TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 4374 - 4385

Questions Classification According to Bloom's Taxonomy using Universal Dependency and Word Net

Thing Thing Goh[1,2*], Hassan Mohamed[2], Nor Azliana Akmal Jamaludin[2], Mohd Nazri Ismail[2], H. S. Chua[1]

[1]*School of Engineering, KDU University College, Selangor, Malaysia*

[2]*Faculty of Defense Science and Technology, Universiti Pertahanan Nasional Malaysia (UPNM), Kuala Lumpur, Malaysia*

ttgoh@kdu.edu.my, hassan@upnm.edu.my, azliana@upnm.edu.my, m.nazri@upnm.edu.my,hs.chua@kdu.edu.my

Article Info
Volume 82
Page Number: 4374 - 4385
Publication Issue:
January-February 2020

Article History
Article Received: *18 May 2019*
Revised: *14 July 2019*
Accepted: *22 December 2019*
Publication: *22 January 2020*

Abstract:

Question Classification (QC) based on Bloom's Taxonomy has been widely accepted and used as a guideline in designing a holistic set of examination questions that consists of various cognitive levels. However, many discrepancies happened in QC due to inconsistence or misclassification of questions to Bloom'slevel. This paper proposes a system that can analyze the examination questions and determine the appropriate Bloom's levels using syntactical and semantic approach. Universal Dependency (UD) that implies Natural Language Processing (NLP) technique is used to identify the important keywords and verbs. Then, WordNet similarity algorithm with Natural Language Toolkit (NLTK) is used to identify the questions category according to the Bloom's Taxonomy. This research focuses on Science, Technology, Engineering, and Mathematics (STEM) examination questions. At present, a set of 100 questions is used and preliminary result indicates both Universal Dependency and WordNet similarity algorithms being able to categorize successfully the questions based on Bloom's Taxonomy.

Keywords: *Question classification (QC), Bloom's Taxonomy, Natural Language Processing (NLP), Universal Dependency, WordNet similarity, Natural Language Toolkit (NLTK).*

1 Introduction

Outcome-based Education (OBE) is an education theory that focuses on what students should learn (outcome) and able to apply after the learning process [1]. In order to ensure the students achieving the defined outcomes, Malaysia Qualification Agency (MQA) has implemented a Malaysia Qualifications Framework (MQF) to regulate the quality and standard of higher education providers. This framework is defined as an instrument developing and classifying qualifications based on a set of criteria that is agreed nationally since 2008. Five learning outcomes clusters have been defined and used as a qualification guideline for all Technical and Vocational Education and Training (TVET) to attain their programmes accreditation by MQA. Programme is referred to any courses or series of subjects or modules offered by TVET that is structured for a specified duration and learning volume to achieve the stated learning outcomes, which usually leads to an award of a qualification such a diploma certification, a bachelor or master degree and etc. Then the Programme Outcomes

January - February 2020
ISSN: 0193 - 4120 Page No. 4374 - 4385

(POs) are the specified knowledge, skills, altitude and abilities that the programme students should acquire and demonstrate upon graduation [2].

The five clusters include cognitive, functional and work skills, and specific industry-appropriate competencies. All POs need to be aligned to these five clusters set by MQA. The achievement of POs depends on the learning outcomes (LOs) of each subject or module in the programme. The LOs are the specified knowledge, skills, altitude and abilities that the students should acquire upon the completion of a period of study in the subjects or modules. Meanwhile, the achievement of LOs depends heavily on the score obtained in each assessment components. The assessment components are tutorials, tests, projects and assignments that are given to students during the period of a semester. It includes also the mid-term examination and the final examination in the semester. In order to attain the POs and LOs achievement result, the assessment components must be designed in line with the LOs set for the modules. Questions set in each assessment components play an important role to help students attaining the desired LOs. Bloom's Taxonomy has been widely used as a guideline in designing a holistic examination questions which consists of various cognitive levels [3]. The Bloom's Taxonomy was created by Benjamin Bloom during the 1950s and is a way to categorize the levels of reasoning skills required in classroom situations. There are six levels in the taxonomy, each requiring a higher level of abstraction from the students. Teachers should attempt to move students up the taxonomy as they progress in their knowledge [4].

Most of the time, educators categorize the examination questions for each assessment into the Bloom's levels manually basedon their understanding of Bloom's levels which may dissimilar from one educator to others[5]. The classification of questions is usually based on verbs used in the examination questions. The verbs are extracted from the examination questions and then mapped to the verbs list in the Bloom's Taxonomy or Bloom's level. However, some of the Bloom's verbs are ambiguous when the verbs fall into more than one category of Bloom's Taxonomy. Therefore, it is tedious and problematic to categorize the examination questions contained such verb are often inconsistently categorized by different educators. Generally, the current practice of Institution of Higher Learning (IHL) in Malaysia is that all examination questions need to be moderated by one or two academics in order to reduce the discrepancy of mapping. However, this work consumes a lot of time and shows inconsistency among academics. Therefore, this research is carried out to propose a framework that can analyse the examination questions and determine the appropriate Bloom's level using Natural Language Processing (NLP)'s semantic approach, the Universal Dependency (UD).

2 Literature Review

2.1 Education Taxonomy

Bloom's Taxonomy is widely accepted and used as an important framework to guide educators in developing a holistic assessment and promoting higher forms of thinking in education [6]. This Taxonomy was introduced by Benjamin Bloom and his research team in year 1956. The Bloom's Taxonomy consists of six cognitive levels - Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation.

This framework was then modified by Anderson and team in year 2001. The revised Bloom's was changed from noun to verb forms [7]. The Bloom's verbs were reorganised also. The change of terminology was to indicate action because thinking implied active engagements. For example, the lowest level of original Bloom Taxonomy, "knowledge" inaccurately described a category of thinking. Thus, it was replaced with the verb "remembering". Besides, the top two levels were swapped. The revised taxonomy had swapped the "evaluation" stage down a level and the "creating" was revised to the highest level. In the "evaluating" level, students needed to defend, support, justify and evaluate their opinion, while at the highest level, the "creating", they needed to generate new ideas, create new products, or construct new points of view. It was revised in such a way because it was able to reflect the increase of complexity of thinking. Creative thinking was considered a more complex form of thinking comparing to the evaluating which was not necessarily involving creative thinking.

The cognitive level started from the lowest level – "remembering" and increasingly moved to complex and abstract higher levels. Students were expected to master the lower levels before moving into higher levels. Furthermore, the noun list of Bloom's had been revised into a verb list byAnderson et al.[8]. The list of verbs was shown as Table 1.

Table 1. Anderson's Revisions on Bloom's Taxonomy verbs[24].

Category	Cognitive Verb list of Anderson Taxonomy	
	Description	Verb list
Creating	Builds a structure or pattern from diverse elements	categorizes, combines, compiles, composes, creates, devises, designs, explains, generates, modifies, organizes, plans, rearranges, reconstructs
Evaluating	Make judgments about the value of ideas or materials	appraises, compares, concludes, contrasts, criticizes, critiques, defends, describes, discriminates, evaluates, explains, interprets
Analysing	Separates material or concepts into component parts	analyses, breaks down, compares, contrasts, diagrams, deconstructs, differentiates, discriminates, distinguishes, identifies, illustrates
Applying	Use a concept in a new situation orunprompted use of an abstraction	applies, changes, computes, constructs, demonstrates, discovers, manipulates, modifies, operates, predicts
Understanding	Comprehending the meaning, and interpretation of instructions and problems.	comprehends, converts, defends, distinguishes, estimates, explains, extends, generalizes, gives an example, infers, interprets, paraphrases
Remembering	Recall or retrieve previous learned information.	defines, describes, identifies, knows, labels, lists, matches, names, outlines, recalls, recognizes, reproduces, selects, states

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 4374 - 4385

2.2NLP for Examination Question Classification

The classification of questions generally based on the verbs used in the questions. The verbs are compared with the Bloom's verbs list to determine the cognitive level of questions. Many researchers have worked on the automation of examination questions classification based on Bloom's Taxonomy using the Natural Language Processing (NLP) approach. NLP is a form of artificial intelligence that helps machine to read and understand text created by human.

NLP began in 1950 and focused in text information retrieval (IR) with large volumes of indexes and search for text. It was started to implement in the word-for-word Russian-to-English machine translation in World War II [9]. With the revolution of computer programming and lexical corpus, NLP had widely implemented in variety interactive applications such as smartphone assistants, online banking and retail self-service tools in automatic translation programs. NLP techniques incorporated a variety of method including rule-based, semantics, and machine learning to extract entities, relationships and understand context, which enabled an understanding of what being said and written, in a comprehensive way.

In area of question classification, many previous researches had relied on matching the keywords in the questions against a set of rules [10]. In 2002, Pinto et al.[11] had proposed a statistical approach with keywords matching to improve the performance and efficiency of Question Answering from web data using semi-structured metadata (QuASM). Language models such as unigram and bigram model had been used to discover the probability of a question given a question class. Question retrieved from web data was classified into question classes according to its entity such as time, location, area, person, organization, money and percent. Then, the entity tagged in the question was used to determine the likely answer to a question from the answer entities.

Chang et al.[12]had attempted to classify the examination questions based on the Bloom's Taxonomy using keywords matching in 2009. It compared the verbs of the examination questions with the verbs of Bloom's Taxonomy to find any exact verbs matching with weightage. However, itwas not accurate because it categorized the questions based on exact keywords matching without considering the semantic of the keywords especially when the keywords can be mapped to many Bloom's levels.

Later, Omar et al.[3]and Haris et al.[13] had proposed an automated analysis of the examination questions to determine the appropriate category based on Bloom's Taxonomy using the rule-based approach based on syntactic structure of the questions. This approach had applied NLP techniques (text pre-processing) to identify the important keywords and verbs. Text pre-processing techniques were stop-words removal, stemming, lemmatization and POS (Part-of-Speech) tagging using NLTK. Rules had to be developed according to the tagging pattern of the sentences. The approach had achieved satisfactory result with macro F1 of 0.77. However, it would need more rules and greater training questions in order to improve the accuracy of the system. It became a tedious task when more rules were needed to be maintained. Furthermore, this approach's accuracy had been restricted by the style of questions. The different style of question had resulted in different syntax and then affected the precision of the result.

In 2014, Biawas et al.[14]had proposed a rule-based approach also for question classification based on syntactic pattern in the sentence. He adopted 2000 questions from Li et al.[19] data set in his research. He employed Stanford's POS tagger to determine the syntactic structure of the

TEST Engineering & Management
January - February 2020
ISSN: 0193 - 4120 Page No. 4374 - 4385

questions and then categorised the questions into 3 categories. He found that similar types of questions having the same syntactic structure. Comparing to the result obtained by Li et al.[19]with a 2-layered taxonomy of 6 course grain and 50 fine grained categories, less number of question categories (3 categories) could achieve a more satisfactory and better result.

In 2016, Ramesh et al.[15] had developed an automated system to generate LO annotation and analysed the content of question paper using NLP with rule-based approach. Results were classified into 4 types - true positive, true negative, false positive and false negative - with confusion matrix.

In overall, rule-based approaches classified question into an organized group by using a set of handcrafted linguistic rules. It could be time consuming and tedious to maintain a numerous set of rules set for a complex system. A lot of analysis and testing was needed in order to ensure a high accuracy result.

In 1980s, Machine learning (ML) approaches in NLP had become more prominent [9]. Machine learning (ML) term is coined by Arthur Samuel in 1959. He defined ML as a field of study that gives computers the ability to learn without being explicitly programmed [16].Mitchell [17] defined ML in detail as a computer program is said to learn from experience E with respect to some class or tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E. Thus, the ML approach in question classification defines the task T asquestion classification according Bloom Taxonomy, the performance P as the classification result while the experience E asthe past questions classification data and result. The machine learning approaches are used to overcome the defect of rule based classification. They are Support Vector Machine (SVM), Nearest Neighbours (NN), Naïve Bayes(NB), Decision Tree (DT), Sparse Network of Winnows (SNoW). ML has been divided into supervised ML and unsupervised ML. Supervised ML is the machine trained with pre-defined dataset. The questions tagged with associated target responses are given to the machine in prior to predict the correct response when new questions were posted. Based on the training dataset, the machine can classify accurately when given a new set of data. On the other hand, unsupervised ML finds the patterns and relationship in the database without labelling. The machine identifies the set of rules by itself and makes decision.

In 2002, Li et al.[18] had presented a machine learning approach to classify the questions into two-layered taxonomy in order to find an accurate answer to a question given a large collection of text. Questions were categorized into six coarse classes and fifty fine classes based on two-layered taxonomy. Primitive features such as words, POS tag, chunks, name entities, head chunks and semantically related words were identified from the questions and analysed to determine the "type" of answer to be expected. It had achieved 98.8% precision for coarse classes and 95% for fine classes. They had proposed to investigate further the application of deeper semantic analysis to feature extraction. Later, Li et al.[19] had used more semantic information sources including name entities, WordNet senses, class-specific related words, and distributional similaritybased categories in question classification. It had achieved best accuracy of 89.3% on a test set of 1000 questions with training set of 21,500 questions.

Zhang et al.[20] did a research on automatic question classification with five types of machine learning algorithms: Nearest Neighbours, Naïve Bayes, Decision Tree, Sparse Network of Winnows, and Support Vector Machine (SVM). The experiment results showed that the SVM obtained the best accuracy which was 79.2% comparing to the other four methods. The SVM

TEST Engineering & Management
January - February 2020
ISSN: 0193 - 4120 Page No. 4374 - 4385

based on tree kernel was applied in this research. It could bring about 20% errors reduction. However, the accuracy could be improved only with large training set.

In 2008, Huang et al.[21]had proposed to classify question with machine learning approach (SVM and Maximum Entropy models) based on word feature and semantic feature extracted from Wh-questions. The SVM and Maximum Entropy Models were used as a classifier in this research. Once the head words were extracted, Hypernyms feature from WordNet of head words were used in classification. Promising result with 89.2% accuracy was obtained using SVM.

Kusuma et al.[22]had suggested a method to classify Indonesian language question items automatically based on the revised Bloom Taxonomy levels. Keywords were extracted based on lexical and syntactical features. The SVM algorithm was implemented to classify the questions with better accuracy result. Lexical feature was done by counting all the wh-words and question length in the sentence. Keyword feature extraction was done by counting all frequency of keywords in questions. POS tagging feature extraction was done by counting words amount of verb, adjective, noun and symbol used in the questions. The features extraction output was classified by using the algorithm of Support Vector Machine (SVM).

Osman et al.[6] had compared different machine learning methods: Naïve Bayes, Support Vector Machine (SVM), logistic regression and decision trees to automatically classify examination questions based on the cognitive levels of Bloom's taxonomy. A total of 600 computer science examination questions were collected. Then each question was tokenized after converting the question to lower case and removing the punctuation. After the NLP processes, a computationally easier term selection approach was used to select the most representative terms in the question. Subsequently, a series of experiments comparing and evaluating different machine learning algorithms to classify the examination questions were carried out. The SVM with unigram features gave the best classification result which was 0.7667.

According to the researches of Li et al.[18][19], Zhang et al.[20], andHuang et al.[21], manual construct set of rules to map with question type was not efficient in maintaining and upgrading. Machine learning was more flexible to reconstruct because it could be trained on a new taxonomy in short time comparing to rules based classification.

According to Osman et al.[6], significant result could be obtained with a huge amount of data available in the research. However, it was a challenge to get enough data for the experiment especially in collecting examination questions. Researchers proposed also to investigate further the semantic knowledge in question classification.

In 2010, Cutrone et al.[23] had proposed a system in automating the assessment process of open questions by using POS (Part-of-Speech) tagging and WordNet database to evaluate students' answer. The system was able to process the answer containing single sentence. The testing result was promising. However, the evaluation did not consider the depth of knowledge expected. Answer might be mis-graded when the given answer was more technical while the supplied answer was in opposite, non-technical.

Jayakodi et al.[24] had proposed a rule-based examination question classification with WordNet and cosine similarity algorithm in 2015. Questions were processed using NLP techniques such as tokenization, lemmatization and tagging before applying semantic analysis technique. Verbs were extracted from the questions and used to identify the similarities comparing with the taxonomy verbs. However, WordNet similarity algorithm alone could not produce a more

January - February 2020
ISSN: 0193 - 4120 Page No. 4374 - 4385

accurate classification result. Thus, cosine similarity was used to improve the accuracy of the classification. Based on the identified tag patterns in the questions, the questions tag pattern and the matching tag pattern in the database were tested to identify the cosine similarity of the pattern. With the hybrid of syntactic and semantic approaches, it had successfully classified the questions into the correct Bloom’s categories. However, it was suggested that the outcome of the research could be improved further after analysing a large number of examination questions from different disciplines.

Recently, deep learning had been used in NLP for designing question answering system. Many works were using neural network based to find word embedding that captured the similarity among words. Mikolov et al.[25] had proposed a word embedding model named Word2Vec to find word representations which was good for predicting certain words using its surrounding words. The Word2Vec and GloVe were two most frequently used in word embedding.

Minaee et al.[26] had proposed a model for question answering using a deep similarity network. The research started with learning a vector representation of questions and answers using Doc2Vec approach. Then, a deep similarity neural network was trained with large scale of public questions and answers database to find the similarity score of pair of answer and question. Finally, the network was used to find the answer to a given question by search over a set of answer candidates by retrieving the highest similarity score. Good performance was obtained using this approach. However, large-scale of database was needed to train the model using a Deep Similarity Neural Network. It would be challenging to implement this approach in examination questions classification unless there was a big pool of questions in the same area that could be collected.

Overall, rule-based approaches are commonly used in examination questions classification. However, the questions are restricted to certain syntactic pattern and more rules are to be set to achieve better accuracy. It is not easy to maintain a huge set of rules. Thus, machine learning has been proposed by researchers to achieve better result in examination questions classification. The SVM has been the popular approach used in examination classification. However, the accuracy can only improve with large training set of examination questions. In order to develop a system that can classify examination questions from different areas or subjects, the system needs to be trained with huge number of questions from each area or subject the system involved before it can be implemented.

Many researchers have suggested to research further using semantic approach. The hybrid approach using WordNet and Cosine similarity proposed by Jayakodi et al.[24]in examination questions classification has successfully classified the questions into the correct Bloom’s categories. However, the research focuses in one discipline only.

In real implementation, apart from being able to classify questions to the correct Bloom’s Taxonomy, the questions classification system should fulfil two requirements. First, system should be able to classify questions based on different areas in order to provide full coverage on all modules offered. Second, it should be able to provide accurate result even in small number of questions training set. Number of examination questions for each module is small compare to huge amount in questions classification in search engine. Thus, questions classification with semantic and syntactic approach using WordNet and Stanford Parser with Universal Dependency is proposed to classify the examination questions based on Bloom’s Taxonomy.

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 4374 - 4385

3 Research Methodology

In this research, the NLP's syntactic rule-basedand semantic approach is adopted in classifying examination questions into correct Bloom's Taxonomy cognitive levels. The Stanford Parser Universal Dependency is used to identify the keywords in the questions while WordNet similarity is used to measure the semantic similarity between the keywords from the questions and the Bloom's Taxonomy keywords. The measured semantic similarity result is used to classify the questions based on the Bloom's Taxonomy.

An overview diagram of the proposed question classification approach is illustrated in Figure 1. The approach uses pipeline interactionthatstarts with Questions Extraction, thenQuestion Segmentation using Natural Language Toolkit (NLTK)'s sentence and word tokenization, followed by Verb Extraction using POSTagging and StopWord. The extracted verb is then fed into the Verb vs Bloom Verbs to obtain similarity value using WordNet similarity. The Keyword Extraction will extract the keywords using Stanford Universal Dependency and then feed the keywords to component – Keywords bs Bloom Verbs to obtain the WordNet similarity result. The final component is Question Classification to Bloom Taxonomy where the similarity result of the verb and keywords are summarized and identified the Bloom Taxonomy levels.

200 final examination questions of an engineering school with different modules are collected as testing set. The testing set are classified manually by a group of subject experts who has more than 10 years of experience in teaching the Engineering subjects. Then, the proposed framework is used to classify the testing set questions automatically based on the Bloom's Taxonomy using the NLP's semantic and syntactic approach - WordNet and Stanford Parser with Universal Dependency.

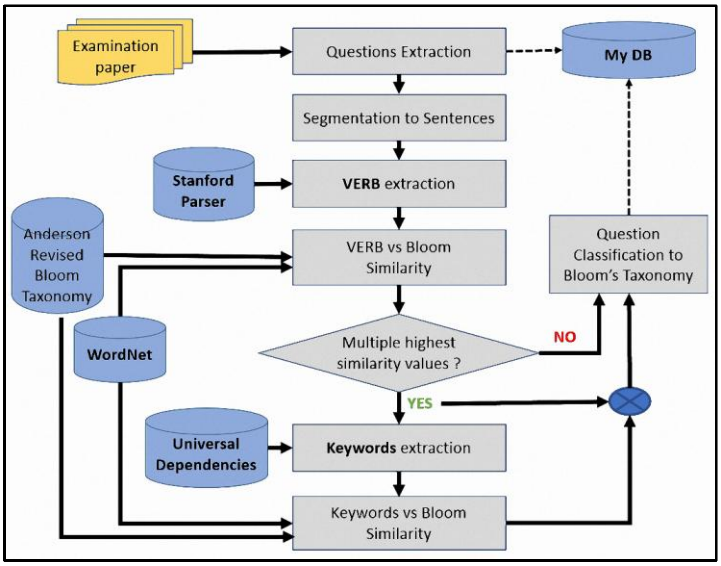

Figure 1. Block diagram of the proposed system with hybrid approach in Question Classification.

3.1Question Extraction

An input to the framework is a set of final examination questions. The questions can be a single sentence type of questions like Q2 example in Table 2, or multiple sentences type of questions like example Q1 in Table 2. The multiple sentences type of questions needs to be segmented into individual sentences. The Q1 is a question consisting two sentences that needs to be broken down to S1 and S2. The question segmentation is using Natural Language Toolkit (NLTK) 's sentence tokenization. NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to corpora and lexical resources such as WordNet and Stanford Parser, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing and etc.

Table 2.Q1 is multiple sentences type of question while Q2 is single sentence type of question.

Q1: A transverse traveling wave on a cord is represented by D = 0.22 m sin (5.6 radm-1 x + 34 rads-1 t), where D and x are in meters and t is in seconds. For this wave, compute the amplitude, frequency and wavelength. **Q1S1**: A transverse traveling wave on a cord is represented by D = 0.22 m sin (5.6 radm-1 x + 34 rads-1 t), where D and x are in meters and t is in seconds. **Q1S2**: For this wave, compute the amplitude, frequency and wavelength.
Q2: Explain your design of circuit.

January - February 2020
ISSN: 0193 - 4120 Page No. 4374 - 4385

3.2Verb Extraction

After tokenization and tagging, the verbs of each question's sentences are extracted. The verbs in the sentences are the words tagged with "VB". As for S1 and S2 sentences, the verb to be extracted is "compute" only. Then the extracted verbs from the sentences are compared with the verbs in Bloom's verbs list using WordNet similarity.

The Bloom's verbs list is listed in Table 1. In this research, the lowest Bloom's Taxonomy level – "remembering" is labelled as Bloom's Taxonomy Level 1; next level – "understanding" is labelled as Level 2, and so on until the highest level – "creating" as Level 6. Thus, the extracted verbs are compared with each Bloom's verb of each Bloom's Taxonomy level.

3.3WordNet Similarity

The WordNet similarity outputs a similarity value when comparing an extracted verb with a Bloom's verbs. The similarity value is ranging from 0.000 to 1.000. The higher similarity value indicates the higher similarity between the comparing verbs. The similarity value of 1.000 means the comparing verbs being identical or exact same verb. On the other hand, the similarity value of 0.000 means both comparing verbs being not related to each other.

Figure 2 is the result of WordNet similarity when comparing the only extracted verb – "compute" with the Bloom's verbs by Bloom's Taxonomy levels. The illustrated similarity values are the highest similarity value of each Bloom's Taxonomy level. Then, the Bloom's Taxonomy level with highest similarity value is classified as the extracted verb's Bloom's Taxonomy level as well as the Bloom's Taxonomy level of the question. Figure 2 showed the question Q1 being classified as Bloom's Taxonomy level 3 – "applying".

```
***********************************************
['compute']
***********************************************

words = compute - Bloom1     maxvalue List1 =  0.333

words = compute - Bloom2     maxvalue List2 =  0.857

words = compute - Bloom3     maxvalue List3 =  1.000

words = compute - Bloom4     maxvalue List4 =  0.857

words = compute - Bloom5     maxvalue List5 =  0.333

words = compute - Bloom6     maxvalue List6 =  0.667
```

Figure2. WordNet similarity result when comparing the verb – "compute" with Bloom's verbs. The result maps the "compute' verb to Bloom's Taxonomy level – "applying".

Some verbs of the Bloom's verbs list are existing in multiple Bloom's Taxonomy levels. For example, the Bloom's verb – "compare" is found in Bloom's Taxonomy Level 4 and Level 5 also. If an extracted verb has the highest WordNet similarity value with this Bloom's verb – "compare", the extracted verb and then the question should be classified as the Bloom's Taxonomy Level 4 or Level 5. As for this scenario, the Stanford Universal Dependency (UD) approach is used to justify the correct Bloom's Taxonomy level.

3.4Universal Dependency (UD)

The Universal Dependency (UD) is provided by the Stanford Parser also. It outputs a dependency tree that is widely used in Natural Language Processing. It relates a head word to a dependent word based on a uniform notation of triples. According to Natalia Silveira, the Stanford Parser produces high quality dependency annotations in assessing the accuracy of the automatic dependency conversion tool [27].

The question Q2 in Table 2 is an example which the only extracted verb – "Explain" gets the same highest WordNet similarity value, 1.000 in Bloom's Taxonomy in Level 2, 5 and 6, as illustrated in Figure 3. Then the Q2 is input to the Universal Dependency (UD) to output a dependency tree as illustrated in Figure 4. The

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 4374 - 4385

extracted verb normally is the root word or the head word, and the direct dependent word is “design”. This direct dependent word is noted as keyword.Once the keyword is identified by the UD, the keyword is fed into WordNet to find the similarity values with the Bloom’s verbs of each Taxonomy level.

```
************************************************
['explain']
************************************************

words = explain - Bloom1     maxvalue List1 =  0.799
words = explain - Bloom2     maxvalue List2 =  1.000
words = explain - Bloom3     maxvalue List3 =  0.727
words = explain - Bloom4     maxvalue List4 =  0.833
words = explain - Bloom5     maxvalue List5 =  1.000
words = explain - Bloom6     maxvalue List6 =  1.000
```

Figure3. WordNet similarity result when comparing the verb – “explain” with Bloom’s verbs. Bloom Level 2, 5 and 6 have the maximum similarity value of 1.

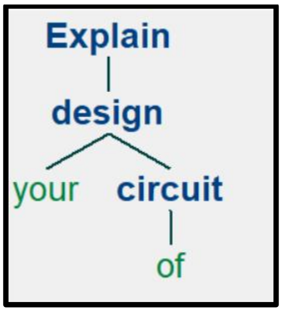

Figure4. Stanford Universal Dependency generates the dependency tree indicating the root word – “Explain” and then the dependant word – “design”.

As a result, there are 2 sets of Wordnetsimilarity values obtained. First set is the similarity values of the verb – “explain” with the Bloom’s verbs that obtained earlier. Second set is the similarity values of thekeyword – “design” with the Bloom’s verbs. The total similarity value of each Bloom level is calculated by summing the same Bloom’s level similarity value from first set data and second set data.The Bloom level with the highest total similarity value is classified as Bloom level of the question.

Figure 5 is theWordNet similarity result of the extractedverb – “explain”(first set) andthe keyword – “design” (second set). The total similarity value is calculated by summing the value of the same Bloom’s level. The total similarity value is 1.625, 1.600 and 2.000 with respective to Bloom’s level 2, 5 and 6. The Bloom’s level 6 obtained highest total value and then the Q2 question is classified as Bloom’s Taxonomy Level 6.

```
************************************************
['explain', 'design']
************************************************

words = explain - Bloom2     maxvalue List2 =  1.000
words = explain - Bloom5     maxvalue List5 =  1.000
words = explain - Bloom6     maxvalue List6 =  1.000

words = design - Bloom2     maxvalue List2 =  0.625
words = design - Bloom5     maxvalue List5 =  0.600
words = design - Bloom6     maxvalue List6 =  1.000

List sum =  [1.625, 1.600, 2.000]
```

Figure5. WordNet similarity result when comparing the dependant word – “difference” with Bloom’s verbs. The Bloom’s Taxonomy Level 6 has the highest value.

3.5Question Classification

The framework is designed to get the similarity value of the verb extracted from the question with the Bloom’s verb list. The Bloom’s level with highest similarity value is then classified as the question’s Bloom level. If multiple Bloom’s levels getting same highest similarity value, the keyword of the question needs to be identified and then the keyword’s similarity value. The system sums both verb and keyword similarity values and classified the Bloom’s level with highest sum value as the question’s Bloom’s level.

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 4374 - 4385

4 Conclusion

Question Classification (QC) based on Bloom's Taxonomy has been widely accepted and used as a guideline for all Science, Technology, Engineering, and Mathematics (STEM) offered by the Technical and Vocational Education and Training (TVET) in Malaysia. However, many discrepancies happened in QC due to inconsistence or misclassification of questions to Bloom's level. This research proposes a framework to classify the questions based on Bloom's Taxonomy with a hybrid of syntactic and semantic Natural Language Processing (NLP) approach using WordNet and Stanford Parser with Universal Dependency. The proposed framework has demonstrated the hybrid approach capable to classified question with the Bloom Taxonomy accurately. The system needs to be evaluated further with the suggested 200 final examination questions for the accuracy check and enhancement.

Acknowledgement

The authors acknowledge the financial support on the publication of this paper by the Universiti Pertahanan Nasional Malaysia (UPNM) and KDU University College, Utropolis Glenmarie. Postgraduate Research Centre (PGRC). Malaysia.

References

[1] Mat Isa, C.M., M.Saman, H., Tahir, T., & Mukri, M.: Understanding of outcome-based education (OBE) implementation by civil engineering students in Malaysia. 2017 IEEE 9th International Conference on Engineering Education (ICEED), 96-100 (2017)

[2] MQA (2017): Code of Practice for Programme Accreditation 2nd Edition. Malaysian Qualification Agency (MQA), Cyberjaya (2017)

[3] Omar, N., Haris, S.S., Hassan, R., Arshad, H., Rahmat, M., Zainal, N.F.A.,& Zulkifli, R.: Automated Analysis of Exam Questions According to Bloom's Taxonomy. Procedia - Social and Behavioral Sciences, 59, 297-303 (2012)

[4] Bloom, B.S.: Taxonomy of Educational Objectives. Vol. 1: Cognitive Domain. New York: McKay (1956)

[5] Yusof, N., & Chai, J.H.: 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 (2010)

[6] Osman, A., &Yahya, A.: Classifications of Exam Questions using linguistically-Motivated Features: A case study based on Bloom's Taxonomy. The Sixth International Arab Conference on Quality Assurance in Higher Education (IACQA), 467-474 (2016)

[7] Tutkun, O.F., Guzel, D., Koroğlu, M., & Ilhan, H.: Bloom's Revized Taxonomy and Critics on It. The Online Journal of Counselling and Education 1 (3), 23-30 (2012)

[8] Anderson, L.W., & Krathwohl, D.R.: A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. New York : Longman (2001)

[9] Nadkarni, P.M., Ohno-Machado, L., & Chapman, W.W.: Natural language processing: an introduction. Journal of the American Medical Informatics Association (JAMIA), 18(5), 544-551 (2011)

[10] Sangodiah, A., Muniandy, M., & Heng, L.E.: Question classification using statistical approach: A complete review. Journal of Theoretical and Applied Information Technology, 71(3), 386-395 (2015)

[11] Pinto, D., Branstein, M., Coleman, R., Croft, W.B., King, M., Li, W., & Wei, X.: QuASM: A System for Question Answering Using Semi-Structured Data. Proceding of ACM/IEEE Joint Conference on Digital Libraries (JCDL), 46-55 (2002)

[12] Chang, W.C., & Chung, M.S.: Automatic applying Bloom's taxonomy to classify and analysis the cognition level of English question items. 2009 Joint Conferences on Pervasive Computing (JCPC), 727-734 (2009)

[13] Haris, S.S., & Omar, N.: A rule-based approach in Bloom's Taxonomy question classification through natural language processing. 2012 7th International Conference on Computing and

Convergence Technology (ICCCT), 410-414 (2012)
[14] Biswas, P., Sharan, A., & Kumar, R.: Question Classification using syntactic and rule based approach. 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 1033-1038 (2014)
[15] Ramesh, R., Mukundan, S., & Iyer, S.: Annotating the Domain Ontology of a Course with ITS Syllabus and Learning Objectives. 2016 International Conference on Learning and Teaching in Computing and Engineering (LaTICE), 130-131 (2016)
[16] Samuel, A. L.: Some Studies in Machine Learning Using the Game of Checker. IBM Journal of Research and Development, 3(3), 200-229 (1959)
[17] Mitchell, T.: Machine Learning. McGraw Hill, 2 (1997)
[18] Li, X., & Roth, D.: Learning question classifiers. COLING 2002: The 19th International Conference on Computational Linguistics, 1, 1-7 (2002)
[19] Li, X., & Roth, D.: Learning question classifiers: the role of semantic information. Natural Language Engineering, 12(3), 229-249 (2006)
[20] Zhang, D., & Lee, W.S.: Question classification using support vector machines. SIGIR '03 Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval, 26-32 (2003)
[21] 21.Huang, Z., Tint, M., & Qin, Z.: Question Classification using Head Words and their Hypernyms. 2008 Conference on Empirical Methods in Natural Language Processing (EMNLP), 927-936 (2008)
[22] Kusuma, S.F., Siahaan, D, & Yuhana, U.L.: Automatic Indonesia's questions classification based on bloom's taxonomy using Natural Language Processing a preliminary study. 2015 International Conference on Information Technology Systems and Innovation (ICITSI), 1-6 (2015)
[23] Cutrone, L.A., & Chang, M.: Automarking: Automatic Assessment of Open Questions. 2010 10th IEEE International Conference on Advanced Learning Technologies, 143-147 (2010)
[24] Jayakodi, K., Bandara, M., & Perera, P.: An automatic classifier for exam questions in Engineering: A process for Bloom's taxonomy. 2015 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), 297-303 (2015)
[25] Mikolov, T., Chen, K., Corrado, G., & Dean, J.: Efficient Estimation of Word Representations in Vector Space. Proceedings of the International Conference on Learning Representations (ICLR 2013), 1-12 (2013)
[26] 26.Minaee, S., & Liu, Z.: Automatic question-answering using a deep similarity neural network. 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), 923-927 (2017)
[27] Silveira, N., Dozat, T., Marneffe, M., Bowman, S., Connor, M., Bauer, J., & Maning, C.: A Gold Standard Dependency Corpus for English. Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014), 2897-2904 (2014)