

## Jobseeker-industry matching system using automated keyword selection and visualization approach

Norhaslinda Kamaruddin, Abdul Wahab Abdul Rahman, Ramizah Amirah Mohd Lawi

Advanced Analytics Engineering Center, Faculty of Computer and Mathematical Sciences,

Universiti Teknologi MARA, Malaysia

---

### Article Info

#### Article history:

Received Oct 8, 2018

Revised Dec 6, 2018

Accepted Dec 15, 2018

---

#### Keywords:

Advert filtering

Automated keyword extraction

Job searching

Skill matching

Visualization

---

### ABSTRACT

Learning opportunities are available with the accessibility of new learning technologies, discovery of untraditional learning pathways and awareness of the importance of connecting current knowledge with new learning. Such situation allows the expansion in the number of courses, programs and professional certifications offered to the students resulting to the increment of the number of graduates annually. The graduates then employed by the industry for executing the job. However, there is a growing concern about the increment of unemployed graduates in the job market. One of the reasons of the mismatch between graduates' skills and employers' needs is that the jobseekers tend to choose wrong job because they are overwhelmed by the choices and typically they just randomly send the application because it is time consuming to filter relevant advert. Such action may have repercussion to the industry because the employers need to select relevant candidates to fill up the post from the unfiltered pile of applications making the selection process lengthy and time consuming. In this paper we proposed an automated approach to match the graduates' and employers' needs using a hybrid of text mining and visualization approach to facilitate jobseekers' task of relevant job application. The important keywords are automatically extracted based on the frequency of the word used in the adverts. Then, the graduates' skills are matched from their personalized profile. Relevant visualization approaches are incorporated to facilitate the selection. It is practical and feasible for the proposed approach to be incorporated in job searching websites that can optimize jobseekers and employers time and effort for a suitable match.

Copyright © 2019 Institute of Advanced Engineering and Science.  
All rights reserved.

---

### Corresponding Author:

Norhaslinda Kamaruddin,  
Advanced Analytics Engineering Center,  
Faculty of Computer and Mathematical Sciences,  
Universiti Teknologi MARA,  
Shah Alam, Selangor, Malaysia.  
Email: norhaslinda@tmsk.uitm.edu.my

---

## 1. INTRODUCTION

Lifelong learning concept allows learners to educate themselves regardless on their stage of career and life. With the uncertainty of the current economic condition, the need for continuing studies is needed for upskilling for better work performance, seeking a higher salary and better opportunities, upgrading the technology understanding and improving marketability. This trend can be observed with high number of higher education institutions (HEI) in Malaysia that is recorded to be 699 in 2018 [1]. The higher education institutions in Malaysia includes public universities (IPTA), private higher education institutions (IPTS), polytechnics, community colleges and technical and vocational training institutions. The detailed number of the HEIs is presented in Figure 1(a). Although the number of IPTAs is maintained, the number of IPTS is

skyrocketed. This is because the demand is high despite the significantly charged higher fees when compared to public institutions. Garwe (2016) conducted a study in Zimbabwe and reported that the main factor of influencing students' choices, are namely; access and opportunity, promotional information and marketing, reference or influence by others, quality of teaching and learning, fees and cost structure and academic reputation and recognition [2]. Such situation is also applied in Malaysia. Hence, it is not surprising to see the number of graduates is also increased.

The graduates will then enter the job market and apply for the relevant posts. The opportunity to seek a relevant job become more challenging because the graduates need to compete with each other to seek the job. As the number of graduate increases, the demand and competition for a job will be more stringent. Based on the graduate tracer study conducted by Ministry of Higher Education [1], the trend for graduates for ICT sector increases exponentially as illustrated in Figure 1(b). The number of graduates in ICT sector ranging from 22,642 to 27,735 graduates that comprises of approximately 8.69% from average of total number of graduates in Malaysia (275,465 graduates) from 2013 to 2017. In the graduate tracer study, Malaysia graduate employment status is also presented [1]. Graduate employment status can be divided into five categories, which are; employed, further study, upgrading skills, waiting for work placement and unemployed. Table 1 shows the total number of graduates and their employment status from 2013 to 2017. From Table 1, it is observed that 25% to 30% graduates are unemployed and from the same report, the reasons of unemployment are discussed. Almost 70% to 75% of the participants stated that the reason that they are unemployed is because they are still seeking for a job ranging from 37,316 graduates in 2017 to 39,864 graduates in 2013. Other reasons given are; taking a break, waiting for placement to further study, responsibility towards family, jobs offered not suitable, choose not to work, not interested to work, lack of self-confidence to face working environment, health problem, refuse to move to another place and other reasons. The distribution of the graduate tracer result of the unemployment factors is depicted in Figure 2.

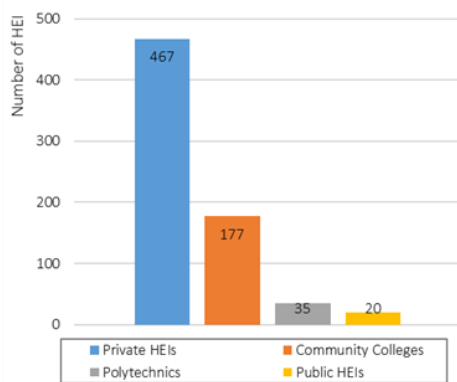


Figure 1(a). Number of malaysia higher education institute (hei) by year [1]

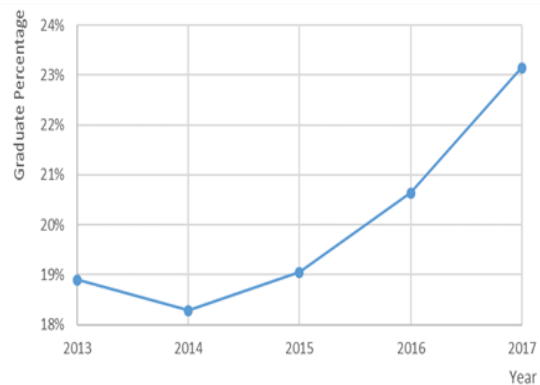


Figure 1(b). Percentage of Malaysian graduate in ICT sector from 2013 and 2017 [1]

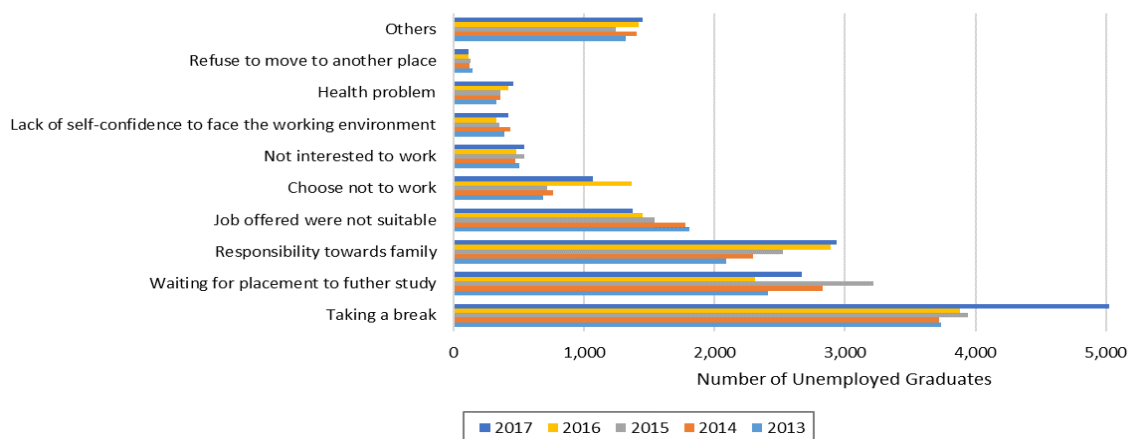


Figure 2. Factors of graduate unemployment in Malaysia from 2013 and 2018 [1]

Table 1. Malaysian Graduates Employability Status by Year [1]

Year	Employed	Further Study	Upgrading Skills	Waiting for Work Placement	Unemployment	Total Graduates
2013	101,286	42,443	2,911	12,908	53,282	212,830
2014	101,619	43,836	3,223	8,941	52,219	209,838
2015	121,740	40,067	3,776	9,133	54,852	229,568
2016	134,561	34,510	5,394	9,619	54,103	238,187
2017	141,257	43,495	5,068	11,906	53,373	255,099

From Figure 2, it is obvious that graduates are still unemployed because they are still trying to find a suitable job. The job searching process is tedious and time consuming because jobseeker needs to evaluate the relevancy of the advert before applying and for the potential employer to gauge the suitability of the candidate. There is a need to match between the graduates' skill sets with the need of the industry for the job posted. To date, there is only a basic filter provided by the job advertisement websites for job selection by specifying level of education, location, specification and minimum salary. In this paper, we are proposing a practical approach to match the graduates' skill sets and the need of the potential employer by using an automated keyword extraction and incorporates visualization to facilitate selection. For simplification, we focus on ICT web-related jobs position.

This paper is organized in the following manner. Section 2 describes on available job searching websites and different approaches used to extract relevant keywords from documents. Section 3 presents the overall framework of the proposed approach. As this work is still work in progress, a prototype of the Job Matching System is illustrated as well. To conclude the paper, Section 4 provides summary and future direction of the research.

## 2. LITERATURE REVIEW

Graduates typically will finding jobs by registering their profiles to job searching websites such as JobStreet.com [3], LinkedIn [4], Monster [5] and others. These websites provide a platform for the jobseeker to find a suitable job and potential employers to advertise job or position that they need to hire. Some of the website provide a personal assistance to the jobseeker as part of their service to facilitate the process of selecting the relevant job for the applicants. The jobseeker profile is available for the company to further analysed. Moreover, the jobseekers can also view the company profile to help them to choose the most suitable career path for them. JobStreet.com and LinkedIn need the users to sign in before they can use the services. On the contrary, Indeed allows user to directly search the relevant job by specifying the job title, keywords and company as well as the location of the job offered. Users can then create or upload resume and sign in to ensure the security of the created resume. LinkedIn offers connection to various user social media accounts for better reachability and visibility. Moreover, JobStreet.com segregated job in many fields such as illustrated in Figure 3. In this paper we are focusing on web related jobs that contains the process of web development. The five jobs that can be considered as the web related jobs, are namely; software developer, web developer, software engineer, Net developer and PHP developer.

Once scope has been set, we collected 100 job advertisements that are related to the web related job. The relevant keywords are manually extracted from the skills requirement. This is to find correlation between skills and job advertised and to see whether there are similar requirements needed by multiply job adverts. The skills are mapped onto a Vern diagram in Figure 4 to show interdependency of the skills and the job. It is noticed that there are general skills that are needed by all web related job requirements such as MS SQL, .Net and ASP with minimum qualification of diploma and/or degree in the related fields. Furthermore, it is also observed that there are specific needs for certain skill for different job. For instance, Shell Script programming is very much needed for web developer as compared to .Net developer. It also shows that it is important to prepare oneself for the skills needed by the industry for that particular job to increase chances of securing the job advertised. However, manual skill mapping is prone to error and very time consuming. Data redundancy and missing data problem will always compromise the accuracy of the mapping. Hence, an automated keyword extraction needs to be used to simplify the process.

Using the correct keywords for job searching increases the chances to find relevant job advertisements and getting shortlisted for an interview. However, jobseekers sometimes are not sure of the right keywords of the job that they want to secure. They may use brute-force approach to test any keywords that may be relevant and such approach is time consuming and energy intensive. Hence, many researchers have proposed many techniques to automatically extract relevant keywords from text documents. This process is to select words and phrases from the text documents that can give the gist of the intended information without the human involvement [6]. Once relevant keywords are extracted, the information can

be used to find a better match for job search queries. Bharti et al. [7] categorized extraction system into four classes; namely, simple statistical, linguists, machine learning and hybrid approaches. Simple statistic approach system looks from the perspective of the raw document such as frequency of the word used, location of the word in the document and is the simplest approach to implement compared to other approaches making the processing time is kept to minimal. The linguists approach incorporates the understanding of language analysis such as lexical, syntactic, discourse and semantic of the word used. The machine learning approach uses the power of classifiers such as support vector machine, naïve bayes and deep learning to understand the contents of the adverts and uses weight to match between the jobseeker application and the job advertisement. In addition, hybrid approach combines two or more previous approaches and utilize the strength of the selected approaches.

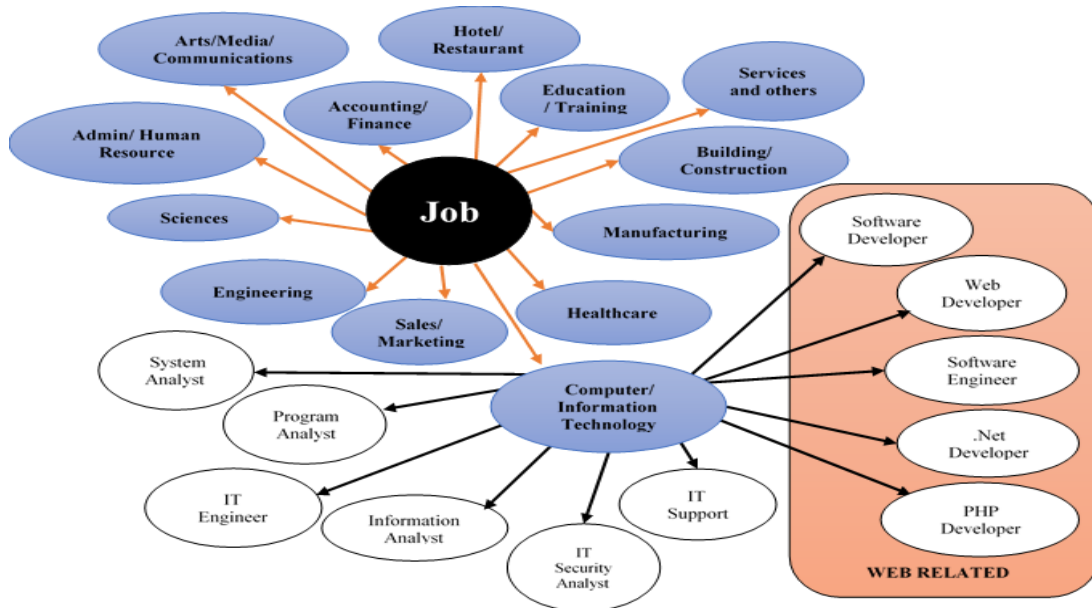


Figure 3. Classification of job based on jobstreet.com perspective [3]

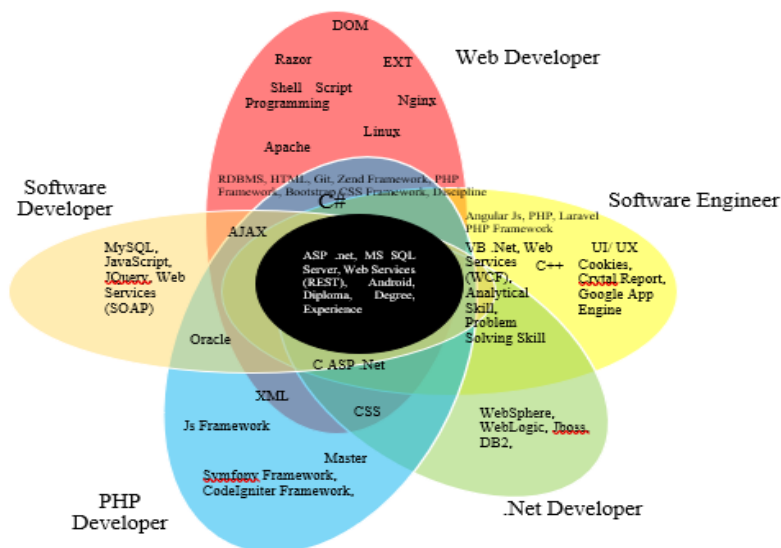


Figure 4. Job requirements interdependency based on extracted job description

Amato et al. [8] attempted to automate the resume management for matching candidate profiles with job description by comparing several techniques such as linear SVC, rule-based and Latent Dirichlet

*Jobseeker-industry matching system using automated keyword selection and... (Norhaslinda Kamaruddin)*

Allocation (LDA) to classify job advertisement against an occupation classifier used by the Italian National Institute of Statistic (ISTAT). They reported that LDA performed consistently well for average precision, recall and FScore as compared to machine learning and rule-based approach. Linear SVC needs sufficient training data to yield good performance and rule-based approach need extra attention by the expert for the rule development. Hauff and Gousios [9] proposed a pipeline that automatize job matching advertisements to developers by looking at three main components, namely; concept extraction from job advertisements and social coding user data, concept weighting and concept matching. Basic concept of Term Frequency Inverse Document Frequency (TF-IDF) is employed for concept weighing. TF-IDF gives low weight for concept that is common and appear in many documents and high weight to concepts that occur many times in a document but rarely across entire corpus. The result shows that there is a substantial overlap between the entities extracted from job advertisements and the entities extracted from developer profiles and the linear correlation between the number of times a concept appears in developer profiles vs job adverts is  $r=0.49$ . In more recent work, Muthyala et al. [10] discussed a methodology that improves user job searching experience by adding skill set and company attribute filters. The TF-IDF weighting is used to calculate the frequencies for all unique skills of the document itself. Then, several similarity measurements were used to measure similarity of the two documents based on their feature vector. The job search result shows that the filtered jobs are ranked using a relevance score derived from a weighted combination of skill sets and companies external factors. Hence, in this work we are using TF-IDF to extract relevant keywords.

### 3. RESEARCH METHODOLOGY

Our method consists of several components that requires parsing, interpreting and normalising semi/unstructured data gathered from jobseeker resume and job advertisement to create a recommendation engine that will be presented in a job searching website. Crawler will capture the posted job advertisements. In our preliminary work, 100 job advertisements are collected for five web-related jobs, namely; web developer, software developer, software engineer, PHP developer and .Net developer. The skill sets required by the companies will be extracted using TF-IDF. The jobseekers resume is recorded, and the jobseeker acquired skill set are also extracted. The acquired jobseeker and the needed company skill sets are ranked using feature ranking method. This is to increase the performance of the recommender system. For this preliminary work, a simple frequency calculation is used. Then, a concept matching is implemented to determine the similarity of the advertised job and jobseeker profile. We propose to use cosine similarity method. The proposed job recommender system workflow is presented in Figure 5.

The result will then be in the range of percentage where the higher score indicates higher similarity. For the ease of selection, a visualization approach is used for the job selection website as depicted in Figure 6. Information such as job title, location, skills needed and expected salary are revealed to the job seeker for them to make a well-informed decision and presented in the form of dashboard.

### 4. CONCLUSION

The extraction of relevant information for job selection is not a trivial task. Enabling automated keyword extraction from jobseeker resume and job advertisement may facilitate jobseeker to find relevant job in a minimal time and help companies to get better candidates to be considered for the job. Although the work presented is only preliminary work, it shows potential to be embedded in the current job searching websites. Further works need to be incorporated to ensure the success of the job recommender system. Such system can be empowered with lifelong learning to foster the continuous development and improvement of the knowledge and skills needed for employment and personal fulfilment [11].

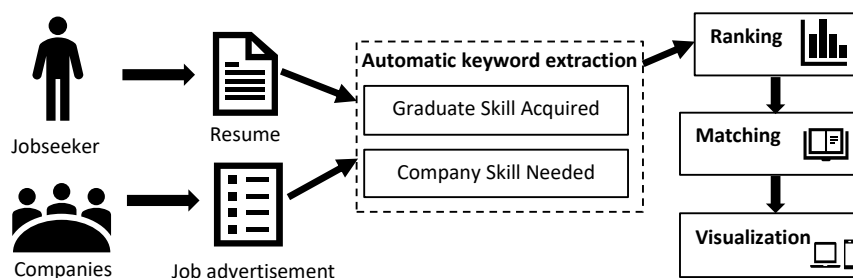


Figure 5. The proposed job recommender system workflow

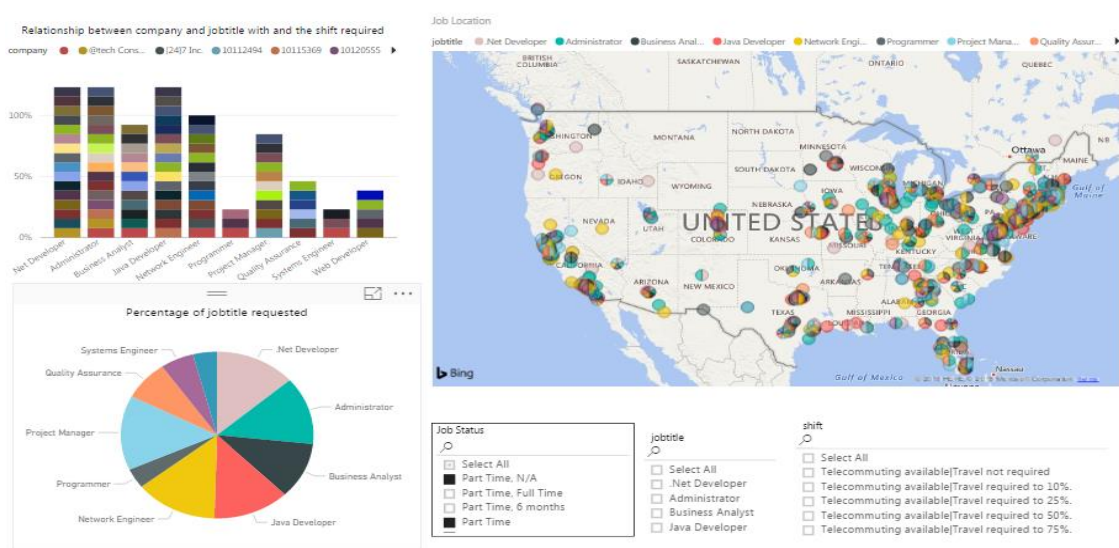


Figure 6. Example of job recommender website based on visualization approach

**ACKNOWLEDGEMENT**

The authors would like to thank Universiti Teknologi MARA (UiTM), International Islamic University Malaysia (IIUM) and Ministry of Higher Education Malaysia (MOHE) for providing financial support through the Transdisciplinary Research Grant Scheme TRGS (TRGS16-04-02-002) to conduct the work published in this paper.

**REFERENCES**

- [1] B. Ahmad, "Laporan Kajian Pengesanan Graduan," 2016. Available: <https://www.mohe.gov.my/muat-turun/awam/statistik/190-laporan-kajian-pengesanan-graduan>.
- [2] Garwe E. C., "Increase in the Demand for Private Higher Education: Unmasking the "Paradox","" *International Journal of Educational Management*, vol/issue: 30(2), pp. 232-51, 2016.
- [3] "Jobstreet.com,"<https://www.jobstreet.com.my/en/job-search/job-vacancy.php?key=ict&location=&specialization=191%2C192%2C193&area=&salary=&ojs=3&src=12>.
- [4] <https://www.linkedin.com>
- [5] <https://www.monster.com.my/>
- [6] Zhang C., "Automatic Keyword Extraction from Documents using Conditional Random Fields," *Journal of Computational Information Systems*, vol/issue: 4(3), pp. 1169-80, 2008.
- [7] Bharti S. K. and Babu K. S., "Automatic Keyword Extraction for Text Summarization: A Survey," *ArXiv preprint arXiv:1704.03242*, 2017.
- [8] Amato F., et al., "Classification of Web Job Advertisements: A Case Study," *Proceedings of the 23rd Italian Symposium on Advanced Database Systems (SEBD 2015)*. Gaeta, Italy, pp. 144-151, 2015.
- [9] Hauff C. and Gousios G., "Matching GitHub Developer Profiles to Job Advertisements," *Proceedings of the 12th Working Conference on Mining Software Repositories*. Florence, Italy, pp. 362-366, 2015.
- [10] Muthyala R., et al., "Data-driven Job Search Engine Using Skills and Company Attribute Filters," *Proceedings of IEEE International Conference on Data Mining Workshops (ICDMW)*, Louisiana, US, pp. 199-206, 2017.
- [11] Hamiz M., et al., "Assessment Analytic Theoretical Framework Based on Learners' Continuous Learning Improvement," *Indonesian Journal of Electrical Engineering and Computer Science*, vol/issue: 11(2), pp. 682-687, 2018.
- [12] Muthyala R, Wood S, Jin Y, Qin Y, Gao H, Rai A. Data-driven Job Search Engine Using Skills and Company Attribute Filters. *Proceedings of IEEE International Conference on Data Mining Workshops (ICDMW)*, Louisiana, US. 2017; 199-206.
- [13] Lawrence S, Lee Giles C, Bollacker K. Digital libraries and autonomous citation indexing. *IEEE Journals on Computer*. vol. 32, no. 6, pp. 67-71, June 1999.
- [14] Yih W T, Goodman J, Carvalho V R. Finding advertising keywords on web pages. In *Proceedings of the 15th international conference on World Wide Web (WWW '06)*. ACM, New York, NY, USA, 2006, 213-222.
- [15] Hamiz M, Bakri M, Kamaruddin N, Mohamed A. Assessment Analytic Theoretical Framework Based on Learners' Continuous Learning Improvement. *Indonesian Journal of Electrical Engineering and Computer Science*. 2018;11(2):682-687.