Center for Open Access in Science ▪ https://www.centerprode.com/ojit.html
Open Journal for Information Technology, 2021, *4*(2), 77-84.
ISSN (Online) 2620-0627 ▪ https://doi.org/10.32591/coas.ojit.0402.04077n

A Literature Review on Automatic Generation of Examinations

Peter Ndegwa Ndirangu & Elizaphan Maina Muuro

The Kenyatta University, Nairobi, KENYA
Computing and Information Technology Department

John M. Kihoro

The Co-operative University of Kenya, Nairobi, KENYA
Directorate of Computing and eLearning

Received: 28 September 2021 ▪ Accepted: 3 December 2021 ▪ Published Online: 30 December 2021

Abstract

The examination is a key activity in determining what the learner has gained from the study. Institutions of higher learning (IHL) perform this activity through various assessment methods (test/examination, practical, etc.). The world today is focused on automation of exam generation which is ongoing with dire need during this period of the COVID-19 pandemic when education is greatly affected, leading to embracing online learning and examination. A text/exam comprises questions and answers that focus on evaluation to determine the student's conversant level in the area of study. Each question has a cognitive level as described by (Armstrong, 2016) in the revised Bloom's taxonomy. Questions chosen have cognitive levels based on the level of study and standardization of the exam. There is, therefore, a need to consider the question's cognitive level along with other factors when generating an examination by incorporating deep learning algorithms.

Keywords: natural language processing, MLA – machine learning algorithm, AI – artificial intelligence.

1. Introduction

Over time, there has been a notable increase in the number of students joining tertiary institutions. To manage the increase, institutions have responded by creating flexible learning patterns including introducing e-learning and embracing technology in digitizing the majority of the work involved. Automation has been adopted to enhance efficiency and effectiveness at a reduced timeframe. Flexible learning pattern calls for flexible examination pattern thus examiners are challenged to re-think an approach to cater to this need. The solution is to semi or fully automate the examination process to minimize human intervention and increase efficiency and effectiveness.

There has been a challenge in developing the examinations as questions and answers are not readily generated. Researchers have indulged in the automatic questions generation with the majority focusing on the multiple-choice questions and "wh" questions as demonstrated by

Correspondence: Peter Ndegwa Ndirangu, The Kenyatta University, School of Engineering and Technology, Nairobi, KENYA. E-mail: ndejanpeter@gmail.com.

(Ali et al., 2018). Question cognitive level, weight, and topic coverage are key factors to consider when setting exams. Most of the researchers focused on vocabulary assessment and understanding while few studies check question complexity based on the complete spectrum of Bloom's taxonomy. Little has been done on the use of Bloom's taxonomy in exam generation.

2. Methodology

This study is constrained to the classification of questions for generation of examinations purposes.

2.1 *Research questions*

1. What are the processes involved when setting examination?
2. What technologies have been utilized to automate examination generation?
3. What machine learning algorithms have been used in questions classification?
4. How can artificial intelligence be incorporated in question classification to better examination generation?

2.2 *Data sources*

Four sources of data have been considered: IEEE, Science Direct, Google Scholar, and Springer.

Out of the 475 documents that were sourced, 184 were considered most relevant. A query criterion "Automatic or automated generation of Examinations or test or exam or questions classification." This search was done from December 2019 to June 2021. It was made flexible to accommodate many items. The summary of results is as below:

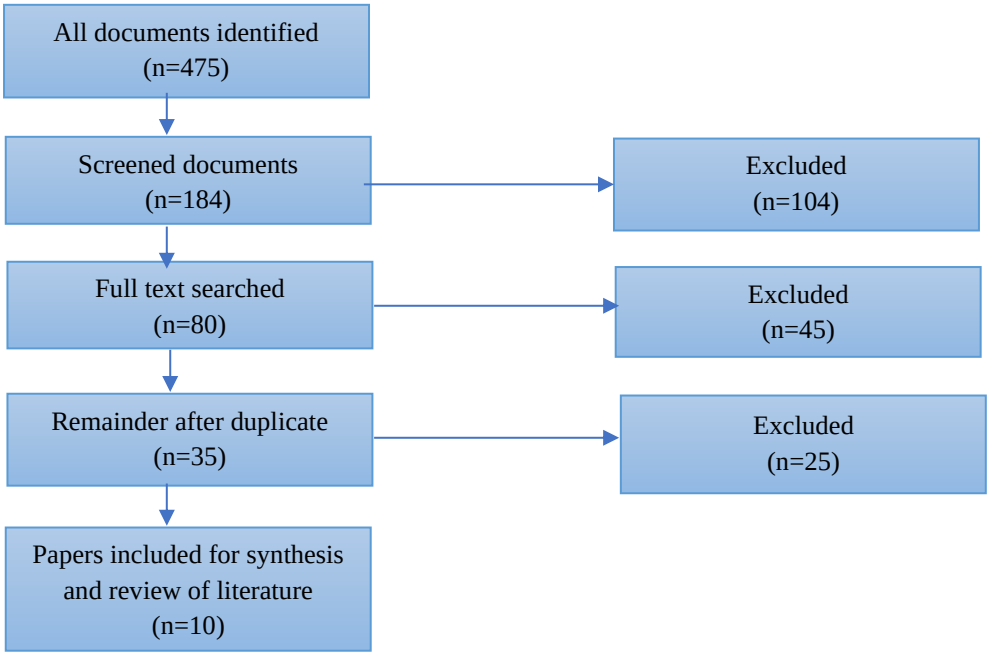

Figure 1. Summary of the query criteria

3. Examinations and test papers

3.1 *Standard examination*

There are many types of assessment or "testing" to access student's learning curves, however, the written examination is the most common approach used by any higher education institution for students' assessment (Omar et al., 2012). An exam contains questions that some studies have sought to classify exam based on Bloom's taxonomy (Abduljabbar & Omar, 2015). In recent times, there have been attempts to classify the questions using Bloom's taxonomy by various researchers using diverse methods. The question cognitive level has been determined using techniques like machine learning, role-based approach among others.

IHL offers studies in one or more levels of study ranging from hands-on skills to professional programs. A question can be examined at various levels of study to test learning. There is, therefore, a need to analyze exam questions to fulfill the requirements by different levels of education such as Bachelor's degree or Master's level (Mohammedid & Omar, 2020).

Content validity, scorer reliability, discrimination, and objectivity are the four principles identified by Johnson (2001) that constitute a standard examination. **Content validity** is representative coverage of the whole course. **Scorer reliability** directs that if the script is subjected to two different examiners, they should arrive at the same score, i.e. there shouldn't be a significant statistical difference in score. There should be a way to differentiate the achievers and weak students to avoid **discrimination**. **Objectivity** prescribes that the test should be fair to all irrespective of age, gender, religion, or any other natural distinction. Examiners should ensure that the test to students in the same level or class test similar concepts and are sensitive to questions cognitive levels to enforce the Objectivity demand.

3.2 *Exam questions*

Questions bear different difficulty levels (Krathwohl, 2002). The difficulty levels build in increasing order from basic, rote memorization to higher (more difficult and sophisticated) levels of critical thinking skills. Therefore, question cognitive levels must be put into consideration during examination generation to facilitate standardization. Failure to consider this may lead to an imbalanced test, i.e., containing many sophisticated questions making it hard for the students or vice versa.

4. A revision of Bloom's taxonomy

Bloom's taxonomy revised edition by (Krathwohl, 2002), breaks the cognitive domain into six levels:

(a) *Remember* – This level entails remembering what is learned.

(b) *Understand* – It is the ability to interpret and comprehend in such a way that one can state the problem in their own words.

(c) *Apply* – Ability to use the concept to solve a new problem.

(d) *Analyze* – This is critically breaking down of information into parts guided by motives or causes and developing inferences that support generalizations.

(e) *Synthesis* – This is the ability to come up with something new by putting information together in a special manner or proposing alternative methods.

(f) *Evaluate* – This is the ability to develop justification and defending an opinion by making judgement about information, the validity of ideas, or the quality of work based on a set of criteria.

5. Review of literature

5.1 *Setting examination*

This is the process of preparing questions to use in assessing the concept taught (Ogula et al., 2006). Ogula is of the view that all the processes of setting exams should be made internally. The processes involved in setting exams are exam setting, moderation, vetting by the external examiner, printing, and proofreading. All these processes consume valuable time and at times may subject exams to leakages if mishandled.

A quality exam should factor in the six Bloom's cognitive domains of knowledge (Bloom, 1994); knowledge, comprehension, application, analysis, synthesis, and evaluation. An exam consists of two sections, the questions' part, and the part of the answer. Questions have properties like mark(s), topic, and complexity (cognitive level). Marks are assigned to each question and determine the weight of the question in the exam. This assignment is influenced by the level of study and the question's complexity. A question examines an area of study (topic) and its complexity indicates the cognitive level. A question can be classified as very simple, simple, moderate, hard, or very hard. Each question should aim to test a certain cognitive level as described in the revised Bloom's taxonomy.

5.2 *Automation techniques in questions classification*

The issue of classifying exam questions based on Bloom's taxonomy has received considerable critical attention in recent years. To handle this task, researchers use different techniques and features (Omar et al., 2012). In this study, MLA and NLP are used. The machine learning algorithms used are K-Nearest Neighbors (KNN), Logistic Regression (LR), and Support Vector Machine (SVM). The study aimed to combine two features: word2vec and TFPOS-IDF (W2VTFPOS-IDF).

Verbs and actions were used to demonstrate different levels of learning (Diab & Sartawi, 2017). The solution was based on the classification of the action verb of the questions or learning outcome statement (LOS), to classify the whole question or LOS into a more accurate level. Action verb classification algorithm was applied on the verb lists from questions and LOS to compute the maximum similarity for every level of the cognitive domain. A rules-based approach was used. The study was concluded by the finding that the approach can be used to provide more accurate verbs and in turn, provide more accurate intended mental skills.

A document analysis method was used by (Karamustafaoglu et al., 2011). The research noted that teachers were asking many questions at the first three levels of Bloom's taxonomy. This study indicated that most teachers fear that the student may not pass the test and therefore resolve to set questions on the low cognitive levels. It concludes by recommending consideration to the questions at the higher cognitive levels to facilitate critical thinking. Surface learning is entertained by assessment strategies that reward low-level outcomes (Buick, 2011).

A comparative study of SVM and K-NN was done by (Patil & Shreyas, 2018) in an attempt to achieve better performance and high quality. Grammar and context checks were applied. The classification was used to test the student level and skills gained compared to Bloom's taxonomy cognitive levels.

Support Vector Machines (SVM) algorithm was used by (Yahya et al., 2012). The classification algorithm was divided into three steps; text representation, SVMs classifiers construction, and SVMs classifiers evaluation. This technique was evaluated by varying the frequency of stop words. The research observed that an increase in the number of words used to represent the question lowers the performance of SVM. It concluded that the number of stop

words should be more than one for a good performance and that reducing the number of stop words does not significantly improve performance.

5.3 *Automation techniques in exam generation*

Computer technology is rapidly changing. This has, therefore, contributed to the development of ideas and algorithms. A computer system can be made to simulate the process of generating exams. Such a system needs to coherently accommodate the discussed items to successfully examine learning. They include; cognitive level, topic, and weight/mark(s).

Despite the need to automate the process of exam generation in institutions, the success of the system must always fulfill certain parameters. Approach, tools, and algorithms used in the development phase play a significant role in fulfilling the addressed need. The quality of E-Systems is determined by views and usages (Nabil et al., 2011).

The question bank is the storage area for the questions fed into the system. Filtering criteria may be adopted which include; exam paper generation process, exclude/include past semester, the total number of items per paper, item complexity, maximum items per topic, paper topic settings, test paper generation, items analysis as described by (Yusof et al., 2017).

An automated paper generation system done by (Bhirangi & Bhoir, 2016) focused on controlled access, questions randomization, and user roles. The use of the cognitive level is not clearly outlined. The software was developed using Java programming language and MySQL database for storage. The algorithm used is improved on the randomization of questions.

Artificial intelligence, randomization, and backtracking are the algorithms used by (Cen et al., 2010) in their project to automate the exam generation process. Technologies used in this system are the MVC pattern in JSP view, JavaBean models, the Servlet Controller, MySQL, CSS + DIV for layout, and JavaScript. Cognitive level and questions weight are not addressed. JSP and Java Servlets are being replaced by emerging technologies. The system produces a word document that can be edited and sometimes loses layout due to compatibility issues.

The cognitive level is used by (Joshi et al., n.d.) in their e-system. Two algorithms; random selection and backtracking, are used. The use of artificial intelligence is not clear. The weight of the questions is computed as a percentage.

Generation of examination should indulge in the use of Natural Language Processing (NLP) as recommended by (Joshi et al., n.d.). This would focus on understanding the question's cognitive levels and prevent a question from being used most frequently.

Package exams developed by (Grun & Zeileis, 2009), provides software infrastructure for scalable exams, associated self-study materials, and joint development. The software used maintenance, variation, and correction as design principles. Technologies used are Latex and R. Questions were separated into answers and solutions sections. Some meta information is collected. Question and a solution description are encapsulated in Latex. In this approach every exercise is contained in a separate sweave file, therefore you need separate files for each. This method was used to make a custom application for processing statistical exams.

Natural Language Processing is used to process text and Named Entity Recognizer and Semantic Role Labeler are used to identify the semantic relation (Rakangor & Ghodasara, 2015). The main focus was to generate simple questions that are true or false or require a one-word answer.

An online system by (Hameed & Abdullatif, 2017) utilized web-based technologies; PHP, MySQL database. Three types of questions were taken into consideration which is true/false,

multiple choices, and image matching. This system did not factor in artificial intelligence, cognitive level, or even question weight.

A rule-based classification approach was used to classify exams by (Kumara et al., 2019; Kumara, Brahmana & Paik, 2019). The model established enabled adjustment of the paper quantitatively. Though the model worked to classify the questions using cognitive levels the research concluded by recommending the introduction of machine learning techniques to increase performance.

6. Conclusion

Examination plays a key role in evaluating what the student has learned and requires to be performed with high precision. Examiners should come up with questions that are sensitive to the cognitive levels outlined by Bloom's taxonomy to ensure that the levels form part of consideration during exam generation. The process of questions classification can be automated by utilizing advancement in technology that presents the world with techniques in AI specifically ML and NLP. A combination of these technologies is resourceful in predicting the question's cognitive levels and realization of a standard examination.

Acknowledgements

It is with the guidance of Prof. John M. Kihoro (The Co-operative University of Kenya) and Dr. Elizaphan Maina (The Kenyatta University) that this research has been successful. This research was supported by the National Research Fund 2016/2017 grant award under the multidisciplinary-multi-institutional category involving The Kenyatta University, The University of Nairobi, and The Co-operative University of Kenya.

The authors declare no competing interests.

References

Abduljabbar, D. A., & Omar, N. (2015). Exam questions classification based on Bloom's taxonomy cognitive level using classifiers combination. *Journal of Theoretical and Applied Information Technology*.

Ali, N. A., Eassa, F., & Hamed, E. (2018). Adaptive E-Learning System Based on Personalized Learning Style. *Journal of Fundamental and Applied Sciences*, *10*(4), 246-251.

Armstrong, P. (2016). Bloom's taxonomy. Vanderbilt University Center for Teaching.

Bhirangi, R., & Bhoir, S. (2016). Automated question paper generation system. *International Journal of Emerging Research in Management &Technology*.

Bloom, B. S. (1994). Reflections on the development and use of the taxonomy in Anderson, Lorin W. & Lauren A. Sosniak, Eds.

Buick, J. M. (2011). Physics assessment and the development of a taxonomy. *European J of Physics Education*, *2*(1).

Cen, G., Dong, Y., Gao, W., Yu, L., See, S., Wang, Q., Yang, Y., & Jiang, H. (2010). A implementation of an automatic examination paper generation system. *Mathematical and Computer Modelling*. https://doi.org/10.1016/j.mcm.2009.11.010

Grun, B., & Zeileis, A. (2009). Automatic Generation of Exams in R. *Journal of Statistical Software*. https://doi.org/http://dx.doi.org/10.18637/jss.v029.i10

Hameed, M. R., & Abdullatif, F. A. (2017). Online examination system. *IARJSET*. https://doi.org/10.17148/IARJSET.2017.4321

Joshi, A., Joshi, M., & Doiphode, S. (n.d.). *A survey on question paper generation system*. https://www.ijcaonline.org/proceedings/ncrenb2016/number1/25549-4014.

Karamustafaoglu, S., Karamustafaoglu, O., Bacanak, A., & Degirmenci, S. (2011). Classification of biology exam questions as to bloom. *Energy Education Science and Technology Part B: Social and Educational Studies*.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*. https://doi.org/10.1207/s15430421tip4104_2

Kumara, B. T. G. S., Brahmana, A., & Paik, I. (2019). Bloom's taxonomy and rules based question analysis approach for measuring the quality of examination papers. *International Journal of Knowledge Engineering*. https://doi.org/10.18178/ijke.2019.5.1.11

Mohammedid, M., & Omar, N. (2020). Question classification based on Bloom's taxonomy cognitive domain using modified TF-IDF and word2vec. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0230442

Nabil, D., Mosad, A., & Hefny, H. A. (2011). Web-Based Applications quality factors: A survey and a proposed conceptual model. *Egyptian Informatics Journal*. https://doi.org/10.1016/j.eij.2011.09.003

Ogula, P. A., Muchoki, F., M., Dimba, M., & Machyo, C. (2006). *Practical guide to teaching practice for students and lecturers*. Catholic University of Eastern Africa: Nairobi.

Omar, N., Haris, S. S., Hassan, R., Arshad, H., Rahmat, M., Zainal, N. F. A., & Zulkifli, R. (2012). Automated analysis of exam questions according to Bloom's taxonomy. *Procedia - Social and Behavioral Sciences*. https://doi.org/10.1016/j.sbspro.2012.09.278

Patil, S. K., & Shreyas, M. M. (2018). A comparative study of question bank classification based on revised Bloom's taxonomy using SVM and K-NN. *2017 2nd International Conference on Emerging Computation and Information Technologies*, ICECIT 2017. https://doi.org/10.1109/ICECIT.2017.8453305

Rakangor, S., & Ghodasara, Y. R. (2015). Literature review of automatic question generation systems. *International Journal of Scientific and Research Publications*.

Yahya, A. A., Toukal, Z., & Osman, A. (2012). Bloom's taxonomy-based classification for item bank questions using support vector machines. *Studies in Computational Intelligence*. https://doi.org/10.1007/978-3-642-30732-4-17

Yusof, S. M., Lim, T. M., Png, L., Khatab, Z. A., & Singh, H. K. D. (2017). Building an efficient and effective test management system in an ODL institution. *Journal of Learning for Development*, 4(2), 211-220.


C O A S