

## Generating Indonesian Question Automatically Based on Bloom's Taxonomy Using Template Based Method

Selvia Ferdiana Kusuma<sup>\*1</sup>, Rinanza Zulmy Alhamri<sup>2</sup>

<sup>1,2</sup>Politeknik Kediri

selvia@poltek-kediri.ac.id<sup>\*1</sup>, rinanza@poltek-kediri.ac.id<sup>2</sup>

### Abstract

*In education field, evaluation is needed to know the extent to which the learning process has been done. The evaluation process can be done through the provision of questions with varying degrees of difficulty. However, making questions with varying degrees of difficulty is not easy. Someone must understand the whole new materials to make the question. If there are a lot of materials, it takes a little time to change them to be a question. Therefore, it is necessary to automate the question generation process, in order to facilitate and accelerate the question generation process. This research introduces a template-based method to generate questions based on New Bloom's Taxonomy. There were 4 stages in this research, dataset collection, pattern identification process, question generating process & classification, and final evaluation process result. The dataset consists of 60 samples of paragraphs that derived from 9 courses of study courses Informatics Engineering. The 60 paragraphs produced 278 sentences and 654 questions. The proposed method is capable of producing an accuracy of 81.65% to generate questions using New Bloom's Taxonomy classification. So it can be concluded that the proposed method can be used to generate questions with varying difficulty levels in accordance with New Bloom's Taxonomy.*

**Keywords:** Question generation, Bloom's Taxonomy, Template Based

### 1. Introduction

In education field, evaluation is needed to know the extent to which the learning process has been done. The evaluation process can be done by answering questions with different difficulty levels [1]. One theory that can be used as a basis for making questions with different difficulty levels is Bloom's Taxonomy [2]. Bloom's Taxonomy classifies students' cognitive abilities into 6 levels, which are remembering, understanding, applying, analysing, evaluating, and creating [3]. Bloom's Taxonomy has keywords that can be used as reference in generating questions with different difficulty levels. However, the question-generation process independently is not easy. People must understand all the materials that have been learned, in order to make some questions from the materials [4]. Surely, this would be difficult for a student. For a teacher the process of generate this question is also not always easy, generating a lot of questions will certainly take a lot of time because they must understand the material and then turn it into various types of questions. In addition, there is sometimes discrepancy between teachers when determining the difficulty level of the question.

Therefore, it is necessary to have a method that can be used to generate questions automatically. The use of question generators automatically also reduces the dependence of an expert in question generation [5]. So far there have been some previous studies related to question generation automatically. Ginanjar & Purnamasari [6] generate automatic questioning for Indonesian Text that contains compound sentences. The method used consists of two stages: pre-process and main process. Pre-process is an early stage method to get the named word and entity type for every word in the sentence, while the main process is a syntactic analysis process consisting of the process of identifying phrases with pattern matching methods, the process of determining grammatical functions by the rule-based method and generating the question process with template based method. The dataset used is 30 input text taken from the School Book of Electronic Kemendikbud RI. Results of question generation accuracy is 71.06%. Accuracy results obtained in this study is less than the maximum because the method proposed in this study has not been able to identify the overall type of compound sentence, so the question generated sometimes not appropriate. Suwarningsih, Supriana, & Purwarianti [7] proposed a template-

---

Kusuma, S., & Alhamri, R. (2018). *Generating Indonesian Question Automatically Based on Bloom's Taxonomy Using Template Based Method*. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 3(2). doi:<http://dx.doi.org/10.22219/kinetik.v3i2.650>

Receive May 25, 2018; Revise July 07, 2018; Accepted August 14, 2018

based approach derived from a combination of Semantic Role Labelling (SRL) and Predicate Argument (PA) to generate question generation. Question generation process are done into these stages: (1) the selection of the content of the sentence category (Problem, Patient, Intervention, Compare, Control, Outcome, Organs (PPPICCOODTQ)), (2) identification of the question transformation, and (3) construction answer questions. This research yields an accuracy score of 80% for simple questions and 71% for complex questions. This research still has many ambiguous rules, so the resulting accuracy is less than the maximum. Although it is classified as having high accuracy, the proposed method can be used only to generate questions on the medical domain. Swali, Palan, & Shah [8] generate questions from a paragraph using a semantic approach and a syntactic approach. This study produces simple and complex questions. Simple questions result from sentences that do not have conjugations, whereas complex questions are produced from a sentence that has conjugations. Each sentence will be selected first whether it can be a question or not. The selection process is done by identifying features owned by every sentence. The features used are the first sentence, the last sentence, the words that often appear, the length of the sentence, the number of nouns, adverbs and hyphens possessed. This research has not been able to make inquiries from sentences that have pronouns. Blstak & Viera [9] conduct question generation based on template based method derived from sentence structure analysis. This study yields a more effective pattern than previous studies because it has fewer patterns but has better results in generating questions. Pattern in this study was made through a combination of POS Tags, Named Entity Relation (NER), and Super Sense Tags (SST). Although it has good accuracy but the use of combining patterns makes the computing process more complex. Basuki & Kusuma [10] proposed a method to generate Open Domain Indonesian Question by using syntactical approach. That research incorporates four stages, namely: the identification of declarative sentence for 8 coarse-class and 19 fine-class sentences, the classification of features for coarse- class sentence and the classification rules for fine-class sentence, the identification of question patterns, and the extraction of sentence's components as well as the rule generation of questions. According to the implementation and experiment, the findings show that the accuracy of coarse-class classification reaches 83.26% by using the SMO classifier and the accuracy of proposed fine-class classification reaches 92%. The generated questions are categorized into three types, namely: TRUE, UNDERSTANDABLE, and FALSE. The accuracy of generated TRUE and UNDERSTANDABLE questions reaches 88.66%.

Based on the literature studies as described previously, it is known that the entire previous research has not been studied about the generation of questions that will generate questions with different difficulty levels according to the difficulty level of Bloom's Taxonomy. Previous research cannot be used to generate questions according to Bloom's Taxonomy because the process of classifying the proposed question or method cannot support the generation of questions that have difficulty levels according to Bloom's Taxonomy.

## **2. Research Method**

There were 4 stages that had been done in this research dataset collection, pattern identification process, generating process & classification of question and final evaluation process result. More detail explanation of these steps were explained in sub chapters 2.1 to 2.4.

### **2.1 Dataset Collection**

The dataset used was a collection of paragraphs of Informatics Engineering courses. Dataset obtained from textbooks and materials which made by lecturers Polytechnic Kediri. The courses used as datasets were Database, Computer Network, Data Mining, Android Programming, Microcontroller Programming, Algorithms and Data Structures, Object Oriented Programming, Web Design, Web Programming. The use of different courses was intended to check whether question generation can automatically be done for various courses in Informatics Engineering. This study used 60 paragraphs of samples from 9 different types.

### **2.2 Pattern Identification**

The process of identifying sentence patterns was needed to be able to determine the rules at the time of question generation as well as the classification of difficulty levels for each question. Each entered paragraph was extracted into sentence units. Each sentence was analysed by an expert to see possible patterns that can be raised. If the sentence was detected and can be made to be a question, it continued to the next step, but if it can be made into a question, the sentence was ignored. The determination of the question or keyword pattern was done by identifying the

argument in each sentence through the identification of the argument. Hence, it was known which part will be chosen as the candidate for being the question. This identification process divided the sentence into 3: arg1, keyword and arg2. Arg1 represented a sentence that existed before the keyword and arg2 represented the sentence after the keyword. An example of the identification result in the phrase " Pengukuran dengan menggunakan em sangat populer karena skalabilitasnya yang menyebabkan kita dapat dengan mudah mengubah pengukuran untuk seluruh elemen yang ada hanya dengan mengubah ukuran teks dasar." ("measurement using 'em' is so popular because of the scalability that makes us can change all the element measurement only by changing the basic text size easily" - trans) Shown in Figure 1.

Pengukuran dengan menggunakan em sangat populer karena skalabilitasnya yang menyebabkan kita dapat dengan mudah mengubah pengukuran untuk seluruh elemen yang ada hanya dengan mengubah ukuran teks dasar	Arg1 Keyword Arg2
---	-------------------------

Figure 1. Sentence Identification Result

After the identification of the questioning pattern completed, then it continued to identify the difficulty level of the question. The identification of the difficulty level of the question was based on the keywords and patterns in the question. The process of making the rule to identify the level of difficulty of the question began with the process of classification of questions that have been collected in accordance with the six levels of difficulty that existed in new bloom's taxonomy by expert. Expert will categorize the question into the remembering level if the answer of the question is a memorable sentence, categorized into the level of understanding if the answer is an explanatory sentence, categorized into the level of applying if the answer is an implementation, categorized into the level of analysing if the answer is an analysis, categorized into the level of evaluating if the answer refers to the evaluation sentence, and then categorized as the creating level if the answer is produce something original. The next process was the identification of keywords and patterns of each of the questions that have been grouped. It was these keywords and patterns that were used as the basis for the classification of questions during the test. After the identification process complete, then the data were stored on the database question pattern. The process of determining this pattern was done by expert in the field of education who have had experience of teaching at least 10 years. An example of the resulting pattern is shown in Table 1.

Table 1. Template Resulted

Id	Level	Keyword 1	Keyword 2	Question word	Location of the question	Question target	Answer target
1	2	Adalah (is/are)	-	Adalah (is/are)	B	Arg1	Arg2
2	1	Disebut (is/are)	-	Disebut (is/are)	B	Arg2	Arg1
17	4	Karena (because)	-	Mengapa (Why)	D	Arg1	Arg2
56	3	Memiliki (has/have)	-	Memiliki (has/have)	B	Arg1	Arg2

The Id represented the id rule that became the identifier of the rule. Level represented the degree of difficulty that existed in the rule, keyword 1 represented the keyword being the identifier of the rule. Keyword 2 represented another word if the keyword consisted of 2 letters. The question word was the word used to clarify the question. The location of the question word represented the place where the question was put "D" which meant to be in front and B meant at the back of the designated arg. The target question was used to determine which argument (arg) the question was raised. While the target answer was used to determine the sentence that was the answer.

### 2.3 Process of Question Generation and Classification

Question generation process was done based on pattern identification on every sentence to be questioned. The use of pattern matching was chosen because it was considered easy to be implemented, had high accuracy and did not require any additional resources or tools. Scheme of generating questions and answers that have been classified according to the degree of difficulty in the new taxonomy of bloom was shown in Figure 2. Paragraphs entered by the user were broken down into per sentence. Each of these sentences was identified. The process of checking the question generation pattern was done through checking the existing keywords in the sentence. Keywords found were then matched against a list of existing patterns. The process of classifying questions was done through keyword identification and pattern identification of each question. Each question that has been raised then identified the level of difficulty based on the keywords and patterns that the question has. After the identification process is completed, then formed questions that have been classified based on the level of difficulty.

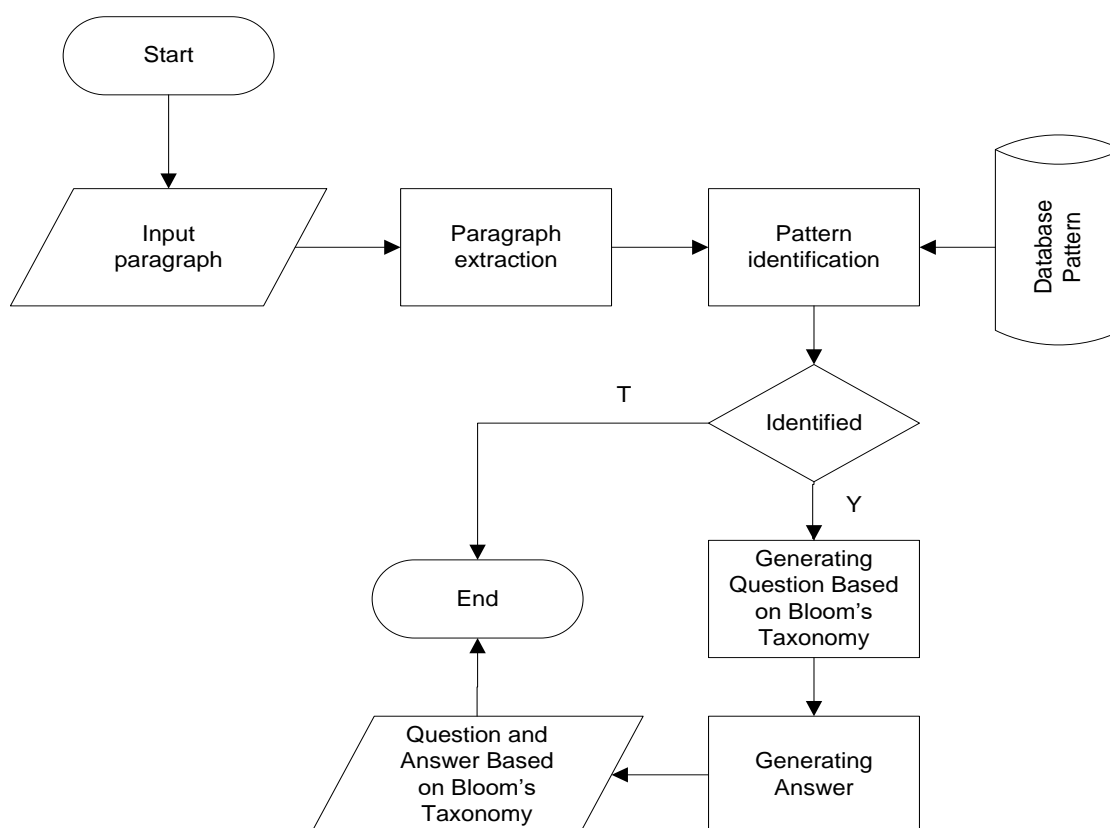


Figure 2. Scheme of Question Generation and Answer Classification Based on Difficulty Levels in Bloom's Taxonomy

The examples of the results of the implementation process were shown in Figure 3, Table 2 and Table 3. Because the second sentence used pronouns that refer to the previous subject, then the sentence was adjusted first so that later the question generated was not ambiguous. Adjustment was to replace the pronoun to be the subject referred. The second sentence changed to " Berbeda dari dua jenis baterai pendahulunya, baterai jenis Li-ion tidak lagi memiliki memory effect dan dapat diisi ulang sebelum baterai belum benar-benar kosong " ("It is different from two types of battery predecessor, Li-ion battery doesn't have memory effect and can be recharged before it is empty"). The subject usually resided before the explanatory keyword like "adalah" ("is/are") or "merupakan" ("is/are"). Therefore, the introduction of the original subject was done through the introduction of the word before the keyword.

Table 2 showed that in one sentence it was possible to contain 3 types of patterns. As in the example, the phrase " Li-ion adalah singkatan dari Lithium-Ion, yang umumnya banyak digunakan berbagai gadget." ("Li-ion stands for Lithium-Ion, which are commonly used for gadget") It can contain 3 types of patterns that were pattern number 1,2 and 72. Table 3 showed

the generating results of questions and answers of identification patterns that have been done in Table 2.

Paragraph Input:

Li-ion adalah singkatan dari Lithium-Ion, yang umumnya banyak digunakan berbagai gadget. Berbeda dari dua jenis baterai pendahulunya, baterai jenis ini tidak lagi memiliki memory effect dan dapat diisi ulang sebelum baterai belum benar-benar kosong. Faktor lain yang dapat memicu menurunnya performa baterai ini salah satunya kelebihan waktu saat charging, atau istilah kerennya overcharge. (*“Li-ion stands for Lithium-Ion, which are commonly used for gadget. It is different from two types of battery predecessor, Li-ion battery doesn't have memory effect and can be recharged before it is empty. Another factor that can trigger decreasing of the battery performance is overcharge”*)

The result of paragraph extraction process:

1. Li-ion adalah singkatan dari Lithium-Ion, yang umumnya banyak digunakan berbagai gadget. (*“Li-ion stands for Lithium-Ion, which are commonly used for gadget”*)
2. Berbeda dari dua jenis baterai pendahulunya, baterai jenis Li-ion tidak lagi memiliki memory effect dan dapat diisi ulang sebelum baterai belum benar-benar kosong. (*“It is different from two types of battery predecessor, Li-ion battery doesn't have memory effect and can be recharged before it is empty”*)
3. Faktor lain yang dapat memicu menurunnya performa baterai ini salah satunya kelebihan waktu saat charging, atau istilah kerennya overcharge. (*“Another factor that can trigger decreasing of the battery performance is overcharge”*)

Figure 3. Process of Sentence Extraction

Table 2. Pattern Identification Result

Id	Sentence	Information
1,2,72	Li-ion adalah singkatan dari Lithium-Ion, yang umumnya banyak digunakan berbagai gadget. ( <i>“Li-ion is stands for Lithium-Ion, which are commonly used for gadget”</i> )	Arg1 Keyword Arg2
56	Berbeda dari dua jenis baterai pendahulunya, baterai jenis Li-ion tidak lagi memiliki memory effect dan dapat diisi ulang sebelum baterai belum benar-benar kosong ( <i>“It is different from two types of battery predecessor, Li-ion batterry doesn't have memory effect and can be recharged before it is empty”</i> )	Arg1 Keyword Arg2

## 2.4 Evaluation

Evaluation process was needed to see the success rate of a developed method. The success rate of the developed method can be seen from the achievement of the accuracy value owned, the greater the value of the accuracy the better the method developed. The process of calculating accuracy was shown in Equation 1.

$$Accuracy = \frac{\text{The amount of correct data}}{\text{The amount of all data}} \quad (1)$$

## 3. Result and Discussion

This research using a dataset of 60 samples of paragraphs derived from 9 courses of study courses Informatics Engineering. The 60 paragraphs produced 278 sentences and 654 questions with varying degrees of difficulty according to the degree of difficulty in the bloom's taxonomy. The templates used to automate question generation automatically amount to 64 sets of

templates. The entire question generated was then validated by the designated expert. Validation was done to ensure the generated questions are completely understandable. A question is said to be valid if the question can be understood meaning well, so it can be answered. The 534 questions were declared to be valid and 120 questions declared invalid, so the accuracy obtained for 81.65%. The accuracy of the results obtained was less than the maximum because of the structure of questions that were not in accordance with the structure of the question in general that was used so that there was a discrepancy when put into the template question. From this research note that the result of the spread of uneven question patterns, keywords that were often used in the question generation process are "adalah, dengan, untuk, merupakan, digunakan, oleh, memiliki". This study also shows that this method generates more questions at the level of remembering and understanding.

#### 4. Conclusion

The method of analysis of proposed sentence structure is successfully used to generate questions and answers automatically. The result of the question raised has been classified based on the difficulty level of the taxonomy of bloom. This study proved to increase the number of question variations, because of 278 sentences can be generated 654 questions with an accuracy of 81.65%. This study can recognize the use of pronouns. But the recognizable pronoun is still limited to "ini" and "itu". In the future, researchers will try to improve the value of accuracy and optimize the use of patterns so that the quality of this question generation program can be even better.

#### Acknowledgment

The authors say thank you to Kemenristek Dikti for supporting this research and all of participants who have helped in the process of collecting datasets, classification and validation in this research.

#### References

- [1] M. Liu, V. Rus, and L. Liu, "Automatic Chinese Factual Question Generation," In IEEE, Pp. 1–12, 2016.
- [2] 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 International Conference On Information Technology Systems And Innovation (ICITSI), Pp. 1–6, 2015.
- [3] A. Yahya and A. Osama, "Automatic Classification Of Questions Into Bloom's Cognitive Levels Using Support Vector Machines," In IEEE, 2011.
- [4] Y. Xu, A. Goldie, and S. Seneff, "Automatic Question Generation and Answer Judging," In Spoken Language System Group, Pp. 1–4, 2008.
- [5] H. Ali, Y. Chali, and S. Hasan, "Automation of Question Generation from Sentences," In Proceedings Of The Third Workshop On Question Generation, Pp. 58–67, 2009.
- [6] D. D. Ginanjar and K. K. Purnamasari, "Pembangkitan Pertanyaan Otomatis Untuk Teks Berbahasa Indonesia Yang Mengandung Kalimat Majemuk," Elib-Unikom, Pp. 1–7, 2016.
- [7] W. Suwarningsih And I. Supriana, "Pattern Discovery Using Qg For Question-Answering Pairs," International Journal on Electrical Engineering and Informatics, Vol. 8, No. 2, Pp. 237–252, 2016.
- [8] D. Swali, J. Palan, And I. Shah, "Automatic Question Generation From Paragraph," International Journal of Advance Engineering and Research Development, Vol. 3, No. 12, Pp. 73–78, 2016.
- [9] M. Blšták And V. Rozinajová, "Automatic Question Generation Based On Analysis Of Sentence Structure," In Springer International Publishing Switzerland, Pp. 223–230, 2016.
- [10] S. Basuki and S. Kusuma, "Automatic Question Generation For 5w-1h Open Domain Of Indonesian Questions By Using Syntactical Template-Based Features," JATIT (Journal of Theoretical and Applied Information Technology), Vol. 96, No. 12, Pp. 3908–3923, 2018.

Table 3. Question and Answer Generation Result

Id Pattern	Level	Question	Answer
1	2	Li-ion adalah... (Li-ion is....)	singkatan dari Lithium-Ion, yang umumnya banyak digunakan berbagai gadget. (stands for Lithium-Ion, which is generally widely used various gadgets.)
2	1	Singkatan dari Lithium-Ion, yang umumnya banyak digunakan berbagai gadget disebut? (stands for Lithium-Ion, which is generally widely used various gadgets called ..)	Li-ion.
56	2	Berbeda dari dua jenis baterai pendahulunya, baterai jenis Li-ion tidak lagi memiliki? (Different from the two types of battery of its predecessor, Li-ion type battery no longer has?)	memory effect dan dapat diisi ulang sebelum baterai belum benar-benar kosong. (memory effect and can be recharged before the battery has not been completely empty.)
72	2	Jelaskan apa yang dimaksud Li-ion? (Explain what is a Li-ion?)	singkatan dari Lithium-Ion, yang umumnya banyak digunakan berbagai gadget. (stands for Lithium-Ion, which is generally widely used various gadgets.)

