11-2019

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

Heba M Ismail

جامعة الإمارات العربيـة المتحدة
United Arab Emirates University

United Arab Emirates University

College of Information Technology

A FRAMEWORK FOR PERSONALIZED CONTENT RECOMMENDATIONS TO SUPPORT INFORMAL LEARNING IN MASSIVELY DIVERSE INFORMATION WIKIS

Heba M. Ismail

This dissertation is submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Under the Supervision of Professor Saad Harous

November 2019

Declaration of Original Work

I, Heba M. Ismail, the undersigned, a graduate student at the United Arab Emirates University (UAEU), and the author of this dissertation entitled "*A Framework for Personalized Content Recommendations to Support Informal Learning in Massively Diverse Information Wikis*", hereby, solemnly declare that this dissertation is my own original research work that has been done and prepared by me under the supervision of Professor Saad Harous, in the College of Information Technology at UAEU. This work has not previously been presented or published or formed the basis for the award of any academic degree, diploma or a similar title at this or any other university. Any materials borrowed from other sources (whether published or unpublished) and relied upon or included in my dissertation have been properly cited and acknowledged in accordance with appropriate academic conventions. I further declare that there is no potential conflict of interest with respect to the research, data collection, authorship, presentation and/or publication of this dissertation.

Student's Signature: ____________ Date: 18/Nov/2019

Advisory Committee

1) Advisor: Professor Saad Harous

Title: Professor

Department of Computer Science and Software Engineering

College of Information Technology

2) Member: Professor M. Adel Serhani

Title: Professor

Department of Information Systems & Security

College of Information Technology

3) Member: Dr Mohammed Mehedy Masud

Title: Associate Professor

Department of Information Systems & Security

College of Information Technology

Approval of the Doctorate Dissertation

This Doctorate Dissertation is approved by the following Examining Committee Members:

1) Advisor: Professor Saad Harous

Title: Professor

Department of Computer Science and Software Engineering

College of Information Technology

Signature ______________________ Date 17/11/2019

2) Member: Dr Elarbi Badidi

Title: Associate Professor

Department of Computer Science and Software Engineering

College of Information Technology

Signature ______________________ Date 17/11/2019

3) Member: Dr Amir Ahmad

Title: Assistant Professor

Department of Information Systems &Security

College of Information Technology

Signature ______________________ Date 17/11/2019

4) Member (External Examiner): Professor Luiz Fernando Capretz

Title: Professor

Department of Electrical and Computer Engineering

University of Western Ontario, Canada

Signature ______________________ Date 17/11/2019

This Doctorate Dissertation is accepted by:

Dean of the College of Information Technology: Professor Khaled Shuaib

Signature ______________________ Date 22-12-2019

Dean of the College of Graduate Studies: Professor Ali Almarzoqi

Signature Ali Hassan ______________________ Date 2/2/2020

Copy 1 of 5

Abstract

Personalization has proved to achieve better learning outcomes by adapting to specific learners' needs, interests, and/or preferences. Traditionally, most personalized learning software systems focused on formal learning. However, learning personalization is not only desirable for formal learning, it is also required for informal learning, which is self-directed, does not follow a specified curriculum, and does not lead to formal qualifications. Wikis among other informal learning platforms are found to attract an increasing attention for informal learning, especially Wikipedia. The nature of wikis enables learners to freely navigate the learning environment and independently construct knowledge without being forced to follow a predefined learning path in accordance with the constructivist learning theory. Nevertheless, navigation on information wikis suffer from several limitations. To support informal learning on Wikipedia and similar environments, it is important to provide easy and fast access to relevant content. Recommendation systems (RSs) have long been used to effectively provide useful recommendations in different technology enhanced learning (TEL) contexts. However, the massive diversity of unstructured content as well as user base on such information-oriented websites poses major challenges when designing recommendation models for similar environments. In addition to these challenges, evaluation of TEL recommender systems for informal learning is rather a challenging activity due to the inherent difficulty in measuring the impact of recommendations on informal learning with the absence of formal assessment and commonly used learning analytics. In this research, a personalized content recommendation framework (PCRF) for information wikis as well as an evaluation framework that can be used to evaluate the impact of personalized content recommendations on informal learning from wikis are proposed. The presented recommendation framework models learners' interests by continuously extrapolating topical navigation graphs from learners' free navigation and applying graph structural analysis algorithms to extract interesting topics for individual users. Then, it integrates learners' interest models with fuzzy thesauri for personalized content recommendations. Our evaluation approach encompasses two main activities. First, the impact of personalized recommendations on informal learning is evaluated by assessing conceptual knowledge in users' feedback. Second, web analytics data is analyzed to get an insight into users' progress and focus throughout the test session. Our evaluation revealed that PCRF generates highly relevant recommendations that are adaptive to changes in user's interest using the HARD model with rank-based mean average precision (MAP@k) scores ranging between 100% and 86.4%. In addition, evaluation of informal learning revealed that users who used Wikipedia with personalized support could achieve higher scores on conceptual knowledge assessment with average score of 14.9 compared to 10.0 for the students who used the encyclopedia without any recommendations. The 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. In addition, they were also able to make use of a larger number of concepts and were able to make comparisons and state relations between concepts.

Keywords: Information Filtering, Information Wikis, Informal Learning, Personalized Content Recommendations, Recommender Systems, Wikipedia, Evaluation, Web Analytics.

Title and Abstract (in Arabic)

نموذج لتوصيات المحتوى الشخصي لدعم التعلم غير الرسمي في شبكات المعلومات الضخمة

الملخص

لقد أثبت التخصيص تحقيق نتائج تعليمية أفضل من خلال التكيف مع احتياجات واهتمامات و/أو تفضيلات المتعلمين المحددة. عادة ما تركز معظم أنظمة برامج التعلم الشخصية على التعلم الرسمي. ومع ذلك، فإن تخصيص التعلم ليس مرغوبًا فيه فقط للتعلم الرسمي، بل هو مطلوب أيضًا للتعلم غير الرسمي، الموجه ذاتيًا، ولا يتبع منهجًا محددًا ولا يؤدي إلى مؤهلات رسمية. أشارت عدد من الدراسات والإحصاءات إلى أن الويكي من بين منصات التعلم غير الرسمية الأخرى يجذب اهتمام متزايد للتعلم غير الرسمي، وخاصة ويكيبيديا. تمكّن طبيعة الويكي المتعلمين من التصفح بحرية في بيئة التعلم وبناء المعرفة بشكل مستقل دون إجبارهم على اتباع مسار تعليمي محدد مسبقًا وفقًا لنظرية التعلم البنائية. ومع ذلك، يعاني التصفح على شبكات الويكي من مشكلات متعددة. لذلك لدعم التعلم غير الرسمي على ويكيبيديا والبيئات المشابهة، من المهم توفير وصول سهل وسريع إلى المحتوى ذي الصلة. منذ فترة طويلة تستخدم أنظمة التوصية (RSs) لتقديم توصيات مفيدة بشكل فعال في سياقات التعلم المحسن التكنولوجية المختلفة (TEL). ومع ذلك، فإن التنوع الهائل للمحتوى غير المهيكل بالإضافة إلى قاعدة المستخدمين على مثل هذه المواقع يفرض تحديات كبيرة عند تصميم نماذج توصية لبيئات مماثلة. بالإضافة إلى هذه التحديات، يعتبر تقييم أنظمة التوصية للتعلم غير الرسمي مهمة صعبًا جداً نظرًا للصعوبة المتأصلة في قياس تأثير التوصيات على التعلم غير الرسمي مع عدم وجود تقييم رسمي أومؤشرات أداء التعلم الشائعة الاستخدام. في هذا البحث، نقترح نموذج فعال لعمل توصيات المحتوى المخصصة (PCRF) يتناسب مع بيئة الويكي بالإضافة إلى إطار للتقييم يمكن استخدامه لتقييم تأثير توصيات المحتوى المخصص على التعلم غير الرسمي من الويكي. يعمل النموذج المقترح على دراسة اهتمامات الدارسين من خلال الاستقراء المستمر لخرائط التصفح وتطبيق خوارزميات التحليل الهيكلي لخرائط التصفح لاستخراج الموضوعات المهمة للمستخدمين الفرديين. بعد ذلك، يدمج نماذج اهتمامات الدارسين مع المواضيع ذات الصلة لعمل توصيات المحتوى المخصصة. يشمل نهج التقييم الخاص بنا نشاطين رئيسيين. أولاً، نقوم بتقييم تأثير التوصيات

المخصصة على التعلم غير الرسمي من خلال تقييم المعارف المكتسبة في تعليقات المستخدمين. ثانيًا نقوم بتحليل بيانات إحصاءات الويب للحصول على نظرة ثاقبة على تقدم المستخدمين و تركيزهم خلال جلسة الاختبار. كشف تقييمنا أن PCRF يقدم توصيات عالية الدقة تتكيف مع التغييرات في اهتمامات المستخدم باستخدام نموذج HARD الذي تتراوح معدل دقته بين MAP@k=٪100 و MAP@k=%86.4. بالإضافة إلى ذلك، كشف تقييم التعليم غير الرسمي أن المستخدمين الذين استخدموا ويكيبيديا مع دعم شخصي يمكنهم تحقيق درجات أعلى في تقييم المعارف بمتوسط 14.9 مقارنة بـ 10.0 للطلاب الذين استخدموا الموسوعة دون أي توصيات. يوضح تحليل بيانات إحصاءات الويب أن المستخدمين الذين استخدموا ويكيبيديا مع توصيات مخصصة زاروا عددًا أكبر من الصفحات ذات الصلة مقارنة بمجموعة التحكم، 644 مقابل 226 على التوالي. بالإضافة إلى ذلك، كانوا أيضًا قادرين على الاستفادة من عدد أكبر من المفاهيم وكانوا قادرين على إجراء مقارنات وشرح علاقات بين المفاهيم.

مفاهيم البحث الرئيسية: فلترة المعلومات، ويكي المعلومات، التعلم غير الرسمي، توصيات المحتوى الشخصي، أنظمة التوصية، ويكيبيديا، التقييم، تحليلات الويب.

Acknowledgements

All Praise is due to ALLAH S.W.T the Almighty, for giving me the blessing, the strength, the chance, and endurance to complete this study.

First, I would like to express my deepest gratitude to my Professor and advisor Professor Saad Harous for his continuous and unlimited support of my PhD study and related research since I joined the PhD program at UAEU until I finally successfully defended my dissertation. His sincere encouragement, patience, motivation, and immense knowledge helped me throughout my study. I’m especially thankful for his guidance and very constructive feedback during the writing of this dissertation. I’m very proud and lucky to be his student.

In addition to my advisor, I would like to thank the rest of my dissertation committee: Dr. Elarabi Badidi, Dr. Amer Ahmad, and Professor Luiz Fernando, for their insightful comments, and very constructive feedback that made my dissertation better.

Very special thanks and gratitude also go to my advisory committee: Professor M.Adel Serhani and Dr. M.Mehedy Masud who provided very constructive comments and valuable feedback on my PhD proposal. I’m especially very indebted to Professor Adel Serhani who kindly offered sincere advice and unconditional support. I consider myself very lucky to have him in my advisory committee as well as to have completed my last few semesters in my PhD study during his duty as the assistant dean for research and graduate studies. His genuine cooperation and guidance were major success factors in my PhD study.

My appreciation also goes out to all my professors who taught me courses during my PhD study, in addition to Professor Saad Harous and Dr M.Mehedy Masud, I'm very thankful to Professor Nazar Zaki, Dr Hany Al Ashwal, Professor Najah Abed AbuAli, and Dr Salah Bouktif. Thanks for your advices and great knowledge.

I am also deeply indebted to my great, amazing mother, husband, and siblings, for their unconditional love and continuous support, genuine encouragement, sincere prayers, and unlimited trust. Without you all, my PhD journey would have been impossible.

My acknowledgements are incomplete without thanking my lovely friend, Haliemeh Sweidan, for her unlimited emotional support and cooperation that helped me in the most difficult times of my study. My deepest thanks also go out to my amazing friends, Reem Alsaadi, Suha Jlala, and Ray Albarazi, for their support and cooperation throughout my study. Without you all, my PhD journey would have been very difficult.

Dedication

To my wonderful parents who raised me up to be the person I am today

To my precious and joyful son, Saeed Ismail

Table of Contents

List of Tables

List of Figures

List of Abbreviations

CB	Content-based Recommendation
CF	Collaborative Filtering
EWC	English Wikipedia Clickstream
LDA	Latent Dirichlet Allocation
LOD	Linked Open Data
PCRF	personalized content recommendation framework
RS	Recommender System
SFV	semantic feature vectors
TEL	Technology Enhanced Learning
TF-IDF	Term Frequency Inverse Document Frequency

Chapter 1: Introduction

1.1 Motivation

Personalization in various contexts is seen to provide different types of gains [1-4]. In learning contexts, personalization has proved to achieve better learning outcomes by adapting to specific learners' needs, interests, and/or preferences [3, 5, 6]. Traditionally, the majority of personalized learning software systems focused on formal learning [7-12]. Formal learning software systems attempt to model formal education normally delivered at schools or colleges by defining specific learning content aligned with a curriculum, learning outcomes, and assessments. However, learning personalization is not only desirable for formal learning, it is also required for informal learning which is self-directed, does not follow a specified curriculum, and does not lead to a formal qualification [13]. Studies of informal learning reveal that up to 90% of adults are engaged in hundreds of hours of informal learning [14]. It has also been estimated that up to 70% of learning in the workplace is informal [15]. Many research works recently investigated how online information sharing platforms such as wikis and blogs can contribute to informal learning [16-18]. Wikis among other informal learning platforms are recently experiencing an increasing demand for informal learning, especially Wikipedia [19-23]. As of today, Wikipedia contains more than 157,000,000 articles in 302 languages among which 37,000,000 articles are in English [24]. This makes Wikipedia one of the greatest sources of knowledge on the web. Additionally, a study that targeted high school students at six campuses in the U.S. between April and May 2009, had shown that up to 82% of students in higher education turn to Wikipedia to give their research a jump start, and 76% of students use Wikipedia to find the meaning of terms in certain topics [25]. Therefore, an effective personalization approach that supports informal learning from wikis is desirable.

1.2 Problem Overview

To support informal learning on diverse information wikis with 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 [26]. The set of user characteristics modeled in a user model 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 [27-30], emotions [31], preferences [32], and context [33] 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 as seen in tutoring systems [34], or online courses [35]. The fact that these formal systems deliver a very specific content for a very specific learner group 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, often referred to as adaptive hypermedia, that dealt with huge bulk of diverse information such as online encyclopedias [36].

Considering the context of information wikis and specifically Wikipedia's context, one method to specify user interest is through keyword-based search. However, in many cases, users may fail to identify representative keywords. Another method to specify user interest is through hyperlinks. This method is powerful but may divert the user away from the main topic of interest. In addition, links mentioned in an article cannot fully cover all related articles in the whole corpus. One of the reasons is because there is no term describing related articles within the current article or simply because some links might not be working. Additionally, the vast diversity of content and user base poses major challenges on modeling users' interests. Typically, on massive information wikis, users do not belong to a specific age group or educational level. They do not also have common learning objectives. Individual users may in fact have multiple different objectives every time they use the wiki. Consequently, users' interests are diverse, changing, and do not generate a definite recurrent pattern. Therefore, an adaptive user-centric interest model is required to provide easy and fast access to relevant content on similar environments.

Recommendation systems (RSs) have long been used to effectively provide user-centric interest models and deliver useful recommendations in different technology enhanced learning (TEL) contexts [37,38]. TEL RSs have been used primarily to recommend additional learning resources within online courses or other learning environments making access to useful resources faster and easier [39]. Furthermore, TEL RSs can recommend to learners effective learning paths [40], or peers learners [41], which is a central recommendation task for distance education settings where learners usually feel isolated.

The most commonly used techniques for TEL RSs are collaborative filtering (CF), and content-based filtering (CB) [38]. CF approaches recommend items primarily based on similarities between users [42]. 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 [42]. However, converting unstructured text into feature vector representation eliminates essential latent semantic relationships that exist in original text. Additionally, in massively diverse environments, the size of items' feature space is likely to be very large resulting in highly sparse user and item profiles which is sometimes referred to as the "*curse of dimensionality*" problem [43]. Sparsity causes major accuracy issues. Reported research work in TEL RSs shows interesting results especially in online learning environments with focused learning objectives and well-defined learning content and learners' base. However, there remain some major challenges inherent in delivering recommendations for massively diverse unstructured content with a heterogeneous user base as seen in Wikipedia and similar environments.

Therefore, different variations of content-based recommendation models have been used to address these challenges. For example, Sriurai et al. [44] used the Latent Dirichlet Allocation (LDA) algorithm for topic-based recommendations, and Adline & Mahalakshmi [45] proposed an article quality framework to classify and recommend Wikipedia articles into readable, learnable, and referable format. Other researchers started to utilize new variations of search algorithms to deliver structural recommendations [46]. In structural recommendation techniques, content or/and users are represented using graphs. Graph search and ranking algorithms are then used to recommend nodes, links, or different combinations of both. A recent research study by Schwarzer et al.[47] 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.

On the other hand, The evaluation of recommender systems in general is a complicated task, because of i) the diversity of different measures that need to be considered, e.g. accuracy, novelty, scalability, serendipity [48], ii) the availability/unavailability and adequacy/inadequacy of benchmark datasets, and iii) the number of users that such evaluations may require. In addition to these factors, evaluation of TEL recommender systems for informal learning is quite a challenging activity due to the inherent difficulty in measuring the impact of recommendations on informal learning with the absence of formal assessment and commonly used learning analytics.

To this end, since we are addressing personalized informal learning, there is a need to model an effective personalized content recommendation framework for massively diverse information wikis such as Wikipedia as well as evaluate the impact of recommendations on informal learning. Therefore, our research objectives are:

+ To model and develop an effective personalized content recommendation framework to support informal learning in massively diverse information wikis.

+ To design an evaluation framework suitable to assess the impact of personalized recommendations on informal learning in information wikis.

In view of these objective, there are number of challenges that we need to address. In the following section we briefly describe the research challenges.

1.3 Challenges Inherent in Designing Recommendations for Massively Diverse Information Wikis

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

1.3.1 Learner Modeling Challenges

Typically, on wiki environments such as Wikipedia, users do not follow consistent patterns of interest over a long period of time. Rather, users are more likely to change their interests over sessions or sometimes within a single session. In recent research, Rodi et al.[49] analyzed the English Wikipedia Clickstream (EWC) dataset gathered during 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 massively diverse information wikis, it is important to account for *changes or evolvements* in the user interest.

Additionally, West and Leskovec [50], have compared human navigation in information networks such as Wikipedia with that of software agents and found that humans, when navigating within an information 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, i.e. information accessible from the current page, possibly using links. Therefore, to help users make the best use of local information, it is important to give them local access to *relevant* information through personalized recommendations.

However, articles in massively diverse information wikis form a scale-free network [51]. 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 out 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 barely based on links or references.

1.3.2 Learning Content Processing Challenges

A variety of learning content representations can be used in personalized learning software systems. In addition to learning objects [2,3], ontologies [52], or more recently Linked Open Data (LOD) [53], a huge amount of learning content on the web is available in the form of unstructured free text. Typically, this is the kind of text we find in blogs, wikis, forums, and social media websites.

Unstructured texts suffer from several complications. Unlike structured data or formal knowledge representations, there are no predefined features and attributes with well-defined values. Unstructured text may contain any number of various words. Additionally, converting unstructured text into feature vector representation, especially in massively diverse environments, results in sparsity and curse of dimensionality problem [43]. Even in the simplest setting, it is likely to have a sparse matrix with thousands of rows and columns most of which are zeros [43].

Several approaches were proposed in the literature to account for semantics in the text. Most of these approaches can be classified into two categories: contextual approaches, and conceptual approaches. Conceptual approaches of semantic analysis rely on external semantic knowledgebases such as ontologies and semantic networks. Conceptual semantic approaches are limited by their underlying knowledgebases and require large amount of manual efforts during the knowledgebase creation and validation phase. In contextual approaches, statistical analysis of the relationships between terms in the text are analyzed. These relationships are mainly co-occurrences. These approaches tend to be more flexible given the possibility of automation. Hence, given the massively diverse nature of Wikipedia's unstructured content, an effective contextual semantic analysis approach capable of alleviating the sparsity challenge is required to support personalized content recommendations.

1.4 Research Questions

Considering the research objectives and challenges we need to answer the following research questions:

- Q1: How can the changing learner interest be modeled effectively and adaptively in massively diverse information wikis?
- Q2: Which recommendation model can effectively deliver personalized content recommendations on massively diverse information wikis?
- Q3: Which evaluation approach can be used to assess the impact of the proposed approach on informal learning?

1.5 Methodology

In this research, a personalized content recommendation framework (PCRF) for Wikipedia content in addition to an evaluation framework that can be used to evaluate the impact of personalized recommendations on informal learning are designed and developed. User studies are designed to evaluate the effectiveness of the proposed approach.

The PCRF first captures raw learning interests for every individual learner in a topical navigation graph (TNG) by tracking individual learning sessions. The learner navigation is modeled as a directed multigraph, TNG (V, E). Every vertex, V, in TNG corresponds to a topic, topics are 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. [54], 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 [55]. The resulting set of ranked and semantically relevant topics represents the final personalized content recommendations.

The proposed framework is composed of four main modules: session tracking, TNG analyzer, personalization, and semantic analysis modules. Figure 1 illustrates a high-level conceptualization of the proposed framework which was first presented at ACM UMAP18 [56]. 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.

The evaluation of informal learning encompasses two main activities. First, the impact of personalized recommendations on informal learning is evaluated by assessing conceptual knowledge in users' feedback. Second, web analytics data is analyzed to get an insight into users' progress and focus as well as propose an evaluation framework based on web analytics that can be used to evaluate informal learning on similar environments.

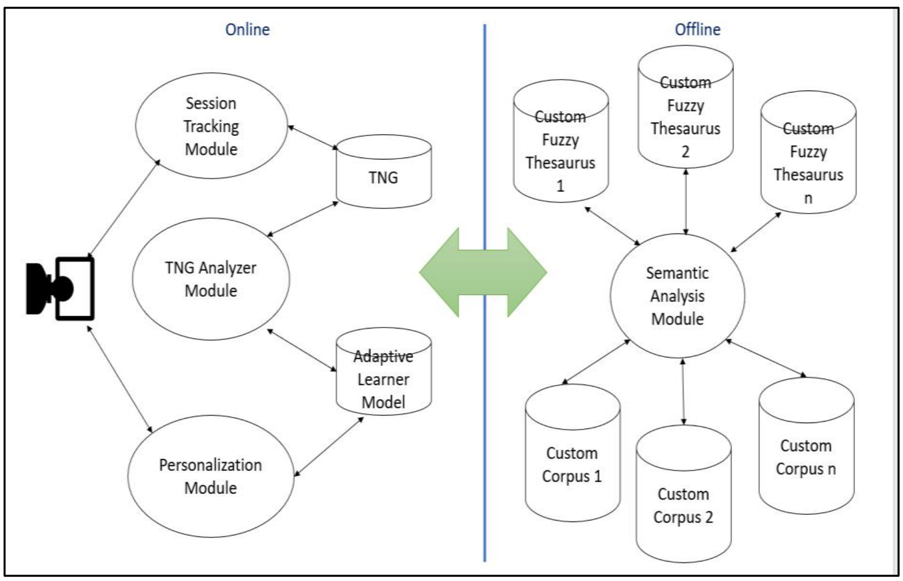


Figure 1: High-level conceptualization of the proposed personalized content recommendation framework

1.6 Dissertation Overview

Chapter 2 presents background knowledge that is fundamental for understanding the concepts, techniques, and methods used in this dissertation. Chapter 3 reviews state-of-the-art research work in learning personalization, user interest modeling, and research field recommender systems. Chapter 4 describes modeling user interests based on adaptive topical navigational graphs. Chapter 5 covers the semantic analysis module in detail. Chapter 6 describes the proposed framework. Chapter 7 introduces our evaluation approach. Finally, findings are summarized, and future directions are highlighted in Chapter 8.

1.7 Research Tasks and Summary of Contributions

To answer the research questions, the following research tasks are carried out:

- T1: Survey related work.

  There are many publications related to learning personalization, user modeling, and recommender systems. Extensive review of related work is conducted. As a result, major components of personalized learning systems, challenges, taxonomies of the field, and a reusable software architecture for personalized learning systems [57] are identified. Also, the shortcomings in commonly used interest modeling approaches and available recommender systems for Wikipedia are highlighted.

- T2: Model and develop an effective learner interest modeling approach adaptive to changing interests in massively diverse hypermedia environments.

  Based on the literature review, a user interest model based on adaptive topical navigational graphs is modeled. The proposed user interest model is personalized to individual users and is effective in capturing changes in user interests during navigation sessions. The proposed user interest modeling approach is explained in Chapter 5 as part of the full content personalization framework.

- T4: Model and develop an effective semantic analysis technique suitable for massively diverse unstructured text found in massively diverse information wikis.

Based on our literature review, an effective semantic analysis approach based on concepts from fuzzy set information retrieval model is modeled and developed. The proposed technique uses fuzzy thesauri to generate feature vector representations of different language units, i.e. words, topics etc. which can be used for text mining, recommendations, and other tasks involving the use of unstructured text. In massively diverse environments as Wikipedia, converting unstructured text into feature vector representation result in sparsity and curse of dimensionality problems with many rows and columns represented with zeros. This intern hinders the accuracy of semantic analysis. A very well-known text mining task that suffers from sparsity is Twitter sentiment analysis. We implement the proposed technique in the context of recommender system as well as Twitter sentiment analysis to assess the applicability of the proposed technique in multiple contexts. Our preliminary results in Twitter sentiment analysis using fuzzy set-based feature vectors are published in ISCMI16 [58], the complete Twitter Fuzzy Set-based Sentiment Analysis Framework and evaluations are published in Soft Computing Journal [59], and the semantic analysis tasks based on fuzzy thesauri related to recommender systems are accepted for publication in IEEE Access.

- T5: Model and develop a personalized content recommender system based on user’s navigation graph and fuzzy thesaurus.

  Using the proposed learner model and semantic analysis technique, an effective personalized content recommendation framework to support informal learning in massively diverse information wikis is modeled and developed. High-level conceptualization of the proposed framework is published in ACM UMAP18 [56].

Detailed design, implementation, and evaluation of the proposed framework is published in IEEE Access.

T6: Develop evaluation methods and metrics to Assess Informal Learning on wiki environments.

An approach to evaluate the impact of personalized recommendations on informal learning is proposed and developed. First, the impact of personalized recommendations on informal learning is evaluated by assessing conceptual knowledge in users' feedback. An assessment rubric is designed, adapted from concept map-based rubric for conceptual knowledge assessment, then, user studies are conducted and the impact of personalized recommendations on informal learning is evaluated. Second, web analytics data is analyzed to get an insight into users' progress and focus through-out the test sessions and an evaluation framework based on web analytics data is proposed. Results of conceptual knowledge assessment is published in EDUCON19 [60]. The proposed evaluation framework accepted for publication in iJEP Journal.

Chapter 2: Background

2.1 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." [61]

This definition opens the field of recommender systems to any application that computes a user-specific utility, covering many areas of applications.

To identify users' information needs and match these needs with useful items, researchers proposed several recommendation classes such as collaborative filtering and content-based filtering, as well as knowledge-based, citation-based, context-aware, and rule-based recommendations, and many more [62-66]. However, the following three classes are considered to be most appropriate for differentiating the approaches in the field of recommender systems in information-oriented websites:

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

2.1.1 Recommendation Classes

<u>1. Content-based</u>

Content-based filtering (CB) is one of the most extensively used and studied recommendation approaches [43]. A vital task of CB is the user modeling process, in which the interests of users are inferred from the items that users interacted with. "Items" are mostly textual, for instance books [67], research papers [68], or webpages [69]. "Interaction" is typically recognized through actions such as downloading, buying, authoring, or tagging an item. Items are represented by a content or document model containing the items' descriptive attributes which are commonly called features. Features are typically word-based, i.e. single words, phrases, n-grams, etc.

Typically, only the most descriptive features are used to model an item and users. These features are ideally weighted generating weighted feature vectors of items and users. 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 a suitable similarity measure, e.g. Cosine.

CB has several advantages. For instance, CB allows a more individual personalization so the recommender system can determine the best recommendations for each user individually, rather than be limited by what other like-minded users like. CB also requires less labor 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 [43].

2. Collaborative Filtering

The term "collaborative filtering" (CF) was coined in 1992 by Goldberg et al., who suggested that "information filtering can be more effective when humans are involved in the filtering process" [70]. However, the type of collaborative filtering known today was introduced two years later for the GroupLens project by Resnick et al. [71]. They assumed that users usually like what other like-minded users like, whereas two users were considered like-minded when they rated items similarly. Therefore, when like-minded users were identified, items that one user rated positively and not yet seen or rated by the other like-minded user, were recommended to the other user, and vice versa.

In contrast to CB, CF offers three advantages. First, CF is content independent, i.e. no complex item processing is required [63]. Second, because the ratings are done by humans either explicitly through ratings or likes and dislikes or implicitly through recurrent visits other navigational indicators, CF considers real quality assessments [63]. Finally, CF is supposed to provide serendipitous, i.e. surprising and unexpected, recommendations because recommendations are not based on item similarity but on user similarity [72], [73].

A major drawback, however, in CF is the “cold start problem,” which may occur in three situations [63]: new users with no rating or navigation history, new items that have not yet received any ratings or impressions from users, and new communities or disciplines. If a new user rates few or no items, the system cannot find like-minded users and therefore cannot provide recommendations. If an item is new in the system and has not yet been rated by at least one user, it cannot be recommended. In a new community, no users have rated items, so no recommendations can be made and as a result, the incentive for users to rate items is low.

Additionally, computational time complexity for CF algorithms tends to be higher than for CB [63]. Collaborative filtering in general is less scalable and requires more offline data processing than CB. 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. [74] point out that CF creates similar users and Sundar et al. [75] criticize that collaborative filtering dictates opinions. This drawback of CF conflicts with the massive diversity of Wikipedia's content and users. Finally, a key challenge in CF is synonymy [6]. Synonymy arises when an item is represented with two or more different names. In such cases, the recommender cannot identify whether the terms represent different items or the same item. For example, a collaborative filtering recommender system will treat "comedy movie" and "comedy film" differently. The diversity and variability of descriptive terms are much greater than commonly thought; hence, the extreme usage of synonym words reduces the performance of CF. In CF, item's contents are thoroughly overlooked, and the algorithms do not consider the latent association between items. However, considering information-oriented websites, semantic associations in the content are vital.

3. Structural Recommendation in Networks

Enormous amount of data can be organized in the form of a graph or a network. The Web itself is a huge network of Web pages. In recent years, many personalized conceptions of search have evolved, where the Web pages recommended to users are based on personal interests. Many search engine providers, such as Google, now provide the ability to determine personalized results. This problem is exactly equivalent to that of ranking nodes in networks with the use of personalized preferences [46]. These are referred to as structural recommendations as they are generated based on structural analysis of networks.

Several structural elements of a network can be recommended. Each of these different types of structural recommendation may have a different set of applications in different scenarios. The two major categories of structural recommendation models are: link-based recommendations, and node-based recommendations. Each one is explained in detail:

1. Node-based Recommendations: In this case, the quality of nodes is judged by their incoming links, and the personalized relevance of nodes is judged by their context. This problem is very closely related to that of search engines. A major observation is that the traditional perception of search in such engines does not distinguish between various users, and is therefore, not personalized to a specific user. In search engines, Web pages (or nodes in the Web graph) are ranked based on their authority and their content. Little emphasis is placed on the identity of the user performing the search. However, notions such as personalized PageRank [76], [77], were eventually developed that can tailor the results to various interests. These forms of personalization incorporate context into the ranking by modifying the traditional notion of PageRank with context-specific personalization [46].

2. Recommending links: In many social networks, such as Facebook, it is important to increase the connectivity of the network. Therefore, users are often recommended potential friends. This problem is equivalent to that of recommending potential links in a network [78]. Several ranking methods are used for link prediction. Additionally, matrix factorization methods can also be adapted to link prediction [79].

Structural recommendation model can be seen as the most suitable model to the context of the research problem given the possibility of incorporating contextual information, i.e. semantics, as well as adapting to changing user's interests inferred through structural analysis of users' generated navigation graphs. Our proposed approach is explained further in Chapter 5.

2.1.2 Evaluating Recommender Systems

When evaluating a recommender system, three experimental settings are expected: offline experiment, user studies and online experiment [42]. Figure 2 illustrates evaluation settings for RS. Each one is explained briefly in the following sections.

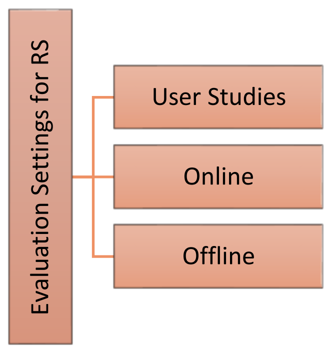


Figure 2: Classification of evaluation settings for RS

1. Offline evaluation

Offline evaluations typically measure the accuracy of a recommender system based on historical data, i.e. benchmark data, with a ground-truth [80]. Measures of precision at position n (P@n) is often used to express how many items of the ground-truth are recommended within the top n recommendations. Other common evaluation metrics include recall, F-measure, mean reciprocal rank (MRR), normalized discounted cumulative gain (nDCG), mean absolute error, and root mean square error. Offline evaluations are also sometimes used to evaluate aspects such as novelty or serendipity of recommendations [72]. This is the simplest evaluation settings of recommendation, but it requires representative benchmark data. Absence of historical data with ground truth inhibits the ability of using this type of evaluation.

2. Online evaluation

Online evaluations started in online advertising and e-commerce applications. They measure the acceptance rates of recommendations in *real-world* recommender systems. Acceptance rates are often measured by click-through rates (CTR), i.e. the ratio of clicked recommendations to displayed recommendations. For instance, if a recommender system displays 10,000 recommendations and 500 are clicked, the CTR is 5%. This method is time consuming and requires very large number of participants. It may last for months or years.

3. User studies

User studies typically measure user feedback through explicit ratings. Users receive recommendations generated by several recommendation methods, then they give explicit feedback on the recommendations' quality, and the approach with the highest average rating is considered most effective [42]. Subjects are typically asked to quantify their overall satisfaction with the recommendations or give a qualitative feedback through questionnaires. User studies are favored in user-centric designs [81]. A major advantage of user studies is that they allow for collecting information about user interaction as well as testing different scenarios. However, user studies are expensive to conduct, time consuming, and require very good design of the test environment, participants' selection criteria, and experimental variables identification.

2.2 Wikipedia

Wikis are interlinked web pages based on the hypertext system of storing and modifying information. Each page can store information and is easily viewed, edited, and commented on by other people using a web browser [20]. This nature of wikis enables learners to freely navigate the learning environment and independently construct knowledge without being forced to follow a predefined learning path in accordance with the constructivist learning theory [82].

A wiki is implemented using a wiki engine. A wiki engine is a form of content management system, but it differs from most other such systems in that the content is created without any defined owner, and wikis have little inherent structure, allowing structure to develop according to the needs of the users.

The online encyclopedia project Wikipedia is the most popular wiki-based website, and is one of the most widely viewed sites in the world, having been ranked in the top ten since 2007 to date [83].

2.2.1 Content and Users

Wikipedia is a multilingual, web-based, free-content encyclopedia project supported by the Wikimedia Foundation and based on a model of openly editable content. Wikipedia is populated collaboratively by largely anonymous volunteers who write without pay.

Since its creation on January 15, 2001, Wikipedia has grown rapidly into one of the largest reference websites, attracting 374 million unique visitors monthly as of September 2015 [84]. As of today, there are more than 157,000,000+ articles in 302 languages among which 37,000,000+ articles are in English (Figure 3 and Figure 4) [24]. This makes Wikipedia an attractive environment for informal learning.

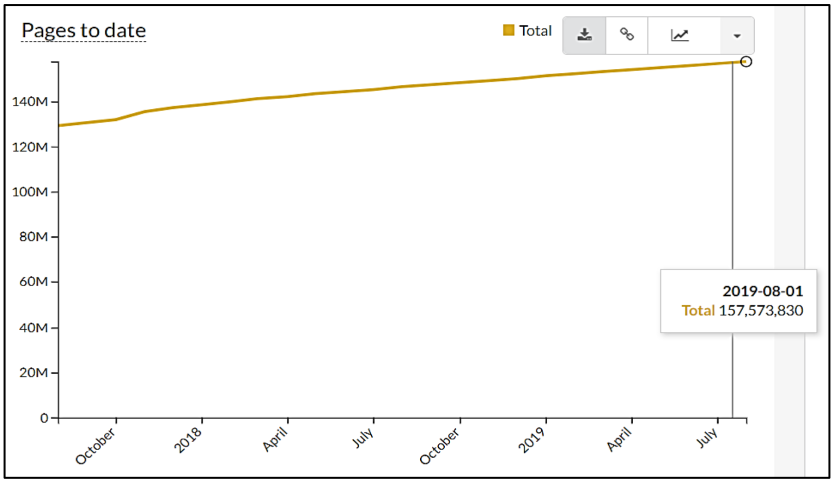


Figure 3: Wikipedia content up to 1st August 2019

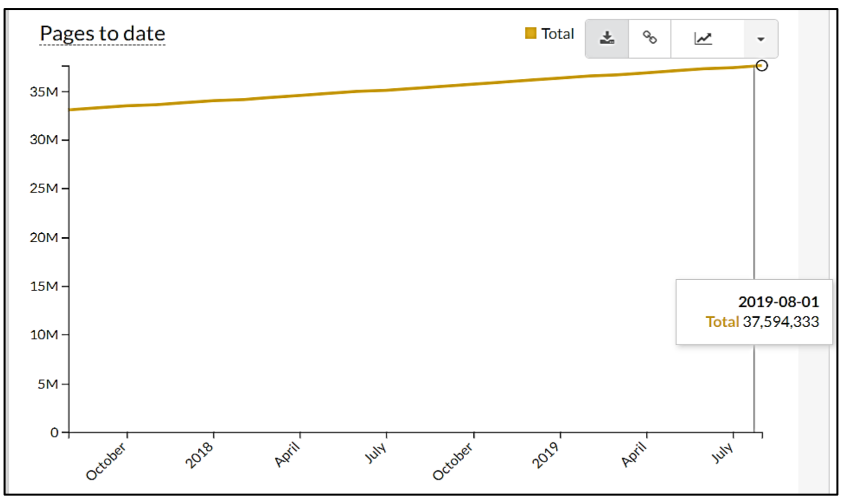


Figure 4: Wikipedia English content up to 1st August 2019

2.2.2 Structure

In his paper, Watts [85], defines "small world network" as a navigable network that is highly connected and in such a network each pair or almost each pair of nodes is connected by a short path. More formally, a "small world network" forms a scale-free network whose degree distribution follows a power law. Smaller number of nodes have the highest degree in the network. If you look at the power distribution (Figure 5) you can see a tail that is very condensed, the left tail, and a tail that is very sparse, that is the right tail. The nodes on the left tail with the highest connectivity are usually called hubs.

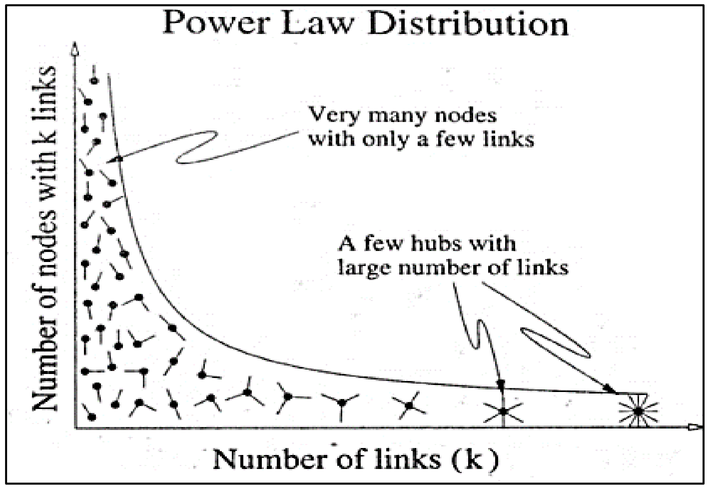


Figure 5: A Power distribution

Denis [51], analyzed Wikipedia's network structure and found that Wikipedia's articles were found to form a scale-free network (Figure 6a). That is few articles are highly connected and thus most commonly linked to other articles whereas many articles are poorly connected and thus relevant information can be missed out when recommending articles based on links only.

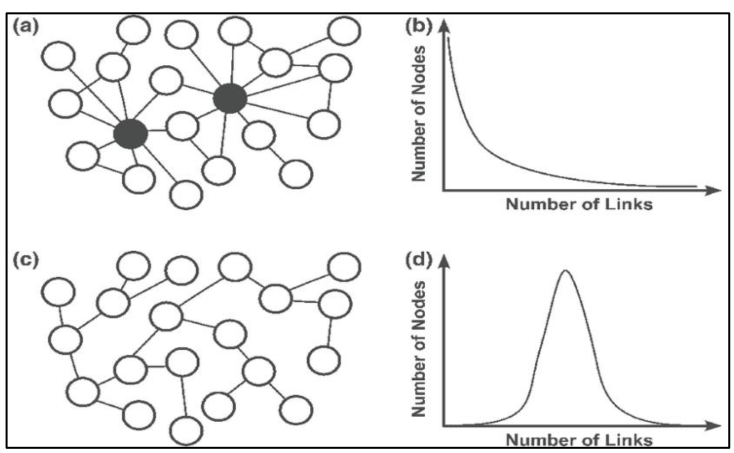


Figure 6: Scale-free (a) vs random network (c)

Chapter 3: Literature Review

This chapter introduces state-of-the-art research work related to learning personalization in general as well as to the specific areas of interest modeling, technology-enhanced learning recommender systems, and Wikipedia recommender systems. It starts by defining the main concepts in learning personalization from a software engineering perspective, then it moves to reviewing the different components of personalized learning software systems highlighting the different techniques used, features, challenges and limitations, and identifying where our research fits among other personalized learning software systems. Then, it highlights different aspects related to modeling users' interests in information-oriented website. Finally, it briefly introduces technology-enhanced recommender systems and focus on Wikipedia recommender systems.

3.1 Why Do We Need Learning Personalization?

Learners have always learned in their own unique and variable ways. However, teaching has traditionally followed a one-size-fits-all approach. Conventionally educators had followed a learning model called cohort-based model, that is characterized by relatively large numbers of students moving, as a group and at the same rate, through the curriculum, physical facilities, and teachers [86]. A major disadvantage of the cohort-based method, given that the model was designed specifically to serve students in groups, is that individual learning needs can never be fully addressed resulting in less effective education. Given that people think in different ways, have different preferences and learn at different paces, many psychologist and cognitive scientists stressed the importance of learning personalization for a more effective education [86]. Considerable educational changes have been made to address learners' personal differences. Nevertheless, given the many variable attributes of learning personalization, learning personalization could not be fully accomplished without technology. As proposed by the American Personalized Learning Initiative, personalized learning at its general sense requires not only a shift in the design of schooling, but also a leveraging of modern technologies. Personalization cannot take place at scale without technology [87]. In the following sections, we provide a brief overview of personalized learning software systems and highlight, where applicable, where our research problem fits.

3.2 Survey of Personalized Learning Software Systems

According to the U.S. Department of Education learning personalization is defined as: "Instruction is paced to learner's needs, tailored to learner's preferences, and tailored to the specific interests of different learners [5]." However, interpretations of different elements of the definition may vary widely depending on the context in which they are implemented [88]. We present in the following sections definitions and explanations of learning personalization specific to the technological context.

In order to limit the assumptions about personalized learning software systems, a precise explanation of the term ''learning personalization" in the context of software systems is given first. We define "Learning Experience" in a software system, adapted from Wang's [89], as the sequence of learning resource accesses, where resources refer to any learning resource that can be implemented in a software environment. For example, learning environments could be hypermedia environment, game environment, specialized simulated training environment, etc. Learning resources may include online courses, e-books, instructions, assessments, learning activities, and so on. Accordingly, personalized learning software systems are learning systems that tailor learning resources accesses within the software environment to a user model. In this context the user model reflects the needs, preferences, interests and pace of learning of an individual learner. We do not treat each aspect of the user characteristics separately, rather, a representative model of the user, i.e. learner, is used to accomplish the personalization process within a software system. Table 1 presents a list of definitions followed in this research, which can be considered as a glossary of learning personalization software systems.

Table 1: Glossary of learning personalization software systems

Term	*Definition*
Learning Experience	The sequence of learning resource accesses in a software learning environment
Software Learning Environment	Hypermedia environment, game environment, specialized training environment, etc.
Learning Resource	Any learning resource that can be implemented in a software environment such as online courses, e-books, instructions, assessments, game quests, and so on. These can be modelled using any knowledge representations such as learning objects, ontologies, linked open data, or data representations such as relational database, semi-structured data, or even unstructured plain text.
User Model	A software model reflecting the needs, preferences, interests and pace of learning of an individual learner using any profiling mechanism.
Personalized Learning Software Systems	Learning systems that tailor learning resources accesses within the software environment to a user model.

Following is a brief review and explanation of the main components of personalized learning software systems: learning environments, learning resources, and learner models.

3.2.1 Software Learning Environment

Various terms are used interchangeably to refer to a wide variety of computerized learning environments, such as e-learning, online learning, mobile learning, game-based learning, virtual learning environments, and tutoring systems. The rationale for using one term or another depends on the perspective from which we analyze the learning environment. Sometimes learning environments are characterized by the type of technology used to implement them, by the interaction model used, or by the learning approach. For example, we may use the term "mobile learning system" to refer to any type of computerized learning system implemented using mobile technologies; this may include an educational game, a specialized training application, or a tutoring application. Alternatively, an e-learning system is more likely to leverage the features of web technologies, this in turn may include online educational games, online courses, or webinars. On the other hand, a learning system that implements one-on-one instructions and assessments mimics a human tutor and is referred to as a tutoring system. Tutoring systems can implement direct instructions and assessments in a virtual learning environment or embed and conceal instructions in a game-based learning environment. We can see now how different terms can refer to the same learning software system depending on the perspective. The type of technology and interaction model provide not only different categorizations of learning systems, but also variable attributes and features for personalization. For example, mobile devices can provide context related data that support personalization, such as location, e.g. [90-91]; game-based learning environments provide rich interaction models helpful in modeling the learner skills and preferences, e.g. [92-94].

Figure 7 represents a classification of software learning environments according to the learning approach, interaction model, and technological framework as explained above. Table 2 provides a brief explanation of each category of software learning environments listed in Figure 7.

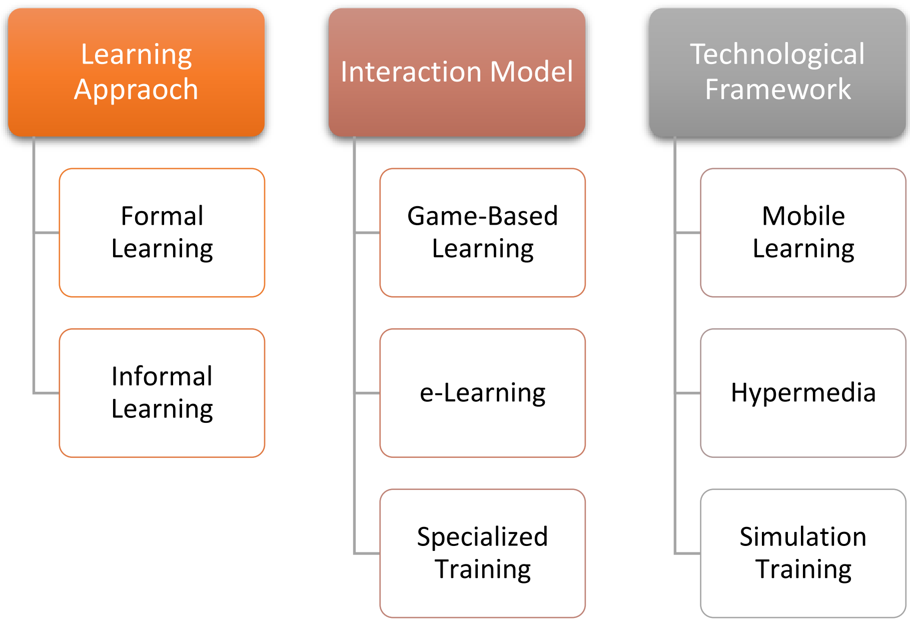


Figure 7: Classification of software learning environments according to the learning approach, interaction model, and technological framework

In this research, however, software learning environments are categorized according to the learning approach. These two main learning approaches are considered:

1. Formal Learning Software Systems
2. Informal Learning Software Systems

Table 2: Explanation of different types of software learning environments

Perspective	*Software Learning Environemnt*	*Definition*
Learning Appraoch	*Formal*	Memics the type of learning carried out at formal educational institutions by providing a well-defined learning content aligned with a curriculum and learning outcomes and evaluates through assessments. Can lead to a qualification or be part of a formal educational system.For examples, tutoring systems and online courses.
	Informal	Offers learning content or activities that are not necessarily aligned with a curriculum and doesn't lead to qualification. Assessment is usually not carried out. For example, online games, information wikis, professional blogs.
Interaction	*Game Based Learning*	Describes an approach to teaching, where students explore relevant aspect of games in a learning context designed by teachers.
	e-Learning	e-Learning is learning utilizing electronic technologies to access educational curriculum outside of a traditional classroom. In most cases, it refers to a course, program or degree delivered completely online.
	Specialized Training	A form of training that puts the learning in virtual environemnt memicing real-life situation through which they can acquire new skills.
Technology	*Mobile*	Mobile technology is the technology used for cellular communication.
	Hypermedia	Hypermedia, an extension of the term hypertext, is a nonlinear medium of information that includes graphics, audio, video, plain text and hyperlinks. The WWW (World Wide Web) is a classic example of hypermedia
	Simulation	Simulation trainings are used as a tool to teach trainees about the skills needed in the real world. It provides a lifelike point-of-care learning experience, and has been widely applied in fields such as aviation, the military, and healthcare.

1. Formal Learning Software Systems

A considerable number of research work in the field of Computer-Assisted Learning emphasizes the importance of embedding good pedagogical design relevant to some learning theories and instructional design methods to ensure effective learning, e.g. [95-97]. According to these assumptions, fully formal learning computer systems were developed attempting to model learning processes and actitivities similar to the ones carried out in class room. In such cases well-defined learning content, learning outcomes and assessment measures are implemented in the learning computer system [10, 11, 92].

Most of the fully formal learning systems attempt to model the human tutor and are called tutoring systems [10, 92]. Tutoring systems are implemented using different technologies, e.g. mobile technologies [12], web technologies [9], and are designed with variable interaction models, e.g. game-based tutoring systems [7], online courses [8], and many others. In these systems personalization is accomplished mainly by modeling skill level, i.e. mapping learning content suitable to the skill level of the learner based on some predefined assesment measures. Additionally some research efforts focused on modeling the learner learning style providing more sophisticated cognitive personalization that maps suitable representations of learning content, as well as, suitable types of learning activities to the learner's learning style [32, 98]. Nevertheless, these learning systems are constrained by a specific content, learning outcomes, and assessment measures that make them suitable for only specific domains, e.g. specific subject matters, specific profesional tarining programs, and specific curriculums, or specific group of learners, e.g. primary students, high schoolers, or professional workers. Furthermore, learners are expected to be interested in the predefined content, given that they are using these particular systems to learn a specific subject and earn a certain qualification or master a certain competency. However, there are cases where learners are interested in multiple different topics or they have just started to experience new interests while they are learning about a specific subject. Using predefined content, instructions and assessment measures may ensure mastery of a specific subject matter, but, hinders adaptivity and limit personalization to learners' changing needs and interests in the general context. As a result, informal learning systems were introduced to support formal learning systems and give more flexibility and freedom to learners.

2. Informal Learning Software Systems

Informal learning is self-directed, does not follow a specified curriculum, and does not lead to formal qualifications [13]. This form of learning is sometimes used to support formal learning activities. For example, e-Learning recomemnder systems [99], and webquests [100] are used to support formal learning.

However, in its broader form, informal learning systems, allow learners to choose what they need to learn anywhere and anytime not restricted to predefined curriculum or assessements measures. This type of learning mimics the natural process of knowledge acquisition in human beings. We explore, observe, acquire knowledge and keep accumulating knowledge in certain areas of interest following learning methods that suit us the most. One common example of informal learning environemts are knowledge sharing systems used in some companies to promote cooperation and knowledge sharing among workers in the workplace [101]. Studies of informal learning reveal that up to 90% of adults are engaged in hundreds of hours of informal learning [14]. It has also been estimated that up to 70% of learning in the workplace is informal [15].

Moreover, recently many research works invistigated how social media networks such as Facebook and knowledge wikis can contribute to informal learning as tools of knowledge sharing and acquisition [16-18, 102]. Wikis among other platforms gained most of the attention [19-23].

Informal learning can be thought of as the most comprehensive type of learning as it covers all types of knowledge and is open to all types of learners. In such contexts, the main driver of learners to learn is their need and interest to learn. This is the type of learning environment addressed in this work.

3.2.2 Learning Resources

A variety of learning resources can be used in personalized learning software systems. Learning resources may include various components, such as online courses, e-books, instructions, assessments, learning activities, etc. Some research works rely on fully structured representation of learning resources such as relational databases, allowing for common database selection and retrieval operations based on some personalized selection conditions or constraints [29, 90]. Furthermore, structured data representation facilitates easy conversion into features' vectors representation which is commonly used in datamining-based approaches for training classification [67], clustering [99], or regression models [93] in personalized learning systems.

In addition to structured data representation, many research works use advanced knowledge representations in the form of learning objects [2, 3], ontologies [52], or more recently Linked Open Data (LOD) [53]. Chiappe defined Learning Objects as: "A digital self-contained and reusable entity, with a clear educational purpose, with at least three internal and editable components: content, learning activities and elements of context. The learning objects must have an external structure of information to facilitate their identification, storage and retrieval: the metadata. [103]." Ontologies are formal representations of taxonomies and concepts, essentially defining the structure of knowledge for various domains such that the *nouns* represent *classes of objects* and the *verbs* represent *relations* between the objects. These learning resource representations, commonly used in the semantic web, are characterized by standardized representations based on formal description of concepts, terms, and relationships within a given knowledge domain allowing for knowledge reusability. For example, the Resource Description Framework (RDF) provides a formal vocabulary for describing properties and classes of RDF-based learning objects. Web Ontology Language (OWL) is based on RDF formalism and is used to describe properties and classes of ontologies. The main limitations of using these knowledge representations are the domain dependency and the development cost.

On the other hand, there is huge amount of information available on the web in unstructured text format, i.e. free text. Typically, this is the kind of text found in blogs, wikis, forums, and social media websites. Considerable research works focus on supporting learning by using unstructured text publishing platforms such as blogs [104, 105], wikis [19], online forums [106], and social media networks such as Facebook [107]. There are many challenges inherent in the processing and analysis of unstructured text. First, unlike structured data, there are no predefined features with known and well-defined values. Second, unstructured text may have the same word used in several ways and in different contexts implying different meaning, i.e. polysemous words, or may have many words referring to the same exact meaning, i.e. synonymous words, causing redundancy and inconsistencies. This type of content is addressed in this research. Figure 8 represents classification of the main learning content types and representations used in personalized learning software systems as explained earlier.

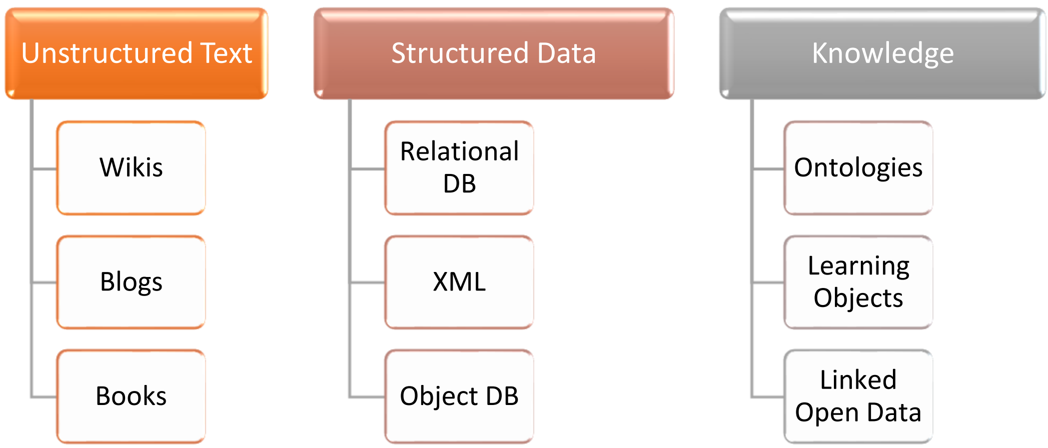


Figure 8: Classification of different learning content used in personalized learning software systems with examples

3.2.3 Learner Modeling

User modeling is the process of inferring information about users by analyzing users' characteristics, choices, or behavior [26]. User models are required by many personalized systems such as personalized search engines [2], eCommerce personalized applications [4], and more importantly for us, personalized learning systems [93, 94, 108]. Since personalization is concerned with tailoring content or some system's functions to specific user's traits, hence, without a user model, there is no personalization possible. So, how can we build learners' models? When creating a learner model, four main questions need to be answered:

1. What aspects of the learner need to be modeled?
2. What data can be used to infer the required model?
3. How will data be collected?

4. How will the model be created?

Since we are focusing on personalized learning systems, we will be addressing these questions about learners. For personalized learning systems and as stated by the definition of personalized learning, presented in section 3.2, we need to model needs, interests, preferences, and pace about an individual learner to accomplish personalization. However, a learner model may cover all or some of these aspects depending on the type of system and level of personalization required. Profiling data input methods range between automatic/implicit and collaborative/explicit [109, 110]. In automatic profiling, learner's characteristics are derived automatically, either from historical data or by monitoring learner's interaction with the system such as: click logs, browse history, cache logs, mouse clicks, eye tracking, and cookies. Whereas in collaborative/explicit profiling, the learner is prompted to input profiling data either through questionnaire or other input mechanisms. In recent studies focusing on modeling context and psychomotor skills, GPS and sensors technology are commonly used to implicitly collect data related to location, temperature, body positions, or eye gaze [90, 91, 111].

Many early efforts in learner modeling used stereotypes to map learners' skill level into pre-defined categories. Stereotyping is a technique used to build models of users through clusters/groups of characteristics or attributes that define number of assumptions about the user's personality, skills, background, or preferences. So, for example, one might know that if someone is a judge, he or she is probably - over forty, well-educated, reasonably pro-establishment, fairly affluent, honest, and well-respected in the community. Some of the earliest examples of stereotype-based personalized learning systems are KNOME [29] and GRUNDY [112]. In these systems, each stereotype incorporates a number of characteristics about the learner, as well as, implies a number of assumptions. In KNOME users were stereotyped into skill-level categories such as “novice user” or “expert user” based on their mastery level in using UNIX command. In GRUNDY stereotypes were used to model books’ preferences in its most basic level. For example, a “Doctor” stereotype implies that the learner is well-educated and prefers specific type of books. Even though, stereotypes were easy to define and implement, as well as, had provided reasonable learner’s models in the past, they were very limited, not adaptive and, in some cases, superficial. More logical and scientific approaches to learner’s skill modeling mainly adopted in tutoring systems were Cognitive Tutors (CT) [28], Constraint-Based Modeling (CBM) [34], and knowledge spaces [27]. In Cognitive Tutors and Constraint Based Modeling, the focus is problem solving skills, the skills are represented as rules (CT), and predicates (CBM), which bear a strong formal similarity. In (CT), a skill is considered correctly applied by the student when a rule is matched to student performance actions. In the case of (CBM), a skill is considered mastered when a predicate is matched over student responses. Whereas, the theory of knowledge spaces indicates which knowledge states can be reached from a given knowledge state, based on inference relations among items supporting efficient curriculum sequencing. The main advantage to curriculum sequencing over (CT) and (CBM) lies in tailoring the learning content based on an accurate assessment of a large array of skills with the least possible amount of evidence. The two major limitations to these skill modeling methods are the need for substantial expert human intervention to define rules, measures, and assessments of skills or different states of knowledge for curriculum sequencing, and the absence of affective factors that strongly influence a learner’s preferences to learning. For personalized formal learning systems that are bounded by predefined learning outcomes, ignoring learners' preferences can be considered as a major drawback reducing the effectiveness of the system and hindering the adaptability. For example, "we may want to know if the learner is bored or frustrated, what is the appropriate moment to switch from drill and practice to explanations and theoretical material. Human tutors are well acquainted with factors like the student's attitude and motivation towards learning a given topic and their critical effect on the learning outcome [113]."

In response to these limitations inherent in stereotypes, or rule-based formal learner modeling approaches, various approaches were introduced based on techniques and concepts commonly used in datamining. Datamining techniques such as classification, clustering, and statistical analysis provide many opportunities for learner modeling combining more than one aspect at a time. Typically for cognitive personality analysis and identification, traits are identified using questionnaires containing descriptive items that accurately reflect the traits of interest [31], which can be used to personalize learning content presentation, instruction mechanism, or any relevant components in the learning environment. In addition, emotions represent a sort of reactions to the perception of a specific (external or internal) event, accompanied by mental, behavioral and physiological changes [114]. They have been defined in a huge variety of ways and there is no agreed-upon theory that explains them. However, "there exist many modalities for affect detection (e.g., spoken and written language, video including facial expression, body posture and movement, physiological signals, tactile interaction data), which can either use a discrete (in terms of specific emotions) or a continuous (in terms of degrees of valence and arousal) representation model [31]". These can be used to define attributes that facilitate the identification of a learner's current state of emotion and taking relevant adaptation actions accordingly using datamining techniques. Moreover, skill levels have become easier to define and detect using datamining classification and clustering techniques. For example, Nascimento et al. [93] implemented logistic regression to classify learners into "literate" vs. "illiterate" based on some fixed attributes. Moreover, in controlled informal settings, datamining, was also used to elicit learner's interests and needs, especially in information and knowledge retrieval (e.g. retrieving books [99], retrieving learning objects in online learning environments [3]). Datamining techniques helped reduce expert human intervention, in terms of defining skill-based rules and allowed for more adaptive modeling. However, datamining approaches still require the identification of relevant attributes as well as representative historical data which most of the time requires manual annotation. Table 3 presents a summary of user modeling approaches explained earlier.

To this, it can be seen that in personalized learning systems where specific learning content or specific learning instructions are typically being personalized for specific user group, characteristics such as knowledge and skill-level [27-29, 93], emotions [31], preferences [32], and context [90] are dominant. These characteristics, especially learner knowledge, are often important in formal learning systems that deliver predefined content and attempt to achieve well-defined learning outcomes such as tutoring systems [34], or online courses [35]. The fact that these formal systems deliver very specific content for a very specific learner base creates no demand for personalized interest modeling. Traditionally, learners using these personalized formal learning systems came with an interest to use and learn the specialized content delivered in that system. 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, often referred to as adaptive hypermedia, that dealt with huge bulk of diverse information such as online encyclopedias [36]. In the following section different approaches for user interest modeling in adaptive hypermedia environments are reviewed.

Table 3: Summary of some of the most common learner modeling approaches in the literature

	Learner Characteristics			
Components of the modeling approach	*Skills*	*Preferences*	*Needs*	*Interests*
Data Used	***Explicit***: Answers to questions, number of mistakes or correct answers, feedback to questionnaires, …etc. ***Implicit***: Time required to complete a learning task, number of times user seek help or look for hints, invalid navigation within the learning environment, etc.	***Explicit***: User choices and feedback to questionnaires such as psychometric analysis tests. ***Implicit***: Inferred knowledge from learner navigation depending on choices of learning tasks, preferred images, activities, navigation patterns …etc.	***Explicit***: User choices and feedback to questionnaires. ***Implicit:*** Visited pages, clicked items, …etc.	***Explicit***: User choices and feedback to questionnaires. ***Implicit:*** Visited pages, clicked items, …etc.
Collection technique	Mainly through user assessment mapped to some pre-defined measures, functions, or rules.	Mainly through user interaction. Log files, keystrokes, mouse clicks, … etc.	Mainly through user interaction. Log files, keystrokes, mouse clicks, … etc.	Mainly through user interaction. Log files, keystrokes, mouse clicks, … etc.
Modeling Technique	Stereotypes Procedural -Cognitive Tutors Declarative -Constraint-Based Modeling (CBM) Knowledge Spaces Data mining approaches: clustering, classification, or association rules.	Stereotypes Rule-based Data mining approaches: clustering, classification, or association rules.	Explicit mapping. Information retrieval approaches Recommendation approaches	Explicit mapping. Information retrieval approaches Recommendation approaches

3.3 User Interest Modeling in Information-oriented Hypermedia Environments

In 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" [115]. Methods and techniques used to model user's interests in information-oriented hypermedia environments varied widely over time. A number of research works done in interest modeling is reviewed. This review excludes research works that use user ratings or user likes/dislikes to model interest. It also excludes models of interest that rely on contextual data such as location, speed, or time as found in context-aware systems. Here in this research, the term "context" is used to refer to the semantic context implying the meaning of the text and not the physical context.

Early efforts in user interest modeling focused on the keyword level [116]. 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 [117], preferences [118], topics [119], or Folksonomies, also known as social tagging [120]. More recently, work in this line explored approaches of data visualization to support information exploration by visually suggesting relevant keywords. Work in this field propose query suggestions [121], negative relevance feedback as used in Intent Radar [122], or visualization as seen in AdaptiveVIBE [123] and SearchLens [124], which include two dimensional visualizations of documents and their relation to the user's inferred interests. Figure 9 shows a screenshot from Adaptive VIBES.

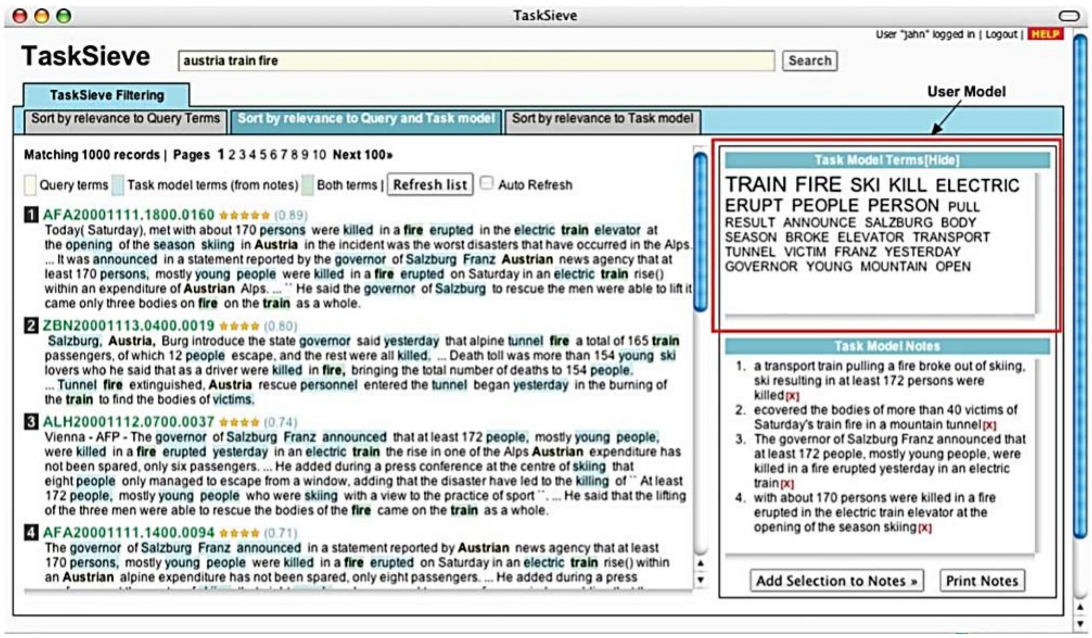


Figure 9: Screenshot from Adaptive VIBE

However, interest modeling approaches relying on keywords defined explicitly by users suffer from many limitations that were highlighted in a number of research studies [116], [125], [126]. 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 [127-131]. 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 [132]. Figure 10 shows a keyword vector user interest model grouped into categories.

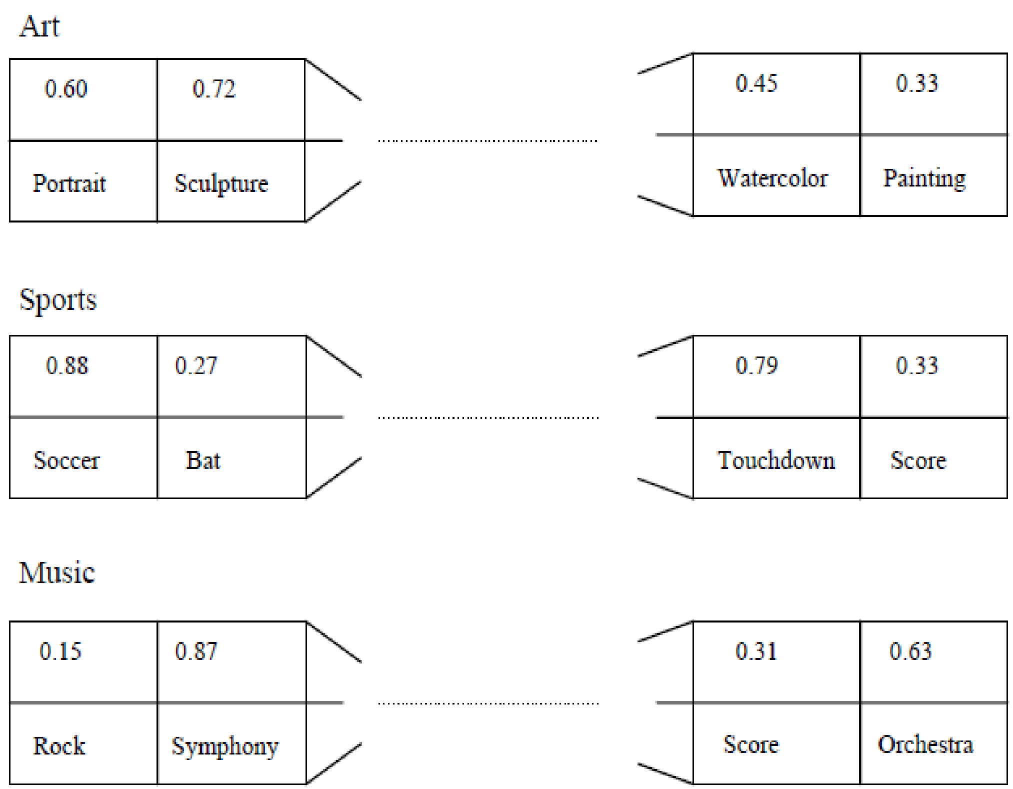


Figure 10: A keyword vector user interest model

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 [133-135], and concept vectors [136], and ontologies [137-139]. In semantic networks interest models, each node represents a concept or a word, and each edge has a weight that reflect the relationship between concepts in the semantic network. Additionally, context attribute can be added to enrich the semantic network. Concept-based profiles are similar to semantic network-based profile in the sense that both are represented by conceptual nodes and relationships between those nodes. However, in concept-based profiles, the nodes represent abstract topics considered interesting to the user, rather than specific words or sets of related words.

Semantics-rich user profiles have an advantage over keyword-based profiles 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 predefined knowledgebases. Moreover, these approaches cannot be considered highly adaptive to changing user interests. Table 4 presents summary of some of the interest modeling approaches discussed in this section.

User-centered and adaptive interest modeling approaches began in recommender systems (RSs) [37]. The field of recommender systems focused on learning and education is called technology-enhanced learning recommender system (TEL-RecSys).

Table 4: Summary of user interest modeling approaches in information-oriented websites

User Interest Model		***Research Work***
Keyword-based	***Explicit/ User-defined***	[119], [118], [117], [121], [122], [123] , [124]
	Implicit/ Corpus-based	[127], [128], [129], [130], [131]
Semantics-rich	***Semantic Networks***	[133], [134], [135]
	Ontology	[137], [138], [139]
	Concept Hierarchies	[136]

3.4 Technology Enhanced Learning Recommender Systems

Many technology-enhanced learning (TEL) systems utilize different types of recommender engines to support learning [37]. As classified by Drachsler et al. [38], TEL recommender systems reported in the literature support various tasks such as finding good learning content [140], [39], suggesting the most effective paths through a plethora of learning resources to achieve a certain competence [141], [40], or suggesting peers learners, which is very central recommendation task for distance education settings where learners usually feel isolated and sometimes demotivated [41].

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

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

3.5 Wikipedia Recommender Systems

Several research papers 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 recommendation on Wikipedia is concerned with recommending editing tasks to authors as proposed in [142], [143], and [144]. In this research, article recommendation models are focused on.

Some recommendation models have been proposed to provide article recommendations in Wikipedia. For example, Sriurai et al. [44] used the Latent Dirichlet Allocation (LDA) algorithm to generate topic-based recommendations. The proposed topic-based model is used to generate topic features which are used to classify articles against topics using LDA. The model was evaluated with an unspecified number of articles by 5 assessors. Each assessor was given a number of recommended articles and linked articles, i.e., linked through hyperlinks within articles, and asked to give a relevance score from 1 to 5. The average relevance score for recommended articles surpasses the relevance score of the linked articles by 1.2. The approach is neither designed to generate personalized recommendations, nor accounts for changing interests. Rather, fixed recommendations are presented to all readers following a pre-built topic distribution that depends on the page links.

In addition to the new variations of content-based recommendations, researchers started to utilize new variations of search algorithms to deliver structural recommendations [46]. In structural recommendation techniques, content or/and users are represented using graphs. Graph search and ranking algorithms are then used to recommend nodes, links, or different combinations of both. A recent research study by Schwarzer et al. [47] 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.

However, Wikipedia's articles were found to form a scale-free network [51]. 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 out when recommending articles merely based on links. Therefore, there is a need to adaptively model the changing interests as well as recommend articles based on semantic relevance, not just barely based on links or references. To this end, our research objective is to design and implement an effective learner interest modeling approach to facilitate personalized content recommendation on Wikipedia.

Chapter 4: Fuzzy Set-based Feature Vector Representation for Efficient Semantic Analysis

4.1 Background

As explained earlier, unstructured texts suffer from several complications. First, unlike structured data, there are no predefined features with known and well-defined values. Unstructured text may contain any number of various words. Second, unstructured text may have the same word used in several ways and in different contexts implying different meanings (polysemous words) or may have many words referring to the same exact meaning (synonymous words) causing redundancy and inconsistencies. Third, in some unstructured text contexts, as seen in informal social networks it is common to use special characters, emoticons, and abbreviations that add noise to the text and at the same time may add high value if analyzed carefully. Various approaches proposed in the literature to account for semantic in the text. Most of these approaches can be classified into two categories: contextual semantic approaches, and conceptual semantic approaches. Conceptual approaches of semantic analysis rely on external semantic knowledge bases such as ontologies and semantic networks. Although conceptual semantic analysis might be more comprehensive in terms of concepts diversity along with their semantic relevance, it is still limited by their underlying knowledge bases and requires large amount of manual effort during the knowledgebase creation and validation phase. In contrast, contextual approaches utilize statistical analysis of the relationships between terms in the text to infer semantic. These approaches tend to be more flexible given the possibility of automation.

Contextual approaches, however, require the unstructured text to be converted into a suitable structured representation. To do this, many preprocessing techniques such as tokenization, stopwords removal, stemming, and trimming are proposed in the literature [145, 146]. After completing preprocessing of unstructured text, it can be converted into a structured format by selecting effective document representation model to calculate semantic similarity between different text units such as words, sentences, paragraphs, and full documents. Document models reported in the literature for contextual semantic analysis can be roughly classified into two major categories: vector-based models, and corpus-based models [147].

Vector space model (VSM) or bag of words (BoW) model is an algebraic model for representing text documents as vectors of text identifiers such as index terms. It is most commonly used in information filtering and information retrieval context [148].

In VSM/BoW documents, d, and queries, q, are represented as vectors such that:

$$d = \{w_{1,} w_{2,} \dots, w_n\}, where\ n\ is\ the\ number\ of\ terms\ in\ d$$

$$q = \{w_{1,} w_{2,} \dots, w_n\}, where\ n\ is\ the\ number\ of\ terms\ in\ q$$

Each dimension corresponds to a separate term. If a term occurs in the document, its value in the vector is non-zero. Several different methods to compute these values, also known as weights, have been developed such as frequency, polarity, and co-occurrence. One of the most commonly known weighting schemes is term frequency inverse document frequency, TF-IDF [149].

The definition of a term varies depending on the problem being addressed. Terms can be single words, keywords, phrases, or paragraphs. Vector operations can be then used to compare documents with queries using metrics such as cosine and dot product which are considered semantic similarity measures. Unfortunately, VSM representation scheme has its own limitations. Some of these are: high dimensionality of the representation resulting in sparsity problems, and theoretically it is assumed that terms are statistically independent resulting in loss of correlation with adjacent words and loss of semantic relationships that exist among the terms in a document.

In contrast, corpus-based document model analyzes relationships between a set of documents and the terms they contain then produces a set of concepts related to the documents and terms. The underlying idea is that the aggregation of all the word contexts in which a given word does or does not appear provides a set of mutual constraints that largely determines the similarity of meaning of words and sets of words to each other. The most commonly known example of corpus-based document representation for semantic analysis is The Latent Semantic Analysis (LSA) [150-152]. It uses Singular Value Decomposition (SVD) to find the semantic representations of words by analyzing the statistical relationships among words in a large corpus of text. When LSA is used to compute sentence similarity, a vector for each sentence is formed in the reduced-dimensional space; similarity is then measured by the cosine of the angle between their corresponding row vectors. The dimension size of the word by context matrix is limited and fixed to several hundred because of the computational limit of SVD. As a result, the vector is fixed and is thus likely to be a very sparse representation of a short text such as a sentence.

To overcome these problems, a hybrid representation of text document based on concepts from fuzzy set theory is proposed.

Our approach uses a fuzzy relationship to generate a matrix of terms and their semantic relationships. We refer to this matrix as “fuzzy thesaurus” throughout our research. This matrix indicates how similar individual terms are, *term-term similarity*, where terms are distinct text units (i.e. single words). Then, this fuzzy thesaurus is used to populate a document vector of various types of terms (i.e. single words, phrases, topics, …etc.) where the value of each term in the vector indicates the fuzzy relationship between the term and the document, *term-document similarity*. This hybrid document representation can be then used to calculate the semantic similarity between text documents.

In the following sections, first, concepts related to fuzzy set information retrieval model are introduced, then, the proposed semantic analysis approach is explained, and finally, we present experiments using the proposed approach for Twitter sentiment analysis being one of the most challenging text mining tasks given the very short size of text documents, i.e. Tweets, which always result in major sparsity issues. In Chapter 5 the application of this approach in the context of personalized content recommendations is itroduced.

4.2 Fuzzy Set Information Retrieval Model

Fuzzy set theory relies on two main principles: sets are not crisp (boundaries of the sets are ambiguous or fuzzy), and elements belong to the fuzzy set at different levels of membership [153]. Language sentences and documents are typical examples of fuzzy sets. A fuzzy set IR model is adopted to determine the degree of membership between every keyword in a sentence and a fuzzy set that contains different words, each of which belongs to the set at some degree of membership. The degrees of similarity or membership, also referred to as the correlation factors among words, are given by a function which assigns a value in the range [0, 1] to any two words. Hence, if two sentences contain many terms that belong to the same fuzzy sets at a high degree of membership then the two sentences are similar. There are several methods to define the correlation factors among different words; for example, (i) word connection calculates the correlation of any two words w_1 and w_2 by counting the number of documents in a collection where both w_1 and w_2 appear together, (ii) keyword co-occurrence, not only considers the number of documents in a collection where both words w_1 and w_2 appear together, but it also considers the frequency of co-occurrence of both w_1 and w_2 in a document, and (iii) distance, considers the frequency of occurrence as well as the distance, which is measured by the number of words, between w_1 and w_2 within a document [55].

Ogawa et al., [154] adopted a fuzzy set IR model to determine whether a keyword in a sentence belongs to a fuzzy set that contains words with different levels of similarities among them. They called the fuzzy set a keyword-connection-matrix and defined it as a type of thesaurus that describes relations between keywords by assigning similarity grades restricted to the interval [0, 1]. Yerra et al. [155] used the same keyword-connection-matrix proposed by Ogawa et al. [154] to detect similar HTML documents. They compared every keyword, k, in a sentence, i, with every keyword, w, in a document, d, and calculated a word-sentence similarity, $\mu_{k,d}$, using a fuzzy association. The average of all μ-values is calculated to yield the overall similarity, Sim (i,d), between i and d.

A similar approach will be adopted in this research and this will be further explained in subsequent section.

4.3 Fuzzy Set-based Feature Vector Representation

The proposed approach for generating feature vectors based on fuzzy set is composed of two main tasks:

1. Building a fuzzy thesaurus of terms
2. Using the fuzzy thesaurus to populate vectors of terms where terms can be distinct words, topics, phrases ...etc.

The process is illustrated in Figure 11.

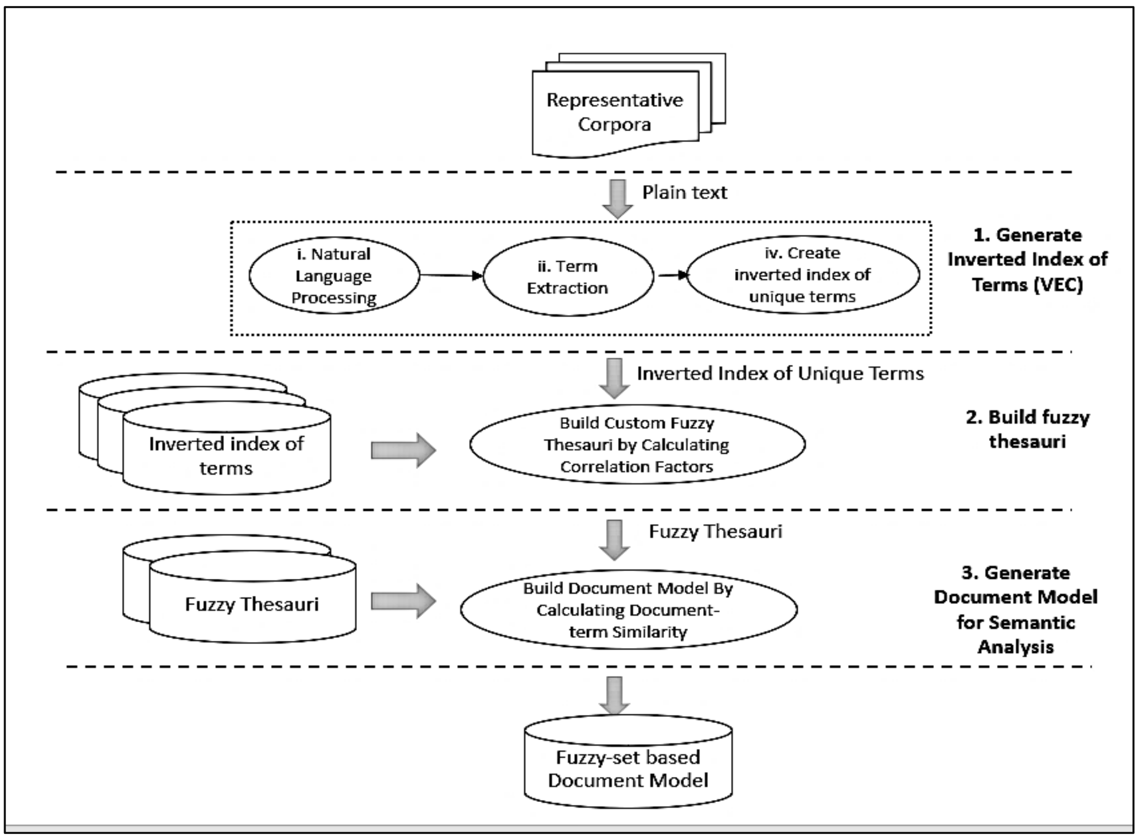


Figure 11: The process of building fuzzy-set based document models (Feature Vectors)

4.3.1 Building a Fuzzy Thesaurus

The first step in generating feature vectors, or term vectors, based on fuzzy sets is to build the fuzzy thesaurus that defines the semantic similarity between distinct terms in the main corpora. For this task, a representative corpus is required. In addition, natural language preprocessing is carried out to prepare the text to be used in the fuzzy thesaurus.

Once preprocessing task is completed, *inverted terms index is created.* An inverted word index, *VEC*, is a set of vectors where each vector indicates for each unique word, or term, in the corpus: the documents in which it appears, and its positions, i.e. occurrences, in that document. For example, Table 5 and Table 6 show the vectors of stemmed words “mobil” and “comput”, vec(mobil) & vec(comput), in the inverted index of words.

Table 5: 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 6: Inverted Index for "Comput", vec(Comput)

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

Using inverted indices of words, i.e. Terms, custom fuzzy thesauri are built that defines the semantic similarity between each two distinct words for every corpus by calculating the distance correlation factors between each two distinct words in the corpus using Equations (1), (2), and (3).

We choose the distance correlation factor since it has empirically been proved to achieve the best results in the information retrieval context with an accuracy rate of 94% compared to 47% for the keyword-connection factor and 52% for the co-occurrence factor [55]. 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 keywords across all documents within a single corpus: the frequency of co-occurrence and relative distance in a single document (C_{ij}), Equation (1), the normalized value (nC_{ij}), Equation (2), and finally the distance correlation factor (Cf_{ij}), Equation (3).

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

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

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

Where $distance(x, y) = |Position(x) - Position(y)| + 1$ 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 & 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.

$|vec\ (wrd_i)|\ \&\ |vec(wrd_j)|$ represent the number of words 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 text units and documents in a corpus.

4.3.2 Generating Feature Vectors based on Fuzzy Sets

Once the fuzzy thesaurus is built it is used to generate the semantic feature vectors using the following equation:

$$\mu_{F,d} = 1 - \Pi\left(1 - Cf_{ij}\right) \tag{4}$$

Where F is a feature and d is a document. Feature is a text feature which can be a keyword, topic, distinct word, a phrase …etc. A document is defined in each problem as the full text. It can refer to a Tweet, a web page, or a learning resource.

The major strengths of this representation are that they are not limited to specific domains or predefined seed terms or entities, do not require any manual annotation, don not force any limit on sentence or document size, and they account for full and partial similarity being constructed based on fuzzy sets thus the resulting document representation are less sparse.

In the following section, a fully automated method for building semantic Twitter feature vectors for machine learning sentiment analysis based on a fuzzy thesaurus and sentiment replacement is defined. The proposed method measures the semantic similarity of Tweets with features in the feature space instead of simply using occurrences or frequencies. By measuring the semantic similarity, we account for the sentiment of the context instead of just counting sentiment words. This is primarily important in Twitter given the informal writing style that may use positive words to ironically express negative feelings and vice versa. In addition, this method produces less sparse datasets.

The major contributions of this work are summarized in the following four points:

1. Outline a framework for semantic Twitter sentiment analysis based on a fuzzy thesaurus and sentiment replacement.
2. Show that using a fuzzy thesaurus can incorporate semantic relationships for Twitter sentiment analysis and increase the accuracy of sentiment analysis.
3. Show that using a fuzzy thesaurus to represent semantic relationships yields some improvement over other representations including frequency, presence or polarity, and term frequency inverse document frequency (TF-IDF).

4.4 Twitter Sentiment Analysis based on Semantic Feature Vectors

In this section, we introduce a new method for generating semantic feature vectors with reduced dimensionality for Twitter sentiment classification from raw Twitter data. Twitter data can be collected using the Twitter API https://dev.twitter.com/rest/public), or can be benchmark data, which is publicly available for experiments and research such as the datasets used in *Section 4.4.4.* Sentiment replacement is used to reduce the dimensionality of the feature space as well as the fuzzy thesaurus is used to incorporate semantics. The proposed method consists of the following three main tasks, highlighted with a gray rectangle in Figure 12:

1. Sentiment replacement.
2. Feature extraction and reduction.
3. Feature vectors generation based on semantic similarities.

The generated semantic feature vectors are then used to train any machine learning classifier for sentiment classification task. We show later, in *Section 4.4.4,* classification results of Multinomial Naïve Bayes (MNB), Bernoulli Naïve Bayes (BNB), and Support Vector Machines (SVM) classifiers. In the following sub-sections, the three main tasks in the proposed method are explained.

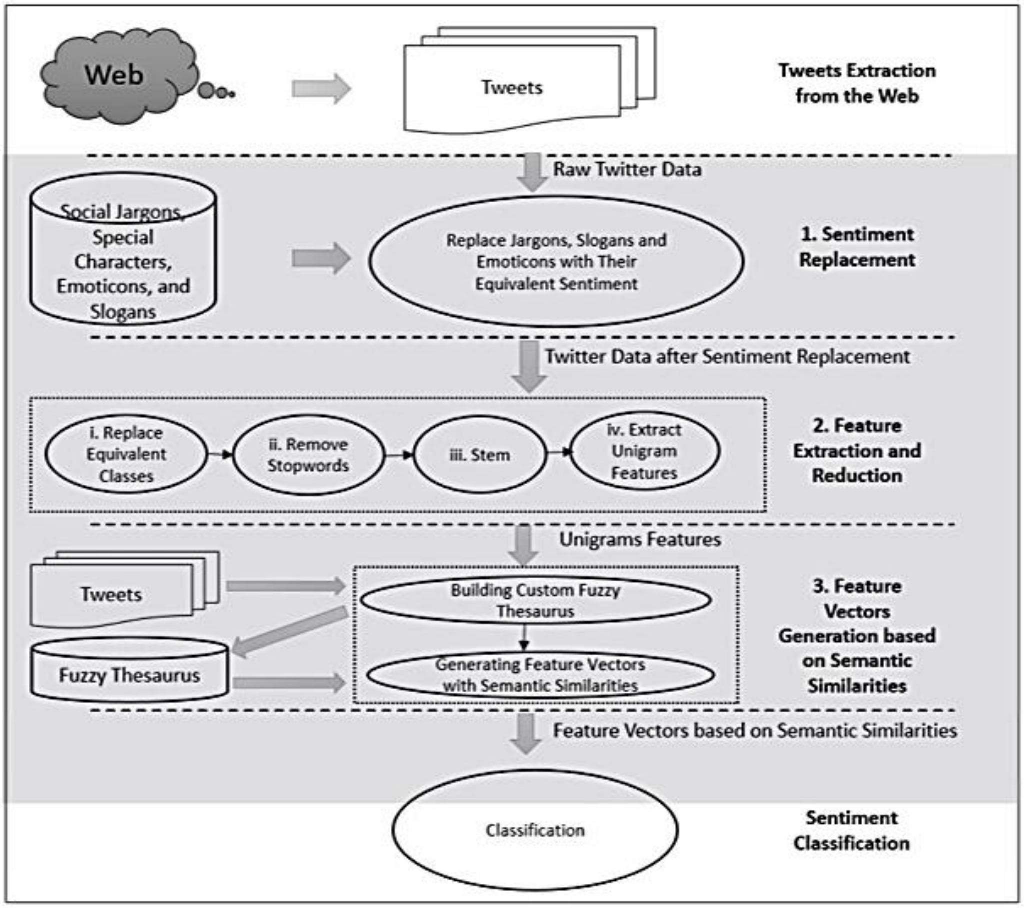


Figure 12: Semantic sentiment classification based on sentiment replacement and fuzzy thesaurus

4.4.1 Sentiment Replacement

Sentiment replacement is achieved via a program that interfaces with a publicly available Twitter slogan, special characters, emoticons, and abbreviation list. In available sentiment lexicons, only proper and formal words are considered. However, in social networks, the use of slogans, emoticons, and abbreviations is very common, and it adds strong indication of the sentiment of the text. These abbreviations and slogans might be removed through natural language processing stages during preprocessing, especially special characters and emoticons, cutting out useful sentiment indicators. Thus, we perform sentiment replacement of slogans and abbreviations before the preprocessing phase. For example, “loool” is replaced with “Happy”. All emoticons are replaced with their equivalent sentiment word. For example, “☺” is replaced with “Happy” and “☹” “:/”, “: \” are replaced with “Sad”.

4.4.2 Feature Extraction and Reduction

Once the sentiment replacement is done, natural language processing [146] of the Twitter data is performed. Generally, unstructured texts cannot be directly processed by classifiers and learning algorithms. In addition, Twitter data is full of peculiarities due to the informal writing style commonly used on Twitter resulting in more noisy text. Thus, we perform a number of natural language processing tasks, which proved effective in previous studies [156-158]. They have become a common practice in Twitter preprocessing for sentiment classification, to transform the Twitter unstructured text into a ‘bag-of-words’ model with a reduced number of features that is manageable by classification algorithms. The following preprocessing tasks are performed in order:

1. Equivalence classes replacement such that:
    - All Twitter usernames which start with @ symbol, are replaced with the term “USERNAME”.
    - All URL links in the corpus are replaced with the term “URL”.
    - Reduce the number of letters that are repeated more than twice in all words. For example, the word “loooooveeee” becomes “loovee” after reduction.
    - Remove all Twitter hashtags which start with the symbol “#”.

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," "by."
3. Stemming [159]: 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. Bag-of-words extraction: we choose unigram features since they can be directly used with the fuzzy association rule as in Equation (4). Typically, in a unigram representation, each single word in the corpus is treated as a feature.

After completing the preprocessing tasks, a custom fuzzy thesaurus is built and is used to generate feature vectors based on semantic similarities that is later used for sentiment classification. The process of building the custom fuzzy thesaurus and generating the feature vectors based on semantic similarities is explained in the following section.

4.4.3 Feature Vectors Generation Based on Semantic Similarities

This phase encompasses two main activities:

1. Building the fuzzy thesaurus.
2. Generating semantic feature vectors.

In subsequent section, each activity is explained in detail.

<u>1. Building the Custom Fuzzy Thesaurus</u>

We build the custom fuzzy thesaurus that defines the semantic similarity between each two distinct words in the Twitter corpus by calculating the distance correlation factors between each two distinct words in the corpus using Equations (1), (2), and (3) explained earlier.

Unigram features, generated in the previous step, are now used to generate vectors of all distinct preprocessed words in the Twitter corpus along with the documents' IDs in which they appear. A Tweet is considered a document in this context, and their positions in every document.

Using the vectors of distinct words in the Twitter corpus, we define for every pair of keywords across all documents: the frequency of co-occurrence and relative distance in a single document (Cij), Equation (1), the normalized value (nCij), Equation (2), and finally the distance correlation factor (Cfij), Equation (3).

As a result, a matrix of all distinct words and their semantic relationships is constructed. This matrix is the custom fuzzy thesaurus which is used to measure the partial similarity and exact match between attributes in the feature space and single terms in each single Tweet.

2. Generating Feature Vectors with Semantic Similarities

Once the fuzzy thesaurus is constructed, every feature, f_i, is compared with every word, wrd_j, in a Tweet, d, to retrieve the corresponding distance correlation factor Cf_{ij} from the custom fuzzy thesaurus which indicates the word-word semantic similarity.

Once a feature, f_i, is compared to each word, wrdj, in a given Tweet, d, the semantic similarity between the feature and the whole Tweet is calculated using Equation (4), which indicates the word-sentence semantic similarity. This is performed for each feature in the feature space against each single Tweet in the corpus as illustrated in Figure 13.

By doing so, we account for the semantic relationship between each feature with each single Tweet in the corpus allowing for analyzing the overall context instead of just considering the occurrence or the frequency of features in each Tweet.

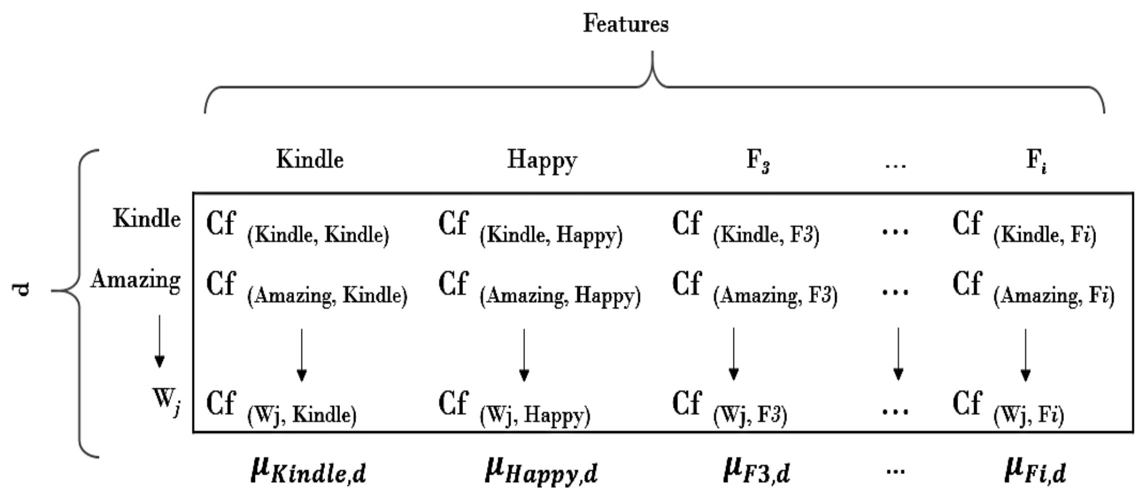


Figure 13: Calculating the word-sentence semantic similarity (μ_(F,d)) between each feature (Fi) in the feature space and each Tweet (d) in the Twitter corpus.

4.4.4 Experimental Work

In this section, the benchmark datasets used in the experiments, the baselines, sentiment replacement and preprocessing, classification based on a fuzzy thesaurus, and finally evaluation measures are introduced.

1. Dataset

We use the STS-Gold Tweet[1] dataset and the Stanford Twitter Sentiment (STS)[2,3] testing dataset to evaluate the effectiveness of the proposed method (Table 7). The STS-Gold Tweet dataset contains 2032 randomly collected Tweets, manually annotated into positive and negative by three annotators. All the annotators agree on the sentiment of the Tweets in the dataset. The Stanford Twitter Sentiment testing set consists of 359 Tweets collected by searching Twitter API with specific queries including products names, companies and people. They are also manually annotated into positive and negative. The original Stanford training dataset is not used, because it is automatically annotated using emoticons. Although automatic sentiment annotation of Tweets using emoticons is fast, its accuracy is arguable because emoticons might not reflect the actual sentiment of Tweets [160]. Another limitation of the Stanford original training set is that the set was automatically annotated based on emoticons, but then the emoticons were removed, hence, if we train a classifier on the Stanford training dataset it will not recognize the emoticons that were initially used for class labeling. Therefore, in this study, only the STS testing dataset, and STS_Gold Tweet dataset are considered applying a 10-fold cross validation to both.

Table 7: Statistics of the Twitter datasets used in this research

Dataset	Number of Tweets	Positive	Negative	Type
STS-Gold Tweet	2032	632	1400	10-fold cross validation
Stanford Twitter Sentiment (STS) – Testing Set	359	182	177	10-fold cross validation

[1] STS-Gold dataset can be requested from the authors at: http://kmi.open.ac.uk/people/member/hassan-saif
[2] Stanford dataset official page: http://help.sentiment140.com/for-students
[3] Stanford testing and training datasets can be downloaded from: https://docs.google.com/file/d/0B04GJPshIjmPRnZManQwWEdTZjg/edit

2. Baselines

We compare the performance of our approach using the fuzzy thesaurus and sentiment replacement against the baselines described below. Even though word unigrams are the simplest features used for sentiment analysis of Tweets data, there is evidence that using n-gram features may hinder the accuracy of Twitter sentiment analysis due to the large number of infrequent words and that unigrams produce better accuracy results [161], [162]. In addition, models trained from word unigrams outperform random classifiers by a decent margin of 20% [163]; hence, only unigram features are used. Sentiment replacement is not performed for the baselines.

A) First Baseline - Unigrams Features with Polarity

We use the NB classifiers and the SVM classifier trained from word unigrams on polarity dataset as our first baseline model. Polarity indicates whether a feature occurs or not in a Tweet.

B) Second Baseline - Unigrams Features with Frequencies

We use the NB classifiers and the SVM classifier trained from word unigrams on frequency dataset as our second baseline model. Frequencies indicate how many times a feature occurs in a Tweet.

C) Third Baseline - Unigrams Features with TF/IDF

We use the NB classifiers and the SVM classifier trained from word unigrams on a term frequency inverse document frequency (TF-IDF) dataset as our third baseline model.

TF-IDF is a measure that is intended to reflect how important a word is to a document in a collection or corpus. TF-IDF is calculated as follows:

- *TF(t,d) = Term Frequency(t,d)*: is the number of times that term t occurs in document d.
- *IDF(t,D) = Inverse Term Frequency(t,D):* measures the importance of term t in all documents (D); this measure is obtained by dividing the total number of documents (N) by the number of documents containing the term (DF), and then taking the logarithm of that quotient.

$$IDF(t,D) = \log_2 (N/DF)$$

- Finally, the weight is obtained by multiplying the two measures:

$$TF\text{-}IDF(t,d) = TF(t,d) * IDF(t,D)$$

<u>3. Sentiment Replacement, Preprocessing and Feature Reduction</u>

Initially all slogans and abbreviations that have sentiment meaning are searched in the raw Twitter corpus and are replaced with their sentiment equivalence following the slogan list available in [164]. Once the sentiment replacement is done, natural language processing [146] of the Twitter data is performed. In Table 8 lists of APIs and techniques used for preprocessing and feature extraction.

Table 8: APIs and techniques applied for NLP

Natural Language Processing Task	API/Technique
Stopwords Removal	Apache Lucene Core 5.3.0[4]
Stemming	Porter Stemming Algorithm[5]
Unigram Extraction	Apache Lucene Core 5.3.0
Equivalence Class Replacement	Java Regex[6]

[4] https://lucene.apache.org/core/
[5] https://tartarus.org/martin/PorterStemmer/java.txt
[6] More about regex can be found at: https://docs.oracle.com/javase/tutorial/essential/regex/

To illustrate the impact of sentiment replacement on reducing feature space dimensionality, Table 9 summarizes the effect of preprocessing and feature reduction on reducing dimensionality of the original feature space on the Stanford Testing dataset. After completing all the preprocessing steps, the feature space size is reduced by 41.26%. The most significant contributor to the feature space dimensionality reduction is the sentiment replacement of slogans, abbreviations, and emoticons. The same steps are applied to the STS_Gold dataset.

Table 9: Effect of preprocessing and feature reduction on the feature space size of STS

Pre-Processing / Feature Reduction	Feature Space Size	% of Reduction
None	2455	0%
Sentiment Replacement of Slogans, Abbreviations, and Emoticons	1593	35.11%
User Names	1605	34.62%
URL	1614	34.26%
Hashtags	1678	31.65%
Repeated Letters	1682	31.49%
All	1442	41.26%

4. Sentiment Classification

We developed a Java program using JDK 8 and JRE 8 on a 2.6 GHz PC running Windows 10 to build the fuzzy thesaurus and generate the semantic feature vectors (SFV) from a Twitter corpus. Figure 14 shows the algorithm for generating semantic feature vectors (SFV). The algorithm takes as inputs the following:

1. Twitter data consisting primarily of messages and sentiment class. Additional data can be present such as user ID, hashtags, queries, etc., which will be preprocessed during natural language processing phases.
2. List of slogans, abbreviations and emoticons with their corresponding sentiment meaning.

In the implementation, the following data structures are used for inputs:

1. *T*: Twitter data represented in a LinkedList of String arrays, where, each node *d* holds a single Tweet from the Twitter corpus.
2. *ASEL*: slogans, abbreviations and emoticons represented in a String array. The ASEL available in [164] is used.

In the intermediate steps, features, F, are represented using LinkedList of Strings, the fuzzy thesaurus composed of all Cf values is represented using a hash table, and word-document-position vectors (WDPV), illustrated earlier in Table 5 and Table 6, are represented using user-defined data types. As an output, the algorithm returns semantic feature vectors (SFV) and exports them to a comma-separated file ready for classification. Subsequently, we used Weka 3.8 [165] to train the classification model and tested it with a 10-fold cross validation.

Table 10 to Table 13 show the classification results of BNB, MNB, and SVM classifiers trained on unigrams with polarities, frequencies, TF-IDF, and Semantic Feature Vectors (SFV) using a 10-fold cross validation before and after applying an Information Gain (IG) attribute selection filter. For sentiment mining, this size of corpus may not provide sufficient coverage of representative sentiment terms and contexts. Therefore, we choose to apply attribute selection filter to eliminate the effect of sentimentally insignificant attributes. Information Gain (IG) is used to select subsets of features that are highly correlated with the class while having low inter-correlation. In other words, the features with the highest information gain are selected and those with very low information gain are removed from the feature space [166].

Algorithm 1 Generating Feature Vectors based on Semantic Similarities

```
for each word wrd_j ∈ T do
    if wrd_j ∈ ASEL then,
      replace wrd_j with equivalent sentiment
    end if
end for
for each Tweet, d = (d_1, d_2, ..., d_n) ∈ T do
    replace equivalent classes
    remove stopwords
    perform stemming
    generate word-document-position vectors → WDPV
    extract unigram features → F
end for
for each word wrd_j ∈ WDPV do
    for each word wrd_i ∈ WDPV do
      if wrd_j & wrd_i 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
      end if
    end for
end for
for each Tweet, d = (d_1, d_2, ..., d_n) ∈ T do
    for each Feature, f= (f_1, f_2, ..., f_i) ∈ F do
      calculate Feature-Tweet semantic similarity → μ_{Fi,dn}
    end for
end for
return semantic feature vectors (SFV)
```

Figure 14: Generating feature vectors based on semantic similarities

5. Evaluation Measures

The type of classification we conduct on Twitter is a typical form of a binary classification in which the input, Tweet, is to be classified into one, and only one, of two non-overlapping classes (positive, negative). There exist several performance measures used with binary classifiers in different areas of application such as F-Score, Precision, Recall, and Specificity.

Opinion or sentiment mining deals with meanings that are most of the time indirect (i.e., implied) and complex (i.e., opinions and emotions are not easy to interpret from text).

So far, there is no consensus on the choice of measures used to evaluate the performance of classifiers in opinion, subjectivity and sentiment analysis [167]. However, it was found that most of the work on sentiment analysis uses accuracy as the measure of overall effectiveness of a classifier in sentiment analysis [157, 158, 168, 169]. Two more useful metrics are added, precision and recall, that measure class agreement of the data labels with the positive labels given by the classifier and effectiveness of a classifier to identify positive labels respectively. Results are discussed in the following section.

6. Discussion and Comparison with Previous Work

Based on the results, semantic feature vectors (SFV) consistently achieved the best accuracies with different classifiers, i.e., SVM and MNB, on the Stanford Testing Dataset compared to the polarity and frequency feature vectors, using the full feature space as illustrated in Table 10, or using selected features as illustrated in Table 11. The TF-IDF feature vectors, however, outperformed the semantic feature vectors using the full feature space. Yet, semantic feature vectors significantly outperformed TF-IDF feature vectors on selected features.

Using the larger STS_Gold dataset, semantic feature vectors (SFV) achieve slightly better or comparable results to the baselines as illustrated in Table 12 with the full feature space. However, by using selected features, semantic feature vectors (SFV) significantly outperform all the baselines using different classifiers, SVM, BNB, and MNB (Table 13). It is worth noticing that with a larger dataset, the classification accuracy drops significantly with the TF-IDF-based datasets. On the other hand, with the SFV-based datasets, the classification accuracy remains consistent at acceptable levels. Consistent levels of accuracies are desirable especially in sentiment analysis of social networks since the size of data is usually very large. Moreover, it is noticeable that semantic feature vectors (SFV) always achieve the best results with significant improvement in accuracy with highly correlated set of features with the class label, i.e., those features that are expected to be strongly defining the semantics of the Tweet. Other dataset representations, e.g., polarity, do not exhibit comparable improvement.

Our results compare favorably with other research work conducted on similar datasets. Go et al. [158] achieved the maximum accuracy of 83% using MaxEnt trained on a combination of unigrams and bigrams using the Stanford Dataset. Our method outperforms the original results produced by Go et al. with maximum accuracy of 84.96 % using SVM classifier. Amongst other research work that compared their results with the Stanford STS Dataset, Speriosu et al.[170] tested on a subset of the Stanford Twitter Sentiment test set with 75 negative and 108 positive Tweets. They reported the best accuracy of 84.7% using label propagation on a rather complicated graph that has users, Tweets, word unigrams, word bigrams, hashtags, and emoticons as its nodes. Also, our results outperform Speriou’s results using a simpler logic.

Table 10: – Stanford Testing Set - all features

Unigrams – 1442 Features		**BNB**	**SVM**	**MNB**
Polarity-Based Baseline	Accuracy	76.60 %	74.37%	79.38 %
	Recall	0.766	0.744	0.794
	Precision	0.766	0.744	0.795
Frequency-Based Baseline	Accuracy	74.37%	71.86%	79.94 %
	Recall	0.744	0.719	0.799
	Precision	0.745	0.719	0.8
TF/IDF- Based Baseline	Accuracy	76.88%	77.99%	81.89%
	Recall	0.769	0.780	0.819
	Precision	0.769	0.780	0.819
Semantic Feature Vectors - SFV	Accuracy	**71.87%**	**74.65%**	**80.78%**
	Recall	0.719	0.747	0.808
	Precision	0.719	0.747	0.809

Table 11: - Stanford Testing Set – selected features.

Unigrams – Selected Features using (IG)		**BNB**	**SVM**	**MNB**
Polarity-Based Baseline	Accuracy	80.2%	81%	81.62%
	Recall	0.802	0.811	0.816
	Precision	0.844	0.855	0.855
Frequency-Based Baseline	Accuracy	77.15 %	79.10%	82.17 %
	Recall	0.772	0.791	0.822
	Precision	0.785	0.828	0.851
TF/IDF- Based Baseline	Accuracy	80.78%	81.89%	81.62%
	Recall	0.808	0.819	0.816
	Precision	0.842	0.846	0.850
Semantic Feature Vectors - SFV	Accuracy	**<u>77.99 %</u>**	**<u>84.96 %</u>**	**<u>83.29 %</u>**
	Recall	0.78	0.85	0.833
	Precision	0.808	0.869	0.856

Table 12: - STS Gold - all features.

Unigrams – 3850 Features		**BNB**	**SVM**	**MNB**
Polarity-Based Baseline	Accuracy	75.78 %	80.17 %	81.1 %
	Recall	0.758	0.802	0.811
	Precision	0.75	0.796	0.807
Frequency-Based Baseline	Accuracy	74.60%	81.25 %	80.70 %
	Recall	0.746	0.813	0.807
	Precision	0.747	0.808	0.806
TF/IDF- Based Baseline	Accuracy	64.03%	79.33%	77.41%
	Recall	0.640	0.793	0.774
	Precision	0.729	0.787	0.768
Semantic Feature Vectors - SFV	Accuracy	**<u>73.75 %</u>**	**<u>80.5 %</u>**	**<u>80.44 %</u>**
	Recall	0.737	0.805	0.804
	Precision	0.774	0.804	0.808

Table 13: - STS Gold – selected features.

Unigrams – Selected Features using (IG)		**BNB**	**SVM**	**MNB**
Polarity-Based Baseline	Accuracy	75.29 %	79.87 %	80.56 %
	Recall	0.753	0.799	0.806
	Precision	0.74	0.796	0.813
Frequency-Based Baseline	Accuracy	77.21 %	81.5 %	82.03 %
	Recall	0.772	0.815	0.820
	Precision	0.763	0.815	0.823
TF/IDF- Based Baseline	Accuracy	79.23%	77.36 %	75.49%
	Recall	0.792	0.774	0.755
	Precision	0.785	0.780	0.804
Semantic Feature Vectors - SFV	Accuracy	**<u>80.54%</u>**	**<u>81 %</u>**	**<u>82.17 %</u>**
	Recall	0.805	0.809	0.822
	Precision	0.801	0.807	0.818

4.4.5 Conclusion

Twitter is one of the most popular social networks where users can express their opinions about a boundless number of topics. This wealth of public opinion attracts vast interest in sentiment analysis of Twitter data. Machine learning approaches for sentiment analysis rely on feature vectors extraction to represent the most relevant and important text features that can be used to train classifiers, such as Naïve Bayes (NB) and Support Vector Machines (SVMs). Feature vector extraction eliminates many semantic relationships in the text. Yet, in many cases, the sentiment conveyed by a word is implicitly associated with the semantics of its context. Several methods reported in the literature for incorporating semantics in sentiment analysis suffer from several drawbacks including costly manual intervention, domain dependence, and limited predefined knowledge bases.

In our research, fuzzy thesaurus can be used for constructing Twitter feature vectors for sentiment classification. The experimental results show that the semantic feature vectors (SFV) consistently produce better results than the baselines. Also, comparison with previous work shows that the proposed method outperforms other methods reported in the literature using the same benchmark data.

In the following chapter, we explore how fuzzy thesaurus can be used effectively to analyze semantics in the research problem. Fuzzy thesaurus is used for semantic analysis at a larger scale for personalized recommendations in massively diverse information wikis.

Chapter 5: A Framework for Personalized Content Recommendations to Support Informal Learning in Massively Diverse Information Wikis

5.1 Background

Considerable research efforts were made to extrapolate and analyze navigation behavior of web users [171-174], not necessarily specific to learning contexts. Outcomes of these studies mainly support better design and structuring of web pages on websites for improved accessibility and usability. In these cases, individual user's navigation pattern is not the concern, rather, results are usually used to analyze interesting topics, web pages, and websites' features as perceived by large numbers of users to provide better browsing experiences for millions of users. On the other hand, some research works analyzing learners' navigation behavior on the web attempted to understand how different navigation patterns can relate to different learners' attributes [175-177]. Outcomes of these studies support the assumption that different learners adopt different navigational patterns based on some cognitive differences. For example, Jens and Thomas [176] found that learners classified as "Explorers" tend to "jump" more to create their own path of learning, while learners classified as "Observers" tend to follow the suggested path by clicking on the "Next" button. Moreover, West and Leskovec [50] have compared human navigation in information networks such as Wikipedia with that of software agents and found that humans, when navigating within an information network, have expectations about what links should exist and base a high-level reasoning plan upon this, and then use local information to navigate through the network.

These studies analyzing users' navigation on the web lead to some useful conclusions that can help in redesigning websites for better usability or redesigning learning environments to cater for different cognitive styles of users. They also support the assumption that users' navigation can unveil important user traits and characteristics that can be used for personalization purposes.

5.2 Modeling Users' Interests based on Topical Navigation Graphs

Massive amounts of information can be organized in some sort of graph structure. For example, webpages in the World Wide Web, quests in a game, users and content in a social network, courses in an educational program, or topics learned from a specific lesson. In these networks (or graphs), each node represents an entity or a piece of information, and each link represents a tie or relationship between two entities. Considerable research works focused on investigating these graphs to infer useful information in various fields of applications. Page et al. [178] in their seminal paper "Bringing order to the web", introduced the PageRank algorithm for analyzing the web as a network of interconnected webpages and assigning ranks to webpages based on web users' accesses which had revolutionized searches on the web. Different variations of PageRank algorithm were introduced, e.g. [179] and [180], and numerous applications to infer useful knowledge from graphs were introduced, e.g. [170] and [181]. In our research we are mainly interested in *topical graphs* generated through learners' free navigation to infer some insight into what learners are interested to learn. There exist some research efforts in different domains focusing on utilizing topical graphs for eliciting important knowledge about specific users [182] and [183]. Beal et al. [183] utilized mind maps generated by researchers based on Docear's research paper system to provide content-based research papers recommendations considering only the content of the mind map without analyzing the structure of the mind map. Docear's mind map-based research paper recommender system proved to be more successful than citation-based and keyword-based recommender systems used in other research papers management systems providing more insight and better understanding of what researcher are interested to learn. On the other hand, Zualkernan et al. [182] proposed that the closer two concepts in the user's topic map are the closer their semantic relationship will be and hence the more similar their search results should be. In addition, Leak et al. [184], [54] studied further concept map's structural influences considering incoming and outgoing connections and proposed three models that helped assigning structural or topological weights to every concept in the map and validated their models with comprehensive user studies. These studies provided evidence on the effectiveness of topical graph structures or topologies in eliciting weighted values reflecting individual user's priorities or rankings of different topics.

In all these research works [182-184, 54], topical graphs are created by users requiring the user to explicitly and frequently inputting his/her topical graphs into the system which is time and effort consuming. However, in the proposed framework, topical navigational graphs are implicitly extrapolated and analyzed, without the user intervention, mainly based on a user's free behavior on the learning environment. The proposed method is based on the following assumptions:

1. A learner's dynamic behavior can be used to dynamically model the learner.
2. In informal learning environments, the most common type of behavior is navigation. According to [174], navigational related events, which brought the total number of events to 31,134 representing 73% of all generated events.

3. We define navigation as the traversal process of moving from one learning resource to another.
4. Learning resources are webpages identified by their topics as well as access requests.
5. Navigation is characterized by the sequence of learning resource accesses, thus, by a sequence of topics.
6. Navigation is modeled per one learning session.

5.3 Proposed Personalized Content Recommendations Framework (PCRF)

The PCRF first captures raw learning interests for every individual learner in a topical navigation graph (TNG) by tracking individual learning sessions. The learner navigation is modeled as a directed multigraph, TNG (V, E). Every vertex ,V, in TNG corresponds to a topic, topics are 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. [54], 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 [55]. The resulting set of ranked and semantically relevant topics represents the final personalized content recommendations.

Our framework is composed of four main modules: session tracking, TNG analyzer, personalization, and semantic analysis modules. Figure 15 illustrates our conceptualization of the proposed framework. The semantic analysis module is designed to be used offline to build and process custom corpora and generate inverted indices of topics used online by the personalization module to generate personalized content recommendations based on the learner models generated by the TNG Analyzer module. Each module is described in the following sections. Table 14 lists and defines the main concepts used in this research.

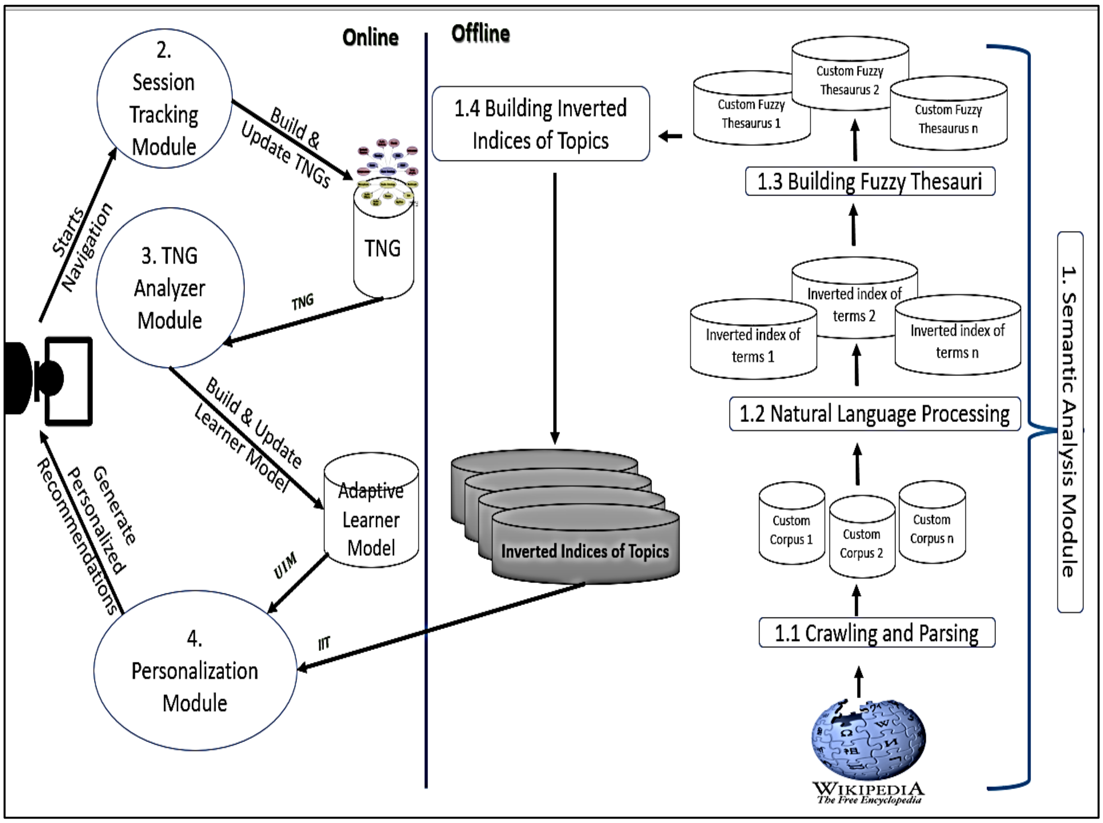


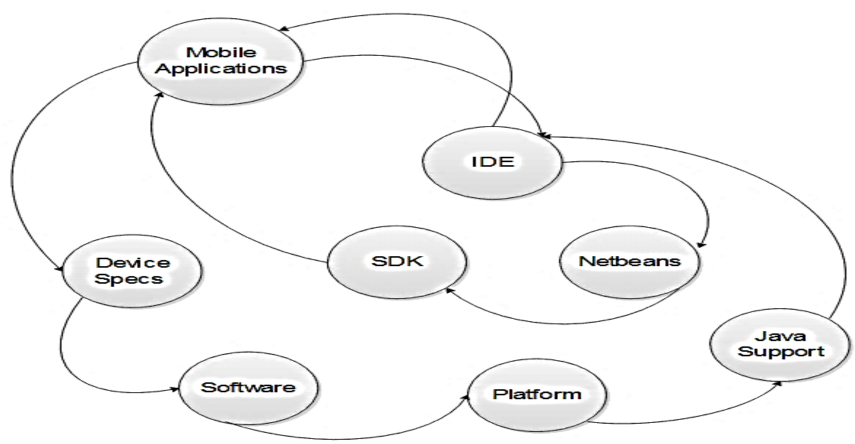
Figure 15: Proposed personalized content recommendation framework (PCRF)

Table 14: 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:* ▪ $W(v) = \left(\alpha \,.\, o(v) + \beta \,.\, i\,(v)\right).\left(\frac{1}{d(v)+1}\right)^{\frac{1}{\delta}}$ *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*, $W(uim_i)$.	▪ $PCR_i = W(\{uim_i : uim_i \in \mathrm{UIM}_i\}) * Sim(\{l_i : l_i \in \mathrm{UIM}_i\}, \{d_n : d_n \in IIT\})$

Table 15 illustrates a motivating example of a user navigating a website about mobile application development. This example is used to illustrate the different phases in the proposed framework.

Table 15: 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.
		

5.3.1 Semantic Analysis Module

We perform semantic analysis using fuzzy thesauri built based on fuzzy set information retrieval model as explained in Chapter 4. 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 17. The algorithm is explained in Figure 19. The process explained here can be used for any other context, i.e. other than Wikipedia recommender systems, because Wikipedia is considered to be a comprehensive and representative corpus especially for English language.

First, custom corpora are extracted from Wikipedia for each main topic category as classified by Wikipedia using a web scraper application. Figure 16 illustrates the 22 main topic categories under which all Wikipedia content is classified as of March 2018. The purpose of these categories is to group major topic classifications in one place, for greater ease and for reference of users and editors of Wikipedia. From this step, a custom corpus is generated 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>, is extracted. Index pages are excluded from the corpora as they do not have any learning content[7]. The third step aims to generate inverted indices of unique terms that can be used to build the fuzzy thesauri. At this stage, natural language processing [146] of the

[7] Processed corpus will be made available online for future experiments.

custom text-based corpora is performed. Generally, unstructured texts cannot be directly processed for semantic analysis. Thus, several natural language processing tasks are performed, which proved effective and have become a common practice for unstructured text preprocessing for semantic analysis.

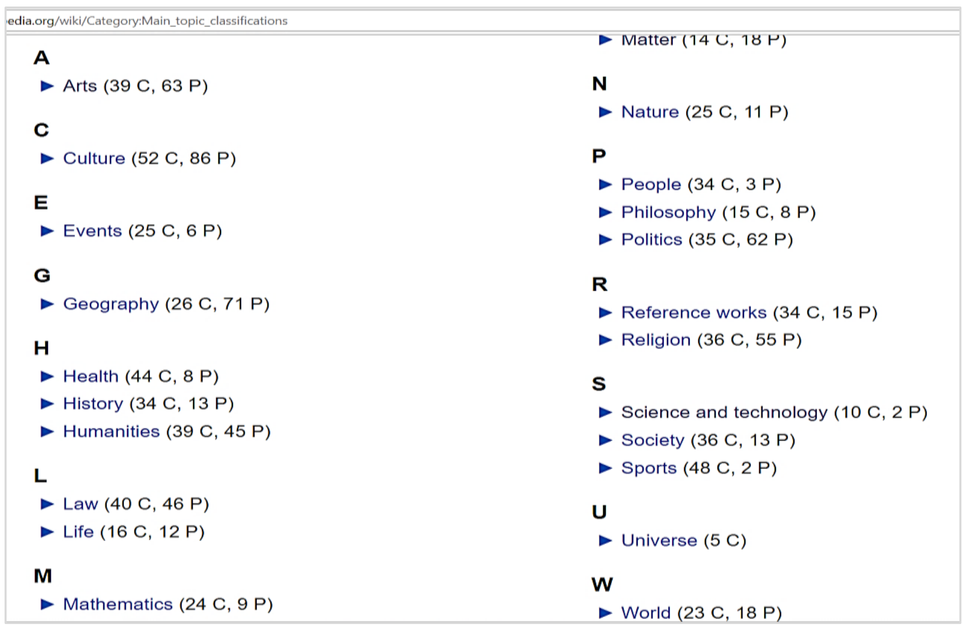


Figure 16: Wikipedia main topic taxonomy as of March 2018

We perform the following preprocessing tasks in order:

1. *Tokenization*: all documents are converted into vectors of raw unprocessed terms, 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", "by".

3. *Stemming* [159]: 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. Detailed explanation of inverted word indices is presented in Chapter 4, Section 4.3.1.

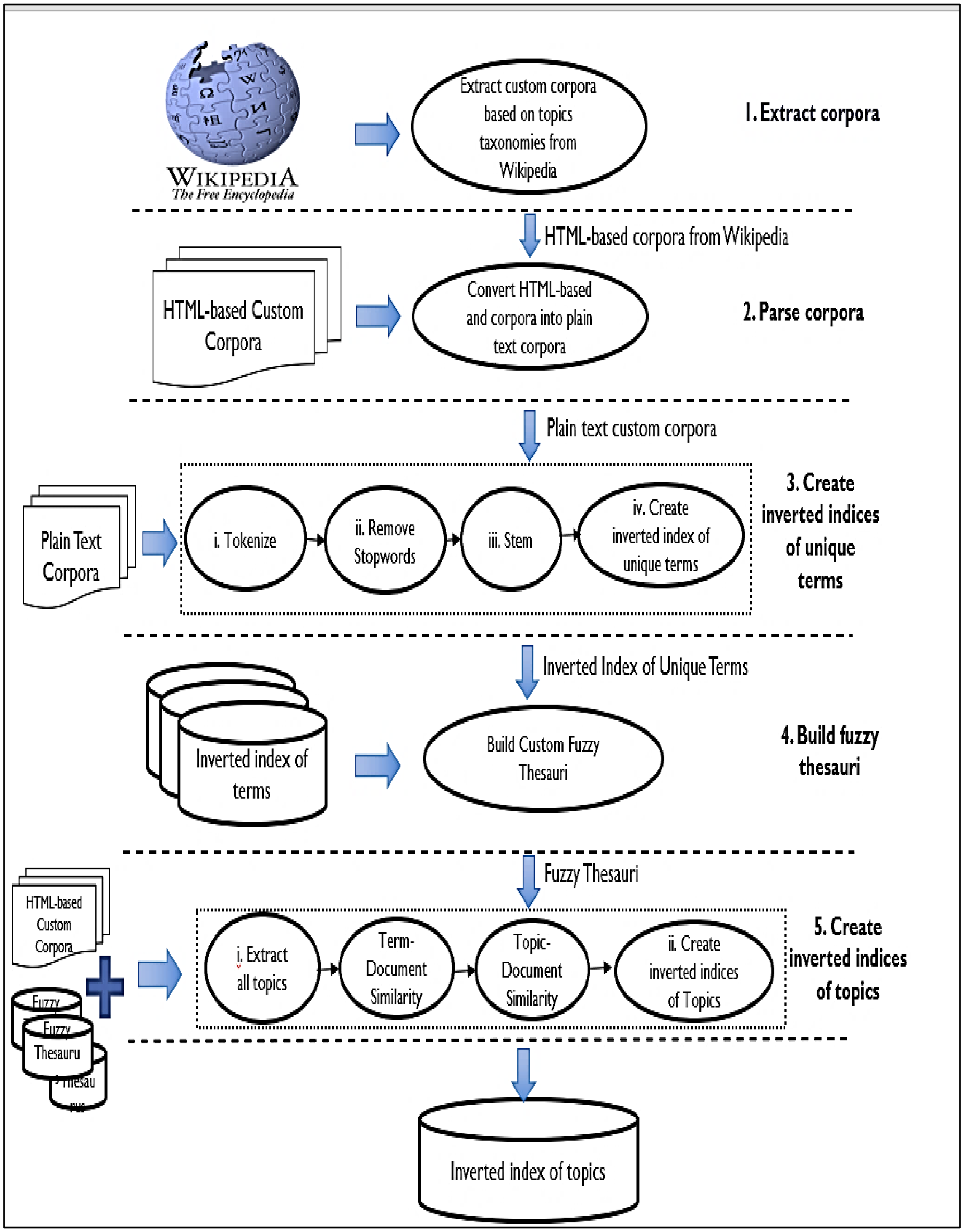


Figure 17: Process of building the fuzzy thesauri and generating the inverted indices of topic

In step four, a custom fuzzy thesaurus is built that defines the semantic similarity between each two distinct words for each custom wiki corpus. Using the inverted indices of distinct terms, *VEC*, we define for every pair of keywords across all documents within a single corpus: the frequency of co-occurrence and relative distance in a single document (C_{ij}), Equation (1), the normalized value (nC_{ij}), Equation (2), and finally the distance correlation factor (Cf_{ij}), Equation (3), explained earlier in Chapter 4, Section 4.3.1. 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 aims to generate 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 18.

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

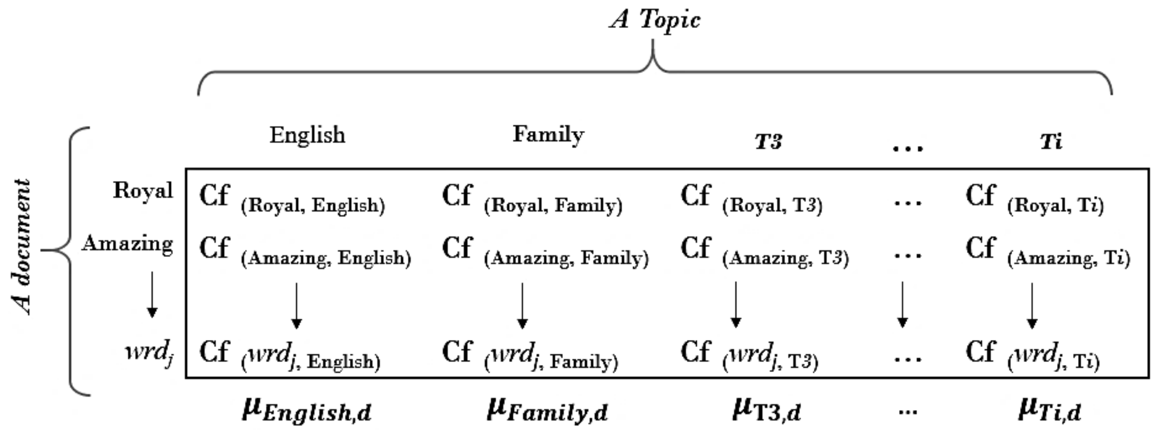


Figure 18: Calculating the Term-document semantic similarity (μ_(T,d)) between each term (Ti) in a given topicn and each document (dn) in the wiki corpus

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, $Sim\ (\text{topic}_n, d_n)$ as follows:

$$\boldsymbol{Topic - Document\ Similarity} \ = \ Sim\ (\text{topic}_n, d_n) = \frac{\mu_{T_1,d_n} + \mu_{T_2,d_n} + \cdots + \mu_{T_i,d_n}}{i} \tag{5}$$

This value is 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 16 shows sample entry in the inverted topic index for the topic “Amazon River”.

Table 16: 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

Algorithm 2 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 19: Algorithm for generating inverted indices of topics

5.3.2 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. The learner navigation is modeled as a directed multigraph, TNG (V, E). Multigraph is used 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, continuous throughout the learning session and unconstrained by time. In Table 15 for example, it can be seen 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.

5.3.3 TNG Analyzer Module

We adapt The Hub-Authority and Root-Distance Model (HARD), and The Connectivity Root-Distance Model (CRD) concept maps' topological analysis models from Leak et al. [184], [54] 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 frequently visited by a user might be of a special interest compared to topics or webpages 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 this research, concepts are treated 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, the process of identifying the structural characteristics and calculating the structural weights is explained using the same example illustrated in Table 15 to demonstrate the different phases of *TNG* analysis.

1. Structural Characteristics Definition

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 in the graph illustrated in Table 15, 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. Look at Table 17 for position characteristics for some nodes in "Mobile Applications" graph presented in Table 15.

Table 17: 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, in the HARD model, nodes are characterized as hub, authority, and upper nodes. In [184] HITS iterative algorithm is adapted to calculate the relative hub, authority, and upper nodes' positional weights. Leak et al. in [184] 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 currently analyzed in the graph. Hence, the weight of node *q* is be expressed as *w(q)* and the link between node *q* and node *p* is be 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 every step later on.

Step 3: *Calculate Hub_Weight such that: Hub_Weight of a node, p, is the sum of Authority_weight of all nodes,* $q_1, q_{2,\dots,} q_n$ *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* $q_1, q_{2,\dots,} q_n$ *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:*

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

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:*

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

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*

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

2. Topological Weights Calculations

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

For the CRD Model:

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

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, parameters α, β, and γ reflect the influences of different roles a node can play in *TNG*. In [54], a hill-climbing algorithm was used to determine the best parameter settings for the CRD and the HARD models which gave the best fit between the models and user data (Table 18).

Table 18: Best fit values for parameters α, β, and γ for CRD and HARD models

Model	α	β	γ
CRD	0.930	4.959	3.603
HARD	0.0	2.235	1.764

Based on the generated weights for the topics in the navigation graph, the nodes with highest weights are selected to represent the topics of most interest to the learner forming a user interest mode, *UIM*. Table 19 shows structural weights of the topics in the "mobile applications" navigation graph. It can be seen how CRD model gives highest weight to the root topic, "Mobile Application", that was first visited by the user compared to node, "Netbean", ranked top by HARD Model because of its higher aggregate hub weight coming from important hubs in the graph namely, "IDE" and "Mobile Application". A question that needs to be answered in this research is whether CRD or HARD models provide better ranking of recommendations as perceived by users.

Table 19: Structural weights of different nodes in the navigation pattern – "Mobile Application Navigation Graph"

Node	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

5.3.4 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, $W(umi_i)$ as follows:

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

Therefore, learning documents with higher semantic similarities to topics in the user model (*UIM*) are retrieved and form a set of ranked personalized content recommendations. Adaptation is accomplished through continuous update of *TNG* as well as *UM* and, accordingly, the structural weights, hence, the personalized topics. The algorithm is explained in Figure 20. Figure 21 and Figure 22 illustrate how user interests are elicited from a user's navigation on our test environment (http://www.theknowledge.site). Differences in structural weights, selected topics, and personalized recommendations become more significant as the size of the navigation graph grows.

Algorithm 3 Generating Personalized Content Recommendations (PCR)

1. **Input:**
IIT: inverted indices of topics
Output:
*PCR_i: Personalized Content Recommendations for user **i***
2. **while** *(user in the wiki domain)*
*Initialize **TNG_i: TNG_i = (V, E)***
OnPageLoad *// HTML event*
3. **Procedure** Generate_Personalized_Content_Rec (***IIT***)
4. *Extract main topic from the webpage → **l_i***
5. **If ($l_i \in IIT$)** *// Webpage contains valid learning topic*
6. *Update **E***
7. *Calculate **W(v)*** *// Using CRD and HARD models*
8. *Update **V_i**: **v** =* $(\boldsymbol{l_i}, \boldsymbol{W}(\boldsymbol{v}))$
9. *Update **TNG_i**: **TNG_i = (V_i, E)***
10. *Update **UIM_i**:* $\boldsymbol{UIM_i} \subseteq \boldsymbol{V_i} : \forall\, \boldsymbol{uim_i} \in \boldsymbol{UIM} \ \wedge\ \forall\, \boldsymbol{v_i} \in (\boldsymbol{V} - \boldsymbol{UIM}),\ \boldsymbol{W}(\boldsymbol{uim_i}) > \boldsymbol{W}(\boldsymbol{v_i})$
11. *Generate* $\boldsymbol{PCR_i} = \boldsymbol{W}(\{\boldsymbol{uim_i}: \boldsymbol{uim_i} \in \mathbf{UIM}_i\}) * \boldsymbol{Sim}(\{\boldsymbol{l_i} : \boldsymbol{l_i} \in \mathbf{UIM}_i\}, \{\boldsymbol{d_n} : \boldsymbol{d_n} \in \boldsymbol{IIT}\})$
12.
13. **end if**
14. **return PCR**
15. **end while**

Figure 20: Algorithm to generate PCRF

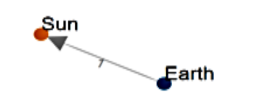
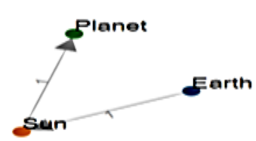
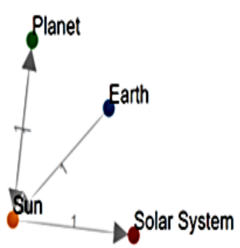
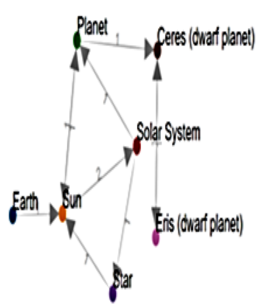
User #1 TNG	Graph Structure	HARD Topological Weights	Selected Topics of Interests	Recommended Topics
	V = { Earth, Sun} E = { (Earth, Sun)}	W (Earth) = 0.83 W (Sun) = 1.00	Earth Sun	1. Earth's Atmosphere 2. Flat Earth
	V = { Earth, Sun, Planet} E = { (Earth, Sun), (Sun, Planet)}	W(Earth)= 0.53 W(Sun)= 0.75 W(Planet) = 1	Earth Sun Planet	1. Ceres (Dwarf Planet) 2. Definition of Planet 3. Earth's Atmosphere
	V = { Earth, Sun, Planet, Solar System} E = { (Earth, Sun), (Sun, Planet), (Planet, Sun), (Sun, Solar System)}	W(Earth)= 0.38 W(Sun)= 0.66 W(Planet) = 0.92 W(Solar System) = 1	Sun Planet Solar System	1. Solar Eclipse. 2. Ceres (Dwarf Planet). 3. Timeline of Discovery of Solar System planets. 4. Earth's Atmosphere.
	V = { Earth, Sun, Planet, Solar System, Star, Ceres (Dwarf Planet), Eris (Dwarf Planet)} E = { (Earth, Sun), (Sun, Planet), (Planet, Sun), (Sun, Solar System), (Solar System, Star), (Star, Sun), (Sun, Solar System), (Solar System, Planets), (Planets, Ceres (Dwarf Planet)), (Ceres (Dwarf Planet), Eris (Dwarf Planet), (Eris (Dwarf Planet), Ceres (dwarf planet)) }	W(Earth)= 0.25 W(Sun)= 0.58 W(Planet) = 0.69 W(Solar System) = 0.80 W(Star)= 0.87 W(Ceres (dwarf planet))= 0.94 W(Eris (dwarf planet))= 1	Eris (Dwarf planet) Ceres (dwarf planet) Star	1. Ceres (Dwarf Planet). 2. Eris (Dwarf Planet). 3. Binary Star 4. Solar Eclipse 5. Ceres

Figure 21: Illustration of user interest modeling and personalized content recommendations using HARD Model

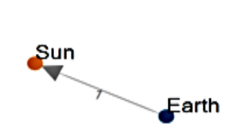
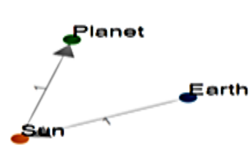
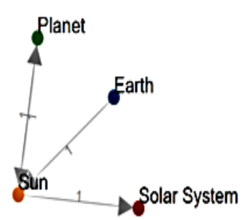
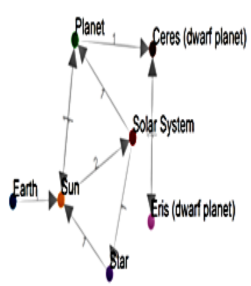
User #1 TNG	Graph Structure	CRD Topological Weights	Selected Topics of Interests	Recommended Topics
	V = { Earth, Sun} E = { (Earth, Sun)}	W (Earth) = 1 W (Sun) = 0.055	Earth Sun	1. Earth's Atmosphere 2. Flat Earth
	V = { Earth, Sun, Planet} E = { (Earth, Sun), (Sun, Planet)}	W(Earth)= 1 W(Sun)= 0.5 W(Planet) = 0.037	Earth Sun Planet	1. Earth's Atmosphere 2. Flat Earth 3. Ceres (Dwarf Planet).
	V = { Earth, Sun, Planet, Solar System} E = { (Earth, Sun), (Sun, Planet), (Planet, Sun), (Sun, Solar System)}	W(Earth)= 1 W(Sun)= 1 W(Planet) = 0.74 W(Solar System) = 0.018	Sun Earth Planet	1. Earth's Atmosphere 2. Flat Earth 3. Ceres (Dwarf Planet). 4. Solar Eclipse.
	V = { Earth, Sun, Planet, Solar System, Star, Ceres (Dwarf Planet), Eris (Dwarf Planet)} E = { (Earth, Sun), (Sun, Planet), (Planet, Sun), (Sun, Solar System), (Solar System, Star), (Star, Sun), (Sun, Solar System), (Solar System, Planets), (Planets, Ceres (Dwarf Planet)), (Ceres (Dwarf Planet), Eris (Dwarf Planet), (Eris (Dwarf Planet), Ceres (dwarf planet)) }	W(Earth)= 0.64 W(Sun)= 1 W(Planet) = 0.64 W(Solar System) = 0.22 W(Star)= 0.1 W(Ceres (dwarf planet))= 0.06 W(Eris (dwarf planet))= 0.05	Sun Earth Planet	1. Earth's Atmosphere 2. Flat Earth 3. Ceres (Dwarf Planet). 4. Solar Eclipse. 5. Binary Star

Figure 22: Illustration of user interest modeling and personalized content recommendations using CRD Model

5.4 Implementation of PCRF

5.4.1 Online Module

PCRF online module is a web application hosted on Apache web server. PCRF's code is primarily written in JavaScript and PHP. However, the semantic analysis module is primarily written in JAVA and run on desktop. For now, inverted indices of topics are upload manually to web servers. However, this communication can be made automatic in the future with a Web service. Table 20 and Table 21 list the most important 'get' and 'post' calls. Figure 23 illustrates the three-tier architecture of PCRF. As shown in Figure 23, *on every page load* in the client side (browser):

1. A javascript call goes to XMLHttpRequest object.
2. HTTP Request is sent to the web server by XMLHttpRequest object.
3. Calculate_and_List.js script extracts the current topic and initiates a *post* request to visitor.php to store the current visited topic by the user.
4. Calculate_and_List.js script also initiates two get requests to retrieve and generate recommendations according to the designated model.
5. Web server interacts with the database using PHP scripts to save visitors' data, retrieve visitors' data, and retrieve recommendations from IIT.
6. Data is retrieved from database.
7. Web server sends JSON data to the XMLHttpRequest callback function.
8. HTML and CSS data is rendered on the browser to display personalized recommendations.

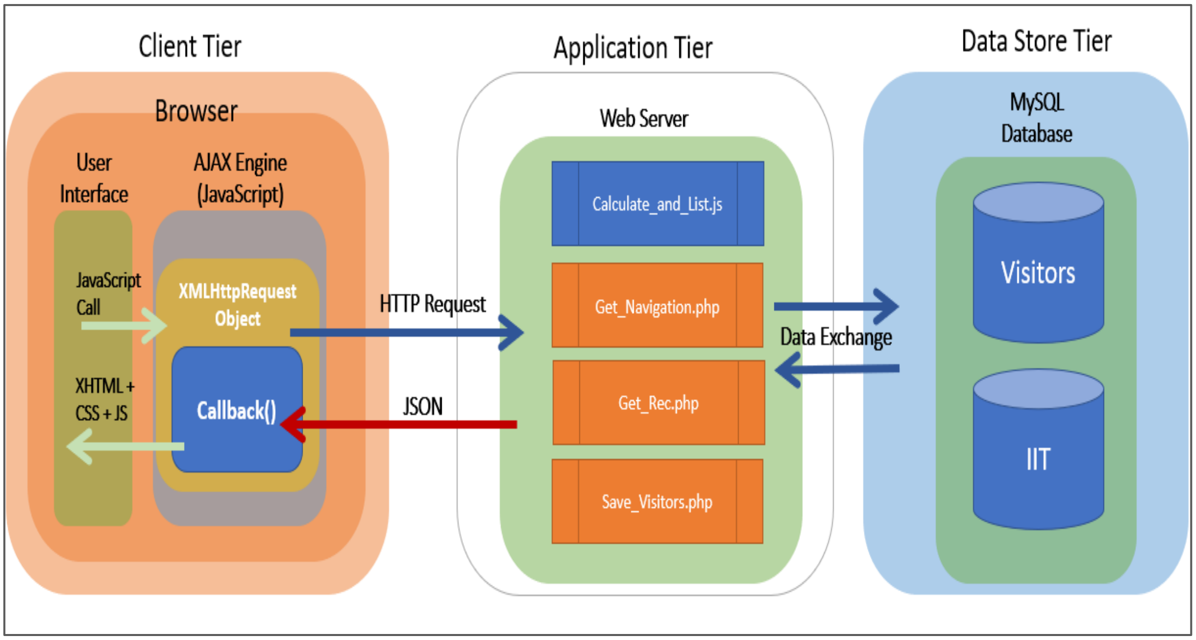


Figure 23: Architecture of PCRF – online module

Table 20: Get requests

Web Services Called with Get Request	Explanation
Get_Navigation	Retrieve user navigation from Visitors tables in MySQL DB up to current page
Get_Recommendations_with_HARD	Retrieve topics from IIT sorted/ranked using HARD ranking scheme
Get_Recommendations_with_CRD	Retrieve topics from IIT sorted/ranked using CRD ranking scheme
Get_Recommendations_with_Top	Retrieve topics from IIT sorted/ranked using Top-K ranking scheme

Table 21: Post requests

Web Services Called with Post Requests	Explanation
Save_visitors	On every click extract the current topic and save it in the Visitors table in MySQL DB

Chrome console's network analyzer is used to analyze the performance of PCRF. "Calculate_and_List.js" script is chosen to be the focus of our analysis, because, it is the script responsible for calling all php scripts to post data and retrieve data. It is also the script responsible for building and analyzing the navigation graphs and generating recommendations. Figure 24 and Figure 25 shows network analyzer's results for "Calculate_and_List.js" script on HARD-based website (www.hardtest.site) and CRD-based website (www.crdtest.site) respectively. Cache is disabled during testing. The test considers navigation graph size starting from one node and up to twenty nodes (1=< |TNG|=<20).

Figure 26 shows response time in seconds in the y-axis against navigation graph size in the x-axis for both websites. Results show that PCRF has very good to excellent performance. Response time for PCRF on both websites is, to a certain limit, consistent and does not increase according to the size of the navigation graph. Furthermore, in incidents were response time is slightly greater than the average, for instance, on CRD-based website, second test case generated the greatest response time of 3.35s with navigation graph containing two nodes only, the cause of the long response time is the TTFB delay which can be a result of temporary network issue (Figure 27).

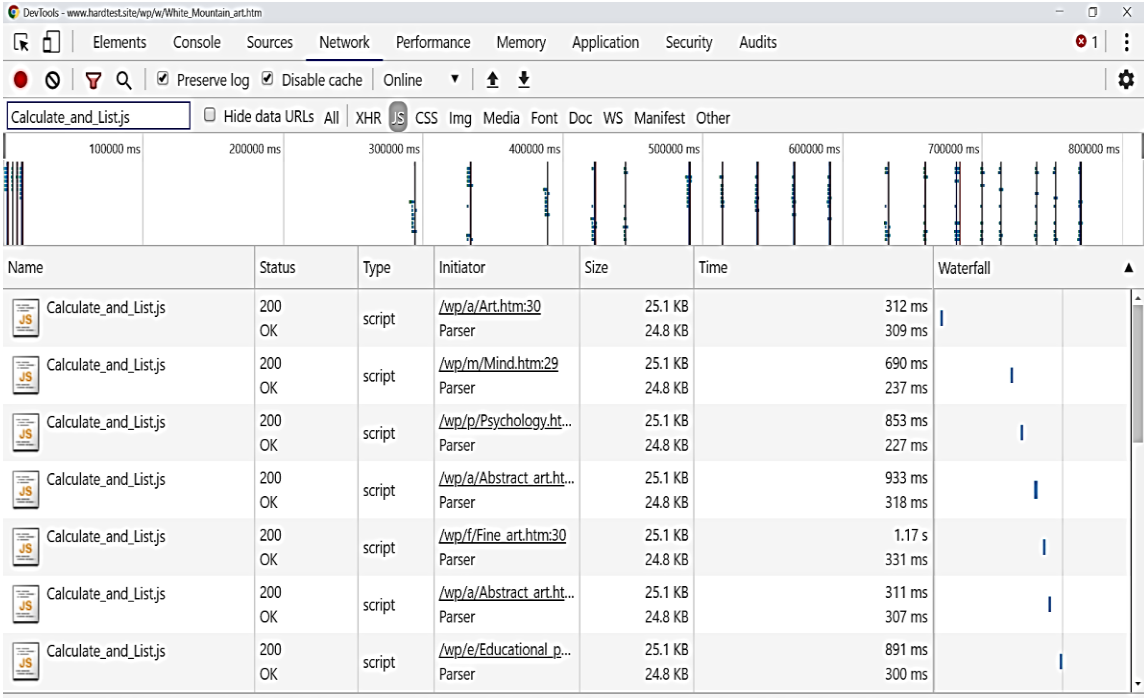


Figure 24: Chrome console's network analyzer's results of the response time analysis of Calculate_and_List.js on www.hardtest.site

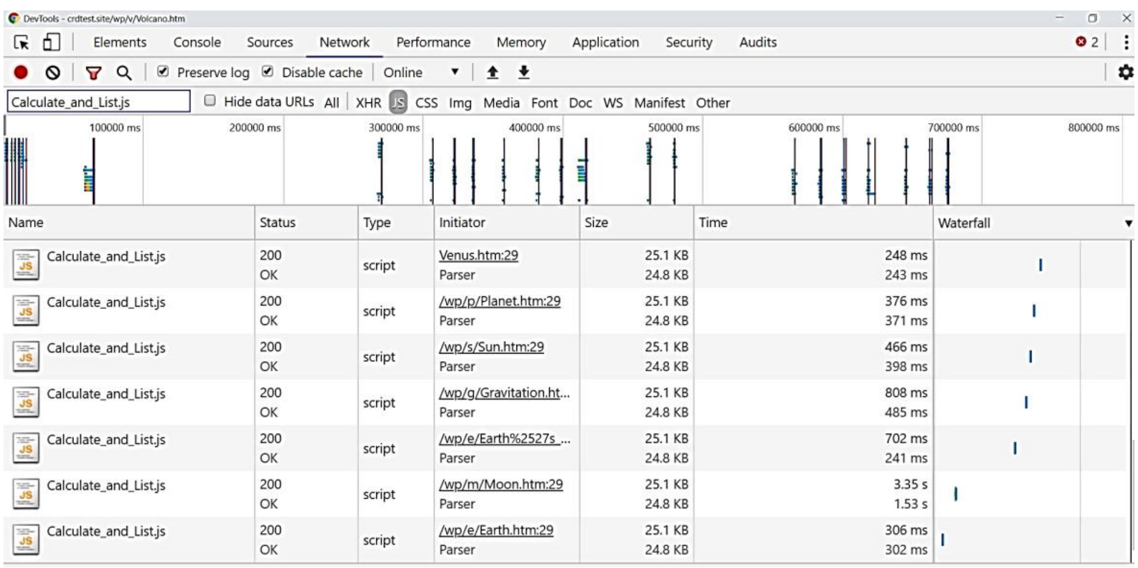


Figure 25: Chrome console's network analyzer's results of the response time analysis of Calculate_and_List.js on www.crdtest.site

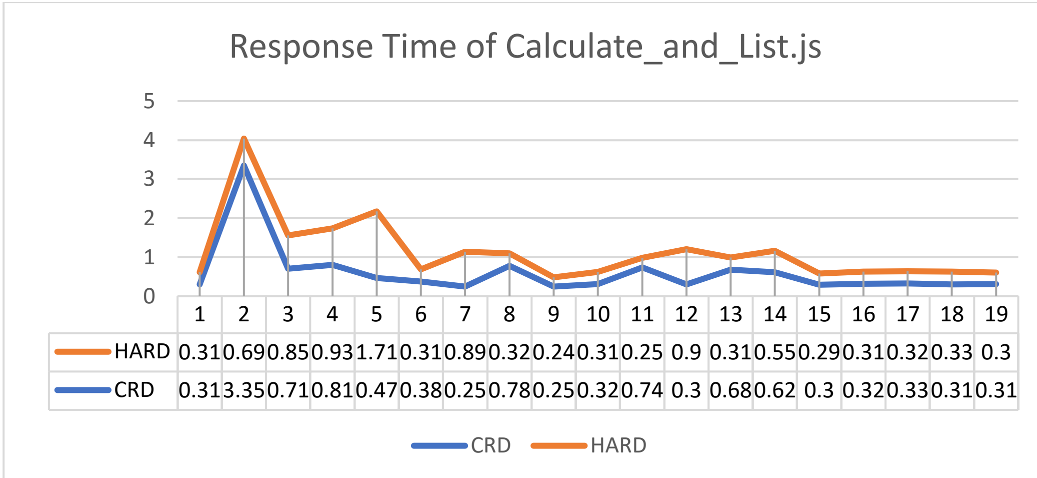


Figure 26: Response time of Calculate_and_List.js on CRD and HARD based websites

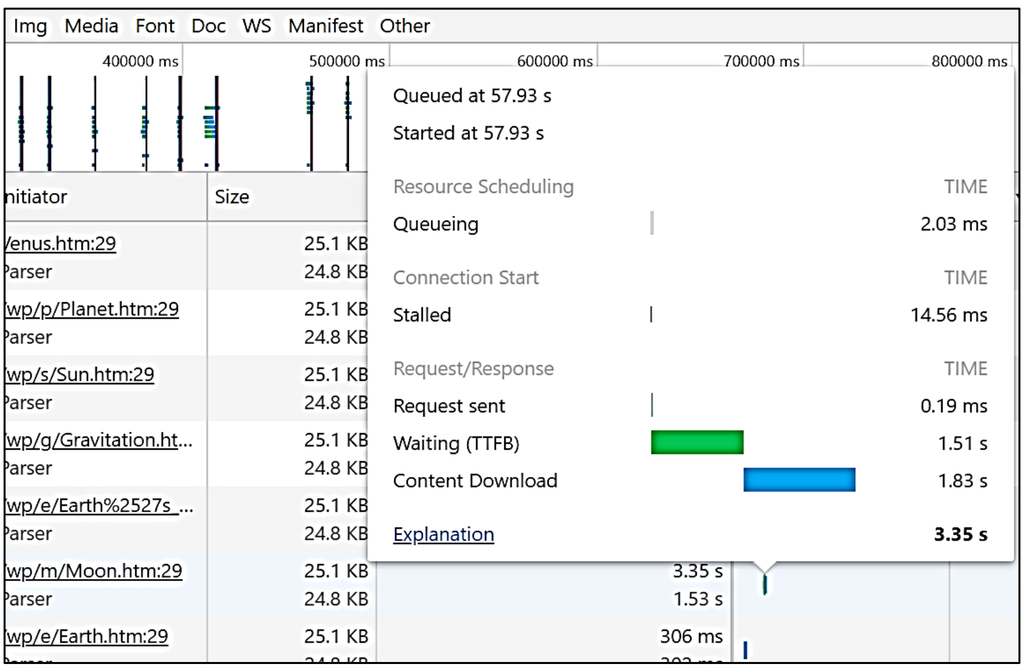


Figure 27: Delay caused by TTFB for the longest response time

5.4.2 Offline Module

We developed four Java programs using JDK 8 and JRE 8. These are:

1. *A parser program* to extract plain text from HTML document. Text surrounded by paragraph tags, <p>, and title tags, <title>, is extracted.
    a. *Input*: HTML-based corpus
    b. *Output*: plain text corpus and list of topics (for Wikipedia titles are extracted to represent topics)

2. *A text pre-processor program* for stopwords removal, tokenization, and stemming.
    a. *Input*: plain text corpus
    b. *Output*: processed text corpus

3. *A fuzzy thesaurus builder program*.
    a. *Input*: processed text corpus
    b. *Output*: Fuzzy thesaurus → the fuzzy thesaurus is stored in a Hash Table and export it into a serializable file.

4. *A topic-document index builder program*.
    a. *Input*: Fuzzy thesaurus, list of topics, and plain text corpus
    b. *Output*: inverted index of topics → it is exported into a comma separated file.

The first two programs run on a Core (TM) i7-6500U 2.6 GHz PC with 16 GB memory running Windows 10. Five GB for JVM memory is allocated. Table 22 lists the APIs used in the first two programs.

Table 22: APIs used for different natural language processing tasks in the offline module

Parsing Natural Language Processing Tasks	API/Technique
Stopwords Removal	Apache Lucene Core 5.3.0[8]
Stemming	Porter Stemming Algorithm[9]
Tokenization	Apache Lucene Core 5.3.0
Text parsing form HTML pages	jsoup Java HTML Parser[10]

To build fuzzy thesauri for experimental purposes an HTML-based corpus of School Wikipedia[11] is used. Even though, the source, i.e. Wikispeedia Game, provides plain text corpus, it was important to perform parsing and natural language processing tasks from scratch using the HTML-based corpus. The Wikispeedia's plain text corpus contains noisy data that was not properly removed during the parsing phase. For instance, if we examine the resulted plain text document for the Webpage "Action Potential" produced by our programs, Figure 28, and the text document produced by Wikispeedia, Figure 29, it can be seen that the latter contains headers such as "Overview", and caption or classification labels/tags such as "#copyright" and "2007 Schools Wikipedia Selection. Related subjects: General Biology". This extra data hinders the precision of correlation factors calculated for fuzzy thesauri, because distance, i.e. number of terms between every two distinct terms, is of vital interest in this task. Hence, these extra terms result in inaccurate distance measures that do not reflect the actual correlation between terms in the text.

[8] https://lucene.apache.org/core/
[9] https://tartarus.org/martin/PorterStemmer/java.txt
[10] https://jsoup.org/
[11] http://snap.stanford.edu/data/wikispeedia.html

An action potential is a wave of electrical discharge that travels along the membrane of a cell. Action potentials are an essential feature of animal life, rapidly carrying information within and between tissues. They are also exhibited by some plants. Action potentials can be created by many types of cells, but are used most extensively by the nervous system for communication between neurons and to transmit information from neurons to other body tissues such as muscles and glands.
Action potentials are not the same in all cell types and can even vary in their properties at different locations in the same cell. For example, cardiac action potentials are significantly different from the action potentials in most neurons. This article is particularly concerned with the "typical" action potential of axons.
A voltage, or difference in electrostatic potential, always exists between the inside and outside of a cell. This results from the distribution of ions across the cell membrane and from the permeability of the membrane to these ions. The voltage of an inactive cell stays at a negative value (inside relative to outside the cell) and varies little. When the membrane of an excitable cell is depolarized beyond a threshold, the cell will undergo (or "fire") an action potential, often called a "spike" (see Threshold and initiation).
An action potential is a rapid swing in the polarity of the voltage from negative to positive and back, the entire cycle lasting a few milliseconds. Each cycleâ€"and therefore each action potentialâ€"has a rising phase, a falling phase, and finally an undershoot (see Action potential phases). In specialized muscle cells of the heart, such as cardiac pacemaker cells, a plateau phase of intermediate voltage may precede the falling phase, extending the action potential duration into hundreds of milliseconds.
Action potentials are measured with the recording techniques of electrophysiology and more recently with neurochips containing EOSFETs. An oscilloscope recording the membrane potential from a single point on an axon shows each stage of the action potential as the wave passes. These phases trace an arc that resembles a distorted sine wave. Its amplitude depends on whether the action potential wave has reached that point on the membrane or has

Figure 28: Plain text document produced by our program for "Action Potential" Webpage

```
#copyright

Action potential

2007 Schools Wikipedia Selection. Related subjects: General
Biology

   A. A schematic view of an idealized action potential
illustrates its
   various phases as the action potential passes a point on a
cell
   membrane. B. Actual recordings of action potentials are often
distorted
   compared to the schematic view because of variations in
   electrophysiological techniques used to make the recording
   Enlarge
   A. A schematic view of an idealized action potential
illustrates its
   various phases as the action potential passes a point on a
cell
   membrane. B. Actual recordings of action potentials are often
distorted
   compared to the schematic view because of variations in
   electrophysiological techniques used to make the recording

   An action potential is a wave of electrical discharge that
travels
   along the membrane of a cell. Action potentials are an
essential
   feature of animal life, rapidly carrying information within
and between
   tissues. They are also exhibited by some plants. Action
potentials can
   be created by many types of cells, but are used most
extensively by the
   nervous system for communication between neurons and to
transmit
   information from neurons to other body tissues such as muscles
and
   glands.

   Action potentials are not the same in all cell types and can
even vary
   in their properties at different locations in the same cell.
For
   example, cardiac action potentials are significantly different
from the
   action potentials in most neurons. This article is
particularly
   concerned with the "typical" action potential of axons.

Overview

   A voltage, or difference in electrostatic potential, always
exists
   between the inside and outside of a cell. This results from
```

Figure 29: Plain text document produced by Wikispeedia for "Action Potential" Webpage

The resulted plain text corpus for the School Wikipedia contains 5,232 documents and 148,946 distinct stemmed terms. With this size of terms/features space, the resulted fuzzy thesaurus contains $148{,}946^2$ entries. This huge number of correlation factors values requires very large JVM memory size. Ideally, more than 5 GB. Therefore, the process of building the fuzzy thesaurus and the inverted indices of topics was not possible on the same machine that was used for parsing and natural language processing tasks.

To build the fuzzy thesaurus and inverted index of topics, high-performance computing (HPC) was used. The HPC nodes run the Linux CentOS operating system and are accessed remotely through a secure shell client. This is a small application that enables connection to a remote computer via SSH (Secure SHell), a cryptographic network protocol. Because Windows is used, HPC was accessed using PuTTY, which is a popular third-party client that may be downloaded through the developer's website. A home directory and two compute nodes "SemanticRecNode2", and "InvertedTopicIndexNode2" are created to compile and run the third and fourth programs listed earlier in this section. 230 GB memory is assigned for the nodes. Compilation and build process of fuzzy thesaurus and inverted index of topics completed in two minutes.

Chapter 6: Evaluating PCRF

The proposed framework 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 commonly referred to as offline evaluation [185]. However, evaluations of systems involving user models cannot and should not be separated from actual users [186]. 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 [81]. A major advantage of user studies is that they allow for collecting information about user interaction as well as testing different scenarios. Therefore, user studies are designed to evaluate the effectiveness of the proposed approach.

6.1 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. User studies are designed following two main strategies. The first strategy aims at evaluating the quality of personalized recommendations, so in the first treatment, Figure 30, three user groups are considered. Two user groups using the websites with personalized recommendations using CRD and HARD models, and one user group using the website with recommendations generated based on popularity model as a baseline.

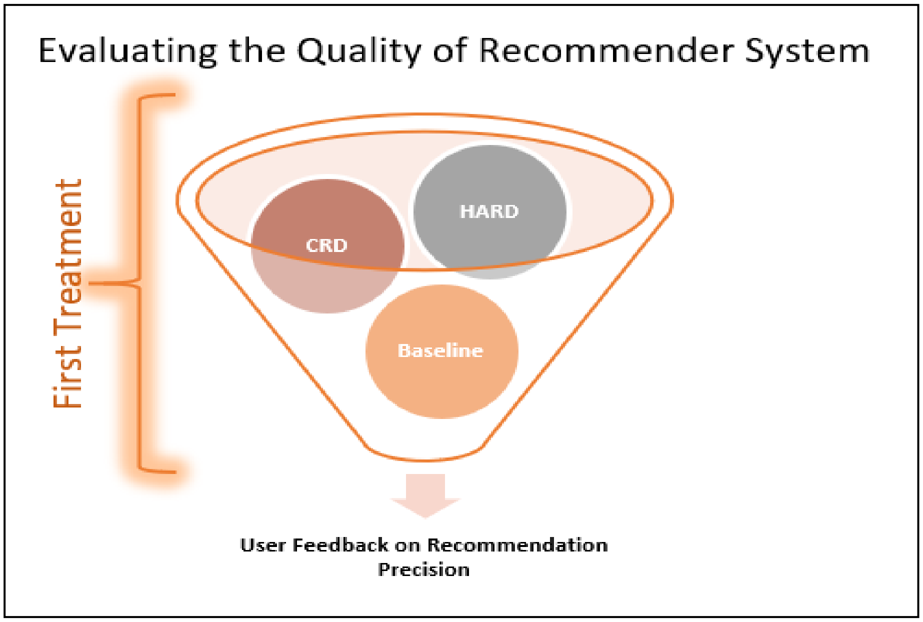


Figure 30: User study - First treatment

The second strategy aims at evaluating the impact of personalized recommendations on informal learning, so in the second treatment, Figure 31, four user groups are considered. 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.

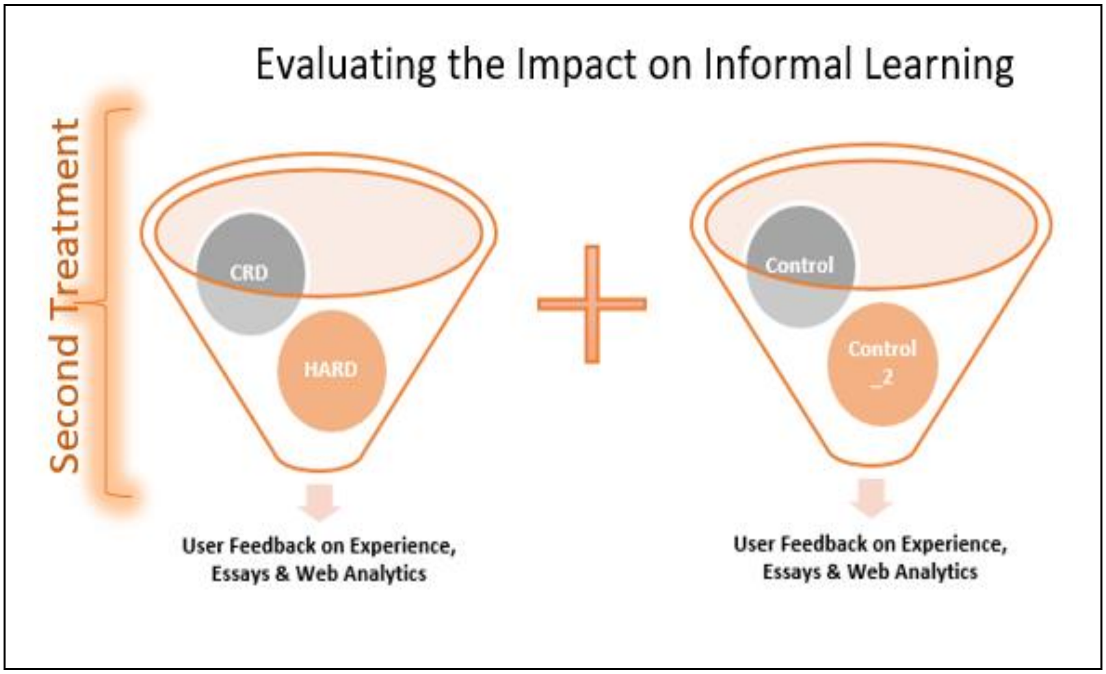


Figure 31: User study - Second treatment

6.2 Technological Framework

To run our user studies, four web-based encyclopedias are developed. The four websites are equipped with user navigation's tracking and analysis algorithms, the proposed personalized content recommendation engine, and popularity-based recommendation engine. The online test encyclopedias are listed in Table 23. 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 32 shows screenshots from the website.

Table 23: 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 model (baseline)

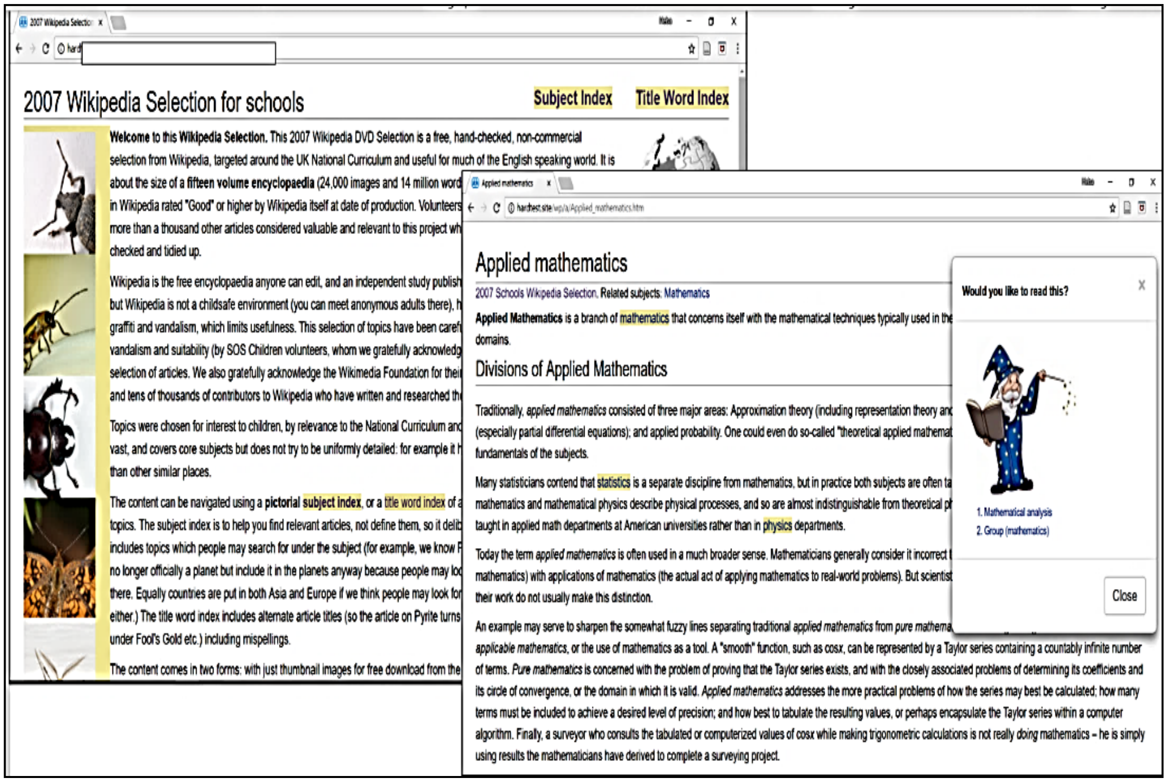


Figure 32: Screenshot from the test environment

6.3 Evaluation Metrics

6.3.1 Metrics to Evaluate the Quality of Recommender System

To evaluate the effectiveness of the proposed PCRF, the rank-based Mean Average Precision, $MAP@k$, is used 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 the experiments, $MAP@k$ is used 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:

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

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, $MAP@k(U)$.

6.3.2 Metrics to Evaluate the Impact of Recommendation on Informal Learning

In the evaluation of informal learning, three types of metrics are used: user-centric qualitative metrics to evaluate the user-perceived effectiveness of the personalized recommendations, objective educational metrics to evaluate the impact 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*, two metrics are evaluated. These have been commonly used in the literature [187]:

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*, conceptual knowledge assessment is considered given that we are evaluating informal learning. In informal learning, no specific curriculum is followed, neither predefined learning outcomes upon which learners can be evaluated. 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 [188]. Therefore, to evaluate informal learning, a conceptual knowledge assessment rubric is designed. This rubric is adapted from concept map-based rubrics[12]. The rubric used is a simplified rubric aimed at assessing conceptual knowledge in essays for primary students. Essays are assessed against five criteria: structure, relationships, exploratory, communication, and writing quality. Essays are 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. Our proposed rubric is illustrated in Figure 33 .Finally, *web analytics data* is used to analyze the general navigational patterns of each user group. Topics' frequencies of visited web pages are analyzed to find out whether a certain test group is focused, distracted, or not focused on the main topic of experiment.

6.4 Learning Content

We use content from the 2007 Wikipedia DVD Selection[13], 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

[12] https://teach.its.uiowa.edu/sites/teach.its.uiowa.edu/files/docs/docs/Concept_Map_Rubrics_ed.pdf
[13] https://en.wikipedia.org/wiki/Wikipedia:Wikipedia_CD_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. Table 24 shows the subject categories under Wikipedia school selection.

Table 24: The subject categories under the Wikipedia Selection for Schools

Category	Articles	Category	Articles
Art	74	Business Studies	88
Citizenship	224	Countries	220
Design and Technology	250	Everyday life	380
Geography	650	History	400
IT	64	Language and literature	196
Mathematics	45	Music	140
People	680	Religion	146
Science	1068		

Criteria	Excellent (4 Marks)	Good (3 Marks)	Adequate (2 Marks)	unacceptable to review (1 Marks)	Earned marks
Structure	Non-linear structure that provides a **very complete picture** of your concepts	Non-linear structure that provides a **complete picture** of your concepts	Non-linear structure that provides a **rough picture** of your concepts	**Inappropriate** structure	/4
Relationships	**There are 3 concepts or more** about the same topic • the concepts are very effectively connected, i.e. cause-and-effect comparisons	**There are 1 – 2 concepts** about the same topic • the concepts are connected with clear and accurate relationships, i.e. comparisons of factual knowledge	**There are 1 – 2 concepts** • relative importance of concepts is indicated, i.e. accurate use of factual knowledge	**Only 1 concept** • no differentiation between concepts • no evidence of meaningful relationships	/4
Exploratory	Essay shows **complex** thinking about the concepts and the meaningful relationship between concepts related to main topic, subtopics as well as relevant topics • essay is focused on main topic, subtopics as well as relevant topics	Essay shows **effective** thinking about the concepts and the meaningful relationships between concepts related to main topic and subtopics • essay is focused on the main topic as well as on subtopics	Essay shows **definite** thinking about concepts related to main topic • essay is focused on the main topic	Thinking process is not clear • essay is not focused on the main topic	/4
Communication	Information is presented clearly and reflects a **high level** of understanding of the topic	Information is presented clearly and reflects **good level** of understanding of the topic	Information is presented clearly and reflects **basic level** of understanding of the topic	Information is not clear and reflects **no understanding** of the topic	/4
Writing Quality	**No spelling mistakes** and the essay is **easy to read**	**Few spelling mistakes** and the essay is **easy to read**	**Few spelling mistakes,** but the essay is **readable with some effort**	**Many spelling** mistakes. The essay is **not readable**	/4
Total score					/20
Student name and group					

Figure 33: Conceptual knowledge rubric

6.5 Data Collection Techniques

Multiple data collection tools are used. For instance, questionnaires are used 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 (See Appendix C for questionnaires). In addition, participants are asked to submit essays related to the topic of space. Moreover, tracking scripts are run to collect navigation-related data.

6.6 Participants

Experiments were carried out 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. These are: Control, HARD, CRD, Baseline, and Control_2. Balanced participation from both male and female students are received. 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 34. Ethical approvals and consent forms are in Appendix A and Appendix B respectively.

Test groups underwent two different treatments following the two strategies explained earlier in Figure 30 and Figure 31. Further details related to test procedure and methods are explained in following sections.

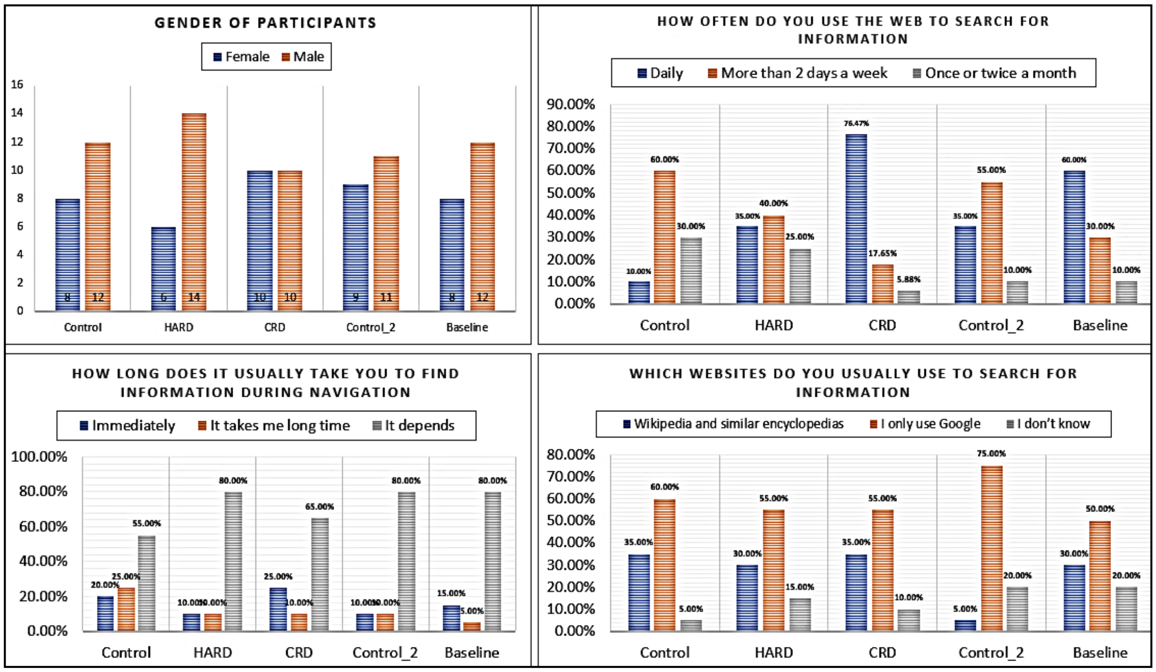


Figure 34: Demographics of test groups

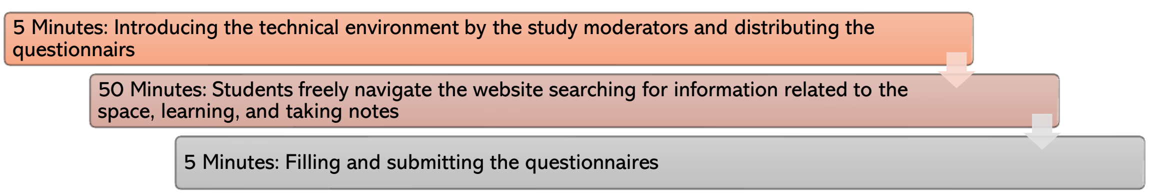


Figure 35: Test session procedure

6.7 Procedure

A writing challenge was announced among year-five students. In the announcement, the students were invited 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 challenge flyer is available in Appendix D. The experiments were carried out during term three of the school year by then the participants had covered enough material related to space as part of their science subject. This information was confirmed 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, factors of previous experiences and minimum required skill levels are controlled. These commonly impact any learning process. Furthermore, a fixed design for all the test sessions in terms of time, location, class setup, and duration is forced 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, all the test experiments are carried 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 in Table 23. All test session followed the same structure as explained in Figure 35

6.8 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 against the baseline (i.e. the popularity model). The selected groups are: CRD, Baseline, and HARD. Each user group had twenty students. Students were asked 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. The recommendations' pop up window was designed 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 need to evaluate the relevance of recommendations is towards the end of their test session where the navigation graph size would have increased above five, 5 << |TNG|. Here also the students are instructed to give their feedback at five different times, as they would have done at the beginning of the test session. Students' feedback on recommendations' relevance was recorded to evaluate the *precision* as well as *adaptivity*. Students' feedback on recommendations along with complete precision calculations are presented in Appendix G. 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 since 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. One-way ANOVA for multiple means is used to measure the statistical significance of the results at alpha level

5% (α = 0.05). Results are found to be statistically significant with *P-Value* = 0.0, (p-value< 0.05) at the beginning as well as at the end of the test session. Hypothesis statistical analysis of mean average precision (MAP@K) is presented in Appendix H. Also, the Tukey method is used for pairwise comparison to further test the statistical significance between every two models' performance with alpha level 5% (α = 0.05). At the beginning of the test session with small size of navigation graphs (|TNG|=<5), the difference between CRD and HARD turn to be insignificant with *P-Value* = 0.895, (p-value > 0.05). However, At the end of the test session with large size of navigation graphs (5<<|TNG|), the difference between CRD and HARD turn to be statistically significant with *P-Value* = 0.0, (p-value < 0.05).

6.8.1 Discussion of the Results of First Treatment

Results of the 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 it can be seen in Figure 36, CRD based recommendations starts as the most precise among all 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 decrease until it reaches 80.35% compared to HARD and the Baseline which both exhibit better adaptability to changes in user interests. Figure 36 shows that up to rank five, with number of topics equals five, HARD model consistently maintains reasonable precision with *MAP@k* score ranges between 85% and 91.25%. The baseline, which does not implement any personalization logic starts so imprecise as 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 the highest number of visits by the users in the current test session.

Examining the performance of the three systems towards the end of the test session as shown in Figure 37, 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. In contrast, 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 performance continue to improve towards the end of the test session but with much less precision compared to HARD or CRD. Figure 38, Figure 39, and Figure 40 show values of MAP@K for HARD, CRD, and Baseline systems respectively at the beginning and at the end of the test session.

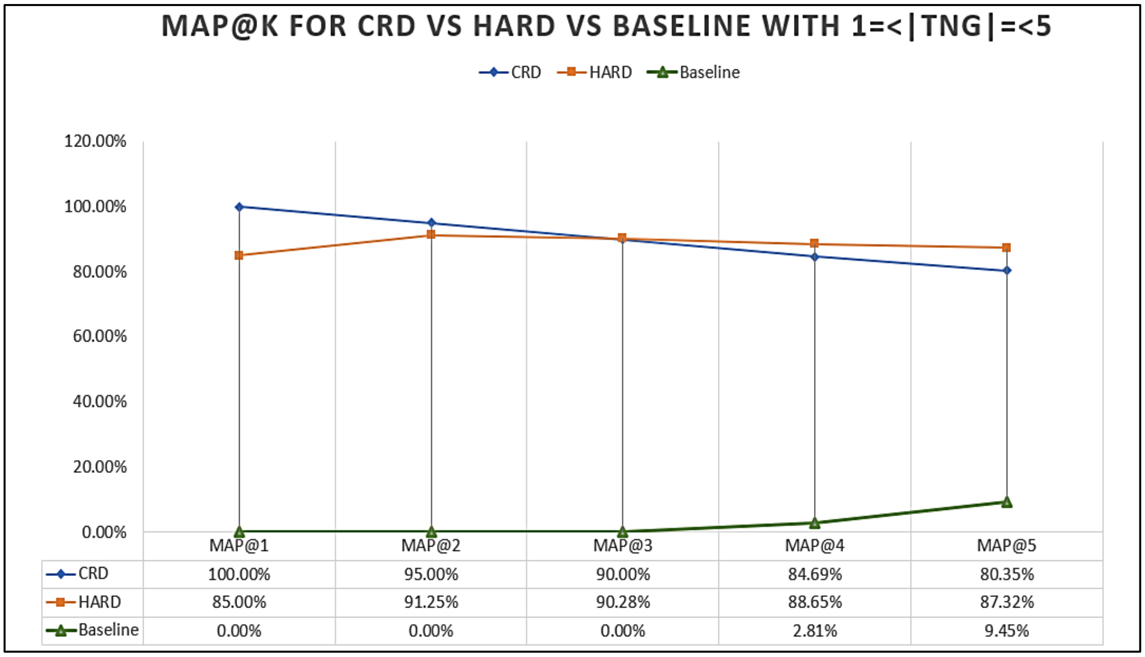


Figure 36: Cross systems MAP@K at the beginning of the test session

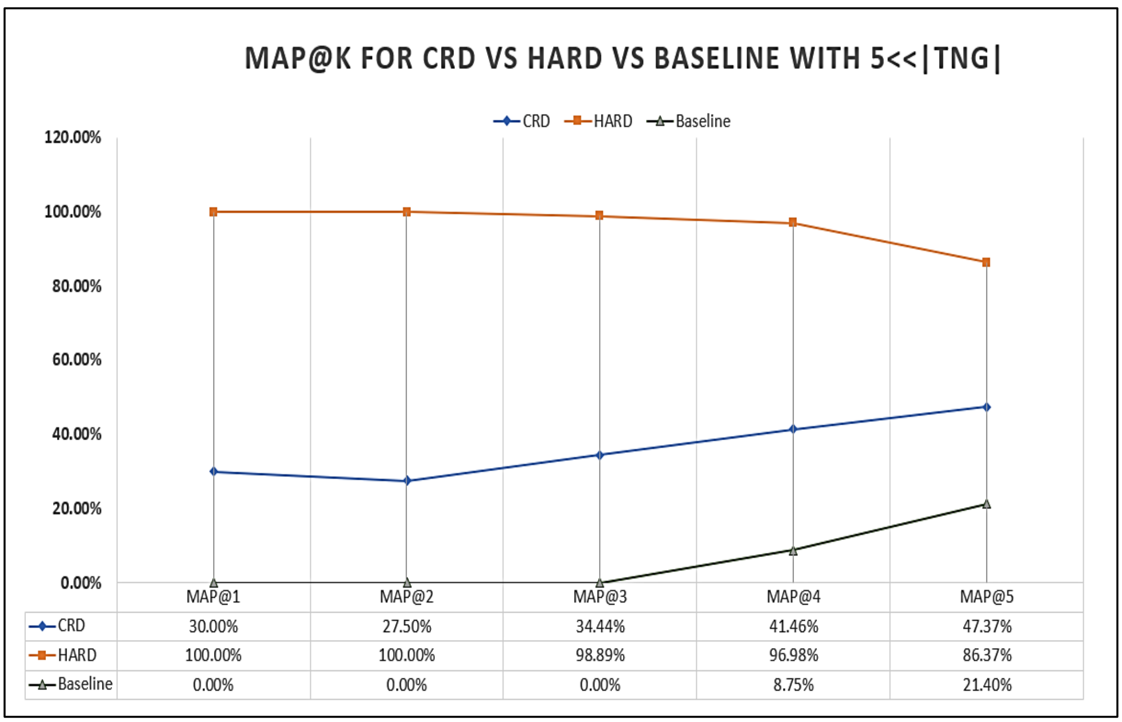
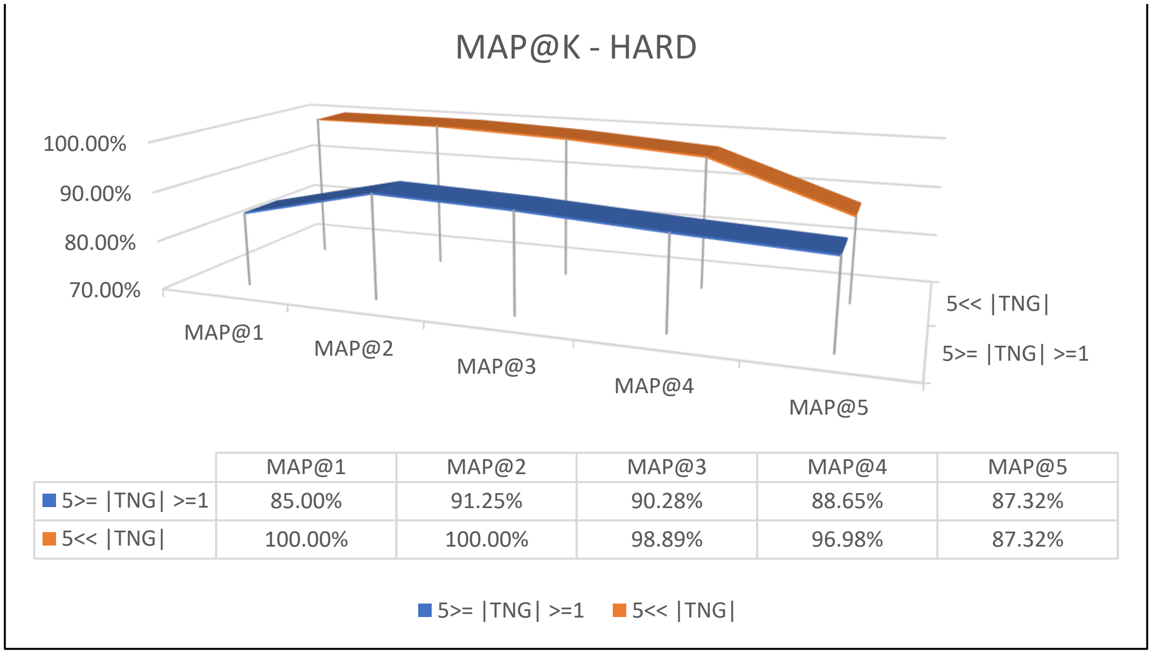
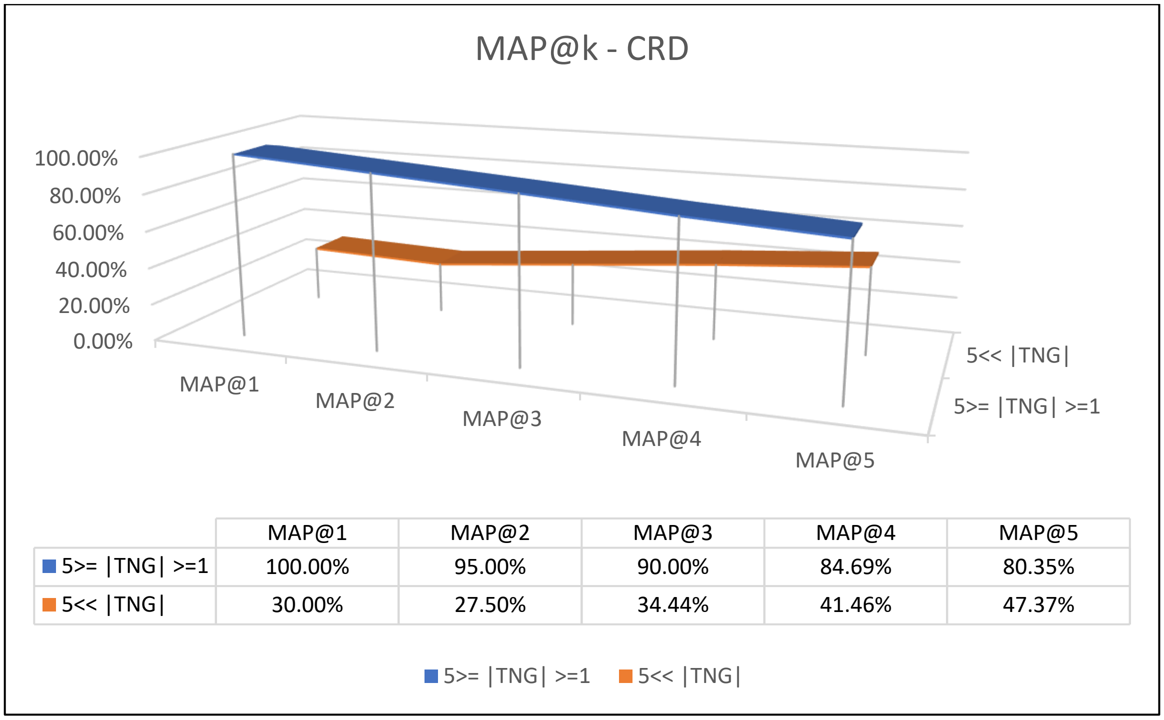


Figure 37: Cross systems MAP@K at the end of the test session



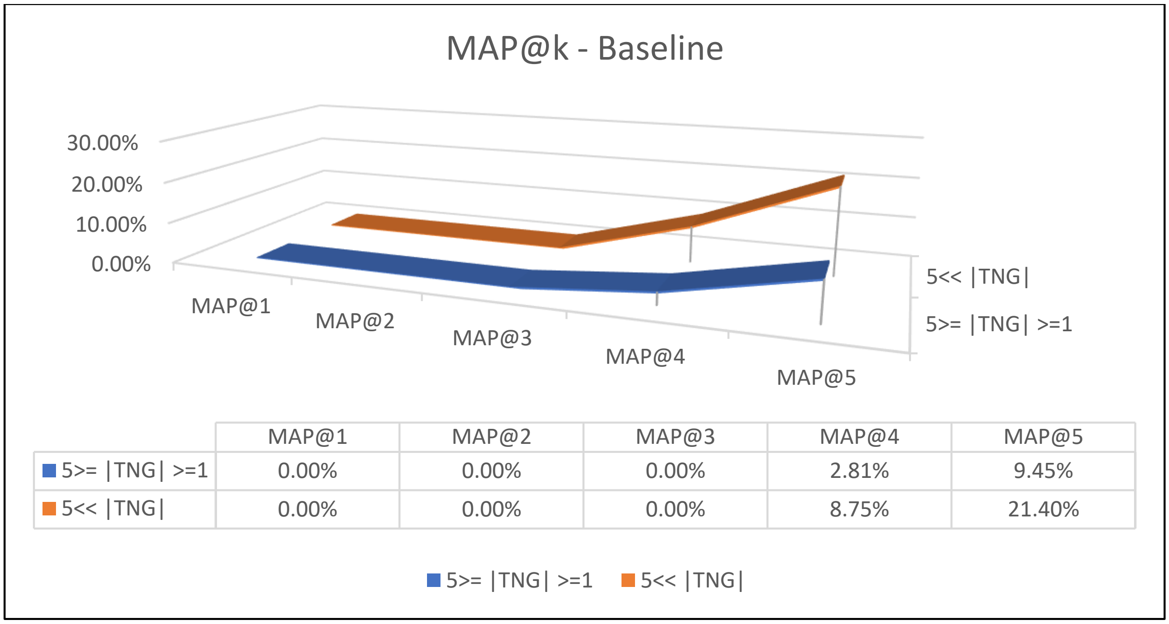
	MAP@1	MAP@2	MAP@3	MAP@4	MAP@5
5>= |TNG| >=1	85.00%	91.25%	90.28%	88.65%	87.32%
5<< |TNG|	100.00%	100.00%	98.89%	96.98%	87.32%

Figure 38: MAP@K for HARD model at the beginning and at the end of the test session



	MAP@1	MAP@2	MAP@3	MAP@4	MAP@5
5>= |TNG| >=1	100.00%	95.00%	90.00%	84.69%	80.35%
5<< |TNG|	30.00%	27.50%	34.44%	41.46%	47.37%

Figure 39: MAP@K for CRD model at the beginning and at the end of the test session



	MAP@1	MAP@2	MAP@3	MAP@4	MAP@5
5>= |TNG| >=1	0.00%	0.00%	0.00%	2.81%	9.45%
5<< |TNG|	0.00%	0.00%	0.00%	8.75%	21.40%

Figure 40: MAP@K for Baseline at the beginning and at the end of the test session

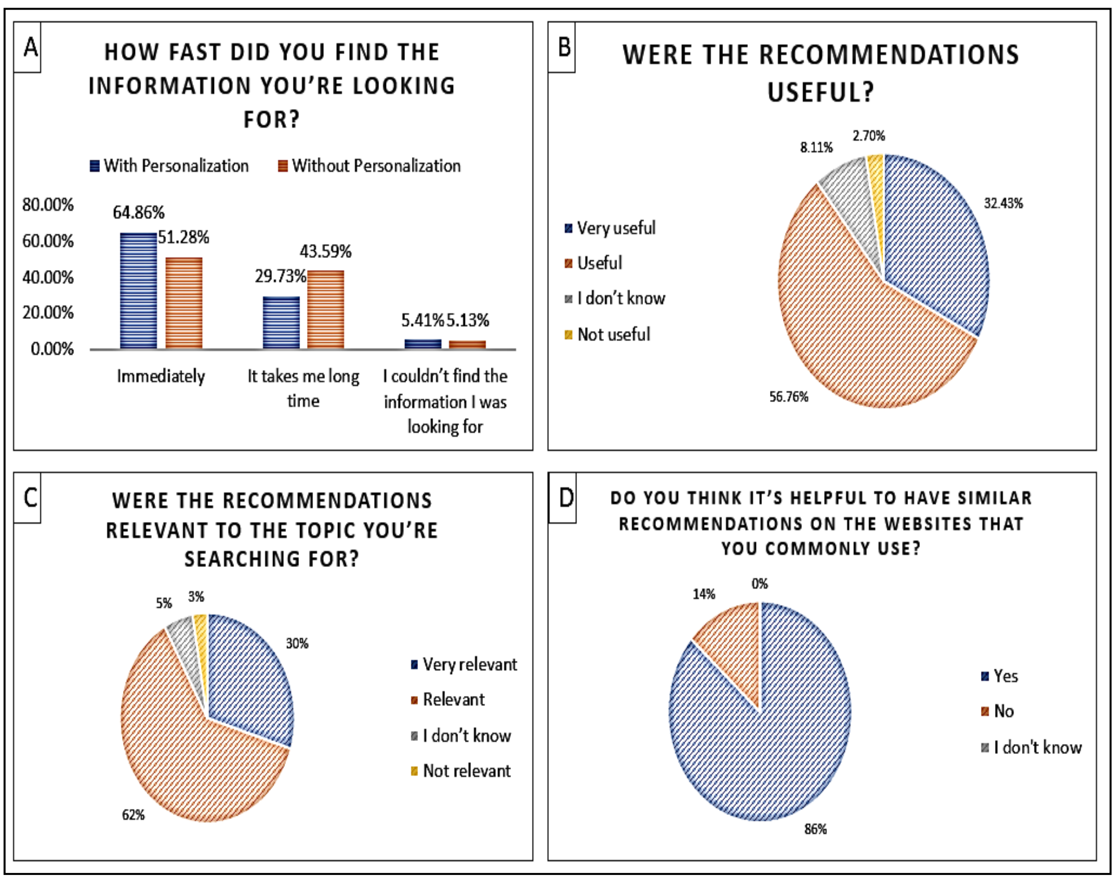


Figure 41: Results of user experience questionnaires

Looking at exemplary individual users' feedback in Table 25 ('1' indicates relevant and '0' indicates irrelevant) for CRD users, it can be seen that for user, U15, who seem 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, looking at two exemplary users on the HARD website in Table 26, user, U4, who seem to be very focused from the beginning of his/her navigation and user, U9, who seem to be changing interest over navigation session. It can be seen that HARD model immediately accommodates the changes and generates precise recommendations to user, U9, with reasonable precision at the beginning of the navigation session, then becomes very precise towards the end of the navigation session when 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 can be seen from the user feedback.

Table 25: 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 26: 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%				

6.9 Second Treatment of the User Study – Assessing the Impact of Recommendations on Informal Learning

Four user groups are selected to evaluate the impact of personalized recommendations on informal learning. These are: CRD, HARD, Control, and Control_2. Responses of CRD and HARD groups are grouped into "*with personalization*" group, and the responses of Control and Control_2 groups are grouped 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. Expressive responses are used rather than points as it is found 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, only 22 essays were selected (11 from the personalized support group and 11 from the control group) for the assessment of informal learning and excluded 10 submissions that are entirely copied from the online encyclopedia. Prizes were awarded for the best three essays. Sample from the control group is shown in Appendix E. Sample from the personalized-Support group is presented in Appendix F.

6.9.1 Discussion of the 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, students were asked 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. The questionnaire revealed 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 41 (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 41 (C). In addition, considering the overall user satisfaction criteria, results show that more than 90% of the 40 users who used the encyclopedia with personalized recommendations found the recommendations to be useful, and more than 80% thought that it would be helpful to have similar recommendations on other websites that they commonly used for information search as shown in Figure 41 (B) and (D) respectively.

<u>2. Evaluating Informal Learning</u>

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 to 10.0 for the students who used the encyclopedia without recommendations as shown in Table 27. 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). Hypothesis statistical analysis of essays' scores is presented in Appendix I. Moreover, the assessors found that participants who used the encyclopedia with personalized recommendations were able to make use of a larger number of concepts, make comparisons, and state relations between concepts.

Table 27: 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

3. Web Analytics-based Evaluation

Web analytics is the measurement, collection, analysis and reporting of web data for purposes of understanding and optimizing web usage [189]. With the inapplicability of formal assessment of learning in informal learning settings it is difficult to collect commonly used learning analytics for evaluation purposes. Therefore, we decide to examine the possibility of using web analytics data, which can be generated from any typical web navigation session, to induce some helpful insights about learners' performance. An initial design of an evaluation framework based on web analytics data is proposed as illustrated in Figure 42, that can be used to evaluate informal learning in similar environments.

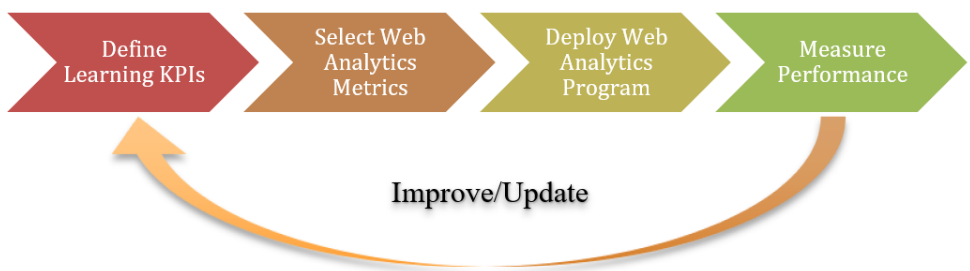


Figure 42: Web Analytics-based Evaluation Framework

In the following sections, different activities involved in the web analytics-based evaluation are explained.

A) Defining Key Performance Indicators (KPIs)

KPIs are "the critical (key) indicators of progress toward an intended result. KPIs provide a focus for strategic and operational improvement, create an analytical basis for decision making and help focus attention on what matters most [190]".

Considering the context of informal learning on information-oriented websites such as Wikipedia, users typically visit the website to learn about diverse topics of interest for various purposes. Additionally, users may have a new learning objective for every new visit to the website. Thus, our objective here is to maximize the value of each visit by providing faster and easier access to relevant content. Therefore, the required KPIs in this context should help us measure and quantify whether users of the website succeed to gain adequate access to relevant content in every visit.

Accordingly, the following three KPIs are considered for each user every time he/she visits the website:

1. *The frequency of relevant topics visited by the user:* this KPI is quantified at the document level, i.e. the main topic of each document/webpage is considered, which can be indicated by the page title in the context of information wikis.
2. *The frequency of relevant keywords in the visited pages:* the main keywords are extracted from the collection of visited pages for each user. Term Frequency Inverse Document Frequency, TF-IDF, is used to measure the importance of individual keywords in the collection. At a high level, a TF-IDF weight finds the words that have the highest ratio of occurring in the current document vs the frequency of occurring in the larger set of documents. As a result, terms that have very high frequency in all the documents in a certain collection will end up having very low TF-IDF, hence, they do not represent important keywords. Whereas, terms that receive high frequency at the document level compared to low frequency at the collection level will have very high TF-IDF scores and as such are considered important keywords. Afterwards, keywords

undergo semantic relevance test to select relevant keywords which can be used to quantify the frequency of relevant keywords.

3. *The frequency of relevant phrases in the visited pages* similar TF-IDF approach explained in KPI number two is applied at the phrase level. The phrase is considered to be composed of two terms.

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

B) Selecting Web Analytics Metrics

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

1. *Hits:* represent the total number of requests made to the server during a given time period, e.g. month, day, hour.
2. *Files*: represent the total number of hits (requests) that actually resulted in something being sent back to the user. That is, not all hits will send data, such as 404-Not Found requests and requests for pages that are already in the browsers cache. So, by looking at the difference between hits and files, a rough indication of repeat visitors can be obtained, as the greater the difference between the two, the more people are requesting pages they already have cached, i.e. have viewed already.
3. *Pages (Views):* are those URLs that would be considered the actual page being requested, and not all the individual items that make it up such as graphics and audio clips. This metric is sometimes called impressions, and defaults to any URL that has an extension of ".htm", ".html" or ".cgi".

4. Visits: occur when some remote site makes a request for a page on a server for the first time. If the same site keeps making requests within a given timeout period, they will all be considered part of the same Visit. If the site makes a request to a server, and the length of time since the last request is greater than the specified timeout period, common default is 30 minutes, a new Visit is started and counted, and the sequence repeats. Since only pages will trigger a visit, remote sites that link to graphic and other non- page URLs will not be counted in the visit totals, reducing the number of false visits.

5. Sites: is the number of unique IP addresses/hostnames that make requests to a server.

6. *Kbytes (KB):* is 1024 bytes (1 Kilobyte). It is used to show the amount of data that is transferred between the server and the remote machine, based on the data found in the server log.

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

C) Choosing and Deploying Web Analytics Program

We evaluated three web analytics programs, namely, Webalizer[14], AWStats[15], and Google Analytics[16]. Google Analytics is a client-side analytics tool for which data is collected by a JavaScript code added to the website's HTML pages. Whereas, the first two are server-side. That is, they use the data contained in the server logs. Google Analytics is excluded since already a number of Java Scripts are run on the test environments for

[14] http://www.webalizer.org/
[15] https://awstats.sourceforge.io/
[16] https://analytics.google.com/analytics/web/#/

tracking navigation graphs and for personalized recommendations. Hence, AWStats is selected as it gives full list of visited URLs that can be easily used for scrapping and processing required to quantify the KPIs mentioned earlier.

Using the page metric, for each user group, viewed pages during the test session are identified by applying time and date filters to AWStats setups. Then, a web scrapper application is run to extract viewed pages found in the AWStats' web analytics log files of both groups. During scrapping repeated extraction of pages is allowed. Repeated page views are counted as they give an indication of the amount of attention a user gives to a specific topic. Table 28 illustrates an example of AWStats page view analytics which is used in the evaluation.

Table 28: Snapshot from Page view analytics using AWStats

136 different pages-url	Viewed	Average size	Entry	Exit
/wp/a/Acetic_acid.htm	115	73.73 KB	110	42
/wp/p/Prehistoric_man.htm	55	41.94 KB	48	22
/wp/s/Sodium_sulfate.htm	47	42.40 KB	43	25
/wp/c/Calcium_chloride.htm	40	38.83 KB	37	16

D) Performance Evaluation based on Web Analytics Data

Analysis of web analytics data revealed that users, who used the encyclopedia with personalized support, navigated more articles related to their topics of interest compared to participants who used the encyclopedia without any personalized support. Users in the control group navigated a total of 226 articles compared to 644 articles navigated by the users in the personalized support group. These numbers include repeated views to the same articles. Manual analysis of the visited articles by both groups revealed that users in the control group were generally focused but visited less diverse topics related to "space" and some of them visited a few irrelevant topics such as “art” and “children charity”. However, the other group of users visited more diverse pages related to “space”. This might have resulted in helping the students who used the online encyclopedia with personalized support to use a larger number of related concepts and state relations among concepts. It can be seen as well in Table 27 that the students in the personalized support group submitted essays of more various topics compared to the control group students who submitted limited number of topics, mainly focused on “Black Hole” and “Neptune”.

Moreover, by performing keyword extraction and phrase extraction on the collection of visited pages of both groups a further validation on the observations highlighted by the manual analysis can be obtained. Table 29 shows statistics on viewed pages, frequency of extracted keywords, and frequency of extracted phrases.

Table 29: Statistics of visited Pages extracted from users' web analytics logs.

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

By considering the twenty highest frequency keywords and phrases of both groups, it can be seen that, for both groups, the top 50 keywords are mostly relevant to the topic of space. This gives a good indication that users were focused on the topic of space. However, the frequency of top keywords viewed by the personalized-support group significantly surpasses the frequency in control group as illustrated in Figure 43 and Figure 44 For example, “Earth” keyword’s frequency is 9,441 in the personalized support group compared to 3,600 in the control group. This in turn, indicates that for the personalized support group more relevant articles related to “earth”, which is an important topic in the space, were visited by the personalized support group. These results reinforce the manual analysis carried earlier.

Furthermore, by analyzing the top 50 phrases extracted from the navigated pages' collection, it can be seen that almost all the top phrases are related to the topic of the 'space' which gives a further validation to the previous observations as illustrated in Figure 45 and Figure 46. In addition, the frequencies of top phrases in the personalized support group surpasses by far the frequencies in the control group. For example, the frequency of "Solar System" is 1,314 in the control group compared to 4,176 in the personalized support group. These statistics validate further our earlier observations.

Finally, it can be concluded 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.

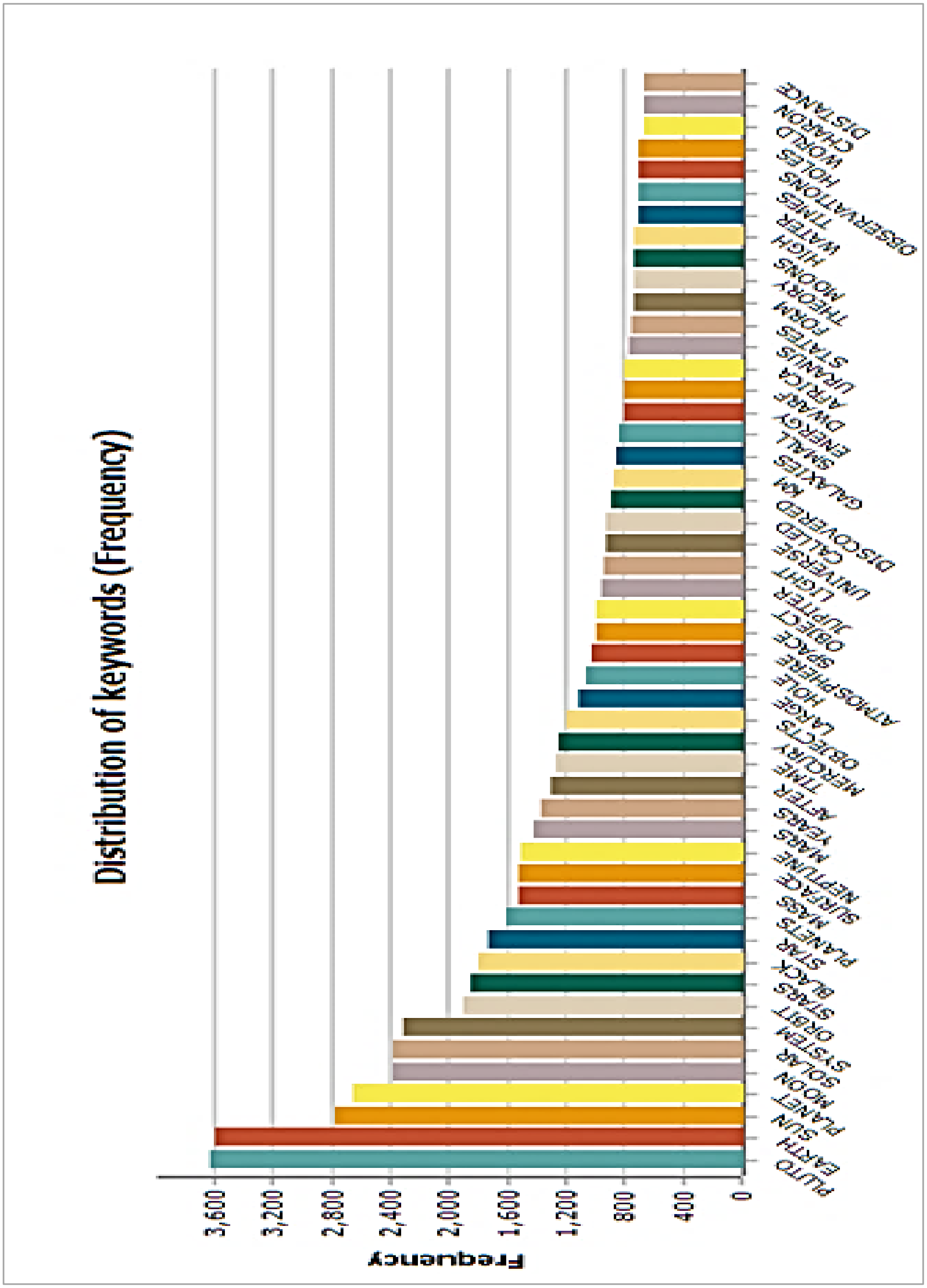


Figure 43: Distribution of keywords for control group

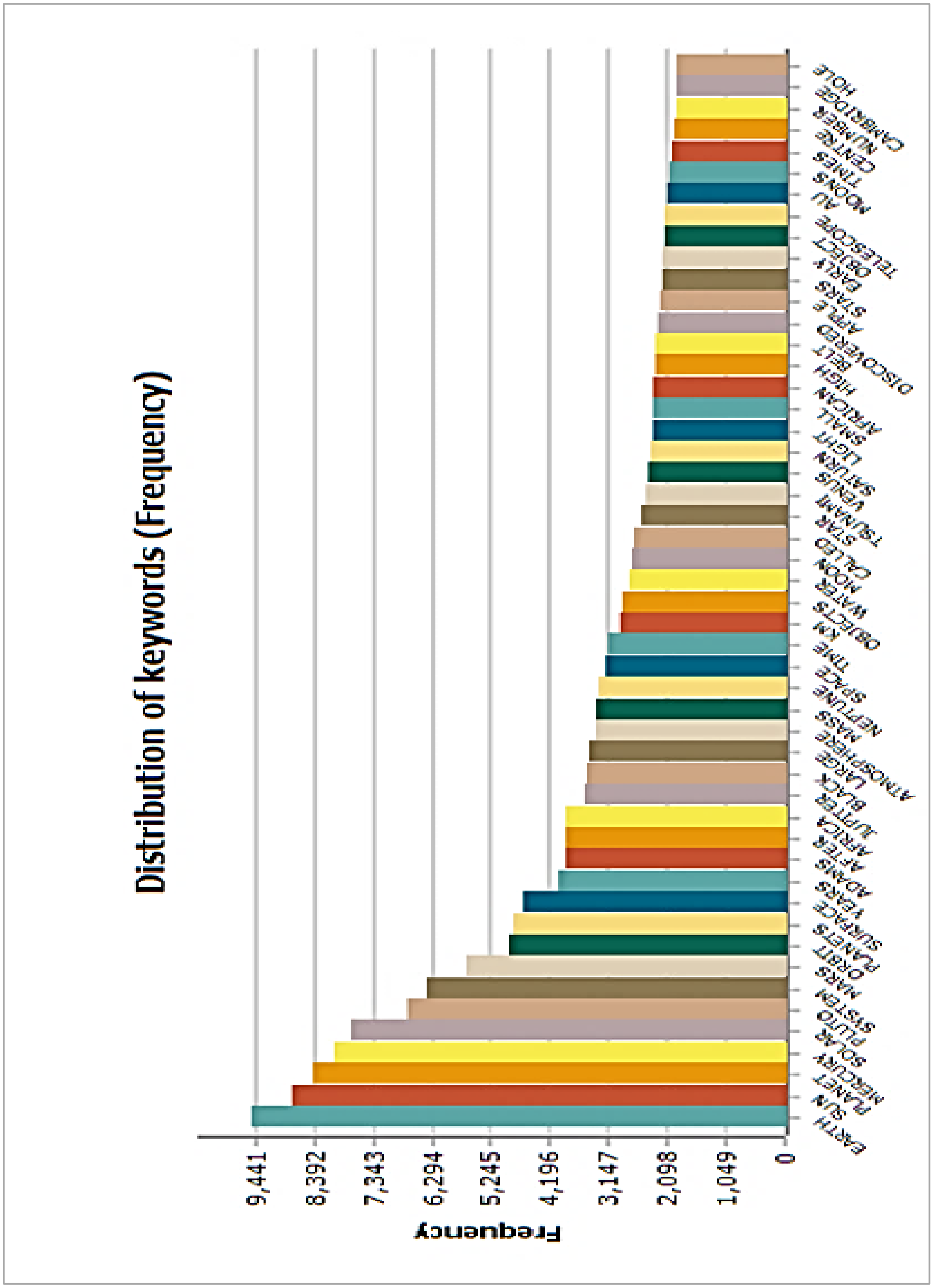


Figure 44: Distribution of keywords for personalized support group

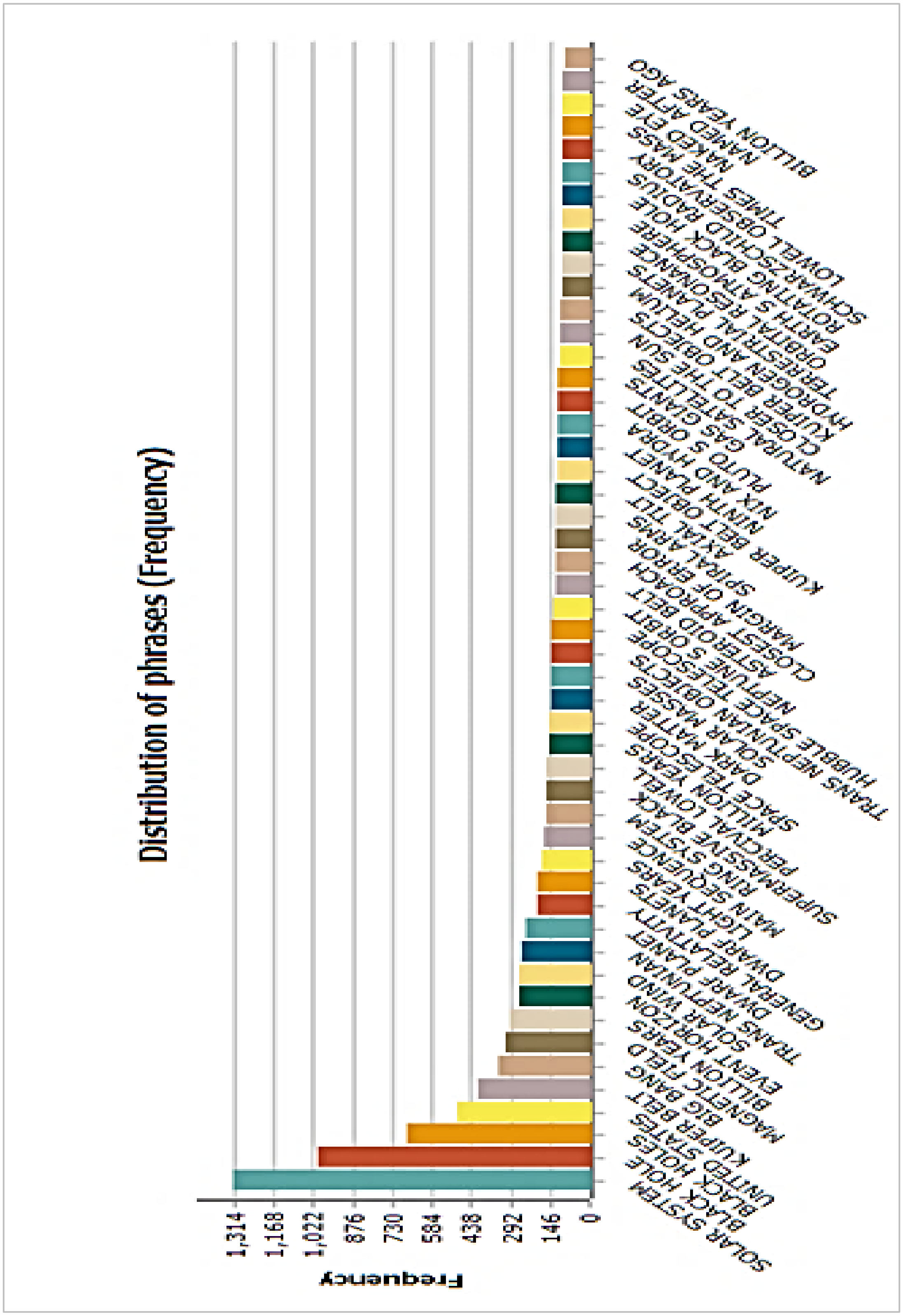


Figure 45: Distribution of phrases for control group

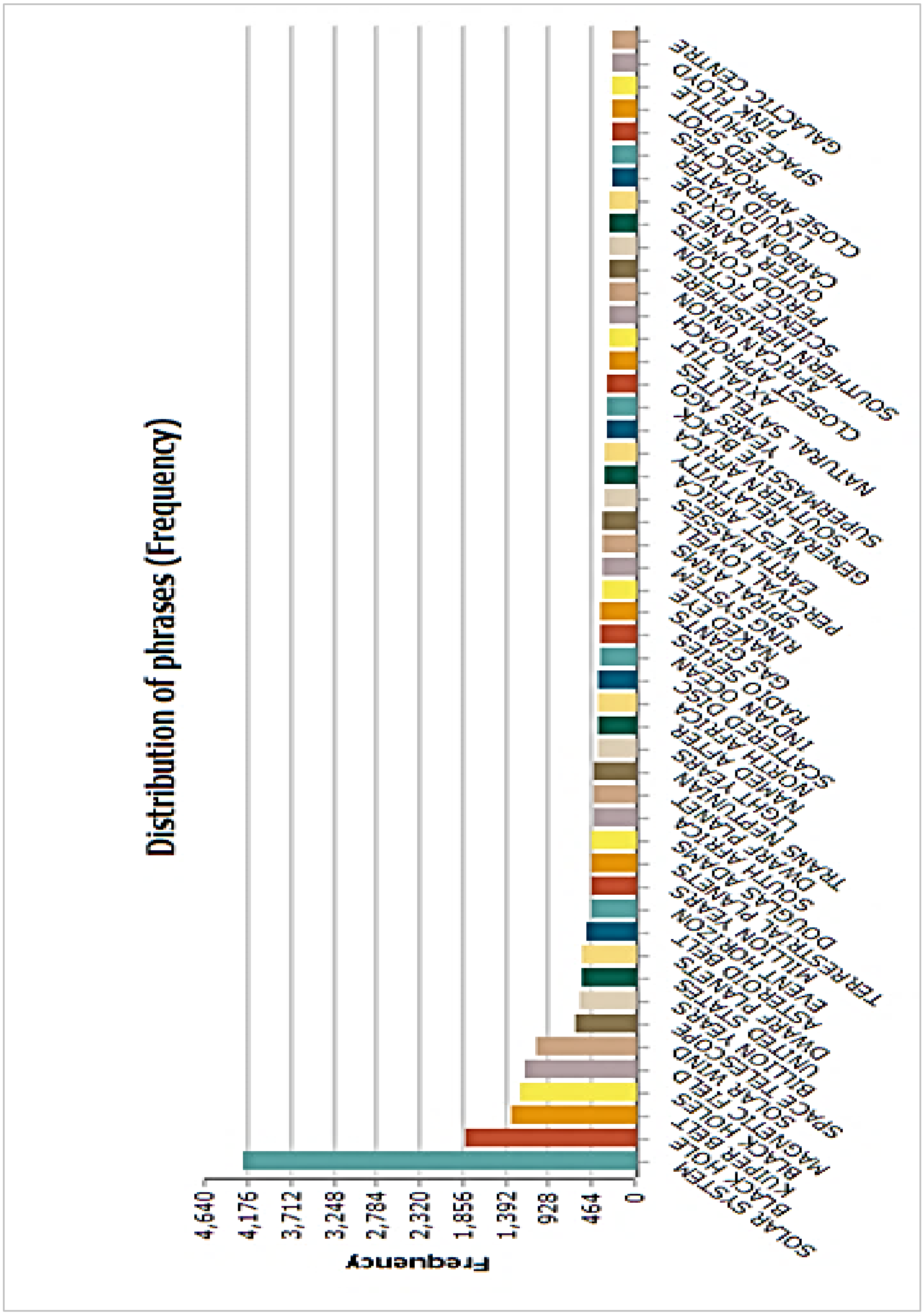


Figure 46: Distribution of phrases for personalized support group

Chapter 7: Conclusion and Future Work

Personalized learning advantages have become evident through research and practice. Even though, most of early efforts in personalized learning focused on formal learning, there is a growing undeniable demand for personalized informal learning. Wikis, especially Wikipedia, are experiencing an enormous attention for informal learning. The nature of wikis allows users to freely navigate and construct knowledge without being forced to follow a predefined learning path or curriculums. However, several limitations are associated with link-based navigation and keyword-based search hindering users' ability to adequately reach relevant content. As a result, there is a need to facilitate easy and fast navigation of relevant content to support informal learning from information wikis.

Additionally, evaluation of informal learning in similar environment is a challenging task due to absence of formal assessments and learning analytics. Consequently, there is a need to define evaluation metrics and tools of informal learning on similar environments.

This dissertation proposed an effective personalized content recommendation framework as well as an evaluation framework based on web analytics. User studies were designed to assess informal learning from Wikipedia.

7.1 Summary of Contributions

• Glossary and taxonomies of personalized learning systems, architectural components, and major challenges.

A comprehensive, systematic review of personalized learning software systems is presented. In the review, glossary of terms, taxonomies of software learning environments, learning content, and learner modeling approaches are presented. The strengths and drawbacks of different personalized learning software systems components are highlighted. Also, a reusable software architecture for personalized learning systems [57] is proposed. This can help in early design stages of personalized learning software system. Finally, a comparison and classification of commonly used user interest models in information-oriented websites and specifically on Wikipedia is presented.

• An effective semantic analysis technique suitable for massively diverse unstructured text found in massively diverse information wikis.

An effective semantic analysis approach based on concepts from fuzzy set information retrieval model is designed and developed. The proposed technique uses fuzzy thesauri to generate feature vector representations of different language units, i.e. words, topics … etc. which can be used for text mining, recommendations, and other tasks involving the use of unstructured text. The proposed technique is implemented in the context of recommender systems as well as sentiment analysis to assess the applicability of the proposed technique in multiple contexts with different document sizes. The preliminary results in Twitter sentiment analysis using fuzzy set-based feature vectors are published in ISCMI16 [58], the complete Twitter Fuzzy Set-based Sentiment Analysis Framework and evaluations are published in Soft

Computing Journal [59], and the semantic analysis tasks based on fuzzy thesauri related to recommender systems is published IEEE Access.

• A personalized content recommender system based on user's navigation graph and fuzzy thesaurus.

A user interest model based on topical navigation graphs is proposed. The proposed model is effective in capturing changes in user interests during navigation sessions. By integrating this user interest model with the proposed semantic analysis technique based on fuzzy sets, an effective personalized content recommendation framework to support informal learning in massively diverse information wikis is designed and developed. The evaluation reveals 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%. High-level conceptualization of the proposed framework is published in ACM UMAP18 [56]. Detailed design, implementation, and evaluation of the proposed framework is accepted for publication in IEEE Access.

• Evaluation methods and metrics to assess informal learning on wiki environments.

We design an approach to evaluate the impact of personalized recommendations on informal learning. First, the impact of personalized recommendations on informal learning is evaluated by assessing conceptual knowledge in users' feedback. An assessment rubric is designed, adapted from concept map-based rubric for conceptual knowledge assessment, then, user studies are designed and run to evaluate the impact of personalized recommendations on informal learning. Second, web analytics data is analyzed to get an insight into users' progress and focus throughout the test sessions and propose an evaluation framework based on web analytics. The evaluation reveals that the personalized content recommendations enhances user experience on Wikipedia. Evaluation of informal learning show that users who used Wikipedia with personalized recommendations achieve higher scores on conceptual knowledge assessment compared to those who used Wikipedia without recommendations. Furthermore, they can make use of larger number of concepts, make comparisons, and state relations between concepts. Web analytics-based evaluation show that those who used Wikipedia with personalized recommendations can make use of a larger number of relevant keywords and phrases. Results of conceptual knowledge assessment is published in EDUCON19 [60]. The proposed evaluation framework is accepted for publication in iJEP Journal.

7.2 Future Work

- Information wikis offer flexible and attractive environments for informal learning. Currently, many corporates are implementing wikis to foster knowledge sharing among employees. Personalized recommendations can aid in recommending relevant articles without the need to conduct explicit search. This can facilitate fast and easy access to useful information as well as help save employees' time and efforts. Additionally, personalized recommendations can help recommending colleagues viewing similar topics or working on similar subjects that can encourage collaboration among employees in the workplace.
- Software environments with similar properties of wikis' users and content can benefit from the proposed framework. For example, online libraries can enhance readers' experience by implementing personalized recommendation of textual content.

Currently, most library recommendation systems implement content-based recommendation models trained on various combination of index attributes such as author, subject, publisher…etc. These types of recommendations are powerful in making recommendations of specific books. However, it will be very helpful to provide also recommendations within books. For instance, section-level recommendations, or chapter-level recommendations for readers while they are reading online. The proposed framework with topical navigation graphs can be adapted to provide this type of recommendations. It can be also used in social networks to effectively provide personalized content recommendations.

- Web analytics have long been used to provide valuable insights specifically for e-marketing purposes. A major advantage of web analytics over other analytics approaches is that analytics can be inferred automatically from web usage data without any explicit intervention from the user. This dissertation has shown that mining web data analytics can also provide rich information that can be used to evaluate informal learning. Evaluation of informal learning is so not trivial task with the absence of assessments and predefined learning outcomes. As a result, giving feedback to learners, or enhancing the learners experience based on any type of indicators is not easy. A comprehensive evaluation framework can be built on top or as an extension to the framework proposed in this dissertation to provide feedback to learners or provide corrective feedback to the recommendation framework to improve the quality of recommendations.

References

[1] Google Inc. 2016, "How Ads Work - Report," Google , 2016.

[2] Biletskiy, Y., Baghi, H., Keleberda, I., & Fleming, M., "An adjustable personalization of search and delivery of learning objects to learners.," *Expert Systems with Applications,* vol. 36, no. 5, pp. 9113-9120, 2009.

[3] Khrib, M. K., Jemn, M., & Nasraoui, O, "Automatic recommendations for e-learning personalization based on web usage mining techniques and information retrieval," in *Eighth IEEE International Conference on Advanced Learning Technologies, ICALT'08.*, 2008.

[4] Zhang, Y., & Jiao, J. R. , "An associative classification-based recommendation system for personalization in B2C e-commerce applications.," *Expert Systems with Applications,* vol. 33, no. 2, pp. 357-367, 2007.

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

[6] Bloom, B. S. , "The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring.," *Educational Researcher,* vol. 13, no. 6, pp. 4-16, 1984.

[7] Mills, C., & Dalgarno, B., "A conceptual model for game-based intelligent tutoring systems.," in *Proceedings of the 2007 Australasian Society for Computers in Learning in Tertiary Education*, 2007.

[8] Schiaffino, S., Garcia, P., & Amandi, A., "eTeacher: Providing personalized assistance to e-learning students.," *Computers & Education,* vol. 51, no. 4, pp. 1744-1754, 2008.

[9] Hong, C. M., Chen, C. M., Chang, M. H., & Chen, S. C., "Intelligent web-based tutoring system with personalized learning path guidance.," in *Seventh IEEE International Conference on Advanced Learning Technologies*, 2007.

[10] Verdú, E., Regueras, L. M., Gal, E., de Castro, J. P., Verdú, M. J., & Kohen-Vacs, D. , "Integration of an intelligent tutoring system in a course of computer network design.," *Educational Technology Research and Development,,* vol. 65, no. 3, pp. 653-677, 2017.

[11] Zhang, B., & Jia, J., "Evaluating an Intelligent Tutoring System for Personalized Math Teaching," in *2017 International Symposium on Educational Technology (ISET)*, 2017.

[12] Motiwalla, L. F., "Mobile learning: A framework and evaluation.," *Computers & Education,* vol. 49, no. 3, pp. 581-596, 2007.

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

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

[15] Streumer, J. N., & Kho, M., "The world of work-related learning," *Work-related learning,* pp. 3-49, 2006.

[16] Selwyn, N., "Web 2.0 applications as alternative environments for informal learning-a critical review," in *International Expert Meeting on ICT and Educational Performance*, 2007.

[17] García-Peñalvo, F. J., Colomo-Palacios, R., & Lytras, M. D. , "Informal learning in work environments: training with the Social Web in the workplace," *Behaviour & Information Technology,* vol. 31, no. 8, pp. 753-755, 2012.

[18] Coll, S. D., & Treagust, D. , "Blended learning environment: an approach to enhance students's learning experiences outside school," *MIER Journal of Educational Studies, Trends and Practices,* vol. 7, no. 2, pp. 121-134, 2018.

[19] Zheng, B., Niiya, M., & Warschauer, M., "Wikis and collaborative learning in higher education.," *Technology, Pedagogy and Education,* vol. 24, no. 3, pp. 357-374, 2015.

[20] Neumann, D. L., & Hood, M., "The effects of using a wiki on student engagement and learning of report writing skills in a university statistics course.," *Australasian Journal of Educational Technology,* vol. 25, no. 3, pp. 382-398, 2009.

[21] Kimmerle, J., Moskaliuk, J., & Cress, U., "Using wikis for learning and knowledge building: Results of an experimental study.," *Journal of Educational Technology & Society,* vol. 14, no. 4, pp. 138–148, 2011.

[22] Chu, S. K. W., Zhang, Y., Chen, K., Chan, C. K., Lee, C. W. Y., Zou, E., & Lau, W., "The effectiveness of wikis for project-based learning in different disciplines in higher education.," *The internet and higher education,* vol. 33, no. 1, pp. 49-60, 2017.

[23] Sharp, L. A., & Whaley, B., "Wikis as Online Collaborative Learning Experiences:"A Different Kind of Brainstorming"," *Adult Learning,* vol. 29, no. 3, pp. 1-11, 2018.

[24] Wikimedia , "Wikimedia Statistics," 2019. [Online]. Available: https://stats.wikimedia.org/v2/#/all-wikipedia-projects. [Accessed August 2019].

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

[26] Zukerman, I., & Albrecht, D. W., "Predictive statistical models for user modeling.," *User Modeling and User-Adapted Interaction,* vol. 11, no. 1-2, pp. 5-18, 2001.

[27] Albert, D., & Hockemeyer, C., "Applying demand analysis of a set of test problems for developing adaptive courses.," in *Computers in Education*, 2002.

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

[29] Chin, D. N. , "KNOME: Modeling what the user knows in UC.," *User Models in Dialog Systems,* pp. 74-107, 1989.

[30] Nascimento HP, Marques LB, de Souza DG, Salgado FM, Bessa RQ, "A AIED Game to help children with learning disabilities in literacy in the Portuguese language," in *SBC - Proceedings of SBGames*, 2012.

[31] Santos, O. C., "Emotions and personality in adaptive e-learning systems: an affective computing perspective.," in *Emotions and Personality in Personalized Services*, Cham, Springer, 2016, pp. 263-285.

[32] Klašnja-Milićević, A., Vesin, B., Ivanović, M., & Budimac, Z, "E-Learning personalization based on hybrid recommendation strategy and learning style identification," *Computers & Education,* vol. 56, no. 3, pp. 885-899, 2011.

[33] Petersen, S. A., & Markiewicz, J. K, "PALLAS: personalised language learning on mobile devices.," in *Wireless, Mobile, and Ubiquitous Technology in Education*, 2008.

[34] Mitrovic, A., Mayo, M., Suraweera, P., & Martin, B., "Constraint-based tutors: a success story.," in *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, 2001.

[35] Hagerty, G., & Smith, S., "Using the web-based interactive software ALEKS to enhance college algebra.," *Mathematics & Computer Education*, vol. 39, no. 3, pp. 183-194, 2005.

[36] Brusilovsky, P., & Millán, E., "User models for adaptive hypermedia and adaptive educational systems.," *The Adaptive Web,* pp. 3-53, 2007.

[37] Manouselis, N., Drachsler, H., Vuorikari, R., Hummel, H., & Koper, R., "Recommender systems in technology enhanced learning," *Recommender Systems Handbook,* Boston, Springer, pp. 387-415, 2011.

[38] Drachsler, H., Verbert, K., Santos, O. C., & Manouselis, N. , "Panorama of recommender systems to support learning," in *Recommender Systems Handbook*, Boston, Springer, pp. 421-451, 2015.

[39] Xiao, J., Wang, M., Jiang, B., & Li, J., "A personalized recommendation system with combinational algorithm for online learning.," J*ournal of Ambient Intelligence and Humanized Computing,* vol. 9, no. 3, pp. 667-677, 2018.

[40] Dwivedi, P., Kant, V., & Bharadwaj, K. K., "Learning path recommendation based on modified variable length genetic algorithm.," *Education and Information Technologies,* vol. 23, no. 2, pp. 819-836, 2018.

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

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

[43] Lops, P., De Gemmis, M., & Semeraro, G., " Content-based recommender systems: State of the art and trends.," in *Recommender systems handbook*, Boston, Springer, 2011, pp. 73-105.

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

[45] Adline, A. A., & Mahalakshmi, G. S., "A recommender system for wiki pages: Usage based rating approach," in *Recent Trends In Information Technology (ICRTIT)*, 2012.

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

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

[48] Silveira, T., Zhang, M., Lin, X., Liu, Y., & Ma, S., "How good your recommender system is? A survey on evaluations in recommendation," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 5, pp. 813-831, 2019.

[49] Rodi, G. C., Loreto, V., & Tria, F., "Search strategies of Wikipedia readers," *PloS one*, vol. 12, no. 2, 2017, https://doi.org/10.1371/journal.pone.0170746.

[50] West, R., & Leskovec, J., "Automatic Versus Human Navigation in Information Networks," in *ICWSM*, 2012.

[51] Helic, D, "Analyzing user click paths in a wikipedia navigation game," in *MIPRO*, 2012.

[52] Gaeta, M., Orciuoli, F., & Ritrovato, P. , "Advanced ontology management system for personalised e-Learning," *Knowledge-Based Systems*, vol. 22, no. 4, pp. 292-301, 2009.

[53] Di Noia, T., Mirizzi, R., Ostuni, V. C., Romito, D., & Zanker, M., "Linked open data to support content-based recommender systems.," in *Proceedings of the 8th International Conference on Semantic Systems*, 2012.

[54] Leake, D., Maguitman, A., & Reichherzer, T., "Understanding knowledge models: Modeling assessment of concept importance in concept maps.," in *Twenty-sixth Annual Conference of The Cognitive Science Society*, vol. 26, no. 26. 2004.

[55] Garcia, I., & Ng, Y. K., "Eliminating Redundant and Less-Informative RSS News Articles Based on Word Similarity and a Fuzzy Equivalence Relation. In Tools with Artificial Intelligence," in *ICTAI'06*, 2006.

[56] Ismail, H., "WikiRec: A Personalized Content Recommendation Framework to Support Informal Learning in Wikis.," in *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*, Singapore, 2018.

[57] Ismail, H., & Belkhouche, B., "A Reusable Software Architecture for Personalized Learning Systems.," in *2018 International Conference on Innovations in Information Technology (IIT)*, 2018.

[58] Ismail, H. M., Zaki, N., & Belkhouche, B., "Using custom fuzzy thesaurus to incorporate semantic and reduce data sparsity for twitter sentiment analysis.," in *3rd International Conference on Soft Computing & Machine Intelligence (ISCMI)*, 2016.

[59] Ismail, H. M., Belkhouche, B., & Zaki, N., "Semantic Twitter sentiment analysis based on a fuzzy thesaurus," *Soft Computing*, vol. 22, no. 18, pp. 6011-6024, 2018.

[60] Ismail, H., & Belkhouche, B., "Evaluating the Impact of Personalized Content Recommendations on Informal Learning from Wikipedia.," in *IEEE Global Engineering Education Conference (EDUCON)*, 2019.

[61] Burke, R., "Hybrid recommender systems: Survey and experiments.," *User Modeling and User-Adapted Interaction*, vol. 12, no. 4, pp. 331-370, 2002.

[62] Lops, P., De Gemmis, M., & Semeraro, G., "Content-based recommender systems: State of the art and trends," *In Recommender Systems Handbook*, pp. 73-105, 2011.

[63] Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S., "Collaborative filtering recommender systems," in *The Adaptive Web*, Springer, pp. 291-324, 2007.

[64] Adomavicius, G., & Tuzhilin, A., "Context-aware recommender systems.," in *Recommender systems handbook*, Springer, pp. 217-253, 2011.

[65] Burke, R., "Knowledge-based recommender systems.," *Encyclopedia of Library and Information Systems*, vol. 69, no. Supplement 32, pp. 175-186, 2000.

[66] Abel, F., Bittencourt, I. I., Henze, N., Krause, D., & Vassileva, J., "A rule-based recommender system for online discussion forums.," in *International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*, 2008.

[67] Mooney, R. J., & Roy, L., "Content-based book recommending using learning for text categorization," in *Proceedings of the fifth ACM conference on Digital libraries*, 2000.

[68] Philip, S., Shola, P., & Ovye, A., "Application of content-based approach in research paper recommendation system for a digital library.," *International Journal of Advanced Computer Science and Applications*, vol. 5, no. 10, pp. 49-176, 2014.

[69] AlMurtadha, Y., Sulaiman, M. N., Mustapha, N., & Udzir, N. I., "Improved web page recommender system based on web usage mining.," in *Proceedings of the 3rd International Conference on Computing and Informatics (ICOCI)*, 2011.

[70] Goldberg, D., Nichols, D., Oki, B. M., & Terry, D., "Using collaborative filtering to weave an information tapestry.," *Communications of the ACM*, vol. 35, no. 12, pp. 61-71, 1992.

[71] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J, "GroupLens: an open architecture for collaborative filtering of netnews," in *In Proceedings of the 1994 ACM conference on Computer supported cooperative work*, 1994.

[72] Ge, M., Delgado-Battenfeld, C., & Jannach, D., "Beyond accuracy: evaluating recommender systems by coverage and serendipity.," in *Proceedings of the fourth ACM conference on Recommender systems*, 2010.

[73] Zhang, Y. C., Séaghdha, D. Ó., Quercia, D., & Jambor, T., "Auralist: introducing serendipity into music recommendation.," in *Proceedings of the fifth ACM international conference on Web search and data mining*, 2012.

[74] Torres, R., McNee, S. M., Abel, M., Konstan, J. A., & Riedl, J., "Enhancing digital libraries with TechLens+.," in *Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries*, 2004.

[75] Sundar, S. S., Oeldorf-Hirsch, A., & Xu, Q., "The bandwagon effect of collaborative filtering technology.," in *In CHI'08 extended abstracts on Human factors in computing systems*, 2008.

[76] Kim, H. N., & El Saddik, A., "Personalized PageRank vectors for tag recommendations: inside FolkRank.," in *Proceedings of the fifth ACM conference on Recommender systems*, 2011.

[77] Lee, S., Song, S. I., Kahng, M., Lee, D., & Lee, S. G., "Random walk based entity ranking on graph for multidimensional recommendation.," in *Proceedings of the fifth ACM conference on Recommender systems*, 2011.

[78] Liben-Nowell, D., & Kleinberg, J., "The link-prediction problem for social networks.," *journal of the Association for Information Science and Technology,* vol. 58, no. 7, pp. 1019-1031, 2007.

[79] Al Hasan, M., & Zaki, M. J., "A survey of link prediction in social networks.," in *Social network data analytics* , Boston, Springer, 2011, pp. 243-275.

[80] Jannach, D., Lerche, L., Gedikli, F., & Bonnin, G., "What recommenders recommend–an analysis of accuracy, popularity, and sales diversity effects.," in *International Conference on User Modeling, Adaptation, and Personalization*, 2013.

[81] Cremonesi, P., Garzotto, F., Negro, S., Papadopoulos, A. V., & Turrin, R.Cremonesi, P., Garzotto, F., Negro, S., Papadopoulos, A. V., & Turrin, R., "Looking for “good” recommendations: A comparative evaluation of recommender systems.," in *IFIP Conference on Human-Computer Interaction*, Berlin, 2011.

[82] Piaget, J., Language and Thought of the Child: Selected Works, vol. 5, Routledge, 2005.

[83] Alexa, "The top 500 sites on the web," Alexa, 2019. [Online]. Available: https://www.alexa.com/topsites. [Accessed September 2019].

[84] Wikipedia, "About Wikipedia," Wikimedia, 2015. [Online]. Available: https://en.wikipedia.org/wiki/Wikipedia:About. [Accessed 15 September 2018].

[85] Watts, D. J., & Strogatz, S. H., "Collective dynamics of ‘small-world’networks.," *Nature,* vol. 393, no. 6684, pp. 440-442, 1998.

[86] Bell, J. S., & Mitchell, R, "Competency-based versus traditional cohort-based technical education: a comparison of students' perceptions," *Journal of Career and Technical Education,* vol. 17, no. 1, pp. 5-22, 2000.

[87] Wolf, M. A, "Innovate to Educate: System [Re]Design for Personalized Learning: A report from the 2010 symposium," Software & Information Industry Association (SIIA), Washington, 2010.

[88] Kihlstrom, J, "How Students Learn and How We Can Help Them," How Students Learn, 2011.

[89] Wang, F. H., "On extracting recommendation knowledge for personalized web-based learning based on ant colony optimization with segmented-goal and meta-control strategies.," *Expert Systems with Applications,* vol. 7, no. 39, pp. 6446-6453, 2012.

[90] S. A. &. M. J. K. Petersen, "PALLAS: personalised language learning on mobile devices.," in *Wireless, Mobile, and Ubiquitous Technology in Education*, 2008.

[91] Chen, C. M., & Li, Y. L., "Personalised context-aware ubiquitous learning system for supporting effective English vocabulary learning.," *Interactive Learning Environments,* vol. 18, no. 4, pp. 341-364, 2010.

[92] Chen, C. M., Chung, C. J., & Chen, M. C., "Personalized Intelligent English Vocabulary Learning System Based on Item Response Theory and Learning Memory Cycle.," in *IEEE International Conference on Systems, Man and Cybernetics, 2006. SMC'06.*, 2006.

[93] Nascimento, H. P. C. P. A., Marques, L. B., de Souza, D. G., Salgado, F. M., & Bessa, R. Q., "A AIED Game to help children with learning disabilities in literacy in the Portuguese language," in *SBC - Proceedings of SBGames*, 2012.

[94] Jung, J., & Graf, S, "An approach for personalized web-based vocabulary learning through word association games," in *International Symposium on Applications and the Internet*, SAINT, 2008.

[95] Van Staalduinen, J. P., & de Freitas, S., "A game-based learning framework: Linking game design and learning.," in *Learning to play: exploring the future of education with video games*, 2011.

[96] Kiili, K., "Digital game-based learning: Towards an experiential gaming model.," *The Internet and higher education,* vol. 8, no. 1, pp. 13-24., 2005.

[97] Rovai, A. P., "A constructivist approach to online college learning.," *The internet and higher Education,* vol. 7, no. 2, pp. 79-93, 2004.

[98] Cha, H. J., Kim, Y. S., Park, S. H., Yoon, T. B., Jung, Y. M., & Lee, J. H., "Learning styles diagnosis based on user interface behaviors for the customization of learning interfaces in an intelligent tutoring system.," in *International Conference on Intelligent Tutoring Systems*, 2006.

[99] Bhatia, L., & Prasad, S. S, "COPAL—Cognitive personalized aid for learning," in *Cognitive Computing and Information Processing (CCIP)*, 2015.

[100] MacGregor, S. K., & Lou, Y., "Web-based learning: How task scaffolding and web site design support knowledge acquisition.," *Journal of Research on Technology in Education,* vol. 37, no. 2, pp. 161-175, 2004.

[101] Touré, C., Michel, C., & Marty, J. C. , "Towards extending traditional informal learning tools in the workplace with social functionalities.," *International Journal of Learning Technology,* vol. 12, no. 2, pp. 119-150, 2017.

[102] Dabbagh, N., & Kitsantas, A., "Personal Learning Environments, social media, and self-regulated learning: A natural formula for connecting formal and informal learning.," *The Internet and higher education,* vol. 15, no. 1, pp. 3-8, 2012.

[103] Laverde, A. C., Cifuentes, Y. S., & Rodriguez, H. Y. R., "Toward an instructional design model based on learning objects.," *Educational Technology Research and Development,* vol. 55, no. 6, pp. 671-681, 2007.

[104] Churchill, D., " Educational applications of Web 2.0: Using blogs to support teaching and learning.," *British journal of educational technology,* vol. 40, no. 1, pp. 179-183, 2009.

[105] Yang, J. C., Quadir, B., Chen, N. S., & Miao, Q. , " Effects of online presence on learning performance in a blog-based online course.," *The Internet and Higher Education,* vol. 30, pp. 11-20, 2016.

[106] Coll, C., Rochera, M. J., & de Gispert, I. , "Supporting online collaborative learning in small groups: Teacher feedback on learning content, academic task and social participation.," *Computers & Education,* vol. 75, pp. 53-64, 2014.

[107] Akbari, E., Pilot, A., & Simons, P. R. J., "Autonomy, competence, and relatedness in foreign language learning through Facebook.," *Computers in Human Behavior,* vol. 48, pp. 126-134, 2015.

[108] Su, P. H., Wang, Y. B., Yu, T. H., & Lee, L. S. , "A dialogue game framework with personalized training using reinforcement learning for computer-assisted language learning.," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013.

[109] Abbey, S., Joglekar, S., & Bedekar, M., "Comparison of Personalisation Systems," in *Sixth International Conference on Emerging Trends in Engineering and Technolog*, 2013.

[110] Ghazali, A. S. M., Noor, S. F. M., & Saad, S., "Review of personalized learning approaches and methods in e-learning environment.," in *Electrical Engineering and Informatics (ICEEI),*, 2015.

[111] Raptis, G. E., Katsini, C., Belk, M., Fidas, C., Samaras, G., & Avouris, N., "Using eye gaze data and visual activities to infer human cognitive styles: method and feasibility studies," in *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, 2017.

[112] Rich, E. , "Users are individuals: individualizing user models," *International journal of man-machine studies,* vol. 18, no. 3, pp. 199-214., 1983.

[113] Desmarais, M. C., & d Baker, R. S., "A review of recent advances in learner and skill modeling in intelligent learning environments.," *User Modeling and User-Adapted Interaction,* vol. 29, no. 1-2, pp. 9-38, 2012.

[114] Mauss, I. B., & Robinson, M. D. , "Measures of emotion: A review.," *Cognition and emotion,* vol. 23, no. 2, pp. 209-237, 2009.

[115] Dictionary, C., Cambridge advanced learner’s dictionary, Cambridge : PONS-Worterbucher, Klett Ernst Verlag GmbH, 2008.

[116] Hearst, M. , Search user interfaces, Cambridge : Cambridge University Press, 2009.

[117] Glover, E. J., Lawrence, S., Birmingham, W. P., & Giles, C. L., "Architecture of a metasearch engine that supports user information needs.," in *Proceedings of the eighth international conference on Information and knowledge management*, 1999.

[118] Glover, E. J., Lawrence, S., Gordon, M. D., Birmingham, W. P., & Giles, C. L., "Web Search---Your Way," *Communications of the ACM,* vol. 44, no. 12, pp. 97-102, 2001.

[119] Bernstein, M. S., Suh, B., Hong, L., Chen, J., Kairam, S., & Chi, E. H., "Eddi: interactive topic-based browsing of social status streams.," in *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, 2010.

[120] Szomszor, M., Alani, H., Cantador, I., O'Hara, K., & Shadbolt, N., "Semantic modelling of user interests based on cross-folksonomy analysis.," in *International Semantic Web Conference*, 2008.

[121] Kelly, D., Gyllstrom, K., & Bailey, E. W., "A comparison of query and term suggestion features for interactive searching.," in *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, 2009.

[122] Peltonen, J., Strahl, J., & Floréen, P., "Negative relevance feedback for exploratory search with visual interactive intent modeling.," in *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, 2017.

[123] Ahn, J. W., & Brusilovsky, P., "Adaptive visualization of search results: Bringing user models to visual analytics.," *Information Visualization,* vol. 8, no. 3, pp. 167-179, 2009.

[124] Chang, J. C., Hahn, N., Perer, A., & Kittur, A., "SearchLens: composing and capturing complex user interests for exploratory search," in *IUI*, 2019.

[125] Teevan, J., Alvarado, C., Ackerman, M. S., & Karger, D. R., "The perfect search engine is not enough: a study of orienteering behavior in directed search," in *ACM SIGCHI conference on Human factors in computing systems*, 2004.

[126] White, R. W., & Drucker, S. M. , "Investigating behavioral variability in web search.," in *ACM 16th international conference on World Wide Web*, 2007.

[127] Lieberman, H., "Letizia: An agent that assists web browsing," in *ACM IJCAI*, 1995.

[128] Moukas, A. , "Amalthaea information discovery and filtering using a multiagent evolving ecosystem.," *Applied Artificial Intelligence,* vol. 11, no. 5, pp. 437-457, 1997.

[129] Sakagami, H., & Kamba, T., "Learning personal preferences on online newspaper articles from user behaviors.," *Computer Networks and ISDN Systems,* vol. 29, no. 8-13, pp. 1447-1455, 1997.

[130] Pazzani, M. J., Muramatsu, J., & Billsus, D., "Syskill & Webert: Identifying interesting web sites," *AAAI/IAAI,* vol. 1, pp. 54-61, 1996.

[131] Chen, L., & Sycara, K. , "WebMate: A personal agent for browsing and searching," *Agents,* vol. 98, pp. 132-139, 1998.

[132] Chowdhury, G. G., Introduction to modern information retrieval., Facet publishing, 2010.

[133] Asnicar, F. A., & Tasso, C., "ifWeb: a prototype of user model-based intelligent agent for document filtering and navigation in the world wide web.," in *Sixth International Conference on User Modeling*, 1997.

[134] Stefani, A., & Strapparava, C., "Exploiting NLP techniques to build user model for Web sites: the use of WordNet in SiteIF Project.," in *2nd Workshop on Adaptive Systems and User Modeling on the WWW*, 1999.

[135] Gentili, G., Micarelli, A., & Sciarrone, F., "Infoweb: An adaptive information filtering system for the cultural heritage domain.," *Applied Artificial Intelligence,* vol. 17, no. 8-9, pp. 715-744, 2003.

[136] Tanudjaja, F., & Mui, L., "Persona: A contextualized and personalized web search.," in *In Proceedings of the 35th Annual Hawaii International Conference on System Sciences*, 2002.

[137] De Luca, E. W., Nürnberger, A., & Von-Guericke, O., "Ontology-based semantic online classification of documents: Supporting users in searching the web.," in *Proc. of the European Symposium on Intelligent Technologies*, 2004.

[138] Sieg, A., Mobasher, B., & Burke, R. D., "Learning ontology-based user profiles: A semantic approach to personalized web search.," *IEEE Intelligent Informatics Bulletin,* vol. 8, no. 1, pp. 7-18, 2007.

[139] Challam, V., Gauch, S., & Chandramouli, A., "Contextual search using ontology-based user profiles," in *Large Scale Semantic Access to Content (Text, Image, Video, and Sound)*, 2007.

[140] Fazeli, S., Loni, B., Drachsler, H., & Sloep, P., "Which recommender system can best fit social learning platforms?," in *European Conference on Technology Enhanced Learning*, Cham, 2014.

[141] Bieliková, M., Šimko, M., Barla, M., Tvarožek, J., Labaj, M., Móro, R., ... & Ševcech, J., "ALEF: From application to platform for adaptive collaborative learning," in *Recommender Systems for Technology Enhanced Learning* , New York, NY, 2014.

[142] Cosley, D., Frankowski, D., Terveen, L., & Riedl, J., "SuggestBot: using intelligent task routing to help people find work in wikipedia.," in *Proceedings of the 12th international conference on Intelligent user interfaces*, 2007.

[143] Wulczyn, E., West, R., Zia, L., & Leskovec, J., "Growing wikipedia across languages via recommendation.," in *Proceedings of the 25th International Conference on World Wide Web*, 2016.

[144] Huang, E., Kim, H. J., & Ko, H., "Task Recommendation on Wikipedia.," *Data Processing,* 2010.

[145] Vijayarani, S., Ilamathi, M. J., & Nithya, M. , "Preprocessing techniques for text mining-an overview," *International Journal of Computer Science & Communication Networks,* vol. 5, no. 1, pp. 7-16, 2015.

[146] Kao, A., & Poteet, S. R. (Eds.)., Natural Language Processing and Text Mining., Springer Science & Business Media, 2007.

[147] Islam, A., & Inkpen, D., "Semantic text similarity using corpus-based word similarity and string similarity.," *ACM Transactions on Knowledge Discovery from Data (TKDD),* vol. 2, no. 2, p. 10, 2008.

[148] Meadow, C. T., Boyce, B. R., & Kraft, D. H., Text information retrieval systems, San Diego, CA: Academic Press, 1992.

[149] Salton, G., Wong, A., & Yang, C. S., "A vector space model for automatic indexing.," *Communications of the ACM,* vol. 18, no. 11, pp. 613-620, 1975.

[150] Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R., "Indexing by latent semantic analysis.," *Journal of the American society for information science,* vol. 41, no. 6, pp. 391-407, 1990.

[151] Landauer, T. K., Foltz, P. W., & Laham, D., "An introduction to latent semantic analysis.," *Discourse processes,* vol. 25, no. 2-3, pp. 259-284, 1998.

[152] Abdellatif, T. M., Capretz, L. F., & Ho, D., "Automatic recall of software lessons learned for software project managers," *Information and Software Technology,* vol. 115, pp. 44-57, 2019.

[153] Zadeh, "Fuzzy Sets," *Information and Control,* vol. 8, pp. 338-353, 1965.

[154] Ogawa, Y., Morita, T., & Kobayashi, K., "A Fuzzy Document Retrieval System Using The Keyword Connection Matrix and a Learning Method," *Fuzzy Sets and Systems,* vol. 39, no. 2, pp. 163-179, 1991.

[155] Yerra, R., & Ng, Y. K., "Detecting Similar HTML Documents Using a Fuzzy Set Information Retrieval Approach," in *Granular Computing IEEE International Conference,* 2005.

[156] Lee, B. Pang and L., "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval,* 2008.

[157] Abbasi, A., Chen, H., & Salem, A., "Sentiment Analysis in Multiple Languages: Features selection for Opinion Classification in Web Forums," *ACM Transactions on Information Systems (TOIS),* pp. 1-34, 2008.

[158] Go, A., Bhayani, R., & Huang, L., "Twitter Sentiment Classification using Distant Supervision," Stanford Digital Library Technologies Projects, 2009.

[159] Porter, M. F, "An Algorithm for Suffix Stripping," *Program,* vol. 14, no. 3, pp. 130-137, 1980.

[160] Saif, H., Fernandez, M., He, Y., & Alani, H., "Evaluation Datasets for Twitter Sentiment Analysis A Survey and A New Dataset, The STS-Gold," in *Interantional Workshop on Emotion and Sentiment in Social and Expressive Media: Approaches and Perspectives from AI (ESSEM 2013),* Italy, 2013.

[161] Barbosa, L., & Feng, J., "Robust Sentiment Detection on Twitter from Biased and Noisy Data.," in *the 23rd International Conference on Computational Linguistics*, 2010.

[162] Ismail, H. M., Harous, S., & Belkhouche, B., "A Comparative Analysis of Machine Learning Classifiers for Twitter Sentiment Analysis.," in *17th International Conference on Intelligent Text Processing and Computational Linguistics-CICLing*, 2016.

[163] Agarwal, A., Xie, B., Vovsha, I., Rambow, O., "Sentiment Analysis of Twitter Data," in *Proceedings of the workshop on languages in social media*, 2011.

[164] "LOL, OMG and ILY: 60 of The Dominating Abbreviations," Just English, 2014. [Online]. Available: http://justenglish.me/2014/07/18/lol-omg-and-ily-60-of-the-dominating-abbreviations/. [Accessed November 2015].

[165] Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J., Data Mining: Practical Machine Learning Tools and Techniques., Morgan Kaufmann, 2016.

[166] Hotho, A., Nürnberger, A., & Paaß, G., "A Brief Survey of Text Mining.," *Ldv Forum*, vol. 20, no. 1, pp. 19-62, 2005.

[167] Sokolova, M., & Lapalme, G., "A systematic analysis of performance measures for classification tasks.," *Information Processing & Management*, vol. 45, no. 4, pp. 427-437, 2009.

[168] Bollen, J., Mao, H., & Pepe, A., "Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena," in *International AAAI Conference on Weblogs and Social Media*, Barcelona, 2011.

[169] Cambria, E., Speer, R., Havasi, C., & Hussain, A., "SenticNet: A Publicly Available Semantic Resource for Opinion Mining.," *AAAI fall symposium: commonsense knowledge*, vol. 10, 2010.

[170] Speriosu, M., Sudan, N., Upadhyay, S., & Baldridge, J. , "Twitter Polarity Classification with Label Propagation over Lexical Links and the Follower Graph," in *Conference on Empirical Methods in Natural Language Processing*, UK, 2011.

[171] Borges, J., & Levene, M., "Data mining of user navigation patterns," in *International Workshop on Web Usage Analysis and User Profiling*, Berlin, 1999.

[172] Smith, K. A., & Ng, A. , "Web page clustering using a self-organizing map of user navigation patterns," *Decision Support Systems,* vol. 35, no. 2, pp. 245-256, 2003.

[173] Cadez, I., Heckerman, D., Meek, C., Smyth, P., & White, S, "Model-based clustering and visualization of navigation patterns on a web site," *Data Mining and Knowledge Discovery,* vol. 7, no. 4, pp. 399-424, 2003.

[174] Catledge, L. D., & Pitkow, J. E. (1995), "Characterizing browsing strategies in the World-Wide Web," *Computer Networks and ISDN systems,* vol. 27, no. 6, pp. 1065-1073, 1995.

[175] Chen, S. Y., & Macredie, R. D., "Cognitive styles and hypermedia navigation: Development of a learning model.," *Journal of the Association for Information Science and Technology,* vol. 53, no. 1, pp. 3-15, 2002.

[176] Liegle, J. O., & Janicki, T. N. , "The effect of learning styles on the navigation needs of Web-based learners.," *Computers in human behavior,* vol. 22, no. 5, pp. 885-898, 2006.

[177] Graf, S., & Liu, T. C., "Analysis of learners' navigational behaviour and their learning styles in an online course.," *Journal of Computer Assisted Learning,* vol. 26, no. 2, pp. 116-131, 2010.

[178] Page, L., Brin, S., Motwani, R., & Winograd, T., "The PageRank citation ranking: Bringing order to the web.," Stanford InfoLab., 1999.

[179] Haveliwala, T. H., "Topic-sensitive pagerank.," in *In Proceedings of the 11th international conference on World Wide Web,* 2002.

[180] Andersen, R., Chung, F., & Lang, K., "Local graph partitioning using pagerank vectors.," in *FOCS'06. 47th Annual IEEE Symposium on Foundations of Computer Science*, 2006.

[181] Zaki, N., Berengueres, J., & Efimov, D., "ProRank: a method for detecting protein complexes.," in *Proceedings of the 14th annual conference on Genetic and evolutionary computation*, 2012.

[182] Zualkernan, I. A., AbuJayyab, M. A., & Ghanam, Y. A. , "An alignment equation for using mind maps to filter learning queries from Google.," in *Sixth IEEE International Conference on Advanced Learning Technologies* , 2006.

[183] Beel, J., Langer, S., Genzmehr, M., & Nürnberger, A., "Introducing Docear's research paper recommender system.," in *In Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries*, 2013.

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

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

[186] Chin, D. N. , "Empirical evaluation of user models and user-adapted systems.," *User modeling and user-adapted interaction*, vol. 11, no. 1-2., pp. 181-194, 2001.

[187] Pu, P., Chen, L., & Hu, R., "A user-centric evaluation framework for recommender systems," in *fifth ACM conference on Recommender systems*, 2011.

[188] Cañas, A. J., Novak, J. D., & Vanhear, J., "Concept map-based knowledge assessment tasks and their scoring criteria: An overview.," PP. 80-88, 2012.

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

[190] Parmenter, D, Key performance indicators: developing, implementing, and using winning KPIs., John Wiley & Sons, 2015.

[191] Ismail, Heba M., Boumediene Belkhouche, and Nazar Zaki., "Semantic Twitter sentiment analysis based on a fuzzy thesaurus," *Soft Computing*, vol. 22, no. 18, pp. 6011-6024, 2018.

[192] Ismail, Heba, Saad Harous, and Boumediene Belkhouche., "A Comparative Analysis of Machine Learning Classifiers for Twitter Sentiment Analysis," *Research in Computing Science*, vol. 110, pp. 71-83, 2016.

[193] Ismail, Heba, Boumediene Belkhouche, and Saad Harous., "Mining Web Analytics Data for Information Wikis to Evaluate Informal Learning," *International Journal of Engineering Pedagogy (iJEP),* p. Accepted, 2019.

[194] Ismail, Heba, Boumediene Belkhouche, and Saad Harous, "A Framework for Personalized Content Recommendations to Support Informal Learning in Massively Diverse Information Wikis," *IEEE Acess,* p. Submitted, 2019.

[195] Ismail, Heba, and Boumediene Belkhouche., "Evaluating the Impact of Personalized Content Recommendations on Informal Learning from Wikipedia," in *2019 IEEE Global Engineering Education Conference (EDUCON)*, Dubai, 2019.

[196] Ismail, Heba., "WikiRec: A Personalized Content Recommendation Framework to Support Informal Learning in Wikis," in *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*, Singapore, 2018.

[197] Ismail, Heba, and Boumediene Belkhouche, "A Reusable Software Architecture for Personalized Learning Systems," in *2018 International Conference on Innovations in Information Technology (IIT)*, 2018.

[198] Belkhouche, B., Ismail, H., & Ramsi, F. , "A Model to Support Outside Classroom Learning.," in *2019 IEEE Global Engineering Education Conference (EDUCON)* , Dubai, 2019.

[199] Ismail, Heba M., Nazar Zaki, and Boumediene Belkhouche, "Using custom fuzzy thesaurus to incorporate semantic and reduce data sparsity for twitter sentiment analysis," in *2016 3rd International Conference on Soft Computing & Machine Intelligence (ISCMI)*, 2016.

[200] Ismail, Heba M., Saad Harous, and Boumediene Belkhouche, "Review of personalized language learning systems.," in *2016 12th International Conference on Innovations in Information Technology (IIT)*, 2016.

[201] Belkhouche, Boumediene, and Heba Ismail., "Personalized learning.," in *2016 IEEE Global Engineering Education Conference (EDUCON)*, 2016.

List of Publications

Journal Articles

[1] Ismail, Heba, Boumediene Belkhouche, and Saad Harous. 2019. "Framework for Personalized Content Recommendations to Support Informal Learning in Massively Diverse Information Wikis." IEEE Acess (IEEE) 7: 172752-172773. doi:10.1109/ACCESS.2019.2956284.

[2] Ismail, Heba, Boumediene Belkhouche, and Saad Harous. 2020. "Mining Web Analytics Data for Information Wikis to Evaluate Informal Learning." International Journal of Engineering Pedagogy (iJEP) 20 (1).

[3] Ismail, Heba M., Boumediene Belkhouche, and Nazar Zaki. 2018. "Semantic Twitter sentiment analysis based on a fuzzy thesaurus." Soft Computing (Springer) 22 (18): 6011-6024.

[4] Ismail, Heba, Saad Harous, and Boumediene Belkhouche. 2016. "A Comparative Analysis of Machine Learning Classifiers for Twitter Sentiment Analysis." Research in Computing Science 110: 71-83.

Conference Papers

[1] Ismail, Heba, and Boumediene Belkhouche. 2019. "Evaluating the Impact of Personalized Content Recommendations on Informal Learning from Wikipedia." 2019 IEEE Global Engineering Education Conference (EDUCON). Dubai: IEEE. 943-952.

[2] Ismail, Heba. 2018. "WikiRec: A Personalized Content Recommendation Framework to Support Informal Learning in Wikis." Proceedings of the 26th

Conference on User Modeling, Adaptation and Personalization. Singapore: ACM. 273-276.

[3] Ismail, Heba, and Boumediene Belkhouche. 2018. "A Reusable Software Architecture for Personalized Learning Systems." 2018 International Conference on Innovations in Information Technology (IIT). IEEE. 105-110.

[4] Ismail, Heba M., Nazar Zaki, and Boumediene Belkhouche. 2016. "Using custom fuzzy thesaurus to incorporate semantic and reduce data sparsity for twitter sentiment analysis." 2016 3rd International Conference on Soft Computing & Machine Intelligence (ISCMI). Dubai: IEEE. 47-52.

[5] Ismail, Heba M., Saad Harous, and Boumediene Belkhouche. 2016. "Review of personalized language learning systems." 2016 12th International Conference on Innovations in Information Technology (IIT). Dubai: IEEE. 1-6.

[6] Ismail, Heba, and Boumediene Belkhouche. 2018. "A Reusable Software Architecture for Personalized Learning Systems." 2018 International Conference on Innovations in Information Technology (IIT). Dubai: IEEE. 105-110.

[7] Belkhouche, B., Ismail, H., & Ramsi, F. . 2019. "A Model to Support Outside Classroom Learning. ." 2019 IEEE Global Engineering Education Conference (EDUCON) . Dubai: IEEE. 1069-1078.

[8] Belkhouche, Boumediene, and Heba Ismail. 2016. "Personalized learning." 2016 IEEE Global Engineering Education Conference (EDUCON). Abu Dhabi: IEEE. 848-853.

Appendices

Appendix – A

SSREC Approval

Social Sciences Research Ethics Committee
-Approval-

Proposal number:	*ERS_2018_5726*
Title of Project:	Learning Content Personalization Framework for Technology-Enhanced Adaptive Learning
PI:	Heba Ismail
Co-PI:	Dr Boumediene Belkhouche

The above proposal has been reviewed by:

☒ one member of the Social Sciences REC
☐ two members of the *Social Sciences REC*

And the decision is:

☒ Favourable
☐ Favourable with Additional Conditions
☐ Provisional Opinion
☐ Unfavourable Opinion
☐ No Opinion (Proportionate Review* only)

Reason:

After evaluating this proposal, we see no major ethical concerns. Therefore, the proposal is approved for one year.

Please ensure that you indicate to research participants that your study has received ethical approval from UAE University by referring to the proposal number.

Name (Chair or designee): Clara Morgan

Clara Morgan

Signature

March 10, 2018
Date

Appendix – B

Consent Forms

United Arab Emirates University

College of Information Technology
Department of Computer Science and Software Engineering

Technology-Enhanced Adaptive Learning Research

<u>Participant Informed Consent Form</u>

Participant/Student Name: ____

Kindly note that this is a test of the educational software framework. We are **not** testing your son/daughter. We want to find out what aspects of the software are confusing/appealing to the students, so we can make it better. We will not videotape or audiotape the session. We may publish our results from this and other sessions in our research reports, but all our results are anonymous. Identity of students are not used at all in our research report.

Research Overview
This is a study about technology-enhanced learning, intended for students who use the internet to search for information that support their learning. Our goal is to design smart technological solutions that are expected to identify and model the learning needs of learners and provide adapted support that can improve the effectiveness and efficiency of learning experiences, with and without the involvement of the teacher.

Procedure
In the test session students will be working with our testing online encyclopaedia designed specifically for school students, targeted around the UK National Curriculum, and useful for much of the English-speaking world. Students are expected to search for information about a specific topic and at the end submit a short writeup about the topic. Detailed instructions will be explained during the session.

All information collected in the session belongs to the college of Information Technology at UAE University and will be used for research purposes only.

Statement of Informed Consent

I have read the description of the study and I voluntarily agree for my son/daughter to participate in the study.

Parent Print Name: ____________________________

Signature: ____________________________

Date: ____________________________

Appendix – C

Questionnaires

Learning Content Personalization Framework for Technology-Enhanced Adaptive Learning

(Students' Demographics)

Please complete the following questions: *(you can ask questions anytime)*

1. Gender	
F	
M	

2. Do you search the internet for information	
Yes	
No	

If your answer to question number 2 is Yes, please complete the following questions, if your answer is No, please skip.

3. How often do you use the internet to search for information	
Daily	
More than 2 days a week	
Once or twice a month	

4. How fast do you find the information you're looking for?	
Immediately	
It takes me long time	
It depends on the topic, sometimes I find the information quickly and sometimes it takes long time.	

5. What is the most commonly source of your information on the web?	
Wikipedia and similar encyclopedias	
I only use Google	
I don't know	

6. Are you excited about this research study?	
So excited	
Not too excited	
I don't know	

Please submit the completed questionnaire to the study administrator at the end of the experiment along with the other papers.

Thank you for completing the questionnaire!

😊

Learning Content Personalization Framework for Technology-Enhanced Adaptive Learning

(Students' Feedback)

Please complete the following questions: *(you can ask questions anytime)*

1. Please write down the group color.
Color:

2. How fast did you find the information you're looking for?	
Immediately	☐
It takes me long time	☐
I couldn't find the information I was looking for	☐

3. Do you have recommendations on your website?	
Yes	☐
No	☐

If your answer to question number 3 is Yes, please complete the following questions, if your answer is No, please skip the remaining questions and submit your questionnaire forms.

4. Were the recommendations useful?	
Very useful	☐
useful	☐
I don't know	☐
Not useful	☐

5. Were the recommendations relevant to the topic you're searching for?	
Very relevant	☐
relevant	☐
I don't know	☐
Not relevant	☐

6. Do you think it's helpful to have similar recommendations on the websites that you commonly use?	
Yes	☐
No	☐
I don't know	☐

7. Please indicate whether the displayed recommendations are relevant or not as instructed by the moderator.	
After 5 minutes	Feedback
After 40 minutes	Feedback

Please submit the completed questionnaire to the study administrator at the end of the experiment along with the other papers.

Thank you for completing the questionnaire! 😊

Appendix – D

Writing Challenge

Writing Challenge

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.

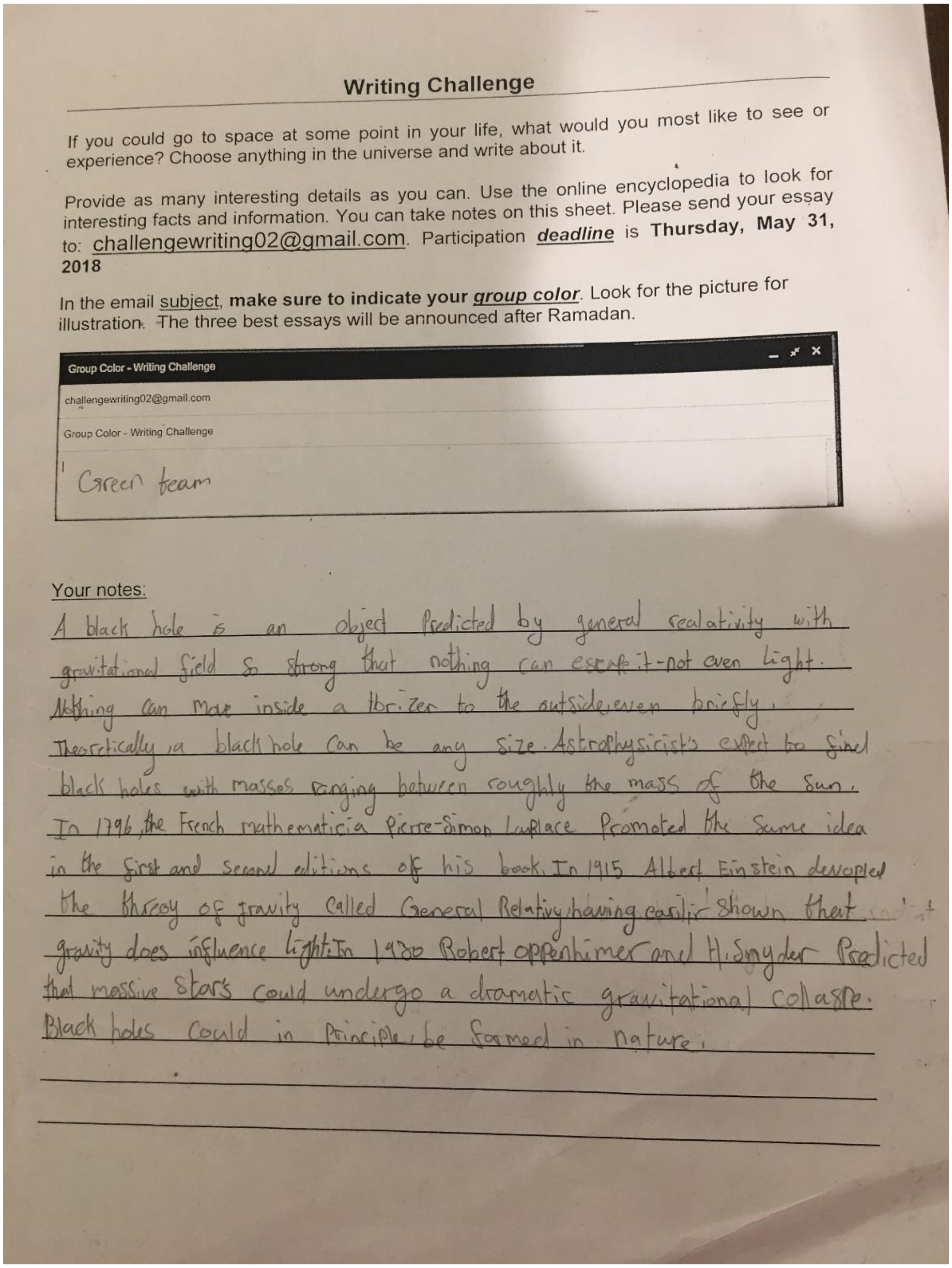
Provide as many interesting details as you can. Use the online encyclopedia to look for interesting facts and information. You can take notes on this sheet. Please send your essay to: challengewriting02@gmail.com. Participation ***deadline*** is **Thursday, May 31, 2018**

In the email subject, **make sure to indicate your *group color***. Look for the picture for illustration. The three best essays will be announced after Ramadan.

Your notes:

Appendix – E

An example of one student participation in the writing challenge – Control Group



Writing Challenge

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.

Provide as many interesting details as you can. Use the online encyclopedia to look for interesting facts and information. You can take notes on this sheet. Please send your essay to: challengewriting02@gmail.com. Participation ***deadline*** is **Thursday, May 31, 2018**

In the email subject, **make sure to indicate your *group color***. Look for the picture for illustration. The three best essays will be announced after Ramadan.

Group Color - Writing Challenge

challengewriting02@gmail.com

Group Color - Writing Challenge

Green team

Your notes:

A black hole is an object Predicted by general realativity with gravitational field so strong that nothing can escape it-not even light. Nothing can move inside a horizen to the outside,even briefly. Theoretically,a black hole can be any size. Astrophysicist's expect to find black holes with masses ranging between roughly the mass of the Sun. In 1796,the French mathematicia Pierre-Simon Laplace Promoted the Same idea in the first and second editions of his book. In 1915 Albert Einstein develpled the threoy of gravity called General Relativity,having earlir shown that gravity does influence light. In 1930 Robert oppenhimer and H. Snyder Predicted that massive stars could undergo a dramatic gravitational collaspe. Black holes could in Principle,be formed in nature.

Appendix – F

An example of one student participation in the writing challenge – Personalized Support Group

A Trip to MARS

Mars is **the second smallest planet** in the solar system after mercury. Mars is at a distance of more than 142 million miles from the Sun. And it is the fourth planet from the sun. Mars **can be spotted with the naked eye** during the night time from the surface of earth, thanks to its reddish appearance. **Galileo Galilei** observed Mars in 1609 with a basic telescope. Mars is also referred to as the "Red Planet" because it has high content of iron oxide which makes it appear reddish. Mars has only two moons as compared to Uranus' twenty seven. They are named:**Phobos and Deimos**. These moons are small and are irregularly shaped.

The ancient Babylonians first created the week and divided it into seven days. They named each day of the week upon the seven known bodies in the sky: the Sun, the Moon, Mars, Mercury, Venus, Jupiter and Saturn. Hence, **Tuesday is the day of Mars.** Because of the red color of the planet, Mars is also associated with aggression. The name of the month **March also derives from Mars.**

Mars has low atmospheric pressure on its surface which is the reason why liquid water cannot exist on its surface for long. If we compare the density of Mars with that of the Earth's, we would find that it is 100 times less dense than Earth. Mars' gravitational force is weaker than that of the earth by 66%. Thus, anything would **weight more on Mars as compared to that on Earth.** Mars takes twice the time Earth takes to complete a full revolution around the sun. The orbit of Mars and Earth are also not in the same plane. There is a tilt of a few degrees between the orbits of the two neighboring planets.

The amount of land surface available on Mars is almost equal to that available on the Earth. Mars **has all four seasons as that of Earth.** However, each season on the Red Planet lasts twice as long as that on earth. The length of one day on Mars is almost identical to that of the Earth. Unlike the Earth, **on Mars sunsets are blue.** Mars has sufficient ice on its polar region. If the ice on its south pole melts, the resulting water will be sufficient to cover the planet's entire surface to a depth of 11 meters. **Mars, in addition to Earth, is the only other planet that has polar ice caps.** It's Northern cap is called – Planum Boreum, and Southern cap is called – Planum Australe. The **tallest mountain in the Solar System** – Olympus Mons – is on Mars. It is more than 21 km high (about three times taller than the Mt. Everest) and more than 600 km in diameter. The **longest and the deepest canyon** in the entire solar system is on Mars – "Valles Marineris" (maximum length – 4000 km, maximum width – 200 km, and maximum depth – 7 km). Compared to the Grand Canyon, it is 4 times deeper and 6 times longer.

Mars is being explored for over four decades now because it is considered as a favorable place for human existence. And the signs of water on the planet has supported the belief that human life can exist on this red planet. NASA is **the only space exploration agency** that has managed to land on Mars so far. Viking 1 and Viking 2 are the two spacecraft that were sent by NASA in 1975 on Mars with the objective to study its surface and to gather important information about its composition and structure. (Launch dates: August 20, 1975 (Viking 1); September 9, 1975 (Viking 2)). The **first photograph of Mars** taken from its surface was taken by Viking 1 on July 20 1976.Currently, **two rovers from NASA** named, Opportunity and Curiosity, are exploring the surface of Mars.

NASA has **plans to create an Earth Independent colony on Mars** by the end of 2030. "Seek Signs Of Life" is the exploration strategy that NASA is currently following to find out the possibilities of life on Mars in the past or present. In an attempt to find the possibility of life on Mars, scientists are typically interested in finding the evidence of water and organics – the chemical building blocks of life. According to NASA, it would roughly take **2 and a half years to make a round trip to Mars from Earth.** This also includes the time that the astronauts would have to allow Mars and Earth to re-align for the return trip.

Scientist have tried to grow plants in soil that mimics Martian soil and they have succeeded in their **attempt to grow tomatoes, peas, and rye** from the said soil. Now, this is going to help those wishing to spend some time on the Red Planet. **Worm-like aliens have been recently spotted on Mars** surface by NASA JPL's Mars orbiter. This spacecraft has been orbiting Mars for the past 11 years.

The Hope Mars Mission or Emirates Mars Mission is a space exploration probe mission to Mars, set to be launched by the United Arab Emirates in 2020. I am 10 years now. In 20 years I would like to go to Mars to experience some of the fantastic things I wrote above which I found in my research. I love to see the worm-like aliens.

[redacted]

Appendix – G

Users' Feedback on Recommendations

CRD First Observation

CRD	Number of nodes in navigation graph:	5>= |TNG| >=1

Recommendations@K	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15	U16	U17	U18	U19	U20
Recommendations@1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Recommendations@2	1 0	1 0	1 1	1 0	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	0 1
Recommendations@3	1 0 1	1 0 1	0 1 1	1 0 1	0 1 1	1 0 1	0 1 1	1 0 1	0 0 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 0	1 1 1	1 1 1	0 1 1
Recommendations@4	0 0 1 1	1 0 1 1	0 0 1 1	1 1 1 1	0 0 0 1	0 0 1 1	0 0 1 1	1 0 1 1	0 0 1 1	0 0 1 1	1 1 1 1	1 0 1 1	0 0 1 1	0 0 1 1	1 1 1 1	1 1 1 1	1 1 0 1	1 1 1 0	1 1 1 1	0 1 1 1
Recommendations@5	0 0 1 1 1	1 0 1 1 1	0 0 1 1 1	1 1 1 1 1	0 0 0 1 1	0 0 1 1 1	0 0 1 1 1	1 0 1 1 0	0 0 1 1 1	0 0 1 1 0	0 0 1 1 1	0 0 1 1 1	0 0 1 1 0	0 0 1 1 1	1 1 1 1 0	0 0 1 1 1	0 0 0 1 1	1 1 1 0 0	1 1 1 1 1	0 1 1 1 1

P@K	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15	U16	U17	U18	U19	U20
P@1	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
P@2	50.00%	50.00%	100.00%	50.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	50.00%
P@3	66.67%	66.67%	66.67%	66.67%	66.67%	66.67%	66.67%	66.67%	33.33%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	66.67%	100.00%	100.00%	66.67%
P@4	50.00%	75.00%	50.00%	100.00%	25.00%	50.00%	50.00%	75.00%	50.00%	50.00%	100.00%	75.00%	50.00%	50.00%	100.00%	100.00%	75.00%	75.00%	100.00%	75.00%
P@5	60.00%	80.00%	60.00%	100.00%	40.00%	60.00%	60.00%	60.00%	60.00%	40.00%	60.00%	60.00%	40.00%	60.00%	80.00%	60.00%	40.00%	60.00%	100.00%	80.00%
AP@1	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
AP@2	75.00%	75.00%	100.00%	75.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	75.00%
AP@3	72.22%	72.22%	88.89%	72.22%	88.89%	88.89%	88.89%	88.89%	77.78%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	88.89%	100.00%	100.00%	72.22%
AP@4	66.67%	72.92%	79.17%	79.17%	72.92%	79.17%	79.17%	85.42%	70.83%	87.50%	100.00%	93.75%	87.50%	87.50%	100.00%	100.00%	85.42%	93.75%	100.00%	72.92%
AP@5	65.33%	74.33%	75.33%	83.33%	66.33%	75.33%	75.33%	80.33%	68.67%	78.00%	92.00%	87.00%	78.00%	82.00%	96.00%	92.00%	76.33%	87.00%	100.00%	74.33%

MAP@1		MAP@2		MAP@3		MAP@4		MAP@5	
MAP@1	100.00%	MAP@2	95.00%	MAP@3	90.00%	MAP@4	84.69%	MAP@5	80.35%

CRD Second Observation

CRD	Number of nodes in navigation graph:	5<< |TNG|

Recommendations@K	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15	U16	U17	U18	U19	U20
Recommendations@1	0	1	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0
Recommendations@2	0 0	1 0	0 0	1 0	0 0	0 0	0 0	1 0	0 0	0 0	0 0	0 0	0 0	0 0	1 1	0 0	0 0	1 1	1 1	0 1
Recommendations@3	0 0 1	1 0 1	0 0 1	1 0 1	0 0 0	0 0 1	0 0 1	1 0 1	0 0 1	0 0 1	0 0 1	0 0 1	0 0 1	0 0 1	1 1 1	0 0 1	0 0 1	1 1 1	1 1 1	0 1 1
Recommendations@4	0 0 1 1	1 0 1 1	0 0 1 1	1 0 1 1	0 0 0 1	0 0 1 1	0 0 1 1	1 0 1 1	0 0 1 1	0 1 1 1	0 0 1 1	0 0 1 1	0 1 1 1	0 0 1 1	1 1 1 1	0 0 1 1	0 0 1 1	1 1 1 0	1 1 1 1	0 1 1 1
Recommendations@5	0 0 1 1 1	1 0 1 1 1	0 0 1 1 1	1 0 1 1 1	0 0 0 1 1	0 0 1 1 1	0 0 1 1 1	1 0 1 1 1	0 0 1 1 1	0 1 1 1 1	0 1 1 1 1	0 0 1 1 1	0 1 1 1 0	0 1 1 1 1	1 1 1 1 0	0 0 1 1 1	0 0 1 1 1	1 1 1 1 1	1 1 1 1 1	0 1 1 1 1

P@K	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15	U16	U17	U18	U19	U20
P@1	0.00%	100.00%	0.00%	100.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	100.00%	100.00%	0.00%
P@2	0.00%	50.00%	0.00%	50.00%	0.00%	0.00%	0.00%	50.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	100.00%	100.00%	50.00%
P@3	33.33%	66.67%	33.33%	66.67%	0.00%	33.33%	33.33%	66.67%	33.33%	33.33%	33.33%	33.33%	33.33%	33.33%	100.00%	33.33%	33.33%	100.00%	100.00%	66.67%
P@4	50.00%	75.00%	50.00%	75.00%	25.00%	50.00%	50.00%	75.00%	50.00%	75.00%	50.00%	50.00%	75.00%	50.00%	100.00%	50.00%	50.00%	75.00%	100.00%	75.00%
P@5	60.00%	80.00%	60.00%	80.00%	40.00%	60.00%	60.00%	80.00%	60.00%	80.00%	80.00%	60.00%	60.00%	80.00%	80.00%	60.00%	60.00%	100.00%	100.00%	80.00%
AP@1	0.00%	100.00%	0.00%	100.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	100.00%	100.00%	0.00%
AP@2	0.00%	75.00%	0.00%	75.00%	0.00%	0.00%	0.00%	75.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	100.00%	100.00%	25.00%
AP@3	11.11%	72.22%	11.11%	72.22%	0.00%	11.11%	11.11%	72.22%	11.11%	11.11%	11.11%	11.11%	11.11%	11.11%	100.00%	11.11%	11.11%	100.00%	100.00%	38.89%
AP@4	20.83%	72.92%	20.83%	72.92%	6.25%	20.83%	20.83%	72.92%	20.83%	27.08%	20.83%	20.83%	27.08%	20.83%	100.00%	20.83%	20.83%	93.75%	100.00%	47.92%
AP@5	28.67%	74.33%	28.67%	74.33%	13.00%	28.67%	28.67%	74.33%	28.67%	37.67%	32.67%	28.67%	33.67%	32.67%	96.00%	28.67%	28.67%	95.00%	100.00%	54.33%

MAP@1		MAP@2		MAP@3		MAP@4		MAP@5	
MAP@1	30.00%	MAP@2	27.50%	MAP@3	34.44%	MAP@4	41.46%	MAP@5	47.37%

HARD First Observation

HARD	Number of nodes in navigation graph:	5>= |TNG| >=1

Recommendations@K	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15	U16	U17	U18	U19	U20
Recommendations@1	1	1	1	1	1	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1
Recommendations@2	1 1	1 0	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1
Recommendations@3	1 1 1	1 1 1	0 1 1	1 1 1	1 1 1	1 0 1	0 1 1	1 0 1	0 0 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 0	1 1 1	1 1 1	1 1 1
Recommendations@4	1 1 1 1	1 1 1 1	1 1 1 0	1 1 0 1	1 1 0 0	0 1 1 1	1 0 1 0	0 1 1 1	1 1 1 1	1 1 1 1	1 1 1 1	1 0 1 1	0 0 1 1	1 1 1 1	1 1 1 1	1 1 1 1	1 1 0 1	1 1 1 0	1 1 1 1	1 1 1 1
Recommendations@5	1 1 1 1 1	1 1 1 1 0	1 1 1 1 0	1 1 1 1 0	1 1 1 1 1	1 1 1 1 0	1 1 1 1 1	1 1 1 1 0	1 1 1 0 1	1 1 0 1 1	1 0 1 1 1	1 1 1 1 0	1 0 1 1 0	0 0 1 1 1	1 1 1 1 0	0 0 1 1 1	1 1 1 1 1	1 1 1 1 0	1 1 1 1 1	0 1 1 1 1

P@K	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15	U16	U17	U18	U19	U20
P@1	100.00%	100.00%	100.00%	100.00%	100.00%	0.00%	100.00%	0.00%	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
P@2	100.00%	50.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
P@3	100.00%	100.00%	66.67%	100.00%	100.00%	66.67%	66.67%	66.67%	33.33%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	66.67%	100.00%	100.00%	100.00%
P@4	100.00%	100.00%	75.00%	75.00%	50.00%	75.00%	50.00%	75.00%	100.00%	100.00%	100.00%	75.00%	50.00%	100.00%	100.00%	100.00%	75.00%	75.00%	100.00%	100.00%
P@5	100.00%	80.00%	80.00%	80.00%	100.00%	80.00%	100.00%	80.00%	80.00%	80.00%	80.00%	80.00%	60.00%	60.00%	80.00%	60.00%	100.00%	80.00%	100.00%	80.00%
AP@1	100.00%	100.00%	100.00%	100.00%	100.00%	0.00%	100.00%	0.00%	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
AP@2	100.00%	75.00%	100.00%	100.00%	100.00%	50.00%	100.00%	50.00%	50.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
AP@3	100.00%	83.33%	88.89%	100.00%	100.00%	55.56%	88.89%	55.56%	44.44%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	88.89%	100.00%	100.00%	100.00%
AP@4	100.00%	87.50%	85.42%	93.75%	87.50%	60.42%	79.17%	60.42%	58.33%	100.00%	100.00%	93.75%	87.50%	100.00%	100.00%	100.00%	85.42%	93.75%	100.00%	100.00%
AP@5	100.00%	86.00%	84.33%	91.00%	90.00%	64.33%	83.33%	64.33%	62.67%	96.00%	96.00%	91.00%	82.00%	92.00%	96.00%	92.00%	88.33%	91.00%	100.00%	96.00%

MAP@1	85.00%	MAP@2	91.25%	MAP@3	90.28%	MAP@4	88.65%	MAP@5	87.32%

HARD Second Observation

HARD	Number of nodes in navigation graph:	5<< |TNG|

Recommendations@K	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15	U16	U17	U18	U19	U20
Recommendations@1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Recommendations@2	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1
Recommendations@3	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 0 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 0	1 1 1	1 1 1	1 1 1
Recommendations@4	1 1 1 1	1 1 1 1	1 1 1 0	1 1 1 1	1 1 1 0	1 1 1 1	1 0 1 0	1 1 1 1	1 1 1 1	1 1 1 1	1 1 1 1	1 0 1 1	1 1 1 1	1 1 1 1	1 1 1 1	1 1 1 1	1 1 0 1	1 1 1 0	1 1 1 1	1 1 1 1
Recommendations@5	1 1 1 1 1	1 1 1 1 0	1 1 1 1 0	1 1 1 1 0	1 1 1 1 1	1 1 1 1 0	1 1 1 1 1	1 1 1 1 0	1 1 1 0 1	1 1 0 1 1	1 1 1 1 1	1 1 1 1 0	1 1 1 1 0	1 1 1 1 1	1 1 1 1 0	1 1 1 1 1	1 1 1 1 1	1 1 1 1 0	1 1 1 1 1	1 1 1 1 1

P@K	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15	U16	U17	U18	U19	U20
P@1	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
P@2	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
P@3	100.00%	100.00%	100.00%	100.00%	100.00%	66.67%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	66.67%	100.00%	100.00%	100.00%
P@4	100.00%	100.00%	75.00%	100.00%	75.00%	100.00%	50.00%	100.00%	100.00%	100.00%	100.00%	75.00%	100.00%	100.00%	100.00%	100.00%	75.00%	75.00%	100.00%	100.00%
P@5	100.00%	80.00%	80.00%	80.00%	100.00%	80.00%	100.00%	80.00%	80.00%	80.00%	100.00%	80.00%	80.00%	100.00%	80.00%	100.00%	100.00%	80.00%	100.00%	100.00%
AP@1	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
AP@2	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
AP@3	100.00%	100.00%	100.00%	100.00%	100.00%	88.89%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	88.89%	100.00%	100.00%	100.00%
AP@4	100.00%	100.00%	93.75%	100.00%	93.75%	91.67%	87.50%	100.00%	100.00%	100.00%	100.00%	93.75%	100.00%	100.00%	100.00%	100.00%	85.42%	93.75%	100.00%	100.00%
AP@5	100.00%	96.00%	91.00%	96.00%	95.00%	89.33%	90.00%	96.00%	96.00%	96.00%	100.00%	91.00%	96.00%	100.00%	96.00%	100.00%	88.33%	91.00%	100.00%	100.00%

MAP@1	MAP@2	MAP@3	MAP@4	MAP@5
100.00%	100.00%	98.89%	96.98%	95.38%

Baseline First Observation

Baseline	Number of nodes in navigation graph:	5>= |TNG| >=1

Recommendations@K	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15	U16	U17	U18	U19	U20
Recommendations@1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Recommendations@2	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
Recommendations@3	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
Recommendations@4	0 0 0 0	0 0 0 1	0 0 0 0	0 0 0 1	0 0 0 0	0 0 0 1	0 0 0 0	0 0 0 1	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 1	0 0 0 0	0 0 0 1	0 0 0 0	0 0 0 1	0 0 0 0	0 0 0 1	0 0 0 0	0 0 0 1
Recommendations@5	0 0 0 1 1	0 0 0 1 1	0 0 0 1 1	0 0 0 1 1	0 0 0 1 1	0 0 0 1 1	0 0 0 1 1	0 0 0 1 1	0 0 0 0 1	0 0 0 1 1	0 0 0 0 1	0 0 0 1 1	0 0 0 0 0	0 0 0 1 1	0 0 0 1 1	0 0 0 1 1	0 0 0 1 1	0 0 0 1 1	0 0 0 1 1	0 0 0 1 1

P@K	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15	U16	U17	U18	U19	U20
P@1	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
P@2	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
P@3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
P@4	0.00%	25.00%	0.00%	25.00%	0.00%	25.00%	0.00%	25.00%	0.00%	0.00%	0.00%	25.00%	0.00%	25.00%	0.00%	25.00%	0.00%	25.00%	0.00%	25.00%
P@5	40.00%	40.00%	40.00%	40.00%	40.00%	40.00%	40.00%	40.00%	20.00%	40.00%	20.00%	40.00%	0.00%	40.00%	40.00%	40.00%	40.00%	40.00%	40.00%	40.00%
AP@1	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
AP@2	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
AP@3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
AP@4	0.00%	6.25%	0.00%	6.25%	0.00%	6.25%	0.00%	6.25%	0.00%	0.00%	0.00%	6.25%	0.00%	6.25%	0.00%	6.25%	0.00%	6.25%	0.00%	6.25%
AP@5	8.00%	13.00%	8.00%	13.00%	8.00%	13.00%	8.00%	13.00%	4.00%	8.00%	4.00%	13.00%	0.00%	13.00%	8.00%	13.00%	8.00%	13.00%	8.00%	13.00%

MAP@1		MAP@2		MAP@3		MAP@4		MAP@5	
	0.00%		0.00%		0.00%		2.81%		9.45%

Baseline Second Observation

Baseline	Number of nodes in navigation graph:	5<< |TNG|

Recommendations@K	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15	U16	U17	U18	U19	U20
Recommendations@1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Recommendations@2	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
Recommendations@3	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
Recommendations@4	0 1 0 0	1 0 0 1	0 0 0 0	0 1 1 1	0 0 1 1	0 0 1 1	0 0 0 0	0 0 0 1	0 0 0 0	1 0 0 0	0 0 0 0	0 1 1 1	0 0 0 0	1 1 1 1	0 1 1 0	0 1 1 1	0 0 0 0	0 0 0 1	0 0 0 0	1 1 0 1
Recommendations@5	0 1 1 1 1	1 1 1 1 1	0 1 1 1 1	0 1 1 1 1	0 0 0 1 1	0 0 1 1 1	0 0 0 1 1	1 1 1 1 1	0 0 0 0 1	1 1 1 1 1	0 0 0 0 1	0 0 1 1 1	0 0 1 1 1	1 1 1 1 1	1 1 1 1 1	0 1 1 1 1	1 1 1 1 1	1 1 0 1 1	1 0 0 1 1	1 1 0 1 1

P@K	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15	U16	U17	U18	U19	U20
P@1	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
P@2	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
P@3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
P@4	25.00%	50.00%	0.00%	75.00%	50.00%	50.00%	0.00%	25.00%	0.00%	25.00%	0.00%	75.00%	0.00%	100.00%	50.00%	75.00%	0.00%	25.00%	0.00%	75.00%
P@5	80.00%	100.00%	80.00%	80.00%	40.00%	60.00%	40.00%	100.00%	20.00%	100.00%	20.00%	60.00%	60.00%	100.00%	100.00%	80.00%	100.00%	80.00%	60.00%	80.00%
AP@1	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
AP@2	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
AP@3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
AP@4	6.25%	12.50%	0.00%	18.75%	12.50%	12.50%	0.00%	6.25%	0.00%	6.25%	0.00%	18.75%	0.00%	25.00%	12.50%	18.75%	0.00%	6.25%	0.00%	18.75%
AP@5	21.00%	30.00%	16.00%	31.00%	18.00%	22.00%	8.00%	25.00%	4.00%	25.00%	4.00%	27.00%	12.00%	40.00%	30.00%	31.00%	20.00%	21.00%	12.00%	31.00%

MAP@1	0.00%	MAP@2	0.00%	MAP@3	0.00%	MAP@4	8.75%	MAP@5	21.40%

Appendix – H

Hypothesis Test for Mean Average Precision

Hypothesis test – one-way ANOVA for multiple means, Factors (CRD, HARD, Baseline) Alpha = 0.05, and 1=<|TNG|=< 5

Method – One-Way ANOVA for Multiple Means

CRD	HARD	Baseline	FITS	FITS_1	FITS_2	RESI	RESI_1	RESI_2
100.00%	85.00%	0.00%	0.900075	0.884981	0.024525	0.099925	-0.03498	-0.02452
95.00%	91.25%	0.00%	0.900075	0.884981	0.024525	0.049925	0.027519	-0.02452
90.00%	90.28%	0.00%	0.900075	0.884981	0.024525	-7.5E-05	0.017797	-0.02452
84.69%	88.65%	2.81%	0.900075	0.884981	0.024525	-0.0532	0.001478	0.0036
80.35%	87.32%	9.45%	0.900075	0.884981	0.024525	-0.09657	-0.01181	0.069975

Null hypothesis	**All means are equal (the three recommendation models perform similarly)**	***Factors: CRD, HRAD, Baseline*** ***Means: MAP@K***
Alternative hypothesis	**Not all means are equal (the three recommendation models perform differently)**	
Significance level	**$\alpha = 0.05$**	

Equal variances were assumed for the analysis.

Factor Information

Factor	Levels	Values
Factor	3	CRD, HARD, Baseline

Analysis of Variance

Source	P-Value	Explanation
Factor	**0.000**	P value is less than alpha (0.05) which means that we can reject null hypothesis with 95% confidence and the factors (i.e. the three methods, CRD, HARD, Baseline) indeed result in different means (i.e. different MAP@K)

Means

Factor	N	Mean	StDev	95% CI
CRD	5	0.9001	0.0785	(0.8484, 0.9518)
HARD	5	0.8850	0.0247	(0.8333, 0.9367)
Baseline	5	0.0245	0.0410	(-0.0272, 0.0762)

Grouping Information Using the Tukey Method and 95% Confidence (Compares each two methods separately)

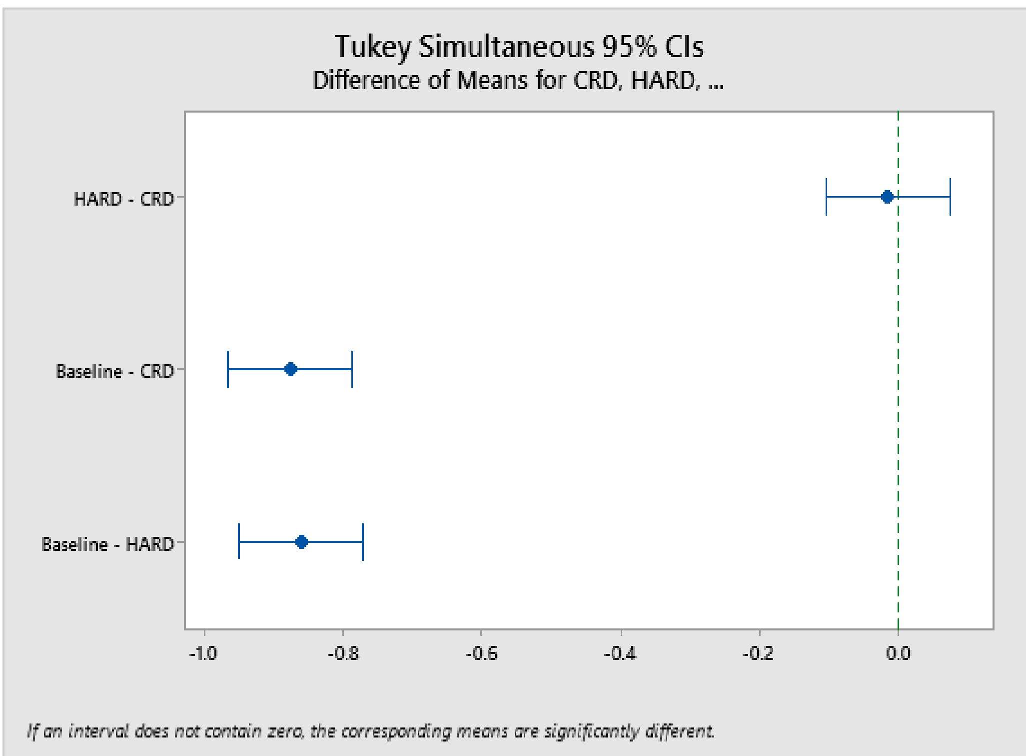
Factor	N	Mean	Grouping	
CRD	5	0.9001	A	
HARD	5	0.8850	A	
Baseline	5	0.0245		B

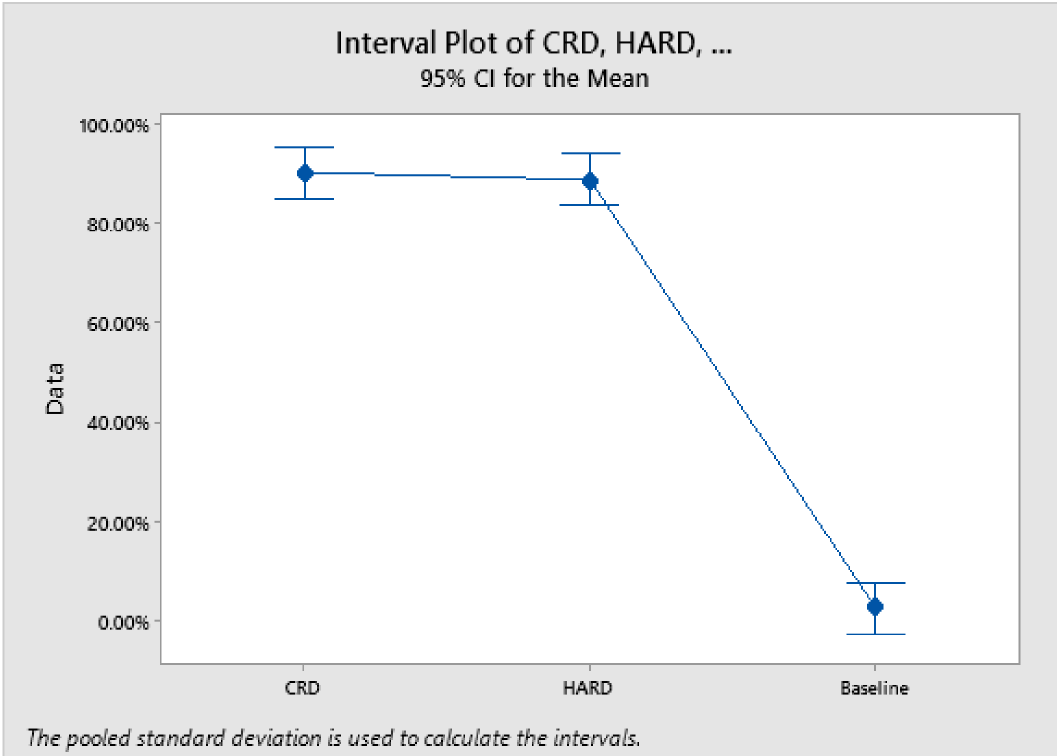
Means that do not share a letter are significantly different.

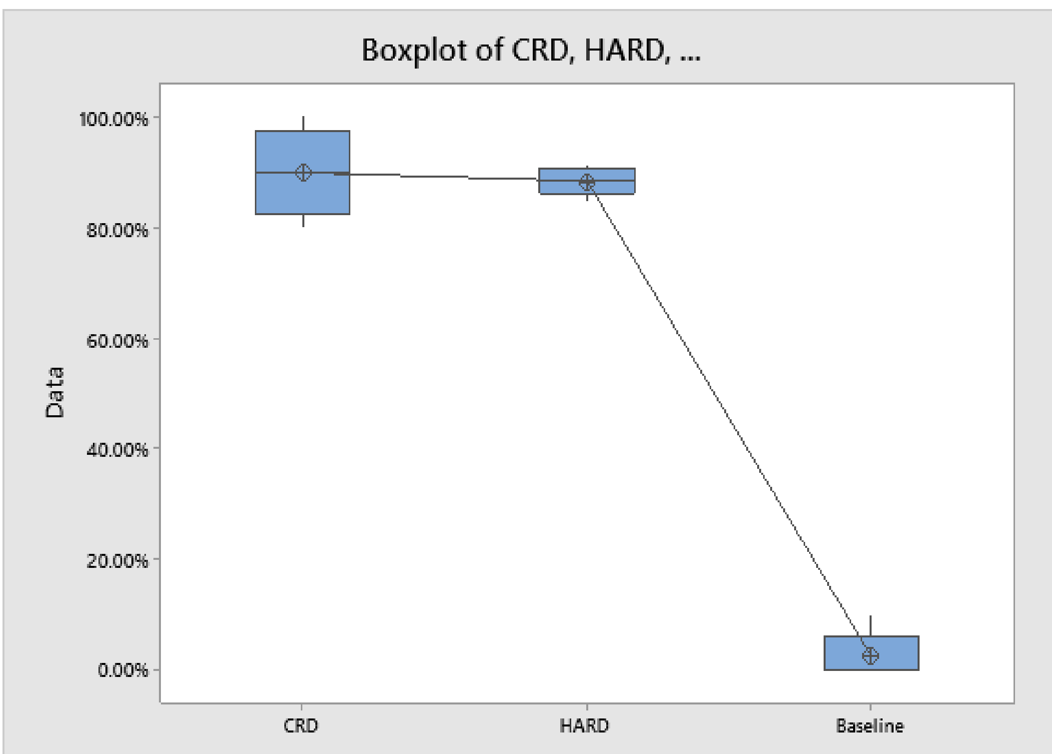
Tukey Simultaneous Tests for Differences of Means

Difference of Levels	Adjusted P-Value	Explanation
HARD – CRD	**0.895**	The difference between CRD & HARD at |TNG|=5 is ***Not*** statistically significant
Baseline – CRD	**0.000**	The difference between CRD & Baseline at |TNG|=5 is statistically significant
Baseline – HARD	**0.000**	The difference between HARD & Base at |TNG|=5 is statistically significant

Individual confidence level = 97.94%




Interval Plot of CRD, HARD, ...
95% CI for the Mean
Data
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
CRD
HARD
Baseline
The pooled standard deviation is used to calculate the intervals.


Boxplot of CRD, HARD, ...
Data
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
CRD
HARD
Baseline

Hypothesis test – one way ANOVA for multiple means, Factors (CRD, HARD, Baseline) Alpha = 0.05, and 5<<|TNG|

Method – One-Way ANOVA for Multiple Means

CRD	HARD	Baseline	FITS	FITS_1	FITS_2	RESI	RESI_1	RESI_2
30.00%	100.00%	0.00%	0.361539	0.966369	0.0603	-0.06154	0.033631	-0.0603
27.50%	100.00%	0.00%	0.361539	0.966369	0.0603	-0.08654	0.033631	-0.0603
34.44%	98.89%	0.00%	0.361539	0.966369	0.0603	-0.01709	0.022519	-0.0603
41.46%	96.98%	8.75%	0.361539	0.966369	0.0603	0.053044	0.003422	0.0272
47.37%	87.32%	21.40%	0.361539	0.966369	0.0603	0.112128	-0.0932	0.1537

Null hypothesis	**All means are equal (i.e. The three recommendation models perform similarly)**	***Factors: CRD, HRAD, Baseline*** ***Means: MAP@K***
Alternative hypothesis	**Not all means are equal (i.e. the three recommendation models perform differently)**	
Significance level	**α = 0.05**	

Equal variances were assumed for the analysis.

Factor Information

Factor	Levels	Values
Factor	3	CRD, HARD, Baseline

Analysis of Variance

Source	P-Value	Explanation
Factor	**0.000**	P value is less than alpha (0.05) which means that we can reject null hypothesis with 95% confidence and the factors (i.e. the three methods, CRD, HARD, Baseline) indeed result in different means (i.e. MAP@K)

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.0783656	96.65%	96.10%	94.77%

Means

Factor	N	Mean	StDev	95% CI
CRD	5	0.3615	0.0821	(0.2852, 0.4379)
HARD	5	0.9664	0.0535	(0.8900, 1.0427)
Baseline	5	0.0603	0.0939	(-0.0161, 0.1367)

Pooled StDev = 0.0783656

Tukey Pairwise Comparisons

Grouping Information Using the Tukey Method and 95% Confidence

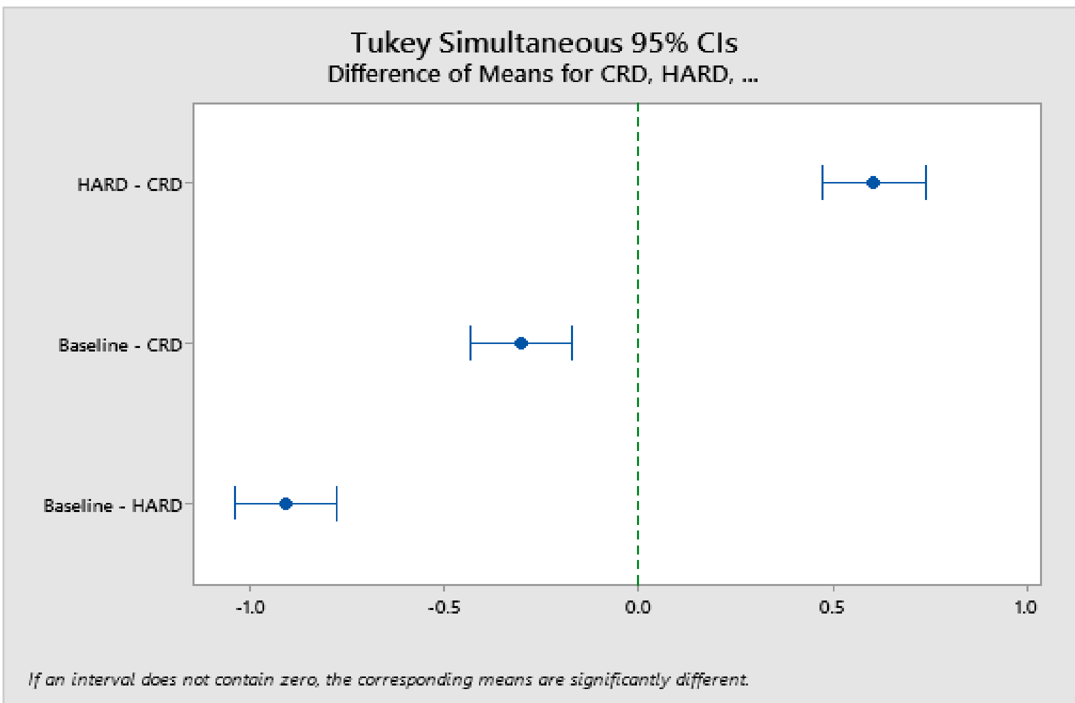
Factor	N	Mean	Grouping		
HARD	5	0.9664	A		
CRD	5	0.3615		B	
Baseline	5	0.0603			C

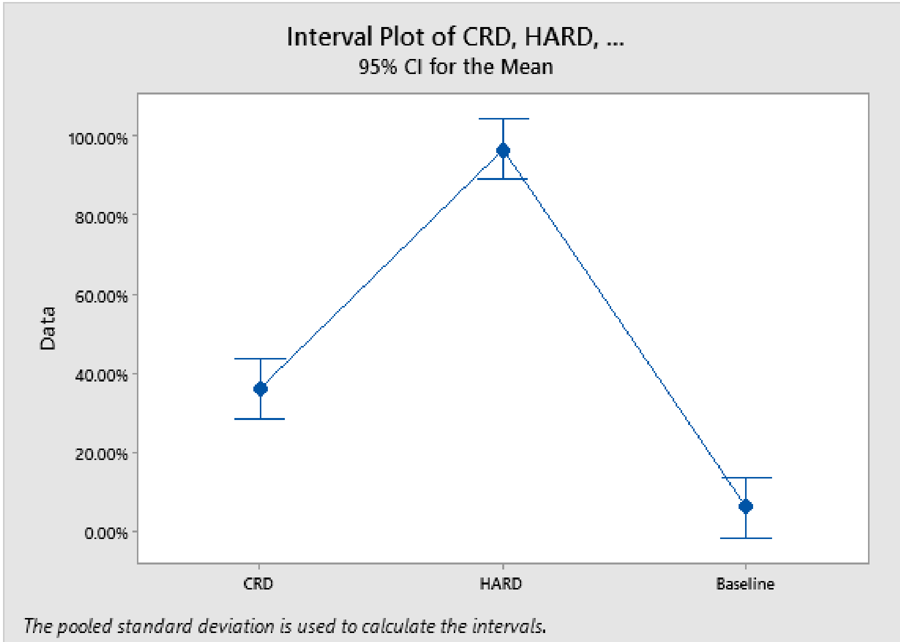
Means that do not share a letter are significantly different.

Tukey Simultaneous Tests for Differences of Means

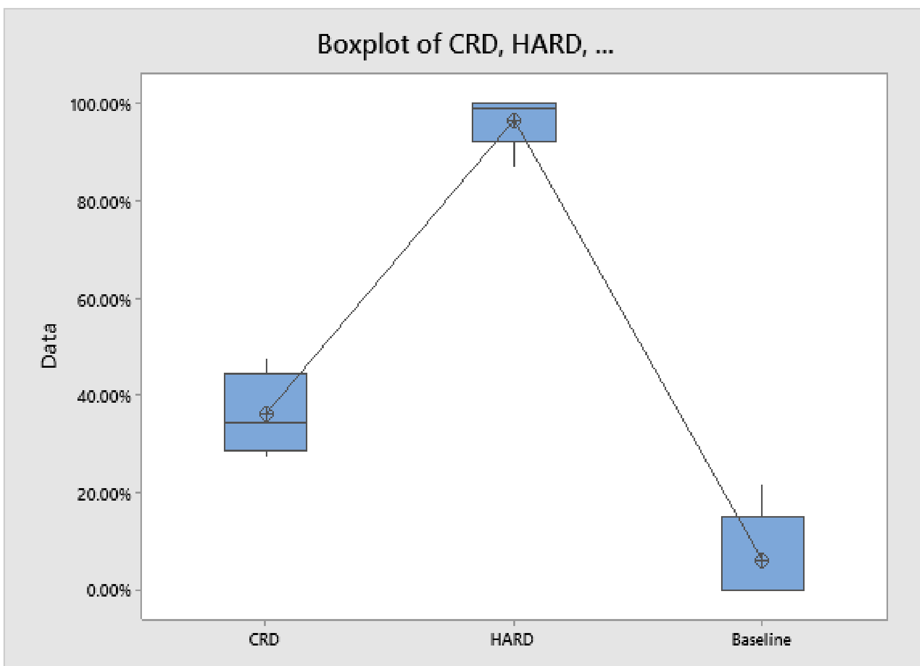
Difference of Levels	Adjusted P-Value	Explanation
HARD – CRD	0.000	The difference between CRD & HARD at 5<<|TNG| is statistically significant
Baseline – CRD	0.000	The difference between CRD & Baseline at 5<<|TNG| is statistically significant
Baseline – HARD	0.000	The difference between HARD & Base at 5<<|TNG| is statistically significant

Individual confidence level = 97.94%





Interval Plot of CRD, HARD, ...
95% CI for the Mean
Data
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
CRD
HARD
Baseline
The pooled standard deviation is used to calculate the intervals.



Boxplot of CRD, HARD, ...
Data
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
CRD
HARD
Baseline

Appendix – I

Hypothesis Test for Conceptual knowledge Assessment

Method – t-Test for small independent samples – sample size <30

μ_1: mean of With Personalization
μ_2: mean of No Personalization
Difference: μ_1 - μ_2

Equal variances are not assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
With Personalization	11	15.00	1.67	0.50
No Personalization	11	10.18	1.78	0.54

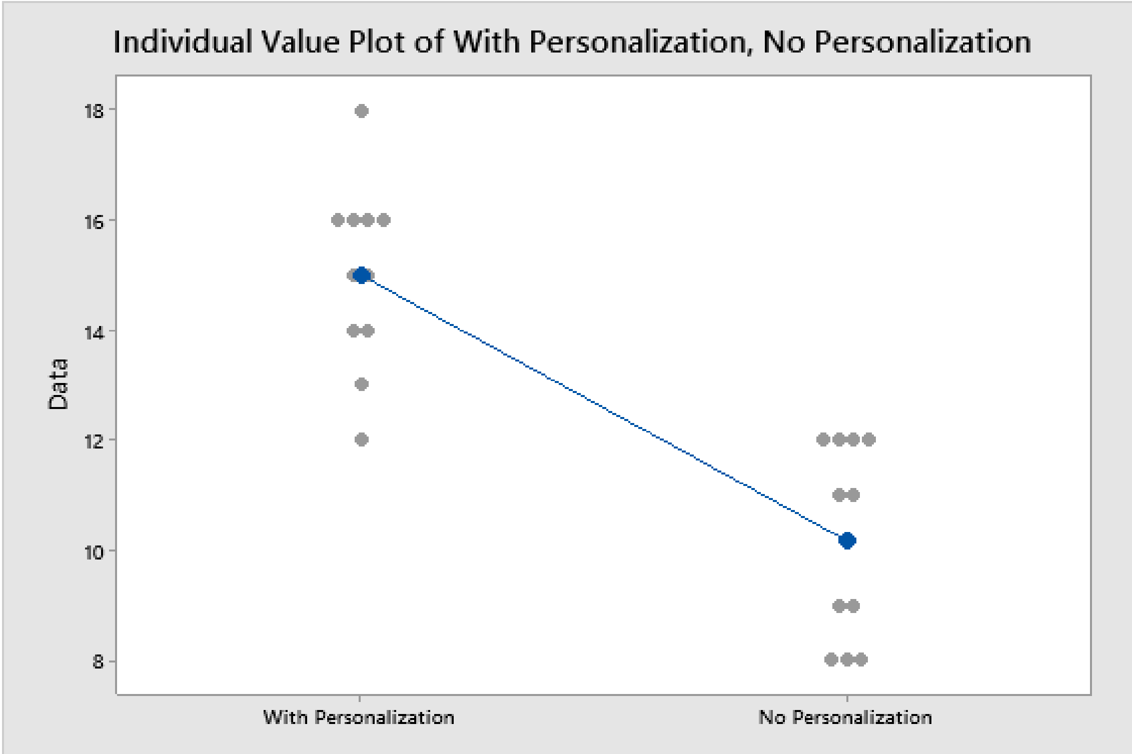
Estimation for Difference

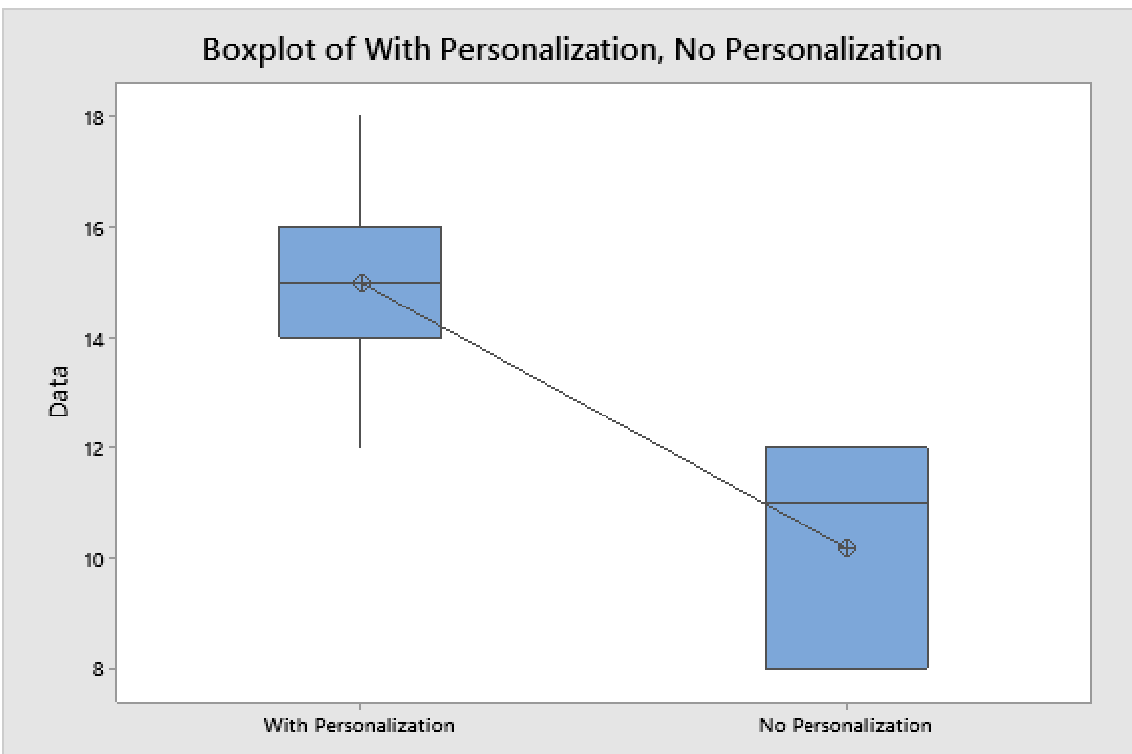
Difference	95% CI for Difference
4.818	(3.277, 6.359)

Test

Null hypothesis	H_0: $\mu_1 - \mu_2 = 0$
Alternative hypothesis	H_1: $\mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
6.54	19	0.000


Individual Value Plot of With Personalization, No Personalization
Data
18
16
14
12
10
8
With Personalization
No Personalization


Boxplot of With Personalization, No Personalization
Data
18
16
14
12
10
8
With Personalization
No Personalization