THE MODEL OF EXAMINATION QUESTION CLASSIFICATION BASED ON BLOOM'S TAXONOMY USING SEMANTIC SIMILARITY TECHNIQUE

GOH THING THING

DOCTOR OF PHILOSOPHY
(COMPUTER SCIENCE)

UNIVERSITI PERTAHANAN NASIONAL MALAYSIA

2024

THE MODEL OF EXAMINATION QUESTION CLASSIFICATION BASED ON BLOOM'S TAXONOMY USING SEMANTIC SIMILARITY TECHNIQUE

GOH THING THING

This thesis has been submitted to the Centre for Graduate Studies, Universiti Pertahanan Nasional Malaysia in fulfilment of the requirements for the Doctor of Philosophy (Computer Science).

January 2024

ABSTRACT

Bloom's Taxonomy (BT) has generally been used as a guideline in designing a holistic set of examination questions that comprise various cognitive levels. It has been emphasised by Engineering Accreditation Council Malaysia (EAC) and Malaysian Qualifications Agency (MQA) to regulate the quality and standard of education provided by setting the assessment questions aligned with the Course Learning Outcomes (CLO). However, there are inconsistencies in the classification of final examination questions based on Bloom's Taxonomy. This is because it is manually conducted by academics and is susceptible to discrepancies in the understanding of BT among academics. Most of the research work focused on single-sentence questions that were not based on real examination questions. While previous research has explored examination question classification using a semantic approach, it encountered challenges in achieving high accuracy, which is greater than 80%. Therefore, this research aims to introduce a model to perform examination question classification based on BT using a semantic approach with real examination question and striving to attain an accuracy exceeding 80%. A Question Classification Model (QCM) was developed in this research using Natural Language Processing (NLP) approaches, such as the Stanford POS (Part-Of-Speech) tagger, to preprocess the examination questions into word tokens. Subsequently, Stanford Parser Universal Dependencies (UD) was used to identify the important verbs in the examination questions that reflect the thinking action. This was followed by a comparison between the identified verbs and the list of BT verbs using the WordNet Similarity approach. Moreover, this research has studied, evaluated and enhanced each approach to achieve the best performance for the QCM. Overall, the developed QCM achieved a recorded

accuracy rate of 83% in the classification of a set of 200 examination questions based on BT. This research helps to control the assessment quality to meet the classification and fulfil the requirements of Outcome-Based Education (OBE) standards.

ABSTRAK

Klasifikasi soalan berdasarkan Taksonomi Bloom (BT) telah diterima secara meluas dan digunakan sebagai garis panduan dalam penyediaan soalan peperiksaan holistik yang terdiri daripada pelbagai domain kognitif. Pentingnya penggunaan BT telah ditekankan oleh Majlis Pengiktirafan Kejuruteraan Malaysia (EAC) dan Agensi Kelayakan Malaysia (MQA) dalam mengawal standard dan kualiti pendidikan dengan menentukan soalan penilaian yang selaras dengan Hasil Pembelajaran Kursus (CLO). Walau bagaimanapun, terdapat banyak percanggahan pendapat dalam pengelasan soalan berdasarkan BT sekiranya pengelasan dilakukan secara manual oleh staf akademik. Selain itu, kebanyakan kerja penyelidikan tertumpu kepada soalan tunggal yang tidak berasaskan soalan peperiksaan yang sebenarnya. Sementara itu, penyelidikan terdahulu telah meneroka pengelasan soalan peperiksaan menggunakan pendekatan kesamaan. Tetapi, ianya menghadapi cabaran dalam mencapai kejituan yang tinggi, iaitu melebihi 80%. Oleh itu, penyelidikan ini bertujuan untuk memperkenalkan model kerja pengelasan soalan peperiksaan berdasarkan BT menggunakan pendekatan semantik yang menggunakan soalan peperiksaan sebenar, di samping mencapai kejituan melebihi 80%. Model Kerja Pengelasan Soalan (QCM) telah dibangunkan dengan menggunakan pendekatan Pemprosesan Bahasa Semula Jadi (NLP), seperti pemberian tanda Stanford POS (Part-Of-Speech), untuk memproses soalan peperiksaan kepada token perkataan. Seterusnya, Stanford Parser Universal Dependencies (UD) digunakan untuk mengenal pasti kata kerja penting dalam soalan peperiksaan yang mencerminkan tindakan berfikir. Ini diikuti dengan perbandingan antara kata kerja yang dikenal pasti daripada soalan peperiksaan dan senarai kata kerja BT menggunakan pendekatan

Kesamaan WordNet. Selain itu, penyelidikan ini mengkaji, menilai dan memperbaiki setiap pendekatan untuk mencapai prestasi terbaik untuk QCM. Secara keseluruhannya, QCM yang dibangunkan mencapai kadar kejituan sebanyak 83% dalam pengelasan set soalan peperiksaan berdasarkan BT. Kajian ini membantu mengawal kualiti penilaian untuk memenuhi pengelasan dan mematuhi keperluan piawaian Pendidikan Berasaskan Hasil (OBE).

ACKNOWLEDGEMENTS

This dissertation is born out of a journey that began in 2017. This journey would not have been possible without the support of my family, supervisors and friends.

First of all, I would like to thank my family for the full support given all these years. I am especially grateful to my hubby, who supported me emotionally and financially.

I would like to express my gratitude to Dr Nor Azliana Akmal binti Jamaludin, my esteemed supervisor, for all the guidance, support and motivation she provided to me throughout my doctoral studies. I could not have imagined having a better advisor on my journey.

My gratitude also goes to Dr Hassan bin Mohamed and Prof. Dr Mohd Nazri bin Ismail for their insightful comments and encouragement in helping me succeed in my studies. I am also grateful to Prof. Dr Aidy Ali, support staff from UPNM and my friends. Without their support over the past few years, I would not have been able to finish my studies.

Last but not least, I am extending my gratitude to the Faculty of Defence Science and Technology, UPNM for the funding opportunity to undertake my studies.

APPROVAL

The Examination Committee has met on 4th October 2023 to conduct the final examination of Goh Thing Thing on her degree thesis entitled "The Model of Examination Question Classification based on Bloom's Taxonomy using Semantic Similarity Technique".

The committee recommends that the student is awarded the Doctor of Philosophy (Computer Science).

Members of the Examination Committee are as follows:

PROF. TS. DR. OMAR BIN ZAKARIA
Faculty of Defence Science and Technology
Universiti Pertahanan Nasional Malaysia
(Chairman)

PROF. MADYA DR. MOHD RIZAL BIN MOHD ISA
Faculty of Defence Science and Technology
Universiti Pertahanan Nasional Malaysia
(Internal Examiner)

PROF. DR. MOHD NORDIN BIN ABDUL RAHMAN
Fakulti Informatik dan Komputeran
Universiti Sultan Zainal Abidin
(External Examiner)

PROF. MADYA DR. SITI SOPHIAYATI BINTI YUHANIZ
Razak Faculty of Technology and Informatics
Universiti Teknologi Malaysia
(External Examiner)

APPROVAL

This thesis was submitted to the Senate of Universiti Pertahanan Nasional Malaysia and has been accepted as fulfilment of the requirements for the degree of the **Doctor of Philosophy (Computer Science)**. The members of the Supervisory Committee were as follows.

Dr. Nor Azliana Akmal binti Jamaludin
Fakulti Sains Komputer dan Teknologi
Universiti Pertahanan Nasional Malaysia
(Main Supervisor)

Dr. Hassan bin Mohamed
Fakulti Sains Komputer dan Teknologi
Universiti Pertahanan Nasional Malaysia
(Co-supervisor)

Prof. Dr. Mohd Nazri bin Ismail
Fakulti Sains Komputer dan Teknologi
Universiti Pertahanan Nasional Malaysia
(Co-supervisor)

UNIVERSITI PERTAHANAN NASIONAL MALAYSIA

DECLARATION OF THESIS

Student's full name : GOH THING THING

Date of birth : 29 JULY 1976

Title : THE MODEL OF EXAMINATION QUESTION CLASSIFICATION BASED ON BLOOM'S TAXONOMY USING SEMANTIC SIMILARITY TECHNIQUE

Academic session : 2017/2018

I hereby declare that the work in this thesis is my own except for quotations and summaries which have been duly acknowledged.

I further declare that this thesis is classified as:

☐ **CONFIDENTIAL** (Contains confidential information under the official Secret Act 1972)*

☐ **RESTRICTED** (Contains restricted information as specified by the organisation where research was done)*

☐ **OPEN ACCESS** I agree that my thesis to be published as online open access (full text)

I acknowledge that Universiti Pertahanan Nasional Malaysia reserves the right as follows.

1. The thesis is the property of Universiti Pertahanan Nasional Malaysia.
2. The library of Universiti Pertahanan Nasional Malaysia has the right to make copies for the purpose of research only.
3. The library has the right to make copies of the thesis for academic exchange.

_____ _____

Signature **Signature of Supervisor/Dean of CGS/

_____ _____

IC/Passport No. **Name of Supervisor/Dean of CGS/

Date: Date:

Note: *If the thesis is CONFIDENTIAL OR RESTRICTED, please attach the letter from the organisation stating the period and reasons for confidentiality and restriction.

** Witness

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES

LIST OF ABBREVIATIONS

UPNM	Universiti Pertahanan Nasional Malaysia
BT	Bloom's Taxonomy
BV	Bloom's Verb
CLO	Course Learning Outcomes
ERR1	Error 1
ERR2	Error 2
ML	Machine Learning
ME	Maximum Entropy
NLP	Natural Language Processing
RW	Root Word
JCN	Jiang, Conrath measure
KW	Keyword
LCH	Leakcock and Chodorow
left3words	"english-left3words-distsim" model
LIN	Lin measure
POS	Part-Of-Speech
PLO	Programme Learning Outcomes
QC	Question Classification
QCM	Question Classification Model
QP	Question Pre-processing
SVM	Support Vector Machine
TP	True Positive
UD	Stanford Parser Universal Dependencies
VE	Verbs Extraction
WUP	Wu and Palmer's Similarity

CHAPTER 1

INTRODUCTION

This chapter introduces Outcome-based Education which focuses on the outcomes that are commonly classified based on Bloom's Taxonomy (BT). This is followed by the examination questions which are a common tool used to access the outcomes. Thus, the examination questions should be classified based on BT and in sequence to align with the outcome classification also. However, various discrepancies occur in question classification due to the inconsistency of questions in BT.

1.1 Problem Background

Outcome-based Education (OBE) is an education theory that focuses on outcomes – what students should learn and their ability to apply the lesson after the learning process (Qadir et al., 2020). To ensure that the students achieve the defined outcomes, Malaysian Qualifications Agency (MQA) has implemented a Malaysia Qualifications Framework (MQF) to regulate the quality and standard of higher education provided. This model is defined as an instrument in developing and

classifying qualifications based on a set of criteria that have been nationally agreed upon since 2008.

Five learning outcome clusters have been defined and used as a qualification guideline for all Technical and Vocational Education and Training (TVET) to obtain their programmes accreditation by MQA. Each programme is structured for a specified duration and learning volume to achieve the stated Programme Learning Outcomes (PLO). These factors contribute to an award of a qualification, such as a diploma certification, a Bachelor's degree, or a Master's degree among others.

The PLOs are the specified knowledge, skills, attitude, and abilities to be acquired and demonstrated by students upon graduation (MQA, 2017). All PLOs should be aligned to the learning outcome clusters set by MQA. The achievement of PLOs is based on the Course Learning Outcomes (CLOs) of each module in the programme. The achievement of CLOs is heavily dependent on the score obtained in each assessment component.

The assessment components are the final examination, tests, projects and assignments distributed to students within a semester. To achieve the CLOs followed by PLOs, the assessment components should be designed in line with the CLOs set for the modules. Furthermore, the questions set in each assessment component plays an important role in assisting students in attaining the targeted CLOs where each CLO is set with different thinking order based on Bloom's Taxonomy (BT). Specifically, lower-order thinking questions emphasise foundation skills and practices, while

higher-order thinking questions involve more complex thinking and creative problem-solving.

Established by Benjamin Bloom's during the 1950s, BT is a method of categorising the levels of reasoning skills required in classroom situations. Six levels in the BT are present, with each requiring a higher level of abstraction from the students. Notably, BT has been widely used as a guideline in designing a holistic examination question, which consists of various cognitive levels (Omar et al., 2012). Teachers are recommended to move students up the BT levels as they progress in their knowledge (Forehand, 2010).

The examination questions classification shall be aligned with the CLO which was defined during the curriculum design. Each CLO is specified with an action verb that corresponds to a Bloom's Taxonomy (BT) verb. Example of a course learning outcome, "Describe concept, principles and theories relating to area of physics and engineering". The action verb "Describe" aligned with the BT verbs from level 2 which is "Understanding". Consequently, academician shall design the examination question which is able to reveal the achievement of the respective course learning outcome.

The current practice in most universities, academicians manually categorise the examination questions for each assessment into the BT levels based on their understanding of BT levels. The question classification was done manually which may vary from one academician to another (Yusof & Chai, 2010; Sulaiman et al., 2020; Jayakodi et al., 2016). Academicians with a clear understanding of the

cognitive levels of BT are able to create sets of examination questions in line with the targeted CLOs. However, most academicians are not familiar with the proper implementation of the cognitive levels of BT (Contreras et al., 2021; Kumara et al., 2019).

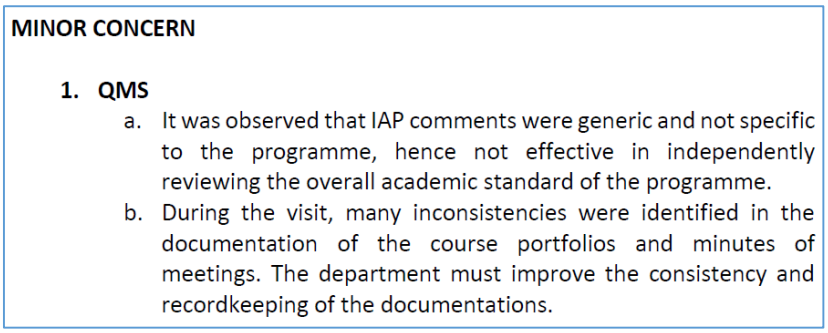
When formulating examination questions, some of the academicians do not follow the prescribed list of BT verbs as a reference. For example, a question such as "Determine the mass and weight of the air contained in a room with dimensions of 6m x 6m x 8m." has the action verb "Determine" that is not in the BT verb list. Therefore, academicians may classify such questions into Bloom's levels based on their individual interpretation, leading to the inconsistency in the classification process.

In another case, the question "Calculate the potential difference between points a and b in Figure 3." was supposed to classify to third level of BT, "Applying", given that it necessitates students to perform calculations for the potential difference, however, it was classified to the fourth level of BT, "Analysing". The inconsistency in classification was the result of the academician's interpretation, which expected students to engage in an analytical process by referring to the provided diagram.

In addition to the misclassification due to lack of understanding of OBE implementation, ambiguity in BT verbs can also contribute to misclassification. The classification of questions based on BT is commonly based on the verbs used in the examination questions. The verbs are extracted from the examination questions and mapped to the list of verbs in BT. However, some of Bloom's verbs are ambiguous

when the verbs fall into more than one level of BT (Osadi et al., 2017; Das et al., 2020; Shaikh et al., 2021). Therefore, issues are present in the BT classification of examination questions with these verbs that caused inconsistent BT classification by different academicians in the same examination questions.

As a result of the misclassification, concern was raised in the accreditation summary report 2022 for Bachelor of Mechanical Engineering, UOW Malaysia KDU University College by Engineering Accreditation Council (EAC), as shown in Figure 1.1.



MINOR CONCERN

1. **QMS**
   a. It was observed that IAP comments were generic and not specific to the programme, hence not effective in independently reviewing the overall academic standard of the programme.
   b. During the visit, many inconsistencies were identified in the documentation of the course portfolios and minutes of meetings. The department must improve the consistency and recordkeeping of the documentations.

Figure 1. 1 Accreditation Feedback from EAC for Bachelor of Mechanical Engineering, UOW Malaysia KDU UC

In Malaysia's Institutions of Higher Learning (IHL), the common practice involves appointing one or two academicians as moderators during the examination question moderation process to reduce the inconsistency in question classification. The moderators are responsible to verify the alignment of the examination questions with the CLOs for the respective modules. Feedback is provided to the module lecturer, who is then required to revise the examination questions to ensure all examination questions align with the CLOs. After the revision, the examination paper undergoes a second round of review by a different moderator. This process has

effectively minimised the misclassification of examination questions, however, it is time-consuming.

In recent years, there has been an increase of interest among the researchers in automating the examination question classification based on BT cognitive domain. Various techniques have been used such as syntactic features, which focused on the structure and the pattern of the question and semantic features which consider the meaning of significant verbs identified from the question. Natural Language Processing (NLP) was commonly used to identify the action verb from the examination question. Machine learning or semantic similarity was used to classify the examination question. Some researchers (Sangodiah et al., 2017; Sulaiman et al., 2020; Contreras et al., 2021) have employed machine learning and semantic similarity (Jayakodi et al., 2016; Diab & Sartawi, 2017) to classify questions, but machine learning requires large question datasets for accurate results. This led researchers to suggest using a semantic approach for better outcomes. However, the semantic approach alone has not achieved a good result. Thus, researchers continue to use both machine learning and semantic approaches in their research (Mohammed & Omar, 2020; Mohamed et al., 2019).

Gathering a substantial set of examination questions can be challenging and most research studies have not used the real-exam question. They mostly focused on single-sentence question (Jayakodi et al., 2016; Sangodiah et al., 2017; Mohammed & Omar, 2018). However, in practice, exam questions are often longer and more complex, particularly those involving case studies or scenarios with higher thinking order.