Journal of Theoretical and Applied Information Technology
30th June 2015. Vol.76. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# BLOOM'S TAXONOMY QUESTION CATEGORIZATION USING RULES AND N-GRAM APPROACH

[1]SYAHIDAH SUFI HARIS, [2]NAZLIA OMAR
[1]Center for Artificial Intelligence Technology (CAIT), FTSM, UKM, Bangi
[2] Center for Artificial Intelligence Technology (CAIT), FTSM, UKM, Bangi
E-mail: [1]syahidahsufi@gmail.com, [2]nazlia@ukm.edu.my

**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. An experiment was designed to investigate student's cognitive level, by developing rules to determine the categorization of questions based on Bloom's Taxonomy (BT). A sample of 135 questions collected from final examination past questions from FTSM, UKM. All questions has been analyzed by Computer Science subject matter experts to identify cognitive category based on BT. Rules are developed by analyzing the syntactic structure from the text questions. Next, some adjustment are made to utilize hybrid ability of rules and statistical approach. 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. The advantage of this approach is that statistical classifier will assist the categorization when questions are not categorized by the rules. This approach gives better flexiblity when a set of 64 rules are developed for programming question domain. The result yeilds 86% for the average $F_1$ for the hybrid technique. The outcome of this study suggest that the combined technique is capable of identifying the correct cognitive category of BT.
**Keywords:** *Programming questions, Bloom's Taxonomy, hybrid technique, categorizing question*

## 1. INTRODUCTION

Written examination is an all time favourite option for educators to assess their students' knowledge [1]. However, developing examination questions are always a challenging task especially when educators are trying to produce high quality and a reasonable one to match the variety of cognitive level of students [2, 3, 4]. Therefore, educators simply use cognitive domain of BT to determine the cognitive levels of examination question [2, 3, 4]. Apart from that, educators have their own style to decide which category is suitable for each questions made [5, 6].

BT is hierarchical type framework for educational objectives and was developed by Benjamin S. Bloom and his team [7]. By using the taxonomy, students are expected to show certain behaviours implying their way of thinking, communicating and responses as an outcome from the learning they obtain in class.

## 2. QUESTION'S CATEGORIZATION

Question categorization means selecting specific category from predefined question category and assign it to a given question [12]. There are a few methods chosen by researchers to classify or categorize question: Structure Induction, Neural network, Support Vector Machines (SVM), Maximum Entropy model (MEM) and rule-based approach. Structure induction extraction through Alignment-based Learning (ABL) is used in [13]. ABL classifier will find a structure or pattern of a text and change it to readable regular expression in order to classify node's information in trie classifier.

Kocik [14] describes how the MEM can help classification of questions into one of the system components of Question Answering (QA). Experiments were carried out to classify questions to the answers. An accuracy of 85.4% is achieved on fine-grained classes. Chang and Chung [4] introduced a system that can analyse and classify English question into BT. The online system fully rely on the verb or keyword exist in the question. A database is used to store the identified keywords. Fei et. al [15] has studied the effectiveness of question classification by using learning neural network. They introduced Backpropagation Learning Algorithm as text classifier to classify question in E-Learning system by using artificial neural network. Yusof and Chai [16] have put forward the effectiveness of question classification

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

by using question classification module in their work. The experiment is carried out with artificial neural network trained using Scaled Conjugate Gradient Algorithm. Three set of features are involved; Document frequency (DF), Category frequency-Document frequency (CF-DF) and Whole Document feature.

There are also attempts to make good use of semantic meaning of student's response to evaluate open-ended question in E-Commerce domain [17]. Few researchers have demonstrated question classification by using support vector machines (SVM). The training dataset of question are converted into vector feature for SVM training to identify the 11 categories of open-ended questions [18]. Once support vector model ready, it then tested with test dataset. The highest accuracy is 92.9% with 100% precision value comes from Significance. While the lowest come from Example class with accuracy of 55.6% and 50% precision.

In order to classify open-ended question in their research, a modified version of tree kernel was introduced to help SVM to take control on syntactic structure of questions [19]. The tree kernels are able to identify weight of the tree fragments based on their depth while looking for question's focus. The results of experiment shows that after 5500 samples kernel trained under coarse-grained definition, the accuracy from word linear kernel is 85.8%, n-gram linear kernel is 87.4% and tree kernel is 90%.
According to Liu et. al [20] high accuracy can be obtained by performing six crucial steps in order to classify Chinese question. The steps are; word segmentation, keyword extraction, bag-of-words, head phrase, syntactic features and semantic features. They obtained consistently 92.35% for testing accuracy.

Classification of questions into BT of cognitive domain for E-Learning system [21] has been evaluated by experimenting with SVM performances. The SVM-light has been used as a tool to classify question. The accuracy obtained is 87.4%, precision of 85.83% and recall of 29.1%. The poor result for recall is due to inadequate size of dataset and in addition to the lack of question terms.

Research done by Panicker et. al [22] shows that question can be categorized by using machine learning approaches such as SVM and Naïve Bayes (NB) with good results if only a huge amount of data involved. In this case, the training set consists of 1500 questions for each 20 categories of newsgroup. SVM obtained precision 0.95 for both category 1 and 2 while NB obtained precision 1 for category 1 and 0.58 for category 2. This suggests that SVM performed as effective classifier as more data added.

All the researches mentioned above concerns on how to categorize or classify question into certain category using purely machine learning approach. Most of them used quite huge amount of data for training and therefore, triggers the results. However, this paper uses small dataset. N-gram has been suggested as statistical classifier to replace other classifier. N-gram is chosen mainly because its ability to process small data training.

## 3. CATEGORIZING QUESTION INTO BLOOM'S TAXONOMY OF COGNITIVE DOMAIN

BT is used to design various courses like Engineering and evaluate the subject [2] to ensure student's mastery of cognitive [8, 9]. According to Jones et. al [3], the best and reasonable question is questions that contain various cognitive categories to harmonize different level of student's cognitive ability.

Combination of cognitive category for every question can create 'express' question [2, 7, 8]. For example, a particular question may contain both Knowledge and Comprehension. The pairing of cognitive category will help reduce questions and student's stress when answering question. The various cognitive categories are also believed to ensure student's mastery and ability to think [1, 2, 3].

Some researchers attempted to relate how BT can be applied to examination questions. [2, 3, 6, 10] describes the involvement each of cognitive levels with good examples of programming examination questions. For example, Starr et. al [11] demonstrates application of revisited BT into first year computer science programme.

## 4. THE HYBRID TECHNIQUE

Examination question is a short text, contains less than 200 words and its features are scattered due to its sparseness [23, 24]. Experiments by Yahya and Osman [21] proved that categorization of question with purely statistical approach will causes poor results if small dataset used; in addition to the fact that question has less word. While the results are impressive if huge dataset involved [22].

Based on the nature of short text, syntactic structures of question have been considered as another option to categorize it. Both rules and n-gram intergrated together as a part of the technique to categorize the questions. This technique may

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

improve the weaknesses of both techniques by combining the advantages of both of it.

Rules are chosen because rules can help to analyze the question structure systematically. Rules are developed by combining part-of-speech (POS), regular expression and specific keyword that exist in the training set. Table 1 presents the tags used to determine sentence structure. Table 2 shows the collection of annotation to be tagged to natural text after text preprocessing (e.g text normalization and POS tagging). Annotations are developed to make grouping of an adjacent word and we can duplicate to other sentences. The annotation also can help defining the correct tagging for dataset. These annotations are made based on the syntactic analysis of questions and rules are created by combining it.

*Table 1. Part-of-speech with tags*

| Part-of-Speech | Explanation |
|---|---|
| NP | Noun Phrase |
| VP | Verb Phrase |
| PP | Preposition Phrase |
| JJ | Adjective |
| CD | Cardinal number |
| SYM | Symbol |
| ADVP | Adjective phrase |
| WRB/Wh | Why, what, when and how |

*Table 2. Annotation for tagging certain phrases for dataset*

| Part-of-Speech | Explanation |
|---|---|
| IS | Informative statement |
| TW | Trigger words |
| CF | Name (noun) for class, method of function |
| NF | Not class, method or function's name |
| CDSYM | Cardinal number and symbol |
| TW(Wh) | What is, what are |
| PP(CD) | Preposition phrase with cardinal number |
| NP(CD) | Noun phrase with cardinal number |

**a. Rules implementation**

Rules for each cognitive category are different. There is 64 rules successfully developed and covers all six category of BT (Knowledge, Comprehension, Application, Analysis, Synthesis and Evaluation). The rules for each category can be express in different way. The following is to demonstrate the rules for every category.

Rules for **Knowledge**

Main Rule: {<NF> + (<CF>?) + (<NP>|<PP>)}

i. NF + CF + NP
ii. NF + CF + PP
iii. NF + NP
iv. NF + PP

Example:

Question: "What is Encapsulation?"

| What/WP is/VBZ | Encapsulation/NNP |
|---|---|
| TW(Wh) ← NF | NP |

In the example above, TW(Wh) becoming NF is basically because the whole sentence is asking for definition of Encapsulation. 'Encapsulation' is not a class, method or function name.

Rules for **Comprehension**

Rule: {<IS> + <CF>? + <NF> + <NP>? + <TW>? + (<NP>|<PP>|<VP>)}

i. IS + CF + NF + NP + TW + NP*
ii. IS + CF + NF + NP + TW + PP*
iii. IS + NF + NP*
iv. IS + NF + TW + VP
v. IS + NF + NP + TW + NP
vi. IS + NF + NP + TW + PP*

Example:

Question: "Sort the above list using the selection sort algorithm. Show each of the passes of the sorting phase."

| Sort/VB ... algorithm/ NN | Show/ VB .... | passes/N NS | sorting/V BG phase/N N |
|---|---|---|---|
| IS | NF | TW | VP |

The rule for the question is IS + NF + TW +VP. The rule contains informative statement, non-function verb, trigger words and verb phrase.

Rules for **Application**

Rule: {(<IS>)* + <TW(Wh)> + <\w+CDSYM>}

This rule may contain more than one IS, follow by trigger words with tag Wh and combination of cardinal number with symbols.

Example:

Question: Suppose that `x` and `y` is `int` variables. Consider the following statements. What is the value of `y` if `x = 6`?

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

| Suppose/VB …. statements/NNS ./. | What is | value of y if x = 6 |
|---|---|---|
| IS | TW(Wh) | \w+CDSYM |

The rule for the question is IS + TW(Wh) + \w+CDSYM. The rule contains informative statement and trigger words with 'what is' for TW(Wh). \w+CDSYM contains more than one word followed by existence of cardinal number and symbols.

Rules for **Analysis**
Rule: {<NF> + (<NP>?) + (<TW>|<TW(Wh)>) + (<NN>?)}

i. NF + (TW | TW(Wh))
ii. NF + (TW | TW (Wh)) + NN
iii. NF + NP + TW + NN
iv. NF + NP + TW(Wh) + NN
v. NF + NP + (TW | TW(Wh))
vi. NF + NP + TW

Example:
Question: "Show/VB how/WRB the/DT split/NN and/CC mergesort/NN algorithm/NN processes/VBZ the/DT following/JJ input/NN"

| Show /VB | how/WRB the/DT split/NN and/CC mergesort/NN algorithm/NN processes/VBZ | following/JJ input/NN |
|---|---|---|
| NF | [TW(Wh)] | NN |

The rule for the question is NF + TW(Wh) + NN. The rule contains non-function verb, for word 'how' and 'quicksort processes' as TW(Wh) and noun phrase.

Rules for **Synthesis**
Rule: {<NF> + <TW> + <CF> + <VP>}

This rule contains non-function verb, followed by TW, noun for class, and verb phrase. This rule can't be modified or expanded to be other rules.

Example:
Question: "Write a program that uses nested loops to print the following output:"

| Write /VB | program/NN that/WDT uses/VBZ | nested/JJ loops/NNS | print/VB ... output/NN |
|---|---|---|---|
| NF | TW | CF | VP |

This level only has one rule and cannot be expand more. The following question shows that it contains non-function verb, method or class. It is followed by word with adjective JJ and noun phrase (NN) as TW, and a few of noun phrases which are noun for class and verb phrase.

Rules for **Evaluation**
Rule: {(<TW(Wh)> + <PP>)* + <TW(Wh)> + <NP>?)}

i. TW(Wh) + PP* + TW(Wh)
ii. TW(Wh) + PP* + TW(Wh) + NP

Example:
Question: In the context of the JVM, what is a 'just in time" (JIT) compiler? Explain how a JIT compiler can improve the performance of the JVM.

| ... what/WDT is/VBZ... | just/RB in/IN time/NN | how/WRB a/DT JIT/NNP compiler/NN | can/MD improve/VB ... JVM/NNP |
|---|---|---|---|
| TW(Wh) | PP | TW(Wh) | PP |

The rule shows that the question contains trigger words with 'what' for TW(Wh), prepositional phrase, trigger words with 'how' for TW(Wh), preprositional phrase.

Questions from test set or token will be input to prototype one by one. Token will undergo data pre-processing to extract syntactic structure from the question. Rules in rules database will try to recognize the syntactic structure and match it with the suitable and similar form. Once the suitable rules found, a specific category will be assign to the particular question with the syntactic structure found. While, rules in development stage, every rules will be revised manual, to ensure the rules fit question from training set. If rules not match with the syntactic structure, it will be pass to N-gram for next step (refer Figure 1).

Journal of Theoretical and Applied Information Technology
30th June 2015. Vol.76. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

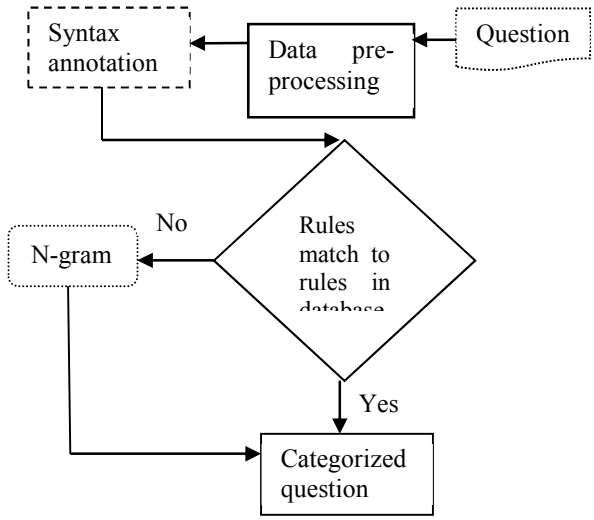


*Figure 1. The categorization procedure with rule-based approach*

**b. Statistical Method**

The algorithm adopts N-gram as it is able to find the probability to predict words. Profile for both training and test document are created and converted to vector form for data presentation. The profiles will carry the category of each question. Frequency in both documents is calculated with Tf-idf and the degree of similarity is calculated using Cosine Similarity. Then, the suitable category is then assigned to test dataset. (refer Figure 2).

## 5. RESULTS AND EVALUATION

For the experiments, 100 questions are selected for training and 35 questions are used for testing. All these questions are related to the Programming domain. To evaluate the question categorization using the combined technique, three major experiments are performed. The categorization uses rule-based approach, N-gram and a combination of both methods. Results from the three experiments are then compared. Precision, recall and F1 are used as it is the suitable measures for information extraction. The following is the results for three experiments of question categorization (Table 3, 4 and 5). Based on the average F1 from rule-based experiment, the precision and recall from this experiment is higher than F1 in N-gram experiment. However, N-gram experiment has good performance for Application and Evaluation. As for last experiment, it shows better performance than the other two experiments.

The significant changes are spotted at Knowledge and Application of the combined experiment. The value of F1 for combined approach is 1.34 times better than rule-based experiment and 1.7 times better than N-gram experiment. This means that the proposed approach has improved the performances of rules and N-gram by the combination of these method.

*Table 3: Result for question categorization with rule-based*

| Cognitive Categories | Recall | Precision | F-measure |
|---|---|---|---|
| Knowledge | 1 | 0.33 | 0.5 |
| Comprehension | 1 | 0.83 | 0.91 |
| Application | 0.6 | 1 | 0.75 |
| Analysis | 0.67 | 1 | 0.8 |
| Synthesis | 1 | 1 | 1 |
| Evaluation | 0.75 | 1 | 0.86 |
| Average | 0.84 | 0.86 | 0.80 |

*Table 4: Results for question categorization with N-gram as classifier*

| Cognitive Categories | Recall | Precision | F-measure |
|---|---|---|---|
| Knowledge | 1 | 0.25 | 0.4 |
| Comprehension | 0.4 | 0.67 | 0.5 |
| Application | 0.9 | 1 | 0.95 |
| Analysis | 0.5 | 0.75 | 0.6 |
| Synthesis | 0.57 | 1 | 0.73 |
| Evaluation | 0.75 | 1 | 0.86 |
| Average | 0.69 | 0.77 | 0.67 |

*Table 5. Results for question categorization with hybrid techniques as classifier*

| Cognitive Categories | Recall | Precision | F-measure |
|---|---|---|---|
| Knowledge | 1 | 0.5 | 0.67 |
| Comprehension | 1 | 0.83 | 0.91 |
| Application | 0.9 | 1 | 0.95 |
| Analysis | 0.67 | 1 | 0.8 |
| Synthesis | 1 | 1 | 1 |
| Evaluation | 0.75 | 1 | 0.86 |
| Average | 0.89 | 0.88 | 0.86 |

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

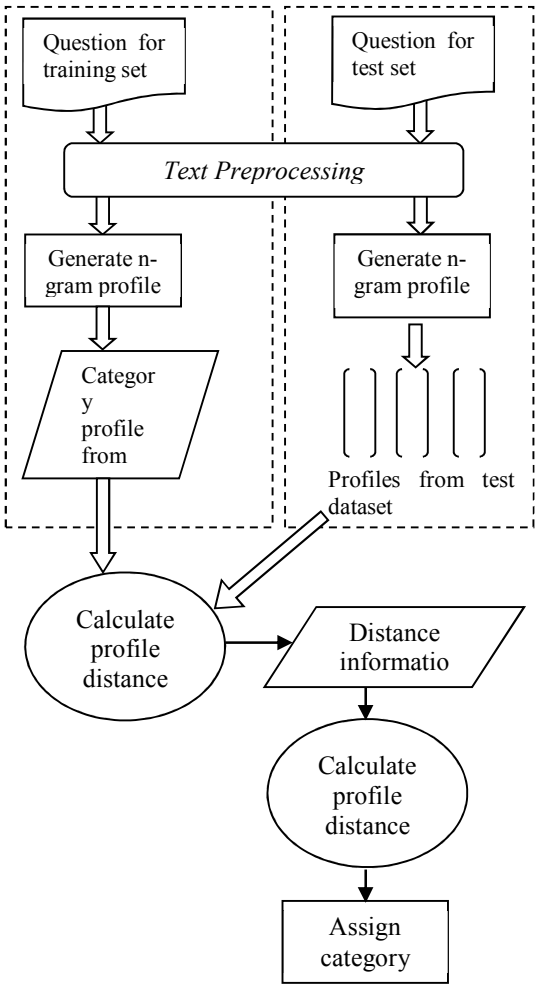


*Figure 2. Flowchart For Categorization With N-Gram*

Synthesis achieved the value of 1 for all results for both rules and combined experiment because this category has various forms of syntax and many rules has been developed for this category and contribute to the classification of the questions. During the combined experiment conducted, the rules classify the questions prior to N-gram. Therefore, the performance for questions from Synthesis is attributed to the quality rules developed.

## 6. CONCLUSION

This paper demonstrates the use of rules and N-gram classifier in the categorization of examination question. The rule-based method helps to determine the question category based on the requested procedure. It is a method that can provide good result for certain situation. However, it may be quite tedious as it can be time consuming. N-gram classifier calculates the gram token's probability and compares the token with reference document that has n-gram profile.

The combination of both methods (rules and N-gram) has proved that this method performs well to the categorized question. The combination method takes the advantages of N-gram to overcome rules weaknesses while rules overcome N-gram weaknesses by categorizing questions with predefined instructions in the rules. This method could well be applied in various Engineering courses with some modification to the rules.

## 7. REFERENCES:

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

[2] T. Scott, "Bloom's Taxonomy Applied to Testing in Computer Science Classes", *Consortium for Computing Sciences in College (CCSC)*. Rocky Mountain Conference, Vol. 19, No. 1, October 2003, pp. 267-274.

[3] K. O. Jones, J. Harland, J. M. V. Reid and R. Bartlett, "Relationship Between Examination Questions and Bloom's Taxonomy", *Frontiers in Education Conference*, FIE '09. 39th IEEE, October 2009. pp. 1-6.

[4] Wen-Chih. Chang and Ming-Shun. Chung, "Automatic Applying Bloom's Taxonomy to Classify and Analysis the Cognition Level of English Question Items", 2009 *Joint Conferences on Pervasive Computing (JCPC)*. December 2009, pp. 727-734.

[5] C.G. Johnson and U. Fuller, "Is Bloom's Taxonomy is Appropriate for Computer Science", *Proceedings of the 6th Baltic Sea conference on Computing education research*: Koli Calling 2006, pp. 115-118.

[6] E. Thompson, A. Luxton-Reilly, J. L. Whalley, Hu, M., and P. Robbins, "Bloom's Taxonomy for CS Assessment", *Proceeding Tenth Australasian Computing Education Conference (ACE2008)*, Wollongong, Australia. pp. 155-162.

[7] B. S. Bloom, "Taxonomy of Educational Objectives, Book 1 Cognitive Domain", *London*: Longman Group Limited.

[8] R. Lister, & J. Leaney, "Introductory Programming, Criterion-Referencing, and Bloom", *Proceeding of the 34th SIGCSE Technical Symposium on Computer Science Education*, University of Technology, Sydney,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Australia. Vol. 35, No. 1, January 2003, pp. 143-147.

[9] D. Oliver, T. Dobele, M. Greber, and T. Roberts, "This course has a Bloom rating of 3.9", *Sixth Australasian computing education Conference (ACE2004)* Australasian Computer Society, Vol. 30, 2004, pp. 227-231.

[10] N. Omar, S. S. Haris, R. Hassan, H. Arshad, M. Rahmat, N. F. A. Zainal and R. Zulkifli, "Automated Analysis of Exam Questions According to Bloom's Taxonomy", *Procedia - Social and Behavioral Sciences*, Vol. 59, October 2012. pp. 297-303.

[11] C. W. Starr, B. Manaris, and R. H. Stalvey, 2008. "Bloom's taxonomy revisited: specifying assessable learning objectives in computer science", *SIGCSE Bulletin*, Vol 40, No. 1 March 2008, pp. 261-265.

[12] K. Hacioglu and W. Ward, "Question Classification with Support Vector Machines and Error Correcting Codes", *Proceedings of HLT-NAACL 2003 (NAACL-Short '03)*, Vol. 2, 2003, pp. 28-30.

[13] M. van Zaanen, L. A. Augusto, and D. Mollá, 2005. "Question Classification by Structure Induction", *Proceeding of Nineteenth International Joint Conference on Artificial Intelligent (IJCAI-2005)*. Vol. 19, pp. 1638-1639.

[14] K. Kocik, "Question Classification Using Maximum Entropy Models", *Honours Thesis*, University of Sydney, 2004.

[15] T. Fei, J. H. Wei, C. T. Kim and Q. Tian, "Question Classification for E-learning by Artificial Neural Network", *Proceedings of the 2003 Joint Conference of the Fourth International* Conference on Information, Communications and Signal Processing and Fourth Pacific-Rim Conference on Multimedia (ICICS-PCM 2003), Vol. 3, December 2003, pp. 1757-1761.

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

[17] L. Cutrone, and M. Chang, "Automarking: Automatic Assessment of Open Questions", *10th IEEE International Conference on Advance Learning Technologies*, July 2010, pp. 143-147.

[18] J. Bullington, I. Endres, and M. Rahman, "Open Ended Question Classification Using Support Vector Machines", *In Proceedings of the Eighteenth Midwest Artificial Intelligent and Cognitive Science Conference (MAICS-2007)*. April 2007.

[19] D. Zhang, and W. S. Lee, "Question Classification using Support Vector Machines", *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '03, Singapore. 2003, pp. 26-32.

[20] B. Liu, Z. F. Hao, X. W. Yang, and X. D. Lin, "Chinese Question Classification with Support Vector Machine", *IJCSNS International Journal of Computer Science and Network Security*, Vol. 6, No. 7, July 2006, pp. 231-240.

[21] A. A. Yahya and A. Osman, "Automatic Classification of Question into Bloom's Cognitive Levels using Support Vector Machines", *The International Arab Conference on Information Technology*, Naif Arab University for Security Science (NAUSS), December 2011, pp. 1-6.

[22] A. D. Panicker, U. Athira and S. Venkitakrishnan, "Question Classification using Machine Learning Approaches", *International Journal of Computer Applications*, Vol. 48, No. 13, June 2012, pp. 1-4.

[23] F. Zhou, F. Zhang, Bingru Yang & Yu Xingang, "Research on Short Text Classification Algorithm Based on Statistics and Rules", 2010 *Third International Symposium on Electronic Commerce and Security*, July 2010, pp. 3-10.

[24] J. Wang, L. Li, and F. Ren, "An Improved Method of Keywords Extraction Based on Short Technology Text", *International Conference on Natural Language Processing and Knowledge Engineering (NLP-KE)*. August 2010, pp. 1-6.