IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 1, 2019, accepted November 18, 2019, date of publication November 27, 2019,
date of current version December 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2956284

Framework for Personalized Content Recommendations to Support Informal Learning in Massively Diverse Information Wikis

HEBA M. ISMAIL, BOUMEDIENE BELKHOUCHE, AND SAAD HAROUS
Department of Computer Science and Software Engineering, UAE University, Al Ain, United Arab Emirates
Corresponding author: Heba M. Ismail (hebaismail20@gmail.com)

ABSTRACT Even though, most of the early efforts in personalized learning focused on formal learning, there is a growing undeniable demand for personalized support for informal learning. Wikis among other information-oriented platforms are experiencing an increasing attention for informal learning, especially Wikipedia. Link-based navigation and keyword-based search methods used on wiki environments suffer from many limitations. To support informal learning on these environments, it is important to provide easy and fast access to relevant content. However, the massive diversity of unstructured content and user base on these environments pose major challenges when designing recommendation models. To the best of our knowledge, no effective personalized content recommendation approach has yet been defined to support informal learning on wikis. We propose an effective personalized content recommendation framework (PCRF) in addition to an evaluation framework that can be used to evaluate the impact of personalized content recommendations on informal learning from wikis. PCRF implements an efficient structural recommendation model by integrating fuzzy thesauri with adaptive users' interest models generated using structural analysis of topical navigational graphs. We design user studies with multiple strategies and treatments to evaluate the effectiveness of the framework and assess the impact of recommendations on informal learning. Experiments show that PCRF generates highly relevant recommendations adaptive to changes in users' interests using the HARD model with MAP@k scores 86.4–100%. An evaluation of informal learning revealed that users of Wikipedia with personalized support could achieve higher scores on a conceptual knowledge assessment with an average score of 14.9 compared with 10.0 for users who used the encyclopedia without recommendations. Results confirm that PCRF can effectively support informal learning on wikis and similar environments.

INDEX TERMS Information filtering, information wikis, informal learning, personalized content recommendations, recommender systems, Wikipedia.

I. INTRODUCTION

Personalized learning advantages have become evident through research and practice [1]. Even though, most of the early efforts in personalized learning focused on formal learning, there is a growing undeniable demand for personalized support for informal learning [2], [3]. In contrast to formal learning, informal learning is independent, self-paced, does not align with a specific curriculum, and does not lead to a formal qualification [4]. Wikis, especially Wikipedia, are experiencing an enormous attention for informal learning [5]–[7]. A study conducted in 2009 revealed that up to 82% of high school students in the U.S. turn to Wikipedia to give their research a head start, and up to 76% of them use Wikipedia to find the definition of terms in specific subjects [8]. As of today, Wikipedia contains more than 157,000,000 articles in 302 languages among which 37,000,000 articles are in English [9]. This makes it a huge knowledge repository for informal learning. To support informal learning on diverse information wikis with a heterogeneous user base, it is important to effectively provide fast and easy access to relevant content. This can be primarily accomplished with a suitable user model. User models are fundamental components in personalized systems in general. These models define important user characteristics that are used to adapt and personalize relevant content [10]. The set of user characteristics modeled

The associate editor coordinating the review of this manuscript and approving it for publication was Arianna Dulizia.

depends on the type of content being personalized as well as on the objective of the personalization system. In personalized learning systems where learning content is typically being personalized, characteristics such as knowledge and skill-level [11]–[14], emotions [15], preferences [16], and context [17] are usually modeled. These characteristics, especially learner knowledge, are often important in formal learning systems that deliver predefined content and attempt to achieve well-defined learning outcomes. For instance, tutoring systems [18], and online courses [19]. The fact that these formal systems deliver a very specific content for a very specific learner base creates no demand for personalized interest modeling. Traditionally, learners using these personalized formal learning systems come with an interest to use and learn the specialized content delivered in these systems. However, user interests have always constituted the most essential aspect of user models, sometimes competing for user knowledge, for adaptive and personalized information retrieval and filtering systems. This is often referred to as adaptive hypermedia, which deals with a huge bulk of diverse information, such as online encyclopedias [20]. To this end, considering the context of information wikis and specifically Wikipedia's context, one method to specify users' interests is through keyword-based search. However, in many cases, users may fail to identify representative keywords. Another method to specify users' interests is through hyperlinks. Hyperlinks can divert the user away from the main topic of interest. Additionally, links mentioned in a specific Web Page cannot completely cover all relevant Web Pages in the whole corpus. That is because there might be no terms describing a relevant Web Page within the current Page or simply because some links are not working. Furthermore, the vast diversity of Wikipedia's content and user base raises major challenges in modeling users' interests. Recommendation systems (RSs) are used to support user-centric interest modeling and deliver useful recommendations in various areas of applications. For instance, in technology enhanced learning (TEL) contexts, RSs are commonly used to recommend courses, peers, or learning paths [21]–[23]. The most commonly used techniques for TEL RSs are collaborative filtering (CF), and content-based filtering (CB) [22]. CF approaches recommend items primarily based on similarities between users. CF approaches identify similarities by analyzing recurring patterns of interests. Hence, these approaches might not be successful in dealing with changing and diverse, or non-recurring users' interests as seen on Wikipedia. In contrast, CB approaches use item's descriptive features to recommend new items with similar attributes. However, converting unstructured text into feature vector representation eliminates essential latent semantic relationships that exist in original text. Even though, current TEL RSs research shows promising results, especially in online learning environments with focused learning objectives and well-defined learning content and learners' base, it yet needs to offer new RSs models suitable for massively diverse unstructured content with a heterogenous user base as seen in Wikipedia. A new variation of CB approaches has been used to address these challenges. For instance, Sriurai *et al.* [24] utilized the Latent Dirichlet Allocation (LDA) to generate topic-based content recommendations, and Adline and Mahalakshmi [25] used an article quality framework to classify and recommend Wikipedia articles into readable, learnable, and referable. In other studies, a modification to search algorithms was proposed to deliver structural recommendations [26]. In structural recommendation techniques, content or/and users are represented using graphs. Structural analysis of graphs is then used to recommend nodes, links, or different combinations of both. For instance, Schwarzer *et al.* [27] proposed a structural recommendation framework for Wikipedia articles based on a modified form of Co-Citation Proximity Analysis (CPA). However, these recommendation models lack personalization, do not support adaptive user modeling, and have not evaluated the impact of recommendations on learning. Nevertheless, we are concerned with personalized support for informal learning, hence, there is a need to model a personalized content recommendation framework for massively diverse wikis as well as evaluate the impact of recommendations on informal learning. We propose a personalized structural recommendation framework for massively diverse information wikis. In addition, we propose an evaluation strategy to evaluate the impact of personalized recommendations on informal learning from wikis. Our conceptual design and preliminary results were presented in EDUCON19 [28]. This paper provides extensive details, formulas, and algorithms related to the proposed framework which can be used to reproduce the same framework on any information-oriented environment with similar properties. Furthermore, a comprehensive evaluation and discussion of experimental results are presented.

The rest of this article is organized as follows. In Section II, we explore major challenges related to modeling learners in massively diverse wikis as well as personalizing unstructured text as seen on similar information wikis. Then, we introduce an overview of related literature in Section III. In Section IV, we explain the major components of the proposed framework. In Section V, we describe our evaluation process, the experimental setup, and the results of our experiment. The work is concluded in Section VI.

II. CHALLENGES RELATED TO MASSIVELY DIVERSE INFORMATION WIKIS

In this section we introduce some challenges related to modeling learners and processing content on Wikipedia that accentuated the need for the proposed personalized content recommendation framework.

A. LEARNER MODELING CHALLENGES

Typically, in wiki environments, users do not follow consistent patterns of interest over a long period of time. Rather, users are more likely to change their interests over navigation sessions or sometimes within a single session. Rodi *et al.* [29] analyzed the English Wikipedia Clickstream (EWC) dataset collected in February 2015 and found that Wikipedia readers do not have a well-defined target in mind. Rather, they start with highly abstract topics and then look at more detailed and focused topics as they continue navigation. These results characterize users' navigation on Wikipedia as being *exploratory* rather than *definite*. Therefore, to model learners' interests on information wikis, it is important to *adapt* to *changes* in users' interest. Additionally, West and Leskovec [30] compared human navigation in information-oriented environments with that of software agents and found that humans, when navigating within an information-oriented network, have expectations about what links should exist next and base a high-level reasoning plan upon this, and then use local information to navigate through the network. These studies suggest that the longer users navigate the information network the more focused they become on their target and they tend to do this through local information usually using links. Therefore, to help users make the best use of local information, it is important to make sure it is relevant.

However, articles in massively diverse information wikis form a scale-free network [31]. That is, some articles are highly connected forming hubs and thus most commonly linked to other articles whereas many articles are not highly connected, and thus, relevant information can be missed when recommending articles merely based on links. Therefore, to personalize content recommendations on information wikis, there is a need to adaptively model the changing interests as well as recommend articles based on semantic relevance, not just based on links or references.

B. LEARNING CONTENT PROCESSING CHALLENGES

A variety of learning content representations can be used in personalized learning software systems. In addition to learning objects [32], [33], ontologies [34], or more recently Linked Open Data (LOD) [35], a considerable amount of learning content on the web is available in the form of unstructured free text. Typically, we find this in blogs, wikis, forums, and on social media websites. Unstructured texts suffer from several complications. In contrast to structured data or formal knowledge representations, unstructured texts have no fixed attributes with predefined values. It may contain any number of different words. Several methods were proposed to represent semantics in unstructured text. Most of these methods can be classified into two classes: contextual methods, and conceptual methods. Conceptual methods of semantic analysis rely on external semantic knowledgebases. For instance, ontologies and semantic networks. Conceptual semantic methods are limited by their underlying knowledgebases and require large amount of manual efforts during the knowledgebase creation and validation phase. In contextual methods, the relationships between terms in the text are analyzed statistically. These relationships are mainly co-occurrences. These methods tend to be more flexible given the possibility of automation. Hence, we use contextual semantic analysis with fuzzy thesauri. Fuzzy thesauri are commonly used in information retrieval and have demonstrated effective results in a variety of contexts [36], [37]. Their main advantages come from domain-independence, and easy automation. Fuzzy thesauri are built based on concepts from the fuzzy set information retrieval model [38], [39].

III. BACKGROUND AND LITERATURE REVIEW

A. USER INTEREST MODELING IN INFORMATION-ORIENTED HYPERMEDIA ENVIRONMENTS

In the Cambridge Advanced Learner's Dictionary, ''interest'' is defined as ''the activities that you enjoy doing and the subjects that you like to spend time learning about'' [40]. Methods and techniques used to model the interests of users in information-oriented hypermedia environments have considerably varied over time. We review several studies conducted on interest modeling. Our review excludes studies that use user ratings or user likes/dislikes to model interest as well as models of interest that rely on contextual data such as location, speed, or time as found in context-aware systems. In this research, we use the term context to refer to the semantic context, implying the meaning of the text and not the physical context. TABLE 1 presents summary of some of the interest modeling approaches discussed in this section.

TABLE 1. Summary of user interest modeling approaches in information-oriented websites.

User Interest Model		Research Work
Keyword-based	*Explicit/ User-defined*	[41], [42] , [43], [44], [45], [46] , [47]
	Implicit/ Corpus-based	[48], [49], [50], [51], [52]
Semantics-rich	*Semantic Networks*	[53], [54], [55]
	Ontology	[56], [57], [58]
	Concept Hierarchies	[59]

Early efforts in user interest modeling focused on the keyword level [60]. Keywords representing user interests could be collected explicitly from the user or implicitly extracted from the documents navigated by the user. Keywords expressed explicitly by users remain the simplest and most common despite the various limitations associated with this approach. As a result, many efforts focused on improving on explicit keyword-based interest models by permitting users to better specify their interests through additional context information such as categories [43], preferences [42], topics [41], or Folksonomies, also known as social tagging [61]. More recently, work in this line explored approaches of data visualization to support information exploration by visually suggesting relevant keywords. Work in this field proposes query suggestions [44], negative relevance feedback as used in Intent Radar [45], or visualization as seen in Adaptive-VIBE [46] and SearchLens [47]. However, interest modeling approaches relying on keywords defined explicitly by users suffer from many limitations that were highlighted in a number of research studies [60], [62], [63]. For example, users may fail to use the right keywords, some keywords may have different meanings in different contexts, and distinct keywords do not convey the level of importance of interests a user has in a certain subject. Alternatively, weighted vectors of keywords implicitly extracted from navigated documents were used to relief the user from having to choose the right keywords, and to give some sort of weighting to different keywords in the user profile [48]–[52]. The keywords in the profile are extracted from documents visited by the user during browsing, or web pages bookmarked or saved by the user. Corpus-based statistics such as Term Frequency Inverse Document Frequency (TF-IDF) are commonly used to weight keywords in the weighted vectors' user profile [64]. However, being derived and weighted automatically from corpus, weighted vectors are ineffective in dealing with continuously changing user interests and might contain inaccurate keywords that are not interesting for the user, yet, are highly weighted according to corpus statistics. Additionally, weighted vectors might over weigh less-interesting keywords, or under weigh more-interesting keywords based on corpus statistics. Moreover, keywords extracted from text are extracted from their context as well, resulting, sometimes, in ambiguities. To address limitations associated with explicit keywords and keyword vectors, researchers used semantics-rich representations such as semantic networks [53]–[55], concept vectors [59], and ontologies [56]–[58]. In semantic networks-based interest models, each node represents a concept or a word, and each edge has a weight that reflects the relationship between concepts in the semantic network. Additionally, context attribute can be added to enrich the semantic network. Concept-based models are similar to semantic network-based models. That is, both are characterized by concept nodes and relationships between those nodes. However, in concept-based models, the nodes are abstract topics that are found interesting by the user. Semantics-rich user models have an advantage over keyword-based models because they can explicitly model the relationships between particular words and higher-level concepts. However. These approaches are more difficult to build compared to keyword-based models, in many cases manual identification and mapping of concepts and relationships are required, and they are restricted by their underlying knowledgebases. Furthermore, these approaches are not adaptive to changing users' interests. User-centered and adaptive interest modeling approaches began in recommender systems (RSs) [21].

B. RECOMMENDER SYSTEMS

Recommender systems are defined as: *"any system that produces individualized recommendations as output or has the effect of guiding the user in a personalized way to interesting or useful objects in a large space of possible options [65]"*. This definition expands the field of recommender systems to any application that computes a user-specific utility, covering many areas of applications. To identify users' needs and map these needs to suitable items, researchers proposed several recommendation classes [66]–[70]. However, we consider the following three classes to be the most appropriate for differentiating the approaches in the field of recommender systems in information-oriented educational websites:

1. Content-based (CB)
2. Collaborative filtering (CF)
3. Structural recommendations in networks

CB is one of the most extensively used and studied recommendation approaches [66]. A crucial task of CB is the user modeling activity, in which the interests of users are extrapolated from the items that users interacted with. "Items" can be books [71], research papers [72], or webpages [73]. Items are represented by a content or document model containing the items' descriptive attributes which we commonly refer to as features. Features are typically word-based, i.e. single terms, phrases, or n-grams. The user model typically consists of the features of a user's items. To find recommendations, the user model and candidate items are compared in the vector space model and similarities are calculated with suitable similarity measures, e.g. Cosine. CB has several advantages. For instance, CB allows a more personalized recommendations by defining items' descriptive features, rather than be limited by what other like-minded users like. CB also requires fewer manual efforts since user models can be created automatically. However, considering the context of massively diverse information wikis, the process of transforming unstructured content into feature vector representation of distinct terms result in many issues. *First*, contextual features found in original text are removed. Terms are extracted from their context eliminating essential latent semantic relationships. *Second*, generated datasets are likely to be very sparse with very huge feature space resulting in computational complexities and inaccuracies.

In contrast to CB, CF assumes that users usually like what other like-minded users like, where two users are considered like-minded when they rate items similarly. Therefore, when like-minded users are identified, items that one user rate positively and not yet seen or rated by the other like-minded user, are recommended to the other user, and vice versa. CF offers three advantages. *First*, CF is content-independent [67]. *Second*, CF considers real quality measurements [67]. *Finally*, CF is supposed to provide serendipitous, i.e. surprising, recommendations because recommendations are not based on item similarity but on user similarity [74], [75]. A major drawback, however, in CF is the "cold start problem," which may occur in three situations [67]: new users with no rating or navigation history, new items that have not yet received any ratings or impressions from users, and new topics or disciplines. This hinders the applicability of CF approaches to provide relevant recommendations to new users or recommend new topics. Additionally, computational time complexity for CF algorithms tends to be higher than for CB [67]. This in turn limits the applicability of CF algorithms for contexts in which item space or user base is massively large as seen in Wikipedia and similar environments. Moreover, Torres et al. point out that CF creates similar users [76] and Sundar et al. criticize that collaborative filtering dictates opinions [77]. This drawback of CF conflicts with the massive diversity of Wikipedia's content and user base. Therefore, different variations of recommendation models are required to address the challenges associated with massively diverse content and user base. We focus on structural recommendations for this objective.

Structural recommendations are generated based on structural analysis of networks. The two major categories of structural recommendation models are: link-based recommendations, and node-based recommendations [26]. For node-based recommendations, the quality of nodes is determined by their incoming edges, and the personalized relevance of nodes is determined by their context [78], [79]. For link-based recommendations, potential links are recommended [80]. Several ranking methods are used for link prediction. Additionally, matrix factorization methods can also be adapted for link prediction [81]. Structural recommendations do not rely on identifying like-minded users or inferring similarities between items' features. As a result, it can help alleviate the limitations associated with CF and CB for personalized content recommendations in massively diverse environments. In Section IV we introduce our structural recommendation framework.

C. WIKIPEDIA RECOMMENDER SYSTEMS

Many studies focused on designing recommendation models for Wikipedia. These can be classified according to the item being recommended into two categories: article recommendation models, and task recommendation models. Task recommendations on Wikipedia are concerned with recommending editing tasks to authors as proposed in [82], [83], and [84]. In our research we are focused on article recommendation models.

Several recommendation models were proposed to deliver article recommendations in Wikipedia. For instance, Sriurai *et al.* [24] proposed a topic-based model to classify articles against topic features using LDA. The model was evaluated with an unspecified number of articles by five assessors. The average relevance score for recommended articles exceeded the relevance score of the linked articles by 1.2. The approach is not designed to generate personalized recommendations. It considers users navigating the same page to have the same interests. Hence, fixed recommendations are provided to all readers following a preset topic distribution. In addition, those recommendations were not employed in any learning activity to evaluate their impact learning.

In contrast, Adline & Mahalakshmi [25] proposed a more sophisticated article quality framework to classify and recommend Wikipedia articles into three usage categories: (1) readable, (2) learnable, and (3) referable, based on some article quality attributes. In their evaluation, fifty 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. The system was found accurate in detecting the usage category of various articles. The proposed framework treats all users equally and does not account for personalization. It treats Wikipedia as a comprehensive source of knowledge and categorizes the articles based on usage purposes accordingly. However, Wikipedia articles are not meant to be an eventual source of knowledge. They are meant to give a lead start. Moreover, evaluating the quality of articles is a valuable contribution, but, assuming that users with certain learning objectives would prefer to view articles of specific quality measures is a very strong assumption that was neither verified in the paper, nor in the literature.

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.

Consequently, because we are concerned with personalized support for informal learning, there is a need to model an effective personalized content recommendation framework for Wikipedia as well as evaluate the impact of recommendations on informal learning.

IV. PERSONALIZED CONTENT RECOMMENDATIONS FRAMEWORK (PCRF)

PCRF 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 multigraph, TNG (V, E). Every vertex,V, in TNG corresponds to a topic, modeled at the page level, and every edge, E, in TNG corresponds to a navigational action. Then, structural topical graph analysis algorithms, adapted from Leak *et al.* [85], 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 used as a user model to recommend semantically relevant topics based on fuzzy thesauri. The fuzzy thesauri are built based on concepts from fuzzy set information retrieval model [38]. The resulting set of ranked and semantically relevant topics represents the final personalized content recommendations.

Our framework comprises four main modules: session tracking, TNG analyzer, personalization, and 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 which are used online by the personalization module to generate personalized content recommendations based on the learner models generated by the TNG Analyzer module. We describe

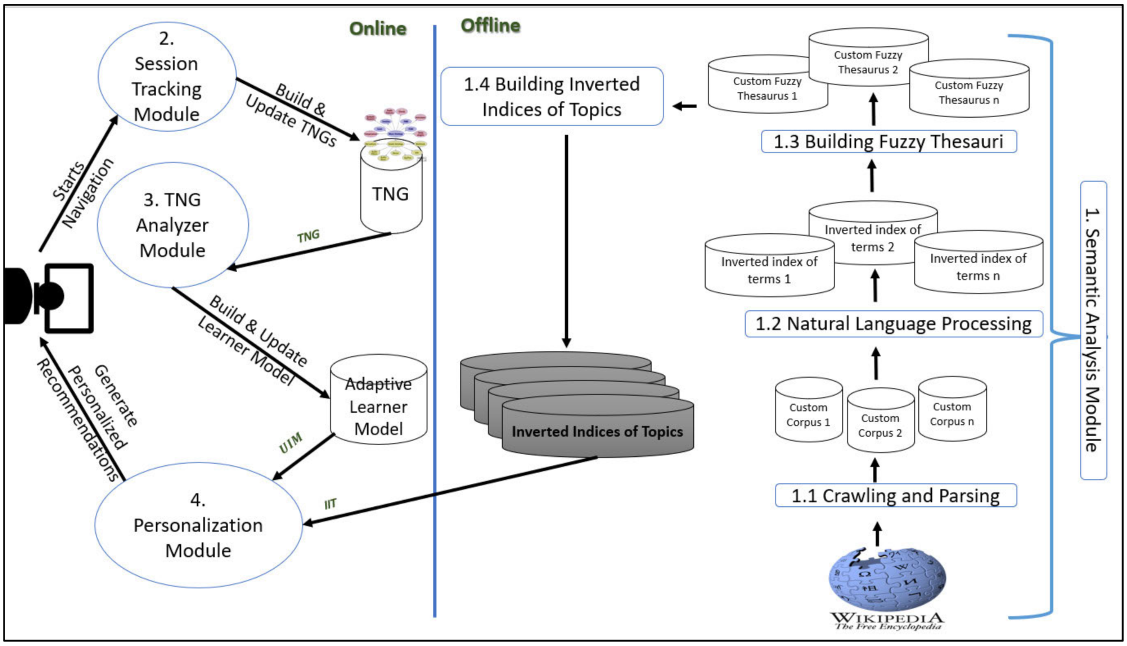


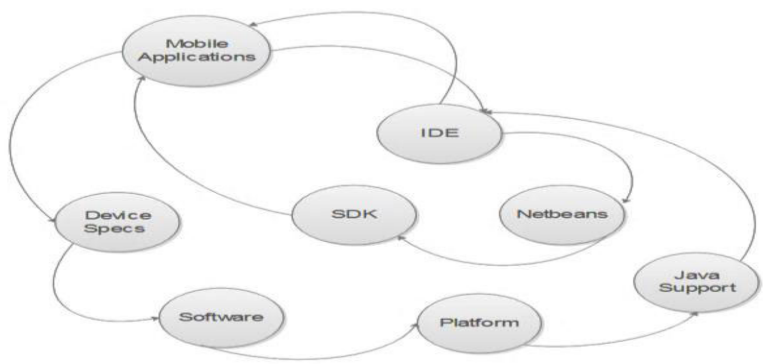
FIGURE 1. Proposed personalized content recommendation framework.

TABLE 2. Defining the main concepts of the proposed framework.

Concepts	Description	Definitions
A Learning Session	A learning session is a sequence of learning resource accesses related to the same user. It starts when the user accesses the domain of the wiki and ends when the user leaves the domain. There is no time constraint. The learning session is represented using a multigraph data structure which we name it a Topical Navigation Graph, TNG, where vertices, V, are weighted topics of interest, and edges, E, are multiple navigational actions between vertices.	▪ $A\ Learning\ Session \stackrel{\text{def}}{=} TNG$ ▪ $TNG = (V, E)$ ▪ *V is a set of weighted vertices, representing the user's topics of interest.* ▪ *E is a multiset of directed edges between nodes in V.*
A Learning Resource	A learning resource is a webpage containing learning content in a wiki. In our work, a learning resource is represented by a topic (the main topic of the web page). A topic is depicted by a vertex in the topical navigation graph. Every vertex, v, has a label, l, and a weight, $W(v)$, such that and the set of visited learning resources is V.	▪ $v = (l, W(v))$ ▪ l *is a label representing the topic of the vertex* ▪ $W(v)$ *is the structural weight of the vertex*
A Structural Weight	A structural weight defines the rank of a vertex in the user navigation graph based on graph structural characteristics only. Two models, HARD and CRD, for graph structural analysis are used to calculate weights. These models are explained later.	*For the CRD Model:* ▪ $W(v) = (\alpha . o(v) + \beta . i(v)).\left(\frac{1}{d(v)+1}\right)^{\frac{1}{\delta}}$ *and for the HARD model:* ▪ $W(v) = \alpha . h(v) + \beta . a(v) + \gamma . u(v)$
A User Interest Model	User interests are defined as topics that receive higher weights in the learner navigation graph after applying structural analysis of TNGs. Hence, for a user, i, a user interest model, UIM_i, is represented using a subset of V_i that belongs to TNG_i for that particular user.	▪ $UIM_i \subseteq V_i : \forall\, uim_i \in UIM \ \wedge\ \forall\, v_i \in (V - UIM),\ W(uim_i) > W(v_i)$
Personalized Recommendations	Personalized content recommendations for user i, PCR_i, can be obtained by mapping topics, l_i, from the user model of user i, UIM_i, to semantically similar learning resources or documents, d, in the inverted index of topics, IIT. Ranking of the personalized recommendations can be achieved using the weights of topics in the user model of user i, $W(uim_i)$.	▪ $PCR_i = W(\{uim_i : uim_i \in \mathrm{UIM}_i\}) * Sim(\{l_i : l_i \in \mathrm{UIM}_i\}, \{d_n : d_n \in IIT\})$

each module in the following sections. Table 2 lists and defines the main concepts we use in our research. Table 3 illustrates a motivating example of a user navigating a website about "Mobile Applications Development". We use this example to illustrate the different phases of the process.

TABLE 3. Illustrating the process of learner modeling using a graph with a motivating scenario.

Scenario	Navigation History	TNG_i
Scenario # 1	Mobile Applications → IDE → Netbeans → SDK → Mobile Application → Device Specs → Software → Platform → Java Support → IDE → Mobile Applications	V_{CRD} = {(Mobile Applications,1.0),(Netbeans, 0.333), (SDK, 0.125), (Device Specs, 0.25), (Sofwtare, 0.167), (Platform, 0.125), (JAVA Support,0.1), (IDE, 0.5)} E = {(Mobile Applications, IDE), (IDE, Netbeans), (Netbeans,SDK), (SDK, Mobile Applications), (Mobile Applications, Device Specs), (Device Specs, Software), (Software, Platform), (Platform, Java Support), (Java Support, IDE), (IDE, Mobile Applications)}
	UIM_i	PCR_i
	UIM_i = {(Mobile Applications,1.0), (IDE, 0.5), (Device Specs, 0.25)}	1. Developing Mobile Applications 2. IDEs for Symbian 3. Adjusting Mobile Specifications from apps.
Illustration		

A. SEMANTIC ANALYSIS MODULE

We perform semantic analysis using fuzzy thesauri built based on fuzzy set information retrieval model. The objective of this module is to generate inverted indices of topics that can be used to associate semantically relevant documents to topics that are found interesting to the learner in the learner model. The complete process of building the fuzzy thesauri and generating the inverted indices of topics is explained in **Figure 2**. The algorithm is explained in **Figure 4**.

First, custom corpora are extracted from Wikipedia for each main topic category as classified by Wikipedia using a web scraper application. From this step we get a custom corpus for each main topic such as science, art, culture, etc. These corpora are represented in HTML. Thus, the ***second*** step in the process is to convert all HTML-based corpora into plain text corpora. Only content within paragraph tags, <p>, and title tags, <title>, are extracted. Index pages are excluded from the corpora as they do not have any learning content.[1]

The ***third*** step aims at generating inverted indices of unique terms that can be used to build the fuzzy thesauri. At this stage, natural language processing [86] of the custom text-based corpora is performed. Generally, unstructured texts cannot be directly processed for semantic analysis. Thus, we perform several natural language processing tasks which proved effective and have become a common practice for unstructured text preprocessing for semantic analysis.

We perform the following preprocessing tasks in order:

1. ***Tokenization***: all documents are converted into vectors of raw unprocessed terms, i.e. tokens.
2. ***Stopwords removal***: stopwords usually refer to the most common words in a language and are considered to have little meaning, for example in English some stopwords are: "a," "an," "and," "are," "as," "at," "be," "but," and "by."
3. ***Stemming*** [87]: this is a process of eliminating the most common morphological and inflectional endings from words in a language with the assumption that all words derived from the same stem share the same meaning.
4. ***Inverted words index creation***: an inverted word index, *VEC*, is a set of vectors where each vector indicates for each unique word in the corpus: the documents in which it appears, and its positions, i.e. occurrences, in that document. For example, Table 4 and Table 5 show the vectors of stemmed words "mobil" and "comput", *vec(mobil) & vec(comput)*, in the inverted index of words.

In step ***four***, we build a custom fuzzy thesaurus that defines the semantic similarity between each two distinct words in each corpus by calculating the distance correlation factors using equations (1), (2), and (3).

[1] Processed corpus can be requested from correspondence author via email.

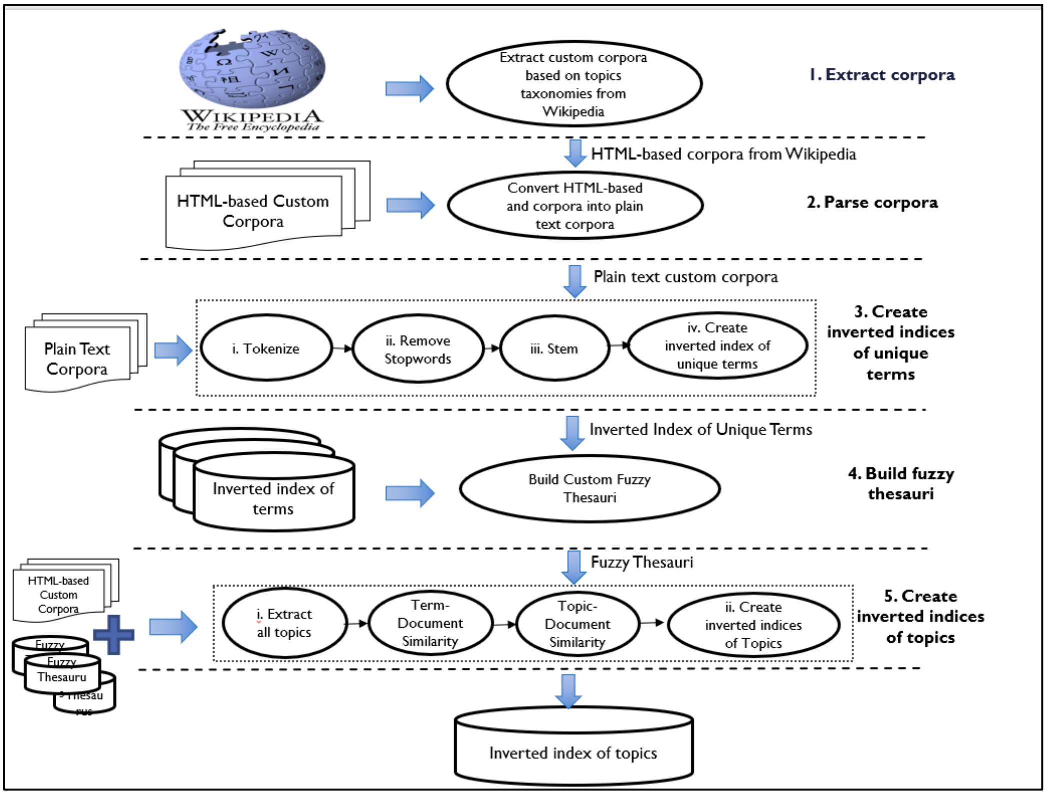


FIGURE 2. Process of building the fuzzy thesauri and generating the inverted indices of topic.

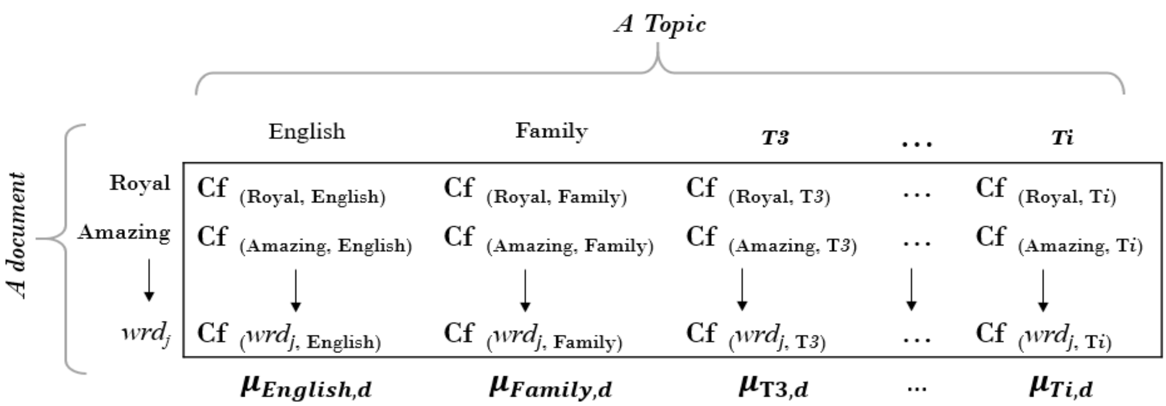


FIGURE 3. Calculating the Term-document semantic similarity (μ_(T,d)) between each term (Ti) in a given topicn and each document (d_n) in the wiki corpus.

TABLE 4. Inverted Index for "Mobil", vec(Mobil).

Term	Document ID	Position
Mobil	d_1	1, 8
	d_2	12, 30
	d_3	7, 10, 27
	d_4	30

TABLE 5. Inverted Index for "Comput", vec(Comput).

Term	Document ID	Position
Comput	d_1	5,30
	d_5	17, 38
	d_6	10, 29

We choose the distance correlation factor because it has empirically proved to achieve the best results in the information retrieval context with an accuracy rate of 94% compared with 47% for the keyword-connection factor and 52% for the co-occurrence factor [38]. That is because distance correlation factors account for frequency and co-occurrence at the same time.

Using the inverted indices of distinct terms, *VEC*, we define for every pair of terms across all documents within a single corpus: the frequency of co-occurrence and relative

```
Algorithm 1 Generating Inverted Indices of Topics (IIT)
Input:
    WikiC: Wikipedia Corpus, WikiC = {d1, ..., dn}
Output:
    IIT: Inverted indices of topics
Procedure GENERATE_INVERTED_INDICES_OF_TOPIC (WikiC)
for each Document, dn = (wrd1, wrd2, ..., wrdj) ∈ WikiC do
  | Extract main topic → topic
  | Update set of topics → Topic
  | Extract distinct words → wrd
  | Remove stopwords
  | Stem distinct words
  | Generate inverted indices of stemmed words → VEC
end for
for each vec(wrdj) ∈ VEC do
  | for each vec (wrdi) ∈ VEC do
  |   | if (wrdj & wrdi appear in the same document) then,
  |   |   | Calculate the frequency of co-occurrence and relative distance in a single document → Cij
  |   |   | Calculate the normalized value → nCij
  |   |   | Calculate the distance correlation factor → Cfij
  |   |   | Update fuzzy thesaurus → FuzTh
  |   | end if
  | end for
end for
for each Document, dn = (wrd1, wrd2, ..., wrdj) ∈ WikiC do
  | for each Topic, topicn = (T1, T2, ..., Ti) ∈ Topic do
  |   | Calculate Term-Document semantic similarity → μ_(Ti, dn)
  |   | Calculate Topic-Document semantic similarity → Sim (topicn, dn)
  | end for
end for
return IIT
```

FIGURE 4. Algorithm for generating Inverted indices of topics.

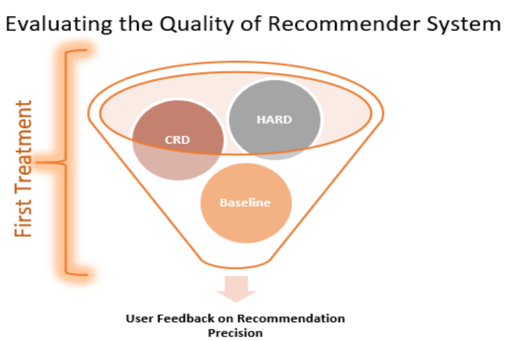


FIGURE 5. User Study - First Treatment.

distance in a single document (C_{ij}), (1), the normalized value (nC_{ij}), (2), and finally the distance correlation factor (Cf_{ij}), (3).

$$C_{i,j} = \sum_{x \in vec(wrd_i), y \in vec(wrd_j)} \frac{1}{distance\,(x, y)} \quad (1)$$

$$nC_{i,j} = \frac{C_{i,j}}{|vec\,(wrd_i)| \times |vec\,(wrd_j)|} \quad (2)$$

$$Cf_{i,j} = \frac{\sum_{m=1}^{k} nC_{i,j}}{k} \quad (3)$$

where $distance\ (x, y) = |\ Position(x)\ \text{-}\ Position(y)| + 1$ is the distance, i.e. the number of words between word x and y in a single document, where x is an element of $vec\,(wrd_i)$ and y is an element of $vec\,(wrd_j)$. $vec\,(wrd_i)$ and $vec\,(wrd_j)$ are the sets of all occurrences of words wrd_i and wrd_j in a single document, d. To calculate the frequency of co-occurrence and relative distance in a single document we sum up the inverse distance of every two occurrences of wrd_i and wrd_j in that common document. For example, the words "mobil" and "comput" appear together in d_1, hence, $vec(mobil) = \{1,8\}$, $vec(comput) = \{5,30\}$, and $C_{mobil,comput} = (1/distance(1,5) + 1/distance(1,30) + 1/distance(8,5) + 1/d(8,30))$. If they appear together in other documents, then we have to repeat the same calculation for every common document as well.

$|vec\,(wrd_i)|$ and $|vec\,(wrd_j)|$ represent the number of occurrences of words wrd_i and wrd_j in $vec\,(wrd_i)$ and $vec\,(wrd_j)$, respectively, i.e. the frequency of wrd_i and wrd_j in a common document, d. For example, $|vec\,(mobil)| = |vec\,(comput)| = 2$ in d_1. Hence, to calculate the normalized frequency of co-occurrence and relative distance for "mobil" and "comput" in d_1 we compute $nC_{mobil,comput} = C_{mobil,comput}$ / (2*2).

The index, m, ranges over $1 \leq m \leq k$ and represents the m_{th} document out of the k documents in which both wrd_i and wrd_j occur together. For the words "mobil" and "comput" the values of m and k are equal, $m = k = 1$. By dividing the sum of normalized values by the number of common documents between every two words in the corpus, distance correlation factors, Cf, are calculated relevant to the size of the corpus. As a result, a matrix of all distinct words and their semantic relationships is constructed. This matrix is the custom fuzzy thesaurus, *FuzTh*, which is used to measure the semantic similarity between different topics of interest in the learner model and in the Wiki.

The ***fifth*** step of the process aims at generating inverted indices of topics, *IIT*. In this phase, main topics, i.e. topics at the webpage or document level, are extracted from the wiki corpora. Topic extraction algorithms such as Probabilistic Latent Semantic Analysis (PLSA) and Latent Dirichlet allocation (LDA) can be used to generate a set of distinct topics, *Topic*. Next, every term, T_i, in every topic, $topic_n$, is compared with every word, wrd_j, in a document, d_n, to retrieve the corresponding distance correlation factor, Cf_{ij}, from the custom fuzzy thesaurus, *FuzTh*, created earlier, which indicates the *word-word* semantic similarity. Once a term, T_i, is compared to each word, wrd_j, in a given document, d_n, the semantic similarity between the term and the whole document, $\mu_{T,d}$, is calculated using Equation (4), which indicates the ***Term-Document semantic similarity***. This is done for each term, T_i, in a given topic, $topic_n$, against a given document, d_n, in the corpus as illustrated in **Figure 3**.

$$\textbf{\textit{Term}} - \textbf{\textit{Document Semantic Similarity}} = \mu\_(T_i, d_n) = 1 - \Pi(1 - Cf_{i,j}) \quad (4)$$

The average of all μ-values for a given topic, $topic_n$, and a given document is calculated to yield the overall similarity between the topic, $topic_n$, and the document, d_n, $Sim\left(\text{topic}_n, d_n\right)$ as follows:

$$\textbf{\textit{Topic}} - \textbf{\textit{Document Similarity}} = Sim\left(\text{topic}_n, d_n\right) = \frac{\mu_{T_1,d_n} + \mu_{T_2,d_n} + \ldots + \mu_{T_i,d_n}}{i} \quad (5)$$

This value will be calculated for all the topics extracted from the wiki corpus against all documents in the corpus to generate an inverted index of topics against documents, *IIT*. An inverted topic index indicates, for each unique topic in the corpus: the documents that are semantically similar and the corresponding semantic similarity value. Table 6 shows sample entry in the inverted topic index for the topic "Amazon River".

TABLE 6. Sample entry in the inverted topic index for the topic "Amazon River".

Topic	Documents	Topic_Document_Similarity
Amazon River	Chew Valley Lake	0.091634
	Colorado River	0.333333
	Columbia River	0.333333
	Congo River	0.333333

B. SESSION TRACKING MODULE

The session tracking module first captures topics of interests from a learner's navigation session in a topical navigation graph, *TNG*. A learning session starts when the learner first accesses the wiki domain and ends when the learner leaves the wiki domain. We model the learner navigation as a directed multigraph, *TNG* (V, E). We are using multigraph since users can go back and forth visiting the same page repeatedly as many times as they want. Every vertex, $v \in V$, 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, $e \in E$, 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 and unconstrained by time. In Table 3 for example we can see how the navigation history of the user is depicted into a TNG structure composed of a set of weighted vertices, each element is a pair of a label and a weight, and a multiset of edges.

C. 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.* [85] Leake and Maguitman [88] to calculate topics' structural weights relevant to individual learners' navigation graphs.

The CRD Model was used by Leak to analyze concept maps' structure based on two observations. ***First***, concepts with higher connectivity, the number of incoming and outgoing connections, may be more important. ***Second***, the root concept, typically located at the top of a concept map, tends to be the most general and inclusive concept. This suggests that concept importance may increase with proximity to the root concept. We find these two observations very relevant and applicable to the navigation behavior of web users. Generally, topics or webpages that are more often visited by a user might be of a special interest compared to topics or webpages that are visited once or very few times in a single navigation session. Moreover, the first visited topic or webpage which act as the root of the TNG might be of a special interest to the user and thus pages that are more closely connected to the root topic might be more important. On the other hand, while CRD Model performs a local analysis, considering only immediate neighbors, HARD Model performs a global analysis on the influences of the concepts on each other. Its analysis centers on three different types of concepts that may be found in a concept map as well as in any web navigation graph:

1. ***Authorities*** are concepts that have multiple incoming connections from hub nodes.
2. ***Hubs*** are concepts that have multiple outgoing connections to authority nodes.
3. ***Upper*** nodes include the root concept and concepts closest to the root concept.

In the context of our research we treat concepts as topics navigated by the user which are depicted as nodes in the topical navigation graph.

The analysis of the structural weights goes through two steps:

1. ***First***, the structural characteristics of each topical node in TNG need to be defined as per the selected model.
2. ***Second***, using the structural characteristics, the relative node's weight W(v) is calculated.

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*. In the following sections we explain the process of identifying the structural characteristics and calculating the structural weights using the same example illustrated in Table 3 to demonstrate the different phases of *TNG* analysis.

1) DEFINING THE STRUCTURAL CHARACTERISTICS

At this stage the navigation graph is analyzed, and structural characteristics of each node is defined as per every model, i.e. CRD, HARD. For example, by applying the CRD model, considering the graph illustrated in Table 3, the node "SDK" is one step away from the root, hence, it has a distance of $d(SDK) = 1$, as well as connectivity of $o(SDK) = 1$, and $i(SDK) = 1$. Table 7 shows position characteristics for some nodes in "Mobile Applications" graph presented in Table 3.

TABLE 7. Position characteristics for some nodes in the "mobile applications' TNG" as per CRD model.

Node Label	Incoming Connections	Outgoing Connections	Distance to Root
Mobile Applications	2	2	0
Device Specifications	1	1	1
Software	1	1	2

Then, using the HARD model, nodes are characterized as hub, authority, and upper nodes. In [88], HITS iterative algorithm was adapted to calculate the relative hub, authority, and upper nodes' positional weights. Leake and Maguitman [88] proved that the proposed algorithm produces positional weights which are ensured to reach a fixed point, converge, after a number of iterations equivalent to the number of nodes in the corresponding concept map. Henceforth, the algorithm to calculate hub, authority, and upper structural weight values of TNG's nodes follows steps 1 to 9:

Step 1: *Set all node's weights w(v) to 1 such that:*

$$hub\_weight = 1$$
$$authority\_weight = 1$$
$$upper\_weight = 1$$

In the following steps, E, refers to the set of edges in the *TNG* graph, q and p, represent any two nodes being currently analyzed in the graph. Hence, the weight of node q is expressed as $w(q)$ and the link between node q and node p is represented as (p,q).

Step 2: *Normalize weights such that:*

$$\sum_{\substack{(v)\in TNG \\ w\in\{authority\_weight, \\ hub\_weight, \\ upper\_weight\}}} w(v)^2 = 1$$

To ensure that this constraint is met, in every step of this algorithm the structural weights, e.g. *Hub_Weight, Authority_Weight, Upper_Weight*, value for every node is divided by the sum of the squares of all corresponding structural weight values in the graph. This is further explained in the following steps.

Step 3: *Calculate Hub_weight such that: Hub_weight of a node, p, is the sum of Authority_weight of all nodes, $q_1, q_{2,\ldots,}q_n$ pointed to by the current node, p such that:*

$$\text{Hub\_Weight}(p) = \sum_{(p,q)\in E} \text{Authority\_Weight}(q)$$

Step 4: *Normalize Hub_Weight to match the constraint in step 2 as:*

$$\text{Hub\_Weight}(p) = \frac{\text{Hub\_Weight}(p)}{\sum_{v\in TNG} (\text{Hub\_Weight}(v))^2}$$

Step 5: *Calculate authority weight such that: Authority_weight of a node, p, is the sum of Hub_weight of all nodes $q_1, q_{2,\ldots,}q_n$ pointing at the current authority such that:*

$$\text{Authority\_Weight}(p) = \sum_{(p,q)\in E} \text{Hub\_Weight}(q)$$

Step 6: *Normalize Authority_weight to match the constraint in step 2 as:*

$$\text{Authority\_Weight (p)} = \frac{\text{Authority\_Weight (p)}}{\sum_{v\in TNG} (\text{Authority\_Weight}(v))^2}$$

Step 7: *Repeat steps 3 to 6 until weights converge. Normally it's repeated as many times as the number of nodes in the graph.*

Step 8: *Calculate Upper node weight as:*

$$\text{Upper\_Weight}(p) = \begin{cases} 1 & if \nexists\,(p,q) \in E \\ \sum_{(q,p)\in E} \text{Upper\_Weight}(q)^2 \end{cases}$$

That is if the node is one level from the root node then assign a weight of one, otherwise sum up the square of *Upper_Weight* of nodes between the current node and the root until the root node is reached then sum up the value of one.

Step 9:*Normalize Upper_Weight according to the constraint in step 2 until they converge as follows:*

$$\text{Upper\_Weight (p)} = \frac{\text{Upper\_Weight (p)}}{\sum_{v\in TNG} (\text{Upper\_Weight}(v))^2}$$

Algorithm 3 Generating Personalized Content Recommendations (PCR)

Input:
IIT: inverted indices of topics
Output:
*PCR_i: Personalized Content Recommendations for user **i***
while *(user in the wiki domain)*
  Initialize TNG_i: $TNG_i = (V, E)$
  OnPageLoad *// HTML event*
  Procedure Generate_Personalized_Content_Rec (***IIT***)
    Extract main topic from the webpage → l_i
    If ($l_i \in IIT$) *// Webpage contains valid learning topic*
      *Update **E***
      Calculate $W(v)$ // Using CRD and HARD models
      Update V_i: $v = (l_i, W(v))$
      Update TNG_i: $TNG_i = (V_i, E)$
      Update UIM_i : $UIM_i \subseteq V_i : \forall\, uim_i \in UIM \;\wedge\; \forall\, v_i \in (V - UIM),\; W(uim_i) > W(v_i)$
      Generate $PCR_i = W(\{uim_i : uim_i \in \mathrm{UIM}_i\}) * Sim(\{l_i : l_i \in \mathrm{UIM}_i\}, \{d_n : d_n \in IIT\})$

    end if
  return PCR
end while

FIGURE 6. Algorithm to generate PCR.

2) CALCULATING THE TOPOLOGICAL WEIGHTS

After defining the structural characteristics of every topic in the TNG using the two different models, CRD, HARD. Now, the topic weight that reflects its importance in the mind of the user can be calculated as follows:

For the CRD Model:

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

and for the HARD model:

$$W(V) = \alpha.h(v) + \beta.a(v) + \gamma.u(v)$$

The CRD Model's parameters α, β, and γ determine influence of the incoming connections, outgoing connections, and distance to the root. The formula implies that the higher a topic's connectivity and the shorter its distance to the root topic the larger its weight. For the HARD Model's parameters α, β, and γ reflect the influences of the different roles a node can play in *TNG*.

Based on the generated weights for the topics in the navigation graph, we can select the nodes with highest weights to represent the topics of most interest to the learner forming a user interest model, *UIM*. Table 8 shows structural weights of the topics in the "Mobile Applications" navigation graph. We can see how CRD model gives highest weight to the root topic "Mobile Application" that was first visited by the user compared to node "Netbean" being ranked top by HARD Model because of its higher aggregate hub weight coming from important hubs in the graph namely, "IDE" and "Mobile Application".

D. PERSONALIZATION MODULE

Personalized content recommendations for user i, PCR_i , can be obtained by mapping topics, l_i, from the user model of user i, UIM_i, to semantically similar learning resources or documents, d, in the inverted index of topics, IIT. Ranking of the personalized recommendations can be achieved using the weights of topics in the user model of user i, $W(uim_i)$ as:

$$PCR_i = W\left(\{uim_i : uim_i \in \mathrm{UIM}_i\}\right) * Sim\left(\{l_i : l_i \in \mathrm{UIM}_i\}, \{d_n : d_n \in IIT\}\right) \quad (6)$$

Therefore, learning documents with higher semantic similarities to topics in the user mode (*UIM*) are retrieved and form a set of ranked personalized content recommendations. Adaptation is accomplished through continuous update of *TNG* as well as *UIM* and, accordingly, the structural weights, hence, the personalized topics. The algorithm is explained in **Figure 6**.

TABLE 8. Structural weights of different nodes in the navigation pattern - program output.

	CRD	HARD
W(Mobile Application)	1.0	0.992
W(IDE)	0.5	0.994
W(Netbean)	0.333	1.0
W(SDK)	0.125	0.988
W(Java Support)	0.1	0.987
W(Platform)	0.125	0.981
W(Software)	0.166	0.972
W(Device Spec)	0.25	0.955

V. EVALUATION

The proposed approach aims 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:

1. Evaluating the quality of personalized content recommendations.
2. 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 which is commonly referred to as offline evaluation [89]. However, it is believed that evaluations of systems involving user models cannot and should not be separated from actual users [90]. 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 [91]. A major advantage of user studies is that they allow for collecting information about user interaction as well as testing different scenarios. Therefore, we designed user studies to evaluate the effectiveness of the proposed approach.

A. USER STUDY DESIGN

We implemented four websites with content from school Wikipedia. One website without any personalized support, two websites with personalized recommendations ranked using CRD and HARD models, and a website with recommendations generated based on popularity model as the baseline. We designed user studies following two main strategies.

The first strategy aims at evaluating the quality of personalized recommendations, so in our first treatment, **Figure 5**, we focus on three user groups. Two user groups using the websites with personalized recommendations using CRD and HARD models, and one user group using the website with recommendations generated using popularity model as a baseline.

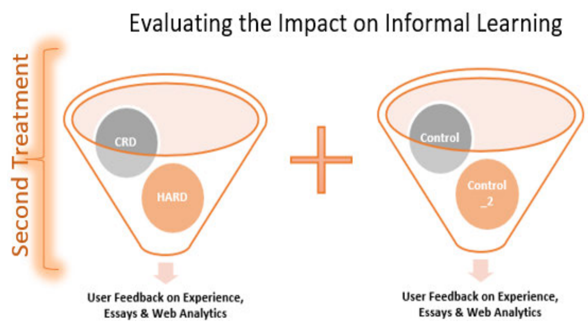


FIGURE 7. User Study - second treatment.

The second strategy aims at evaluating the impact of personalized recommendations on informal learning, so in our second treatment, **Figure 7**, we focus on four user groups. Two user groups using the websites without recommendations, Control and Control_2, and two user groups using the website with personalized recommendations based on CRD and HARD models.

B. TECHNOLOGICAL FRAMEWORK

To run our user studies, we developed four web-based encyclopedias equipped with user navigation's tracking and analysis algorithms, the proposed personalized content recommendation engine using CRD and HARD models, and popularity-based recommendation engine as a baseline. Our online test encyclopedias are listed in Table 9. The four 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. **Figure 8** shows screenshots from the website.

TABLE 9. Test websites.

Website URL	Type
www.theknowledge.site	No personalized support
www.hardtest.site	Personalized content recommendations ranked using HARD model
www.crdtest.site	Personalized content recommendations ranked using CRD model
www.basetest.site	Recommendations based on popularity

C. EVALUATION METRICS

1) METRICS TO EVALUATE THE QUALITY OF RECOMMENDER SYSTEM

To evaluate the effectiveness of the proposed PCRF, we use the rank-based Mean Average Precision, $MAP@k$, score to quantify recommendation quality at different ranks, k. Generally, $MAP@k$ quantifies the precision at the system level by calculating the mean of the average precision scores for a set of queries at different ranks up to k.

In our experiments, we use $MAP@k$ to calculate mean average precision scores for a set of users, U, in a user group using the same system. Hence, $MAP@k$ is calculated as:

$$\text{MAP@k (U)} = \frac{1}{|\text{U}|} * \left[\sum_{u=1}^{u=|U|}\left[\frac{1}{m} * \sum_{k=1}^{m} P@k\right]\right]$$

In this equation, $P@k$, denotes the precision at rank k for an individual user. For example, if user u_1 received a set of three recommendations and found the first two to be relevant and the third one to be irrelevant such that the user rating matrix is [1,1,0], where one indicates relevant and zero indicates irrelevant, then $P@1 = (1/1) = 1$, $P@2 = (2/2) = 1$, and $P@3 = (2/3) = 0.67$. Then, the average precision up to a rank $k = m$ for a single user is calculated as $AP@k = \frac{1}{m} * \sum_{k=1}^{m} P@k$. So, for user u_1, $AP@3 = [1/3^*(1+1+0.67)] = 0.89$. Finally, the mean of the average precisions of all users in a user group is calculated to quantify the recommendations quality at the system level for that user group, $MAP@k\,(U)$.

2) METRICS TO EVALUATE THE IMPACT OF RECOMMENDATION ON INFORMAL LEAENING

In our evaluation of informal learning we use three types of metrics: user-centric qualitative metrics to evaluate the user-perceived effectiveness of the personalized recommendations, objective educational metrics to evaluate the impact

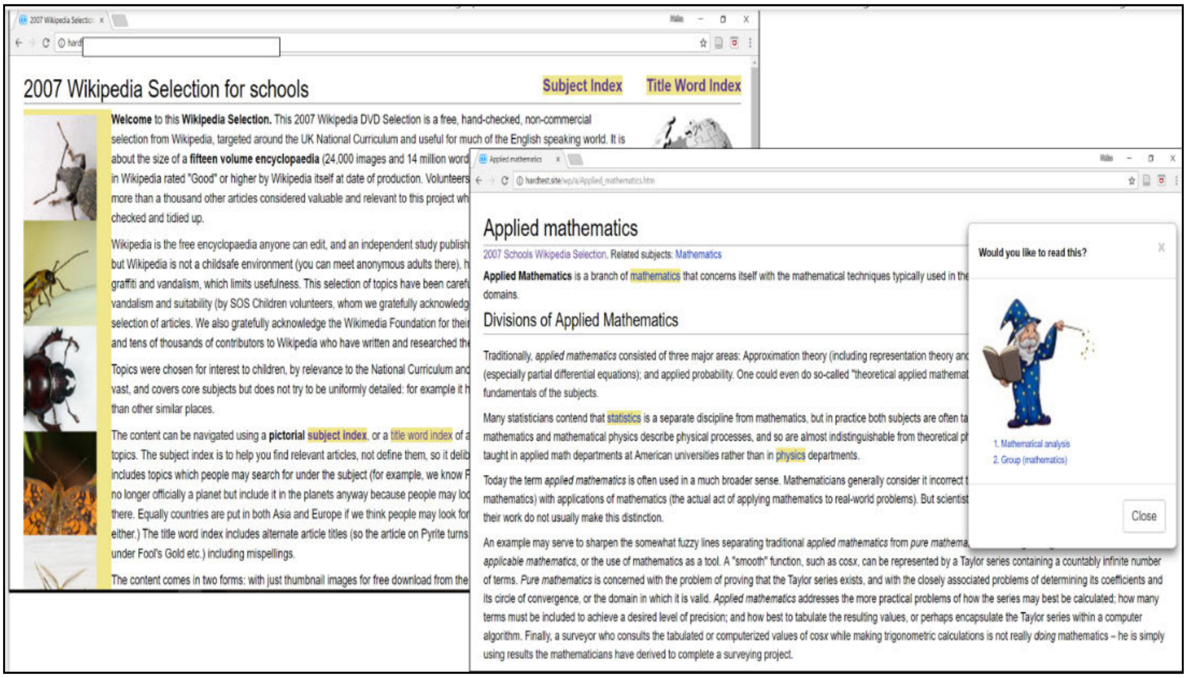


FIGURE 8. Screenshot from the test environment.

of recommendations on learning, and web analytics to get an insight into learners' focus and attention during the experiment.

For the ***user-centric qualitative metrics***, we evaluate two metrics that have been commonly used in the literature [92]:

1) *Perceived accuracy or relevance*: how much the recommendations match the users' interests, preferences, and tastes.
2) *Overall users' satisfaction*: the global users' feeling of the experience with the RS.

For ***educational metrics***, we focus on conceptual knowledge assessment as we are evaluating informal learning. That is, we are not following a curriculum or predefined learning outcomes upon which we can evaluate learners. Knowledge assessment allows measuring the general outcomes of learning and determines the effectiveness of the learning process. As knowledge structures cannot be observed directly, various indirect methods are used instead. Concept maps (CM) are one of such methods [93]. Therefore, to evaluate informal learning, we design a conceptual knowledge assessment rubric adapted from concept map-based rubrics.[2] The rubric we use 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 such as number of correct concepts used, complexity of concepts, number of relationships between concepts, the ability of learners to explain some comparisons between concepts, …etc. the full rubric is available in our EDUCON paper [28]. Finally, we use ***web analytics data*** to analyze the general navigational patterns of each user group. We examine visited pages and topics' frequencies to find out whether a certain test group is focused, distracted, or not focused on the main topic of experiment.

D. 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. This selection of topics has been carefully chosen, tidied up, and checked for vandalism and suitability for school students. The content can be navigated using a pictorial subject index, or a title word index of all topics.

E. DATA COLLECTION TECHNIQUES

We use multiple data collection tools. 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 the topic of space. Additionally, we run tracking scripts to collect navigation-related data.

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

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

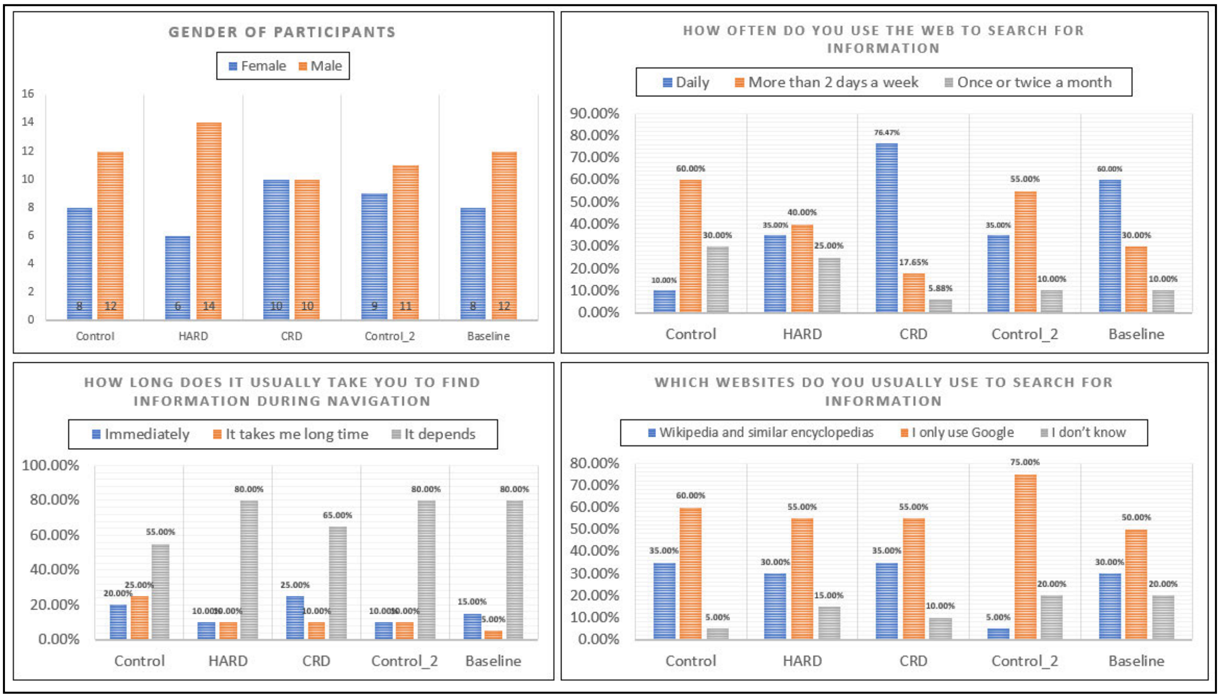


FIGURE 9. Demographics of test groups.

F. PARTICIPANTS

Experiments were conducted at a local private school teaching the UK National Curriculum. All year-five students were invited to participate in the experiments. Therefore, all participants' ages range between nine and ten years old. Consent forms were sent to interested students' parents to allow their children to participate in the experiments. A total of one hundred students from year-five participated in the experiments. Students were randomly assigned into five test groups each composed of twenty students. Test groups are: Control, HARD, CRD, Baseline, and Control_2. We had balanced participation from both male and female students. All participants use the internet to search for information at different levels of usage. Most of the students use either google or Wikipedia to search for information, hence, participants are familiar with web search and are familiar with the technological environment of the experiment. Demographics of participants per test group are summarized in **Figure 9**.

Test groups underwent two different treatments following the two strategies explained earlier. Details are explained in following sections.

G. PROCEDURE

A writing challenge was announced among year-five 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 at term three of the school year by then the participants had covered enough material related to "Space" as part of their science subject. We confirmed this information from teachers to ensure participants' familiarity with the topic of the experiments as 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 experiments took place on five consecutive days in the middle of the school day during the second break hour. The variable factors were limited to website setups in terms of recommendations' logic as explained earlier. All test session followed the same structure as explained in **Figure 10**.

H. FIRST TREATMENT OF THE USER STUDY – ASSESSING THE QUALITY OF RECOMMENDATION SYSTEMS

Three user groups were selected to evaluate the quality of the proposed recommender system. These are: CRD, Baseline,

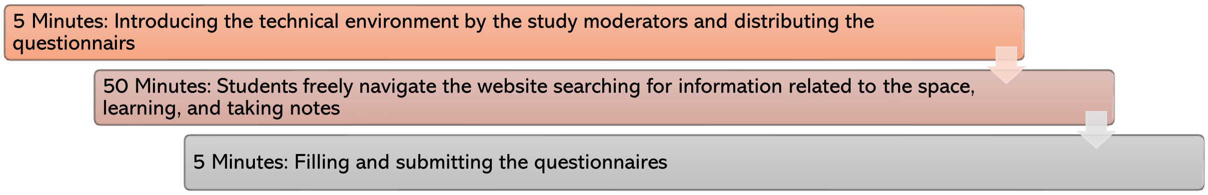


FIGURE 10. **Test session procedure.**

and HARD. Each user group had twenty students. We asked the students to evaluate the relevance of recommendations at two times during the test session. ***The first time*** is at the beginning of the test session at which the students would have at least visited one page, hence, the size of navigation graphs is between one and five, 1 =< |TNG| =< 5. The students are instructed to give their first feedback five times. That is, when the recommendation list contains one recommendation, P@1, then two recommendations, P@2, then three, P@3, and up to five recommendations, P@5. We designed the recommendations' pop up window in such a way that displays increasing number of recommendations at the beginning of the test session. That is, it displays one recommendation, then two, then three, and up to five, so as not to confuse the users. ***The second time*** the students had to evaluate the relevance of recommendations was towards the end of their test session where the navigation graph size would have increased above five, $5 \ll |TNG|$. Again, at that time the students were instructed to give their feedback at five different times as they would have done at the beginning of the test session. We record students' feedback on recommendations' relevance to evaluate the ***precision*** as well as ***adaptivity***. As explained earlier, users of similar information-oriented websites tend to exhibit an exploratory behavior and are likely to change interest during their navigation. In that sense, a successful recommender system should not only recommend relevant topics but also promptly adapt to changes in users' interest.

We use the rank-based mean average precision, *MAP@k*, as a metric because it gives good evaluation of both relevance as well as accuracy of ranking at the system level. We hypothesize that the three systems' *MAP@K* scores will not be equal. We use one-way ANOVA for multiple means to measure the statistical significance of our results at alpha level 5%, $\alpha = 0.05$. All results are statically significant with $P\text{-}Value = 0.0$ (P-Value < 0.05).

Results of First Treatment: Results of our evaluation reveal that indeed the three recommendation systems generate recommendations at different levels of precision over the test sessions and differ in their adaptivity. At the beginning of the test session, as shown in **Figure 11**, CRD-based recommendations starts as the most precise among all other systems with $MAP@1 = 100\%$ at the first rank compared with $MAP@1 = 85\%$ and $MAP@1 = 0.0\%$ for HARD and the Baseline respectively. However, as the users continue navigation, CRD fails to promptly adapt to changes in users' interests and its precision continues to drop until it reaches 80.35% compared with HARD and the Baseline which both exhibit better adaptability to changes in user interests. **Figure 11** shows that up to rank five, with number of topics equals five, HARD model consistently maintains reasonable precision with *MAP@k* scores ranging between 85% and 91.25%. The Baseline which does not implement any personalization logic starts so imprecise because it displays recommendations that are popular on the website which are apparently not relevant to the test topic. Yet, as users continue to navigate and click on relevant topics during the test session, it starts to display some relevant recommendations that had received higher number of visits by the users in the current test session.

Analyzing the performance of the three systems towards the end of the test session as shown in **Figure 12**, HARD-based recommendations turn to be the most precise and the most adaptive with *MAP@k* scores ranging between 100% and 86.4%. HARD system exhibited consistent performance in terms of precision throughout the test session. Whereas, CRD system's performance dropped significantly towards the end of the test session with MAP@K scores ranging between 27.5% and 47.4%. Baseline system's performance continue to improve towards the end of the test session by accommodating the current trend in user interests, but with much less precision compared with HARD or CRD.

Considering exemplary users' feedback ('0' for irrelevant and '1' for relevant) from CRD group, TABLE 10, we can see that for user, U15, who seems to be determined from the beginning on his/her topic of interest, CRD gives very precise recommendations repeatedly. However, for user, U3, who seems to be unsure about the topic of interest from the beginning, CRD fails to adapt to changes in interest. This could be a result of the ranking logic of CRD that places a very high weight for the root node which is the first node in the navigation graph of the user. Hence, if the user is not very clear about his/her target right from the beginning and is rather exploring some topics searching for the main topic of interest, which is the typical case for information-oriented websites' users, CRD might not be very successful in delivering precise recommendations at the top of the recommendation list. On the other hand, if we look at two exemplary users on the HARD website, user, U4, who seems to be very focused from the beginning of his/her navigation, and user, U9, who seems to be changing interests during navigation session. We find that HARD model immediately accommodates the changes and generates precise recommendations to user, U9,

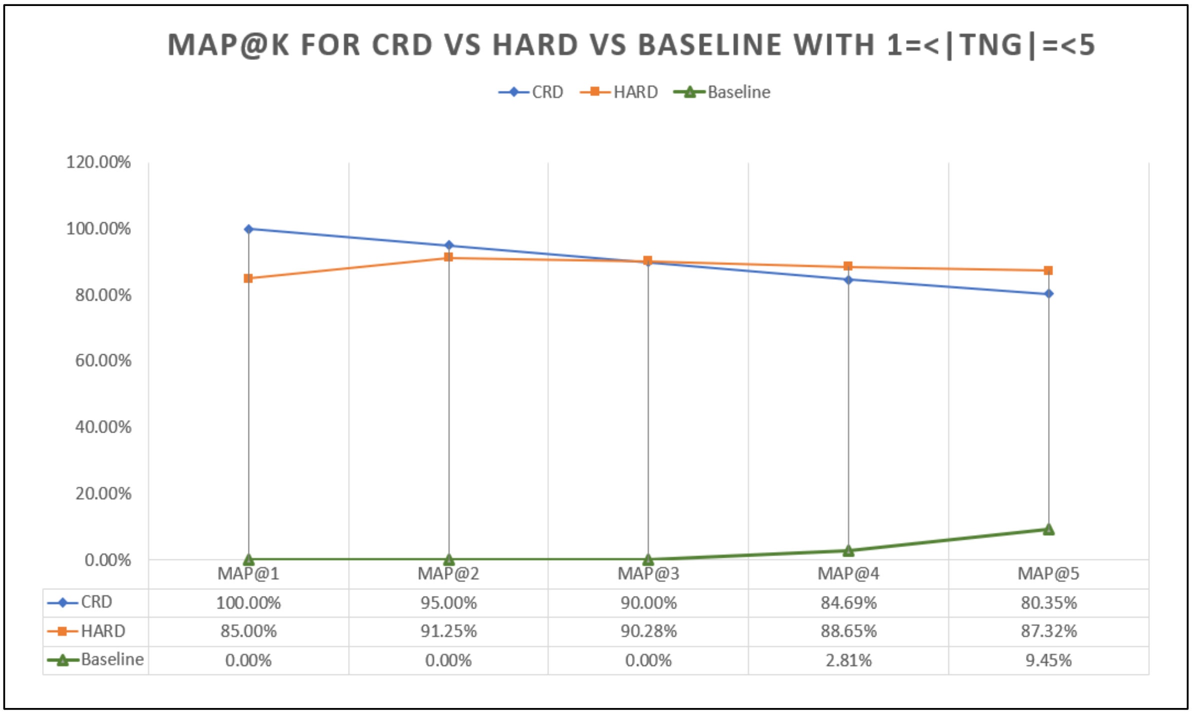


FIGURE 11. Cross systems MAP@K at the beginning of the test session.

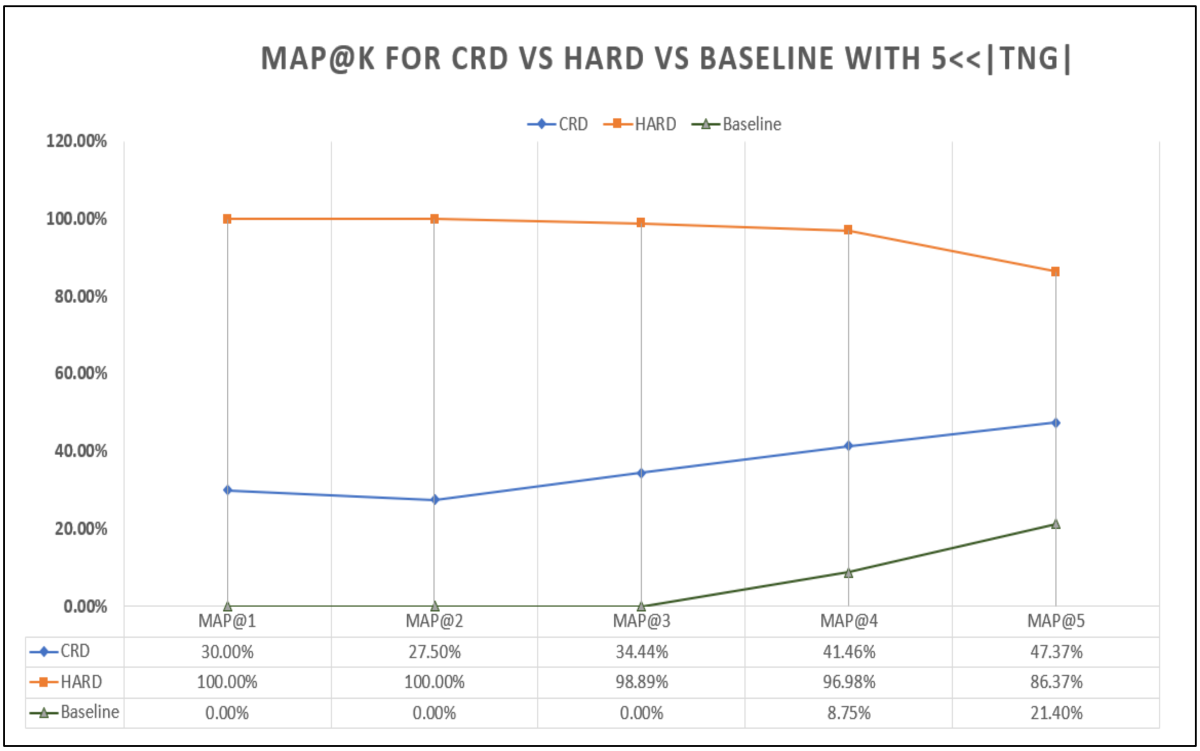


FIGURE 12. Cross systems MAP@K at the end of the test session.

with reasonable precision at the beginning of the navigation session, then becomes very precise towards the end of the navigation session as the user's interest becomes more well-defined giving comparable experience to both users: U4, who started with clear interests, and, U9, who started a bit unsure as we can see from the user feedback, TABLE 11.

TABLE 10. Exemplary user feedback for CRD system.

CRD User Group

1=<|TNG|=<5

Recommendations@K	U3					U15				
Recommendations@1	1					1				
Recommendations@2	1	1				1	1			
Recommendations@3	0	1	1			1	1	1		
Recommendations@4	0	0	1	1		1	1	1	1	
Recommendations@5	0	0	1	1	1	1	1	1	1	0

P@K	U3	U15
P@1	100.00%	100.00%
P@2	100.00%	100.00%
P@3	66.67%	100.00%
P@4	50.00%	100.00%
P@5	60.00%	80.00%

AP@K	U3	U15
AP@1	100.00%	100.00%
AP@2	100.00%	100.00%
AP@3	88.89%	100.00%
AP@4	79.17%	100.00%
AP@5	75.33%	96.00%

5<<|TNG|

Recommendations@K	U3					U15				
Recommendations@1	0					1				
Recommendations@2	0	0				1	1			
Recommendations@3	0	0	1			1	1	1		
Recommendations@4	0	0	1	1		1	1	1	1	
Recommendations@5	0	0	1	1	1	1	1	1	1	0

P@K	U3	U15
P@1	0.00%	100.00%
P@2	0.00%	100.00%
P@3	33.33%	100.00%
P@4	50.00%	100.00%
P@5	60.00%	80.00%

AP@K	U3	U15
AP@1	0.00%	100.00%
AP@2	0.00%	100.00%
AP@3	11.11%	100.00%
AP@4	20.83%	100.00%
AP@5	28.67%	96.00%

TABLE 11. Exemplary user feedback for HARD system.

HARD User Group

1=<|TNG|=<5

Recommendations@K	U4					U9				
Recommendations@1	1					0				
Recommendations@2	1	1				1	1			
Recommendations@3	1	1	1			0	0	1		
Recommendations@4	1	1	0	1		1	1	1	1	
Recommendations@5	1	1	1	1	0	1	1	1	0	1

P@K	U4	U9
P@1	100.00%	0.00%
P@2	100.00%	100.00%
P@3	100.00%	33.33%
P@4	75.00%	100.00%
P@5	80.00%	80.00%

AP@K	U4	U9
AP@1	100.00%	0.00%
AP@2	100.00%	50.00%
AP@3	100.00%	44.44%
AP@4	93.75%	58.33%
AP@5	91.00%	62.67%

5<<|TNG|

Recommendations@K	U4					U9				
Recommendations@1	1					1				
Recommendations@2	1	1				1	1			
Recommendations@3	1	1	1			1	1	1		
Recommendations@4	1	1	1	1		1	1	1	1	
Recommendations@5	1	1	1	1	0	1	1	1	0	1

P@K	U4	U9
P@1	100.00%	100.00%
P@2	100.00%	100.00%
P@3	100.00%	100.00%
P@4	100.00%	100.00%
P@5	80.00%	80.00%

AP@K	U4	U9
AP@1	100.00%	100.00%
AP@2	100.00%	100.00%
AP@3	100.00%	100.00%
AP@4	100.00%	100.00%
AP@5	96.00%	96.00%

I. SECOND TREATMENT OF THE USER STUDY – ASSESSING THE IMPACT OF RECOMMENDATIONS ON INFORMAL LEARNING

Four user groups were selected to evaluate the impact of personalized recommendations on informal learning, namely, CRD, HARD, Control, and Control_2. We grouped the responses of CRD and HARD groups into "***with personalization***" group, and the responses of Control and Control_2 groups into ***"without personalization***" group. 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.

VI. RESULTS OF SECOND TREATMENT

1) 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 13 (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 13 (C).** In addition, considering the overall user satisfaction criteria, results show that more than 90% of the 40 users who used the encyclopedia

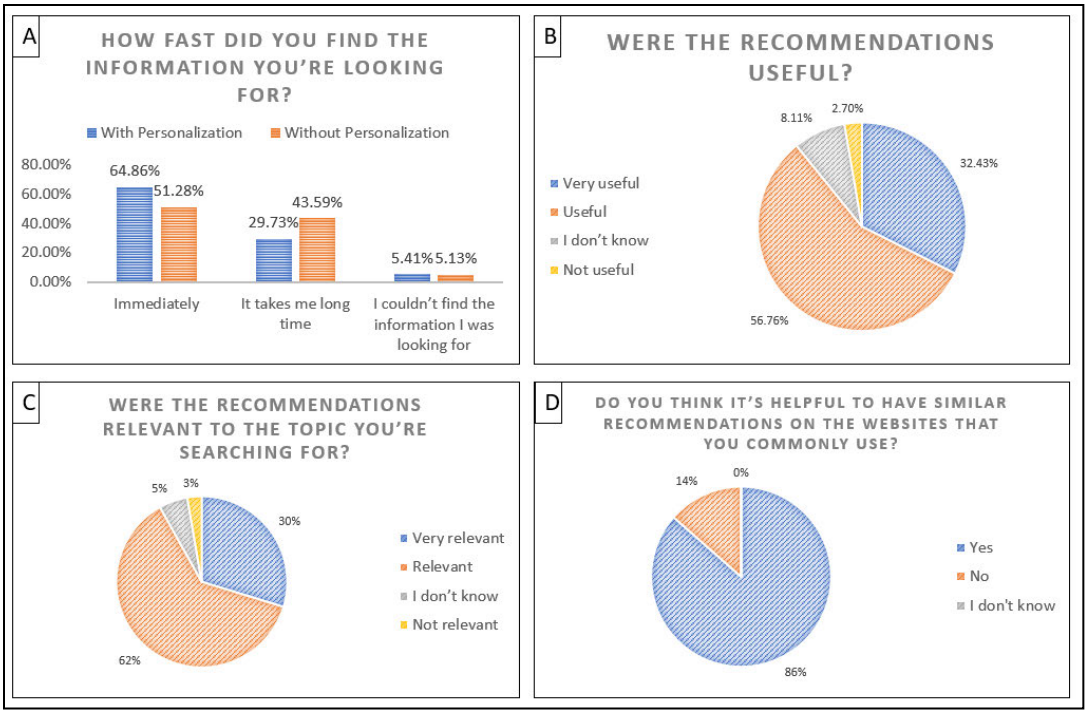


FIGURE 13. Results of user experience questionnaires.

TABLE 12. 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

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 13 (B)** and **(D)** respectively.

2) EVALUATING INFORMAL LEARNING

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 could achieve 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 with 10.0 for the students who used the encyclopedia without recommendations as shown in Table 12. The results are statistically significant at alpha level 5%, $\alpha = 0.05$, using *t-Test* for small independent samples with *P-Value* $= 0.0$ (P-Value < 0.05). Additionally, 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.

3) WEB ANALYTICS

Analysis of web analytics data revealed that users who used the encyclopedia with personalized support navigated more articles related to "Space" compared with 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. 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 of the users 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 12 that the students in the personalized support group submitted essays of various topics compared to the control group students who submitted limited number of topics, mainly focused on ''Black Hole'' and ''Neptune''. Overall, we can 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.

VII. CONCLUSION

In this paper, we proposed an effective personalized content recommendation framework (PCRF). PCRF implements efficient structural recommendation model by integrating fuzzy thesauri with adaptive users' interest models generated using structural analysis of topical navigational graphs.

We designed user studies with two strategies to evaluate the quality of PCRF as well as the impact of personalized recommendations on informal learning from Wikipedia.

Our evaluation revealed that PCRF generates highly relevant recommendations that are adaptive to changes in user's interest using the HARD model with $MAP@k$ scores ranging between 100% and 86.4%.

In addition, evaluation of informal learning revealed that users who used Wikipedia with personalized support achieved higher scores on the conceptual knowledge assessment with average score of 14.9 compared with 10.0 for the students who used the encyclopedia without any recommendations. In addition, they could make use of a larger number of concepts and were able to make comparisons and state relations between concepts. Manual analysis of web analytics data show that users who used Wikipedia with personalized recommendations visited larger number of relevant pages compared to the control group, 644 vs 226 respectively. Results confirm that PCRF can support informal learning in information wikis.

PCRF can also be adapted to provide personalized recommendations in similar environments. For instance, it can be used to provide within-book recommendations. For example, section-level, chapter-level, or paragraph-level recommendations. PCRF can be used in library systems, social forums, expert blogs, corporate knowledge sharing systems.

In the future, PCRF can be improved to provide different categories of recommendations in a hybrid model. Additionally, we intend also to explore the possibility of designing an automated evaluation framework suitable to evaluate informal learning without the need for any type of assessments using web analytics data. Web analytics data can give valuable insights on users' focus, interests, and progress. We intend to explore possible data mining tasks that can be applied to web analytics data to give users feedback on their progress and help improve the quality of recommendations.

REFERENCES

[1] B. Bray and K. McClaskey, ''A step-by-step guide to personalize learning,'' *Learn. Lead. Technol.*, vol. 40, no. 7, pp. 12–19, 2013.

[2] S. B. Merriam, R. S. Caffarella, and L. M. Baumgartner, *Learning in Adulthood: A Comprehensive Guide*. Hoboken, NJ, USA: Wiley, 2012.

[3] J. N. Streumer and M. Kho, ''The world of work-related learning,'' in *Work-Related Learning*. Dordrecht, The Netherlands: Springer, 2006, pp. 3–49.

[4] J. Misko, ''Combining formal, non-formal and informal learning for workforce skill development,'' ERIC, Paris, France, 2008.

[5] B. Zheng, M. Niiya, and M. Warschauer, ''Wikis and collaborative learning in higher education,'' *Technol., Pedagogy Educ.*, vol. 24, no. 3, pp. 357–374, 2015.

[6] D. L. Neumann and M. Hood, ''The effects of using a wiki on student engagement and learning of report writing skills in a University statistics course,'' *Australas. J. Educ. Technol.*, vol. 25, no. 3, pp. 382–398, 2009.

[7] J. Kimmerle, J. Moskaliuk, and U. Cress, ''Using wikis for learning and knowledge building: Results of an experimental study,'' *J. Educ. Technol. Soc.*, vol. 14, no. 4, p. 138, 2011.

[8] A. Head and M. Eisenberg, ''How today's college students use Wikipedia for course-related research,'' *Peer Rev. J. Internet*, vol. 15, no. 3, pp. 1–15, 2010.

[9] Wikimedia. (2019). *Wikimedia Statistics*. Accessed: Aug. 2019. [Online]. Available: https://stats.wikimedia.org/v2/#/all-wikipedia-projects

[10] I. Zukerman and D. W. Albrecht, ''Predictive statistical models for user modeling,'' *User Model. User-Adapted Interact.*, vol. 11, nos. 1–2, pp. 5–18, 2001.

[11] D. Albert and C. Hockemeyer, ''Applying demand analysis of a set of test problems for developing adaptive courses,'' in *Proc. Int. Conf. Comput. Educ.*, Dec. 2002, pp. 69–70.

[12] R. K. Sawyer, *The Cambridge Handbook of the Learning Sciences*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[13] D. N. Chin, ''KNOME: Modeling what the user knows in UC,'' in *User Models in Dialog Systems*. Berlin, Germany: Springer, 1989, pp. 74–107.

[14] H. P. Nascimento, L. B. Marques, D. G. de Souza, F. M. Salgado, and R. Q. Bessa, ''A AIED game to help children with learning disabilities in literacy in the Portuguese language,'' in *Proc. SBGames*, 2012, pp. 1–10.

[15] O. C. Santos, ''Emotions and personality in adaptive e-Learning systems: An affective computing perspective,'' in *Emotions and Personality in Personalized Services*. Cham, Switzerland: Springer, 2016, pp. 263–285.

[16] A. Klašnja-Milićević, B. Vesin, M. Ivanović, and Z. Budimac, ''E-Learning personalization based on hybrid recommendation strategy and learning style identification,'' *Comput. Edu.*, vol. 56, no. 3, pp. 885–899, 2011.

[17] S. A. Petersen and J.-K. Markiewicz, ''PALLAS: Personalised language learning on mobile devices,'' in *Proc. IEEE Int. Conf. Wireless, Mobile, Ubiquitous Technol. Educ.*, Mar. 2008, pp. 52–59.

[18] A. Mitrovic, M. Mayo, P. Suraweera, and B. Martin, ''Constraint-based tutors: A success story,'' in *Proc. Int. Conf. Ind., Eng. Other Appl. Appl. Intell. Syst.*, 2001, pp. 931–940.

[19] G. Hagerty and S. Smith, ''Using the Web-based interactive software ALEKS to enhance college algebra,'' *Math. Comput. Educ.*, vol. 39, no. 3, pp. 183–194, 2005.

[20] P. Brusilovsky and E. Millán, ''User models for adaptive hypermedia and adaptive educational systems,'' in *The Adaptive Web*. Berlin, Germany: Springer, 2007, pp. 3–53.

[21] N. Manouselis, H. Drachsler, R. Vuorikari, H. Hummel, and R. Koper, ''Recommender systems in technology enhanced learning,'' in *Recommender Systems Handbook*. Boston, MA, USA: Springer, 2011, pp. 387–415.

[22] H. Drachsler, K. Verbert, O. C. Santos, and N. Manouselis, ''Panorama of recommender systems to support learning,'' in *Recommender Systems Handbook*. Boston, MA, USA: Springer, 2015, pp. 421–451.

[23] J. Xiao, M. Wang, B. Jiang, and J. Li, ''A personalized recommendation system with combinational algorithm for online learning,'' *J. Ambient Intell. Hum. Comput.*, vol. 9, no. 3, pp. 667–677, 2018.

[24] W. Sriurai, P. Meesad, and C. Haruechaiyasak, "Recommending related articles in Wikipedia via a topic-based model," in *Proc. IICS*, 2006, pp. 1–10.
[25] A. A. A. Adline and G. S. Mahalakshmi, "A recommender system for wiki pages: Usage based rating approach," in *Proc. Int. Conf. Recent Trends Inf. Technol. (ICRTIT)*, Apr. 2012, pp. 553–558.
[26] C. C. Aggarwal, "Structural recommendations in networks," in *Recommender Systems*. Cham, Switzerland: Springer, 2016, pp. 309–334.
[27] M. Schwarzer, M. Schubotz, N. Meuschke, C. Breitinger, V. Markl, and B. Gipp, "Evaluating link-based recommendations for Wikipedia," in *Proc. 16th ACM/IEEE-CS Joint Conf. Digit. Libraries*, Jun. 2016, pp. 191–200.
[28] H. Ismail and B. Belkhouche, "Evaluating the impact of personalized content recommendations on informal learning from Wikipedia," in *Proc. IEEE Global Eng. Educ. Conf. (EDUCON)*, Apr. 2019, pp. 943–952.
[29] G. C. Rodi, V. Loreto, and F. Tria, "Search strategies of Wikipedia readers," *PLoS ONE*, vol. 12, no. 2, 2017, Art. no. e0170746.
[30] R. West and J. Leskovec, "Automatic versus human navigation in information networks," in *Proc. ICWSM*, 2012, pp. 1–8.
[31] D. Helic, "Analyzing user click paths in a Wikipedia navigation game," in *Proc. 35th Int. Conv. MIPRO*, May 2012, pp. 374–379.
[32] Y. Biletskiy, H. Baghi, I. Keleberda, and M. Fleming, "An adjustable personalization of search and delivery of learning objects to learners," *Expert Syst. Appl.*, vol. 36, no. 5, pp. 9113–9120, 2009.
[33] M. K. Khribi, M. Jemn, and O. Nasraoui, "Automatic recommendations for E-learning personalization based on Web usage mining techniques and information retrieval," in *Proc. 8th IEEE Int. Conf. Adv. Learn. Technol. (ICALT)*, Jul. 2008, pp. 241–245.
[34] M. Gaeta, F. Orciuoli, and P. Ritrovato, "Advanced ontology management system for personalised e-Learning," *Knowl.-Based Syst.*, vol. 22, no. 4, pp. 292–301, 2009.
[35] T. Di Noia, R. Mirizzi, V. C. Ostuni, D. Romito, and M. Zanker, "Linked open data to support content-based recommender systems," in *Proc. 8th Int. Conf. Semantic Syst.*, 2012, pp. 1–8.
[36] Y. Ogawa, T. Morita, and K. Kobayashi, "A fuzzy document retrieval system using the keyword connection matrix and a learning method," *Fuzzy Sets Syst.*, vol. 39, no. 2, pp. 163–179, 1991.
[37] R. Yerra and Y.-K. Ng, "Detecting similar HTML documents using a fuzzy set information retrieval approach," in *Proc. IEEE Int. Conf. Granular Comput.*, Jul. 2005, pp. 693–699.
[38] I. Garcia and Y.-K. Ng, "Eliminating redundant and less-informative RSS news articles based on word similarity and a fuzzy equivalence relation," in *Proc. 18th IEEE Int. Conf. Tools Artif. Intell. (ICTAI)*, Nov. 2006, pp. 465–473.
[39] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, Jun. 1965.
[40] C. Dictionary, *Cambridge Advanced Learner's Dictionary*. Cambridge, U.K.: Klett Ernst Verlag GmbH, 2008.
[41] M. S. Bernstein, B. Suh, L. Hong, J. Chen, S. Kairam, and E. H. Chi, "Eddi: Interactive topic-based browsing of social status streams," in *Proc. 23rd Annu. ACM Symp. User Interface Softw. Technol.*, 2010, pp. 303–312.
[42] E. J. Glover, S. Lawrence, M. D. Gordon, W. P. Birmingham, and C. L. Giles, "Web search—Your way," *Commun. ACM*, vol. 44, no. 12, pp. 97–102, 2001.
[43] E. J. Glover, S. Lawrence, W. P. Birmingham, and C. L. Giles, "Architecture of a metasearch engine that supports user information needs," in *Proc. 8th Int. Conf. Inf. Knowl. Manage.*, 1999, pp. 210–216.
[44] D. Kelly, K. Gyllstrom, and E. W. Bailey, "A comparison of query and term suggestion features for interactive searching," in *Proc. 32nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2009, pp. 371–378.
[45] J. Peltonen, J. Strahl, and P. Floréen, "Negative relevance feedback for exploratory search with visual interactive intent modeling," in *Proc. 22nd Int. Conf. Intell. User Interfaces*, 2017, pp. 149–159.
[46] J.-W. Ahn and P. Brusilovsky, "Adaptive visualization of search results: Bringing user models to visual analytics," *Inf. Vis.*, vol. 8, no. 3, pp. 167–179, 2009.
[47] J. C. Chang, N. Hahn, A. Perer, and A. Kittur, "SearchLens: Composing and capturing complex user interests for exploratory search," in *Proc. IUI*, 2019, pp. 1–12.
[48] H. Lieberman, "Letizia: An agent that assists Web browsing," in *Proc. ACM IJCAI*, 1995, pp. 924–929.
[49] A. Moukas, "Amalthaea information discovery and filtering using a multiagent evolving ecosystem," *Appl. Artif. Intell.*, vol. 11, no. 5, pp. 437–457, 1997.
[50] H. Sakagami and T. Kamba, "Learning personal preferences on online newspaper articles from user behaviors," *Comput. Netw. ISDN Syst.*, vol. 29, nos. 8–13, pp. 1447–1455, 1997.
[51] M. J. Pazzani, J. Muramatsu, and D. Billsus, "Syskill & Webert: Identifying interesting Web sites," in *Proc. AAAI/IAAI*, vol. 1, 1996, pp. 54–61.
[52] L. Chen and K. Sycara, "WebMate: A personal agent for browsing and searching," *Agents*, vol. 98, pp. 132–139, May 1998.
[53] F. A. Asnicar and C. Tasso, "ifWeb: A prototype of user model-based intelligent agent for document filtering and navigation in the world wide Web," in *Proc. 6th Int. Conf. User Modeling*, 1997, pp. 1–5.
[54] A. Stefani and C. Strapparava, "Exploiting NLP techniques to build user model for Web sites: The use of WordNet in SiteIF project," in *Proc. 2nd Workshop Adapt. Syst. User Modeling WWW*, 1999.
[55] G. Gentili, A. Micarelli, and F. Sciarrone, "InfoWeb: An adaptive information filtering system for the cultural heritage domain," *Appl. Artif. Intell.*, vol. 17, nos. 8–9, pp. 715–744, 2003.
[56] E. W. De Luca, A. Nürnberger, and O. Von-Guericke, "Ontology-based semantic online classification of documents: Supporting users in searching the Web," in *Proc. Eur. Symp. Intell. Technol.*, 2004, pp. 1–9.
[57] A. Sieg, B. Mobasher, and R. D. Burke, "Learning ontology-based user profiles: A semantic approach to personalized Web search," *IEEE Intell. Inform. Bull.*, vol. 8, no. 1, pp. 7–18, Nov. 2007.
[58] V. Challam, S. Gauch, and A. Gauch, "Contextual search using ontology-based user profiles," in *Proc. Large Scale Semantic Access Content (Text, Image, Video, Sound)*, 2007, pp. 612–617.
[59] F. Tanudjaja and L. Mui, "Persona: A contextualized and personalized Web search," in *Proc. 35th Annu. Hawaii Int. Conf. Syst. Sci.*, Jan. 2002, pp. 1232–1240.
[60] M. Hearst, *Search User Interfaces*. Cambridge, U.K.: Cambridge Univ. Press, 2009.
[61] M. Szomszor, H. Alani, I. O. Cantador, K. O'Hara, and N. Shadbolt, "Semantic modelling of user interests based on cross-folksonomy analysis," in *Proc. Int. Semantic Web Conf.*, 2008, pp. 632–648.
[62] J. Teevan, C. Alvarado, M. S. Ackerman, and D. R. Karger, "The perfect search engine is not enough: A study of orienteering behavior in directed search," in *Proc. ACM SIGCHI Conf. Hum. Factors Comput. Syst.*, 2004, pp. 415–422.
[63] R. W. White and S. M. Drucker, "Investigating behavioral variability in Web search," in *Proc. ACM 16th Int. Conf. World Wide Web*, 2007, pp. 21–30.
[64] G. G. Chowdhury, *Introduction to Modern Information Retrieval*. London, U.K.: Facet Publishing, 2010.
[65] R. Burke, "Hybrid recommender systems: Survey and experiments," *User Model. User-Adapted Interact.*, vol. 12, no. 4, pp. 331–370, Nov. 2002.
[66] P. Lops, M. de Gemmis, and G. Semeraro, "Content-based recommender systems: State of the art and trends," in *Recommender Systems Handbook*. Boston, MA, USA: Springer, 2011, pp. 73–105.
[67] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen, "Collaborative filtering recommender systems," in *The Adaptive Web*. Berlin, Germany: Springer, 2007, pp. 291–324.
[68] G. Adomavicius and A. Tuzhilin, "Context-aware recommender systems," in *Recommender Systems Handbook*. Boston, MA, USA: Springer, 2011, pp. 217–253.
[69] R. Burke, "Knowledge-based recommender systems," *Encyclopedia Library Inf. Syst.*, vol. 69, no. Suppl. 32, pp. 175–186, 2000.
[70] F. Abel, I. I. Bittencourt, N. Henze, D. Krause, and J. Vassileva, "A rule-based recommender system for online discussion forums," in *Proc. Int. Conf. Adapt. Hypermedia Adapt. Web-Based Syst.*, 2008, pp. 12–21.
[71] R. J. Mooney and L. Roy, "Content-based book recommending using learning for text categorization," in *Proc. 5th ACM Conf. Digit. Libraries*, 2000, pp. 195–204.
[72] S. Philip, P. Shola, and A. Ovye, "Application of content-based approach in research paper recommendation system for a digital library," *Int. J. Adv. Comput. Sci. Appl.*, vol. 5, no. 10, pp. 1–4, 2014.

[73] Y. Al Murtadha, M. N. Sulaiman, N. Mustapha, and N. I. Udzir, "Improved Web page recommender system based on Web usage mining," in *Proc. 3rd Int. Conf. Comput. Inform. (ICOCI)*, 2011, pp. 1–5.

[74] M. Ge, C. Delgado-Battenfeld, and D. Jannach, "Beyond accuracy: Evaluating recommender systems by coverage and serendipity," in *Proc. 4th ACM Conf. Rec. Syst.*, 2010, pp. 257–260.

[75] Y. C. Zhang, D. Ó. Séaghdha, D. Quercia, and T. Jambor, "Auralist: Introducing serendipity into music recommendation," in *Proc. 5th ACM Int. Conf. Web Search Data Mining*, 2012, pp. 13–22.

[76] R. Torres, S. M. McNee, M. Abel, J. A. Konstan, and J. Riedl, "Enhancing digital libraries with TechLens+," in *Proc. 4th ACM/IEEE-CS Joint Conf. Digit. Libraries*, 2004, pp. 228–236.

[77] S. S. Sundar, A. Oeldorf-Hirsch, and Q. Xu, "The bandwagon effect of collaborative filtering technology," in *Proc. Extended Abstr. Hum. Factors Comput. Syst. (CHI)*, 2008, pp. 3453–3458.

[78] H.-N. Kim and S. El Saddik, "Personalized PageRank vectors for tag recommendations: Inside FolkRank," in *Proc. 5th ACM Conf. Rec. Syst.*, 2011, pp. 45–52.

[79] S. Lee, S.-I. Song, M. Kahng, D. Lee, and S.-G. Lee, "Random walk based entity ranking on graph for multidimensional recommendation," in *Proc. 5th ACM Conf. Rec. Syst.*, 2011, pp. 93–100.

[80] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *J. Assoc. Inf. Sci. Technol.*, vol. 58, no. 7, pp. 1019–1031, 2007.

[81] M. Al Hasan and M. J. Zaki, "A survey of link prediction in social networks," in *Social Network Data Analytics*. Boston, MA, USA: Springer, 2011, pp. 243–275.

[82] D. Cosley, D. Frankowski, L. Terveen, and J. Riedl, "SuggestBot: Using intelligent task routing to help people find work in wikipedia," in *Proc. 12th Int. Conf. Intell. User Interfaces*, 2007, pp. 32–41.

[83] E. Wulczyn, R. West, L. Zia, and J. Leskovec, "Growing Wikipedia across languages via recommendation," in *Proc. 25th Int. Conf. World Wide Web*, 2016, pp. 975–985.

[84] E. Huang, H. J. Kim, and H. Ko, "Task recommendation on Wikipedia," *Data Process.*, 2010.

[85] D. Leake, A. Maguitman, and T. Reichherzer, "Understanding knowledge models: Modeling assessment of concept importance in concept maps," in *Proc. 26th Annu. Conf. Cogn. Sci. Soc.*, 2004, pp. 1–7.

[86] A. Kao and S. R. Poteet, *Natural Language Processing and Text Mining*. Springer, 2007.

[87] M. F. Porter, "An algorithm for suffix stripping," *Program*, vol. 14, no. 3, pp. 130–137, 1980.

[88] D. B. Leake and A. Maguitman, "Combining concept mapping with CBR: Towards experience-based support for knowledge modeling," in *Proc. 14th Int. Florida Artif. Intell. Res. Soc. Conf.*, 2011, pp. 1–5.

[89] C. C. Aggarwal, *Recommender Systems*. New York, NY, USA: Springer, 2016.

[90] D. N. Chin, "Empirical evaluation of user models and user-adapted systems," *User Model. User-Adapted Interact.*, vol. 11, nos. 1–2, pp. 181–194, 2001.

[91] P. Cremonesi, F. Garzotto, S. Negro, A. V. Papadopoulos, and R. Turrin, "Looking for 'good' recommendations: A comparative evaluation of recommender systems," in *Proc. IFIP Conf. Hum.-Comput. Interact.*, Berlin, Germany, 2011, pp. 152–168.

[92] P. Pu, L. Chen, and R. Hu, "A user-centric evaluation framework for recommender systems," in *Proc. 5th ACM Conf. Rec. Syst.*, 2011, pp. 157–164.

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