

Improvement of Information Technology Career Path by using Apriori and Ontology on Mobile Application

Sumitra Nuanmeesri, Lap Poomhiran

Abstract: *In Thailand, the joblessness rate is continually expanding for new graduates. But information technology careers are nevertheless popular in business in this digital age. Finding a way to support students to know the career direction that is suitable for them and is ready before going out for a job is therefore very valuable. However, most students do not know which occupation is best for them and are concern about the need to educate themselves and build up information that is suitable for their current career. This paper presents the Thai mobile application for predicting a student's career based on subject results and guidance of training courses for information technology career path using the Apriori algorithm and ontology technique with the longest matching approach in the process of Thai word segmentation. The developed mobile application was tested in black box by experts and evaluated the satisfaction by users. The result shows that the developed mobile application was the highest effective with the high consensus for the evaluation of user satisfaction.*

Keywords: *Apriori, association rules, future career, mobile application, ontology, recommendation.*

I. INTRODUCTION

Nowadays, several graduates in higher education do not have a job or work. In some instances, the works or jobs do not fit with the subjects they graduated. According to statistics [1], the unemployment rate of new graduates in Thailand has increased by almost 370,000 people. But information technology careers are nevertheless popular in business due to lack of staff and relatively high recruitment rates in this digital age. The reason might be that the qualifications of graduates do not meet with job requirements or that the subjects in the curriculum of the university do not keep up with the fast-changing technology at present.

The research has the concept of predicting a student's career based on subject results and recommends training courses available on the internet to students by analyzing the relationship of the career qualifications and the courses of the training. The result will help students to have the potential and readiness to apply for jobs and obtain jobs that match their ability, including adding knowledge to keep up with technological developments to create more

opportunities for students to get jobs in the future. It is a reflection of the university that has developed higher graduate employment conditions and has made progress in the information technology labor market for the country.

This study therefore develops and tests a model 1) to predict the future career for the student by using grade points of subjects to build the association rules, and 2) to suggest the training courses which are suitable for career or job qualifications base on ontology technique.

II. RELATED WORKS

The exploration of the literature reviews which are related to this research shows that Apriori is a popular algorithm used to find information based on association rules [2]. It examines macro-level connections and then takes transactions into account by defining objects and analyzing items at each level. By Apriori algorithm, the set of items or attributes that can be identified by the most frequent itemset of occurring together in a dataset. In order to reduce the search space, this algorithm uses two phases: 'join', and 'prune'. The algorithm for the joining phase is to create the itemset k+1 out of each itemset by attaching each itemset to itself. Next phase, it will check the number of an individual itemset in the database. If the nominee itemset is not found at least of minimum support value, the itemset is rare and thus excluded. In accordance with the grading mark and study length, the aim of Cameranesi and Diamantini [3] is mining the data attributes for three kind of careers as following: high efficiency careers, moderate careers, and low efficiency careers. Hodigère and Bilimoria [4] aim to forecast career by using multimodal private information of student in mining technique.

An ontology technique can describe the concepts and relationships between a certain reality and the intended meaning of vocabulary words as knowledge pool [5]. The ontology template can be generated by some software or program such as Protégé [6] platform which stores the knowledge acquired from data source on the internet and integrated the dialog manager in chatbot for mobile application [7]. in the step of the ontology process. There are no markers for word boundaries or delimiters between multiple words and between sentences because the limitation of Thai is different from English. The longest matching technique can be applied to separate each word in Thai [7]. In addition, for the purpose of data collection, the classification theory is used as the basis for the grouping and division of information into primary

Revised Manuscript Received on January 5, 2020

* Correspondence Author

Sumitra Nuanmeesri *, Assistant Professor in Information Technology, Faculty of Science and Technology, Suan Sunandha Rajabhat University, Thailand. Email: Sumitra.nu@ssru.ac.th

Lap Poomhiran, Ph.D. student in Information Technology, Faculty of Information Technology, King Mongkut's University of Technology North Bangkok, Thailand. Email: lap_p@windowlive.com

Improvement of Information Technology Career Path by using Apriori and Ontology on Mobile Application

classes; the data category was also divided into subclasses for content analysis, subject knowledge scope and theory of cognitive systems. According to Reinberger et al. [8] suggests that a parsed medical corpus pattern matching and clustering algorithms with dependence classes, creating semantical words and semantic connections.

The great information overload in the context of mobile applications has led to the development of mobile recommender systems. Many of them focus on point-of interest, travel, tourism, games and media. Most popular mobile recommender systems today are tourist guides, route recommendations and information recommendation. Mobile application can be utilized for the constructed ontology for recommendation system. To develop the mobile application for academic supporting the current students in junior [9]. This research purpose to apply the association rules based on Apriori algorithm and ontology technique to forecast the career path [10] with training courses guidance.

III. RESEARCH METHODOLOGY

The research methods adopted in the development process for the improvement information technology career path using Apriori and ontology on mobile application for students, include the following steps:

A. Requirement analyzing

Before built the mobile application, to gather and analyze the requirements of the user, we used the rapid application development (RAD) [11] methodology. RAD can reduce the cost of time to develop and resources. It consists of four stages: requirement planning, system design, development, and cutover stages. RAD model is represented as Figure 1.



Fig. 1. The model of RAD process [7]

In this process, we have planned and conducted surveys to gather user requirements and capabilities of the mobile application. The outcomes of the specification and planning process are used by evaluating the necessary components of the system for the design and implementation [7][12].

Design the system with consideration of the on-screen display style, the simplicity of use, and make it easy to understand. In this research, efficiency is increased and outcomes are improved by separating the main three components: the user interface, the word segmentation, rules base, the ontology and database. The user interface consists of three main functions: predicting the future career, guiding the training courses that related their career, and displaying the webpage of training courses.

B. Design and building the association rules using Apriori algorithm

In this work, the association rules were generated from data which has subjects and student's grade points data. This

information was stored in the academic database system between 2011 and 2019. We scope the only course for a bachelor degree in Information Technology Department, Faculty of Science and Technology at Suan Sunandha Rajabhat University. There are 25 subjects and 1,161 student records who have a job in total. The data was category into five jobs: multimedia and graphics, support and services, network securities, developer, and other job. And it was separated into two parts. First, student records between 2011 to 2016 are the training set for generating the association rules. Last, the remaining data is for measuring the system's efficiency and accuracy.

The data in the first part was analyzed by using the Apriori algorithm in Python language programming version 3.6.5 to generate the association rules. The parameter of minimum support and minimum metric values were 0.1 and 0.9 respectively. The result of the operation shows that there are 30 rules were discovered. In considering the rules with confidence was 100%, there are 14 rules for three careers which are related to subjects and student's grade point 'A' as shown in Table I, II, and III.

The meaning for each rule is the student must have achieved all the grades in a subject that appears in grade point 'A' and is predicted to be a developer job after graduation. For example, according to the association rule number 5th in Table I, if student has got grade 'A' of the Object Orientation System Analysis and Design subject, and got grade 'A' of the Fundamental of Mathematics subject, this student career is a developer.

Table- I: The association rules for 'Developer' job

Subject code and name	Association rules no.							
	1	2	3	4	5	6	7	8
INT1104 (Computer Programming Development)	A						A	A
INT1503 (Object Orientation System Analysis and Design)		A			A	A	A	
INT2104 (Object Orientation Programming Development)			A			A		A
INT2403 (Web Application Development)	A	A	A	A		A	A	A
MAT1201 (Fundamental of Mathematics)				A	A			

Table- II: The association rules for 'Multimedia and graphics' job

Subject code and name	Association rules no.	
	9	10
INT2402 (Web Design)		A
INT3107 (Applied for Computer Graphic)	A	
INT3110 (Multimedia Technology)		A

Table- III: The association rules for 'Network securities' job

Subject code and name	Association rules no.			
	11	12	13	14
INT2303 (Management and Design Network System)	A			A
INT2304 (Information Security)		A		
INT2305 (Data Communication and Networks)			A	A



All student data from the three-academic years between 2017 and 2019 have been truly studying this process. There are 342 student records in the second part of the data was applied to measuring the accuracy of association rules. These data were not used to stage of formulation of the rules. To measuring the performance of 14 association rules, the accuracy was calculated in (1).

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

Where

TP: In a certain subject, high grades are usually true career.

FP: In a certain subject, high grades are usually false career.

TN: In a certain subject, low grades are usually true career.

FN: In a certain subject, low grades are usually false career.

C. Design and building the ontology structure

The knowledge of jobs received from the design process and the creation of linking rules using the Apriori algorithm can collect keywords and show the extent of information. In this research used data and information related with job requirements of company on internet and organizations or companies that offer IT training courses to students were used to gather 1,500 documents. The information about job qualifications and training courses related to careers were gathering from the website in Thailand by using web scraping technique. These documents were approved by five experts in information technology and Thai education, and fifty users who are students with Information Technology Department. Moreover, to estimate the definition and relationship of the validity of ontology applied in this framework, this ontology was evaluated by experts and users.

The feature extraction process is the necessity to extract relevant information from the text and metadata on website. There are four steps as the following:

1. The collection of data and pre-processing: the documents of the Thai job qualifications and training courses were gathered by using a service we called 'web scraping engine'. It was developed in the Python version 3.6.5 for gathering, scraping, cleansing, transforming, and improving data to be appropriate for analysis.

2. Word segmentation processing: the information from web scraping engine were separated to each word by using the longest matching technique.

3. Keyword extracting and grouping: all keywords of job qualifications and training courses were extracted from the documents and grouped in kind of word such as noun, pronoun, verb, preposition, conjunction, adjective, and number.

4. Named entity recognition processing: entity extraction classified named entities that are present in a keyword into categories which has meaningful units for matching jobs and training courses then store into application database which is designed in this work.

Figure 2 shows the system architecture diagram of main processes.



Fig. 2. The system architecture diagrams.

D. Mobile application development

The results of the ontology process are stored in the ontology database which is designed in this work. We develop the mobile application in cross-platform for both iOS and Android operating system by using ionic framework and integrated to the web server which is develop in PHP. This application was built for students in Information Technology Department at Suan Sunandha Rajabhat University, when the users input their student code, the system will predict the future career and suggest the training courses which have link to webpages that scrapped from internet and a link direct to original website.

Figure 3 and Figure 4 display user interface which is related to describe the features of jobs and IT training course by student on mobile application in Thai.

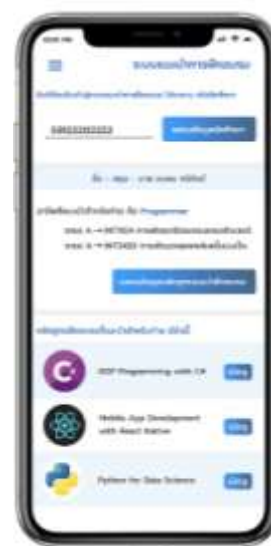


Fig. 3. Example of training course recommendations.

Improvement of Information Technology Career Path by using Apriori and Ontology on Mobile Application



Fig. 4. Example of training course details.

E. Evaluation the ontology and mobile application

In the ontology process for recognizing and classifying the keywords, the Term Frequency-Inverse Document Frequency (TF-IDF) is used to assess the relative frequency of words by a different proportion of the word over the whole corpus of the text in a specific document. According to information from a set of documents, the extraction of information is a major process involving the removal from sources of certain entities, events, and relations. In a series of documents TF-IDF was used to determine the weight of a word by measuring the term frequency in documents submitted in formulas as (2) and (3) [13]:

$$W_{i,j} = TF_{i,j} \times IDF_i \quad (2)$$

$$IDF_i = \log \frac{N}{DF_i} \quad (3)$$

Where

$TF_{i,j}$ is total amount of frequency of i in j .

DF_i is total amount of documents in i .

N is total amount of documents.

In this work, after building the ontology, the efficiency of text classification was evaluated by values of accuracy, recall, precision, and F-measure. The Accuracy, recall, precision, and F-measure can be representing as (1), (4), (5), and (6) [14][15] respectively.

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$F\text{-measure} = \frac{2 \times Precision \times Recall}{Precision+Recall} \quad (6)$$

Where

TP: job qualifications are match the future career, and suggested the related training courses.

TN: job qualifications are match the future career, and suggested the unrelated training courses.

FP: job qualifications are not match the future career, and suggested the related training courses.

FN: job qualifications are not match the future career, and suggested the unrelated training courses.

The results found that the relationships between the future career and the aforementioned training courses point to the future career of graduates had been influenced by training courses. It gave students the chance to get work directly within the degree of graduated to meet the business and job requirements. Consequently, we hope that this work will be extremely useful in the suggestion of training courses in order to help students' qualifications better.

The effectiveness of the mobile application was evaluated by black-box testing which is a technique that disinterested the internal process of parts or components in software [16] such as the programming structure, software internal details. It focusses only inputs and output of application or software. The scope of black box testing was separated into five assessment indicators: user requirements test, functions test, usability test, system performance test, and system securities test. There are five experts in information technology and business to assess the system and questionnaires.

Finally, the efficiency of the user satisfaction for the mobile application was evaluated by a sample of 50 people who are students in the 3rd and 4th year of the Information Technology Department. The scope of the questionnaires consists of the accuracy for the recommendation, integrity of functionality on the mobile application, easy to use, useful to users, and system performance. The survey results were analyzed following the Likert scale [17] to level parameters, as shown in Table IV, to assess the mean and standard deviation value [9][18].

Table- IV: The Likert scale for evaluation

Scale	Range of Weighted Mean	Level of Effective
5	4.51 – 5.00	The highest
4	3.51 – 4.50	The high
3	2.51 – 3.50	The medium
2	1.51 – 2.50	The little
1	1.00 – 1.50	The least

IV. EXPERIMENTAL RESULTS

There are the results of effectiveness for the mobile application shown as the following.

A. Effectiveness of the mobile application evaluated by experts

According to five indicators of black box testing, the result showed that the “User requirements test” and “Usability test” indicator had the highest mean value at 4.60 together with the standard deviation value at 0.55. The overall mean for experts was 4.44 and the overall standard deviation was 0.51, showing that the developed mobile application was the highest effective. The result of evaluation for acceptance of the mobile application that could be detected by five experts in while being used has



high consensus. All of values the interquartile range no more than 1 and the quartile deviation no more than 0.5 in five assessment indicators shows as Table V.

Table- V: The black box testing output of the effectiveness of the mobile application and acceptance by experts

Assessment Indicators	Mean	SD	Quartiles			Interquartile Rang	Quartile Deviation
			Q1	Median	Q3		
Experts							
1. User requirements test	4.60	0.55	4	5	5	1	0.5
2. Functions test	4.40	0.55	4	4	5	1	0.5
3. Usability test	4.60	0.55	4	5	5	1	0.5
4. System performance test	4.40	0.55	4	4	5	1	0.5
5. System securities test	4.20	0.45	4	4	4	0	0
Total	4.44	0.51	4	4	5	1	0.5

B. Effectiveness of the mobile application with user satisfaction

The research evaluated the satisfaction of mobile application by 50 users who are students in the 3rd and 4th year of the Information Technology Department. The effectiveness evaluation results of the mobile application, focusing on the satisfaction testing, showed that, among the five indicators, “Easy to use” criteria had the highest mean value at 4.76 together with the standard deviation value at 0.43. The overall mean and standard deviation for experts were 4.72 and 0.45 respectively. The scores show that the developed mobile application was the highest effective and the result of evaluation for acceptance of the mobile application had high consensus. All of values the interquartile range no more than 1 and the quartile deviation no more than 0.5 in five assessment indicators shows as Table VI.

Table- VI: The black box testing result of the user satisfaction and acceptance while being used

Assessment Indicators	Mean	SD	Quartiles			Interquartile Rang	Quartile Deviation
			Q1	Median	Q3		
Experts							
1. The accuracy for the recommendation	4.70	0.46	4	5	5	1	0.5
2. Integrity of functionality	4.72	0.45	4	5	5	1	0.5
3. Easy to use	4.76	0.43	5	5	5	0	0
4. Useful to users	4.72	0.45	4	5	5	1	0.5
5. System performance	4.70	0.46	4	5	5	1	0.5
Total	4.72	0.45	4	5	5	1	0.5

V. CONCLUSION

This research presents the development of the Thai mobile application for predicting future careers and guidelines the training courses in information technology using the Apriori algorithm and ontology technique for 3rd and 4th year students to prepare their abilities for their future careers. Apriori is used to find association rules that can help find the relationship between grades in subject and students' future careers. In field of information technology, future careers can be cluster as five careers. There are only three careers that

were generated to the 14 association rules which have 100 percent confidence value. These rules have established guidelines for students to find a job that is sufficient to improve their opportunity to be recruited.

The ontology technique is created to handle terms and relationships to describe the features of students' future careers and information technology training courses on the internet. The longest matching technique was applied in the process of Thai word segmentation. The black box testing technique and questionnaires were needed to evaluate system. The results of the evaluation by experts and users were found this mobile application with the highest effective. It can be said that the mobile application could be efficiently recommend the courses of training in the field of information technology to students.

There is some suggestion from users. Students request the university for participation in training courses to improve their skills or register for free topics that suit their qualifications in the university. This will create opportunities for students to get a job directly in graduate programs, which is an indicator of the quality of the higher education that allows and supports students successfully.

ACKNOWLEDGMENT

We are grateful for the Institute for Research and Development, Suan Sunandha Rajabhat University, who supported and gave an opportunity in this research.

REFERENCES

1. ThaiQuote (2019, October 15). A half-million graduate’s joblessness in 2019, a big problem for the government that cannot ignore. [Online]. Available: <https://www.thaiquote.org/content/227505>
2. D. N. Goswami, A. Chaturvedi, and C. S. Raghuvanshi, “An Algorithm for Frequent Pattern Mining Based On Apriori,” *International Journal on Computer Science and Engineering*, vol. 2, no. 4, 2010, pp. 942–947.
3. M. Cameranesi, and C. Diamantini, “Students' Careers Analysis: a Process Mining Approach,” In *Proceedings of WIMS'17*, (Amantea, Italy, June 2017), 2017, pp. 19–22.
4. R. Hodigere, and D. Bilimoria, “Constructing professional resource networks from career biographical database,” *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2012, pp. 1242–1247.
5. J. Allen, *Natural Language Understanding*, 2nd, Ed., Delhi: Pearson Education, 2004.
6. A. Vegesna, P. Jain, and D. Porwal, “Ontology based Mobile application (for e-commerce website),” *International Journal of Computer Applications*, vol. 179, no.14, 2018, pp. 51–55.
7. S. Nuanmeesri, and L. Poomhiran, “Ontology based Thai chatbot on social media marketing for community enterprise,” *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, no. 4, 2019, pp. 10153–10158.
8. M. L. Reinberger, P. Spyns, A. J. Pretorius, and W. Daelemans, “Automatic Initiation of an Ontology,” *Springer-Verlag In Meersman*, 2014, pp. 600–617.
9. S. Nuanmeesri, “Development alumni relationship management using Thai speech recognition on mobile application,” *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, no. 4, 2019, pp. 9915–9923.
10. M. Sodanil, S. Chotirat, L. Poomhiran, and K. Viriyapant, “Guideline for academic support of student career path using mining algorithm,” In *Proceedings of the 2019 3rd International Conference on Natural Language Processing and Information Retrieval*, (Tokushima, Japan), 2019, pp. 133–137.
11. Wavemaker (2019). Rapid Application Development Model. [Online]. Available:



Improvement of Information Technology Career Path by using Apriori and Ontology on Mobile Application

<https://www.wavemaker.com/rapid-application-development-model>

12. K. Kularbphetong, "Developing the Thai Regional Dialect Based on Semi-automatic Technique," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, no. 2, 2019, pp. 2842–2846.
13. H. C. Wu, R. W. P. Luk, K. F. Wong, and K. L. Kwok. "Interpreting TF-IDF term weights as making relevance decisions," *ACM Transactions on Information Systems*, vol. 26, no. 3, 2008.
14. M. Sokolova, and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing and Management*, vol. 45, 2009, pp. 427–437.
15. S. Nuanmeesri, "Mobile application for the purpose of marketing, product distribution and location-based logistics for elderly farmers," *Applied Computing and Informatics*, in press, vol. 9, 2019.
16. L. Williams. *Testing Overview and Black-Box Testing Techniques*, 2006, pp. 37.
17. R. Likert, *A technique for the measurement of attitudes*. Archives of Psychology, 1932, pp. 1–55.
18. S. Nuanmeesri, "The augmented reality for teaching Thai students about the human heart," *International Journal of Emerging Technologies in Learning*, vol. 13, no. 6, 2018, pp. 99–112.

AUTHORS PROFILE



Sumitra Nuanmeesri was born in Ratchaburi, Thailand. She has obtained a Ph.D. in Information Technology at King Mongkut's University of Technology North Bangkok (KMUTNB), Thailand. She is Assistant Professor and lecturer in Information Technology Department, Faculty of Science and Technology, at Suan Sunandha Rajabhat University (SSRU) in Bangkok. She has taught web design and programming, object-oriented programming, and internet marketing. Her research interests include mobile application programming, information and image retrieval, big data and mining, machine learning, recognition of text and speech, supply chain management system, augmented and virtual reality development.



Lap Poomhiran is currently a Ph.D. student in Information Technology, Faculty of Information Technology, King Mongkut's University of Technology North Bangkok (KMUTNB), Thailand. His research interests include web and mobile programming, augmented reality (AR) and virtual reality (VR).