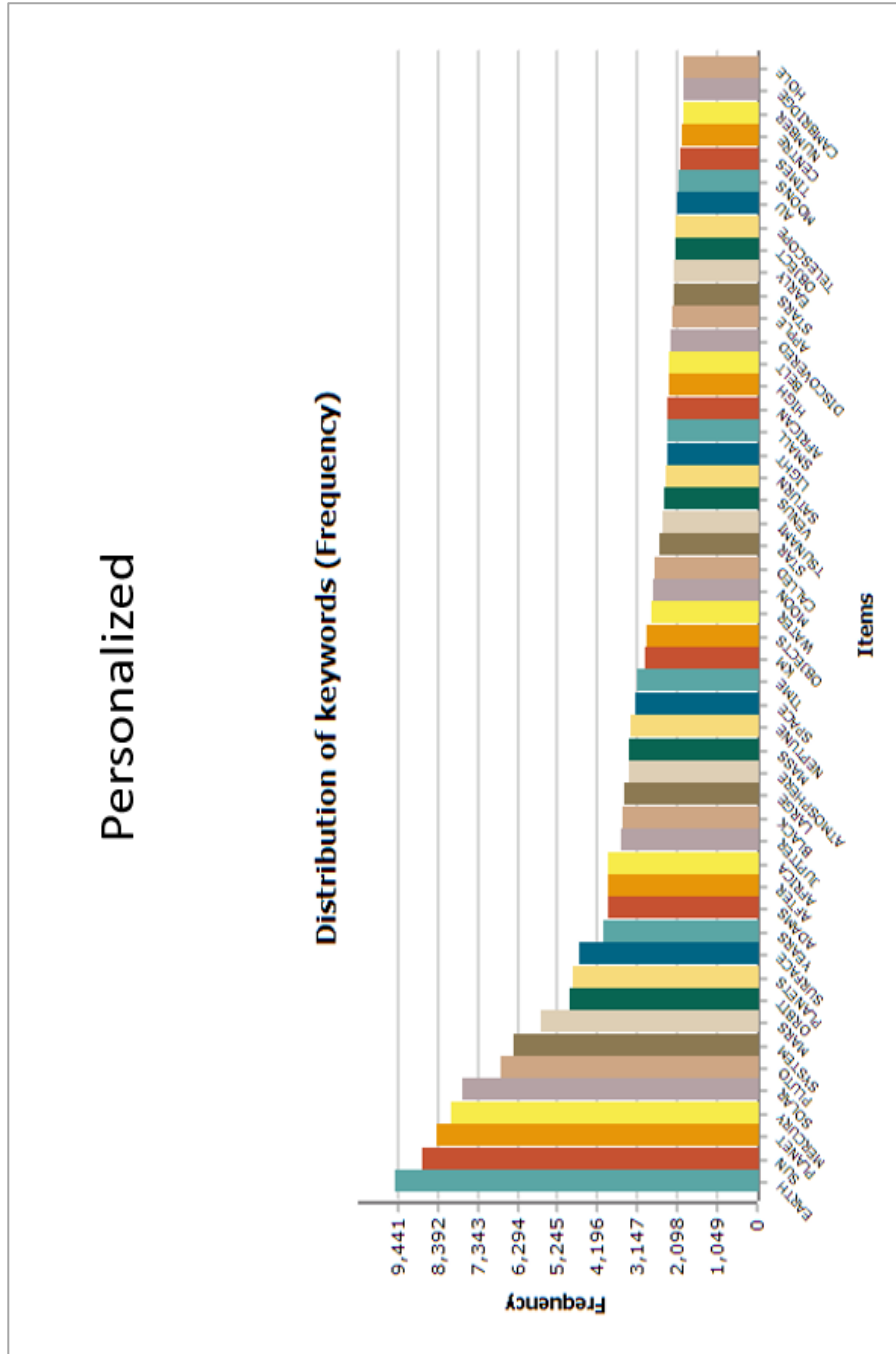
Fig. 6. Distribution of keywords for control group

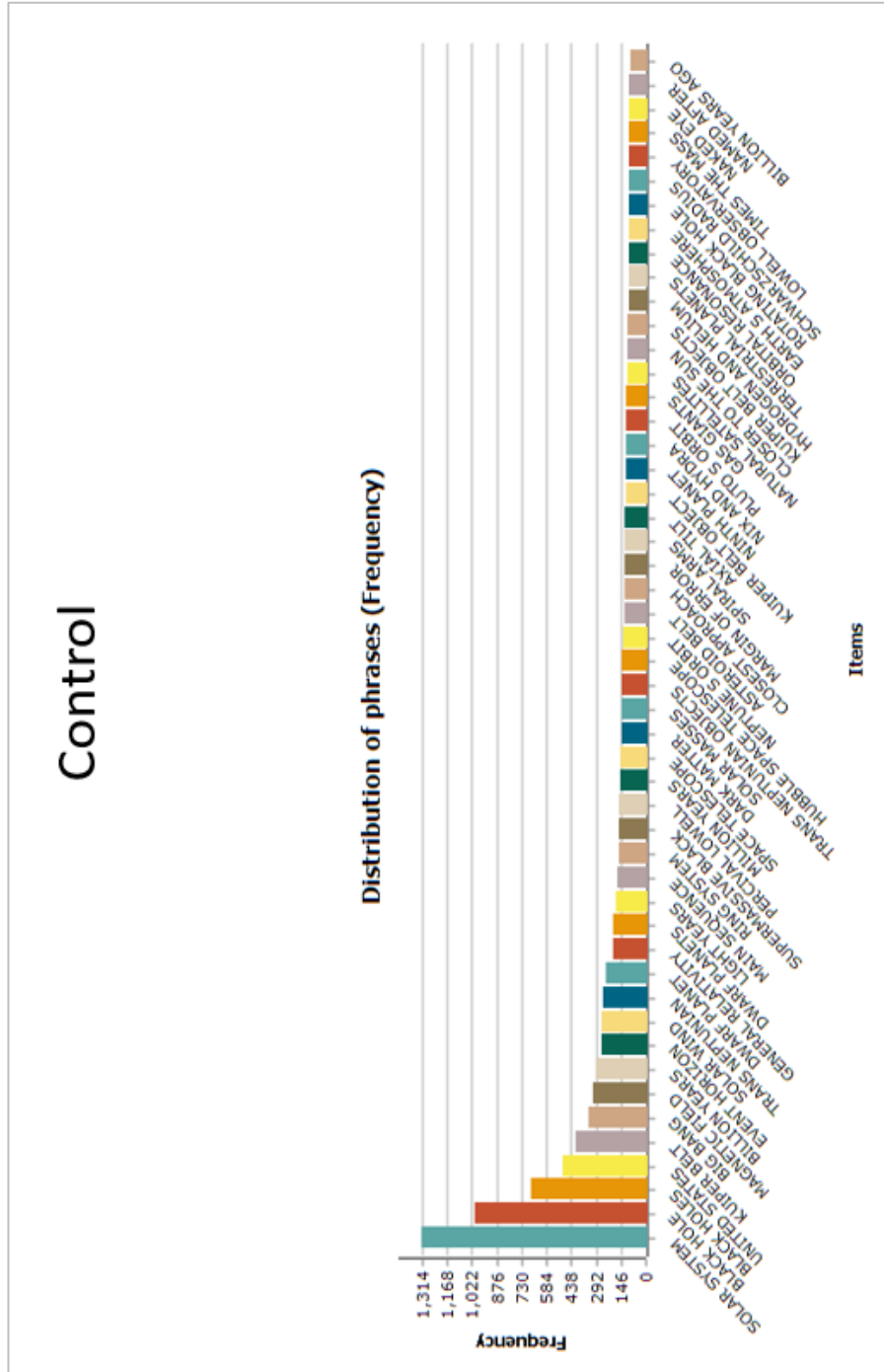**Fig. 7.** Distribution of keywords for personalized support group

Fig. 8. Distribution of phrases for control group

Fig. 9. Distribution of phrases for personalized support group

Furthermore, by analyzing the top 50 phrases extracted from the navigated pages' collection, we can see that almost all the top phrases are related to the topic of the space which gives a further validation to the previous observations as illustrated in Figure 8 and Figure 9. In addition, the frequencies of top phrases in the personalized support group surpasses by far the frequencies in the control group. For example, the frequency of "Solar System" is 1,314 in the control group compared to 4,176 in the personalized support group. These statistics validate further our earlier observations.

Finally, we conclude that personalized content recommendations effectively support informal learning from Wikipedia or other information website. That is because they provide easier and faster access to relevant information as well as help learners to be more focused on their topics of interest.

5 Conclusion

Information wikis and especially Wikipedia are attracting enormous attention for informal learning. Several limitations are associated with link-based navigation and keyword-based search. As a result, a framework that would support easy and fast navigation of relevant content is required to support informal learning from information wikis. Additionally, evaluation of informal learning in similar environment is a challenging task due to absence of formal assessments and learning analytics. In this paper, we present an effective personalized content recommendation framework as well as we propose an evaluation framework based on web analytics. We design user studies to asses informal learning from Wikipedia.

Our evaluation reveals that the personalized content recommendations enhances user experience on Wikipedia. Evaluation of informal learning show that users who used Wikipedia with personalized recommendations achieve higher scores on conceptual knowledge assessment compared to those who used Wikipedia without recommendations. Furthermore, they make use of larger number of concepts, make comparisons, and state relations between concepts. Web analytics-based evaluation show that those who used Wikipedia with personalized recommendations can make use of a larger number of relevant keywords and phrases.

In the future, we intend to model a comprehensive evaluation framework based on web analytics that can be used to give users insight into their progress as well as help improving content recommendations.

6 References

[1] Bray, B., & McClaskey, K, "A STEP-BY-STEP Guide to Personalize Learning," 2013.

[2] Campbell, L. O., Heller, S., & DeMara, R. F., "Implementing Student-Created Video in Engineering: An Active Learning Approach for Exam Preparedness." International Journal of Engineering Pedagogy, vol. 9, no. 4, 2019. https://doi.org/10.3991/ijep.v9i4.10363

[3] Albert, D., & Hockemeyer, C., "Applying demand analysis of a set of test problems for developing adaptive courses." in Computers in Education, 2002. https://doi.org/10.1109/cie.2002.1185866

[4] Koedinger, K. R., & Corbett, A., "Cognitive tutors." in The Cambridge handbook of the learning sciences, 2006, pp. 61-77.

[5] Misko, J., Combining Formal, Non-Formal and Informal Learning for Workforce Skill Development, Adelaide, Australia: National Centre for Vocational Education Research Ltd, 2008.

[6] Merriam, S. B., Caffarella, R. S., & Baumgartner, L. M., Learning in adulthood: A comprehensive guide, John Wiley & Sons, 2012.

[7] Streumer, J. N., & Kho, M., "The world of work-related learning," Work-related learning, pp. 3-49, 2006. https://doi.org/10.1007/1-4020-3939-5_1

[8] Goh, W. W., "Can Wiki be used to facilitate critical thinking? A qualitative approach." International Journal of Engineering Pedagogy (iJEP), vol. 2, no. 4, pp. 18-23, 2012. https://doi.org/10.3991/ijep.v2i4.2261

[9] Zheng, B., Niiya, M., & Warschauer, M., "Wikis and collaborative learning in higher education.," Technology, Pedagogy and Education, vol. 24, no. 3, pp. 357-374, 2015. https://doi.org/10.1080/1475939x.2014.948041

[10] Neumann, D. L., & Hood, M., "The effects of using a wiki on student engagement and learning of report writing skills in a university statistics course.," Australasian Journal of Educational Technology, vol. 25, no. 3, 2009. https://doi.org/10.14742/ajet.1141

[11] Kimmerle, J., Moskaliuk, J., & Cress, U., "Using wikis for learning and knowledge building: Results of an experimental study.," Journal of Educational Technology & Society, vol. 14, no. 4, p. 138, 2011.

[12] Head, A. J., & Eisenberg, M. B., "How today's college students use Wikipedia for course-related research," Peer Reviewed Journal of The Internet, vol. 15, no. 3, 2010.

[13] Wikipedia, "About Wikipedia," Wikimedia, 2015. [Online]. Available: https://en.wikipedia.org/wiki/Wikipedia:About. [Accessed 15 September 2018]. https://doi.org/10.1515/9783110376357-002

[14] Manouselis, N., Drachsler, H., Vuorikari, R., Hummel, H., & Koper, R., "Recommender systems in technology enhanced learning," Recommender systems handbook, pp. 387-415, 2011. https://doi.org/10.1007/978-0-387-85820-3_12

[15] Drachsler, H., Verbert, K., Santos, O. C., & Manouselis, N., "Panorama of recommender systems to support learning," in Recommender systems handbook, Boston, Springer, 2015, pp. 421-451. https://doi.org/10.1007/978-1-4899-7637-6_12

[16] Silveira, T., Zhang, M., Lin, X., Liu, Y., & Ma, S., "How good your recommender system is? A survey on evaluations in recommendation," International Journal of Machine Learning and Cybernetics, vol. 10, no. 5, pp. 813-831, 2019. https://doi.org/10.1007/s13042-017-0762-9

[17] Ismail, H., & Belkhouche, B., "Evaluating the Impact of Personalized Content Recommendations on Informal Learning from Wikipedia." in IEEE Global Engineering Education Conference (EDUCON), 2019. https://doi.org/10.1109/educon.2019.8725052

[18] Ricci, F., Rokach, L., & Shapira, B., Introduction to recommender systems handbook, Springer US, 2011.

[19] Fazeli, S., Loni, B., Drachsler, H., & Sloep, P., "Which recommender system can best fit social learning platforms?," in European Conference on Technology Enhanced Learning, Cham, 2014. https://doi.org/10.1007/978-3-319-11200-8_7

[20] Xiao, J., Wang, M., Jiang, B., & Li, J., "A personalized recommendation system with combinational algorithm for online learning.," ournal of Ambient Intelligence and Humanized Computing, vol. 9, no. 3, pp. 667-677, 2018. https://doi.org/10.1007/s12652-017-0466-8

[21] Bieliková, M., Šimko, M., Barla, M., Tvarožek, J., Labaj, M., Móro, R & Ševcech, J., "ALEF: From application to platform for adaptive collaborative learning," in Recommender

Systems for Technology Enhanced Learning, New York, NY, 2014. https://doi.org/10.1007/978-1-4939-0530-0_10

[22] Dwivedi, P., Kant, V., & Bharadwaj, K. K., "Learning path recommendation based on modified variable length genetic algorithm.," Education and Information Technologies, vol. 23, no. 2, pp. 819-836, 2018. https://doi.org/10.1007/s10639-017-9637-7

[23] Nowakowski, S., Ognjanović, I., Grandbastien, M., Jovanovic, J., & Šendelj, R., "Two recommending strategies to enhance online presence in personal learning environments.," in Recommender Systems for Technology Enhanced Learning, 2014. https://doi.org/10.1007/978-1-4939-0530-0_11

[24] Sriurai, W., Meesad, P., & Haruechaiyasak, C., "Recommending Related Articles in Wikipedia via a Topic-Based Model," in IICS, 2006.

[25] Adline, A. A., & Mahalakshmi, G. S., "A recommender system for wiki pages: Usage based rating approach," in Recent Trends In Information Technology (ICRTIT), 2012. https://doi.org/10.1109/icrtit.2012.6206831

[26] Aggarwal, C. C., "Structural Recommendations in Networks," in Recommender systems, Springer, 2016, pp. 309-334.

[27] Schwarzer, M., Schubotz, M., Meuschke, N., Breitinger, C., Markl, V., & Gipp, B., "Evaluating link-based recommendations for Wikipedia," in Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries, 2016. https://doi.org/10.1145/2910896.2910908

[28] Leake, D., Maguitman, A., & Reichherzer, T., "Understanding knowledge models: Modeling assessment of concept importance in concept maps.," in twenty-sixth annual conference of the cognitive science society, 2004.

[29] Garcia, I., & Ng, Y. K., "Eliminating Redundant and Less-Informative RSS News Articles Based on Word Similarity and a Fuzzy Equivalence Relation. In Tools with Artificial Intelligence," in ICTAI'06, 2006. https://doi.org/10.1109/ictai.2006.54

[30] Ogawa, Y., Morita, T., & Kobayashi, K., "A Fuzzy Document Retrieval System Using The Keyword Connection Matrix and a Learning Method," Fuzzy Sets and Systems, vol. 39, no. 2, pp. 163-179, 1991. https://doi.org/10.1016/0165-0114(91)90210-h

[31] Yerra, R., & Ng, Y. K., "Detecting Similar HTML Documents Using a Fuzzy Set Information Retrieval Approach," in Granular Computing IEEE International Conference, 2005. https://doi.org/10.1109/grc.2005.1547380.

[32] Leake, D. B., & Maguitman, A., "Combining concept mapping with CBR: towards experience-based support for knowledge modeling.," in fourteenth international florida artificial intelligence research society conference, Florida, 2011.

[33] Aggarwal, C. C., Recommender systems, NYm USA: Springer International Publishing, 2016.

[34] Chin, D. N., "Empirical evaluation of user models and user-adapted systems.," User modeling and user-adapted interaction, vol. 11, no. 1-2, pp. 181-194, 2001. https://doi.org/10.1023/a:1011127315884

[35] Cremonesi, P., Garzotto, F., Negro, S., Papadopoulos, A. V., & Turrin, R.Cremonesi, P., Garzotto, F., Negro, S., Papadopoulos, A. V., & Turrin, R., "Looking for "good" recommendations: A comparative evaluation of recommender systems." in IFIP Conference on Human-Computer Interaction, Berlin, 2011. https://doi.org/10.1145/1979742.1979896

[36] Pu, P., Chen, L., & Hu, R., "A user-centric evaluation framework for recommender systems," in fifth ACM conference on Recommender systems, 2011. https://doi.org/10.1145/2043932.2043962

[37] Cañas, A. J., Novak, J. D., & Vanhear, J., "Concept map-based knowledge assessment tasks and their scoring criteria: An overview.," 2012.

[38] Burby, J., Brown, A., & WAA Standards Committee, "Web analytics definitions," Web Analytics Association, Washington DC, 2008.

[39] Parmenter, D, Key performance indicators: developing, implementing, and using winning KPIs. John Wiley & Sons, 2015. https://doi.org/10.1002/9781119019855

7   Authors

Heba Ismail is a PhD candidate in the department of computer science and software engineering in the College of Information Technology at UAE University. She worked on multiple research projects related to text mining, social network analysis, information retrieval, and natural language processing. Currently, her research interest is focused on user modeling, recommender systems for technology enhanced learning, and personalized learning systems. Email: hebaismail20@gmail.com

Boumediene Belkhouche currently works at the Primitive Learning, École Buissonnière. Boumediene does research in Algorithms, Programming Languages and Software Engineering. Their current project is 'personalized learning.' Email: bbelkhouche@uaeu.ac.ae

Saad Harous obtained his PhD in computer science from Case Western Reserve University, Cleveland, OH, USA in 1991. He has more than 30 years of experience in teaching and research in three different countries: USA, Oman and UAE. He is currently a Professor at the College of Information Technology, in the United Arab Emirates University. His teaching interests include programming, data structures, design and analysis of algorithms, operating systems and networks. His research interests include parallel and distributed computing, P2P delivery architectures, wireless networks and the use of computers in education and processing Arabic language. He has published more than 150 journal and conference papers. He is an IEEE senior member. Email: harous@uaeu.ac.ae