

## Automatic Question Generation Using Software Agents for Technical Institutions

Shivank Pandey<sup>1</sup>, K.C. Rajeswari<sup>2</sup>

### Abstract

*In the attempt of producing quality graduates, a major factor that needs a considerable amount of attention is the institutions evaluation system. Now, in order to produce quality result their examination system must be very effective, questions must be able to assess students in every domain. Preparation of question paper with high standard that really kindles the student's thinking ability is very challengeable task that need to be performed by the academicians. There arises a need for automatic question generation. For all the existing automatic question generating systems, the problem lies either in eliminating the user's role or in developing factual questions based on Bloom's taxonomy. To overcome these issues, in the proposed system, the focus is to take input in form of a text file from user which contains of the text upon which the user desires to fetch questions; the output is produced in form of a text file containing questions based on Bloom's taxonomy. The entire process is carried out by software agents, which eliminates the major problems of existing systems.*

### Keywords

*Automatic Question Generation, Bloom's taxonomy, Document processing agent, Information classification agent, Question generation.*

### 1. Introduction

In the ongoing advent of computer based technology There are many changes being made till now in various fields that tend to move from manual systems to automated systems. These automatic systems help us with much cost and time efficient solutions. In the education field, the academicians are majorly dependent on their own for generating questions for various examinations.

**Shivank Pandey**, Department of Computer Science and Engineering, Sona College of Technology, Salem, Tamil Nadu, India.

**K.C. Rajeswari**, Department of Computer Science and Engineering, Sona College of Technology, Salem, Tamil Nadu, India.

However several successful attempts have been made for the development of automated assessment systems. The work done in field of AQG, focuses primarily on generation of simple conceptual questions, like who is the president of India?, when was first plane invented? Or what is meant by the term 'cosmology'? This may not prove to be very efficient for judging the students learning. So in order to effectively assess the students, the first step is to design the question paper which covers all the necessary elements to test his/her knowledge. One such taxonomy was proposed by Bloom (1956), who described level of learning to be categorized into six different cognitive domains, namely

1. Knowledge
2. Comprehension
3. Application
4. Analysis
5. Synthesis
6. Evaluation

The advantage of generating questions based on Bloom's taxonomy enables to generate the questions that help to assess learning ability of the students. That means if the tutor wants to assess the student's learning about the topic at the basic level, the questions to be asked should be taken from knowledge level, similarly if he wants to evaluate the advanced level knowledge of student, he must select questions based on higher levels of Bloom's taxonomy like analysis or synthesis. The proposed framework helps in question generation by deploying agents, the agents will perform various operations like document processing, information classification and question generation. Thus system may also be termed as a multi agent system.

### 2. Literature Review

In 2006, Ittizar Aldabe et al. [1] made an attempt to generate automatic questions named ArikIturri, the approach was based on Corpora and NLP techniques, and the information source for the system was the linguistically analyzed real corpora, represented in XML mark-up language. The benefit of the system was its ability to reject the ill formed questions also.

In 2006, Li-Chun Sung et al. [2] proposed a design for Automatic Quiz Generation for Ubiquitous English E-learning System from a given English text to test learner. This was based on quiz generating system for comprehension of text content and English skills. It parses an English text into a semantic network which enhances it by the help of intrinsic and extrinsic knowledge. The question generator takes the refined semantic network and the knowledge base of WordNet and Google to generate questions of Quiz.

In 2010, Ming Liu et al. [3] proposed automatic question generation for literature review writing support. He took the literature review as input text and then the extracted syntactic and semantic features of text were used for question generation, which were based on a rule repository consisting of match pattern and question templates. However, the questions were not simple text based questions, they were based on categories like opinion, result, system, etc. which tested the cognitive learning skills of the learner. Another similar work done by them was G-ASKS [4], where the similar approach was used for academic writing to generate questions based on Graesser & Person taxonomy, here the Citation Classification was done with the help of Naive Bayes Classifier.

In 2010, Memoona Naz et al. [5] developed a utility based Agent for question paper generation which also considered the difficulty level of questions while generating question paper, its prerequisite is the database of questions, from which the agent selects questions according to the difficulty level determined by the examiner. Another work for question generation was done in 2011, Gutl et al. [6] approach was based on extraction of important concepts from textual learning content from which they developed different types of questions such as single choice, multiple choice and completion exercises. The work makes use of statistical, structural and semantic methods of NLP as well as rule based Artificial intelligence solution, however they used WordNet instead of domain knowledge for concept selection. In 2012, K. Sathiyamurthy and T.V. Geetha [7] proposed automatic question generation from documents for E-learning was focused on question generation based on Bloom's taxonomy [8], here the documents are collected over the web using Google search and then those ranked documents are processed for noun/phrase extraction which serve as

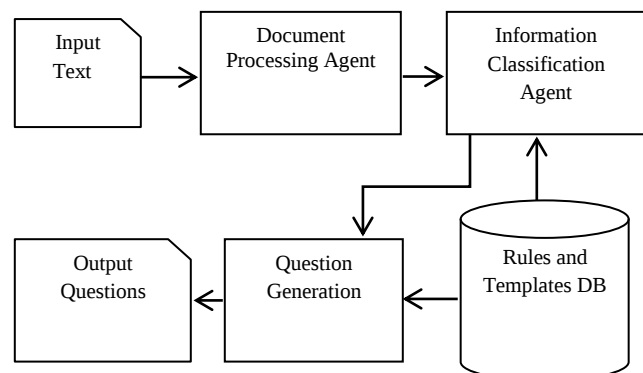
the keywords for question templates, which were designed for different level of Bloom's taxonomy.

In 2013, Virendra Kumar et al. [9] developed a question generator system based on use of Stanford parser, here the entities present in the sentence are identified by Named Entity Recognition (NER) technique. However the system works only at sentence level and also the generated questions are simple WH based questions. Another implementation of automatic question generation module was done in 2012 by Laszlo Bednarik and Laszlo Kovacs [10] where they developed algorithms for clustering and classification tasks according to the set parameters. After text preprocessing, the words were classified and keywords were selected. Later, the related candidate words were selected using word clusters, then these candidate words were grammatically aligned for question generation.

In 2012, Ming Liu, Rafael A. Calvo, Anindito Aditomo, and Luiz Augusto Pizzato [11] developed another question generating system where they first extracted key phrases from students' literature review papers. Each key phrase was then matched with a Wikipedia article and classified into one of five abstract concept categories: Research Field, Technology, System, Term, and Other. Using the content of the matched Wikipedia article, the system then constructed a conceptual graph structure representation for each key phrase and the questions were then generated based on the structure.

### 3. Proposed System

#### A. Architecture:



**Figure 1: Proposed System Architecture**

**INPUT:** Text file with some text.

**OUTPUT:** Text file containing questions.

### B. Architecture Description:

The system takes the input of a text file from user, which is first processed by the DPA (Document Processing Agent) which extracts the words from the ranked list of words extracted from the input text, the list is based on the occurrence of words. The output of DPA is fed as input into the ICA (Information Classification Agent) whose task is to classify the input on the basis of Bloom's taxonomy levels by the help of rules repository. Then it is processed by the Question Generation Module to the final stage of generating questions with the help of question templates from the database, the output is in the form of questions stored in a text file. The agents are implemented in java language.

### C. Document Processing:

The work of processing the input text document is performed by the Document Processing Agent. It first processes the input by the Tree Tagger tool, which produces the output in the form of one word per line, along with the tag and the lemma of that particular token. The second step is that the processed output is ranked based upon the frequency of occurrence of each word. Finally stemming process is done to get the proper keyword, this is accomplished by filtering all the suffix and prefix out of the given word. The entire process is carried out by the agent, which eliminates the work of the human in the processing module.

### D. Information Classification:

In this step, the Information Classification Agent (ICA) takes the input from DPA in the form of list of selected keywords after the process of stemming and word count, the ICA then finds the Bloom's category of those words, by searching appropriate action verb in the repository which fits with the given keyword.

### E. Question Generation:

This is the final step in the question generation process, here the question generation module takes the output of ICA as input to generate questions. The process is a template based approach, which fits the selected keywords in the question template according to the Bloom's levels.

**Example of questions generated for Knowledge Level:** This is a sample of input and its expected output by the system.

**Input Text:** An entity is a "thing" or "object" in the real world that is distinguishable from all other objects. For example, each person in an enterprise is

an entity. An entity set is a set of entities of the same type that share the same properties, or attributes. The set of all persons who are customers at a given bank, for example, can be defined as the entity set customer. The attributes in an entity set can be of following type, Simple, composite, Single-valued, multivalued, or Derived attribute.

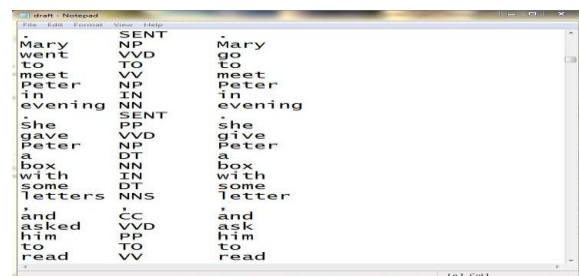
### Output Questions:

1. Define entity.
2. Describe entity set with example.
3. Relate entity set with entity
4. Give an example of entity.
5. List all the type of attribute in entity set.

## 4. Results

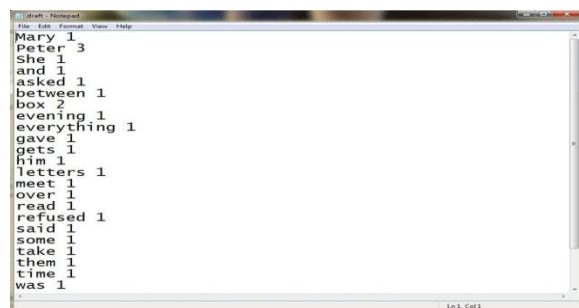
**Document Processing Agent:** It does the preprocessing on the given input text document and produces following outputs.

Output produced by tree tagger: The input text is analysed in the form of one word at a time and is displayed along with its form.



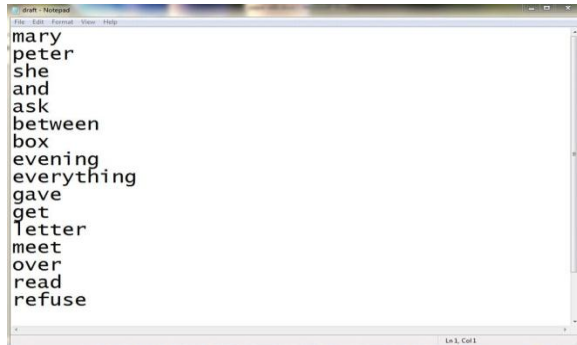
**Figure 2: Tree Tagger Output**

Output of word count process: Here the occurrence of each word is counted and displayed in numerical form along with the word.



**Figure 3: Word Count Output**

Output of Stemming process: The result of this process is obtained by reducing each word to its proper noun/adjective form.



**Figure 4: Stemmed Output**

The results for Information classification agent and question generation module are yet to be obtained, as the system is under development.

## 5. Conclusion

The paper has presented a novel approach for automatically generating questions from text, which combines the agent technology with the flexible question templates. The proposed system will thus be able to effectively judge the cognitive level of student's learning. Use of templates based upon Bloom's Taxonomy has also helped to generate conceptual level questions. The successful implementation of the system has made the question generation task of the technical institutions very easy, which can be used by them to test the students in all the aspects, and has also benefitted the students to test their level of understanding of the topic, after reading.

## 6. Future Work

The work can be enhanced further for advanced level question generation based on Bloom's Taxonomy levels like analysis, synthesis and evaluation. It may also consider the mathematical and derivational type of question generation in future. Another focus is on use of 'the overgenerate and rank' approach [12] that still remains unused in agent based Automatic Question Generating Systems, which will help the examiner to select questions based upon their ranks that will enhance the standard of the question paper.

## References

- [1] Itziar Aldabe, Maddalen Lopez de Lacalle, Montse Maritxalar, Edurne Martinez2, and Larraitx Uria, "ArikIturri: An Automatic Question Generator Based on Corpora and NLP Techniques", ITS 2006, LNCS 4053, pp. 584 – 594, 2006.
- [2] Li-Chun Sung, Yi-Chien Lin, Meng Chang Chen, "The Design of Automatic Quiz Generation for Ubiquitous English ELearning System", 2006.
- [3] Ming Liu, Rafael A. Calvo and Vasile Rus, "Automatic Question Generation for Literature Review Writing Support" (2010).
- [4] Ming Liu, Rafael A. Calvo, Vasile Rus, "G-Asks: An Intelligent Automatic Question Generation System for Academic Writing Support", ICITF, Dialogue and Discourse 3(2) (2012), pp. 101–124.
- [5] Memoona Naz, M. Aslam and Ehtesham-ul-haq Dar, "Utility Based Agent For Test Paper Generation" in International Journal of Multidisciplinary Sciences and Engineering, vol. 1, no. 1, September 2010.
- [6] Gutl C., Lankmayr K., Weinhofer J. and Hofler M., 'Enhanced automatic question creator – EAQC: concept, development and evaluation of an automatic test item creation tool to foster modern e-education', Electronic Journal of e-Learning, Vol. 9, No. 1 (2011).
- [7] K. Sathiyamurthy and T.V. Geetha, "Automatic question generation from documents for E-learning", Int. J. Metadata, Semantics and Ontologies, Vol. 7, No. 1, 2012.
- [8] Bloom, B. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive Domain. New York, Longmans, Green and Co, Vol. 1.
- [9] Virendra Kumar, Imran Khan and Vikas Choudhary, "A Question Generator System Using Stanford Parsing", International Journal of Engineering Research and Development, Volume 7, Issue 2 (May 2013), PP. 01-05.
- [10] Laszlo Bednarik, Laszlo Kovacs, "Implementation and assessment of the automatic question generation module", 3rd IEEE International Conference on Cognitive Info communications, December 2-5, 2012, Kosice, Slovakia.
- [11] Ming Liu, Rafael A. Calvo and Luiz Augusto Pizzato, "Using Wikipedia and Conceptual Graph Structures to Generate Questions for Academic Writing Support", IEEE Transactions on Learning Technologies, VOL.5, NO.3, 2012.
- [12] Heilman, M., Noah, S.A, "Question generation via overgenerating transformations and ranking" in the 14th Annual Conference on Artificial Intelligence in Education Workshop on Question Generation (2009).



**Kurubarahalli Chinnasamy Rajeswari** is at present working as Assistant professor senior grade at Sona College of technology, Salem, India. She received her B.E. degree in Electronics and Communication Engineering from Bharathiar University, Tamil Nadu, India in 2002 and M.E. degree in Computer Science and Engineering from Anna University, Chennai, Tamil Nadu, India in 2006. The focus of her research dissertation is on the enrichment of spontaneous speech synthesis. Her general research interest include natural language processing, artificial intelligence, automatic question generating systems, speech processing. She is associated with the Computer Society of India and Indian Society for Technical Education as a member. She has co-authored a book on artificial intelligence in 2009 under Umayam publications. She also holds the credit of 12 publications across various international, national conferences and journals.



**Shivank Pandey** was born in Uttar Pradesh, India on 10<sup>th</sup> July, 1991. He received his B.Tech degree in Computer Engineering from Rajasthan technical University, India in 2012. In 2012, he enrolled for Masters in Engineering in Software Engineering from Anna University, Chennai, India. At present, he is a teaching assistant at the Department of Computer Science, Sona College of technology, Salem, India. His research interests are agent based technology, intelligent tutoring systems, knowledge representation, data mining and software engineering.