Education and Information Technologies

The Official Journal of the IFIP Technical Committee on Education

Volume 26, issue 2, March 2021

54 articles in this issue

Editorial for EAIT issue 2, 2021

Arthur Tatnall

EditorialNotes | Published: 10 March 2021 | Pages: 1339 - 1352

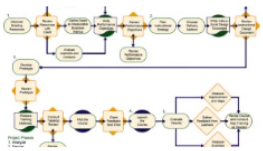
The effects of robotics training on students' creativity and learning in physics

Alireza Badeleh

OriginalPaper | Published: 14 August 2019 | Pages: 1353 - 1365

This is part of 1 collection

A literature review: efficacy of online learning courses for higher education institution using meta-analysis



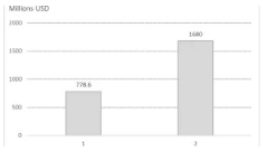
Mayleen Dorcas B. Castro & Gilbert M. Tumibay

OriginalPaper | Published: 04 November 2019

Pages: 1367 - 1385

This is part of 1 collection

Dimensions of robotic education quality: teachers' perspectives as teaching assistants in Thai elementary schools



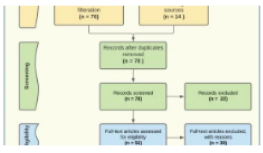
Suparoek Chootongchai, Noawanit Songkram & Krerk Piromsopa

OriginalPaper | Published: 17 December 2019

Pages: 1387 - 1407

This is part of 1 collection

Cultural challenges eLearners from the GCC countries face when enrolled in Western educational institutions: A thematic literature review



Nafeth Al Hashlamoun

OriginalPaper | Published: 27 August 2020

Pages: 1409 - 1422

The impact of a technological intervention program on

You have access to our articles

For authors

Submission guidelines

Language editing services

Ethics & disclosures

Open Access fees and funding

Contact the journal

Submit manuscript

Working on a manuscript?

Avoid the most common mistakes and prepare your manuscript for journal editors.

Learn more →

Explore

Online first articles

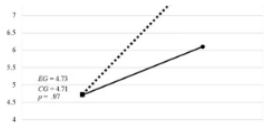
Volumes and issues

Collections

Sign up for alerts

Advertisement

[The impact of a technological mind-mind program on learning mathematical skills]



Omar Cóndor-Herrera & Carlos Ramos-Galarza

OriginalPaper | Published: 27 August 2020 |

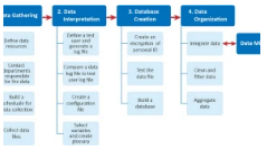
Pages: 1423 - 1433

[Examining PRESERVICE teachers' use of SMARTBOARD and pc tablets in lessons]

Murat Yalman & Bulent Basaran

OriginalPaper | Published: 01 September 2020 | Pages: 1435 - 1453

[Behind the scenes of educational data mining]



Yael Feldman-Maggor, Sagiv Barhoom ... Inbal Tuvi-Arad

OriginalPaper | Published: 01 September 2020 |

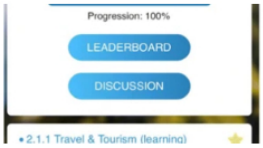
Pages: 1455 - 1470

[Effect of learning management system on Student's performance in educational measurement and evaluation]

Basil C. E. Oguguo, Fadip Audu Nannim ... Augustina C. Nzeadibe

OriginalPaper | Published: 02 September 2020 | Pages: 1471 - 1483

[Incorporating new literacies in designing a mobile learning application for secondary/middle school students]

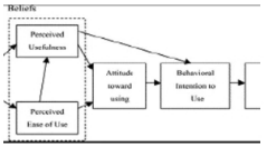


Chiew Hong Ng, Noi Keng Koh & Hua Loon Ling

OriginalPaper | Published: 02 September 2020 |

Pages: 1485 - 1504

[The two perfect scorers for technology acceptance]



Pritika Reddy, Kaylash Chaudhary ... Ronil Chand

OriginalPaper | Published: 03 September 2020 |

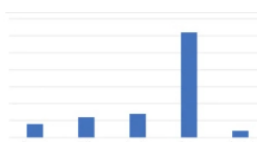
Pages: 1505 - 1526

[A machine learning approximation of the 2015 Portuguese high school student grades: A hybrid approach]

Ricardo Costa-Mendes, Tiago Oliveira ... Frederico Cruz-Jesus

OriginalPaper | Open Access | Published: 05 September 2020 | Pages: 1527 - 1547

[Systematic literature review of STEM self-study related ITSs]

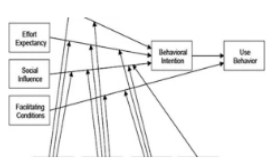


Asmaa Alabdulhadi & Maha Faisal

OriginalPaper | Published: 07 September 2020 |

Pages: 1549 - 1588

[Toward a model for acceptance of MOOCs in higher education: the modified UTAUT model for Saudi Arabia]



Maryam Altalhi

OriginalPaper | Published: 07 September 2020 |

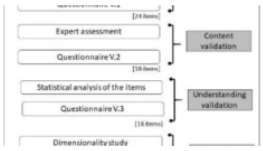
Pages: 1589 - 1605

[Teaching-learning conceptions, teaching motivation, and perceptions towards ICT: A research in Turkish public high schools]

Gökhan Baş & Muhammet Baştuğ

OriginalPaper | Published: 08 September 2020 | Pages: 1607 - 1625

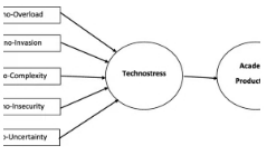
[Design and validation of an instrument of self-perception regarding the lecturers' use of ICT resources: to teach, evaluate and research](#)



Francisco D. Guillén-Gámez & María J. Mayorga-Fernández

OriginalPaper | Published: 08 September 2020 | Pages: 1627 - 1646

[Impact of technostress on academic productivity of university students](#)



Pallavi Upadhyaya & Vrinda

OriginalPaper | Open Access | Published: 11 September 2020 | Pages: 1647 - 1664

["I'm not against online teaching, but what about us?": ICT in Ghana post Covid-19](#)

Michael Agyemang Adarkwah

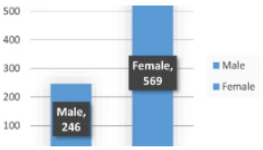
OriginalPaper | Published: 16 September 2020 | Pages: 1665 - 1685

['Dreaming in colour': disabled higher education students' perspectives on improving design practices that would enable them to benefit from their use of technologies](#)

Jane Seale, Chetz Colwell ... Dorit Olenik-Shemesh

OriginalPaper | Open Access | Published: 16 September 2020 | Pages: 1687 - 1719

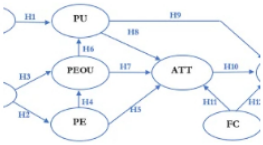
[College students information security awareness: a comparison between smartphones and computers](#)



Nashrawan Taha & Laila Dahabiyeh

OriginalPaper | Published: 16 September 2020 | Pages: 1721 - 1736

[Fun first, useful later: Mobile learning acceptance among secondary school students in Indonesia](#)



Ahmad R. Pratama

OriginalPaper | Published: 17 September 2020 | Pages: 1737 - 1753

[Influencing factors in M-learning adoption in higher education](#)

Abeer Qashou

OriginalPaper | Published: 17 September 2020 | Pages: 1755 - 1785

[Perceptions of E-learning among undergraduates and academic staff of higher educational institutions in north-eastern Nigeria](#)

Muhammed Kuliya & Sani Usman

OriginalPaper | Published: 17 September 2020 | Pages: 1787 - 1811

[To WeChat or to more chat during learning? The relationship between WeChat and learning from the perspective of university students](#)

Rujing Hou, Shenghao Han ... Chubing Zhang

OriginalPaper | Published: 18 September 2020 | Pages: 1813 - 1832

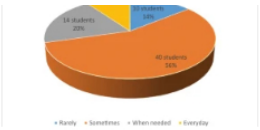
[Knowledge sharing technologies in higher education:](#)

Preferences of CIS students in Cyprus

Fezile Ozdamli & Nadire Cavus

OriginalPaper | Published: 18 September 2020

Pages: 1833 - 1846



The effectiveness of a virtual fitness trainer app in motivating and engaging students for fitness activity by applying motor learning theory

Nur Azlina Mohamed Mokmin & Nurullizam Jamiat

OriginalPaper | Published: 18 September 2020

Pages: 1847 - 1864

Evaluating the impact of ICT on teaching and learning: A study of Palestinian students' and teachers' perceptions

Husam Qaddumi, Brendan Bartram & Ali L. Qashmar

OriginalPaper | Published: 19 September 2020 | Pages: 1865 - 1876

Investigation into Nomophobia amongst Turkish pre-service teachers

Nese Sevim-Cirak & Omer Faruk Islim

OriginalPaper | Published: 19 September 2020 | Pages: 1877 - 1895

Parental mediation of adolescent Internet use: Combining strategies to promote awareness, autonomy and self-regulation in preparing youth for life on the web

Nili Steinfeld

OriginalPaper | Published: 22 September 2020 | Pages: 1897 - 1920

Augmented reality (AR) as a learning material in special needs education

Hasan Köse & Nevin Güner-Yildiz

OriginalPaper | Published: 23 September 2020 | Pages: 1921 - 1936

Feelings of belonging or alienation and social emotional perceptions of immigrant youths in the digital age, in comparison with native-born youths

Gila Cohen Zilka

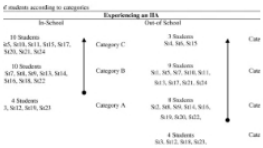
OriginalPaper | Published: 24 September 2020 | Pages: 1937 - 1954

Secondary vocational education students' expressed experiences *of* and approaches *to* information interaction activities within digital environments: a Phenomenographic study

Fernando Bolaños & Álvaro Salinas

OriginalPaper | Published: 24 September 2020

Pages: 1955 - 1975

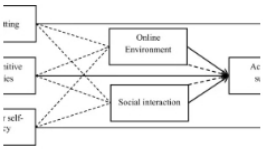


Examination of influence of e-learning on academic success on the example of Bosnia and Herzegovina

Edisa Puška, Adisa Ejubović ... Adis Puška

OriginalPaper | Published: 25 September 2020

Pages: 1977 - 1994

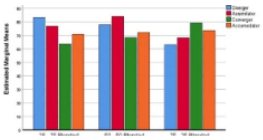


An exploration into EFL learners' writing skills via mobile-based dynamic assessment

Saman Ebadi & Saba Bashir

OriginalPaper | Published: 26 September 2020 | Pages: 1995 - 2016

[Online and face-to-face composition in forming the professional competencies of technical teacher candidates with various learning style types](#)

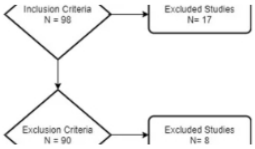


Hamonangan Tambunan, Marsangkap Silitonga & Uli Basa Sidabutar

OriginalPaper | Published: 28 September 2020 |

Pages: 2017 - 2031

[Context aware mobile learning: A systematic mapping study](#)



Bimal Aklesh Kumar, Bibhya Sharma & Elisa Yumi Nakagawa

OriginalPaper | Published: 28 September 2020 |

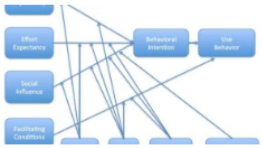
Pages: 2033 - 2052

[A review of technologies for collaborative online information seeking: On the contribution of collaborative argumentation](#)

Elisabeth Mayweg-Paus, Maria Zimmermann ... Niels Pinkwart

OriginalPaper | Open Access | Published: 30 September 2020 | Pages: 2053 - 2089

[User acceptance of learning innovation: A structural equation modelling based on the GUAM framework](#)



A. C. Obienu & F. I. Amadin

OriginalPaper | Published: 01 October 2020 |

Pages: 2091 - 2123

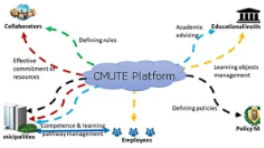
[Playfulness in the classroom: Gamification favor the learning of pharmacology](#)

Francislaine Aparecida dos Reis Lívero, Gustavo Ratti da Silva ... Evellyn Claudia Wietzikoski Lovato

OriginalPaper | Published: 02 October 2020 |

Pages: 2125 - 2141

[Cloud computing and semantic web technologies for ubiquitous management of smart cities-related competences](#)

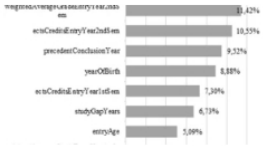


Omiros Iatrellis, Theodor Panagiotakopoulos ... Achilles Kameas

OriginalPaper | Published: 04 October 2020 |

Pages: 2143 - 2164

[A data-driven approach to predict first-year students' academic success in higher education institutions](#)



Paulo Diniz Gil, Susana da Cruz Martins ... Joana Martinho Costa

OriginalPaper | Published: 06 October 2020 |

Pages: 2165 - 2190

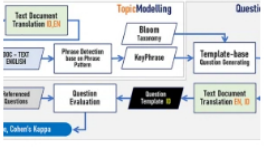
[Identifying the differentiation practices of virtual school teachers](#)

Dennis Beck & Jennifer Beasley

OriginalPaper | Published: 08 October 2020 |

Pages: 2191 - 2205

[Question generation model based on key-phrase, context-free grammar, and Bloom's taxonomy](#)

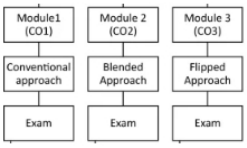


Bambang Dwi Wijanarko, Yaya Heryadi ... Widodo Budiharto

OriginalPaper | Published: 10 October 2020 |

Pages: 2207 - 2223

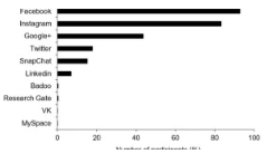
[Evaluation of course outcome attainment of engineering course with traditional, blended and flipped classroom approaches](#)



C. M. Vivek & P. Ramkumar

OriginalPaper | Published: 13 October 2020 |
Pages: 2225 - 2231

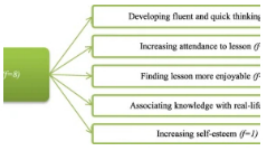
[The usage of, and confidence in, social media as study sources among undergraduate students: A cross-sectional survey comparing it with traditional study sources](#)



Ricardo Borges Viana, Alex Vito Santos Neves-Silva ... Claudio Andre Barbosa de Lira

OriginalPaper | Published: 15 October 2020 |
Pages: 2233 - 2252

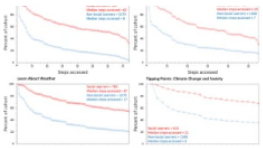
[The effect of virtual learning environments designed according to problem-based learning approach to students' success, problem-solving skills, and motivations](#)



Seda Akti Aslan & Kemal Duruhan

OriginalPaper | Published: 15 October 2020 |
Pages: 2253 - 2283

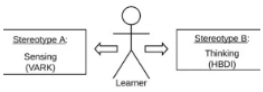
[Influence of social learning on the completion rate of massive online open courses](#)



R. A. Crane & S. Comley

OriginalPaper | Open Access | Published: 21 October 2020 |
Pages: 2285 - 2293

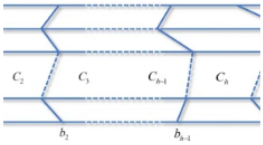
[Personalized tutoring through a stereotype student model incorporating a hybrid learning style instrument](#)



Christos Troussas, Konstantina Chrysafiadi & Maria Virvou

OriginalPaper | Published: 26 October 2020 |
Pages: 2295 - 2307

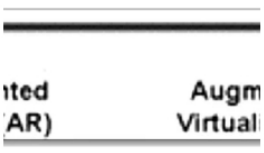
[Learner classification based on interaction data in E-learning environments: the ELECTRE TRI method](#)



Muhittin Şahin, Aydın Ulucan & Halil Yurdugül

OriginalPaper | Published: 26 October 2020 |
Pages: 2309 - 2326

[Facilitating the development of Preservice teachers' proportional reasoning in geometric similarity problems using augmented reality activities](#)



Muhammet Arican & Bilal Özçakir

OriginalPaper | Published: 28 October 2020 |
Pages: 2327 - 2353

← Previous 1 2 Next →

Discover content

Journals A-Z

Books A-Z

Publish with us

Publish your research

Open access publishing

Products and services

Our products

Librarians

Societies

Partners and advertisers

Our imprints

Springer

Nature Portfolio

BMC

Palgrave Macmillan

Apress

Your privacy choices/Manage cookies | Your US state privacy rights | Accessibility statement | Terms and conditions | Privacy policy | Help and support

13.251.9.252

6763 SpringerLink Indonesia eJourna Consortium - Higher Education (3000122892) - Universitas Kristen Maranatha (3000181437) - 11741 SpringerLink Indonesia eJourn Consortium (3000951794)

SPRINGER NATURE

© 2023 Springer Nature

Education and Information Technologies
https://doi.org/10.1007/s10639-020-10356-4

Question generation model based on key-phrase, context-free grammar, and Bloom's taxonomy

Bambang Dwi Wijanarko[1,2] · Yaya Heryadi[1] · Hapnes Toba[3] · Widodo Budiharto[1]

Received: 3 July 2020 / Accepted: 7 October 2020/Published online: 10 October 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Automated question generation is a task to generate questions from structured or unstructured data. The increasing popularity of online learning in recent years has given momentum to automated question generation in education field for facilitating learning process, learning material retrieval, and computer-based testing. This paper report on the development of question generation framework based on key-phrase method for online learning with a constraint that the generated questions should comply with the learning outcomes and skills from Bloom's Taxonomy. The proposed method was tested using learning materials of Software Engineering course for undergraduate level written in Bahasa Indonesia obtained from Bina Nusantara's (Binus's) Online Learning repository. Using one-semester lecture material, this study generated 92,608 essay-type questions from 6-level Bloom's Taxonomy which were further sampled randomly to obtain 120 question samples for method evaluation. Performance evaluation using average Bilingual Evaluation Understudy (BLEU) involving five independent reviewers toward samples of these questions achieved 0.921 and 0.6 Cohen's Kappa. The relevance of Bloom's Taxonomy level of the generated questions was evaluated by means of classification model with 0.99 accuracy. The results indicate that not only are the generated questions well understood and agreed by the reviewers, they are also relevant to the expected Bloom's Taxonomy level there for the questions can be delivered to students in the respected course delivery and evaluation.

Keywords Bloom's taxonomy · Context-free grammar · Keyphrase extraction · Question generation · Question template

✉ Bambang Dwi Wijanarko
bwijanarko@binus.edu

Extended author information available on the last page of the article

1 Introduction

Question generation (QG) is a task to generate question from structured or unstructured data. This task is an interesting Natural Language Processing problem with wide potential applications in many domains. In education field, for instance, QG plays some important roles in learning process and computer-based testing (Herranen and Aksela 2019). In learning process, question generation can be used to assist students in an assisted learning system with their comprehension on learning material. By means of assisted learning system, students learn to respond to the system generated questions about facts and ideas related to the topics being studied.

In the recent years, research on question generation in education field has grown rapidly due to the availability of large scale training dataset, the advent of deep learning, and the development of GPU-based computing technology. The research topic range from enhancing theoretical aspect of question generation to improving the varieties of question generating approaches and methods (Al-Yahya 2011; Lindberg and Sc 2013; Lindberg and Sc 2013; Duan et al. 2017; Rodrigues 2017; Ye and Wang 2018; Elsahar et al. 2018; Huang et al. 2018), Automatic QG system in education is adapted from general QG with an additional constraint: the generated questions should comply with learning objectives and skills that educators set for their students in particular topic. The learning objectives of a particular course are commonly set base on Bloom's Taxonomy (Gleason 2018; Quintana and Tan 2019).

So far many QG studies have used general datasets that are not learning content and produce factual or shallow questions, whereas education and learning require in-depth questions that are following Bloom's taxonomy. Research opportunities in the QG field are still open (Heilman et al. 2011; Adamson et al. 2013; Jouault et al. 2016). The limitation of this research is that the questions generated by QG are only for lecture material that is more theoretical, but not for materials that use numbers and calculations, schemes, or program code such as mathematics, theoretical computer science, systems programming, etc.

One of the topics which attracts the attention of researchers is how to efficiently generate questions which comply with higher-level learning objectives and skills. According to (Willis et al. 2019), the main research challenge is how to generate questions to support effective teaching especially to promote critical thinking, retention, and context involvement.

Adressing this problem (Singh et al. 2017) suggest a method to generate question based-on noun phrases. The method comprises two stages. The first is extracting key-phrases from input texts. The second is generating questions based on the extracted key-phrase. The key-phrase based methods is potensial because they have ability to capture the context of the syntax and semantics of the input text. However, this method is not practical for real input text as it is not scalable to be applied for large volume of input text.

Several attempts to overcome the problem of key-phrase extraction have been proposed by (Meng et al. 2017) who proposed a question generation method based on a generative encoder-decoder model. Another method was proposed by (Yang et al. 2017) which produces key-phrases based on the semantic meaning of texts, and the question generation method use tags and linguistic rules.

One prominent QG method is construction model proposed by (Subramanian et al. 2018). The proposed model uses the probability of word order in the document as key-phrases which then acts as a target for the construction of question generation.

Although many studies on question generating for education purposes have been conducted, to the best of our knowledge, there are only few reports on keyphrase-based method which uses several levels of Bloom's taxonomy as constraint in generating questions. Therefore, the objective of this study is to propose a method that generates questions based on key-phrase which embed Bloom's taxonomic in selecting contexts for constructing questions.

Research questions that need to be analyzed with various reference methods such as information extraction, template-based questions, semantic roles, and evaluating the Question Generation method use various metrics to measure the performance of the text generation model to find solutions to research problems, namely:

1) How to identify KeyPhrase to find hidden information in semi-unstructured documents.
2) How to extract Context and Phrase from a text document to design Bloom's Taxonomy based questions.

2 Related work

A prominent definition of question generation was proposed by (Rus et al. 2008) Question generation is defined as the automatic generation of questions from inputs such as text, raw data, and knowledge bases. The generated questions can be in such form as factual-type, Yes/No-type, or why-type questions.

Another definition is proposed by (Yao et al. 2012). Question generation is defined as the task of generating reasonable questions from various input data such as text, database, or semantic representation. Further, (Yao et al. 2012; Nema et al. 2019) defined question generation as a system to generate rational questions from structured or unstructured data.

According to (Graesser et al. 2009; Heilman and Smith 2010; and Jouault et al. 2016; Kumar et al. 2018), various question generating systems can be categorized broadly into several categories namely:

1) Shallow question generating: QG systems designed to generate fact questions. For example question about: "who", "what", "where", "which one", "how many", and "yes/no answer". Despite the simplicity, these questions do not support deep learning and discussions (Adamson et al. 2013).
2) Deep question generating: QG systems designed to generate questions that need logical thinking to answer. For examples, questions which begin with "why", "why not", what if", "what if not", and "how"

A study by (Jouault et al. 2016; Divate and Salgaonkar 2017) concluded that there is a potential gap between the knowledge of QG system and that of human experts used to generate questions based on input text. This becomes an issue when some human experts were asked to evaluate questions generated by a QG system. While a QG

system only generates questions base on explisit representation of information from the input text, human experts can generate questions might have deep knowledge not written in the text input. To address this issue, (Jouault et al. 2016) proposed a scheme to be used by human experts to evaluate deep generated questions produced by a QG system (see Table 1).

QG approach for general applications that exploits features extracted from input text has gained wide research interests resulting in a vast number of methods. A study by (Han et al. 2018) proposed topic phrase extraction method based on features provided from the input text as a basis for question generation. A method proposed (Chao and Li 2018; Kurdi et al. 2019). refined the previous method by adding contextual information to the extracted topic phrase. (Li et al., 2019) showed some evidences that context helps improving performance of question generation.

Topic extraction is also used by (Xie et al. 2017) to set the subject and predicate in the generated questions. (Tong et al. 2019) proposed a method that uses a word and expands the word with the context structure of the word context in the input text.

In the area of QG for education, (Diab and Sartawi 2017) proposed a method that uses the semantic relationship between verbs in the generated questions and learning outcomes to evaluate compliance of a generated questions with a particular Bloom's taxonomy levels. Another study results reported by (Yang et al. 2017) proposed tags and linguistic rules to extract features as a basis for generating questions.

A study by (Emu et al. 2017) proposed a method base on structure of phrases extracting important features in the TF × IDF document representation matrix. A similar method using embedding key phrases to extract unique key phrases from scientific articles and ranking key phrases using PageRank was proposed by (Mandal et al. 2018).

The advent of deep learning algorithms has motivated many researchers to use the algorithm for QG. A recent research used deep learning model to study end-to-end

Table. 1 Deep question evaluation scheme

Question category	Evaluation criteria
C1: Questions asking facts.	Remembering
C2: Questions asking causal relations.	Comprehension
C3: Application	Questions are more complex than C2 but does not require complete integrated knowledge like C4. It requires an understanding of the topic of the questions and its context.
C4: Analysis	Questions requiring integrated knowedge of the topic as a whole. It requires knowledge of the topic of the questions as well as a general understanding of the main topic important events and theor context.
C5: Evaluation	Questions requiring a deep thinking. It requires having understanding of global history and the relations between the topic and other historical topics.

Source: (Jouault et al. 2016; Bloom 1956)

nerves to produce pairs of questions and answers through context paragraph input (Willis et al. 2019).

A study by (Gan and Yu 2018) concluded that QG is not merely the syntactical problem but more to relation between syntac-semantic. The author argue that machine-learning based methods produce low accuracy due to its limitation to capture phrase structure, scope of syllabus, difficulty level, and cognitive level as guided by Bloom's taxonomy (Kale and Kiwelekar 2013).

3 Research method

3.1 Research framework

The research framework of this study is shown in Fig. 1. The research framework consist of: (i) translating text document to English, (ii) Phrase extraction using Context-free grammar, (iii) Question generation, and (iv) Generating question evaluation using human experts.

3.2 Dataset

The source of dataset for this study is Binus Online Learning repository. The input dataset comprises course syllabus, lecture materials, case study materials, student-lecturer discussion forum log, and student grade database for software engineering course for undergraduate program. The input textual dataset contains 127,299 characters, 19,673 words, 1347 sentences, and 225 paragraphs.

3.3 Topic extraction

Every sentence should contain a topic that reflect the semantic meaning of the sentence. The topic of a sentence is typically an object or a noun phrase. Hence, topic detection in this study is implemented by tokenization to extract noun phrases as key-phrase of a sentence. The candidates noun phrases are then used as the object of the questions combined with verbs derived from Bloom's taxonomy.

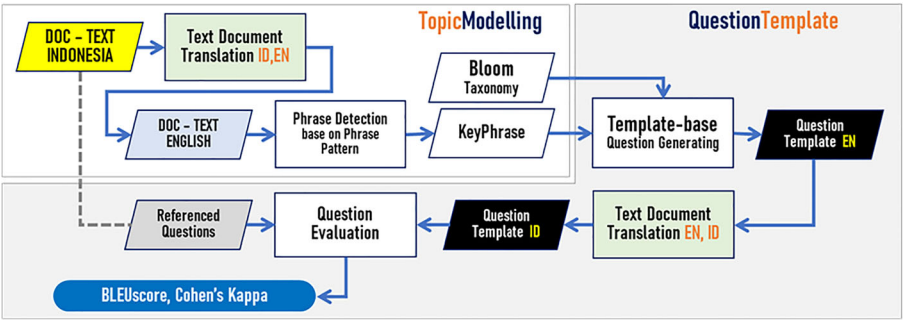


Fig. 1 Research framework

3.4 Key-phrase selection

Key-phrases can be viewed as high-level representation of a text document. They have been widely used as input for several document analysis and modeling such as document summarization, clustering, and topic analysis. Automatic keyphrase extraction, however, is a challenging task to capture the main topics of documents. Some prominent studies proposed various methods from text mining field including indexing, clustering, and summarisation.

A study by (Mikolov et al. 2013), for example, proposed a method based on probability of the two-word sequence which is measured as follows:

$$score\left(w_i, w_j\right) = \frac{count\left(w_i\, w_j\right) - \delta}{count\left(w_i\right) \times count\left(w_j\right)} \tag{1}$$

where: $count(w_i, w_j)$ is the number of a sequence of word w_i followed by w_j, $count(w_i)$ the number of word w_i, $count(w_j)$ the number of word w_j in the input document, and δ is a constant to prevent too many phrases formed by some less frequent words.

Another method is proposed by (Turney 2003) which treated the problem as a classification task. In the proposed method, the author proposed Naive Bayes model as the key-phrase classifier. The proposed method comprises three steps.

1) Document pre-processing: deleting stop words, deleting non-alpha-numerics, separating text into phrases, and stemming phrases.
2) Extracting the TF × IDF score from all sample phrases. This score is a standard metric in information retrieval to measure how specific the P phrase in document D formulated as:

$$TF \times IDF(P, D) = Pr[ph\ in\ D\ is\ P] \times (-logPr[P\ in\ a\ doc]) \tag{2}$$

where, $Pr[ph\ in\ D\ is\ P]$ is the probability of the phrase P contained in document D, and $logPr[P\ in\ a\ doc]$ is the document in corpus containing P (excluding D).

3) Calculating the distance of each sample phrase from the beginning of the corresponding document. The distance is calculated as the number of words preceding the first appearance, divided by the number of words in the document. The resulting feature is a number between 0 and 1 that represents the proportion of documents before the first appearance of the phrase. In this method the attribute, $TF \times IDF$ and the distance between phrases are

assumed to be independent because the $TF \times IDF$ phrase is discrete with a T value and the distance is D. Then, the probability of a unique phrase is:

$$Pr[key|T,D] = \frac{Pr[T|key] \times Pr[D|key] \times Pr[key]}{Pr[T,D]} \quad (3)$$

Where $Pr[T|\,key]$ is the probability that a Keyphrase has a $TF \times IDF$ with score T, $Pr[D|\,key]$ the probability that it has a distance D. $Pr[T|\,key]$ is the probability a priori that the phrase is Keyphrase, and $Pr[T, D]$ is the normalization factor, so that the value of $Pr[key|\,T, D]$ is between zero and one.

Figure 2 shows the relationship between phrases and the sentences containing the phrases. Each phrase is identified using a postag to determine the phrase category as nouns (NN), adjective (JJ) or verb (VB). Phrases that have been marked with a postag become key phrase and will determine the types of question. Key-phrases of the type NN will become context questions with operational verbs with correspond to Bloom's taxonomic levels.

If both noun-phrase and adjective phrase are found in a sentence, then both of them will be used to form a question. However, if there is only one adjective key phrase, the noun phrase from the previous sentence will be used to form a question.

3.5 Question template construction

Context-Free Grammar (CFG) is a class of formal grammar defined as a quadruple $G = (V, \Sigma, P, S)$, where: V is a set of non-terminal symbols. Σ is a set of terminal symbols such that $V \cap \Sigma = \varnothing$; P is a set of rules in a form of $P : V \rightarrow (V \cup \Sigma)^*$, and S is a start symbol (Hopcroft et al. 2001). In formal language, CFG rules can be seen as a set

SENTENCE	POSTAG
Software Engineering is a branch of science that focuses on theories, methods or supporting tools that develop software professionally.	[('software', 'NN'), ('engineering', 'NN')]
The software development stage from requirements analysis, planning, design, software creation, to maintenance, the costs incurred at the maintenance stage take up the largest portion.	[('maintenance', 'NN'), ('stage', 'NN')] [('requirements', 'NNS'), ('analysis', 'NN')] [('software', 'NN'), ('creation', 'NN')] [('software', 'NN'), ('development', 'NN'), ('stage', 'NN')]
Some people consider the life cycle model to be more general for a more specific category of methodology and software development processes to refer to the process chosen by an organization.	[('life', 'NN'), ('cycle', 'NN'), ('model', 'NN')] [('software', 'NN'), ('development', 'NN'), ('processes', 'NNS')] [('specific', 'JJ'), ('category', 'NN')]
The software development methodology also known as the HR framework does not appear until.	[('software', 'NN'), ('development', 'NN'), ('methodology', 'NN')]
According to Elliotts, the life cycle of SDLC system development can be considered as the oldest formal methodological framework for building information systems.	[('formal', 'JJ'), ('methodological', 'JJ'), ('framework', 'NN')] [('information', 'NN'), ('systems', 'NNS')] [('system', 'NN'), ('development', 'NN')]
The main idea of SDLC is to pursue the development of information systems in a very deliberate, structured and methodical manner, requiring every stage of the life cycle from the initial idea to the delivery of the system which must be carried out rigidly	[('information', 'NN'), ('systems', 'NNS')] [('initial', 'JJ'), ('idea', 'NN')] [('methodical', 'JJ'), ('manner', 'NN')]
Methodologies, processes and frameworks from proscriptive steps that can be used directly by organizational inactivity to flexible frameworks that the organization uses to produce a series of specific steps tailored to the needs of the project or group.	[('organization', 'NN'), ('uses', 'NNS')] [('organizational', 'JJ'), ('inactivity', 'NN')] [('proscriptive', 'JJ'), ('steps', 'NNS')] [('specific', 'JJ'), ('steps', 'NNS')]
It should be noted that since DSDM came in, all the methodologies listed above except the RUP have become agile methodologies, but many organizations, especially the government, still use a pre-agile process that is often a waterfall or similar.	[('agile', 'NN'), ('methodologies', 'NNS')] [('pre-agile', 'JJ'), ('process', 'NN')]
The development team can also approve the programming environment, such as which integrated development environment is used, and one or more dominant programming paradigms, programming style rules, or choice of software libraries or software frameworks.	[('development', 'NN'), ('environment', 'NN')] [('development', 'NN'), ('team', 'NN')] [('software', 'NN'), ('frameworks', 'NNS')] [('software', 'NN'), ('libraries', 'NNS')]
Software design can refer to all activities involved in conceptualizing, framing implementing, commissioning, and finally modifying systems or complex activities that follow the requirements specifications and before programming.	[('complex', 'JJ'), ('activities', 'NNS')] [('requirements', 'NNS'), ('specifications', 'NNS')]

Fig. 2 Sentence Key-phrase PostTag

of written recursive rules that are used to produce strings based on specified patterns. Therefore, CFG is the basis for the syntax of programming languages that provide efficient algorithms and parsing in programming languages, such as: files and data flow (Moraes et al. 2018).

In this study, CFG is used to develop question templates based on the structure of operational phrases and verbs in the Bloom's taxonomy. In this study, Structure of the question is constructed using the verb Bloom's Taxonomy and is paired with two adjacent key-phrases from the sentence in the document. The assumptions as follows:

1. Noun-phrases in general have a unique meaning or have a clear context,
2. Adjective can not stand alone; it is an explanation of a noun-phrase.

3.6 The construction of questions uses context (noun-phrase), topic (key-phrase), Bloom's taxonomy

A set of Question Templates is constructed using the following steps. The first step is, selecting key phrases in the form of noun (NN) and adjective (JJ). Adjective key-phrases are paired with noun-phrases from the previous phase. Then, the questions are formed automatically with key-phrases parsed into the procedure for making template questions (Fig. 4).

Based on the question construction in Fig. 3, the number of questions generated from the Question Generation algorithm execution is determined by the number of key-phrases found from the input document. The results of experiments on the questions generated, showed that on average there are 2 key-phrases consisting of 1 to 3 words, and an average sentence length of 140 words. Finally, the automated phrase-based question construction algorithm generated more than 92,608 questions classified in 6 Bloom taxonomic levels (see Table 2).

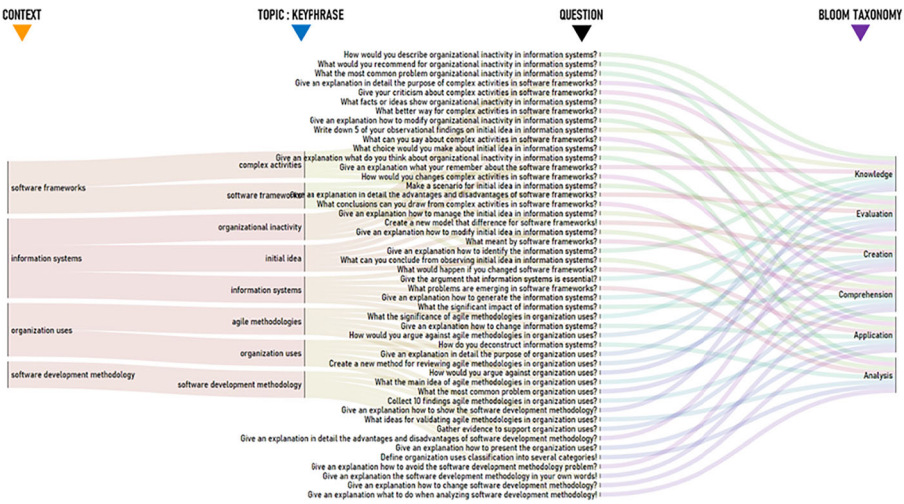


Fig. 3 Question template construction

Table 2 Output from question template algorithm

No	ID_Quest	KEY-Phrase	Question Generating	BLOOM Level
1	20_04	software engineering	What can observe about software engineering?	Comprehension
2	30_11	software engineering	What information is useful for software engineering?	Application
3	40_01	software engineering	How to sort the parts software engineering?	Analysis
4	50_06	software engineering	What changes to software engineering as recommendation?	Evaluation
5	60_08	software engineering	Determine the formula to solve the problem of software engineering!	Creation
6	20_06	software engineering	What is the significant impact of software engineering?	Comprehension
7	30_04	software engineering	How to modify software engineering?	Application
8	40_05	software engineering	Find a number of ways to problem-solving of software engineering!	Analysis
9	50_12	software engineering	How to overcome software engineering weaknesses?	Evaluation
10	50_07	important part	How to handle important part in Software Re-engineering?	Evaluation
11	60_02	special attention	What must change to revise special attention in software engineering?	Creation
12	10_03	state expenditures	How to identify the state expenditures?	Knowledge
13	30_10	state expenditures	What are the instructions for state expenditures?	Application
14	60_07	state expenditures	Develop a proposal that would produce state expenditures!	Creation
15	50_12	state expenditures	How to overcome state expenditures weaknesses?	Evaluation
16	40_01	state expenditures	How to sort the parts state expenditures?	Analysis
17	40_08	significant portion	What was the underlying problem with significant portion in state expenditures?	Analysis
18	20_07	software development	What software development are most popular?	Comprehension
19	30_05	software development	Base on experience implement for the development of software development?	Application
20	40_06	software development	What are some of the problems of software development?	Analysis
21	50_12	software development	How to overcome software development weaknesses?	Evaluation
22	40_11	overall development	What we must know about overall development in software development?	Analysis
23	10_08	software development	How to indicate a software development?	Knowledge
24	60_01		What alternative to suggest for software development?	Creation

Table 2 (continued)

No	ID_Quest	KEY-Phrase	Question Generating	BLOOM Level
		software development		
25	20_02	requirements analysis	How to express requirements analysis?	Comprehension
26	30_02	requirements analysis	How to present requirements analysis?	Application
27	40_02	requirements analysis	What do you infer about requirements analysis?	Analysis
28	50_05	requirements analysis	How to effectively is requirements analysis?	Evaluation
29	60_04	requirements analysis	What can invent requirements analysis?	Creation
30	20_06	software creation	What is the significant impact of software creation?	Comprehension

4 Model performance evaluation

4.1 Quality evaluation of the generated questions

This evaluation stage aims to evaluate the quality of the generated questions compared to the manually generated questions (reference questions) by human experts. In this study, this evaluation step involve five human independent reviewers who have competence in the field of study or discussion materials used as questions. Each reviewer independently evaluated a number of generated question samples. The measurement used BLEU score and Cohen's Kappa metrics.

4.2 Evaluating the effect of context on question quality

This evaluation step aims to evaluate the effect of context on the generated question quality using statistical test. The hypotheses to be tested are: (i) significant of mean BLEU test from questions without context, (ii) significant of mean BLEU test from questions with context, and (iii) significant difference of mean BLEU test from questions with context and without context.

In this study, the design of this performance evaluation is implemented as follows:

1) The tested questions are those generated by the system using question template represented by context-free grammar rules and each input text document is represented by N-gram.
2) BLEU score is computed based on system generated question and referenced question constructed manually by each reviewer.
3) The number of independent reviewer is 5. Each reviewer understands the course materials.
4) The number of question samples is 120 chosen randomly from the total of 92,608 generated questions

4.3 Evaluation the predicted Bloom's taxonomy of the generated questions

This test aims to evaluate how good the proposed key-phrase based model is in generating questions that comply with Bloom's taxonomy levels. This task is viewed as a classification problem using input data as a labeled data set $S = \{x_i, y_i\}_1^n$ where $x_i \in X$ is a representation of question, $y_i \in Y$ is Bloom's taxonomy level of a questions, X and Y are domain of question representation and question label respectively, and n is the number of sample data. It is assumed that there is a mapping $F : X \rightarrow Y$.

A classification model based on machine learning G is trained using supervised learning technique to approximate F so that a new question is generated? q then its Bloom's taxonomy level can be predicted as:

$$G(q) = \widehat{F}(q)$$

Performance of the trained classification model is measured using accuracy, precision, and recall metrics.

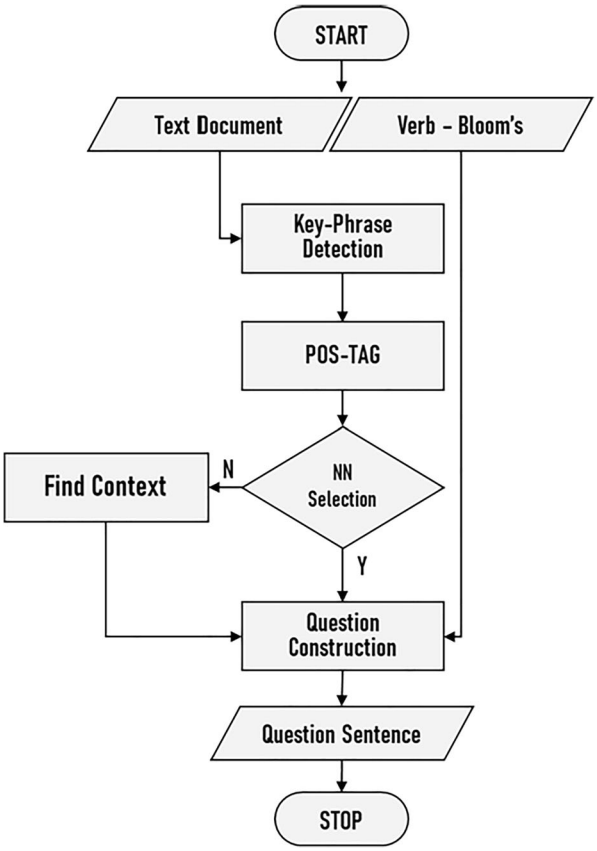


Fig. 4 Question construction process flow

5 Research result and discussion

5.1 Evaluation of the generated question quality

The average BLEUscore obtained from questions The score is 0.921 context seems higher than questions without context 0.861. This shows that the Question Generation Algorithm is able to generate questions using words that resemble the questions compiled by reviewers when making similar questions. (see Table 3).

The Cohen's Kappa score of questions using the 0.627 context seems smaller than the questions without the 0.760 context. However, both of these scores are in the good category, which means that the reviewers have a good agreement in capturing the context of the question.

The measurements of BLEUscore and Cohen's Kappa involve human reviewers who have competence in the field of study or discussion material used as questions.

5.2 Evaluating the effect of context to question quality

The statistical testing evaluation the following results:

Test #1:

$$H_0 : \overline{B}_{woc} = 0, H_1 : \overline{B}_{woc} > 0,$$

With number of samples $(n) = 5$, mean of BLEU score without context $\left(\overline{B}_{woC}\right) = 0.86$, standard deviation $(s_{woC}) = 0.03$,

$Z = \frac{\overline{B}_{woC}}{s_{woC}/\sqrt{n}} = 64.10 > 1.96$ (α= 0.10).

Conclusion: mean of BLEU score without context is not equal to 0.

Test #2:

$$H_0 : \overline{B}_{wc} = 0,$$

$$H_1 : \overline{B}_{wc} > 0,$$

With number of samples $(n) = 5$, mean of BLEU score with context $\left(\overline{B}_{wc}\right) = 0.92$, standard deviation $(s_{wC}) = 0.03$,

$Z = \frac{\overline{B}_{wc}}{s_{wC}/\sqrt{n}} = 68.57 > 1.96$ (α = 0.10).

Conclusion: mean of BLEU score with context is not equal to 0.

Table 3 ComputedScore BLEU dan Cohen's Kappa

(+) CONTEXT							(−) CONTEXT						
N-Gram	N-Gram	BLEU 1	BLEU 2	BLEU 3	BLEU 4	BLEU 5	AVERAGE	BLEU 1	BLEU 2	BLEU 3	BLEU 4	BLEU 5	AVERAGE
1	1	0.9363	0.9209	0.9219	0.9387	0.9699	0.9375	0.9646	0.9543	0.9510	0.9551	0.9882	0.9626
2	2	0.9040	0.8861	0.8825	0.9039	0.9508	0.9055	0.9475	0.9318	0.9275	0.9345	0.9830	0.9449
3	3	0.8757	0.8587	0.8516	0.8772	0.9341	0.8794	0.9346	0.9155	0.9089	0.9190	0.9786	0.9313
4	4	0.8522	0.8366	0.8298	0.8549	0.9190	0.8585	0.9238	0.9022	0.8950	0.9072	0.9755	0.9208
5	5	0.8316	0.8173	0.8125	0.8350	0.9055	0.8404	0.9145	0.8907	0.8832	0.8972	0.9732	0.9118
6	6	0.8130	0.7990	0.7975	0.8165	0.8925	0.8237	0.9057	0.8801	0.8727	0.8883	0.9716	0.9037
7	7	0.7980	0.7848	0.7859	0.8011	0.8824	0.8104	0.8984	0.8705	0.8637	0.8812	0.9715	0.8970
8	8	0.8171	0.8037	0.8094	0.8213	0.9115	0.8326	0.8965	0.8664	0.8607	0.8800	0.9771	0.8961
AVERAGE	AVERAGE	0.8535	0.8384	0.8364	0.8561	0.9207	0.8610	0.9232	0.9014	0.8953	0.9078	0.9773	0.9210
Cohhen Kappa with CONTEXT: 0.62728							Cohen Kappa with-out CONTEXT: 0.76067						

Test #2:

$$H_0 : \overline{B}_{wc} - \overline{B}_{woc} = 0,$$

$$H_1 : \overline{B}_{wc} - \overline{B}_{woc} > 0,$$

With number of samples (n) = 5, mean of BLEU score without context $\left(\overline{B}_{woC}\right) = 0.86$, BLEU score with context $\left(\overline{B}_{wc}\right) = 0.92$, $s_{woC} = s_{wC} = 0.03$,

$Z = \frac{\overline{B}_{wc} - \overline{B}_{woC}}{\sqrt{\left(\frac{s_{wc}{}^2}{n} + \frac{s_{woC}{}^2}{n}\right)}} = 3.16 > 1.96$ (α = 0.10).

Conclusion: mean of BLEU score with context is higher than that without the context.

5.3 Evaluation the predicted Bloom's taxonomy of the generated questions

The following model are trained using 74,881 (80% of the whole generated questions as input dataset) as training dataset and 18,721 samples (20% of the input dataset) as testing dataset. Performances of several classification models are summarized into the following table Table 4.

6 Conclusions and future work

Question generation in the online learning field is a task to generate questions which are relevant to the text input documents and Bloom's Taxonomy of the expected learning output. This task remains a challenging problem.

The main challenges are as follows: (i) there exists a potential gap between the knowledge that question generation system and human experts use to generate questions based on input text, and (ii) the generated questions are expected to promote critical thinking, retention, and context involvement of the students.

This study proposes a question generation model based on key-phrase extracted from text input documents through syntactic parsing using predefined context-free grammar rules. The proposed model was tested using learning materials from Bina

Table 4 Testing performance of classifier model

Model	Acc.	Prec.	Recall	F1
Decision Tree	1.00	1.00	0.99	0.99
Random Forest	0.99	0.99	0.99	0.99
Gradient Boosting	0.99	0.99	0.99	0.99

As can be seen from the above table, the proposed key-phrase model can generate questions with strong relevance to Bloom's taxonomy

Nusantara University's Online Repository that generates 92,608 essay-type questions from 6-level Bloom's Taxonomy.

Performance evaluation using average Bilingual Evaluation Understudy (BLEU) involving five independent reviewers toward samples of these questions achieved 0.921 and 0.6 Cohen's Kappa. The relevance of Bloom's Taxonomy level of the generated questions is evaluated by means of classification model with 0.99 accuracy. These results indicate not only were that the generated questions well understood and agreed by the reviewers they were also relevant to the expected Bloom's Taxonomy level so that the questions can be delivered to students in the respected course delivery and evaluation. Generated questions have been tested in the software engineering course and are proven (1) to assist teachers in guiding asincronous learning, namely online discussions, (2) to assist teachers in providing learning practice questions.

In future work, we plan to investigate a method of extracting answers from corpus text based on machine-generated questions. We also plan to investigate the performance of a questioner model with a variety of Encoder-Decoder to extract answers to questions.

Acknowlegment This research was supported partially by the Directorate General of Research and Development, Ministry of Research, Technology, and Indonesian Higher Education, as part of the Research Doctoral Dissertation Research Grant to Binus University entitled "Generating Questions Automatically in Online Learning Using Encoders-Decoders Using Attention" with Contract number: 049 / LL3 / PG / 2020.

References

Adamson, D., Bhartiya, D., Gujral, B., Kedia, R., Singh, A., & Rosé, C. P. (2013). Automatically generating discussion questions. In *AIED* (pp. 81–90). https://doi.org/10.1007/978-3-642-39112-5_9.

Al-Yahya, M. (2011). OntoQue: A question generation engine for educational assesment based on domain ontologies. In *Proceedings of the 2011 11th IEEE International Conference on Advanced Learning Technologies, ICALT 2011* (pp. 393–395). IEEE. https://doi.org/10.1109/ICALT.2011.124.

Bloom, B.S. (1956). Taxonomy of educational objectives: The classification of educational goals.

Chao, Z., & Li, L. (2018). The combination of context information to enhance simple question answering. In *Proceedings - 2018 5th international conference on behavioral, economic, and socio-cultural computing, BESC 2018* (pp. 109–114). https://doi.org/10.1109/BESC.2018.8697305.

Diab, S., & Sartawi, B. (2017). Classification of questions and learning outcome statements (LOS) into Bloom's taxonomy (BT) by similarity measurements towards extracting of learning outcome from learning material. *International Journal of Managing Information Technology, 9*(2), 01–12. https://doi.org/10.5121/ijmit.2017.9201.

Divate, M., & Salgaonkar, A. (2017). Automatic question generation approaches and evaluation techniques. *Current Science* (00113891), 113(9).

Duan, N., Tang, D., Chen, P., & Zhou, M. (2017). Question generation for question answering. In *EMNLP 2017 - Conference on Empirical Methods in Natural Language Processing, Proceedings* (pp. 877–885). https://doi.org/10.18653/v1/d17-1090.

Elsahar, H., Gravier, C., & Laforest, F. F. (2018). Zero-shot question generation from knowledge graphs for unseen predicates and entity types. *NAACL-HLT*, *abs/1802.0*, 218–228. https://doi.org/10.18653/v1/n18-1020.

Emu, I. H., Ahmed, A. U., Islam, M., Al Mamun, S., & Uddin, A. (2017). An efficient approach for Keyphrase extraction from English document. *International Journal of Intelligent Systems and Applications, 9*(12), 59–66. https://doi.org/10.5815/ijisa.2017.12.06.

Gan, W., & Yu, X. (2018). Automatic understanding and formalization of natural language geometry problems using syntax-semantics models. *ICIC International, 14*(1), 83–98.

Gleason, N. W. (2018). *Higher education in the era of the fourth industrial revolution*. https://doi.org/10.1007/978-981-13-0194-0.

Graesser, A. C., Rus, V., Cai, Z., & Hu, X. (2009). Question answering and generation. *Applied Natural Language Processing: Identification, Investigation and Resolution, 1*, 1–16. https://doi.org/10.4018/978-1-60960-741-8.ch001.

Han, H., Yang, F., Huang, J., & Zhou, B. (2018). CFXGBoost: Topic phrase extraction based on context features and XGBoost for knowledge base question answering. In *ICNC-FSKD 2017 - 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery* (pp. 2397–2402). IEEE. https://doi.org/10.1109/FSKD.2017.8393148.

Heilman, M., & Smith, N. A. (2010). Good question! Statistical ranking for question generation. NAACL HLT 2010 - Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Proceedings of the Main Conference, 609–617.

Heilman, M., Aleven, V., Cohen, W. W., Litman, D. J., & Smith, N. A. (2011). Automatic factual question generation from text. In *Dissertation* (p. 203). Retrieved from www.lti.cs.cmu.edu

Herranen, J., & Aksela, M. (2019). Student-question-based inquiry in science education. *Studies in Science Education, 55*(1), 1–36. https://doi.org/10.1080/03057267.2019.1658059.

Hopcroft, J. E., Motwani, R., & Ullman, J. D. (2001). Introduction to automata theory, languages, and computation, 2nd edition. *ACM SIGACT News, 32*, 60–65. https://doi.org/10.1145/568438.568455.

Huang, Y.-T., Chen, M. C., & Sun, Y. S. (2018). Bringing personalized learning into computer-aided question generation. *CoRR*, *abs/1808.0*. Retrieved from http://arxiv.org/abs/1808.09735

Jouault, C., Seta, K., & Hayashi, Y. (2016). Content-dependent question generation using LOD for history learning in open learning space. *New Generation Computing, 34*(4), 367–394. https://doi.org/10.1007/s00354-016-0404-x.

Kale, V. M., & Kiwelekar, A. W. (2013). An algorithm for question paper template generation in question paper generation system. In *2013 The International Conference on Technological Advances in Electrical, Electronics and Computer Engineering, TAEECE 2013* (pp. 256–261). https://doi.org/10.1109/TAEECE.2013.6557281.

Kumar, V., Ramakrishnan, G., & Li, Y.-F. (2018). A framework for automatic question generation from text using deep reinforcement learning. *CoRR*, *abs/1808.0*. Retrieved from http://arxiv.org/abs/1808.04961

Kurdi, G., Leo, J., Parsia, B., Sattler, U., & Al-Emari, S. (2019). A systematic review of automatic question generation for educational purposes. *International Journal of Artificial Intelligence in Education., 30*, 121–204. https://doi.org/10.1007/s40593-019-00186-y.

Li, J., Gao, Y., Bing, L., King, I., & Lyu, M. R. (2019). Improving question generation with to the point context. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 3214–3224). Stroudsburg, PA, USA: Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1317.

Lindberg, D., & Sc, B. (2013). Automatic question generation from text for self-directed learning by.

Mandal, S., Mahata, S. K., & Das, D. (2018). Preparing Bengali-English Code-Mixed Corpus for Sentiment Analysis of Indian Languages. Retrieved from http://arxiv.org/abs/1803.04000

Meng, R., Zhao, S., Han, S., He, D., Brusilovsky, P., & Chi, Y. (2017). Deep Keyphrase generation. *CoRR*, *abs/1704.0*, 582–592. https://doi.org/10.18653/v1/P17-1054.

Mikolov, T., Chen, K., Corrado, G. S., Dean, J., Sutskever, I., Chen, K., … Dean, J. (2013). Distributed representations of words and phrases and their compositionality.

Moraes, S., Godbole, A., & Gharpure, P. (2018). Affinity analysis for context-free grammars. In *IEEE International Conference on Power, Control, Signals and Instrumentation Engineering, ICPCSI 2017* (pp. 2471–2474). IEEE. https://doi.org/10.1109/ICPCSI.2017.8392161.

Nema, P., Mohankumar, A. K., Khapra, M. M., Srinivasan, B. V., & Ravindran, B. (2019). Let's ask again: Refine network for automatic question generation. Retrieved from http://arxiv.org/abs/1909.05355

Quintana, R. M., & Tan, Y. (2019). Characterizing MOOC pedagogies: Exploring tools and methods for learning designers and researchers. *Online Learning Journal*, *23*(4), 62–84. https://doi.org/10.24059/olj.v23i4.2084.

Rodrigues, H. P. (2017). Learning semantic patterns for question generation and question answering.

Rus, V., Graesser, A., & Cai, Z. (2008). Question generation: Example of a multi-year evaluation campaign. *Workshop on the Question Generation Shared Task and Evaluation Challenge*, (January). Retrieved from http://www.cs.memphis.edu/~vrus/questiongeneration/5-RusEtAl-QG08.pdf

Singh, S., Tiwari, S., Varshney, A., & Sharma, A. (2017). Unsupervised key-phrase extraction using noun phrases. *International Journal of Computer Applications, 162*(1), 1–5. https://doi.org/10.5120/ijca2017913171.

Subramanian, S., Wang, T., Yuan, X., Zhang, S., Trischler, A., & Bengio, Y. (2018). Neural models for key phrase extraction and question generation. In *QA@ACL* (pp. 78–88). Association for Computational Linguistics. https://doi.org/10.18653/v1/w18-2609

Tong, P., Zhang, Q., & Yao, J. (2019). Leveraging domain context for question answering over knowledge graph. *Data Science and Engineering, 4*(4), 323–335. https://doi.org/10.1007/s41019-019-00109-w.

Turney, P. D. (2003). Coherent Keyphrase extraction via web mining. In *IJCAI International Joint Conference on Artificial Intelligence.*

Willis, A., Davis, G., Ruan, S., Manoharan, L., Landay, J., & Brunskill, E. (2019). Key phrase extraction for generating educational question-answer pairs. In *Proceedings of the 6th* 2019 ACM conference on learning at scale*,* L@S *2019*. https://doi.org/10.1145/3330430.3333636.

Xie, Z., Zeng, Z., Zhou, G., & Wang, W. (2017). Topic enhanced deep structured semantic models for knowledge base question answering. *SCIENCE CHINA Information Sciences, 60*(11). https://doi.org/10.1007/s11432-017-9136-x.

Yang, Z., Hu, J., Salakhutdinov, R., & Cohen, W. W. (2017). Semi-supervised QA with generative domain-adaptive nets. In *ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*. https://doi.org/10.18653/v1/P17-1096.

Yao, X., Bouma, G., Zhang, Y., Piwek, P., & Boyer, K. E. (2012). Semantic-based question generation and implementation. *Dialogue & Discourse*, *3*(2), 11–42. https://doi.org/10.5087/dad.2012.202.

Ye, H., & Wang, L. (2018). Semi-supervised learning for Neural Keyphrase generation. Retrieved from http://arxiv.org/abs/1808.06773

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Bambang Dwi Wijanarko [1,2] · Yaya Heryadi [1] · Hapnes Toba [3] · Widodo Budiharto [1]

[1] Computer Science Department, BINUS Graduate Program-Doctor of Computer Science, Bina Nusantara University, Jakarta, Indonesia

[2] Computer Science Department, BINUS Online Learning, Bina Nusantara University, Jakarta, Indonesia

[3] Faculty of Information Technology, Maranatha Christian University, Bandung, Indonesia