Personalized Recommender System for Technology Enhanced Learning using Learners' Metacognitive Activities

by

Lydia Odilinye

M.Sc., University of Ibadan, 2010
B.Sc., Covenant University, 2006

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
School of Computing Science
Faculty of Applied Sciences

SIMON FRASER UNIVERSITY
Fall 2019

Approval

Name: **Lydia Odilinye**

Degree: **Doctor of Philosophy (Computer Science)**

Title: **Personalized Recommender System for Technology Enhanced Learning using Learners' Metacognitive Activities**

Examining Committee:

Chair: Nick Sumner
Assistant Professor

Fred Popowich
Senior Supervisor
Professor

Anoop Sarkar
Supervisor
Professor

Marek Hatala
Internal Examiner
Professor
School of Interactive Arts and Technology

Cristina Conati
External Examiner
Professor
Department of Computer Science
University of British Columbia

Date Defended: **August 27, 2019**

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Abstract

Learning, an active cognitive activity, differs from one learner to another. This therefore suggests the need for personalized learning. Recommender systems in this context can be seen as a resourceful tool to provide appropriate learning materials that are tailored to the (personalized) learning needs and goals of the learner; and also to enhance learning. The development of personalized recommender systems typically involves a *learner model* component, which is used to capture and store the personal information, preferences and other characteristics of the learner. While reading, learners engage in number of metacognitive activities e.g. text marking/creating highlights. These metacognitive interactions could serve as useful information for the learner model, to achieve personalization. In addition, the use of a probabilistic topic modeling based document retrieval (Latent Dirichlet Indexing) method makes it possible to provide finer grained multiple but topically related documents to facilitate learning.

The current study investigates the effectiveness of using the highlights (a metacognitive activity) a learner makes while reading, as a preference elicitation method for the learner model. It also investigates the use of the Latent Dirichlet Indexing model to provide relevant recommendation of textual learning materials that enhance the personalized learning experiences of learners in a task-oriented activity. The experimental design allows the comparison of the performance, learner experience, learner interaction, and a number of other subjective analysis measures among two groups conditions; where one group receives recommendations based on the proposed methodology, and the second group receive random recommendations. The recommender system is integrated with nStudy, an online learning platform that provides a number of annotation tools (e.g. highlighting, tags) that support metacognitive activities.

Findings show that the highlights learners create while reading serve as an appropriate input mechanism to guide personalized learning recommendations. Specifically, there was a significant difference in the learners' evaluation of the recommendation quality and accuracy between the two group conditions. The findings revealed that the learners in the experimental had positive perception of the recommendation quality and accuracy, which is also correlated to the user experience, and interaction.

Dedication

To my darling daughter Eliana, whom I had and nurtured during the course of this writing.

Acknowledgments

I cannot express enough thanks to my supervisor Dr. Fred Popowich for his immense support of my research and study, for his patience, understanding and encouragement throughout this program. Also, I deeply appreciate the meaningful learning opportunities he provided, which has taught me more than I could ever give him credit for. I could not have imagined having a better advisor and mentor. I would also like to thank Dr. Annop Sarkar for his support and candid critical insights.

The astounding family I am blessed with has been the strong physical pillar behind this accomplishment. A million thanks to: my loving parents for the unending love, support, prayers and care; my supportive husband for the encouragement, outstanding assistance and love; my siblings for your amazing encouragement and prayers; my daughter for the inspiration and drive to complete the program.

Words would fail me to truly appreciate the amazing research team of the EdPsych lab: Dr. Phil Winne, Dr. John Nesbit, Zahia Marzouk, Jovita Vytasek, Mladen Rakovic, Kenny Tang, Donya Samadi, Daniel Chang, and Michael Lin, thank you very much for the wonderful assistance with the experimental design and conducting the user study.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Teaching and learning are activities that humans have been performing throughout time. These activities have also been influenced by advances in technology, whether it be the printing press, the computer or the Internet. The difference between the previous technologies used (e.g. painting, writing, film) and digital systems of today is that digital systems are interactive. That is, computers and mobile devices can provide materials in a variety of media, and can respond to the learners (Duval et al., 2017). Technology Enhanced Learning (TEL) according to Duval et al., (2017), harnesses the power of interactivity and has the potential to enhance what is learned, how we learn and how we teach. TEL therefore can be described as the application of information and communication technologies to support and enhance all forms of teaching and learning activities (Kirkwood and Price, 2014). The term *Technology Enhanced Learning* is sometimes used as a synonym of the other terms: *e-learning and online learning*. E-learning and online learning are sometimes referred to as learning through technology, however, in this thesis the term TEL refers to educational technologies that enhance (personalized) learning.

With the rapidly increasing amount of learning materials and resources available online, it is becoming more difficult for learners to find appropriate information or learning material to satisfy their needs. Many studies report information overload as one of the main problems that learners encounter in online learning and when searching for the "right" information to satisfy their needs (Manouselis et al., 2011). Searching for relevant information is considered a pivotal activity in teaching and learning (Drachsler, 2009). Therefore, in the context of TEL, a recommendation system (technology) is considered a resourceful software tool that could be used to identify interesting learning materials from a large pool of resources. Also, recommendation systems are able to reduce the burden of information overload by recommending the "right" information at the right time and in the right format (media) of the learner's interest.

In general, recommender systems provide suggestions of objects to a user (Basu et al., 1998). They are widely used in the e-commerce and e-learning domains (Wang et al., 2010). In the e-commerce domain, recommendation is a popular personalization technology where the objects for recommendation are products (such as movies, books) and the users are the customers. In the e-learning domain however, the objects for recommendation are learning materials, and the users are the learners or students. The main objective of the recommender systems, regardless of the domain in which it is applied, is to predict objects that are of interest or relevant to a user. Recommender systems rely heavily on information related to a given domain. The dependency restricts the possibility of applying the recommendation strategy used in one domain to another. For example, personalized recommendation in the e-commerce setting is based on what other users like/ratings, while in the e-learning/TEL setting, it is argued that to achieve personalized recommendations, the recommendation tool should also take into account other features peculiar to each learner, such as: current learning goal, prior knowledge, and other learner characteristics (Drachsler, 2009).

Learning is an active cognitive activity that differs from one learner to another (Shishehchi et al., 2011); each learner has individual needs and particular requirements. Some learners may be highly self-motivated and learn by exploring while other students prefer some specific guidance in a structured way. Therefore, the development of a personalized learning recommender system that would cater to the peculiarity of each learner in TEL is considered important. Personalized learning describes the search for, and the recommendation of, potential learning activities that are the most suitable to the individual learner (or learner group) (Drachsler, 2009). Personalized learning is said to occur when the learning activities has been designed to fit the needs, goals, talents, and interests of the learners (Klasnja-Milicevic, 2011). Learning according to (Drachsler, 2009), is no longer a part of childhood and youth alone, but is becoming a lifelong process. It is also not limited to the context of a regular school or university campus, but also includes informal learning, professional learning at work, personal development, and learning in Massive Open Online Courses (MOOCs). This leads to the notion of self-directed learning. Self-directed learning can be described as a process in which individuals take the initiative, with or without the help of others, in diagnosing their learning needs, formulating learning goals, identifying human and material resources for learning, choosing and implementing appropriate learning strategies, and evaluating learning outcomes (Knowles, 1975).

Self-directed learning can be regarded as a skill, where the individual must know how to set goals, what is needed to achieve those goals, and how to actually attain these goals. Thus, self-directed learning is considered a useful skill for lifelong learning. Developments in educational technology have supported the creation of complex Technology Enhanced Learning Environments (TELE), which provide learners with rich opportunities to use digital technologies to interact with, to configure and to control their learning environments, to communicate with other learners, and to receive quick feedback from all the actors involved

(Persico and Steffens, 2017). TELEs are able to offer learners freedom and choice thus providing them with the opportunity to make strategic decisions about their own learning – this allows learners to practice self-directed learning (Persico and Steffens, 2017). Therefore, to facilitate self-directed learning, a personal learning recommender system would be required to assist learners with determining available learning activities, materials and resources that would match their personal needs, preferences, prior knowledge and current situation to attain their learning goals.

Persico and Steffens (2017) identified three important areas in which TELEs could be beneficial to self-directed learning: metacognition, personalization and assessment. The term metacognition was coined by Flavell (1971) and can be defined as "cognition about cognition", that is, the knowledge concerning one,s own cognitive processes. Metacognition, in TELEs is facilitated by the capability for online learning systems to keep track of the learning dynamics of each learner, and this allows learners to go back to their previous actions and reflect on their learning processes, strategies and progress. The concept of personalization is concerned with the possibility for the learner to control and configure their own environment in order to make the learning process optimal, while the concept of assessment presents a means to provide feedback to the learner based on his/her performance to improve and accelerate learning (Sadler, 1998). In this thesis, we will be investigating the benefits of metacognition and personalization to self-directed learning in TELEs. Specifically, we will apply some metacognitive and personalization strategies to provide personalized learning recommendations in a TELE to support self-directed learning.

With respect to recommendation in the TEL environment, the concept of metacognition can be captured by the strategic selection of recommended learning materials and the strategic processing of the selected materials (Zhou and Xu, 2012). There exists a number of metacognitive activities a learner could engage in while learning, in this thesis however, we are focusing on the metacognitive activities related to reading comprehension, because it constitutes the most common context where learning occurs (Zhou and Xu, 2012). Based on the metacognitive activities associated with reading, recommendations would be made to facilitate comprehension, recall and deeper text processing. The metacognitive reading strategies: organization, note taking, underlining/ highlighting are learning strategies that are focused on reading and are essential to the learning process. This is because they could help a learner find connections within a body of new information; pay attention to, encode new material and provides external storage of information for later studies; determine portions of a body of text that are important to learn and what is trivial respectively (Ormrod, 2012). These metacognitive and learning strategies also support the notion of personalization, because the strategies allow the learners to take control over the learning process. Therefore, it is assumed that TELEs that includes the possibility to take notes, bookmark, highlight portions of the content favors the practice of self-directed learning (Persico and Steffens 2017).

As aforementioned, personalized learning recommender systems design in education, should take into account learners' preferences and characteristics such as learning goals, prior knowledge and a number of other factors. While the development of recommender systems that provide recommendations by taking some aspects of the learners' characteristics is challenging, a number of systems have been developed to this effect. Drachsler et al. (2015) reviewed 82 recommender systems in TEL along their 15 years of existence (2000-2014). The reviewed systems were classified into seven exclusive clusters according to their characteristics (methodology) and analyzed for their contribution to the evolution of recommender systems in the TEL research field. The seven clusters identified include recommender systems designed using the collaborative filtering technique, collaborative filtering with TEL domain particularities, educational constraint as source of information, non-collaborative filtering techniques, contextual information, assessing the impact of recommendation, and recommendation of courses. These user centered design approaches although successful in their development lacked the ability to foster communication and metacognition (Drachsler et al., 2015).

In this thesis, we examine the possibility of the design of a recommender system using the metacognitive activities a learner engages in while reading to achieve personalized learning; where the metacognitive study activities include creating bookmarks, highlighting portions of a text, taking notes, tags, among others. There are several reasons why learners engage in metacognitive activities while reading (e.g highlighting/text marking). From a text processing perspective, the act of deciding what to highlight and otherwise could enable the learner to process the textual information at a deeper and more evaluative level than they would when simply reading it (Craik and Lockhart 1972; Nist and Hogrebe 1987). Another reason is that highlighting could make the marked portion of text more memorable since it stands out from the rest of the unmarked text, which could facilitate recall. Finally, the highlighted text could serve as a guide for later studies (Kornell and Bjork 2007). Based on these reasons we consider the highlights learners' make as a reflective practice that provides insight to the learner's comprehension, and could also reveal the information seeking needs of the learner. By considering the highlights as a information seeking need, the highlights learners make serve as input to the recommender system, which it uses to make recommendations. Typically, the information seeking needs of a learner are expressed using a search query in a search engine. However, search queries do not always return relevant information, due to reasons such as poor queries and natural language ambiguities (Batista, 2007). Therefore, a more cognitive approach to identifying the learners' information needs could be achieved by examining the interactions the learner engages in with the learning materials.

Our recommendation approach also covers two (metacognition and personalization) of the three important areas in which technology enhanced learning environments can be beneficial to self-directed learning identified by Persico and Steffens (2017). The third important area of assessment has been previously published. In which automatic question generation from text was adapted to an online and self-directed learning platform (Odilinye et al., 2015). The automatic questions generated were aligned to the specified pedagogical goals and to a learner's model and provided formative assessment to the learner. More details are included in chapter 4. To identify the "right" learning material for a learner based on the metacognitive activities, we propose to use probabilistic topic modeling approach to analyze and identify the topics/themes of interest to the learner from his/her interactions. Using the information inferred from the learner's interaction, the system is then able to identify and retrieve appropriate learning resources that match the learner's information seeking needs.

1.2 Problem Statement

The popular approaches adopted in recommender systems include collaborative filtering, content-based methods, knowledge-based methods and hybrid techniques (Drachsler, 2009). In terms of methods used for personalized recommendations, the collaborative filtering recommendation technique is the most widely used in the TEL field (Drachsler et al., 2015). Collaborative filtering aggregates ratings or recommendations of objects, determines the commonalities between learners based on their rating, and generates new recommendations based on a comparison amongst them. However, this method suffers from a number of limitations: the cold start problem, rating sparsity and scalability. Because the collaborative filtering method relies heavily on users' ratings, the cold start problem for example exists for new users and new items with no and/or few ratings, the system may not be able to make accurate recommendations. This has led to increased interest in hybrid systems; as hybrid systems leverage the additional information such as content, to solve these limitations.

The adaptation of recommender systems (a popular commercial technology in e-commerce) to the education domain has found a rather limited use of the methodologies and processes for making recommendations used in the e-commerce domain. In contrast to purchasing products online, it is not sufficient to recommend learning materials that other learners like. Learners differ from each other, so also their learning style, preferences, prior knowledge and other characteristics differ. Therefore, in the TEL domain it is imperative that the recommendation strategy takes into account the peculiarities of each learner. While it still remains a challenge to design algorithms and interfaces that take the peculiarities of learners into account, a number of approaches have been developed to characterize each learner by creating a learner model. The learner model stores information of the various characteristics of the learner, and based on the learner model, appropriate recommendations are suggested. Some of the methods that have been used to identify and build a model that represents the learner's needs, learning style and other characteristics include: rule-based methods, collaborative filtering, association rule mining, hybrid methods (combination of two or more methods), among others.

A wealth of research has shown that metacognition plays a crucial role in the promotion of effective learning. In most of the e-learning environment designs, however, meta-cognitive strategies have generally been neglected, and therefore, satisfactory uses of these strategies have rarely been realized (Zhou and Xu, 2012). Most learners are not even aware of what they have been studying (Kurt, 2007). If the learning system could automatically guide and intelligently recommend learning activities or strategies to facilitate student monitoring and control of their learning, it would favor and improve their learning process and performance. Unfortunately, nearly no e-learning systems to date have attempted to do so. In this thesis, we explore the use of metacognitive strategies to provide personalized learning to learners, we propose the development of a recommender system takes into consideration the meta-cognitive activities of a learner.

There are a wide range of metacognitive activities a learner may engage in such as goal setting, self-evaluation, self-monitoring and reading strategies. We are however, focusing on the metacognitive activities related to reading comprehension, because it constitutes the most common context where learning occurs (Zhou and Xu, 2012). Also, the metacognitive interactions of a learner while reading could be considered activities that support personalization because the learner has control over the choices made in using metacognitive tools to aid learning. Learning environments that provide tools to create bookmarks, highlight text, take notes are said to support reading metacognitive activities and personalization (Persico and Steffens 2017).

In this thesis we aim to develop a personalized learning recommender system. The design of personalized learning recommender system entails the creation of a learner model which obtains or infers the learner's characteristics such as prior knowledge and learning style, based on which recommendations are made. Instead of using this approach, we make use of the metacognitive activities that the learner engages in while reading to make recommendations. The metacognitive activities a learner may engage in while reading include: creating bookmarks, highlights, note taking, creating tags, among others. We view these activities as a reflective practice that could provide insight to the learner's comprehension, and could also reveal the information seeking needs of the learner. Therefore, the learners' interactions serve as input to the recommender system, for appropriate recommendations.

To achieve the development of a robust recommender system which also leverages the contextual information, we deploy probabilistic topic models techniques to infer and analyze the data obtained from the metacognitive activities of the learners, to provide appropriate recommendations. By using probabilistic topic models, we are able to obtain finer grained recommendations, this is because it is based on the content of the learning material.

The overall objectives with the hybrid approach are to (a) improve the personalized learning experience of the learner, (b) provide finer grained recommendation of topical related items, which also takes into consideration the meta-cognitive activities of the learner.

1.3 Scope, Limitations and Assumptions of the Study

This study is intended to investigate the usability and suitability of a novel feature – learners' highlights from text, a metacognitive reading activity – as information in the *learner model* and user preference elicitation method to guide recommendation. It is intended as proof of concept to determine if the method works as judged by users of the system. As such, the user study conducted measured the users' subjective experiences and the effects of the system's methodology on the user experience and interaction. Also, although the experimental design of the study included an educational task, the outcome of the task was not evaluated in this study because the focus for this first step is to conduct extensive investigation that the system works.

Two main assumptions were made in this study regarding the system's methodology and evaluation. To evaluate the adequacy and suitability of the learner-generated highlights to guide recommendations, we compared the participants of the study preference to using highlights to typing a search query. Here, we assume that the participants being undergraduate students are (a) somewhat familiar with the use of search queries to retrieve documents from the Internet, (b) they experience some challenges of using search queries (e.g. poor queries and natural language ambiguities). Based on this assumption, the design of the user study did not include a test condition in which the participants' use queries to receive recommendations.

The idea to deploy topic search for document retrieval and recommendations that go beyond traditional approaches to information retrieval (keyword search) using probabilistic topic modeling techniques, is based on a second assumption that the metacognitive reading activities of a learner may include (a) multiple events such as multiple highlights, notes, tags (b) the events (highlights in particular) may span different topics/themes. Based on this assumption, the Latent Dirichlet Indexing model by Wang et al., (2010) provided topic-based recommendations inferred from the metacognitive activities of the learner.

1.4 Thesis Focus and Key Contributions

The research activity of this thesis is focused on the development of a recommender system (technology) to support learning. The recommender system is designed as an integrated tool in a learning environment that supports self-directed learning. The unique features and contributions of the recommender system includes:

- Personalized learning experience
  As earlier stated, learning is a cognitive activity and differs from learner to learner. Therefore, to cater to the peculiar need of each learner, the personalized environment is considered very important. Methods of achieving personalized learning typically involves the creation of a learner model that infers and or stores information of the

learner such as learning style, prior knowledge. In this thesis, we present a novel method – using the learner,s highlights, a meta-cognitive reading activity to achieve personalization.

- Integration with an online learning platform
  Given that the proposed recommender system relies heavily of the metacognitive activities of learners, the recommender system developed is integrated with an online learning platform nStudy (Beaudoin and Winne, 2009), that supports learning, collaboration and research. The nStudy online learning platform also provides tools learners need to record, catalog, analyze, organize, view and synthesize selected information for tasks of any scope and information in any subject area (Beaudoin and Winne, 2009). Some of the tools supported by nStudy include creating bookmarks, adding notes, tags, highlighting portions of a text. nStudy also keeps detailed log data of all the activities learners engage in during a study session. This provides the availability of a technology enhanced learning environment that supports metacognition as well as personalized learning recommendations.

- Accurate finding of recommended learning materials
  The methodology we adopt for the identification and retrieval of the "right" documents for recommendation allows for accurate and appropriate recommended learning materials. Our approach is based on the assumption that a document contains multiple topics. Probabilistic topic modeling is used to infer the topics and themes contained in the metacognitive activities and feedback from the learner. Thus, we are able to identify learning materials that satisfy the learner's information seeking needs, which may span multiple topics.

- Alternative to search queries, preference elicitation method, learner model
  We present an alternative to search queries for recommendations. Typically, the information seeking needs of a learner are expressed using a search query in a search engine. However, search queries do not always return relevant information, due to reasons such as poor queries and natural language ambiguities (Batista, 2007). Therefore, a more cognitive approach to identifying the learners' information needs could be achieved by examining the interactions the learner engages in with the learning materials. Also, this information encapsulates the learner's preference and learning goal which makes up the data in the learner model to achieve personalized recommendations.

- Facilitating Metacognition
  Given that recommendations are generated based on the highlights the learner makes. This encourages the learner to make highlights and makes them pay attention to the highlights created. In this context, the recommender system could be said to facilitate metacognition; because the process of determining whether a text segment should

be highlighted or otherwise has metacognitive components; where the metacognitive part involves (a) monitoring, applying and continuously adjusting standards, and (b) controlling processes leading to mark text or not (Marzouk, 2018).

- Real time recommendations as a service, good user interface design
  Recommendations are made in real time. As the learner interacts with the learning material, the interactions are sent, analyzed in real time, based on which recommendations are generated. The user design interface as well as recommendation strategy are intended to enhance the learning experience of the learners. This is achieved by making the reducing the cognitive load, steps and processes to receiving recommendations.

1.5 Thesis Outline

This thesis is organized as follows.

Chapter 2 reviews the main existing approaches to recommendation in general and specifically the recommendation approaches used in TEL environments. In chapter 3 and 4, we provide details on the techniques adopted in the design of the recommender system: metacognitive activities and probabilistic topic models. Chapter 3 discusses the probabilistic topic model used in the system design – Latent Dirichlet Allocation (LDA), and chapter 4 describes learner behavior in self-directed learning context, within technology enhanced learning environments. In chapter 5, we discuss in detail our methodology for the design and development of the recommender system. Chapter 6 provides the findings, evaluation and results of the experiments and user studies conducted. Chapter 7 concludes the thesis. It gives a summary of the methods used and well as the achieved results, and identifies future work to be done.

Chapter 2

Recommender Systems for TEL: State of the art

The popular approaches in recommender systems in general are collaborative filtering, content-based method, knowledge-based method and hybrid methods (Syed, 2017). These techniques have also been used to develop recommender systems to enhance teaching and personalized learning. In section 2.1, we provide details on each of the recommendation approaches as well as their limitations and in section 2.2, we review literature on recommender systems in education in general and specifically personalized learning recommender systems.

2.1 Recommendation Approaches

2.1.1 Collaborative Filtering

The collaborative filtering (CF) method is the most popular technique used in recommender systems. Collaborative filtering technique is often regarded as a social-based approach which makes use of the collective behavior of a collection of learners to make recommendation. CF based algorithms provide recommendations or predictions based on the opinions of other like-minded users. The opinion of users can be obtained explicitly from the users or by using some implicit measures. Opinions can be explicitly given by the user as a rating score, generally within a certain numerical scale, or can be implicitly derived e.g. from purchase records, analyzing timing logs, mining web hyperlinks, among others (Sarwar, 2001). The relationship between users and items is primarily expressed in terms of ratings provided by the users, the ratings are exploited for recommendations and to predict the rating a user would provide for a specific item. Recommendation in CF entails the analysis of the relationship between users and the interdependencies among products to identify new user-item associations (Koren, 2008).

A CF algorithm makes recommendations in two main steps (a) Similarity computation (b) Prediction computation. Similarity computation estimates the closeness between users and items based on the ratings. The information obtained from computing the similarity between users and items are used for predicting the items to be recommended to users as well as for predicting the rating a user would give an item, in the prediction computation step. CF recommender systems can either be item-based, user-based or stereotypes filtering. The main advantages of collaborative filtering technique are that they use information that is provided bottom-up by user ratings, the method is domain-independent and requires no content analysis,and the quality of the recommendation increases over time (Herlocker et al., 2004).

However, the CF technique suffers two main limitations. The cold start problem is a limitation that occurs due to the fact that CF techniques depend on sufficient user behavior from the past to make recommendations. This problem also occurs when new users or items are added. The new users problem occurs because new users to the system have not made ratings yet in order to get accurate recommendations. New items added to the system also have to be rated by a sufficient number of users before they are recommended (new item problem). The second limitation is the sparsity of the past user actions in a network. Since the CF technique deals with community-driven information, they support popular tastes more strongly than unpopular tastes. The learners with an unusual taste may get less qualitative recommendations, and learners with usual taste are unlikely to get unpopular items of high quality recommended (Drachsler et al., 2009).

2.1.1.1 Item-Based Collaborative Filtering

The item-based approach correlates the set of items the target user has rated and computes how similar they are to target items, and then identifies the most similar items to the set of items the target user had previously rated. Similarities between the items are also computed, and once the most similar items are found, the prediction is then computed by taking a weighted average of the target user's ratings on these similar items (Sarwar 2001). For example, if two items (e.g. $item_A$, $item_C$) have the same user ratings and the items are highly correlated, if a user likes/rates $item_C$, the system recommends $item_A$ to the user because they are highly correlated and have similar user ratings.

2.1.1.2 User-Based Collaborative Filtering

User-based techniques correlate users by mining their ratings and then recommend new items that were preferred by similar users. The user-based approach is primarily used to predict a user's opinion for an item, using the opinion of similar users. Similarity between users is computed by looking at the overlap in opinions of a target user for items and correlates then to the opinions of other users (Drachsler et al., 2009). For example, $user_1$ rates three items ($item_A$, $item_B$, $item_C$), and another user $user_2$ rates two($item_B$, $item_C$) of the three items $user_1$ with similar ratings. Since there is an overlap in the ratings of the users, $item_A$ is recommended to $user_2$.

2.1.1.3 Stereotype/Demographic Filtering

In the stereotype filtering technique, items are recommended to similar users based on their mutual attributes. In comparison to the user-based and item-based techniques, the advantages of stereotype filtering is that it is domain-independent, and does not require a large amount of historical data in order to provide recommendations (Drachsler et al., 2009). Therefore, the method does not suffer from the cold start problem, and is able to recommend similar but yet unknown items and have learners discover preferable items by 'serendipity', and is an accurate way to allocate learners into groups when no behavior data is available (Drachsler et al., 2009). The method however suffers from a limitation: obtaining stereotypical information from the users. Such information has to be collected in dialog with the users and stored in user profiles, and when insufficient information is collected from the users, this could affect recommendations(Drachsler et al., 2009). For example, if there are two users with the same features (e.g. demographic features), and one of the users likes an item, that same item would be recommended to the second user based on the mutual attributes both users share.

2.1.2 Content-Based Method

Content based techniques also referred to as information-based approaches, make use of the information about individual users or items for recommendation. The basic approach of content-based filtering is to compare the content of already consumed items (e.g., a list of news articles) with new items that can potentially be recommended to the user, i.e., to find items that are similar to those already consumed (positively rated) by the user. To achieve this, the content-based method entails creating a profile for each user to characterize their preferences, also the items contain content descriptions (e.g. keywords or categories) (Felfernig et al., 2014). Items are recommended to users by associating the users to items with matching/similar attributes. That is, recommendation of a new item is based on the similarity of the item's features or attributes to the user's characteristics; where the basic idea behind content-based recommender systems is to recommend items that are similar with what user liked before (Ma 2016). The two main types of content-based techniques are case-based reasoning and attribute-based systems.

2.1.2.1 Case-based Reasoning

Case-based reasoning recommender systems recommend items with highest correlation to items the user liked before. This is achieved by computing the similarity between items, and similarity of the items is based on the attributes of the items. Case-based reasoning can also be described as a problem-solving paradigm that faces a new problem by retrieving past cases (experiences) that solve similar problems and reusing them in the new problem situation (Wilson and Leake, 2001). For example, if the attributes (e.g. learning goal) of two items ($item_A$, $item_C$) are very similar, and a user likes $item_A$, the recommender system would also recommend $item_C$, due to the similarities of the item's attributes. The advantages of case-based methods are domain-independence, the system does not require content analysis, and the quality of the recommendation improves over time when the users have rated more items. The disadvantage of this method includes the new user problem, sparsity (previously discussed) and overspecialization – which occurs because only the items that are highly correlated with the user profile or interest can be recommended.

2.1.2.2 Attribute-based Techniques

Attribute-based recommender systems recommend items by matching an item's attribute(s) to the user's profile. This technique only takes the user and item attributes into account for recommendation. The users profile which describe their learning needs are mapped directly to the description of the items available, and items with high correlation are recommended. For example, if a user's profile indicates a learning goal X, and an item ($item_A$) has been characterized to contain the same learning goal X, $item_A$ is therefore recommended to the user because of the match. A major advantage of the attribute-based technique therefore is it does not suffer from the cold start problem because behavior data about the user us not needed a priori.

2.1.3 Knowledge-Based Method

The knowledge based approach does not rely on the user ratings and item descriptions like in the methods we previously discussed, rather it makes use of deep semantic knowledge about the items to make recommendations. This approach aggregates the knowledge about the users and items then applies this knowledge to generate recommendations (Felfernig et al., 2014). The semantic information provided for the items describes the item in finer details that allows the information to be exploited in a different way For example a learning material on *accounting* may include detailed semantic information such as: duration of consumption, complexity of learning material, associated topics, average user ratings, and prerequisites. Thus, knowledge-based recommender systems do not attempt to build long-term generalizations about their users, but rather they prefer to generate a recommendation based on matching between user's needs, preferences and set of items available (Felfernig et al., 2014).

Knowledge-based recommendation systems make use of three types of knowledge – knowledge about the users, knowledge about the items, and knowledge about the matching between the item and user's needs. The knowledge about the user is captured in a *user model*, where the user articulates his/her preferences, knowledge of the items are obtained from the deep semantic knowledge which describe the items, and the knowledge of matching the items to user needs is based a set of rules (constraints) or similarity metrics based on which, depending on the user requirements determines the items to be recommended.

Knowledge-based recommender system can answer the question: how can special items meet the special user's need by using knowledge? This question is addressed by exploiting the deep knowledge about the domain to determine the best solution for user's need (Draschler et al., 2009). Using knowledge-based approaches, the relationship between user's need and recommended items can be explicitly modeled in an underlying knowledge base. Advantages of the knowledge-based approach are that it does not suffer the cold-start problem, the sparsity problem, and also the overspecialization problem since the approach to recommendation does not rely on user rating but rather statistical evidence.

2.1.4 Hybrid Methods

The motivation for hybrid recommendation systems is the opportunity to achieve greater accuracy from the combination of two or more recommendation approaches. There are four main design types of hybrid recommendation which are: weighted, mixed, cascade and switching (Felfernig et al., 2014). Weighted hybrid recommendation is based on the idea of generating recommendations by combining the results of individual recommendation approaches. For example, a hybrid system that includes two individual recommendation approaches – content-based recommender system and collaborative filtering recommender systems, recommendation is obtained by summing the item's scores of the individual approaches, and items with the highest overall score are presented to the user.

Mixed hybrid recommendation is based on the idea that the results of the individual recommender systems are shown in an integrated result. That is, item scores can be on the basis of, for example, the zipper principle; where items with the highest collaborative filtering prediction value gets the highest overall score and items with the highest content-based prediction value is assigned the second best overall score, and so forth (Felfernig et al., 2014). The idea behind the cascade hybrid recommendation is that the individual recommender system used exploit the recommendation of the upstream recommender system as a basis for deriving their own recommendation. For example, a hybrid recommender system that consist of two recommenders Q and U, after the first recommender system Q has computed its results, they are forwarded to the second recommender system U as input, for further analysis.

The switching hybrid recommendation denotes an approach where depending on the situation, a specific recommendation approach is chosen and used. For example, in a hybrid recommendation which makes used of the collaborative filtering and content-based methods for recommendation, when a new user who has no rating data is added to the system, the system makes use of the content based recommender to make recommendations, because the collaborative filtering technique is able to provide recommendations using rating information.

2.2 Related Work on Recommender Systems in TEL

In this section, we present related work on recommender systems in TEL. We limit the papers reviewed and discussed to those related to personalized learning recommender systems. Section 2.2.1 discusses the methods that have been used to develop personalized learning recommender system, and section 2.2.3 presents the various educational tasks that recommender systems have been used to address.

2.2.1 Personalized Learning Recommender systems

The basic idea behind personalized learning recommendation is the need to provide recommendations that meet the specific learning needs and peculiarities of the learner to enhance the learning experience. The development of personalized learning recommender systems of necessity includes a *learner model* which is used to obtain/infer information about the learner. The learner profile is used to capture information about the learner's characteristics such as the learning goal, learning style, prior knowledge, and the information obtained is used to guide recommendation. A number of methods (algorithms) have been exploited to utilize the information about the learner in the learner model to achieve personalized learning recommendations. In this section, we strictly review papers that discuss the methods that have been used to achieve personalization learning, the learner's characteristics that makes up the learner model, and how the learner model is used for the task of recommending learning resources. The papers reviewed are categorized according to the method deployed for personalization.

2.2.1.1 Stereotype Approach (Collaborative Filtering)

As aforementioned, the collaborative filtering technique is the most widely used technique to develop recommender systems in general. However, compared with resource recommendation in e-commerce systems, users in e-learning systems have learning preferences and characteristics that the current e-commerce algorithms do not leverage effectively to generate recommendations Salehi (2013). To address the requirements of resource recommendation in TEL settings, a number of systems have been designed that make use of the collaborative filtering method enhanced with multi-attribute criteria of educational resources.

Tsai et al. (2006) proposed an adaptive personalized ranking mechanism that recommends SCORM-compliant learning objects from repositories in the Internet. The recommendation process uses two algorithms: a preference-based algorithm and a neighbor-interest-based algorithm in determining and ranking the degree of relevance of learning objects to a user's intention, and enables learners find suitable learning objects easily. Given that learning objects may be composed of various media, such as text documents, audio/video clips, pictures or flash, and different learners may prefer different presentations and strategies of a same learning object, the preference-based algorithm is used to bias the recommendations with the learner's preferences. That is, learning objects tending to suit a learner's preference more will get higher priorities when recommended to the learner. This algorithm is used together with the neighbor-interest-based algorithm (used in the collaborative filtering method) to identify learners with similar profiles; where the learner profile consists of the learner's feedback (ratings) on the learning object. The output from both algorithms are aggregated and used to retrieve and rank the learning objects to be presented to the learner.

Salehi (2013) developed a recommender system for learning resources based on the collaborative filtering technique enhanced with the learner's navigation history data. The system begins by obtaining rating information in order to extract the implicit attributes of resources and learners using matrix factorization algorithms. Also the BIDE algorithm is used to discover sequential patterns of resource accessing for improving the recommendation quality. To build the learner's model, the ratings as well as the server usage logs of learners collected periodically are used. The original usage logs are preprocessed, cleaned and used to build a Learner Tree (LT) for each learner. The LT is introduced to take into account explicit multi-attribute of resources, time-variant multi-preference of learner and learners' rating matrix simultaneously. Thereafter, learners are clustered and access patterns of learning resources are mined. The recommendation process makes use of two approaches: implicit and explicit collaborative filtering recommendation and sequential based recommendation, and produces two recommendation sets. The two recommendation sets are combined for the final recommendation that would be presented to the learner.

Gomez-Albarran and Jimenez-Diaz (2009) developed a hybrid recommender system that incorporates collaborative recommendation capabilities in an educational repository. The recommender interface allows the students to retrieve a set of relevant learning resources after posing a query. The query result is an ordered list of documents, where priority is given to documents that: are similar to the target query, adapted to the student's knowledge level (student profile), and are relevant to other students with a similar knowledge level (student ratings). The recommendation methods used are the collaborative filtering and case-based reasoning approach (CBR). A learner model is built using the learner's rating scores, learner profile and learning goals. The system assumes that the student profile and goals are not static but evolve in time, therefore, a collaborative filtering approach that considers the ratings made by similar students is used to store the learner profile and learning goal the learner has when a rating is given – this allows the system to keep track of the learner's evolution over time. From a CBR point of view, the incorporation of this information represents a kind of learning and can used to refine the recommendation process.

Tan (2010) describes a framework for recommendation system based on collaborative filtering in e-learning. The framework proposed consisted of five parts: data collection, data ETL, model generation, strategy configuration, and service supply. The created modules in this system are the recommendation model database, recommendation system database, recommendation management, data/model management. Using the collaborative filtering technique, recommendation is based on the correlation between that learner and other learners who have studied courses from the E-Learning platform. The recommender system analyses the learners' requirements then run corresponding recommendation algorithms to recommend some courses to a learner.

Tiffany (2013) presents a recommendation system for an evolving e-learning based on the system's observation of its learners and the ratings given by the learners. The system supports studying and learning an advanced course on data mining and web mining. The system is web based and consists of two collaborations: collaboration between a system and the user, and collaboration between the system and the open Web. The system infers relevant information about a learner based on their learning characteristic such as the web pages/documents opened, and ratings. The recommendation process makes use of collaborative filtering and data clustering, where the data clustering is performed to obtain the learner model and is used in combination with the user ratings to provide personalized recommendations.

Chatti et al. (2009) implemented a recommender system that leverages the advantages of social tagging. Sixteen tag-based collaborative filtering recommendation algorithms were proposed and used for the recommendation of learning materials in a Personal Learning Environment (PLE) settings. The 16 algorithms entailed user-based (users with similar tagging behavior share same learning interests) and item-based (items that have similar annotations/tags have same topic/content) recommendations. The extra information for the tags and the tagging behavior of learners were used in combination with the learners rating to correlate similar users and similar items, in order to provide recommendations that matches the learner model which comprises of the learner's rating, tags used, and tagging behavior.

In the papers examined, a number of learners' characteristics were explored to provide extra information that were used in combination with the social information (ratings) from the collaborative filtering technique to build the learner model, for the generation of personalized recommendations. Although the systems developed using these methods and model achieved improved performance and accuracy, the bottlenecks and limitations of the collaborative filtering technique – cold start problem, new-item / new-user problems were still reported.

2.2.1.2 Ontology based Recommendations

Ontologies, a semantic Web technology facilitates knowledge sharing, reuse, communication, collaboration and construction of knowledge rich and intensive systems (Ge et al., 2012), and has been adapted in the design of recommender systems for TEL in a number of ways: to model the learner's profile, to structure learning materials, to denote the semantic relationship between learning materials, among others.

In order to recommend appropriate learning resources to different learners, Shen and Shen (2004) developed a personalized recommender system whose recommendation mechanism is based on competency gap analysis and sequencing specifications. The sequencing rules form the basis of the recommendation system, which defines how the learners navigate the learning resource, and how different learning resources are selected for different learners based on the learner's profile. The ontology of the content of a learning resource (e.g. a course material) is first created. This involves splitting the learning material into smaller units (e.g. based on the concepts/topics in the resource), defining the interrelation and dependencies of the concepts. A learner profile is used to capture information about the characteristics of each learner as well as the learning objective competency. Before learning resources are presented to the learner, the learner's competency is updated by performing competency gap analysis which compares the learner's competency and the objective competency, based on the result of this analysis, the appropriate set of sequencing rules is determined on how to present the recommendation and guide how the learner navigates the resources.

Kerkiri and Manitsaris (2009) proposed an ontology framework to design a recommender system which makes use of three ontologies to describe the learners' profile, learning resources and reputation metadata. The learners of the system initially are expected to provide their personal details, preferences and learning style. This information make up the learner's profile and is stored using the ontology format, where each learner is denoted as a vector having k properties describing his profile. The learning resources are also denoted using the ontology format, and a third ontology – reputation ontology is used to collect both implicit and explicit evaluations from the learner's about the resources. Instances of the reputation ontology are used to connect the learner's ontology and the learning resource ontology, and the content of the reputation ontology is similar to the user rating – the learner's opinion about the learning resource. Personalized recommendation is achieved by aligning the characteristics of the learner to the properties of the learning resources and also leveraging data from the reputation ontology, along with the rules of the ontologies. The rules specify how the recommended resources are presented to the learner based on their learning preferences.

The ontology-based personalized recommendation system framework proposed by Ge et al., (2012) included a domain ontology constructed by integrating multi-resource heterogeneous information; a user's interest ontology generated by analyzing the user's demographic characteristics, personal preferences, and navigational history information; and an automatic retrieval specification and expansion method is utilized to categorize the information queried by a user. Personalized recommendations are achieved by matching the results of the domain ontology, user's query requests and interest ontology. Using this methodology, the recommender system is said to be able to suggest appropriate information to the learner who is likely interested in the related topics.

Yuan et al., (2012) proposed a new ontology-based user modeling method for personalized recommendation. The user model makes use of concept nodes to describe the user interests, and can be generated in three steps: first, the web server logs are analyzed to obtain the users' access scores on the leaf nodes of ontology concept hierarchy tree; second, ontology reasoning technology is used as well as access scores on the leaf nodes in the ontology tree in order to get the access scores on the non-leaf nodes, third, a merge of the access score vector on leaf nodes and score vector on non-leaf nodes is performed to build the ontology-based user model. The system also included a domain ontology constructed using the OWL Web Ontology Language. Personalized recommendation is generated automatically by matching the user model ontology to the domain ontology, using similarity algorithms.

The previous work we have examined so far mainly made use of the learner's preferences for personalized recommendations of learning resources, however, Cheng et al. (2018), notes that personalized learning path is one of the most promising personalization solutions for e-learning. An ontology-based learning path recommendation method is proposed by the authors which utilizes the learners' knowledge mastery to generate an ontology-based personalized learning path and appropriate learning resources. Because knowledge mastery changes during the learning period, an update mechanism of the learner ontology is proposed by simulating learners' actual knowledge growth process to keep the model up to date. Three ontologies are used by the system: domain ontology, knowledge mastery ontology and learning order ontology. The personalized learning path generation process is started after a learner completes a learning activity and rates his/her performance. The strategy for recommendation is based on level promoting and level matching. Level promoting occurs when the learners move to a higher level of difficulty when they have achieved sufficient improvements in their current level – achieved by using the information in the knowledge mastery ontology. Level matching is that learners are arranged to study materials whose difficulty levels correspond to their current level, using the learner's preferences together with the learner oder ontology.

The advantage of the ontology and educational standards infrastructure is that it is based on a well-established mechanism that makes the information machine-interpretable and allows syntactic and semantic interoperability among web applications. Using these infrastructures to represent the components of a recommender system (domain, learner model) makes it easier to retrieve resources as well as matching or correlating users to items and similar users, this is because the properties of a component are described through metadata and not on its actual content.

2.2.1.3 Concept Maps

Concept mapping (Novak and Gowin 1984) has been widely used to externalize knowledge, conduct knowledge construction (Leake et al. 2003), share knowledge, and compare knowledge to advance human learning and understanding (Leake et al. 2004). In concept mapping, either the subject constructs a two dimensional, visually-based representation of concepts and their relationships (Leake et al. 2003) or it can be created automatically. In the field of recommender systems in TEL, concept maps have been used to capture the learner model, as well as the domain (learning resources).

A recommender system, Customized Learning Service for Concept Knowledge (CLICK) designed by Okoye et al. (2012) was designed to provide digital library resources recommendations based on users concept knowledge demonstrated through automated evaluation and approximation of their knowledge, from essay writing. The CLICK system is made up of four components: the domain concept map generator module, student concept map generator module, misconception identification module and instructional plan generation module. The domain concept map is generated from a library resource, while the student concept map is generated from an essay the student writes, from which the system also diagnoses the learner's incorrect, incomplete and fragmented conceptual knowledge by comparing the student's concept map with a reference domain concept map. The recommendation module takes as input, the list of the student's misconceptions, the domain concept map and student's concept map. It then generates a concept graph for each of the concept maps, and the maximum sub-graph as the similarity measure to generate recommendation of learning resources that addresses the learner's misconceptions.

Kardan et al. (2006) developed a hybrid recommender system based on concept maps and collaborative tagging. The tags and concept maps are generated manually by the learner and are used to capture the learner's knowledge. The study resources are stored in a repository from which learners can find and read materials of their choice. After reading, the learner may tag the reading material with one or more keywords that demonstrate his knowledge of the content and/or may create a concept map; where each node of the concept map contains a tag that the learner used in the previous stage to describe the resource. Similarity measures are used to compute the similarities between the tags and concept maps. The matching process is performed to determine what should be recommended to the learner. For example, if tags are used in the learner's concept maps, but are not used to describe the relationship with other maps, then the system recommends a concept map in which these tags are related. Otherwise, a concept map composed of more tags or a different concept map is recommended.

2.2.1.4 Data Mining, Web Mining

The use of web mining techniques to build an agent that could recommend online learning resources, activities or shortcuts based on learners' Web navigation history can be used to improve learning as well as provide personalized recommendations (Zaiane, 2002). This can be achieved by using Web usage mining, which performs mining on web data, particularly data stored in logs managed by the web servers. The web log provides raw traces of the learners' navigation and activities on the website (Zaiane, 2002). A number of recommender systems have leveraged this information to create learner models, and generate recommendations.

Khribi et al. (2009) proposed the structure for an automatic personalization approach aiming to provide online automatic recommendations for active learners without requiring their explicit feedback. Recommended learning resources are computed based on the current learner's recent navigation history, as well as exploiting similarities and differences among learners' preferences and educational content. The framework consists of two modules: an off-line module which pre-processes the data to build learner and content models, and an online module which uses the learner and content models to generate a recommendation list. The recommendation process begins with offline mining of the learners' models based on Web usage mining techniques, which clusters the data on the learners' web sessions. Each cluster contains similar sessions and shows similar interests of different learners and can be viewed as one learner's model. Mining of association rules from the clustered sessions is then performed and the learner's preferences and interests are extracted. The generation of a recommendation list process combines both content based approach and collaborative filtering methods.

A Web mining tool and recommender engine developed by Romero et al. (2009) was integrated into an open source general-purpose adaptive hypermedia system called AHA! The process of Web personalization is based on Web usage mining and consists of three phases: data preparation, pattern discovery and recommendation. The first two phases are performed offline and the last phase is performed online. To build the learner model, the system uses the student's information stored in Web log files. It applies sequential mining algorithm over all the learner's navigation sessions to discover the most frequent navigational pattern and the data is clustered; where each cluster corresponds to the learner model. In the recommendation phase, the recommender engine is activated each time that a learner visits a Web page, the engine classifies the student into one of the clusters. Finally, recommendation is generated according to the rules in the cluster. So, only the rules of the corresponding cluster are used to match the current Web page (concept) in order to obtain the current list of recommended links.

Hsu (2010) developed an online personalized recommendation system for English language learning. Recommendation is based on the combination of content based filtering, collaborative filtering and data mining techniques. The system recommends the best and suitable English course to a learner with the different interests. In this system, all learners classify to some groups with the similar study behavior by using clustering algorithm. Each lesson in each group has one initial score. The content based approach is responsible for setting these scores to respective lessons. Association rule mining technique analyzes the association of lessons in each segment and adjusts the score of each lesson of each student.

2.2.1.5 Fuzzy Logic Theory and Item Response Theory

Fuzzy set theory and logic provide a way to quantify the non-stochastic uncertainty that is induced from subjectivity, vagueness and imprecision (Zadeh, 1994). Compared with traditional statistical methods, the application of fuzzy logic theory in recommender systems has a number of benefits (Hsu, 1998): the membership function in fuzzy theory is designed to handle vagueness and imprecision, thus making it more reliable; the membership function can be continuous, which is more accurate in representing the attributes of items and user preferences; and the fuzzy mathematical method is easy to perform once the membership functions of attribute have been defined.

Lu (2011) proposed a personalized learning recommender system that recommends suitable learning materials to all learners with different learning style, learning need and knowledge background of learner. The system framework consists of four main components such as getting student information, identifying the student requirements, learning material matching analysis and generating recommendations. This framework is connected to the user interface and applies the student database and learning material database. Furthermore, supported by a student requirement model and fuzzy matching rule to discover associations between learner requirements and learning material, two related technologies were developed under the framework. The first relates to learner's needs by using the multi attribute evaluation method and the second one is a fuzzy matching method to find the appropriate learning materials according to a learner's need. The multi-background learners can use this system as online learning.

Chen and Duh (2008) proposed a personalized e-learning system based on Item Response Theory. A learner's information and the course's information are stored in separate databases. The architecture of the system designed included a course recommendation agent which played the role of offering suitable course materials to a learner from the course database. This recommendation performs according to the learner ability in a selected course unit. A feedback agent, another component of the system's architecture, was used to obtain information about the learner's ability according to the learner database and course database. The authors also extended their work and developed an intelligent tutoring system that includes the personalized recommender system. This system also works based on fuzzy Item Response theory, that includes an off-line courseware modeling process, four intelligent agents such as learning interface agent, feedback agent, courseware recommendation agent, and courseware management agent and four databases include the user account database, user profile database, courseware database and teacher account database. One of the main components in this system is a recommendation agent. Recommendation agent evaluates the learner's ability due to the learner's feedback by using the proposed fuzzy item response theory; also it recommends the appropriate courseware to a learner. Learner information in this system covers learner ability, learner response and learning paths.

Hseih et al. (2013) developed a personalized remedial learning system to assist learners in remedial learning. The system has four major components: the learner testing component, inference module, learning style analysis and learning path and remedial materials recommender. The proposed system adopted fuzzy logic theory to determine appropriate learning path as well as learning resources based on the learners' misconceptions found in a preceding quiz. The learner model consists of the learning style and prior knowledge of the learner. Each learner using the system is given a questionnaire to analyze their learning style, thereafter the learner is given a quiz which is used to obtain the learners prior knowledge and to identify their misconceptions, and stores them in the learner portfolio repository. The inference module uses fuzzy logic to infer appropriate learning paths for each of the learners' misconceptions. Concepts in each course are constructed into a learning path, the recommender system selects the most suitable remedial materials for a learner based on the learner model to facilitate more efficient remedial learning. Based on the generated learning path, the Learning Style Analysis then retrieves the related remedial materials that satisfy the learners' preferences and stores them in the remedial materials repository.

Wang (2011), deployed a fuzzy knowledge extraction model to extract personalized recommendation knowledge by discovering effective learning paths from past learning experiences through an ant colony optimization model. The proposed approach imitates the natural ants, which share the paths they have found leading to food by scattering pheromone along the paths. Learners play the role of ants, scattering trail marks in a proper way according to their learning performances along the learning paths characterized by specific learning contexts. These trail marks can then be used to discover effective learning paths for learners with specific learning styles and competency. The learner model is described in terms of competency and learning style; symbolized using a fuzzy set theory. The recommendation task is modeled as finding best learning paths for different types of learners in different knowledge subspaces, consisting of nodes and edges, dictating the learning paths of learners with particular learning contexts, indexed by competency levels and learning characteristics. Each node in a knowledge subspace represents a learning resource. When the learner accesses a resource item he moves to the corresponding knowledge node. A trail mark proportional to the learner's membership is obtained/updated, referred to as a local update. A global update is performed when the competency of the learner changes and is affirmed through feedbacks. As the learner moves through the knowledge subspaces, recommendations are provided by considering the membership degree to which the learner belongs to as well as the learning path that she/ he has currently passed.

2.2.1.6 Social Navigation and Metadata

The idea of social navigation in virtual spaces is borrowed from the long established principle that people tend to follow other people's trails in space. Social navigation in general, allows users to help each other and generally serve as a navigational aid (Tancheva and Koennecke, 2008). In the context of TEL recommender systems, social navigation techniques could be used to provide valuable help in guiding users to the most useful information. Social navigation entails processing and analyzing traces, log data of past user behavior and using the assembled information to guide future users (Brusilovsky et al., 2010). This information has also been used to create a learner model and leveraged to provide personalized recommendations.

Brusilovsky et al. (2010) designed a social navigation infrastructure for an educational digital library, *Ensemble* – a computing portal to provide access to learning materials and resources for education in the Science, Technology, Engineering and Mathematics (STEM) disciplines at all age and education levels. The infrastructure keeps track of the various action of the portal users such as resource browsing, rating, commenting, tagging, and fragment extraction and composition. These information is accumulated in the learner model server, and makes it available for social navigation services. The collected data is then process in two levels: portal and group levels; where the portal level integrates traces of all portal users, while the group level integrates actions of a specific community or group of users. Ensemble supports a number of feedback tools for learners such as comments, tagging, ratings, fragment selection. These information is collected and stored social navigation component of the system and is used in two ways: to indicate user interest in various items, and to allow the system to identify items that are similar from the user behavior perspective. This data is used to identify similar resources and efficient learning paths.

Shelton et al. (2010) developed *Folksemantic*, a platform that integrates OpenCourseWare search, Open Educational Resource recommendations, and social network functionality into a single open source project. The system keeps track of the users' attention metadata which includes RSS feeds, clicks, shares, comments and amount of time spent on a web page. The information is collected and stored in the learner model, and is used as the basis for providing personalized recommendations. Folksemantic is a content-based recommendation system that recommends related resources based on the semantic relatedness of their metadata. The system begins the recommendation process by identifying the top 20 most semantically related items to the learner's metadata. Recommendation scores are computed and assigned to items on the recommendation list. The higher the recommendation score, the higher the item appears on the final recommendation list. Also, the recommendation score is used to order/rank the items on the list.

Adaptive Learning Framework (ALEF) designed by Simko et al. (2010) was intended to deliver educational resources to learners. Bielikova et al. (2014) enhanced the functionality of the system by making the system deliver tailored learning experience via personalized recommendation and enabling learners to collaborate and actively participate in learning using interactive educational components. The ALEF architecture consists of three components: domain model, user model and the framework components. In the domain model, the learning resources are described using domain relevant terms, while the user model is used to store the interaction of users with domain elements. The user model contains information such as: which learning the learner visited, how much time was spent reading it, exercises completed, among others. The framework components comprises reusable and extendable tools – annotation tool, feedback tool and widgets – which are used to collect data for the user model. Recommendation of learning resources is based on a hybrid approach using switching and mixed hybridization methods. The learner characteristics that are related to domain concepts (such as concept knowledge) represented by relevant domain terms are used to select appropriate learning resources.

A number of methods to develop personalized learning recommender systems which includes a learner model have been reviewed and discussed in this section. As earlier mentioned, the scope of the papers reviewed were streamlined to those on personalized recommendation of learning resources. From the review of literature, it can be observed that there are no recommender systems as at yet that leverage the metacognitive activities the learner engages in while actively learning to make recommendations. Therefore, we propose to use this information to build a hybrid personalized learning recommender system that also deploys topic modeling approaches to identify appropriate learning resources.

2.2.2 Recommendation Tasks in TEL

In this section, we provide an overview of the different tasks that recommender systems have been used to accomplish in the TEL domain. While the scope of this thesis is focused on the recommendation of personalized learning resources, the methodology behind the recommender system we developed could be adapted to other recommendation tasks.

2.2.2.1 Find Learning Resources

The identification of learning resources is a task every learner is faced with during a learning session. Due to the ever increasing amount of resources available online, it is becoming a challenge for learners to find and make a decision on appropriate learning resources for their learning needs. Therefore, one of the tasks recommender systems in the TEL domain perform is to guide learners to appropriate learning materials. This entails methods that automatically infer the learners' learning needs and intelligently generate and recommend learning materials that would improve the learning experience.

2.2.2.2 Recommend Learning Activities

To encourage active learning, recommender systems in the TEL domain have been used to dynamically generate and suggest suitable learning activities to learners. The Internet, wireless technologies, as well as mobile devices can be used to motivate learners in different contexts and active ways. For example, making it possible to perform learning interactions with online educational resources through hand-held devices, which could also provide suggestions on the various learning activities that can be engaged in (Martin and Carro, 2009).

Depending on different criteria (e.g. learner characteristics, context), in addition to suggesting learning activities, the system could also suggest appropriate workspaces to support the corresponding learning activities, which can also be dynamically generated. These activities could be done individually or collaboratively as a group.

2.2.2.3 Find Learning Sequence

Learning sequence also known as curriculum sequencing (CS) is an important concern in TEL. It refers to the ordering of recommended learning materials, appropriately sequenced to match a particular learning process (Sentance and Csizmadia 2017). It replaces the rigid, general and 'one size fits all' course structure set by experts with a more flexible and personalized learning sequence. To obtain appropriate learning sequence for a learner, the learner's behavior, knowledge levels, learning styles, learning capabilities as well as curriculum related prerequisites constraints are taken into consideration (Al-Muhaideb and Menai 2011). Various techniques such as statistical, evolutionary computation have been utilized to find optimal learning path sequence that satisfies the pedagogical structure as well as learner specific needs.

2.2.2.4 Find Learning Path

The recommendation task of finding an appropriate learning path for a learner is similar to the learning sequencing task. The identification of learning materials arranged in an appropriate order with a starting and ending point, rather than a sequence of unordered learning materials is referred to as a learning path. The recommended sequence is matched to the learner preferences for enhancing their learning capabilities (Dwivedi et al., 2017). Also, the length of recommended sequence is typically not fixed but varies for each learner because learners differ in their preferences, knowledge levels, learning styles, emotions, etc.

2.2.2.5 Find Peer Learners

The infrastructure peer-to-peer (P2P) networks, promote users' participation through blogs, wikis and folksonomies, which makes those who access the web a potential consumer and producer of personal knowledge and experiences. Web 2.0 and P2P networks make personalization more effective and efficient since a learner can access large amounts of learning materials from different peers, for instance by contacting peers with similar profiles. The concept of *trust* is used together with the task of finding peer learners; where using trust relationships among peers could lead to the identification of a learning path (Carchiolo et al., 2010). In addition to trust, the process of peer selection could be performed using two methods: similarity and expertise. Similarity is used to model the real world behavior of finding peers having similar profiles, while expertise allows the identification of peers with expertise in a given context (Carchiolo et al., 2010).

2.2.2.6 Predict Learning Performance

The prediction of learning performance is the problem of predicting the learner's ability (usually estimated by a score metric) in solving tasks when interacting with a learning system. The prediction of learning performance identifies how the user learns and adapts to new problems. In the recommender system context, predicting learning performance has been considered as a rating prediction problem; where the learner, task, and performance information could be modeled as the user, item, and rating, respectively (Thai-Nghe et al., 2012).

2.3 Summary

In this chapter, we examined the four widely used recommendation approaches: collaborative filtering, content-based methods, knowledge based methods, and hybrid methods. These methods have been used to achieve the task of recommendation in various context, domains, and fields such as, product, movie recommendations. However, in the domain of education, where the focus has been to provide personalized recommendations, these approaches have not been suitable for the task. A learner model which captures and stores information about the learner is an important component of a recommender system that would be able to provide personalized learner-specific recommendations.

A review of the methods that have been used to achieve personalized learning recommendations in the TEL domain revealed six techniques: stereotype methods, ontologies, concept maps, data mining/web mining, fuzzy logic theory, and social navigation data. From the review of literature, it can be observed that recommender systems that leverage the metacognitive activities the learner engages in while actively learning to guide recommendations have not been done yet. Therefore, we propose to the development of hybrid personalized learning recommender system that uses the data obtained from the learners' metacognitive activities to build a learner model and provide personalized recommendations.

In the next chapter, we discuss the topic modeling method proposed and used (for document retrieval) in combination with the learners' metacognitive activities to provide recommendations.

Chapter 3

Probabilistic Topic Modeling

Topic models in machine learning and natural language processing (NLP) describe the set of algorithms for discovering the hidden thematic structure of large collections of documents. With the vast amount of digital information available today, topic models are useful in organizing, analyzing and summarizing large collections of documents according to the underlying discovered thematic structure. Topic models can be adapted to different kinds of data. For example, topic models have been used to find patterns in genetic data, images, and social networks (Blei, 2012; Liu et al., 2016; Nolaso and Olieira, 2018).

Probabilistic topic modeling involves statistical methods that are used to analyze and annotate large collections of documents based on the topical structure of the collection. Topic models have also been used for document representation which is a crucial part of information retrieval, and has been proven to be a promising method for information retrieval (Wei and Croft, 2006). In this chapter, we discuss a popular approach to topic modeling that has been widely used to model and extract topics from large document collections – Latent Dirichlet Allocation (discussed in section 3.1), we propose an LDA-based model for document representation as well as retrieval in the recommender system developed (discussed in section 3.2).

3.1 Latent Dirichlet Allocation

Topic models are probabilistic models for uncovering the underlying semantic structure of a document collection based on a hierarchical Bayesian analysis of the original text (Blei and Lafferty, 2009). Latent Dirichlet Allocation (LDA) is a topic modeling approach aimed at automatically discovering the topics from a collection of documents. Similar to the pLSI model, LDA is based on the intuition that documents exhibit multiple topics (it is sometimes referred to as the Bayesian version of the pLSI model). Given a collection of documents, LDA assumes that there are K (latent) topics, and each of the words in the documents are drawn from the K topics, with different proportions (mixture components). A topic here is defined to be a distribution over a fixed vocabulary of terms. LDA is a generative probabilistic modeling approach, in which the data is treated as arising from a generative process that includes hidden variables.

Hidden variable models are structured distributions in which observed data interact with hidden random variables. The model defines a joint probability distribution over both the observed and hidden random variables. The hidden variable is inferred using probabilistic posterior inference. The observed data are the words in each of the documents in the collection and the hidden variables are the topics (thematic structure of the corpus and the proportion of the topics each document exhibits).

Similar to the bag-of-words assumption that the order of words in a document can be neglected, the LDA model assumes that the orderings of the documents in a collection can be neglected. That is, the documents are exchangeable. To take into account exchangeable representations for documents and words, LDA considers mixture models that would be used to capture the exchangebility of both words and documents (Blei et al., 2003). With the exchangeability assumption, LDA captures significant intra-document statistical structures using the mixing distribution.

3.1.1 Statistical Assumption

As described earlier, the goal of topic modeling is to automatically discover the topics (K *topics*) from a collection of documents. The documents are observed, while the topic structure (characterized as the topics, per-document topic distributions, and the per-document per-word topic assignments) is the hidden structure. To achieve the task of topic identification and extraction from large document collections, the LDA model makes two basic assumptions. (1) A document can be characterized as a mixture of multiple topics in different proportions, and (2) the documents in a collection are exchangeable. That is, the probability of the document collection is invariant to permutation.

Based on these assumptions, together with the observed variables (the words in the documents), the computational task for the LDA model is to use the observed documents to infer the hidden topic structure.

3.1.2 LDA: Algorithm

LDA, a parametric Bayesian approach to topic modeling, can be modeled as a generative process – which determines the hidden structure (topics) responsible for generating the observed variables (words) in the collection. This way, LDA reveals the interaction between the observed words in the documents and the hidden topic structure. To mathematically illustrate the LDA model, let d be a document defined as a sequence of N words, $d = (w_1, w_2, \ldots, w_N)$, where w_N denotes the nth word in the sequence. A corpus C is defined as a collection of M documents, $C = \{d_1, d_2, \ldots, d_M\}$. LDA infers hidden k topics from a corpus by inferring the topic mixture $\theta_d = \{\theta_1, \theta_2, \ldots, \theta_k\}$ at the document level, and at the word level, a set of N topic-words is identified for each topic defined as $z = (z_1, z_2, \ldots, z_N)$.

A vocabulary index set $\{1, 2, \ldots, V\}$ is maintained to indicate whether a word is used or not, that is, $w^j = 1$ if the jth word of the vocabulary list is used, $w^j = 0$ otherwise.

The LDA model as a generative process for each document in the collection is described as follows;

1. For each topic,
   (a) Draw a distribution over topics $\beta_{k=1\ldots K} \sim Dirichlet_V(\eta)$
2. For each document,
   (a) Draw a vector of topic proportions $\theta_{d=1\ldots M} \sim Dirichlet_K(\alpha)$
   (b) For each word,
      i. Draw a topic assignment $Z_{d,n} \sim \text{Multinomial}(\theta_d)$
      ii. Draw a word $W_{d,n} \sim \text{Multinomial}(\beta_k)$

LDA contains two Dirichlet random variables: the topic proportions θ are distributions over topic indices $\{1, ..., k\}$; and the topics β are distributions over the vocabulary. $\theta_d \sim \text{Dir}(\alpha)$ is a K-dimensional vector, the collection of all prior weights of topic k in a document. $\beta_k \sim \text{Dir}(\eta)$ is a V-dimensional vector, the collection of all prior weights of a word w in a topic. $Z_{d,n} \sim \text{Multinomial}(\theta_d)$, and $W_{d,n} \sim \text{Multinomial}(\beta_k)$. $\theta \in \mathbb{R}^K$, $\alpha \in \mathbb{R}^K$ and the dimension K are assumed to be known. The conditional probability of the jth word in the vocabulary list, given that the kth topic is selected, is denoted by $\beta_{kj} = p(w^j = 1|z^k = 1)$. Its maximum likelihood estimator can be obtained from a posterior probability distribution. The matrix of the conditional probabilities is denoted by $\beta = [\beta_{kj}] \in \mathbb{R}^{K\times N}$. The joint prior probability distribution $p(\theta, z, d|\alpha, \beta)$ is expressed as:

$$p(\theta, z, d|\alpha, \beta) = p(\theta|\alpha) \prod_{n=1}^{N} p(z_n|\theta)p(w_n|z_n, \beta) \tag{3.1}$$

where $p(z_n|\theta) = \theta_k$ for the unique k that indicates whether the kth topic is used to select the nth word in the document. The marginal probability of the word appearance can be derived by integrating over θ and summing over z on the prior, represented as:

$$\begin{aligned} p(d|\alpha, \beta) &= \int \sum_{\{z_n\}} p(\theta, z, d|\alpha, \beta) d\theta \\ &= \int p(\theta|\alpha) \sum_{\{z_n\}} \prod_{n=1}^{N} p(z_n|\theta)p(w_n|z_n, \beta) d\theta \\ &= \int \prod_{n=1}^{N} \sum_{z_n} p(z_n|\theta)p(w_n|z_n, \beta) d\theta \end{aligned} \tag{3.2}$$

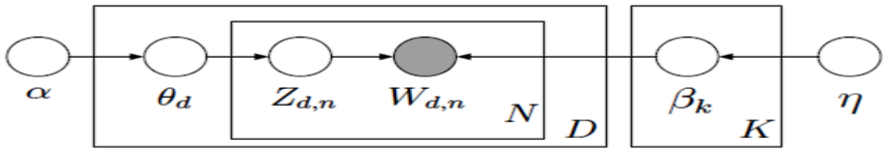


Figure 3.1: LDA as a graphical model (Blei et al., 2003)

The corpus probability is given by:

$$p(C|\alpha,\beta) = \prod_{d=1}^{M} \int p(\theta_d|\alpha) \prod_{i=1}^{N_d} \sum_{z_{dn}} p(z_{dn}|\theta_d) p(w_{dn}|z_{dn},\beta) d\theta_d \tag{3.3}$$

where θ_d's are variables in the document level and w_{dn} and z_{dn} are variables in the word level. Using equation (3.1) and (3.2), the posterior distribution of the hidden variables θ and z are given by:

$$p(\theta, z|d,\alpha,\beta) = \frac{p(\theta, z, d|\alpha,\beta)}{p(d|\alpha,\beta)} \tag{3.4}$$

The hidden topical structure of a collection is represented in the hidden random variables: the topics $\beta_{1:K}$, the per-document topic proportions $\theta_{1:D}$, and the per-word topic assignments $Z_{1:D,1:N}$ (Blei and Lafferty, 2009). These variables are inferred by approximate inference method because exact inference computation is intractable. To overcome this difficulty, (Blei et al., 2003) introduced free variational parameters γ and ϕ for Dirichlet and multinomial distribution, respectively, and defined variational distribution.

$$q(\theta, z|\gamma,\phi) = q(\theta|\gamma) \prod_{n=1}^{N} q(z_n|\phi_n) \tag{3.5}$$

The maximum likelihood estimators of α and β are calculated by the EM algorithm using Jensen's inequality to estimate the lower bound on the log-likelihood of the variational distribution $q(\theta, z|\gamma, \phi)$.

Figure 3.1 illustrates a graphical model representation of the LDA model; where the nodes are random variables and the edges denote dependence between random variables. The shaded node denotes the observed random variables while the unshaded nodes represent the hidden random variables. The rectangular boxes are *plate notations*, which denote replication. Let K be a specified number of topics, V the size of the vocabulary, D the number of documents in the collection, N the number of unique words, α a positive K-vector, and η a scalar. The variables in figure 3.1 represent: α Dirichlet parameter, θ_d per-document distribution, $Z_{d,n}$ per-word topic assignment, $W_{d,n}$ observed words, β_k topics, and η topic hyperparameter.

3.2 Topic Model Based Document Retrieval

As mentioned in chapter 1, for the retrieval of documents, our proposed recommender system makes use of topic models. Document Retrieval refers to the computerized process of producing a list of documents that are relevant to an inquirer's request by comparing the user's request to an automatically produced index of the textual content of documents in the system (Liddy, 2005). A number of theoretical models have been developed and used for document retrieval such as Boolean, Vector Space, Probabilistic, and Language Modeling models. These models match the terms in a user query to the index terms that represent a document, and rank the matched documents for retrieval. Document Indexing (DI) is considered a crucial technique to retrieve significant information for users (Choi and Lee, 2010). Using DI, text documents are converted into index terms or document features that can be trivially analyzed by computers. With an appropriate ranking function or retrieval model, and has been shown to be effective for document retrieval (Croft et al., 2010).

Document Modeling (DM) refers to the task of finding inherent structures and features in documents. Document modeling can be applied to DI because the features of a document can be regarded as index terms. Concept models (a DM method) analyze documents in a concept space by considering a hidden layer of the interweaving relationship between terms, which allows it to overcome the difficulties of the Vector Space Model (VSM) using concept representation for documents (Wang et al., 2010). The concept model is based on the assumption that terms appearing frequently in a document are likely to be related to each other. A deterministic approach in concept modeling is Latent Semantic Analysis (Deerwester et al., 1990), which derives conceptual features that can be utilized for automatic document indexing and retrieval.

Probabilistic concept modeling, developed as an attempt to relax the assumption made in the concept model that each document is generated from a single topic, is a probabilistic approach to model documents in terms of latent concepts – a mixture of topics (Wang et al., 2010). Examples of generative probabilistic concept models are Probabilistic Latent Semantic Analysis (Deerwester et al., 1990), Probabilistic Latent Semantic Indexing (Hofmann, 1999), and LDA (Blei et al., 2003). There are a number of search algorithms incorporating some form of topic model in information retrieval e.g. cluster-based retrieval, pLSI, and more recently LDA (Wei and Croft, 2006). These techniques have also been used for document representation. Typically, documents are represented as a "bag of words", that is, the words are assumed to occur independently. However, to capture the relationships between words, topic models can be used which group the words in a corpus into themes/topics.

LDA, a widely used approach to building topic models is based on a formal generative model of documents, and has been extensively studied considering its feasibility and effectiveness for information retrieval. In the experiments conducted by Wei and Croft (2006) which compared LDA-based models to the pLSI model for information retrieval, they noted that the LDA model possesses consistent generative semantics by treating the topic mixture distribution as a k-parameter hidden random variable rather than a large set of individual parameters which are explicitly linked to the training set; thus LDA overcomes the overfitting problem and the problem of generating new documents in pLSI. In comparison to the cluster model, Wei and Croft (2006) also noted that the LDA model allows a document to contain a mixture of topics, relaxing the assumption made in the cluster model that each document is generated from only one topic. This assumption may be too limited to effectively model a large collection of documents; in contrast, the LDA model allows a document to exhibit multiple topics to different degrees, thus being more flexible.

Furthermore, Yi and Allan (2009) explored the utility of different types of topic models for retrieval purposes. The effectiveness of different types of topic models were evaluated within retrieval approaches. Their work reports that: topic models are effective for document smoothing, and more rigorous topic models such as Latent Dirichlet Allocation provide gains over cluster-based models, and smoothing documents by using their similar documents is as effective as smoothing them by using topic models.

A number of approaches have been implemented to include topic models in information retrieval. For example, a document is represented by itself and the topics to which it belongs, in which the computation of $P(w|D)$ also incorporates the probabilities of the topics. Another approach makes use of query expansion, where a query is computed based on its related topic using topic models (Yi and Allan, 2009). However, in this thesis we make use of the document indexing and retrieval method – *Indexing by Latent Dirichlet Allocation* (LDI), developed by Wang et al., (2010), details of this method are discussed in section 3.2.1.

3.2.1 Document Indexing and Retrieval

Indexing and retrieval using probabilistic concept models are based on the assumption that the concepts are distributed differently in relevant and non-relevant documents. Latent Dirichlet Allocation, an alternative generative probabilistic concept model makes use of a symmetric Dirichlet prior, to resolve the limitations associated with pLSI (Wallach et al., 2009). LDA addresses the topic-based structural analysis of corpora, and thus it can be regarded as a model for topic search. LDA has mainly been used in document modeling and classification in a few research studies, such as (Wei and Croft, 2006; Azzoardi et al., 2004), which applied LDA in the context of query searches. Although these studies focused on methods of avoiding the assignment of a zero value to the conditional probability of a query given a document, they have shown the effectiveness of LDA in retrieval tasks.

Given that LDA models documents as a mixture of topics, the LDI method leverages this approach for representing documents in a topic space where the topics can be seen as index terms for indexing. In the next sections, we discuss how the LDI method defines terms and document representations in the topic space for indexing, and second, how the representations can be applied for the document retrieval task.

3.2.1.1 Document Representation in Topic Space

1. Probability of a topic given a word
   The LDI method directly uses the β matrix of the LDA model to construct explicit document representations associated with topics. The conditional probability β_{jk} in LDA is considered as the selection probability of the word w^j given a topic (concept) z^k; which represents the probability of a word given a specific topic and is used to identify words that are associated with a topic. However, for the characterization of the probability of a topic given a word, the word representation is defined in topic space, $W^j \in \mathbb{R}^K$. The k^{th} component W_j^k of W_j represents the probability of word w^j embodying the kth concept z^k, using Bayes' rule, this can be represented as:

$$W_j^k = p(z^k = 1|w^j = 1) = \frac{p(w^j = 1|z^k = 1)p(z^k = 1)}{\sum_{h=1}^{K} p(w^j = 1|z^h = 1)p(z^h = 1)} \tag{3.6}$$

   The LDI method makes the assumption that the probability of a topic selection is uniformly generated i.e. $p(z^h = 1) = p(z^k = 1)$, since the topics are typically unordered. With this assumption, the probability of a word w^j corresponding to a topic z^k is given as:

$$W_j^k = \frac{\beta_{jk}}{\sum_{h=1}^{K} \beta_{jh}} \tag{3.7}$$

2. Probability of topic given a document
   Next, the documents are also represented in the topic space $D^i \in \mathbb{R}^k$. The k^{th} D_i^k of D_i represents the probability of a topic z^k, given a document d^i is expressed as:

$$\begin{aligned} D_i^k = p(z^k|d_i) &= \sum_{w^j \in d_i} p(z^k|w^j, d_i)p(w^j|d_i) \\ &= \sum_{w^j \in d_i} p(z^k|w^j)p(w^j|d_i) \end{aligned} \tag{3.8}$$

   Here also, the LDI method assumes that the conditional probability $p(z^k|w^j, d_i) = p(z^k|w^j)$. This assumption is based on the LDA assumption that there is a fixed number of underlying topics that are used to generate the words in documents (Croft et al., 2010). In other words, it is assumed that the words in topic space do not depend on which document it is used in, but on the topic it is generated from. An approximation of $\hat{D}_i^k$ of D_i^k can be obtained by substituting $p(w^j|d_i)$ with $\hat{p}(w^j|d_i)$:

$$\hat{p}(w^j|d_i) = \frac{\eta_{ij}}{N_{di}} \tag{3.9}$$

where η_{ij} denotes the number of occurrences of word w^j in document d_i and N_{di} denotes the number of words in document d_i, i.e. $N_{di} = \sum_{j=1}^{V} \eta_{ij}$. By substituting equation (3.9) in (3.8), the approximation of $\hat{D}_i^k$ becomes:

$$\begin{aligned} D_i^k \approxeq \hat{D}_i^k =& \frac{\sum_{w^j \in d_i} p(z^k|w^j)\eta_{ij}}{N_{di}} \\ =& \frac{\sum_{w^j \in d_i} W_j^k \eta_{ij}}{N_{di}} \end{aligned} \tag{3.10}$$

In general, a document includes various words that are used to explain key topics in the document. The definition of document probability in equation (3.10) captures the topical features of words in the document. This definition is distinguished from the usual definition of the probability of a document in LDA, which assumes that document probability is the same as the probability of the simultaneous occurrence of all words used in the document. The new definition overcomes the difficulty that is associated with the latter definition in which the probability of a document considerably depends on the length of the document. If topics are regarded as index terms, document representation in the topic space can therefore be utilized for automatic document indexing (Wang et al., 2010).

3.2.1.2 Similarity between Document and Query

Given the definition in the previous subsection, each term can be represented in the topic space, i.e. $W_j^k = \left\{W_j^1, W_j^2, \ldots, W_j^K\right\}$. The similarity between two terms w^s and w^t is defined as:

$$\begin{aligned} p(w^s, w^t) &= \overrightarrow{W}_s \cdot \overrightarrow{W}_t \\ &= \sum_{k=1}^{K} \frac{p(z^k|w^s)p(z^k|w^t)}{\sum_{h=1}^{K} p(z^h|w^d) \sum_{h=1}^{K} p(z^h|w^t)} \\ &= \sum_{k=1}^{K} \frac{\beta_{sk}\beta_{tk}}{\sum_{h=1}^{K} \beta_{sh} \sum_{h=1}^{K} \beta_{th}} \end{aligned} \tag{3.11}$$

where $\overrightarrow{W_s} = \frac{W_s}{\|W_s\|}$, $\overrightarrow{W_t} = \frac{W_t}{\|W_t\|}$ and $\overrightarrow{W_s} \cdot \overrightarrow{W_t} = \left\langle \frac{W_s}{\|W_s\|}, \frac{W_t}{\|W_t\|} \right\rangle$.

The above similarity measure quantifies the proximity of two terms in topic space in terms of the cosine value of the angle between them. Thus, in general, the similarity between two distinct terms in general does not equal to zero, which mitigates the problem of synonymy. This also alleviates the problem of polysemy since each term has multiple topical interpretations, due to the representation in a topic space (Wang et al., 2010). Analogously, similarity measures to compare two documents and to compare a term and a document can be defined as:

$$p(d_s, d_t) = \overrightarrow{D_s} \cdot \overrightarrow{D_t} \approxeq \overrightarrow{\hat{D}_s} \cdot \overrightarrow{\hat{D}_t} \tag{3.12}$$

and

$$p(w^s, d_t) = \overrightarrow{W_s} \cdot \overrightarrow{D_t} \approxeq \overrightarrow{W_s} \cdot \overrightarrow{\hat{D}_t} \tag{3.13}$$

respectively.

The query as a pseudo-document that contains a set of query terms $Q = \{q_1, q_2, \ldots, q_L\}$. Similar to equation (3.6), the probability vector of the query with respect to the kth topic can be defined in the concept space as:

$$\begin{aligned} Q^k = p(z^k|Q) &= \sum_{q_j \in Q} p(z^k|q_j, Q)p(q_j|Q) \\ &\approxeq \frac{\sum_{q_j \in Q} p(z^k|q_j)}{L} \end{aligned} \tag{3.14}$$

and the similarity between query Q and document d_s is measured by:

$$p(d_s, Q) = \overrightarrow{D_s} \cdot \overrightarrow{Q} \approxeq \overrightarrow{\hat{D}_s} \cdot \overrightarrow{\hat{Q}} \tag{3.15}$$

where $\overrightarrow{Q} = \{Q^1, Q^2, \ldots, Q^K\}$ in the topic space. Representing a query as a probability vector is made possible owing to the definition of the document probability vector in equation (3.6). The probability vector represents the characteristics of the words, documents, and queries in the topic space. An advantage of the LDI method is that an unseen training query can be treated coherently as a document in the training set (Wang et al., 2010). This feature is pertinent to LDA, however, is not implemented in other automatic indexing methods such as pLSI (Hofmann, 2001). As with the LDA model, the size of topic space K plays an important role in the LDI method also. In LDA, the topic size K determines the degree of abstraction of information, i.e. the larger the value K is, the finer is the segmentation of information.

3.3 LDI Model Evaluation

To evaluate the LDI model, Wang et al., (2010) compared the model to a tf-idf weight-based VSM model, LDA-based document modeling (LBDM), LSI, and pLSI on four data set collections: CRAN, MED, CISI, CACM. Each data set contains a corpus, list of queries and corresponding relevant documents. Figure 3.2 shows the precision-recall curves of the tested methods on the four standard data sets. The numerical values of the x-axis denote the recall of labeled relevant documents, while the numerical values of the y-axis represent the precision of retrieved documents. Compared to TFIDF, LSI, pLSI, and LBDM, the authors noted that the LDI model achieved the best performance, except for small intervals in recalls on MED and CRAN. In the experiments on CRAN, CISI, and CACM, LDI had a higher precision for a high recall regime – a property seems more valuable in practice since the documents in higher positions are viewed by more users (Wang et al., 2010).

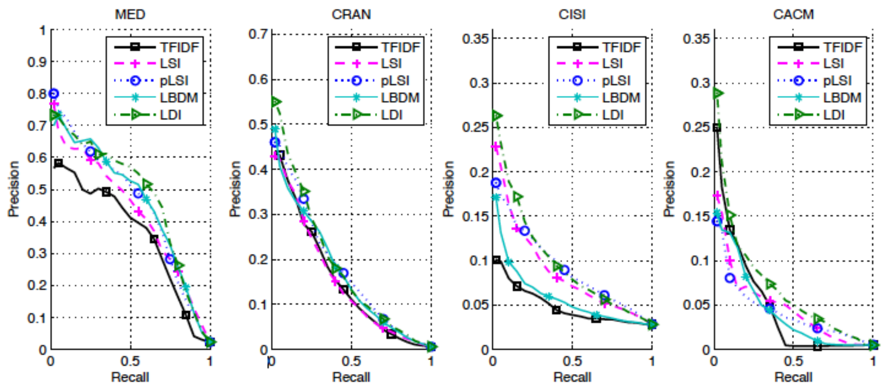


Figure 3.2: Precision-Recall Curves for TFIDF, LSI, pLSI, LBDM, and LDI on MED, CRAN, CISI, and CACM.(Wang et al., 2010)

The proposed LDI performed better than the indexing methods TFIDF, LSI, and pLSI, and showed a comparable performance with the retrieval model LBDM. In addition to the performance superiority to the LBDM, the proposed LDI also has advantages in the following two aspects: (1) the topical representation of a document for automatic indexing is proposed in our method, while it is not introduced by the LBDM; and (2) no smoothing parameters need to be tuned in our method whereas additional parameters need to be trained in the LBDM (Wang et al., 2010).

In conclusion, according to the authors (Wang et al., 2010), the LDI method is a viable automatic document indexing method for information retrieval. Also, since LDI is a generative model, it can be employed in the information systems without labeled items, such as search engines, as to retrieve relevant documents according to the user input queries. Therefore, in this thesis, for the identification and retrieval of relevant reading resources, we deployed the LDI model.

Chapter 4

Metacognition and Self-Regulated Learning

The traditional role of the classroom teacher is significantly changing with time. According to Williamson (2015), this change is due to the developments in the definition of learning. Previously, learning was associated with knowledge absorption, however, it is now recognized as the active construction of knowledge (de Jager et al., 2005). Today, it is argued that learning should be about how to deal with (new) challenging situations or problems and not necessarily to regurgitate or apply objective facts only Boud (2001). This understanding of learning has resulted in a number of pedagogical developments, one being the increased need for learners to be self-directed (Williamson, 2015), and thus learning how to learn has become an important educational issue (Vermunt, 1995). Self-directed learning enhances the learner's ability to 'manage self', and this competency encompasses other concepts such as self-motivation, self- belief, solving problems, working independently, setting goals and assessing one's own learning. It is advocated that the skill self-directed learning should be taught and encouraged at all levels of education as it has been shown that learners who can manage themselves demonstrate resourcefulness, reliability and resiliency (Ministry of Education, 2007).

Self-directed learning, self-planned learning, self-education, self-regulated learning, independent learning and open learning are concepts that (loosely speaking) define the same phenomenon (Dagal and Bayindir, 2016). Self-directed learning can be described as a process in which individuals take the initiative, with or without the help of others, in diagnosing their learning needs, formulating learning goals, identifying human and material resources for learning, choosing and implementing appropriate learning strategies, and evaluating learning outcomes (Knowles, 1975). According to Zimmerman (1986, 1989), a learner can be described as self-regulated to the degree that they are metacognitively, motivationally and behaviorally active participants of their learning process, where they initiate and direct their efforts to acquire knowledge and skills without relying on other agents (e.g. parents, teachers) for instructions. Although, the two concepts: self-regulated learning and self-directed learning can be argued to be different, they share similarities such as learners' independence, active engagement, goal-oriented behavior and control of the learning process (Loyen et al., 2008; Pilling-Cormick and Garrison 2007). These similarities are the underlying learning conditions that this thesis address with providing an appropriate TEL environment which facilitates learning and assists the learner in the achievement of the learning goal by the recommendation of relevant learning resources.

Self-regulated learning has three components: motivation, metacognition, and learning behavior (Zimmerman, 1990; Dignath et al., 2008). In the context of self-regulated learning, motivation refers to a learner's self-efficacy and autonomy, and is closely linked to a learner's goal (Bolhuis 2003; Zimmerman 1990). The achievement of the learning goal helps the learner pay attention to the learning process and deploy appropriate cognitive strategies (Bolhuis, 2003). The term metacognition in relation to self-regulated learning refers to a learner's ability to think consciously about their cognition and have control over their cognitive processes (Zimmerman, 1989). Metacognition can be defined as monitoring and controlling cognition, in other words, the individual's thinking of his/her own cognitive processes (Flavel 1989; Dinsinmore et al., 2008). Metacognition can be categorized in two groups: knowledge about cognition, and metacognitive skills. Knowledge about cognition relates to learners having knowledge about their own cognitive activities and cognitive strategies (Dagal and Bayindir, 2016) while metacognitive skills define the processes that organize and control cognition, such as: planning, monitoring and evaluating (Boekaerts, 1999).

Recently, the concept of metacognition has emerged as a popular area of study. A possible reason for this is that there exists a positive relationship between metacognitive elements and academic success (Bagceci and Sarica, 2011). Meta-cognition has been identified to be closely related to learning processes, learning goal attainment and academic success because it makes the learner aware of their thinking processes and the ability to control their cognitive system (Baltaci and Akpinar, 2011; Vrugt and Oort, 2008). Learners with high metacognitive awareness are therefore aware of what they know and do not know, and are successful at planning, information management, contracting strategies, monitoring, debugging and evaluating. Thus, it can be said that metacognition has a positive effect on learning success and goal achievement (Schraw and Sperling-Dennison, 1994; Zimmerman and Schunk, 2001).

Learning behavior, a third component of self-regulated learning, entails the decisions and actions the learners make in order to optimize their learning environment (ZImmerman, 1990). This includes making available appropriate tools and resources that empowers the learner, give the learner control over the learning process, as well as fosters self-direction, collaboration and cooperation that encourage the idea of learning community (Bolhuis and Voeten, 2001; Watson, 2004). Based on this notion, we hypothesize that a TEL environment equipped with adequate metacognitive tools and that allows for collaboration would truly facilitate self-regulated learning. In the next section, we discuss metacognitive standards and study strategies that learners deploy during learning.

4.1 Metacognitive Standards, Reading Strategies

Academic learning has been regarded as a self-regulated process where learners use and direct their thoughts, feelings and actions to achieve learning goals (Zimmerman, 2002). The process of self-regulated learning consists of sub-processes which includes goal setting, planning and selecting strategies, monitoring and evaluating performance (Winne and Hadwin, 1998). Monitoring and evaluating one's learning is a metacognitive process that is integral to self-regulated learning (Tobbias and Everson, 2000; Winne, 2001). Based on this perspective, the process of monitoring requires the use of a set of metacognitive standards which could be used to compare the current status or condition of learning (Winne, 2001). The metacognitive standards are created by the learner and are influenced by both external task and cognitive conditions (Winne and Hadwin, 1998). Task conditions include task instructions, resources available for the learner, and the learning environment. Cognitive conditions include factors such as learners' past learning experiences, personal dispositions, beliefs about knowledge and knowing, domain knowledge, and knowledge of study tactics and strategies (Marzouk, 2018).

Reading is a complex cognitive process involving various sub processes (Mckeown and Beck, 2009) some of which are metacognitive. Choosing when to use a reading strategy, how to use it, evaluating and monitoring comprehension, and controlling strategy and use are examples of metacognitive processes taking place while reading (Baker and Brown, 1984). There are a number of metacognitive reading activities that a learner may engage in; also a number of systems that support self-regulated learning have been developed to facilitate some of the metacognitive reading strategies and activities. In the next section, we review some of the existing systems designed to support metacognitive reading activities.

4.1.1 Metacognitive Reading Activities in Self Regulated Learning Setting

There exists a number of learning environments or platforms that have been designed to support self regulated learning. MetaTutor, a hypermedia learning environment designed to detect, model, trace, and foster students' self regulated learning about human body systems such as the circulatory, digestive, and nervous systems Azevedo et al., (2008). MetaTutor entails four phases to train students on SRL processes and to learn about the various human body systems. The phases include (1) modeling of key SRL processes, (2) discrimination task where learners choose between good and poor use of these processes, (3) a detection task where learners get to see video clips of human agents engage in similar learning tasks and are asked to stop the video whenever they see the use of a SRL processes (and then they have to indicate the process from a list), and (4) the actual learning environment used to learn about the biological system. Overall, the system was designed to provide adaptive human scaffolding, that addresses both the content of the domain and the processes of self regulated learning, which enhances students' learning about challenging science topics with hypermedia. A Narrative-centered learning environment for eighth grade Microbiology, Crystal Island (Rowe et al., 2009) was designed to provide significant potential for enhancing students' learning experiences. It deploys engaging interactive narrative experiences that are pedagogically effective and tailored to an individual student to provide problem-solving guidance that simultaneously enhances students' self-efficacy for self regulated learning. Due to their interactive nature, narrative-centered learning environments are designed to cope with a wide range of actions a student may perform, thus providing them with a strong sense of control – important for supporting motivation and self regulation. Some of the actions a student can perform with Crystal Island includes: pick up and manipulate objects, take notes, view posters, operate lab equipment, and talk with non-player characters.

Self explanation, a metacognitive reading strategy describes the process of explaining text to one's self either orally or in writing. These explanations are generally based on information contained in the discourse context and can be initiated while reading (McNamara and Magliano, 2008). A number of frameworks, systems have been developed to encourage and initiate self-explanation while reading for a number of domains (e.g maths). Self-Evaluation Coach (SE-Coach) developed by Conati and VanLehn (2000) is a scaffolding tool meant to encourage students to spontaneously self-explain. This is achieved at two levels: the first level of scaffolding is provided by a masking mechanism that presents different parts of an example covered by grey boxes, each corresponding to a "unit" of information. When the student moves the mouse over a box, it disappears, revealing the text or graphics under it. The second level of scaffolding is provided through specific prompts to self-explain. Whenever the student unmasks a piece of the example, if it contains an idea worthy of explanation the interface will append a button labeled "self-explain". Pressing the button produces simple prompts to initiate self-explanations in terms of domain principles. Tajika et al., (2007) examined the effect of self-explanation on word problem solving. The results of user study performed Tajika et al., (2007) showed that students in the self-explanation group outperformed students in the other groups on both the ratio word problem test and on the transfer test. In addition, high explainers who generated more self-explanations relating to deep understanding of worked out examples outperformed low explainers on both ratio word problem and transfer tests. NORMIT, a constraint-based tutor developed by Mitrovic (2002) teaches data normalization and supports self-explanation. In comparison to other self-explanation systems, NORMIT requires an explanation from users for each action that is performed for the first time and not at every step. For subsequent actions of the the same type, an explanation is required only if it is performed incorrectly. The results of the user study performed by Mitrovic (2002) revealed that self-explanation increased problem solving skills and better conceptual knowledge. These results show the effectiveness of self-explanation as a powerful and useful metacognitive reading technique to enhance learning.

Therefore, in this thesis we examine the use of other metacognitive reading activities and strategies to facilitate personalized learning. Text marking, a reading strategy has been identified as the most preferred strategy among college students (Gier et al, 2009). Many college students reported that marking textbooks increased concentration, enhanced comprehension, and facilitated review (Nist and Kirby, 1986), and is perceived as effortless, requires no training and minimizes material to study and review (Blanchard and Mikkleson, 1987). Text marking is also widely promoted, study skills courses at schools and universities advocate text marking as an effective study strategy (Wade and Trathen, 1989). Other common reading strategies that learners engage in are knowledge/concept maps and note taking (Marzouk, 2018). While a number of research have been done to investigate the efficacy of each of these study strategies (comparing and contrasting study techniques) on reading comprehension, information retention, recall, knowledge transfer, among others, in this thesis, we investigate the use of a study strategy (text marking) for recommendation of learning materials.

Although text marking involves both cognitive and metacognitive processes, adopting Winne's (2001) "if-then" view of a reading strategy provides deeper understanding of what takes place when a reader is interacting with information. According to this view, if a learner judges a set of criteria is satisfied then a study strategy is applied (Marzouk, 2018). The judgment component of this sequence is a metacognitive act because it involves learner's thinking about and using self-created standards to guide learners' cognition about the text (Winne, 2001). Applying this definition to text marking, when learners are reading and marking text they use metacognitive criteria to identify which information merits marking (Marzouk, 2018). Given a reading objective (e.g., "Read the following text and mark important information"). Learners use the reading objectives to create standards to guide judgment when reading text about which text to mark and during study, learners use metacognitive standards created in relation to objectives to judge whether to mark. If learners are not required to mark overtly, learners still use these standards to metacognitively judge whether a text segment is worthy of attention (Marzouk, 2018).

Information worthy of attention does not necessarily mean it is important. Therefore, a distinction between two key concepts: importance and relevance, is imperative. According to McCrudden and Schraw (2007) "relevance is the degree to which a text segment is germane to a specific task or goal, whereas importance is the degree to which a segment contains essential information needed to understand a text" (p.114). Important segments in text often are cued by the author (e.g., by typographical cues, order of presentation). Thus importance is text-related. Relevance, on the other hand, is determined by learner's objectives or standards. It is a text-external phenomenon (McCrudden and Schraw, 2007).

A relevant text segment does not need to be important. Based on McCrudden and Schraw's (2007) concept of relevance, learners generate metacognitive standards in relation to cognitive and task conditions, then use those standards to judge whether a text segment deserves marking (Marzouk, 2018). Marked text indicates that a learner judged it relevant. Thus, the process of determining whether a text segment should be marked or otherwise has both cognitive and metacognitive components. The cognitive component includes text encoding processes and accessing prior knowledge to comprehend what is read; the metacognitive part involves (a) monitoring, applying and continuously adjusting standards, and (b) controlling processes leading to mark text or not (Marzouk, 2018).

Having established the metacognitive aspect of text marking as a study strategy, we investigate its use as a learner model, preference elicitation method, for the recommendation of personalized learning materials. More details on this is provided in chapter 5.

4.2 Self-regulated learning and Motivation

In the context of self-regulated learning, motivation refers to a learner's self-efficacy and autonomy, and is closely linked to a learner's goal (Bolhuis 2003; Zimmerman 1990). The achievement of the learning goal helps the learner pay attention to the learning process and deploy appropriate cognitive strategies (Bolhuis, 2003). Learner motivation involves learners' goals for the task and their beliefs about the importance and interest of the task. Essentially it concerns learners' reasons for doing a task, that is, the learners' individual answers to the question, "Why am I doing this task?" (Pintrich, 1990). Self Determination Theory (Deci and Ryan, 1985) distinguishes between intrinsic and extrinsic motivation based on the different goals or reasons a learner obtains a goal for completing a task (Ryan and Deci, 2000).

When a learner is inherently interested in a task, the learner is said to be intrinsically motivated (Ryan and Deci, 2000; Murphy and Alexander, 2000). According to Ryan and Deci, (2000), intrinsic motivation results in high-quality learning and creativity, a goal of modern education (Boekaerts et al., 2006). A learner's interest in a learning task becomes intrinsic when he or she considers the task itself rewarding (Code et al., 2006). Intrinsic motivation is a natural motivational tendency and is a critical element in cognitive, social, and physical development because it is through acting on one's inherent interests that one grows in knowledge and skills (Murphy and Alexander, 2000). On the other hand, when a learner is extrinsically motivated about a particular task, they perform the task to accomplish a goal set by another individual rather than an inherent goal (Murphy and Alexander, 2000). The main difference between the constructs of intrinsic and extrinsic motivation is the perceived control or influence is either external or internal e.g. task instruction (instructor goals) – where the learning goals and objectives are set by the instructor, and learner goals – where the learning goals are set by the learner (Tilstra and McMaster, 2013).

Having goals and objectives is important for learning as they define “where you are headed and how to demonstrate when you have arrived” (Kaufman, 2000). Goal orientation of learning encapsulates the reasons a learner performs a task and it assists with evaluating the learner’s performance on the task (Pintrich 1990). Two major goal orientations exist in the literature: 1) mastery orientation, also called task-goal orientation and learning-goal orientation and 2) performance-goal orientations, also called ego orientation and ability-goal orientation (Ames, 1992). Learners’ different goal orientations are key to understanding their varying approaches to regulate their learning in a particular task. Students who adopt a mastery goal orientation are theorized to persist, deeply elaborate study material, and experience enhanced task enjoyment (Dewek, 1998). Students who adopt a performance goal orientation are theorized to process study materials less deeply, experience decreased task enjoyment, and withdraw effort in the face of failure (Dewek, 1998).

As has been mentioned, learners’ motivation involves setting learning goals and objectives for the task; which could be done by the learner or by task instruction. According to Smith and Regan (1993), task instructions that include reading objectives are perceived to explicitly describe what the learners should know or be able to do at the completion of instruction. This enables learners, each having unique cognitive conditions, to use the learning goals and objectives specified, to construct personal (metacognitive) standards to decide which information is worthy of attention. While reading and studying, learners use these metacognitive standards to guide search and selection processes to locate information that merits attention (Winne and Hadwin, 1998). To test the recommendation system proposed and developed in this thesis, a user study was performed (more details of the user study are discussed in chapter 5). The user study involved a reading task in a self-regulated learning setting. The learning goal (reading objective) for the task was specified using task instructions.

The inclusion of the reasons (learning goal) why a learner is pursuing a task also allows for the integration of the achievement motivation into the model of self-regulated learning (Pintrich and Schunk, 1996). Intelligent self-regulation requires that the learner has in mind some goals to be achieved, against which performance can be compared and assessed. The feedback from the learner’s assessment gives useful information on the learner’s present state of learning and performance in relation to the learning goals and objectives (Nicol and Macfarlane-Dick, 2005). In the next section, we discuss the different types of assessments in self-regulating setting and how they facilitate learning and comprehension.

4.3 Assessments and Self-Regulated Learning

Assessment through posing questions is an integral part of learning. It can be defined as the process of gathering data to better understand the strengths and weaknesses of students’ learning (Harris and Hodges, 1995). Assessment can also be described as a reflective process where learners use criteria to evaluate their performance and determine how to improve (Siegesmund, 2017). It makes available important data that can be used to measure the progress of learning with respect to the learning goals and objectives. Research has also shown that learners need assessments to learn, regardless of whether they are posed by teachers or formulated by the students themselves (Morgan and Saxton, 1994), and also that assessments increase comprehension for learners (Rittle-Johnson, 2006).

Assessment plays an important role in the process of learning and motivation. There are three types of assessment, and it has been argued that the type of assessment tasks learners perform determines how learners will approach the learning task and what study behaviors they use. In the words of higher education scholar John Biggs, "what and how students learn depends to a major extent on how they think they will be assessed." (Biggs, 1999, p. 141). Given the importance of assessment for learning, it is important to consider and include appropriate mechanisms to measure learning progress or otherwise especially in a self-regulated learning setting. Well-designed assessment methods provide valuable information about the student's learning. They tell us what students learned, how well they learned it, and where they struggled. Assessment therefore becomes a lens for understanding student learning, identifying invisible barriers, and helping us to improve our teaching approaches.

In the next sections, we discuss the types of assessments and how they can be used to facilitate learning.

4.3.1 Diagnostic Assessment: Assessment as Learning

Diagnostic assessments (assessment as learning) emphasize the role of the student, not only as a contributor to the assessment and learning process, but also as the critical connector between them (Lorna, 2003). It is also considered as a process in metacognition, which occurs when learners monitor what they are learning and use the feedback from this monitoring to make adjustments, adaptations, and even major changes in what they understand (Lorna, 2003). Types of diagnostic assessments include pre-tests (on content and abilities), self-assessments (identifying skills and competencies), discussion board responses (on content-specific prompts), interviews (brief, private, 10-minute interview of each student).

4.3.2 Formative Assessment: Assessment for Learning

Formative assessment refers to assessment that is specifically intended to generate feedback on performance to improve and accelerate learning (Sadler, 1998). It provides feedback and information during the instructional process, while learning is taking place, and while learning is occurring. Formative assessment measures student progress – a primary focus of the assessment, to identify areas that the learner may need improvement. Also, a central argument is that, formative assessment and feedback should be used to empower students as self-regulated learners and to improve learning (Siegesmund, 2017).

4.3.3 Summative Assessment: Assessment of Learning

Summative assessment takes place after the learning has been completed and provides information and feedback that sums up the teaching and learning process. When the information from an assessment is used solely to make a judgment about level of competence or achievement, it is a summative assessment. Summative assessment is more task-oriented and assesses the learner upon completion of the task, whereas formative assessment focuses on the process toward completing the task. In summative assessment, once the task is completed, no further revisions can be made. However, if the learners are allowed to make revisions, the assessment becomes formative, where students can take advantage of the opportunity to improve their learning.

In self-regulated learning settings and platforms, natural language generation techniques have been deployed to automatically generate questions from text as assessment procedures for learners. Odilinye et al., (2015) developed methods that generate questions from the learning material a learner engages in that focuses on pedagogical goals. The method entailed aligning assessments with learning objectives (pedagogical goals) which provides evidence to both the instructor and learner that the learner has obtained the required knowledge gain from the learning material. The method developed by Odilinye et al., (2015) could serve as both formative and summative assessments.

As a formative assessment, the question generation method generated questions based on a learner model (to provide adaptive feedback to the learner), as well as the instructor goals (questions aligned to the learning objective). The method examined the difference (missed items) between the specified instructor goals and learner model (obtained from the learner's interaction with the learning material). This technique aimed at prompting the learner to focus on the important key terms of the document. For example, if the difference between the learner model and instructor goal is not an empty set, questions generated from the non-empty set may help the learner pay more attention to the missed key terms of the instructor goal. As a summative assessment, the automatic question generation module generated questions based on pedagogical goals as specified by the instructor goals. This technique aimed at facilitating the learning process by generating questions based on the pedagogical goals of the learning material, thus ensuring that the learner derives the necessary knowledge and skill from the reading exercise.

The automatic question generation module is not a part of this thesis, since the questions generated by the technique were evaluated by both human experts as well as automatic methods, ROUGE (an automatic summarization evaluation technique, Lin and Hovy, 2003). In the next chapter, we discuss our proposed methodology for the recommendation of textual learning materials using the metacognitive activities of the learner as well as the LDI model. The next chapter also discuss the experimental design and the user study conducted.

Chapter 5

Methodology

In this chapter, we discuss the proposed methodology adopted for the design and development of the recommender system. This initial version of the recommender system is designed to recommend educational textual documents only. To evaluate the recommender system, a user-based approach is included. The experimental design and procedure of the user study as well as the educational research objectives of the study are discussed in this chapter.

5.1 Experimental Design

The recommender system developed for this study is intended to assist learners' select appropriate textual documents for task-oriented reading. Task-oriented reading is an activity where an individual reads to meet a goal (McCrudden and Schraw, 2007), which may be provided by an instructor or a self-directed reading goal. Such readings may involve multiple documents. Task-oriented reading of multiple documents therefore requires the ability to search for and identify relevant resources that facilitate the completion of the reading task. In this context, a recommender system is considered a resourceful tool that could be used to identify relevant documents from a large pool of documents. A recommender system for educational purposes therefore should be tailored to support the learners' information seeking needs as well as enhance the learners' learning experience.

The recommender system uses as input, the learners' metacognitive activities to make recommendations and the retrieval of relevant learning materials is based on probabilistic topic modeling, specifically the LDI model (discussed in chapter 3). The process of recommendation incorporates the metacognitive activities of the learner to provide personalized recommendations that aims to address the learner's information seeking needs which may be captured during the process of reading. We refer to the learners' interaction with a learning material as *metacognitive activities*. During a reading session, learners' interact with the reading material in a number of ways, which may include marking a portion of the text (create highlights), make note. These interactions/meta-cognitive activities serve

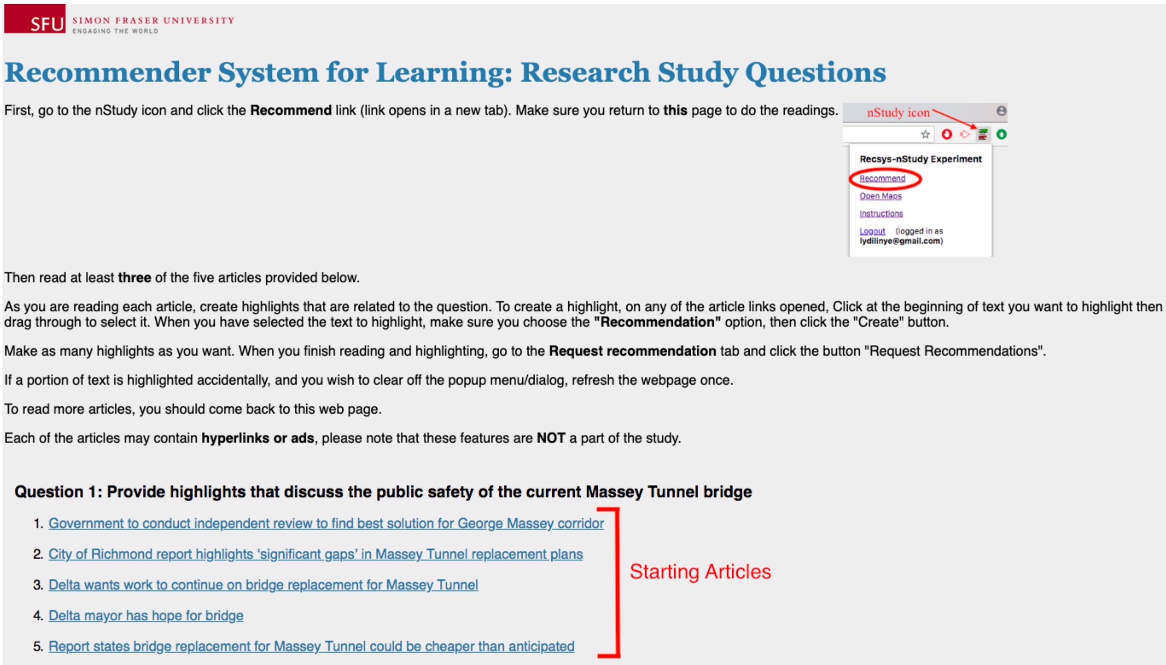


Figure 5.1: Instructions, nStudy icon and the starting articles of the user study.

as the user model and preference elicitation method (input) to the recommender system, and is a novel method/input mechanism for recommendations.

The recommendation process also leverages the capability of topic models to discover the latent topics contained in a collection of documents, as well as the topics contained in the metacognitive activities of the learners. By using the topic modeling approach (LDI model) for document retrieval, the recommender system is able to provide recommendations to the learner based on the topic similarity of their metacognitive activities to documents in the collection. The use of the learners' metacognitive activities as input to the recommender system is intended to replace search queries. Typically, the information seeking needs of a learner are expressed using a search query in a search engine. However, search queries do not always return relevant information, due to reasons such as poor queries and natural language ambiguities (Batista, 2007). Therefore, a more cognitive approach to identifying the learners' information needs can be achieved by examining the interactions the learner engages in with the learning materials. The metacognitive activities of the learner (e.g. highlights, notes, comments) usually contain more information than search queries text, this gives the learner the freedom to express his/her information seeking needs in different ways without limitations.

Using the metacognitive activities of learners as input to the recommender system could also imply that the information contained in the learners interaction may span multiple topics. Therefore, by deploying the LDI model for document retrieval, the system is robust enough to recommend a range of topically diverse documents that are based on the topics

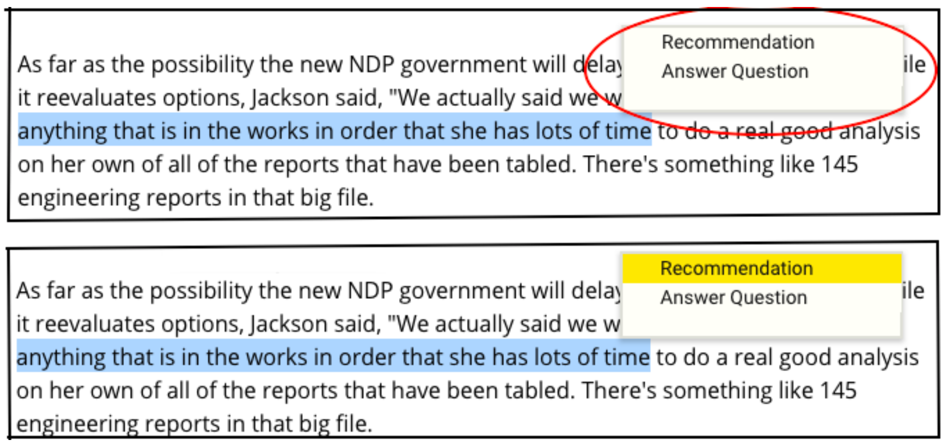


Figure 5.2: Set of tags used in the study.

contained in the learners' metacognitive activities, which are reflective and relevant to the learners' information seeking needs. For an example to compare recommendations based on a user's query and the default document retrieval method (likelihood model) and our proposed approach – recommendation based on the learners' metacognitive activities and the LDI model for document retrieval, given a reading task to "write an essay on the *cost* and *environmental* implications of the Massey tunnel." Assuming the user types the following query on a search engine: "cost and environmental implication of the Massey tunnel." The likelihood document retrieval model would provide documents that discuss both the cost and environmental issues (which may not exist or be adequate to complete the task). However, our proposed approach, the learner isn't required to type in a search query but rather interact naturally with the learning material e.g. by creating highlights (text marking). Assuming the learner creates the following highlights: "financial information and detailed cost estimates of proposed bridge released", "conduct an independent technical review of the bridge." The system designed takes all the learner's highlights and treats them as a pseudo document. Then, it infers the topics contained in the pseudo document, and identifies two topics: "cost information of the Massey Tunnel" and "environmental assessment of the Massey tunnel." Based on the topics inferred, it recommends to the learner documents that discusses each of the topics.

Therefore, our proposed methodology for recommendation first, removes the extra task the learner has to perform by not only typing a search query but also identifying the right search query that would give relevant documents (this may take several rounds of refining the search query). Second, using topic models for document retrieval makes it possible to provide multiple but finer grained topically related documents that would be useful to complete the reading task. The learning platform that was used is nStudy (Beaudoin and Winne, 2009). nStudy is an online tool that supports learning and research. nStudy is a web application offers learners a wide array of tools for identifying and operating on information

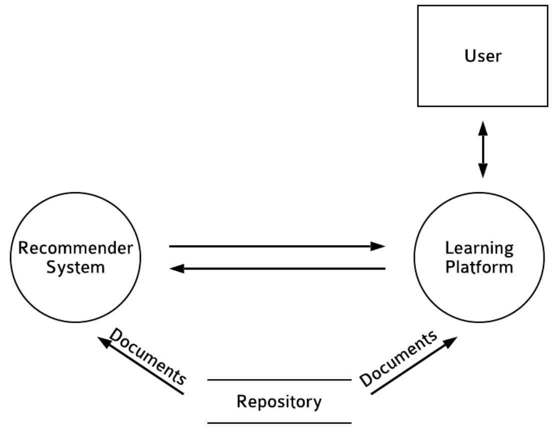


Figure 5.3: Pictorial description of the system and data flow.

they study. nStudy is designed to provide learners and researchers tools to explore their learning skills, metacognition and self-regulated learning. As learners use nStudy's tools to study information in the Internet, nStudy logs fine-grained, time-stamped trace data that reflect the cognitive and metacognitive events in self-regulated learning. It provides a wide range of metacognitive reading tools that allows a learner interact with an online document or web page the way they would with a paper version. Some of the tools nStudy provides includes: highlights (text marking), notes taking, bookmarking, and tags. nStudy provides the annotation tools that allows learners interact with an online document as they would with a paper version. The annotation tools allow the learners record, organize, view the documents they read. Some of the annotation tools that nStudy provides are highlights (text marking), tags, notes. For this study, the tools that are used are the highlights and tags. The nStudy interface would also be used to display the reading task, task instructions, and the participants would be able to read the articles as well as make annotations on the articles using the interface.

As a learner uses nStudy, the software records very fine grained time stamps of all the activities performed by the learner, such as bookmarking the web sites visited, records the notes created, as well as the information operated on (e.g. text highlighted, tags). To obtain data on the learners' interaction, the recommender system developed is integrated with nStudy. As a learner reads articles, make highlights or create tags on nStudy, the details of the portion of text highlighted and/or tag created are sent to the recommender system in real time. Based on the meta-cognitive activities of the learner the recommender system receives, it is able to make personalized recommendations that are tailored to the learner's interests. In this case, the feedback obtained from the learner serves as a prompt (search query) that signals to the recommender system the kind of items the learner may be interested in.

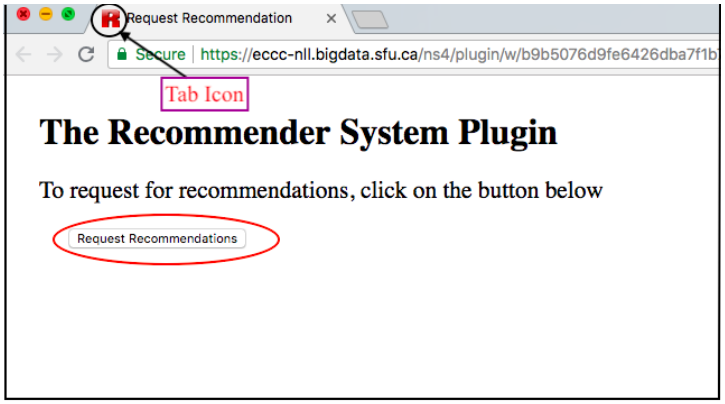


Figure 5.4: Recommender system plugin to request for recommendations.

5.1.1 Description of the System

The recommender system was designed as a plug-in to be integrated to the nStudy learning platform (and could be extended to other online learning platform that support metacognitive reading activities). Figure 5.3 depicts the system's architecture and data flow. To use the system, the learner is first required to either create an account or sign in. This step is important to create a record of all the activities the learner engages in, and is done via nStudy's icon (see Figure 5.1). Thereafter the learner commences searching for documents, reading and interacting with the documents read. In the case of the user study we conducted, after the participants creates and account and/or signs in, the demographics questionnaire is administered online immediately. Upon completion of the questionnaire, a web page containing a brief instruction on what to do, as well as the list of *starting articles* for the task is displayed. The starting articles (which were chosen randomly) are the initial articles the learner reads and interacts with to commence reading session. As the learner reads the articles, they perform metacognitive reading activities (text marking, tags). There may be various reasons a learner creates a highlight. The use of tags could be used to specify the reason. This study adopts the use of tags, however, the set of tags were limited to two, shown in Figure 5.2. This is because to complete the educational task of the study, learners had to put together an ensemble of highlights in lieu of writing an actual essay. Therefore, the tags were used to distinguish between the highlights created to complete the task and otherwise (for recommendation).

Upon creating at least one highlight, and the learner would like to receive more articles (new recommendations), the recommend plug-in is opened and the *request recommendation* button is clicked (see Figure 5.4). The new list of recommended articles is opened on a new tab which contains articles related to the topics/themes inferred from the highlights the learner had made (see Figure 5.5). The recommendation list includes: the title of the article, its URL, a dynamic summary (for each of the articles) and an explanation module – a set of keywords extracted from the highlights the learner created based on which

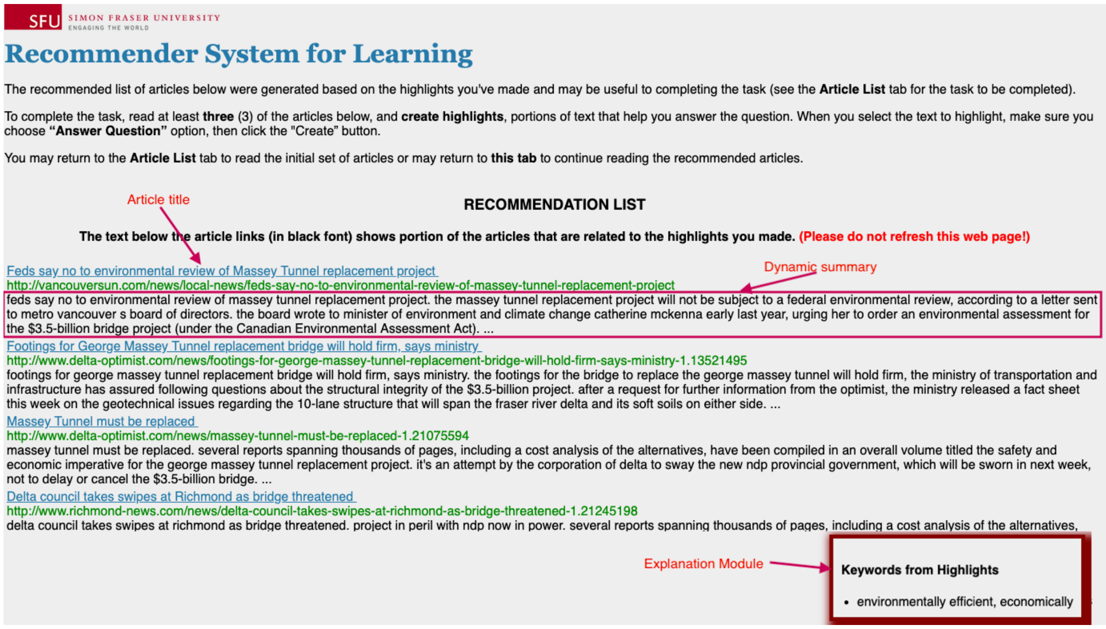


Figure 5.5: List of recommended articles showing the dynamic summary, and explanation module.

the recommendation was generated. The learner could go through the process of reading-highlighting-request recommendation for any number of time and when satisfied with the highlights created to complete the task, the feedback questionnaire is administered for the learner's evaluation of the system.

5.2 User Study Procedure

There has been an increased interest in more user-centered evaluation metrics for recommender systems such as those mentioned in (McNee et al., 2006a). Therefore, a user study was conducted to include a user-based evaluation of the quality of the recommender system and also to attempt to answer the research questions of the thesis (discussed in section 5.3). The participants were 49 undergraduate students, 27 females and 22 males with various disciplinary majors attending a university in Western Canada. Ages ranged from 18 to 26 years (M =21, SD=2.67). All participants were recruited via an advertisement posted in a busy spot on campus.

During a user study session, participants are welcomed to the lab and given the consent letter to sign. Then the participant is assigned an ID which is reflective of which experimental condition the participant belongs. The researcher briefly explains the steps and details of the study to the participant, and thereafter the session begins with completion of the demographics questionnaire. Participants studied as long as they desired. The study session took approximately an hour. When participants finished both questions in the task, they completed the feedback questionnaire and were compensated $18 for taking part in the study.

5.2.1 Treatment Conditions: Variations in Recommendation Strategy

There are two treatment conditions involved in the user study: (a) random recommendation, and (b) experimental recommendation. Participants would be randomly assigned to one of the two treatment groups. Both groups received recommendations from the collection of articles used in the study. However, the *random recommendation* group received randomly selected recommendations using a random number generator module, while the *experimental recommendation* group received recommendations based on this thesis proposed methodology (LDI model and the learners' meta-cognitive activities).

5.2.2 Questionnaires

Two questionnaires were included in the study, a demographics questionnaire, used to obtain some personal information of the participants (e.g. domain knowledge), and a feedback questionnaire where the participants are asked to provide their perception about various aspects of the recommender system. The demographics questionnaire is presented before the participant begin the study task, while the feedback questionnaire is administered after the participant completes the task. The Likert scale is used to design questions in the feedback questionnaire, and participants are expected to select a rating on a scale that ranges from one extreme to another: "strongly agree" to "strongly disagree."

5.2.3 Educational Task and Instructions

The educational task for the user study required that the participants complete two essay-like questions. Participants were not required to write out actual essays but rather provide an ensemble of highlights (marked texts) that would be suitable to answer the questions. For reasons such as different writing capabilities of learners in general, we decided not to ask the learner to write actual essays, but to put together highlights from the (recommended) articles they read that would be adequate to complete the tasks. Before commencing the task, detailed instructions were specified, providing a step-by-step guide on what is expected, and how to complete the task. The instructions were carefully stated to not interfere with the natural way the learners behave when studying, to be able to simulate and obtain best results that depict the 'true' way learners interact with learning materials for a task-oriented activity. To begin the reading process for the task, the participants were presented with a list of five randomly selected articles from the collection of articles (we refer to these articles as the *entry articles*). Only the entry articles titles are presented at this point, and the titles are clickable links to the corresponding web pages. The participants decide on which of the entry articles to commence reading, and when to create highlights.

5.2.4 Materials

The task for the study is about the George Massey tunnel. To obtain a collection of news articles on the George Massey tunnel, we crawled 16 online news websites between August 8 and September 30, 2017. A total of 95 articles that discuss various issues on the George Massey tunnel replacement project were obtained. The collection of articles had an average of four paragraphs, had a title, and the content of the news article was structured.

The Latent Dirichlet Allocation (LDA) model was used to infer the topics in the collection of news articles. A Java-based package of the LDA model, MALLET (McCallum, 2002), was used. Nine distinct topics were identified. The number of topics were determined experimentally, by trying out other values and examining the topics output. The output of LDA model was used for document retrieval phase of the recommendation task using the LDI model (more details discussed in chapter 3, section 3.2).

5.2.5 Recommendation Process

The study entailed the use of two types of highlights (a) highlights that represent the learners' interaction with the learning material (b) ensemble of highlights for the outcome of the educational task. The highlights from (a) would be used as input to the recommender system, while the highlights from (b) would be evaluated with respect to the research questions (see section 5.3). Having multiple purposes for the highlights annotation, the need to differentiate between the highlights created is imperative. Tags were therefore used to differentiate between the types of highlights.

Two tags are associated with the highlight annotations for this study. When a learner selects a portion of text for highlighting, a pop-up menu appears, from which the learner chooses the appropriate tag for the highlight: either the highlight is meant for recommendation or to answer the question. All the events and activities performed by the participants are recorded and logged by nStudy. Also, the highlights events for recommendation are saved and sent to the recommender system in real time. Whenever the learner feels to request for further articles, s/he clicks on a "Request for recommendation" button. The recommended list of articles opens in a new tab browser and contains a list of articles that are topically similar and related to the highlights the participants created. The document retrieval process is based on the Indexing by Latent Dirichlet Allocation (LDI) model by Wang et al., (2010) discussed in chapter 3.

5.2.6 Presentation of Recommendation

When the list of articles for recommendation has been computed by the system, the recommended list of articles together with instructions to guide the learner is opened on a new tab on the opened browser the learner makes use of (see Figure 5.5). Only the top-N articles are presented to the learner at each time; where $5 \geq N \leq 7$ – his range for the number of articles a recommendation list should contain was decided on primarily due to the corpus size. A ranking function was used to decide the top-N articles to be presented. Based on the assumption that the metacognitive activities of a learner may include (a) multiple events such as multiple highlights, notes, tags (b) the metacognitive activities span different topics/themes, the ranking function first classifies the recommendations generated into topics inferred from the metacognitive activities of the learner. Second, the ranking function orders the recommendation based on the ratio of highlights to topics (i.e. mapping each highlight to a topic). Lastly, the number of topics inferred from the learners' metacognitive activities together with the ratio of highlights associated with each topic determines the number of articles from each topic to be presented.

For example, a learner makes ten highlights events ($E = 10$) for recommendation during the study session, and three topics T_1, T_2, T_3 are inferred from the learner's highlights; where five of the highlights are associated with T_2, three are associated with T_1 and two highlights with T_3. The ranking function then orders the recommended list based on the ratio of the topics: T_2, T_1, T_3. The number of articles per topic to be displayed in the top-N presented is done by (i) dividing the total number of highlights events E by 5 (the minimum value of N), (ii) the quotient from the division is used to further divide the number of highlights associated with the topics inferred. The result of the computation is rounded up to the nearest whole number and used to determine the number of articles per topic among the top-N articles to be presented. Therefore, going with the example, the recommended list of articles that would be presented to the learner would be ranked as following: three articles from T_2, two articles from T_1, and one article from T_3. Here, we assume again that any article from the topics would be relevant to the learner's highlight and without having to perform (key)word mapping.

For each the recommended article in the list, a dynamic summary is included, aiming to display the portion of the article that has the most utility with respect to the learners' information need inferred from the highlights. Due to reasons such as topic overlap (an article having more than one significant topic) and reoccurring similarity of the key terms from the learners' highlights and the topic-words, an article may be recommended to the learner more than once. If this occurs, the article would be visually marked to indicate that it has been previously seen and/or operated on.

It has also been recognized that many recommender systems functioned as black boxes, providing no transparency into the working of the recommendation process, nor offering any additional information to accompany the recommendations beyond the recommendations themselves (Herlocker et al., 2000). Therefore, the inclusion of an explanation component aims to justify why an item is recommended. Explanations in this light provides transparency, exposing the reasoning behind a recommendation. We included explanation together with the presentation of the recommended list of articles. The explanation consists of a set of keywords extracted from the learners' metacognitive activities, based on which the recommender system provided the recommendations.

5.3 Evaluation

The standard evaluation measures used to assess recommender systems can be split into three categories: online metrics, offline metrics and user feedback. The online metrics assesses how the learner interacts with the recommendation. For example, the metric *click-through rate* would be used to measure the number of articles the learner clicks/reads from the list of recommended articles, *session success rate* would be used to measure the number of articles the learner operated on for the completion of the task. The offline metrics measures the relevance of the recommended articles to the information needs of the learner, such as precision, recall, F-score. The user's feedback on the recommender system is obtained by administering questionnaires at the end of a user study session.

In the context of recommendation systems, recommending top-N items to the user is a widely used approach. Therefore, similar to modern information systems, a more useful metric is precision and recall metrics of the first N items instead of all the items. Thus the notion of precision and recall at k where k is a user definable integer that is set by the user to match the top-N recommendations objective; where precision is the proportion of recommendations that are good recommendations, and recall is the proportion of good recommendations that appear in top recommendations. Computing precision and recall values requires a gold standard or ground truth judgment of relevance document collection (e.g. TREC), which contains large number of queries and binary classification of documents that are relevant or non-relevant with respect to the query.

Information retrieval/filtering methods (including recommender systems) typically map a query to a ranked list of retrieved documents, however, our approach to recommendation maps multiple "queries" (expressed in the learner's metacognitive activities) to an ordered list of retrieved articles. Therefore, the standard offline metrics (precision, recall) as well as the gold standard data collections are not suitable for evaluating the recommender system. We therefore limit our evaluation to online metrics and user feedback. These metrics have been combined into a user-centric framework for evaluating recommender systems by Knijnenburg et al., (2012) depicted in figure 5.6.

The framework provides insight into the relationships between the general concepts that play a role in the user experience of recommender systems and consists of six interrelated conceptual components:

Objective System Analysis: The Objective System Aspects (OSAs) are the aspects of the system that are to be evaluated such as the algorithm, input mechanism, and output/presentation mechanism.

Subjective System Analysis: The Subjective System Analysis (SSAs) are regarded as the

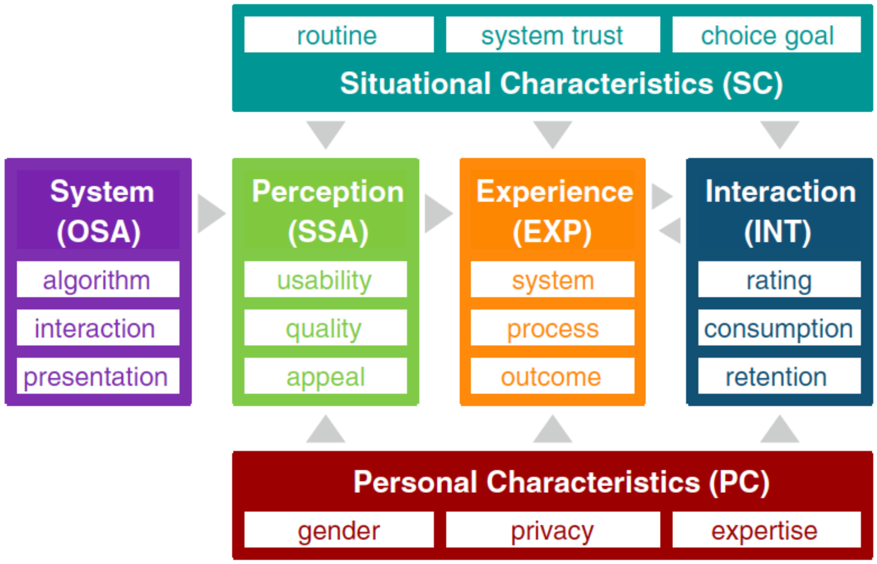


Figure 5.6: User-centric evaluation of recommender systems by Knijnenburg et al., (2012).

mediating variables that attempt to explain the effects of the OSAs on the user experience and interaction. SSAs are measured using questionnaires administered to the participants during or after interacting with the recommender system. SSAs measurements help establish whether the users perceive aspects of the OSA, independently of their evaluation of the aspect. For example, if an improvement made to a recommender system does not lead to the expected increase in user satisfaction, SSA measurements can be used to find out if users simply did not notice the improvement, or if they noticed it but did not really like it. SSAs therfore, mediate the effects of OSAs on the user experience (Knijnenburg and Willemsen, 2015).

User Experience: The user experience (EXP) are the users' self evaluations of the effectiveness of the different aspects of the recommender system such as the system's usefulness, appropriateness of the recommended items. User experience is also measured with questionnaires.

User Interaction: The user interaction (INT) factors objectively measure how the user interacted with the system. This can be done by e.g. logging the users' clicks, number of recommended items inspected, their rating feedback.

Personal and Situational Characteristics: The personal characteristics (PCs) and situational characteristics (SCs) of a user are factors that can influence the users' evaluation of the SSAs, EXP and INT with the recommender system. PCs such as domain knowledge, gender have been shown to affect SSA measures, also SCs (e.g. trust in technology) have been shown to effect the INT variables (Knijnenburg et al., 2012).

The evaluation framework was adapted as a conceptual guideline for developing hypotheses investigated. The variables measured (discussed in section 5.3.1) was done by a post-experiment questionnaire; where the question items was based on the evaluation framework as designed by Knijnenburg et al., (2012).

5.3.1 Variables for Analysis

The independent variable and dependent variables measured by the questionnaire are outlined below.

5.3.1.1 Independent Variable

The independent variables are the objective aspects of the system which would be manipulated in the different experimental conditions discussed in section 5.2.1, and allow us to measure the differences in outcomes between the conditions.

System's Methodology
Using the variations in the experimental conditions in this study, we seek to investigate the effect of the recommender system's methodology: (a) preference elicitation method – learners' highlights (b) document retrieval method – LDI model, on the dependent variables discussed below.

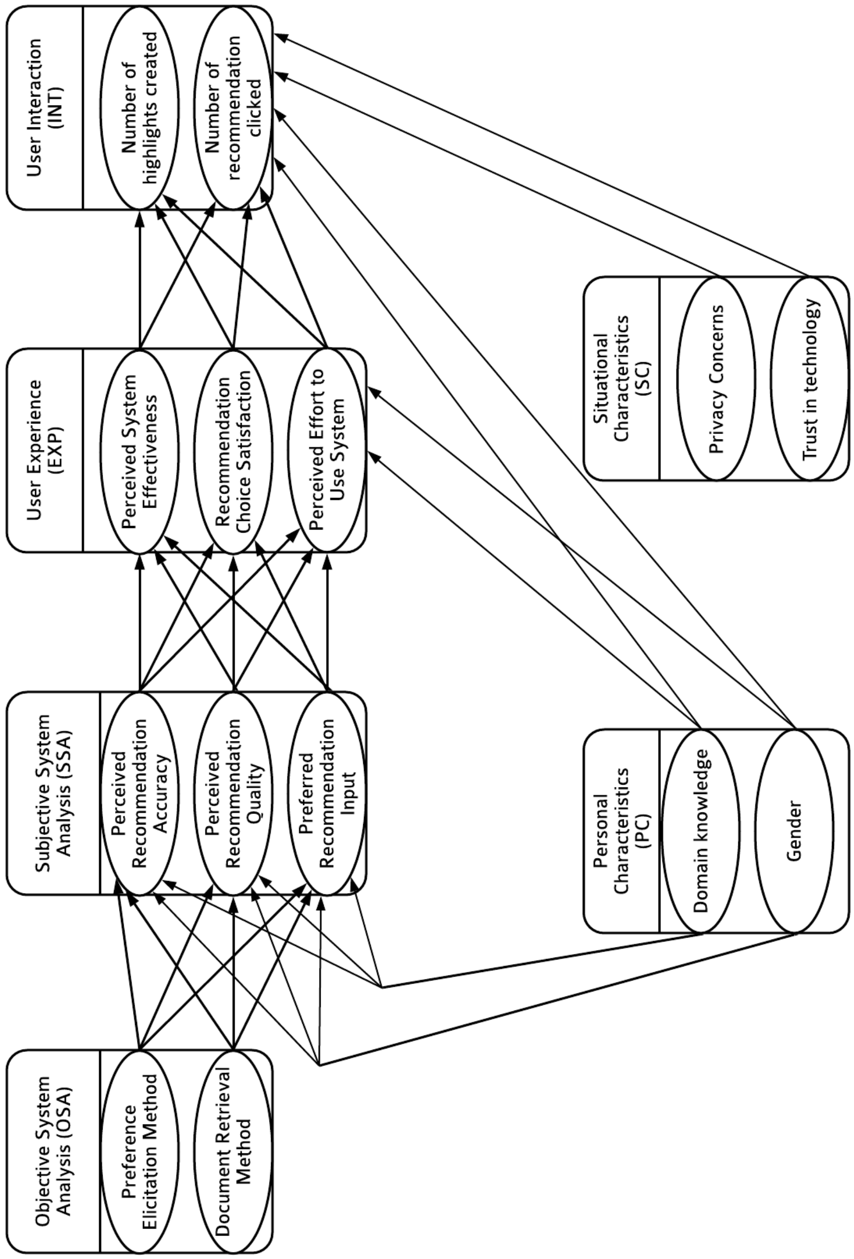
5.3.1.2 Dependent Variables

The dependent variables are intended to measure the user interaction (INT) and user experience (EXP). A combination of the observed behavior of the participants with the recommendations, and the subjective analysis of the system from the users' point of view (obtained from the post-experiment questionnaire) would be used. As stated by Knijnenburg et al., (2012) it may not be possible to explain the effects of the independent variables on the dependent variables directly. For example, if it is observed that participants are more satisfied or behave differently between experimental conditions, these observations may need to be justified (Knijnenburg et al., 2010). The authors have shown that the user experience and user interaction with a system may be explained using a number of mediating variables (SSAs), which could be analyzed to measure the effect of the different experimental conditions (manipulation of the system) on the user experience and user interaction. The mediating variables explain how and why the effects come about.

Figure 5.7 shows the concepts (OSA, SSA, EXP, INT, PC and SC) of the procedure as well as the variables measured for each concept. The arrows show relations between the concepts or variables.

The mediating variables can be used to test the hypothesized effect of the independent variables (does the objective analysis of the system correlate with the subjective analysis of the system) and also provide explanations for the effects between experimental conditions. The mediating variables for this study are:

Figure 5.7: Concepts, variables and hypothesized relationships for the user-centric evaluation of the recommender system, based on Knijnenburg et al., (2012) framework



A. Perceived recommendation quality: Is a subjective measure of the relevance of the recommendations the system provides. It measures the participants' perception of the quality of the recommendations received, for the different methodologies used in the experimental conditions. The participants' perceptions of the recommendation quality may be useful in predicting the effects of the recommender system's objective aspects on the user experience and user interaction.

B. Perceived recommendation accuracy: Measures subjectively how the system is able to provide recommendations (using the LDI model) that fit the information seeking needs of the participants, as well as recommend appropriate articles for the completion of the educational task. Similar to the perceived recommendation quality variable, perceived recommendation accuracy can be used to test the effects on the independent variables, dependent variables as well as the study outcome. In this case, perceived recommendation accuracy can also be used to predict the effects of the recommender system's aspects on the user interactions: number of recommended articles clicked, number of recommended articles operated on.

C. Preferred elicitation method: Measures the participants' perceptions on the preference elicitation method used. This variable is also assumed to be also useful in investigating the effects of the system's methodology on the users' experience and interaction.

The variables to measure user experience are obtained from the questionnaire, they are:

- Perceived system effectiveness: Measures the usefulness of the recommender system in providing valuable personalized recommendations to the participants.
- Recommendation choice satisfaction
- Effort to use the system

The variables to measure user interaction are obtained from the logged data of the participants' engagements with the recommendations, they include:

- Number of recommended articles clicked
- Number of recommended articles operated on

As aforementioned, a number of other variables (commonly referred to as personal and situational characteristics) have been shown to affect the users' subjective analysis of the system as well as the user experience (Knijnenburg et al., 2012). These variables are obtained from the questionnaire administered to the participants of the study.

- Domain knowledge: Measures how the prior knowledge of the participant influences perceived recommendation accuracy, perceived recommendation quality and user experience.

– Gender: Investigates whether gender as a PC influences the user perception of the recommendation accuracy, recommendation quality, user interaction and user experience.

– Trust in technology: Measures how issues relating to the participants' trust in technology influences perceived recommendation accuracy, perceived recommendation quality and user experience.

– Privacy concerns: Measures how issues relating to the participants' privacy with using the system and/or providing data to the system influences perceived recommendation accuracy, perceived recommendation quality and user experience.

5.4 Research Questions and Hypotheses

Based on the proposed methodology for recommendation of textual documents, we seek to investigate and address the following research questions in the user study:

1. Is the learner-generated highlight appropriate data for the learner model, and suitable as a preference elicitation method for recommendation? The use of highlights to guide recommendation is novel. Therefore, we seek to determine if it is an appropriate input mechanism for recommendation. To answer this question, we hypothesize the following:

   H1.1: There is a significant difference in the preference elicitation method between the two group conditions.

   H1.2: The preference elicitation method has a positive effect on the user experience.

2. What components of the framework affect the user experience? We seek to examine the effects of the subjective system aspects, personal characteristics and situational characteristics of the evaluation framework on the user experience. To answer this question, the effects of these components are examined individually.

   2.1: Subjective system aspects on user experience

   2.2: Personal characteristics on the user perception and user experience.

   2.3: Situational Characteristics on the user perception and user experience.

3. What components of the framework affect the user interaction? This research questions examine the effect(s) of the preference elicitation methods, user experience, personal characteristics, and situational characteristics on the user interaction. To answer the question, the effects of these components are examined individually.

   3.1: Effects of the user experience on user interaction.

   3.2: Effects of personal and situational characteristics on user interaction.

3.3: Effects of the preference elicitation method on user interaction.

5.4.1 Data Sources

To provide answers to the research questions and hypotheses of the study, three data sources from (and about) the user are used and analyzed.

1. Logged data
   The logged data comprises the observable behavior and activities of the learner during the study. The behavior of the learner that would be analyzed in this study are: the number of recommended articles the participant clicks, and the number of articles the participant operated on (e.g. creates highlights, notes, tags). Also, the content of the metacognitive activities of the participants would be used for data analysis (e.g. the text highlighted, the text of the tag created). nStudy, the learning platform being used for the study records very fine grained time stamps of all the activities performed by the learner. Therefore, the logged data nStudy records for each participant would be used.

2. Questionnaires
   Before the reading activities the participants are administered a demographics questionnaire. Also, upon completion of the tasks, another (feedback) questionnaire is administered, used to obtain the participant's feedback about the recommender system. Both questionnaires are administered online and is intended to obtain some personal information of the participants as well as to measure each participant's experience and a number of subjective analysis of the recommender system (e.g perceived recommender quality).

3. Study outcome
   The study outcome is an ensemble of highlights each participant puts together to answer the educational tasks. With the study outcome, it would be possible to assess the educational significance of the study. However, this study is intended as a "proof of concept" that the system works, as judged by the users before investigating what and how much learners using your system actually benefit by using it.

5.5 Statistical Evaluation

We adopted a pragmatic procedure for the recommender system's evaluation that is based on the framework by Knijnenburg et al., (2012). This entails analyzing the data by testing the significance and size of a subset of the effects. For each hypothesis, the data is analyzed to determines whether there is a statistically significant difference between the means in two experimental groups using independent t-test. Also, the correlation of the variables (and concepts), and effect size of each effect is also computed.

In the next chapter, we discuss the findings and results obtained from the data analysis in relation to the research questions and hypotheses.

Chapter 6

Results and Findings

This research involved the design and development of a recommender system to support TEL using a novel feature – the learners' highlights, a metacognitive reading activity – to guide recommendation. The main questions investigated are: (1) Is the learner-generated highlight appropriate data for the learner model and preference elicitation method for recommendation? Which entails investigating three components: subjective system aspects, personal characteristics and situational characteristics (2) What components of the framework affect the user experience? (3) What components of the framework affect the user interaction? Also entails examining four components: user experience, preference elicitation method, personal characteristics and situational characteristics.

The post-experiment questionnaire contained thirty questions targeted at eight categories: perceived recommendation quality, perceived recommendation accuracy, preferred recommendation input method, perceived system effectiveness, recommendation choice satisfaction, perceived effort to use the system, personal characteristics, and situational characteristics. To determine if the questions fit the categories intended, exploratory factor analysis with Varimax rotation was performed to extract factors from the observed variables.

To test the internal consistency of the items, each of the categories, Cronbach's alpha reliability test was performed, which measures how closely related the set of items are as a group. Due to space limitations, Table 6.1 shows a sample of the factor analysis loadings for the questions of the questionnaire, and Appendix A contains the complete table. The results of Cronbach's alpha reliability tests reveals that the items in the categories/factors are internally consistent; where a reliability coefficient of 0.70 or higher is considered "acceptable." To answer the research questions and test the hypotheses of the study, we adopted a pragmatic procedure which entails computing independent sample T-test (statistics), p-value (significance), effect size tests (Cohen's d), and regression analysis for each effect. We discuss the results and findings in the next sections.

Table 6.1: Sample of the constructs and their measurement.

Question	**Variable**	**Loading**
Recommendation Quality (Cronbach's alpha: 0.920)		
I liked the articles recommended	like_articles	0.848
The system provided valuable recommendations	valuable_articles	0.861
The system had too many irrelevant recommendations	too_many_irrelevant_articles	0.792
I didn't like any of the recommended articles	didn't_like_articles	0.804
Recommendation Accuracy (Cronbach's alpha: 0.879)		
The articles were well chosen based on my highlights	well_chosen articles	0.811
The recommended articles were relevant to completing the task	relevant_articles	0.893
I would give most of the articles recommended a high rating	high_rating_articles	0.846
The list of recommended articles was appealing	articles_appealing	0.799
Preferred Elicitation Method (Cronbach's alpha: 0.813)		
I prefer to use highlights for recommendation highlights	prefer_highlighting	0.823
I would have preferred typing search queries for recommendation	prefer_typing	0.914
Recommender System Effectiveness (Cronbach's alpha: 0.905)		
I would recommend the system to others	recommend_system	0.823
The system made me aware of my highlights completing the task	relevant_articles	0.742
I can find better articles using the recommender system	high_rating_articles	0.895
I can find better articles without the recommender system	articles_appealing	0.806
Recommendation Choice Satisfaction (Cronbach's alpha: 0.798)		
I enjoyed reading the articles I selected	enjoyed_reading articles	0.938
The articles were appropriate for the task	task_appropriate articles	0.892
The chosen articles fit my preference	articles_fit_my preference	0.837
The articles I read were a waste of my time	articles_waste of_time	0.826

Variable	Group A Mean_responses	Group B Mean_responses	T-Values
prefer_highlighting	4.58	2.65	$t(47) = 3.57$ $p < 0.001$ $d = 1.540$
prefer_typing	2.28	3.86	$t(47) = -1.75$ $p = 0.006$ $d = 1.332$

Table 6.2: Mean responses, t-test results of the preference elicitation method variables

6.1 Is the learner-generated highlight appropriate data for the learner model, and suitable as a preference elicitation method for recommendation?

Given that the use of highlights (learners' metacognitive activities) to guide recommendation is novel, this research question therefore attempts to investigate if it is an appropriate input mechanism for recommendation. It is important to note here that two types of highlights/text marking was used in the study: highlights for recommendation and highlights to answer the task/question. To differentiate between the highlight types, after the learner selects a portion of text to highlight, a pop-up menu containing two tag options is opened. The learner then labels the highlight with the appropriate tag as intended. The set of highlights analyzed to answer this research question are the *highlights for recommendation*.

The participants received no training on how to create highlights using the learning platform nStudy, and were not restricted on the number of highlights to be created nor the number of times they could request for recommendations. From the data analyzed, for both group conditions (experimental group: Group A, and control group: Group B), the average number of highlights created was 34.76 for participants in Group A, and 11.52 for participants in Group B. The average number of words a highlight consisted of 11.8 words. The participants in Group A requested recommendation on average 2.7 times while participants in Group B requested for recommendations an average of 5.3 times. Further analysis on the content and features of the subsequent highlights participants made showed that the highlights got "better" after each iteration of recommendation requested. To judge what a *good* highlight is, we came up with three qualities: (a) the highlight should contain keywords that are related to task (b) the highlights are among the pre-identified texts in the articles that the researcher had identified to generate optimal recommendations to complete the task (c) the highlight contained less than 15 words and does not span multiple sentences. The threshold for the number of words is based on the assumption that search queries typed contains an average of 8 words and do not exceed 15 words (Hurn, 2009).

Variable	**like_ articles**	**valuable articles**	**too_many irrelevant_articles**	**didn't_like _articles**	**prefer_ highlights**	**prefer_ typing**
like_articles	1					
valuable_ articles	0.87	1				
too_many_ irrelevant_ articles	-0.86	-0.90	1			
did't_like articles	-0.90	-0.86	0.84	1		
prefer_ highlights	**0.86**	**0.79**	**-0.74**	**-0.72**	1	
prefer_ typing	**-0.79**	**-0.85**	**0.84**	**0.76**	-0.82	1

Table 6.3: Correlation coefficients of the preference elicitation method and perceived recommendation quality variables

The responses to the feedback questionnaire also showed that 92.3% of the participants in Group A reported to prefer the use of highlights to guide recommendations, while 51.8% of the participants in Group B noted their preference to use search queries to guide recommendation. To explain the difference in the preference elicitation method between the two groups, as well as examine the effect (if any) on the user experience, and user interaction, a number of hypotheses were tested (discussed below) using independent sample t-test, regression analysis and correlation tests. It is important to note that based on our assumptions regarding the participants' familiarity and challenges with search queries (discussed in section 1.3), the experimental design of this study did not include a treatment condition to specifically test the use of search queries. Therefore the outcome of the results may not validate the effectiveness of using highlights vs. typing a search query.

- **H1.1:** *There is a significant difference in the preference elicitation method between the two group conditions.* This hypothesis tests whether there is a significant difference between the means of the two groups. Two items were used to collect the participants preferences. Table 6.2 shows the mean responses to the items measuring this concept.

  Results of the t-test also show that this difference is significant with a large effect size[1] for both variables *prefer_highlighting*: [$t(47) = 3.57$, $p < .001$, $d = 1.54$] and

[1]A typical threshold for significance is p < .05, meaning that the chance of incorrectly rejecting the null hypothesis of no effect is smaller than 5%. Accepted interpretations of effect size are: small/weak effect: d=0.2, medium effect: d=0.5, large/strong effect: d≥0.8

Variable	well_ chosen	relevant articles	high_rating articles	articles_ appealing	prefer_ highlights	prefer_ typing
well_chosen_articles	1					
relevant_ articles	0.80	1				
high_rating_articles	-0.87	-0.92	1			
articles_ appealing	-0.82	-0.80	0.77	1		
prefer_ highlights	**0.82**	**0.79**	**0.75**	**0.80**	1	
prefer_ typing	**-0.58**	**-0.73**	**-0.65**	**-0.72**	-0.79	1

Table 6.4: Correlation coefficients of the preference elicitation method and perceived recommendation accuracy variables

prefer_typing[2]: [$t(47)$ = -1.75, p = .006, d = 1.33]. Therefore, we accept the hypothesis that there is a significant difference in the preference elicitation method between the 2 group conditions; where participants in the experimental group (Group A) had a greater preference for the use of highlights to guide recommendation, and participants in the control group (Group B) had a greater preference for using search queries to guide recommendations.

A possible explanation for the results obtained could be due to the (perceived) quality and/or accuracy of the recommendations received. Group B participants received random recommendations, which implies that the highlights they created weren't taken into account nor used in the recommendation generation process, and otherwise for participants in Group A. To verify this assumption, we performed a Pearson correlation test to determine if the preferred elicitation method is related to the perceived recommendation quality (see Table 6.3) and recommendation accuracy (Table 6.4).

Four variables were used to measure the perception of the participants regarding the recommendation quality coded [3] as: *like_articles*, *valuable_articles*, *irrelevant_articles*, and *didn't_like_articles*, which when correlated with the preferred elicitation method variables reveals that the two concepts are very strongly and significantly correlated ($|r| > .7$, $p < .001$). *like_articles* and *valuable_articles* variables are positively related

[2]A negative t-value shows a reversal in the directionality of the effect being studied, but has no impact on the significance of the difference between groups of data.

[3]Table 6.1 shows a sample of the questionnaire items, and Appendix A shows the full questionnaire item wordings and their abbreviations.

Variable	Group A Mean_responses	Group B Mean_responses	T-Values
like_articles	4.30	2.37	t(47) = 3.42, p = .001, d = 1.45
valuable_articles	4.45	2.46	t(47) = 3.55, p = .001, d = 1.47
too_many_irrelevant articles	1.76	3.71	t(47) = -2.33, p < .001, d = 1.11
didn't_like_articles	1.40	3.38	t(47) = -1.91, p < .001, d = 1.48

Table 6.5: Mean responses and t-test results of the perceived recommendation quality variables

to *prefer_highlighting*, while *irrelevant_articles*, and *didn't_like_articles* variables are negatively related to *prefer_highlighting* with correlation coefficients ($r = 0.86$, $p < .001$), ($r = 0.79$, $p < .001$), ($r = -0.74$, $p < .001$) and ($r = -0.72$, $p < .001$) respectively. The opposite relationship is obtained for the *prefer_typing* variable, where *like_articles* and *valuable_articles* variables are negatively related to *prefer_typing*, and *irrelevant_articles*, and *didn't_like _articles* variables are positively related to *prefer_typing*, with correlation coefficients ($r = -0.79$, $p < .001$), ($r = -0.85$, $p < .001$), ($r = 0.84$, $p < .001$) and ($r = 0.76$, $p < .001$) respectively.

The recommendation accuracy concept is also measured by four variables coded as: *articles_well_chosen*, *articles_task_relevant*, *articles_high_rating*, and *rec_list_appealing*. Correlating the variables with the preference elicitation concept variables: *prefer_highlighting* and *prefer_typing*, the results showed that indeed the two concepts are also very strongly and significantly correlated ($|r| > .5$, $p < .001$). The four recommendation accuracy concept variables were strongly positively and significantly correlated with *prefer_highlighting* with correlation coefficients ($r = 0.82$, $p < .001$), ($r = 0.79$, $p < .001$), ($r = 0.75$, $p < .001$) and ($r = 0.80$, $p < .001$) respectively, and strongly negatively correlated with *prefer_typing* with correlation coefficients ($r = -0.58$, $p < .001$), ($r = -0.73$, $p < .001$), ($r = -0.65$, $p < .001$) and ($r = -0.72$, $p < .001$) respectively.

Regression analysis is also performed to determine which variable of the preferred elicitation method contributes to the perceived recommendation quality and accuracy. The output of the regression model suggests that the *prefer_highlighting* variable alone contributes (prefer_typing variable is excluded from the model) to both perceived recommendation quality and accuracy [$R^2 = .526$, $F(1, 47) = 51.20$, $p < .001$] and [$R^2 = .375$, $F(1, 47) = 43.08$, $p < .001$] respectively. The results obtained from the analyses confirms the assumption that the difference in the preferred elicitation method between the two groups is as a result of the perceived recommendation quality and accuracy. That is, the participants in Group B's preference for typing search queries as opposed to using highlights to guide recommendation is linked to the quality and accuracy of the recommendations they received.

– **H1.2**: *The preference elicitation method has a positive effect on the user experience.* This hypothesis assumes that the preference elicitation method influences the user experience. For example, if a user is dissatisfied (or otherwise) with the use of highlights to guide recommendation, it could affect his/her experience with, and evaluation of the recommender system. Knijnenburg et al., (2011) suggest that the user experience measurements should distinguish the evaluation objects of the recommender system: the process, outcome, and system itself. *perceived effort to use system* is the process-related experience concept that assesses the effort and time required to operate the system; *perceived system effectiveness* is the system-related experience concept that measures the users' evaluation of the recommender system's effectiveness, and *recommendation choice satisfaction* is the outcome-related experience concept that measures the users' satisfaction with the chosen items read.

  To determine whether the user experience is influenced by the preference elicitation method, correlation and regression analysis were conducted to examine the relationship between the concepts that measure the user experience and the preference elicitation method. Pearson correlation test results shown in Table 6.11 reveal that the preference elicitation method concepts are significantly related to the user experience concepts ($|r| \geq .45, p \leq 0.003$). More specifically, the variables measuring the user experience concepts are positively related to the *prefer_highlights* variable, and negatively correlated to the *prefer_typing* variable. Table 6.6 summarizes the results of the regression analysis. The *prefer_highlighting* variable was observed to have contributed to two of the three concepts that measured the user experience (*recommendation choice satisfaction* and *perceived system effectiveness*) with coefficients $R^2 = .542$, $F(1, 47) = 38.82$, $p = .001$, and $R^2 = .317$, $F(1, 47) = 15.05$, $p < .001$ respectively; while the *prefer_typing* variable was excluded from the model. However, neither *prefer_highlighting* nor *prefer_typing* variables contributed to the *effort to use system* concept.

  As earlier mentioned, the *effort to use system* concept assesses the system's *process* of generating recommendations, while *recommendation choice satisfaction* assesses the outcome – recommendations produced and *perceived system effectiveness* evaluates the system. Therefore, it is understandable that the elicitation method may not nec-

User experience concepts	Variable	R^2	B	β	*Sig.*
Preference Elicitation Method					
recommendation_ choice_satisfaction	prefer_highlighting	0.542	0.037	0.67	.000
perceived_system _effectiveness	prefer_highlighting	0.317	0.031	0.46	.001
Perceived Recommendation Quality					
recommendation_ choice_satisfaction	like_articles	0.772	0.068	0.791	.000
	valuable_articles		0.055	0.638	
perceived_system _effectiveness	like_articles	0.469	0.076	0.553	.000
	valuable_articles		0.61	0.457	
Perceived Recommendation Accuracy					
recommendation_ choice_satisfaction	well_chosen_articles	0.732	0.059	0.643	.000
	relevant_articles		0.055	0.594	
	high_rating_articles		0.046	0.476	
	appealing_articles		0.172	0.438	
perceived_system _effectiveness	well_chosen_articles	0.211	0.035	0.439	.001
	relevant_articles		0.066	0.421	
	high_rating_articles		0.051	0.372	
	appealing_articles		0.010	0.182	

Table 6.6: Regression weights of the subjective system aspect predictors method on the user experience

Variable	Group A Mean_responses	Group B Mean_responses	T-Values
well_chosen articles	4.32	2.38	t(47) = 3.32, p = .005, d = 1.62
relevant_ articles	4.44	2.50	t(47) = 2.85, p < .001, d = 1.26
high_rating articles	4.14	2.54	t(47) = 2.37, p < .001, d = 1.85
articles_ appealing	4.08	2.42	t(47) = 2.00, p < .001, d = 1.71

Table 6.7: Mean responses and t-test results of the perceived recommendation accuracy variables

Variable	like articles	valu -able articles	irrele -vant articles	didn't_ like articles	well_ chosen articles	relev -ant articles	high_ rating articles	appea ling articles
like articles	1							
valuable articles	0.87	1						
irrelevant articles	-0.86	-0.90	1					
didn't_like articles	-0.90	-0.86	0.84	1				
well_chosen_articles	0.80	0.77	-0.78	-0.76	1			
relevant articles	0.78	0.83	-0.81	-0.74	0.80	1		
high_rating_articles	0.72	0.76	-0.79	-0.70	0.77	0.87	1	
appealing articles	0.81	0.84	-0.80	-0.85	0.72	0.69	0.67	1

Table 6.8: Correlation coefficients of the perceived recommendation quality and perceived recommendation accuracy variables

essarily influence or affect the process by which the system generates recommendation, the process by which the system generated recommendations rather has to do with the underlying algorithms and structure of the recommender system. However, for the outcome-related concept, *recommendation choice satisfaction*, which is directly influenced by the elicitation method, and *perceived system effectiveness* concept, the regression analysis indicates that the *prefer_highlighting* variable has a significant positive regression weight. Therefore, we conclude that the use of highlights to guide recommendations had a positive effect on some aspects of the user experience.

6.2 What components of the framework affect the user experience?

This research question examines the effects of the subjective system aspects, personal characteristics and situational characteristics of the evaluation framework on the user experience.

6.2.1 Subjective System Aspects on User Experience

The subjective aspects of the system provides an evaluation of the objective aspects (e.g. algorithm) as perceived by the users. It also provides explanations on the resulting user experience (Knijnenburg et al., 2011). For example, does a high perceived recommendation quality lead to enhanced user experience? To examine the effect, using the hypotheses below, we examine the relationship between the two concepts (subjective system aspect and user experience), compute independent t-tests on the two group conditions, and perform regression analysis to identify the predictors.

- **H2.1**: *There is significant difference in the perceived recommendation quality between the two group conditions.* To determine whether participants of the two groups judge the recommendation quality differently, independent sample t-test is also performed on the four variables that were used to measure the participants' perceptions on the recommendation quality (*perceived recommendation quality*). This test is intended to assess the algorithm behind the recommendations used in the two group conditions are significantly different; where participants in Group A received recommendations based on their metacognitive activities and the LDI model, and participants in Group B received random recommendations. More specifically, the variables measuring the concept *perceived recommendation quality* focused on assessing if the articles generated for recommendation were appropriate and relevant to completing the task. Table 6.5 shows the mean responses to the variables measuring this concept for both groups.

  The results of the t-test shows that this difference is significant with a large effect size for the four variables measuring the concept: *like_articles* $[t(47) = 3.42, p = .001, d = 1.45]$, *valuable_articles* $[t(47) = 3.55, p = .001, d = 1.47]$, *irrelevant_articles* $[t(47) = -2.33, p < .001, d = 1.11]$, and *didn't_like_articles* $[t(47) = -1.91, p < .001, d = 1.48]$.

- **H2.2**: *There is significant difference in the perceived recommendation accuracy between the two group conditions.* Similar to the hypothesis above, independent sample t-test is performed to further determine if the algorithm behind the recommendations used in the two group conditions are significantly different; where participants in Group A received recommendations based on their metacognitive activeness and the LDI model, and participants in Group B received random recommendations. The variables measuring the *perceived recommendation accuracy* concept measure the participants' ratings of the recommendation they received.

  The results of the t-test and the mean responses on the participants' perception on the recommendation accuracy is shown in Table 6.7. The t-test result reveals that the difference in the perceived recommendation accuracy between the two group conditions is significant with a large effect size for the four variables measuring the concept: *well_chosen_articles* $[t(47) = 3.32, p = .005, d = 1.62]$, *relevant_articles* $[t(47) = 2.85, p < .001, d = 1.62]$, *high_rating_articles* $[t(47) = 2.37, p < .001, d = 1.85]$, and *appealing_ articles* $[t(47) = 2.00, p < .001, d = 1.71]$.

Concept	Variable	Group A Mean	Group B Mean	T-test Values
recommendation choice_satisfaction	enjoyed_reading articles	4.36	2.63	t(47) = 4.37, $p < .001$, d = 2.39
	task_appropriate articles	4.64	2.42	t(47) = 3.01, $p < .001$, d = 2.66
	articles_fit_prefer ence	4.12	2.58	t(47) = 2.40, p = .003, d = 1.85
	articles_waste_of time	1.76	3.54	t(47) = -4.21, p = .001, d = 2.21
recommender_system effectiveness	recommend_sys tem	4.92	3.83	t(47) = 3.10, p = .341,
	sys_aware_highli ghts	4.28	2.50	t(47) = 2.42, p = .126,
	sys_good_articles	4.32	2.59	t(47) = 4.62, $p < .001$, d = 2.43
	I_can_find_better articles	1.28	3.97	t(47) = -2.36, p = .001, d = 1.71
effort_to_use_system	sys_easy_to_use articles	4.52	4.39	t(47) = 3.32, p = .274,
	invest_lot_of effort	1.92	2.42	t(47) = 2.85, $p < .001$, d = 1.54
	recs_take_too_ much_time	1.78	1.58	t(47) = 2.37, $p < .111$,

Table 6.9: Mean responses and t-test results of the user experience concepts and variables

Variable	Perceived Recommendation Quality			
	like_articles	valuable article	irrelevant article	didn't_like article
Recommender System Effectiveness				
recommend_system	0.89	0.88	-0.69	-0.51
sys_aware_highlights	0.79	0.74	-0.58	-0.64
sys_good_articles	0.82	0.88	-0.73	-0.79
I_can_find_better_articles	-0.71	-0.69	0.77	0.71
Recommendation Choice Satisfaction				
enjoyed_reading_articles	0.88	0.87	-0.81	-0.83
task_appropriate_articles	0.91	0.93	-0.92	-0.91
articles_fit_preference	0.89	0.84	-0.67	-0.63
articles_waste_of_time	-0.81	-0.80	0.74	0.71
Effort to Use System				
sys_easy_to_use	0.84	0.82	-0.75	-0.77
invest_lot_of_effort	-0.73	-0.76	0.64	0.70
recs_take_too_much_time	0.71	0.78	-0.66	-0.62

Table 6.10: Correlation coefficients of the perceived recommendation quality and the user experience variables

To determine whether the participants' perception of the recommender system's quality is related to the perceived recommendation accuracy, correlation and multiple regression analyses were performed on the variables measuring the two concepts. The results of the correlation test shows that both concepts are very strongly and significantly correlated with correlation coefficients of $|r| > .6$ and $p < .001$, while the regression model produces $R^2 = .82$, $F(1, 47) = 58.43$, $p < .001$ excluded *irrelevant_articles* and *didn't_like_articles* variables from being predictors . Table 6.8 contains more details of the analysis.

– **H2.3**: *The users' subjective system analysis of the recommender system affects the user experience.* Having established that there is a significant difference in the perceived recommendation quality, accuracy and preferred elicitation method (subjective system aspects) between the two group conditions, this hypothesis investigates how the perceptions result in specific user experience (*recommendation choice satisfaction*, *recommender system effectiveness*, *perceived effort to use system*). For example, whether a higher perceived recommendation quality leads to a higher satisfaction with the recommendations. Computing independent t-test on the concepts of the subjective system aspects and the user experience reveals that for the four variables that measure the *recommendation choice satisfaction* concept: *enjoyed_reading_articles*, *task_appropriate_articles*, and *articles_fit_preference*, *articles_waste_of_time*, there is a significant difference between the two groups; where the t-values for the variables are $[t(47) = 4.37, p < .001, d = 2.39]$, $[t(47) = 3.01, p < .001, d = 2.66]$, $[t(47) = 2.40, p = .003, d = 1.85]$, $[t(47) = -4.21, p < .001, d = 2.21]$ respectively. The analysis also revealed that two of the four variables that measure the *recommender system effectiveness* are significantly different *sys_good_articles* $[t(47) = 4.62, p < .001, d = 2.43]$, *I_can_find_better_articles* $[t(47) = -2.36, p = .004, d = 1.71]$ between the groups; and for one variable (*invest_lot_of_effort*) of the *perceived effort to use system* concept showed a significant difference between the two groups with t-values $[t(47) = 2.85, p < .001, d = 1.54]$. The variables of the *recommender system effectiveness* that weren't significantly different between the two groups are *recommend_system*, and *sys_aware_highlights*. This suggests that participants in both groups may have been satisfied with these aspects of the system despite the variation in the method behind the recommendations generated. The results further support the notion that the use of highlights, may be suitable to guide recommendation, and also facilitates metacognition. Similarly, the two variables (*system_easy_to_use* and *recs_take_too_much_time*) of the *perceived effort to use system* concept that weren't significantly different between the two groups reveals that the system is easy to use (where the mean responses from both groups are greater than 4.3), and it doesn't take too long to provide recommendation (the mean responses from both groups are less than 2.0).

In general, the results obtained from the correlation tests (shown in Tables 6.10 and 6.11) reveal that there is a strong and significant relationship between the variables representing the user experience concepts and the participants' perceived recommendation quality and accuracy, with correlation coefficients ($|r| \geq .50, p \leq 0.005$). The variables *system_easy_to_use* and *recs_take_too _much_time* are positively related to perceived recommendation quality and perceived recommendation accuracy concepts, while the *invest_lot_of_effort* variable is negatively related. Three of the four *perceived system effectiveness* variables *recommend_system*, *sys_aware_highlights*, and *sys_good_articles* are also positively related to the perceived recommendation quality and perceived recommendation accuracy concepts while the variable *I_can_find _better_articles* is negatively related. A similar relationship is observed for the *perceived choice satisfaction* variables; where variables *enjoyed_reading_articles*, *task_appro priate_ articles*, and *articles_fit_preference* are positively related, and the variable *articles_waste _of_time* is negatively related to the perceived recommendation quality and perceived recommendation accuracy concepts.

The result of the regression model reveals that only two of the four variables that measure *perceived recommendation quality* influence the *recommender system effectiveness* and *recommendation choice satisfaction* concepts of the user experience. The variables/predictors *like_articles* and *valuable_articles* produced $R^2 = .46$, $F(1, 47) =$

Variable	Perceived Recommendation Accuracy well_chosen articles	task_relevant articles	high_rating articles	appealing articles	Elicitation Method prefer highlights	prefer typing
Recommender System Effectiveness						
recommend system	0.87	0.83	0.73	0.83	0.71	-0.52
sys_aware highlights	0.82	0.73	0.76	0.84	0.84	-0.45
sys_good articles	0.78	0.80	0.75	0.81	0.76	-0.60
I_can_find better_articles	-0.65	-0.68	-0.71	-0.69	-0.68	0.51
Recommendation Choice Satisfaction						
enjoyed_reading_articles	0.77	0.74	0.72	0.86	0.75	-0.57
task_appropriate_articles	0.81	0.84	0.77	0.82	0.72	-0.55
articles_fit preference	0.66	0.77	0.73	0.69	0.65	-0.53
articles_waste of_time	-0.68	-0.70	-0.58	-0.73	-0.71	0.50
Effort to Use System						
sys_easy_to use	0.85	0.81	0.77	0.81	0.79	0.74
invest_lot_of effort	-0.62	-0.71	-0.67	-0.69	-0.71	-0.55
recs_take_too much_time	0.76	0.62	0.51	0.69	0.77	-0.51

Table 6.11: Correlation coefficients of the perceived recommendation accuracy, preference elicitation method and the user experience variables

64.39, $p < .001$ for *recommender system effectiveness*, and $R^2 = .77$, $F(1,47) = 53.18$, $p < .001$ for *recommendation choice satisfaction* concepts. All the variables that measure perceived recommendation accuracy were observed to be predictors of the regression model for the *recommender system effectiveness* concept which generated $R^2 = .21$, $F(1,47) = 18.25$, $p < .001$, and $R^2 = .73$, $F(1,47) = 82.91$, $p < .001$ for *recommendation choice satisfaction*. However, none of the variables measuring perceived recommendation quality and accuracy was discovered to influence the *effort to use system* concept. Table 6.6 shows the details of the regression analysis.

Therefore, based on the results of the analyses performed, we accept the hypothesis that the subjective system aspects positively affect some of the concepts that measure user experience. More specifically, for the three subjective system aspects the following was discovered: (a) *prefer_highlights* variable of the preference elicitation method influenced the *recommender system effectiveness* and *recommendation choice satisfaction* concepts, but did not influence the *effort to use system* concept (b) *like_articles* and *valuable_articles* of the *perceived recommendation quality* also influenced the *recommender system effectiveness* and *recommendation choice satisfaction* concepts, but did not influence the *effort to use system* concept (c) all the four variables of the *perceived recommendation accuracy* concepts influenced the *recommender system effectiveness* and *recommendation choice satisfaction* concepts, but also did not influence the *effort to use system* concept. This implies that none of the subjective system aspects contributed to the user's perceived effort to use the system.

6.2.2 Personal Characteristics on the user perception and user experience

This research questions examines if some of the personal characteristics of the users affect their perception of the quality, accuracy of the recommender system and/or the user experience. As popularly mentioned in literature, the personal characteristics of users that influence perception and user experience are *gender* and *domain knowledge*. These personal characteristics were obtained from the pre-experiment (demographics) questionnaire administered to the participants. From the data analyzed, a total 22 males and 27 females participated in the user study experiment. The questionnaire item used to collect information about the domain knowledge of the participants had three response categories, represented as: *novice*, *intermediate*, and *expert*. 91.8% of the participants indicated to be not familiar (novice) with the domain (source reading text), while for the intermediate and expert levels, 4.1% of the participants indicated to belong to each of the categories.

Given that the data obtained for each of the domain knowledge categories wasn't significantly sufficient, we are not able to assess the effects of the domain knowledge on the participants' perceptions of the recommendation quality, accuracy and user experience. Independent t-tests were performed to measure whether there is a significant difference between the perception of males and females on the recommendation quality, accuracy and

Table 6.12: Mean responses and t-test results of the personal characteristics of the users' on user experience and perception concepts and variables

Concept	**Variable**	**Male Mean**	**Female Mean**	**T-test Values**
recommendation quality	like_articles	3.39	3.31	t(47) = 1.23, p = .179,
	valuable_articles	3.52	3.42	t(47) = 1..67, p = .150,
	irrelevant articles	2.70	2.73	t(47) = -1.09, p = .157
	didn't_like articles	2.35	2.38	t(47) = -0.91, p = .381
recommendation accuracy	well_chosen articles	3.13	3.58	t(47) = 1.29, p = .169
	relevant articles	3.39	3.58	t(47) = 1.49, p = .284
	high_rating articles	3.21	3.38	t(47) = 1.58, p = .188
	appealing articles	3.34	3.00	t(47) = 1.01, p = .224
recommendation choice_satisfaction	enjoyed_reading articles	4.36	2.63	t(47) = 1.51, p = .375
	task_appropriate articles	4.64	2.42	t(47) = 1.09 , p = .193
	articles_fit_prefer ence	4.12	2.58	t(47) = 0.76, p = .245
	articles_waste_of time	1.76	3.54	t(47) = 1.21, p = .183
recommender_system effectiveness	recommend_sys tem	3.39	4.00	t(47) = 0..86, p = .341,
	sys_aware_highli ghts	4.09	3.88	t(47) = 1.73, p = .126
	sys_good_articles	3.57	3.31	t(47) = 1.99, p = .248
	I_can_find_better articles	3.04	3.19	t(47) = 0.61, p = .117
effort_to_use_system	sys_easy_to_use articles	4.47	4.35	t(47) = 1.65, p = .278,
	invest_lot_of_effo rt	2.22	2.12	t(47) = 0.49, p = .108
	recs_take_too_ much_time	1.74	1.77	t(47) = -2.15, p < .136

user experience. The results obtained from this data analysis revealed that there is no significant difference between the perception and user experience of the males and females that took part in the user study, table 6.12 shows the details of the analysis. Therefore, to answer the research question, we conclude that the personal characteristics of the participants had no effect on user experience nor perceptions; where the *gender* variable revealed no effects, and 91.8% of the participants were novices to the domain of the educational task to be completed.

6.2.3 Situational Characteristics on the user perception and user experience

Similar to the analysis performed on the participants' personal characteristics, this research questions examines if some of the situational characteristics of the users have an effect on their perception of the quality, accuracy of the recommender system and/or the user experience. The situational characteristics of users that have been widely noted to influence the perception and user experience are *trust in technology* and *privacy concerns*. The situational characteristics of the participants were obtained from the post-experiment (feedback) questionnaire administered. 71.6% of the participants reported to have general trust in technology and do not have privacy concerns with the system, while 28.4% reported not to have trust in technology and had privacy concerns with the system.

Computing independent t-tests on the two groups of users (those who trust in technology and those who don't) reveals that there is no significant difference in the user experience and perception between the group of users. Also, there is no significant difference in the user experience between the users who have privacy concerns with the system and those who don't. Table 6.13 summarizes the results of the analysis. Therefore we can conclude that the situational characteristics of the participants had no effect in the perception of the recommendation quality and accuracy assessments, and the user experience.

6.2.4 Relationship between the User Experience concepts

To determine how the three concepts (*recommendation_choice_satisfaction*, *recommender_system_effectiveness*, and *effort_to_use_system*) that measure the user experience relate and contribute to each other, correlation and regression analysis were performed. For simplicity, three composite variables were created for each concept and used for the analysis and the results of the analyses revealed that: (a) in the regression model for *recommendation_choice_satisfaction*, only *recommender_system_effectiveness* is a predictor (*effort_to_use_system* is excluded) and produced [$R^2 = .461$, $F(1, 47) = 18.366$, $p < .001$]. Furthermore, the correlation test also showed a positive and significant correlation between the concepts having coefficients of [$r = 0.659$, $p < .001$]. (b) *effort_to_use_system* concepts is a predictor to the *recommender_system_effectiveness* concept, which produced

[$R^2 = .314$, $F(1, 47) = 6.972$, $p < .001$], and the correlation test also showed a positive and significant correlation between the concepts having coefficients of [$r = 0.648$, $p < .001$].

In sum, this section (section 6.2) investigated the effects and influence of three components (subjective system aspects, personal characteristics, and situational characteristics) of the user-centric framework deployed to evaluate the recommender system on the user experience of the participants that participated in the user study experiments. In the data analysis performed in sections 6.2.2 and 6.2.3, we also examined whether the personal and situational characteristics of the users also affects their perception of the recommendations (as an indirect link to the user experience), because as discussed in section 6.2.1, we discovered that the perception influences the user experience. The results of the data analysis performed showed that only the subjective system aspects component contributed to the user experience.

6.3 What components of the framework affects the user interaction?

Similar to the data analysis conducted in section 6.2, we seek to examine the effect(s) of the preference elicitation methods, user experience, personal characteristics, and situational characteristics on the user interaction. Two variables are used to measure the concept of user interaction: *number of recommended articles clicked*, and *number of highlights created*; where the *number of recommended articles clicked* variable is computed for each participant and represents the sum of all the articles clicked on among the recommended articles they received, and *number of highlights created* variable also computed for each participant is the sum of all the highlights-tags created on all the articles clicked on, from among the list of recommendations received. Information about the user interaction is obtained from the logged data for each participant. The effect of each component on the user interaction is discussed in the subsections below.

6.3.1 Effects of the User Experience on User Interaction

Here, we hypothesize that given the two group conditions provide recommendations using different approaches, the difference in the methodology should be reflective in users' engagement with the recommendations provided. That is, since the participants in Group B receive random recommendations, there may be fewer relevant and task-appropriate articles to read and make highlights from, thus the number of clicks and articles operated on might be fewer compared to Group A participants who receive recommendations that are tailored to the highlights created. To determine whether the perception of the users has any effect on the user interaction, the hypotheses below investigate the effects and relationships of a number of concepts and variables on the user interaction.

Table 6.13: Mean responses and t-test results of the situational characteristics of the users' on user experience and perception concepts and variables

Concept	**Variable**	**Male Mean**	**Female Mean**	**T-test Values**
recommendation quality	like_articles	3.39	3.31	t(47) = 1.23, p = .179,
	valuable_articles	3.52	3.42	t(47) = 1..67, p = .150,
	irrelevant articles	2.70	2.73	t(47) = -1.09, p = .157
	didn't_like articles	2.35	2.38	t(47) = -0.91, p = .381
recommendation accuracy	well_chosen articles	3.13	3.58	t(47) = 1.29, p = .169
	relevant articles	3.39	3.58	t(47) = 1.49, p = .284
	high_rating articles	3.21	3.38	t(47) = 1.58, p = .188
	appealing articles	3.34	3.00	t(47) = 1.01, p = .224
recommendation choice_satisfaction	enjoyed_reading articles	4.36	2.63	t(47) = 1.51, p = .375
	task_appropriate articles	4.64	2.42	t(47) = 1.09, p = .193
	articles_fit_prefer ence	4.12	2.58	t(47) = 0.76, p = .245
	articles_waste_of time	1.76	3.54	t(47) = 1.21, p = .183
recommender_system effectiveness	recommend_sys tem	3.39	4.00	t(47) = 0..86, p = .341,
	sys_aware_highli ghts	4.09	3.88	t(47) = 1.73, p = .126
	sys_good_articles	3.57	3.31	t(47) = 1.99, p = .248
	I_can_find_better articles	3.04	3.19	t(47) = 0.61, p = .117
effort_to_use_system	sys_easy_to_use	4.47	4.35	t(47) = 1.65, p = .278,
	invest_lot_of_effo rt	2.22	2.12	t(47) = 0.49, p = .108
	recs_take_too_ much_time	1.74	1.77	t(47) = -2.15, p < .136

Variable	Group A Mean_responses	Group B Mean_responses	T-Values
rec_articles_clicked	8.86	3.74	t(47) = 5.47 p = 0.01 d = 3.23
highlights_created	34.76	11.52	t(47) = 4.91 p = 0.04 d = 2.61

Table 6.14: Mean responses, t-test results of the user interaction variables

- **H3.1**: *There is significant difference in the user interaction between the two group conditions.* The results of the descriptive statistics analysis of the user interaction variables revealed that Group A participants had an average of 8.86 *number of recommended articles clicked*, and 33.32 *number of highlights created*, while Group B participants had an average of 3.74 *number of recommended articles clicked*, and 12.38 *number of highlights created*. Using t-test analysis, we further examined whether there is a significant difference between the user interaction variables between the two group conditions. The results of the analysis confirms the hypothesis that there is a significant difference between the user interaction of the two groups; where *number of recommended articles clicked*: [$t(47) = 5.47$, $p = 0.01$, $d = 3.23$] and *number of highlights created*: [$t(47) = 4.91$, $p = 0.04$, $d = 2.61$]. Table 6.14 shows the details of the analysis.

- **H3.2**: *The user experience concepts affect the user interaction.* Correlation test and regression analysis are performed to determine the relationship between the concepts as well as to identify which of the three concepts that measure the user experience contributed to the user interaction variables. The results of the correlation test suggests that the user experience concepts are positively and significantly related to the user interaction, while the results of the regression analysis indicates that at least one variable each, of the three concepts that measure the user experience are predictors in the user interaction. The variables are: *task_appropriate_articles* of the recommendation choice satisfaction concept which produced [$R^2 = .573$, $F(1, 47) = 65.08$, $p < .001$], *sys_good_articles* of the recommender system effectiveness concept produced [$R^2 = .527$, $F(1, 47) = 52.29$, $p < .001$], and *invest_lot_of_effort* of the effort to use the system concept produced [$R^2 = .198$, $F(1, 47) = 5.12$, $p = .003$].

  To measure the effect of each concept as a unit, composite variables were created for each of the user experience concepts. Computing regression analysis on the composite variables to examine the effects on the user interaction gave the following results

Concept	**Correlation with user interaction**	R^2	Sig	B	β
recommendation_choice satisfaction	0.765	0.434	$< .001$	0.079	0.66
recommender_system effectiveness	0.659	0.275	$< .001$	0.058	0.53
effort_to_use_system	0.371	0.098	$< .001$	0.025	0.30
prefer_highlighting		0.339	$< .001$	0.079	0.58

Table 6.15: Correlation coefficients and regression weights on the effects of user experience and preferred elicitation method on user interaction

(shown in Table 6.15): *recommendation choice satisfaction*: [$R^2 = .434$, $F(1,47) = 36.02$, $p < .001$], *recommender system effectiveness*: [$R^2 = .275$, $F(1,47) = 27.29$, $p < .001$], and *effort to use system*: [$R^2 = .098$, $F(1,47) = 4.48$, $p < .001$]. Based on these results, we accept the hypothesis that the user experience influences the user interaction. The variable that contributes the most to the user interaction is the *task_appropriate_articles* of the *recommendation choice satisfaction* concept, which hints that make more metacognitive interactions when provided with appropriate learning materials.

6.3.2 Effects of Personal, Situational Characteristics on User Interaction

As aforementioned in section 6.2.2, the only personal characteristics variable we have sufficient data for analysis is *gender: M/F*. To determine if there is significant difference between the gender variable and the user interaction, an independent t-test was performed. The result of the test showed that there is no significant difference between the user interaction of the males and females that took part in the experiment.

Independent t-tests and regression analysis are performed to investigate the effects of the situational characteristics (trust in technology and privacy concerns) on the user interaction variables. The results (more details in Table 6.16) reveal that (a) there is a significant difference between the interaction of users who trust in technology and those who don't. This is also true for users who had privacy concerns with the system and those who didn't, which produced t-values [$t(47) = 6.32$, $p < 0.001$, $d = 2.89$] and [$t(47) = 5.01$, $p < 0.001$, $d = 2.77$] respectively. The regression analysis generated *trust in technology*: [$R^2 = .314$, $F(1,47) = 16.18$, $p < .001$], *privacy concerns*: [$R^2 = .248$, $F(1,47) = 10.33$, $p < .001$]. These results show that the situational characteristics of the users affect the user interaction, where the lack of *trust in technology* and having *privacy concerns* about the system disclosing personal information influenced the user interactions, and otherwise.

Concept	T-Values	Regression Weights			
		R^2	*Sig*	B	β
trust_in_tech	$t(47) = 6.32$ $p < 0.001$ $d = 2.89$	0.314	$< .001$	0.031	0.42
privacy_concerns	$t(47) = 5.01$ $p < 0.001$ $d = 2.77$	0.248	$< .001$	0.020	0.27

Table 6.16: T-values and regression weights on the effects of the users' situational characteristics on user interaction

6.3.3 Effect of Preference Elicitation Method on User Interaction

Similar to the user experience, we hypothesize that the preference elicitation method also has a positive effect on the user interaction, where the user interaction refers to the participant's engagement with the recommendations. Since there are two types of highlights involved in the user experiments; where the first set of highlights are created to guide recommendations and the second set of highlights made on the list of recommendations to complete the educational task, the *user interaction* concept strictly considers the highlights created to complete the educational task (*num_highlights_created)* as well as the number of recommendations clicked (*num_recs_clicked*).

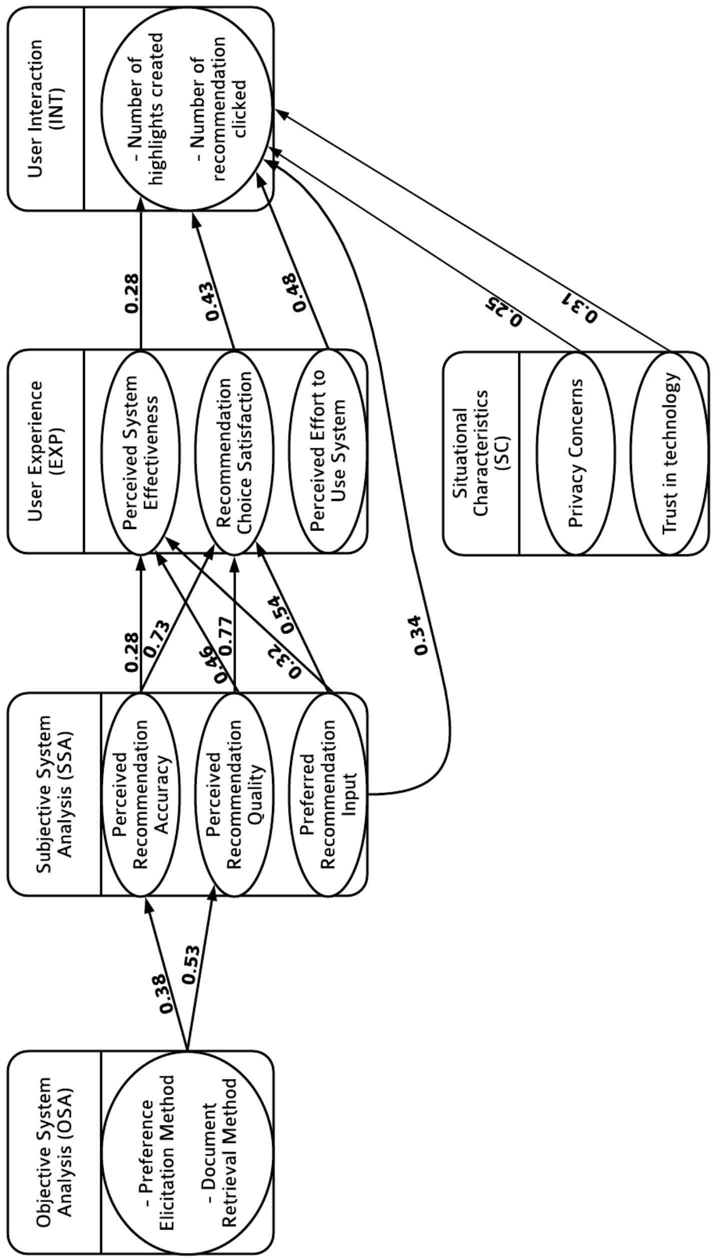
Computing correlation and regression analysis shows that the *user interaction* concept is significantly related to *preference elicitation method*; where *prefer_highlighting* variable is positively correlated having coefficients ($r = .614$, $p < .001$), and *prefer_typing* is negatively correlated ($r = -.404$, $p < .004$). The regression model obtained further revealed that only the *prefer_highlighting* variable is a predictor (prefer_typing is removed). This indicates that the *prefer_highlighting* variable contributes to the user interactions [$R^2 = .339$, $F(1, 47) = 24.06$, $p < .001$]. Therefore, we conclude that the user interaction was affected or influenced by the use of highlights.

6.4 Summary

Independent t-tests, correlation tests, regression analysis, as well as the computation of effect sizes were performed in the data analysis process to investigate the three research questions the study entailed. The first research question examined the use of the metacognitive activities of the learner as an appropriate preference elicitation method to guide recommendations, for which 92.3% of the participants in Group A reported to prefer the use of highlights to guide recommendations, and 48.2% of the participants in Group B noted their preference for highlights. Further investigation revealed that there is a significant difference in the preference elicitation method between the two group conditions; where participants in the experimental group (Group A) had a greater preference for the use of highlights to guide recommendation, and participants in the control group (Group B) had a greater preference for using search queries to guide recommendations. Regression analysis and correlation tests confirmed that the difference in the preference elicitation method between the two groups is a function of the perceived recommendation quality and accuracy.

The second research question examined the components of the framework that influenced the user experience. The components investigated were subjective system aspects, personal characteristics and situational characteristics. Results from the analysis suggests that the three components in some way affect the user experience. The data analyses on what components influenced the user interaction revealed that user experience, preferred elicitation method and situational characteristics are the components that influence the user interaction. Based on the results of the various data analysis conducted, Figure 5.6 (hypothesized relationship among the framework's components) is revised (see Figure 6.1) to depict the actual relationships as identified from the data analysis.

Figure 6.1: Graphical representation of the relationships between the concepts, variables evaluated of the recommender system,.



Chapter 7

Conclusion and Future Work

A personalized learning recommender system was designed to support self regulated learning. A recommender system in the context of learning can be considered as a useful tool for finding relevant documents among the vast amount of materials available on the Internet, as well as alleviating information overload which has been identified as one of the main problems learners encounter in online learning and when searching for the "right" information to satisfy their needs. This also makes the provision of personalized learning recommendations imperative as the system is able to tailor the learning activities to fit the needs, goals, talents, and interests of the learner. The development of personalized learning recommender systems therefore should include a *learner model* which is used to obtain/infer information about the learner. The learner model has widely been used to capture and store information about the learner's characteristics such as the learning goal, learning style, prior knowledge, and the information collected is used to guide recommendation.

A number of methods (algorithms) have been exploited to utilize the information about the learner in the learner model to achieve personalized learning recommendations such as concept maps, fuzzy logic, among others. In this thesis, we investigated the possibility of using highlights, one of the many metacognitive reading activity a learner engages in while reading to achieve personalization. Other metacognitive activities learner may engage in while reading include self-explanation, creating bookmarks, highlighting (text marking), taking notes, tags, among others. Specifically, the metacognitive activity used in this study is highlights, and the metacognitive aspects of creating highlights lies in the act of the learner judging whether to mark and what to mark. For the retrieval of relevant documents, the Latent Dirichlet Index (LDI) model is used. The LDI method leverages probabilistic topic modeling approaches for representing documents in a topic space where the topics can be seen as index terms. Together with the metacognitive activities, the recommender system provides recommendations that are related to the learners' information seeking needs.

Indexing and retrieval using probabilistic concept models are based on the assumption that the concepts are distributed differently in relevant and non-relevant documents. The LDI model is based on a popular topic modeling approach: Latent Dirichlet Allocation

(LDA), which models a document as a mixture of topics allows for the retrieval of documents based on the topical relatedness. In a small pilot study involving 10 graduate students, we asked the participants to evaluate different sets of articles retrieved using the LDI model and the standard query likelihood retrieval model. 80% of the participants reported that the sets of articles retrieved using the LDI model were better in terms of providing finer grained, but topically related articles.

The recommender system was integrated into an online learning platform, *nStudy*. nStudy supports learning, collaboration and research, annotation tools that learners need to record, catalog, analyze, organize, view and synthesize selected information for tasks of any scope and information in any subject area (Beaudoin and Winne, 2009). Some of the annotation tools supported by nStudy are creating bookmarks, adding notes, tags, highlighting portions of a text, as well as maintaining detailed log data of all the activities learners engage in during a study session. Thus, nStudy platform provided the platform that supports self-regulated learning, tools to facilitate metacognitive activities, and log data necessary for analysis. To evaluate the system developed, a user-centric framework for evaluation model was adopted by Knijnenburg et al., (2011).

The user study involved 49 participants who were randomly assigned to one of the two group conditions involved in the study. To investigate the effect of the system's methodology (the use of the highlights created as a preference elicitation method to guide recommendations and the use of the LDI model for document retrieval) on the user experience and user interaction as it relates to personalized learning, two groups were created. Participants in the first group, Group A, received recommendations based on the recommender system's methodology, while participants in the second group, Group B, received random recommendations. Three main research questions and a number of hypothesis were tested and analyzed using the information collected from the user study via two sets of questionnaires administered to participants during the experiments.

A pre-experiment (demographics) questionnaire and post-experiment (feedback) questionnaire were administered to each user that took part in the study. The data collected from the feedback questionnaire contained the users perceptions and assessments of the recommender system. The questions in the questionnaire using exploratory factor analysis can be categorized into eight groups: perceived recommendation quality, perceived recommendation accuracy, preferred elicitation method, recommendation choice satisfaction, recommender system effectiveness, effort to use system, general trust in technology, and privacy concerns. A pragmatic approach to evaluate the data was adopted which entailed performing independent t-tests, correlation tests, regression analysis, as well as the computation of effect sizes.

The results of the first research question which examined the user's perception on the use of the highlights as an appropriate preference elicitation method to guide recommendations revealed that that the use of highlights is considered suitable to guide recommendation, and also facilitates metacognition. The second research question examined the components of the framework that influenced the user experience. Results from the data analysis suggests that three components in some way affect the user experience they are: the subjective system aspects, personal characteristics and situational characteristics. The third research question also investigated the components that affected the user interaction. Three components were identified to contribute to the user interaction concepts they are: user experience, situational characteristics, and preference elicitation method.

7.1 Future Work

The design, implementation, analysis, and evaluation of our personalized learning recommender system to support self-regulated learning provides a first step of what needs to be examined as the basis for investigating the effects on learners' achievements when they use the recommender system. While the system works, as judged by the users, investigating what and how much learners using the system actually benefit by using it is an important next step for potential future work. Given that the educational task involves a study outcome (in our study, the outcome is an ensemble of highlights each participant puts together to answer the educational tasks). With the study outcome of the task,together with the other data sources the system records and generates (e.g. log data), it would be possible to assess wide range of educational values, as we will now outline.

A personalized learning recommending system has the potential to allow a learner to develop a deeper and more balanced understanding of the topic they are researching. This could be based not only on the highlighting that the learner has provided, but also on the other data that could be obtained from the log data nStudy records. For example, the amount of time a user spends reading/viewing a document could be suggestive that the document is relevant to the topic being researched. Furthermore, the incorporation of a question generation module, which assesses the learner based on pedagogical goals or the learning objectives of the reading session could be used to assess the learner's level of understanding on the topic they are researching.

Given a task oriented activity which for example involves writing an essay, or similar to the task involved in our user study (an ensemble of highlights to complete an educational task), a potential way to investigate how much learners benefit from using the recommender system is by assessing the essays drafted before and after using the system. This could be done by incorporating a pre-test and post-test activity into the study; where each user drafts an essay based on a topic before using the recommender system (pre-test), and after completing the reading session which comprises documents recommended by the system, the user is asked to draft another essay based on the same topic as in the pre-test. The comparison of the outcomes of the pre-test and post-test would provide an assessment of whether or not the recommender system was helpful in making users draft better essays.

Additionally, a more focused and interesting assessment of the educational significance with the user experience concepts could examine "if when provided with appropriate recommendations, are students able to identify why the recommendations are appropriate and find the answer to the question?" where the research objective would be to discover whether if students are able to identify and highlight relevant information from the recommended articles that would be necessary to complete an educational task.

Finally, future research can also investigate how the integration of other technology enhanced learning functionalities might be included in a personalized learning recommendation system. Some of the TEL functionalities that could be included to enhance learning are: collaborative learning and question generation module. According to Sansivero (2016), collaborative or active learning is a methodology that transforms that traditional lecture or teacher focused classroom into a student or learning centered room. In a collaborative learning setting, the students work together to help each other understand a content, solve problems or create projects and products with the instructor working as a moderator or facilitator. Collaborative spaces in education trickled down from corporate "flex/open workspaces." They were designed based on the understanding that interactivity and collaboration in small groups produces stronger solutions that would have not been reached individually and encourages sharing of research for enhanced learning. Further, it encourages trust building, communication, practical learning/application, and acceptance and enhances problem-solving skills, therefore could be a potentially valuable add-on feature for the system.

Assessment through posing questions is considered an integral part of learning. It can be leveraged to gather data that would be helpful to better understand the strengths and weaknesses of students' learning (Harris and Hodges, 1995). As a reflective process in which learners evaluate their performance and determine how to improve, an automatic question module could make available, important data that can be used to measure the progress of learning with respect to the learning goals and objectives. Research has also shown that learners need assessments to learn, regardless of whether they are posed by teachers or formulated by the students themselves (Morgan and Saxton, 1994), and also that assessments increase comprehension for learners (Rittle-Johnson, 2006). Therefore, the inclusion of an assessment module could be beneficial.

Bibliography

[1] Alexander, J. M., K. E Johnson, J. Albano, T. Freygang, B. Scott. Relations between intelligence and the development of metaconceptual knowledge. Metacognition & Learning, Vol.1, 51-57, 2006.

[2] Azevedo, R., Johnson, A., Chauncey, A., & Burkett, C. Self-regulated Learning with MetaTutor: Advancing the Science of Learning with MetaCognitive Tools. In M. S. Khine & I.M. Saleh (Eds.), New Science of Learning: Computers, Cognition, and Collaboration in Education, pp 225-247. Springer, 2010.

[3] Bagceci, B., Dos, B., R. SarÄśca. An analysis of metacognitive awareness levels and academic achievement of primary school students, Mustafa Kemal University Journal of Social Sciences Institute, Vol.8, No.16, 551-556, 2011.

[4] M. Baltaci, B. AkpÄśnar. The effect of web based instruction on the metacognition awareness levels of learners. Mustafa Kemal University Journal of Social Sciences Institute, Vol.8, No.16, 319-333, 2011.

[5] Batista, H. (2007). 7 Reasons Why Search Engines Don't Return Relevant Results 100% of the Time. https://moz.com/blog/7-reasons-why-search-engines-dont-return-relevant-results-100-of-the-time. Accessed March, 2019.

[6] Beaudoin, L. and Winne, P.H. nStudy: An internet tool to support learning, collaboration and researching learning strategies. CELC-2009 http://learningkit.sfu.ca/lucb/celc-2009/nstudy.pdf.

[7] Bielikova, M., Simko, M., Barla, M., Tvarozek, J., Labaj, M., Mro, R., Srba, I., and Sevcech, J.: ALEF: from Application to Platform for Adaptive Collaborative Learning. Special issue on Recommender Systems for Technology Enhanced Learning: Research Trends and Applications, Springer Berlin (2014).

[8] Boekaerts, M., Els, D., and V. Paul, "Goal-directed behavior an contextual factors in the classroom: An innovative approach to the study of multiple goals," Educational Psychologist, vol. 41, pp. 33-51, 2006.

[9] Boekaerts, M. Self-regulated learning: where we are today, International Journal of Educational Research, Vol.31, 445-457, 1999.

[10] Bolhuis, S., and Voeten, M. J. M. (2001). Toward self-directed learning in secondary schools: What do teachers do? Teaching and Teacher Education, 17(7), 837-855. doi:10.1016/S0742- 051X(01)00034-8

[11] Brusilovsky, P., Cassel, L.N., Delcambre, L.M.L., Fox, E.A., Furuta, R., Garcia, D.D., Shipman III, F.M., Yudelson, M.: Social navigation for educational digital libraries, Procedia Computer Science, 1(2), 2889-2897 (2010).

[12] Case, L .P., Harris, K. R., S. Graham. Improving the mathematical problem solving of students with learning disabilities: Self-regulated strategy development. The Journal of Special Education, Vol.26, 1-19, 1992.

[13] Carchiolo, V., Longheu, A., Malgeri, M.: Reliable peers and useful resources: Searching for the best personalised learning path in a trust- and recommendation-aware environment, Information Sciences, Volume 180, Issue 10, pp. 1893-1907 (2010), ISSN 0020-0255, http://dx.doi.org/10.1016/j.ins.2009.12.023.

[14] Mohamed Amine Chatti, Vlatko Lukarov, Hendrik Thus, Arham Muslim, Ahmed Mohamed Fahmy Yousef, Usman Wahid, Christoph Greven, Arnab Chakrabarti & Ulrik Schroeder Learning Analytics: Challenges and Future Research Directions. 2014: eleed, Iss. 10.

[15] Chen, C.M., Duh, L.-J.: Personalized web-based tutoring system based on fuzzy item response theory, Expert Systems with Applications, Volume 34, Issue 4, May 2008, pp. 2298-2315, ISSN 0957-4174, http://dx.doi.org/10.1016/j.eswa.2007.03.010 (2008).

[16] Code, J.R., MacAllister, K., Gress, L.Z., Nesbit, J.C.: Self-Regulated Learning, Motivation and Goal Theory: Implications for Instructional Design and E-Learning. In Proceeding Sixth IEEE international conference on Advanced Learning Technologies, 2006. doi: 10.1109/ICALT.2006.1652581 (2006).

[17] Conati, C., Vanlehn, K. Toward Computer-Based Support of Meta-Cognitive Skills: a Computational Framework to Coach Self-Explanation. In International journal of Artificial Intelligence in Education (2000), 11, 398-415.

[18] Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. Journal of verbal learning and verbal behavior, 11(6), 671-684. http://dx.doi.org/10.1016/S0022-5371(72)80001-X

[19] Dagal, A.B., Bayindir., D. The investigation of the relationship between the level of metacognition awareness, self-directed learning readiness and academic achievement of preschool teacher candidates. Universal Journal of Educational research 4(11): 2533-2540, 2016, DOI: 10.13189/ujer.2016.041106. https://files.eric.ed.gov/fulltext/EJ1118705.pdf

[20] Deci, E., and Ryan, R. Intrinsic motivation and self determination in human behavior. New York, NY: Plenum, 1985

[21] Desoete, A., Roeyers H. Off-line metacognition a domain-specific retardation in young children with learning disabilities. Learning Disability Quarterly, Vol.25, 123-139, 2002.

[22] Dewek, C. "The development of early self-conceptions: Their relevance for motivational processes," in Motivation and self-regulation across the life span, J. Heckhausen and C. Dweck, Eds. New York, NY: Cambridge University Press, 1998, pp. 257-280.

[23] Dinsmore, D.L., P.A. Alexander, S.M. Loughlin. Focusing the conceptual lens on metacognition, self-regulation and self-regulated learning, Educational Psychology Review, 20, 391-409, 2008.

[24] Dunning,D.A., K. Johnson, J. Ehrlinger, J. Kruger, Why people fail to recognize their own incompetence. Current Directions in Psychological Science, Vol.12, No.3, 83-87, 2003.

[25] Drachsler H., Verbert K., Santos O.C., Manouselis N. (2015) Panorama of Recommender Systems to Support Learning. In: Ricci F., Rokach L., Shapira B. (eds) Recommender Systems Handbook. Springer, Boston, MA.

[26] Duval, E., Sharples, M., and Sutherland, R. Research Themes in Technology Enhanced Learning. In E. Duval, M. Sharples & R. Sutherland (Eds.), Technology Enhanced Learning: Research Themes (pp.1–10). Springer, Cham Switzerland (2017).

[27] Dwivedi, P., Kant, V. & Bharadwaj, K.K. Learning path recommendation based on modified variable length genetic algorithm. Educ Inf Technol (2018) 23: 819. https://doi-org.proxy.lib.sfu.ca/10.1007/s10639-017-9637-7

[28] Earl, L. (2003) Assessment as Learning: Using Classroom Assessment to Maximise Student Learning. Thousand Oaks, CA, Corwin Press. http://web.uvic.ca/ thopper/iweb09/GillPaul/Site/Assessment_files/Assessment.pdf

[29] Estefania, M. and Carro, R. M. . Supporting the Development of Mobile Adaptive Learning Environments: A Case Study. Published in: IEEE Transactions on Learning Technologies (Volume: 2 , Issue: 1 , Jan.-March 2009).

[30] Felfernig, A., Jeran, M., & Ninaus, G., Reinfrank, F., Reiterer, S., Stettinger, M. Basic Approaches in Recommendation Systems. In M. Robillard, W. Maalej, R. Walker, T. Zimmermann (Eds.), Recommendation Systems in Software Engineering, Springer (2014). DOI: 10.1007/978-3-642-45135-5_2

[31] Flavel, J.H. Metacognition and cognitive monitoring; A new area of cognitive-developmental inquiry. American Psychologist, 34(10), 906-911, 1979.

[32] Flavell, J. H. (1971). First discussant's comments: What is memory development the development of? Human Development, 14(4), 272-278.

[33] Gomez-Albarran, M., Jimenez-Diaz, G.: Recommendation and Students' Authoring in Repositories of Learning Objects: A Case-Based Reasoning Approach. International Journal of Emerging Technologies in Learning (iJET) 4(1), 35-40 (2009).

[34] Herlocker, J. L., Konstan, J. A., and Riedl, J. (2000). Explaining collaborative filtering recommendations. In ACM conference on Computer supported cooperative work, pages 241-250.

[35] More words in average search query: Hitwise. Web page accessed May 25, 2019. https://www.dmnews.com/customer-experience/article/13040684/more-words-in-average-search-query-hitwise

[36] Huang, Y.-M., Huang, T.-C., Wang, K.-T., Hwang, W.-Y.: A Markov-based Recommendation Model for Exploring the Transfer of Learning on the Web. Educational Technology and Society, 12(2),144-162 (2009).

[37] Hsieh, T.-C., Lee, M.-C., Su, C.-Y.: Designing and implementing a personalized remedial learning system for enhancing the programming learning. Educational Technology & Society 16(4): 32-46 (2013).

[38] Hsu, C.-K., Hwang, G.-J., Chang, C.-K.: A personalized recommendation-based mobile learning approach to improving the reading performance of EFL students, Computers & Education, Volume 63, April 2013, pp. 327-336, ISSN 0360-1315, http://dx.doi.org/10.1016/j.compedu.2012.12.004 (2013).

[39] Kardan, A. A., Abbaspour, S., Hendijanifard, F. A Hybrid Recommender System for E-learning Environments Based on Concept Maps and Collaborative Tagging.

[40] Kerkiri, T., Manitsaris, A., Mavridis, I.: How e-learning systems may benefit from ontologies and recommendation methods to efficiently personalise resources. IJKL 5(3/4): 347-370 (2009).

[41] Khribi, M.K., Jemni, M., Nasraoui, O.: Automatic Recommendations for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval. Educational Technology and Society, 12(4), pp. 30-42 (2009).

[42] Kirkwood, Adrian and Price, Linda (2014). Technology-enhanced learning and teaching in higher education: what is ‘enhanced’ and how do we know? A critical literature review. Learning, Media and Technology, 39(1) pp. 6-36.

[43] Klasnja-Milicevic, A., Vesin, B., Ivanovic, M. & Budimac, Z. (2011). E-Learning personalization based on hybrid recommendation strategy and learning style identification. Computers & Education, 56(3), 885-899. Elsevier Ltd.

[44] Knijnenburg, B.P., Willemsen, M.C., Gantner, Z., Soncu, H., Newell, C. Explaining the user experience of recommender systems. User Modeling and User-Adapted Interaction 22(4-5), 441-504 (2012). DOI 10.1007/s11257-011-9118-4

[45] Knijnenburg, B.P.,Willemsen, M.C., Hirtbach, S. Receiving recommendations and providing feedback: The user-experience of a recommender system. In: F. Buccafurri, G. Semeraro (eds.) E-Commerce andWeb Technologies, vol. 61, pp. 207-216. Springer, Berlin, Heidelberg (2010). DOI 10.1007/978-3-642-15208-5_19

[46] Knijnenburg, B.P.,Willemsen, M.C. Evaluating Recommender Systems with User Experiments. In: Ricci F., Rokach L., Shapira B. (eds) Recommender Systems Handbook. Springer, Boston, MA (2015). DOI 10.1007/978-1-4899-7637-6_9

[47] Knijnenburg, B.P., Willemsen, M.C., Kobsa, A.: A pragmatic procedure to support the user-centric evaluation of recommender systems. In: 5th ACM Conference on Recommender Systems, pp. 321-324. Chicago, IL (2011a).

[48] Knijnenburg, B.P., Meesters, L., Marrow, P., Bouwhuis, D.: User-centric evaluation framework for multimedia recommender systems. In: First International Conference on User Centric Media, LNICST, vol. 40, pp. 366–369. Venice, Spain (2010a).

[49] Knijnenburg, B.P., Willemsen, M.C., Hirtbach, S.: Receiving recommendations and providing feedback: The user-experience of a recommender system. In: 11th International Conference on Electronic Commerce and Web Technologies, LNBIP, vol. 61, pp. 207-216. Bilbao, Spain (2010b).

[50] Knijnenburg, B.P.,Willemsen, M.C.: The effect of preference elicitation methods on the user experience of a recommender system. In: 28th International Conference on Human Factors in Computing Systems, pp. 3457–3462. Atlanta, GA (2010).

[51] Knowles, M. (1975). Self-Directed Learning: A Guide for Learners and Teachers. Chicago, IL: Follett Publishing Company.

[52] Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. Psychonomic Bulletin & Review, 14(2), 219–224. http://dx.doi.org/10.3758/BF03194055

[53] Koren, Y., R. Bell, and C. Volinsky. Matrix Factorization Techniques for Recommender Systems. IEEE Computer, 42(8):42–49, 2009.

[54] Liddy, E.D. Document Retrieval, Automatic. Published in the Encyclopedia of Language & Linguistics, 2nd Edition. Elsevier Limited. 2005.

[55] Ma, K. Content-based Recommender System for Movie Website. M.Sc thesis (2016).

[56] McNamara, D. S., & Magliano, J. P. (2009). Self-explanation and metacognition. In D. Hacker, J. Donlosky, & A. C. Graesser (Eds.), Handbook of metacognition in education (pp. 60-81). New York: Taylor & Francis.

[57] Manouselis, N., Drachsler, H., Vuorikari, R., Hummel, H., and Koper, R.: Recommender systems in technology enhanced learning. In: Rokach, L., Shapira, B., Kantor, P., Ricci, F., editor, Recommender Systems Handbook: A Complete Guide for Research Scientists & Practitioners, pp. 387–409. Springer (2011).

[58] Martin, E., Carro, R.M.: Supporting the Development of Mobile Adaptive Learning Environments: A Case Study. TLT 2(1): 23-36 (2009).

[59] Marzouk, Z.: Text Marking: A Metacognitive Perspective. PhD thesis (2018).

[60] McCallum, Kachites, A. "MALLET: A Machine Learning for Language Toolkit." http://mallet.cs.umass.edu. 2002.

[61] McCrudden, M. T. and Schraw, G., Relevance and goal-focusing in text processing. Educational Psychology Review, 19, 113-139, 2007.

[62] Mitrovic, T., Supporting Self-Explanation in a Data Normalization Tutor. Supplementary Proceedings of AIED2003, 2003.

[63] Murphy, P., and P. Alexander, "A motivated exploration of motivation terminology," Contemporary Educational Psychology, vol. 25, pp. 3-53, 2000.

[64] Ministry of Education. (2007). The New Zealand Curriculum. Learning Media Ltd. Wellington, New Zealand.

[65] Nava Tintarev. Explaining Recommendations (2010). http://navatintarev.com/papers/Nava%20Tintarev_PhD_Thesis_(2010).pdf

[66] Nist, S.L. & Hogrebe, M.C. (1987). The role of underlining and annotating in remembering textual information. Reading Research and Instruction, 27(1), 12-25. http://dx.doi.org/10.1080/19388078709557922

[67] Odilinye, L., Popowich, F., Zhang, E., Nesbit, J.C., & Winne, P. H. (2015). Aligning automatically generated questions to instructor goals and learner behaviour. In Proceedings of the 2015 IEEE International Conference on Semantic Computing (pp. 216-223). ICSC 2015: Anaheim, CA.

[68] Ormrod, J. (2012). Human learning. (6th ed.). Upper Saddle River, NJ: Pearson Education, Inc.

[69] Persico, D., and Steffens, K. Self-Regulated Learning in Technology Enhanced Learning Environments. In E. Duval, M. Sharples & R. Sutherland (Eds.), Technology Enhanced Learning: Research Themes (pp.1–10). Springer, Cham Switzerland (2017).

[70] Romero, C., Ventura, S., Zafra, A., De Bra, P.: Applying Web usage mining for personalizing hyperlinks in Web-based adaptive educational systems. Computers & Education 53(3), pp. 828–840 (2009).

[71] Rowe, P. J., Mott, B. W., McQuiggan, S. W., Robison, J. L., Lee, S., & Lester, C. J. (2009). Crystal Island: A Narrative Centered Learning Environment for Eighth Grade Microbiology. Workshop on Intelligent Educational Games. Brighton, UK, pp. 11âĂŞ21.

[72] Ryan, R., and Deci, E. “Intrinsic and extrinsic motivations: Classic definitions and new directions," Contemporary Educational Psychology, vol. 25, pp. 54–67, 2000.

[73] Sadler, D.R. (1998) Formative assessment: revisiting the territory, Assessment in Education, 5(1), 77-84.

[74] Sansivero, G.: Challenges and Advantages of collaborative learning: Developing Workforce Readiness in Students. https://www.seenmagazine.us/Articles/Article-Detail/articleid/5613/challenges-amp-advantages-of-collaborative-learning Web page accessed July, 2019.

[75] Salehi, M.: Application of implicit and explicit attribute based collaborative filtering and BIDE for learning resource recommendation, Data & Knowledge Engineering, Volume 87, September 2013, pp. 130–145, ISSN 0169-023X, http://dx.doi.org/10.1016/j.datak.2013.07.001 (2013).

[76] Sarwar, B., G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In 10th International Conference on World Wide Web, pages 285âĂŞ295, 2001.

[77] Schraw, R. Sperling-Dennison. Assessing metacognitive awareness. Contemporary Educational Psychology, Vol.19, 460–470, 1994.
Schraw (Ed.), Issues in the measurement of metacognition (pp. 147–222). Lincoln, NE: Buros Institute of Mental Measurements and Erlbaum Associates.

[78] Shen, L., Shen, R. Learning content recommendation service based-on simple sequencing specification. In: Liu W et al. (eds) Lecture notes in computer science, pp. 363-370 (2004).

[79] Shishehchi, S., Banihashem, S. Y., MatZin, N.A., and Mohd, S.A. Review of Personalized Recommendation Techniques for Learners in E-learning Systems. In International Conference on Semantic Technology and Information Retrieval 28-29 June 2011, Putrajaya, Malaysia (2011).

[80] Tajika, H., Nakatsu, N., Nozaki, H., Ewald, N., & Shunichi, M. (2007). Effects of self-explanation as a metacognitive strategy for solving mathematical word problems. Japanese Psychological Research, 49, 222âĂŞ233.

[81] Tancheva, K. and Koennecke, J. Social Navigation, Recommender Systems, and Libraries.

[82] Tilstra, J., and McMaster, K. L. (2013). Cognitive processes of middle grade readers when reading expository text with an assigned goal. Learning and Individual Differences, 28, 66–74. http://dx.doi.org/10.1016/j.lindif.2013.09.011

[83] Tobias, S. and Everson, H.T. (2000). Assessing metacognitive knowledge monitoring. In G.

[84] Tsai, K.H., Chiu, T.K., Lee, M.C., Wang, T.I.: A learning objects recommendation model based on the preference and ontological approaches. In: Proc. of 6th International Conference on Advanced Learning Technologies (ICALT'06). IEEE Computer Society Press (2006).

[85] Veenman, M. V., Van Hout-Wolters, B. H., Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. Metacognition and learning, 1(1), 3–14. http://dx.doi.org/10.1007/s11409-006-6893-0

[86] Veenman, M. V. J., Beishuizen, J. J. Intellectual and metacognitive skills of novices while studying texts under conditions of text difficulty and time constraint. Learning and Instruction, Vol.14, 621-640, 2004.

[87] Vrugt, A., F. Oort. Metacognition, achievement goals, study strategies and academic achievement: pathways to achievement. Metacognition Learning, V ol.30, 123-146, 2008.

[88] Wade, S.E. Trathen, W. (1989). Effect of self selected study methods on learning. Journal of Educational Psychology, 8(1), 40-4. http://dx.doi.org/10.1037/0022-0663.81.1.40

[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 System Application, 39(7), pp. 6446-6453 (2012).

[90] Watson, V. (2004). Principles of effective practice in supporting students to become self-regulated learners. Paper Presented at NZARE Conference, Turning the Kaleidoscope, Wellington, New Zealand. Retrieved: 02/08/15 Retrieved from: http://www.nzcer.org.nz/system/files/14343.pdf

[91] Wei X., Croft W. B. (2006). LDA-based document models for ad-hoc retrieval. Published in SIGIR '06 Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, pages 178-185, Seattle, WA, USA, August 06 - 11, 2006.

[92] Williamson, G. Self-regulated learning: an overview of metacognition, motivation and behaviour. Journal of Initial teacher inquiry (2015). Volume 1. https://pdfs.semanticscholar.org/1e1c/2d16922e667f9d39368b47904aac1506f6bf.pdf

[93] Wilson, D.C. and Leake, D.B.. Maintaining case-based reasoners: dimensions and directions, Computational Intelligence, vol. 17, n. 2, pp. 196-213, May 2001. (doi:10.1111/0824-7935.00140).

[94] Winne, P. H., and Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. Graesser (Eds.), Metacognition in educational theoryand practice (pp. 277âĂŞ304). Hillsdale, NJ: Lawrence Erlbaum.

[95] Yi X., Allan J. (2009) A Comparative Study of Utilizing Topic Models for Information Retrieval. In: Boughanem M., Berrut C., Mothe J., Soule-Dupuy C. (eds) Advances in Information Retrieval. ECIR 2009. Lecture Notes in Computer Science, vol 5478. Springer, Berlin, Heidelberg. DOI: https://doi.org/10.1007/978-3-642-00958-7_6

[96] Zaiane, O.R.: Building a recommender agent for e-learning systems. Computers in Education, 2002. vol.1, 3-6, doi: 10.1109/CIE.2002.1185862 (2002).

[97] Zimmerman, B.J., D.H. Schunk, Reflections on theories of self-regulated learning and academic achievement, In B.J. Zimmerman, & D.H. Schunk, (Edi.)Self-regulated learning and academic achievement, Theoretical perspectives (pp.289-307). Routledge, 2001.

[98] Zhou, X., Xu, H. Challenges for Educational Recommender Systems. In C.O. Santos (Ed.), Educational Recommender Systems and Technologies: Practices and Challenges, pp. 281–289 (2012).

Appendix A

Feedback Questionnaire Items and Measurements

Question	Variable	Loading
Recommendation Quality (Cronbach's alpha: 0.920)		
I liked the articles recommended	like_articles	0.848
The system provided valuable recommendations	valuable_articles	0.861
The system had too many irrelevant recommendations	too_many_irrelevant_articles	0.792
I didn't like any of the recommended articles	didn't_like_articles	0.804
Recommendation Accuracy (Cronbach's alpha: 0.879)		
The articles were well chosen based on my highlights	well_chosen articles	0.811
The recommended articles were relevant to completing the task	relevant_articles	0.893
I would give most of the articles recommended a high rating	high_rating_articles	0.846
The list of recommended articles was appealing	articles_appealing	0.799
Preferred Elicitation Method (Cronbach's alpha: 0.813)		
I prefer to use highlights for recommendation highlights	prefer_highlighting	0.823
I would have preferred typing search queries for recommendation	prefer_typing	0.914
Recommender System Effectiveness (Cronbach's alpha: 0.905)		
I would recommend the system to others	recommend_system	0.823

The system made me aware of my highlights completing the task	relevant_articles	0.742
I can find better articles using the recommender system	high_rating_art icles	0.895
I can find better articles without the recommender system	articles_appeal ing	0.806
Recommendation Choice Satisfaction (Cronbach's alpha: 0.798)		
I enjoyed reading the articles I selected	enjoyed_reading articles	0.938
The articles were appropriate for the task	task_appropriate articles	0.892
The chosen articles fit my preference	articles_fit_my preference	0.837
The articles I read were a waste of my time	articles_waste of_time	0.826
Recommender System Effectiveness (Cronbach's alpha: 0.852)		
I would recommend the system to others	recommend_system	0.768
The system made me aware of my highlights	sys_aware highlights	0.875
I can find better articles using the recommender system	sys_good_articles	0.810
I can find better articles without the recommender system	I_can_find better_articles	0.794
Effort to Use System (Cronbach's alpha: 0.797)		
The system is easy to use	sys_easy_to_use	0.888
I have to invest a lot of effort to use the system	invest_lot_of effort	0.723
It takes too much time before the system provides recommendation	rec_takes_too much_time	0.699
General Trust in Technology (Cronbach's alpha: 0.838)		
Technology rarely works for me	tech_rarely_works	0.903
I'm less confident when I use technology	less_confident	0.842
I prefer technology, it works for me	tech_works	0.810
Privacy Concerns (Cronbach's alpha: 0.711)		
I feel confident that the system respects my privacy	respect_privacy	0.879
I'm afraid the system discloses private information	disclose_privacy	0.816

about me		
The system invades my privacy	invades_privacy	0.788

Appendix B

User Study Articles Titles

About the Massey Tunnel Replacement Project
Analysis of George Massey Tunnel Replacement Project
Environmental review of the Massey Tunnel Replacement Project
BC NDP questions Massey Tunnel replacement project
BC NDP would side with mayors on Massey Tunnel replacement project
Bridging the gap
Debt could add $8 billion to Massey bridge cost, says NDP
Delta gets bridge meeting
Delta decries 'rotting' Massey tunnel, says new bridge best, safest option
Delta wants work to continue on bridge replacement for Massey Tunnel
Delta making a case for bridge
Delta urging bridge replacement project to proceed
Delta council takes swipes at Richmond as bridge threatened
Delta mayor concerned NDP and Greens will kill bridge
Delta mayor defends support for proposed bridge to replace Massey tunnel
Delta mayor to discuss Massey Tunnel replacement with transportation minister
Delta mayor meets with transportation minister
Delta mayor has hope for bridge
Feds say no to environmental review of Massey Tunnel replacement project
Financing costs for Massey Tunnel replacement total $8 billion on $3.5-billion project
Get global experts to review tunnel
Change in government rekindles Massey Bridge debate
Richmond urges premier to halt Massey tunnel replacement bridge
Liberals are still committed to bridge
Metro Vancouver Mayors reject replacement of Massey Tunnel with 10 lane bridge

Metro Vancouver to ask BC NDP government about status of bridge replacement for George Massey Tunnel
Massey Tunnel replacement: Metro Vancouver board wants more time to study impacts
Massey Tunnel replacement project brings out critics and protesters
Massey Tunnel must be replaced
Massey Tunnel to be replaced with BC's largest toll bridge
Massey Tunnel: Let's avoid another sad chapter in our 'do nothing novel'
Massey Tunnel replacement project given environmental green light
BC's Massey tunnel replacement project officially begins
Massey tunnel replacement project off to a rocky start
Bridge is best replacement for Massey Tunnel
Footings for George Massey Tunnel replacement bridge will hold firm, says ministry
Bridge project in jeopardy
New span promised to meet first responders' Massey Bridge concerns
What is going to happen to the bridge proposed to replace the Massey Tunnel
Protesters crash groundbreaking for bridge to replace Massey Tunnel
The Public Safety and Economic Imperative for the George Massey Tunnel Replacement Project
Metro board asks new government to reconsider Massey Tunnel project
Richmond and Delta worlds apart on Massey Tunnel project
City of Richmond report highlights 'significant gaps' in Massey Tunnel replacement plans
Richmond councillor wants NDP to take new direction on proposed bridge to replace Massey Tunnel
Silent majority must let Horgan know
Uncertain future for bridge
Metro Vancouver mayors agree on need for new 8-lane bridge
Vancouver's mayor wants Massey Replacement replacement soon
George Massey Tunnel review
Report suggests BC government go back to drawing board Massey Tunnel
Massey Tunnel will not be replaced with 10-lane bridge
No final decision on Massey Tunnel replacement yet, BC government
New plan for Massey Tunnel crossing expected in 2020
Mayors call for immediate action on Massey tunnel replacement
Replacement Bridge unlikely for George Massey Tunnel
New span promised to meet first responders' Massey Bridge concerns
What is going to happen to the bridge proposed to replace the Massey Tunnel
Further delays hamper B.C's Massey Tunnel replacement

B.C. technical review of George Massey Tunnel Replacement
Metro board asks new government to reconsider Massey Tunnel project
Richmond and Delta worlds apart on Massey Tunnel project
City of Richmond report highlights ‘significant gaps’ in Massey Tunnel replacement plans
Richmond councilor wants NDP to take new direction on proposed bridge to replace Massey Tunnel
PCA disappointed by George Massey Tunnel review
Five options for George Massey Tunnel replacement
At-risk Frogs, Owls may be harmed by the George Massey Tunnel replacement project
George Massey Tunnel replacement could be cheaper than projected
Replacing Massey Tunnel remains top priority, Delta Mayor

Appendix C

Analysis of the User Interactions

1. The user Interactions of Group A participants: 25 participants

Num of articles clicked	**Num of highlights created**
6	27
5	19
7	30
8	43
6	31
5	27
8	37
7	32
7	39
8	46
6	35
7	40
3	9
9	52
8	42
9	36
7	31
6	22
7	41
8	39
9	47
5	19
6	28

7	32
8	29
Summary	
Avg. num of highlights	34.8
Avg. num of requests for recommendation	2.7

2. Group B participants interactions: 24 participants

Num of articles clicked	**Num of highlights created**
2	9
1	0
5	23
2	12
1	7
6	22
2	5
3	10
0	0
6	19
5	17
3	7
0	0
3	12
2	2
7	21
3	17
3	15
4	22
7	19
3	11
2	16
3	14
1	7

Summary	
Avg. num of highlights	11.5
Avg. num of requests for recommendation	5.3