Available online at www.sciencedirect.com
SciVerse ScienceDirect
ELSEVIER
Procedia - Social and Behavioral Sciences 59 (2012) 297 – 303
Procedia Social and Behavioral Sciences

UKM Teaching and Learning Congress 2011

Automated analysis of exam questions according to bloom's taxonomy

Nazlia Omar[a,]*, Syahidah Sufi Haris[a], Rosilah Hassan[a], Haslina Arshad[a], Masura Rahmat[a], Noor Faridatul Ainun Zainal[a] & Rozli Zulkifli[b]

[a]*Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia*
[b]*Faculty of Engineering and Build Environment, Universiti Kebangsaan Malaysia*

Abstract

Bloom's Taxonomy is a classification of learning objectives within education that educators set for students. The cognitive domain within this taxonomy is designed to verify a student's cognitive level during a written examination. Educators may sometimes face the challenge in analysing whether their examination questions comply within the requirements of the Bloom's taxonomy at different cognitive levels. This paper proposes an automated analysis of the exam questions to determine the appropriate category based on this taxonomy. This rule-based approach applies Natural Language Processing (NLP) techniques to identify important keywords and verbs, which may assist in the identification of the category of a question. This work focuses on the computer programming subject domain. At present, a set of 100 questions (70 training set and 30 test set) is used in the research. Preliminary results indicate that the rules may successfully assist in the identification of the Bloom's taxonomy category correctly in the exam questions.

Keywords: Bloom's taxonomy; natural language processing; rule-based

1. Introduction

There are many types of assessment or 'testing' to access student's learning curves. However, written examination is the most common approach used by any higher education institutions for students' assessment. Question is an element that is intertwined with the examination. Questions raised in the paper plays an important role in efforts to test the students' overall cognitive levels held each semester. Effective style of questioning as described by Swart (2010) is always an issue to help students attend to the desired learning outcome. Furthermore, to make it effective, balancing between lower and higher-level question is a must Swart (2010). Bloom's Taxonomy, created by Bloom (1956), has been widely accepted as a guideline in designing reasonable examination questions belonging to various cognitive levels. The hierarchical models of Bloom's are widely used in education fields (Chang & Chung, 2009) constructing questions (Lister & Leaney, 2003), to ensure balancing and student cognitive mastery (Oliver et al.,

* Corresponding author. Tel.: +6-03- 8921-6733; fax: +6-03-8925-6184
E-mail address: no@ftsm.ukm.my

 doi:10.1016/j.sbspro.2012.09.278

2004). From the computer science domain itself, the taxonomy improves curricular design and assessments (Scoot, 2003).

Normally, academicians would categorise a question according to the Bloom’s cognitive level manually. However, according to Yusof and Chai (2010), not all can identify the cognitive level of a question correctly. This may lead to miscategorizing of the exam questions and subsequently may fail to meet the examination standard required for the subject. In addition, some academicians also show no significant agreement on how to use Bloom's taxonomy in educating students (Johnson & Fuller, 2006).

The aim of this paper is to propose a rule-based approach in determining the Bloom’s taxonomy cognitive level of examination questions through natural language processing. Exam questions will be analyzed and each question will be categorized based on the Bloom’s taxonomy cognitive level. The scope of the work is limited to computer programming domain. This will assist the academicians in setting up suitable exam questions according to the requirements.

2. Related Work

Much work (Swart, 2010; Scott, 2003; Thompson et al., 2008; Chang & Chung, 2009) has attempted to classify exam questions based on the Bloom’s taxonomy. However, there has not been much attempt in using natural language processing techniques to solve this problem. Chang & Chung (2009) presented an online test system to classify and analyse the cognitive level of Bloom’s taxonomy to English questions. The system accepts the exam question as an input, which will then be segmented. This system has a database where various verbs of Bloom's taxonomy are stored. The database includes verbs with lower-case and capital letters. The system then compares all the verb tenses present in the questions. When a keyword is found in the test item, then the particular question belongs to the keyword. Weightage for the question is applied if any of Bloom's category shares the same verb. The authors provide four match situations to indicate matching items; Correct Match Items, Partial Match Items, No Keyword items and No Match Items. Result shows that keywords show efficiency only to 'Knowledge' level of Bloom's.

Previous researchers proposed a model to classify question items with artificial neural network approach that applies different feature method (Yusof & Chai, 2010). The model is trained using the scaled conjugate gradient learning algorithm. Several data processing techniques are applied to a feature set and then the content of a question is transformed into a numeric form called a feature vector. In order to perform text classification, three types of feature set are used i.e. whole feature set, the Document Frequency (DF) and Category Frequency-Document Frequency (CF-DF). A question item which consist of 274 questions were selected for processing. From the system, out of the three feature sets, DF reductions gave more efficient result with the combination of classification and convergence time.

Automarking (Cutrone & Chang, 2010), a learning management system, is capable of automarking once students submit their answers online. Through natural language processing, the student's answer is evaluated with semantic meaning. This is done through text pre-processing phase where the semantic meaning get 'special space'. The product of pre-processing phase is the canonical form. Comparisons between the canonical from the student’s response and the correct answer are compared to achieve the level of equivalence. Finally, appropriate grade values will be given. However, the system is unable to analyse multiple sentences based on the overall meaning.

Although all the works above incorporate Bloom's taxonomy in their work, they do not categorise question based on the semantic of the text. A work from Chang and Chung (2009) is based on keyword matching while keywords are varied over researchers. Question categorization should imply the nature of the question and how the questions can help educators to identify the learner's cognitive level.

3. Bloom’s Cognitive Domain

Cognitive domain of Bloom's Taxonomy (Bloom, 1956) is one of the three domains that were introduced by Benjamin Bloom in 1950s. This domain is designed to verify a student's cognitive quality during written examination. The famous Bloom's taxonomy consists of six levels i.e. knowledge, comprehension, application, analysis, synthesis and evaluation (Bloom, 1956).

The following describe each levels of Bloom's Taxonomy:

3.1 Knowledge-level:

Also known as recalling of data (Bloom, 1956). Scott (2003) refers it as 'rote learning' or 'memorization'. This level serves as the lower level or the beginning level of the hierarchy. It is a level where students remember or memorize facts or recall the knowledge they learn before.

The questions for programming in this category have the criteria of recalling specific input from previous lessons, defining or describing computing terms, methodology and process, stating relevance description for a subject area, concept or term and listing explicitly information from questions (Scott, 2003).

Examples:

a) List all the nodes in the left subtree of node J.

b) Describe the key properties of a binary tree.

c) Define method in JAVA.

3.2 Comprehension-level:

Bloom (1956) describes this level as grasping the meaning of information. The ability to interpret, translating, extrapolating, classifying, explaining are the concepts of these levels. The questions for programming (Thompson et al., 2003) in this category could be translating algorithm (e.g.; write output of a program), explaining the processes and flows of program and providing examples to illustrate a concept or an algorithm.

Examples:

a) What is the output of the following code segment?

b) Explain in words what happens in the following C++ code.

3.3 Application-level:

Application is defined by applying the concept to a certain scenario (Starr et al., 2008). The questions for programming in this category have the following criteria: understand the concept and use it to a new algorithm and modifying controls.

Examples:

a) Declare a variable, employees to represent the records of 120 employees.

b) Modify the given 'for' loop into 'while' loop.

3.4 Analysis-level:

This level requires students to breakdown information into simpler parts and analyse each of it. This may imply drawing a relationship, assumptions, distinguish or classifying the parts. According to Thompson et al. (2008), programming questions should contain the following: subdivide programming algorithm into classes, components or methods; systematize elements to achieve objective; recognize components of a development and distinguish non-related components or needs. In addition, it should be able to explain what exactly happens to memory when the codes are executed line by line.

Examples:

a) Outline how class `BookList` could be implemented using an array.

b) Given the following postfix notation:

*12 9 + 9 *5 3 / - =*

Using the stack implementation to evaluate the above postfix notation, show the memory configuration of the stack after the token at number 2 is read by the compiler.

3.5 Synthesis-level:

If a student achieves this level, the student should be able to integrate and combine ideas or concepts by rearranging components into a new whole (a product, plan, pattern or proposal) (Bloom, 1956). Scott (2003)

suggests programming questions for this level should instruct student to write codes based on previous level by writing a complete program or create new alternative methods or algorithm to solve a problem.
Examples:

a) *Write the definition of the function OutputTime if the statements from the lines 22 to 34 were to be performed in a function.*
b) *Write a program that prompts the user to input the masses of the bodies and the distance between the bodies. The program then outputs the force between the bodies.*

3.6 Evaluation-level:

This is a final level where judging, criticism, supporting or defending own stand involves. Thompson et al. (2008) discuss this level in Bloom's Taxonomy for CS Assessment. According to them, programming question is interpreted by checking codes if the code fits the requirement for testing strategy. This level also includes commenting quality of codes based on standards or execution criteria.
Example:

a) *Justify the concept of inheritance and give the sample of code to illustrate your answer.*

Diverse study suggests various Bloom's verb but this work presents the ambiguity problem (Chang & Chung, 2009). Some of the keyword may appear in other level as well (Jones et al., 2009). Nevertheless, there are no exact standard verb keywords for each level so far. As a result, some researchers took their own initiative to provide related keywords (Chang & Chung, 2009). Mastering lower levels is a prerequisite before students are able to move to more difficult levels (Ranganathan & Nygard, 2010; Chang & Chung, 2009). However, it is not a good practice to directly assume a specific cognitive level for a question simply because a similar verb is found (Thompson et al., 2008)

For instance, consider the following question:

Write a complete C++ program that sums the values contained in the array K with the size of 12 elements. The values are: 1,3,5,4,7,2,99,16, 45, 67, 89, and 45. The result should display:

''The total value of all array elements is…………''.

The question above requests the student to write an array of K with size of 12 in C++, with values for it are provided. We might assume that this question can be matched with the keyword **Write** because the word 'Write' appears in it. 'Write' can either be in Knowledge or Synthesis (will be discussed later). If we take a closer look at it, the question requires us to formulate a program written in C++ that can sum all 12 values contained in the array of K. Therefore, the suitable cognitive level for the question is Synthesis. A student with this level of learning should take previously learned concepts and apply them together to create something new (Scott, 2003). Table 1 summarizes each levels of Bloom's taxonomy as outlined by Bloom et al. (1956).

Table 1. Summary of categories in Bloom's Taxonomy

Bloom's category	Definition	Sample keywords / verb
Knowledge	Draws out factual answer, testing recall & recognition of specific facts.	Recall, Tell, List, State and Name.
Comprehension	Understand the meaning of the information.	Arrange, Explain, Classify, Translate, Distinguish and Demonstrate.
Application	Ability of applying knowledge to actual situations (new and concrete).	Modify, Apply, Operate, Prepare and Illustrate.
Analysis	To break down into parts, or forms. Make a relation to the assumptions, classify and distinguishes.	Distinguish, Examine, Identify and Categorize.
Synthesis	Rearrange component ideas into a new whole. Develop a pattern or structure from diverse elements.	Create, Combine, Develop, Rewrite and Compile.
Evaluation	Discriminate the value using definite criteria and	Appraise, Critique, Decide,

make comparisons.	Evaluate and Judge.

4. Research Methodology

In this work, a rule-based approach is adopted in classifying the question items into their corresponding Bloom's cognitive level. The test items are a collection of examination questions in Programming subjects obtained from the Faculty of Technology and Information Science, UKM. The training set consists of 70 examination questions and the test dataset comprises of 30 questions. Only written final examination question are taken for test items. All of questions were manually categorized by a group of subject matter expert in programming domain. The system will classify each of questions automatically to their corresponding verbs from the Taxonomy with the assistance of the developed rules. In order to determine the category of questions, this work excluded difficulty level of each question as a measuring factor.

5. Pre-processing

Text pre-processing is a method in natural language processing to make the computer understand the structure or content of the text. It will allow us to make the text more readable and easy to use for later process. Text pre-processing involves processes such as stopwords removal, stemming, lemmatization and POS tagging. In this work, stopwords removal is applied to the question in order to make the text more readable for later process. Following this, each word will then be tagged using a tagger. In this research, NLTK tagger (Bird et al., 2009) is used to tag the exam questions. To illustrate the tagging process, consider the following sentence:

"*Outline how class ArrayList could be implemented using an array*.",

The tagged output is: Outline/VB how/WRB class/NN ArrayList/NN could/MD be/VB implemented/VBN using/VBG an/DT array/NN./.

The tagger will help to identify important nouns and verbs, which may be important in determining the question's category. In addition, the sentence pattern may assist in the correct identification of the question's category. After tagging, some rules will be applied according to question's structure.

6. Rules Development

Through this research, a rule-based approach is adopted in determining the category of an examination question based on the Bloom's taxonomy. The rules are developed from a training set which consists of 70 examination questions in the programming subjects. There are two conditions where the rules will be applied:

- The rules will distinguish the suitable keyword for each question depending on its category.
- Help to choose the correct category if the keyword shares more than one category. For example, *Summarize* may fall under Comprehension or Synthesis category.

After analysing all the questions in the training set, the questions' patterns show that most of them start with a verb. However, only some of it starts with Wh-pronoun, a determiner, preposition or subordinating conjunction, a noun and an adverb. Before rules can be applied, specific patterns should be identified from the questions item. The following will demonstrate how the pattern and rules are developed after POS tagging is applied.

Question: Write down the output of the following program:

Question with tag: Write/VB down/RB output/VB following/JJ program/NN :/:

Pattern: /VB (1st word), /VB (3rd word)

Each verb in the question will be captured. The verb 'Write' appears as its keyword. Based on Bloom's Taxonomy, Write can be categorised into two categories: **Knowledge** and **Synthesis**.

The following algorithm illustrates the process of matching the rules based on the question:

```
FOR each of question,

    FOR each sentence, read into an array. Split into words.
            IF pattern is found
                If the keyword "Write" is found
```

```
                IF found:
                        Apply Rule1
                                Assign weight to Knowledge
                        Apply Rule2
                                Assign weight to Synthesis
                Choose the greater value or positive value
                Assign question category
                Store in database
    FOR EACH_match in pattern :
            print join (keyword, category, question)
```

Based on the algorithm, the question can be applied to two different rules i.e Rule 1 and Rule 2. Rule 1 states that the questions fall under the 'Knowledge' category meanwhile Rule 2 states that it can be categorised under the 'Synthesis' category. This raises a conflict as to which category the question should fall into. When this situation occurs, there is a need to introduce 'category weighting' to assist in the decision. The next subsection explains the assignment of weights for the category.

7. Category Weighting

As mentioned before, simply relying on a keyword found in the question does not necessarily means that a correct Bloom's taxonomy category or cognitive level can be determined automatically. Based on the given scenario, a question may fall into more than one category. Thus, to overcome this problem, weights are assigned to the conflicting categories. The weight is calculated based on question's category from subject matter experts (SMEs). For example, based on the experts, the previous question can be assigned to the following weights: 0.3 for Knowledge and 0.7 for Synthesis.

Based on these weights, the proposed system will choose the higher weight i.e. 0.7 in which the question will be categorised as Synthesis. Currently, we are still working on the training set to obtain the optimised weights for each question and its patterns. Figure 1 shows the overall process in determining the Bloom's taxonomy category of a given question.

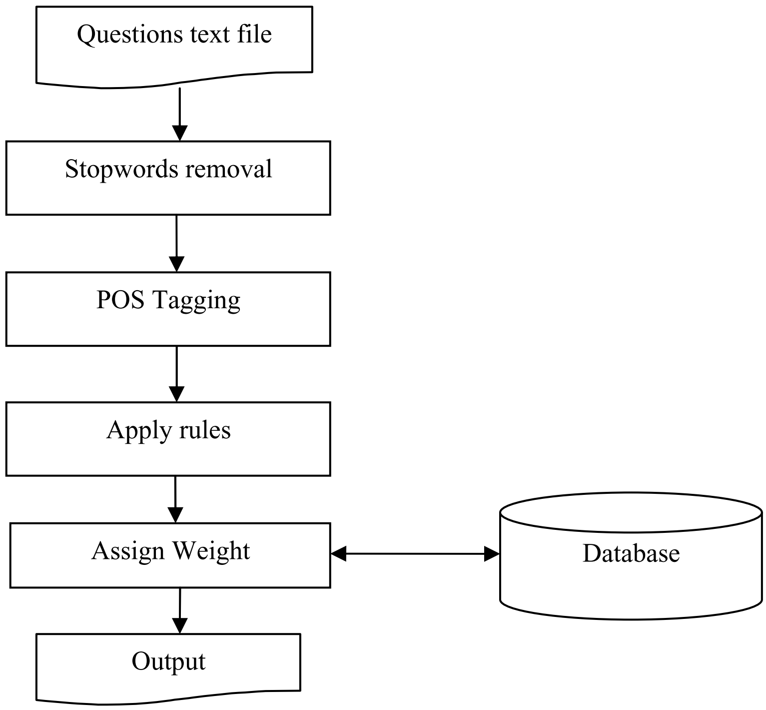

Figure 1. Processes involved in determining Bloom's category

8. Conclusion

Bloom's Taxonomy is a classification of learning objectives within education that educators set for students. We depicted a concept to automate the process of categorising examination question according to Bloom's Taxonomy based on its cognitive levels. The formation of rules may improve the accuracy of the result. For future work, more rules will be developed and tested to enhance the system's effectiveness. Thus, further testing has been our main interest in the near future.

References

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*, O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.USA.

Bloom, B.S. *Taxonomy of Educational Objectives – Handbook 1 Cognitive Domain*, London. Longman. (1956).

Cutrone, L., & Chang, M. (2010). Automarking: Automatic Assessment of Open Questions, *2010 10th IEEE International Conference on Advance Learning Technologies*. 143-147.

Johnson C.G. & U. Fuller. (2006). Is Bloom's Taxonomy is Appropriate for Computer Science, Koli Calling 2006, *Koli National Park, Finland*. 115-118.

Jones, K. O., Harland, J., Reid, J. M.V., Barlett, R. (2009) . Relationship Between Examination Questions and Bloom's Taxonomy, *39th ASEE/IEEE Frontiers in Education Conference*. October 18-21, 2009. San Antonio, TX. W1G-1 – W1G-6.

Lister, R., & Leaney, J. (2003). Introductory Programming, Criterion-referencing and bloom. SIGCSE '03: *Proceeding of the 34th SIGCSE Technical Symposium on Computer Science Education, ACM Press*. 143-147.

Oliver, D., Dobele, T., Greber, M., & Roberts, T. (2004). This course has a Bloom rating of 3.9, *Proceedings of the sixth conference on Australasian computing education, Australasian Computing Society Inc* – Volume 30, Dunedin, New Zealand, 227-231.

Ranganathan, P., & Nygard, K. A. (2010). Bloom's Online Assessment Test (BOAT) to assess student learning outcome in a distance engineering education course. *2010 2nd International Conference on Computer Engineering and Technology*. 159-161.

Scott, T. (2003). Bloom's Taxonomy Applied to Testing in Computer Science Classes. *Consortium for Computing Science in Colleges: Rocky Mountain Conference*. (October 2003) 267-274.

Starr, C.W., Manaris, B., & Stalvey, R.H. (2008). Bloom's Taxonomy Revisited: Specifying Assessable Learning Objectives in Computer Science. *SIGCSE '08*. Portland, Oregon, USA. March 12-15, 261-265.

Swart, A.J. (2010). Evaluation of Final Examination Papers in Engineering: A Case Study Using Bloom's Taxonomy, *IEEE Transactions on Education*, (May 2010) Vol. 53, No.2 257-264.

Thompson, E., Luxton-Reilly, A., Whalley, J. L. Hu, M., P. Robbins. (2008). Bloom's Taxonomy for CS Assessment. *Proceeding Tenth Australasian Computing Education Conference (ACE 2008)*, Wollongong, Australia. 155-162.

Wen-Chih Chang, Ming-Shun Chung. (2009). Automatic Applying Bloom's Taxonomy to Classify and Analysis the Cognition Level of English Question Items. *IEEE*. 727-733.

Yusof, N., & Chai, J. H. (2010). Determination of Bloom's Cognitive Level of Question Items using Artificial Neural Network. *2010 10th International Conference on Intelligent Systems Design and Applications*. 866-870.