AUTHOR: MBIPOM, B.

TITLE: Knowledge driven approaches to e-learning recommendation.

YEAR: 2018.

Knowledge Driven Approaches to e-Learning Recommendation

Blessing Mbipom

A thesis submitted in partial fulfilment of the
requirements of the
Robert Gordon University
for the degree of Doctor of Philosophy

May 2018

Abstract

Learners often have difficulty finding and retrieving relevant learning materials to support their learning goals because of two main challenges. The vocabulary learners use to describe their goals is different from that used by domain experts in teaching materials. This challenge causes a semantic gap. Learners lack sufficient knowledge about the domain they are trying to learn about, so are unable to assemble effective keywords that identify what they wish to learn. This problem presents an intent gap. The work presented in this thesis focuses on addressing the semantic and intent gaps that learners face during an e-Learning recommendation task.

The semantic gap is addressed by introducing a method that automatically creates background knowledge in the form of a set of rich learning-focused concepts related to the selected learning domain. The knowledge of teaching experts contained in e-Books is used as a guide to identify important domain concepts. The concepts represent important topics that learners should be interested in. An approach is developed which leverages the concept vocabulary for representing learning materials and this influences retrieval during the recommendation of new learning materials. The effectiveness of our approach is evaluated on a dataset of Machine Learning and Data Mining papers, and our approach outperforms benchmark methods. The results confirm that incorporating background knowledge into the representation of learning materials provides a shared vocabulary between experts and learners, and this enables the recommendation of relevant materials.

We address the intent gap by developing an approach which leverages the background knowledge to identify important learning concepts that are employed for refining learners' queries. This approach enables us to automatically identify concepts that are similar to queries, and take advantage of distinctive concept terms for refining learners' queries. Using the refined query allows the search to focus on documents that contain topics which are relevant to the learner. An e-Learning recommender system is developed to evaluate the success of our approach using a collection of learner queries and a dataset of Machine Learning and Data Mining learning materials. Users with different levels of expertise are employed for the evaluation. Results from experts, competent users and beginners all showed that using our method produced documents that were consistently more relevant to learners than when the standard method was used. The results show the benefits in using our knowledge driven approaches to help learners find relevant learning materials.

Keywords: e-Learning Recommendation, Semantic Gap, Intent Gap, Query Refinement, Background Knowledge

Acknowledgments

I am sincerely grateful to God for granting me the ability to go through this PhD programme.

I am thankful to my supervisors, Professor Susan Craw and Dr. Stewart Massie for their guidance throughout the period of this research. I am thankful for all the weekly discussions we had, these were very useful and helped to sharpen my thinking. I am particularly grateful for the time given to go through my work each time the need arose. Thank you for being an excellent supervisory team to me.

I wish to acknowledge the people I have met at Robert Gordon University who have been a part of my research journey. I am grateful to the staff and technical team from the School of Computing Science and Digital Media for their support during my PhD. I am thankful to the members of the Artificial Intelligence research group for the opportunities to present my work and for the valuable feedback received each time. Thank you for giving your time to be part of the user evaluation.

I am thankful to the Daramolas who have been my family in Aberdeen, thank you for your unwavering support and care throughout the course of my studies. I am grateful to my friend, Mayowa for being a valuable part of this journey, thanks for all the chats, lunches, and for being there. Many thanks go to Iyke, who was ever ready to help with any technical issues throughout the period of my research.

Special appreciation goes to my family; I am very grateful to my Dad, Late Professor Ebong Mbipom for being a great inspiration, and to my Mum, Professor Grace Mbipom who keeps giving me the support and encouragement I need, thank you for being there. I am grateful to my eldest Sister, Dr Rachel, for her valuable advice, and encouragement throughout this journey. I acknowledge my siblings for their love, prayers and support during this PhD programme, I am thankful to have you all.

Finally, I am especially grateful to the University of Calabar, Nigeria for the immense support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Artificial Intelligence (AI) is a field of Computer Science that focuses on building intelligent machines that operate like humans do. AI has the potential to support education in the development of innovative technologies to enhance teaching and learning (Popenici & Kerr 2017). AI methods have been applied to assist in the teaching process by providing timely and relevant feedback to learners. For example, in (Goel & Joyner 2017), an AI teaching assistant called Jill Watson was developed to automatically answer frequently asked questions from learners, so that the teaching assistants could focus on more creative aspects of the e-Learning course. In developing the AI teaching assistant, a knowledge base of previous question and answer pairs were analyzed and put in relevant categories. So, when a learner asked a new question, the AI assistant was able to map the question to a correct category and provide a relevant answer to the learner. Their results suggest that learners preferred the e-Learning course with the AI teaching assistant over the traditional course. Similar AI techniques have been applied to create many AI agents to assist the teaching process in e-Learning courses (Goel & Joyner 2016).

E-learning involves the use of computing technologies for teaching and learning, thus giving learners the opportunity to learn anytime and from anywhere. This may be combined with traditional face-to-face approaches as in blended learning (Bonk & Graham 2012) or completed by electronic means as in distance learning (Moore & Kearsley 2011). Learning resources may be delivered through personal computers or laptops as in Computer Assisted Instruction (Pilli & Aksu 2013) or via mobile devices as in M-learning (Crompton 2014). This research develops underlying models to support the retrieval and recommendation of e-Learning resources that can be applied to any of these scenarios.

There is currently a substantial amount of e-Learning resources available to learners on the

Web such as through open access resources and online courses (Clarà & Barberà 2013). However, learners are overloaded with information from the Web and are often discouraged by the time spent in finding relevant learning resources to support their learning goals (Chen, Niu, Zhao & Li 2014). We propose e-Learning Recommendation as a solution for addressing this challenge. e-Learning recommendation enables learners to find relevant learning materials from the substantial collection of available learning materials.

Recommendation in e-Learning differs from recommendation in other areas for various reasons. First, in e-Learning recommendation, one is not able to recommend a learning resource to a learner simply because other learners with similar preferences liked the learning resource. Instead the learner's goals have to be taken into account before recommendation is made. So, an important feature that distinguishes e-Learning recommendation from other areas is the learning goals that learners have. A learner's goals are usually expressed through a query. However, learners have insufficient domain knowledge, so they are often unable to ask effective queries to retrieve relevant resources that meet their goals. Second, learning resources do not typically have a set of clearly defined features, unlike items in other forms of recommendation. This is because learning resources contain a substantial amount of unstructured text. So, learning resources are not suitably indexed for retrieval. The difficulty of dealing with unstructured text creates a challenge for e-Learning recommendation when trying to find and retrieve relevant learning resources.

1.1 Research Motivation

This research is motivated by two issues associated with e-Learning recommendation. First, the unfamiliar vocabulary used by domain experts when describing learning concepts in teaching materials is usually different from the vocabulary used by learners when looking for materials. This scenario creates a semantic gap between experts and learners (Millard, Gibbins, Michaelides & Weal 2005). This semantic gap makes it difficult to find relevant learning materials for the learners. Second, learners lack sufficient knowledge about the domain they are researching, so are unable to assemble effective keywords that convey what they are intending to learn (Liu, Kim & Creel 2013). This problem results in an intent gap for the learners.

Figure 1.1 illustrates that there exists the semantic and intent gap between domain experts and learners. On one side of the gap, there are experts who are the authors of most of the learning resources used by learners. These experts have a good knowledge of the domain they are working in. However, the vocabulary used by the experts often differs from the vocabulary used by learners. This situation describes the vocabulary mismatch and highlights the semantic gap between domain experts and learners.

On the other side of the gap, there are learners who wish to learn various things. The learning goals of the learners can be expressed through their queries. The fact that they are learners means they are not likely to know how to describe the topics they actually wish to learn about, and this makes it difficult to map learners' queries to relevant learning resources. This situation highlights the intent gap we seek to address in this research.

Figure 1.1: Bridging the gap between learners and domain experts

This research aims to develop methods that support e-Learning Recommendation in order to bridge the intent and semantic gaps between learners and domain experts. This project will investigate how to harness the knowledge of domain experts to develop techniques that can provide a rich vocabulary that can be employed for representing learning resources and for refining learners' queries. The intuition behind our approach is that the domain experts have an idea of the kind of topics that learners in a given learning domain should be interested in. So we can harness the knowledge of experts to help learners find relevant documents.

The knowledge from domain experts can be used as a guide for building background knowledge that captures a good coverage of important learning topics. The vocabulary contained in the background knowledge can be used to support learners when they are trying to find relevant learning materials, given the semantic gap faced by learners. The background knowledge would be useful for learning about relevant topics, this is particularly helpful for learners because of the intent gap they face, as learners often have difficulty determining which topics are relevant.

In building our method, we will leverage the knowledge of domain experts as a guide to build knowledge driven approaches for e-Learning recommendation. DeepQA applies a similar approach to reason on medical reports in order to improve diagnosis in the medical domain (Ferrucci, Levas, Bagchi, Gondek & Mueller 2013). In this thesis, our developed techniques are demonstrated in Machine Learning and Data Mining, however the techniques we describe can be applied to other learning domains.

The following research question will guide the investigation in this research: How can we develop techniques that employ e-Learning recommendation approaches to help learners find relevant documents, given the large amounts of learning materials that are currently available and the difficulty that learners have in asking effective queries? More specifically, the following questions will be addressed:

1. How can we mine expert teaching materials to create background knowledge that supports e-Learning recommendation?

2. How can we capture an effective vocabulary from background knowledge to represent learning materials and make them more accessible to learners?

3. How can we leverage background knowledge to identify important learning topics for refining learners' queries when seeking relevant learning materials?

1.2 Research Aims and Objectives

This research aims to address the semantic and intent gaps faced by learners in order to help learners find relevant learning materials. The research will explore knowledge driven approaches for supporting e-Learning recommendation that would enable learners to find relevant learning materials, given the growing availability of learning resources on the Web. An approach that automatically creates background knowledge in the form of a set of rich learning topics which can be used to support e-Learning will be explored. This thesis will investigate how the background knowledge can be used to address the semantic gap by providing a shared vocabulary for experts and learners, with the aim of making learning materials more accessible to learners. Another challenge this research seeks to address is the intent gap. This is a challenging task because learners are often unable to express their intention when searching for relevant learning materials because of their lack of knowledge of relevant topics. The intent gap is particularly true of learners. So, we will explore effective techniques of addressing the intent gap. The developed techniques will be incorporated in a recommender system, to allow for the demonstration of an e-Learning recommendation task. So that users can make use of the system and judge the relevance of recommendations made to them. This would allow us to measure the effectiveness of the developed techniques.

In order to achieve the aims of this research, the following objectives have been identified:

1. To identify the challenges within e-Learning recommendation by performing a critical review of research in recommendation with a focus on issues in e-Learning recommendation.

2. To address the semantic gap by exploring how to provide a shared vocabulary between domain experts and learners in order to enable learners find relevant materials.

3. To address the intent gap by exploring effective methods to help learners identify relevant topics in order to support learners to ask useful queries when looking for learning materials.

This research will analyse relevant literature to identify the main challenges to focus on. Objective 1 will help with a critical review of relevant literature. In the e-Learning domain, the key components are learning resources and learners. Each of these two components will be further examined in order to address the semantic and intent gaps. Addressing the semantic gap will entail creating a suitable method that will enable the learning materials to be more accessible to learners. So objective 2 will help with addressing this challenge. Objective 3 will entail developing techniques to support learners to ask effective queries when searching for relevant resources.

1.3 Thesis Overview

The rest of the thesis chapters are organised as follows. Chapter 2 provides a critical review of previous work on recommendation. The key aspects involved in recommendation are highlighted. Further, the issues that differentiate e-Learning recommendation from other kinds of recommendation are discussed. The representation of learning resources and query refinement are two issues discussed in this review. Previous approaches used to address such problems are explored.

Chapters 3 and 4 focus on addressing the semantic gap by building background knowledge which contains important learning topics. In chapter 3, we introduce a method that automatically creates background knowledge using the knowledge from domain experts. The background knowledge provides a vocabulary for representing learning materials. The work presented in chapter 3 has been published in (Mbipom, Craw & Massie 2016). In chapter 4 we refine the method used for creating the background knowledge, this results in a richer vocabulary of learning topics. An enhanced method for representing documents is developed. The domain concepts from the richer vocabulary presented in this chapter are used in the rest of this thesis. The work discussed in chapter 4 has been published in (Mbipom, Craw & Massie 2018).

Chapters 5 and 6 aim to address the intent gap by exploring techniques to support learners to identify relevant topics when asking queries. In chapter 5 the rich representation from the previous chapter is used to develop query refinement approaches. Further, this chapter presents an e-Learning recommender system that is developed to demonstrate an e-Learning recommendation task using refined queries. The system architecture and the technologies employed for building the system are presented. In chapter 6, a user evaluation of the developed query refinement method is presented. The e-Learning recommender system developed in the previous chapter is employed for the user evaluation task. The design of the user evaluation, the metrics used and the results are discussed. The evaluation examines the relevance of the recommendations made using the developed query refinement approaches. The work presented in chapters 5 and 6 has been published in (Mbipom, Massie & Craw 2018).

Chapter 7 contains the conclusions which sum up the key contributions and achievements of this research. In addition, the limitations of this research are presented in this chapter, and a discussion of potential future directions for this research.

Chapter 2

Literature Survey

Recommendation enables a user to find relevant items or services from a substantial collection of items. The key components in recommendation are the user, the items for recommendation and a need that the user has. Specifying the need to be met is often a difficult task for the user. A user's need can be specified explicitly or implicitly or through a combination of both. For example, when searching for an item, a user explicitly specifies a need as the input to the search engine, and the item returned to the user should be relevant to that need. In search by exploration systems, a user often has a vague idea or an example of this need, so the user can browse through a collection of available items as a means of clarifying what the need is and what is available in the collection. The user can then select what is suitable from the available items. In recommendation systems, a user's need can be implicitly determined through the preferences, interests or context of that user.

The recommendation approach used can be pull-based or push-based (Chandramouli, Levandoski, Eldawy & Mokbel 2011). The pull-based recommendation method usually requires some input from a user before items are recommended to the user. This is the kind of approach adopted in e-Learning recommendation systems, where a user explicitly presents a need in the form of a query and then receives some recommendations that are relevant to that query. On the other hand, the push-based recommendation systems push items to the user. This action can be based on a number of different factors such as the user's location, the time of day, and the user's interests. For example, a user who is in the city centre of Edinburgh in August, can be recommended some upcoming events to attend around the city, because there are usually a large number of tourists at that time of the year attending the Fringe Festival. The push-based recommendation approach can be beneficial for e-commerce recommender systems, as they can influence users who receive such push notifications to actually purchase the item.

One challenge associated with the pull-based approach is that a recommendation is usually triggered by a request in the form of a query from a user. This poses a problem particularly for e-Learning recommendation systems, because learners are not always able to clearly specify what they need as they are often new to the topic they are trying to learn about. So crafting effective queries is difficult for learners. This challenge will be further explored.

The content-based and collaborative filtering methods are two ways often used to filter items when making recommendations to users (Kantor, Rokach, Ricci & Shapira 2011). The content-based approach compares the content of items with the user's need before recommending items to users. The content-based approach is useful for e-Learning recommendation which aims to meet the learner's need captured in the query. Although the content-based approach typically assumes that a user would most likely be interested in similar items to those liked previously, in e-Learning recommendation, this is not the case because one does not want to recommend learning resources that are similar to those previously read by a learner. Instead, the learner's need is used as the basis for finding relevant items. On the other hand, the collaborative filtering method relies on the preferences of similar users when making predictions for a user. This approach is not helpful for e-Learning recommendation because each user's need has to be taken into account.

This chapter presents an overview of recommendation systems underlining the main aspects of such systems. A landscape of e-Learning Recommendation is presented highlighting the key components involved. Further, some challenges associated with e-Learning recommendation are explored, and the approaches usually employed to address the identified challenges are examined.

2.1 Recommendation

Recommendation entails predicting relevant items for a user from a large collection of items. Systems that employ recommendation techniques are called recommender systems. These systems are very useful when the number of items to select from are more than the user's ability to search through them (Jannach, Zanker, Felfernig & Friedrich 2010). Hence, recommender systems are said to assist in managing the information overload problem by offering users only relevant items from a substantial amount of items (Schafer, Konstan & Riedi 1999, Park, Kim, Choi & Kim 2012). For example, a user looking for a book on Amazon may not be able to ask an effective query that conveys what they need. So the recommendation system in Amazon suggests books that may be relevant to meet the user's need from the collection of books available. Recommender systems have been widely applied in movie recommendation (Harper & Konstan 2016); music recommendation (Horsburgh, Craw & Massie 2015); and in e-commerce (Linden, Smith & York 2003). Such systems are able to make predictions based on the interests and preferences of users, often captured through ratings or user activities. The focus of this research is on e-Learning recommendation (Manouselis, Drachsler, Verbert & Duval 2013). In e-Learning recommendation, one cannot rely on the preferences or interests of learners for making recommendations, because the recommendations are usually dependent on the need of the user captured in a query.

The key components involved in recommendation are illustrated in Figure 2.1. A user has features such as the needs, preferences and interests that often influence the kind of items that are recommended to the user. Another key component of recommendation is the items. The items have a fixed set of features that can be used to describe them. For example, a camera can have features such as the price, brand, battery type, and megapixels. An item's features are quite useful for identifying the item when making predictions. Some items can also contain metadata which is additional information about the item. For example, a video can have metadata such as a creator, duration and genre. In addition to the features and metadata, some items can have a description which contains more information about the item. This description is usually presented as text, and it can be difficult to use because of the challenges of working with unstructured data.

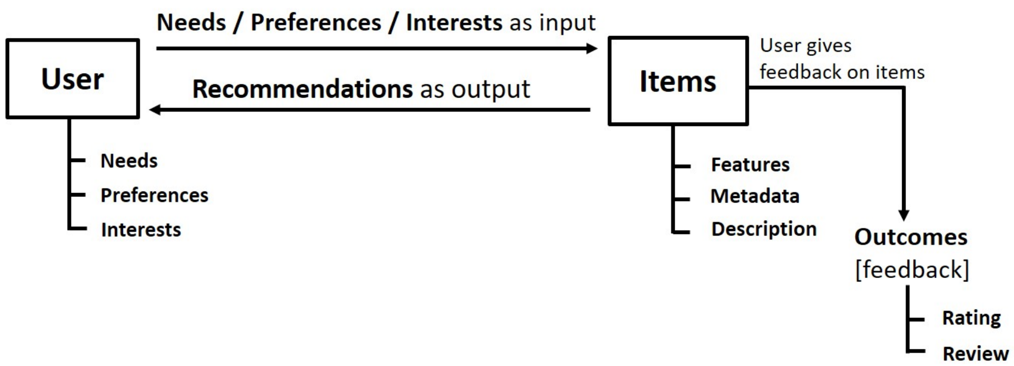


Figure 2.1: Overview of recommendation

After a user interacts with recommended items, the user can give some feedback in the form of a rating or review. The feedback is an outcome that acts as a measure of the user's satisfaction with the recommendations. A user's rating can be used for determining the preferences of the user, because a rating captures what items the user liked or disliked. Similarly, the rating can be used to identify users who have similar tastes or interests, and recommendations can be made to such users on the basis of their similarity to other users. A review can also be used to determine the important aspects of an item. For example, product reviews are employed in (Dong, Schaal, O'Mahony, McCarthy & Smyth 2013) to create a feature representation for products before they are recommended. The reviews are employed because, users would mention features that are important to them within the reviews.

2.1.1 Users

Users are the people who employ the services offered by recommender systems. Each user has certain preferences which determines the kind of recommendations that are made to that user. A user can also have similar preferences to other users. There are typically two types of users, these are the white-sheep and the gray-sheep users (Ghazanfar & Prgel-Bennett 2014). The white-sheep users have similar preferences with other users, so they have a high correlation with other users, while the gray-sheep users often have a low correlation with other users. It is often more difficult to predict relevant items for the gray-sheep users because their preferences are often different from the preferences of other users, so it is hard to find users that are similar to them.

A recommendation approach that makes use of the similarity between users for making predictions is the collaborative filtering (CF) approach (Ricci, Rokach & Shapira 2011). In this approach an item can be recommended to a user if other users with similar preferences have consumed the item and given it a good rating. Hence we have, "users who like this, also like that". In this approach, the outcomes such as ratings or reviews associated with an item is often quite useful when determining the recommendations to make to other users.

The CF approach can be item-based as employed in Amazon (Linden et al. 2003), where the method aims to find other similar items and use this as a basis for making predictions; or it can be user-based as employed in MovieLens (Harper & Konstan 2016), where the aim of the method is to find similar users and make predictions based on their preferences. Amazon employs an item-to-item collaborative filtering approach when making recommendations to a user. So, the correlation between items in a ratings matrix is computed. Items that have a high correlation with the items already purchased by the user are then recommended to the user. Amazon introduced the item-based approach as a means of tackling the scalability problems they faced with user-based CF. The number of items is often less than the number of users in a recommender, so it is more scalable to compute the similarity between items and use this for recommendation. Items do not change as often as users, so item-item similarity can be calculated prior to recommendation and used when required.

MovieLens employs a Collaborative filtering approach when recommending movies to its users. Figure 2.2 shows a screen shot from MovieLens recommender system. A user's preferences are collected through their ratings provided for movies, an example is shown in the section labelled "your ratings". The preferences that the user has provided is then used to identify other users that have also watched and provided ratings for the movies. So similar users can be identified based on the kind of movies they have watched. Recommendations can then be made to the user based on the movies that other users with similar preferences liked.

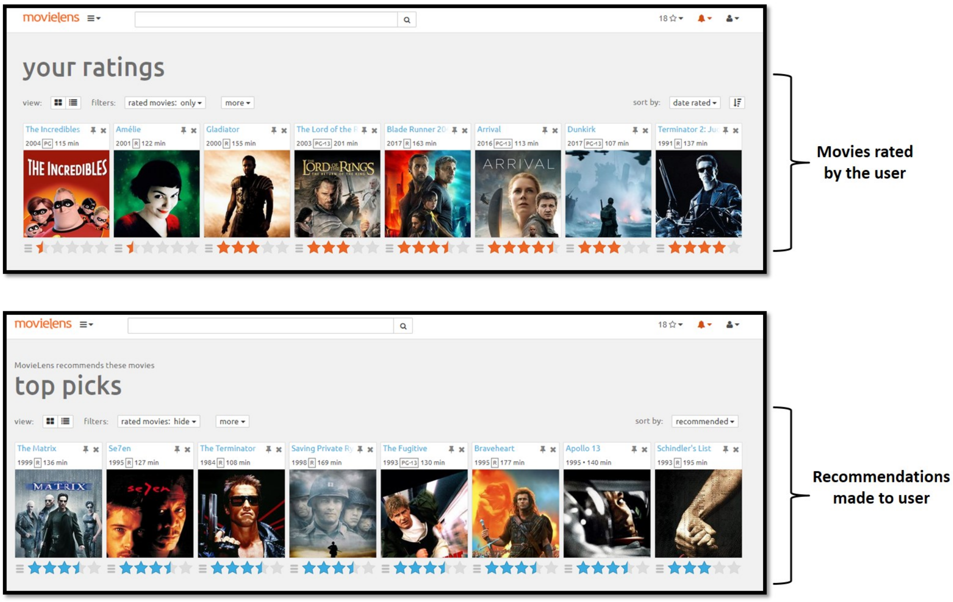


Figure 2.2: Collaborative filtering approach adapted from MovieLens

Data sparsity is one issue that affects recommender systems that use the CF approach, because this approach relies on large amounts of data to make suitable predictions (Hu, Li & Chao 2012). This data sparsity problem can result in cold-start when trying to make recommendations. The issue of cold-start can happen in either of two ways (Lika, Kolomvatsos & Hadjiefthymiades 2014). One way cold start presents itself is through the new-user problem (Adomavicius & Tuzhilin 2005), when a new user recently joins a recommender system, and has provided insufficient ratings. For example, a new user, Alice may have provided a similar rating to that provided by Bob, and a different rating to another user Carol, for an item they all purchased. The next time the system makes a recommendation for Alice, Bob's preferences would be used to make the predictions for Alice, because of the limited information available for Alice. When this happens, it is difficult for the system to make suitable recommendations for Alice, because of the few ratings she has provided. So the limited information about a new user causes the new user problem.

Another way cold-start happens is through the new-item problem (Hauger, Tso & Schmidt-Thieme 2008). For example, a movie recommender using a CF approach may not recognise that users who rated "*Guardians of the Galaxy*" highly may also like a newly released film, "*Thor: Ragnarok*", even if both films have a Superhero fiction genre and are from Marvel studios. This is because the newer film has insufficient ratings, for the system to make an effective prediction about it. This example, highlights the new-item problem of the CF approach in recommenders.

2.1.2 Items

Items represent various products or services that recommender systems offer to users. Items may be journal articles such as those recommended by CiteSeerx (Li, Councill, Lee & Giles 2006), or movies, such as those recommended by Netflix (Bennett & Lanning 2007), or the videos recommended by YouTube (Burgess & Green 2009). For each of these examples mentioned, there exists a large number of items to choose from, hence the need for a recommender system. The description of an item is important in certain recommendation scenarios and this may form the basis of recommendation for that item. In such a scenario, items that are similar to what a user has consumed previously may be recommended. This approach is employed in NewsWeeder (Lang 1995), where similar news items are recommended to a user based on previously read news items.

A recommendation approach that uses the description of items for making predictions is the content-based recommendation approach (Lops, Gemmis & Semeraro 2011). In this approach, a new item is recommended to a user if the description of the item is similar to items that the user has consumed previously. So the features of items play a key role in this approach when trying to make recommendations. NewsWeeder employs a content-based approach for recommending news items (Lang 1995). In NewsWeeder, a user navigates through the system and selects a news item of interest and reads it. The system then builds a profile of this user based on the news items which the user likes and offers more news items in similar categories to the user. In the content-based approach, it is likely that a user who usually reads articles about technology, would be interested in reading articles containing the latest trends in the technology category, so the system can make such recommendations to the user. Figure 2.3 illustrates that ScienceDirect also uses a content-Based approach for recommending related journal articles to a user (Tenopir, Wang, Zhang, Simmons & Pollard 2008).

In the content-based recommendation approach, the content descriptions of the items is an important feature to use for making predictions. Figure 2.3 shows the system recommending other

Figure 2.3: Content-based recommendation approach adapted from ScienceDirect

similar articles to a user who had previously used the article “Recommender systems survey”. Notice that the original article the user read contained the keywords : “Recommender” and “system”. So when new articles are being recommended to this user, it is observed that the three new articles recommended all contain these keywords. This example thus highlights the importance of the description of items in this approach. So a suitable representation of the item would be useful to enable effective recommendation.

In NewsWeeder and ScienceDirect, the recommendations made to users are usually based on the history of what the user has already consumed. The new recommendations are often more items related to the previous item. In e-Learning, a learner would not necessarily be looking to see more of the same kind of resources that had been read before, as this can make the learner lose interest in the course. However, a knowledge of the features of the learning materials that a learner has consumed previously can be useful in e-Learning for selecting a set of suitable recommendations for that learner. For example, if a learner preferred learning materials that were presented as videos, it could be an indication of the kind of materials to choose when making recommendations for such a learner. So, the content-based approach would be useful for exploiting the content of learning materials when making recommendations to learners.

One issue that recommender systems face is the “long tail” problem. This is usually caused by many items that have little or no ratings, thus making it almost impossible to select such items for recommendation because they are not very popular. Park & Tuzhilin (2008) propose a method for leveraging the items in the long tail, by clustering them and using the combined ratings for such items rather than the few ratings per item. Other researchers employ the content of items to determine the similarity of items that have low ratings (Schein, Popescul, Ungar & Pennock 2002). Successfully leveraging learning materials in the "long tail" can be seen as introducing some diversity in recommendations, and this can be useful for e-Learning recommendation, because predicting only popular materials can be boring for a learner.

2.1.3 Recommendation Issues for e-Learning

The data driven nature of the recommendation approaches pose challenges for e-Learning recommendation. A drawback of the collaborative filtering approach lies in the difficulty of getting adequate number of learners to capture sufficient feedback. In (Zaíane 2002), a CF method is adopted. Association rule mining techniques are used to build a model that captures the interactions of learners within an e-Learning system. Association rules are used here because of their success in analysing market baskets, such as: "users who bought this, also bought that". In the e-Learning domain, similar approaches can be used to predict the next relevant part of the system to navigate to, based on what other similar learners usually do. So, the developed model is then used to suggest other learning materials and learning paths through the system to a learner, in order to improve the learner's performance and navigation within the system, and enable them to find relevant learning materials.

A collaborative filtering approach is also used in Tan, Guo & Li (2008) to design an e-Learning recommender. In developing this system, one assumption made is that learners that have similar browsing or feedback patterns would be interested in similar materials. The similarity between learners is computed based on their browsing patterns and feedback, and this is used as a basis for recommending courses that other similar learners had studied. One drawback of the CF method is that in e-Learning recommendation, the similarity between learners found in feedback such as ratings is not sufficient for making recommendation to other learners. This is because the new learner often has learning goals which should be considered when making recommendations. Another limitation of systems that use CF techniques is that the features of the items for recommendation are not considered, because the emphasis in this approach is the preference of users captured through their ratings.

On the other hand, the content-based recommendation approach often learns from a large number of learning resources before making effective predictions to a learner. A drawback of the content-based method is that it has no knowledge of history of other learners, and it is not personalised. This approach does not learn from previous recommendations made, so the same recommendation can be made to different learners if the document seems appropriate for the need of a particular learner, even if such a document was not useful to previous learners. One solution to this would be a way of capturing feedback such as ratings from users, and incorporating it in the recommendation process. In (Ghauth & Abdullah 2009), a content-based recommendation method that incorporates feedback when recommending similar learning materials to learners is presented. In their work, learning materials that are similar in content to previously used materials and have been rated highly by other learners are recommended. So, elements of the content-based and collaborative filtering approaches are combined for recommendation.

Another drawback of systems that implement the content-based recommendation approach is that the feedback from other similar users is not primarily taken into account, because the focus of the content-based approach is on the item description used for recommendation. A way of addressing these limitations has given rise to the development of hybrid recommendation approaches (Burke 2002), that aim to suitably combine the strengths of the different approaches.

Hybrid recommendation approaches are often achieved by exploiting item descriptions as demonstrated in the content-based recommendation approach in order to cater for the insufficient user-rating information provided by the CF approach. Melville, Mooney & Nagarajan (2002) combine CF and content-based approaches to develop a hybrid method as a way of addressing the data sparsity problem that affects the CF approach. The rating data is combined with content-based predictions for movies, thus enriching the data that is available for making predictions. This produces better results than when any of the separate recommendation approaches are used alone. Tkalcic, Kunaver, Košir & Tasic (2011) make the assumption that users with similar personalities should also have similar preferences for items. So the features of the users are employed for computing the similarity between users even before an item is rated by the user. Thus providing more data to use for making predictions.

2.2 e-Learning Recommendation

e-Learning recommendation typically involves a query received from the learner, as an input; a collection of learning resources as the items from which to make recommendations; and selected resources recommended to the learner, as an output. The User-Items pair in recommender systems can be mapped to the Learner-Learning Resources pair in e-Learning recommender systems as illustrated in Figure 2.4 which is the e-Learning version of Figure 2.1. Each learner has features such as the Goals, which is what the learner hopes to achieve after learning a topic. A learner also has abilities and prior knowledge which can determine what that learner is able to learn currently. For example, a learner may have achieved a previous qualification which means the learner is competent in some field of learning. The learner can also have preferences and learning styles which means the learner would prefer certain kinds of learning materials. For example, if a learner was a visual learner, then such a learner would prefer materials such as videos.

Learning styles can be a controversial issue because some learners can demonstrate certain styles given the kind of learning materials that are available to them, hence such learners tend to develop a coping strategy to help them interact with the learning materials. On the other hand, some learners may have a preference for materials that are presented to them in certain ways, and this can impact the way that they learn. In such scenarios, capturing learning styles has a potential to yield some benefit for helping the learner to comprehend the content of the learning materials.

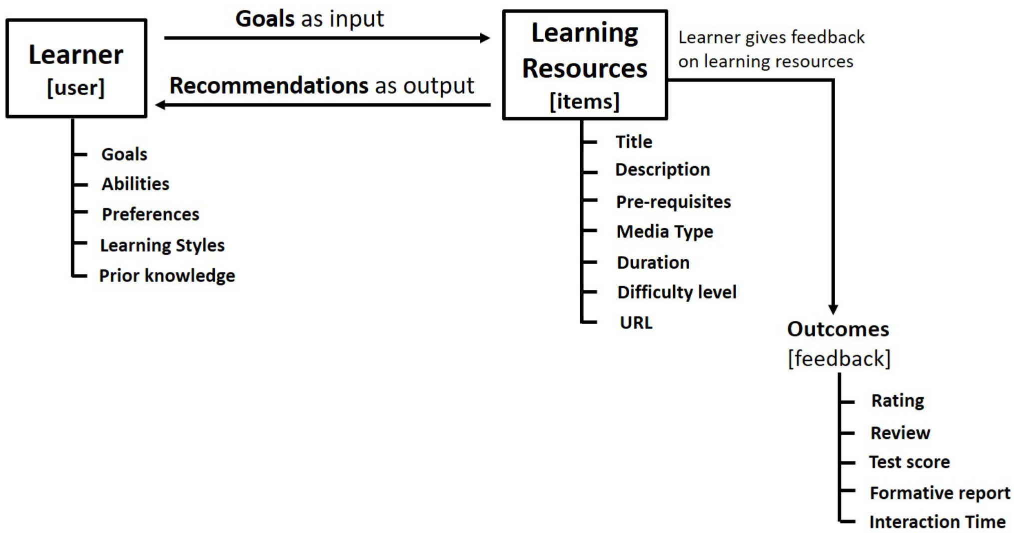


Figure 2.4: Landscape of e-Learning recommendation

Learning resources are another key component of the e-Learning recommendation landscape illustrated in Figure 2.4. The learning materials are the items for recommendation which contain what a learner wishes to learn. The learning resources often have features that can be used to describe it. These features will be further explored.

Outcomes are the feedback produced from the learner's interaction with the recommended learning resources, such as a rating or review. Although this component is generated after recommendations have been made, there is the potential to use outcomes such as ratings within the collaborative filtering approach where the rating patterns of users helps to determine their preferences, which are then used as the basis of recommendation. The learners, learning resources and outcomes shown in Figure 2.4 will be discussed in the following subsections.

2.2.1 Learners

Learners are the users in an e-Learning recommender system. Learners have a complex set of features that can be used to describe them, as illustrated in Figure 2.4. The features of a learner that can be considered in e-Learning is different from the features considered for users in other forms of recommendation such as Music or e-commerce because of the nature of e-Learning recommendation. A way of collecting the information about a learner is through Learner modelling, which entails creating a representation that captures the information associated with a learner. The learner model is generated to assist in making useful recommendations to the learner (Shiri, Aïmeur & Frasson 1998), because it can contain information such as the goals, preferences and learning styles of a learner. The learner model is useful for the CF approach where the similarity between learners is important for making predictions. In e-Learning recommendation, learners are more willing to invest some time to help generate such information, because of the investment they will be making when learning. In contrast, for e-commerce, users are often trying to purchase items quickly, and can be put off if they have to spend time generating information about themselves. Various authors differ in the level of detail and in the approach used to capture information for modelling learners. The 3 approaches often used to model learners are the static, dynamic and hybrid approaches.

In the **static learner modelling** approach, questionnaires can be used to gather learner information prior to making recommendations to learners. The learners in (Hothi & Hall 1998) completed a questionnaire containing questions such as: skill level on the Web, Microsoft programs used, course specific topics, number of hours were spent weekly using a computer, and watching TV. The results from this were used to assign learners to one of 4 pre-defined groups, describing a learner as a novice or expert on the system based on their knowledge of the system or the content shown to them. The learner model was used to determine what learning resources were given to learners in each group. Using the static learner model, it was difficult to correctly classify learners who were neither experts nor novices, because they could not successfully determine the ability of a learner using this approach. However, they found that learners in similar groups had similar working patterns. This finding is useful for collaborative filtering methods where the behaviours of similar users can form the basis of the recommendations made to learners.

The learning style of learners is captured in (Peter, Bacon & Dastbaz 2010) using the VARK model. This is because the VARK model has a distinct classification of learning preferences such as Visual, Aural, Read/Write and Kinesthetic. These preferences can be used for selecting suitable learning materials for learners. Having a knowledge of the learning style of learners can influence the kind of resources recommended to learners because the learning style captures the preferences of learners in terms of the kind of materials that a given learner would be interested in. This can be used in a CF approach where learners that have similar preferences can be recommended learning materials that other similar learners have liked.

Existing questionnaires can also be used to gather learner information. The Unified Learning Style Model (ULSM) presented in (Popescu 2010) provides a means of capturing the learning preferences of learners in categories such as their preference in processing and organisation of information, their perception such as visual/verbal, and their pacing. The model is created from a combination of the well known models such as the Felder and Silverman model (Felder & Silverman 1988) and the VARK model (Fleming 1995), thus providing a single model that can be employed for capturing information about a learner. Although an existing questionnaire can be task specific and not easily adaptable, employing it allows one to reuse the questions that have been used to generate learner information previously thus reducing the time and effort required in creating a new one. In (Giuliano, Moser, Poremba, Jones, Martin & Slaughter 2014), the ULSM is adapted and used together with the Learning Preference Questionnaire to help identify the preferences of learners in a Pharmacy course. The identified preferences were applied to capture the study techniques of the learners.

One problem of the static approach is the kind of information generated. Fixed learner models are created which often capture the description of a few learners because it is challenging to find a large number of users for capturing their data. The static approach to modelling learners is quite limited in its functionality because one relies on the learner for generating the information. However, it can be a starting point when creating a representation of the learner, so that such information can be used to identify similar learners and make recommendations based on their preference. This is particularly useful for systems that adopt collaborative filtering approaches.

In the **dynamic learner modelling** approach, the activities of a learner are assessed while learning. The observed interactions can then be used to build a model of that learner. This can be done through tracing by keeping a record of a learner's interactions during a learning process (Cordier, Mascret & Mille 2009). A learner's actions can be observed during online courses as demonstrated in e-Teacher (Schiaffino, Garcia & Amandi 2008). Similarly, a log of the learner's activities can be generated as employed in (Conati & Kardan 2013). The possibility of modelling the behaviour of a learner while navigating through an e-Learning environment is presented in (Conati & Kardan 2013). This behaviour captures information about the learning ability of the learner. The learners can then be classified based on their ability. A CF recommendation approach would benefit from such a method, where similar users can be determined using the information captured in their learner models. This information can then be used to make recommendations for other new learners.

Although the dynamic learner modelling approach allows feedback to be provided to learners during the learning process (Benabdellah, Gharbi & Bellajkih 2013), one problem of using dynamic learner modelling only is that a learner's background information, and prior skills are not captured (Brusilovsky 1996). de Rosis, De Carolis & Pizzutilo (1994) suggests that it is better to elicit such information from the learner as a means of improving the reliability of the system.

The **hybrid learner modelling** allows prior knowledge about a learner to be captured using its static component as presented in (González, Burguillo & Llamas 2006, Kritikou, Demestichas, Adamopoulou, Demestichas, Theologou & Paradia 2008) while its dynamic component allows a learner's interactions within the learning system to be evaluated as described in (Benabdellah et al. 2013). Figure 2.5 illustrates a hybrid approach which would suitably combine the strengths of the static and dynamic approaches.

A hybrid approach can begin by gathering information from a new learner explicitly as employed in the static method. The learner's performance on tasks is then observed with the help of the dynamic method. The results from the static and dynamic methods are combined to create a hybrid model of the learner. This model can then be used to recommend suitable materials for such a learner. An iterative approach can involve an assessment of the learner's performance after the recommendation was made. For example, Klašnja-Milićević, Vesin, Ivanović & Budimac (2011) present a system that uses the FLSM model to capture the learning style of a learner based on how they process, perceive, receive, and understand information. The learners are then clustered based on their respective learning styles. Each learner's activity within the learning system is observed from web logs. The system gives each learner a rating to indicate their level of knowledge for a given learning material. The system assigned ratings for learners are compared to generate the

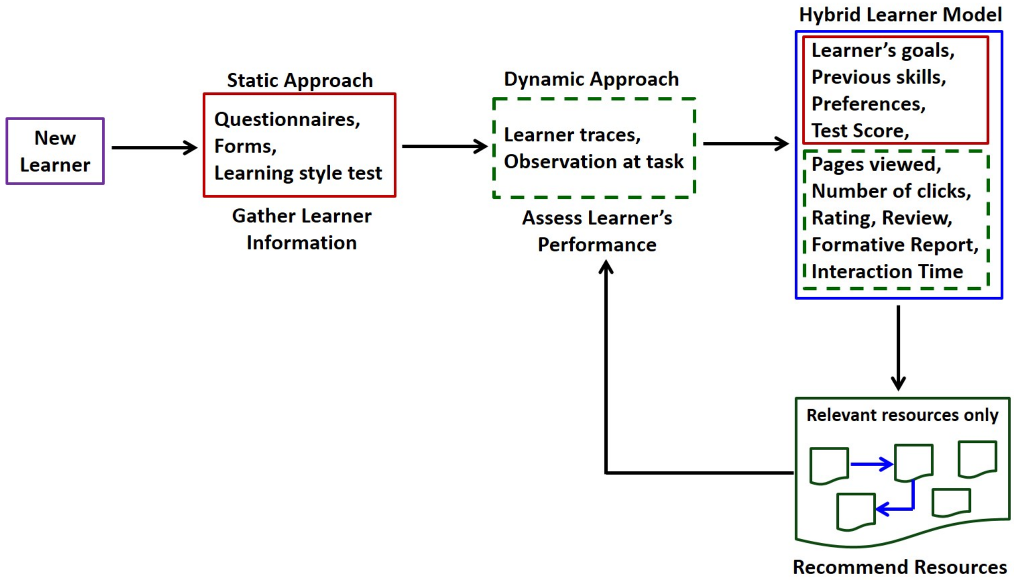


Figure 2.5: Hybrid approach to learner modelling

similarity between learners. The system then recommends relevant links to a learner based on how similar that learner is to other learners. An advantage of the hybrid learner modelling approach is that recommendations made to a learner can be adapted based on the observations made.

One challenge with learner modelling in general is that a large amount of information has to be captured from many learners to build a meaningful representation that can be used to make effective recommendations (Di Bitonto, Laterza, Roselli & Rossano 2010). This presents a problem because it is not always possible to have a large number of learners. Further creating learner models can be demanding as it involves a substantial amount of learner input.

2.2.2 Learning Resources

Learning resources are the items for recommendation. Each Learning resource can be described using its content or metadata features such as its title, description, publication date, author, duration, pre-requisites (Shen & Shen 2004), duration, difficulty level (Yarandi, Tawil & Jahankhani 2011), URL, and media type. The media type can be text such as PowerPoint presentations (Ghauth & Abdullah 2009), or videos such as webcasts or YouTube videos.

The metadata associated with the resources is often used as a means of describing the learning materials. Some metadata standards often employed are the Dublin Core Metadata Standard (Weibel, Kunze, Lagoze & Wolf 1998) and the IEEE Learning Object Model (LOM) (IEEE 2002). The metadata allows learning resources to be identified more easily and enhances their reuse

(Hakala 2000). In (Anibaldi, Jaques, Celli, Stellato & Keizer 2015) Dublin Core elements are used to develop metadata for agricultural publications on the AGRIS repository (Fogarolli, Brickley, Anibaldi & Keizer 2011). Although Dublin Core elements are useful for defining metadata, its structure and depth depends on the developer using it. In Holzinger, Kleinberger & Müller (2001), the IEEE LOM is used to describe learning materials. However the complex nature of the model can present a challenge for its users, in determining the entities to use for optimal results.

Another standard often used to build e-Learning resources is the Sharable Content Object Reference Model (SCORM) presented by Bohl, Scheuhase, Sengler & Winand (2002). The SCORM allows resources to be compatible and reusable across different learning management systems (LMS). SCORM allows the interactions of learners with resources to be captured, and this is useful for generating data about the engagement of learners with resources.

The metadata standards are useful guides for defining structured data that describe learning resources, however only relevant elements from these standards may be included by authors because of the complexity of most of the standards.

Two key features of learning resources which often contain text that can be used for representing the learning materials are its title and description. Dealing with this text is challenging because of its unstructured nature. This makes e-Learning recommendation challenging because the descriptive features of learning resources need to be suitably indexed for effective recommendation, unlike products in e-commerce recommender systems which often have a fixed set of features. Usually, when learners try to find the right learning resources, they do not often use the right vocabulary to search for the resources. This presents a semantic gap, where a different vocabulary is used by learners and the authors of learning materials. A possible solution to addressing the semantic gap is by building a suitable method for representing learning materials in order to enhance the retrieval and recommendation of relevant resources to learners.

2.2.3 Outcomes

Outcomes capture the interactions of users with items, giving an indication of how suitable recommended items were for users. For example, the feedback from users in e-commerce systems can be captured in product reviews, and these reviews represent the outcome after using the product. The user reviews can be reused when deciding what products to recommend to other users. For example in (Chaves, Gomes & Pedron 2012), 1500 reviews from 50 hotels were analysed and this showed 3 key features that were often mentioned by users across the reviews, thus giving an indication of the features that were important to users. The results of the study highlighted the features that managers should focus on.

In the e-Learning domain an outcome can be seen as an indication of how suitable a learning resource is for a learner. For example, the feedback generated after a learner has read a recommended resource forms the outcome. Table 2.1 contains the outcomes that exist in recommenders. The first column of the table shows the type of outcome, the second column states the domain the outcome is commonly used in, the third column states if it is easy to capture the outcome, the fourth column states if the outcome is knowledge rich and the last column contains some references as examples of where such outcomes are used. Ratings and reviews are often used in e-commerce recommenders, but they have the potential of enhancing the recommendations made in the e-Learning domain (Bobadilla, Hernando & Arroyo 2011). A test score, a formative report, and the interaction time are often used in e-Learning systems (Kritikou et al. 2008, Vasilyeva, Pechenizkiy & De Bra 2008, Benabdellah et al. 2013).

Table 2.1: Types of outcomes in recommenders

Type	**Domain**	**Easy to Capture**	**Knowledge Rich**	**References**
Ratings	e-Learning/ e-commerce	Yes	No	(Hu & Liu 2004, Bennett & Lanning 2007, Bobadilla et al. 2011)
Reviews	e-Learning/ e-commerce	No	Yes	(O'Mahony & Smyth 2010, Aciar, Zhang, Simoff & Debenham 2007, Chaves et al. 2012)
Test Score	e-Learning	Yes	No	(Kritikou et al. 2008, Vasilyeva et al. 2008, Benabdellah et al. 2013)
Formative Report	e-Learning	No	Yes	(Wang 2007)
Interaction Time	e-Learning	Yes	No	(Conati & Kardan 2013, Biswas 2012)

A rating: captures a user's assessment on an item that the user has consumed. The most basic form of a rating can be "thumbs up" or "thumbs down" (Hu & Liu 2004). Where thumbs up indicates that the user likes the recommendation and thumbs down indicates otherwise. It may be more refined as a score on a given scale. The ratings of items by users can be used in a collaborative filtering approach to recommend to users as discussed in section 2.1.1.

In the e-Learning domain, a rating captures a learner's assessment on a learning resource. It is represented as a score on a given scale. In (Bobadilla et al. 2011), a scale from 0 to 5 is used to capture the views of learners on certain aspects of their learning system as well as the performance of the lecturers. It is noted that a rating is a learner's means of informing others about their own learning experience. This is useful for systems that apply the CF approach, because ratings can be used to capture the preferences of learners. Although a rating is often easier to collect from users than other forms of outcomes, a rating is not knowledge rich because a single score is used as a summary of all the different aspects of an item.

A review: is the means by which a user evaluates the recommendations made. It captures a user's opinion on different aspects of an item that a user has interacted with. A review is richer than a rating because it contains more information and the user can give reasons for liking or disliking a particular part of an item or the entire item. However it is more difficult to capture and use because of the different forms which they can take. For example Hu & Liu (2004) summarize different reviews of users by classifying the reviews and also extracting product features from them. The text in a review may also be analysed using sentiment analysis to determine if the review was positive or negative (de Albornoz, Plaza, Gervás & Díaz 2011, Pang & Lee 2008).

Unlike a rating, a review may contain reasons why a recommendation may be highly or poorly rated. In (Aciar et al. 2007), the reviews of users are captured and reused to make better recommendations to other users. Important item features are often contained in reviews and this may be an indication of the features producers should focus on, for example in (Chaves et al. 2012) a study involving 1500 hotel reviews revealed three features that were often used across the reviews. Similarly, in (Dong, O'Mahony, Schaal, McCarthy & Smyth 2013) product features are extracted from user reviews and used to enhance the recommendations made to other users. In e-Learning, a learner's review on a resource is important because it is an indication of a learner's experience while interacting with the resource and potentially it can be used when recommending that resource to other similar learners.

A test score: is the result of a test or an assessment and it is also referred to as knowledge-of-response by Vasilyeva et al. (2008), because it indicates a proportion of the learner's responses that are correct. In an e-Learning scenario, a test score captures the performance of a learner after interacting with a resource (Kritikou et al. 2008). Although a test score is objective and easy to capture (Benabdellah et al. 2013), it is not knowledge rich because a low test score may only suggest that a resource was not suitable for a learner but it does not say which parts were difficult. In order to reuse this test score, a learning resource associated with a high test score may be recommended to learners with similar preferences, because a similar learner achieved a high test score from using the resource. However, in e-Learning recommendation a learner's goals should be considered before recommendations are made.

A formative report: is usually generated after an assessment or test has been taken. It is similar to a test score because it also captures a learner's performance on a resource. However, a formative report gives a more detailed description about how suitable a particular resource was for a given learner. Although a formative report can be demanding to capture, it is a good indication of which parts of a recommendation a learner may have had difficulty with. For example, if a formative report indicates that a given learner had difficulty with a particular resource, when reusing this experience, the resource recommended to other new learners may be different because of the knowledge contained in the outcome associated with the experience of a previous learner.

Interaction time: This captures the actual time a learner spent interacting with a given resource. In capturing interaction time, the time interval between a learner's activities within a system can be captured as demonstrated in (Conati & Kardan 2013), alternatively the time learners take to complete given tasks can also be recorded (Biswas 2012). The various durations captured can be compared with threshold values and this can be used to give useful feedback regarding a learner's interaction with a learning material as is done in (Kritikou et al. 2008). For example, if a learner spent longer time than expected on a resource, it may be an indication that the learner found the material interesting. Using interaction time as feedback should be done with caution because, it does not capture situations such as if a learner was distracted during learning.

An outcome is a form of feedback which gives an indication of how useful a recommendation was to a user. Although the feedback captures a single user's interaction with a recommended item, there is the potential to capture the feedback from many learners and reuse it to help other new learners. However, this would require inputs from many learners before sufficient feedback is generated. Alternatively, recommendation approaches that rely on the content of learning materials rather than the outcomes of many learners can be adopted.

2.2.4 Challenges with e-Learning Recommendation

Finding relevant materials for a learner is a more challenging task than recommending items to users because of the complex set of features that can be used to represent learners and learning resources. One key feature that learners have is their learning goals. A learner's goals can be captured through the query which is the input to the recommender system. The query gives us an idea of what the learner wants to achieve. However, learners do not often know how to effectively convey what they wish to learn because they lack sufficient knowledge of the domain. This highlights the intent gap faced by learners, and presents a challenge for successful recommendation to be made. One way of addressing this challenge is by creating a method that supports the refinement of queries by representing the query using the vocabulary of the domain, with the aim of making relevant recommendations to learners. In other forms of recommendation, a user does not necessarily require goals. This is because the user's preferences or browsing history is sufficient for making predictions for the user.

In e-Learning recommendation, the preferences of a user captured in user-user similarity is usually not sufficient for making predictions. Often, the goals of that learner have to be taken into account for a relevant recommendation to be made. So this distinguishes e-Learning recommendation from other standard forms of recommendation where a CF method can be applied to recommend materials that learners with similar preferences rated highly. Further, the prior knowledge and level of expertise that a learner has is also a feature that can determine the kind of recommendations made, as these skills need to be considered to provide learning materials at the right level to the learner.

The time commitment that learners invest to study a course is another important factor which makes e-Learning recommendation challenging. The time cannot be recovered if the learning material recommended was not suitable for the learners. There is a bigger commitment made by learners to learn a new topic, than the time commitment made when buying an item in e-commerce, or when listening to a new track in music recommendation. For e-commerce, the user has the option of returning the item if the user is not pleased, but in e-Learning the time spent in studying an unsuitable material is lost.

In content-based systems such as NewsWeeder, users often want more of the same kind of news article. Also in movie recommenders, users may want more of the same genre of movies. However, in e-Learning recommendation, learners need different learning materials to the ones already studied. However, if learning styles are used in an e-Learning system, then the resources that suit the learning styles of the learner should be considered.

Learning resources are not like products in e-commerce that have a fixed set of features that describe them. Instead, a large number of learning resources are largely text, which is unstructured data. This makes creating representations for learning materials challenging, unlike feature representations for products. The content-based recommendation approach which makes predictions based on item descriptions can be used to handle e-Learning recommendation. The next section will further explore some of these challenges, and propose ways by which they can be addressed.

2.3 Addressing e-Learning Recommendation Challenges

Learners represent the users of an e-Learning recommendation system. Learners usually have learning goals that they wish to achieve when using an e-Learning system. These goals can be captured through queries. However, learners often have difficulty asking an effective query that conveys what their learning goals are because of two reasons. First, they lack sufficient knowledge about the domain they are researching, so are unable to assemble effective keywords that identify what they wish to learn (Liu et al. 2013). This problem results in an intent gap. Second, the vocabulary used by teaching experts is often different from that used by learners (Millard et al. 2005). This presents a semantic gap.

Learning resources contain the information a learner wishes to learn. Learning resources are often unstructured text, and are not suitably indexed for retrieval, thus making e-Learning recommendation a challenging task. A possible solution to addressing this challenge is the creation of effective representations that capture the important concepts within learning resources. However, building suitable representations for learning resources in e-Learning environments is a difficult task, largely because of the challenge of dealing with unstructured text. The approaches used for refining learners' queries and representing learning resources will be explored in the following sections.

2.3.1 Refining Learners' Queries

Refining a learner's query entails modifying the initial query into a new query that better captures the learner's goals. Figure 2.6 illustrates two key methods often used for refining queries. These methods typically involve using internal knowledge sources or using external knowledge sources for the refinement process. The lower segment of Figure 2.6 shows the knowledge sources usually employed by each of the two main approaches. A document collection such as one containing learning resources is often used as the internal knowledge source while a domain knowledge source such as Wikipedia or DBpedia can be used as external knowledge sources. Both methods can be implemented manually or automatically, with explicit or implicit feedback required. These approaches are further discussed in the following sections.

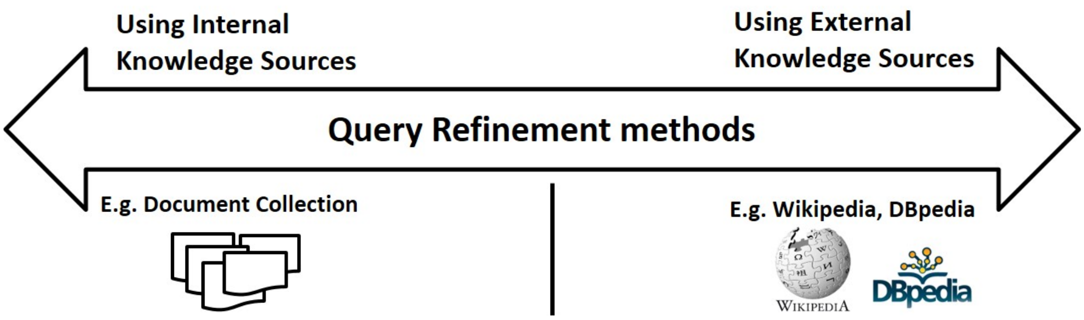


Figure 2.6: Approaches used for query refinement

Using Internal Knowledge Sources to Refine Queries

One approach to query refinement can be done introspectively, where internal knowledge from a document collection is used as a means of feedback. The feedback can be generated explicitly or implicitly and then used to refine the query. Generating feedback explicitly involves some input from the learner. In the explicit approach, when a query is received from a learner, an initial set of documents considered to be relevant to the query are retrieved. The learner has to explicitly judge which documents were relevant by selecting such documents. Based on the learner's input, terms from the selected documents can then be used to refine the initial query, and a second search performed on the document collection using the refined query. Although this method has the potential to improve relevance, the drawback of this approach is the user involvement needed for relevance judgement. This is because it can be challenging to have users generate feedback during a search process (Anick 2003).

On the other hand, the feedback can be generated implicitly. This implicit approach would involve conducting a search on a collection of documents, and automatically selecting terms from an initial set of documents considered to be relevant. The selected terms would then be used for refining the query. Although the learner is not required to give feedback for selecting relevant documents, this approach relies on the relevance of the initial set of documents retrieved. Using internal knowledge for query refinement is similar to pseudo relevance feedback, where an initial set of relevant documents are found, then the top *k* documents are used to refine the query in order to improve retrieval performance.

Figure 2.7 shows the general idea for using internal knowledge from documents for refining queries as demonstrated in (Vélez, Weiss, Sheldon & Gifford 1997). In this work, when a simple query is received from a user, the query is run against a document collection. If a few documents are retrieved, the user is allowed to browse through the documents. However, if too many documents are retrieved, such that the documents are more than what a user can browse through, then terms are automatically suggested to the user. An explicit approach is adopted where the user has to choose terms from the suggestions made in order to refine the query. The user can refine the query by either adding the terms to the initial query to focus the query or by changing the query to achieve a broader view of the query. Wu & Fang (2013) attempt to reduce the time often needed in the 2-stage retrieval that is usually done in pseudo-relevant feedback. Given that the refined query is a version of the initial query, the results of the initial retrieval set are used to reduce the time needed to process the refined query in the second round of retrieval. The method shows an improvement in the time taken for the retrievals.

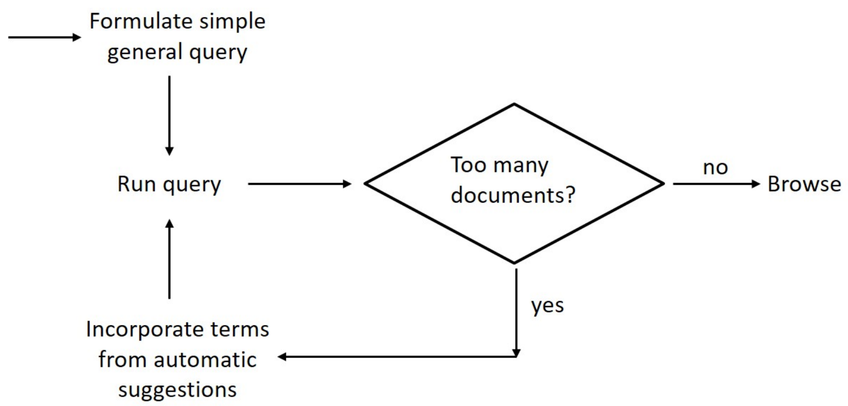


Figure 2.7: An example of refining queries using internal knowledge sources

A drawback of approaches that use internal knowledge to refine queries is that the search results may be directed towards a few documents, and this can be harmful if the documents are only about specific topics. Further, the retrieval performance for difficult queries can be affected if the initial retrieval set contains some irrelevant documents (Li, Luk, Ho & Chung 2007).

Using External Knowledge Sources to Refine Queries

Another approach to query refinement involves using external knowledge sources for refining queries. In this approach, the terms from domain sources are used to refine queries. A source with a good coverage is usually recommended for this task. Domain knowledge sources have been used to identify potentially relevant terms to use for refining queries. Sources such as DBpedia (Meij, Bron, Hollink, Huurnink & De Rijke 2009), and Wikipedia (Meij & de Rijke 2010, He & Ounis 2007) have been employed.

In (Meij et al. 2009), DBpedia concepts and the corresponding Wikipedia concept descriptions are used as the knowledge sources to identify concepts that are contained in queries. A supervised machine learning approach is employed to choose which concepts are particularly relevant for the queries. A similar approach is used in (Meij, Bron, Hollink, Huurnink & de Rijke 2011), where a set of candidate DBpedia concepts are selected, then a supervised machine learning approach is employed to choose relevant concepts to the query. Assessors are employed to manually map queries to relevant DBpedia concepts, and the result from this mapping and the query features are then used to train a model to predict relevant concepts. The identified concepts can be used to generate domain contextual information for enriching the initial query. The best performance was obtained when the full query was used to search for relevant concepts. The developed model is able to identify relevant concepts, such that in most instances the first concept suggested was relevant to the query.

Wikipedia is used as the knowledge source for refining queries in (Meij & de Rijke 2010). A supervised machine learning technique is also applied. Annotators are employed to manually identify relevant Wikipedia articles for queries. Their annotations are used to train a model which is later used to make predictions of relevant concepts for new queries. The relevant concepts for a query are used to generate terms that can be added to the initial query as a way of refining the query. They found that shorter queries were often more ambiguous. In such scenarios, adding focused terms from the external knowledge sources yielded better performance than when terms from the document collection were used for refinement. He & Ounis (2007) create a large external knowledge source using a combination of Wikipedia and a TREC collection. Wikipedia is used because of the concept descriptions it contains, while the TREC collection contains Newswire articles. The intuition is that using a large knowledge source for refining queries should improve retrieval, because there are likely to be more relevant documents to use for refinement. Their results confirm the benefit of using an external knowledge source when refining queries.

The effectiveness of using external knowledge sources for query refinement has been demonstrated in previous work (Bendersky, Metzler & Croft 2012, Diaz & Metzler 2006). For example, (Bendersky et al. 2012) found the use of multiple knowledge sources for refining queries to be more effective than when a single knowledge source was used. Diaz & Metzler (2006) show that using external knowledge for refining queries performed better than when internal knowledge from the collection was used for refinement. They suggest that the external knowledge source is not usually affected by the uncertainty that occurs when a document collection is used as the internal knowledge source.

However, one potential challenge in using external knowledge sources for query refinement is the possibility of query drift, where the refined query deviates from the original query. This can happen when noisy terms are used to refine the query. In (Xu, Jones & Wang 2009), a Wikipedia article that is most descriptive of the query is selected and used to refine the query as a way of reducing the effect of using noisy terms. Their results show an improvement in the quality of the initial document used for refinement.

2.3.2 Representation of Learning Resources

Text representation allows natural language text to be analysed so that meaningful information can be derived from the text. Text is often unstructured, so creating suitable representations for text can be challenging. This is because there are no fixed set of features by which the text can be represented. Two broad approaches often used to address the challenge of text representation are illustrated in Figure 2.8. These are knowledge-light methods, such as topic models (Blei & McAuliffe 2007, Chen & Liu 2014); and knowledge-rich methods, such as those that take advantage of ontologies for creating representations for text (Boyce & Pahl 2007, Yarandi et al. 2011). In Figure 2.8, the lower row of items identifies a range of knowledge sources that can be employed in creating representations for text. The knowledge-light methods use corpus-based knowledge sources. These methods are easier to develop and they do not rely heavily on external knowledge sources, hence they are knowledge-light. On the other hand, the knowledge-rich methods use structured knowledge sources. Such methods are more time consuming to build as they often need the creation of knowledge structures, however they are designed to provide useful representations.

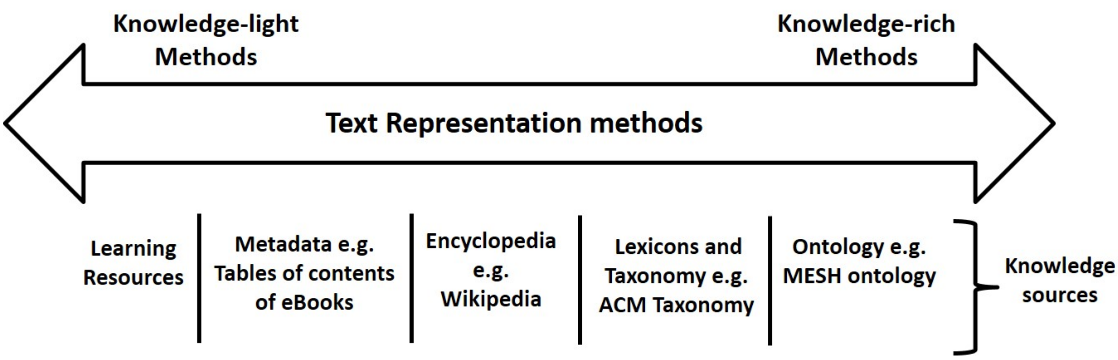


Figure 2.8: Approaches used for representing text

Knowledge-light Text Representation Methods

Knowledge-light representation approaches usually involve the use of statistical models to identify important topics from a corpus. The identified topics are often keywords, key phrases or a combination of both. Keywords can be extracted from documents in a corpus by applying clustering techniques to identify the most informative topics from documents. The identified topics form descriptive tags which can then be used for modelling users within a recommendation system as employed by Timonen, Toivanen, Kasari, Teng, Cheng & He (2012). A summary of methods which apply supervised and unsupervised techniques to identify keywords from text are presented in (Beliga, Meštrović & Martinčić-Ipšić 2015). A demonstration of how key phrases are automatically extracted from documents in order to provide useful summaries for the text is presented in (Witten, Paynter, Frank, Gutwin & Nevill-Manning 1999). Similarly, key phrases which describe the main ideas from news articles are automatically extracted using statistical techniques by Dostál & Jezek (2011). Coenen, Leng, Sanderson & Wang (2007) showed that using a combination of keywords and phrases was better than using only keywords when representing text.

A combination of keywords and key phrases would be further explored when creating a representation for learning resources. This is because important topics in a domain may not only appear as words, but sometimes occur as phrases. For example, the phrase "decision trees" is an important phrase in the Machine Learning and Data Mining domain. The individual words "decision' and "trees" may not be as important as having these two words together. Hence the advantage of using an approach that combines keywords and key phrases when representing text.

The important topics in the corpus-based method can be extracted from different text sources such as: learning resources, metadata, news articles and encyclopedia e.g. Wikipedia or DBpedia. A collection of learning materials are used as the text source for extracting topics in (Rodrigues, Antunes, Gomes, Santos, Barbeira & Carvalho 2007). Natural Language Processing (NLP) techniques are applied to the learning materials to help with identifying topics. The terms in titles and headings of sections in the materials are considered to be more important. Similarly, a collection of academic papers as well as the abstracts of academic papers are used as a text source for generating keywords in (Yang, Chen, Cai, Huang & Leung 2016). Further, WordNet (Fellbaum 2005) is used for POS tagging, in addition to standard NLP approaches as part of the steps involved in identifying important topics.

Metadata of learning materials can also be used as a text source for creating representations as demonstrated in (Bousbahi & Chorfi 2015). The features used for representation are extracted from the metadata. This provides high level features that would be less noisy than using free-text. However, such a source can be quite limited in the amount of important topics contained. Metadata is also used by Dietze, Yu, Giordano, Kaldoudi, Dovrolis & Taibi (2012). In their work, linked data methods are applied to the metadata of learning resources as a means of organising the learning resources, so that documents from different sources can be linked in a structured way and found easily. In (Dostál & Jezek 2011), news articles are the text source used for extracting important topics. Encyclopedia sources such as Wikipedia are useful text sources for identifying important topics as demonstrated in (Qureshi, O'Riordan & Pasi 2014). In a related work, the content of academic Web pages are combined with the titles of Wikipedia articles to generate a text source for identifying important topics for a chosen domain (Qureshi et al. 2014).

Wikipedia is also used by Milne & Witten (2008) for disambiguating words from Newswire text in order to link the words to the relevant Wikipedia article. Similarly, another encyclopedia, DBpedia has been used as a text source to support the identification of important topics from social media text. For example, NLP methods are applied to the words from the social media post to generate a set of topics in (Muñoz-García, García-Silva, Corcho, Higuera Hernández & Navarro 2011). Each identified topic is mapped to a set of DBpedia entities as a means of verifying the topic and removing any ambiguity. The encyclopedia sources cover information about a large range of topics in a variety of languages, hence their use for representation tasks.

One drawback of the knowledge-light text representation method is that, it relies on the coverage of the document collection used, and so the topics produced may not be representative of the domain. A way of addressing the issue with limited coverage, is the use of encyclopedia sources such as Wikipedia and DBpedia. This helps to enrich the available corpus that is used for representing text. The use of such encyclopedias will be explored for developing a suitable representation for learning materials.

Knowledge-rich Text Representation Methods

Knowledge-rich methods are often structured representations that capture relationships between domain concepts. This often involves using a taxonomy such as the ACM taxonomy, reusing an existing ontology such as the MESH ontology, or creating a new ontology. An ontology defines the relationships between concepts, often represented in hierarchical form as they would appear in the real world (Boyce & Pahl 2007). An existing ontology for programming language developed in (Sosnovsky & Gavrilova 2006) inspires the work in (Ruiz-Iniesta, Jimenez-Diaz & Gomez-

Albarran 2014), which reuses an ontology to create a representation for learning resources. The developed structure contains knowledge about important concepts, as well as a description of concepts, and connections between concepts such as pre-requisites and subsequents. Although an existing ontology is used, domain experts are still employed for validating the developed structure.

An approach often adopted when building new ontologies is Methontology (Fernández-López, Gómez-Pérez & Juristo 1997). This aim of the approach is to take in as input some text such as learning materials and produce an ontology as output. It begins with a specification which includes information such as the purpose and end-users of the ontology. Another stage of building the ontology is knowledge acquisition which entails extracting relevant knowledge from sources such as books and even domain experts. The conceptualization stage is a key part of the method which involves building a collection of domain concepts and verbs that would be used in the ontology. The integration phase often involves reusing any existing definitions or meta-data from other sources, while the implementation phase entails using an ontology editor such as Protégé to create the ontology (Knublauch, Fergerson, Noy & Musen 2004).

A new ontology is created as a means of representing the learning resources within an e-Learning repository in (Nasraoui & Zhuhadar 2010). Information about the courses is used to identify the important concepts for the domain. The open source ontology editor, Protégé is used for building the Ontology. Likewise, a new ontology is developed to capture the domain concepts contained in e-Learning courses in (Panagiotis, Ioannis, Christos & Achilles 2016). The new structure is represented using a tool which is often used for such a task, the Web Ontology Language (McGuinness & Van Harmelen 2004). As a way of speeding up the process of building ontologies, a language that can be used to identify patterns such as semantic relations in text is presented in (Ghadfi, Béchet & Berio 2014). Further, a tool for creating ontologies from text is enhanced. However, analysing the text that is used for such a task often involves more runtime.

Although ontologies are designed to have a good coverage of their domains, the output is still dependent on the view of its builders and, because of handcrafting, existing ontologies cannot easily be adapted to new domains. Creating a new ontology is often a time consuming task, requiring the inputs of domain experts. On the other hand, reusing an existing ontology can reduce the time taken to build one from scratch, but existing ontologies are often insufficient for solving a new problem and they often require some alignment in order to suit the new task. e-Learning is dynamic because new resources are becoming available regularly, and so using fixed ontologies limits the potential to incorporate new content.

2.3.3 Popular approaches for representing text

Learning resources often contain text, and this text is often unstructured. So finding and retrieving relevant documents can be challenging. Document retrieval usually relies on the content of the document for making predictions about the relevance of the document. One way of addressing this challenge is by creating suitable representations for the text to enable the retrieval of relevant documents. Some approaches often used for representing text are presented.

Bag of Words representation

Bag of Words (BOW) refers to a collection of the words extracted from a document. The order that the words appear is not considered. The Vector Space Model (VSM) can be applied to the extracted words to create a feature vector for the document, where each word is a feature that describes the document, and the value of the word is its weight. There are 3 commonly used weighting schemes in the VSM. First, the binary value that captures the presence or absence of a word, as a 1 or 0 respectively. Second, the Term Frequency (TF), which captures how many times a term appears in a document. Finally, the Term Frequency - Inverse Document Frequency (TF-IDF) weight which combines TF and IDF (Sparck Jones 1972). The IDF computes the number of documents in a collection that contain the given word. The IDF captures the importance of a word in a document collection by reducing the weight of common terms and emphasising the weight of rare terms. Hence the use of the TF-IDF weighting in many Information Retrieval systems.

A key step before creating a representation for text is often that of pre-processing which helps to prepare a document for indexing. Pre-processing often involves stages such as stop word removal, stemming or lemmatisation, and tokenisation. Stopwords are common words such as “a”, “is”, “the” that are are found across all documents but do not contribute to important information in documents, hence stopwords are often removed during pre-processing. The English stopwords [1] and SMART stopwords (Salton 1971) are 2 sets of stopwords that are often used.

Stemming and lemmatisation aim to reduce words to their base. Stemming takes a harsh approach to this task, while lemmatisation tries to ensure that its output is still a meaningful word in its dictionary form. A common stemming algorithm often used is the Porter stemming (Porter 1980), and a common lemmatiser used is the WordNet Lemmatiser. Applying Porter stemming to the word “organising” produces “organis”, while applying the WordNet Lemmatiser produces “organise”. Stemming is often harmful for precision but can increase recall. Manning, Raghavan

[1] http://snowball.tartarus.org/algorithms/english/stop.txt

& Schütze (2008) suggest that the advantage of performing lemmatisation is minimal for retrieval. However, choosing the technique to use depends on the task one is performing. Tokenisation entails splitting up a sequence of text from a document into individual words referred to as tokens. For example: the sentence "cat sat down" is split as "cat" "sat" "down" containing 3 tokens. The tokens can then be used for indexing the document.

Apache Lucene (Hatcher, Gospodnetic & McCandless 2004) is a commonly used framework for indexing text documents. A more recent framework developed using Lucene is Elasticsearch (Kuć & Rogoziński 2015). The relevance scoring for matching documents is computed by applying Lucene's practical scoring function, which uses ideas from the Boolean model, VSM and TF-IDF weighting. Lucene's scoring function as implemented in Elasticsearch is shown in Equation 2.1. Given a query, q the relevance score of a document, d to the query is given as:

$$Score\,(q,d) = coord\,(q,d) \cdot queryNorm\,(q) \cdot \sum_{t \in q} \left(\sqrt{tf\,(t,d)} \cdot idf\,(t)^2 \cdot boost(t) \cdot norm(d) \right) \tag{2.1}$$

The following aspects can be used to describe the scoring function:

Term importance: $tf\,(t,d)$ is the term frequency, measuring how often a term, t occurs in a document, d. Normally, tf gives a lot of importance to long documents, so in Elasticsearch the square root of tf is taken as a scaling measure to cope with this effect. $idf\,(t)$ is the inverse document frequency, measuring how often term, t occurs in the document collection. The weight of a frequently occurring term is reduced using idf. The idf for a term is given as the logarithm of the number of documents in the collection divided by the number of documents the term occurs in.

Overlap of terms: $coord\,(q,d)$ is the coordination factor that counts the number of terms from the query, q that appear in a document, d. So documents with a higher percentage of query terms are rewarded, as such documents have a higher chance of being a good match for the query.

Weighting: $boost\,(t)$ is the boost factor used to increase the importance of each term, t in a field. $queryNorm\,(q)$ is a factor used for normalizing the query, q. $norm\,(d)$ is the field-length normalization factor that is based on the number of terms in a field of a document, d. The norm considers the length of the fields in a document, so shorter fields such as "title" are given higher weights than longer fields, such as "description". The norm is computed as the inverse square root of the number of terms in the field. The position of a term does not affect the computation of the norm. The function is designed so that using default values of the parameters is still effective for scoring.

Semantic Enhancements

Latent Semantic Indexing (LSI) (Deerwester, Dumais, Furnas, Landauer & Harshman 1990) and Latent Dirichlet allocation (LDA) (Blei, Ng & Jordan 2003) are other methods often used to represent text. Both LSI and LDA can be employed either for semantic enhancement using co-occurrence, by identifying words that have semantic relationships in a document collection; or for Topic modelling by identifying topics contained in a document collection and mapping the text to the identified topics.

The intuition behind LSI is that words which occur within the same context would likely have similar meanings. LSI uses Singular-value decomposition (SVD) to reduce the number of terms of a term-document matrix while maintaining the similarity structure within documents. A large term-document matrix is used to build a space that puts terms and documents that are closely related near each other. Terms can be compared using the dot product between term vectors. Similar terms have a dot product value closer to 1, while dissimilar terms have similarity values closer to 0. The created space can then be used for indexing. For example, query terms can be mapped to a position within this space, then documents in that neighborhood would be retrieved.

The basic assumption of LDA is that documents are created from a set of topics. These topics produce words in proportion to the probability distribution of that topic. The topic probabilities can then form a representation for the document. One drawback is that LDA models can be computationally expensive when the amount of text is large.

A method that overcomes LDA's drawback is Word2Vec (Mikolov, Chen, Corrado & Dean 2013). Word2Vec is a representation that learns the context in which words occur. Given a document collection as input, neural networks are used to create a vector space such that each distinct word in the collection is represented by a vector in that space. The word vectors with common contexts are placed closer to each other. An advantage of Word2Vec is the word embeddings it creates, which allows the representation to consider different levels of similarity between words.

2.4 Summary

Recommendation in e-Learning aims to help learners to find relevant learning resources that meet their learning goals. This is not easy because learners usually have difficulty finding the right learning materials and the learning resources are often unstructured text, and are not suitably indexed for retrieval. This makes finding a relevant document difficult. Hence the need for an effective method of refining learners' queries and representing learning materials to improve recommendation.

Queries can be refined using internal or external knowledge sources. A drawback of using internal knowledge sources is the limitation that can arise from the scope of the document collection used. This can be harmful if the documents used for refinement contain irrelevant ones. External knowledge sources such as encyclopedias have been useful domain sources for enriching learners' queries during refinement particularly for short queries which can be ambiguous. The query refinement method we explore would draw insight from methods that use external knowledge sources for refining queries.

Learning resources are often unstructured text. Two broad ways of representing learning materials are knowledge-light and knowledge-rich approaches. Knowledge-light methods usually treat learning resources as a bag-of-words. Methods such as VSM, and LSI can be applied to create a representation for the learning materials. The drawback of such approaches is that it relies on the corpus used to create a representation. So the output produced tends to be limited, and does not capture a good coverage of the domain. Knowledge-rich approaches employ structured representations such as taxonomies and ontologies to support the representation of learning materials. Adopting this approach is often expensive given the time and expertise required to create a structure such as an ontology.

The focus of this research is creating solutions to enable learners find relevant learning materials. This will be done by building knowledge-driven approaches that takes advantage of the knowledge of teaching experts to identify important domain concepts. The identified concepts will be enriched with knowledge from encyclopedias for the creation of background knowledge that can be employed to support the refinement of learners' queries and for representing learning materials.

Chapter 3

Research Methodology

This research aims to create effective techniques for e-Learning recommendation. The review of literature identified the semantic gap and the intent gap as two key issues that make e-Learning recommendation challenging. In order to address the semantic gap, we will explore effective methods for representing learning resources, with the aim of making the resources more accessible. The intent gap will be addressed by exploring effective query refinement methods, to help learners identify relevant topics and find documents that meet their needs. Figure 3.1 contains an overview of the research methodology.

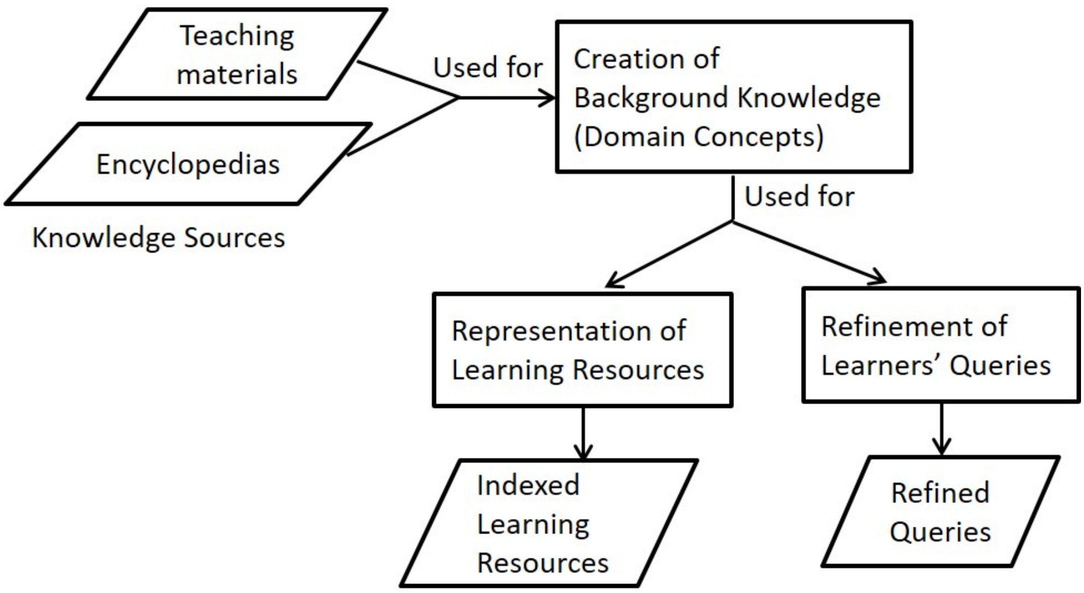


Figure 3.1: Research methodology overview

The knowledge from teaching materials and encyclopedias are leveraged to create background knowledge, which is then used for resource representation and query refinement. We adopt this approach because the teaching materials are written by experts that are knowledgeable in their respective domains and these teaching experts know what learners should be interested in. The encyclopedia source provides rich descriptions for the domain concepts in the background knowledge. The background knowledge provides a shared vocabulary for learners and teaching experts. We will employ the background knowledge for two main tasks.

The first task will be the representation of learning resources. The domain concepts contained in the background knowledge provide a rich vocabulary for representing learning resources. This allows us to focus the search on documents that contain relevant concepts. The result will be indexed learning resources, that should be more accessible during document retrieval.

The second task will be the refinement of learners' queries. When queries are received from learners, we can take advantage of the domain concepts contained in the background knowledge for identifying what concepts are most similar to learners' queries. The concept vocabulary from the identified concepts can then be used to refine the queries. The result of this will be refined queries that should be more effective when used to search for learning materials.

This chapter presents the creation of background knowledge. In the rest of this chapter and the next, we will discuss how the background knowledge is used for representing learning resources. Then, in the following two chapters, we will explore how the background knowledge is used for the refinement of learners' queries. We will also discuss the evaluation of our developed methods throughout these chapters.

3.1 Knowledge Representation of e-Learning Documents

Recommendation in e-Learning is challenging because learning materials are often unstructured text, and so are not suitably indexed for retrieval. One solution to this challenge is the creation of a knowledge-rich representation that contains a good coverage of relevant topics. Such a collection of domain topics can then be used to support the representation of learning materials. This can be useful for highlighting the relevant topics contained in learning materials.

A related challenge in e-Learning recommendation occurs when learners try to find relevant materials that meet their learning goals. There is often a mismatch in the vocabulary used by learners when searching, and that used by teaching experts who author the learning materials. This presents a semantic gap. In order to bridge this semantic gap, a method that automatically creates custom background knowledge is introduced.

Background knowledge refers to information about a domain that is useful for general understanding and problem-solving (Zhang, Liu & Cole 2013). In e-Learning, background knowledge can be employed to influence the retrieval of relevant learning materials for learners. Background knowledge can be captured as a set of domain concepts, each representing an important topic in the domain. For example, in a domain for e-Learning such as Machine Learning, one would find topics such as Classification, Clustering and Regression. Each of these topics would be represented by a concept, in the form of a concept label and a pseudo-document which provides a rich description for the concept. The domain concepts can then be used to underpin the representation of learning materials.

3.2 Background Knowledge

The process involved in creating background knowledge is illustrated in Figure 3.2. Domain knowledge sources are required as input to the process, and a structured collection of teaching materials and an encyclopedia source are used. Ngrams are automatically extracted from the structured collection to provide a set of potential concept labels, and then a domain lexicon is used to validate the extracted ngrams in order to ensure that the ngrams are also being used in another information source. The encyclopedia provides candidate pages that become the concept label and discovered text for the ngrams. The output from this process is a set of domain concepts, each comprising a concept label and an associated pseudo-document.

The knowledge sources and process involved in creating a background knowledge representation for an e-Learning domain are presented in the following sections. Two approaches that have been developed using the background knowledge representation are described. The developed methods are then applied to an e-Learning recommendation task.

3.2.1 Knowledge Sources

Two knowledge sources are used as initial inputs for discovering concept labels. A structured collection of teaching materials provides a source for extracting important topics identified by teaching experts in the domain, while a domain lexicon provides a broader but more detailed coverage of the relevant topics in the domain. Books are highlighted as a contributing factor which is often linked to the success of teaching and learning (Agrawal, Chakraborty, Gollapudi, Kannan & Kenthapadi 2012). The authors of books are usually experts in their respective domains,

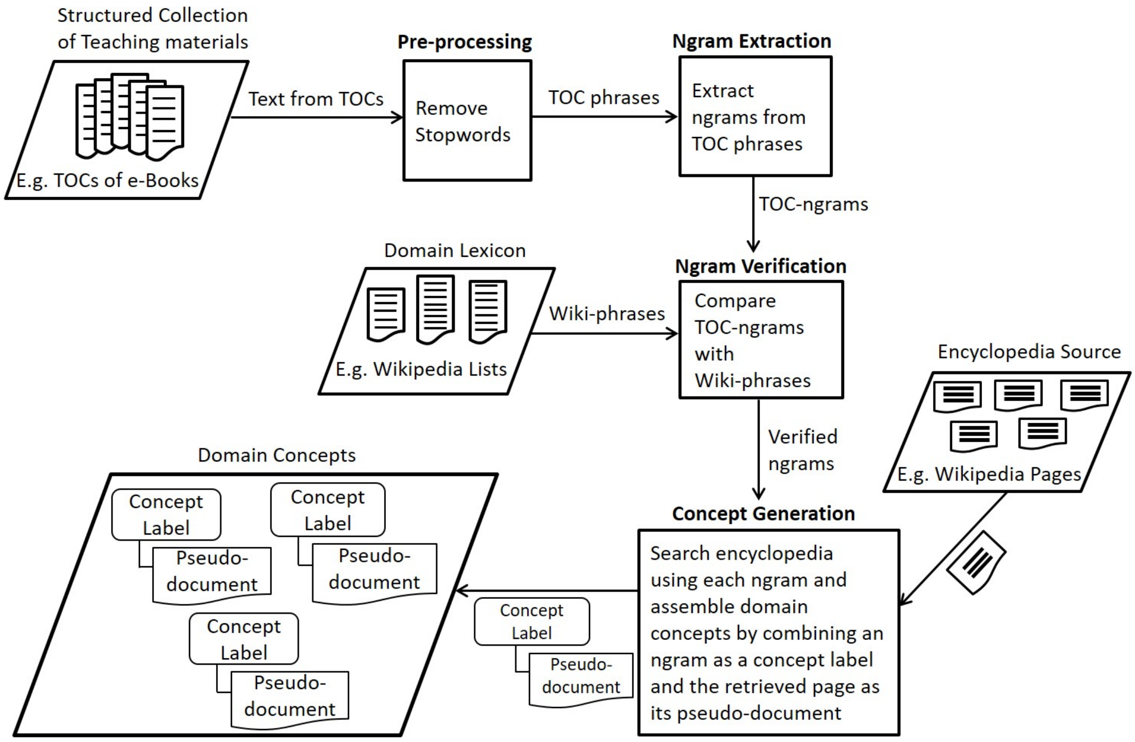


Figure 3.2: An overview of the background knowledge creation process

and they carefully design books to contain important topics that learners should be interested in.

The domain lexicon is used to verify that the concept labels identified from the teaching materials are directly relevant. Thereafter, an encyclopedia source, such as Wikipedia pages, is searched and this provides the relevant text to form a pseudo-document for each verified concept label. The final output from this process is the set of concepts each comprising a concept label and an associated pseudo-document.

The approach is demonstrated with learning materials from Machine Learning and Data Mining. The collection of teaching materials used are e-Books. The Tables-of-Contents (TOCs) of the e-Books are used as a structured knowledge source. A summary of the e-Books used is shown in Table 3.1. The first column contains the title of the e-Books and the surname of the authors, while the second column contains the number of Google Scholar citations for each e-Book, as at the time this research was done. Two Google Scholar queries: “Introduction to data mining textbook” and “Introduction to machine learning textbook” guided the selection process, and 20 e-Books that met all of the following 3 criteria were chosen. First, the book should be about the domain. Second, there should be Google Scholar citations for the book. Finally, the book should be accessible on the Web.

Table 3.1: Summary of e-Books used

Book Title & Author	**Cites**
Machine learning; Mitchell	264
Introduction to machine learning; Alpaydin	2621
Machine learning a probabilistic perspective; Murphy	1059
Introduction to machine learning; Kodratoff	159
Gaussian processes for machine learning; Rasmussen & Williams	5365
Introduction to machine learning; Smola & Vishwanathan	38
Machine learning, neural and statistical classification; Michie, Spiegelhalter, & Taylor	2899
Introduction to machine learning; Nilsson	155
A First Encounter with Machine Learning; Welling	7
Bayesian reasoning and machine learning; Barber	271
Foundations of machine learning; Mohri, Rostamizadeh, & Talwalkar	197
Data mining-practical machine learning tools and techniques; Witten & Frank	27098
Data mining concepts models and techniques; Gorunescu	244
Web data mining; Liu	1596
An introduction to data mining; Larose	1371
Data mining concepts and techniques; Han & Kamber	22856
Introduction to data mining; Tan, Steinbach, & Kumar	6887
Principles of data mining; Bramer	402
Introduction to data mining for the life sciences; Sullivan	15
Data mining concepts methods and applications; Yin, Kaku, Tang, & Zhu	23

Wikipedia is used to create a domain lexicon because it contains articles for many domains (Völkel, Krötzsch, Vrandecic, Haller & Studer 2006), and the contributions of many people (Yang & Lai 2010), so this provides the coverage needed for the lexicon. Yang & Lai (2010) set out to find what motivated users to contribute freely to Wikipedia to create such a large knowledge base. They found that one of the motivating factors was the sense of achievement users had when making a contribution to the knowledge base. Their study confirms similar conclusions by (Nov 2007) that most users are pleased to share what they know with others. These findings give some explanation to the growth of this Encyclopedia source.

Wikipedia concepts are used in (Gabrilovich & Markovitch 2007) to provide meaning for natural language texts because of the large amounts of concepts available. Also, Zheng, Li, Huang & Zhu (2010) exploit Wikipedia as a knowledge base for linking entities found in unstructured text to Wikipedia articles in order to provide some descriptive information for the entities. Similarly, in this thesis, Wikipedia is exploited to build background knowledge that can be used for representing unstructured text from learning materials to enable the recommendation of relevant documents. The lexicon is generated from all the available Wikipedia sources for the Machine Learning and

Data Mining domain. There are 2 available sources for this domain, so we use both sources. First, the phrases in the *contents* and *overview* sections of the chosen domain are extracted to form a topic list. Second, a list containing the titles of articles related to the domain is added to the topic list to assemble the lexicon. Overall, the domain lexicon consists of a set of 664 Wiki-phrases.

Using Wikipedia as a knowledge source can present the challenge of the provenance of some of the contributors. However, we note that in this research, the TOCs of e-Books are used as the starting point for identifying concepts for creating our background knowledge. Wikipedia provides a description of the concepts already identified from the e-Books, because the e-Books have known authors. The articles on Wikipedia are open to review as other contributors can edit content that is not consistent. Further, any disputed article entries can be settled through a discussion page associated with each entry. The study by Giles (2005) suggests that the editing feature helps to improve the quality of articles in the Encyclopedia. These findings give us confidence to use the sources chosen for creating our background knowledge.

3.2.2 Knowledge Extraction Process

Knowledge extraction refers to the process of creating knowledge from knowledge sources. The sources used as input for the knowledge extraction task can be structured sources such as the TOCs of e-Books, or unstructured sources such as the descriptive text from Wikipedia pages. The knowledge created at the end of the extraction can then be used for further tasks such as representing documents and refining queries. The following sections discuss the stages involved in extracting knowledge from the sources earlier identified. The goal is to create background knowledge that can be used for a task such as the representation of learning materials, in order to support the retrieval and recommendation of relevant documents.

Generating Potential Domain Concept Labels

In the first stage of the process, the text from the TOCs is pre-processed. Characters such as punctuations, symbols, and numbers are removed from the TOCs, so that only words are used for generating concept labels. After this, 2 sets of stopwords are removed. First, a standard English stopwords list[1], which allows common words to be removed while still retaining a good set of words for generating the concept labels. The second stopwords are an additional set of words which are referred to as TOC-stopwords. This contains: structural words, such as *chapter* and

[1] http://snowball.tartarus.org/algorithms/english/stop.txt

appendix, which relate to the structure of the TOCs; roman numerals, such as *xxiv* and *xxxv*, which are used to indicate the sections in a TOC; and words, such as *introduction* and *conclusion*, which describe parts of a learning material and are generic across domains.

Stemming is not used in this task during pre-processing because of the following reasons. First, from our earlier experiments we found that when searching an encyclopedia source with the stemmed form of words, relevant results would not be returned, so it was better to use the full form of the words. In addition, the background knowledge would be used for query refinement, so the stemmed words would not be helpful for such a task. The output from pre-processing is a set of TOC phrases. In the next stage, ngram extraction is applied to the TOC phrases to generate all 1-3 grams across the entire set of TOC phrases. The output from this process are TOC-ngrams containing a set of 2038 unigrams, 5405 bigrams and 6133 trigrams, which are used as the potential domain concept labels. Many irrelevant ngrams are generated from the TOCs because all 1-3 grams have been generated from the TOCs.

Verifying Concept Labels using Domain Lexicon

The TOC-ngrams are first verified using a domain lexicon to confirm which of the ngrams are relevant for the domain. The domain lexicon contains a set of 664 Wiki-phrases, each of which are pre-processed by removing non-alphanumeric characters. Table 3.2 contains a spread of the Wiki-phrases in the domain lexicon. The first column of the table shows the name of the ngram, while the second column contains the number of Wiki-phrases in that category and the last column contains the proportion of the ngrams as a percentage. The 84% of the Wiki-phrases shown in Table 3.2 that are 1-3 grams are used for verification of the TOC-ngrams. This is because there is a reduction in the number of ngrams when n is greater than 3 as shown in Table 3.2.

Table 3.2: Spread of Wiki-phrases used for verifying concept labels

Ngram	**Number**	**%**
Unigrams	64	10%
bigrams	280	42%
trigrams	213	32%
4grams	78	12%
5grams	16	2%
6grams	8	1%
7grams	2	0%
8grams	1	0%
9grams	1	0%
10grams	1	0%

The comparison of TOC-ngrams with the domain lexicon identifies the potential domain concept labels that are actually being used to describe aspects of the chosen domain in Wikipedia. During verification, ngrams referring directly to the title of the domain, e.g. *machine learning* and *data mining*, are not included in the Wiki-phrases because the aim is to generate concept labels that describe specific topics within the domain. Overall, a set of 17 unigrams, 58 bigrams and 15 trigrams are verified as potential concept labels. Bigrams yield the highest number of ngrams, which indicates that bigrams are particularly useful for describing topics in this domain.

Domain Concept Generation

Domain concepts are generated after a second verification step is applied to the ngrams returned from the previous stage. Each ngram is retained as a concept label if all of 3 criteria are met. Firstly, if a Wikipedia page describing the ngram exists. Secondly, if the text describing the ngram is not contained as part of the page describing another ngram. Thirdly, if the ngram is not a synonym of another ngram. For the third criteria, if two ngrams are synonyms, the ngram with the higher frequency is retained as a concept label while its synonym automatically remains part of the extracted text. For example, 2 ngrams *cluster analysis* and *clustering* are regarded as synonyms in Wikipedia, so the text associated with them is the same. The label *clustering* is retained as the concept label because it occurs more frequently in the TOCs, and its synonym, *cluster analysis* is contained as part of the discovered text.

Each concept label is used to search automatically on Wikipedia pages in order to generate a domain concept. The search returns discovered text that forms a pseudo-document which includes the concept label. If the search does not return any text for a given concept label, then a manual check is done to confirm whether the concept label is regarded as a synonym of another concept label in Wikipedia. If the concept label is a synonym, then the concept label would already be included in the description of its synonym. Each concept label and pseudo-document pair make up a domain concept. Overall, 73 domain concepts are generated. Each pseudo-document is pre-processed using standard techniques such as removal of English stopwords and Porter stemming. The terms from the pseudo-documents form the concept vocabulary that can now be used to represent learning resources.

3.3 Harnessing Background Knowledge for Representation

The background knowledge contains a rich representation of the learning domain. By harnessing this knowledge for representing learning resources, one would expect to retrieve documents based on the domain concepts that they contain. The domain concepts are designed to be effective for e-Learning, because they are assembled from the TOCs of teaching materials (Agrawal et al. 2012). This section presents two approaches that have been developed by employing the background knowledge in the representation of learning materials.

3.3.1 A Concept Based Representation Approach

Representing documents with the concept vocabulary allows retrieval to focus on the concepts contained in the documents. This section introduces the CONCEPTBASED document representation method (CONCEPTBASED). The phases involved in this method are illustrated in Figures 3.3 & 3.4. Figure 3.3a is a term-concept matrix created using the concept vocabulary, $t_1 \ldots t_c$, from the pseudo-documents of concepts, $C_1 \ldots C_m$. The columns are the pseudo-documents of concepts. While Figure 3.3b is a term-document matrix created using the concept vocabulary, $t_1 \ldots t_c$, and its columns are a set of documents $D_1 \ldots D_n$ from the collection of learning materials to be represented. The entry into each matrix in Figure 3.3 is the TF-IDF weighting (Salton & Buckley 1988). In Figure 3.3a, c_{ij} is the TF-IDF of term t_i in concept C_j, and in Figure 3.3b, d_{ik} is the TF-IDF of t_i in D_k.

	C_1	...	C_j	...	C_m
t_1					
...					
t_i			c_{ij}		
...					
t_c					

(a) Term-concept matrix

	D_1	...	D_k	...	D_n
t_1					
...					
t_i			d_{ik}		
...					
t_c					

(b) Term-document matrix

Figure 3.3: Term matrices for concepts and documents

Firstly, in Figure 3.3, the concept vocabulary, is used to create a term-concept matrix and a term-document matrix. Next, documents $D_1 \ldots D_n$ are represented with respect to concepts by computing the cosine similarity of the term vectors for concepts and documents. The output is the concept-document matrix shown in Figure 3.4a, where y_{jk} is the cosine similarity of the vertical shaded term vectors for C_j and D_k from Figures 3.3a and 3.3b respectively. Finally, the document similarity is generated by computing the cosine similarity of concept-vectors for documents. Figure 3.4b shows z_{km}, which is the cosine similarity of the concept-vectors for D_k and D_m from Figure 3.4a. So, the CONCEPTBASED document representation approach uses the representation and similarity in Figure 3.4 to influence retrieval. The expectation is to retrieve documents that are similar based on the domain concepts that they contain.

	D_1	...	D_k	...	D_m	...	D_n
C_1							
...							
C_j			y_{jk}		y_{jm}		
...							
C_m							

(a) Concept-document matrix representation

	D_1	...	D_m	...	D_n
D_1	1				
...					
D_k			z_{km}		
...					
D_n					1

(b) Document-document similarity

Figure 3.4: Document representation and similarity using the CONCEPTBASED approach

3.3.2 Augmenting the Representation of Learning Resources

This section introduces a CONCEPTBASED Augmented document representation, (CB-AUG) that is developed to augment the representation of learning materials using the concept vocabulary. The CB-AUG approach exploits the relative distribution of the vocabulary in the concept and document spaces for the representation of learning resources. The TF-IDF weight of a term changes depending on its relative frequency in both spaces.

First, the domain concepts, $C_1 \ldots C_m$ generated in section 3.2.2, and the documents that are to be represented, $D_1 \ldots D_n$, are merged to form a corpus. Next, a term-document matrix with TF-IDF weighting is created using all the terms, $t_1 \ldots t_T$ from the vocabulary of the merged corpus as shown in Figure 3.5a. For example, entry q_{ik} is the TF-IDF weight of term t_i in D_k. If t_i has a lower relative frequency in the concept space compared to the document space, then the weight q_{ik} is boosted. So, distinctive terms from the concept space will get boosted. Although the overlap of terms from both spaces are useful for altering the term weights, it is valuable to keep all the terms from the document space because this provides a richer vocabulary. The shaded term vectors for

$D_1 \ldots D_n$ in Figure 3.5a form a term-document matrix for documents whose term weights have been influenced by the presence of terms from the concept vocabulary.

	C_1	...	C_j	...	C_m	D_1	...	D_j	D_k	...	D_n
t_1											
...											
t_i			p_{ij}						q_{ik}		
...											
t_T											

(a) Augmented term-document matrix representation

	D_1	...	D_k	...	D_n
D_1	1				
...					
D_j			r_{jk}		
...					
D_n					1

(b) Augmented document similarity

Figure 3.5: Representation and similarity of documents using the augmented approach

Finally, the document similarity in Figure 3.5b, is generated by computing the cosine similarity between the augmented term vectors for $D_1 \ldots D_n$. Entry r_{jk} is the cosine similarity of the term vectors for documents, D_j and D_k from Figure 3.5a. The CB-AUG method exploits the vocabulary in the concept and document spaces to influence the retrieval and recommendation of documents.

3.4 Evaluating Learning Resource Representation

An e-Learning recommendation task is simulated to enable the evaluation of the developed document representation methods. A collection of topic-labelled learning resources are used for the evaluation. These are papers from Microsoft Academic Search, in which the author-defined keywords associated with each paper identifies the topics they contain (Hands 2012). The author-defined keywords represent what relevance would mean in an e-Learning domain and these are used for judging document relevance. Although academic papers are not ideal learning resources, these papers act as a collection of e-Learning resources because of their availability, and the author-defined keywords associated with each paper. Using a query-by-example scenario, the relevance of a retrieved document is evaluated by considering the overlap of author-defined keywords in the query document. This evaluation approach provides a means of measuring the ability of the developed methods to identify relevant documents.

3.4.1 Dataset

The dataset used for this evaluation contains 217 Machine Learning and Data Mining papers. This dataset is referred to as dataset 1. A distribution of the keywords per document is shown in Figure 3.6. The x-axis shows the documents in the collection, the documents in this figure are sorted based on the number of keywords they contain. The y-axis shows the number of keywords.

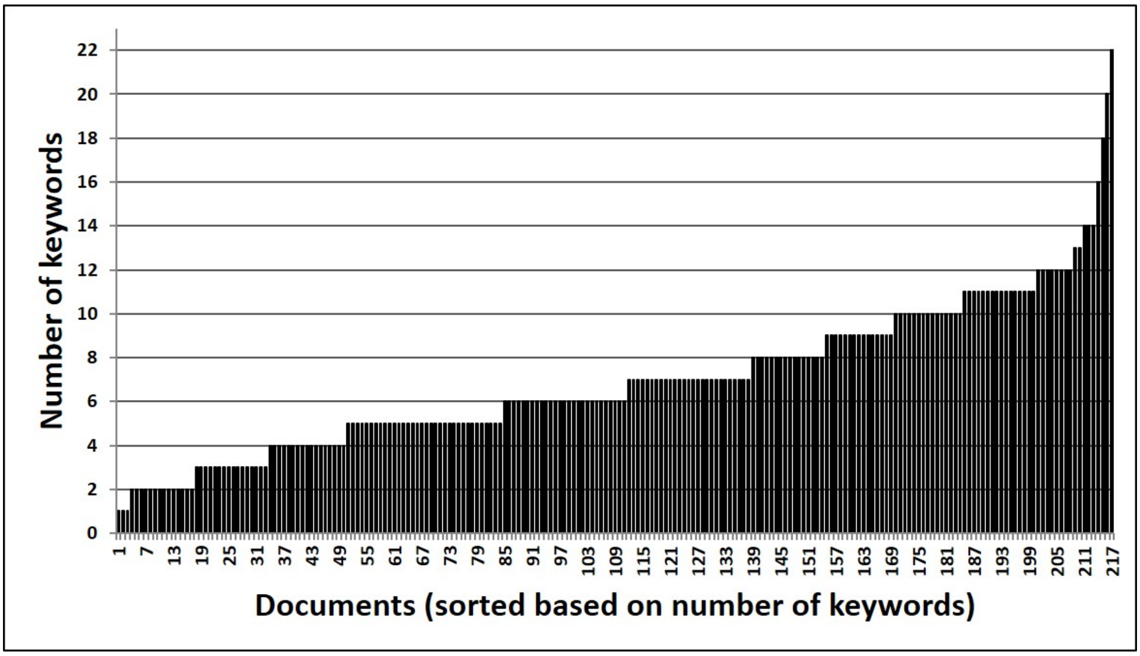


Figure 3.6: Number of keywords per document in dataset 1

The minimum number of keywords is 1, while the maximum number of keywords is 22. So every document used has at least one author-defined keyword. There are 903 unique keywords, and 1,497 keywords in total. In this evaluation, a retrieved document is considered to be relevant based on the proportion of its keywords that overlap with the keywords of the query document. In order to determine a suitable overlap threshold for this task, the overlap scores for all the document pairs are computed. The number of unique pairs for a document collection can be computed as:

$$\text{Number of unique pairs(n)} = \frac{n^2 - n}{2} \tag{3.1}$$

where n is the number of documents in the collection. Given that there are 217 documents in the collection used, there are $23,436$ unique pairs for all the 217 document pairs as given by Equation 3.1. In these entries, $20,251$ have an overlap of zero, meaning that there is no overlap in 86% of the data. So only 14% of the data have an overlap of keywords.

Figure 3.7 shows a summary of the overlap scores for the 14% of the document pairs with a non-zero overlap score. The x-axis contains the proportion of pairs for 14% of the data, while the y-axis contains the overlap threshold value. The minimum overlap threshold value is 0.0 while the maximum overlap value is 1.0 for this collection.

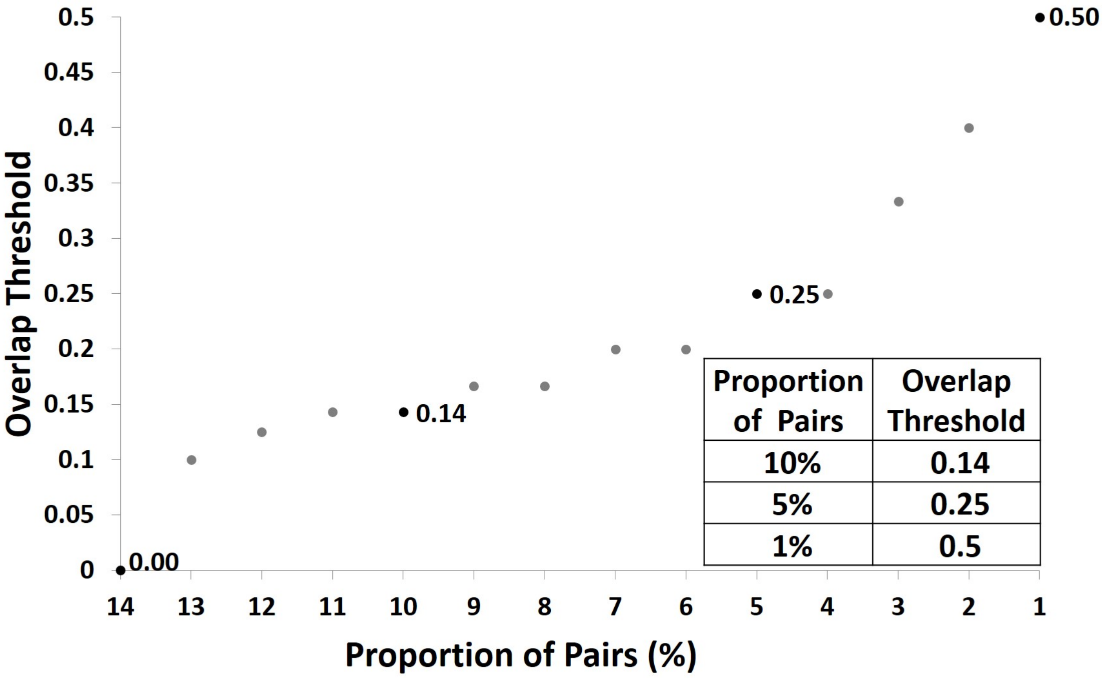


Figure 3.7: Overlap of document pairs in dataset 1

There are only 14% of the document pairs that have some author-defined keywords in common, which gives us an idea of how challenging this task is. The table inserted in Figure 3.7 contains the overlap threshold values for 10%, 5% and 1% of the data. It is observed that, the higher the overlap threshold value, the lower the proportion of documents considered to be relevant. The lower proportion of pairs is an indication that the document pairs being compared have fewer keywords in common.

There are 10% of document pairs with overlap scores ≥ 0.14, while 5% of document pairs have overlap scores which are ≥ 0.25, and 1% of document pairs with overlap scores ≥ 0.5. For experiments with this first dataset, 0.14 and 0.25 are used as thresholds, thus avoiding extreme values that would allow either very many or very few of the documents to be considered as relevant. Using these threshold values still allows us to have a challenging retrieval task.

3.4.2 Experimental Design

Evaluations using human evaluators are expensive, so the author-defined keywords are employed for judging the relevance of a document. The keywords are used to define an overlap metric. Given a query document Q with a set of keywords K_Q, and a retrieved document R with its set of keywords K_R, the relevance of R to Q is based on the overlap of K_R with K_Q.
The overlap is computed as:

$$Overlap(K_Q, K_R) = \frac{|K_Q \cap K_R|}{min\left(|K_Q|, |K_R|\right)} \tag{3.2}$$

A retrieval is considered to be relevant by setting an overlap threshold. So, if the overlap between K_Q and K_R meets the threshold, then K_R is considered to be relevant.

3.4.3 Evaluation Metric

The interest is in the topmost documents retrieved, because the aim is for the top recommendations to be relevant. Precision@n is used to determine the proportion of retrieved documents that are relevant. It is given by:

$$Precision@n = \frac{|retrievedDocuments \cap relevantDocuments|}{n} \tag{3.3}$$

where, n is the number of documents retrieved each time, *retrievedDocuments* are the documents that are retrieved, and *relevantDocuments* are those documents that are considered to be relevant, meaning that the documents have an overlap that is greater than the threshold.

3.4.4 Results and Discussion

This section presents the results of evaluating different document representation methods. The performance of the following methods are compared:

- CONCEPTBASED document representation method, (CONCEPTBASED) which represents documents using the domain concepts (Section 3.3.1).
- CONCEPTBASED Augmented document representation method, (CB-AUG) which uses the term distribution in the concept vocabulary to influence the weight of terms in the document vocabulary (Section 3.3.2).

- BOW method is a standard Information Retrieval method where documents are represented using the terms from the document space only with TF-IDF weighting. BOW is used as the benchmark method.

- RANDOM method has been included to give an idea of the relationship between the threshold and the precision values.

For all the methods, the documents are first pre-processed by removing English stopwords and applying Porter stemming. Then, after representation, a similarity-based retrieval is employed using cosine similarity. The methods are evaluated using a leave-one-out retrieval. The performance of CONCEPTBASED and CB-AUG are compared against that of BOW.

Figure 3.8 shows the precision of the methods given an overlap threshold of 0.14. The number of recommendations (n) is shown on the x-axis, while the average precision@n is shown on the y-axis. The number of recommendations range from 1 to 10, because our interest is in the top 10 recommendations retrieved. Typically one would focus on the earlier retrievals because these should contain documents that are more likely to be relevant.

The results from the RANDOM (▲) method are consistent with the relationship between the threshold and the proportion of data as shown in Figure 3.7. When an overlap threshold of 0.14 is used, the RANDOM method has an average precision of about 0.1, which is 10%. Recall that there were 10% of document pairs with overlap scores ≥ 0.14. So these results from RANDOM are consistent based on the data used.

Overall, the CONCEPTBASED augmented method (■) performs better than the BOW($\times$), CONCEPTBASED(●) and RANDOM methods. The BOW method performs well because the document vocabulary used in BOW is large, but the vocabulary used in the CONCEPTBASED method may be too limited. The complexity of the representation method in CB-AUG overcomes the limitation faced by the CONCEPTBASED method. It is observed that the graphs for CB-AUG, BOW, and CONCEPTBASED fall as the number of recommendations, *n* increases. This behaviour is as expected because the earlier retrievals are more likely to be relevant. However, the overlap of CB-AUG and BOW at higher values of n may be because the documents retrieved by both methods are drawn from the same neighbourhoods. Further, the CB-AUG method may be better at ranking the relevant documents it retrieves. These results show that augmenting the representation of documents with a larger concept vocabulary, as done in CB-AUG, is a better way of employing the background knowledge for representing learning materials.

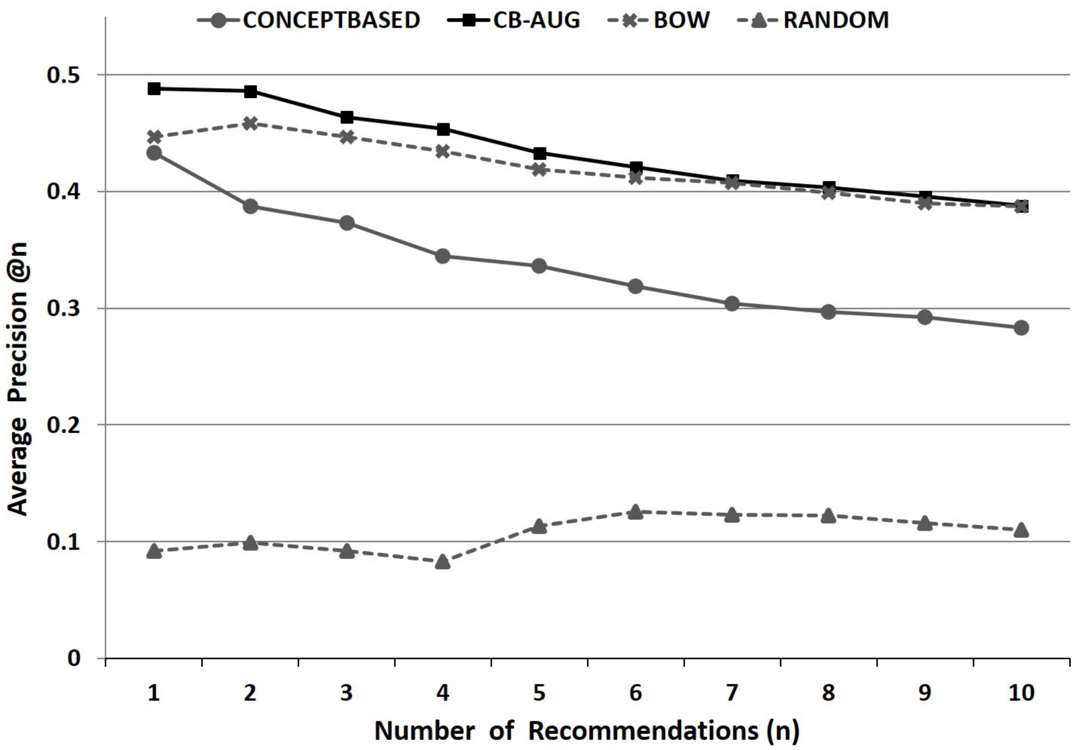


Figure 3.8: Precision of the methods at an overlap threshold of 0.14

Figure 3.9 illustrates the precision of the methods at an overlap threshold of 0.25. The x-axis contains the number of recommendations *n*, while the y-axis contains the average precision at the different values of n. The relative performance at a threshold of 0.25 is similar to the performance at 0.14 however, this is a more challenging threshold for all the methods. Again, the performance of RANDOM at a threshold of 0.25 is consistent with the relationship between the threshold and the proportion of data. At this tougher threshold of 0.25, the average precision for RANDOM is about 0.05. Recall from Figure 3.7 that there were only 5% of document pairs with overlap scores ≥ 0.25. Hence the results of RANDOM are consistent with the threshold used for this dataset. There is an unexpected behaviour observed from CB-AUG and BOW at this more challenging threshold. Both methods do not perform well on the first retrieval, but improve at the second retrieval. For values of *n* from 2 to 10, the graphs for CB-AUG and BOW fall as the value of *n* increases, which is as expected. The performance of CONCEPTBASED is as before with a gradual decrease as the number of recommendations increases. Again the limited vocabulary used in CONCEPTBASED limits its performance in this retrieval task.

Generally, the results show that the CONCEPTBASED augmented document representation method is able to identify relevant learning resources by highlighting the concepts they contain,

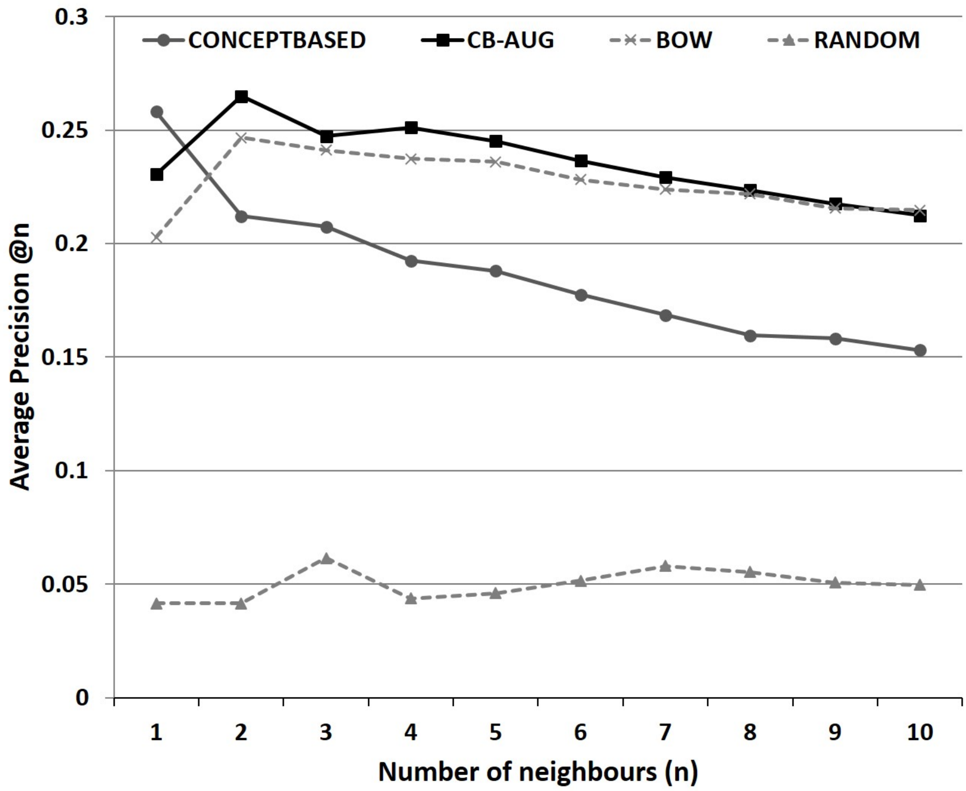


Figure 3.9: Precision of the methods at an overlap threshold of 0.25.

and this is important in e-Learning. The graphs show that employing a knowledge driven approach to support the representation of learning resources is useful for e-Learning recommendation.

3.5 Summary

Finding relevant learning materials to recommend to learners within e-Learning recommendation tasks can be challenging. This is because the learning materials are often unstructured text, and so are not easily indexed for retrieval. Hence the need for a suitable method of representing learning materials with the aim of improving recommendation. Furthermore, the vocabulary used in learning materials by domain experts is usually different from the vocabulary used by learners when trying to find relevant materials. The mismatch in vocabulary presents a semantic gap.

A step is taken to bridge the semantic gap by creating a method that automatically creates custom background knowledge in the form of a set of rich concepts related to the selected learning domain. The domain-specific background knowledge is created by exploiting a structured collection of teaching materials as a guide for identifying important learning concepts. The identified concepts are enriched with descriptive text from an encyclopedia source. Discovered text from the encyclopedia forms a pseudo-document for each concept. These pseudo-documents are used to extend the coverage and richness of the representation. So, each concept is made up of a concept label and an associated pseudo-document. The concept-space consists of the vocabulary from the concepts which is employed for document representation.

The developed background knowledge captures both key topics highlighted by the e-Book TOCs that are useful for teaching, and additional vocabulary related to these topics. So, the concept space provides a vocabulary and focus that is based on teaching materials with provenance. The CONCEPTBASED document representation method takes advantage of similar distributions in the concept and document spaces to define a concept term driven representation. CONCEPTBASED focuses on the concept space, by using only the concept vocabulary, however this vocabulary is from a limited number of concepts, so it is too restricted for concept-based distinctiveness.

The CONCEPTBASED augmented document representation method exploits differences between distributions of document terms in the concept and document spaces, in order to boost the influence of terms that are distinctive in a few concepts. The evaluation results confirm that augmenting the representation of learning resources with a knowledge-rich representation as done in CB-AUG improves e-Learning recommendation. The larger vocabulary from both concepts and documents has been focused by the use of the vocabulary from the concept space.

Chapter 4

Enhanced Representation

A suitable representation for an e-Learning domain should have a good coverage of relevant topics from the domain. This would allow for an approach that caters for recommendations that meet learners' queries which can be varied. One issue highlighted in the results from Chapter 3 was that the concept generation method produced a few concepts resulting in a limited concept vocabulary. In this chapter, the challenge associated with the representation of learning materials is further explored. The concept generation method used in the previous chapter is enhanced to improve our background knowledge and increase the coverage of the concept vocabulary. An enhanced method for representing documents is developed. In addition, the performance of the developed method on a larger dataset is examined.

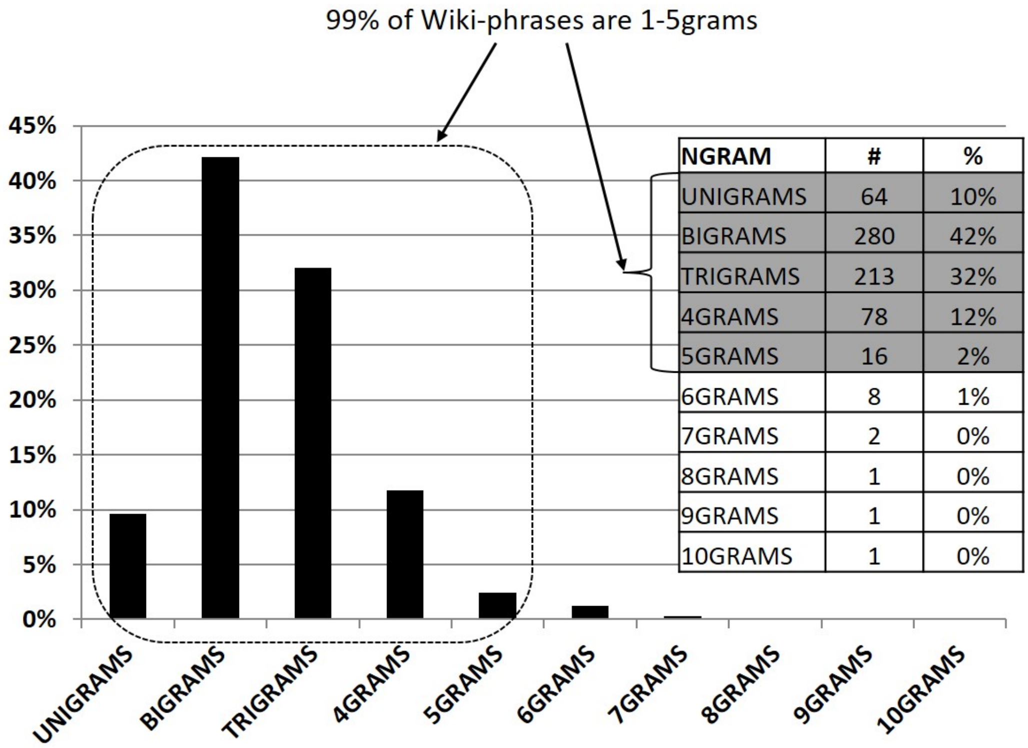
4.1 Enriching the Domain Concepts

Domain concepts are potentially very useful for representing learning resources, because they contain important topics that describe a domain. The advantage in using domain concepts for representing learning resources is that the concept vocabulary allows the retrieval of the represented resources to focus on the domain concepts contained in learning resources, and this is useful for e-Learning recommendation. In this section, the method used previously for generating domain concepts is refined to address the issue of a limited concept vocabulary. The generation of a larger concept vocabulary which provides a better coverage of the learning domain is explored with the aim of creating a richer knowledge source that can be employed for tasks such as the representation of learning resources.

4.1.1 Enhanced Concept Representation

In refining the concept generation method, the phases of the knowledge extraction process described in section 3.2.2 are applied. In addition to the TOC stopwords, the SMART stopwords (Salton 1971) are also removed during pre-processing. This allows words that do not contribute to learning terms to be removed while retaining a good set of words for generating the domain concepts. Words referring to the name of the domain used for demonstration such as: *machine*, *learning*, *data*, and *mining* are not removed during pre-processing, because it was observed that removing these words before ngram generation prevents other relevant ngrams which contain these words such as *instance based learning* or *reinforcement learning*, from being identified.

Figure 4.1 shows a distribution of the Wiki-phrases in the domain lexicon. The x-axis contains the Wiki-phrases based on the number of n-grams contained in the phrases. This ranges from 1 to 10 representing unigrams to 10grams respectively. The y-axis contains the proportion in percentage terms for each of the ngrams. The embedded table shows the name of the ngram, the number of Wiki-phrases in that category and the proportion of the ngrams in percentage respectively. In Figure 4.1, the dotted lines around the bars and the shaded portion of the embedded table are highlighting that 99% of the Wiki-phrases are 1-5grams.



NGRAM	#	%
UNIGRAMS	64	10%
BIGRAMS	280	42%
TRIGRAMS	213	32%
4GRAMS	78	12%
5GRAMS	16	2%
6GRAMS	8	1%
7GRAMS	2	0%
8GRAMS	1	0%
9GRAMS	1	0%
10GRAMS	1	0%

Figure 4.1: Distribution of Wiki-phrases

Ngram extraction is increased to generate 1-5grams from the TOC-phrases because, the distribution of the Wiki-phrases in Figure 4.1 showed that 99% of phrases are 1-5grams; so this allows the number of concepts that can be generated to increase. Ngram extraction is applied to the TOC-phrases to produce the following TOC-ngrams: 2467 Unigrams; 5387 Bigrams; 3625 Trigrams; 1668 Fourgrams; and 576 Fivegrams. The TOC-ngrams are verified by comparing the TOC-ngrams with the Wiki-phrases. The verification allows us to identify a set of potential concept labels that are actually being used to describe aspects of the domain in another knowledge source, such as Wikipedia. The verification produces a set of potential concept labels containing 24 Unigrams; 96 Bigrams; 38 Trigrams; 6 Fourgrams; and no Fivegrams.

A second verification step is applied to the potential concept labels. In this verification, each potential concept label is retained as a concept label if all of 3 criteria are met. Firstly, if a Wikipedia page describing the concept label exists. Secondly, if the description of the concept label is not contained as part of the page describing another concept label. Thirdly, if the concept label is not a synonym of another concept label. The verified concept labels are used to search Wikipedia pages in order to generate a domain concept. The search returns discovered text that forms a pseudo-document and a concept label. Overall, the enhanced method produces 150 domain concepts that pass the second verification, each having a concept label and pseudo-document pair. The pseudo-document terms are pre-processed using standard techniques of English stop-word removal and Porter Stemming. These terms now form the concept vocabulary of the domain concepts for the enhanced background knowledge, which will be employed in the enhanced CONCEPTBASED* document representation method.

4.1.2 Recommendation using the Enhanced Concept Representation

The enhanced CONCEPTBASED* document representation approach, employs the richer concept vocabulary of the enhanced background knowledge for representing documents. It is expected that the representation created using the enhanced CONCEPTBASED* method provides a better coverage of the learning domain because of the richer concepts it contains. The aim is to address the issue of the limited concepts contained in the initial CONCEPTBASED method. For recommendation using the enhanced CONCEPTBASED* method, the same representation and document similarity as the previous CONCEPTBASED method illustrated in Figures 3.3 & 3.4 is used, but this time with a richer concept vocabulary. So documents are represented with respect to concepts by computing the cosine similarity of term vectors for concepts and documents to produce a concept document matrix. Then, the similarity between documents can be generated by computing the similarity between respective concept vectors for documents.

By using the enhanced CONCEPTBASED* method for representing documents, we expect to retrieve documents that are similar based on the concepts that they contain, and this is obtained from a document-document similarity matrix as previously shown in Figure 3.4b. A standard approach of representing documents is to define the document similarity based on the term document matrix illustrated in Figure 3.3b, but this exploits the document vocabulary only. In the enhanced CONCEPTBASED* approach, more emphasis is put on the domain concepts, so the concept document matrix illustrated in Figure 3.4a is used, to underpin the similarity between documents. The enhanced CONCEPTBASED* method combines the focus from learning concepts with the breadth of a richer set of domain concepts when representing documents.

4.1.3 Evaluating the Enhanced Concept Representation Approach

This section investigates whether the domain concepts generated using an enhanced approach such as the enhanced CONCEPTBASED* approach is better for representing documents than the concepts generated using the initial CONCEPTBASED method. The same evaluation method and dataset 1 containing 217 Machine Learning and Data Mining papers, presented in Section 3.4 is employed here. An overlap threshold of 0.14 is used because there are 10% of document pairs in this dataset with overlap scores ≥ 0.14. A leave-one-out retrieval is employed for evaluating the performance of the different methods.

Figure 4.2, contains the results of the evaluation. The number of recommendations is shown on the x-axis, while the average precision@n is shown on the y-axis. The performance of the enhanced CONCEPTBASED* method (♦), and the CONCEPTBASED augmented method, CB-AUG (■) are shown by the black lines, while the performance of CONCEPTBASED(●) is shown by the bold gray line. BOW(×) is included as the benchmark, while RANDOM(▲) is included to give an idea of the relationship between the threshold used and the precision values. The performance of BOW(×) and RANDOM(▲) are shown using broken gray lines.

It is observed that the graphs of all the methods fall as the number of recommendations, *n* increases. This is expected as earlier retrievals are more likely to be relevant. The CONCEPTBASED* method performs better than CB-AUG on dataset 1. We can recall that CB-AUG had the best performance on dataset 1, however using the richer concepts contained in CONCEPTBASED*, results in a better performance for CONCEPTBASED*. The enhanced CONCEPTBASED* method

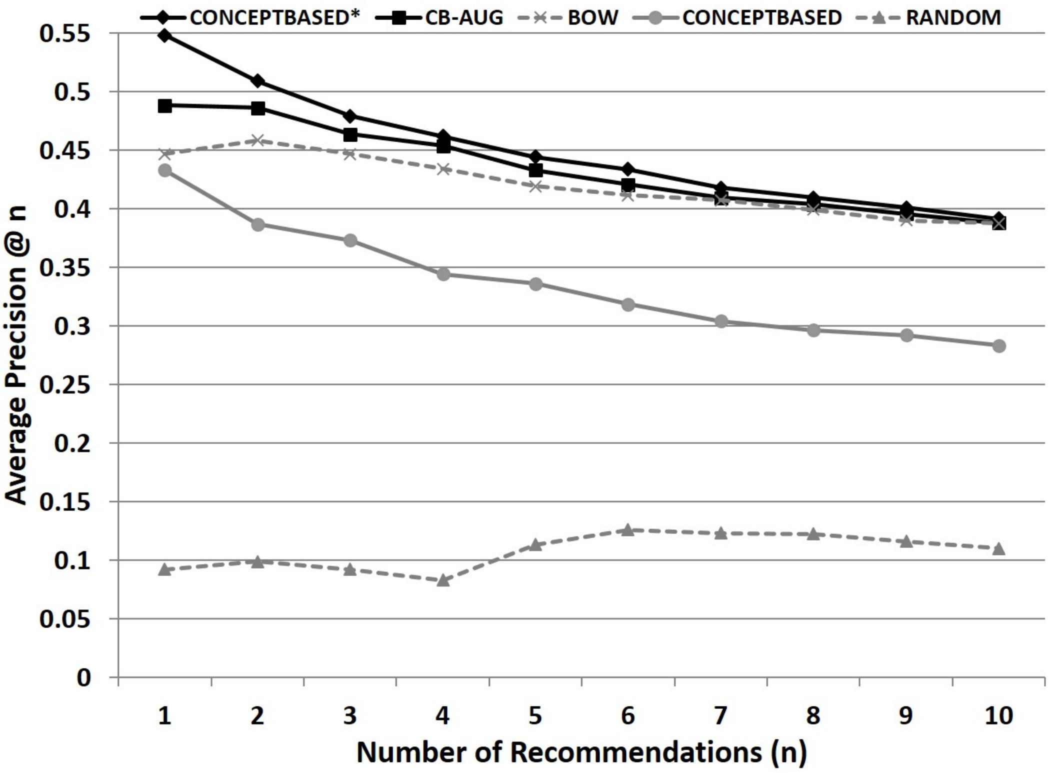


Figure 4.2: Comparing CONCEPTBASED*, CONCEPTBASED and other representation methods at a threshold of 0.14

also outperforms CONCEPTBASED, showing that the enhanced method provides better coverage of the topics in the domain.

Overall, CONCEPTBASED* outperforms CB-AUG, CONCEPTBASED, BOW, and RANDOM, by producing better recommendations for all values of n. This performance shows the advantage of using the richer concept vocabulary for representing learning materials. The results confirm that the enhanced CONCEPTBASED* approach contains concepts that have a better coverage of the learning topics than the CONCEPTBASED method which has a limited set of concepts. By using the domain concepts from the enhanced CONCEPTBASED* method, we are able to address the issue of a poor performance from CONCEPTBASED which had a limited concept vocabulary for representing documents. So we adopt the concept vocabulary from the enhanced domain concepts as a knowledge-rich representation for the Machine Learning and Data Mining domain.

4.1.4 Exploring a Larger Concept Vocabulary

Given that the enhanced CONCEPTBASED* representation method with a larger concept vocabulary performs better than the CONCEPTBASED representation method that had a few concepts.

The performance of a representation method using a much larger concept vocabulary is explored.

Wikipedia contains concepts for many learning domains including the Machine Learning and Data Mining domain. We extract all the Machine Learning and Data Mining concepts that have associated descriptions in Wikipedia to form a set of concepts labels and associated pseudo documents. There are 579 concepts that meet this criteria. The terms from this set of concepts now form the concept vocabulary for the Wikipedia concepts. This vocabulary can be used for representing documents in a method which is referred to as WIKICONCEPT. The implementation of the WIKICONCEPT method is the same as the CONCEPTBASED method, however the concepts used in WIKICONCEPT are 579 concepts from Wikipedia.

In order to measure the performance of the WIKICONCEPT representation, the evaluation method and dataset 1 presented in §3.4 is used with an overlap threshold of 0.14. The evaluation is performed using a leave-one-out retrieval. Figure 4.3 contains the results of the comparison between the enhanced CONCEPTBASED* representation and the WIKICONCEPT representation. The x-axis contains the number of recommendations (n), where n ranges from 1 to 10. The y-axis contains the average precision at the different values of n. Both graphs fall as the value of n increases, which is an expected behaviour, as we would expect earlier retrievals to contain documents that are more likely to be relevant. It is observed that the CONCEPTBASED* method outperforms the WIKICONCEPT method at the different values of recommendations from 1 to 10.

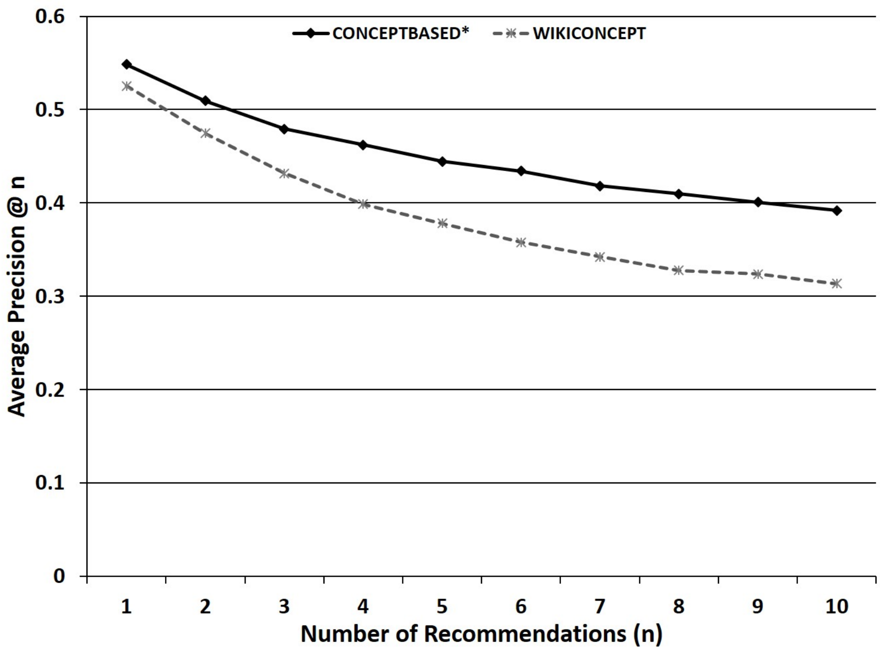


Figure 4.3: Comparing CONCEPTBASED* and WIKICONCEPT at a threshold of 0.14

WIKICONCEPT contains concepts more than 3 times the size of concepts in CONCEPTBASED*. However, the larger concept vocabulary of WIKICONCEPT is not helpful for representation. This may be because the large number of Wikipedia concepts makes it difficult to clearly distinguish distinct concepts. Some of these concepts may be synonyms or specializations of other concepts, thus making it harder to differentiate concepts. CONCEPTBASED* has a good coverage of the domain and performs better at distinguishing relevant concepts in the documents.

4.2 Examining the Representation Methods on a Larger Dataset

The performance of our developed concept based document representation methods are compared against that of the standard BOW approach on a larger dataset, to examine the effectiveness of the concept based methods on a larger document collection. We use a second dataset which we refer to as dataset 2, for the following experiments.

Dataset 2 contains 1000 Machine Learning and Data Mining papers also from Microsoft Academic Research. Figure 4.4 contains a distribution of the keywords per document in dataset 2. The x-axis shows the documents which are sorted based on the number of keywords they contain. The y-axis contains the number of keywords. In this dataset, there are 3063 unique keywords, and 4551 keywords in total. We take advantage of these author-defined keywords for judging the relevance of retrievals made.

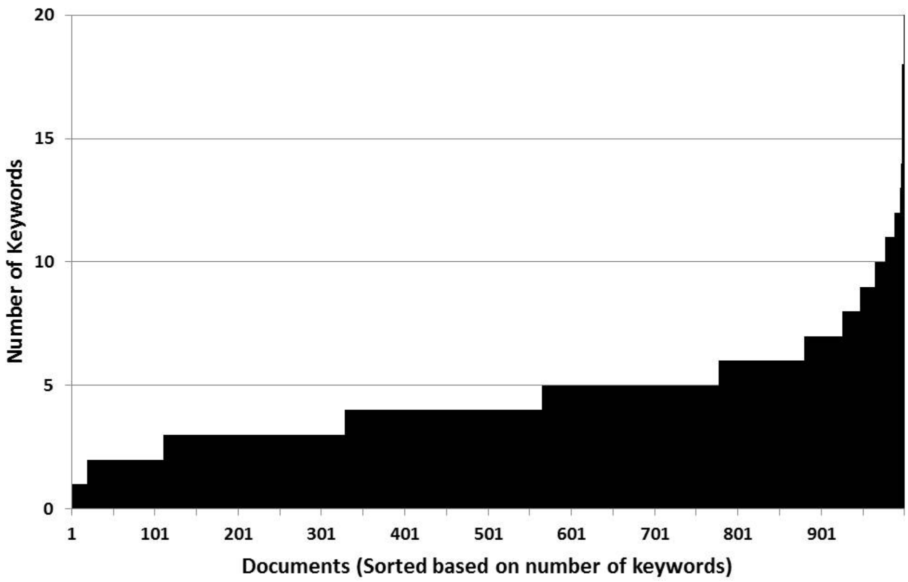


Figure 4.4: Number of keywords per document in dataset 2

A retrieved document is considered to be relevant based on the proportion of its keywords that overlap with the keywords of the query document. The overlap profile of documents in the collection enable us to choose a suitable threshold for the experiments. There are $499,500$ entries for the 1000 document pairs, and $480,129$ entries are zero, meaning that there is no overlap in 96% of the data. So only 4% of the data have an overlap of keywords. Figure 4.5 shows a summary of the overlap profile for the 4% of the document pairs that have an overlap of keywords. The x-axis shows the proportion of pairs in percentage, for those with non-zero values. While, the y-axis contains the overlap threshold. The embedded table in Figure 4.5 contains the overlap threshold values for 3%, 2% and 1% of the data. For example, there are 3% of document pairs with overlap scores ≥ 0.2. Notice that, the higher the overlap threshold value, the lower the proportion of documents considered to be relevant.

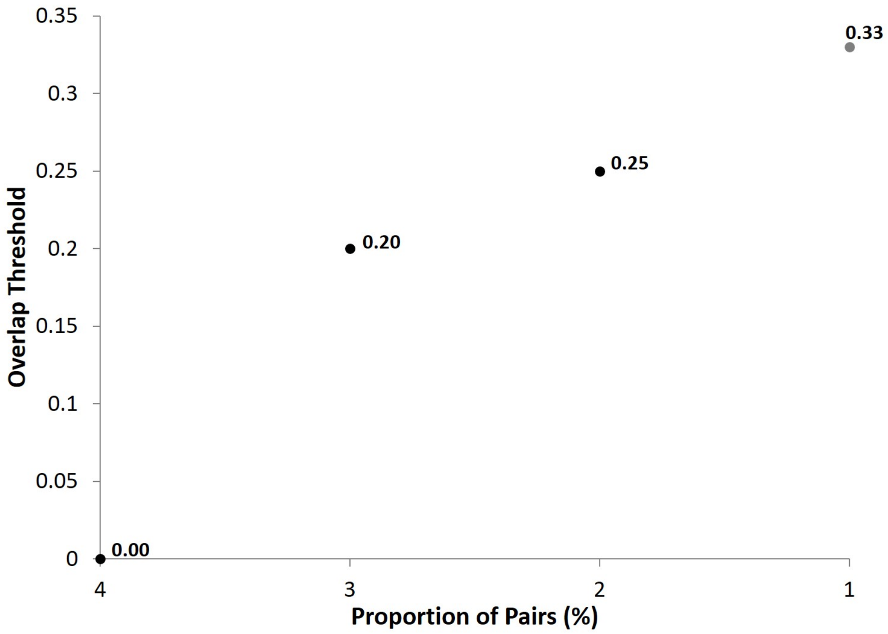


Figure 4.5: Overlap of document pairs in dataset 2

The same evaluation method presented in §3.4.2 is employed here. A leave-one-out retrieval method is applied, and precision@n as in Equation 3.3 is used to determine the proportion of relevant documents retrieved. With dataset 2, we use thresholds of 0.2 and 0.25 thus preventing values that allow either too many or few documents to be considered as relevant. The results of comparing our methods with the standard method are shown in Figure 4.6. The number of recommendations is shown on the x-axis and the average precision@n is on the y-axis. The average precision values are based on the overlap of keywords between document pairs and the threshold value used in the experiment. RANDOM(▲) gives an idea of the relationship between the threshold and the precision values, and the results are consistent with the overlap profile in Figure 4.5. The average precision values for all the methods are generally lower for dataset 2. We recall that fewer documents in this dataset share common keywords, hence the general reduction in average precision values for all the methods.

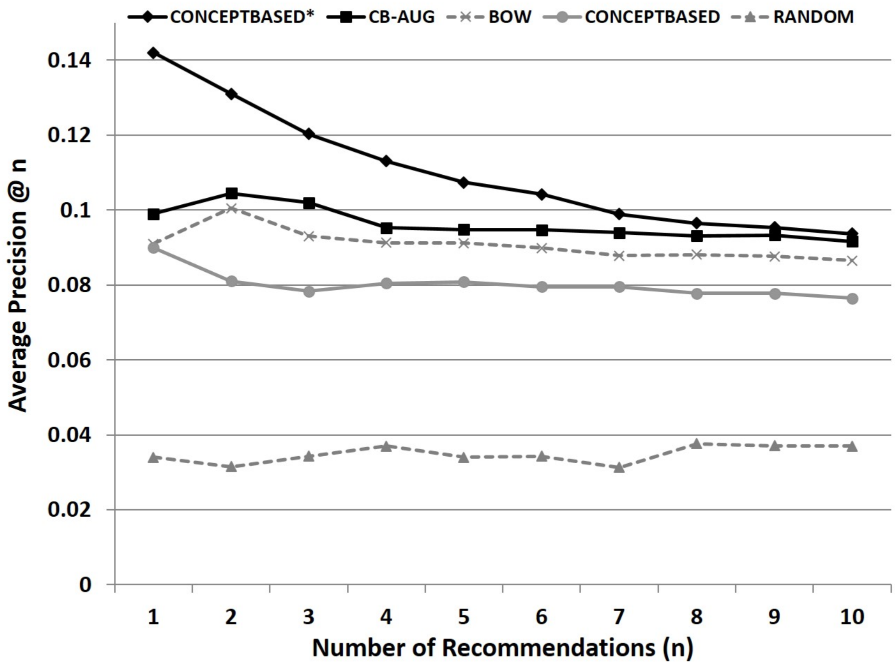


Figure 4.6: Precision of the methods at overlap threshold of 0.2

On this bigger dataset, CONCEPTBASED*(♦) outperforms CB-AUG(■), BOW(×), CONCEPTBASED(●), and RANDOM thus confirming that using a richer and focused vocabulary to represent documents is useful for e-Learning recommendation. The results also show CB-AUG performing better than BOW, again confirming that augmenting the representation of learning resources with domain concepts is better than using the content of documents only for representation in e-Learning recommendation.

Figure 4.7 shows the results of experiments run at a threshold of 0.25. The performance for all the methods drops at this tougher threshold. The relative performance at 0.25 is similar to the performance at 0.2. Again CONCEPTBASED*(♦) outperforms the other methods.

Our results show that we are able to leverage the vocabulary from the enhanced CONCEPTBASED* method which is a larger vocabulary with a better coverage of the domain, thus allowing our method to influence the retrieval and recommendation of relevant learning resources.

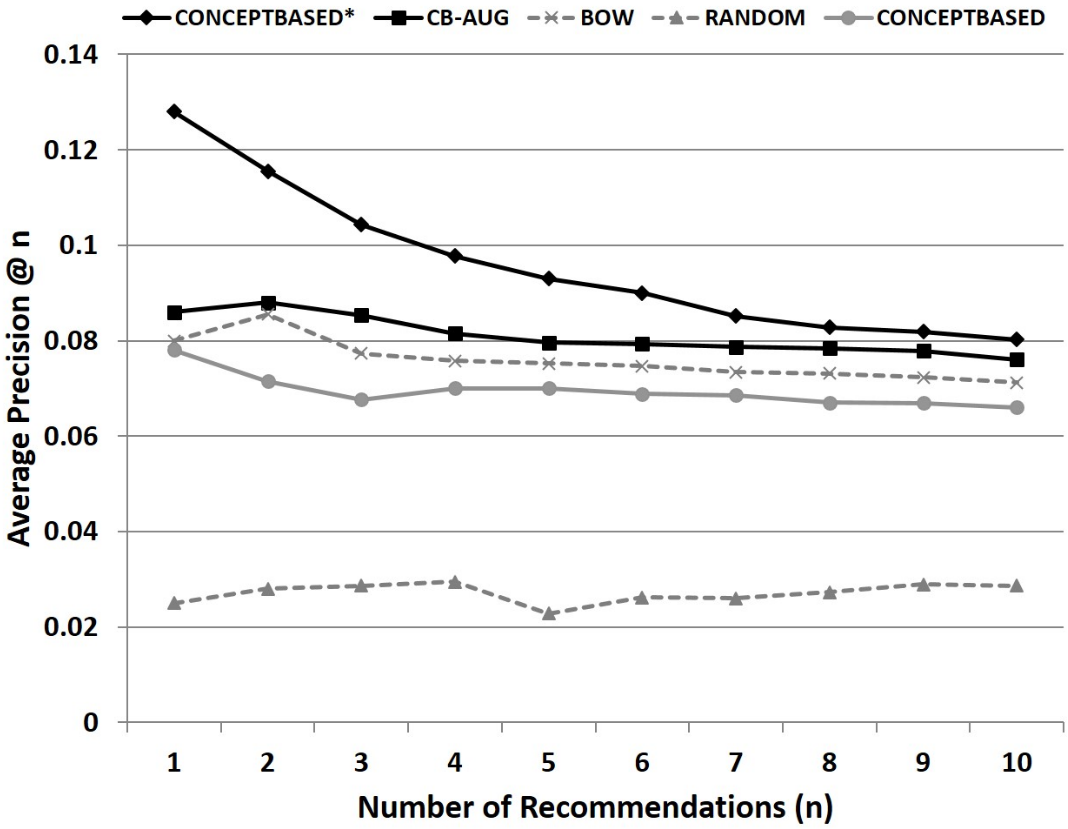


Figure 4.7: Precision of the methods at overlap threshold of 0.25

4.3 Scaling Representation Methods

The use of a larger dataset presents the challenge of dealing with an increased vocabulary. So in this section we explore whether a selection of potentially useful concept terms can be employed for representing documents while still maintaining the effectiveness of our representation methods. The concept terms contained in the documents are examined to give us an idea of the spread of concepts within documents. This distribution of concept terms will allow us to know the proportion of such terms that are useful during the representation process.

An example is shown in Figure 4.8 which is the distribution of 20681 concept terms from CONCEPTBASED*, that are contained in the documents from dataset 1. The x-axis shows the number of documents the concept terms occur in, while the y-axis shows the term frequency in percentage. The embedded table contains a distribution of the concept terms in documents. The first column corresponds to the values in the x-axis, while the second column corresponds to the y-axis in the graph. This distribution shows us that not all concept terms will be useful during the representation process, because there are many documents that do not contain such terms.

It is observed that 63% of the concept terms are not contained in any document, and 9% are contained in only one document, while 5% of concept terms can be found in two documents in the collection. The frequency of concepts terms continues to decrease as the number of documents increases. So the graph does not show individual frequency values for documents greater than 15 and less than 217. This is because the focus is on the earlier portion of the graph, where there are higher frequency values, so the graph only shows results from 0 to 15 documents. A total of 9% of concept terms can be found in the set of documents greater than 15 and less than 217. A value of 9% is very small, that is why each individual value beyond 15 is not shown on the graph. There are 217 documents in total in this collection, so we stop at 217.

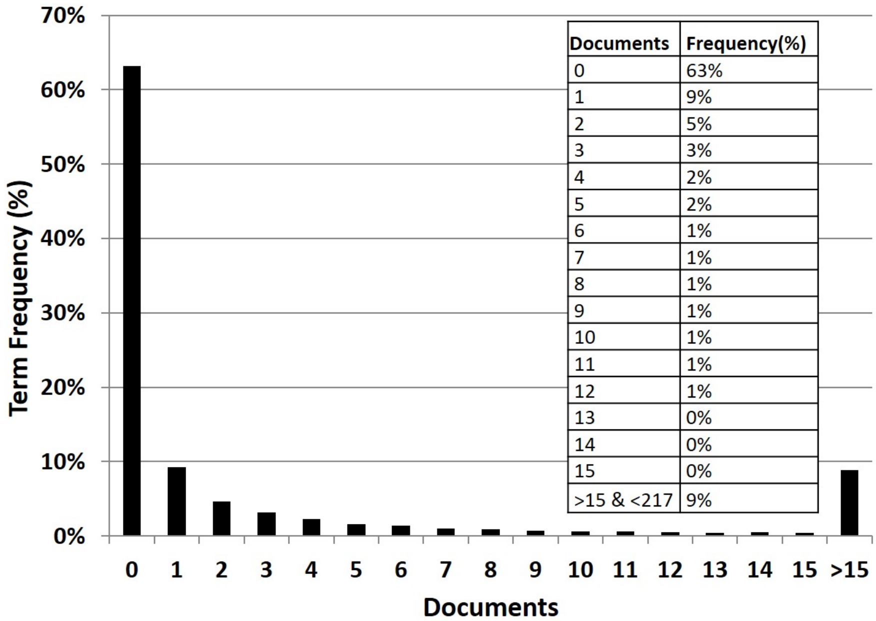


Figure 4.8: Distribution of concept terms contained in a document collection.

A selection of potentially useful terms are examined with the aim of scaling and improving the representation method. This approach will be called the *Scaled* approach. The concept terms that appear in 0 or 1 document, will not be helpful for influencing retrieval. This is because such terms will not be useful for distinguishing documents. So such terms are not used. Terms that appear in all documents, are usually commonly occurring terms, and are often not useful for differentiating between documents. For terms contained in more than 1 document, there is a trade off between the vocabulary size and how effective such terms would be for retrieval. We found retrieval to be effective, when terms from more than 2 documents were used. Hence, we select concept terms that appear in more than 2 documents. The selected concept terms are then used to create a term document matrix for the collection we wish to represent. TF-IDF is used as the weighting scheme when creating the term document matrix because TF-IDF is useful for distinguishing documents. For a document retrieval task, TF-IDF would allow retrieval to focus on the more representative terms.

In order to generate a set of potentially influential concept terms for representation, the highly weighted concept terms are identified from the term document matrix. This is achieved by applying an average TF-IDF method to compute the weight of concept terms in each document within the term document matrix. The average TF-IDF values of the terms are ranked and a percentage of concept terms with high TF-IDF values are selected as the set of concept terms that can then be employed for representing documents. These terms will be called the *Scaled* vocabulary.

4.4 Evaluating the Scaled Representation Approaches

This scaling of the representation methods should be particularly helpful when there is a large number of documents to be represented as scaling should offer some computational gain. A scaled vocabulary is applied to both CONCEPTBASED* and WIKICONCEPT methods using the approach described in §4.3. The output produces the scaled form of both methods, which are referred to as CONCEPTBASED*-SCALED and WIKICONCEPT-SCALED. The evaluation and dataset described in §3.4 is employed. A leave-one-out retrieval is applied to evaluate the performance of the scaled methods against the use of the full vocabulary.

Figure 4.9 contains the results of comparing the full and scaled vocabularies of the CONCEPTBASED* and WIKICONCEPT methods. It is observed that all the graphs are falling as the number of recommendations increases, which is an expected behaviour, as earlier retrievals are more likely to be relevant. For both methods, the results show that using a scaled vocabulary yields better performance than when the entire vocabulary is used for representation. So, a selection of potentially useful terms is more effective for representing documents than using all the terms in the vocabulary. These results demonstrate that scaling is good because the scaled approaches perform as well as, and even better than using the full vocabulary. This shows that scaling is effective for representation of documents.

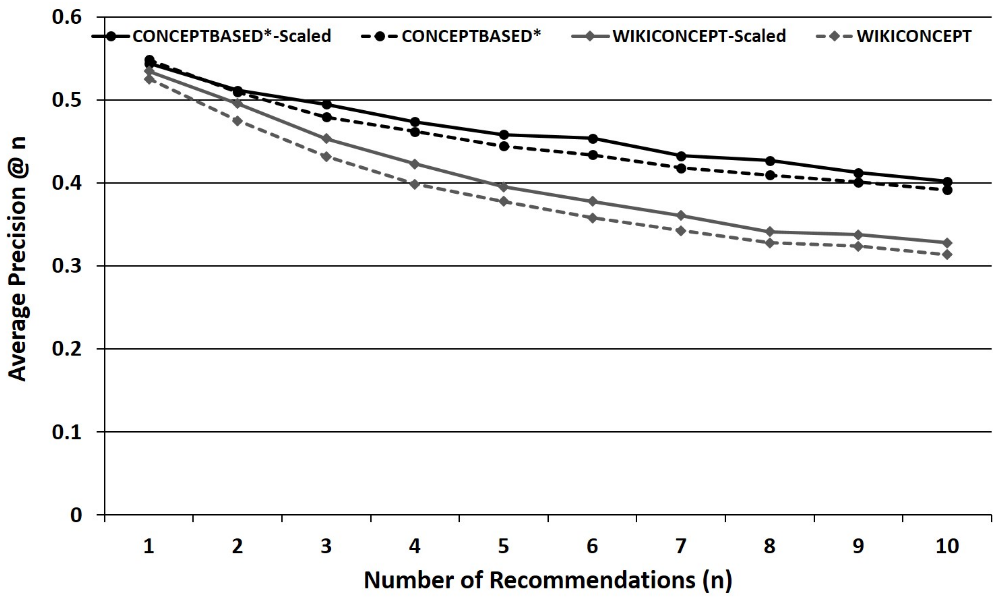


Figure 4.9: Comparing ConceptBased* and WikiConcept on a full and scaled vocabulary at an overlap threshold of 0.14

4.5 Summary

An enhanced representation which provides a better coverage of the learning domain has been examined. The method used to create the background knowledge representation in the previous chapter is refined to generate a richer and focused set of domain concepts. The richer concepts are used to develop an enhanced ConceptBased* document representation approach that is employed for representing learning resources.

It is observed that having a lot of information as seen in BOW, CB-AUG or WikiConcept methods is not effective for creating a suitable representation for learning resources. However, exploiting a rich and focused set of domain concepts is a better approach for representing learning resources. ConceptBased* contains a richer and focused set of domain concepts than the other approaches. Exploiting the richer concept vocabulary from ConceptBased* was shown to be useful for representing learning resources. During the representation process, ConceptBased* is able to focus on the important topics as seen from a learning point of view.

The performance of the methods are examined on a larger dataset. The evaluation results show the improvement in e-Learning recommendation when the richer concept vocabulary is used for representing learning resources. The results also show that the knowledge generated from a structured domain knowledge source such as TOCs is helpful for representing documents. The enhanced set of domain concepts in ConceptBased* will be employed in the rest of the thesis.

Chapter 5

An e-Learning Recommendation Framework

E-Learning recommendation typically involves a learner's query, as an input; a collection of learning resources from which to make recommendations; and selected resources that are recommended to the learner, as an output. Learners often have difficulty asking an effective query, because they lack sufficient domain knowledge to know what topics are relevant for them. This challenge results in an intent gap. A way of addressing this challenge is by creating a suitable method for refining learners' queries with the aim of supporting learners to identify relevant topics and ask effective queries.

Query refinement can be done explicitly or implicitly. The explicit approach to query refinement is often done manually, where a set of concepts that are potentially relevant to the query are shown to the learner. This can be done using a word cloud or an ontology. The learner then selects which concepts are relevant, and the selected concepts are used for refining the query. For example in Figure 5.1, a learner can be shown a word cloud containing a set of concept labels of the most similar domain concepts based on the learner's query. The learner can then select the concepts from this set to be used for refining the query. A word cloud is used in this system because it allows the learner to have a quick view of concepts that are relevant to the query. The size of each concept label in the word cloud is based on the relative frequency of that concept in the set of retrieved documents shown to the learner. Adopting an explicit approach to query refinement often means that a learner would need to be knowledgeable about the learning domain, to decide which concepts are relevant. However, learners do not often have sufficient knowledge of the domain

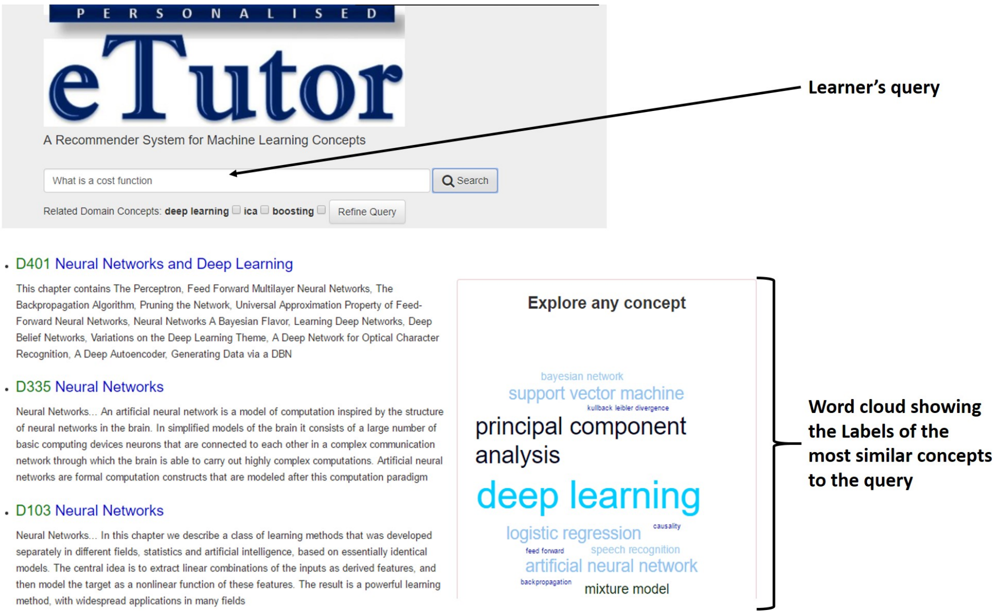


Figure 5.1: Most similar concepts shown to a learner as a word cloud

they are learning about, so adopting this approach poses a challenge. So, the explicit approach does not address the intent gap that learners face.

Alternatively, query refinement can be done implicitly. In this method, concepts that are potentially relevant to a user's query are chosen automatically and used to refine the query. This approach does not need the user's input for choosing relevant concepts. So this helps to address the intent gap in e-Learning recommendation. However, one challenge with the implicit approach to query refinement is that we need to know how much we can generalize or specialize the concepts that would be used for query refinement. This usually involves a way of determining aspects such as: number of relevant concepts to select, what parts of the concepts are chosen, and the number of relevant terms to use for refinement.

An e-Learning recommendation framework has been developed to demonstrate an e-Learning recommendation task. The system architecture employed for building the recommender system is presented together with the technologies used for implementing the system. The document collection used as the items for recommendation are presented. An implicit query refinement approach is created to support the refinement of learners' queries. The developed query refinement approach helps to address the intent gap. The different aspects and knowledge sources used in this approach are examined to address the potential issues identified with implicit query refinement methods. The features of queries are explored for developing the query refinement method with the aim of improving the recommendation made to learners.

5.1 e-Learning Recommender System

There are 3 key components to consider when building a recommender system. These are: the input, the recommendation component that processes the items for recommendation and the output. In this e-Learning recommender system, the query received from the learner is the input. A content-based recommendation approach is used to process a set of e-Learning materials, and the output is a list of recommended learning materials shown to the learner. These components are effectively combined with the aim of recommending relevant learning materials to learners.

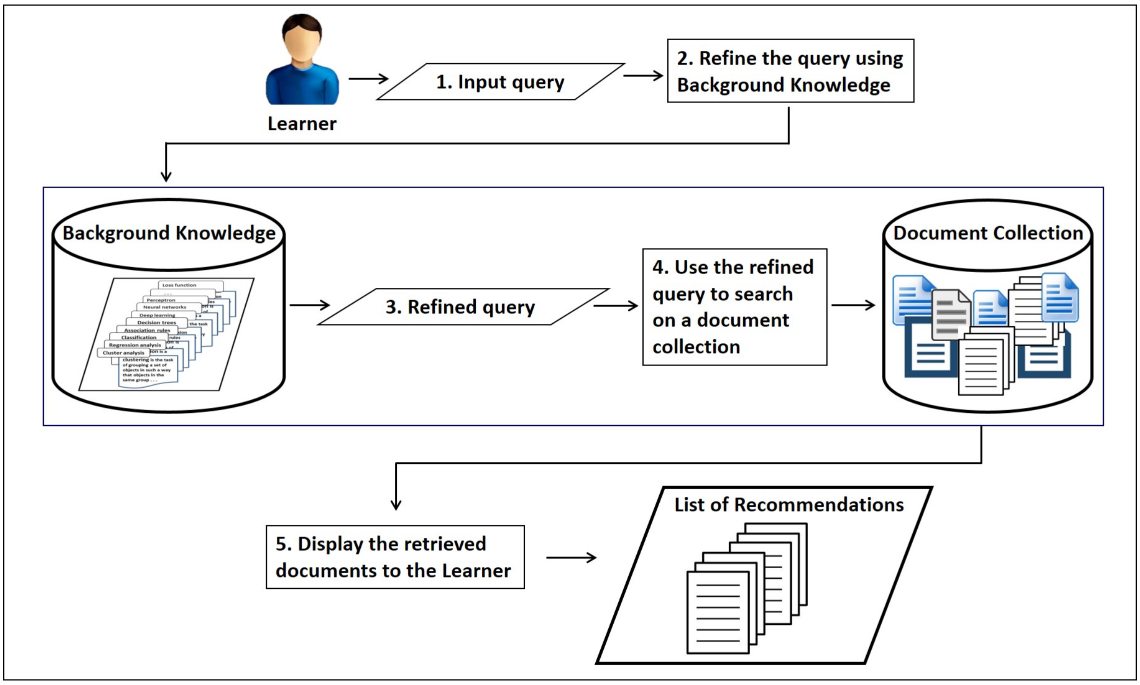


Figure 5.2: System architecture of the e-Learning recommender

Figure 5.2 illustrates the system architecture used in developing the e-Learning recommender system. The process starts at step 1, where a learner inputs a query. In the recommender system, the query received can be represented in different ways with the aim of improving the recommendations made. The terms in the query can be used to search directly on a document collection. This is a Bag-Of-Words (BOW) method, which is a standard document retrieval method. Alternatively, as illustrated in steps 2-4, the developed background knowledge can be employed to produce a concept-based representation of the query, and then a standard document retrieval method applied. In addition, a method that dynamically determines the representation of a query can also be adopted before a standard document retrieval method is applied. The output of this process is shown in step 5 where the list of recommendations are displayed to the user.

5.1.1 Technologies used to build the Recommender System

The e-Learning recommender system is designed as a Single Page Application. The user interface and functions between the client-side and data are created using AngularJS which is a JavaScript framework for building Single Page Applications (Dayley 2014). The design used in the system allows the entire recommendation page to be loaded in the learner's browser after a query is sent. The browser only updates the parts of the page that are needed when a request is made. This allows only the relevant parts of the application to be deployed quickly without loading the entire page each time, thus providing a better experience for learners.

The e-Learning documents and the domain concepts from the background knowledge are indexed using Elasticsearch (Kuć & Rogoziński 2015). Elasticsearch is an open source framework that provides fast indexing for the documents, it is based on Apache Lucene (McCandless, Hatcher & Gospodnetic 2010). Elasticsearch can be installed using a recent version of java, it is started up by typing the "*elasticsearch*" command within its bin directory. To confirm that it is running, its port number is typed into a web browser. For example, one would type "localhost:9200" into a browser. If Elasticsearch is up and running a message including features such as the name of the Elasticsearch instance, the cluster name, id, version number, timestamp, lucene version, and the tagline "*You Know, for Search*" would be shown on the browser.

The data in the system is converted to JavaScript Object Notation (JSON) format [1]. This allows Elasticsearch to effectively parse the data which is stored as name-value pairs. The data is loaded in bulk into the system using the bulk data API. It is important to specify the features contained in the data for the system, so that different features can be suitably indexed. We do not want a numeric feature to be treated as text or vice versa. So we use a mapping to declare the type for each feature in our data. For example, the data may contain a title and description, these two features can be declared as string, while another feature such as a similarity score can be declared using a numeric type such as long. The local server used for hosting the system is the Wampserver 2.5 server [2]. This is started up and ready to serve pages through a Web browser. In this system Google chrome is used as the Web browser.

Figure 5.3 shows the recommender system after all its components have been successfully started up. The system can receive queries from learners and make recommendations based on the queries received. Section a) is the search bar for a learner to input a query. Section b) is the search

[1]http://www.json.org/

[2]http://wampserver.aviatechno.net/

button that activates the search process. Section c) contains the refine query button, while section d) contains the most similar concepts to a query. A learner can select any concept labels to add on to the query as a means of explicitly refining the query, and then click on the refine query button to trigger a new search with the new query. The result is a new list of recommendations. Section e) contains the top recommendation made for the query. Section f) contains a list of recommendations shown to the learner as output. The learner can click on any of these documents to read the learning material. Section g) contains a brief description of a recommended document. The description allows the learner to have a quick idea of the content of a document before selecting a document to read. Clicking on a recommended item opens the document in a modal window, which appears as a page above the current list of recommendations. This design prevents the learner from waiting for a new page to load or completely leaving the list of recommendations to a different page. The learner can read the document and provide a rating for the document. Closing the modal window takes the learner back to the list of recommendations. Section h) shows a word cloud that contains an aggregation of concept labels that are considered to be most relevant to the query. The learner can click on any concept label to add it to the query as a means of explicitly refining the query.

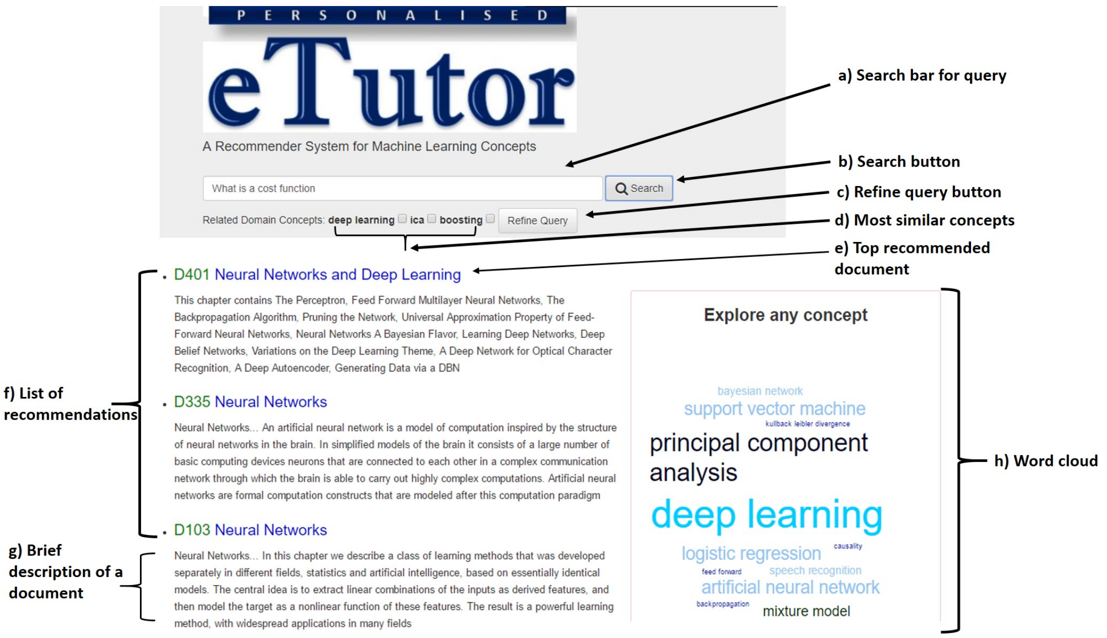


Figure 5.3: The e-Learning recommender system

5.1.2 Collection of e-Learning Documents

A document collection of e-Learning materials is accessed by the system. The documents are 504 chapters of 32 Machine Learning and Data Mining e-Books. This collection of e-Books is different from the e-Books that were used for building the background knowledge representation. This dataset is used because e-Books are designed to be effective for learning and teaching. Furthermore, e-Books have provenance associated with them, given that each book has at least one author associated with it. Hence, our use of e-Books instead of other sources such as Web pages, or blog posts. The collection used is fairly spread across the concepts as shown in Figure 5.4, which highlights the distribution of documents over the concepts. In Figure 5.4, the x-axis shows the 150 domain concepts, while the y-axis shows the normalised average of the cosine similarity between each document and a respective concept. To compute these values, the cosine similarity between Concepts and Documents is first evaluated to produce a Document Concept matrix; the entries into this matrix are row normalised. The average of the cosine similarity for each concept is taken and these values are used to create the scatter plot shown in Figure 5.4.

The outliers shown in Figure 5.4 have different influences when retrieving similar concepts. The outliers to the top show concepts that are very similar to many documents. This can be due to the following reasons. First, such concepts can contain very little descriptive text, so when a comparison is done, it matches many documents in the collection. Second, such concepts may contain text that is common to many of the documents, so again such concepts are similar to many documents. These concepts belonging to the outliers at the top would not be very useful for distinguishing between documents.

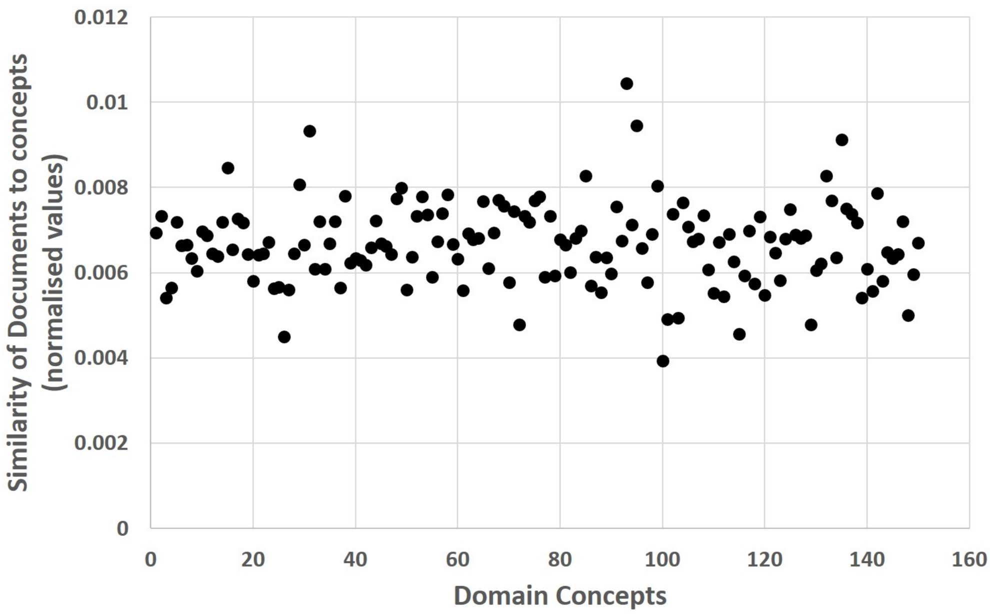


Figure 5.4: Distribution of documents

The outliers at the bottom in Figure 5.4 are from very specialized concepts, so they are not similar to many documents. However, such concepts would be useful for identifying documents that refer to more specialized domain concepts. The middle and bottom sections of the graph are the areas of interest, because these are the concepts that would be very useful for identifying relevant topics contained in documents when a query is received. The graph is not skewed to only a few concepts, so potential bias to a limited number of concepts is avoided. The documents are fairly distributed over the concepts in the chosen domain, so the collection is suitable for use within an e-Learning recommendation task.

5.2 Knowledge Rich Approach to Query Refinement

The enhanced background knowledge presented in Chapter 4 is used to support the refinement of queries as a step towards addressing the intent gap. The background knowledge contains a set of domain concepts for a given domain. Each domain concept has a label and an associated pseudo-document which contains a description of the concept. Domain concepts that are similar to a learner's query are identified from the background knowledge representation and used to influence the refinement of learners' queries. The terms from the pseudo-documents and concept labels of the most similar concepts are put together to form a potential refined query. Highly weighted terms from the potential refined query form the vocabulary that is used to create a refined query. The refined query can then be used to search for learning materials. We expect the materials to be relevant to the learner because the query used has been represented using a vocabulary drawn from learning concepts in the domain.

The CONCEPTBASED query refinement method (CONCEPTBASED-QR), which employs domain concepts to refine queries is presented in the following sections. An example of a refined query is given, and we explore the features of queries to determine suitable representation approaches to adopt for a given query. In order to enable us demonstrate the refinement of queries within an e-Learning recommendation framework, a collection of queries are generated from users. The process involved in generating the query collection is presented.

5.2.1 Refining Queries using Domain Concepts

The background knowledge contains a rich set of domain concepts, each representing an important topic in the domain. Each domain concept is made up of a concept label and an associated pseudo-document. We leverage the domain concepts in the development of the CONCEPTBASED-QR method. The concept vocabulary containing terms t_1 to t_c, from the concepts, C_1 to C_m is used to create a concept term matrix with TF-IDF weighting (Salton & Buckley 1988). TF-IDF is useful for distinguishing concepts within the concept space, hence its use in this method. Each row of the concept term matrix represents a concept, while each column represents a term. The entries into the matrix are TF-IDF weights that have been row normalised as a scaling measure to allow for equal comparison among all the concepts.

In §4.3, we found that a selection of potentially useful terms is effective for scaling up the representation rather than using all the terms in the vocabulary. So we adopt this approach when creating a representation for a refined query. In doing this, the concept terms are ranked based on their average normalised TF-IDF values. The top 10% of terms with the highest TF-IDF values are selected and used to create a new concept term matrix as shown in Figure 5.5. The selected terms t_{c1} to t_{cn}, from the concepts, C_1 to C_m are the set of potential terms that would be used for refining a query.

Terms / Concepts	t_{c1}	t_{c2}	. . .	t_{cn}
C_1	tf-idf			
C_2				
. . .				
C_m				

Figure 5.5: Concept Term Matrix using selected concept terms

When a new query is received from a learner, a search is performed on all the domain concepts. A ranked list of domain concepts that are similar to the query is retrieved. The terms from the term-vectors of the most similar concepts are put together to create a potential refined query. Terms with the highest weights are selected from the potential refined query and added to the initial query to create a refined query. The refined query can be used to search on a document collection, and documents would be retrieved and presented to the learner. We expect the retrieved documents to be relevant because the query used for the search has been generated using domain concepts related to the initial query.

Figure 5.6 contains an illustration of how a refined query is generated. In this example, Cq_1, Cq_2, and Cq_k are the k most similar concepts to the query, while t_{c1}, t_{c2} to t_{cn}, are the selected concept terms. The entries into the matrix are the tf-idf weights of the terms in the respective concepts. While, $SimScore_1$, $SimScore_2$ and $SimScore_k$ are the similarity scores between the query and concepts Cq_1, Cq_2, and Cq_k respectively. The weight of a concept term such as t_{c1} in the potential refined query is generated by computing the weighted sum for that term. An example is shown in Equation 5.1 of how the weight of term t_{c1} is computed. The weighted sum of t_{c1} is achieved by multiplying the weight $SimScore1$ with the tf-idf scores of terms that appear in concept Cq_1, and taking the column sum for t_{c1}. Altering the tf-idf weights of concept terms with their respective similarity scores would allow terms from concepts that are more similar to the query to have more influence in the refined query. The output from this process is a potential refined query containing concept terms, t_{c1} to t_{cn} together with their respective weights.

Terms / **Concepts**	$\mathbf{t_{c1}}$	$\mathbf{t_{c2}}$	...	$\mathbf{t_{cn}}$	**SimScore**
$\mathbf{C_{q1}}$	tf-idf($\mathbf{t_{c1}}$, $\mathbf{C_{q1}}$)				$\mathbf{SimScore_1}$
$\mathbf{C_{q2}}$	tf-idf($\mathbf{t_{c1}}$, $\mathbf{C_{q2}}$)				$\mathbf{SimScore_2}$
$\mathbf{C_{qk}}$	tf-idf($\mathbf{t_{c1}}$, $\mathbf{C_{qk}}$)				$\mathbf{SimScore_k}$
Potential refined query	$\mathbf{t_{c1}}$	$\mathbf{t_{c2}}$	...	$\mathbf{t_{cn}}$	

Figure 5.6: Generating a refined query

$$WeightedSum(t_{c1}) = \sum_{i=1}^{k} tf\text{-}idf(t_{c1}, C_{qi}) \times SimScore_i \tag{5.1}$$

where t_{c1} is a concept term, and $tf\text{-}idf(t_{c1}, C_{qi})$ is the tf-idf score of term, t_{c1} in the i-th concept, C_{qi}, and $SimScore_i$ is the similarity between the query, q and the i-th concept C_{qi}.

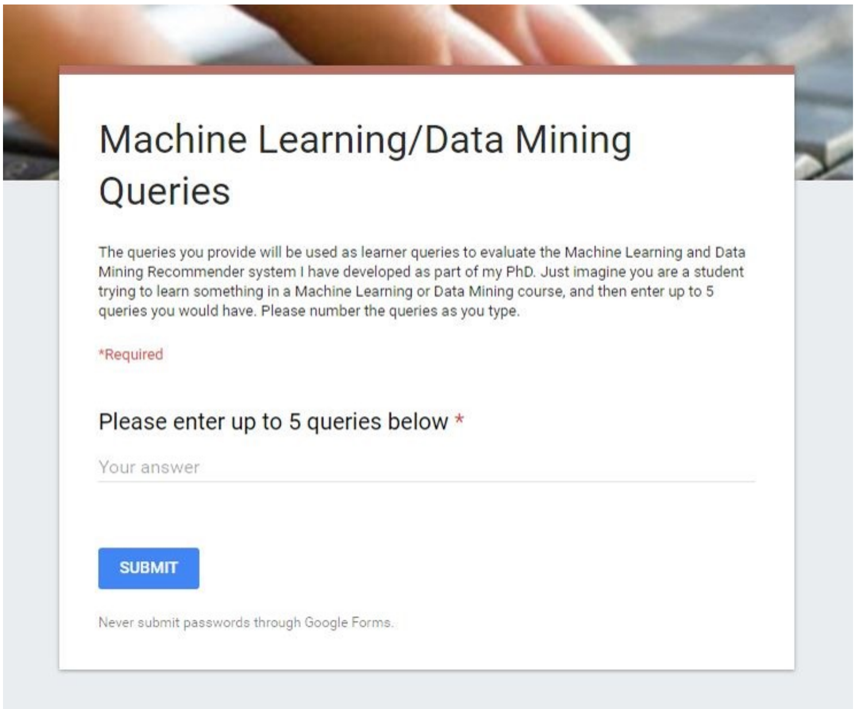
The weight of a term in the potential refined query gives an indication of the importance of the term within the concept space in relation to the given query. We take advantage of this weight by selecting the highly weighted terms from the query document. These terms are then used for generating a refined query. We adopt this approach so that the refined query would not deviate from the initial query because of using noisy terms for refinement (Xu et al. 2009). One challenge this poses is determining a suitable number of concept terms to select for the refined query. In our

ConceptBased-QR method, a set of terms with the highest weights from the potential refined query are selected and added to the initial query to generate a refined query. The initial query is also included as part of the refined query to allow us to maintain the context of the query and prevent a drift in the refined query. The refined query can then be used to perform a search on a document collection. The result is a set of documents we expect would be relevant to the learner, because the query used has been refined using a vocabulary drawn from learning concepts in the domain.

A refined query can be seen as the *initial query* + *generated concept terms*. For example, given an initial query from a learner such as: *"How do you implement gradient descent algorithm?"*. A search is performed on the set of domain concepts and the 3 most similar concepts to this query are: stochastic gradient descent, backpropagation, and winnow algorithm. The terms from these concepts are put together as described in §5.2.1. Ten of the highest weighted terms from an amalgamation of these concepts are: gradient, descent, stochastic, formula, update, momentum, delta, rate, derivative, backpropagation. These terms would then be added to the initial query. So the refined query becomes: *how do you implement gradient descent algorithm gradient descent stochastic formula update momentum delta rate derivative backpropagation*. This refined query is used to search on a document collection.

5.2.2 Generation of Query Collection

The query collection contains realistic learner-focused queries that are used for evaluating the representation methods applied to refine queries in the recommender system. The queries are generated from two sources. First, postgraduate students in the School of Computing Science and Digital Media took part as learners in generating queries. An e-mail specifying the task was sent to them. In order to allow learners to send anonymous responses, and return their queries without seeing what others had typed, a Google form was created to capture the queries. Figure 5.7 shows a screen shot of the Google form created for collecting learner queries. Next, two online sources, Courseras Machine Learning MOOC and Quora, were used to generate additional learner queries. Coursera was chosen because it had recently run a MOOC in the domain we are using for this evaluation. In addition, Quora was selected because it includes Machine Learning and Data Mining as topics it covers. Course specific questions were accessed from Coursera's MOOC, while the open questions in Quora from the Machine Learning and Data Mining topics were chosen.


Machine Learning/Data Mining Queries

The queries you provide will be used as learner queries to evaluate the Machine Learning and Data Mining Recommender system I have developed as part of my PhD. Just imagine you are a student trying to learn something in a Machine Learning or Data Mining course, and then enter up to 5 queries you would have. Please number the queries as you type.

*Required

Please enter up to 5 queries below *

Your answer

SUBMIT

Never submit passwords through Google Forms.

Figure 5.7: Google Form for collecting queries

For each of the query sources used, the aim was to generate realistic learner queries, so queries where the user wanted to learn about a technique, for example: “How does cluster analysis work” were used. Other generic or career-related queries such as “What is it like to be a data scientist at Amazon?”, or queries that were out of scope such as “is there any course on ML?” were not included. There were 25 queries generated from learners and 60 queries generated from online sources, resulting in 85 queries. From this set, 15 of the queries were randomly chosen for parameter tuning while 70 queries were left for the user evaluation. Each query had an Identifier (ID) associated with it. A random number generator was used to generate the IDs of the queries used for parameter tuning.

5.3 Aspects of the Query Refinement Method

In this section we examine two key aspects of the CONCEPTBASED-QR method. First, the vocabulary used for refinement. Second, the parameters used within the CONCEPTBASED-QR method. Some queries generated by learners can be high level and the concept labels are also high level. So we investigate the vocabulary that is used for query refinement, to determine if using only concept labels are sufficient or if using a bigger vocabulary would be better. We wish to determine how specific the vocabulary used for query refinement should be. We compare when only the concept labels are used for refinement with when the descriptions together with the concept labels are used. Using the concept labels only is denoted by **Labels only**, while using the descriptions and labels is denoted by **Descriptions + Labels**. These notations are used to refer to the methods in the following sections.

We expect suitable parameters that will allow us to refine queries effectively and enable us to find relevant documents. Two parameters of the query refinement method are: the number of domain concepts selected and the number of highly weighted terms to select from the potential refined query. The first parameter examined is the number of concepts to select for refinement. It is important to choose a suitable number of concepts for query refinement to avoid deviating from a learner's goal. The number of concepts is needed after an initial query from a learner has been compared with all the domain concepts. The second parameter is the number of highly weighted terms to select from the potential refined query. Selecting a suitable number of terms is necessary in order to avoid using noisy or irrelevant terms for refinement. The selected terms with highest weights would be added to an initial query to generate a refined query.

5.3.1 Experimental Design

The e-Learning recommender presented in Chapter 5 is used for the following experiments. The dataset accessed by the recommender contains 504 chapters of Machine Learning and Data Mining (ML/DM) e-Books. A description of this dataset was presented in §5.1.2. The experiments are evaluated on a collection of 15 learner queries randomly chosen for initial experiments, by using a random number generator.

A document retrieval task is used to evaluate the relevance of the top 3 documents retrieved for each query. The top 3 documents are selected because from our previous document retrieval experiments, we found that earlier retrievals are more likely to be relevant, hence we focus on the first 3 retrievals in this task. The relevance of each retrieved document is evaluated by the researcher using a 5-star rating mechanism as recommended in (Weijters, Cabooter & Schillewaert 2010). Where, 1-star is poor and 5-stars are very good. Experiments are run using the CONCEPTBASED-QR method with the number of concepts set at 1, 3, 5, 7 and 10. These are denoted as CB1 to CB10. For example at CB1 for **Labels only**, the label of the most similar concept to the query is added to the initial query for refinement. For **Descriptions + Labels** at CB1 the most similar concept containing its pseudo-document and label is retrieved and used for query refinement as described in §5.2.1.

5.3.2 Results and Discussion

Figure 5.8 contains the results of a comparison of the part of the domain concepts to use for refinement. The x-axis shows the number of concepts, while the y-axis contains the average ratings for the top 3 most relevant documents retrieved for each variant of the method. The shaded bars in Figure 5.8 show when **Labels only** are used for refinement. For using **Labels only**, there is an increase in the average rating from CB1 to CB3, when two more concept labels are added to an initial query. However, from CB3 to CB10 there is a decrease in the average rating as the number of concepts increase, as CB10 < CB7 < CB5 < CB3. The best performance for using **Labels only** is at CB3, when the three most relevant concept labels are added to an initial query for refinement.

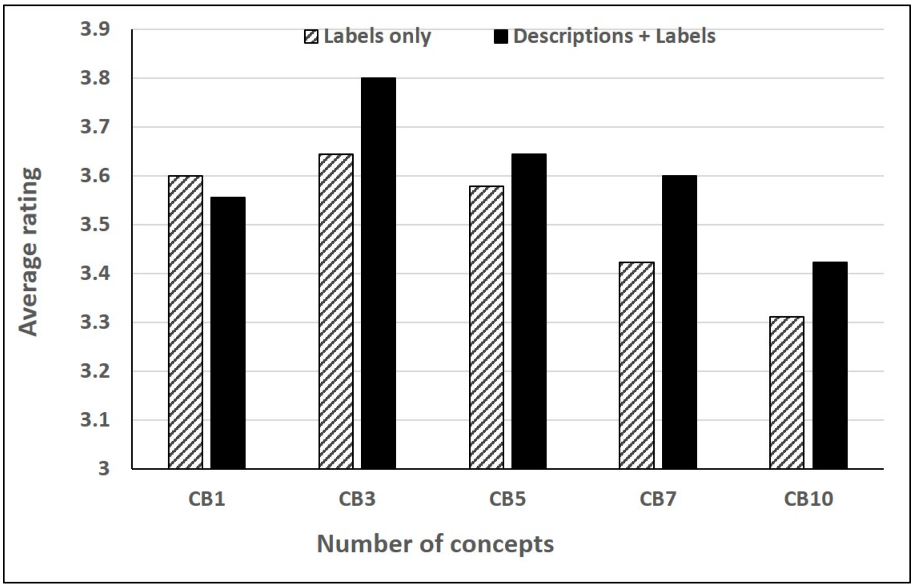


Figure 5.8: Comparison of the vocabulary to use for refinement

The solid dark bar in Figure 5.8 represents when **Descriptions + Labels** are used. For **Descriptions + Labels**, it is observed that using only 1 concept is not very helpful for generating useful terms to refine the query with. The terms from a single concept may be too limited to influence the retrieval of relevant documents. Better performance is observed from CB3 when the descriptions and labels from 3 concepts are used for refinement. So, using up to 3 concepts provides better coverage of the domain. The performance of CB5 reduces compared to that of CB3, but using the terms from 5 concepts is still better than using the terms from only 1 concept for refinement of a query. The performance of this method continues to fall with higher number of concepts, as the performance of CB7 < CB5, and CB10 < CB7. The performance of CB10 even falls below that of CB1. Using the terms from many concepts such as 10 concepts for refinement would mean generalizing the query further. Also, going beyond 10 concepts may not be useful given that there are 150 concepts in this collection of domain concepts used. The best performance for using **Descriptions + Labels** is at CB3, when the descriptions and concept labels from the top 3 domain concepts are used for query refinement.

Concept labels are good to show to users during a query refinement process because the labels contain some terms that learners may recognise. However, the result of the experiments show that using a bigger vocabulary from descriptions and labels is better for representing a query in order to find relevant documents. The best performance is at CB3, when the rich descriptions and Labels from the top 3 concepts are used for refinement. So for refining our queries we would use the **Descriptions + Labels** from the 3 most similar concepts.

The number of highly weighted words to use for query refinement is examined in the following experiment. The method used is **Descriptions + Labels** from CB3 which we found to give the best performance from the previous experiment. The performance of using the top 10, 25, 50, 100, and 250 words are compared in this experiment. Figure 5.9 contains a comparison of the number of words to select from a query document. At 10 words, the average rating is 3.8, but increasing the number of words to 25 improves the average rating to 3.96. The rating reduces to 3.8 as the number of words is increased to 50. The reduction in average rating continues as the number of words are increased to 100 and 250. Using too many words to refine a query would mean generalizing the initial query a lot. The best performance in this experiment is achieved when 25 words are used.

The results of the experiments show that the best performance for the CONCEPTBASED-QR method is when the top 25 words from the **Descriptions + Labels** of the 3 most similar concepts are added to an initial query to generate a refined query. So we adopt these settings for the CONCEPTBASED-QR method.

5.4 Knowledge Source of Pseudo-documents

The pseudo-documents of the domain concepts play a key role during query refinement, because they provide the vocabulary used for refinement. We examine the knowledge source of the descriptive text used within the pseudo-document of each concept. The aim is to generate descriptive text that would be effective for refining queries. The text from 2 sources are compared. These are: Wikipedia pages and DBpedia abstracts. Our pseudo documents currently contain descriptions

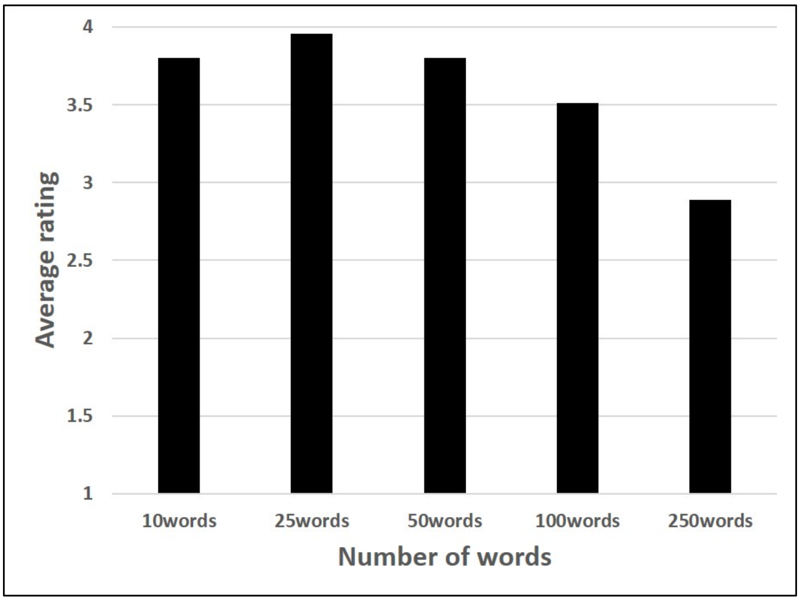


Figure 5.9: Number of words to use for refinement

from Wikipedia pages. Initials experiments show the presence of some noisy terms from the Wikipedia pages, hence our consideration of text from DBpedia abstracts. The DBpedia abstract is created from the first section of the Wikipedia entry for a concept. The DBpedia abstract contains a meaningful description of a concept and it does not contain the noisy and irrelevant terms often found as part of a Wikipedia page for a concept. So we would expect better performance from using DBpedia abstracts given the reduction in noisy terms.

5.4.1 Experimental Design

The performance from using DBpedia and Wikipedia are compared on a small retrieval experiment. The e-Learning recommender system presented in §5.1 is used. The experiments are run using the CONCEPTBASED-QR method with the number of concepts set to 3, and the number of words retrieved from the potential query document is 25 words. We use these parameters because they are the best performing parameters for the CONCEPTBASED-QR method. A random sample of 5 queries from the 15 queries used for parameter tuning are employed in this task. The queries are randomly selected using a random number generator. For each query, the top 5 documents are retrieved. To judge the relevance of retrieved documents to each query, the researcher rates each retrieved document using a 5-star rating mechanism where, 1 is poor and 5 is very good.

Figure 5.10 shows a comparison of the first 10 highly weighted terms retrieved from concepts populated using text from DBpedia and Wikipedia for the 5 sample queries. This gives us an idea of the kind of terms retrieved by the two sources. In Figure 5.10, terms that either appear in the query or in both sources are highlighted in bold. As these are the highly weighted terms, many of them are relevant to the sample query shown. For query 1, 3, 4 and 5 both knowledge sources share two-five words in common. For query 2, DBpedia is the only knowledge source to share a term with one of the query terms, none of the top 10 terms produced from the Wikipedia concepts are common to the query terms. In the Wikipedia output for Query 3, the second and third terms picked up are "displaystyle" and "boldsymbol", which do not appear to be relevant terms for a query such as "deep learning". It is observed that the DBpedia output for this same query is able to avoid such noisy terms.

1	**Query**	**Association rule mining**
	DBpedia	**rule** induction **rules association** paradigms **web mining** extracted transactions itemset
	Wikipedia	**mining rule web** data **rules** frequent delta learning **association** items
2	**Query**	**How to perform fast real time clustering**
	DBpedia	**clustering** cobweb conceptual concept kmeans objects spectral cluster laplacian clusters
	Wikipedia	deep ensemble genetic bma layers crossover bucket fitness mutation bagging
3	**Query**	**Deep learning**
	DBpedia	**deep lazy layers** neural speech networks **layer** recognition moments **learning**
	Wikipedia	**learning** displaystyle boldsymbol **lazy** layers **deep** eager competitive **layer** data
4	**Query**	**How do you implement gradient descent algorithm**
	DBpedia	**gradient descent stochastic** formula update momentum delta rate derivative backpropagation
	Wikipedia	**gradient descent stochastic** frequent apriori winnow gsp sequences optimization genetic
5	**Query**	**What is the vanishing gradient problem**
	DBpedia	**gradient descent** stochastic **formula** gradients momentum rate update **neural** iteration
	Wikipedia	deep ica boosting signals **formula** signal layers pdf kurtosis categorization **neural**

Figure 5.10: Terms from DBpedia abstracts vs terms from Wikipedia pages

5.4.2 Results and Discussion

Figure 5.11 shows a detailed view of the top 5 retrievals made using DBpedia and Wikipedia for each of the 5 sample queries. For query 1, both knowledge sources produce documents that are relevant to the query. Recall from Figure 5.10 that both sources had up to 5 terms in common and they contained all the query terms. Hence, there is such a good retrieval set of documents produced for query 1. This pattern is also similar to the results for query 3, where both sources have a number of terms in common including the query terms. A similar behaviour is seen in queries 4 and 5. Here there are fewer terms in common to both sources, and we also see a slight reduction in the average ratings of documents retrieved for the query. For query 2, the Wikipedia source is not able to retrieve documents that are relevant to the query. Although this behaviour is unexpected, it is not surprising given that none of the terms produced by Wikipedia in Figure 5.10 were relevant to the query. Further investigation to this behaviour from the Wikipedia concepts reveals that when a term such as "clustering" taken from the initial query is used as the query, relevant documents can be found. However, using the full phrase in query 2 poses a difficulty for Wikipedia concepts. On the other hand, DBpedia concepts are able to avoid this problem, and find documents that are relevant to the query.

	Query	DBpedia	DBpedia rating	Wikipedia	Wikipedia rating
1	Association rule mining	1 -> Itemset Mining	4	1 -> Itemset Mining	4
		2 -> Frequent Itemsets	4	2 -> association pattern mining	5
		3 -> association pattern mining	5	3 -> Regional Association Rule Mining and Scoping from Spatial Data	5
		4 -> association pattern mining advanced	5	4 -> association pattern mining advanced	5
		5 -> Association Analysis	5	5 -> Frequent Itemsets	4
2	How to perform fast real time clustering	1 -> Kernel Spectral Clustering and Applications	5	1 -> Genetic Algorithms for Subset Selection in Model	1
		2 -> Clustering	5	2 -> Neural Networks and Deep Learning	1
		3 -> Clustering	5	3 -> Pre Trained Deep Neural Networks A Hybrid	1
		4 -> Clustering both hierarchical and kmeans clustering	5	4 -> Three Classes of Deep Learning Networks	1
		5 -> A Survey of Constrained Clustering	4	5 -> Neural networks	1
3	Deep learning	1 -> Neural Networks and Deep Learning	4	1 -> Neural Networks and Deep Learning	4
		2 -> Deep Autoencoders Unsupervised Learning	4	2 -> Pre Trained Deep Neural Networks A Hybrid	4
		3 -> Pre Trained Deep Neural Networks A Hybrid	4	3 -> Three Classes of Deep Learning Networks	5
		4 -> Three Classes of Deep Learning Networks	5	4 -> Deep Stacking Networks and Variants supervised learning	3
		5 -> Selected Applications in Speech and Audio Processing	2	5 -> Deep Autoencoders Unsupervised Learning	4
4	How do you implement gradient descent algorithm	1 -> Stochastic Gradient Descent The LMS Algorithm and its Family	5	1 -> Stochastic Gradient Descent	5
		2 -> Stochastic Gradient Descent	5	2 -> Stochastic Gradient Descent The LMS Algorithm and its Family	5
		3 -> The Single Neuron as a Classifier	3	3 -> Fitness Meta Modeling	1
		4 -> Neural Networks and Deep Learning	4	4 -> Search and Optimization Methods	1
		5 -> Neural Networks	4	5 -> Efficient learning	1
5	What is the vanishing gradient problem	1 -> Stochastic Gradient Descent The LMS Algorithm and its Family	5	1 -> Neural Networks and Deep Learning	4
		2 -> Stochastic Gradient Descent	5	2 -> Selected Applications in Speech and Audio Processing	1
		3 -> Parameter Learning A Convex Analytic Path	2	3 -> Some Historical Context of Deep Learning	3
		4 -> Neural Networks and Deep Learning	4	4 -> Three Classes of Deep Learning Networks	4
		5 -> Online Learning	1	5 -> Pre Trained Deep Neural Networks A Hybrid	4
		Average rating	**4.16**		**3.08**

Figure 5.11: Top 10 Terms from DBpedia abstracts and Wikipedia pages

Figure 5.12 shows the results for using DBpedia vs Wikipedia respectively as a knowledge source for pseudo-documents. Overall, the documents retrieved for queries refined using DBpedia abstracts have an average rating of 4.16 while those found using Wikipedia concepts have an average rating of 3.08. We can see from this small experiment that there is better performance when the focused abstracts from DBpedia are used as the source of descriptive text for the pseudo-documents of concepts. So we adopt DBpedia abstracts as the knowledge source for the concepts.

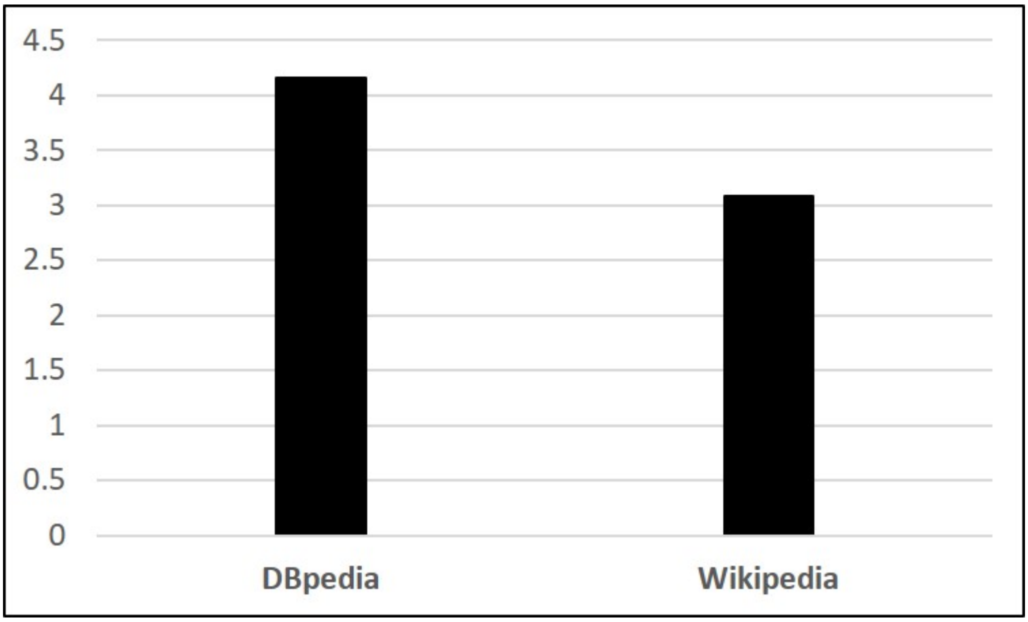


Figure 5.12: Using DBpedia abstracts vs Wikipedia pages as a knowledge source

5.5 Exploring Query Features for a HYBRID Approach

Query refinement can be harmful because the terms used for refining a query can sometimes be too specific or too generic for the query. Initial experiments on refining queries reveals that some queries are high level and well structured, and may not need refinement for relevant documents to be retrieved. Examining the features of queries can be useful for determining when a query may need refinement. A HYBRID query refinement method is introduced which explores query features to decide when to use a standard BOW method or the CONCEPTBASED-QR method. The features explored in developing the HYBRID method are the length of a query, the presence of a concept label in a query, and the similarity of a query to the concepts. The concept label in a query was found to be a dominant feature, so the HYBRID query refinement method takes advantage of the presence of a concept label in a query to make a dynamic choice in determining when to apply the BOW or CONCEPTBASED-QR method to refine a query.

The following two observations provide justification for using the concept label as a determining feature in HYBRID. First, the BOW method tends to work well on queries that clearly identify the topics that are relevant for the domain. These topics are captured in the concept labels contained in the queries. Such queries do not often have a big intent or semantic gap, which is an indication that the queries are from learners that have sufficient knowledge of the domain. So, such queries may not need refinement and can be used as they are to search for documents.

Second, the CONCEPTBASED-QR method is designed to be effective for queries that are vague and not well formed. These queries do not typically contain concept labels. Such queries often have a big intent or semantic gap, meaning that the queries would need some refinement before being applied to a search task. This also gives an indication that such queries are from learners that lack sufficient domain knowledge. So the CONCEPTBASED-QR method can be applied to refine queries that do not contain concept labels. We expect that the HYBRID method would allow us to determine when to do query refinement for a given query.

5.6 Summary

An e-Learning recommendation framework has been developed to demonstrate a recommendation task within the e-Learning domain. The developed system will be employed to evaluate different query refinement approaches. The recommender system is designed as a single page application using the AngularJS framework. This allows users to experience seamless interactions with the system because only relevant pages are deployed to the user each time. The domain concepts and e-Learning documents used within the system are indexed using Elasticsearch which provides fast indexing for the concepts and documents. The documents are retrieved based on their similarity to the query. These documents are presented to the user as recommendations. The developed system allows us to demonstrate an e-Learning recommendation task.

One challenge with e-Learning recommendation is that the query received from a learner does not often sufficiently capture what a learner wants to know. A learner's query tends to be short, vague and ambiguous. Using such a query to search for documents does not usually yield relevant recommendations thus posing a challenge for finding the right e-learning materials. One solution is to generate a better query that can be used to influence the retrieval of relevant learning materials.

Two knowledge driven methods for refining a learner's query have been presented in this chapter. First, the CONCEPTBASED-QR method presented in §5.2.1 uses a vocabulary which is drawn from domain concepts to create a refined query. Second, a HYBRID query refinement method has been developed to take advantage of the ability of the CONCEPTBASED-QR method to cope with vague queries, and the ability of BOW to cater for well-formed queries. So HYBRID dynamically chooses to use either the CONCEPTBASED-QR or BOW method for refining a query based on the features of the query. The refined query from these methods can then be used to search on a document collection, and documents are retrieved and presented to the learner. By using learning concepts from the domain to create a refined query, we expect that the retrieved documents will be relevant because the query used for the search has been generated using domain concepts related to the learner's initial query.

In the next chapter, we will see how the HYBRID and CONCEPTBASED-QR methods are applied to queries in a larger e-Learning recommendation task. Queries refined using the HYBRID and CONCEPTBASED-QR methods will be compared against queries with no refinement in a standard BOW retrieval. User evaluation experiments and results will be presented.

Chapter 6

User Evaluation

The query refinement methods developed in the previous chapter are evaluated by users in an e-Learning recommendation task. The aim in this evaluation is to provide relevant documents to learners, so the recommendations of documents produced should be judged by users. Relevance judgement is subjective and depends on the opinions of individual users taking part in the evaluation. So users are employed for evaluating the recommendations made by the different methods. The users employed for this evaluation differ from typical users normally employed in a standard user evaluation. There are PhD students, researchers, lecturers and professors employed as the users in this evaluation. This is because we need users that have some knowledge about the Machine Learning and Data Mining (ML/DM) topic to judge the relevance of recommendations.

In the following sections the user evaluation design is presented, highlighting the task and the evaluation metrics that are used. The profile of users who took part in the evaluation is discussed in more detail. The evaluation results presented are two-fold. First, results for the relevance of recommendations made using the different methods, and second the results from the qualitative feedback received for the coverage of relevant topics across the documents evaluated.

Interesting findings are made when investigating whether documents considered to be relevant for a query also cover topics that are relevant to the query. These results are obtained by comparing the coverage scores with the ratings scores provided by users.

6.1 User Evaluation Design

The user evaluation was designed to allow users to provide independent ratings on the performance of the different approaches. The design seeks to prevent the introduction of any potential bias to a particular method. The e-Learning recommender system presented in §5 is employed for this user evaluation. The word cloud is not included in the evaluation system because we measure automated query refinement approaches rather than users interacting with the system. The 3 methods compared in this evaluation are:

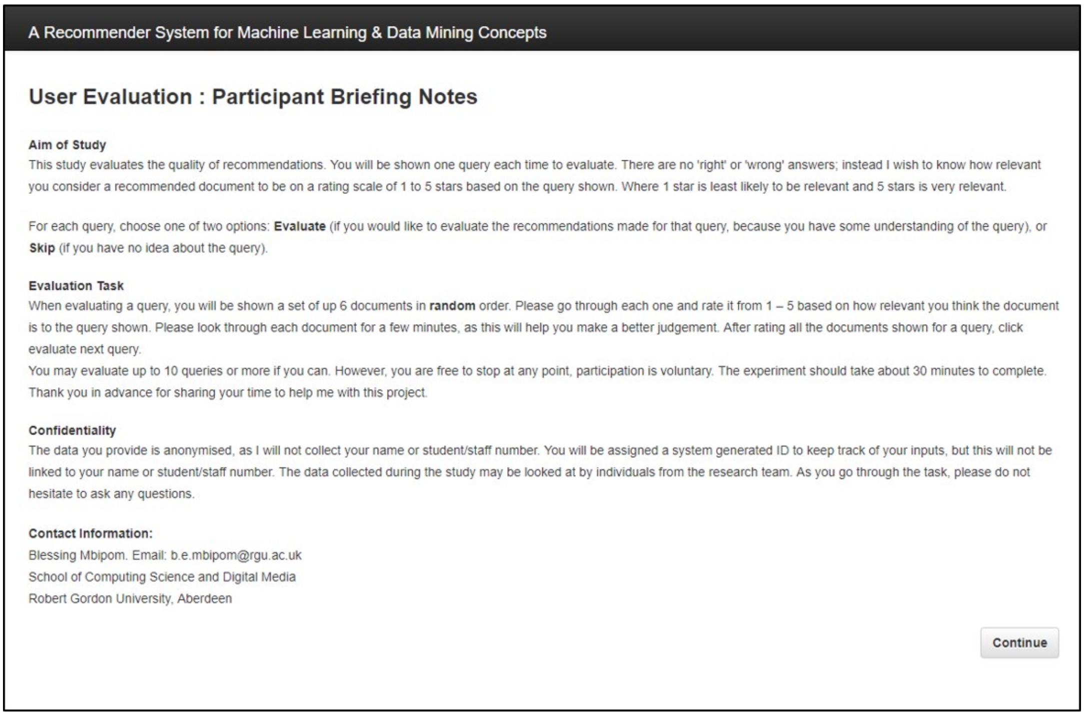
- The Bag-Of-Words method (BOW), which is a standard Information Retrieval method, where a learner's query is represented using the terms in the query only.
- The CONCEPTBASED query refinement method (CONCEPTBASED-QR), which identifies the most similar domain concepts to a query, and uses these identified concepts to create a concept based representation of a query. This method is presented in §5.2.
- The HYBRID query refinement method (HYBRID), which takes advantage of query features to make a dynamic choice in determining when to apply the BOW or CONCEPTBASED-QR methods. This HYBRID method is presented in §5.5.

The e-Learning recommender system discussed in Chapter 5 is used as the evaluation system. The system was deployed on Microsoft Azure (Copeland, Soh, Puca, Manning & Gollob 2015), so the system could be accessible to users online. This allowed users to complete the evaluation at their convenience. The evaluation system was made available online for 8 weeks. A link to the system was shared with researchers working in the ML/DM field across 10 institutions. The learning materials presented in Chapter 5 and a collection of 70 learner queries are given to the users for relevance judgement in this evaluation.

6.1.1 Evaluation Task

An evaluation task typically involves users carrying out certain processes in order to generate some data, while an experiment entails users conducting a set of tasks which produces some results. An evaluation usually involves experiments that measure the performance of different methods. At the start of this evaluation, each user was shown a briefing containing a guide on the evaluation. Figure 6.1 illustrates the briefing notes which contain: the aim of the study; a description of the evaluation task; some information about confidentiality; and the researcher's contact information.

The briefing notes is the first screen shown to users. This is useful, so that the users have clear guidance before they take part in the evaluation. The same notes are shown to all users, so that they have access to the same set of instructions irrespective of when they take part in the evaluation.


A Recommender System for Machine Learning & Data Mining Concepts

User Evaluation : Participant Briefing Notes

Aim of Study
This study evaluates the quality of recommendations. You will be shown one query each time to evaluate. There are no 'right' or 'wrong' answers; instead I wish to know how relevant you consider a recommended document to be on a rating scale of 1 to 5 stars based on the query shown. Where 1 star is least likely to be relevant and 5 stars is very relevant.

For each query, choose one of two options: **Evaluate** (if you would like to evaluate the recommendations made for that query, because you have some understanding of the query), or **Skip** (if you have no idea about the query).

Evaluation Task
When evaluating a query, you will be shown a set of up 6 documents in **random** order. Please go through each one and rate it from 1 – 5 based on how relevant you think the document is to the query shown. Please look through each document for a few minutes, as this will help you make a better judgement. After rating all the documents shown for a query, click evaluate next query.
You may evaluate up to 10 queries or more if you can. However, you are free to stop at any point, participation is voluntary. The experiment should take about 30 minutes to complete. Thank you in advance for sharing your time to help me with this project.

Confidentiality
The data you provide is anonymised, as I will not collect your name or student/staff number. You will be assigned a system generated ID to keep track of your inputs, but this will not be linked to your name or student/staff number. The data collected during the study may be looked at by individuals from the research team. As you go through the task, please do not hesitate to ask any questions.

Contact Information:
Blessing Mbipom. Email: b.e.mbipom@rgu.ac.uk
School of Computing Science and Digital Media
Robert Gordon University, Aberdeen

Continue

Figure 6.1: Briefing notes shown to users

During the study, each user was shown one query at a time to evaluate. Figure 6.2 shows the choice screen for a query. For each query, the user could choose to skip, if the user had no idea about the query; or proceed to evaluate the query because the user felt confident about the query. This allowed each user to evaluate recommendations for queries they were knowledgeable about.

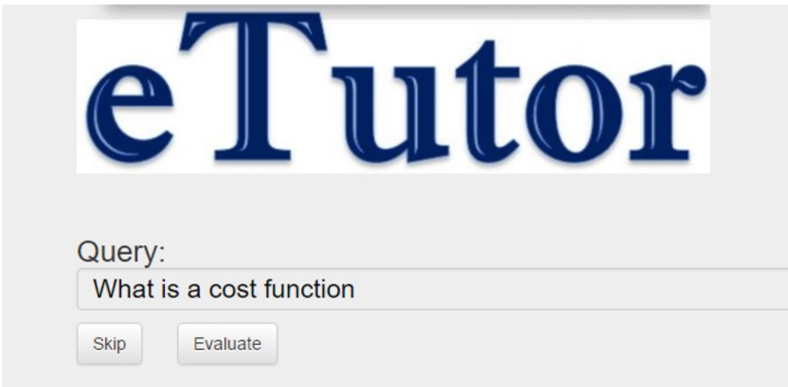


Figure 6.2: Choice screen for a query

When evaluating a query, the user was shown up to 6 retrieved documents in random order. The set of up to 6 documents are the top 3 documents from the CONCEPTBASED-QR and the BOW methods. We select the top 3 documents because previous experiments showed this to be a suitable parameter for judging relevance of retrievals made. Since HYBRID applies either CONCEPTBASED-

QR or BOW, the documents for HYBRID are already included in the retrieval set shown to users, so we do not show documents for HYBRID separately. If less than 6 documents are shown, then this means that the methods have at least one document in common among their top 3 documents. Figure 6.3 shows an example of some recommendations from both methods for a given query. Clicking on a document allows the user to read that document. The user is expected to provide a rating for the document after reading it. The rating provided captures a measure of how relevant the document is to the query being evaluated.

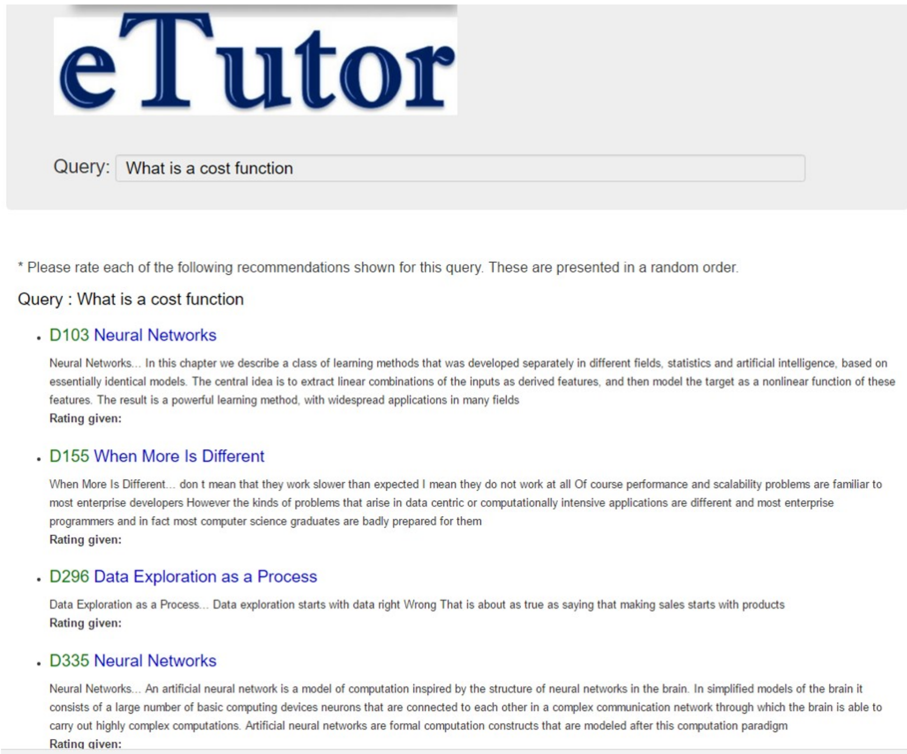


Figure 6.3: List of recommendations

Figure 6.4 illustrates the modal screen used to show a recommended document that the learner has clicked on. The relevance of each document to the query is captured using a rating scale of 1 to 5 stars where 1 is least likely to be relevant and 5 stars is very relevant. A 5-point scale is recommended in (Weijters et al. 2010) for a study such as ours which captures direct summaries of feedback from users as averages. The rating stars as shown in Figure 6.4 were included within the modal that contained the document, so that each user would have an opportunity of interacting with the document before leaving a rating.

It is important that the retrieval set of documents shown to users is presented in a way that avoids any potential bias. Three issues of bias are considered and addressed. First, the users

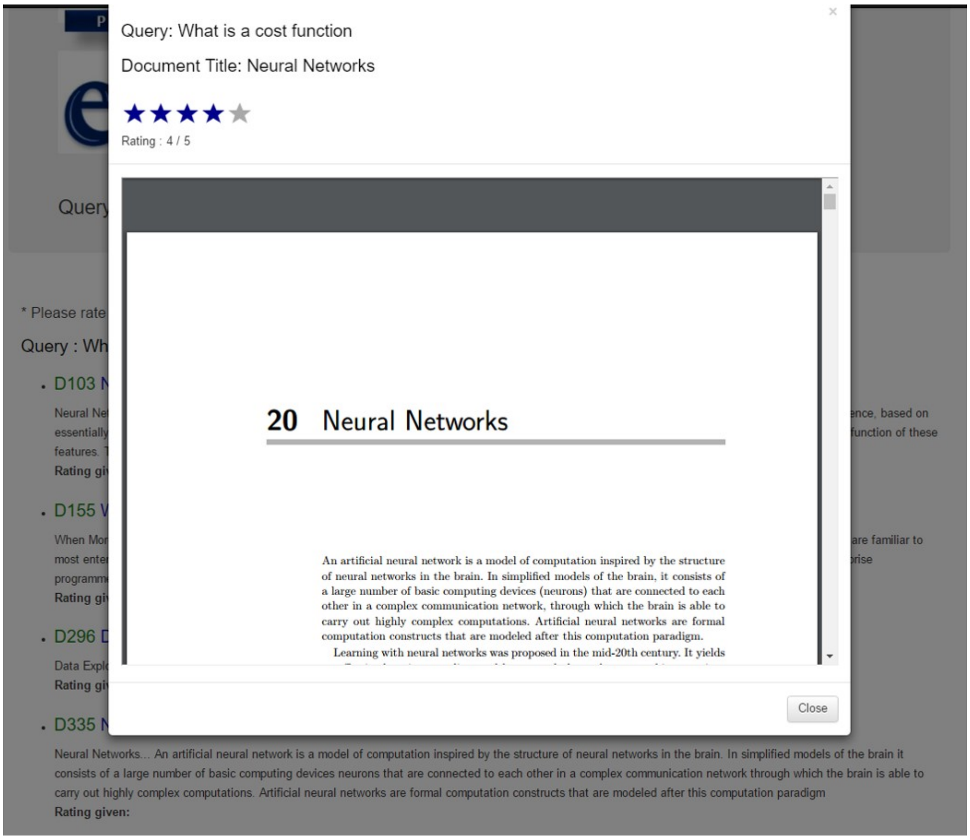


Figure 6.4: Selected document and star-rating

do not know which method produced the recommendations they are evaluating, this prevents a user from favouring one method over the others. Second, the order of documents presented to users is randomized. This rules out the bias of documents shown at the top being considered to be relevant over those lower down the recommendation list. Third, the same user evaluates the recommendations from both BOW and CONCEPTBASED-QR methods for the same query. This ensures that the same user gives an evaluation for both methods at the same time for a given query. This prevents the possibility that we only receive ratings from a positive user for one method, and ratings from a generally negative user for the other method.

6.1.2 Evaluation Metrics

The evaluation uses the ratings given by users across all query-recommendation pairs to measure the performance of the CONCEPTBASED-QR and BOW methods. The following metrics are computed: First, the $user-rating(u)$ which is the average of all the query-recommendation pairs that a user has rated. It is given by:

$$user{-}rating\,(u) = \sum_{|(q,r)_u|} \frac{R_u(q,r)}{|(q,r)_u|} \tag{6.1}$$

where u is a user, (q,r) is a query-recommendation pair, $(q,r)_u$ is the set of (q,r) pairs evaluated by user, u, and R_u is the rating a user has given to a (q,r) pair.

Next, we compute the *rating* which is the average of the ratings from those users who have evaluated the recommendation, r for the query, q. It is given as:

$$rating\,(q,r) = \sum_{u \in U_q} \frac{R_u(q,r)}{|U_q|} \tag{6.2}$$

where (q,r) is a query-recommendation pair, U_q are the users that have evaluated a query, q, and R_u is the rating a user, u has given to a (q,r) pair. The performance of a method is computed by taking the average $rating(q,r)$ across all queries that have been evaluated.

6.2 Users

The users needed for this task should be knowledgeable in the ML/DM domain in order to be suitably qualified to give relevance judgements. This is because the evaluation is not a standard user evaluation where the users are learners. Instead, the users are employed for the purpose of judging the relevance of recommendations made by different methods.

6.2.1 Pre-evaluation Questionnaire

Each user completed a pre-evaluation questionnaire online at the start of the evaluation. The users indicated that they were happy to take part in the evaluation task by ticking a consent box shown to them at the beginning of the questionnaire. The pre-evaluation questionnaire was designed to capture the current role, expertise and experience of the users within the ML/DM field. This design allowed users to provide data about their background, and enabled us to check that each user met the criteria for taking part in the evaluation. Table 6.1 contains the questions that each user was asked. The current role and highest qualification can give us an indication of what each user does currently as a job and their level of education. The options provided for the current role ranged from an *MSc student* to a *Lecturer/Professor*, while the qualifications ranged from *No degree* to a *PhD Degree*. This evaluation needed knowledgeable users, hence the options provided.

Table 6.1: Pre-evaluation questionnaire

Question	**Required**	**Type of Input**	**Options**
Current Role	Yes	Single Selection	MSc Student PhD Student Researcher Lecturer/Professor
University / Institution	No	Text Input	(Filled by users)
Highest Qualification	Yes	Single Selection	No Degree BSc MSc PhD
Experience in ML/DM	Yes	Single Selection	Less than one year One to two years Three To five years Over five years Over ten years
Expertise in ML/DM	Yes	Single Selection	beginner competent expert

A potential bias of having only users from the researcher's university was ruled out by sending the link by email to researchers across different institutions. Researchers were contacted through the International Conference on Case Based Reasoning proceedings list, as their branch of AI means they are more likely to know other researchers who are working in the Machine Learning and Data Mining domain. Users were asked an optional question about their University or institution, to give us an idea of the different institutions that the users taking part were affiliated with. There were responses from users across 10 institutions. Users were asked to make a single selection about their experience and expertise in the ML/DM field. This data would be useful to confirm that the users taking part have some knowledge in ML/DM, because the judgements made by the users would be used to determine the performance of the methods.

Figure 6.5 contains a screen shot of the questionnaire. The data provided through the questionnaire would give us an idea of the experience and expertise of the users. Furthermore, an analysis of the evaluation results using the questionnaire data would allow us to gain valuable insights into the way different users judged the recommendations made by the system.

6.2.2 User Profile

There were responses from 22 users. Figure 6.6 shows the profile of users that took part based on their role, qualification, experience and expertise in the ML/DM domain. For the user roles,

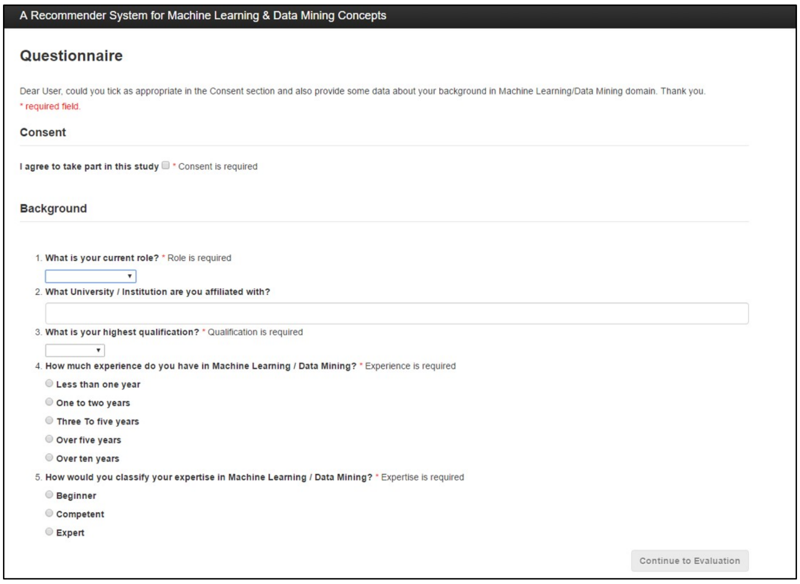


Figure 6.5: Screen shot of the pre-evaluation questionnaire

16 were PhD students, and 3 were Researchers, while 3 were Lecturers or Professors. All the users had at least an MSc degree or higher. There were 3 users with over ten years experience, and another 3 users had over five years experience, while 10 users had between three to five years experience and 5 users had one to two years experience, and only 1 user had less than a year's experience in ML/DM. So majority of users had over three years experience in the domain used for evaluation. This level of experience in the ML/DM topic is useful, because the judgements made should be from people who are conversant with the domain. In terms of their expertise in ML/DM, there were 2 experts, 16 competent users and only 4 beginners. The expertise is based on the user's self-assessment captured in the questionnaire.

The profile in Figure 6.6 shows that most users are competent or expert in the subject. Majority of the users are PhD students, fewer are researchers, and lecturers or professors. So, these users are suitable for this evaluation task because they are the ones that would know best about the learning materials. Hence we can be confident in the judgements that would be provided by them.

6.3 Recommendation Results

Users evaluated 105 queries and provided ratings for 521 query-recommendation (q,r) pairs. On average users evaluated 4.8 queries and provided ratings for 23.7 (q,r) pairs. There were 6 of the 70 queries that were not evaluated. Figure 6.7 shows the spread of ratings for all the (q,r) pairs that were evaluated. In Figure 6.7, the heat maps for the CONCEPTBASED-QR, HYBRID

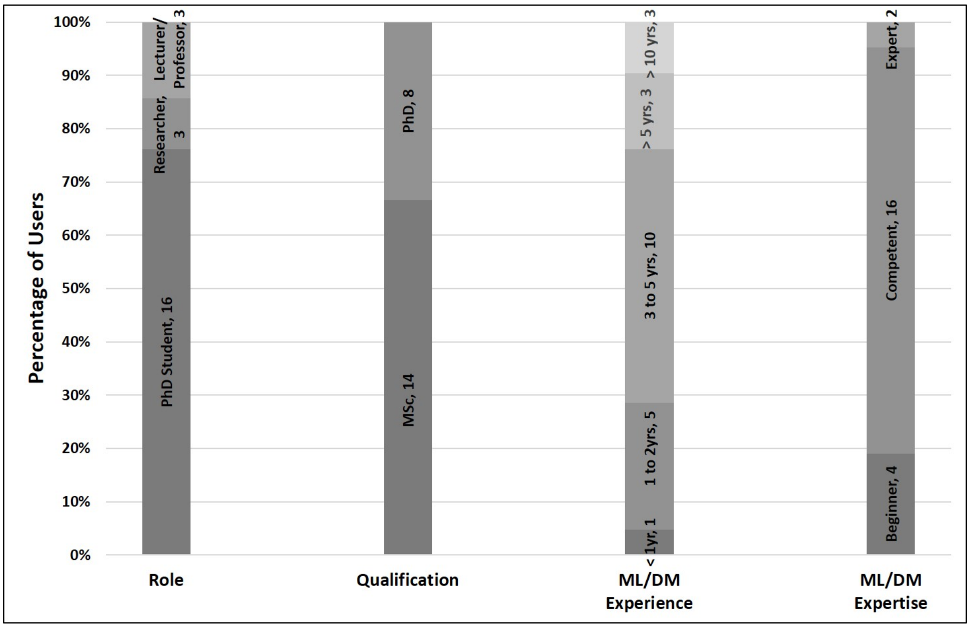


Figure 6.6: Profile of users

and BOW methods are shown respectively. The ratings range from 5 to 1, and the colours are from green for the highest rating of 5, to red for the lowest score, 1. In plotting the heat map, the average rating values per (q,r) were sorted in descending order.

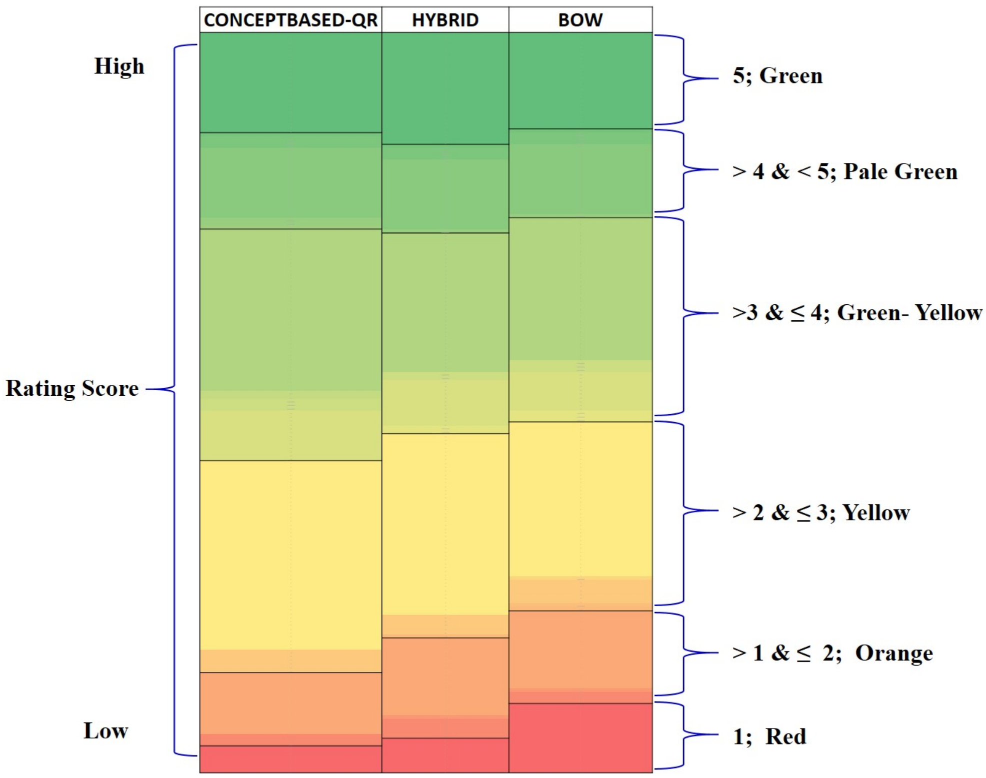


Figure 6.7: Spread of ratings for query-recommendation pairs evaluated

All the 3 heat maps for the methods are plotted in the same way using the actual average rating value given for each (q,r) pair. Lines are included on the heat maps to show a change in the value of the rating. The rating of 5 in the green segment at the top of the heat map means that all the users who evaluated each (q,r) pair in that segment, gave each (q,r) pair a score of 5, while the rating of 1 in the orange segment at the bottom of the heat map means that all the users who evaluated those (q,r) pairs in that segment gave them a rating of 1. In considering high ratings, HYBRID does best in producing documents with high quality ratings followed by CONCEPTBASED-QR and then BOW. So we find that HYBRID is able to correctly identify when to use either BOW or CONCEPTBASED-QR for refining a query in order to produce such good quality documents. For the documents with the lowest ratings, the standard BOW method produces the highest number of documents with very poor ratings. Although the standard BOW method is able to produce some good quality documents, BOW has difficulty preventing poor retrievals.

There is an interesting behaviour across the heat maps from top to bottom, beginning from the green to the red segment. One observation for the green and pale green segments is that HYBRID performs best followed by CONCEPTBASED-QR and then BOW in order to produce relevant documents that are given the highest ratings. There is a change in the green-yellow segment, and the CONCEPTBASED-QR method takes over and outperforms the other methods. In the green-yellow segment, HYBRID has the second best performance and then BOW. As the rating values decrease in the yellow segment, the CONCEPTBASED-QR method continues to outperform the other methods with a much bigger margin. This trend continues till the end of the heat map as seen in the red segment. However, it is noted that for the lower ratings, a good method should only have few or no documents with poor ratings. This is seen by the performance of the CONCEPTBASED-QR and HYBRID methods that have only few (q,r) pairs with low ratings.

Overall CONCEPTBASED-QR has the best performance by having the fewest number of documents with poor ratings. In particular CONCEPTBASED-QR is very good at not presenting poor retrievals to users. So, CONCEPTBASED-QR does well at presenting good retrievals, and it performs best in preventing poor retrievals from being shown. HYBRID also does well at showing high quality retrievals, but does not do as well as CONCEPTBASED-QR in preventing poor retrievals from being shown. BOW can be good, however when BOW gets retrievals wrong, it is really bad. So the performance of BOW is at extremes, where its good retrievals are very good and its bad retrievals are very bad. These results for all the marked areas show that users gave higher ratings to the recommendations made using the HYBRID and CONCEPTBASED-QR methods.

Figure 6.8 is a heat map with a broader view of the spread of ratings for the CONCEPTBASED-QR method. The heat map is sorted twice. First, based on the average rating per (q,r) pairs. Second based on the average user-rating computed using Equation 6.1. The rows contain the ratings for the (q,r) pairs, and the columns contain the ratings given to (q,r) pairs by each user. The column names contain demographic information, indicating if a user is an Expert (E), Competent user (C) or a Beginner (B). The (q,r) pairs with the highest ratings are at the top of the heat map, as shown by the green slots nearer the top. The (q,r) pairs with lower ratings have the red slots and appear near the bottom of the heat map. The queries evaluated had at least one user providing ratings for the (q,r) pairs. The average was computed for (q,r) pairs rated by more than one user.

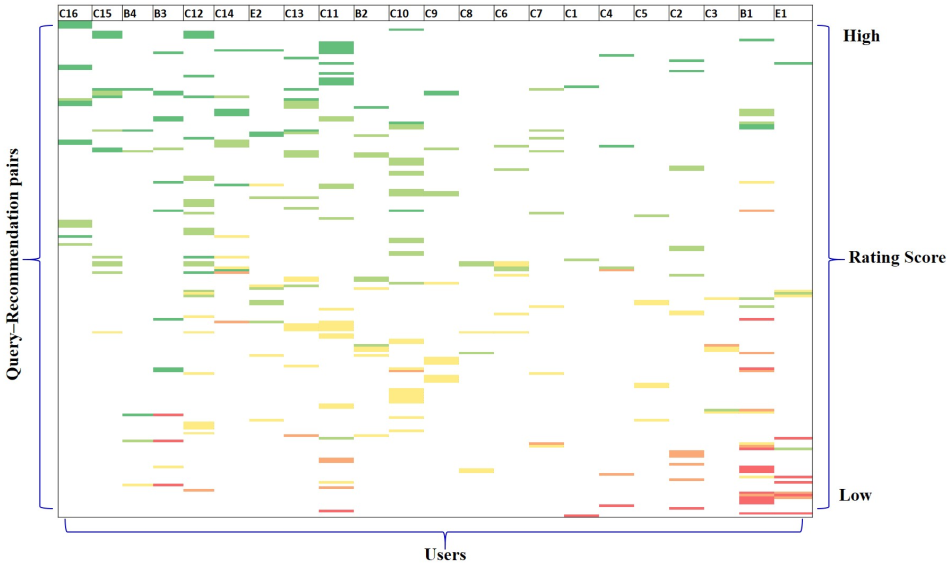


Figure 6.8: Spread of ratings for query-recommendation pairs for CONCEPTBASED-QR

A closer look at some of the (q,r) pairs rated by more than one user showed a good level of consensus for the ratings provided. For example, a query such as: "*How does cluster analysis work?*" received ratings from three competent users and one beginner. For the first (q,r) pair, their ratings were 5, 5, 5, and 4, with average ratings for the top 3 (q,r) pairs being 4.75, 4.5, and 4.25. Another query "*Is it possible to use reinforcement learning to solve any supervised or unsupervised problem?*", has consistently lower ratings from a beginner and a competent user who provide ratings for the (q,r) pairs for this query. The first (q,r) pair has ratings of 1, and 3, while the average on the top three documents are 2,3, and 2. In these two examples we find agreement with the ratings provided by users for the (q,r) pairs evaluated. We see a good level of consensus among the ratings provided for the CONCEPTBASED-QR method, so this allows us to make confident predictions from these results.

6.3.1 Examining Individual User-ratings

The rating for each user is examined to gain insight to the range of ratings given by users. For example if the user-rating for a particular user was 1.0, then such a user gave a rating of 1 to all the documents evaluated. On the other hand, if the user-rating was 5.0 then this means the user gave a rating of 5 to all the documents seen. For these 2 examples, it would mean the users did not really interact with the resources, but merely rated the documents using the extreme values throughout. Such results would be discarded to avoid judging a method wrongly. The user-rating is computed for each user based on Equation 6.1.

A summary of the queries evaluated by users is shown in Table 6.2. Column 1 contains the ID for users, column 2 contains the number of queries evaluated per user, and column 3 contains the number of query-recommendation pairs each user evaluated. The ID contains demographic information, such as Expert (E), Competent user (C) and Beginner (B). For example, C10 represents the 10th competent user, B1 represents the 1st beginner, and E1 represents the first expert. The table is sorted in descending order based on the number of (q,r) pairs that were evaluated.

ID	**# of queries**	**# of (q,r) pairs**
C10	10	55
C12	10	50
B1	10	47
C11	10	41
C13	7	33
C2	5	30
C16	5	26
C14	5	25
C15	5	24
B3	5	24
E1	4	22
C9	4	21
E2	4	21
B2	4	17
C6	3	14
C7	3	13
C4	2	11
C5	2	11
B4	2	10
C8	2	10
C3	2	10
C1	1	6

Table 6.2: Summary of queries evaluated per user

Four users evaluated up to 10 queries each, but the most active user provided ratings for 55 (q,r) pairs. Another user evaluated up to 7 queries. There were five users that evaluated 5 queries each, while four users evaluated 4 queries each. There were two users that evaluated 3 queries each, and five users that evaluated 2 queries each. The least active user evaluated 1 query and rated 6 (q,r) pairs. These results show a spread over the number of queries evaluated by all the users, because a single user did not provide ratings for all or majority of the queries. So issues of potential bias are prevented because the queries evaluated are a contribution from all the users.

Figure 6.9 shows the user-rating scores for all users. The x-axis contains a listing of all the users, while the y-axis contains the average user-rating scores. Looking within the graph from top to bottom, the broken line at the top represents the maximum rating each user gave based on all the (q,r) pairs that user evaluated. It is observed that all but five users had given a rating of 5 to at least one (q,r) pair. The next line is for the median, this is included to give an idea of the middle value of user-rating scores. Below the median is the average user rating for each user. The entire graph is sorted in ascending order based on the average user rating. The final line represents the minimum rating scores given by each user for all the (q,r) pairs they evaluated. From this graph, we see that 12 users had given a minimum rating of 1 to at least one (q,r) pair. These scores show variety in the ratings given by users, so we can trust the judgements made using these ratings.

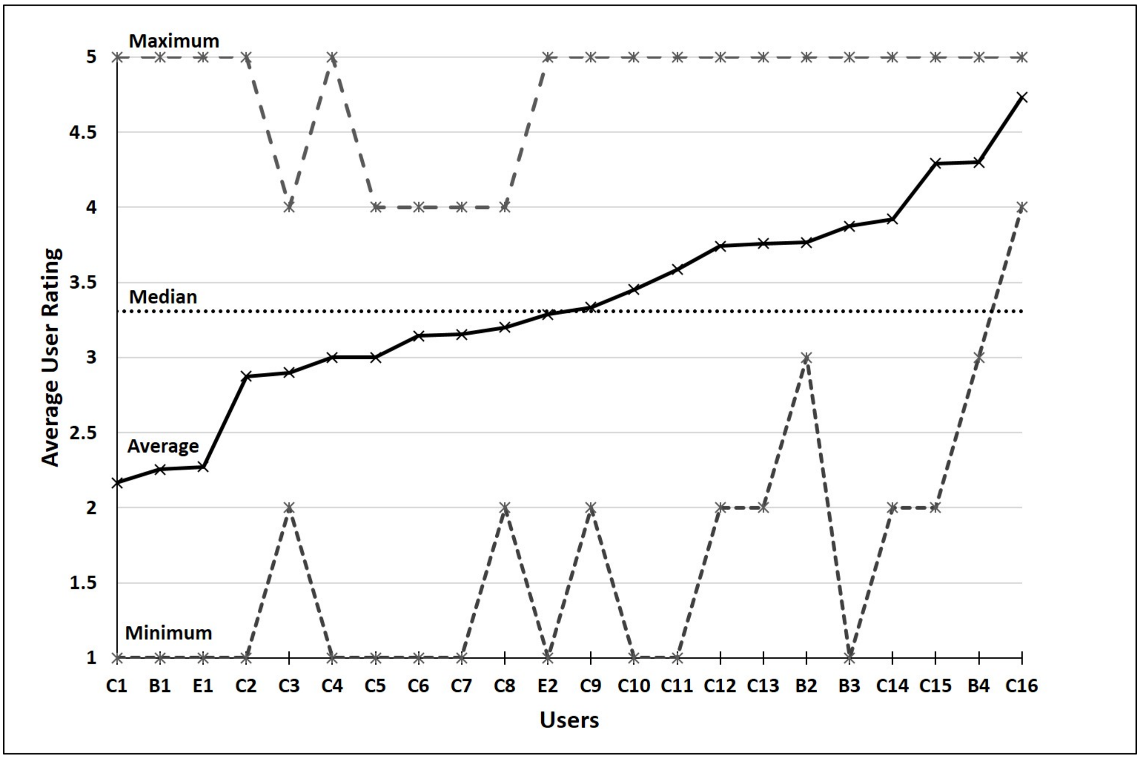


Figure 6.9: Statistics of user rating scores

The average user-rating values shown in the graph are sorted from minimum to maximum, from a total score of 5. The minimum average user rating was 2.17 from a competent user who evaluated only 1 query, while the maximum average user rating was 4.73 from a competent user who evaluated 5 queries. The average user-rating scores range from 2.17 to 4.73. None of the user-rating scores are at the extremes of 1.0 or 5.0 which means no user rated every single document evaluated with the same extreme rating scores of 1 or 5. Further, the users who have the minimum and maximum scores rated only 1 and 5 queries respectively, so there is no bias introduced from the ratings provided by these two users, as majority of the queries evaluated are from users within the middle scores as shown in Figure 6.9. It is observed that there are big differences between the user-rating for all users. This range of user-rating scores highlights the benefit of designing the evaluation so that the same user evaluates both methods at the same time. Thus allowing us to avoid any potential bias of having generally positive users providing ratings for one method, and generally negative users providing ratings for another method.

The median user-rating for this set of users was 3.31, while the average user-rating was 3.36. Most of the user-rating scores are within the median, which shows that the users actually interacted with the documents they were rating. The minimum and maximum values of the average scores are based on only 6 queries evaluated, so we do not need to exclude these values. The minimum average score came from a competent user who evaluated only one query, while the maximum score came from another competent user who evaluated 5 queries. These statistical measures are an indication that individual users had some variation in the ratings for (q,r) pairs they evaluated. These scores give us confidence to use the results generated to make meaningful judgements about the methods being evaluated.

6.3.2 Analysing Results Based on Demographics

The results are analysed based on a user's expertise, to determine if this affects the ratings they provided. The rating for each (q,r) pair is computed using Equation 6.2. Table 6.3 captures the average ratings of all users and the ratings based on expertise of users for each method. This allows us to confirm if there is any agreement among the users irrespective of their expertise.

Table 6.3: Average rating

Method	All Users	Experts	Competent	Beginners
CONCEPTBASED-QR	3.54	3.13	3.66	3.29
HYBRID	3.45	2.71	3.58	3.27
BOW	3.33	2.58	3.46	3.25

For all users, we found that ConceptBased-QR > Hybrid > Bow. The ConceptBased-QR method has the best performance. So, there is benefit when a query that has been refined using domain concepts is employed to search for relevant learning materials. For experts, the average rating scores for all methods are lower, however the experts still agree that the best performance is from the ConceptBased-QR method. We are confident in the results received from experts because they know what topics learners should be interested in. The competent users have higher rating scores across all methods, and they also agree that the ConceptBased-QR method performs better. Although the ratings by the beginners for all the methods are very similar, their rating scores also agree with the other users that the ConceptBased-QR method performs best.

A bigger difference is seen in the ratings provided by the experts, perhaps they tend to be more discriminating about their relevance judgement. It is noted that for documents that are relevant to the query, the experts give high ratings but documents that are not so relevant to the query are given much lower ratings. It is observed that the higher up the level of expertise, the bigger the difference between ConceptBased-QR and Bow. Experts have a stronger preference for the ConceptBased-QR method, followed by the competent users and the beginners.

6.3.3 Average Rating by Position

One would expect the topmost recommendation made to be the most relevant document for a learner's query. In this section, we compare the rating at positions 1 to 3 for the three methods. Table 6.4 contains the results for the average rating at 1, 2, and 3. Rating@1 is the average rating for the topmost (q,r) pair, while Rating@2 is the average rating of the top 2 (q,r) pairs, and Rating@3 is the average rating of the top 3 (q,r) pairs.

Table 6.4: Average rating @ N

Method	**Rating@1**	**Rating@2**	**Rating@3**
ConceptBased-QR	3.53	3.58	3.54
Hybrid	3.45	3.48	3.45
Bow	3.48	3.40	3.33

The results in Table 6.4 show that the ConceptBased-QR and Hybrid methods have slightly higher ratings for the Rating@2 compared to Rating@1. This means that users rated documents at position 2 higher than documents at position 1. This behaviour for the ConceptBased-QR method may be because ConceptBased-QR is designed to be effective for vague queries but in this scenario, the performance of ConceptBased-QR is examined on a combination of queries that are vague as well as those queries that are well-formed. Another reason for this behaviour may be that there can be more than one document that is suitable for a given query. When this happens, users may rate such documents equally or give one a lower rating. The results of HYBRID are a contribution from the results of CONCEPTBASED-QR, hence we see a similar behaviour from the HYBRID method to the behaviour shown by the CONCEPTBASED-QR method in Table 6.4. In the BOW method, the Rating@1 $>$ Rating@2 $>$ Rating@3. We find a decrease in the average ratings as the value of N increases. The BOW method is able to rank documents by finding the most relevant document first.

The CONCEPTBASED-QR method may have less variation in the type of documents it retrieves. However, CONCEPTBASED-QR provides better quality materials than BOW at all the first 3 positions, as the quality of documents retrieved by BOW is less than that of CONCEPTBASED-QR. The CONCEPTBASED-QR method has a good range of relevant documents. BOW does not have such a good range of relevant retrievals, as there is a larger difference between the quality of results found in BOW for the top 3 positions, hence the large variation in the BOW scores.

6.3.4 Preference of Methods

Relevance judgement is subjective and depends on the users who are providing ratings for the documents seen. We wanted to know how many users preferred the recommendations produced from the CONCEPTBASED-QR method, the HYBRID method or the standard BOW method. We use the rating as given in Equation 6.2, and we count how many users rated documents from one method higher than the other. Recall that the users here do not represent the typical learners in a user evaluation. These users are knowledgeable in the ML/DM domain, and their task is to provide relevance judgements for recommendations made by the different methods.

Table 6.5 contains the results for the preference users had for either the CONCEPTBASED-QR method, denoted by CB or the BOW method. Half of the experts preferred CONCEPTBASED-QR over BOW, while the other half thought both methods were the same. There were 14 users that preferred the recommendations produced using the CONCEPTBASED-QR method over those of the BOW method. From these 14 users, there was 1 expert, 12 competent users, and 1 beginner. We can trust the judgement of experts and competent users because they are more knowledgeable about the domain, and they know what documents should be relevant to learners.

Four users preferred the recommendations produced by BOW over the CONCEPTBASED-QR method. However, we note that none of the experts preferred the standard BOW method over

Table 6.5: Preferences of methods: CONCEPTBASED-QR (CB) vs BOW

Preference	Users	Demographics
CB > BOW	14	1 expert, 12 competent, 1 beginner
BOW > CB	4	1 competent, 3 beginners
CB = BOW	4	1 expert, 3 competent

CONCEPTBASED-QR. Recall from Table 6.3, that the scores provided by the beginners for all methods were quite similar. This is because beginners are not very reliable at deciding whether materials are relevant or not. So we cannot rely totally on the judgements provided by beginners. There were 4 users who rated documents seen from both methods equally, as CB = BOW, one of these being an expert and 3 being competent users. Such a rating could have come from users that gave equal scores to all documents seen for a query. Overall, we find that majority of competent users preferred the recommendations produced by the CONCEPTBASED-QR method over BOW.

Table 6.6 shows the results for the preference users had for either CONCEPTBASED-QR or HYBRID. There were 13 users that preferred recommendations made by CONCEPTBASED-QR over that of the HYBRID method. Given that HYBRID is either CONCEPTBASED-QR or BOW. This would mean all the times when a user preferred a CONCEPTBASED-QR recommendation over a BOW recommendation. The 13 users who preferred CONCEPTBASED-QR over HYBRID consist of 2 experts, 10 competent users, and 1 beginner. All but 1 of the users are competent or expert, so we can rely on their judgement. There are 3 competent users and 3 beginners that prefer HYBRID over CONCEPTBASED-QR, meaning such users preferred the BOW recommendations over the CONCEPTBASED-QR recommendations. There were 3 competent users that provided equal ratings for CONCEPTBASED-QR and HYBRID. This means that these 3 users provided the same scores for recommendations by CONCEPTBASED-QR and BOW for a query.

Table 6.6: Preferences of methods: CONCEPTBASED-QR (CB) vs HYBRID

Preference	Users	Demographics
CB > HYBRID	13	2 experts, 10 competent users, 1 beginner
HYBRID > CB	6	3 competent users, 3 beginners
CB = HYBRID	3	3 competent users

The results for the preference users had for either HYBRID or BOW are shown in Table 6.7. For the comparison between HYBRID and BOW, 12 users prefer HYBRID over BOW. These users consist of 1 expert, 10 competent users, and 1 beginner. Since HYBRID is generated from either CONCEPTBASED-QR or BOW, this would be when the recommendations produced by CONCEPTBASED-QR were preferred over those produced by BOW. We find that majority of competent users prefer HYBRID over the standard BOW method.

Table 6.7: Preferences of methods: HYBRID vs BOW

Preference	Users	Demographics
HYBRID > BOW	12	1 expert, 10 competent users, 1 beginner
BOW > HYBRID	5	1 expert, 2 competent users, 2 beginners
HYBRID = BOW	5	4 competent users, 1 beginner

For BOW > HYBRID, 1 expert, 2 competent users, and 2 beginners prefer the BOW method over HYBRID. This would be a situation where the BOW recommendation was better than the CONCEPTBASED-QR recommendation. Such a result can be for a query where BOW was really effective at retrieving documents. We recall from Figure 6.7 that BOW did well at getting some good retrievals, while CONCEPTBASED-QR was best at not recommending bad retrievals. Further, experts rated very good retrievals highly as seen from their scores in Table 6.3. The experts were best at differentiating the ratings given to the documents evaluated. There are 5 users who rated HYBRID and CONCEPTBASED-QR equally, 4 of these are competent users and 1 a beginner. Perhaps the HYBRID method used the CONCEPTBASED-QR method for these queries, hence there is no difference in their performance.

Generally, for the preference of methods, it is observed that majority of competent users prefer the CONCEPTBASED-QR and HYBRID methods over the standard BOW method. We can trust the results from these users given their level of expertise in the ML/DM domain. The competent and expert users know what the learners should want to learn about for a particular query, so we can rely on their relevance judgements.

6.4 Qualitative feedback

Having recommendations with high ratings is good, but recommending documents that cover topics relevant to the query is important in e-Learning recommendation. Ideally, it would be good to identify the coverage of relevant topics contained in the documents recommended by each method. However, our evaluation design required us to randomize the documents shown to users, so documents from all methods are shown together. The coverage of relevant topics in documents are examined to identify two key things. First, the results would allow us to have a measure of the degree of coverage of documents in our collection. Second, the results would inform us about the level of complexity of queries. We capture the coverage by asking users to provide additional feedback after they evaluate each query. This feedback was optional for the users. Figure 6.10

shows the question that was displayed to users after they had evaluated the recommendations for a query. A text box was also provided to capture additional feedback from the users. In this section we present results from the optional comments users provided after evaluating a query and users' responses to the coverage of documents shown.

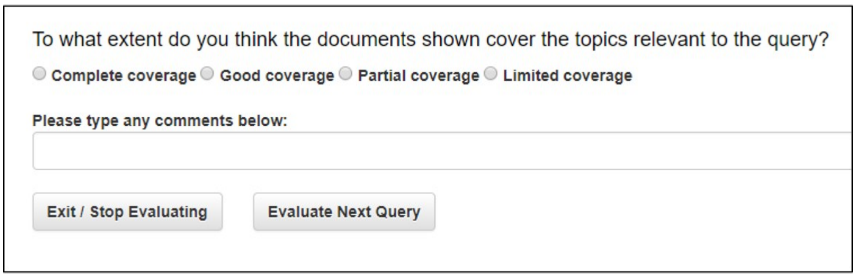


Figure 6.10: Subjective feedback from users

6.4.1 Optional Comments on Recommendations

Users were given the opportunity to provide additional feedback on the coverage of the documents they were shown for each (q,r) pair they had evaluated. Table 6.8 shows the comments received from users. Some contextual information is given to gain more insight from the optional comments provided by users on the coverage of relevant topics contained in the documents shown to users. The first column of Table 6.8 contains the demographics of the users, while the second column contains the coverage scores the users provided, and the third column contains the comments. The comments are sorted based on the coverage column, from complete to partial coverage.

There were responses from 15 users, 8 of the comments had good coverage associated with the documents retrieved, an additional 3 comments were from documents that were given complete coverage. Only 4 comments were associated with partial coverage. There was no comment associated with limited coverage. For the expertise of users who provided comments, all but one of the comments came from competent users. One expert also provided a comment. As all the users were competent or expert, there is a high level of confidence associated with their feedback.

Generally the comments provided were positive. For example, a competent user who evaluated the coverage as complete said: "*Very relevant recommended text most of which I downloaded for personal consumption*". For comments given good coverage ratings, the comments provided by the expert user shows that just one of the top 3 documents retrieved provided good coverage of the topics relevant to the particular query the user evaluated. This could have resulted in the partial coverage score given by the expert user to the documents seen. For the partial coverage

Table 6.8: Comments from users

Expertise	Coverage	Comments
Competent	Complete	Very good retrieval set, Excellent coverage, Relevant topics contained
Competent	Complete	Very relevant recommended text most of which I downloaded for personal consumption
Competent	Complete	A couple of recommendations cover subtopics such as web mining and graph mining, but the rest have sufficient info
Competent	Good	Good coverage of relevant topics
Competent	Good	The top rated documents have good coverage of the query
Competent	Good	Deep learning was covered
Competent	Good	Fair coverage in 4 of the 6 documents
Competent	Good	Overall good coverage of topics
Competent	Good	Relevant topics in documents shown
Competent	Good	Overall a good spread of topics relevant to the query
Expert	Good	Only 1 of the document (D284) addressed the query as I understood it but this document provided good coverage of the topic
Competent	Partial	One document, D389 had relevant topics
Competent	Partial	I expected one article explaining the difference between them
Competent	Partial	Fair coverage overall, 3 of the 6 documents appear to cover relevant topics
Competent	Partial	Some partial coverage of topics

comments, one of the competent users said half of the documents seen appeared to contain relevant topics. Another competent user evaluated a query asking about the difference between two topics, although the user gave partial coverage to the documents, the competent user gave rating scores of 4 and 5 to all the documents recommended for this query. So the documents retrieved were relevant to the query, even though the coverage scores given by the user was partial coverage.

It is observed that any unmet expectation users have based on a query they evaluate can affect the coverage scores they give. This is particularly true of the users who provided comments, given that they are competent and expert users with more knowledge of the learning domain. Such users can be expecting documents that challenge them. However, the individual ratings provided for each (q,r) pair helps to provide useful information about the relevance of the recommendation for the query being evaluated. The comments received from the users was about the retrieval set shown to them, and not about the usability of the evaluation system. However, a user informed the researcher about how easy the evaluation system was to use, because the user did not experience any difficulty navigating through the evaluation system.

6.4.2 Coverage of Relevant Topics

Users were asked to what extent the documents they evaluated covered topics relevant to the query. For this question, a user could make one selection from 4 options: complete, good, partial and limited coverage. Figure 6.11 captures results for the coverage of topics relevant to the query. The x-axis shows the type of coverage, while the y-axis shows the percentage of entries that were assigned to the respective coverage type. There were 50% of entries from users which indicated that the documents had good coverage. An additional 19% of entries said the documents had complete coverage of relevant topics, while 21% of entries said the documents had partial coverage. Only 10% of entries said the documents had limited coverage of relevant topics. So, most of the documents recommended covered topics that were relevant to the queries.

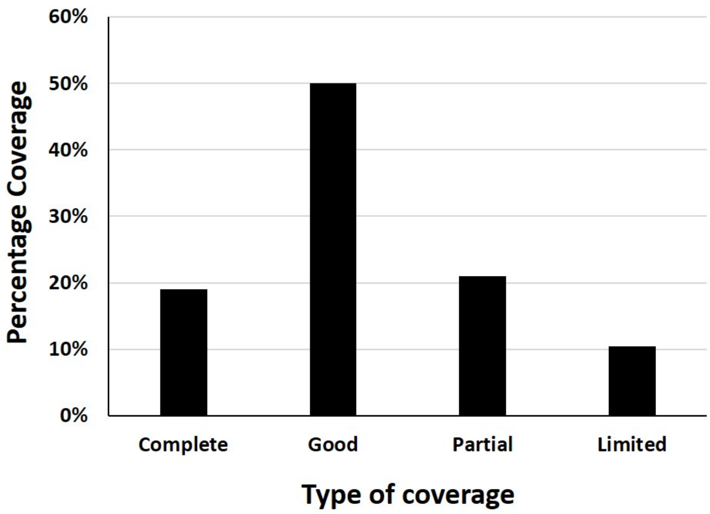


Figure 6.11: Coverage of topics relevant to the query

Figure 6.12 is a heat map of the coverage scores for the queries that were evaluated. The heat map is plotted by converting the coverage scores to numeric values where complete is 4, good is 3, partial is 2, and limited is 1. The colours for the coverage scores are green for complete coverage, pale green for good coverage, yellow-orange for partial coverage, and red for limited coverage.

The extreme values for coverage are 4 and 1, represented by the green and red segments respectively. The green segment for complete coverage means that all the users who provided coverage scores for those particular queries gave it complete coverage. While, the red segment means that all users who evaluated those queries, gave them limited coverage.

Figure 6.13 contains a heat map which shows a broader view of the spread of coverage scores for the queries evaluated. The heat map is sorted twice. First based on the average coverage scores

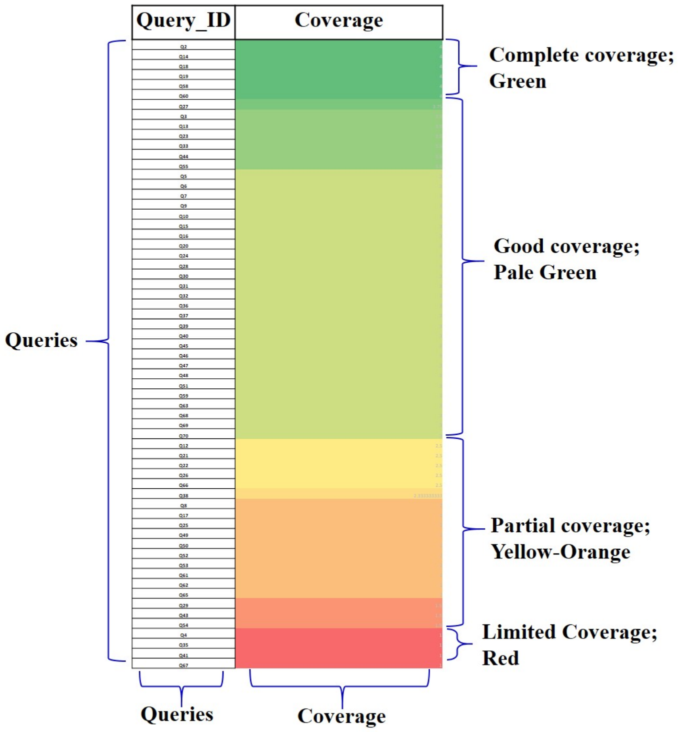


Figure 6.12: Coverage scores per query

per query and second based on the average coverage scores per user. Figure 6.13 captures the queries with the best coverage at the top of the diagram as seen by the green slots nearer the top left, and those with least coverage as seen by the red slots nearer the bottom left of the diagram.

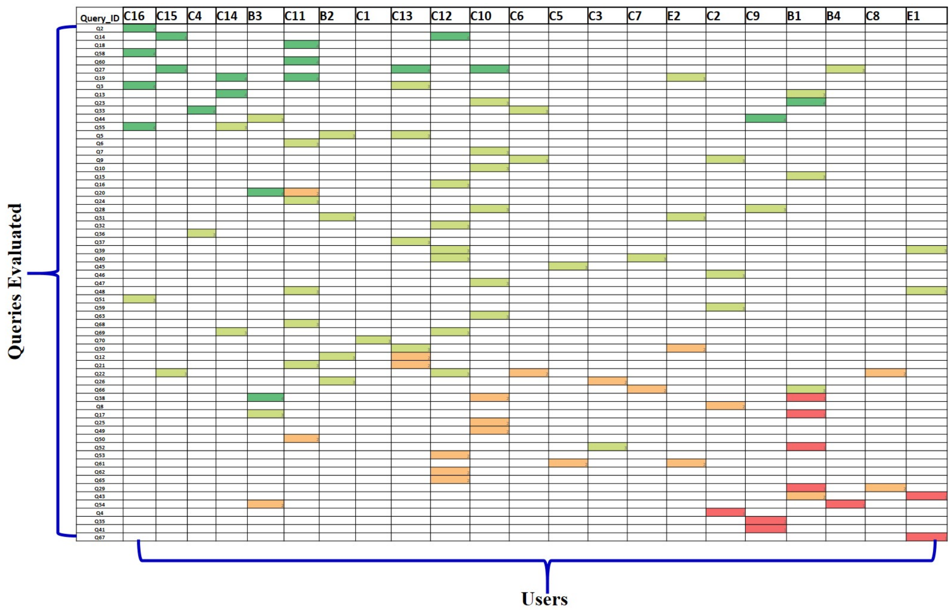


Figure 6.13: Spread of scores for the coverage

The queries rated had consistent coverage scores from more than one user, so this allowed us to gain some more insight to the queries. For example, a query such as: *"How does cluster analysis work?"* had complete coverage scores from three users and a good coverage score from one user. While the query: *"Is it possible to use reinforcement learning to solve any supervised or unsupervised problem?"* had partial coverage from one user and limited coverage from another user. We draw the following insights from the coverage scores. First, some queries may not be well written and so can be hard to understand. Second, the topic contained in some queries may be very specialized, thus making it more challenging to find documents that are relevant to such queries. These results show potential in exploring the features of queries when refining them for e-Learning recommendation.

6.4.3 Coverage vs Rating

In this section we investigate whether a query with a high rating score would also have a high coverage score, while queries with low ratings would be associated with low coverage scores. The coverage scores provided are compared with the ratings users provided for queries. An assumption is that queries that are easy to understand should be easier for retrieving documents that have good coverage of relevant topics, given that the documents are contained in the collection.

To enable us make a comparison between coverage and ratings, we take the average of the rating score for the top 3 query-recommendation pairs for each query, to generate a single rating score for a query. This is because the coverage scores are based on individual queries. Figure 6.14 is a heat map which contains a comparison of the coverage scores given to queries with rating scores. The first column contains the coverage scores for the queries. The following columns contain the rating scores for the CONCEPTBASED-QR, HYBRID and BOW methods respectively.

In plotting the heat map, the coverage scores were sorted in descending order. Each row represents the entry for a single query. The columns for the 3 methods are sorted based on the query IDs, so each row captures the same query across the entire row, for the coverage score and for each of the 3 methods respectively. This allows us to gain insights from specific queries. Figure 6.14 highlights the cases where hybrid gets it right or wrong, by deciding to use either BOW or CONCEPTBASED-QR to refine a query. This is because HYBRID takes on either the colour of BOW or CONCEPTBASED-QR, thus allowing us to see where HYBRID has been generated from. The heat map also allows us to see when HYBRID is better than BOW or CONCEPTBASED-QR.

At the top of the heat map in the green segment, it is observed that majority of the queries with

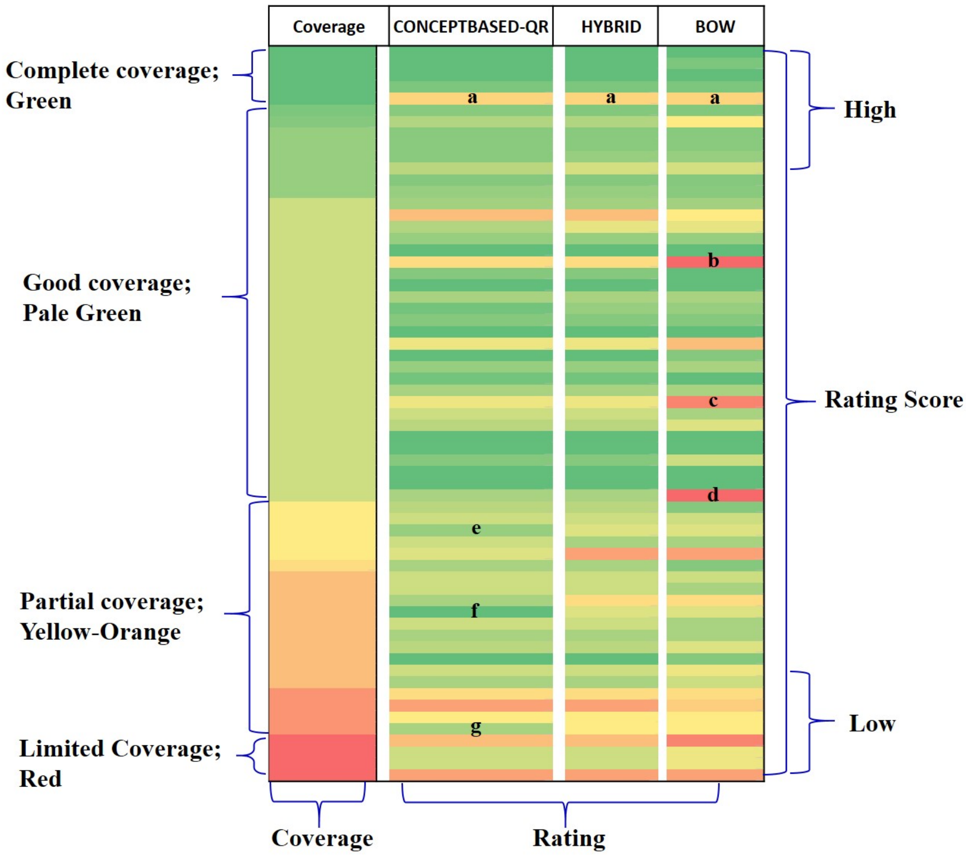


Figure 6.14: Visualisation of the coverage scores per query for each method

high coverage also have high rating scores for all the methods. This result is as one would expect. Where queries that are easy to understand are also easy to find good documents for retrieval, and such documents are contained in the collection. Hence the documents retrieved for such queries are rated highly by users. It is observed on the point marked **a**, that all the methods have the same rating score of 3 for the retrieval produced by that query. Although the rating score given is 3, the users reported that the documents shown for this query had complete coverage of topics relevant to the query. So this gives an indication that users were satisfied with the documents that were recommended for this query.

The standard BOW approach has less consistent performance overall when coverage and ratings are compared. BOW has these queries with bad retrievals, such that all 3 (q, r) pairs have low ratings, hence the full red colour on some of its queries as seen on the points marked **b**, **c**, and **d**. The ratings for these queries are low even though the queries are in the pale green segment which is for queries with good coverage. It is observed that on such queries, CONCEPTBASED-QR is not red, perhaps the documents produced by CONCEPTBASED-QR influenced the coverage score for these queries.

At the bottom of the heat map in the yellow-orange and red segments, notice that queries which have low ratings also tend to have low coverage. Such queries may either be difficult to understand or lacking in relevant documents in the collection. However, at the points marked **e**, **f** and **g**, it is observed that CONCEPTBASED-QR produces better documents with higher ratings compared to HYBRID and BOW. The CONCEPTBASED-QR method takes advantage of its extra domain knowledge to refine queries that are complex and still find relevant documents for such queries. So, CONCEPTBASED-QR is able to perform well on queries that have low coverage scores. These low coverage scores can be a contribution from the documents produced by BOW and HYBRID. Overall, it is seen that queries with high ratings generally have a high coverage score while queries with low ratings tend to have low coverage scores for the CONCEPTBASED-QR and HYBRID methods.

6.5 Summary

A user evaluation of the HYBRID, CONCEPTBASED-QR, and BOW query refinement methods has been presented in this chapter. The evaluation is not a typical user trial, but instead it is relevance judgement which employed knowledgeable users. The design of the user evaluation, the evaluation metrics used and the results have also been discussed. The e-Learning recommender system developed in Chapter 5 was employed for the user evaluation task. The evaluation performed was two-fold. First an evaluation of the relevance of recommendations made by the three methods. Second, an evaluation of the coverage of relevant topics across the documents rated.

A collection of queries and a dataset of Machine Learning and Data Mining documents were used for the evaluation. The evaluation system was hosted online for 8 weeks. There were 22 users who evaluated 105 queries and provided ratings for 521 query-recommendation (q,r) pairs. Users completed a questionnaire at the start of the evaluation, which provided data about their expertise and experience in the Machine Learning and Data Mining domain. All the users had at least an MSc degree. The questionnaire data allowed us to gain useful insights from our results.

The evaluation was designed to allow users to provide unbiased judgements on the methods. The order of documents shown to users were randomized to prevent the bias of earlier documents being regarded as relevant over those shown further down the list. The evaluation design also prevented a potential bias to a method because each user evaluated the (q,r) pairs for a query for all methods without knowing which method was being evaluated. The nature of the HYBRID method meant its rating scores could be generated from the ratings of either the CONCEPTBASED-QR or BOW query refinement methods.

Results for the relevance of recommendations showed that HYBRID did best in producing high quality documents. It was observed that the CONCEPTBASED-QR method is particularly good at preventing documents with potentially low ratings from being retrieved. However, the standard BOW method struggled to prevent documents with low ratings from being retrieved, so BOW had the highest number of poor retrievals. Overall, the CONCEPTBASED-QR method had the best performance by producing many good retrievals, and the fewest poor retrievals. There was a good level of consensus on the judgements provided for each method by users with different levels of expertise. Results from experts, competent users and beginners all showed that using queries refined using the CONCEPTBASED-QR and HYBRID methods to search produced documents that were consistently more relevant to learners than when the standard BOW method was used.

Evaluation results for the coverage of relevant topics across the documents evaluated showed that most of the documents recommended covered topics that were relevant to the query. There were 50% of entries which stated that documents had good coverage of topics for the query. In addition, 19% of entries stated that documents had complete coverage, while 21% had partial coverage and only 10% had limited coverage.

A close examination of some queries that had low ratings as well as low coverage scores, revealed that some learner queries can be difficult to understand as they may not be well written. So this causes a challenge for query refinement. One way of addressing this challenge can be by exploring more query features when designing a HYBRID method. HYBRID uses the features of a query to make a dynamic choice in determining which method to use for refining a query. The results show HYBRID to perform better than the standard approach, thus highlighting the advantage in exploiting query features for determining a suitable query refinement approach to adopt for a query.

A comparison of the relevance and coverage scores generally showed that (q,r) pairs that had high rating scores associated with them also had good coverage scores. User evaluation results demonstrate the effectiveness of using the CONCEPTBASED-QR and HYBRID methods for identifying relevant learning concepts that are employed in refining queries, to help learners find relevant learning materials.

Chapter 7

Conclusions and Future Work

The research presented in this thesis investigates knowledge driven approaches for supporting e-Learning recommendation to enable learners find relevant learning materials. This chapter discusses the contributions made and how this research can be applied to other domains. The achievement of the research objectives are presented and some ways that this research can be taken forward are discussed. The chapter ends with a reminder, highlighting the key insights of this research.

7.1 Contributions

This research has developed techniques that support e-Learning recommendation by helping learners to ask effective queries and find relevant documents from the large amounts of learning materials that are currently available. Two key issues that make e-Learning recommendation challenging were identified. The first issue was the need to provide a shared vocabulary for teaching experts and learners, in order to support the representation of learning materials and enable the materials to be more accessible during recommendation. Tackling this issue enabled us to address the semantic gap that learners face. The second issue was the need to help learners to identify relevant learning topics and craft effective queries when trying to find relevant learning materials. Solving this issue enabled us to address the intent gap faced by learners.

A key contribution of this research is the creation of background knowledge which contains important information that can be employed for general understanding and problem-solving in a given domain. The background knowledge creation process generates a set of domain concepts containing concept labels, and their respective concept descriptions. The background knowledge captures important domain concepts as highlighted by teaching experts, thus providing a shared vocabulary for teaching experts and learners.

The domain concepts are used to underpin the representation of learning materials. Using learning concepts for the representation of learning materials allows the retrieval to focus on documents that contain relevant concepts. The method for representing documents using domain concepts was presented in §4.1. In this approach, the domain concepts are used to underpin the similarity between documents. The evaluation results show that employing domain concepts to represent learning materials supports e-Learning recommendation by enabling relevant materials to be more accessible for retrieval.

The concept vocabulary is also employed in the development of suitable methods for the refinement of learners' queries. By using domain concepts, we are able to help learners to identify relevant learning concepts and use the concept vocabulary to inject knowledge of intent when refining learners' queries. The query refinement method that uses domain concepts is presented in §5.2. The result of applying this method is the creation of effective queries that can be used to find and retrieve relevant materials for learners. A HYBRID query refinement method is developed to cater for queries that can be specific or generic. The HYBRID method automatically determines which method to use for refining a query based on the features of the query. This method is presented in §5.5. The evaluation results show that harnessing the knowledge of domain concepts for refining queries helps learners to seek relevant documents.

The learning domain used to test the methods developed in this research has been Machine Learning and Data Mining, an area in which the author is knowledgeable. However, the methods can easily be applied for e-Learning recommendation in other domains. Given a collection of learning materials, background knowledge for the new domain would be created using data sources such as: TOCs of eBooks for generating concepts; a domain lexicon, for verifying concepts; and an Encyclopedia source, such as Wikipedia and DBpedia for generating concept descriptions. The new background knowledge will then be embedded in the e-Learning recommender system in §5.1, and used for recommendation in the new domain.

7.2 Achievement of the Research Objectives

This section will discuss the extent to which each of the research objectives presented in Section 1.2, has been achieved.

Identify the challenges within e-Learning recommendation by performing a critical review of research in recommendation with a focus on issues in e-Learning recommendation.

The semantic gap and intent gap faced by learners have been identified as two main issues that make e-Learning recommendation challenging. The semantic gap presents itself through the lack of a shared vocabulary between learners and domain experts. While the intent gap happens because learners lack sufficient knowledge about what topics are suitable for them when searching for relevant learning materials.

Standard recommendation approaches should be implemented differently when considering e-Learning recommendation. Typically, collaborative filtering systems rely on the preferences of users with similar interests for making predictions. However, in e-Learning recommendation, the need of a learner captured through a query would have to be considered for suitable recommendations to be made. In content-based systems, users are often recommended items that are similar to those previously consumed. For e-Learning systems, you do not wish to recommend more of the same learning materials that a learner has read. Instead, the recommendation should be based on the learner's current query.

The item-user pair in recommender systems can be mapped to the learning resource-learner pair in e-Learning recommenders. However, finding relevant materials is challenging due to the complex features that are often used to describe learning materials and learners. Learning resources are largely text which presents challenges of dealing with unstructured data and indexing the learning resources for retrieval. Further, the vocabulary used in the resources is often different from that used by learners, thus making it difficult to find and retrieve relevant learning resources for learners. A key feature that is important for learners is the learning goal, which is often captured through a query. This query should be taken into account when making recommendations because it is supposed to capture what a learner wishes to learn. However, learners often find it difficult to craft an effective query because they lack sufficient knowledge of the domain. This research explores approaches that support e-Learning recommendation to enable learners find relevant learning resources.

Address the semantic gap by exploring how to provide a shared vocabulary between domain experts and learners in order to enable learners find relevant materials

The semantic gap is addressed by introducing a novel method that automatically creates background knowledge in the form of a set of rich learning concepts related to the selected learning domain. The identified concepts provide a vocabulary and focus that is based on teaching materials with provenance. The concepts in the background knowledge represent important topics that learners should be interested in. The background knowledge is employed to influence retrieval in the recommendation of new learning materials by leveraging the vocabulary associated with the concepts during the representation process.

A CONCEPTBASED document representation approach employs the concept vocabulary only for representing documents. However the initial CONCEPTBASED document representation approach had a limited number of concepts, so its vocabulary was too restricted for concept-based distinctiveness to be effective. An augmented document representation approach leverages a vocabulary from both concepts and documents for representing documents. The augmented approach exploits differences between distributions of document terms in the concept and document spaces, in order to boost the influence of terms that are distinctive in a few concepts. The vocabulary from both concepts and documents is focused using the vocabulary from the concept space. Evaluation results show that augmenting the representation of learning resources with the concepts addresses the semantic gap by providing a shared vocabulary between learners and experts. This work won the Donald Michie Memorial Award for the Best Technical Paper at the BCS AI International Conference (Mbipom et al. 2016).

The background knowledge is enhanced by refining the method used to generate the domain concepts. The output is a richer set of domain concepts which is used to develop the enhanced CONCEPTBASED* document representation method. The richer concept vocabulary in CONCEPTBASED* provides a better coverage of the domain when employed in the representation and retrieval of documents. The benefit of the enhanced background knowledge is evaluated using a collection of Machine Learning and Data Mining documents. Our approach outperforms benchmark methods, demonstrating the advantage of using background knowledge for representing learning materials which enables learners to find relevant materials during e-Learning recommendation. This work has been published in a Special Issue of the Expert Systems Journal (Mbipom, Craw & Massie 2018).

Address the intent gap by exploring effective methods to help learners identify relevant topics in order to support learners to ask useful queries when finding learning materials

The development of an approach that refines learners' queries by identifying important learning topics is a key contribution that addresses the intent gap. Query refinement is often done explicitly, where a learner has to choose which topics are relevant. Such approaches are difficult because learners do not usually know what topics are relevant. The approach in this research is done implicitly so that we do not rely on learners, who often have insufficient knowledge about what they are looking for. A knowledge-rich representation containing important learning topics has been generated from learning materials in the form of concepts in our background knowledge. The approach employs background knowledge by leveraging concepts that are similar to queries and distinctive concept terms for refining learners' queries. This allows the search using a refined query to focus on topics that should be relevant for a given learner's query. So, the refined queries enable learners to ask effective queries and find relevant learning materials.

A recommender system is built to demonstrate the recommendation of learning materials. The developed system allows us to evaluate the effectiveness of the query refinement approach. A collection of queries and a dataset of Machine Learning and Data Mining documents are used for evaluation. The evaluation is not a standard user trial with learners, because the users had to be knowledgeable in the chosen domain, to be suitably qualified to give relevance judgements. Relevance judgement is subjective because it depends on the opinions of the users taking part in the evaluation. However, we had a good level of consensus on the relevance judgements provided by users with different levels of expertise. The results from experts, competent users and beginners all showed that using learning concepts to refine queries achieved effective queries. The search using our refined queries produced documents that were consistently more relevant than when the standard method was used.

An investigation of the coverage of relevant topics across the query-recommendation pairs showed that most of the recommendations covered topics that were relevant to the query. A comparison of the relevance and coverage scores generally showed that documents that had high rating scores associated with them also had good coverage scores. The evaluation results demonstrate the effectiveness of our approach to support e-Learning recommendation, by recommending relevant learning materials that contain a good coverage of topics that are relevant to the queries evaluated. This work was presented at the EAAI symposium at AAAI (Mbipom, Massie & Craw 2018).

7.3 Future Work

The work in this thesis focused on developing approaches to address the semantic and intent gaps in e-Learning recommendation. The semantic gap is addressed by exploiting background knowledge to provide a shared vocabulary between learners and domain experts. In addressing the intent gap, a method was developed to help learners identify relevant topics and ask effective queries when searching for relevant learning materials. This section discusses some limitations of this research and presents some ways that the work can be taken forward.

7.3.1 Incorporating Structure into the Creation of Background Knowledge

The background knowledge contains important learning topics about a given domain. So it gives a coverage of relevant topics that are useful for learning about a given domain. The identified topics are automatically identified from the domain and used to create the background knowledge. A limitation of the existing method is that the relationships between the identified learning topics are not captured. For example, the approach does not capture if a concept is a sub-concept of another one, so there is not a notion of hierarchy and structure within the created background knowledge. A possible future direction can look at incorporating structure when creating background knowledge. This would be useful for differentiating generic concepts from more specialized concepts.

We present two potential applications for having structure in background knowledge. First, there is potential for query refinement. Using a structured background knowledge for refinement would be useful for determining the extent to which a query should be refined. The investigation of coverage of relevant topics in §6.4.2 revealed that some queries are more specialized than others. For example, if our aim was to narrow down the scope of a query during refinement, then more specialized concepts could be employed for refinement. In such a scenario, a knowledge of the hierarchy of concepts will be beneficial when refining queries. Further, the similarity between concepts can give an indication of how generic or specialized a concept is.

A second application for structure in background knowledge is in supporting search by exploration. This application would allow learners to search for learning materials through a structured view of the learning domain. This will be useful in scenarios where learners have a vague idea of what they wish to learn, but are not sure how to go about searching for the topics. A related application area for this is in intelligent browsing of documents. Where the learner is able to browse through available documents in a collection, and the documents are organised based on the topics in the domain. Further, a learner can see which documents describe a given concept because of the way the documents are arranged. This approach helps to prevent the learner from being overwhelmed with documents but rather provides an intelligent way of finding learning materials.

7.3.2 Improving the HYBRID Query Refinement Method

The HYBRID query refinement approach presented in this research is effective. However, there is the potential for extending the HYBRID method, by exploring the use of a Machine Learning approach for choosing when to refine a query. The development of the proposed method can be modelled as a classification problem. In the proposed method, the features of queries in a large query collection can be extracted and used to build a classifier that automatically determines what method is suitable for refining a query based on the features of that query. Features such as the query length, similarity between the query and the concepts, the presence of a concept label in a query, and a check to determine if the query is the same as a concept label, can be investigated when extending the hybrid approach for refining queries. This approach would need a set of queries for building the model and additional queries for testing the accuracy of the model. The model would predict what method should be used for refining a query based on the query features.

7.3.3 Enriching the CONCEPTBASED Query Refinement Method

The CONCEPTBASED* document representation method uses the concept vocabulary for representing learning materials. In this method, a concept-based representation is applied to the documents as well as to the query, which produces a query document. It is noted that the query is in the form of a document. This approach is used in chapters 3 and 4, and the outcome is a shared vocabulary that enables relevant learning materials to be retrieved. Our CONCEPTBASED query refinement approach identifies the most similar concepts to a query and uses highly weighted concept terms to refine a query. A standard retrieval method is then applied to the refined query. It is noted that the query here is in the form of a short learner query. Our query refinement method is used in chapters 5 and 6, and it is effective for identifying relevant learning materials.

A potential future direction is to create a richer concept-based representation of a learner's query that can produce a query document where the query terms are weighted. The weighting of query terms would capture their importance within the concept space. So, given a new query, the concepts will be searched to determine the relevant concepts to the query. The potentially relevant concepts will be combined to form a query document. This query document will then form the new query that is used to search on a document collection. We expect that using the query document to search should enhance the retrieval of relevant documents because the query document captures a richer concept representation. The CONCEPTBASED* document representation and the CONCEPTBASED query refinement methods can then be incorporated within a recommender system. This would allow us to leverage the rich vocabulary from concepts both for representing documents and for refining queries.

7.4 Conclusions

There are large amounts of e-Learning materials currently available to learners on the Web. However, learners often have difficulty finding relevant materials to meet their learning goals because of the semantic and intent gaps. The vocabulary used by teaching experts is often different from that used by learners, and this results in a semantic gap. Learners lack sufficient knowledge about the domain they are trying to learn about, so are unable to ask effective queries that convey what they wish to learn, and this presents an intent gap.

The work presented in this thesis focuses on addressing the semantic and intent gaps that learners face during e-Learning recommendation. These issues are addressed by creating background knowledge which contains important learning concepts drawn from teaching materials used by domain experts. The background knowledge provides a shared vocabulary between domain experts and learners. The vocabulary is used for representing learning materials, and this improves the retrieval of relevant learning resources.

An e-Learning recommender system is created that employs the background knowledge to identify relevant concepts that are used for the refinement of learners' queries. The concept vocabulary produces an effective query that focuses the search on documents that are relevant to learners. Evaluation results demonstrate the effectiveness of our knowledge driven approaches to support the retrieval in the recommendation of relevant e-Learning materials.

This research has shown how the knowledge from domain experts can be leveraged to build knowledge driven approaches for e-Learning recommendation. The impact of the adoption of the approaches developed in this research can enable the increased engagement of learners with e-Learning materials.

Bibliography

Aciar, S., Zhang, D., Simoff, S. & Debenham, J. (2007). Informed recommender: Basing recommendations on consumer product reviews, *IEEE Intelligent Systems* **22**(3): 39–47.

Adomavicius, G. & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions, *IEEE Transactions on Knowledge and Data Engineering* **17**(6): 734–749.

Agrawal, R., Chakraborty, S., Gollapudi, S., Kannan, A. & Kenthapadi, K. (2012). Quality of textbooks: An empirical study, *ACM Symposium on Computing for Development*, pp. 16:1–16:1.

Anibaldi, S., Jaques, Y., Celli, F., Stellato, A. & Keizer, J. (2015). Migrating bibliographic datasets to the Semantic Web: The AGRIS case, *Semantic Web* **6**(2): 113–120.

Anick, P. (2003). Using terminological feedback for web search refinement: a log-based study, *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 88–95.

Beliga, S., Meštrović, A. & Martinčić-Ipšić, S. (2015). An overview of graph-based keyword extraction methods and approaches, *Journal of Information and Organizational Sciences* **39**(1): 1–20.

Benabdellah, N. C., Gharbi, M. & Bellajkih, M. (2013). Content adaptation and learner profile definition: Ant colony algorithm application, *8th International Conference on Intelligent Systems: Theories and Applications*, IEEE, pp. 1–7.

Bendersky, M., Metzler, D. & Croft, W. B. (2012). Effective query formulation with multiple information sources, *Proceedings of the 5th ACM International Conference on Web Search and Data Mining*, pp. 443–452.

Bennett, J. & Lanning, S. (2007). The Netflix prize, *Proceedings of Knowledge Discovery and Data Mining cup and workshop*, Vol. 2007, p. 35.

Biswas, P. (2012). A brief survey on user modelling in human computer interaction, *Speech, Image and Language Processing for Human Computer Interaction: Multi-Modal Advancements* pp. 1–19.

Blei, D. M. & McAuliffe, J. D. (2007). Supervised topic models, *Neural Information Processing Systems*, pp. 121–128.

Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003). Latent Dirichlet Allocation, *Journal of Machine Learning Research* **3**: 993–1022.

Bobadilla, J., Hernando, A. & Arroyo, A. (2011). E-learning experience using recommender systems, *Proceedings of the 42nd ACM Technical Symposium on Computer Science Education*, ACM, pp. 477–482.

Bohl, O., Scheuhase, J., Sengler, R. & Winand, U. (2002). The Sharable Content Object Reference Model (SCORM)-a critical review, *Proceedings of the IEEE International Conference on Computers in Education*, pp. 950–951.

Bonk, C. J. & Graham, C. R. (2012). *The handbook of blended learning: Global perspectives, local designs*, John Wiley & Sons.

Bousbahi, F. & Chorfi, H. (2015). MOOC-Rec: A case based recommender system for MOOCs, *Procedia - Social and Behavioral Sciences* **195**: 1813 – 1822.

Boyce, S. & Pahl, C. (2007). Developing domain ontologies for course content, *Journal of Educational Technology & Society* **10**(3): 275–288.

Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia, *User Modeling and User-Adapted Interaction* **6**(2-3): 87–129.

Burgess, J. & Green, J. (2009). *YouTube: Online Video and Participatory Culture*, Cambridge: Polity Press.

Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments, *User Modeling and User-Adapted Interaction* **12**(4): 331–370.

Chandramouli, B., Levandoski, J. J., Eldawy, A. & Mokbel, M. F. (2011). StreamRec: a real-time recommender system, *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*, pp. 1243–1246.

Chaves, M. S., Gomes, R. & Pedron, C. (2012). Analysing reviews in the Web 2.0: Small and medium hotels in Portugal, *Tourism Management* **33**(5): 1286 – 1287.

Chen, W., Niu, Z., Zhao, X. & Li, Y. (2014). A hybrid recommendation algorithm adapted in e-learning environments, *World Wide Web* **17**(2): 271–284.

Chen, Z. & Liu, B. (2014). Topic modeling using topics from many domains, lifelong learning and big data, *31st International Conference on Machine Learning*, pp. 703–711.

Clarà, M. & Barberà, E. (2013). Learning online: Massive Open Online Courses (MOOCs), connectivism, and cultural psychology, *Distance Education* **34**(1): 129–136.

Coenen, F., Leng, P., Sanderson, R. & Wang, Y. J. (2007). Statistical identification of key phrases for text classification, *Machine Learning and Data Mining in Pattern Recognition*, Springer, pp. 838–853.

Conati, C. & Kardan, S. (2013). Student Modeling: Supporting Personalized Instruction, from Problem Solving to Exploratory Open-Ended Activities, *AI Magazine* **34**(3): 13–26.

Copeland, M., Soh, J., Puca, A., Manning, M. & Gollob, D. (2015). *Microsoft Azure: Planning, Deploying, and Managing Your Data Center in the Cloud*, Apress.

Cordier, A., Mascret, B. & Mille, A. (2009). Extending case-based reasoning with traces, *In Grand Challenges for Reasoning from Experiences, Workshop at International Joint Conference on Artificial Intelligence*, pp. 23–32.

Crompton, H. (2014). A diachronic overview of technology contributing to mobile learning: A shift towards student-centred pedagogies, *Increasing access through mobile learning* pp. 7–15.

Dayley, B. (2014). *Learning AngularJS*, Pearson Education.

de Albornoz, J. C., Plaza, L., Gervás, P. & Díaz, A. (2011). A joint model of feature mining and sentiment analysis for product review rating, *in* P. Clough, C. Foley, C. Gurrin, G. Jones,

W. Kraaij, H. Lee & V. Mudoch (eds), *Advances in Information Retrieval*, Vol. 6611 of *Lecture Notes in Computer Science*, Springer, pp. 55–66.

de Rosis, F., De Carolis, B. & Pizzutilo, S. (1994). User tailored hypermedia explanations, *Workshop on Adaptive Hypertext and Hypermedia at 4th International Conference on User Modeling*, pp. 169–170.

Deerwester, S., Dumais, S., Furnas, G., Landauer, T. & Harshman, R. (1990). Indexing by latent semantic analysis, *Journal of the American Society for Information Science* **41**(6): 391–407.

Di Bitonto, P., Laterza, M., Roselli, T. & Rossano, V. (2010). A recommendation method for e-learning environments: the rule-based technique, *Journal of e-Learning and Knowledge Society* **6**(3).

Diaz, F. & Metzler, D. (2006). Improving the estimation of relevance models using large external corpora, *Proceedings of the 29th ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 154–161.

Dietze, S., Yu, H. Q., Giordano, D., Kaldoudi, E., Dovrolis, N. & Taibi, D. (2012). Linked education: Interlinking educational resources and the Web of data, *27th Annual ACM Symposium on Applied Computing*, pp. 366–371.

Dong, R., O'Mahony, M. P., Schaal, M., McCarthy, K. & Smyth, B. (2013). Sentimental product recommendation, *Proceedings of the 7th ACM conference on Recommender systems*, pp. 411–414.

Dong, R., Schaal, M., O'Mahony, M. P., McCarthy, K. & Smyth, B. (2013). Opinionated product recommendation, *in* S. J. Delany & S. Ontañón (eds), *Case-Based Reasoning Research and Development*, Vol. 7969, Springer, pp. 44–58.

Dostál, M. & Jezek, K. (2011). Automatic keyphrase extraction based on NLP and statistical methods, *Proceedings of the DATESO: Annual International Workshop on DAtabases, TExts, Specifications and Object*, pp. 140–145.

Felder, R. M. & Silverman, L. K. (1988). Learning and teaching styles in engineering education, *Engineering education* **78**(7): 674–681.

Fellbaum, C. (2005). WordNet and wordnets, *in* K. Brown (ed.), *Encyclopedia of Language and Linguistics*, Elsevier, Oxford, pp. 665–670.

Fernández-López, M., Gómez-Pérez, A. & Juristo, N. (1997). METHONTOLOGY: from ontological art towards ontological engineering, *Proceedings of the AAAI Spring Symposium Series on Ontological Engineering*, American Asociation for Artificial Intelligence, Stanford, USA, pp. 33–40.

Ferrucci, D., Levas, A., Bagchi, S., Gondek, D. & Mueller, E. T. (2013). Watson: Beyond Jeopardy!, *Artificial Intelligence* **199**: 93–105.

Fleming, N. D. (1995). I'm different; not dumb. Modes of presentation (VARK) in the tertiary classroom, *Research and Development in Higher Education, Proceedings of the Annual Conference of the Higher Education and Research Development Society of Australasia*, Vol. 18, pp. 308–313.

Fogarolli, A., Brickley, D., Anibaldi, S. & Keizer, J. (2011). AGRIS-from a bibliographical database to a Web data service on agricultural research information, *Agricultural Information Worldwide* **4**(1): 10–15.

Gabrilovich, E. & Markovitch, S. (2007). Computing semantic relatedness using wikipedia-based explicit semantic analysis, *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, Morgan Kaufmann Publishers, pp. 1606–1611.

Ghadfi, S., Béchet, N. & Berio, G. (2014). Building ontologies from textual resources: A pattern based improvement using deep linguistic information, *Proceedings of the 5th International Conference on Ontology and Semantic Web Patterns-Volume 1302*, CEUR-WS.org, pp. 14–25.

Ghauth, K. I. B. & Abdullah, N. A. (2009). Building an e-learning recommender system using vector space model and good learners average rating, *9th IEEE International Conference on Advanced Learning Technologies*, pp. 194–196.

Ghazanfar, M. A. & Prgel-Bennett, A. (2014). Leveraging clustering approaches to solve the gray-sheep users problem in recommender systems, *Expert Systems with Applications* **41**(7): 3261–3275.

Giles, J. (2005). Internet encyclopaedias go head to head, *Nature* **438**: 900–901.

Giuliano, C. A., Moser, L. R., Poremba, V., Jones, J., Martin, E. T. & Slaughter, R. L. (2014). Use of a unified learning style model in pharmacy curricula, *Currents in Pharmacy Teaching and Learning* **6**(1): 41–57.

Goel, A. K. & Joyner, D. A. (2016). Design of an online course on Knowledge-Based AI, *Proceedings of the 13th AAAI Conference on Artificial Intelligence*, pp. 4089–4094.

Goel, A. K. & Joyner, D. A. (2017). Using AI to Teach AI: Lessons from an Online AI Class, *AI Magazine* **38**(2).

González, C., Burguillo, J. C. & Llamas, M. (2006). A Qualitative Comparison of Techniques for Student Modeling in Intelligent Tutoring Systems, *36th Annual Frontiers in Education Conference*, IEEE, pp. 13–18.

Hakala, J. (2000). Dublin core metadata initiative, *Current Research Information Systems* .

Hands, A. (2012). Microsoft academic search, *Technical Services Quarterly* **29**(3): 251–252.

Harper, F. M. & Konstan, J. A. (2016). The movielens datasets: History and context, *ACM Transactions on Interactive Intelligent Systems (TiiS)* **5**(4): 19.

Hatcher, E., Gospodnetic, O. & McCandless, M. (2004). *Lucene in action*, Manning Publications.

Hauger, S., Tso, K. H. & Schmidt-Thieme, L. (2008). Comparison of recommender system algorithms focusing on the new-item and user-bias problem, *Data Analysis, Machine Learning and Applications*, Springer, pp. 525–532.

He, B. & Ounis, I. (2007). Combining fields for query expansion and adaptive query expansion, *Information Processing & Management* **43**(5): 1294–1307.

Holzinger, A., Kleinberger, T. & Müller, P. (2001). Multimedia learning systems based on IEEE learning object metadata (LOM), *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, Vol. 2001, pp. 772–777.

Horsburgh, B., Craw, S. & Massie, S. (2015). Learning pseudo-tags to augment sparse tagging in hybrid music recommender systems, *Artificial Intelligence* **219**: 25–39.

Hothi, J. & Hall, W. (1998). An Evaluation of Adapted Hypermedia Techniques Using Static User Modelling, *In Proceedings of the 2nd Workshop on Adaptive Hypertext and Hypermedia*, pp. 45–50.

Hu, B., Li, Z. & Chao, W. (2012). *Data Sparsity: A Key Disadvantage of User-Based Collaborative Filtering?*, Springer Berlin Heidelberg, pp. 602–609.

Hu, M. & Liu, B. (2004). Mining and summarizing customer reviews, *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, USA, pp. 168–177.

IEEE (2002). Draft standard for learning object metadata, *IEEE Learning Technology Standards Committee* .

Jannach, D., Zanker, M., Felfernig, A. & Friedrich, G. (2010). *Recommender Systems: An Introduction*, Cambridge University Press.

Kantor, P. B., Rokach, L., Ricci, F. & Shapira, B. (2011). *Recommender systems handbook*, Springer.

Klašnja-Milićević, A., Vesin, B., Ivanović, M. & Budimac, Z. (2011). E-learning personalization based on hybrid recommendation strategy and learning style identification, *Computers & Education* **56**(3): 885–899.

Knublauch, H., Fergerson, R. W., Noy, N. F. & Musen, M. A. (2004). The protégé owl plugin: An open development environment for semantic web applications, *Proceedings of the 3rd International Semantic Web Conference*, Springer Berlin Heidelberg, pp. 229–243.

Kritikou, Y., Demestichas, P., Adamopoulou, E., Demestichas, K., Theologou, M. & Paradia, M. (2008). User profile modeling in the context of web-based learning management systems, *Journal of Network and Computer Applications* **31**(4): 603–627.

Kuć, R. & Rogoziński, M. (2015). *Mastering Elasticsearch*, Packt Publishing Ltd.

Lang, K. (1995). NewsWeeder: Learning to Filter Netnews, *In Proceedings of the 12th International Conference on Machine Learning*, Citeseer, pp. 331–339.

Li, H., Councill, I., Lee, W.-C. & Giles, C. L. (2006). CiteSeerx: an architecture and web service design for an academic document search engine, *In Proceedings of the 15th ACM International conference on World Wide Web*, pp. 883–884.

Li, Y., Luk, W. P. R., Ho, K. S. E. & Chung, F. L. K. (2007). Improving weak ad-hoc queries using Wikipedia as external corpus, *Proceedings of the 30th ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 797–798.

Lika, B., Kolomvatsos, K. & Hadjiefthymiades, S. (2014). Facing the cold start problem in recommender systems, *Expert Systems with Applications* **41**(4): 2065–2073.

Linden, G., Smith, B. & York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering, *IEEE Internet Computing* **7**(1): 76–80.

Liu, J., Kim, C. S. & Creel, C. (2013). Why do users feel search task difficult?, *Proceedings of the Association for Information Science and Technology* **50**(1): 1–4.

Lops, P., Gemmis, M. & Semeraro, G. (2011). Content-Based Recommender Systems: State of the Art and Trends, *in* F. Ricci, L. Rokach, B. Shapira & P. B. Kantor (eds), *Recommender Systems Handbook*, Springer US, pp. 73–105.

Manning, C. D., Raghavan, P. & Schütze, H. (2008). *Introduction to Information Retrieval*, Cambridge University Press.

Manouselis, N., Drachsler, H., Verbert, K. & Duval, E. (2013). *Recommender systems for learning*, Springer.

Mbipom, B., Craw, S. & Massie, S. (2016). Harnessing background knowledge for e-learning recommendation., *Research and Development in Intelligent Systems XXXIII: Incorporating Applications and Innovations in Intelligent Systems XXIV*, Springer, pp. 3–17.

Mbipom, B., Craw, S. & Massie, S. (2018). Improving e-learning recommendation by using background knowledge, *Expert Systems* . doi: 10.1111/exsy.12265.

Mbipom, B., Massie, S. & Craw, S. (2018). An e-learning recommender that helps learners find the right materials, *Proceedings of the 8th Symposium on Educational Advances in Artificial Intelligence*, AAAI Press, New Orleans, LA.

McCandless, M., Hatcher, E. & Gospodnetic, O. (2010). *Lucene in Action*, Manning Publications Co.

McGuinness, D. L. & Van Harmelen, F. (2004). OWL Web Ontology Language Overview, *W3C recommendation* .

Meij, E., Bron, M., Hollink, L., Huurnink, B. & De Rijke, M. (2009). *Learning semantic query suggestions*, Springer Berlin Heidelberg, pp. 424–440.

Meij, E., Bron, M., Hollink, L., Huurnink, B. & de Rijke, M. (2011). Mapping queries to the linking open data cloud: A case study using DBpedia, *Web Semantics: Science, Services and Agents on the World Wide Web* **9**(4): 418–433.

Meij, E. & de Rijke, M. (2010). Supervised query modeling using Wikipedia, *Proceedings of the 33rd ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 875–876.

Melville, P., Mooney, R. J. & Nagarajan, R. (2002). Content-boosted collaborative filtering for improved recommendations, *18th National Conference on Artificial Intelligence*, American Association for Artificial Intelligence, Menlo Park, CA, USA, pp. 187–192.

Mikolov, T., Chen, K., Corrado, G. & Dean, J. (2013). Efficient estimation of word representations in vector space, *International Conference on Learning Representations: Workshops Track* .

Millard, D. E., Gibbins, N. M., Michaelides, D. T. & Weal, M. J. (2005). Mind the semantic gap, *Proceedings of the 16th ACM Conference on Hypertext and Hypermedia*, New York, USA, pp. 54–62.

Milne, D. & Witten, I. H. (2008). Learning to link with Wikipedia, *17th ACM Conference on Information and Knowledge Management*, pp. 509–518.

Moore, M. G. & Kearsley, G. (2011). *Distance education: A systems view of online learning*, Cengage Learning.

Muñoz-García, O., García-Silva, A., Corcho, O., Higuera Hernández, M. & Navarro, C. (2011). Identifying topics in social media posts using DBpedia, *Proceedings of the Networked and Electronic Media Summit*, Informatica, pp. 81–86.

Nasraoui, O. & Zhuhadar, L. (2010). Improving recall and precision of a personalized semantic search engine for e-learning, *4th IEEE International Conference on Digital Society*, pp. 216–221.

Nov, O. (2007). What motivates Wikipedians?, *Communications of the ACM* **50**(11): 60–64.

O'Mahony, M. & Smyth, B. (2010). A classification-based review recommender, *Knowledge-Based Systems* **23**(4): 323 – 329.

Panagiotis, S., Ioannis, P., Christos, G. & Achilles, K. (2016). APLe: Agents for Personalized Learning in distance learning, *7th International Conference on Computer Supported Education*, Springer, pp. 37–56.

Pang, B. & Lee, L. (2008). Opinion mining and sentiment analysis, *Foundations and trends in Information Retrieval* **2**(1-2): 1–135.

Park, D. H., Kim, H. K., Choi, I. Y. & Kim, J. K. (2012). A literature review and classification of recommender systems research, *Expert Systems with Applications* **39**(11): 10059–10072.

Park, Y.-J. & Tuzhilin, A. (2008). The long tail of recommender systems and how to leverage it, *Proceedings of the 2008 ACM Conference on Recommender Systems*, RecSys, pp. 11–18.

Peter, S. E., Bacon, E. & Dastbaz, M. (2010). Personalised and adaptable learning: designing the architecture, *"e" Teaching and Learning Workshop Proceedings*, pp. 11–15.

Pilli, O. & Aksu, M. (2013). The effects of computer-assisted instruction on the achievement, attitudes and retention of fourth grade mathematics students in north cyprus, *Computers & Education* **62**: 62–71.

Popenici, S. A. & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education, *Research and Practice in Technology Enhanced Learning* **12**(1).

Popescu, E. (2010). A unified learning style model for technology-enhanced learning: What, why and how?, *International Journal of Distance Education Technologies* **8**(3): 65–81.

Porter, M. F. (1980). An algorithm for suffix stripping, *Program* **14**(3): 130–137.

Qureshi, M. A., O'Riordan, C. & Pasi, G. (2014). Exploiting Wikipedia to identify domain-specific key terms/phrases from a short-text collection, *Proceedings of the 5th Italian Information Retrieval Workshop*, pp. 63–74.

Ricci, F., Rokach, L. & Shapira, B. (2011). Introduction to Recommender Systems Handbook, *in* F. Ricci, L. Rokach, B. Shapira & P. B. Kantor (eds), *Recommender Systems Handbook*, Springer US, pp. 1–35.

Rodrigues, L., Antunes, B., Gomes, P., Santos, A., Barbeira, J. & Carvalho, R. (2007). Using textual CBR for e-learning content categorization and retrieval, *4th International Conference on Case-Based Reasoning workshop on textual case-based reasoning*.

Ruiz-Iniesta, A., Jimenez-Diaz, G. & Gomez-Albarran, M. (2014). A semantically enriched context-aware OER recommendation strategy and its application to a computer science OER repository, *IEEE Transactions on Education* **57**(4): 255–260.

Salton, G. (1971). *The SMART Retrieval System-Experiments in Automatic Document Processing*, Prentice-Hall, Upper Saddle River, NJ, USA.

Salton, G. & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval, *Information Processing & Management* **24**(5): 513–523.

Schafer, J. B., Konstan, J. & Riedi, J. (1999). Recommender Systems in E-Commerce, *In Proceedings of the 1st ACM Conference on Electronic Commerce*, ACM, pp. 158–166.

Schein, A. I., Popescul, A., Ungar, L. H. & Pennock, D. M. (2002). Methods and metrics for cold-start recommendations, *Proceedings of the 25th Annual International ACM SIGIR conference on Research and Development in Information Retrieval*, pp. 253–260.

Schiaffino, S., Garcia, P. & Amandi, A. (2008). eTeacher: Providing personalized assistance to e-learning students, *Computers & Education* **51**(4): 1744–1754.

Shen, L. & Shen, R. (2004). Learning content recommendation service based-on simple sequencing specification, *Advances in Web-Based Learning - International Conference on Web-Based Learning*, Springer Berlin Heidelberg, pp. 363–370.

Shiri, M. E., Aïmeur, A. E. & Frasson, C. (1998). Student modelling by Case Based Reasoning, *Proceedings of the 4th International Conference on Intelligent Tutoring Systems*, Springer Berlin Heidelberg, pp. 394–403.

Sosnovsky, S. & Gavrilova, T. (2006). Development of educational ontology for C-programming, *International Journal on Information Theories and Applications* **13**(4): 303–308.

Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval, *Journal of documentation* **28**(1): 11–21.

Tan, H., Guo, J. & Li, Y. (2008). E-learning recommendation system, *IEEE International Conference on Computer Science and Software Engineering*, Vol. 5, pp. 430–433.

Tenopir, C., Wang, P., Zhang, Y., Simmons, B. & Pollard, R. (2008). Academic users interactions with ScienceDirect in search tasks: Affective and cognitive behaviors, *Information processing & management* **44**(1): 105–121.

Timonen, M., Toivanen, T., Kasari, M., Teng, Y., Cheng, C. & He, L. (2012). Keyword extraction from short documents using three levels of word evaluation, *International Joint Conference on Knowledge Discovery, Knowledge Engineering, and Knowledge Management*, Springer, pp. 130–146.

Tkalcic, M., Kunaver, M., Košir, A. & Tasic, J. (2011). Addressing the new user problem with a personality based user similarity measure, *1st International Workshop on Decision Making and Recommendation Acceptance Issues in Recommender Systems*, p. 106.

Vasilyeva, E., Pechenizkiy, M. & De Bra, P. (2008). Adaptation of elaborated feedback in e-Learning, *Adaptive Hypermedia and Adaptive Web-Based Systems*, Springer, pp. 235–244.

Vélez, B., Weiss, R., Sheldon, M. A. & Gifford, D. K. (1997). Fast and effective query refinement, *Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Vol. 31, pp. 6–15.

Völkel, M., Krötzsch, M., Vrandecic, D., Haller, H. & Studer, R. (2006). Semantic Wikipedia, *15th ACM International Conference on World Wide Web*, pp. 585–594.

Wang, T. (2007). What strategies are effective for formative assessment in an e-learning environment, *Journal of Computer Assisted Learning* **23**(3): 171–186.

Weibel, S., Kunze, J., Lagoze, C. & Wolf, M. (1998). Dublin core metadata for resource discovery, *Internet Engineering Task Force Request For Comments* **2413**(222): 132.

Weijters, B., Cabooter, E. & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels, *International Journal of Research in Marketing* **27**(3): 236–247.

Witten, I. H., Paynter, G. W., Frank, E., Gutwin, C. & Nevill-Manning, C. G. (1999). KEA: Practical automatic keyphrase extraction, *4th ACM Conference on Digital libraries*, pp. 254–255.

Wu, H. & Fang, H. (2013). An incremental approach to efficient pseudo-relevance feedback, *Proceedings of the 36th ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 553–562.

Xu, Y., Jones, G. J. & Wang, B. (2009). Query dependent pseudo-relevance feedback based on Wikipedia, *Proceedings of the 32nd ACM SIGIR conference on Research and development in Information Retrieval*, pp. 59–66.

Yang, H.-L. & Lai, C.-Y. (2010). Motivations of Wikipedia content contributors, *Computers in Human Behavior* **26**(6): 1377 – 1383.

Yang, K., Chen, Z., Cai, Y., Huang, D. & Leung, H. F. (2016). Improved automatic keyword extraction given more semantic knowledge, *International Conference on Database Systems for Advanced Applications*, Springer, pp. 112–125.

Yarandi, M., Tawil, A.-R. & Jahankhani, H. (2011). Adaptive e-learning system using ontology, *In Proceedings of the 22nd International Workshop on Database and Expert Systems Applications*, pp. 511–516.

Zaíane, O. R. (2002). Building a recommender agent for e-learning systems, *Proceedings of the IEEE International Conference on Computers in Education*, pp. 55–59.

Zhang, X., Liu, J. & Cole, M. (2013). Task topic knowledge vs. background domain knowledge: Impact of two types of knowledge on user search performance, *Advances in Information Systems and Technologies*, Springer, pp. 179–191.

Zheng, Z., Li, F., Huang, M. & Zhu, X. (2010). Learning to link entities with knowledge base, *Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 483–491.

Appendix A

Published Papers

Mbipom, B., Craw, S. & Massie, S. (2016). Harnessing background knowledge for e-learning recommendation. In: Research and Development in Intelligent Systems XXXIII. Springer pp. 3–17. Awarded Donald Michie Memorial Prize for the Best Technical Paper.

Mbipom, B., Massie, S. & Craw, S. (2018). An e-learning recommender that helps learners find the right materials, *Proceedings of the 8th Symposium on Educational Advances in Artificial Intelligence (EAAI-18)*, AAAI Press, New Orleans, LA.

Mbipom, B., Craw, S. & Massie, S. (2018). Improving e-learning recommendation by using background knowledge, *Expert Systems*. doi: 10.1111/exsy.12265.

Harnessing Background Knowledge for E-learning Recommendation

Blessing Mbipom, Susan Craw and Stewart Massie

Abstract The growing availability of good quality, learning-focused content on the Web makes it an excellent source of resources for e-learning systems. However, learners can find it hard to retrieve material well-aligned with their learning goals because of the difficulty in assembling effective keyword searches due to both an inherent lack of domain knowledge, and the unfamiliar vocabulary often employed by domain experts. We take a step towards bridging this semantic gap by introducing a novel method that automatically creates custom background knowledge in the form of a set of rich concepts related to the selected learning domain. Further, we develop a hybrid approach that allows the background knowledge to influence retrieval in the recommendation of new learning materials by leveraging the vocabulary associated with our discovered concepts in the representation process. We evaluate the effectiveness of our approach on a dataset of Machine Learning and Data Mining papers and show it to outperform the benchmark methods.

1 Introduction

There is currently a large amount of e-learning resources available to learners on the Web. However, learners have insufficient knowledge of the learning domain, and are not able to craft good queries to convey what they wish to learn. So, learners are often discouraged by the time spent in finding and assembling relevant resources to meet their learning goals [5]. E-learning recommendation offers a possible solution.

E-learning recommendation typically involves a learner query, as an input; a collection of learning resources from which to make recommendations; and selected resources recommended to the learner, as an output. Recommendation differs from

Blessing Mbipom · Susan Craw · Stewart Massie
School of Computing Science & Digital Media, Robert Gordon University, Aberdeen, UK

Blessing Mbipom e-mail: b.e.mbipom@rgu.ac.uk · Susan Craw e-mail: s.craw@rgu.ac.uk · Stewart Massie e-mail: s.massie@rgu.ac.uk

an information retrieval task because with the latter, the user requires some understanding of the domain in order to ask and receive useful results, but in e-learning, learners do not know enough about the domain. Furthermore, the e-learning resources are often unstructured text, and so are not easily indexed for retrieval [11]. This challenge highlights the need to develop suitable representations for learning resources in order to facilitate their retrieval.

We propose the creation of background knowledge that can be exploited for problem-solving. In building our method, we leverage the knowledge of instructors contained in eBooks as a guide to identify the important domain topics. This knowledge is enriched with information from an encyclopedia source and the output is used to build our background knowledge. DeepQA applies a similar approach to reason on unstructured medical reports in order to improve diagnosis [9]. We demonstrate the techniques in Machine Learning and Data Mining, however the techniques we describe can be applied to other learning domains.

In this paper, we build background knowledge that can be employed in e-learning environments for creating representations that capture the important concepts within learning resources in order to support the recommendation of resources. Our method can also be employed for query expansion and refinement. This would allow learners' queries to be represented using the vocabulary of the domain with the aim of improving retrieval. Alternatively, our approach can enable learners to browse available resources through a guided view of the learning domain.

We make two contributions: firstly, the creation of background knowledge for an e-learning domain. We describe how we take advantage of the knowledge of experts contained in eBooks to build a knowledge-rich representation that is used to enhance recommendation. Secondly, we present a method of harnessing background knowledge to augment the representation of learning resources in order to improve the recommendation of resources. Our results confirm that incorporating background knowledge into the representation improves e-learning recommendation.

This paper is organised as follows: Sect. 2 presents related methods used for representing text; Sect. 3 describes how we exploit information sources to build our background knowledge; Sect. 4 discusses our methods in harnessing a knowledge-rich representation to influence e-learning recommendation; and Sect. 5 presents our evaluation. We conclude in Sect. 6 with insights to further ways of exploiting our background knowledge.

2 Related Work

Finding relevant resources to recommend to learners is a challenge because the resources are often unstructured text, and so are not appropriately indexed to support the effective retrieval of relevant materials. Developing suitable representations to improve the retrieval of resources is a challenging task in e-learning environments [8], because the resources do not have a pre-defined set of features by which they can be indexed. So, e-learning recommendation requires a representation that captures the domain-specific vocabulary contained in learning resources. Two broad approaches are often used to address the challenge of text representation: corpus-based methods such as topic models [6], and structured representations such as those that take advantage of ontologies [4].

Corpus-based methods involve the use of statistical models to identify topics from a corpus. The identified topics are often keywords [2] or phrases [7, 18]. Coenen et al. showed that using a combination of keywords and phrases was better than using only keywords [7]. Topics can be extracted from different text sources such as learning resources [20], metadata [3], and Wikipedia [14]. One drawback of the corpus-based approach is that, it is dependent on the document collection used, so the topics produced may not be representative of the domain. A good coverage of relevant topics is required when generating topics for an e-learning domain, in order to offer recommendations that meet learners' queries which can be varied.

Structured representations capture the relationships between important concepts in a domain. This often entails using an existing ontology [11, 15], or creating a new one [12]. Although ontologies are designed to have a good coverage of their domains, the output is still dependent on the view of its builders, and because of handcrafting, existing ontologies cannot easily be adapted to new domains. E-learning is dynamic because new resources are becoming available regularly, and so using fixed ontologies limits the potential to incorporate new content.

A suitable representation for e-learning resources should have a good coverage of relevant topics from the domain. So, the approach in this paper draws insight from the corpus-based methods and structured representations. We leverage on a structured corpus of teaching materials as a guide for identifying important topics within an e-learning domain. These topics are a combination of keywords and phrases as recommended in [7]. The identified topics are enriched with discovered text from Wikipedia, and this extends the coverage and richness of our representation.

3 Background Knowledge Representation

Background knowledge refers to information about a domain that is useful for general understanding and problem-solving [21]. We attempt to capture background knowledge as a set of domain concepts, each representing an important topic in the domain. For example, in a learning domain, such as Machine Learning, you would find topics such as Classification, Clustering and Regression. Each of these topics would be represented by a concept, in the form of a concept label and a pseudo-document which describes the concept. The concepts can then be used to underpin the representation of e-learning resources.

The process involved in discovering our set of concepts is illustrated in Figure 1. Domain knowledge sources are required as an input to the process, and we use a structured collection of teaching materials and an encyclopedia source. We automatically extract ngrams from our structured collection to provide a set of potential concept labels, and then we use a domain lexicon to validate the extracted ngrams in order to ensure that the ngrams are also being used in another information source. The encyclopedia provides candidate pages that become the concept label and discovered text for the ngrams. The output from this process is a set of concepts, each comprising a label and an associated pseudo-document. The knowledge extraction process is discussed in more detail in the following sections.

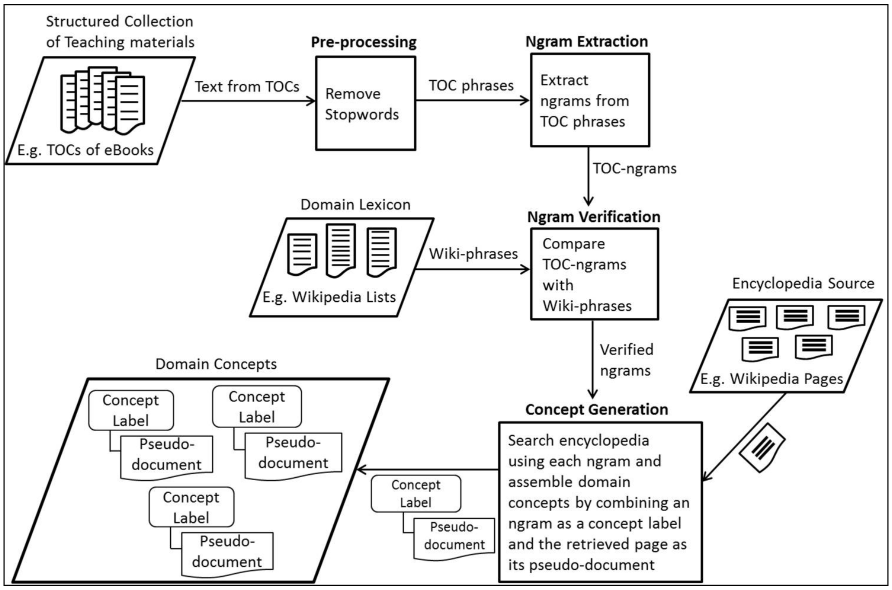


Fig. 1 An overview of the background knowledge creation process

3.1 Knowledge Sources

Two knowledge sources are used as initial inputs for discovering concept labels. A structured collection of teaching materials provides a source for extracting important topics identified by teaching experts in the domain, while a domain lexicon provides a broader but more detailed coverage of the relevant topics in the domain. The lexicon is used to verify that the concept labels identified from the teaching materials are directly relevant. Thereafter, an encyclopedia source, such as Wikipedia pages, is searched and provides the relevant text to form a pseudo-document for each verified concept label. The final output from this process is our set of concepts each comprising a concept label and an associated pseudo-document.

Our approach is demonstrated with learning resources from Machine Learning and Data Mining. We use eBooks as our collection of teaching materials; a summary of the books used is shown in Table 1. Two Google Scholar queries: “Introduction to data mining textbook” and “Introduction to machine learning textbook” guided the selection process, and 20 eBooks that meet all of the following 3 criteria were chosen. Firstly, the book should be about the domain. Secondly, there should be Google Scholar citations for the book. Thirdly, the book should be accessible. We use the Tables-of-Contents (TOCs) of the books as our structured knowledge source.

We use Wikipedia to create our domain lexicon because it contains articles for many learning domains [17], and the contributions of many people [19], so this provides the coverage we need in our lexicon. The lexicon is generated from 2 Wikipedia sources. First, the phrases in the *contents* and *overview* sections of the chosen domain are extracted to form a topic list. In addition, a list containing the titles of articles related to the domain is added to the topic list to assemble our lexicon. Overall, our domain lexicon consists of a set of 664 Wiki-phrases.

Table 1 Summary of eBooks used

Book Title & Author	**Cites**
Machine learning; Mitchell	264
Introduction to machine learning; Alpaydin	2621
Machine learning a probabilistic perspective; Murphy	1059
Introduction to machine learning; Kodratoff	159
Gaussian processes for machine learning; Rasmussen & Williams	5365
Introduction to machine learning; Smola & Vishwanathan	38
Machine learning, neural and statistical classification; Michie, Spiegelhalter, & Taylor	2899
Introduction to machine learning; Nilsson	155
A First Encounter with Machine Learning; Welling	7
Bayesian reasoning and machine learning; Barber	271
Foundations of machine learning; Mohri, Rostamizadeh, & Talwalkar	197
Data mining-practical machine learning tools and techniques; Witten & Frank	27098
Data mining concepts models and techniques; Gorunescu	244
Web data mining; Liu	1596
An introduction to data mining; Larose	1371
Data mining concepts and techniques; Han & Kamber	22856
Introduction to data mining; Tan, Steinbach, & Kumar	6887
Principles of data mining; Bramer	402
Introduction to data mining for the life sciences; Sullivan	15
Data mining concepts methods and applications; Yin, Kaku, Tang, & Zhu	23

3.2 Generating Potential Domain Concept Labels

In the first stage of the process, the text from the TOCs is pre-processed. We remove characters such as punctuation, symbols, and numbers from the TOCs, so that only words are used for generating concept labels. After this, we remove 2 sets of stopwords. First, a standard English stopwords list[1], which allows us to remove com-

[1] http://snowball.tartarus.org/algorithms/english/stop.txt

mon words and still retain a good set of words for generating our concept labels. Our second stopwords are an additional set of words which we refer to as TOC-stopwords. It contains: structural words, such as *chapter* and *appendix*, which relate to the structure of the TOCs; roman numerals, such as *xxiv* and *xxxv*, which are used to indicate the sections in a TOC; and words, such as *introduction* and *conclusion*, which describe parts of a learning material and are generic across domains.

We do not use stemming because we found it harmful during pre-processing. When searching an encyclopedia source with the stemmed form of words, relevant results would not be returned. In addition, we intend to use the background knowledge for query refinement, so stemmed words would not be helpful.

The output from pre-processing is a set of TOC phrases. In the next stage, we apply ngram extraction to the TOC phrases to generate all 1-3 grams across the entire set of TOC phrases. The output from this process are TOC-ngrams containing a set of 2038 unigrams, 5405 bigrams and 6133 trigrams, which are used as the potential domain concept labels. Many irrelevant ngrams are generated from the TOCs because we have simply selected all 1-3 grams.

3.3 Verifying Concept Labels using Domain Lexicon

The TOC-ngrams are first verified using a domain lexicon to confirm which of the ngrams are relevant for the domain. Our domain lexicon contains a set of 664 Wiki-phrases, each of which is pre-processed by removing non-alphanumeric characters. The 84% of the Wiki-phrases that are 1-3 grams are used for verification. The comparison of TOC-ngrams with the domain lexicon identifies the potential domain concept labels that are actually being used to describe aspects of the chosen domain in Wikipedia. During verification, ngrams referring directly to the title of the domain, e.g. *machine learning* and *data mining*, are not included because our aim is to generate concept labels that describe the topics within the domain. In addition, we intend to build pseudo-documents describing the identified labels, and so using the title of the domain would refer to the entire domain rather than specific topics. Overall, a set of 17 unigrams, 58 bigrams and 15 trigrams are verified as potential concept labels. Bigrams yield the highest number of ngrams, which indicates that bigrams are particularly useful for describing topics in this domain.

3.4 Domain Concept Generation

Our domain concepts are generated after a second verification step is applied to the ngrams returned from the previous stage. Each ngram is retained as a concept label if all of 3 criteria are met. Firstly, if a Wikipedia page describing the ngram exists. Secondly, if the text describing the ngram is not contained as part of the page describing another ngram. Thirdly, if the ngram is not a synonym of another ngram. For the third criteria, if two ngrams are synonyms, the ngram with the higher frequency is retained as a concept label while its synonym is retained as part of the extracted text. For example, 2 ngrams *cluster analysis* and *clustering* are regarded as synonyms in Wikipedia, so the text associated with them is the same. The label *clustering* is retained as the concept label because it occurs more frequently in the TOCs, and its synonym, *cluster analysis* is contained as part of the discovered text.

The concept labels are used to search Wikipedia pages in order to generate a domain concept. The search returns discovered text that forms a pseudo-document which includes the concept label. The concept label and pseudo-document pair make up a domain concept. Overall, 73 domain concepts are generated. Each pseudo-document is pre-processed using standard techniques such as removal of English stopwords and Porter stemming [13]. The terms from the pseudo-documents form the concept vocabulary that is now used to represent learning resources.

4 Representation using Background Knowledge

Our background knowledge contains a rich representation of the learning domain and by harnessing this knowledge for representing learning resources, we expect to retrieve documents based on the domain concepts that they contain. The domain concepts are designed to be effective for e-learning, because they are assembled from the TOCs of teaching materials [1]. This section presents two approaches which have been developed by employing our background knowledge in the representation of learning resources.

4.1 The CONCEPTBASED *approach*

Representing documents with the concept vocabulary allows retrieval to focus on the concepts contained in the documents. Figures 2 & 3 illustrate the CONCEPTBASED method. Firstly, in Figure 2, the concept vocabulary, $t_1 \ldots t_c$, from the pseudo-documents of concepts, $C_1 \ldots C_m$, is used to create a term-concept matrix and a term-document matrix using TF-IDF weighting [16]. In Figure 2a, c_{ij} is the TF-IDF of term t_i in concept C_j, while Figure 2b shows d_{ik} which is the TF-IDF of t_i in D_k.

Next, documents D_1 to D_n are represented with respect to concepts by computing the cosine similarity of the term vectors for concepts and documents. The output is the concept-document matrix shown in Figure 3a, where y_{jk} is the cosine similarity of the vertical shaded term vectors for C_j and D_k from Figures 2a and 2b respectively. Finally, the document similarity is generated by computing the cosine similarity of concept-vectors for documents. Figure 3b shows z_{km}, which is the cosine similarity of the concept-vectors for D_k and D_m from Figure 3a.

The CONCEPTBASED approach uses the document representation and similarity in Figure 3. By using the CONCEPTBASED approach we expect to retrieve docu-

	C_1	...	C_j	...	C_m
t_1					
...					
t_i			c_{ij}		
...					
t_C					

(a) Term-concept matrix

	D_1	...	D_k	...	D_n
t_1					
...					
t_i			d_{ik}		
...					
t_C					

(b) Term-document matrix

Fig. 2 Term matrices for concepts and documents

	D_1	...	D_k	...	D_m	...	D_n
C_1							
...							
C_j			y_{jk}		y_{jm}		
...							
C_m							

(a) Concept-document matrix representation

	D_1	...	D_m	...	D_n
D_1	1				
...					
D_k			z_{km}		
...					
D_n					1

(b) Document-document similarity

Fig. 3 Document representation and similarity using the ConceptBased approach

ments that are similar based on the concepts they contain, and this is obtained from the document-document similarity in Figure 3b. A standard approach of representing documents would be to define the document similarity based on the term document matrix in Figure 2b, but this exploits the concept vocabulary only. However, in our approach, we put more emphasis on the domain concepts, so we use the concept document matrix in Figure 3a, to underpin the similarity between documents.

4.2 The **Hybrid** *Approach*

The Hybrid approach exploits the relative distribution of the vocabulary in the concept and document spaces to augment the representation of learning resources with a bigger, but focused, vocabulary. So the TF-IDF weight of a term changes depending on its relative frequency in both spaces.

First, the concepts, C_1 to C_m and the documents we wish to represent, D_1 to D_n, are merged to form a corpus. Next, a term-document matrix with TF-IDF weighting is created using all the terms, t_1 to t_T from the vocabulary of the merged corpus as shown in Figure 4a. For example, entry q_{ik} is the TF-IDF weight of term t_i in D_k. If t_i has a lower relative frequency in the concept space compared to the document space, then the weight q_{ik} is boosted. So, distinctive terms from the concept space will get boosted. Although the overlap of terms from both spaces are useful for altering the term weights, it is valuable to keep all the terms from the document space because this gives us a richer vocabulary. The shaded term vectors for D_1 to D_n in Figure 4a form a term-document matrix for documents whose term weights have been influenced by the presence of terms from the concept vocabulary.

	C_1	...	C_j	...	C_m	D_1	...	D_j	D_k	...	D_n
t_1											
...											
t_i			p_{ij}						q_{ik}		
...											
t_T											

(a) Hybrid term-document matrix representation

	D_1	...	D_k	...	D_n
D_1	1				
...					
D_j			r_{jk}		
...					
D_n					1

(b) Hybrid document similarity

Fig. 4 Representation and similarity of documents using the HYBRID approach

Finally, the document similarity in Figure 4b, is generated by computing the cosine similarity between the augmented term vectors for D_1 to D_n. Entry r_{jk} is the cosine similarity of the term vectors for documents, D_j and D_k from Figure 4a. The HYBRID method exploits the vocabulary in the concept and document spaces to enhance the retrieval of documents.

5 Evaluation

Our methods are evaluated on a collection of topic-labeled learning resources by simulating an e-learning recommendation task. We use a collection from Microsoft Academic Search (MAS)[10], in which the author-defined keywords associated with each paper identifies the topics they contain. The keywords represent what relevance would mean in an e-learning domain and we exploit them for judging document relevance. The papers from MAS act as our e-learning resources, and using a query-by-example scenario, we evaluate the relevance of a retrieved document by considering the overlap of keywords with the query. This evaluation approach allows us to measure the ability of the proposed methods to identify relevant learning resources. The methods compared are:

- CONCEPTBASED represents documents using the domain concepts (Sect. 4.1).
- HYBRID augments the document representation using a contribution of term weights from the concept vocabulary (Sect. 4.2).
- BOW is a standard Information Retrieval method where documents are represented using the terms from the document space only with TF-IDF weighting.

For each of the 3 methods, the documents are first pre-processed by removing English stopwords and applying Porter stemming. Then, after representation, a similarity-based retrieval is employed using cosine similarity.

5.1 Evaluation Method

Evaluations using human evaluators are expensive, so we take advantage of the author-defined keywords for judging the relevance of a document. The keywords are used to define an overlap metric. Given a query document Q with a set of keywords K_Q, and a retrieved document R with its set of keywords K_R, the relevance of R to Q is based on the overlap of K_R with K_Q. The overlap is computed as:

$$Overlap(K_Q, K_R) = \frac{|K_Q \cap K_R|}{min\,(|K_Q|, |K_R|)} \tag{1}$$

We decide if a retrieval is relevant by setting an overlap threshold, and if the overlap between K_Q and K_R meets the threshold, then K_R is considered to be relevant.

Our dataset contains 217 Machine Learning and Data Mining papers, each being 2-32 pages in length. A distribution of the keywords per document is shown in Figure 5, where the documents are sorted based on the number of keywords they contain. There are 903 unique keywords, and 1497 keywords in total.

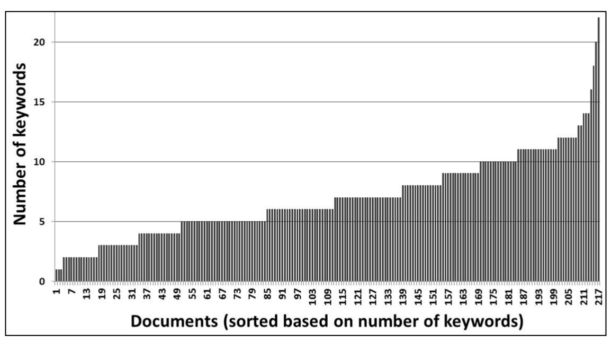


Fig. 5 Number of keywords per Microsoft document.

A summary of the overlap scores for all document pairs is shown in Table 2. There are 23436 entries for the 217 document pairs, and 20251 are zero, meaning that there is no overlap in 86% of the data. So only 14% of the data have an overlap of keywords, indicating that the distribution of keyword overlap is skewed. There are 10% of document pairs with overlap scores that are ≥ 0.14, while 5% are ≥ 0.25.

Table 2 Overlap of document-keywords and the proportion of data

Overlap Coefficient	**Number of Pairs**
Zero	20251 (86%)
Non-zero	3185 (14%)

Proportion of Data	**Overlap Threshold**
10%	0.14
5%	0.25
1%	0.5

The higher the overlap threshold, the more demanding is the relevance test. We use 0.14 and 0.25 as thresholds, thus avoiding the extreme values that would allow either very many or few of the documents to be considered as relevant. Our interest is in the topmost documents retrieved, because we want our top recommendations to be relevant. We use precision@n to determine the proportion of relevant documents retrieved:

$$Precision@n = \frac{|retrievedDocuments \cap relevantDocuments|}{n} \tag{2}$$

where, n is the number of documents retrieved each time, *retrievedDocuments* is the set of documents retrieved, and *relevantDocuments* are those documents that are considered to be relevant i.e. have an overlap that is greater than the threshold.

5.2 Results and Discussion

The methods are evaluated using a leave-one-out retrieval. In Figures 6, the number of recommendations (n) is shown on the x-axis and the average precision@n is shown on the y-axis. RANDOM (▲) has been included to give an idea of the relationship between the threshold and the precision values. RANDOM results are consistent with the relationship between the threshold and the proportion of data in Table 2.

Overall, HYBRID (■) performs better than BOW ($\times$) and CONCEPTBASED (●), showing that augmenting the representation of documents with a bigger, but focused vocabulary, as done in HYBRID, is a better way of harnessing our background knowledge. BOW also performs well because the document vocabulary is large, but the vocabulary used in CONCEPTBASED may be too limited. All the graphs fall as the number of recommendations, *n* increases. This is expected because the earlier retrievals are more likely to be relevant. However, the overlap of HYBRID and BOW at higher values of n may be because the documents retrieved by both methods are drawn from the same neighbourhoods.

The relative performance at a threshold of 0.25 in Figure 7, is similar to the performance at 0.14. However, at this more challenging threshold, HYBRID and BOW

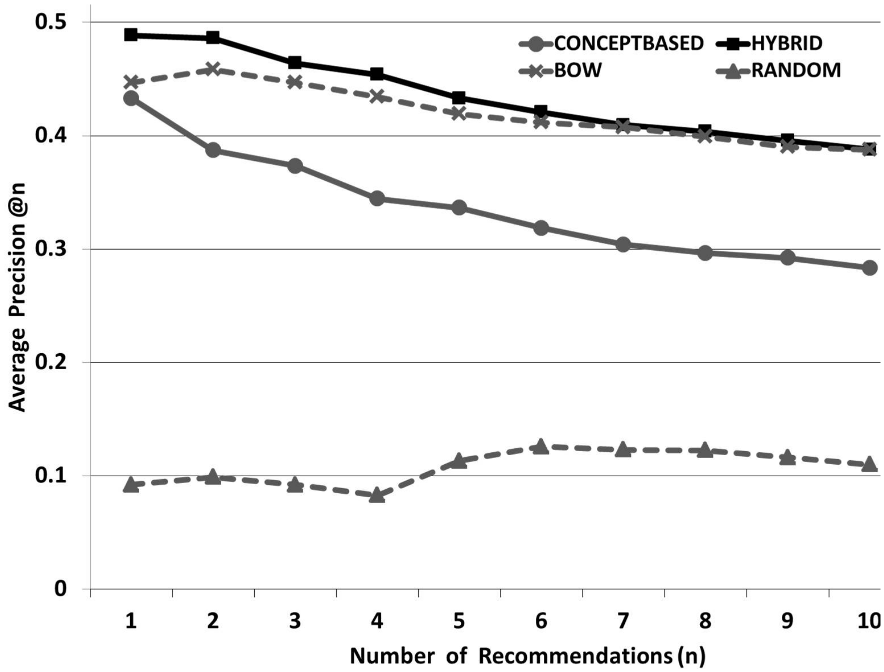


Fig. 6 Precision of the methods at an overlap threshold of 0.14

do not perform well on the first retrieval. This may be due to the size of the vocabulary used by both methods. Generally, the results show that the HYBRID method is able to identify relevant learning resources by highlighting the domain concepts they contain, and this is important in e-learning. The graphs show that augmenting the representation of learning resources with our background knowledge is beneficial for e-learning recommendation.

6 Conclusions

E-learning recommendation is challenging because the learning resources are often unstructured text, and so are not appropriately indexed for retrieval. One solution is the creation of a concept-aware representation that contains a good coverage of relevant topics. In this paper domain-specific background knowledge is built by exploiting a structured collection of teaching materials as a guide for identifying important concepts. We then enrich the identified concepts with discovered text from an encyclopedia source, and use these pseudo-documents to extend the coverage and richness of our representation.

The background knowledge captures both key topics highlighted by the e-book TOCs that are useful for teaching, and additional vocabulary related to these top-

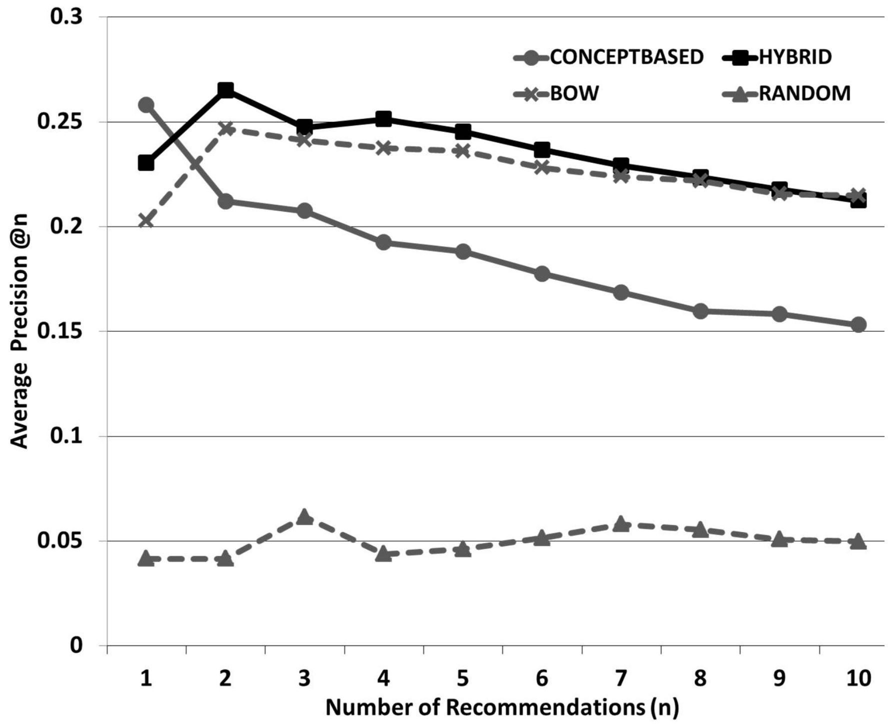


Fig. 7 Precision of the methods at an overlap threshold of 0.25.

ics. The concept space provides a vocabulary and focus that is based on teaching materials with provenance. CONCEPTBASED takes advantage of similar distributions of concept terms in the concept and document spaces to define a concept term driven representation. HYBRID exploits differences between distributions of document terms in the concept and document space, in order to boost the influence of terms that are distinctive in a few concepts.

Our results confirm that augmenting the representation of learning resources with our background knowledge in Hybrid improves e-learning recommendation. The larger vocabulary from both concepts and documents has been focused by the use of the vocabulary in the concept space. Although CONCEPTBASED also focuses on the concept space, by using only concept vocabulary, this vocabulary is too restricted for concept-based distinctiveness to be helpful.

In future, the background knowledge will be exploited to support query expansion and refinement in an e-learning environment. One approach would be to represent learners' queries using the vocabulary from our knowledge-rich representation. Alternatively, our background knowledge can be employed to support search by exploration. This would allow learners to search for resources through a guided view of the learning domain.

References

1. Agrawal, R., Chakraborty, S., Gollapudi, S., Kannan, A., Kenthapadi, K.: Quality of textbooks: An empirical study. In: ACM Symposium on Computing for Development, pp. 16:1–16:1 (2012)
2. Beliga, S., Meštrović, A., Martinčić-Ipšić, S.: An overview of graph-based keyword extraction methods and approaches. Journal of Information and Organizational Sciences **39**(1), 1–20 (2015)
3. Bousbahi, F., Chorfi, H.: MOOC-Rec: A case based recommender system for MOOCs. Procedia - Social and Behavioral Sciences **195**, 1813 – 1822 (2015)
4. Boyce, S., Pahl, C.: Developing domain ontologies for course content. Journal of Educational Technology & Society **10**(3), 275–288 (2007)
5. Chen, W., Niu, Z., Zhao, X., Li, Y.: A hybrid recommendation algorithm adapted in e-learning environments. World Wide Web **17**(2), 271–284 (2014)
6. Chen, Z., Liu, B.: Topic modeling using topics from many domains, lifelong learning and big data. In: 31st International Conference on Machine Learning, pp. 703–711 (2014)
7. Coenen, F., Leng, P., Sanderson, R., Wang, Y.J.: Statistical identification of key phrases for text classification. In: Machine Learning and Data Mining in Pattern Recognition, pp. 838–853. Springer (2007)
8. Dietze, S., Yu, H.Q., Giordano, D., Kaldoudi, E., Dovrolis, N., Taibi, D.: Linked education: Interlinking educational resources and the web of data. In: 27th Annual ACM Symposium on Applied Computing, pp. 366–371 (2012)
9. Ferrucci, D., Levas, A., Bagchi, S., Gondek, D., Mueller, E.T.: Watson: Beyond Jeopardy! Artificial Intelligence **199**, 93–105 (2013)
10. Hands, A.: Microsoft academic search Technical Services Quarterly **29**(3), 251–252 (2012)
11. Nasraoui, O., Zhuhadar, L.: Improving recall and precision of a personalized semantic search engine for e-learning. In: 4th International Conference on Digital Society, pp. 216–221. IEEE (2010)
12. Panagiotis, S., Ioannis, P., Christos, G., Achilles, K.: APLe: Agents for personalized learning in distance learning. In: 7th International Conference on Computer Supported Education, pp. 37–56. Springer (2016)
13. Porter, M.F.: An algorithm for suffix stripping. Program **14**(3), 130–137 (1980)
14. Qureshi, M.A., O'Riordan, C., Pasi, G.: Exploiting Wikipedia to identify domain-specific key terms/phrases from a short-text collection. In: 5th Italian Information Retrieval Workshop, pp. 63–74 (2014)
15. Ruiz-Iniesta, A., Jimenez-Diaz, G., Gomez-Albarran, M.: A semantically enriched context-aware OER recommendation strategy and its application to a computer science OER repository. IEEE Transactions on Education. **57**(4), 255–260 (2014)
16. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. Information Processing & Management **24**(5), 513 – 523 (1988)
17. Völkel, M., Krötzsch, M., Vrandecic, D., Haller, H., Studer, R.: Semantic Wikipedia. In: 15th International Conference on World Wide Web, pp. 585–594. ACM (2006)
18. Witten, I.H., Paynter, G.W., Frank, E., Gutwin, C., Nevill-Manning, C.G.: KEA: Practical automatic keyphrase extraction. In: 4th ACM Conference on Digital libraries, pp. 254–255 (1999)
19. Yang, H.L., Lai, C.Y.: Motivations of Wikipedia content contributors. Computers in Human Behavior **26**(6), 1377 – 1383 (2010)
20. Yang, K., Chen, Z., Cai, Y., Huang, D., Leung, H.: Improved automatic keyword extraction given more semantic knowledge. In: International Conference on Database Systems for Advanced Applications, pp. 112–125. Springer (2016)
21. Zhang, X., Liu, J., Cole, M.: Task topic knowledge vs. background domain knowledge: Impact of two types of knowledge on user search performance. In: Advances in Information Systems and Technologies, pp. 179–191. Springer (2013)

An e-Learning Recommender that Helps Learners Find the Right Materials

Blessing Mbipom and **Stewart Massie** and **Susan Craw**
School of Computing Science & Digital Media
Robert Gordon University
Aberdeen, UK

Abstract

Learning materials are increasingly available on the Web making them an excellent source of information for building e-Learning recommendation systems. However, learners often have difficulty finding the right materials to support their learning goals because they lack sufficient domain knowledge to craft effective queries that convey what they wish to learn. The unfamiliar vocabulary often used by domain experts creates a semantic gap between learners and experts, and also makes it difficult to map a learner's query to relevant learning materials. We build an e-Learning recommender system that uses background knowledge extracted from a collection of teaching materials and encyclopedia sources to support the refinement of learners' queries. Our approach allows us to bridge the gap between learners and teaching experts. We evaluate our method using a collection of realistic learner queries and a dataset of Machine Learning and Data Mining documents. Evaluation results show our method to outperform benchmark approaches and demonstrates its effectiveness in assisting learners to find the right materials.

Introduction

Learners often have difficulty asking an effective query of a search engine for two reasons. First, they lack sufficient knowledge about the domain they are researching, so are unable to assemble effective keywords that identify what they wish to learn (Liu, Kim, and Creel 2013). This problem results in an intent gap. Second, the vocabulary used by teaching experts is often different from that used by learners, as learners often describe their problems in different terms to how experts present the solutions (Millard et al. 2005). This presents a semantic gap.

Artificial intelligence (AI) methods have been applied to assist the teaching process in the design of an online course, by using AI agents to provide feedback to learners (Goel and Joyner 2016). AI techniques are also used in our method to assist the learning process by creating an e-Learning recommender system that provides learners with relevant documents. The developed method allows us to bridge both the intent and semantic gap between learners and domain experts. We address the intent gap by placing a learner's query within the space of learning concepts, and identifying the most similar concepts to use for refining the query. The semantic gap is addressed by leveraging the vocabulary associated with the domain concepts to support the refinement of queries. This allows us to refine the learner's query using the vocabulary of the domain. The effect is to focus the search on relevant documents and improve the recommendations.

We use background knowledge extracted automatically from a structured collection of teaching materials, to influence query refinement by providing a vocabulary for refining the queries. An e-Learning recommender system is built to evaluate the effectiveness of our approach using a collection of realistic learner queries and a dataset of Machine Learning and Data Mining resources. Evaluation results show our method to outperform a standard Bag of Words approach. The results demonstrate that using background knowledge to refine learners' queries supports the learning process by helping students to find relevant documents.

There are two key contributions from this work. First, an effective method for refining a learner's query to support the retrieval of relevant documents. Second, an e-Learning recommender system that assists the learning process and helps learners to find relevant learning materials.

Related Work

A large amount of e-Learning materials is available to learners on the Web. However, learners are often discouraged by the time spent in finding and assembling relevant resources to support their learning goals (Chen et al. 2014). Often, learners are new to the topic they are researching, so they can have difficulty asking effective queries in a search engine. The unfamiliar vocabulary often used in teaching materials poses a challenge to learners trying to find relevant materials. One way of addressing these challenges is by refining queries to improve the recommendation made to learners.

One approach to query refinement is by using internal knowledge from a document collection as a feedback method (Wu and Fang 2013). This approach is similar to pseudo relevance feedback. In this method, an initial set of documents considered to be relevant are found, then terms from these documents are used to refine the query to improve retrieval performance. A drawback of this approach is that search results may be directed towards a few documents, and this can be harmful if the documents are only about specific topics. Further, the retrieval performance for difficult queries can be affected if the initial retrieval set does not contain relevant documents (Li et al. 2007).

Another approach to query refinement involves using external knowledge sources for refining queries (Meij and de Rijke 2010; Meij et al. 2011). This approach entails using terms from domain sources to refine queries (Bendersky, Metzler, and Croft 2012). A knowledge source with a good coverage is usually recommended for this task. Sources such as Wikipedia (He and Ounis 2007; Xu, Jones, and Wang 2009), and DBpedia (Meij et al. 2009) have been used to identify potentially relevant terms to use for refining queries. The effectiveness of this approach has been demonstrated in previous work (Bendersky, Metzler, and Croft 2012). WatsonPaths applies a similar approach to reason over domain knowledge sources for answering queries in the medical domain (Lally et al. 2017). One potential challenge is query drift, where a refined query deviates from the original query (Xu, Jones, and Wang 2009). So, one needs to determine how much knowledge is sufficient for refining a query.

The approach in this paper draws insight from methods that use external knowledge sources. Our challenge is addressed in an e-Learning domain. So, the knowledge source used is drawn from learning concepts from the Tables of contents (TOCs) of e-Books and enriched with descriptive text from DBpedia abstracts. An initial set of similar learning concepts are automatically identified for each query. Highly weighted terms from the identified concepts then form the vocabulary that is used to refine a learner's query.

Background Knowledge

Background knowledge refers to specialized information about a domain that can be used for general understanding and problem solving (Zhang, Liu, and Cole 2013). We attempt to capture background knowledge as a set of domain concepts, each representing an important topic in the domain. For example, in a learning domain, such as Data Mining, you would find topics such as classification, association rules, and regression. Each of these topics is represented by a concept, in the form of a concept label and a pseudo-document which describes the concept.

We use the background knowledge that we developed in earlier work (Mbipom, Craw, and Massie 2016), to underpin e-Learning recommendation for a broad learning topic such as Machine Learning and Data Mining. In this work, the background knowledge is employed to support the refinement of queries. The background knowledge creation process is shown in Figure 1. The input to this process are domain knowledge sources, such as a structured collection of teaching materials and an encyclopedia source. Two knowledge sources are used as initial inputs for discovering domain concepts. First, the TOCs of 20 e-Books are used as a structured collection of teaching materials, which provide a source for extracting important topics identified by teaching experts in the domain. Second, a domain lexicon is used to verify that the concept labels identified from the teaching materials are directly relevant. The lexicon is created from Wikipedia because it contains articles for many learning domains (Zheng et al. 2010), and the contributions of many people (Yang and Lai 2010). The domain lexicon containing 664 Wiki-phrases provides a broader but more detailed coverage of the relevant topics in the domain. Thereafter, an encyclopedia source, such as DBpedia abstracts is searched and this provides the relevant text to form a pseudo-document for each verified concept label. The final output from this process is the background knowledge containing a set of 150 domain concepts each comprising a concept label and an associated pseudo-document.

The concept vocabulary with terms t_1 to t_c, from concepts, C_1 to C_m is used to create a concept term matrix with TF-IDF weighting (Salton and Buckley 1988). TF-IDF is useful for distinguishing concept terms in the concept space, and for identifying concepts that are relevant to queries hence its use in this method. A set of potentially useful concept terms are selected from the concept vocabulary as a means of scaling up the representation. So, we represent the background knowledge by using the top 10% of concept terms that have the highest TF-IDF values. The selected concept terms are used to create a concept term matrix. The selected terms t_{c1} to t_{cn}, from the concepts, C_1 to C_m are the set of potential terms that would be used for refining a query.

Refining Queries using Domain Concepts

The background knowledge representation is used to support the refinement of queries as a step towards addressing the intent and semantic gap learners face. When a new query is received from a learner, a search is performed on all the domain concepts. A ranked list of domain concepts that are similar to the query is retrieved. The terms from the term-vectors of the most similar concepts are put together to create a potential refined query. Terms with the highest weights are selected from the potential refined query and added to the initial query to create a refined query. The refined query can then be used to search on a document collection, and documents would be retrieved and presented to the learner. We expect the retrieved documents to be relevant because the query used for the search has been generated using domain concepts related to the initial query.

Figure 2 contains an illustration of how a refined query is generated. In this example, Cq_1, Cq_2, and Cq_k are the k most similar concepts to the query, while t_{c1} to t_{cn}, are the selected concept terms. The entries into the matrix are the tf-idf weights of the terms in the respective concepts. While, $SimScore_1$, $SimScore_2$ and $SimScore_k$ are the similarity scores between the query and concepts Cq_1, Cq_2, and Cq_k respectively. The weight of a concept term such as t_{c1} in the potential refined query is generated by computing the weighted sum for that term.

Equation 1 shows how the weight of term t_{c1} is computed. The weight t_{c1} is achieved by multiplying the weight $SimScore1$ with the tf-idf scores of terms that appear in concept Cq_1. This is also done for terms that appear in Cq_2 and Cq_k respectively. The column sum for t_{c1} is then computed. Altering the tf-idf weights of concept terms with the respective similarity scores allows terms from concepts that are more similar to the query to have more influence in the refined query. The output is a potential refined query with concept terms, t_{c1} to t_{cn} and their respective weights.

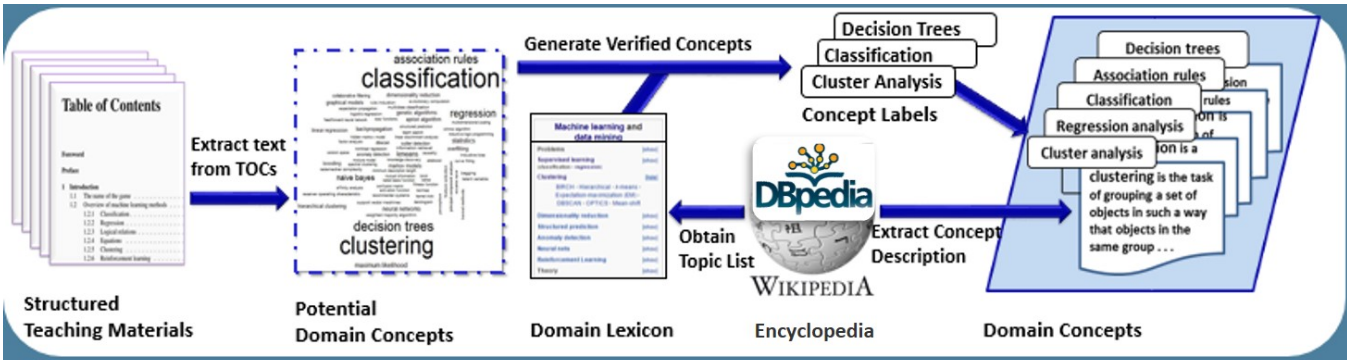


Figure 1: An overview of the background knowledge creation process

Terms / Concepts	t_{c1}	t_{c2}	. . .	t_{cn}	SimScore
C_{q1}	tf-idf(t_{c1}, C_{q1})				$SimScore_1$
C_{q2}	tf-idf(t_{c1}, C_{q2})				$SimScore_2$
C_{qk}	tf-idf(t_{c1}, C_{qk})				$SimScore_k$
Potential refined query	t_{c1}	t_{c2}	. . .	t_{cn}	

Figure 2: Generating a refined query

$$Weight(t_{c1}) = \sum_{i=1}^{k} (\text{tf-idf}(t_{c1}, C_{qi}) \times SimScore_i) \quad (1)$$

where t_{c1} is a concept term, and $\text{tf-idf}(t_{c1}, C_{qi})$ is the tf-idf score of term, t_{c1} in the i-th concept, C_{qi}, and $SimScore_i$ is the similarity between the query, q and the i-th concept C_{qi}.

The weight of a term in the potential refined query gives an indication of the importance of the term within the concept space in relation to the given query. We take advantage of this weight by selecting the highly weighted terms from the potential refined query. These terms are used for generating a refined query. We adopt this approach so that noisy terms would not be included during query refinement (Xu, Jones, and Wang 2009). We include the initial query as part of the refined query to maintain the context of the query.

For example, given an initial query: *"How do you implement gradient descent algorithm?"*. A search is performed on the set of concepts and the 3 most similar concepts to this query are: *stochastic gradient descent*, *backpropagation*, and *winnow algorithm*. The terms from these concepts are put together as described above. We use the top 25 highest weighted terms from an amalgamation of these concepts as the generated concept terms. In this example, the first 10 terms are: gradient, descent, stochastic, formula, update, momentum, delta, rate, derivative, backpropagation. The top 25 terms are then added to the initial query. So, the refined query becomes: *initial query* + *generated concept terms*. The refined query is then used to search for documents.

User Evaluation

An e-Learning recommender system is developed to demonstrate how refined queries can be used for e-Learning recommendation. Three methods are implemented. First our CONCEPTBASED query refinement method which uses the most similar domain concepts to create a concept based representation of a query. Second, the Bag-Of-Words (BOW) method, which is a standard Information Retrieval method, where a learner's query is represented using the terms in the query only. Finally, a HYBRID method which takes advantage of the features in a query to make a dynamic choice in determining when to apply CONCEPTBASED or BOW to refine a query. The evaluation aims to compare the relevance of documents retrieved using CONCEPTBASED and HYBRID methods against the standard BOW method.

Data

The data used for evaluation is drawn from 2 collections. First, a document collection for recommendation and second, a collection of learner queries. The document collection contains 504 chapters from 32 Machine Learning and Data Mining (ML/DM) e-Books. The collection used is fairly distributed across the domain concepts. The query collection contains realistic learner-focused queries which we use for evaluating the system. We used 2 sources to generate our queries. First, postgraduate students in the School of Computing Science and Digital Media took part as learners in generating queries. An e-mail specifying the task was sent to them. In order to allow learners to send anonymous responses, and return their queries without seeing what others had asked, a Google form was created to capture the queries. Second, online sources such as Coursera's Machine Learning MOOC and Quora were used to generate queries. Course specific questions were accessed from Coursera's MOOC, while the open questions in Quora from the Machine Learning and Data Mining topics were chosen.

For the query sources, our aim was to have realistic learner queries, so we used queries where the user wanted to learn about a technique, for example: "what are the various data mining techniques for fraud detection". We did not use generic or career-related queries such as "What is it like to be a data scientist at Amazon?", or queries that were out of scope such as “is there any course on ML?”. Overall, 11 queries are from learners and 59 queries from online sources, resulting in 70 queries.

The evaluation system was deployed using Microsoft Azure (Copeland et al. 2015), so the system could be accessible to users online. The evaluation system was available online for 8 weeks. To generate the recommendations for evaluation, the 70 queries were run using the methods, and the top 3 recommendations from each method was stored. A link to the system was shared with users working in the ML/DM field. Each user completed a questionnaire at the start to provide data about their background in the ML/DM domain. The data users provided gave us an idea of the experience and expertise of the users. An analysis of the evaluation results using the questionnaire data allowed us to gain valuable insights into the way different users judged the recommendations made by the system.

There were responses from 22 users. There were 16 PhD students, 3 Researchers, and 3 Lecturers/Professors. All users had at least an MSc degree or higher. There were 3 users with over 10 years experience in ML/DM, 3 users had over 5 years experience, 10 users had between 3-5 years experience, 5 users had 1-2 years experience, and only 1 user had less than a year’s experience in ML/DM. This level of experience in the ML/DM topic is useful, because the judgements should be from people who know the domain. For expertise, there were 2 experts, 16 competent users and only 4 beginners. So most users are competent or expert.

Design of the User Evaluation

At the start, each user was shown a briefing containing a guide on the evaluation study, the task, a note on confidentiality, and the researcher’s contact information. During the study, a user was shown one query each time to evaluate. For each query, the user could choose to skip, if the user had no idea about the query, or proceed to evaluate because the user had some understanding of the query. This allowed each user to evaluate recommendations for queries they were knowledgeable about. When evaluating a query, the user was shown up to 6 retrieved documents in random order. The set of documents were the top 3 documents from the ConceptBased and BOW methods. Since Hybrid applies either ConceptBased or BOW, the documents for Hybrid are already included in the retrieval set shown to users.

It is important that the retrieval set of documents shown to users is presented in a way that avoids any potential bias. Three issues of bias are considered and addressed. First, the users do not know which method produced the recommendations they are evaluating, this prevents a user from favouring one method over the others. Second, the order of documents presented to users is randomized. This rules out the bias of documents shown at the top being considered to be relevant over those lower down the list. Third, the same user evaluates recommendations from both BOW and ConceptBased methods for the same query. This ensures that the same user gives an evaluation for all methods at the same time for a given query. This prevents the possibility of receiving ratings from a positive user for one method, and ratings from a generally negative user for another method.

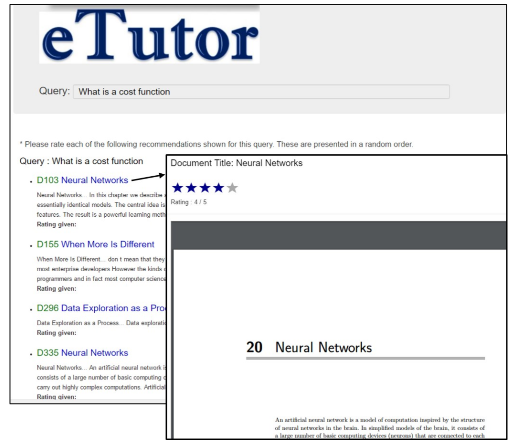


Figure 3: Recommendation screen

Figure 3 shows a page for displaying recommendations, and the screen for a document the learner has selected. The relevance of each document to the query is captured using a rating scale of 1 to 5 stars where 1 is least likely to be relevant and 5 stars is very relevant. The rating stars were included in the page that contained the document, so that each user had an opportunity of reading the document before rating it.

Evaluation Metrics

The evaluation uses the ratings given by users for query-recommendation pairs. We compute the $rating$ for a query-recommendation (q, r) pair as:

$$rating\,(q, r) = \sum_{u \in U_q} \frac{R_u(q, r)}{|U_q|} \qquad (2)$$

where U_q are users that evaluated query q, and R_u is the rating user u has given to a (q, r) pair. Performance of a method is computed by taking the average rating for the queries.

Recommendation Results

Users evaluated 105 queries and provided ratings for 521 query-recommendation (q, r) pairs. There were 6 of the total 70 queries that were not evaluated. Figure 4 shows the spread of ratings for (q, r) pairs that were evaluated as heat maps for the ConceptBased, Hybrid and BOW methods respectively. The ratings range from 5 to 1, for the highest rating of 5, to the lowest rating, 1. In plotting the heat map, the average rating values per (q, r) pair are sorted in descending order. The 3 heat maps are plotted in the same way using the actual average rating value given for each (q, r) pair. Lines are included to show a change in ratings.

In considering high ratings, Hybrid does best in producing documents with high quality ratings followed by ConceptBased and then BOW. Hybrid is able to correctly identify when to use either BOW or ConceptBased for refining a query in order to produce such good quality documents. For (q, r) pairs with the lowest ratings, the standard BOW method produces the highest number of documents with very poor ratings, BOW has difficulty pre-

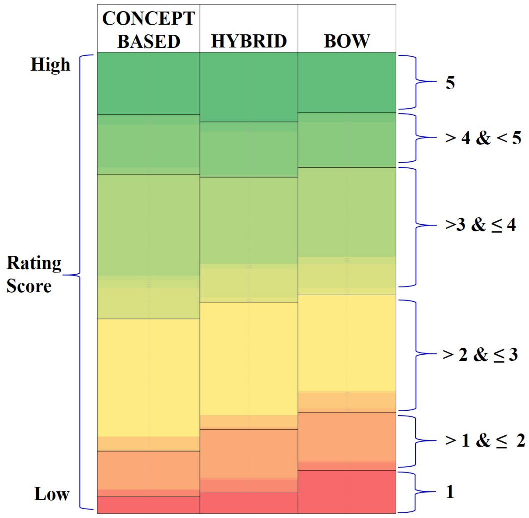


Figure 4: Spread of ratings for query-recommendation pairs

venting poor retrievals from being shown. CONCEPTBASED has the fewest number of (q, r) pairs with poor ratings. In particular CONCEPTBASED is very good at not presenting poor retrievals to users. These results for all the marked areas show that users gave higher ratings to the recommendations made using the HYBRID and CONCEPTBASED methods than those made using the standard BOW method.

We wish to know if a user's expertise affects the ratings they provided. This would allow us to confirm if there is some agreement among the users irrespective of their level of expertise. The rating for each (q, r) pair is computed based on Equation 2. Table 1 contains the average ratings of all users as well as ratings based on expertise of users for each of the 3 methods. For all the users, we find that CONCEPTBASED (CB) > HYBRID > BOW. So, using a CONCEPTBASED representation of a query to find learning materials is better than when HYBRID or BOW is used. For experts, the average rating scores for all methods are lower, nonetheless the experts still agree that the best performance is from the CONCEPTBASED method. We are confident in the results received from experts because they know what topics learners should be interested in. The competent users have higher rating scores across all methods, and they also agree that the CONCEPTBASED method performs better. Although the ratings by the beginners for all the methods are very similar, their rating scores also agree with the other users that the CONCEPTBASED method performs best.

Table 1: Average rating

Method	All	Expert	Competent	Beginner
CB	3.54	3.13	3.66	3.29
HYBRID	3.45	2.71	3.58	3.27
BOW	3.33	2.58	3.46	3.25

Relevance judgement is subjective and depends on the users who are providing ratings for the documents seen. We wanted to know how many users preferred the recommendations produced using either our CONCEPTBASED method (CB) or the standard BOW method. We use the rating as given in Equation 2 for BOW and CB and we count how many users rated documents from one method higher than the other. Table 2 contains the results for the preference users had for either CB or BOW. Half of the experts preferred CONCEPTBASED, while the other half of the experts thought both methods were the same. None of the experts thought that BOW was better than CONCEPTBASED. 14 users preferred the recommendations produced using the CONCEPTBASED method over those of the standard BOW method.

We trust the judgement of experts because they are more knowledgeable about the domain, and they know what documents should be relevant to learners. We can recall from Table 1 that the scores provided by the beginners for all methods were very similar, so we cannot rely totally on the judgements provided by beginners. Four users rated documents seen from both methods equally. This could come from users that gave equal scores to all documents seen for a query. Overall, majority of competent users preferred the documents from CONCEPTBASED over BOW.

Table 2: Preferences of methods

Preference	#	Demographics
CB $>$ BOW	14	1expert,12competent,1 beginner
BOW $>$ CB	4	1 competent, 3 beginners
CB $=$ BOW	4	1 expert, 3 competent

Coverage of Relevant Topics

Having recommendations with high ratings is good, but recommending documents that cover topics relevant to the query is important. We capture the coverage by asking users to provide feedback after they evaluate each query. This feedback is optional for users. Users were asked what extent they thought the documents they were shown covered the topics relevant to the query. The user could make a selection from 4 options: Complete, Good, Partial and Limited coverage. 50% of entries from users said the documents had good coverage. An additional 19% of entries said the documents had complete coverage, while 21% of entries said the documents had partial coverage. Only 10% of entries said the documents had limited coverage. So, most recommendations covered topics relevant to the queries.

Figure 5 contains a heat map which shows a broader view of the spread of coverage scores for the queries evaluated. The heat map is plotted by converting the coverage options to numeric values where Complete is 4, Good is 3, Partial is 2, and Limited is 1. The heat map is sorted twice. First based on the average coverage scores per query and second based on the average coverage scores per user. Figure 5 captures queries with the best coverage at the top, and those with least coverage at the bottom of the diagram.

The queries rated had consistent coverage scores from more than one user, so this allowed us to gain more insight to the queries. For example, a query such as: *"How does cluster analysis work?"* had "complete coverage" from 3 users

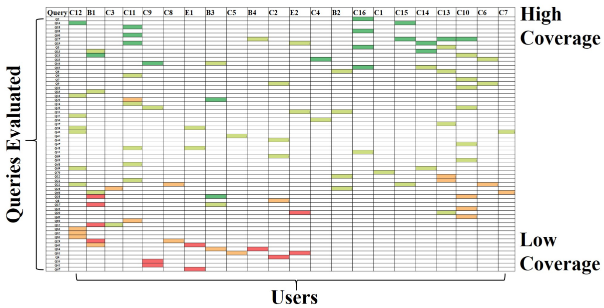


Figure 5: Spread of scores for the coverage

and "good coverage" from 1 user. While the query: *"Is it possible to use reinforcement learning to solve any supervised or unsupervised problem?"* had partial coverage from 1 user and limited coverage from 1 user. We draw the following insights from the coverage. First, some queries may not be well written and so can be hard to understand. Second, the topic contained in some queries may be very specialized. The results show potential in exploring the features of queries when refining them for e-Learning recommendation.

Conclusions

There are large amounts of e-Learning materials available to learners on the Web. However, learners often have difficulty finding relevant materials. Learners are often new to the topic they are researching, and so are unable to create effective queries in a search engine. We have created an e-Learning recommender system that uses background knowledge extracted from a collection of teaching materials to support the refinement of learners' queries. The rich vocabulary from the background knowledge is used to refine the queries and this allows us to focus the search on documents that are relevant to learners.

We use a collection of realistic queries and a dataset of Machine Learning and Data Mining documents for evaluation. Relevance judgement is subjective and depends on the opinions of the users taking part in the evaluation. However, we had a good level of consensus on the relevance judgements provided for each method by users with different levels of expertise. Results from experts, competent users and beginners all showed that using a CONCEPTBASED representation of a query to search produced documents that were consistently more relevant to learners than when the standard method was used. User evaluation results demonstrate the effectiveness of our approach to assist students to find the right learning materials. In future, the background knowledge can be developed to provide a guided view of a learning domain and support intelligent browsing of documents.

References

Bendersky, M.; Metzler, D.; and Croft, W. B. 2012. Effective query formulation with multiple information sources. In *5th ACM International Conference on WSDM*, 443–452.

Chen, W.; Niu, Z.; Zhao, X.; and Li, Y. 2014. A hybrid recommendation algorithm adapted in e-learning environments. *World Wide Web* 17(2):271–284.

Copeland, M.; Soh, J.; Puca, A.; Manning, M.; and Gollob, D. 2015. *Microsoft Azure: Planning, Deploying, and Managing Your Data Center in the Cloud*. Apress.

Goel, A. K., and Joyner, D. A. 2016. Design of an online course on Knowledge-Based AI. In *AAAI*, 4089–4094.

He, B., and Ounis, I. 2007. Combining fields for query expansion and adaptive query expansion. *Information Processing & Management* 43(5):1294–1307.

Lally, A.; Bagchi, S.; Barborak, M. A.; Buchanan, D. W.; Chu-Carroll, J.; Ferrucci, D. A.; Glass, M. R.; Kalyanpur, A.; Mueller, E. T.; Murdock, J. W.; Patwardhan, S.; Prager, J. M.; and Welty, C. A. 2017. WatsonPaths: Scenario-based question answering and inference over unstructured information. *AI Magazine* 38(2).

Li, Y.; Luk, W. P. R.; Ho, K. S. E.; and Chung, F. L. K. 2007. Improving weak ad-hoc queries using Wikipedia as external corpus. In *30th ACM SIGIR*, 797–798.

Liu, J.; Kim, C. S.; and Creel, C. 2013. Why do users feel search task difficult? *ASIS&T* 50(1):1–4.

Mbipom, B.; Craw, S.; and Massie, S. 2016. Harnessing background knowledge for e-learning recommendation. In *36th BCS SGAI*. Springer. 3–17.

Meij, E., and de Rijke, M. 2010. Supervised query modeling using Wikipedia. In *33rd ACM SIGIR*, 875–876.

Meij, E.; Bron, M.; Hollink, L.; Huurnink, B.; and De Rijke, M. 2009. Learning semantic query suggestions. In *The Semantic Web - ISWC*, 424–440. Springer.

Meij, E.; Bron, M.; Hollink, L.; Huurnink, B.; and de Rijke, M. 2011. Mapping queries to the linking open data cloud: A case study using DBpedia. *Web Semantics: Science, Services and Agents on the World Wide Web* 9(4):418–433.

Millard, D. E.; Gibbins, N. M.; Michaelides, D. T.; and Weal, M. J. 2005. Mind the semantic gap. In *16th ACM Conference on Hypertext and Hypermedia*, 54–62.

Salton, G., and Buckley, C. 1988. Term-weighting approaches in automatic text retrieval. *Information Processing & Management* 24(5):513 – 523.

Wu, H., and Fang, H. 2013. An incremental approach to efficient pseudo-relevance feedback. In *36th ACM SIGIR*, 553–562.

Xu, Y.; Jones, G. J.; and Wang, B. 2009. Query dependent pseudo-relevance feedback based on Wikipedia. In *32nd ACM SIGIR*, 59–66.

Yang, H.-L., and Lai, C.-Y. 2010. Motivations of Wikipedia content contributors. *Computers in Human Behavior* 26(6):1377–1383.

Zhang, X.; Liu, J.; and Cole, M. 2013. Task topic knowledge vs. background domain knowledge: Impact of two types of knowledge on user search performance. In *Advances in Information Systems and Technologies*. Springer. 179–191.

Zheng, Z.; Li, F.; Huang, M.; and Zhu, X. 2010. Learning to link entities with knowledge base. In *NAACL HLT*, 483–491.

Improving e-Learning Recommendation by using Background Knowledge

Blessing Mbipom Susan Craw
Stewart Massie
School of Computing Science & Digital Media
Robert Gordon University, Aberdeen, UK
b.e.mbipom/s.craw/s.massie@rgu.ac.uk

Abstract

There is currently a large amount of e-Learning resources available to learners on the Web. However, learners often have difficulty finding and retrieving relevant materials to support their learning goals because they lack the domain knowledge to craft effective queries that convey what they wish to learn. In addition, the unfamiliar vocabulary often used by domain experts makes it difficult to map a learner's query to a relevant learning material. We address these challenges by introducing an innovative method that automatically builds background knowledge for a learning domain. In creating our method, we exploit a structured collection of teaching materials as a guide for identifying the important domain concepts. We enrich the identified concepts with discovered text from an encyclopedia, thereby increasing the richness of our acquired knowledge. We employ the developed background knowledge for influencing the representation and retrieval of learning resources to improve e-Learning recommendation. The effectiveness of our method is evaluated using a collection of Machine Learning and Data Mining papers. Our method outperforms the benchmark, demonstrating the advantage of using background knowledge for improving the representation and recommendation of e-Learning materials.

1 Introduction

Learning-focused content is increasingly available on the Web, thus providing an excellent source of information for building e-Learning systems (Clarà and Barberà, 2013). However, learners often have difficulty finding the right learning materials because they lack the domain knowledge required to formulate effective queries (Chen et al., 2014). In addition, a mismatch in the vocabulary used by learners when crafting their queries and that used by domain experts to describe learning concepts poses a further challenge for systems recommending resources to learners.

Another challenge with e-Learning recommendation is that the learning resources are often unstructured text, and so are not properly indexed for retrieval (Nasraoui and Zhuhadar, 2010). The challenge of dealing with unstructured learning resources creates a difficulty in finding and retrieving relevant learning resources. Hence the need for an effective method of representing learning materials with the aim of improving recommendation.

This paper proposes the automated acquisition of background knowledge about a domain that can then be employed for enhancing e-Learning recommendation. In our method, we create a concept-aware representation that contains a good coverage of relevant topics from the domain. First, we exploit a structured collection of teaching materials as a guide for identifying the important concepts. Next, we enrich the identified concepts with discovered text from an encyclopedia source, thereby increasing the richness of our representation. Our developed method is demonstrated in Machine Learning and Data Mining, although the method we present can be applied to learning materials in other domains.

Other projects such as DeepQA (Ferrucci et al., 2013) and DBpedia (Lehmann et al., 2015) use a range of knowledge-rich representations to enhance retrieval. Such knowledge-rich sources are usually in the form of important topics that describe a domain. While these projects generally rely on handcrafted knowledge sources, they highlight the advantage in exploiting knowledge-rich representations as a basis for improving recommendation.

A good coverage of domain topics is useful for representing learning materials. These domain topics contain rich vocabulary and provide a good knowledge source for mapping learners' queries to learning materials. Thus allowing us to address the mismatch in the vocabulary used by learners and domain experts. We address this issue by introducing a method that automatically creates custom background knowledge in the form of a rich set of domain topics. Further, we explore building a richer vocabulary to achieve a better coverage of the domain, and this method is employed to improve e-Learning recommendation.

We make several contributions in this work. Firstly, the creation of background knowledge for an e-Learning domain. We describe how we take advantage of the knowledge of experts contained in e-Books to build a knowledge-rich representation that is used to enhance recommendation. Secondly, we present a method that harnesses the developed background knowledge to augment the representation of learning resources in order to improve e-Learning recommendation. Finally, we explore a larger concept vocabulary which provides a better coverage of the domain. We refine our method presented in (Mbipom et al., 2016) to generate a richer and focused set of domain concepts. The results from our evaluation show the improvement in e-Learning recommendation when the richer concept vocabulary is used for representing learning resources.

The rest of this paper is organised as follows. In Section 2 we present related text representation approaches that underpin this work. Section 3 describes the development of our background knowledge using available knowledge sources. Section 4 discusses the representation of learning resources using our methods. Then Section 5 presents the evaluation of the learning resource representation. In Section 6 we present our refined method of generating background knowledge with an evaluation using the richer vocabulary and a larger dataset for recommendation. Finally, Section 7 presents our conclusions.

2 Related Work

E-Learning recommendation is challenging because learning resources are often unstructured text, and so are not properly indexed for retrieval. A possible solution to addressing this challenge is the creation of effective representations that capture the content of learning resources. However, building suitable representations for learning resources in e-Learning environments is not easy (Dietze et al., 2012), as the resources do not have a pre-defined set of features by which they can be indexed.

We propose the creation of a knowledge-rich representation that captures the domain-specific vocabulary contained in learning resources. Figure 1 illustrates two broad approaches often used to address the challenge of text representation. These are corpus-based methods, such as topic models (Blei and McAuliffe, 2007; Chen and Liu, 2014); and structured representations, such as those that take advantage of ontologies (Boyce and Pahl, 2007; Yarandi et al., 2011). In Figure 1, the lower row of items identifies various knowledge sources that can be employed to build a range of knowledge-light to knowledge-rich text representation approaches.

Corpus-based methods usually involve the use of statistical models to identify topics from a corpus. The identified topics are often keywords (Beliga et al., 2015; Matsuo and Ishizuka, 2004) or phrases (Coenen et al., 2007; Witten et al., 1999). Coenen et al. showed that using a combination of keywords and phrases was better than using only keywords (Coenen et al., 2007). These topics can be extracted from different text sources such as: learning resources (Rodrigues et al., 2007; Yang et al., 2016), metadata e.g. Tables of contents (Bousbahi and Chorfi, 2015), and Encyclopedia e.g. Wikipedia (Milne and Witten, 2008; Qureshi et al., 2014). A drawback of the corpus-based methods is that, they normally rely on the coverage of the document collection used, so the topics produced may not be representative of the learning domain.

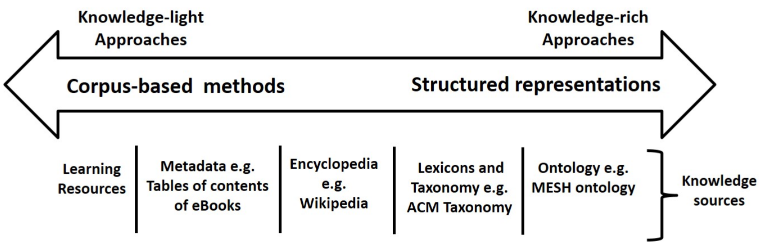


Figure 1: Two broad approaches used for text representation

Structured representations capture relationships between important domain concepts. This often entails using an existing ontology e.g ACM taxonomy (Nasraoui and Zhuhadar, 2010; Ruiz-Iniesta et al., 2014), or creating a new one (Gherasim et al., 2013; Panagiotis et al., 2016). Although ontologies are designed to have a good coverage of their domains, the output is still dependent on the view of its builders and, because of handcrafting, existing ontologies cannot easily be adapted to new domains. e-Learning is dynamic because new resources are becoming available regularly, and so using fixed ontologies limits the potential to incorporate new content.

The approach adopted in this paper draws insight from both the corpus-based methods and structured representations highlighted in Figure 1. We leverage on a structured corpus of teaching materials such as Tables of contents of e-Books, in order to identify important topics in an e-Learning domain. These topics are a combination of keywords and phrases as recommended in (Coenen et al., 2007). The identified topics are then enriched with discovered text from Wikipedia in order to enhance our representation. In addition, we refine the methods developed in previous work (Mbipom et al., 2016) so that we can generate a richer set of relevant topics that provide a good coverage of the learning domain. Consequently, our approach is employed to influence the representation and retrieval of relevant learning resources.

3 Creation of Background Knowledge

Background knowledge refers to information about a domain that is useful for general understanding and problem-solving (Zhang et al., 2013). We attempt to capture background knowledge as a set of domain concepts, each representing an important topic in the domain. For example, in a learning domain, such as Machine Learning, you would find topics such as Classification, Clustering and Regression. Each of these topics would be represented by a concept, in the form of a concept label and a pseudo-document which describes the concept. The concepts can then be used to underpin the representation of e-Learning resources.

Our knowledge extraction process is shown in Figure 2. The input to this process are domain knowledge sources, and we use a structured collection of teaching materials and an encyclopedia source. Next, ngrams are automatically extracted from our structured collection to generate a set of potential concept labels. Then a domain lexicon is used to validate the extracted ngrams to ensure that the ngrams are also being used in another information source. The encyclopedia provides text descriptions for the identified ngrams. The output from this process is a set of domain concepts, each having a concept label and an associated pseudo-document. We discuss the stages of the background knowledge creation in the following sections.

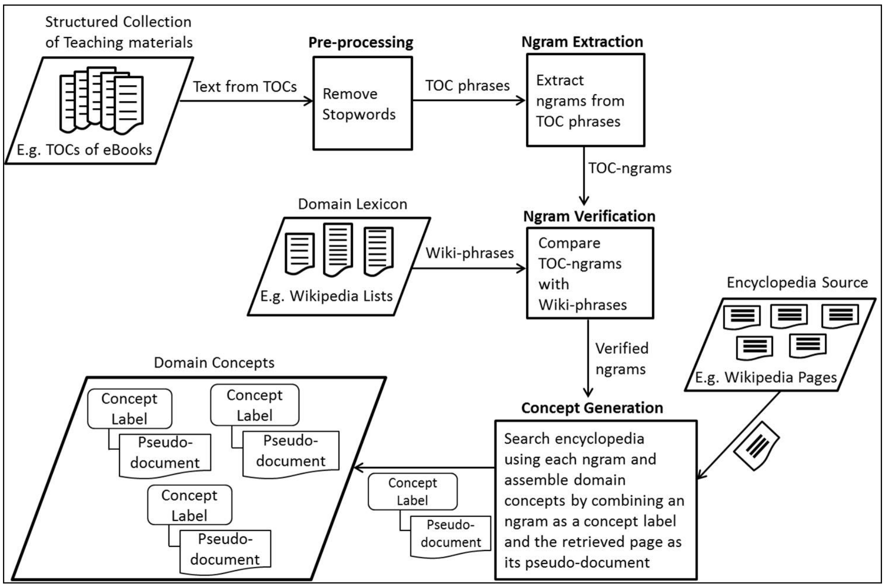


Figure 2: An overview of the background knowledge creation process

3.1 Knowledge Sources

Two knowledge sources are used as initial inputs for discovering concept labels. A structured collection of teaching materials provides a source for extracting important topics identified by teaching experts in the domain, while a domain lexicon provides a broader but more detailed coverage of the relevant topics in the domain. The lexicon is used to verify that the concept labels identified from the teaching materials are directly relevant. Thereafter, an encyclopedia source, such as Wikipedia pages, is searched and provides the relevant text to form a pseudo-document for each verified concept label. The final output from this process is our set of domain concepts each comprising a concept label and an associated pseudo-document.

Our approach is demonstrated with learning resources from Machine Learning and Data Mining. We use e-Books as our collection of teaching materials; a summary of the books used is shown in Table 1. Two Google Scholar queries: "Introduction to data mining textbook" and "Introduction to machine learning textbook" guided the selection process, and 20 e-Books that met all of the following 3 criteria were chosen. Firstly, the book should be about the domain. Secondly, there should be Google Scholar citations for the book. Thirdly, the book should be accessible. We use the Tables-of-Contents (TOCs) of the books as our structured knowledge source.

We use Wikipedia to create our domain lexicon because it contains articles for many learning domains (Völkel et al., 2006; Zheng et al., 2010), and the contributions of many people (Yang and Lai, 2010), so this provides the coverage we need in our lexicon. The lexicon is generated from 2 Wikipedia sources. First, the phrases in the *contents* and *overview* sections of the chosen domain are extracted to form a topic list. Then, a list with the titles of articles related to the domain is added to the topic list to assemble our lexicon. Overall, our domain lexicon contains a set of 664 Wiki-phrases.

Table 1: Summary of e-Books used

Book Title & Author	**Cites**
Machine learning; Mitchell	264
Introduction to machine learning; Alpaydin	2621
Machine learning a probabilistic perspective; Murphy	1059
Introduction to machine learning; Kodratoff	159
Gaussian processes for machine learning; Rasmussen & Williams	5365
Introduction to machine learning; Smola & Vishwanathan	38
Machine learning, neural and statistical classification; Michie, Spiegelhalter, & Taylor	2899
Introduction to machine learning; Nilsson	155
A First Encounter with Machine Learning; Welling	7
Bayesian reasoning and machine learning; Barber	271
Foundations of machine learning; Mohri, Rostamizadeh, & Talwalkar	197
Data mining-practical machine learning tools and techniques; Witten & Frank	27098
Data mining concepts models and techniques; Gorunescu	244
Web data mining; Liu	1596
An introduction to data mining; Larose	1371
Data mining concepts and techniques; Han & Kamber	22856
Introduction to data mining; Tan, Steinbach, & Kumar	6887
Principles of data mining; Bramer	402
Introduction to data mining for the life sciences; Sullivan	15
Data mining concepts methods and applications; Yin, Kaku, Tang, & Zhu	23

3.2 Generating Potential Domain Concepts

In the first stage of the process, the text from the TOCs is pre-processed. We remove punctuations, symbols, and numbers from the TOCs, so that only words are used for generating concept labels. After this, we remove 2 sets of stopwords. First, a standard English stopwords list, which allows us to remove common words and still retain a good set of words for generating our concept labels. Second, an additional set of words which we refer to as TOC-stopwords are removed. It contains: structural words, such as *chapter* and *appendix*, which relate to the structure of the TOCs; roman numerals, such as *xxiv* and *xxxv*, which are used to indicate the sections in a TOC; and words, such as *introduction* and *conclusion*, which describe parts of a learning material and are generic across domains. In addition, words referring directly to the name of the domain used for demonstration are removed, as we wish to generate concepts that describe the domain.

We do not use stemming because we found it harmful during pre-processing. When searching an encyclopedia source with the stemmed form of words, relevant results would not be returned. The output from pre-processing is a set of TOC phrases. In the next stage, we apply ngram extraction to the TOC phrases to generate all 1-3 grams from the entire set of TOC phrases. The output from this process are TOC-ngrams containing a set of 2038 unigrams, 5405 bigrams and 6133 trigrams, which are used as the potential domain concept labels. Many irrelevant ngrams are generated from the TOCs because we have simply selected all 1-3 grams.

3.3 Verifying Concept Labels using Domain Lexicon

A domain lexicon is used to verify the generated TOC-ngrams to confirm which of the ngrams are relevant for the domain. Our domain lexicon contains a set of 664 Wiki-phrases, each of which is pre-processed by removing non-alphanumeric characters. The distribution of Wiki-phrases is shown in Figure 3. The 84% of the Wiki-phrases that are 1-3 grams are used for verification. The comparison of TOC-ngrams with the domain lexicon identifies the potential domain concept labels that are actually being used to describe aspects of the chosen domain in Wikipedia. During verification, ngrams referring directly to the title of the domain, e.g. *machine learning* and *data mining*, are not included in the Wiki-phrases because our aim is to generate concept labels that describe specific topics within the domain. Overall, a set of 17 unigrams, 58 bigrams and 15 trigrams are verified as potential concept labels. Bigrams yield the highest number of ngrams, which indicates that bigrams are particularly useful for describing topics in this domain.

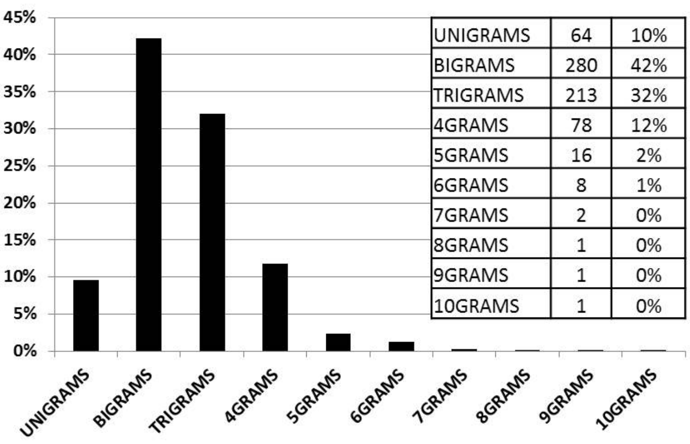


Figure 3: Distribution of Wiki-phrases used for verifying concept labels

3.4 Domain Concept Generation

Our domain concepts are generated after a second verification step is applied to the ngrams returned from the previous stage. Each ngram is retained as a concept label if all of 3 criteria are met. Firstly, if a Wikipedia page describing the ngram exists. Secondly, if the text describing the ngram is not contained as part of the page describing another ngram. Thirdly, if the ngram is not a synonym of another ngram. For the third criteria, if two ngrams are synonyms, the ngram with the higher frequency is retained as a concept label while its synonym is retained as part of the extracted text. For example, 2 ngrams *cluster analysis* and *clustering* are regarded as synonyms in Wikipedia, so the text associated with them is the same. The label *clustering* is retained as the concept label because it occurs more frequently in the TOCs, and its synonym, *cluster analysis* is contained as part of the discovered text.

The concept labels are used to search Wikipedia pages in order to generate a description for the identified concept label. The search returns discovered text that forms a pseudo-document which includes the concept label. So, the concept label and pseudo-document pair make up a domain concept. Overall, 73 domain concepts are generated. Each pseudo-document is pre-processed using standard techniques of English stopwords removal and Porter stemming (Porter, 1980). The pseudo-document terms form the concept vocabulary that can be used to represent resources.

4 Representing Learning Resources Using Background Knowledge

Our background knowledge contains a rich representation of the learning domain and by harnessing this knowledge for representing learning resources, we expect to retrieve documents based on the domain concepts that they contain. These concepts are designed to be effective for e-Learning, because they are assembled from TOCs of teaching materials (Agrawal et al., 2012). We present two approaches which have been developed by employing our background knowledge in the representation of learning resources.

4.1 The CONCEPTBASED Document Representation approach

Representing documents with the concept vocabulary allows retrieval to focus on the concepts contained in the documents. Figures 4 & 5 illustrate the CONCEPTBASED method. Firstly, in Figure 4, the concept vocabulary, $t_1 \dots t_c$, from the pseudo-documents of concepts, $C_1 \dots C_m$, is used to create a term-concept matrix and a term-document matrix using TF-IDF weighting (Salton and Buckley, 1988). In Figure 4a, c_{ij} is the TF-IDF of term t_i in concept C_j, while Figure 4b shows d_{ik} which is the TF-IDF of t_i in D_k.

	C_1	...	C_j	...	C_m
t_1					
...					
t_i			c_{ij}		
...					
t_C					

(a) Term-concept matrix

	D_1	...	D_k	...	D_n
t_1					
...					
t_i			d_{ik}		
...					
t_C					

(b) Term-document matrix

Figure 4: Term matrices for concepts and documents

Next, documents $D_1 \dots D_n$ are represented with respect to concepts by computing the cosine similarity of the term vectors for concepts and documents. The output is the concept-document matrix shown in Figure 5a, where y_{jk} is the cosine similarity of the vertical shaded term vectors for C_j and D_k from Figures 4a and 4b respectively. Finally, the document similarity is generated by computing the cosine similarity of concept-vectors for documents. Figure 5b shows z_{km}, which is the cosine similarity of the concept-vectors for D_k and D_m from Figure 5a. So, the CONCEPTBASED approach uses the document representation and similarity in Figure 5 to influence retrieval. We expect to retrieve documents that are similar based on the domain concepts that they contain.

	D_1	...	D_k	...	D_m	...	D_n
C_1							
...							
C_j			y_{jk}		y_{jm}		
...							
C_m							

(a) Concept-document matrix representation

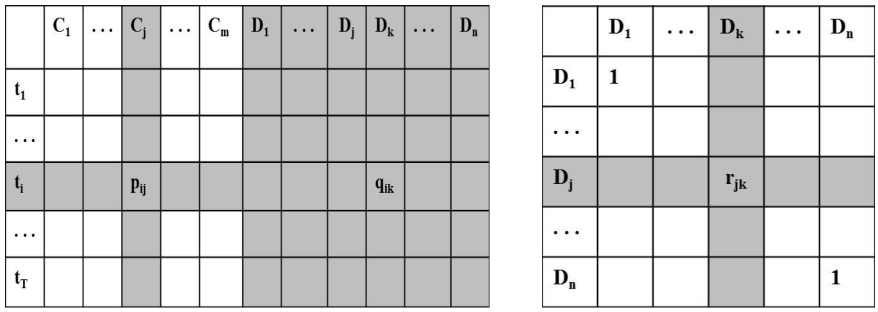
	D_1	...	D_m	...	D_n
D_1	1				
...					
D_k			z_{km}		
...					
D_n					1

(b) Document-document similarity

Figure 5: Document representation and similarity using the CONCEPTBASED approach

4.2 The HYBRID Document Representation Approach

The HYBRID approach exploits the relative distribution of the vocabulary in the concept and document spaces to augment the representation of learning resources with a bigger, but focused, vocabulary as shown in Figure 6. So the TF-IDF weight of a term changes depending on its relative frequency in both spaces. First, our 73 domain concepts, $C_1 \dots C_m$ from section 3.4, and the documents we wish to represent, $D_1 \dots D_n$, are merged to form a corpus. Next, a term-document matrix with TF-IDF weighting is created using all the terms, $t_1 \dots t_T$ from the vocabulary of the merged corpus as shown in Figure 6a. Entry q_{ik} is the TF-IDF weight of term t_i in D_k. If t_i has a lower relative frequency in the concept space compared to the document space, then the weight q_{ik} is boosted. So, distinctive terms from the concept space will get boosted. Although the overlap of terms from both spaces are useful for altering the term weights, it is valuable to keep all the terms from the document space because this gives us a richer vocabulary. The shaded term vectors for $D_1 \ldots D_n$ in Figure 6a form a term-document matrix for documents whose term weights have been influenced by the presence of terms from the concept vocabulary.



(a) Hybrid term-document matrix representation

(b) Hybrid document similarity

Figure 6: Representation and similarity of documents using the HYBRID approach

Finally, the document similarity in Figure 6b, is generated by computing the cosine similarity between the augmented term vectors for $D_1 \ldots D_n$. Entry r_{jk} is the cosine similarity of the term vectors for documents, D_j and D_k from Figure 6a. The HYBRID method exploits the vocabulary in the concept and document spaces to influence the retrieval of documents.

5 Evaluating Learning Resource Representation

Our methods are evaluated on a collection of topic-labelled learning resources by simulating an e-Learning recommendation task. We use a collection from Microsoft Academic Search (MAS)(Hands, 2012), in which the author-defined keywords associated with each paper identifies the topics they contain. The keywords represent what relevance would mean in an e-Learning domain and we exploit them for judging document relevance. The papers from MAS act as our e-Learning resources, and using a query-by-example scenario, we evaluate the relevance of a retrieved document by considering the overlap of keywords with the query. This evaluation approach allows us to measure the ability of the methods to identify relevant learning resources.

We compare the performance of our CONCEPTBASED and HYBRID methods against that of Bag of Words (BOW). The BOW is a standard Information Retrieval method where documents are represented using terms from the document space only with TF-IDF weighting. For each of the 3 methods, the documents are first pre-processed by removing English stopwords and applying Porter stemming. Then, after representation, a similarity-based retrieval is employed using cosine similarity.

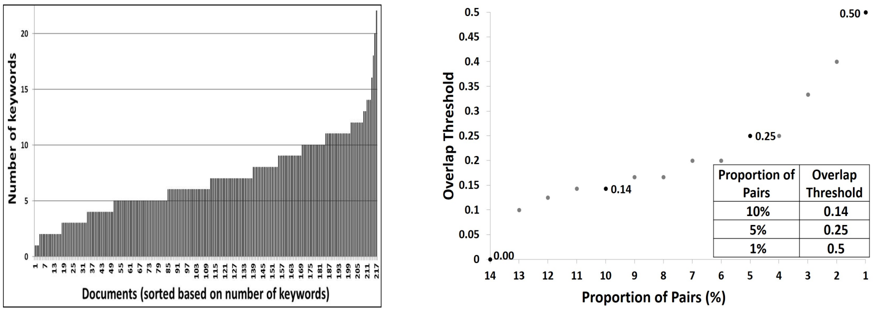
5.1 Evaluation Method and Dataset

Evaluations using human evaluators are expensive, so we take advantage of the author-defined keywords for judging the relevance of a document. The keywords are used to define an overlap metric. Given a query document Q with a set of keywords K_Q, and a retrieved document R with its set of keywords K_R, the relevance of R to Q is based on the overlap of K_R with K_Q. The overlap is computed as:

$$Overlap(K_Q, K_R) = \frac{|K_Q \cap K_R|}{min\left(|K_Q|, |K_R|\right)} \quad (1)$$

We decide if a retrieval is relevant by setting an overlap threshold, and if the overlap between K_Q and K_R meets the threshold, then K_R is considered to be relevant.

Figure 7 shows the number of keywords per document and the overlap of document pairs for the first dataset used. Our first dataset which we refer to as dataset 1 contains 217 Machine Learning and Data Mining papers. A distribution of the keywords per document is shown in Figure 7a, where the documents are sorted based on the number of keywords they contain. There are 903 unique keywords, and 1,497 keywords in total. A summary of the overlap scores for all document pairs is shown in Figure 7b. There are $23,436$ entries for the 217 document pairs, and $20,251$ are zero, meaning that there is no overlap in 86% of the data. So only 14% of the data have an overlap of keywords, indicating that the distribution of keyword overlap is skewed. There are 10% of document pairs with overlap scores ≥ 0.14, and 5% are ≥ 0.25. For experiments with this dataset we use 0.14 and 0.25 as thresholds, thus avoiding extreme values that would allow either very many or few of the documents to be considered as relevant.



(a) # of keywords per Document (b) Overlap of document pairs

Figure 7: Number of keywords per document and overlap profile of document pairs in dataset 1

Our interest is in the topmost documents retrieved, because we want our top recommendations to be relevant. We use precision@n to determine the proportion of relevant documents retrieved:

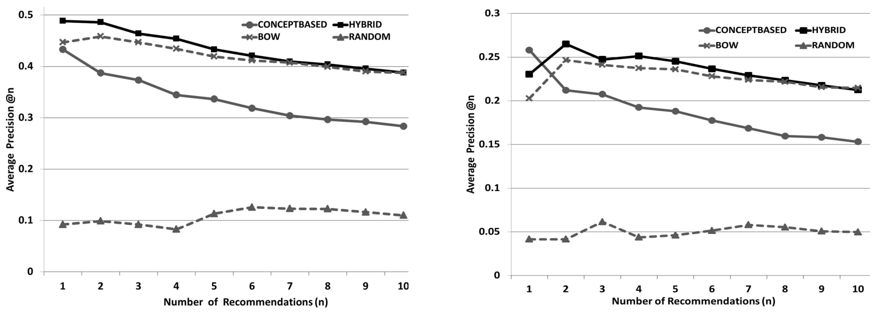
$$Precision@n = \frac{|retrievedDocuments \cap relevantDocuments|}{n} \tag{2}$$

where, n is the number of documents retrieved each time, *retrievedDocuments* is the set of documents retrieved, and *relevantDocuments* are those documents that are considered to be relevant i.e. have an overlap that is greater than the threshold.

5.2 Evaluation Results

The methods are evaluated using a leave-one-out retrieval. In Figure 8, the number of recommendations (n) is shown on the x-axis and the average precision@n is shown on the y-axis. RANDOM(▲) has been included to give an idea of the relationship between the threshold and the precision values. RANDOM results are consistent with the relationship between the threshold and the proportion of data in Figure 7b.

Overall, HYBRID(■) performs better than BOW(×) and CONCEPTBASED(●), showing that augmenting the representation of documents with a bigger, but focused vocabulary, as done in HYBRID, is a better way of harnessing our background knowledge. BOW also performs well because the document vocabulary is large, but the vocabulary used in CONCEPTBASED may be too limited. The complexity of the representation method in HYBRID overcomes the limitation of CONCEPTBASED. All the graphs fall as the number of recommendations, *n* increases. This is expected because the earlier retrievals are more likely to be relevant. However, the overlap of HYBRID and BOW at higher values of n may be because the documents retrieved by both methods are drawn from the same neighbourhoods.



(a) Results at Threshold of 0.14 (b) Results at Threshold of 0.25

Figure 8: Precision of the methods at overlap thresholds of 0.14 and 0.25 on dataset 1

The relative performance at a threshold of 0.25 in Figure 8b, is similar to the performance at 0.14. However, at this more challenging threshold, HYBRID and BOW do not perform well on the first retrieval. Generally, the results show that the HYBRID method is able to identify relevant learning resources by highlighting the domain concepts they contain, and this is important in e-Learning. The graphs show that augmenting the representation of learning resources with our background knowledge is beneficial for e-Learning recommendation.

6 Refined Background Knowledge

One issue with the previous concept generation method is that the concept vocabulary produced was limited. A suitable representation for e-Learning resources should have a good coverage of relevant domain topics. In this section, we discuss the steps taken to refine our method used for generating domain concepts in order to improve our background knowledge and increase the coverage of our concept vocabulary.

6.1 Enriched Domain Concepts

In developing this method, we go through the phases described in sections 3.2 - 3.4. First, in addition to the TOC stopwords, the SMART stopwords (Salton, 1971) are also removed during pre-processing. This allows us to remove words that do not contribute to learning terms, and still retain a good set of words for generating our concepts. Second, words referring to the name of the domain used for demonstration such as: *machine*, *learning*, *data*, and *mining* are not removed during pre-processing, as we observed that removing these words before ngram generation prevents other relevant ngrams such as *instance based learning* or *reinforcement learning*, that contain any of these words from being identified. Third, we increase our ngram extraction to generate 1-5 grams from our TOC-phrases because, a distribution of the Wiki-phrases in Figure 3 showed that 99% of phrases are 1-5grams; this allows us to increase the number of concepts we can generate.

We apply ngram extraction to the TOC-phrases to produce the following TOC-ngrams: 2467 Unigrams; 5387 Bigrams; 3625 Trigrams; 1668 Fourgrams; and 576 Fivegrams. The TOC-ngrams are verified as described in Section 3.3 using the Wiki-phrases to produce a set of potential concept labels containing 24 Unigrams; 96 Bigrams; 38 Trigrams; 6 Fourgrams; and no Fivegrams. A second verification step as described in Section 3.4 is applied to the potential concept labels. This entails using the verified ngrams to search Wikipedia pages in order to generate a domain concept. The search returns discovered text that forms a pseudo-document and a concept label. Overall, our refined method has 150 domain concepts that pass the second verification, each having a concept label and pseudo-document pair. The pseudo-document terms are pre-processed using standard techniques of English stopword removal and Porter Stemming. These terms now form the concept vocabulary of our refined background knowledge which we refer to as the CONCEPTBASED+ method.

6.2 Recommendation using the CONCEPTBASED+ approach

The CONCEPTBASED+ method employs the richer concept vocabulary of our refined background knowledge for representing documents. We expect the representation created using the CONCEPT-BASED+ method to contain a better coverage of the learning domain because of the richer concepts it contains. Our aim is to address the issue of the limited concepts contained in the CONCEPTBASED method. For recommendation using CONCEPTBASED+, we use the same representation and document similarity as the CONCEPTBASED method illustrated in Figures 4 & 5, but with a richer concept vocabulary. So documents are represented with respect to concepts by computing the cosine similarity of term vectors for concepts and documents to produce a concept document matrix. Then, the similarity between documents can be generated by computing the similarity between respective concept vectors for documents.

By using the CONCEPTBASED+ method for representation, we expect to retrieve documents that are similar based on the concepts they contain, and this is obtained from a document-document similarity matrix as shown in Figure 5b. A standard approach of representing documents would be to define the document similarity based on the term document matrix illustrated in Figure 4b, but this exploits the concept vocabulary only. In our approach, we put more emphasis on the domain concepts, so we use the concept document matrix illustrated in Figure 5a, to underpin the similarity between documents. The CONCEPTBASED+ method combines the focus with breadth of a richer set of domain concepts when representing documents.

6.3 Evaluating the Refined Representation

This section investigates whether the domain concepts generated using a refined approach i.e. CONCEPTBASED+ are better for representing documents than concepts generated with a standard method i.e.CONCEPTBASED. The same evaluation method and dataset 1 presented in Section 5.1 is adopted here, and a leave-one-out retrieval is applied for evaluating the methods. In Figure 9, the number of recommendations is shown on the x-axis while the average precision@n is shown on the y-axis. An overlap threshold of 0.14 is used because there are 10% of document pairs in this dataset with overlap scores ≥ 0.14.

The performance of CONCEPTBASED+(♦) is shown by the darker line, and CONCEPTBASED(●) by the gray line. BOW($\times$) is included as the benchmark and RANDOM(▲) gives an idea of the relationship between the threshold used and the precision values. The graphs of all the methods fall as the number of recommendations, n increases. This is expected as earlier retrievals are more likely to be relevant. Overall, CONCEPTBASED+ outperforms CONCEPTBASED, BOW, and RANDOM, by producing better recommendations for all values of n. This performance shows the advantage of using the richer concept vocabulary for representing learning materials. The results confirm that CONCEPTBASED+ contains concepts that have a better coverage of the learning domain than CONCEPTBASED which has a limited set of concepts. So we adopt CONCEPTBASED+ as a background knowledge representation for learning materials in this domain.

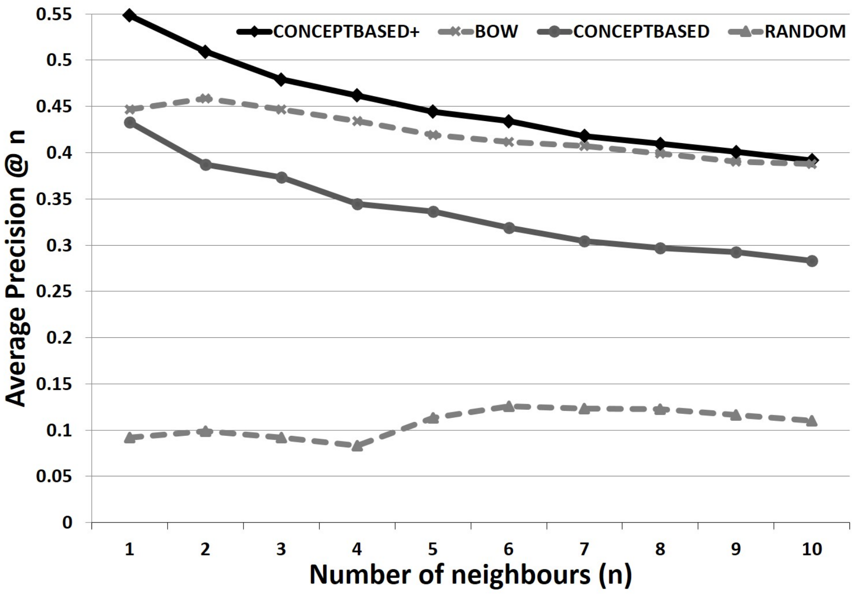
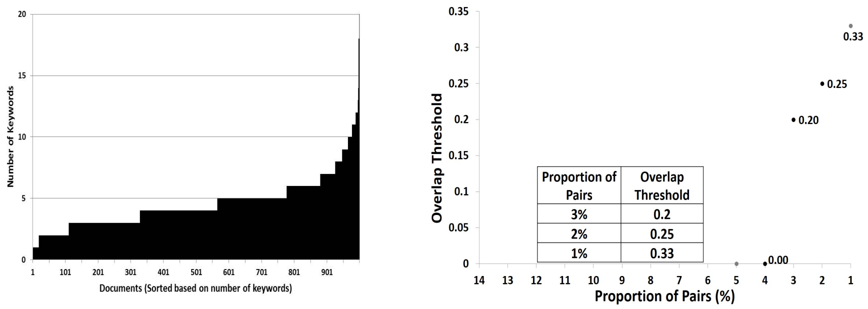


Figure 9: Comparing CONCEPTBASED and CONCEPTBASED+ at a threshold of 0.14 on dataset 1

6.4 Evaluation Using a Larger Dataset

We compare the performance of our HYBRID and CONCEPTBASED+ methods against that of the standard BOW approach on a larger dataset, in order to confirm our findings from the previous experiments. Figure 10 contains the number of keywords per document and the overlap of document pairs for the second dataset used. Our second dataset which we refer to as dataset 2 contains 1000 Machine Learning and Data Mining papers also from Microsoft Academic Research. Figure 10a contains a distribution of the keywords per document, where the documents are sorted based on the number of keywords they contain. There are 3063 unique keywords, and 4551 keywords in total. We take advantage of these author-defined keywords for judging relevance. A summary of the overlap profile of document pairs for dataset 2 is shown in Figure 10b. There are $499,500$ entries for the 1000 document pairs, and $480,129$ entries are zero, meaning that there is no overlap in 96% of the data. So only 4% of the data have an overlap of keywords, indicating that the distribution of keyword overlap is skewed. There are 3% of document pairs with overlap scores ≥ 0.2. The same evaluation method presented in 5.1 is used here. Then a leave-one-out retrieval method is applied, and precision@n as given in Equation 2 is used to determine the proportion of relevant documents retrieved. With dataset 2, we use a threshold of 0.2 thus preventing values that allow either too many or few documents to be considered as relevant. In Figure 11, the number of recommendations is shown on the x-axis and the average precision@n is on the y-axis. The average precision values are based on the overlap of keywords between document pairs and the threshold value chosen for the experiment. RANDOM(▲) gives an idea of the relationship between the threshold and the precision values, and the results are consistent with the overlap profile in Figure 10b.



(a) # of keywords per document (b) Overlap of document pairs

Figure 10: Number of keywords per document and overlap profile of document pairs in dataset 2

On this bigger dataset, CONCEPTBASED+(♦) method outperforms HYBRID(■), BOW(×), and CONCEPTBASED(●), confirming that using a richer and focused vocabulary to represent documents is useful for e-Learning recommendation. The results also show HYBRID performing better than BOW, again confirming that augmenting the representation of learning resources with domain concepts is better than using the content only for e-Learning recommendation. Experiments were also run at thresholds of 0.25 and 0.33 and the relative performance at these thresholds is similar to the performance at 0.2, so the graphs are not shown. Our results show that we are able to leverage on the vocabulary from CONCEPTBASED+ which is not only a larger vocabulary, but one focused on domain concepts, thus allowing our method to influence the retrieval and recommendation of relevant learning resources.

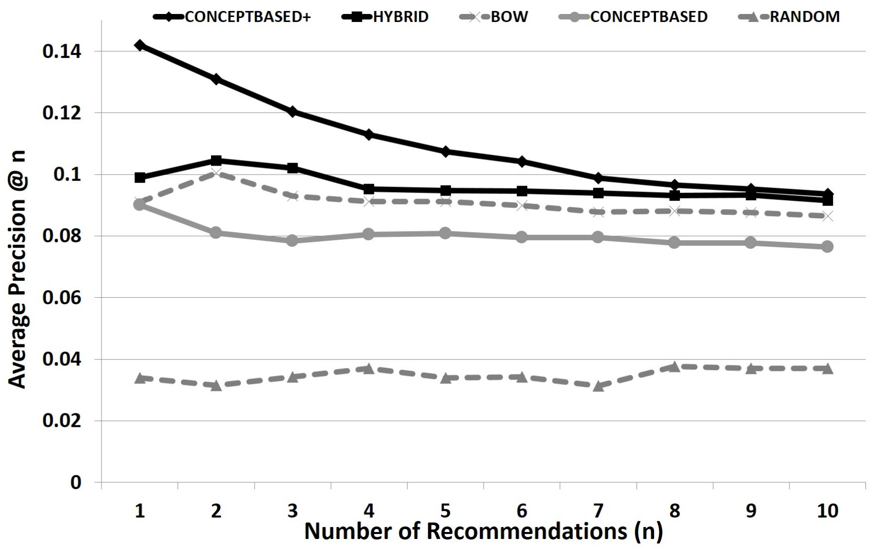


Figure 11: Precision of the methods at overlap threshold of 0.2 on dataset 2

7 Conclusions

The growing availability of e-Learning materials on the Web provides opportunities for learners to easily access new and valuable information. However, finding good materials is difficult because retrieval has to overcome the challenge of ineffective queries often input by learners. e-Learning recommendation offers a possible solution to this difficulty. Though, recommendation in e-Learning environments is challenging because the learning materials are often unstructured text, and so are not properly indexed for retrieval. We address this challenge by creating a method that automatically acquires background knowledge in the form of a rich set of concepts related to the selected learning domain. In building our method, we take advantage of the knowledge of experts contained in the TOCs of e-Books to identify relevant domain topics. By using e-Books we benefit from the provenance associated with these teaching materials. The identified topics are enriched with discovered text from Wikipedia, and this extends the coverage and richness of our representation.

CONCEPTBASED method takes advantage of similar distributions of concept terms in the concept and document spaces to define a concept-term driven representation. Although the concept vocabulary in CONCEPTBASED is limited, HYBRID exploits the relative distribution of the vocabulary in the concept and document spaces to augment the representation of learning resources with a larger vocabulary influenced by domain concepts. CONCEPTBASED+ provides a richer concept vocabulary that allows concept-based distinctiveness to be helpful in the representation and retrieval of documents. This refined method allows us to generate a richer and focused set of domain concepts, which provides a better coverage of the domain. The performance of CONCEPTBASED+ in our evaluation shows the advantage of using the richer concept vocabulary for representing learning materials. Our results confirm an improvement in e-Learning recommendation when a rich concept vocabulary is used for representing learning resources.

References

Agrawal, R., Chakraborty, S., Gollapudi, S., Kannan, A., and Kenthapadi, K. (2012). Quality of textbooks: An empirical study. In *ACM Symposium on Computing for Development*, pages 16:1–16:1. doi: 10.1145/2160601.2160623.

Beliga, S., Meštrović, A., and Martinčić-Ipšić, S. (2015). An overview of graph-based keyword extraction methods and approaches. *Journal of Information and Organizational Sciences*, 39(1):1–20.

Blei, D. M. and McAuliffe, J. D. (2007). Supervised topic models. In *Neural Information Processing Systems*, pages 121–128.

Bousbahi, F. and Chorfi, H. (2015). MOOC-Rec: A case based recommender system for MOOCs. *Procedia - Social and Behavioral Sciences*, 195:1813 – 1822. doi: 10.1016/j.sbspro.2015.06.395.

Boyce, S. and Pahl, C. (2007). Developing domain ontologies for course content. *Journal of Educational Technology & Society*, 10(3):275–288.

Chen, W., Niu, Z., Zhao, X., and Li, Y. (2014). A hybrid recommendation algorithm adapted in e-learning environments. *World Wide Web*, 17(2):271–284. doi: 10.1007/s11280-012-0187-z.

Chen, Z. and Liu, B. (2014). Topic modeling using topics from many domains, lifelong learning and big data. In *31st International Conference on Machine Learning*, pages 703–711.

Clarà, M. and Barberà, E. (2013). Learning online: Massive Open Online Courses (MOOCs), connectivism, and cultural psychology. *Distance Education*, 34(1):129–136. doi: 10.1080/01587919.2013.770428.

Coenen, F., Leng, P., Sanderson, R., and Wang, Y. J. (2007). Statistical identification of key phrases for text classification. In *Machine Learning and Data Mining in Pattern Recognition*, pages 838–853. Springer. doi: 10.1007/978-3-540-73499-4_63.

Dietze, S., Yu, H. Q., Giordano, D., Kaldoudi, E., Dovrolis, N., and Taibi, D. (2012). Linked education: Interlinking educational resources and the Web of data. In *27th Annual ACM Symposium on Applied Computing*, pages 366–371. 10.1145/2245276.2245347.

Ferrucci, D., Levas, A., Bagchi, S., Gondek, D., and Mueller, E. T. (2013). Watson: Beyond Jeopardy! *Artificial Intelligence*, 199:93–105. doi: https://doi.org/10.1016/j.artint.2012.06.009.

Gherasim, T., Harzallah, M., Berio, G., and Kuntz, P. (2013). Methods and tools for automatic construction of ontologies from textual resources: A framework for comparison and its application. In *Advances in Knowledge Discovery and Management*, pages 177–201. Springer. doi: 10.1007/978-3-642-35855-5_9.

Hands, A. (2012). Microsoft academic search. *Technical Services Quarterly*, 29(3):251–252. doi: 10.1080/07317131.2012.682026.

Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., Hellmann, S., Morsey, M., Van Kleef, P., Auer, S., et al. (2015). DBpedia–a large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web*, 6(2):167–195. doi: 10.3233/SW-140134.

Matsuo, Y. and Ishizuka, M. (2004). Keyword extraction from a single document using word co-occurrence statistical information. *International Journal on Artificial Intelligence Tools*, 13(01):157–169. doi: 10.1142/S0218213004001466.

Mbipom, B., Craw, S., and Massie, S. (2016). Harnessing background knowledge for e-learning recommendation. In *Research and Development in Intelligent Systems XXXIII: Incorporating Applications and Innovations in Intelligent Systems XXIV*, pages 3–17. Springer. doi: 10.1007/978-3-319-47175-4_1.

Milne, D. and Witten, I. H. (2008). Learning to link with Wikipedia. In *17th ACM Conference on Information and Knowledge Management*, pages 509–518. doi: 10.1145/1458082.1458150.

Nasraoui, O. and Zhuhadar, L. (2010). Improving recall and precision of a personalized semantic search engine for e-learning. In *4th IEEE International Conference on Digital Society*, pages 216–221. doi: 10.1109/ICDS.2010.63.

Panagiotis, S., Ioannis, P., Christos, G., and Achilles, K. (2016). APLe: Agents for personalized learning in distance learning. In *7th International Conference on Computer Supported Education*, pages 37–56. Springer. doi: 10.1007/978-3-319-29585-5_3.

Porter, M. F. (1980). An algorithm for suffix stripping. *Program*, 14(3):130–137.

Qureshi, M. A., O'Riordan, C., and Pasi, G. (2014). Exploiting Wikipedia to identify domain-specific key terms/phrases from a short-text collection. In *Proceedings of the 5th Italian Information Retrieval Workshop*, pages 63–74.

Rodrigues, L., Antunes, B., Gomes, P., Santos, A., Barbeira, J., and Carvalho, R. (2007). Using textual CBR for e-learning content categorization and retrieval. In *4th International Conference on Case-Based Reasoning workshop on textual case-based reasoning*.

Ruiz-Iniesta, A., Jimenez-Diaz, G., and Gomez-Albarran, M. (2014). A semantically enriched context-aware OER recommendation strategy and its application to a computer science OER repository. *IEEE Transactions on Education*, 57(4):255–260. doi: 10.1109/TE.2014.2309554.

Salton, G. (1971). *The SMART Retrieval System-Experiments in Automatic Document Processing*. Prentice-Hall, Upper Saddle River, NJ, USA. doi: 10.1109/TPC.1972.6591971.

Salton, G. and Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5):513–523. doi: 10.1016/0306-4573(88)90021-0.

Völkel, M., Krötzsch, M., Vrandecic, D., Haller, H., and Studer, R. (2006). Semantic Wikipedia. In *Proceedings of the 15th International Conference on World Wide Web*, pages 585–594. ACM. doi: 10.1145/1135777.1135863.

Witten, I. H., Paynter, G. W., Frank, E., Gutwin, C., and Nevill-Manning, C. G. (1999). KEA: Practical automatic keyphrase extraction. In *4th ACM Conference on Digital libraries*, pages 254–255. doi: 10.1145/313238.313437.

Yang, H.-L. and Lai, C.-Y. (2010). Motivations of Wikipedia content contributors. *Computers in Human Behavior*, 26(6):1377 – 1383. doi: 10.1016/j.chb.2010.04.011.

Yang, K., Chen, Z., Cai, Y., Huang, D., and Leung, H. F. (2016). Improved automatic keyword extraction given more semantic knowledge. In *International Conference on Database Systems for Advanced Applications*, pages 112–125. Springer. doi: 10.1007/978-3-319-32055-7_10.

Yarandi, M., Tawil, A.-R., and Jahankhani, H. (2011). Adaptive e-learning system using ontology. In *In Proceedings of the 22nd International Workshop on Database and Expert Systems Applications*, pages 511–516. doi: 10.1109/DEXA.2011.9.

Zhang, X., Liu, J., and Cole, M. (2013). Task topic knowledge vs. background domain knowledge: Impact of two types of knowledge on user search performance. In *Advances in Information Systems and Technologies*, pages 179–191. Springer. doi: 10.1007/978-3-642-36981-0_17.

Zheng, Z., Li, F., Huang, M., and Zhu, X. (2010). Learning to link entities with knowledge base. In *Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 483–491.

Appendix B

Ethics Approval

STUDENT PROJECT ETHICAL REVIEW (SPER) FORM

The aim of the University's *Research Ethics Policy* is to establish and promote good ethical practice in the conduct of academic research. The questionnaire is intended to enable researchers to undertake an initial self-assessment of ethical issues in their research. Ethical conduct is not primarily a matter of following fixed rules; it depends on researchers developing a considered, flexible and thoughtful practice.

The questionnaire aims to engage researchers discursively with the ethical dimensions of their work and potential ethical issues, and the main focus of any subsequent review is not to 'approve' or 'disapprove' of a project but to make sure that this process has taken place.

The *Research Ethics Policy* is available at www.rgu.ac.uk/research-ethics-policy

Student Name	Blessing Ebong Mbipom
Supervisor	Professor Susan Craw
Project Title	Knowledge Driven Approaches for e-Learning Recommendation
Course of Study	PhD Computing
School/Department	School of computing Science and Digital Media

PART 1: DESCRIPTIVE QUESTIONS

1.	Does the research involve, or does information in the research relate to: *[see Guidance Note 1]*	**Yes**	**No**
	(a) individual human subjects	√	
	(b) groups (e.g. families, communities, crowds)		
	(c) organisations		
	(d) animals?		
	(e) genetically-modified organisms www.rgu.ac.uk/hr/healthsafety/page.cfm?pge=26027#122628		
	Please provide further details:		
	Staff and Students would be contacted to take part in evaluating the developed recommender system.		
2.	Will the research deal with information which is private or confidential? *[see Guidance Note 2]*	**Yes**	**No**
			x
	Please provide further details:		

PART 2: THE IMPACT OF THE RESEARCH			
3.	In the process of doing the research, is there any potential for harm to be done to, or costs to be imposed on: *[see Guidance Note 3(i)]*	**Yes**	**No**
	(a) research participants?		x
	(b) research subjects? *[see Guidance Note 3(ii)]*		x
	(c) you, as the researcher?		x
	(d) third parties? *[see Guidance Note 3(iii)]*		x
	Please state what you believe are the implications of the research:		
4.	When the research is complete, could negative consequences follow:	**Yes**	**No**
	(a) for research subjects		x
	(b) or elsewhere? *[see Guidance Note 4]*		x
	Please state what you believe are the consequences of the research:		

PART 3: ETHICAL PROCEDURES			
5.	Does the research require informed consent or approval from: *[see Guidance Note 5(i)]*	**Yes**	**No**
	(a) research participants?	√	
	(b) research subjects? *[see Guidance Note 5(ii)]*		
	(c) external bodies? *[see Guidance Note 5(iii)]*		
	If you answered yes to any of the above, please explain your answer:		
	Consent would be required from the research participants (staff/students) who would take part in evaluating the recommender system.		

		Yes	No
6.	Are there reasons why research subjects may need safeguards or protection? *[see Guidance Note 6]*		x
	If you answered yes to the above, please state the reasons and indicate the measures to be taken to address them:		
7.	Does the research involve any "regulated work with children" and/or "regulated work with protected adults", therefore requiring membership of the *Protecting Vulnerable Groups (PVG) Scheme*? *[see Guidance Note 7]*	**Yes**	**No**
			x
	[Please note: if the research potentially involves "regulated work", this MUST be raised with your HR Business Partner immediately. In this instance, the Human Resources Department will conduct a detailed assessment and will confirm whether or not PVG Membership is required.]		
	(a) PVG membership is not required.	√	
	(b) PVG membership may be required for working with children.		x
	(c) PVG membership may be required for working with protected adults.		x
	(d) PVG membership may be required for working with both children and protected adults.		x
	If you answered yes to (b), (c) or (d) above, please give further information about the work you will be required to undertake and the nature of the contact with these groups. Please provide as much detail as possible:		
	Are you already a PVG member?	**Yes**	**No**
	If yes, please provide your PVG Scheme number:		
8.	Are specified procedures or safeguards required for recording, management, or storage of data? *[see Guidance Note 8]*	**Yes**	**No**
		√	
	If you answered yes to any of the above, please give details:		
	i) Personally identifying data would not be collected from the participants. ii) The data collected would not be used in ways that go beyond the terms on which it has been given.		

PART 4: THE RESEARCH RELATIONSHIP

		Yes	No
9.	Does the research require you to give or make undertakings to research participants or subjects about the use of data? *[see Guidance Note 9]*	√	
	If you answered yes to the above, please outline the likely undertakings:		
	Likely undertakings would be: i. I understand that data collected during the study may be looked at by individuals from the research team where it is relevant to my taking part in this research. I give permission for these individuals to have access to my data. ii. I agree to take part in this study.		
10.	Is the research likely to be affected by the relationship with a sponsor, funder or employer? *[see Guidance Note 10]*		x
	If you answered yes to the above, please identify how the research may be affected:		

Part 5: Other Issues

		Yes	No
11.	Are there any other ethical issues not covered by this form which you believe you should raise?		x

Statement by Student

I believe that the information I have given in this form is correct, and that I have addressed the ethical issues as fully as possible at this stage.

Signature:		**Date:**	16th February, 2017

If any ethical issues arise during the course of the research, students should complete a further Student Project Ethical Review (SPER) form.

The *Research Ethics Policy* is available at www.rgu.ac.uk/research-ethics-policy

PART 6: TO BE COMPLETED BY THE SUPERVISOR

		Yes	No
12.	Does the research have potentially negative implications for the University? *[see Guidance Note 11]*		x
	If you answered yes to the above, please explain your answer:		
13.	Are any potential conflicts of interest likely to arise in the course of the research? *[see Guidance Note 12]*		x
	If you answered yes to the above, please identify the potential conflicts:		
14.	Are you satisfied that the student has engaged adequately with the ethical implications of the work? *[see Guidance Note 13]*	x	
	If you answered no to the above, please identify the potential issues:		

15.	**Appraisal:** Please select one of the following	
	i. The research project should proceed in its present form – no further action is required	x
	ii. The research project requires ethical approval by the School Ethics Review Panel (SERP) (or equivalent)	
	iii. The research project requires ethical review by the University's Research Ethics Sub-Committee	
	iv. The project needs to be returned to the student for modification prior to further action	
	v. The research project requires ethical review by an external body (N.B. Question 5 above). If this applies, please give these details:	

Title of External Body providing ethical review	
Address of External Body	
Anticipated date when External Body may consider project	

AFFIRMATION BY SUPERVISOR			
I have read the student's responses and have discussed ethical issues arising with the student. I can confirm that, to the best of my understanding, the information presented by the student is correct and appropriate to allow an informed judgement on whether further ethical approval is required.			
Signature:	Susan Grant	**Date:**	**16 February 2017**

The form has been independently scrutinised and reviewed by
Professor Patrik Holt. 20/02/17

Patrik O'Briain Holt.

ⓘ Guidance Note 1

Ethical principles normally apply to information, data, and derivative substances in the same way as they apply to the subjects themselves. Consequently, work with individual financial data is governed by the principles of work with individual human subjects, and work with animal tissue is governed by the principles of work with animals.

[return to Question 1]

ⓘ Guidance Note 2

The Australian National Health and Medical Research Council argues: "Individuals have a sphere of life from which they should be able to exclude any intrusion ... A major application of the concept of privacy is information privacy: the interest of a person in controlling access to and use of any information personal to that person." This principle applies to all information about a person, whether or not it is obtained directly from that person. The area that is private is conventional and culturally defined; in the UK it commonly includes income and family arrangements.

The information obtained in research is not, however, necessarily private. Some material is in the public sphere, which includes published and broadcast material, academic discourse, and the activities of government. Activities undertaken in a public place are public, rather than private, if they are openly displayed (e.g. artistic exhibition or attendance at a public event) or subject to public regulation (e.g. driving)."

[return to Question 2]

ⓘ Guidance Note 3

(i) "Harm" refers to negative consequences beyond those which would occur in the normal course of events. Costs may include putting subjects under stress, causing them anxiety, or even wasting their time. The question asks only about potential harm. Potential harm is not cancelled out by potential benefit. Broader consequences are considered in the following question.
Reviews of information are also subject to ethical consideration. It should never be assumed that no harm can be done to people simply by writing about them.

(ii) "Research subjects" includes not just participants and informants but those about whom data is collected. The term covers any research subject, including humans, animals, and inanimate subject matter.

(iii) The University has a responsibility to avoid putting you at risk, and potentially dangerous situations should always be drawn to the University's attention.

(iv) "Third parties" include any person, group or organisation who may be affected by the process of the research.

[return to Question 3]

ⓘ Guidance Note 4

"Elsewhere" is an open category, intended to include consequences for third parties, sections of the community (e.g. "the voluntary sector"), the economy ("the catering industry") or the environment. ("the national park"), globally, and generalities which are harder to identify (e.g. "animal welfare"). Student researchers should never assume that their work is harmless only because they don't believe others will read it.

[return to Question 4]

ⓘ Guidance Note 5

(i) Research in the public sphere (question 2) may not require the consent or approval of research subjects. The advice of the Canadian Tri-Boards is that "REBs (research ethics boards) should recognize that certain types of research - particularly biographies, artistic criticism or public policy research - may legitimately have a negative effect on organizations or on public figures in, for example, politics, the arts or business. Such research does not require the consent of the subject ... Consent is not required from organizations such as corporations or governments for research about their institutions".

There is a general presumption that consent should be obtained from subjects whenever the information is private. The requirement to seek consent can, however, be waived in certain exceptional cases, for example where there is necessary deception, or where the consent of a subject may jeopardise the welfare of an informant. All such cases require explicit ethical review and an extended justification.

(ii) The consent of research *subjects* cannot be presumed because the consent of *informants* has been obtained. For example, one member of a family cannot necessarily be taken to speak for others, and an employer cannot always give consent on behalf of employees.

(iii) The consent of *external bodies* is required for several types of research, including e.g.

- research relating to the NHS
- research for work with dangerous substances, and
- research involving experimentation with animals.

The existence of external consent does not ethically exclude the project from consideration by the University, or vice-versa. Please provide a brief description of the project as submitted to the external body for ethical review.

[return to Question 5]

ⓘ Guidance Note 6

This may apply, for example, to human subjects who are regarded as vulnerable (e.g. children or prisoners) and to animals. Consent should not be taken as sufficient protection.

[return to Question 6]

ⓘ Guidance Note 7

(i) Regulated work normally involves caring for, supervising or working with individuals who participate in an organised activity. There are two types of regulated work: regulated work with ***children*** and regulated work with ***protected adults*.**

(ii) ***Children*** are all people under the age of 18.

(iii) ***Protected adults*** are individuals aged 16 or over who are provided with (and thus receive) a type of care, support or welfare service. It is a service-based definition and avoids labelling adults on the basis of disability. A person will be a protected adult for the duration that they are receiving the service. Therefore some adults will be protected most of the time (e.g. residents within a care home) whereas others will only be protected for short periods (e.g. whilst receiving medical treatment at a hospital).

(iv) Further details can be found at www.rgu.ac.uk/about/governance/policies-and-legal/disclosure-scotland and www.disclosurescotland.co.uk/pvg/pvg_index.html.
Alternatively, you may want to discuss this with your HR Representative:
https://you.rgu.ac.uk/org/hr/SitePages/Meet%20the%20HR%20Team.aspx.

[return to Question 7]

ⓘ Guidance Note 8

Private data should be presumed to be under the control of the person or organisation to whom it relates. Anonymity is not a sufficient condition for confidentiality. Removing names from a report, or using aggregate data, may not be enough to ensure that respondents cannot be recognised or identified; and even where material is not identifiable except by the person who gave it, using it in ways that go beyond the terms on which it has been given may be a breach of trust.

[return to Question 8]

ⓘ Guidance Note 9
The integrity of the researcher, and the status of future research, requires that such undertakings should be respected. Promises should not be given in circumstances where they cannot be kept. For example, a researcher is not at liberty to conceal criminal activity and consequently cannot offer unconditional confidentiality in a study of such activity.

[return to Question 9]

ⓘ Guidance Note 10
Students who are undertaking research within the context of a work placement or employment should be aware that this is likely to have implications for the research and should identify what those implications are.

Sponsorship includes the grant of access to material by a responsible organisation.

[return to Question 10]

ⓘ Guidance Note 11
The University needs to know if the research may jeopardise its reputation through, for example, work for oppressive governments or other research relationships (e.g. work for tobacco firms) that might compromise or bias the research. Negative consequences in the form of criticism of the University or negative evaluations by students are legitimate potential outcomes.

[return to Question 12]

ⓘ Guidance Note 12
This includes, for example, conflicts between researchers, funders, stakeholders, employers and other research projects.

[return to Question 13]

ⓘ Guidance Note 13
In signifying agreement, principal supervisors are accepting part of the ethical responsibility for the project.

[return to Question 14]