Automated Revised Bloom's Taxonomy Question Categorization Using Rule-Based Approach

Dr. Saritha Namboodiri[1], Aiswarya C.[2]
[1,2]*Sreekrishnapuram V. T. Bhattathiripad College*

***Abstract*—In this work, we have developed a python-based tool that can predict the cognitive level of the questions in a question paper based on Revised Bloom's Taxonomy. We used a training set consisting of 800 questions from the previous year question papers of the University of Calicut. The training set encompasses 120 questions each from the six levels of cognitive learning as specified in the Revised Bloom's Taxanomy. Using the above training set, important keywords, using Parts-Of-Speech tags of the Natural Language Processing Toolkit and other pre-processing NLP techniques, were identified. Chunking Grammer rules were defined from the important Parts-Of-Speech tagged words in the form of regular expression. The RegExpParser implemented the above grammer using Regular Expression Matching and Substitution methods. Our automated tool could classify the test set consisting of 160 questions into its respective cognitive level of learning based on the Revised Bloom's Taxonomy with 85.8% accuracy.**

***Index Terms*—Written examination; Bloom's taxonomy; Revised Bloom's taxonomy; Chunking rules; Rule-based approach; Natural language processing; Parts of Speech tags; Question classification**

I. INTRODUCTION

Written examination is a universal method practiced in most of the educational institutions to evaluate the student's performance. Whether or not the written examination is able to assess the student's ability in their learning objectives very much depends on the questions presented in the examination paper. Therefore, a good and reasonable examination paper must consist of various cognitive levels of questions based on Revised Bloom's Taxonomy. Bloom's taxonomy is a hierarchical model consisting of six levels of learning objectives so as to promote higher forms of thinking in learning and education.

In 1954, Benjamin Bloom, a U.S. psychiatrist, made a significant contribution to education by developing Bloom's Taxanomy, a mechanism to classify and categorize different levels of learning [1, 2]. Bloom's taxonomy is a hierarchical model consisting of six levels of learning objectives so as to promote higher forms of thinking in learning and education based on complexity and specificity. The Original Bloom's Taxonomy created by Benjamin Bloom has undergone many revisions and the framework was lately revised in 2001 by Lorin Anderson and David Krathwohl, yielding the Revised Bloom's Taxonomy [3]. Revised Bloom's Taxonomy allows classifying educational goals, so as to better assess student performance. The present Revised Bloom's Taxonomy is a matrix having the cognitive and knowledge dimensions wherein the conceptually six different levels of cognitive dimension are remembering, understanding, applying, analyzing, evaluating, and creating and the knowledge dimension includes Factual, Conceptual Procedural and Metacognitive Knowledge.

The six different levels of cognitive dimension in Revised Bloom's taxonomy are:

1. Remembering: Refers to recalling previous learned information includes recognizing, listing, describing, retrieving, naming and finding. Examples: Define the purpose of a constructor. List the arithmetic operators in increasing order of precedence.
2. Understanding: Refers to understanding and summarizing by one's own words that is explaining ideas or concepts. Includes summarizing, paraphrasing, explaining etc. Examples: Look at the section of code give below and explain what it does.
3. Applying: Refers to applying a concept in certain

situations that is using information in another familiar situation. Includes implementing, carrying out, using, executing etc. Examples: Given Circle, Square and Triangle class then create a Shape class as a super class of these three classes that includes all the common methods.

4. Analyzing: Refers to breaking information into parts to explore understanding of the whole concepts and their relationships. Includes comparing, organizing, differentiating, deconstructing, interrogating. Examples: Given three methods circle, triangle and square. How the circle method does differ from other two methods?
5. Evaluating: Refers to making judgments based on certain values, criteria or standard and justifying a decision or course of action. Includes checking, hypothesizing, critiquing, experimenting, judging. Example: Which algorithm linear search or binary search is efficient?
6. Creating: Refers to integrating elements together to form a coherent or functional whole as well as reorganizing elements into a new pattern or structure that is generating new ideas or ways of viewing things. Includes designing, constructing, planning, producing, inventing etc. Examples: Create a table named Employee with its required attributes.

Although there are various means of accessing student's learning outcomes, written examination is the widely used approach by any higher education institutions for students' assessment. Question is an element that is intertwined with examination and framing questions in an examination plays an important role in testing the students' overall cognitive levels. Revised Bloom's Taxanomy is widely accepted as a guideline in designing reasonable examination questions so as to ensure balancing of students' cognitive mastery.

Many classification tools have been developed using Bloom's Taxonomy. For the cognitive domain using Support Vector Machines, rule-based approach, Artificial Neural Network, Naïve based and other Machine Learning approaches [4-9]. However, there are limited tools available based on the Revised Bloom's Taxonomy framework. Most of the existing tools based on Revised Bloom's Taxanomy use the concept of word or keyword matching [10]. These tools based on word matching approaches are inefficient as it becomes necessary for the developer to include as many words as possible into the system in order to efficiently classify questions. Moreover, the method fails if the word is not present in the dataset. To overcome these pitfalls, we have developed a python-based tool that automates the Revised Bloom's Taxanomy rules to classify questions from question papers. The tool predicts the cognitive level of the question, given the set of questions in a written examination. A rule-based approach and Natural Language Processing (NLP) is performed to determine the cognitive level of examination questions based on the Revised Bloom's taxonomy [11, 12]. NLP preprocessing is done to tokenize and extract parts of speech tags. With the help of chunking rules and pattern matching techniques, each question is categorized into the respective cognitive level based on the Revised Bloom's Taxonomy.

II. METHODOLOGY

Revised Bloom's Taxonomy is the reference for educators to understand the Educational Learning Objectives. In this work, a rule-based approach along with Natural Language Processing is adopted to classify questions from the question papers into their corresponding Revised Bloom's Taxanomy cognitive level. To automate the process of classifying questions according to Revised Bloom's Taxonomy and classify each question in its respective level, we used a dataset of 800 questions from the previous year question papers of the University of Calicut. The training set consists of 120 examination questions from each of the six levels of learning as specified in the Revised Bloom's Taxanomy. The test dataset is a random collection of 160 questions in Programming subjects obtained from the University of Calicut Previous question papers.

2.1 Pre-processing

To understand the underlying structure of the 800 train set questions, each question passes through NLP text pre-processing comprising of the process of tokenization, lemmatization, stemming, stopwords removal and Parts of Speech (POS) tagging.

- Tokenization: Breaking up a stream of text into individual words or symbols called tokens.
- Stemming: Process of reducing or transforming

inflected words to their base or root words; normal form. For example, converting "flew" and "flying" both into "fly".

- Lemmatization: It is the process of converting words into its base form, normally aiming to remove inflectional endings only and to return to the base or dictionary form of that word, known as the lemma. For example, converting cars into car and boys into boy.
- Stop words Removal: Is the process to eliminate punctuation symbols, special characters and frequently or commonly used words in English language such as a, an, in, on etc.
- Part of Speech Tagging (POS): Is the process of assigning parts of speech tag to each word or token, such as noun, verb, adjective, etc.

2.2 Chunking rules

To add more structure to the question Shallow Parsing (Chunking) was performed following Part-Of-Speech Tagging that results in a group of the words called "chunks". Chunking extracts meaningful short phrases from the Tagged Part-of-Speech question by using a special syntax called RegExp syntax, that delimit the chunks. These rules are then converted to 'regular' regular expressions. This is illustrated using a toy example. Consider a sample question "Define the terms encapsulation and polymorphism". After NLP pre-processing the sample question is transformed as "Define terms encapsulation polymorphism". From the transformed sample question, a flat tree is constructed which is depicted in Figure 1.

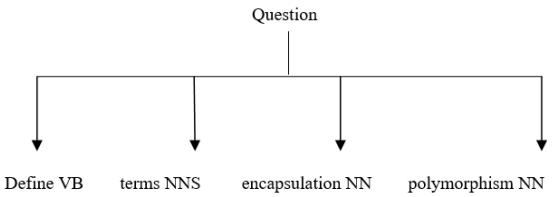

Figure 1. Flat tree of the sample question: "Define the terms encapsulation and polymorphism".

From the flat tree as depicted in Figure 1., chunking rules are formulated as follows:

{<VB><NNS><NN><NN>}

The above rule can also be written as:

Chunk: {<VB><NNS><NN>+}

Chunk: {<VB><N.*>+}

From the 800 questions of the train set, chunking rules are generated forming the Chunking Grammar. RegexpParser implements the Chunking Grammer by using Regular Expression Matching and Substitution methods. The flow of logic is depicted in Figure 1.

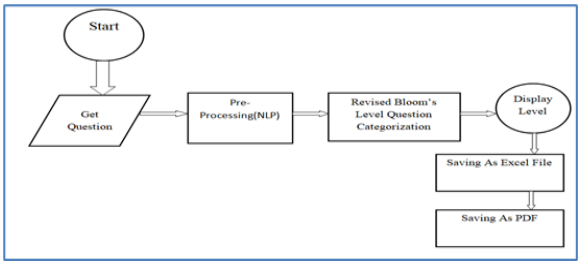

Figure 1: Flow of logic of Revised Bloom's Taxonomy Tool

To test the model, a set of 160 questions test set is used. Each question from the test set is copied to the Text Box next to the Label 'QUESTION' of the user interface. The user interface is depicted in Figure 2.

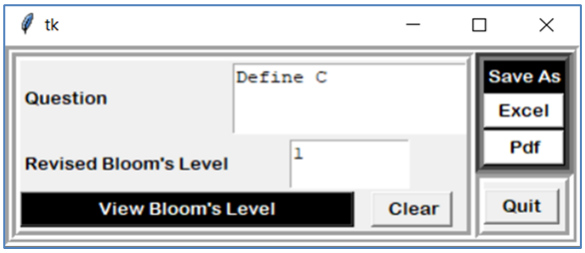

Figure 2. User interface of the tool

The entire set of questions can thus be categorized and stored in an excel sheet. The tool computes the frequency of occurrence of each level thus helping the teachers to prepare Question Papers for Written Examination appropriate to assess the learning outcomes of the learner. The question set can be stored in an Excel or PDF format for future use.

III. RESULTS

The POS that are predefined in the NLP toolkit and used in this work is listed in Table 1.

Table 1. Predefined POS (Parts-Of-Speech) from Natural Language Toolkit (NLP)

Abbreviation	Meaning
CC	Coordinating Conjunction
CD	Cardinal Dgit
DT	Determiner
FW	Foreign Word

IN	Proposition or Subordinating Conjunction
JJ	Adjective
MD	Modal
NN	Noun, Singular
NNS	Noun, Plural
NNP	Proper Noun, Singular
NNPS	Proper Noun, Plural
RB	Adverb
RBR	Adverb, Comparative
RBS	Adverb, Superlative
VB	Verb
VBD	Verb Past Tense
VBN	Verb Past Particle
WDT	Wh-Determiner
WP	Wh-Pronoun (Who)
WRB	Wh-Adverb (How)

The chunking rules for the training dataset are tabulated in Table 2.

Table 2. Chunking Rules formulated from the Train set

	Revised Bloom's Level	Chunking Rules
Remember	1	FT : {<WP><JJ>+} FT1 :{<WP><N.*>+} FT2: {<WP><VBD>+} FT3 :{<WP><VB|VBP|VBG>+} L1 :{(<FT><N.*>+)} L01 :{(<FT2><JJ>+<N.*>+)} L02 :{(<FT2><N.*>+)} L03 :{(<L01><JJ>+<N.*>+)} L04 :{(<L02><IN><N.*>+)} L05 :{(<FT3><JJ>+<N.*>+)} L06 :{(<L05><JJ>+<N.*>+)} L07 :{(<FT3><N.*>+<VBG><N.*>+)} L11 :{(<L1><IN><N.*>+)} L12 :{(<L1><JJ>+<N.*>+)} L13 :{(<L1><VBP><N.*>+)} L14 :{(<FT1><JJ>+<N.*>+)} L15 :{(<FT1><IN><IN>?<N.*>+)} L16 :{(<FT1><IN><JJ>)} L17 :{(<L16><N.*>+)} L18 :{(<L15><VB.*><N.*>+)} L19 :{(<FT1><VBZ>)} L20 :{(<FT1><VB.*><RBR><N.*>)} Level1:{<FT1|L1|L11|L12|L13|L14|L15|L16|L17|L18|L19|L20|L01|L02|L03|L04|L05|L06|L07>}
Understand	2	FT1 :{<WP><N.*>+} FT3 :{<WP><VB|VBP|VBG>+} FT4 :{<WP><VBZ>+} L2 :{(<L1><IN><VBN><N.*>+)} L21 :{(<L1>+<VBD><N.*>+)} L22:{(<FT1><IN><VB.*><N.*>+)} L23 :{(<FT3><VBN><N.*>+)} L24 :{(<FT3><VBD><JJ>)} L25 :{(<FT4><VBP><CD>?<N.*>+)} L26 :{(<FT4><IN><CD>?<VBG>?<N.*>+)} L27 :{(<FT4><JJ>+<N.*>+)} Level2: {<L2|L21|L22|L23|L24|L25|L26|L27>}
Apply	3	FT1 :{<WP><N.*>+} FT5 :{<WP><MD><VB>+<N.*>+} FT6 :{<WP><MD><VB>+<JJ>+} L3 :{(<FT1><MD><VB.*>+)} L31 :{(<L3><IN>?<N.*>*<IN>?<N.*>+)} L32 :{(<L3><JJ>+<N.*>+)} L33 :{(<L27><IN><N.*>+)} L34 :{(<FT5><VBD>+<N.*>*<JJ>+<N.*>+)} L35 :{(<FT5><IN><N.*>+<VB.*>+)} L36 :{(<FT6><N.*>+<VB.*>+<N.*>+)} L37 :{(<L51><VB.*>+<N.*>*<JJ>+<N.*>+)} L38 :{(<L11><VBN>+)} L39 :{(<FT1><VB.*>+<N.*>+)} L3a :{(<L15><IN><VBN>?<N.*>+)} L3b :{(<L05><IN><N.*>+)} Level3:{<L3|L31|L32|L33|L34|L35|L36|L37|L38|L39|L3a|L3b>}
Analyze	4	FT8 :{<WRB><N.*>+} L4 :{(<L1>+<IN><JJ>+<N.*>+)} L41 :{(<L11><JJ>+<N.*>+)} L42 :{(<FT8><JJ>+<N.*>+)} L43 :{(<L42><JJ>+)} L44 :{(<FT8><VB.*>+<RB><JJ>+)} L45 :{(<FT8><VB.*>+)} L46 :{(<L45><JJ>+<N.*>+)} L47 :{(<L55><VBP>)} Level4: {<L4|L41|L42|L43|L44|L45|L46|L47|FT8>}
Evaluate	5	FT3 :{<WP><VB|VBP|VBG>+} FT4 :{<WP><VBZ>+} FT7 :{<WRB><JJ>+} L5 :{(<FT1><VB.*>+<MD><VB.*>+)} L51 :{(<FT3><N.*>+)} L52 :{(<FT3><VBG><N.*>+<IN><N.*>+)} L53 :{(<L51><IN><N.*>+)} L54 :{(<FT4><VBG>?<MD><VB>+<N.*>*)} L55 :{(<FT7><N.*>+)} L56 :{(<FT7><VB.*>+<N.*>+)} Level5 :{<L5|L51|L52|L53|L54|FT5|L55|L56>}
Create	6	FT1 :{<WP><N.*>+} L6 :{(<FT1><VB.*>+<JJ>+<N.*>+)} L61 :{(<L23><JJ>+<N.*>+)} L62 :{(<WRB><RB><VB.*>+<N.*>+)} Level6 :{<L6|L61|L62>}

IV. CONCLUSION

In this work we automated the process of categorizing examination questions based on Revised Bloom's Taxonomy to its cognitive levels. Revised Blooms Taxonomy stresses the need for High level skill testing questions (that help learners to retain information for longer duration) than the Low-level skill addressing the basics of the subject. This python-based tool with the automatic categorization of the questions for a written examination to their cognitive level is helpful to teachers to prepare a well-organized and quality examination paper to assess students' cognitive level. Natural Language Processing together with the rule-based method approach was able to classify questions into their respective cognitive level based on Revised Bloom's Taxonomy with an accuracy of 85.8%.

REFERENCES

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

[2] Wen-Chih Chang, Ming-Shun Chung. (2009). Automatic Applying Bloom's Taxonomy to Classify and Analysis the Cognition Level of English

[3] Anderson, L.W. (Ed.), Krathwohl, D.R. (Ed.), Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., & Wittrock, M.C. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives (Complete edition). New York: Longman.

[4] Hui, C. (2009). Feature Reduction for neural Network in Determining the Bloom's Cognitive Level of Question Items. University Technology Malaysia.

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

[6] Yahya, A., & Osman, A. (2011). Automatic Classification of Questions into Bloom's Cognitive Levels using Support Vector Machines. Najran: Najran University.

[7] Ikonomakis, M., Kotsiantis, S., & Tampakas, V. (2005). Text Classification using Machine Learning Techniques. WSEAS Transactions on Computers, 966-974.

[8] Jungermann, F. (2009). Information Extraction with RapidMiner. Symposium Spractechnologie and humanities, (pp. 50-61). Duisburg.

[9] Sebastiani, F. (2002). Machine Learning in Automated Text Categorization. ACM Computing Surveys, 1-47.

[10] Syahid Ah Sufi Haris, Nazlia Omar. Bloom's Taxonomy Question Categorization Using Rules And N-Gram Approach.

[11] Steven Bird, Ewan Klein, Edward Loper. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit

[12] http://thepeakperformancecenter.com/educationallearning/thinking/bloomstaxonomy/blooms-taxonomy-revised