國立清華大學

NATIONAL TSING HUA UNIVERSITY

碩士論文

On the Automatic Construction of Knowledge-Map from Handouts for MOOC Courses

基於 MOOC 課程講義之自動構建知識地圖

所別　通訊工程研究所

學號　104064522

姓名　黃譯葦 (Yi-Wei Huang)

指導教授　黃能富博士 (Dr. Nen-Fu Huang)

中華民國一〇六年七月

On the Automatic Construction of Knowledge-Map from Handouts for MOOC Courses

Prepared by Yi-Wei Huang

Directed by Prof. Nen-Fu Huang

In Partial Fulfillment of the Requirements

for the Degree of
Master of Science

Institute of Communication Engineering

National Tsing Hua University

Hsinchu, Taiwan 30013, R.O.C.

E-mail: s104064522@m104.nthu.edu.tw

August 9, 2017

Abstract

Massive open online courses (MOOCs) offer valuable opportunities for freedom in learning; however, many learners face cognitive overload and conceptual and navigational disorientation. Especially, when studying in the scenarios with complex material, there is a need for both effective learning strategies and the management of knowledge and information. The knowledge maps are a powerful tool that is have the spatial learning strategies by visualizing the knowledge. In this study, we used handouts to automatically build domain-specific knowledge maps for MOOCs. We considered handouts as conceptual models created by teachers, and we performed text mining to extract keywords from MOOC handouts and used the structure of the handouts to extract relations between keywords. Each knowledge map is based on the structure of the handouts, each consisting of an outline, title, and content. Our experiments of a course prove the proposed the knowledge map generator is faster than manual knowledge map, moreover, it is quite accurate. The average similarity of our system is high at almost 80%, and the highest similarity one is 95.8%. The findings suggest that using handouts to build knowledge maps is feasible. This feature is valuable for learners to quick identify the relation between the concepts in every week. We hope that the proposed system help learners to review, consolidate, and clarify the relation between the concepts. The overall idea and steps will be presented in this study.

Keywords -knowledge maps, learning styles, MOOCs, open learning.

中文摘要

近年來大型線上開放課程（MOOCs）為來自世界各地的使用者帶來學習的機會，然而線上課程使許多使用者面臨認知超載以及概念和導航障礙。尤其是在學習中遇到困難時，需要有效的學習策略以及知識和訊息的管理。在此篇論文裡，我們使用課程講義來自動產生線上課程的知識圖。我們認為線上課程講義是來自教師所創造的課程概念模型，所以我們對課程講義進行文本挖掘、並從中提取關鍵字，之後再利用課程講義的結構來提取關鍵字之間的關係。每一個知識地圖都是基於講義的結構，是由講義的大綱、標題以及內容所組成的。由我們的實驗結果證明得知，我們所提出知識地圖生成器的速度比專家手繪的知識地圖來的快速，並且知識地圖生成器所生成的知識地圖準確性是相對高的。由我們系統所生成的知識地圖平均相似度高達近 80％，而最高相似度可達 95.8％。研究結果表明，使用講義建立知識圖是可行的。這個功能對於學習者來說很有價值，可以快速識別每週知識概念間的關係。我們希望這個系統可以幫助學習者複習、整合與澄清課程概念之間的關係，所有想法與實作步驟會在接下來的內容詳細介紹。

關鍵字：知識地圖，學習系統，MOOCs，開放式學習

Acknowledgement

I would first like to thank my thesis advisor Prof. Nen-Fu Huang of the college of computer science at National Tsing Hua University. The door to Prof. Huang office was always open whenever I ran into a trouble spot or had a question about my research or writing. He consistently allowed this paper to be my own work, but steered me in the right the direction whenever he thought I needed it.

I would particularly like to single out my lab mate, Po-Wen Ou. I want to thank you for your excellent cooperation and for all of the opportunities I was given to conduct my research and further my thesis at National Tsing Hua University.

Finally, I must express my very profound gratitude to my parents and to my partner, lab mate for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Massive open online courses (MOOCs) are open educational resources that are available to learners worldwide free of charge. MOOCs include high-quality instructional videos produced by professors from prestigious universities. Learners share their ideas and reflections in discussion forums and use an online exercise system to evaluate their learning outcomes. Because MOOCs provide with high-quality self-directed online learning environments at low costs, they are considered a contemporary method of learning. MOOCs have been adopted by prestigious educational institutions, such as Stanford, Harvard, and the Massachusetts Institute of Technology, as well as private organizations and individuals, such as Salman Khan. Examples of MOOCs include Coursera [1], edX [2], Udacity [3], and Khan Academy [4].

Although the self-regulated learning structure of MOOCs offers considerable flexibility and the freedom to organize, select, and learn material of interest, many learners face cognitive overload and conceptual and navigational disorientation [5] [6] [7]. Cognitive overload refers to learners being required to process more information than what the human memory can usually hold [8]. MOOCs are not normally personalized or adaptable to learner needs and preferences [9]. This represents a significant challenge in learning environments with large numbers of learners [10]. In addition, MOOCs pose other challenges that may affect learners' success and willingness to complete courses.

In recent years, knowledge maps have been used extensively in education and business. A knowledge map is a visual representation of knowledge originally developed by Holley and Dansereau [11]. Knowledge maps could improve learning when used to summarize information. Evidence suggests that creating or studying summaries enhances the ability to recall summarized ideas [12]. Studying maps rather than text passages assists in the recollection of central ideas and details [13].

Several studies have investigated the use of knowledge maps in open learning environments, and a number of authors have experimented with organizing and presenting learning materials in open learning environments and customizing materials to learners' needs and preferences [13]. Knowledge maps facilitate the building schemata of learning concepts in learners' memories, thereby assisting learners in the learning process [14]. The construction of knowledge maps requires help from domain experts, who are difficult to hire. Few studies have investigated the automatic construction of knowledge maps [15] [16]. Several studies have attempted to automatically build domain knowledge maps for e-learning using text mining techniques [17].

Although some studies have provided valuable information on the relationship between MOOCs and knowledge maps, many factors of this relationship are still yet to be explored. This paper describes an innovative study combining knowledge maps and handouts from MOOCs. We regarded each handout as a conceptual model for a teacher. Many studies have briefly defined conceptual models as models created by professionals such as researchers, teachers, or engineers to facilitate the understanding and teaching of world systems and states of affairs [18].

The present study aimed to determine variations in accuracy between conceptual models and automatically constructed knowledge maps, and use knowledge maps to recommend. This paper proposes a method for the automatic construction of knowledge maps and an example of how to implement the method by using MOOC handouts.

The remainder of this thesis is organized as follows: We first introduce MOOCs platforms and the knowledge map on MOOCs platforms in chapter 2. The proposed method for the automatic construction of knowledge maps is presented in chapter 3. Implementation details are presented in chapter 4. Experiment is presented in chapter 5. Results and analysis are presented in chapter 6. Finally, a conclusion and suggestions for further study are presented in chapter 7.

Chapter 2

Related Work

2.1 MOOCs

MOOCs are centered around class instructor, instructor will usually provide series of lecture video, where learners mainly get knowledges. Besides, exercises, quiz, assignments are also used during the course. In recent year, most of the popular MOOC websites such as Coursera, edX, Udacity, etc.

2.1.1 Coursera

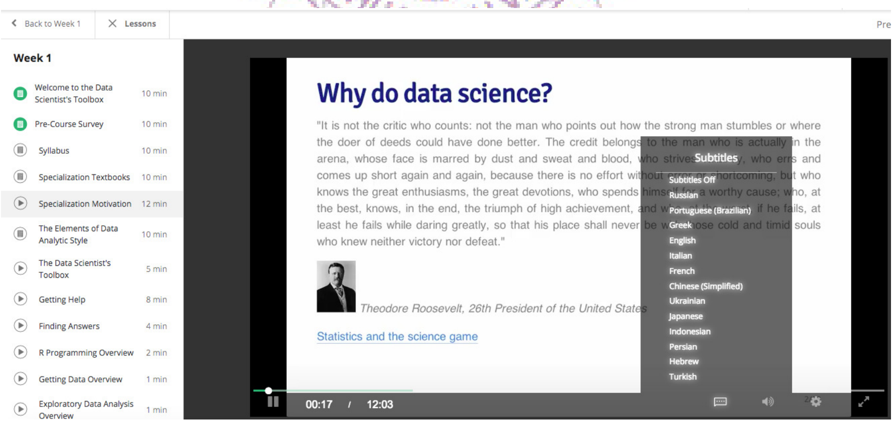

Figure 2.1: Course videos have translated subtitles

Coursera [1] is founded by Stanford professors Andrew Ng and Daphne Koller that offers online courses in April 2012. As of May 2017, Coursera had more than 25 million registered users and more than 2,000 courses. These courses provide quizzes, weekly exercises, peer-graded assignments, and sometimes a final project or exam. An inclusive experience designed for a diverse global audience, including video subtitles in over 30 languages (Fig. 2.1). Coursera provides mobile learning through apps for iOS and Android.

2.1.2 edX

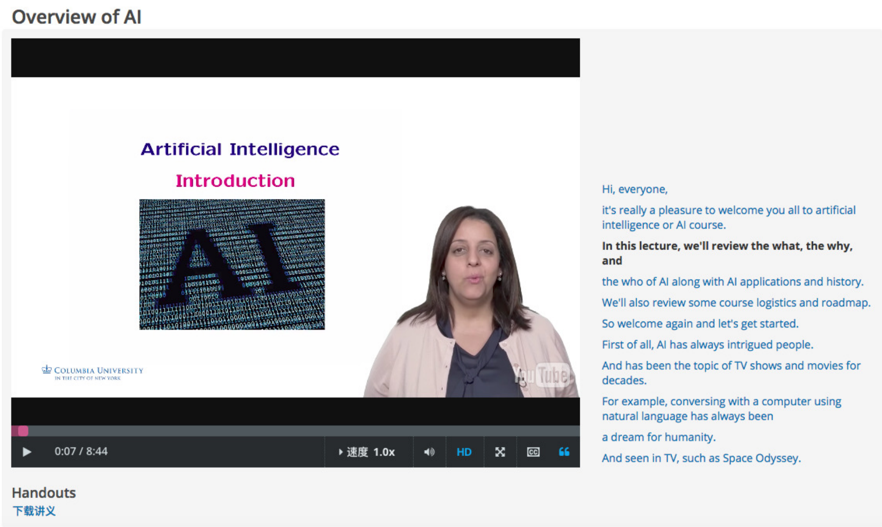

Figure 2.2: Interactive transcripts for edX course video

EdX [2] is founded by the Massachusetts Institute of Technology and Harvard University in May 2012. As of 29 December 2016, edX has around 10 million students taking more than 1,270 courses online. EdX differs from other MOOC providers, such as Coursera and Udacity, in that it is a nonprofit organization and runs on the Open edX [19] open-source software. In addition to educational offerings, edX is utilized for research into learning and distance education by collecting learners' clicks and analyzing the data, as well as collecting demographics from each registrant [20]. EdX course videos have interactive transcripts (Fig. 2.2). Interactive transcripts appear to the right of the screen, and keep pace with the video as it plays.

2.1.3 Udacity

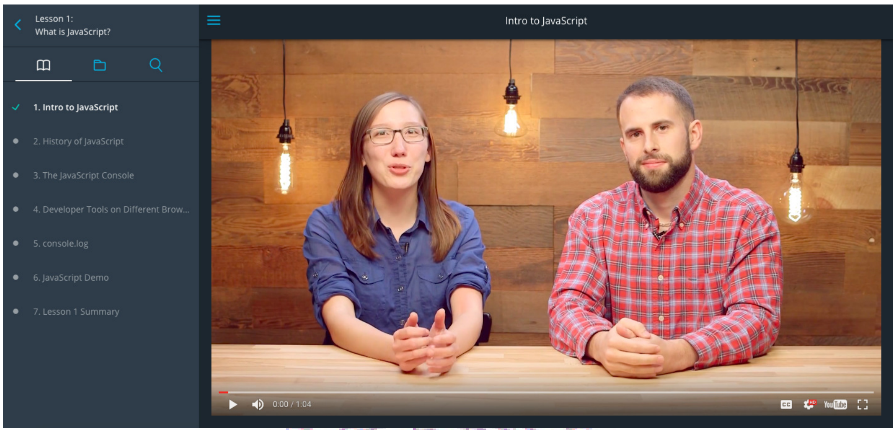

Figure 2.3: Course videos in Udacity

Udacity [3] is a for-profit educational organization founded by Sebastian Thrun, David Stavens, and Mike Sokolsky offering MOOCs. Each course consists of several units comprising video lectures (Fig. 2.3) with closed captioning, in conjunction with integrated quizzes to help students understand concepts and reinforce ideas, as well as follow-up homework which promotes a ”learn by doing” model.

2.1.4 Khan Academy

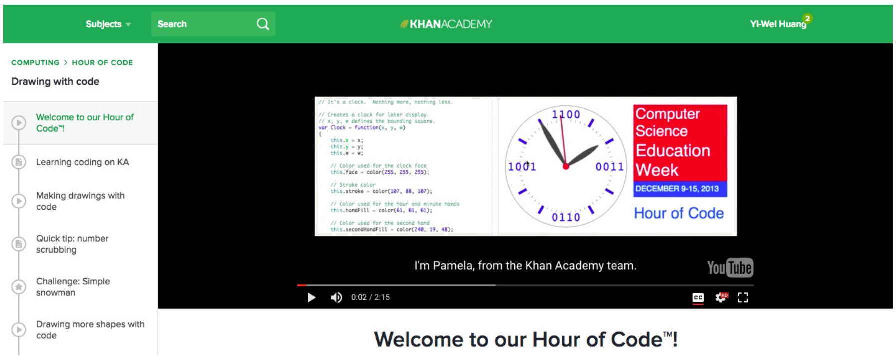

Figure 2.4: Khan Academy built on the videos which are hosted on YouTube

Khan Academy [4] is a nonprofit educational organization created in 2006 by educator Salman Khan with a goal of creating an accessible place for people to be educated. The organization produces short lectures in the form of YouTube videos (Fig. 2.4). Its website also includes supplementary practice exercises and tools for educators. Khan Academy's website aims to provide a personalized learning experience. The website is meant to be used as a supplement to its videos, because it includes other features such as progress tracking, practice exercises, and teaching tools.

2.1.5 Junyi Academy

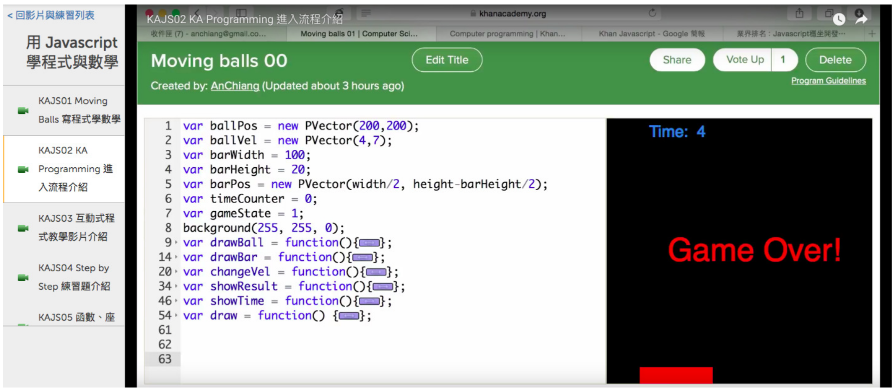

Figure 2.5: Course videos in Junyi Academy

Similar to Khan Academy, Junyi Academy [21] is a nonprofit educational website of the Republic of China (Fig. 2.5). The earlier Junyi Academy translated the online math questions and teaching videos on the Khan Academy. In order to meet the needs of local education, Junyi Academy recorded videos in the localization of the course

2.1.6 FutureLearn

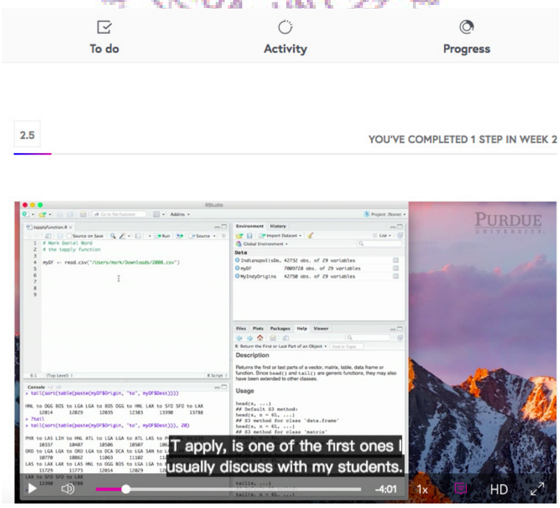

Figure 2.6: Course videos in FutureLearn

FutureLearn [22] is an education platform founded in December 2012. The company is wholly owned by The Open University in England. As of January 2017, included 109 UK and international partners, including non-university partners. FutureLearn's courses span a broad range of topics (Fig. 2.6).

2.1.7 Open2Study

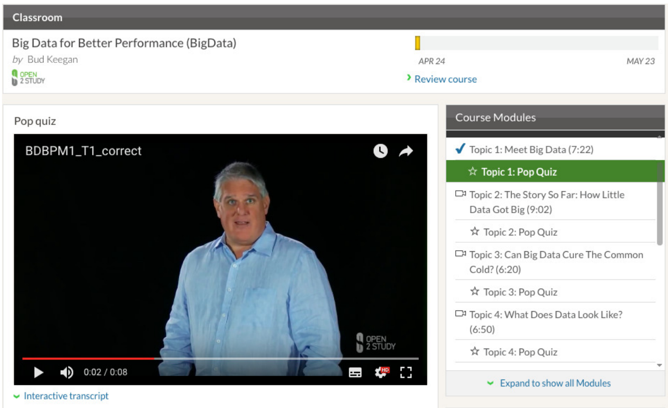

Figure 2.7: A four-week period with online assessments at the end of each of the modules

Open2Study [23] was launched in March 2013 and is a teaching, learning and assessment platform. The learning platform provided by Open2Study consists of weekly modules, which are completed over a four-week period with online assessments at the end of each of the modules (Fig. 2.7). Each course is self-contained, complete with interactive course content and resource materials.

2.1.8 xuetangX

Figure 2.8: Course videos in xuetangX

XuetangX [24] was launched on October 10, 2013, and it is the first Chinese MOOC platform it was initiated by Tsinghua University (Fig. 2.8). About 400 courses are offered on the XuetangX platform [25]. There are over five million registered users, and 6.2 million enrollments.

2.2 Knowledge map on MOOCs platform

2.2.1 Khan Academy

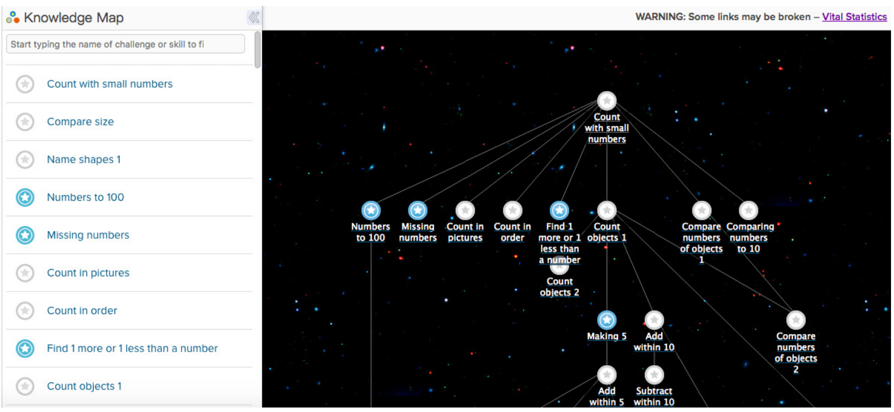

Figure 2.9: This is a section of a Khan Academy knowledge map for math concepts

The Knowledge Map [26] was a feature on Khan Academy which contained a starry night, containing all of the stars (Fig. 2.9). The stars represent lessons. Yellow stars with a blue border are lessons, user is proficient at, green borders mean recommended lessons, and others are lessons that are not recommended. An orange border means a lesson user should review. It also told the user how skills are connected to each other. The Knowledge Map also has a navigation bar, with which students could search for a particular skill. Since the Learning Dashboard came out, the feature is no longer available as of August 2013. Users can still access it here, but it is no longer reachable without the direct link.

2.2.2 Junyi Academy

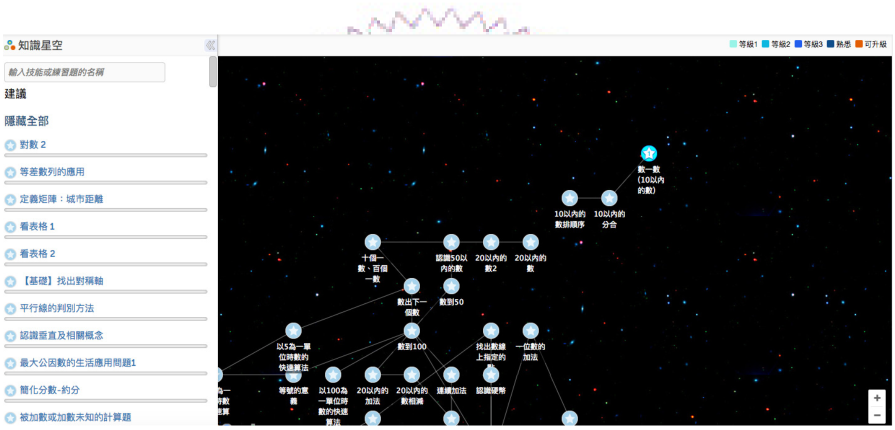

Figure 2.10: This is a section of a Junyi Academy knowledge map for math concepts

Similar to Knowledge Map on Khan Academy, an overview of how all the math exercises tie in together (Fig. 2.10). The knowledge map for math concepts range from all subjects covered in school and for all grades, Kindergarten to High School.

Chapter 3

System Architecture

In this thesis, we designed and implemented a system for the automatic construction of knowledge maps incorporating a current MOOC platform. This chapter introduces the system architecture and data analysis module. Moreover, we will also introduce the website MVC design and the user interface module.

3.1 Layer Structure Overview

This system can be divided into the following four parts: a web server, user interface module, data analysis module, and data server (Fig. 4.6). The web server consists of a web application and a web application programming interface (API). It mainly handles requests from users, enabling them to upload course files and select keywords from handouts. The user interface module contains the analyzed results based on MOOC handouts and presented using knowledge maps. The user interface module also provides manipulation like deleting, moving, adding and exporting. The data analysis module is a pure API server without a front-end view that is responsible for collecting and analyzing course files.

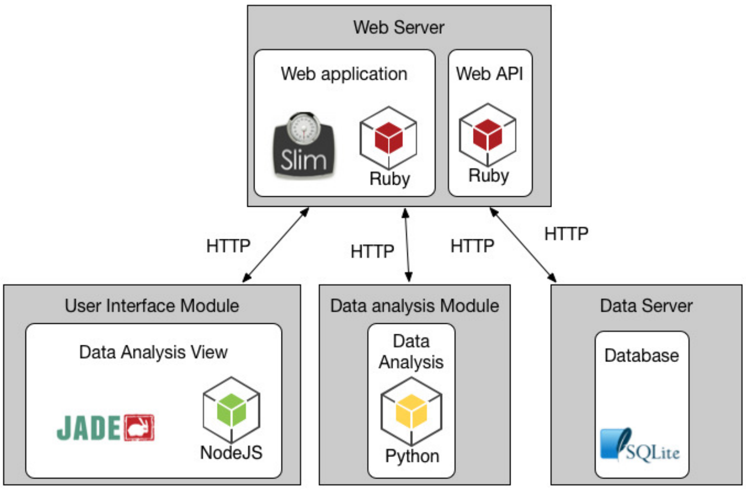

Figure 3.1: System structure of knowledge map generator

3.2 Website MVC design

Model–view–controller (MVC) is a software architectural pattern for implementing user interfaces on computers. The MVC pattern separates the model, the view, and the controller into three separate classes [27] (Fig. 3.2). The controller interprets the mouse and keyboard inputs from the user, informing the model and/or the view to change as appropriate [28]. The view manages the display of information. The model manages the behavior and data of the application domain, responds to requests for information about its state (usually from the view), and responds to instructions to change state (usually from the controller).

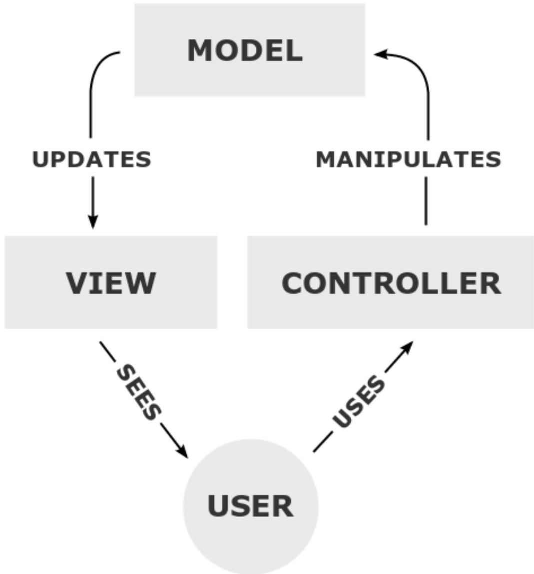

Figure 3.2: Collaboration of MVC elements

3.3 System Architecture

Figure. 3.3 illustrates each part of the system architecture in detail. The system comprises two subsystems: a web server and data analysis server. The web server grants access to the system; a user requests desired websites through a browser. Moreover, the web server contains most of the web application logic and is responsible for front end presentation for the user. We scheduled the task of automatically downloading course files to the data analysis server for analysis. The data analysis module consists of keyword extraction and relation extraction. Keyword extraction is based on the term frequency–inverse document frequency (TF–IDF) method [29], which provides numerical statistics to reflect how crucial a word is to a document in a collection or corpus [30]. The TF–IDF method is often used as a weighting factor for information retrieval, text mining, and user modeling. However, in this study, we analyzed the structure of MOOC handouts and subsequently defined keyword relations.

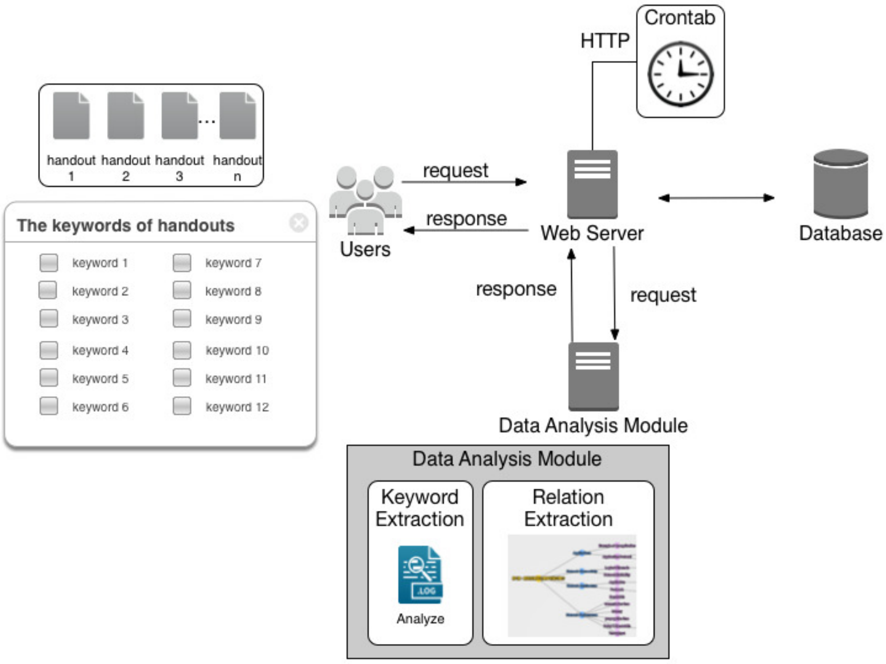

Figure 3.3: System architecture for the web server and the data analysis module

3.4 User interface module

User interface module plays an modificator in the proposed system, in order to correctly show the knowledge maps to students. Figure. 3.4 illustrates each part of the user interface module in detail. We designed four functions for users, in table 5.1 shows the description of each function.

Table 3.1: The description of each function in user interface module

Function	Description
Remove	We provide the function that the concept nodes and links are removed.
Move	We provide the function that the concept nodes and links are moved.
Add	We provide the function that the concept nodes and links are added.
Export	We provide the function that the knowledge map is exported as image.

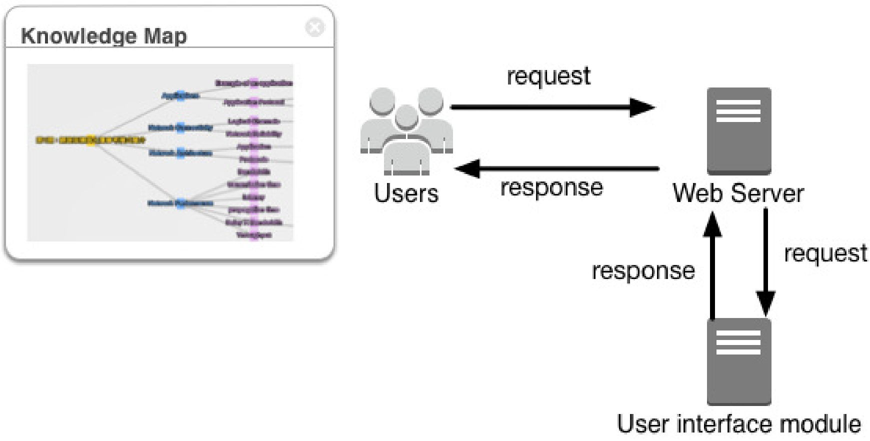

Figure 3.4: System architecture for the user interface module

Chapter 4

System Implementation

This chapter discusses the implementation of the proposed system, focusing on the functions of the web server, the data analysis module, the user interface and the recommender system.

4.1 Web Server

4.1.1 Uploading material

Our web server can provide user multiple file upload support. In this subsection, we show the APIs that web server provided for uploading files that is course handouts. The schema column of table 4.1 pictures out the data structure stored in SQLite [31].

Table 4.1: APIs uploads course handouts

method	URL	purpose	schema
GET	/course_data	shows the course handouts that is uploaded by user	
POST	/file_data	uploads the course handouts by user	filename: String document: String
DELETE	/file_data	deletes the course handouts by user	

4.2 Data Analysis Module

The data analysis module is crucial in the proposed system (Fig. 4.1). We use course handouts to determine the structure and relations of the module. In this subsection, we explain how the effective text and structure of a handout is extracted and how a knowledge map is generated.

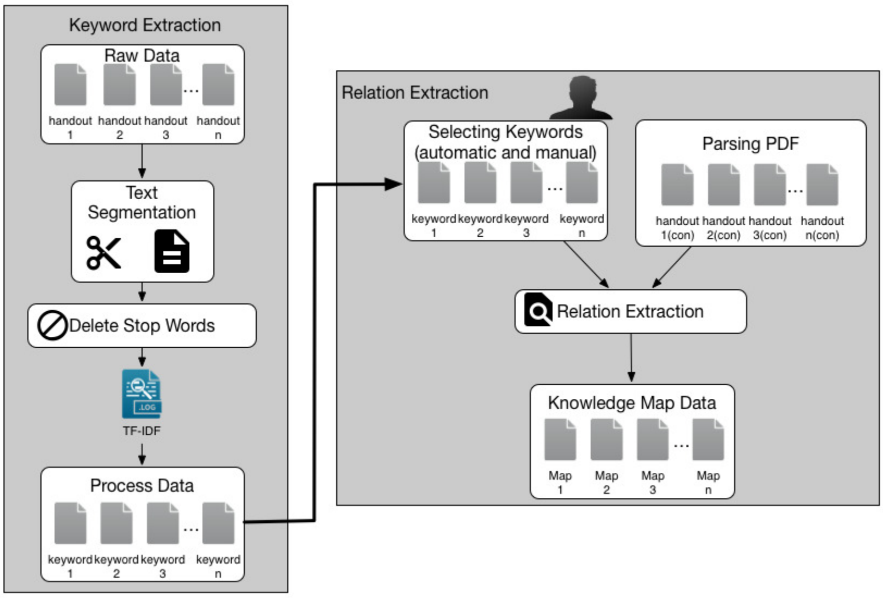

Figure 4.1: Data Analysis Module

4.2.1 Keyword Extraction

In the first phase, keywords are extracted from course handouts. Before extraction, we must complete the following steps. First, we use the text segmentation method, which entails the division of written text into meaningful units such as words, sentences, and topics. Second, we delete stop words, which are words that are filtered out before or after the processing of natural language data (text) [32]. Once these preprocesses have been completed, we use the TF–IDF algorithm to weight terms. Finally, we produce new files containing the keywords and their corresponding weights.

4.2.1.1 Text Segmentation

For English sentences, direct word-division based on space is suitable. For Chinese sentences, some specific word-division method should be applied in order to get meaningful word, so we use Jieba [33] this flexible tool to segment out sentences that are from course handouts. Jieba is an open source python Chinese text segmentation module which provide friendly API for developers. Jieba segment the text by its own dictionary on default, however it allows developers to edit or replace the dictionary which Jieba used to segment text.

4.2.1.2 Delete Stop Words

Stop-words are those meaningless words which makes little contribution to the meaning of the course handout. For each sentence, after we segment it into a bag of words, we examine whether the words contained are in the stop-word list and remove the words that are in the list.

4.2.1.3 TF/IDF

The TF–IDF algorithm is the product of the term frequency tf and inverse document frequency idf. We might count the term frequency, which refers to the number of times each term occurs in each document. However, in cases where document length is highly variable, adjustments are often made. The weight of a term in a document is simply proportional to the term frequency [34]. To calculate the term frequency $tf_{t,d}$, the simplest option is to use the raw count of a term in a document (i.e., the number of times that term t occurs in document d). If we denote the raw count by $f_{t,d}$, the term frequency adjusted for document length is expressed as follows [29]:

$$tf_{t,d} = \frac{f_{t,d}}{\sum_k f_{k,d}} \tag{4.1}$$

The inverse document frequency is a measure of how much information the word provides, or in other words, whether the term is common or rare across all documents. It is calculated based on the logarithmically scaled inverse fraction of the documents that contain the word, which is obtained by dividing the total number of documents by the number of documents containing the term, and subsequently obtaining the logarithm of that quotient, expressed as follows:

$$idf(t,D) = \log \frac{N}{|\{d \in D : t \in d\}|} \tag{4.2}$$

Where:

N: is the total number of documents in the corpus $N = |D|$

$|\{d \in D : t \in d\}|$: is the number of documents where the term t appears (i.e., $tf_{t,d} \neq 0$). A term being absent from the corpus leads to division by zero. Therefore, it is common to adjust the denominator to $1 + |\{d \in D : t \in d\}|$.

Subsequently, tf–idf is calculated as follows:

$$tfidf(t,d,D) = tf(t,d) \cdot idf(t,D) \tag{4.3}$$

A high TF–IDF weight is reached through a high term frequency in the given handouts and a low term frequency in the whole collection of documents. Hence, the weights tend to filter out common terms.

4.2.2 Relation Extraction

After we extract the keywords, we must define relations. Before extraction, we must complete the following steps. First, we must obtain refined keywords by the knowledge map creator. Second, we analyze the construction of the course handout. Once these preprocesses have been completed, we can extract the relations. Finally, we produce new files containing the structure of the knowledge map.

4.2.2.1 Selecting Keywords

After we extract the keywords, the knowledge map creator selects refined keywords from the list of highest ranked terms (Fig. 4.2). The knowledge map creator also selects the refined titles from course handout.

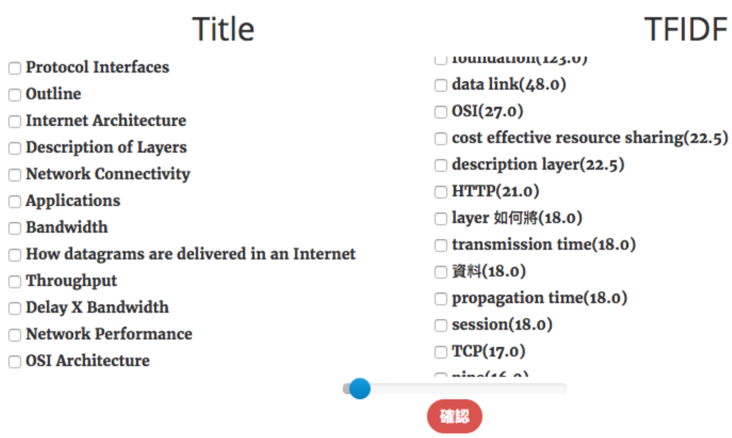

Figure 4.2: The knowledge map creator selects refined keywords

4.2.2.2 Parsing PDF

This section focuses on how to use the PDFMiner to analyze course handouts. Each handout consists of an outline, title, and content. We must complete the following steps. First, we filter out some punctuation marks, bullet points, and images. Second, we analyze the word size in every column and the sentence layers on every page. Finally, we produce analysis results and store files.

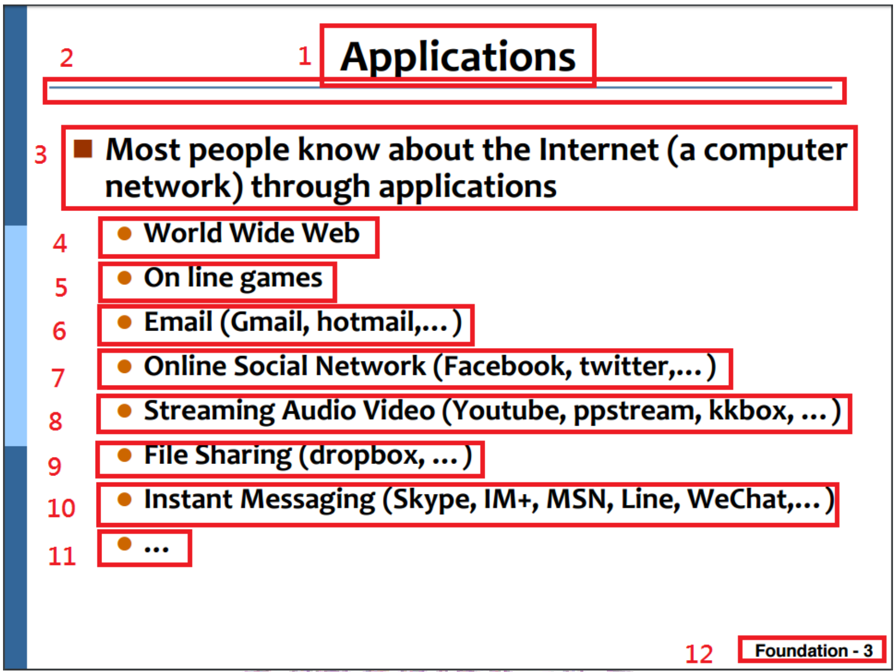

Figure 4.3: Layout analyzer in PDFMiner returns the twelve objects for this page

First, we assume that the handouts are PDF files. PDFMiner [35] is an open source Python package that extracts information from PDF documents and provides a developer-friendly API. The package currently supports Python 2 and can be installed using pip or other Python package management modules. Unlike other PDF-related tools, PDFMiner focuses entirely on obtaining and analyzing text data by enabling a user to obtain the exact location of text on a page and other information such as fonts and lines. We use this flexible tool to determine the structure bases of MOOC handouts. The layout analyzer returns objects for each page in the PDF document (Fig. 4.3).

We extract the corresponding locations and font sizes based on objects. Because the tool does not have a hierarchy function, we designed a function to determine the hierarchy based on the corresponding locations. The hierarchy indicates which layer the sentence belongs to. In addition, we can determine which sentence is higher or lower than the sentence currently being analyzed.

Algorithm 1 The Hierarchical Function of the PDF document

Input: $now\_position$ and $last\_position$
Output: $history\_layer$

if $now\_position > last\_position$ **then**
  $now\_layer = last\_layer + 1$
  add now_layer to $history\_layer$ array
else if $now\_position = last\_position$ **then**
  $now\_layer = last\_layer$
  add now_layer to $history\_layer$ array
else if $now\_position < last\_position$ **then**
  find the same position in $history\_layer$ array
end if

4.2.3 Knowledge Map Generation

We use the refined keywords and parsing files of handouts to create knowledge maps and the name of the course chapter as the root of the knowledge map. The first layer is the outline. We collect the content of the same outline based on the parsing files of handouts. Because every content includes a title, the second layer is the title. The parsing files of handouts contain every sentence matched with layer numbers. Furthermore, we must determine keywords in sentences based on the refined keywords file. We can match keywords to layer numbers, and we can then regard the highest layer number as the third layer. In accordance with the rules, we create the remaining layers. Finally, we generate a knowledge map for each handout.

4.3 User Interface Module

4.3.1 Environment

4.3.1.1 ShareCourse

ShareCourse [36] is a well-known MOOC platform in Chinese-speaking countries. It provides courses online and in mobile applications. ShareCourse provides more than 1000 courses and has more than 60,000 learners registered online. It provides an environment with numerous facilities for MOOC learners, including instructional videos, peer-graded quizzes and assignments, discussion forums, weekly exercises, and certification.

4.3.1.2 Cytoscape.js

Cytoscape.js [37] is an open-source graph theory (a.k.a. network) library written in JavaScript. We use cytoscape to present the result of data analysis module to user, providing them a friendlier and fancier data visualization view.

4.3.1.3 Bootstrap

Bootstrap [38] is an open-source and free front-end web framework for design website and web applications, it contains html, CSS and JavaScript-based template for typography. Useful elements such as button, modal, grid system and more are provided which makes developers create beautiful custom website easily.

4.3.2 Modified Knowledge Map

We set up a platform [39] for teacher or TAs to manage course elements. With this interface, teaching group can create and modify knowledge map. The expert could add concepts or relational link in our interface (Fig. 4.4). The expert also could delete concepts or relational link in our interface.

Figure 4.4: Modified knowledge map.

4.3.3 Knowledge Map Page

We added a tab beside the current learner progress page (Fig. 4.5). When "Knowledge Map" is clicked, the browser sends a request to the server to obtain information regarding which chapters have corresponding knowledge maps.

Figure 4.5: The tab of a "Topics on Investment" course.

Once the button been clicked, the browser will send a post request to user interface module to get knowledge map of each week. The request return data shows the vertical tabs of the "Knowledge Map" page. The vertical tabs are clickable buttons and a learner can select the knowledge map they wish to view (Fig. 4.6).

Figure 4.6: Knowledge map for the week 1 of a "Topics on Investment" course.

Concepts in knowledge map will be a clickable button, and if the concept is clicked, another section contains multiple hyperlink, we call it hyperlink section, will show up at the bottom of knowledge map. Figure 4.7 shows the result we click "money markets" in knowledge map, and a section contains one hyperlinks shows up. The hyperlinks will bring them to which videos and which forums are about this concept.

Figure 4.7: Hyperlinks show up on click the concept in map.

4.4 Recommender System

We would like to give our users a few recommendations in knowledge maps. We recommend suitable videos and forums. Figure 4.7 shows a section contains one hyperlinks shows up. The hyperlinks will bring users to the video is about the concept of the “money markets”. In addition, we also integrate the misconception diagnostic system into our recommender system (Fig. 4.8). We can find the examination question by matching concept in knowledge map. This feature is valuable for teachers to quick set exam papers.

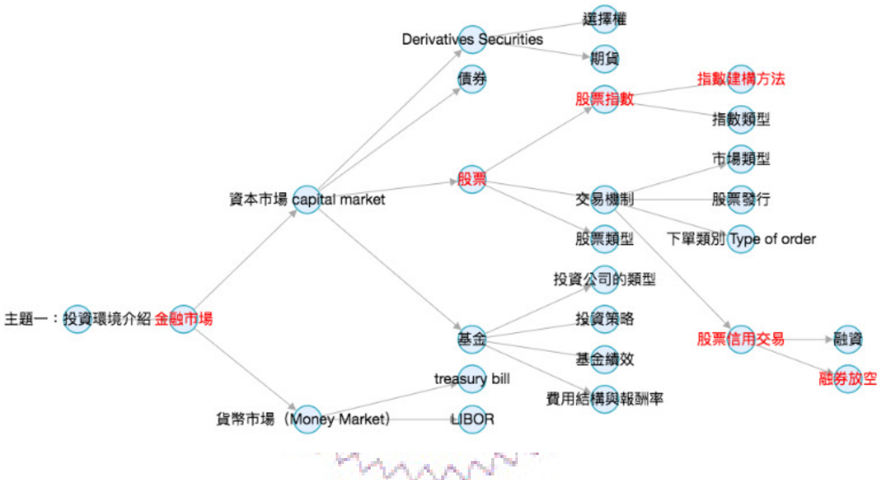

Figure 4.8: Integrating the misconception diagnostic system into our recommender system.

4.4.1 The Recommendation Page

Figure 4.9: The page of the video recommendation.

The video recommendation by matching concept between the video title and the knowledge map. Figure 4.9 shows the result we click the first hyperlink, the system open this video in a new browser window.

Figure 4.10: The topic and the ID of the forum.

The forum recommendation by matching concept between the topic of the forum and the knowledge map. Figure 4.10 shows the topic and the ID of the forum.

Chapter 5

Experiment

Since we are curious about whether knowledge map helps user or not, knowledge map had been applied in one courses on ShareCourse. In this chapter, we will have two sets of experiments were performed to examine our system. The first set of experiments compares knowledge maps created by our system with the expert knowledge maps. The second set of experiments uses questionnaire to collect data. The outcome of the questionnaire is to measure students' use, usability and attitudes to use knowledge map and their perceptions about learning "Topics on Investment" using knowledge map. A "Topics on Investment" course will understand the investment environment home and abroad. The course contents consist the presentation of investment evaluation tool and so on.

5.1 Similarity between Knowledge Maps is calculated

5.1.1 Data Set

We used the handouts of a "Topics on Investment" [40] that can be download on Share-Course. The course last approximately nine weeks, with one to two hours of video lectures a week. Each week has course handouts. Hence, we can collect the nine handouts.

Table 5.1: The details about the handouts of a ”Topics on Investment”

Week	Number of pages	Size
1	72	1.5 MB
2	40	1.2 MB
3	54	1.1 MB
4	49	1.2 MB
5	45	1.2 MB
6	46	1.1 MB
7	38	1.0 MB
8	51	1.1 MB
9	59	1.3 MB

We also used the handouts of an ”Introduction to Computer Networks [41] that can be download on ShareCourse. The course last approximately ten weeks, with one to two hours of video lectures a week. Because this course has two exams, we can collect the eight handouts.

Table 5.2: The details about the handouts of a ”Introduction to Computer Networks”

Week	Number of pages	Size
1	41	2.8 MB
2	52	1.8 MB
3	78	830 KB
4	50	802 KB
5	44	440 KB
6	41	355 KB
7	78	1.2 MB
8	60	640 KB

5.1.2 Method

In this experiment, the automatically constructed knowledge maps was compared to expert knowledge map. We use cosine similarity [42] to calculate the similarity of the path that is from the expert knowledge maps and the automatically constructed knowledge maps. The cosine of two non-zero vectors can be derived by using the Euclidean dot product formula:

$$a \cdot b = \|a\|_2 \|b\|_2 \cos\theta \tag{5.1}$$

Given two vectors of attributes, A and B, the cosine similarity, $\cos(\theta)$, is represented using a dot product and magnitude as

$$similarity = \cos(\theta) = \frac{A \cdot B}{\|a\|_2 \|b\|_2} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{5.2}$$

Where:

A_i and B_i are components of vector A and B respectively.

The resulting similarity ranges from -1 meaning exactly opposite, to 1 meaning exactly the same, with 0 indicating orthogonality (decorrelation), and in-between values indicating intermediate similarity or dissimilarity. In addition, we also use paths of leaves (POL) algorithm [43] to get similarity between the expert knowledge maps and the automatically constructed knowledge maps.

We used the knowledge map for the week 5 of a "Topics on Investment" for an example. The calculating similarity is carried out in four steps:

Step 1: We will convert concepts in knowledge map to letter designations, and create each POL path (Fig. 5.1).

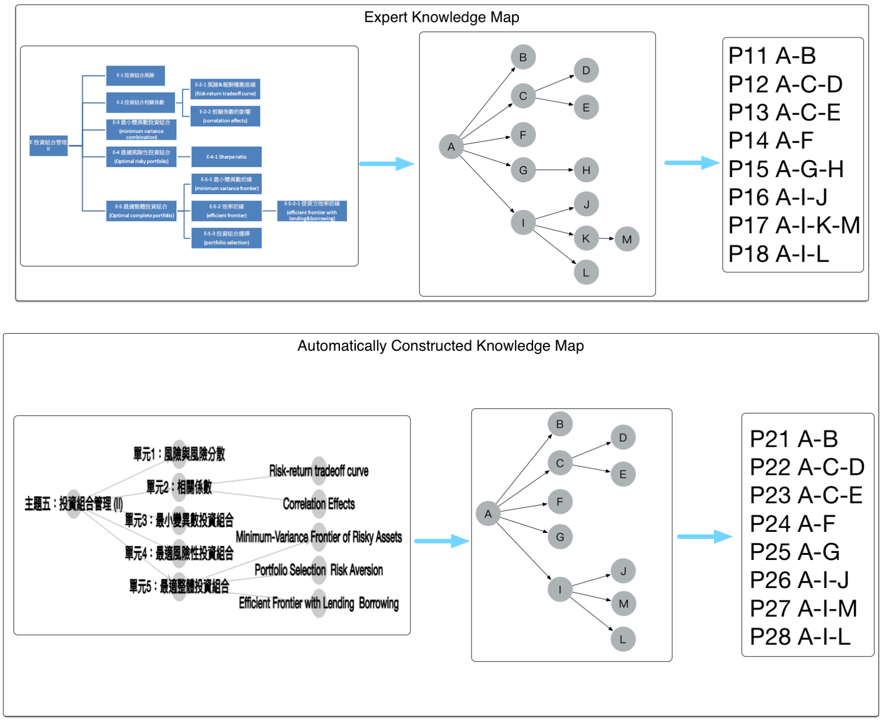

Figure 5.1: Convert concepts in knowledge map to letter designations, and create each POL path.

Step 2: Computing the cosine similarity between P1n and P2i. For instance, P1n is P11 and P2i is P22 (Fig. 5.2).

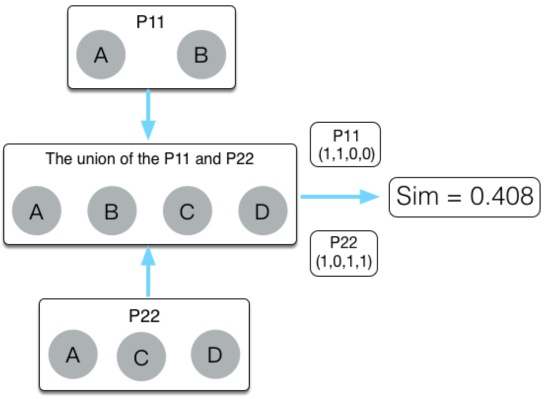

Figure 5.2: Computing the cosine similarity.

Step 3: Creating a matrix between P1n and P2i (Table 5.3).

Table 5.3: The matrix between P1n and P2i.

	P21	P22	P23	P24	P25	P26	P27	P28
P11	1.0	0.408	0.408	0.4998	0.4998	0.408	0.408	0.408
P12	0.408	1.0	0.667	0.408	0.408	0.337	0.337	0.337
P13	0.408	0.667	1.0	0.408	0.408	0.337	0.337	0.337
P14	0.4998	0.408	0.408	1.0	0.4998	0.408	0.408	0.408
P15	0.408	0.337	0.337	0.408	0.816	0.337	0.337	0.337
P16	0.408	0.337	0.337	0.408	0.408	1.0	0.667	0.667
P17	0.354	0.289	0.289	0.354	0.354	0.577	0.866	0.577
P18	0.408	0.337	0.337	0.408	0.408	0.667	0.667	1.0

Step 4: Computing the similarity between the expert knowledge map and the automatically constructed knowledge map, expressed as follows:

$$Sim(Kmap_1, Kmap_2) = \left(\frac{\sum_n [Max(P_{1n}) \times Len(P_{1n})]}{\sum_n Len(P_{1n})} + \frac{\sum_i [Max(P_{2i}) \times Len(P_{2i})]}{\sum_i Len(P_{2i})} \right) / 2 \tag{5.3}$$

For an example, we calculated the final similarity for the week 5 of a "Topics on Investment" (Fig. 5.3).

$$\left(\frac{(1\times2)+(1\times3)+(1\times3)+(1\times2)+(0.816\times3)+(1\times3)+(0.866\times4)+(1\times3)}{2+3+3+2+3+3+4+3} + \frac{(1\times2)+(1\times3)+(1\times3)+(1\times2)+(0.816\times2)+(1\times3)+(0.866\times3)+(1\times3)}{2+3+3+2+2+3+3+3} \right)/2$$
$$= 0.95801$$

Figure 5.3: The final similarity from the week 5 of a "Topics on Investment".

5.2 Experiment Questionnaire

5.2.1 Data Set

The system was successfully tested with students from an undergraduate course. The course is a small private online course (SPOC) [44] refers to a version of a MOOC used locally with on-campus students. While the class strength was thirty-three, thirty-one students finished the questionnaire [45].

5.2.2 Method

Conduct and interpret an internal consistency reliability analysis through Cronbach's alpha. Cronbach's alpha [46] [47] is the most common measure of internal consistency ("reliability"). It is most commonly used when you have multiple Likert questions in a survey/questionnaire that form a scale and you wish to determine if the scale is reliable. If you are concerned with inter-rater reliability, we also have a guide on using Cohen's (K) kappa that you might find useful.

We use a tool to analyze the questionnaire, the tool is statistical product and service solutions (SPSS). SPSS [48] [49] is a comprehensive system for analyzing data. SPSS can take data from almost any type of file and use them to generate tabulated reports, charts, and plots of distributions and trends, descriptive statistics, and complex statistical analysis.

Chapter 6

Results and Analysis

Because we sought to determine whether knowledge maps help teachers and students, a knowledge map was applied in a course on ShareCourse. This chapter presents the results and analysis of the similarity for all weeks of a ”Topics on Investment” and ”Introduction to Computer Networks”. At the end of this chapter also presents the analysis of the questionnaire.

6.1 Analysis of the Similarity for All Weeks of a ”Topics on Investment” and an ”Introduction to Computer Networks”

6.1.1 Topics on Investment

We will show the knowledge maps for the all weeks of a ”Topics on Investment” (Fig. 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9). The knowledge maps respectively created by experts and our system. Table 6.1 shows the similarity for the knowledge maps respectively created by experts and our system.

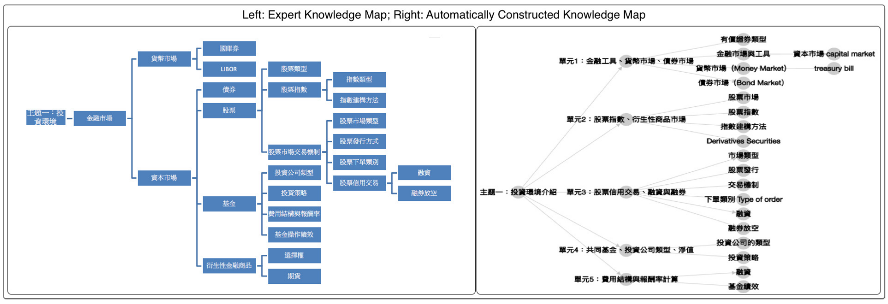

Figure 6.1: The knowledge maps for week 1 of a "Topics on Investment".

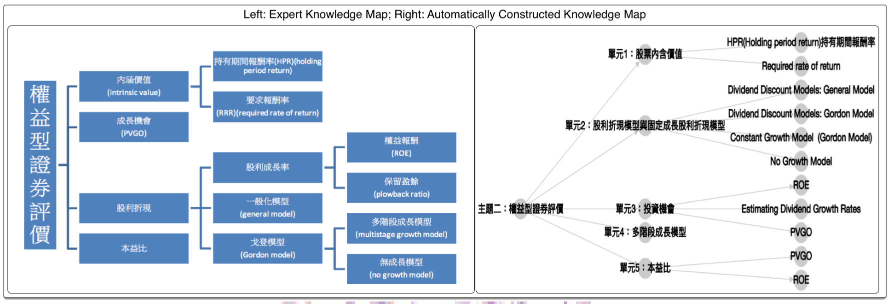

Figure 6.2: The knowledge maps for week 2 of a "Topics on Investment".

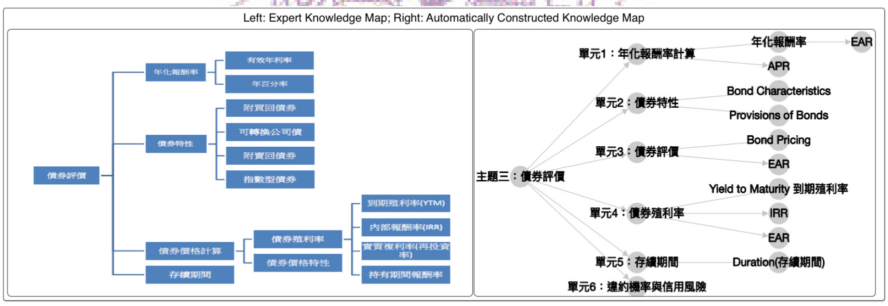

Figure 6.3: The knowledge maps for week 3 of a "Topics on Investment".

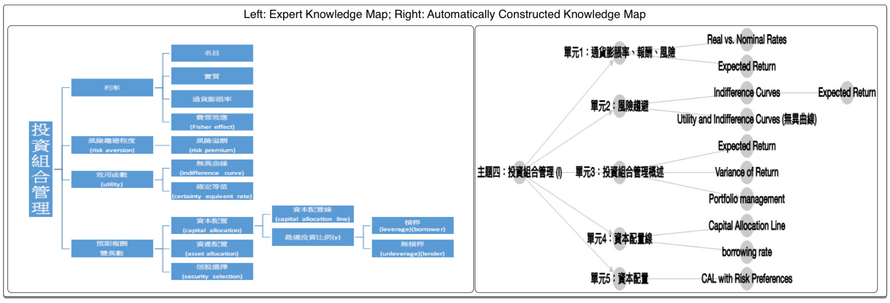

Figure 6.4: The knowledge maps for week 4 of a "Topics on Investment".

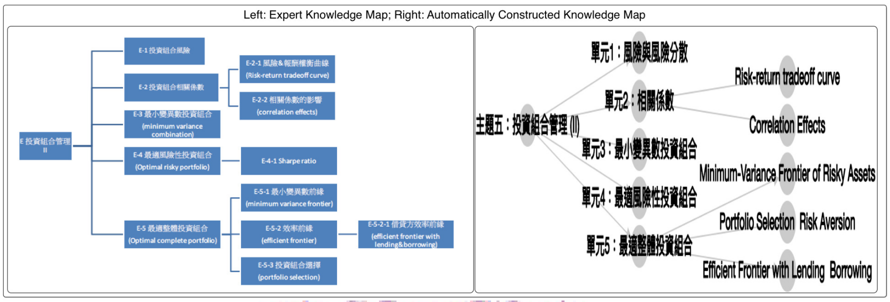

Figure 6.5: The knowledge maps for week 5 of a "Topics on Investment".

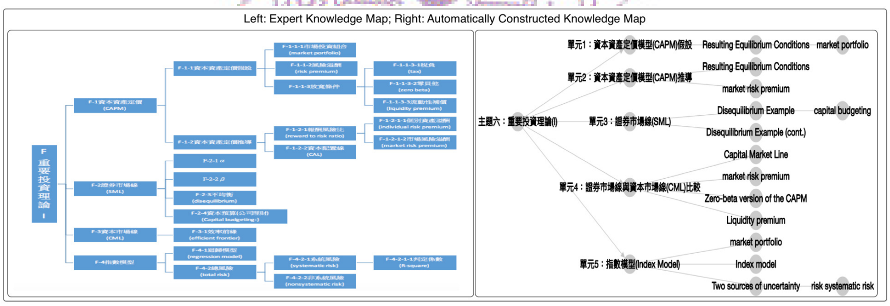

Figure 6.6: the knowledge maps for week 6 of a "Topics on Investment".

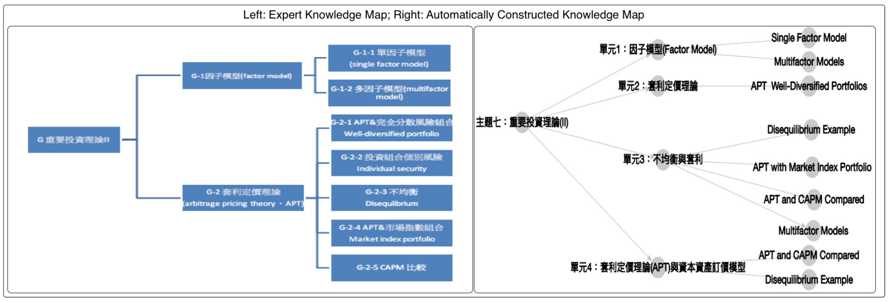

Figure 6.7: The knowledge maps for week 7 of a ”Topics on Investment”.

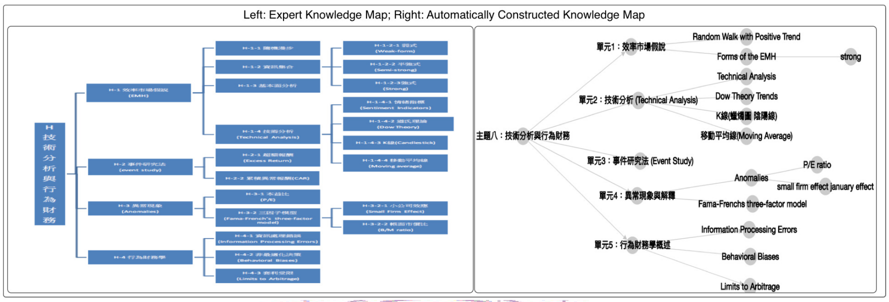

Figure 6.8: The knowledge maps for week 8 of a ”Topics on Investment”.

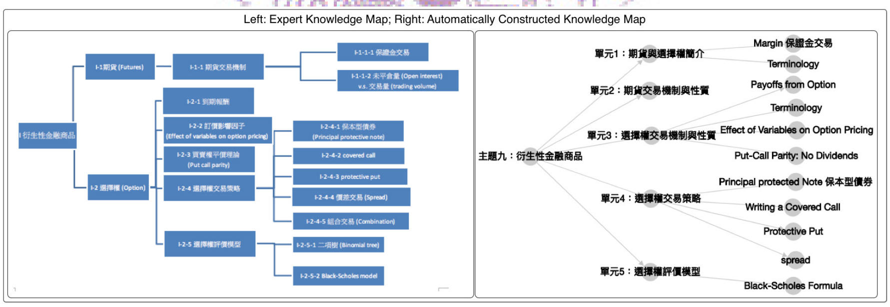

Figure 6.9: The knowledge maps for week 9 of a ”Topics on Investment”.

Table 6.1: The similarity for all weeks of a ”Topics on Investment”.

	Similarity (0.0 - 1.0)
Week 1	0.70399
Week 2	0.816733
Week 3	0.778174
Week 4	0.62299
Week 5	0.95801
Week 6	0.65838
Week 7	0.773826
Week 8	0.87222
Week 9	0.69139
Average	0.76397

We discussed an experimental method that creating a matrix in previous chapter. We used the knowledge map for the week 7 for an example. Table 6.2 shows a matrix for week 7. We choose the largest number as the similarity in each path. We see the paths P11 and P15 that the largest numbers are 1.0 and 0.816 respectively. We see the paths P21 and P25 that the largest numbers are 1.0 and 0.667 respectively. The final similarity is calculated using following formula:

$$Sim(Kmap_1, Kmap_2) = \left(\frac{\sum_n [Max(P_{1n}) \times Len(P_{1n})]}{\sum_n Len(P_{1n})} + \frac{\sum_i [Max(P_{2i}) \times Len(P_{2i})]}{\sum_i Len(P_{2i})} \right) /2 \tag{6.1}$$

Table 6.2: The matrix for the week 7.

	P21	P22	P23	P24	P25	P26	P27	P28	P29
P11	1.0	0.667	0.333	0.408	0.333	0.333	0.333	0.333	0.333
P12	0.667	1.0	0.333	0.408	0.333	0.333	0.333	0.333	0.333
P13	0.333	0.333	1.0	0.408	0.333	0.333	0.333	0.333	0.333
P14	0.333	0.333	0.667	0.408	0.333	0.333	0.333	0.333	0.333
P15	0.333	0.333	0.667	0.816	0.667	0.667	0.667	0.333	0.333
P16	0.333	0.333	0.667	0.408	0.667	0.333	0.333	0.333	0.333
P17	0.333	0.333	0.667	0.408	0.333	0.333	0.333	0.333	0.333

As seen from the figure 6.10, the point G of the path P14 is ”individual security”.

Because the concept was mentioned by the teacher in the video were not explained in the handouts, the similarity is low in the path P14 of table 6.2.

We see the point E, H and K in the automatically constructed knowledge map and the points are about ”arbitrage pricing theory” of the information. Because the point E, H and K are similar, layer one has a few unnecessary concepts. On the basis of the above reasons, the similarity in the path P24 to P29 is lower than the P21 to P23.

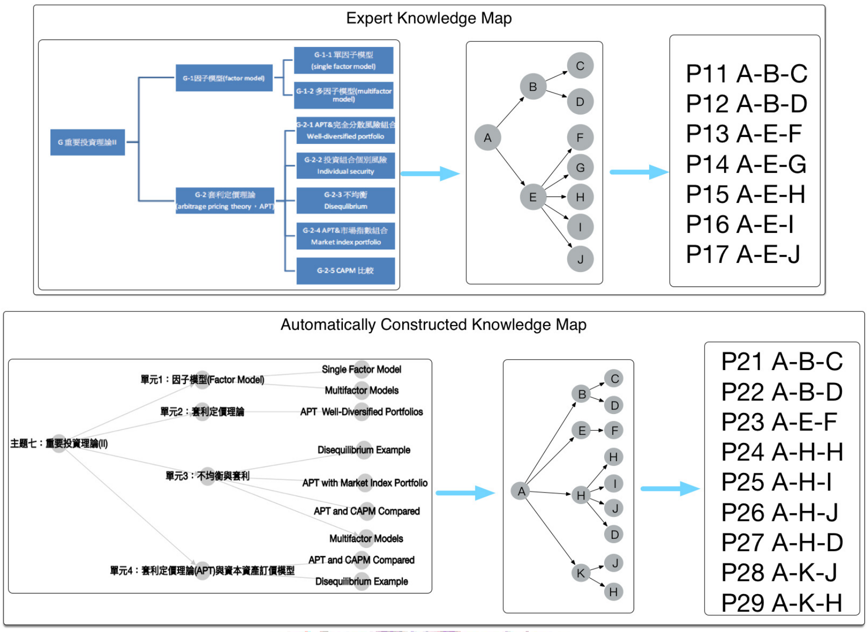

Figure 6.10: Convert concepts in knowledge map to letter designations, and create each POL path.

As seen from the table 6.1, week 5 have the highest similarity. The reason is the concepts and relations in the automatically constructed knowledge map is similar to the content of expert knowledge map. The automatically constructed knowledge map is short of two concepts, ”Sharpe ratio” and ”efficient frontier”, are in layer 2 and layer 3 respectively. Because the architecture is almost correct, the similarity is high.

Week 2 and Week 8 also the high similarity, but they lower than Week 5. The reason is the architecture being approximately correct, but every layer has a few missing concepts. The following reasons reduce the similarity:

1: The knowledge map user selects incorrect concepts.

2: A few concepts that were mentioned by the teacher in the video were not explained in the handouts. Because the teacher wanted to further explain the handout content, new concepts did not appear on the handouts.

The similarity of week 1, week 3 and week 7 is about 0.7. Because the concepts in the architecture of the layer one are similar, layer one has a few unnecessary concepts. The first layer is the outline of handout, so a few unnecessary concepts are crawled.

The similarity of week 4, week 6 and week 9 is about 0.65. These ones have all problems of the above.

6.1.2 Introduction to Computer Networks

We will show the knowledge maps for the all weeks of a ”Introduction to Computer Networks” (Fig. 6.11, 6.12, 6.13, 6.14, 6.15, 6.16, 6.17, 6.18). The knowledge maps respectively created by experts and our system. Table 6.3 shows the similarity for the knowledge maps respectively created by experts and our system.

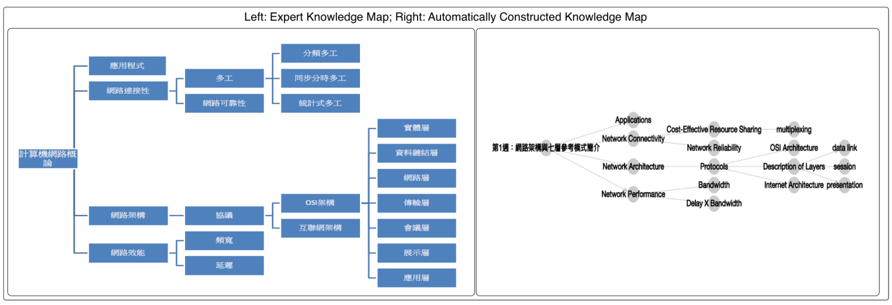

Figure 6.11: The knowledge maps for week 1 of a ”Introduction to Computer Networks”.

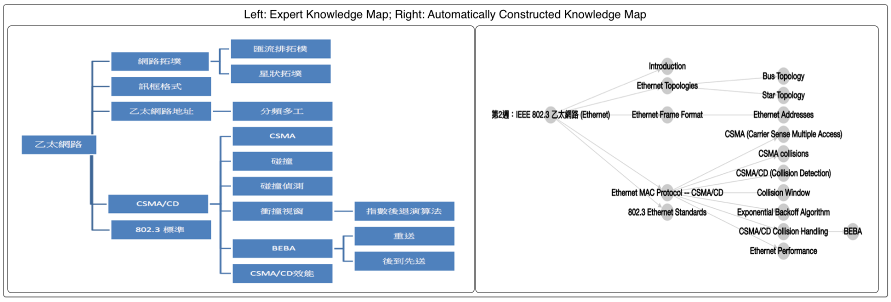

Figure 6.12: The knowledge maps for week 2 of a "Introduction to Computer Networks".

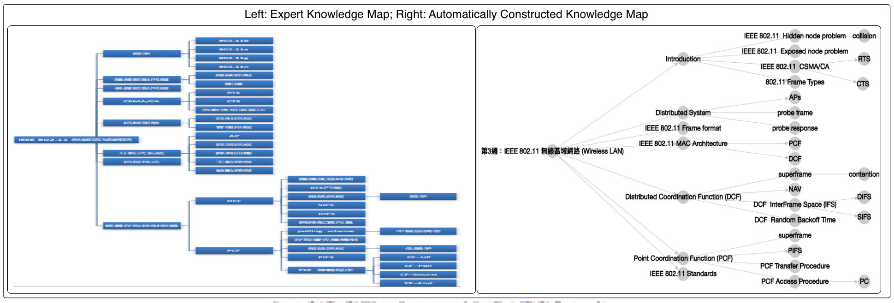

Figure 6.13: The knowledge maps for week 3 of a "Introduction to Computer Networks".

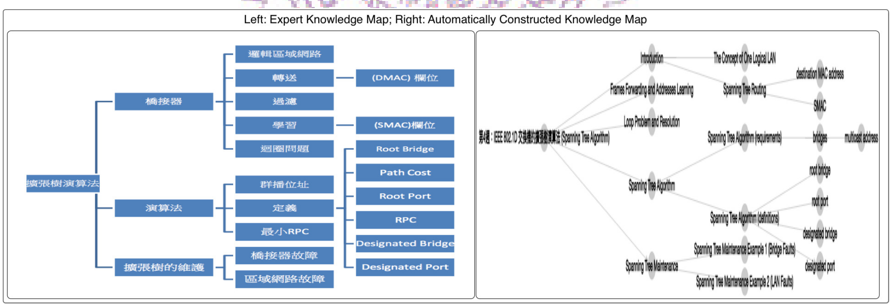

Figure 6.14: The knowledge maps for week 4 of a "Introduction to Computer Networks".

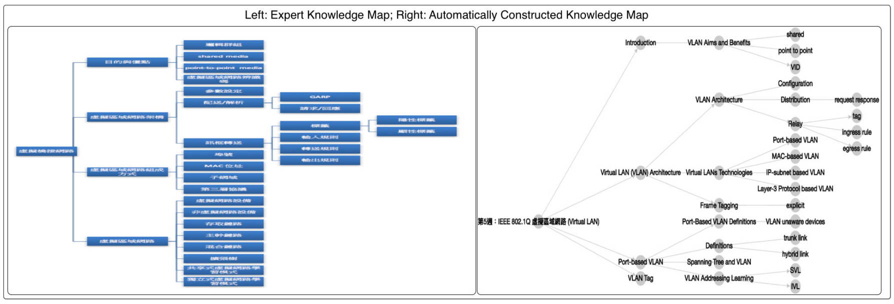

Figure 6.15: The knowledge maps for week 5 of a "Introduction to Computer Networks".

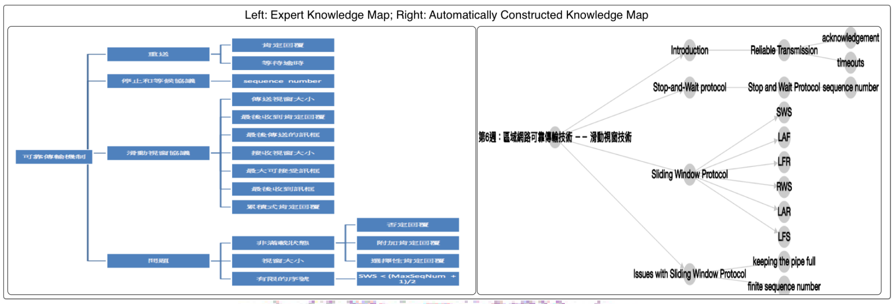

Figure 6.16: The knowledge maps for week 6 of a "Introduction to Computer Networks".

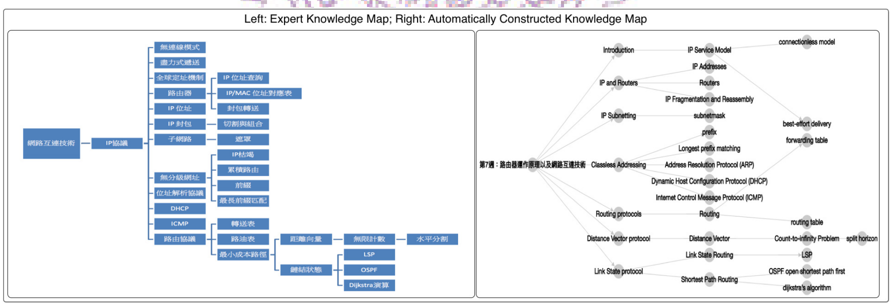

Figure 6.17: The knowledge maps for week 7 of a "Introduction to Computer Networks".

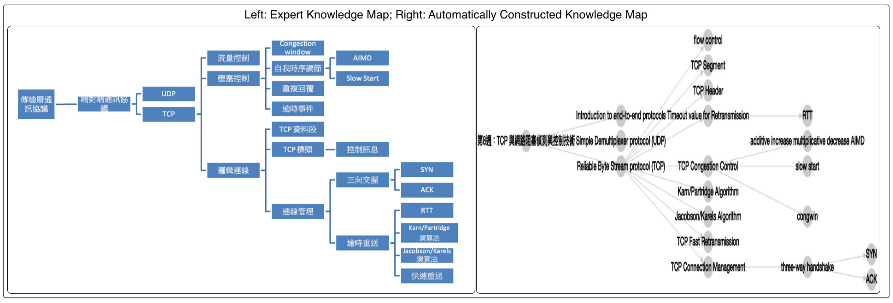

Figure 6.18: The knowledge maps for week 8 of a ”Introduction to Computer Networks”.

Table 6.3: The similarity for all weeks of a ”Introduction to Computer Networks”.

	Similarity (0.0 - 1.0)
Week 1	0.88781
Week 2	0.89935
Week 3	0.81752
Week 4	0.80985
Week 5	0.90202
Week 6	0.89483
Week 7	0.66989
Week 8	0.77223
Average	0.83169

As seen from the table 6.3, week 5 have the highest similarity. The reason is the concepts and relations in the automatically constructed knowledge map is similar to the content of expert knowledge map. Because the architecture is almost correct, the similarity is high.

Week 1, Week 2, Week 3, Week 4 and Week 6 also the high similarity, but they lower than Week 5. The reason is the architecture being approximately correct, but every layer has a few missing concepts.

The similarity of week 8 is about 0.77. Because the concepts in the architecture of the layer one are similar, layer one has a few unnecessary concepts. The first layer is the outline of handout, so a few unnecessary concepts are crawled.

The similarity of week 7 is about 0.66. These ones have all problems of the above.

Despite the above problem, the similarity of our system still high at almost 80%. We only a short amount of time was required to obtain the relevant information from the automatic knowledge map. If we lost some information, the expert could still add concepts to our system. Experts can use this flexible system to quickly create knowledge maps. With the correct keywords and accurate layer assignment, high-precision knowledge maps can be generated.

6.2 Analysis the Questionnaire

The outcome of the questionnaire is to measure students' easy-to-use, usability and attitudes to use knowledge map and their perceptions about learning "Topics on Investment" using knowledge map. Total number of quantify the questions are 19. Cronbach's alpha of the questionnaire is 0.926, the internal consistency is excellence. The 51.6 per cent of filling out the questionnaire is boy. The 77.4 per cent of filling out the questionnaire is sophomore. The 54.9 per cent of filling out the questionnaire is major in technology management. The score of the midterm test is higher than 85 is 41.9 per cent of filling out the questionnaire. The GPA of the first semester is higher than 4 is 25.8 per cent of filling out the questionnaire. The 48.4 per cent of filling out the questionnaire has studied online courses. Figure 6.19 6.22 6.25 6.28 shows the statistical data of a questionnaire.

Figure 6.19 shows the outcome of the questionnaire is to measure students' easy-to-use about learning "Topics on Investment" using knowledge map. Scores range from 3 to 6, and the average is about above 4. Based on the results, the system has given learners excellent service.

題目	個數	不同意(分數)	很同意(分數)	平均數	標準差
1. 「知識地圖」系統很容易使用	31	3	6	4.87	.670
2. 我在很短的時間就學會使用「知識地圖」系統	31	3	6	4.77	.762
3. 「知識地圖」系統的介面規劃符合操作習慣	31	3	6	4.74	.773
4. 「知識地圖」系統呈現的搜尋結果清楚易讀	31	3	6	4.77	.717

Figure 6.19: The outcome of the questionnaire is to measure students' easy-to-use.

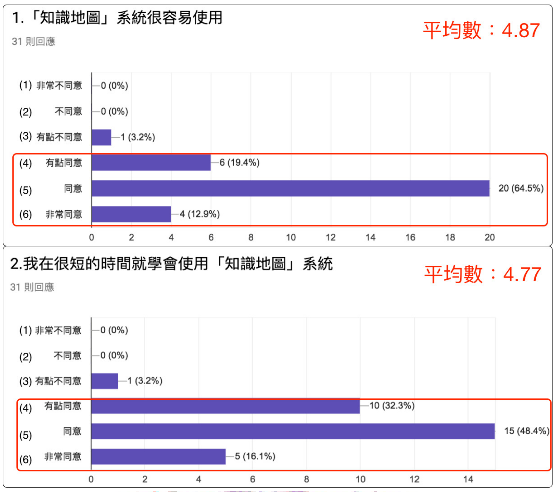

Figure 6.20: The line chart is to measure students' easy-to-use.

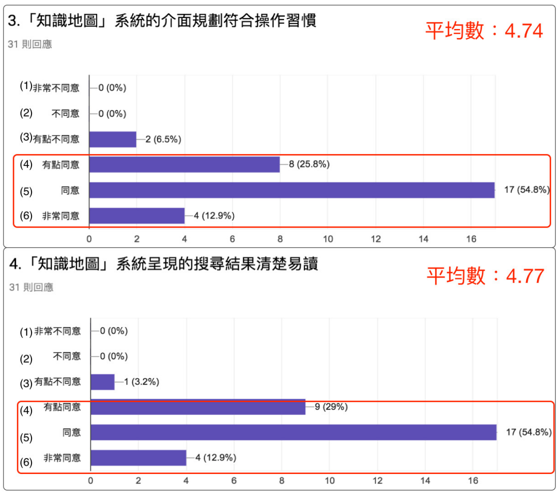

Figure 6.21: The line chart is to measure students' easy-to-use.

Figure 6.22 shows the outcome of the questionnaire is to measure students' usability about learning "Topics on Investment" using knowledge map. Scores range from 2 to 6, and the average is about above 4. Based on the results, the system help learners to understand their disorientation, and provide the sorted main points of the course. The learners also can understand the learning progress.

題目	個數	不同意(分數)	很同意(分數)	平均數	標準差
5. 使用「知識地圖」系統有助於我瞭解所要解決的問題	31	2	6	4.58	.886
6.「知識地圖」系統可以幫助我進行重點整理學習內容	31	3	6	4.52	.811
7.「知識地圖」系統提供一個好的解決問題的方式	31	3	6	4.58	.848
8.「知識地圖」系統對我解決問題是有幫助的	31	3	6	4.55	.723
9.「知識地圖」系統可以檢視自己的學習狀況	31	2	6	4.68	.945

Figure 6.22: The outcome of the questionnaire is to measure students' usability.

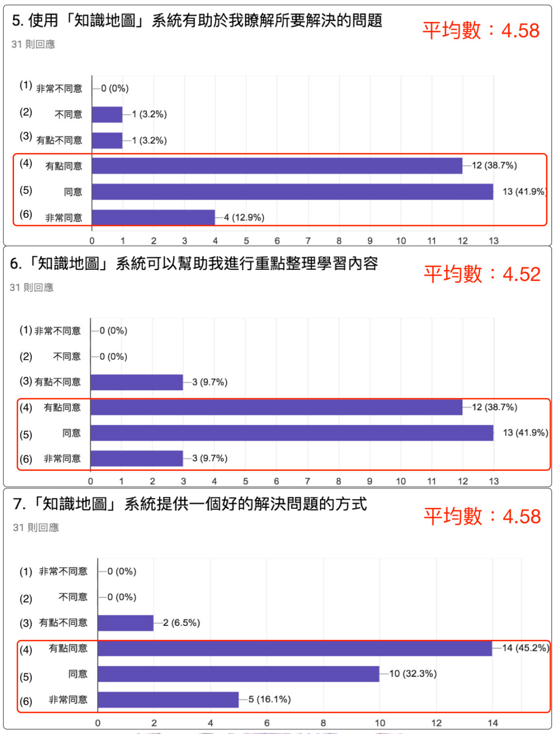

Figure 6.23: The line chart is to measure students’ usability.

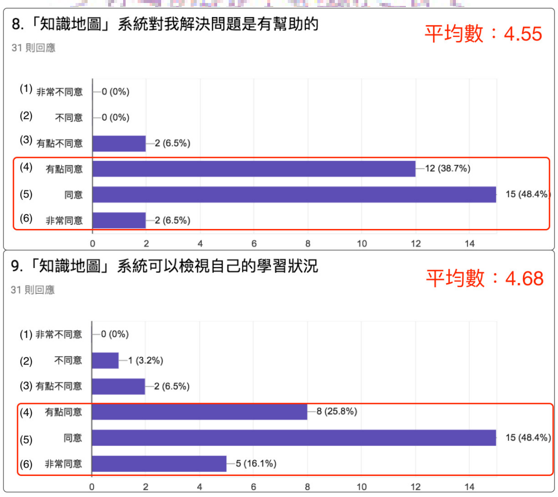

Figure 6.24: The line chart is to measure students’ usability.

Figure 6.25 shows the outcome of the questionnaire is to measure students’ attitudes about learning ”Topics on Investment” using knowledge map. Scores range from 1 to 6, and the average is about above 4. Based on the results, the learning attitudes is positive. The great majority of people want to learn more about the information of the course, and they would like to attempt the other course in the future.

題目	個數	不同意(分數)	很同意(分數)	平均數	標準差
10. 我比以前更有學習網路課程的意願	31	1	6	4.45	1.060
11. 未來我會想嘗試學習其他的網路課程	31	1	6	4.84	1.036
12. 未來我會主動找尋與這個課程相關的資料	31	4	6	4.74	.729
13. 我希望未來還能使用類似的方式進行學習	31	1	6	4.74	1.064

Figure 6.25: The outcome of the questionnaire is to measure students' attitudes.

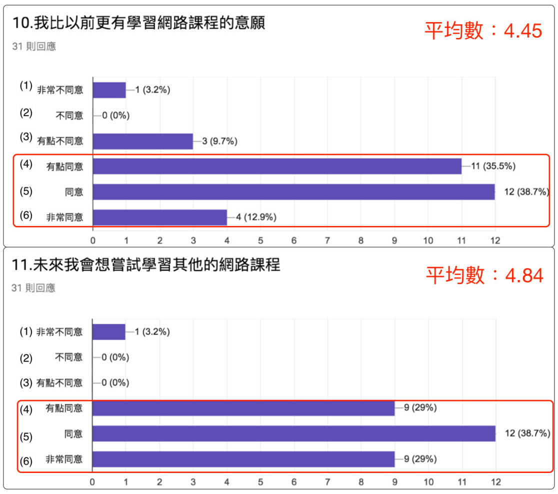

Figure 6.26: The line chart is to measure students' attitudes.

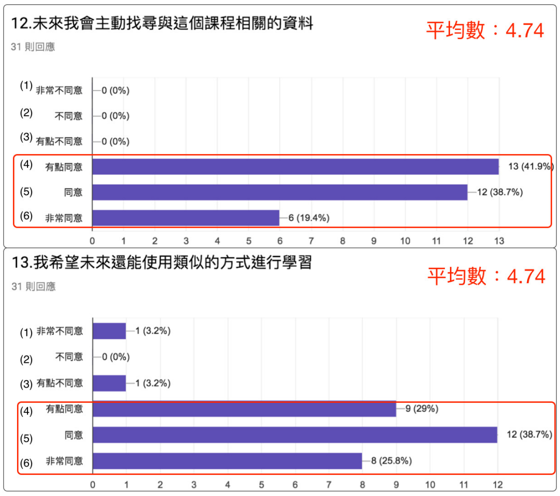

Figure 6.27: The line chart is to measure students' attitudes.

Figure 6.28 shows the outcome of the questionnaire is to measure students' perceptions about learning ”Topics on Investment” using knowledge map. Scores range from 3 to 6, and the average is about above 4. Based on the results, the learners are satisfied with the system, and they are interested to use this system. They can make use of the system to find the related videos and forums.

題目	個數	不同意 (分數)	很同意 (分數)	平均數	標準差
14. 整體而言，我對於使用「知識地圖」系統感到滿意	31	3	6	4.81	.703
15. 我認為「知識地圖」系統很有趣	31	3	6	4.68	.748
16. 我會利用「知識地圖」找相關資源，像是影片、討論區、習題	31	3	6	4.35	.798
17. 在觀看「知識地圖」時，我會確認自己是否理解	31	3	6	4.65	.839
18. 如果我不了解某個「知識地圖」上的課程概念，我會想辦法弄懂	31	3	6	4.61	.919
19. 我願意推薦其他人使用「知識地圖」系統	31	3	6	4.52	.769

Figure 6.28: The outcome of the questionnaire is to measure students' perceptions about learning.

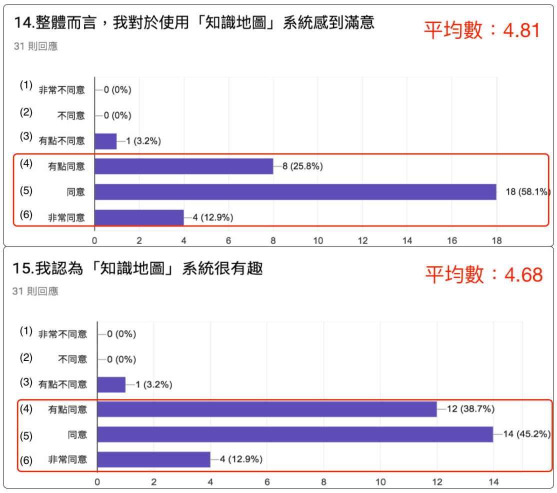

Figure 6.29: The line chart is to measure students' perceptions.

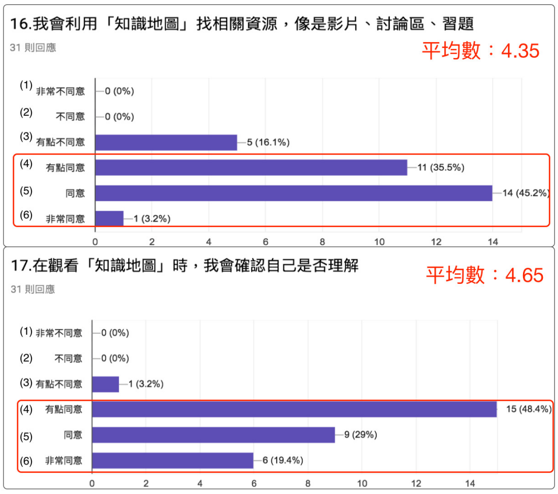

Figure 6.30: The line chart is to measure students' perceptions.

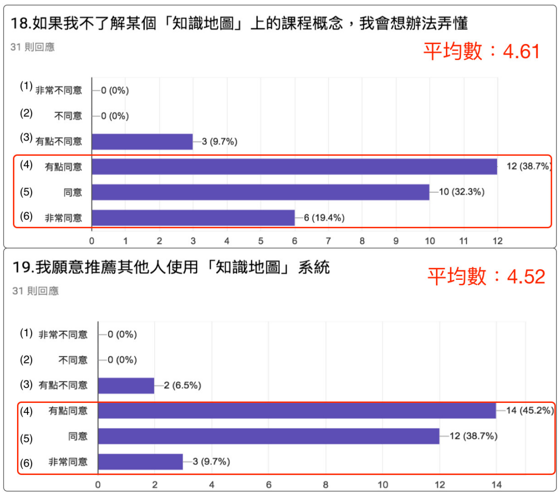

Figure 6.31: The line chart is to measure students' perceptions.

In most of students' opinion, the relation between the two concepts is correct. The students think that the knowledge map can preview before the course, and review after the course. The outcome of the questionnaire is almost positive responses.

Chapter 7

Conclusion and Future Work

Massive open online courses (MOOCs) offer valuable opportunities for freedom in learning; however, many learners face cognitive overload and conceptual and navigational disorientation. According to the results of a meta-analysis study [13], knowledge maps effectively enhance knowledge retention by assisting learners in recalling central ideas and forming summaries based on acquired knowledge. In fact, there's a lot of tool that created knowledge map nowadays, but user have to create manually. Several researchers have investigated the knowledge maps created automatically, but studies have seldom investigated using handouts to build knowledge maps. In this study, we used handouts to automatically build domain-specific knowledge maps for MOOCs. This feature is valuable for learners to quick identify the relation between the concepts in every week. This is also useful for the teacher, they only a short amount of time was required to obtain the relevant information from the automatic knowledge map. To the best of our knowledge, this is the first content based automatic knowledge map and recommendation system supports the Chinese-speaking environment. Furthermore, our experiments of a course prove the proposed the knowledge map generator is faster than manual knowledge map, moreover, it is quite accurate. The average similarity of our system is high at almost 80%, and the highest similarity one is 95.8%. Because the knowledge map creator can modify knowledge map, the modified knowledge map is more accurate than the original. Our system can quickly create and modify knowledge map, and the creator can obtain accurate knowledge map. In addition, the students think that the knowledge map can preview before the course, and review after the course. The findings suggest that using handouts to build knowledge maps is feasible.

Beyond design and implementation result, there is still a lot of improvement we could make to promote the usage or optimize the user perception in the future. Based on our observation, we can append the feature that filter the same synonyms and find the relation between the same synonyms. We can in the process of advancing the similarity by using with the methods such as word2vec. In addition, we want that expansion of the data range to construct more robust knowledge maps for MOOCs, and creation of a recommendation system to increase learning efficiency. We want to contribute to the users, so we can integrate their learning outcomes into knowledge map. From these considerations, we can redesign the intuitive user interface. If we have long document or handouts, we want to understand them quickly. For reasons above, we think we can use artificial intelligence for achieving refined handouts and documents. In addition, we want to be a third party service provider. They only need to upload files, we will provide the architecture of knowledge map.

Bibliography

[1] A. Ng and D. Koller, "Coursera," Retrieved April 16, 2017, from the World Wide Web: https://zh-tw.coursera.org/, 2012.

[2] M. I. of Technology and H. University, "edx," Retrieved April 16, 2017, from the World Wide Web: https://www.edx.org/, 2012.

[3] M. S. Sebastian Thrun, David Stavens, "Udacity," Retrieved April 16, 2017, from the World Wide Web: https://www.udacity.com/, 2012.

[4] S. Khan, "Khan academy," Retrieved April 16, 2017, from the World Wide Web: https://www.khanacademy.org/, 2006.

[5] M. Wang, J. Peng, B. Cheng, H. Zhou, and J. Liu, "Knowledge visualization for self-regulated learning." *Educational Technology & Society*, vol. 14, no. 3, pp. 28–42, 2011.

[6] J. A. Ruipérez-Valiente, P. J. Muñoz-Merino, C. D. Kloos, K. Niemann, M. Scheffel, and M. Wolpers, "Analyzing the impact of using optional activities in self-regulated learning," *IEEE Transactions on Learning Technologies*, vol. 9, no. 3, pp. 231–243, 2016.

[7] A. Sunar, S. White, N. Abdullah, and H. Davis, "How learners' interactions sustain engagement: a mooc case study," *IEEE Transactions on Learning Technologies*, 2016.

[8] C. G. Brinton, M. Chiang, S. Jain, H. Lam, Z. Liu, and F. M. F. Wong, "Learning about social learning in moocs: From statistical analysis to generative model," *IEEE transactions on Learning Technologies*, vol. 7, no. 4, pp. 346–359, 2014.

[9] C. Alario-Hoyos, M. Pérez-Sanagustín, C. Delgado-Kloos, M. Munoz-Organero *et al.*, "Delving into participants' profiles and use of social tools in moocs," *IEEE Transactions on Learning Technologies*, vol. 7, no. 3, pp. 260–266, 2014.

[10] H. A. Fasihuddin, G. D. Skinner, and R. I. Athauda, "Boosting the opportunities of open learning (moocs) through learning theories," *GSTF Journal on Computing (JoC)*, vol. 3, no. 3, p. 112, 2013.

[11] C. D. Holley and D. F. Dansereau, *Spatial learning strategies: Techniques, applications, and related issues.* Academic Press, 2014.

[12] P. W. Foos, "The effect of variations in text summarization opportunities on test performance." *The Journal of experimental education*, vol. 63, no. 2, pp. 89–95, 1995.

[13] J. C. Nesbit and O. O. Adesope, "Learning with concept and knowledge maps: A meta-analysis," *Review of educational research*, vol. 76, no. 3, pp. 413–448, 2006.

[14] H. Fasihuddin, G. Skinner, and R. Athauda, "Knowledge maps in open learning environments: An evaluation from learners\," *Journal of Information Technology and Application in Education*, vol. 4, pp. 18–29, 2015.

[15] C. YongYue and X. HuoSong, “Research on the auto-construction methods of concept map,” in *Intelligent Human-Machine Systems and Cybernetics, 2009. IHMSC’09. International Conference on*, vol. 2. IEEE, 2009, pp. 75–77.

[16] G. P. Jain, V. P. Gurupur, J. L. Schroeder, and E. D. Faulkenberry, “Artificial intelligence-based student learning evaluation: a concept map-based approach for analyzing a student’s understanding of a topic,” *IEEE Transactions on Learning Technologies*, vol. 7, no. 3, pp. 267–279, 2014.

[17] J. H. Lee and A. Segev, “Knowledge maps for e-learning,” *Computers & Education*, vol. 59, no. 2, pp. 353–364, 2012.

[18] I. M. Greca and M. A. Moreira, “Mental models, conceptual models, and modelling,” *International Journal of Science Education*, vol. 22, no. 1, pp. 1–11, 2000.

[19] M. I. of Technology and H. University, “Open edx,” Retrieved April 16, 2017, from the World Wide Web: https://open.edx.org/, 2012.

[20] ——, “edx help center,” Retrieved April 16, 2017, from the World Wide Web: https://support.edx.org/hc/en-us, 2012.

[21] F. X. ZHOU, “Junyi academy,” Retrieved April 16, 2017, from the World Wide Web: https://www.junyiacademy.org/, 2012.

[22] T. O. University, “Futurelearn,” Retrieved April 16, 2017, from the World Wide Web: https://www.futurelearn.com/, 2012.

[23] P. Wappett, “Open universities australia,” Retrieved April 16, 2017, from the World Wide Web: https://www.open2study.com/courses, 2013.

[24] T. University, "Xuetangx," Retrieved April 16, 2017, from the World Wide Web: http://www.xuetangx.com/, 2013.

[25] D. Shah, "Xuetangx: A look at china's first and biggest mooc platform," Retrieved April 16, 2017, from the World Wide Web: https://www.class-central.com/report/xuetangx/, 2016.

[26] "Knowledge map," Retrieved April 16, 2017, from the World Wide Web: http://khanacademy.wikia.com/wiki/Knowledge_Map.

[27] A. Leff and J. T. Rayfield, "Web-application development using the model/view/controller design pattern," in *Proceedings of the 5th IEEE International Conference on Enterprise Distributed Object Computing*, ser. EDOC '01. Washington, DC, USA: IEEE Computer Society, 2001, pp. 118–. [Online]. Available: http://dl.acm.org/citation.cfm?id=645344.650161

[28] Microsoft, "Model-view-controller," Retrieved April 16, 2017, from the World Wide Web: https://msdn.microsoft.com/en-us/library/ff649643.aspx.

[29] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information processing & management*, vol. 24, no. 5, pp. 513–523, 1988.

[30] J. Leskovec, A. Rajaraman, and J. D. Ullman, *Mining of massive datasets*. Cambridge University Press, 2014.

[31] D. R. Hipp, "Sqlite," Retrieved April 16, 2017, from the World Wide Web: https://www.sqlite.org/, 2000.

[32] B. J. Flood, "Historical note: The start of a stop list at biological abstracts," *Journal of the American Society for Information Science*, vol. 50, no. 12, pp. 1066–1066,

1999. [Online]. Available: http://dx.doi.org/10.1002/(SICI)1097-4571(1999)50:12⟨1066::AID-ASI5⟩3.0.CO;2-A

[33] "Jieba," Retrieved April 16, 2017, from the World Wide Web: https://github.com/fxsjy/jieba.

[34] H. P. Luhn, "A statistical approach to mechanized encoding and searching of literary information," *IBM Journal of Research and Development*, vol. 1, no. 4, pp. 309–317, Oct 1957.

[35] Y. Shinyama, "Pdfminer," Retrieved April 16, 2017, from the World Wide Web: https://github.com/euske/pdfminer, 2004.

[36] N. T. University, "Sharecourse," Retrieved April 16, 2017, from the World Wide Web: http://www.sharecourse.net/sharecourse/, 2012.

[37] D. C. at the University of Toronto, "Cytoscape.js," Retrieved April 16, 2017, from the World Wide Web: http://js.cytoscape.org/, 2015.

[38] J. T. Mark Otto, "Bootstrap," Retrieved April 16, 2017, from the World Wide Web: http://getbootstrap.com/.

[39] P.-W. O. Yi-Wei Huang, "Wordclouds," Retrieved June 27, 2017, from the World Wide Web: http://wordclouds.sharecourse.net/, 2017.

[40] N. T. University, "Topics on investment," Retrieved June 21, 2017, from the World Wide Web: http://www.sharecourse.net/sharecourse/course/view/courseInfo/1026, 2012.

[41] ——, "Introduction to computer networks," Retrieved June 21, 2017, from the World Wide Web: http://www.sharecourse.net/sharecourse/course/view/courseInfo/908, 2012.

[42] "Cosine similarity," Retrieved June 21, 2017, from the World Wide Web: https://en.wikipedia.org/wiki/Cosine_similarity.

[43] Z. Yihua, Z. Chaoqun, Z. Tong, W. Longfeng, X. Mali, W. Dongbo, and L. Xiaohui, "The research of recognizing the reviews in webpages based on calculating the similarity of dom-subtrees," *New Technology of Library and Information Service*, vol. 11, p. 012, 2013.

[44] "Small private online course," Retrieved June 21, 2017, from the World Wide Web: https://en.wikipedia.org/wiki/Small_private_online_course.

[45] Y.-W. Huang, "Experiment questionnaire," Retrieved June 21, 2017, from the World Wide Web: https://goo.gl/PdiHov, 2017.

[46] "Cronbach's alpha," Retrieved June 21, 2017, from the World Wide Web: https://en.wikipedia.org/wiki/Cronbach%27s_alpha.

[47] "Cronbach's alpha using spss statistics," Retrieved June 21, 2017, from the World Wide Web: https://statistics.laerd.com/spss-tutorials/cronbachs-alpha-using-spss-statistics.php.

[48] "Ibm spss software," Retrieved June 21, 2017, from the World Wide Web: https://www.ibm.com/analytics/us/en/technology/spss/.

[49] “Spss statistics base 17.0 user’s guide,” Retrieved June 21, 2017, from the World Wide Web: https://www.jou.ufl.edu/assets/researchlab/SPSS-Statistcs-Base-Users-Guide-17.0.pdf.