Natural Language Processing for Mapping Exam Questions to the Cognitive Process Dimension

https://doi.org/10.3991/ijet.v17i13.29095

Hindayati Mustafidah(✉), S. Suwarsito, Tito Pinandita
Universitas Muhammadiyah Purwokerto, Jawa Tengah, Indonesia
h.mustafidah@ump.ac.id

Abstract—Exam questions as a test instrument to measure educational success must have good quality. This quality can be measured by the level of cognition expected of students. The level of cognition reflects the mastery of learning materials as a form of evaluation of the teaching and learning process outlined in the curriculum. For this reason, the exam questions need to be mapped into the cognitive process dimensions, namely based on the categories in the Revised Bloom's Taxonomy (RBT). The mapping method used is Natural Language Processing (NLP) as one of the fields of Artificial Intelligence development. The stages in this mapping are pre-processing, including tokenization, stemming, stop-word removal, and feature extraction using POS Tagging. The output of this mapping process is in the form of categories of test items into RBT: remembering (C1), understanding (C2), applying (C3), analyzing (C4), evaluating (C5), and creating (C6). The classification results obtained information that the exam questions prepared were still dominated at the C2 cognitive level, which was indicated by the use of operational verbs in the understanding category. The results of testing the method used produce an accuracy of 82.22%. Thus, the NLP method can classify test items into Revised Bloom's Taxonomy to determine the dimensions of students' cognitive processes.

Keywords—Natural Language Processing, exam questions, cognitive process dimension, Revised Bloom's Taxonomy

1 Introduction

Education is a system, while learning is a form of educational activity. As a system, education is a continuous activity. It involves many components, namely: a) raw input (students), b) instrumental input (educators, objectives, materials/programs/curriculum, methods, infrastructure, and facilities), c) environmental input (situation and conditions of the educational environment, social, economic, cultural, security conditions), d) process (implementation of education), and e) product (graduates) [1]. The availability of educational components determines the quality of graduates both in quantity and quality and the functioning of each element according to its role in the implementation of education as a system.

An effective and efficient educational process will help assess the process and assessment of learning outcomes; on the other hand, a good and correct, valid, and reliable learning assessment will reveal the level of achievement and actual learning conditions. The relationship between the elements of the educational process was proposed by [2] which became the basis for other general models. The interpretation of Tyler's model, in practice it is implemented into defining learning goals, organizing learning activities, and evaluating the learning process implemented.

Educational objectives are formulated to describe student behavior reflected through instructional objectives in the perspective of the course. The instructional goals are elaborating educational plans [3]. The formulation of the proper educational purposes will determine a good assessment. According to [4], a good assessment must describe the expected learning outcomes. Assessment as a pedagogic ability [5] must be carried out appropriately to get good feedback on the learning carried out. A good assessment will obtain good feedback on the teaching carried out, which means that the quality of education is also good.

The test is given to students as an assessment instrument in the form of items. Ref [6] conveyed the stages of item development, namely: production, preparation, administration, reporting, documentation, and evaluation. Besides, ref [7] said that assessment activities through written tests are traditional and common methods practiced in most educational institutions today. Hence, questions must be given under the content of the courses studied to meet learning objectives. Writing questions is a very challenging step for questions makers. The situation is daring the teacher or items maker to compile quality items that can be used to know the various cognitive levels. Thus, Bloom's Taxonomy or its revision named Revised Bloom's Taxonomy (RBT), has become general references to guide for preparing test items for the teaching and learning process [8].

Interpretation of test results is conducted by making standard terms indicators of what will be measured. These indicators must be derived from the formulated categories and subcategories, classifying educational objectives and the education taxonomy classification. A goal formulation contains a verb which is a learning objective. This verb describes the expected cognitive process, which includes six categories, namely remembering (C1), understanding (C2), applying (C3), analyzing (C4), evaluating (C5), and creating (C6) which previously consisted of knowledge, comprehension, application, analysis, synthesis, and evaluation [9]. The six categories are known as Revised Bloom Taxonomy, as shown in Figure 1.

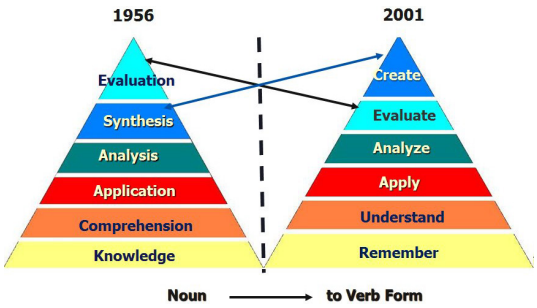

Fig. 1. Bloom's Taxonomy and Revised Bloom's Taxonomy diagram [10]

The cognitive processes dimension in education refers to the RBT for classifying instructional objectives, as shown in Figure 1. There are several other terms for instructional purposes, namely learning objectives, performance goals, or learning objectives. The taxonomy of instructional objectives is a hierarchy that starts from the lowest instructional purposes to the highest level. In other words, goals at a higher level cannot be achieved until the goals at a lower level are achieved.

In the current situation of the COVID-19 pandemic, almost all work is carried out using computer technology [11]. Along with this pandemic period, we are currently in the era of the Industrial Revolution (RI) 4.0 or the Digital Revolution, where the use of computer technology is unavoidable [12]. In this era, large-scale changes in various fields profoundly impact real life [13], not to be separated from the world of education [14]. Applying new methods in the teaching and learning process by integrating information technology is a challenge for an educator/teacher to innovate in the teaching and learning process [15]. The impact of AI developments is also strongly felt in the learning and teaching process [16] and [17], including in conducting assessments in the context of learning evaluation [18] and [19].

Based on the description that has been presented, the problem identified is the relationship between the prepared test questions as a feedback instrument on the achievement of learning objectives and the development of AI technology. The test is carried out to the learning evaluation process through test techniques. The exam questions that are prepared must be able to measure the level of student cognition on the mastery of the lecture material presented. A mapping of exam questions into the dimensions of cognitive processes is needed to determine this level of cognition. Natural Language Processing (NLP) is one of the developments of AI that has proven its impact in various fields of life [20]. NLP can help solve this problem because it can parse a sentence into its parts.

Natural language processing is a research field in computer science, especially AI which focuses on natural or human language. This process translates natural language into data used by computers to learn how to understand the language as outlined in the text [21]. Simultaneously, text processing is a knowledge extraction process in text data [22]. Ref [23] called it a process that typically contains elements of information retrieval, which is one of the methods of text acquisition.

Several studies on computational classifying tests into Bloom's Taxonomy have been carried out. The Inference Engine method was carried out by [24], the artificial neural network method by [25], and the Support Vector Machine (SVM) by [26]. In addition, the use of the rules method with POS tagging was carried out by [27] and in the same year [28] using weighted rules. The SVM method is also carried out by [29] and combines it with Naïve Bayes (NB) and K-Nearest Neighbor (K-NN). Meanwhile, NLP with lexical and syntactic extraction was carried out by [30] for elementary school level questions. Therefore, it is necessary to follow up research on mapping or identify exam questions based on the dimensions of cognitive processes at the tertiary level. Mapping questions into cognitive dimensions are necessary because, based on research [31], there is a significant positive correlation between teaching and cognitive attendance.

2 Method

Items as elements of questions that contain students' thinking abilities after the learning process have a role in the educational process because they determine students' cognitive quality in mastering the lecture material given. Good items must have a level of cognition under the learning design as outlined in the curriculum. We must consider the expected level of student cognition in preparing the items. Before writing the item text, we must do an initial step, namely compiling test specifications in which the test objectives are listed, to measure the level of cognition that can be achieved. However, most of the question makers (lecturers) rarely do it. Based on the performance of NLP, which can extract and map language in the form of text into other formats, this research will map the test questions to the dimensions of cognitive processes, in this case, the category in the Revised Bloom Taxonomy (RBT). This framework is shown in Figure 2. The final result achieved from this process is the type of questions into the cognitive level, namely remembering (C1), understanding (C2), applying (C3), analyzing (C4), evaluating (C5), or creating (C6).

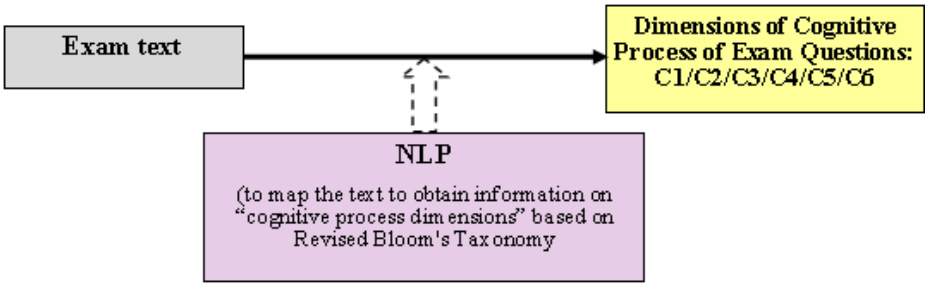

Fig. 2. The framework of the research carried out

2.1 Data acquisition

This research is limited to the written exam questions used as research objects, namely the text of exam questions for various subjects held at the Faculty of Engineering and Science and the Faculty of Teaching and Education. This study needs text data about the end of the 2020-2021 academic year exam. Samples were taken from the questions tested in the study program from both faculties of Universitas Muhammadiyah Purwokerto.

2.2 Research stage

The method used in this study is simpler than the methods that have been used previously, namely applying the basic NLP concept to extract text to find the verb and then looking for the equivalent in the operational verb table (KKO - in Bahasa) [8] and [9]. Based on the framework presented, an outline of the general description of the steps taken in this research is illustrated in Figure 3.

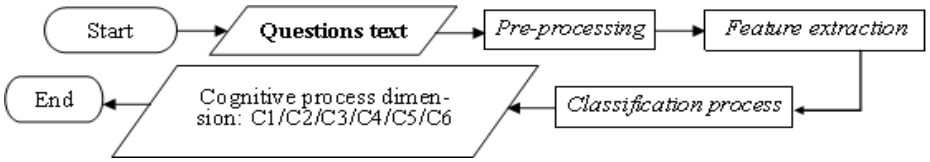

Fig. 3. Research step

The first stage of the model, as in Figure 3, is pre-processing the identification of the test question text under the framework of learning objectives in the RBT. At this stage, we do tokenization, case folding, stemming, and stop-word removal. The tokenization stage splits the input string based on each word that composes it, followed by case folding by reducing all letters to lower case, stemming, to transform the terms contained in a text into its root words, and stop-word removal, which removes non-descriptive words, for example, "*yang* (which)," "*dan* (and)," "*di* (at), "*dari* (from)," "*ke* (to)," and others. This stop-word removal aims to reduce index size and processing time. Furthermore, the feature extraction process is carried out using Part-of-speech (POS) tagging, which assigns POS markers or syntactic classes to each word in the corpus. The verbs resulting from the pre-processing are then matched with the list of operational verbs included in the KKO table. This KKO table contains a list of operational verbs used in RBT which is divided into six cognitive levels.

2.3 Testing design

This NLP-based model was tested for its accuracy in generating information on the level of student cognition based on the RBT using eq. (1). This test compares the number of items that the model can map to the total items [30].

$$AM = \frac{\sum_{i=1}^{n} B_i}{N} x100\% \quad (1)$$

with AM is model accuracy level; B_i is the i-th item that was successfully analyzed by the model and matched the expert's judgment with i = 1, 2, .., n; and N is total items mapped. We used three experts, namely lecturers from three disciplines: exact sciences, social sciences, and education.

3 Results and discussion

3.1 Dataset

The research data is exam questions written in Bahasa from two faculties, namely the Faculty of Engineering and Science and the Faculty of Teaching and Education of Universitas Muhammadiyah Purwokerto, Indonesia. This data is considered sufficiently representative from exact sciences, social sciences, and education. The number of datasets used is 90 items representing 30 items from the three fields of science. Examples of items used in this study are shown in Table 1.

Table 1. List of sample questions in the dataset

No.	Items text
1.	*Jelaskan apa yang dilakukan seorang perekayasa PL setelah kebutuhan-kebutuhan user diakomodasi?* (Explain what a software engineer does after the user's needs are accommodated?)
2.	*Jelaskan perbedaan white box testing dan black box testing!* (Explain the difference between white box testing and black box testing!)
3.	*Informasi apa saja yang dapat diperoleh dari peta RBI? Jelaskan dan berikan contohnya!* (What information can be obtained from the RBI map? Explain and give an example!)
4.	*Apa perbedaan informasi pada peta umum dan peta tematik? Jelaskan dan berikan contohnya!* (What is the difference between information on general maps and thematic maps? Explain and give an example!)
5.	*Jelaskan apa saja komponen pariwisata dan berikan contohnya!* (Explain what the components of tourism are and give an example!)
6.	*Destinasi wisata juga sebaiknya dikembangkan di wilayah saudara tinggal. Bagaimana potensi obyek wisata tersebut? Jelaskan!* (Tourist destinations should also be developed in the area where you live. What is the potential of these tourism objects? Please explain!)
7.	*Apabila Hari Pendidikan Nasional pada tanggal 2 Mei adalah Selasa, HUT Kemerdekaan RI tanggal 17 Agustus pada tahun yang sama adalah …* (If National Education Day on May 2 is Tuesday, the Independence Day of the Republic of Indonesia on August 17 of the same year is…)
8.	*Apa perbedaan yang mencolok antara struktur data stack dan queue? Jelaskan dengan singkat!!* (What is the main difference between stack and queue in data structures? Explain briefly!)
9.	*Apakah ada perbedaan kecepatan proses dari ketiga jenis prosesor yang ada dengan tingkat alpha = 0,025? Beri penjelasan!* (Is there a difference in the processing speed of the three types of processors with alpha = 0.025 level? Give an explanation!)
10.	*Karena data harus dikelompokkan menjadi 2 kelompok yaitu kelompok atas dan kelompok bawah, maka urutkan data tersebut dari tinggi ke rendah (sortir berdasarkan skor total)!* (Because the data must be grouped into 2 groups, namely the upper group and the lower group, then sort the data from high to low (sort by total score)!)
…	
90.	*Buat pohon pencarian berdasarkan graph berikut ini, dimulai dari S dan berakhir di G. Angka di samping node menunjukkan panjang lintasan dari node tersebut ke goal state (G)!* (Create a search tree based on the following graph, starting at S and ending at G!. The number next to the node indicates the path length from that node to the goal state (G).)

3.2 Data processing

The NLP is a natural language identification process so that humans can communicate with computers using human language [32]. The basic process of natural language identification includes tokenizing to set the input string as a word order [33], case folding, stemming that cleans or removes affixes in a word, and stop-word removal to eliminate the unused terms. Stemming in Bahasa (Indonesian language) is better based on a dictionary so that fewer errors are generated than based on rules [32]. The identification process is according to the type using POS Tagging.

The item text data was extracted to find verbs as the basis for classification into the RBT. Therefore, the first step to take is tokenization. The POS Tagging method is used

to obtain verb information. Its results are then calculated using the Bag of Word (BoW) technique to know how many verbs are in each item. Furthermore, a dictionary was built by first doing stemming fromgetting the meaning of the verbs contained in the item text in the operational verb framework (KKO – in Bahasa) in the RBT.

Preprocessing is the initial step in classification that aims to interpret a sentence into a feature vector by converting into words [34]. This step dramatically affects the classification process. The deletion of stop-words is carried out after the stemming process to avoid capitalization/lowercase writing errors. The pre-processing stage is intended to shape the text of the questions into a structured form so that the mapping process to the cognitive dimensions in the RBT can be carried out. This RBT includes six classification categories, namely: remembering (C1), understanding (C2), applying (C3), analyzing (C4), evaluating (C5), and creating (C6). Pre-processing that has been done through the stages tokenization, case folding, stemming, stop-word removal, and POS tagging.

Tokenization is the process of parsing sentences, in this case, the text of the questions, into smaller parts, namely in the form of words. The tokenization stage cuts the input string based on each word that composes it. Before the input string is cut, it is necessary to do case folding to convert the text into a standard form, namely lowercase or lowercase. Tokenization aims to avoid writing errors in letters in the item text. The tokenization stage breaks down strings or sentences in a text item into words, distinguishing certain characters that can be treated as word separators or not. For example, whitespace characters, such as enter, tabulation, space (as word separators) or single quotation characters (‘), period (.), semicolon (;), colon (:) or others that function as non-word separators in general. In this study, the item text as a model input case must have met the correct grammar. Case folding is carried out to avoid errors in writing upper/lower case letters. Case folding is done to change the writing of words into lowercase entirely. It is intended to prevent being case-sensitive to find word similarity.

The stemming process is carried out to produce essential words. This process is used to anticipate ambiguity between nouns and verbs (for example, the basic word "sequence" if it gets an affix can become a sequence, namely a noun, or sort which is a verb). The next step is stop-word removal, which remove unused words. This stage aims to reduce index size and processing time. Words are the syntax in a sentence. Words in Bahasa are not consistently recognized directly because sometimes these words have affixes. To find out the labeling, the affix removal process was carried out. Affixes are additional things attached to words and give rise to new meanings [32]. Affixes are divided into three types: prefixes, suffixes, and confixes.

POS Tagging is the stage of identifying the sentence structure in the item text to determine the verb used. These verbs are the basis for mapping the dimensions of cognitive processes in the RBT. This tagging uses CRFTagger based on Fam-Rashel [35] and [36]. This process is the main key in recognizing the type of word, whether a word is included in the type of noun, verb, adjective, etc.

3.3 Classification process

The first thing to do is feature extraction in the classification process. A feature is a distinguished thing used to classify a question. There are three types features to organize questions, namely lexical, syntactic, and semantic features [37]. This study used only two kinds of feature extraction, namely lexical and syntactic extraction, as shown in Table 2.

Table 2. Types of feature extraction used

Feature	Description	Reason
Lexical	Types of words in the question	as one of the characteristics that can be used to classify questions [37]
	Question length	
Syntactic	Number of verbs	characteristics that can be used to identify the level of Bloom's Taxonomy [27]
	Keywords	keywords can interpret a topic [38]

Lexical features extracted the questions based on the context of the words [39]. One of the text extraction or feature extraction methods used to extract text features based on their appearance in sentences is the Bag of Word (BoW). This process determines how many verbs there are in the item text to indicate the level of cognition needed in working on the questions. As an example of the following computer subject questions:

"Jelaskan apa yang dilakukan seorang perekayasa PL setelah kebutuhan-kebutuhan user diakomodasi?"

After tokenization and case folding is performed, the following results are obtained:

[*'jelaskan', 'apa', 'yang', 'dilakukan', 'seorang', 'perekayasa', 'pl', 'setelah', 'kebutuhan-kebutuhan', 'user', 'diakomodasi', '?'*]

The length of the question or the number of words is 11.

Syntactic features are extracted from the syntax questions. As in the previous example of math problems, the results of extracting the syntactic features from the sample questions using POS Tagging are as follows:

[[('*jelaskan*', 'VB'), ('*apa*', 'WH'), ('*yang*', 'SC'), ('*dilakukan*', 'VB'), ('*seorang*', 'NND'), ('*perekayasa*', 'NN'), ('*pl*', 'NNP'), ('*setelah*', 'SC'), (*'kebutuhan-kebutuhan*', 'NN'), ('*user*', 'NN'), ('*diakomodasi*', 'NN'), ('?', 'Z')]]

The item text contains two verbs.

Defining keywords is essential for the success of item classification. The problem classification process can be carried out if the keywords represent the identified taxonomic level. This study uses keywords that exist [40]. However, some keywords are less specific because they are also at another level of Bloom's taxonomy. If there is one keyword or verb in the question text at a different level in the RBT, it is classified to a higher level [41].

Based on the classification process carried out on the dataset and concerning the operational verb dictionary [8], the results are shown in Figure 4. The level of cognition of C2 occupies the highest percentage, while C6 is the lowest. The highest rate was found in the items tested in social science and education because they used operational

verbs included in the "understanding" category. On the other hand, the C6 class is found in exact science items where students are asked to build a structure or program.

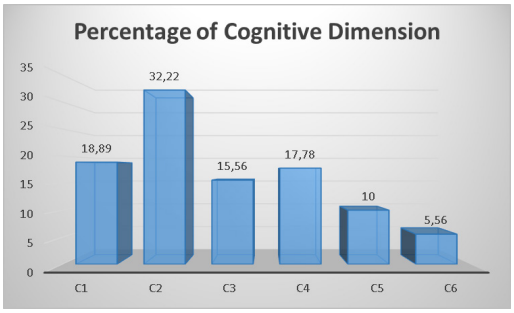

Fig. 4. Results of item feature extraction

This model has been tested by comparing the results of the classification of questions from the model with the NLP method used in this study with expert classification. The testing produces accuracy that can be used as a successfully study benchmark of this method. Based on the test result as in eq. (1), the result obtained from the accuracy of the technique used is:

$$AM = \frac{74}{90} x100\% = 82.22\%$$

This accuracy may be different because the mapping of items into level's taxonomic classification carried out by experts is subjective [42], even though there is already a classification standard or benchmark. This accuracy is relatively higher than [24] at 51%, [25] at 65.9%, and [27], which reached 77% who also used the POS tagging method, even [29] with the K-NN method which reached 81.6%. However, the accuracy achieved in this study is still below the performance of using the SVM method as done by [30], which is 88.6%, [26] at 87.4%, and [29], which achieved 86%. Ref [29] also used the NB method and produces an accuracy of 85%.

4 Conclusion

The testing results show that the model based on the NLP method can classify questions based on the dimensions of cognitive processes in Revised Bloom's Taxonomy. The use of lexical and syntactic features succeeded in classifying questions with accuracy results above 80%. The success of items classification in various fields of science proves that lexical and syntactic features can be used. However, this accuracy may be improved using different methods, such as Naïve Bayes (NB), K-Nearest Neighbor (KNN), or other methods.

5 Acknowledgment

The authors would like to thank the Universitas Muhammadiyah Purwokerto through the Research and Community Service Board, which has funded this research based on contract number A.11-III/700-S.Pj./LPPM/XII/2020.

6 References

[1] A. M. Yusuf, *Asesmen dan Evaluasi Pendidikan: Pilar Penyedia Informasi dan Kegiatan Pengendalian Mutu Pendidikan*. Jakarta: Prenadamedia Group, 2015.

[2] R. W. Tyler, *Basic principles of curriculum and instruction*. Chicago: University of Chicago Press, 2013. https://doi.org/10.7208/chicago/9780226086644.001.0001

[3] Wahyudin, "Perumusan Tujuan Instruksional," 2015. http://file.upi.edu/Direktori/FIP/JUR._PEND._LUAR_SEKOLAH/196009261985031-UYU_WAHYUDIN/Perumusan_tujuan_instruksional.pdf (accessed Apr. 04, 2015).

[4] C. R. Gareis and L. W. Grant, *Teacher-Made Assessments: How to Connect Curriculum, Instruction, and Student Learning*, Second Edi. New York: Taylor & Francis, 2015. https://doi.org/10.4324/9781315773414

[5] N. H. Tien, D. T. Anh, M. Van Luong, N. M. Ngoc, and N. Thanh, "Formative Assessment in the Teacher Education in Vietnam," *J. Hunan Univ. Nat. Sci.*, vol. 47, no. 8, pp. 1–10, 2020.

[6] S. Lane, M. R. Raymond, and T. M. Haladyna, *Handbook of Test Development*, 2nd ed. New York: Routledge Taylor & Francis Group, 2016.

[7] D. A. Abduljabbar and N. Omar, "Exam questions classification based on Bloom's Taxonomy cognitive level using classifiers combination," *J. Theor. Appl. Inf. Technol.*, vol. 78, no. 3, pp. 447–455, 2015.

[8] L. W. Anderson and D. R. Krathwohl, *Platform Framework for Learning, Teaching and Assessment (Revised Bloom's Educational Taxonomy) (Kerangka Landasan untuk Pembelajaran, Pengajaran, dan Asesmen (Revisi Taksonomi Pendidikan Bloom))*, I, diterje. Yogyakarta: Pustaka Pelajar, 2010.

[9] L. W. Anderson and D. R. Krathwohl, *A Taxonomy for Learning, Teaching, and Assessing : A revision of Bloom's Taxonomy of Education Objectives*. United States (US): Addison Wesley Longman, Inc., 2001.

[10] Kemendikbud, *MATERI PELATIHAN IMPLEMENTASI KURIKULUM 2013*. Jakarta: Badan Pengembangan Sumber Daya Manusia Pendidikan dan Kebudayaan dan Penjaminan Mutu Pendidikan, 2014.

[11] Kompasiana, "Peran Teknologi Dalam Pandemi Covid-19," *https://www.kompasiana.com/waju27020/5ea90c97d541df0aac6be9c2/peran-teknologi-dalam-pandemi-covid-19*, 2020 (accessed May 30, 2020).

[12] M. E. Auer, D. Guralnick, and I. Simonics, Teaching and Learning in a Digital World: Proceedings of the 20th International Conference on Interactive Collaborative Learning – Volume 2. Switzerland: Springer International Publishing, 2018. https://doi.org/10.1007/978-3-319-73210-7

[13] A. D. Syafaati, "Revolusi Industri dari Generasi 1.0 hingga 4.0," *https://www.academia.edu/37491240/REVOLUSI_INDUSTRI_DARI_GENERASI_1.0_HINGGA_4.0*, 2019 (accessed Feb. 19, 2019).

[14] M. E. Auer and R. B. Kalyan, *Cyber-physical Systems and Digital Twins: Proceedings of the 16th International Conference on Remote Engineering and Virtual Instrumentation.* Switzerland: Springer International Publishing, 2020. https://doi.org/10.1007/978-3-030-23162-0

[15] E. Sugandi, "Blended Learning Sebagai Solusi Pembelajaran Matematika Bagi Generasi Digital (Blended Learning as a Mathematics Learning Solution for the Digital Generation)," in *National Conference on Mathematics, Science and Education (NACOMSE)*, 2018, vol. 1, no. 01, pp. 225–234.

[16] M. E. Auer, H. Hortsch, and P. Sethakul, *The Impact of the 4th Industrial Revolution on Engineering Education: Proceedings of the 22nd International Conference on Interactive Collaborative Learning (ICL2019) – Volume 2.* Switzerland: Springer International Publishing, 2020. https://doi.org/10.1007/978-3-030-40271-6

[17] Y. Wang and G. Zheng, "Application of Artificial Intelligence in College Dance Teaching and Its Performance Analysis Problems in Dance Teaching Based on AI Application," *Int. J. Emerg. Technol. Learn.*, vol. 15, no. 16, pp. 178–190, 2020. https://doi.org/10.3991/ijet.v15i16.15939

[18] M. E. Auer and T. Tsiatsos, *The Challenges of the Digital Transformation in Education: Proceedings of the 21st International Conference on Interactive Collaborative Learning (ICL2018) - Volume 1.* Switzerland: Springer International Publishing, 2020. https://doi.org/10.1007/978-3-030-11932-4

[19] M. Li and Y. Su, "Evaluation of online teaching quality of basic education based on artificial intelligence," *Int. J. Emerg. Technol. Learn.*, vol. 15, no. 16, pp. 147–161, 2020. https://doi.org/10.3991/ijet.v15i16.15937

[20] T. Munakata, *Fundamentals of the New Artificial Intelligence : Neural, Evolutionary, Fuzzy and More*, Second. London: Springer-Verlag, 2008.

[21] H. Lane, C. Howard, and H. M. Hapke, *Natural Language Processing in Action: Understanding, analyzing, and generating text with Python*. Shelter Island, NY: Manning Publications Co., 2019.

[22] T. Jo, *Text Mining: Concepts, Implementation, and Big Data Challenge*. Switzerland: Springer International Publishing AG, 2019.

[23] G. Ignatow and R. Mihalcea, *An Introduction to Text Mining*. California: SAGE Publications, Inc., 2018.

[24] Y. Chang and H. Chen, "An Automatic Inference System for the Quality Analysis of Test Items Based on the Bloom's Revised Taxonomy," in *Proceeding of International Conference on Machine Learning and Cybernetics*, 2009, vol. 5, no. July, pp. 2852–2856. https://doi.org/10.1109/ICMLC.2009.5212592

[25] N. Yusof and C. J. Hui, "Determination of Bloom's cognitive level of question items using artificial neural network," in *Proceeding of 10th International Conference on Intelligent Systems Design and Applications*, 2010, pp. 866–870. https://doi.org/10.1109/ISDA.2010.5687152

[26] A. A. Yahya and A. Osman, "Automatic classification of questions into Bloom's cognitive levels using support vector machines," *https://www.semanticscholar.org/paper/Automatic-Classification-of-Questions-into-Bloom%27s-Yahya-Osama/1e161bbff7c43710a9061efc9b5dc57bf83c5607?p2df*, 2011. (accessed Nov. 21, 2019).

[27] S. S. Haris and N. Omar, "A rule-based approach in Bloom's Taxonomy question classification through natural language processing," in *7th International Conference on Computing and Convergence Technology (ICCCT)*, 2012, pp. 410–414.

[28] N. Omar *et al.*, "Automated analysis of exam questions according to Bloom's taxonomy," *Procedia-Social Behav. Sci.*, vol. 59, pp. 297–303, 2012. https://doi.org/10.1016/j.sbspro.2012.09.278

[29] M. Mohammed and N. Omar, "Question classification based on Bloom's Taxonomy using enhanced TF-IDF," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 8, no. 4–2, pp. 1679–1685, 2018. https://doi.org/10.18517/ijaseit.8.4-2.6835

[30] S. F. Kusuma, D. Siahaan, and U. L. Yuhana, "Automatic Indonesia's questions classification based on bloom's taxonomy using Natural Language Processing a preliminary study," in *Proceeding of International Conference on Information Technology Systems and Innovation, ICITSI 2015 - Proceedings*, 2015, pp. 1–6. https://doi.org/10.1109/ICITSI.2015.7437696

[31] Z. Sidiropoulou and I. Mavroidis, "The Relation Between the Three Dimensions of the Community of Inquiry and the Learning Styles of Students in a Distance Education Programme.," *iJET*, vol. 14, no. 23, pp. 180–192, 2019. https://doi.org/10.3991/ijet.v14i23.11564

[32] R. A. Sukamto and D. H. Widyantoro, "Indonesian Parsing using Collins's Parser," Bandung, 2009.

[33] B. V. Indriyono, E. Utami, and A. Sunyoto, "Pemanfaatan Algoritma Porter Stemmer Untuk Bahasa Indonesia Dalam Proses Klasifikasi Jenis Buku," *J. Buana Inform.*, vol. 6, no. 4, pp. 301–310, 2015. https://doi.org/10.24002/jbi.v6i4.462

[34] V. Srividhya and R. Anitha, "Evaluating preprocessing techniques in text categorization," *Int. J. Comput. Sci. Appl.*, vol. 47, no. 11, pp. 49–51, 2010.

[35] A. Dinakaramani, F. Rashel, A. Luthfi, and R. Manurung, "Designing an Indonesian part of speech tagset and manually tagged Indonesian corpus," in *2014 International Conference on Asian Language Processing (IALP)*, Oct. 2014, pp. 66–69. https://doi.org/10.1109/IALP.2014.6973519

[36] F. Rashel, A. Luthfi, A. Dinakaramani, and R. Manurung, "Building an Indonesian rule-based part-of-speech tagger," in *International Conference on Asian Language Processing, IALP, October 20-22, 2014*, 2014, pp. 70–73. https://doi.org/10.1109/IALP.2014.6973521

[37] B. Loni, "Enhanced question classification with optimal combination of features," Delft University of Technology, 2011.

[38] N. Kang, C. Domeniconi, and D. Barbará, "Categorization and keyword identification of unlabeled documents," in *Fifth IEEE International Conference on Data Mining (ICDM'05)*, 2005, pp. 4-pp.

[39] S. Banerjee and S. Bandyopadhyay, "Bengali question classification: Towards developing qa system," in *Proceedings of the 3rd Workshop on South and Southeast Asian Natural Language Processing*, 2012, pp. 25–40.

[40] R. Utari, "Taksonomi Bloom: Apa dan Bagaimana M enggunakannya?," 2011.

[41] S. K. Patil and M. M. Shreyas, "A Comparative Study of Question Bank Classification based on Revised Bloom's Taxonomy using SVM and K-NN," in *2017 2nd International Conference On Emerging Computation and Information Technologies (ICECIT)*, 2017, pp. 1–7. https://doi.org/10.1109/ICECIT.2017.8453305

[42] O. F. Tutkun, D. Guzel, M. Koroğlu, and H. Ilhan, "Bloom's Revised Taxonomy and Critics on It," *TOJCE Online J. Couns. Educ.*, vol. 1, no. 3, pp. 23–30, 2012.

7 Authors

Hindayati Mustafidah is a lecturer in Informatics Engineering Department of Universitas Muhammadiyah Purwokerto and member of IEEE. She is a Editor in Chief of informatics journal in Indonesia and has research interests in Artificial Intelligence (email: h.mustafidah@ump.ac.id).

S. Suwarsito is an Assoc. Professor at Faculty of Teaching and Education, Universitas Muhammadiyah Purwokerto, Indonesia. Her research interests are oceanography, hydrography, and fisheries (email: suwarsito@ump.ac.id).

Tito Pinandita is a lecturer in Informatics Engineering Department of Universitas Muhammadiyah Purwokerto and has research interests in mobile programming (email: titop04@gmail.com).

Article submitted 2021-12-24. Resubmitted 2022-04-04. Final acceptance 2022-04-08. Final version published as submitted by the authors.