國立中央大學

資訊工程學系
碩士論文

基於職業技能需求之線上課程推薦系統
An On-line Courses Recommendation System based on Industry Occupation Skill Requirements

研究生：王海慧

指導教授：施國琛教授

中華民國 106 年 6 月

摘要

近年，由於資訊科技的進步，大規模網路開放式課程(MOOCs)於數位學習的研究領域中逐漸流行與普及。利用翻轉教室的概念讓學生自行於線上數位學習平台使用課本、影片、教學等素材進行學習。

然而，在台灣常見的一個問題則是教育與實際職業之間的差距。大學應屆畢業生通常不完全具備產業所需的技能，從大多的情況看來學生們在大學時期所學到只包含常用或較為基本的技能。隨著網際網路的發達，隨手可得的學習素材越多，我們希望能夠鼓勵學生利用大學四年的時間自我學習、充實。而如何推薦學生選擇正確的技能來學習，將是這篇論文所要探討的。

這篇論文提出了一個分群演算法，將 104 人力銀行上的職缺資料蒐集且處理過後，將每個職缺所需的技能統整且分群。讓學生能夠對每個職稱所需的技能分群有大概的了解，進而幫助學生們自我學習這類的技能，提升畢業後找尋工作的成功率。

關鍵字：MOOCs、人力銀行、課程推薦、分群、職業技能。

Abstract

MOOCs had bring us to a higher education with the concept of flipped classrooms, where students make use of the online studying materials such as online textbooks, video tutorials, and all sorts of documents which may take in forms of a web page, online learning platform, educational learning management systems. We see the stupendous potential of MOOCs in education.

However, there has always been a problem that existed in Taiwan that is also often discussed. It is known as the gap between industry and education, which means that the students who has graduated from universities, do not always have the skills that the industries needed. We find that in most cases, students will only have some skills or knowledge about some tools that is listed from the requirements of the industries. The students have plentiful self-studying resources from the internet, we hope to encourage the students to learn and empower themselves by correctly recommend what are the most required skills of their desired occupation. Therefore, this paper proposed a clustering method that shows the results of groups of skills that are commonly needed for a particular type of job

This system hopes to solve the problem known as the gap between industry and education, which the students, who had graduated from universities, do not always have the skills that the industries needed. In most cases, students will only have some skills or knowledge about some tools that is listed from the requirements of the industries.

We encourage the students to self-study the courses according to the course map formed from their desired occupation, and therefore increase the chances that they will get the job offer.

Keywords—MOOCs; Job Hunting; Course Recommender; Clustering; Skills.

Acknowledgements

首先感謝施國琛教授的指導。施教授帶領我踏入 E-Learning 的領域，指出研究方向，並且讓我在研究上有自我發展的空間，能夠完成這篇碩士論文。在撰寫論文的過程中，施教授也適時地進行監督與指導,使本篇碩士論文能夠順利完成。同時也要感謝參與口試的口試委員們，蘇柏齊教授以及孫士韋教授。在教授們的指導以及建議下,本篇碩士論文得以更加完善、嚴謹。

在這兩年的日子裡，感謝實驗室所有人使我的研究生活能夠絢麗與多彩。特別感謝陳禹任學長、邱翊學長、張閔翰學長，給了我許多的建議，幫助我確認研究的方向。感謝同屆的好友們鄭羽婷、黃韻文、鄭凱仁、林書勤。我們一同奮鬥的生活點滴將永不忘記。感謝已經畢業的陳振榮學長、劉書豪學長、莫展霖學長、黃也恩學姊、賀苒學姊、林宥均學長還有碩一學弟們陳璞、王振廷、林廷亮以及蘇炯綸。

當然也要感謝從大學時期就一直指導且幫助我的黃信貿教授、葉建華教授以及王慶生教授。還有我親愛的專題組員黃騰嶢與彭俊龍，摯友賴欣妤、楊博鈞。

最後，感謝我摯愛的母親，因為有你的支持，我才能夠走到這裡。今後我也不會讓您失望的，也感謝舅舅、外婆、表弟妹們的支持，因為有你們，我才有勇氣繼續前進。

Contents

List of Figures

Chapter 1. Introduction

With the rapid development of technology, human's daily life becomes more and more convenient with the growth of technology. This has changed our way of living, the changes that affected traditional learning. In traditional learning, students sit in classrooms listen to the educator on stage to learn. Now days, we have a better way of learning, the concept of the flipped classrooms, where students learn from online resources such as online textbooks, video, tutorials. These learning materials are often found from online learning platforms namely edX, Coursera, Udacity etc., or educational learning management systems from different university or colleges, where educators uploads their course materials to the online learning platform. Students can therefore start their learning progress with the online learning materials provided. This concept of modern learning style is also known as MOOCs, massive open online course, defined as an online course aimed at unlimited participation and open access via the web. While MOOCs becoming popular, more and more online courses will be found on web.

As the increasing amount and different categories of the online courses, it becomes crucial for the students to decide what to study from, which course to take with their limited spare time. Apart from that, another issue that is related students' learning and it is also a big problem we face in Taiwan is that students whom graduated from universities do not always have the necessary skills that the company employer required.

For example, an employee requirement usually consists of a list of required skills or tools, but in the real situation, the majority of the students, whom just graduated from university, will only be able to meet a small portion of the required tools. This is also known as the gap between industry and education. However, it will take a lot of effort and time for our education system to advance to a stage for the gap to be closed. It is then impossible to change the way of education immediately.

Thus, how to solve this problem would be the main contribution of this research. It is mentioned earlier that since the raise of MOOCs, we have the necessary resources for studying, it will be very helpful to advice the students what to study in order for them to get prepared when getting a job. Therefore, in this research, the job employment data of a popular job hunting site, www.104.com.tw, is collected for the required skills or tools references, since it is the actual requirement from the industry. Then, these data is clustered in different clusters to show that some set of skills or tools are often needed for a specific type of job. In our case, this research only focuses on the Computer Science field related occupations. The result clusters will be showed in the form of a directed graph, which can represent a course map, this is to help the students to get an idea of the order to study, they can then find the related MOOC courses of the skills or tools, and self-study in their free time while completing university. This research is to help the students to enhance their own knowledge based on their occupation target in the industry, therefore increases the chances of getting the offer.

1.1 Background

1.1.1 MOOC

A massive open online course, which is an online course aimed at unlimited participation and open access from the web. In addition to traditional course materials such as filmed lectures, readings, and problem sets, many MOOCs provide interactive user forums to support community interactions among students, professors, and teaching assistants, usually called TAs. MOOCs are a recent and widely researched development in distance education which were first introduced in 2006 and emerged as a popular mode of learning in 2012.

Early MOOCs often emphasized open-access features, such as open licensing of content, structure and learning goals, to promote the reuse and remixing of resources. Some later MOOCs use closed licenses for their course materials while maintaining free access for students.

As MOOCs have evolved, there appear to be two distinct types: those that emphasize the connectivist philosophy, and those that resemble more traditional courses. To distinguish the two, Stephen Downes proposed the terms "cMOOC" and "xMOOC".

cMOOCs are based on principles from connectivist pedagogy indicating that material should be aggregated (rather than pre-selected), remixable, re-purposable, and feeding forward, for example, evolving materials should be targeted at future learning. cMOOC instructional design approaches attempt to connect learners to each other to answer questions and/or collaborate on joint projects. This may include emphasizing collaborative development of the MOOC. Andrew Ravenscroft of the London Metropolitan University claimed that connectivist MOOCs better support collaborative dialogue and knowledge building.

xMOOCs have a much more traditional course structure typically with a clearly specified syllabus of recorded lectures and self-test problems. They employ elements of the original MOOC, but are, in effect, branded IT platforms that offer content distribution partnerships to institutions. The instructor is the expert provider of knowledge, and student interactions are usually limited to asking for assistance and advising each other on difficult points.[1]

In this research, our system will suggest the course map of the different type of tools or skills. Therefore, the actual improvement of the students' knowledge will be based on students self-learning, which is more likely to be classified as xMOOCs.

[1] https://en.wikipedia.org/wiki/Massive_open_online_course

1.1.2 edX

edX is a massive open online course, which uses the concept of MOOC mentioned earlier, provider. It hosts online university-level courses in a wide range of disciplines to a worldwide student body, including some courses at no charge. It also conducts research into learning based on how people use its platform. EdX differs from other MOOC providers, such as Coursera and Udacity, in that it is a nonprofit organization and runs on the Open edX open-source software.

The Massachusetts Institute of Technology and Harvard University created edX in May 2012. More than 70 schools, nonprofit organizations, and corporations offer or plan to offer courses on the edX website. As of 29 December 2016, edX has around 10 million students taking more than 1,270 courses online.

EdX courses consist of weekly learning sequences. Each learning sequence is composed of short videos interspersed with interactive learning exercises, where students can immediately practice the concepts from the videos. The courses often include tutorial videos that are similar to small on-campus discussion groups, an online textbook, and an online discussion forum where students can post and review questions and comments to each other and teaching assistants. Where applicable, online laboratories are incorporated into the course. For example, in edX's first MOOC, “a circuits and electronics course”, students built virtual circuits in an online lab.

EdX offers certificates of successful completion and some courses are crediteligible. Whether or not a college or university offers credit for an online course is within the sole discretion of the school. EdX offers a variety of ways to take courses, including verified courses where students have the option to audit the course , which is no cost, or to work toward an edX Verified Certificate, when fees vary by course. For courses announced before December 7, 2015, there was an option to take honor code courses to work toward an Honor Code Certificate, which is also no cost.

EdX also offers XSeries Certificates for completion of a bundled set of two to seven verified courses in a single subject which cost varies depending on the courses.

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. A team of researchers at Harvard and MIT, led by David Pritchard and Lori Breslow, released their initial findings in 2013.

EdX member schools and organizations also conduct their own research using data collected from their courses. Research focuses on improving retention, course completion and learning outcomes in traditional campus courses and online.

EdX has engaged in a number of partnerships with educational institutions in the United States, China, Mongolia, Japan, and more to utilize edX courses in "blended classrooms." In blended learning models, traditional classes include an online interactive component. San Jose State University (SJSU) partnered with edX to offer 6.00xL Introduction to Computer Science and Programming, as a blended course at

SJSU and released an initial report on the project in February 2013. Initial results showed a decrease in failure rates from previous semesters. The percentage of students required to retake the course dropped from 41% under the traditional format to 9% for those taking the edX blended course. In Spring 2013, Bunker Hill Community College and Massachusetts Bay Community College implemented a SPOC, or small private online course. The colleges incorporated an MIT-developed Python programming course on EdX into their campus-based courses, and reported positive results.

Open edX is the open-source platform software developed by EdX and made freely available to other institutions of higher learning that want to make similar offerings. On June 1, 2013, edX open sourced its entire platform. The source code can be found on GitHub.[2]

In this research, edX is the learning platform that is linked from the system, to find the related courses of the tools and the skills. Since it concludes a large variety of courses on its platform, and from all around different universities and colleges.

[2] https://en.wikipedia.org/wiki/EdX

1.1.3 Job hunting

Job hunting, job seeking, or job searching is the act of looking for employment, due to unemployment, underemployment, discontent with a current position, or a desire for a better position. The immediate goal of job seeking is usually to obtain a job interview with an employer which may lead to getting hired. The job hunter or seeker typically first looks for job vacancies or employment opportunities.

Many job seekers research the employers to which they are applying, and some employers see evidence of this as a positive sign of enthusiasm for the position or the company, or as a mark of thoroughness. Information collected might include open positions, full name, locations, web site, business description, year established, revenues, number of employees, stock price if public, name of chief executive officer, major products or services, major competitors, and strengths and weaknesses.[3]

[3] https://en.wikipedia.org/wiki/Job_hunting

1.1.4 Employment website

An employment website is a website that deals specifically with employment or careers. Many employment websites are designed to allow employers to post job requirements for a position to be filled and are commonly known as job boards. Other employment sites offer employer reviews, career and job-search advice, and describe different job descriptions or employers. Through a job website a prospective employee can locate and fill out a job application or submit resumes over the Internet for the advertised position.

The success of jobs search engines in bridging the gap between jobseekers and employers has spawned thousands of job sites, many of which list job opportunities in a specific sector, such as education, health care, hospital management, academics and even in the non-governmental sector. These sites range from broad all-purpose generalist job boards to niche sites that serve various audiences, geographies, and industries. Many industry experts are encouraging jobseekers to concentrate on industry specific sector sites. Types of different employment websites are classified as follows.

1. Job postings

A job board is a website that facilitates job hunting and range from large scale generalist sites to niche job boards for job categories such as engineering, legal, insurance, social work, teaching, mobile app development as well as cross-sector categories such as green jobs, ethical jobs and seasonal jobs. Users can typically deposit their résumés and submit them to potential employers and recruiters for review, while employers and recruiters can post job ads and search for potential employees.

The term job search engine might refer to a job board with a search engine style interface, or to a web site that actually indexes and searches other web sites.

Niche job boards are starting to play a bigger role in providing more targeted job vacancies and employees to the candidate and the employer respectively. Job boards such as airport jobs and federal jobs among others provide a much focused way of eliminating and reducing time to applying to the most appropriate role. USAJobs.gov is the United States' official website for jobs. It gathers job listings from over 500 federal agencies.

2. Metasearch and vertical search engines

Some web sites are simply search engines that collect results from multiple independent job boards. This is an example of both metasearch, since these are search engines which search other search engines, and vertical search, since the searches are limited to a specific topic - job listings.

Some of these new search engines primarily index traditional job boards. These sites aim to provide a "one-stop shop" for job-seekers who don't need to search the underlying job boards. In 2006, tensions developed between the job boards and several scraper sites, with Craigslist banning scrapers from its job classifieds and Monster.com specifically banning scrapers through its adoption of a robots exclusion standard on all its pages while others have embraced them.

Indeed.com, a "job aggregator", collects job postings from employer websites, job boards, online classifieds, and association websites. Simply Hired is another large aggregator collecting job postings from many sources.

LinkUp, website, is a job search engine, "job aggregator", that indexes pages only from employers' websites choosing to bypass traditional job boards entirely. These vertical search engines allow jobseekers to find new positions that may not be advertised on the traditional job boards.

Industry specific posting boards are also appearing. These consolidate all the vacancies in a very specific industry. The largest "niche" job board is Dice.com which focuses on the IT industry. Many industry and professional associations offer members a job posting capability on the association website.

3. Employer review website

An employer review website is a type of employment website where past and current employees post comments about their experiences working for a company or organization. An employer review website usually takes the form of an internet forum. Typical comments are about management, working conditions, and pay.

Although employer review websites may produce links to potential employers, they do not necessarily list vacancies.

4. Pay For Performance (PFP)

The most recent second generation of employment websites, often referred to as pay for performance, PFP, involves charging for membership services rendered to job seekers.

5. Websites providing information and advice for employees, employers and job seekers

Although many sites that provide access to job advertisements include pages with advice about writing resumes and CVs, performing well in interviews, and other topics of interest to job seekers there are sites that specialize in providing information of this kind, rather than job opportunities. One such is “Working” in Canada. It does provide links to the Canadian Job Bank. However, most of its content is information about local labor markets in Canada, requirements for working in various occupations, information about relevant laws and regulations, government services and grants, and so on. Most items could be of interest to people in various roles and conditions including those considering career options, job seekers, employers and employees.[4]

In Taiwan, one of the most popular employment site is called 104, www.104.com.tw, it will be classified as the “Pay For Performance” type of employment website, however, only the employee company will need to be charged to post job requirements. It is completely free for the job seekers to look for jobs and post their curriculum vitae on the website. It is also possible for the employee to directly look for job seeker from their individual curriculum vitae.

[4] https://en.wikipedia.org/wiki/Employment_website

1.2 Motivation

In Taiwan, students at high school spent much effort and time trying to get high results in school exams in order to get in a better university or college. The one thing that most of the parents and student themselves forget about is the actual interest and desire of the students future career. This usually results in student change major during university or college and even change university or college. This process will take a lot of time and determination, retaking examinations to another university or college or changing major, these will be a waste of time and energy. However, even if the student took what they thought the right choice, what they learnt in their university or college do not always meet the industry requirements. Therefore, what to learn to achieve the industry’s requirement is crucial. This is the final goal in this society.

1.3 Thesis Organization

The whole thesis is organized as follow. Related works are discussed in Section II. In Section III, we will talk about the whole method in detail. In Section IV, we will show our experiment setup and results. In the end of this paper, we will discuss conclusion and some future works in Section V.

Chapter 2. Related Work

There are only quite a few of the researches that are related to this concept. Some like [1], which is published in 2011, where the goal is similar as to recommend software skills for jobs in the field of Internet Technology. In this research, they proposed a solution to assist employers when preparing advertisement via identification of suitable soft skills together with its relevancy to that particular job title. Bayesian network is employed to solve this problem because it is suitable for reasoning and decision making under uncertainty. The proposed Bayesian Network is trained using a dataset collected via extracting information from advertisements and also through interview sessions with a few identified experts.

Another part that is important in this research is the clustering algorithm for the skills and tools. K-means would be the basic clustering algorithm to use. However, it would be difficult to determine the vector with this method, and the results might also be affected by the way vectors defined. [2] had a K-means clustering in the situation that the vectors can be defined in the right way. Its objective is to propose an effective clustering technique to group users' sessions by modifying K-means algorithm and suggest a method to compute the distance between sessions based on similarity of their web access path, which takes care of the issue of the user sessions that are of variable length. The basic K-means algorithm initially selects the cluster centroids randomly and finds the new cluster centroid based on the average value obtained within each cluster in each iteration. In the modified K-means algorithm, the old cluster centroid is updated by the delta amount, where, delta is nothing but the average distance value of each cluster.

Therefore, the first step is to define the similarity between the skills and tools data collected. [3] has included various ways of deciding similarity, and even combines them together for improvement. It focuses on keywords search, and designed method for measuring keywords similarity with Jaccard's, N-Gram, Vector space, Average (JNVA) and Jaccard's, N-Gram, Length, Average (JNLA) by using hybrid method; a combination of Jaccard's , N-Gram and Vector Space to make Keywords search practical. These methods are evaluated by three criterions which are precision, recall, and F-measure. The result reveals that the method for measuring keywords similarity with the application of JNVA and JNLA can successfully predict the similarity between keywords query with index words. These methods can be applied in order to develop searching engines performance especially semantic search. [4] presented one such class of item-based recommendation algorithms that first determine the similarities between the various items and then used them to identify the set of items to be recommended. Cosine-Based Similarity and Conditional Probability-Based Similarity are mainly used for calculating similarities in this paper.

There are also some other recommendation system such as [5], where they proposed a practical and effective approach, in which they first analyze the correlation between students' achievement with their employment situation, whether they have obtained an employment or not, according history data. For the data of students obtained employment, they further discover association rules from students' achievement and concrete occupations by data mining. Moreover, for a new or not obtained employment student, they recommend appropriate occupation for him or her based on the above association rules.

Some kind of filtering method was also tried during the research, the most popular would be [6], from Amazon.com. They use recommendation algorithms to personalize the online store for each customer. Rather than matching the user to similar customers, item-to-item collaborative filtering matches each of the user's purchased and rated items to similar items, then combines those similar items into a recommendation list.

As well as some other algorithms used in similar cases, [7] applied data mining techniques to social networks to help users of the social digital media to distinguish these important friends from a large number of friends in their social networks.

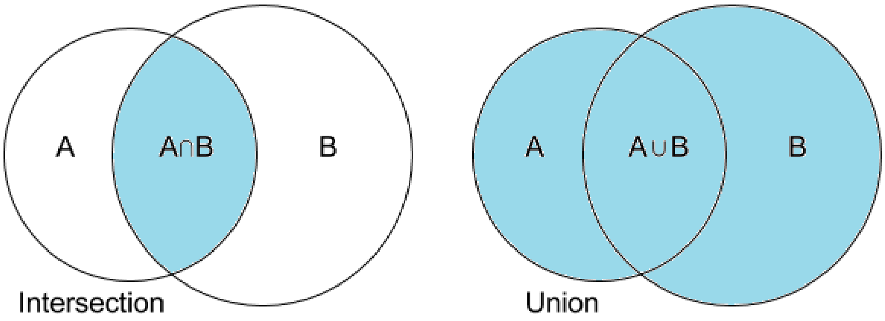
Other researches that also help students to get their desired occupations like [8] used developer profiles form GitHub to match the job advertisement. Employers and HR personnel who may use GitHub to learn more about a developer's skills and interests. They propose a pipeline that automatizes this process and automatically suggests matching job advertisements to developers, based on signals extracting from their activities on GitHub.

From the related works that are studied, some that could have been used in this system will be introduced as follows.

2.1 Keywords Similarity

The keyword search is the simplest and popular method for the search engine, and there are several ways to measure the similarity between keywords. The most adopted methods today are N-Gram, Vector Space, and Jaccard Similarity Coefficient. The Jaccard Similarity Coefficient can effectively measure the similarity of keywords. The contents that in LOR are described via metadata, in this study uses the Jaccard Similarity Coefficient to compare the similarity between metadata.

Jaccard Similarity Coefficient: The Jaccard Similarity Coefficient (known as Jaccard Index) measures the similarity between two objects which with n attributes, and the value is defined as the intersection of the attributes divided by the union of the attributes. [5]For the definition of the union here is any attributes of the object is established, and the definition of the intersection is both attributes of the object required and established. The concept of the intersection and the union are shown in Figure 1.



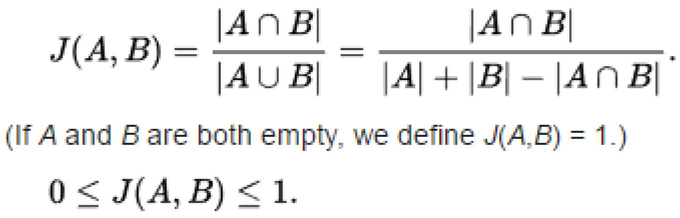
[5] https://en.wikipedia.org/wiki/Jaccard_index

Figure 1 The concept of intersection and union

Attributes Objects	Attribute A	Attribute B	Attribute C	Attribute D
Object A	1	0	0	1
Object B	1	0	1	0
Object C	1	1	0	1
Object D	0	0	0	1
Object E	1	1	1	0

Figure 2 Example of the objects sets of the Jaccard Index

The Jaccard Similarity value of Object A and Object C will be 2/3 =0.67.



$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}.$$

(If A and B are both empty, we define $J(A,B) = 1$.)

$$0 \le J(A,B) \le 1.$$

Chapter 3. Proposed Method

The System Structure is shown on Figure 3. Students are free to choose a desired occupation from the job list in the system, and a course map will be formed from the complete course map that is stored in the database by using the list of required skills of this specific job. Each course on the course map will link to the search results in top online MOOCs site. When there is a new skill that appears from the analyses of sorting the required skills, a notice will be send to all educators who registered for the system, and collect their feedbacks. The feedbacks will then be analyzed to update the course map in the database.

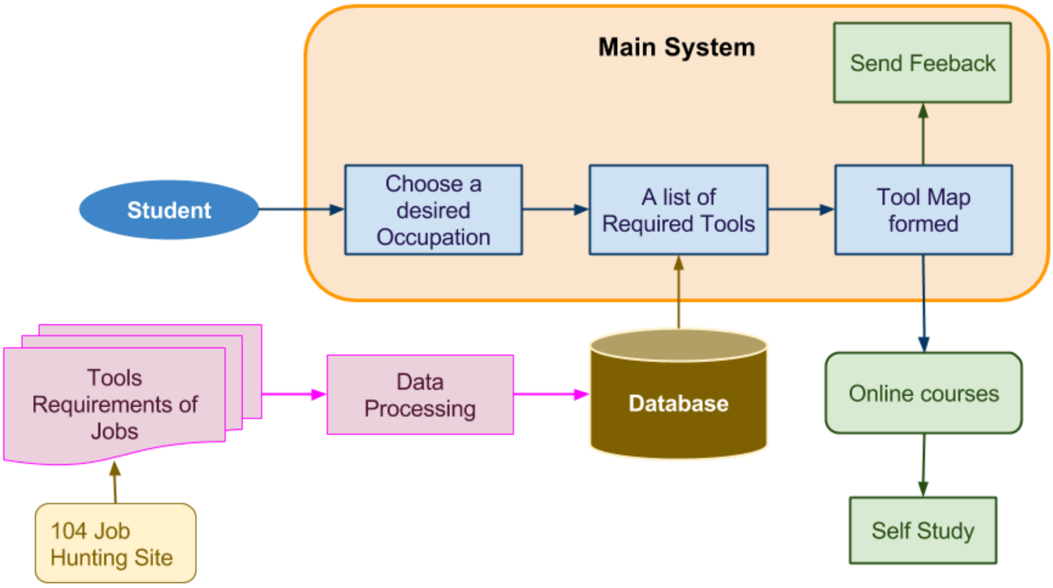
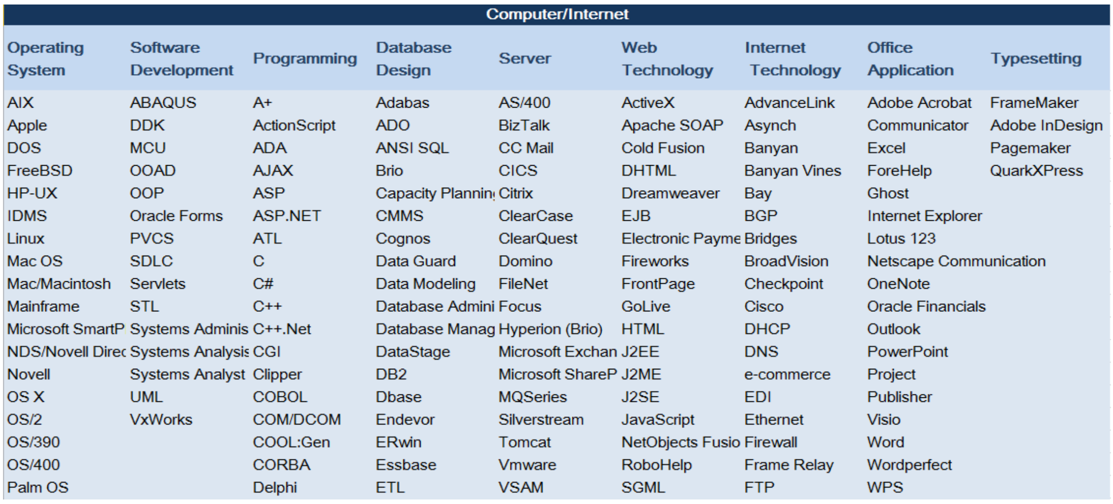


Figure 3 System Structure

3.1 Data Collection

In Taiwan, there are some popular job hunting web sites, which are also referred as an employment agency. An employment agency is an organization which matches employers to employees, where organizations posts job openings on the job hunting web sites with a certain amount of charge. Regular users are allowed to register with their identification number, then, write an own curriculum vitae in a personal page. The form of both data, from the organizations and users, are uploaded uniformly. Therefore, there is a standard input for the required skills. An example of the metadata of part of the "Computer/Internet" section is shown on Figure 4.

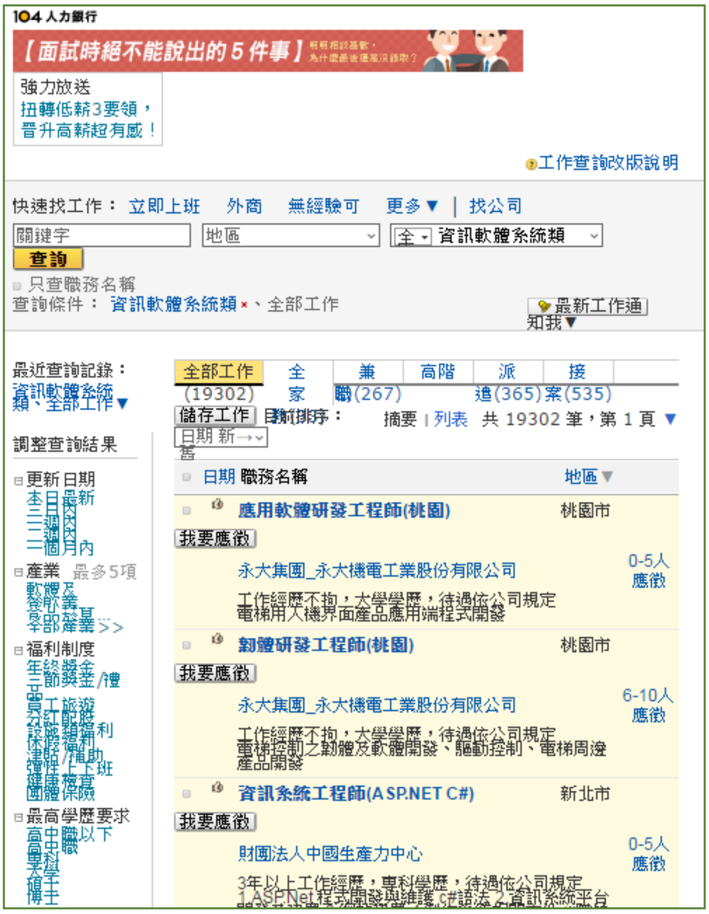


Computer/Internet								
Operating System	Software Development	Programming	Database Design	Server	Web Technology	Internet Technology	Office Application	Typesetting
AIX	ABAQUS	A+	Adabas	AS/400	ActiveX	AdvanceLink	Adobe Acrobat	FrameMaker
Apple	DDK	ActionScript	ADO	BizTalk	Apache SOAP	Asynch	Communicator	Adobe InDesign
DOS	MCU	ADA	ANSI SQL	CC Mail	Cold Fusion	Banyan	Excel	Pagemaker
FreeBSD	OOAD	AJAX	Brio	CICS	DHTML	Banyan Vines	ForeHelp	QuarkXPress
HP-UX	OOP	ASP	Capacity Plannin	Citrix	Dreamweaver	Bay	Ghost	
IDMS	Oracle Forms	ASP.NET	CMMS	ClearCase	EJB	BGP	Internet Explorer	
Linux	PVCS	ATL	Cognos	ClearQuest	Electronic Payme	Bridges	Lotus 123	
Mac OS	SDLC	C	Data Guard	Domino	Fireworks	BroadVision	Netscape Communication	
Mac/Macintosh	Servlets	C#	Data Modeling	FileNet	FrontPage	Checkpoint	OneNote	
Mainframe	STL	C++	Database Admini	Focus	GoLive	Cisco	Oracle Financials	
Microsoft SmartP	Systems Adminis	C++.Net	Database Manag	Hyperion (Brio)	HTML	DHCP	Outlook	
NDS/Novell Direc	Systems Analysis	CGI	DataStage	Microsoft Exchan	J2EE	DNS	PowerPoint	
Novell	Systems Analyst	Clipper	DB2	Microsoft ShareP	J2ME	e-commerce	Project	
OS X	UML	COBOL	Dbase	MQSeries	J2SE	EDI	Publisher	
OS/2	VxWorks	COM/DCOM	Endevor	Silverstream	JavaScript	Ethernet	Visio	
OS/390		COOL:Gen	ERwin	Tomcat	NetObjects Fusio	Firewall	Word	
OS/400		CORBA	Essbase	Vmware	RoboHelp	Frame Relay	Wordperfect	
Palm OS		Delphi	ETL	VSAM	SGML	FTP	WPS	

Figure 4 An example of the metadata of part of the "Computer/Internet" section

Since that the job skills requirements are in standard format, we are then able to collect the data from the job hunting web site by using web spider.

3.1.1 104 Job Hunting Site



104 人力銀行

【面試時絕不能說出的５件事】

強力放送
扭轉低薪3要領，
晉升高薪超有感！

工作查詢改版說明

快速找工作： 立即上班 外商 無經驗可 更多▼ | 找公司

關鍵字 | 地區 | 全 資訊軟體系統類

查詢

□ 只查職務名稱

查詢條件： 資訊軟體系統類 × 、全部工作

最新工作通知我▼

最近查詢記錄：
資訊軟體系統類、全部工作▼

調整查詢結果

⊟ 更新日期
本日最新
三日內
一週內
二週內
一個月內

⊟ 產業 最多5項
軟體及網路相關業
電腦及消費性電子製造業
全部產業>>

⊟ 福利制度
年終獎金
三節獎金/禮品
員工旅遊
分紅配股
設施類福利
休假福利
津貼/補助
彈性上下班
健康檢查
團體保險

⊟ 最高學歷要求
高中職以下
高中職
專科
大學
碩士
博士

全部工作 (19302) | 全家 | 兼職(267) | 高階 | 派遣(365) | 接案(535)

儲存工作 目前排序： 日期 新→舊 摘要 | 列表 共 19302 筆，第 1 頁 ▼

□ 日期 職務名稱 地區▼

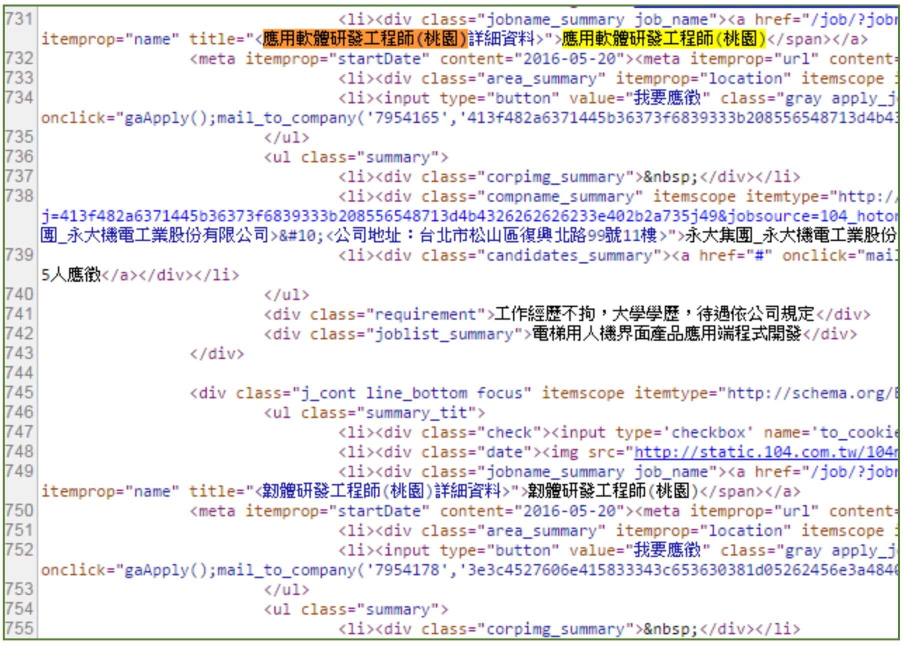
□ 應用軟體研發工程師(桃園) 桃園市
我要應徵
永大集團_永大機電工業股份有限公司 0-5人應徵
工作經歷不拘，大學學歷，待遇依公司規定
電梯用人機界面產品應用端程式開發

□ 韌體研發工程師(桃園) 桃園市
我要應徵
永大集團_永大機電工業股份有限公司 6-10人應徵
工作經歷不拘，大學學歷，待遇依公司規定
電梯控制之韌體及軟體開發、驅動控制、電梯周邊產品開發

□ 資訊系統工程師(ASP.NET C#) 新北市
我要應徵
財團法人中國生產力中心 0-5人應徵
3年以上工作經歷，專科學歷，待遇依公司規定
1.ASP.Net 程式開發與維護 c#語法 2.資訊系統平台

Figure 5 Example of Searching Results under Computer Science Section

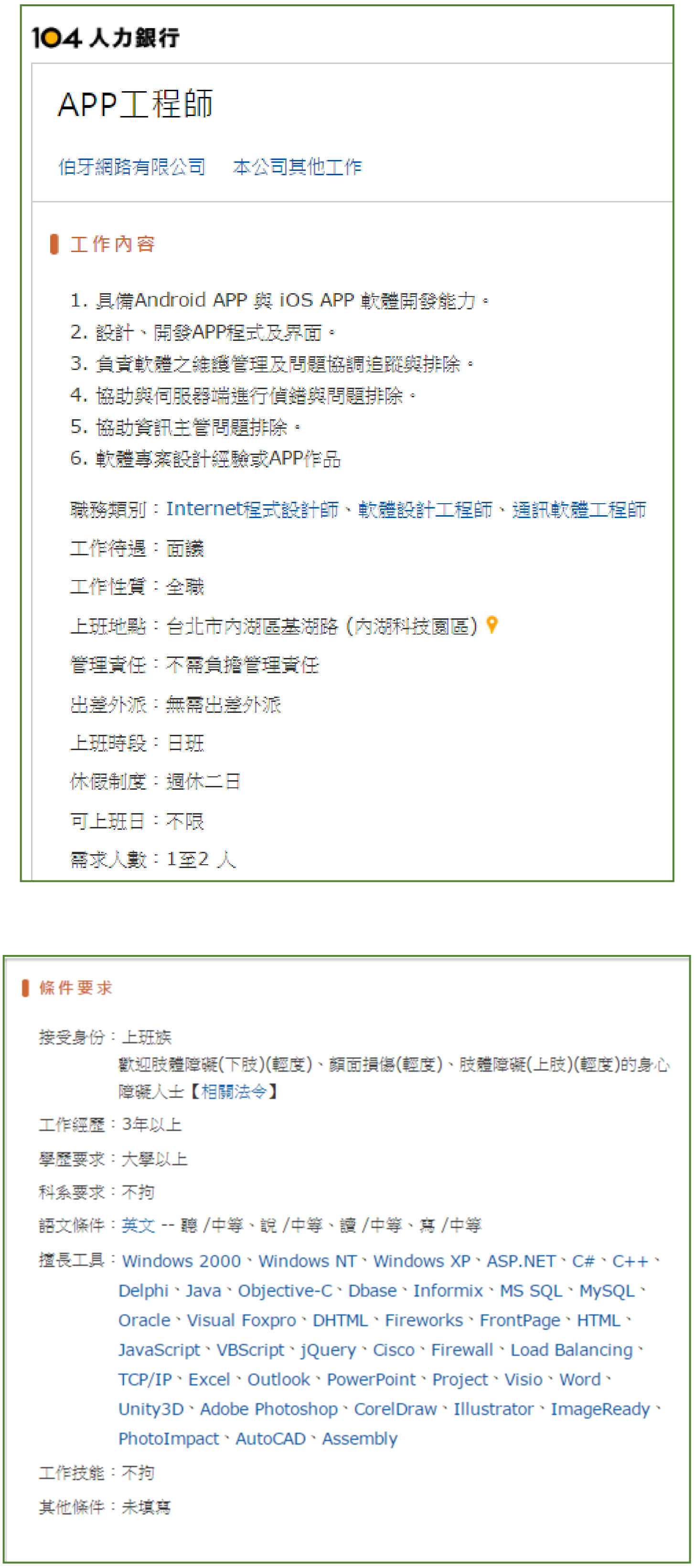
104 Job Hunting Site is one of the most popular job hunting site in Taiwan. The site is well organized and has uniform structured job description page. Figure 5 shows an example of searching results under computer science section. Since the site is structured format, we can use web crawler to retrieve all the information of the job description.



```
                                <li><div class="jobname_summary job_name"><a href="/job/?job
itemprop="name" title="<應用軟體研發工程師(桃園)詳細資料>">應用軟體研發工程師(桃園)</span></a>
                <meta itemprop="startDate" content="2016-05-20"><meta itemprop="url" content
                                <li><div class="area_summary" itemprop="location" itemscope 
                                <li><input type="button" value="我要應徵" class="gray apply_j
onclick="gaApply();mail_to_company('7954165','413f482a6371445b36373f6839333b208556548713d4b4
                        </ul>
                        <ul class="summary">
                                <li><div class="corpimg_summary">&nbsp;</div></li>
                                <li><div class="compname_summary" itemscope itemtype="http:/
j=413f482a6371445b36373f6839333b208556548713d4b4326262626233e402b2a735j49&jobsource=104_hoto
團_永大機電工業股份有限公司>&#10;<公司地址：台北市松山區復興北路99號11樓>">永大集團_永大機電工業股份
                                <li><div class="candidates_summary"><a href="#" onclick="mai
5人應徵</a></div></li>
                        </ul>
                        <div class="requirement">工作經歷不拘，大學學歷，待遇依公司規定</div>
                        <div class="joblist_summary">電梯用人機界面產品應用端程式開發</div>
                </div>

                <div class="j_cont line_bottom focus" itemscope itemtype="http://schema.org/
                        <ul class="summary_tit">
                                <li><div class="check"><input type='checkbox' name='to_cooki
                                <li><div class="date"><img src="http://static.104.com.tw/104
                                <li><div class="jobname_summary job_name"><a href="/job/?job
itemprop="name" title="<韌體研發工程師(桃園)詳細資料>">韌體研發工程師(桃園)</span></a>
                <meta itemprop="startDate" content="2016-05-20"><meta itemprop="url" content
                                <li><div class="area_summary" itemprop="location" itemscope 
                                <li><input type="button" value="我要應徵" class="gray apply_j
onclick="gaApply();mail_to_company('7954178','3e3c4527606e415833343c653630381d05262456e3a484
                        </ul>
                        <ul class="summary">
                                <li><div class="corpimg_summary">&nbsp;</div></li>
```

Figure 6 html of the Searching Results

Figure 6 shows the html file of the searching results under computer science section. The title of the first results is high lightered, we used python and beautiful soup to identify different attribute in the html file, and we can see the high lightered title is under the html tag “title”, by using the similar method, we need to write the code that fit this specific searching results page. Once we get the title and the linked web link for the search results, then we will be able to look at the next level of the searched results, which is the actual content of the job requirement of the specific jobs. The metadata in the job requirements are classified as follows: Job Title, Content, Category, Experience, Education, Major, Language, Tools, Skills and Others.



104 人力銀行

APP工程師

伯牙網路有限公司　本公司其他工作

工作內容

1. 具備Android APP 與 iOS APP 軟體開發能力。
2. 設計、開發APP程式及界面。
3. 負責軟體之維護管理及問題協調追蹤與排除。
4. 協助與伺服器端進行偵錯與問題排除。
5. 協助資訊主管問題排除。
6. 軟體專案設計經驗或APP作品

職務類別：Internet程式設計師、軟體設計工程師、通訊軟體工程師

工作待遇：面議

工作性質：全職

上班地點：台北市內湖區基湖路 (內湖科技園區)

管理責任：不需負擔管理責任

出差外派：無需出差外派

上班時段：日班

休假制度：週休二日

可上班日：不限

需求人數：1至2 人

條件要求

接受身份：上班族
歡迎肢體障礙(下肢)(輕度)、顏面損傷(輕度)、肢體障礙(上肢)(輕度)的身心障礙人士【相關法令】

工作經歷：3年以上

學歷要求：大學以上

科系要求：不拘

語文條件：英文 -- 聽 /中等、說 /中等、讀 /中等、寫 /中等

擅長工具：Windows 2000、Windows NT、Windows XP、ASP.NET、C#、C++、Delphi、Java、Objective-C、Dbase、Informix、MS SQL、MySQL、Oracle、Visual Foxpro、DHTML、Fireworks、FrontPage、HTML、JavaScript、VBScript、jQuery、Cisco、Firewall、Load Balancing、TCP/IP、Excel、Outlook、PowerPoint、Project、Visio、Word、Unity3D、Adobe Photoshop、CorelDraw、Illustrator、ImageReady、PhotoImpact、AutoCAD、Assembly

工作技能：不拘

其他條件：未填寫

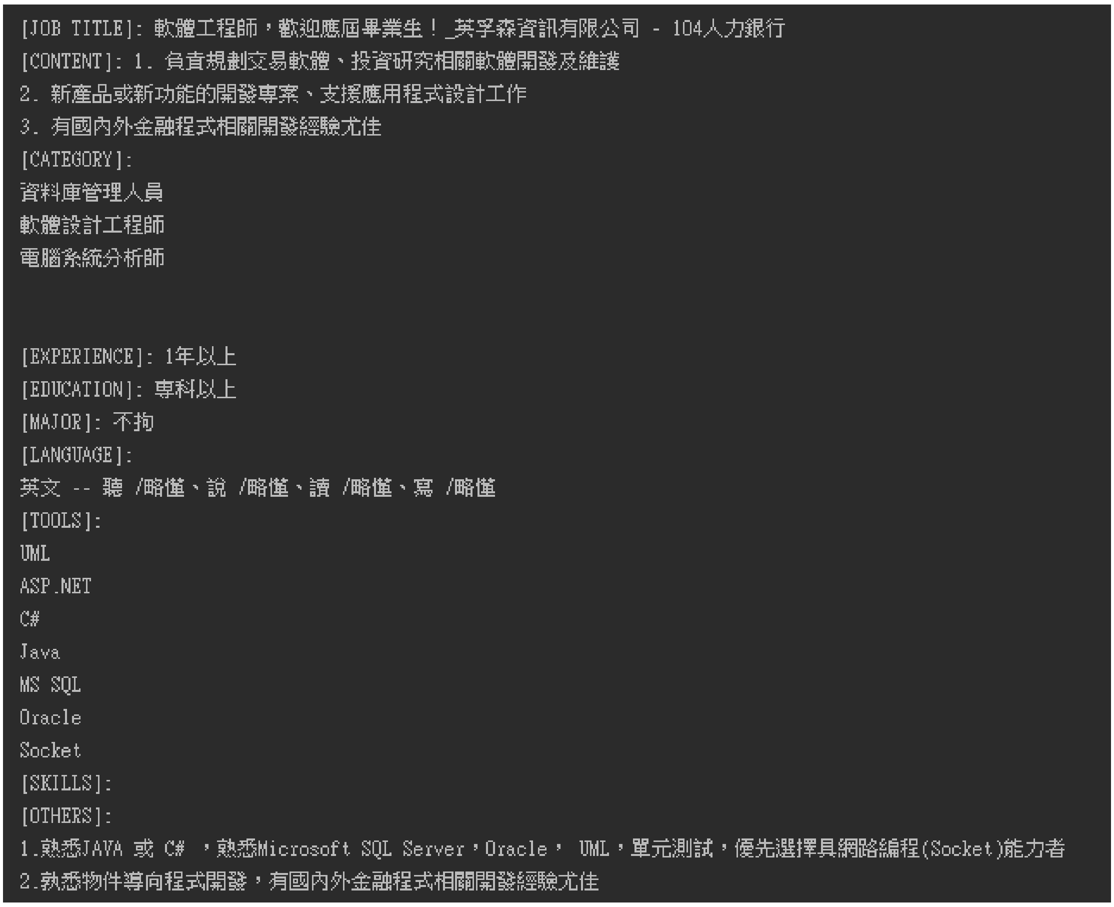
Figure 7　Detailed Job Description Page

An example of a detailed job requirement page is shown on Figure 7. Here we will noticed that the selections of tools are actually chosen from a fixed tick-box of all tools, therefore, we are able to exact the required tool list data from the html data. The extracted tools will be formatted in English. Therefore, the detailed metadata description of the data collected are summarized in Figure 8.

Metadata	**Description**
Job Title	The title of the job that the employer required.
Content	The detailed description that the employer wrote for this specific job opening.
Category	The category of this posted job in standard format.
Experience	The experience requirement of this specific job opening, written in number of years or no experience.
Education	The level of education that this specific job opening required.
Major	The type of major that this specific job opening required.
Language	The level of language skills that this specific job opening required.
Tools	The different kinds of tools required.
Skills	The skills required for this job opening.
Others	Any other extra descriptions.

Figure 8 Detailed Description of the metadata of Job Requirement

Figure 9 show an example of the data collected.



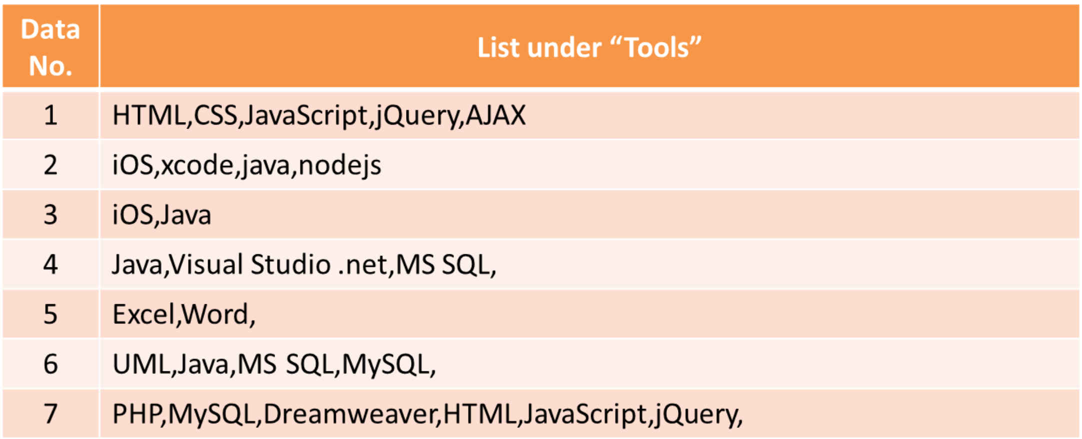
```
[JOB TITLE]: 軟體工程師，歡迎應屆畢業生！_英孚森資訊有限公司 - 104人力銀行
[CONTENT]: 1. 負責規劃交易軟體、投資研究相關軟體開發及維護
2. 新產品或新功能的開發專案、支援應用程式設計工作
3. 有國內外金融程式相關開發經驗尤佳
[CATEGORY]:
資料庫管理人員
軟體設計工程師
電腦系統分析師


[EXPERIENCE]: 1年以上
[EDUCATION]: 專科以上
[MAJOR]: 不拘
[LANGUAGE]:
英文 -- 聽 /略懂、說 /略懂、讀 /略懂、寫 /略懂
[TOOLS]:
UML
ASP.NET
C#
Java
MS SQL
Oracle
Socket
[SKILLS]:
[OTHERS]:
1.熟悉JAVA 或 C# ，熟悉Microsoft SQL Server，Oracle， UML，單元測試，優先選擇具網路編程(Socket)能力者
2.孰悉物件導向程式開發，有國內外金融程式相關開發經驗尤佳
```

Figure 9 An example of the data collected

From the collected data, we can see that the most important part of the data we need is in the Tools section. In the Tools section, all skills that are mostly found in the industry will be listed on this section for the employers to select their required skills.

3.2 Sorting of Required Skills

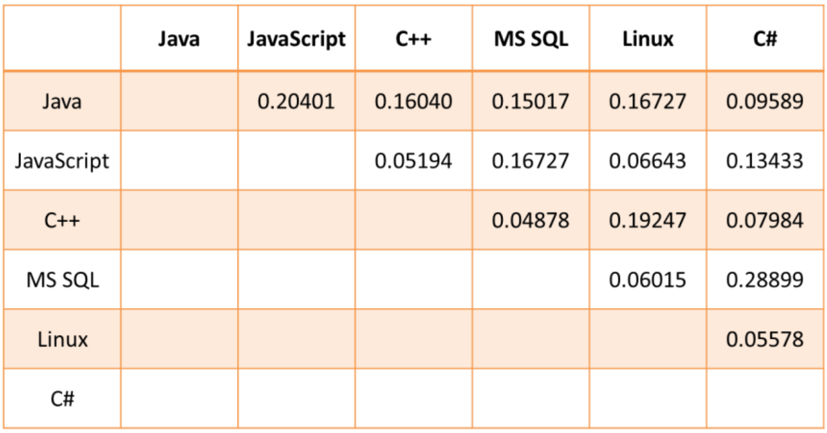
After the data has been collected, the grouping method for the required skills will be one of the major contributions. A certain job title can be posted on the job hunting web site, but this specific job title may be from different organizations, and for example, if we find three different organizations posting a recruit for software engineer, it is unlikely that the required skills for these three organizations will be the same, since different organizations may use different tools to develop.



Data No.	List under “Tools”
1	HTML,CSS,JavaScript,jQuery,AJAX
2	iOS,xcode,java,nodejs
3	iOS,Java
4	Java,Visual Studio .net,MS SQL,
5	Excel,Word,
6	UML,Java,MS SQL,MySQL,
7	PHP,MySQL,Dreamweaver,HTML,JavaScript,jQuery,

Figure 10 Example of the List of tools collected.

A possible method that can be used in this situation is called Jaccard Similarity. The Jaccard Similarity Coefficient is a parameter used to compare characteristic similarity between sets of information. Similarity measurement of Jaccard’s between two example sets is a quotient of sharing characteristic number divided by all characteristic number, where A may equals a set of required skills that does not occur in skill list one and B may equals a set of required skills that does not occur in skill list two.



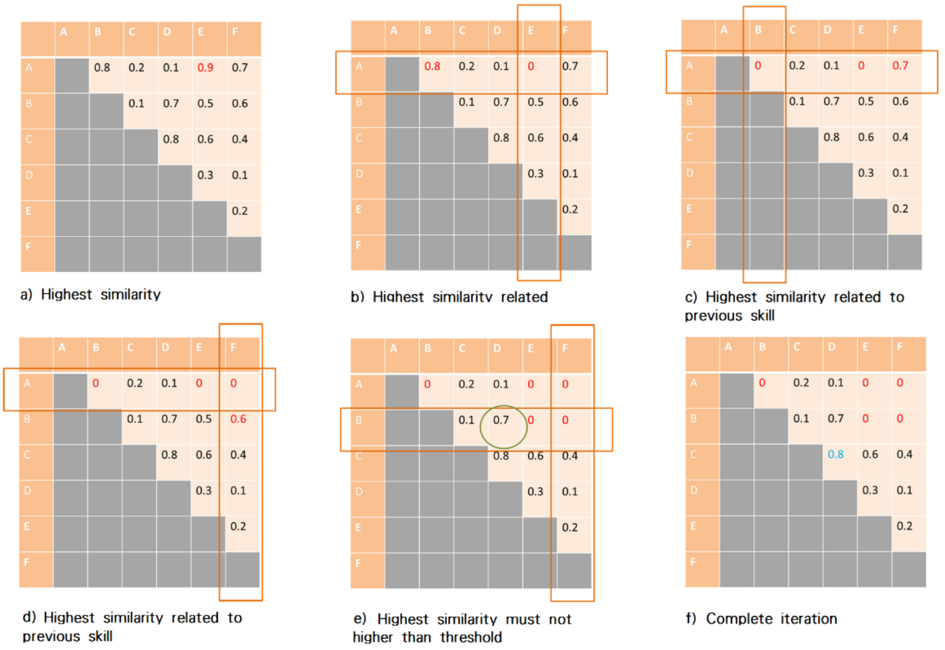
	Java	JavaScript	C++	MS SQL	Linux	C#
Java		0.20401	0.16040	0.15017	0.16727	0.09589
JavaScript			0.05194	0.16727	0.06643	0.13433
C++				0.04878	0.19247	0.07984
MS SQL					0.06015	0.28899
Linux						0.05578
C#						

Figure 11 shows an example table of the Jaccard's Similarity calculated from the data collected.

Even though we have the Jaccard's Similarity calculated, it is difficult to define the vectors for algorithms like K-means. In the similar researched, most of the vectors derived from TF-IDF, which is called term frequency–inverse document frequency, a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus. It is often used as a weighting factor in information retrieval, text mining, and user modeling. Most of the researches find the related data from the job description using TF-IDF. But since we already have a better and organized job hunting site in Taiwan, and the skills or tools are fixed, we will have the direct and correct skills and tools required, without trying to extract useful data from the job description text.

3.3 Clustering of Results

Once we calculate the complete Jaccard's Similarity of the full tool matrix, we can start with the clustering of data. Here, we cannot use commonly seen methods of clustering such as K-means and so forth, because we cannot clearly define a way to measure the vector in this tool matrix. Then, we have tried Hierarchical Clustering on the tool matrix, the results is fine, but with a major problem where the same tool will not appear in another cluster. This clustering method will then not be suitable for our situation, because one skill can be needed as much important together two set of different skills. So, this clustering method will kill the opportunities for one tool to be classified into another tool set. In other words, another method has to be developed in order to suit this set of data. The method are explained below.



a) Highest similarity

b) Highest similarity related

c) Highest similarity related to previous skill

d) Highest similarity related to previous skill

e) Highest similarity must not higher than threshold

f) Complete iteration

Figure 12 Diagram of the clustering algorithm

The steps explanation of the clustering algorithm used is showed on Figure 12. This example consists of the Jaccard's similarity of skills A to F. In order to find the first and the strongest cluster of skills, the pair of skills with the highest similarity is selected in diagram 5a, which is skills A and E with a value of 0.9. During the clustering, a threshold is set and modified to find the best results, in this case, the threshold is set to be 0.5. In each iterations, the value selected must not be lower than the threshold.

The second step would be to find the highest similarity related to both skill A and E, which is 0.8, the similarity between A and B, showed on diagram 5b. The same idea continues in the next steps, which is showed on diagram 5c and 5d. In the next step, showed in diagram 5e, please take note that, the value 0.7, circled in green, is not selected, otherwise the iteration will continue until all the values are selected, this is rule that when we selected the highest similarity value, we do not select the value lower than the threshold. At the same time, we always select the value that lower than the previous selected value.

Diagram 5f shows the end of one complete iteration, this results in the first cluster found. Therefore, the first cluster consists of skills A, B, E and F. The next cluster will start with the value 0.8, which is the similarity between skills C and D marked in blue. By using this algorithm we can solve the problem happened when using Hierarchical Clustering, which is when the same skill do not appear in other clusters, this result will not be suitable in the real situation. When the above example is clustered, the results would be cluster 1{A, B, E, F}; cluster 2 {C, B, D} and cluster 3{C, E, F}. In this case, we will have a result that the same skill is possible to appear in two or more clusters.

```
        Clustering Algorithm

while max Sim value > threshold
        for eachRow in range of All skills
                for eachCol in range of All skills
                        find max Sim value
                        record skills
        if max Sim value > threshold
                find next max Sim value
                add skill to list
end while
```

Figure 13 Pseudo code of the Clustering Algorithm

3.4 Creating the Course Map

Another major contribution is to form a course map from the results of the groups formed, and help the students by recommending the order of the courses to study these required skills. The “Computer Science Curricula 2013, Curriculum Guidelines for Undergraduate Degree Programs in Computer Science, December 20, 2013, The Joint Task Force on Computing Curricula Association for Computing Machinery (ACM) IEEE Computer Society” will studied to find out the inter relations between the courses, and a standard course map will be constructed according to this document. However, it is necessary for the course map to be updated if a new skill is required by the industry.

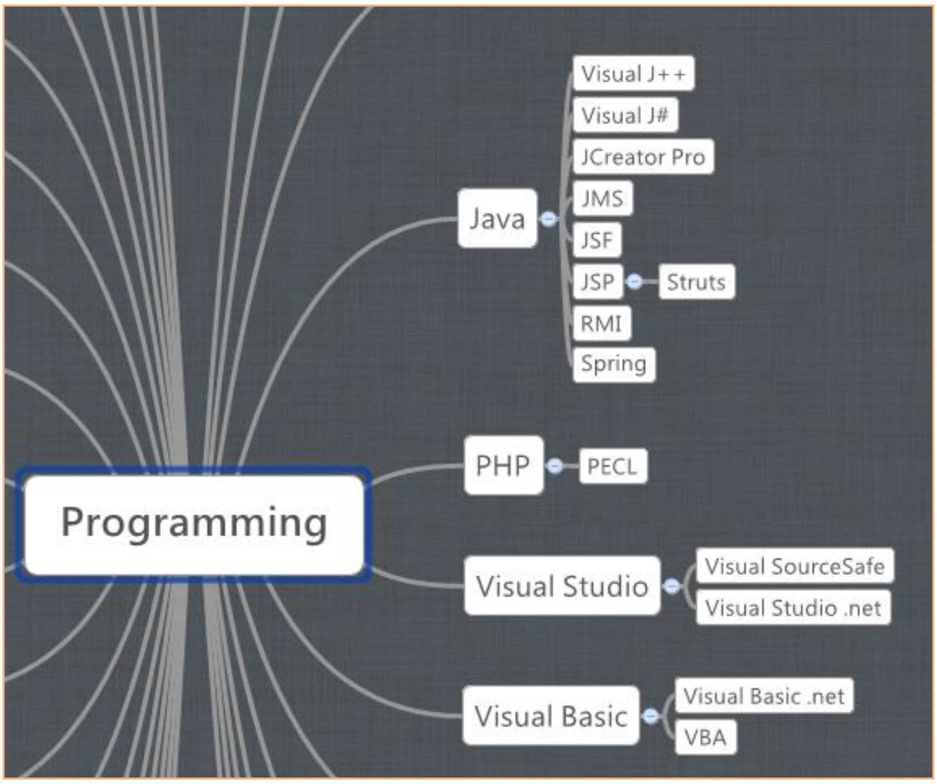


Figure 14 A portion of the Course Map created

A possible method to solve this problem is to implement a system that allows the educators to choose the possible parent nodes on the course map for the new skills. Therefore, whenever a new metadata of a required skill appears in the results of Sorting of Required Skills, a notice will be sent to the educators who are also part of this project or who had registered in this system, to encourage them to choose some possible parent courses for a specific skill based on their own educating experiences. A final result will then be analyzed from all feedbacks of the educators, and the new skill will be updated on the course map according to the final result.

Chapter 4. System Implementation

4.1 Web Crawler

The program is mainly written in Python, the first part of the program is to collect the data from 104 job hunting web site. BeautifulSoup is used to for crawling data from the html tags in the job description. This process is written automatically, which means that the job search section of the computer science will be crawled in one click. The data is first simply stored in json format for future use. Figure 15 shows the concept of web crawler.

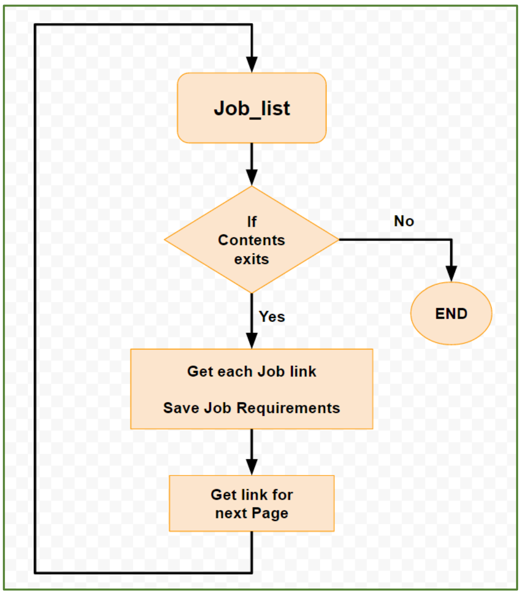
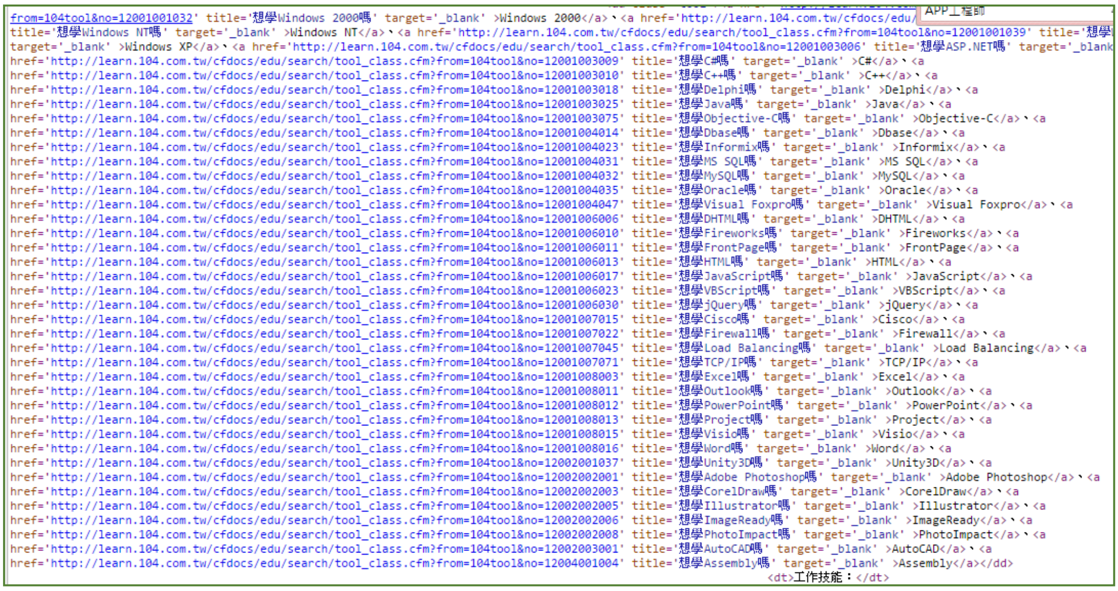


Figure 15 Concept of web crawler.

Data are extracted from the html code of each job description. Figure 16 shows the list of job skills required for a certain job.

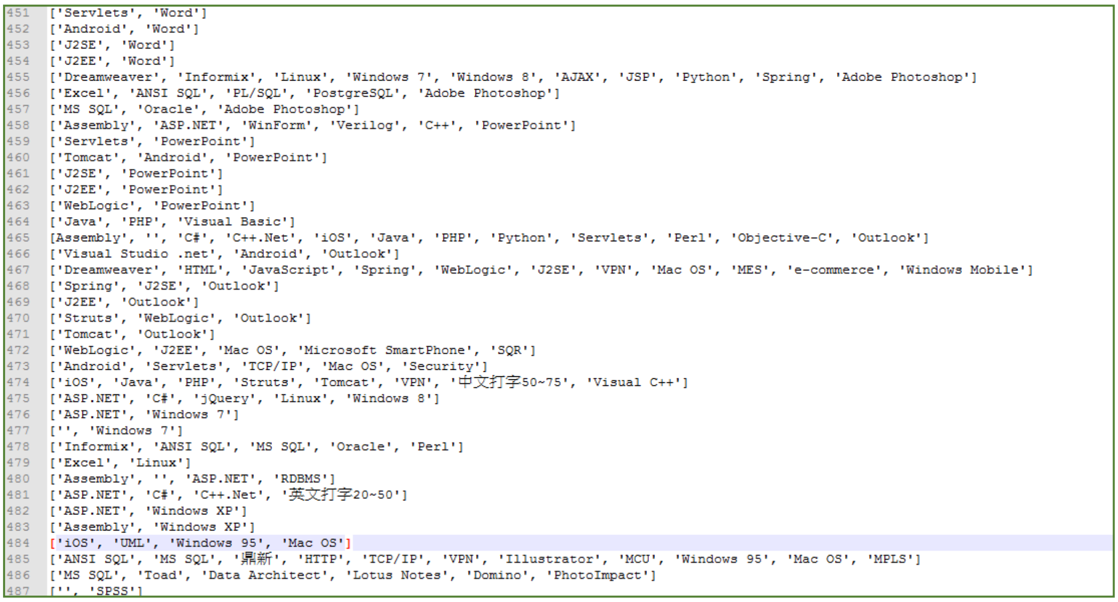


```
from=104tool&no=12001001032' title='想學Windows 2000嗎' target='_blank' >Windows 2000</a>、<a href='http://learn.104.com.tw/cfdocs/edu/
title='想學Windows NT嗎' target='_blank' >Windows NT</a>、<a href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001001039' title='想學
target='_blank' >Windows XP</a>、<a href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001003006' title='想學ASP.NET嗎' target='_blank
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001003009' title='想學C#嗎' target='_blank' >C#</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001003010' title='想學C++嗎' target='_blank' >C++</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001003018' title='想學Delphi嗎' target='_blank' >Delphi</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001003025' title='想學Java嗎' target='_blank' >Java</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001003075' title='想學Objective-C嗎' target='_blank' >Objective-C</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001004014' title='想學Dbase嗎' target='_blank' >Dbase</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001004023' title='想學Informix嗎' target='_blank' >Informix</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001004031' title='想學MS SQL嗎' target='_blank' >MS SQL</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001004032' title='想學MySQL嗎' target='_blank' >MySQL</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001004035' title='想學Oracle嗎' target='_blank' >Oracle</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001004047' title='想學Visual Foxpro嗎' target='_blank' >Visual Foxpro</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001006006' title='想學DHTML嗎' target='_blank' >DHTML</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001006010' title='想學Fireworks嗎' target='_blank' >Fireworks</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001006011' title='想學FrontPage嗎' target='_blank' >FrontPage</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001006013' title='想學HTML嗎' target='_blank' >HTML</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001006017' title='想學JavaScript嗎' target='_blank' >JavaScript</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001006023' title='想學VBScript嗎' target='_blank' >VBScript</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001006030' title='想學jQuery嗎' target='_blank' >jQuery</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001007015' title='想學Cisco嗎' target='_blank' >Cisco</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001007022' title='想學Firewall嗎' target='_blank' >Firewall</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001007045' title='想學Load Balancing嗎' target='_blank' >Load Balancing</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001007071' title='想學TCP/IP嗎' target='_blank' >TCP/IP</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001008003' title='想學Excel嗎' target='_blank' >Excel</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001008011' title='想學Outlook嗎' target='_blank' >Outlook</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001008012' title='想學PowerPoint嗎' target='_blank' >PowerPoint</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001008013' title='想學Project嗎' target='_blank' >Project</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001008015' title='想學Visio嗎' target='_blank' >Visio</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12001008016' title='想學Word嗎' target='_blank' >Word</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12002001037' title='想學Unity3D嗎' target='_blank' >Unity3D</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12002002001' title='想學Adobe Photoshop嗎' target='_blank' >Adobe Photoshop</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12002002003' title='想學CorelDraw嗎' target='_blank' >CorelDraw</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12002002005' title='想學Illustrator嗎' target='_blank' >Illustrator</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12002002006' title='想學ImageReady嗎' target='_blank' >ImageReady</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12002002008' title='想學PhotoImpact嗎' target='_blank' >PhotoImpact</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12002003001' title='想學AutoCAD嗎' target='_blank' >AutoCAD</a>、<a
href='http://learn.104.com.tw/cfdocs/edu/search/tool_class.cfm?from=104tool&no=12004001004' title='想學Assembly嗎' target='_blank' >Assembly</a></dd>
                                                                                                                   <dt>工作技能：</dt>
```

Figure 16 Tools list in html data

4.2 Data Processing

Once the data are collected, the tools lists are processed to find the clusters using Jaccard Similarity and the clustering algorithm. Then, the tools from the clusters are used to trace back the categories of the job requirements, the categories are decided by the most appeared rate, in order to classify and name the category of each cluster. Figure 17 shows a portion of the clustering results. The value of threshold is adjusted while observing from the results.



```
['Servlets', 'Word']
['Android', 'Word']
['J2SE', 'Word']
['J2EE', 'Word']
['Dreamweaver', 'Informix', 'Linux', 'Windows 7', 'Windows 8', 'AJAX', 'JSP', 'Python', 'Spring', 'Adobe Photoshop']
['Excel', 'ANSI SQL', 'PL/SQL', 'PostgreSQL', 'Adobe Photoshop']
['MS SQL', 'Oracle', 'Adobe Photoshop']
['Assembly', 'ASP.NET', 'WinForm', 'Verilog', 'C++', 'PowerPoint']
['Servlets', 'PowerPoint']
['Tomcat', 'Android', 'PowerPoint']
['J2SE', 'PowerPoint']
['J2EE', 'PowerPoint']
['WebLogic', 'PowerPoint']
['Java', 'PHP', 'Visual Basic']
[Assembly', '', 'C#', 'C++.Net', 'iOS', 'Java', 'PHP', 'Python', 'Servlets', 'Perl', 'Objective-C', 'Outlook']
['Visual Studio .net', 'Android', 'Outlook']
['Dreamweaver', 'HTML', 'JavaScript', 'Spring', 'WebLogic', 'J2SE', 'VPN', 'Mac OS', 'MES', 'e-commerce', 'Windows Mobile']
['Spring', 'J2SE', 'Outlook']
['J2EE', 'Outlook']
['Struts', 'WebLogic', 'Outlook']
['Tomcat', 'Outlook']
['WebLogic', 'J2EE', 'Mac OS', 'Microsoft SmartPhone', 'SQR']
['Android', 'Servlets', 'TCP/IP', 'Mac OS', 'Security']
['iOS', 'Java', 'PHP', 'Struts', 'Tomcat', 'VPN', '中文打字50~75', 'Visual C++']
['ASP.NET', 'C#', 'jQuery', 'Linux', 'Windows 8']
['ASP.NET', 'Windows 7']
['', 'Windows 7']
['Informix', 'ANSI SQL', 'MS SQL', 'Oracle', 'Perl']
['Excel', 'Linux']
['Assembly', '', 'ASP.NET', 'RDBMS']
['ASP.NET', 'C#', 'C++.Net', '英文打字20~50']
['ASP.NET', 'Windows XP']
['Assembly', 'Windows XP']
['iOS', 'UML', 'Windows 95', 'Mac OS']
['ANSI SQL', 'MS SQL', '鼎新', 'HTTP', 'TCP/IP', 'VPN', 'Illustrator', 'MCU', 'Windows 95', 'Mac OS', 'MPLS']
['MS SQL', 'Toad', 'Data Architect', 'Lotus Notes', 'Domino', 'PhotoImpact']
['', 'SPSS']
```

Figure 17 Sets of clusters of the Clustering Results

After getting the results of the skill clusters, the categories of the skill clusters are being summed up and calculated back to be the main category of the specific cluster of skills.

4.3 Web Implementation

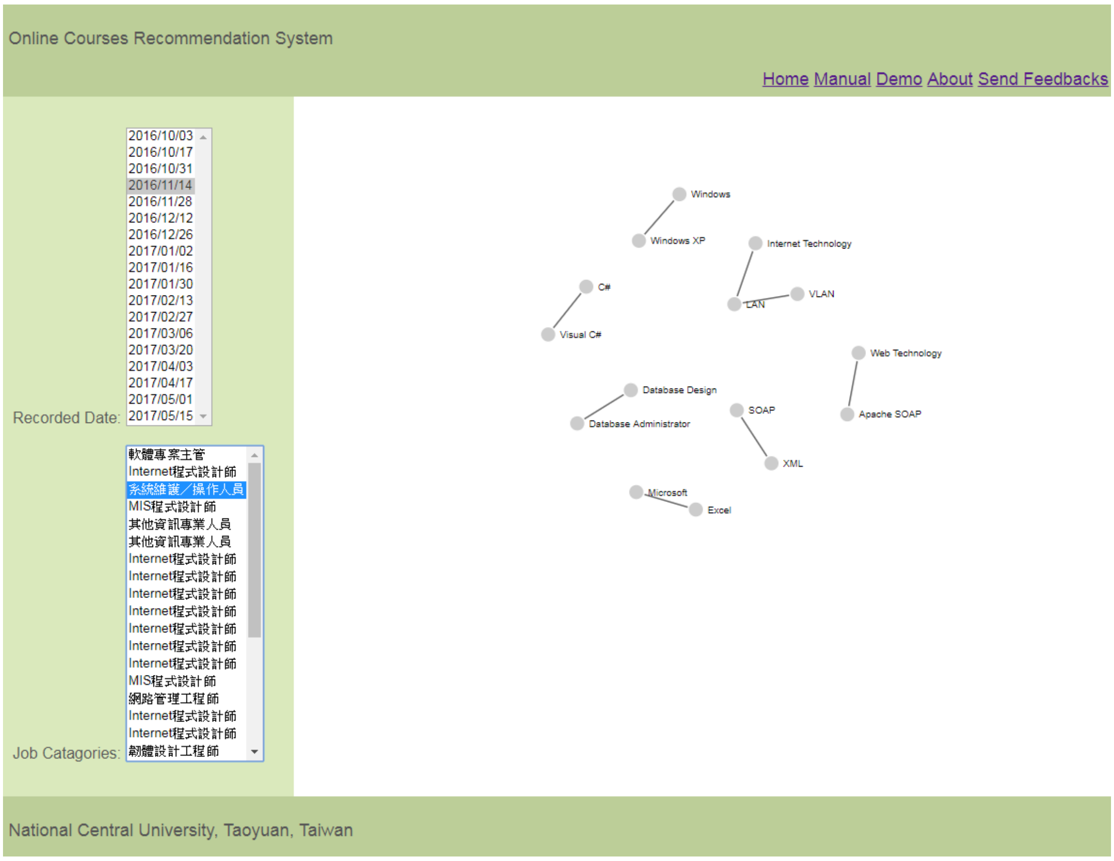
We decided to show this system in the form of a web site, therefore, it is clearer to show the results course maps and easier for the students to access. The web also contains a section for feedback questionnaire, in order to do further analysis and to approve the system. The web is simply implemented on GitHub. GitHub has personal web spaces called GitHub Pages. It can be created by creating a new repository named username.github.io, where username is your username on GitHub.

Chapter 5. Experiment Results and Discussions

First, the data are collected from www.104.com.tw, which is the popular job hunting site in Taiwan, which is collected every two weeks. Each set of the data from the Internet Technology section will have around 1500 of job recruitment. Each job recruitment will have a set of skills, these skills are then clustered using the clustering algorithm mentioned in section III C. The number of total data collected is 18 sets, lasted from October 2016 to May 2017. The data is collected every two weeks.

5.1 Web Presentation

The results will be displayed on a web site. The skills are presented in the form of a course map, this is to give them a guide of the order to learn the courses, although the skills are not always related from the clustering results. Figure 18 shows the web presentation of the clustering results. On the left hand side, a scroll list will allow the selections of different type of clusters. When the type is selected, the course guide map will show on the right hand side. The course guide map showed is derived from the whole map which is created in section III D. Figure 18 shows a small portion of the skills from the whole map.



Online Courses Recommendation System

Home Manual Demo About Send Feedbacks

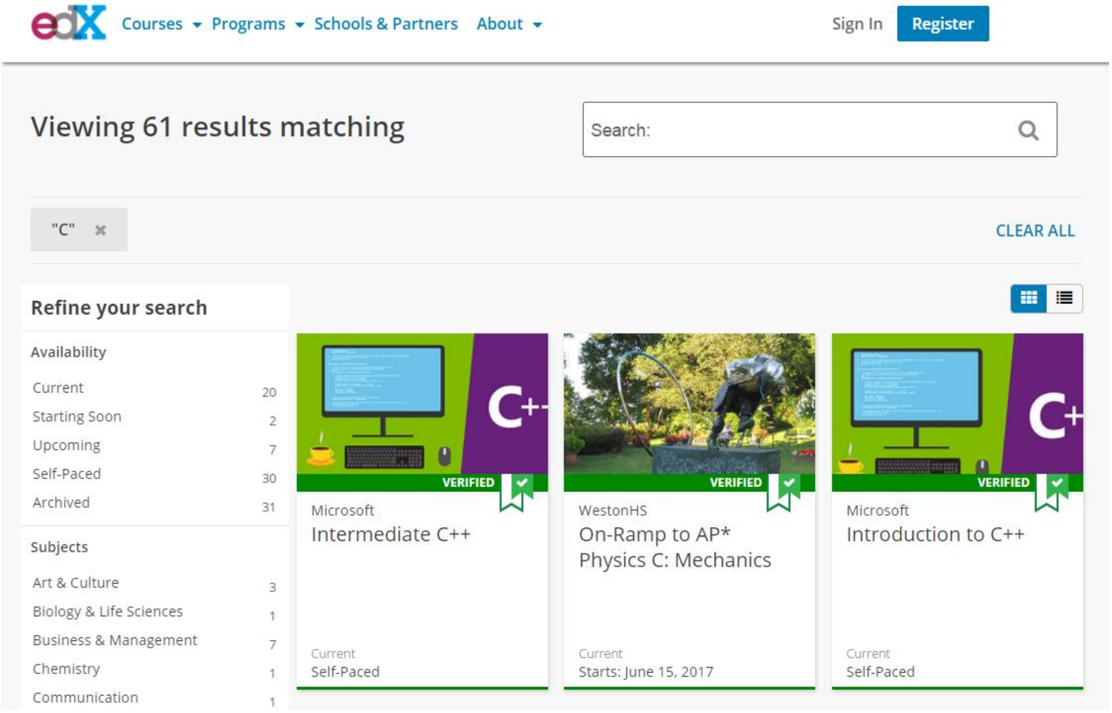
Recorded Date:
2016/10/03
2016/10/17
2016/10/31
2016/11/14
2016/11/28
2016/12/12
2016/12/26
2017/01/02
2017/01/16
2017/01/30
2017/02/13
2017/02/27
2017/03/06
2017/03/20
2017/04/03
2017/04/17
2017/05/01
2017/05/15

Job Catagories:
軟體專案主管
Internet程式設計師
系統維護／操作人員
MIS程式設計師
其他資訊專業人員
其他資訊專業人員
Internet程式設計師
Internet程式設計師
Internet程式設計師
Internet程式設計師
Internet程式設計師
Internet程式設計師
Internet程式設計師
MIS程式設計師
網路管理工程師
Internet程式設計師
Internet程式設計師
韌體設計工程師

National Central University, Taoyuan, Taiwan

Figure 18 Web Presentation of the Clustering Results

When each skill node is clicked, it will direct to another page, where the searched results of the skill on the skill node selected, from the MOOCs courses will be showed. In this system, the search results from the edX website is showed, as in Figure 19. So, it is more convenient for students to go straight to have a glance at the actual course materials. When a new skill appears, the modification system which will send notice to the users, are still in developing stage, and will soon be completed in the future.



Viewing 61 results matching

Refine your search

Availability	
Current	20
Starting Soon	2
Upcoming	7
Self-Paced	30
Archived	31

Subjects	
Art & Culture	3
Biology & Life Sciences	1
Business & Management	7
Chemistry	1
Communication	1

Microsoft
Intermediate C++
Current
Self-Paced

WestonHS
On-Ramp to AP* Physics C: Mechanics
Current
Starts: June 15, 2017

Microsoft
Introduction to C++
Current
Self-Paced

Figure 19 Search results from the edX site when clicking on node C++

Different clustering results of the job types will be showed on the left hand side of the web site. These job types are referred from the categories of the jobs collected. In the computer science related section, there are a total of 11 different categories of clusters, different clusters may belong to the same category.

The actual job detailed references will be showed for the students, on the right hand side of the web site, a search bar will be provided to look for job title more easily.

5.2 Feedbacks

In order to approve that this system will be helpful to the education, a questionnaire is designed for this system as feedbacks. The questionnaire consists of 10 questions divided into two sections. The questions asked and the form of the response are listed on Figure 20.

Section A	Please choose your age.
	Please choose your occupation.
	Please choose your highest degree.
Section B	Do you agree that there is a skill gap between University and Industry?
	Do you think that 104 Job Hunting Site is the most popular in Taiwan?
	Do you think that edX is a good site for MOOC learning?
	Do you think that gather Job Information from Job Hunting Sites will be the best way to understand to real need from the Industry?
	Do you think the results are likely the requirements you see from the industry?
	Do you think that the map presentation will help you with the order to study the skills?
	Do you think this System is helpful?
	Please feel free to give any other suggestions

Figure 20 Questions of the feedback questionnaire

The first section contains questions that intended to understand the respondent's background. The age is divided in tens, so it is clearer to separate the respondent's age difference. The occupations are only divided into three, which are students, professor and society member. The last question in section A is the highest degree of the respondents. The second section is focused on the usefulness of the system, each question is rated on a scale of 1 to 5.

5.3 Results

The results from Section A shows that most of the respondents are undergraduate and master students. Their age ranged from 22 years old to 25 years old mostly. The graph is shown on Figure 21. It is good to see that some society member has also sent their feedback. In this work, the most reliable people for the feedbacks will be the people from the society, since they have already been through the process of finding jobs after graduation, they will be most likely to have experienced the hardness while looking for a desire occupation. The education of the persons from the feedbacks are mostly undergraduate and masters.

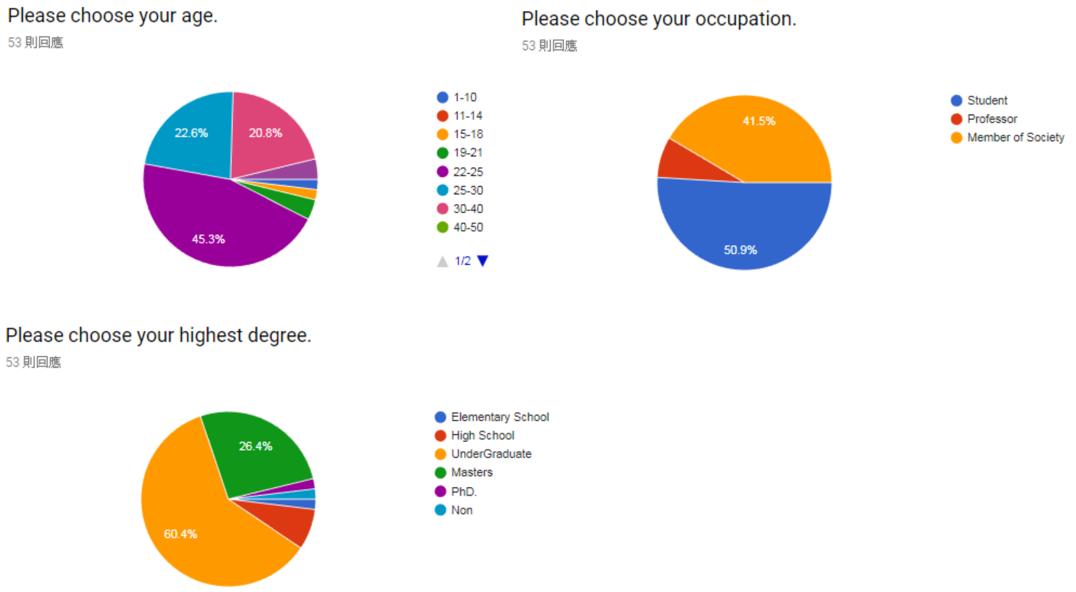


Figure 21 The Feedback results of Section A

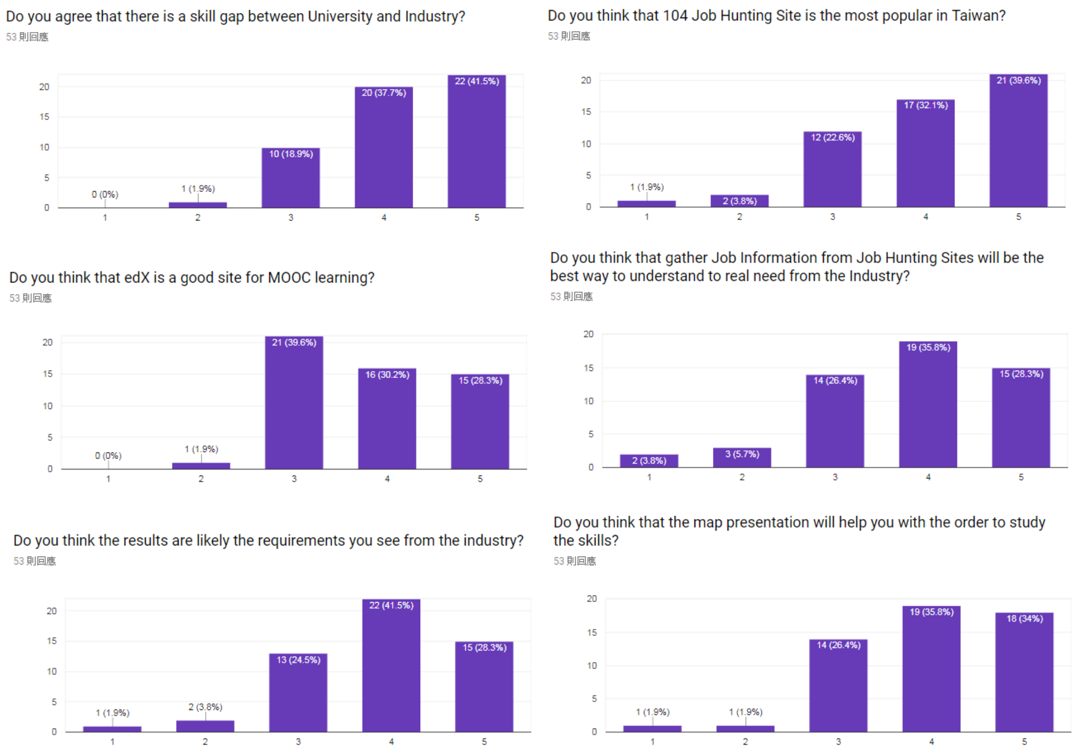
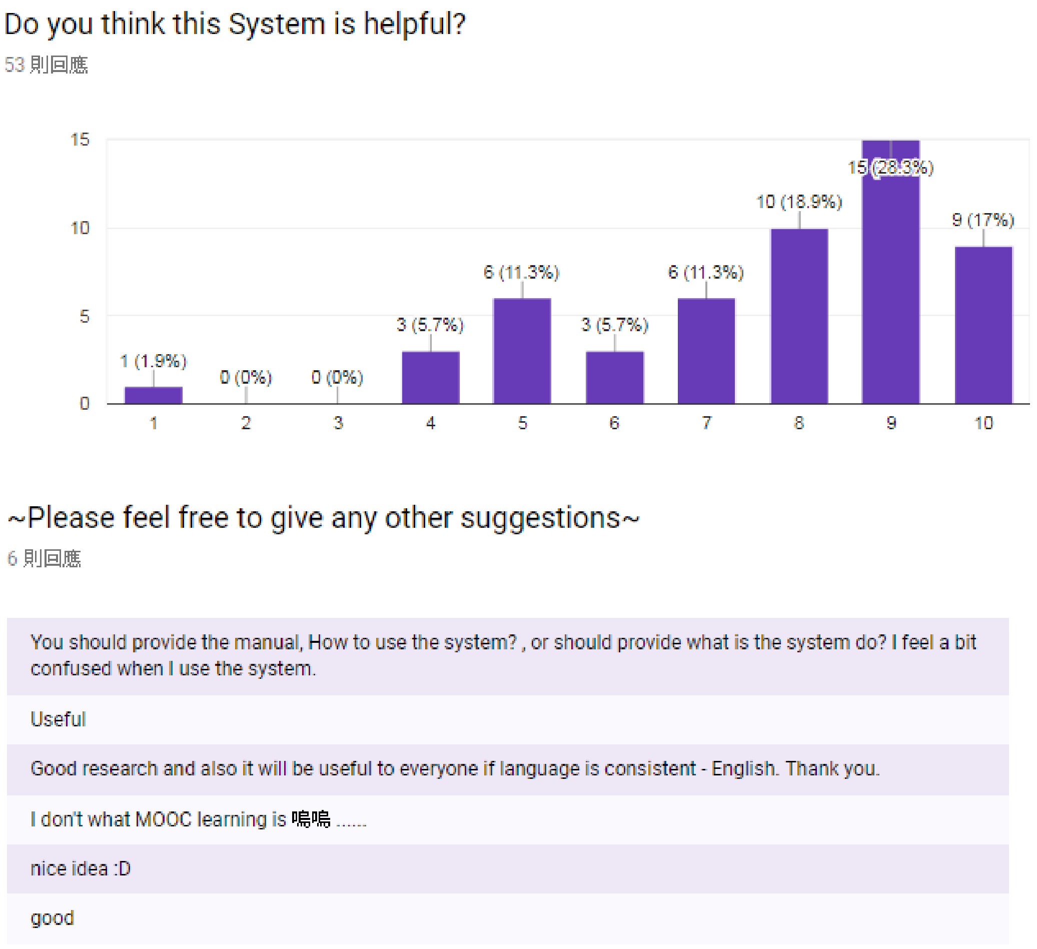


Figure 22 The Feedback results of Section B

The questions based on the usefulness of the system also showed a positive results, which is shown on Figure 22. The last question is the overall rating of the system. In order to be more specific, this question is scaled from 1 to 10. The overall rating of the system had also showed a good results.



Do you think this System is helpful?

53 則回應

~Please feel free to give any other suggestions~

6 則回應

You should provide the manual, How to use the system? , or should provide what is the system do? I feel a bit confused when I use the system.

Useful

Good research and also it will be useful to everyone if language is consistent - English. Thank you.

I don't what MOOC learning is 嗚嗚 ......

nice idea :D

good

Figure 23 Overall rating of the System

Chapter 6. Conclusion and Future Work

This system hopes to solve the problem known as the gap between industry and education, which the students, who had graduated from universities, do not always have the skills that the industries needed. In most cases, students will only have some skills or knowledge about some tools that is listed from the requirements of the industries. We encourage the students to self-study the courses according to the course map formed from their desired occupation, and therefore increase the chances that they will get the job offer.

As from the results, the feedback shows that choosing edX for the learning platform is not quite suitable. It is true that most courses related to the skill node are not always applicable. Therefore, in the future works, the lab members are creating knowledge books from Wikipedia. Using the Wikipedia contents to create their studying unit. This will hope to overcome the fact that not enough learning materials for some certain skills are on the available MOOC platform.

References

[1] A. A. Bakar and C. Y. Ting, “Soft skills recommendation systems for IT jobs: A Bayesian network approach” on Conference on Data Mining and Optimization (DMO), Selangor Malaysia, 28-29 June 2011

[2] G. Poornalatha and Prakash S. Raghavendra, “Web user Session Clustering Using Modified K-Means algorithm”, A. Abraham et al. (Eds.): ACC 2011, Part II, CCIS 191, pp. 243–252, Springer-Verlag Berlin Heidelberg 2011

[3] J. Singthongchai and S. Niwattanakul, “A Method for Measuring Keywords Similarity by Applying Jaccard’s, N-Gram and Vector Space”, Lecture Notes on Information Theory Vol. 1, No. 4, December 2013

[4] G. Karypis, “Evaluation of Item-Based Top-N Recommendation Algorithm”, NSF CCR-9972519, EIA-9986042, ACI-9982274, by Army Research OÆce contract DA/DAAG55-98-1-0441, by the DOE ASCI program, and by Army High Performance Computing Research Center contract number DAAH04-95-C-0008, 2011

[5] Y. Wang, X. Zhang, L. Nan, and D. Wang “Occupation recommendation based on student achievement mining in vocational skill training”, 11th International Conference on Fuzzy Systems and Knowledge Discovery, 2014

[6] G. Linden, B. Smith, and J. York, “Amazon.com recommendations- item-to-item collaborative filtering”, IEEE Internet Computing, January • February, 2003

[7] Juan J. Cameron, Carson Kai-Sang Leung and Syed K. Tanbeer, “Finding Strong Groups of Friends among Friends in Social Networks”, on ninth IEEE International Conference on Dependable, Autonomic and Secure Computing, 2011.

[8] C. Hauff and G. Gousios,”Matching GitHub developer profiles to job” on 12th Working Conference on Mining Software Repositories, 2015

[9] Y. Zhang, C. Yang and Z. Niu, “A Research of Job Recommendation System Based on Collaborative Filtering” on Seventh International Symposium on Computational Intelligence and Design, 2014

[10] J. Wang, Z. Liu and H. Zhao, “Group Recommendation Using Topic Identification in Social Networks” on Sixth International Conference on Intelligent Human-Machine Systems and Cybernetics, 2014

[11] M. S. Chen, J. Han and Philip S. Yu, “Data Mining: An Overview from a Database Perspective” on IEEE Transactions on Knowledge and Data Engineering, Vol 8, No 6, December, 1996

[12] S. Ye, J. Lang and F. Wu, “Crawling Online Social Graphs” on 12th International Asia-Pacific Web Conference, 2010

[13] H. Sun, T. Wu, M. Yan and Y. Wu, “A new item clustering-based collarative filtering approach” on Ninth Web Information Systems and Applications Conference, 2012

[14] Y. Li, J. Zhang and H. Dan, “Text Clustering Based on Domain Ontology and Latent Semantic Analysis”, on International Conference on Asian Language Processing, 2010

[15] Nguyen Viet Anh, Nguyen Viet Ha And Ho Si Dam, “Constructing A Bayesian Belief Network to Generate Learning Path In Adaptive Hypermedia System” , 2008

[16] J. Davies and M. Graff, “Performance in e-learning: online participation and student grades”, British Journal of Educational Technology Vol 36 No 4 2005 657–663, 2005