

Online Learning Portal using NLP

Gaurav Gawade

Department of Information and Technology
Vidyavardhini's college of Engineering and Technology
Vasai, India

Shreyash Mhashilkar

Department of Information and Technology
Vidyavardhini's college of Engineering and Technology
Vasai, India

Niketan Patil

Department of Information and Technology
Vidyavardhini's college of Engineering and Technology
Vasai, India

Prof. Swati Varma

Department of Information and Technology
Vidyavardhini's college of Engineering and Technology
Vasai, India

Abstract:- This paper is devoted to use NLP techniques for Generating Narrative films from Text/PDF to evolve education in elementary and secondary schools as students get exhausted by usual text and are more enthusiastic towards videos. With the help of visual aids this would not only help them better remember the content being delivered but also improve their vocabulary. It also covers automatic question generation based on text analysis with the goal to develop an intelligent framework that should automate or semi-automate the process of quiz and exam question generation. The framework operation is predicated on information retrieval and NLP algorithms. It allows generating "Wh"-type (What? When? How?) questions.

I. INTRODUCTION

Online Learning Portal is basically divided into two modules :Generating Narrative Films form TEXT:-An automatic video generator which creates a video from a textbook chapter (in pdf/text format). The video will be a narration (with subtitles) of the whole chapter while showing (titled) images relevant to the current paragraph. It is good because makes learning fun and interesting. Students will pay more attention in class and retain knowledge much better by associating pictures to information. Listening to the text (with good pronunciation) while watching the subtitles would help students get better at speaking and understanding English. It works well for subjects with a lot of factual information like history, geography, biology, economics etc. where looking at relevant pictures is actually useful. For eg:- Albert Einstein was a German-born theoretical physicist who developed the thought of relativity, one of the two pillars of recent physics.

1) Automatic Question Generation- It is a process in which given an input text to the system it will create reasonable questions from the input as output. The potential benefits of using automated systems to get questions helps reduce the dependency on humans to get questions and other needs related to systems interacting with natural languages.

Question generation are often applied in many fields like intelligent tutoring systems, MCQ generation, FAQ generation etc.

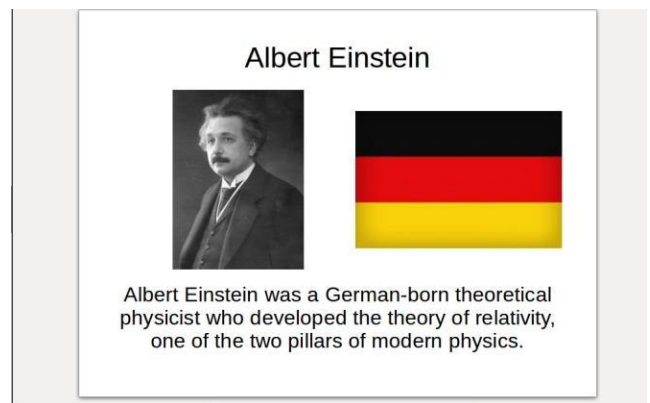


Fig. 1. Snapshot from Narrative film

II. MOTIVATION

Different studies show that the use of technology in schools (as crucial part of education system) has developed new ways of teaching and learning. It enhances learning by providing a better understanding of the subject as well as motivating students. This study was carried out in different institutes to investigate the effectiveness use of technology in Teaching English as a Foreign Language (TEFL) process and if the learners prefer this new way of teaching over conventional methods. Fifty-six students of a secondary school were the subjects of this study. The subjects were divided into two groups, (Experimental and Control). Each group were taught separately, one by using technology in class (e.g; video-projector , power-point and other visual aids), the other through a traditional method such as textbooks. An independent sample t-test was performed and showed that there was a significant difference between the means of two groups. It proved that teaching based on the use of technology had a significant positive effect on learners' performance. Analysis showed that the experimental group learners performed better than the control group. So with the enhanced learning through visual aids students will retain information faster and can evaluate themselves through list of questions that will be asked at the end of the module.

III. RELATED WORK

Intel AI Developer use Natural Language Processing for Smart Question Generation. Automatic question generation is a component of Natural Language Processing (NLP). It is a neighborhood of research where many researchers have presented their work and remains a neighborhood under

research to realize higher accuracy. Many researchers who are been working on the area of automatic question generation through NLP, has developed various techniques and models to get various types of question automatically. Nowadays tutors/professors/teachers (academicians) spend lot of there time generating test papers and quizzes manually. Similarly, students spend tons of your time on self-analysis (self-calibration). Moreover, students are hooked in to their mentors for the self-analysis. Hence, they are working on this NLP area, which has a huge scope of development at this moment. They want to build a computer application system that can help you in calibrating yourself and remove any dependencies on mentors. Here, students can give the input text of whatever material they mentioned, and on this basis they get a group of questions with answers from which they will do a self-analysis (self-calibration). A similar approach is employed by mentors for creating test papers and quizzes. Moreover, online examinations became very fashionable, including many major examinations, like GATE, CAT, and NET. Multiple Choice Questions (MCQ) is extremely easy for evaluations, and its evaluation is implemented through computerized applications in order that results are often declared within a couple of hours, and the evaluation process is 100% pure. By making this computerized application, they can reduce the task of an educator. Much time can be saved if they can know what appropriate questions can be asked for the given input of text. Hence, they want to develop a system which can generate various logical questions from the given text input. For the time being, only humans are capable of accomplishing this.

IV. LITERATURE SURVEY

The automatic question generation (AQG) applications has been classified into two major groups :1) to support the event of dialogue and to develop interactive question and answer system, and 2) to automate the educational assessment. Questions are generated either directly from the expository texts [1], [2] or after the domain topic identification. The first approach, which was proposed by Weizenbaums Eliza program in 1966 [4], was supported towards pattern matching to certain words within the patient's conversation with none understanding of the content. For example, if the patient will be saying, "I am depressed these days" the computer would see the words "I am" and will generate "How long have you been" followed by the remainder of the patient's statement to produce the question, "How long have you been depressed these days?". The second AQG domain in educational assessment has been investigated for many years [5] too. In fact, setting an exam question paper is a very time-consuming task. AQG can significantly reduce the workload of the instructors. The overall AQG process can be described as follows [5], [6]:

- (1) perform a morph syntactic, semantic and/or discourse analysis of the source sentence,
- (2) identify topically important keywords from the sentences for question formulation,
- (3) replace the topically important keyword with a blank or "Wh"question,
- (4) post-process the question and make sure the grammatical correctness.

Interesting examples of these steps have also been discussed in Yao and Zhang, which made use of Minimal Recursive

Semantics. Olney et al. used concept maps to generate questions from text. These approaches focus more on semantics and grammar of the question created which developed a novel approach of using lexical syntactic patterns to form question-answer pairs. Lindberg et al. used semantic labels for identifying the patterns in text in order to formulate the questions. There exists a some kind of AQG tools which has been developed and employed for educational purposes [5], which makes the use of syntactic approach to generate "Wh"-type (What? Who? When?) questions from individual sentences. In addition to Autoquest, there are other systems for "Wh"-question generation using approaches like transformation rules and generating questions supported a given templates . There is also work done in gap filling questions, which is mainly used for vocabulary learning, vocabulary-testing, and language learning. In this paper, for "Wh"-questions generation, pattern matching is employed to a particular extent. This paper pays a special attention to improving the quality of generated questions through supervised learning that is achieved by involving an instructor into ranking the questions. This approach facilitates generating quality questions by removing ones which are either indirectly associated with the text topic or haven't any specific meaning. To implement AQG we use various artificial intelligence and machine learning techniques like rule systems and neural networks. Among neural networks, recurrent neural networks (RNN) are very popular for the natural language processing (NLP) as they allow taking into consideration the text context. For generating 'Wh'-question generation, they have try using Long Short Term Memory (LSTM) networks, which is a modification of RNN. The effectiveness of LSTM networks has been proven in NLP, language understanding, and machine translation applications. All RNNs implement a sequence of repeating modules, which allows remembering previous text. However, in standard RNNs, these modules have a very simple structure, and they are not able to remember long-term facts. LSTM networks also have these modules, but they have more complex structure. An LSTM's chaining module, instead of having only one neural network layer, has four of them, which interact in a special way. Wang et al. Employ bidirectional LSTM RNN for the various automated tagging tasks. Ghosh et al. introduced contextual LSTM network for large scale NLP tasks. This type of network shows good performance in such specific NLP tasks as word prediction, next sentence selection, and sentence topic prediction. In our research, employing LSTM networks allows generating quiz questions that are relevant to the general text topic.

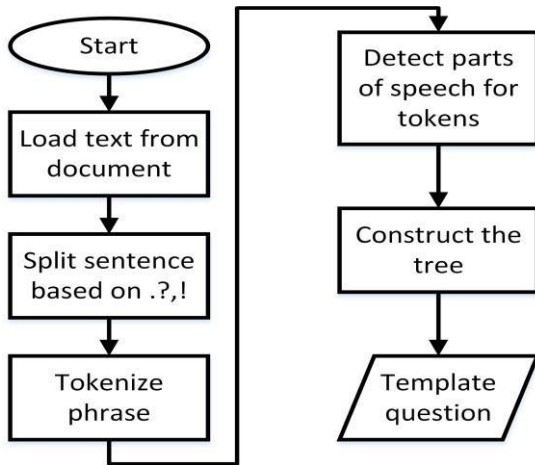


Fig. 2. Training process of question generation

Today technology has a major role to play in pedagogy. When teachers apply this in their classrooms, in fact, they want to attract the students' attention, so that they can enhance effective ways of learning. Microsoft PowerPoint is a software application which is particularly used to present data and information by using text, diagrams with animation, transitional effects, and images, etc in the form of slides. It helps people to understand the idea or topic in front of audience practically and easily. Ozaslan & Maden (2013) concluded in their study that students learned better if the course material was presented through some visual tools[8]. They, also, reported that teachers believed that PowerPoint presentations made the content more captivating; therefore, they helped them to take students attention. The result conducted by Corbeil's study (2007) showed that students exposed to visual presentations preferred them over the textbook presentations [9]; she believed that the students were learning better when their attention was being captured via highlighting, colour, different fonts, and audio visual effects. Presentations are used for purpose of presenting new structures to students, practicing and for reviewing language structures which have already been taught (Segundo & Salazar,2011)[10]. Stepp- Greany (2002) reported to examine, a number of benefits for students related to general use of the technology in classroom which includes increased concentration, improvement in self-concept and proficiency in basic skills, more student-centric learning and involvement in the learning process. Zhao (2007) has conducted a research to investigate different perspectives and experiences of 17 social studies teachers following technology integration training. The research showed that the teachers held a variety of views towards technology integration. These views influenced their application of technology in the classroom. Most teachers were keen to use this technology, experiencing positive experiences with technology integration training, increased their use of technology in the classroom, and using technology more creatively.

V. IMPLEMENTATION

A] Implementation of automatic question generation:

The system will be taking a paragraph as input and will generate relevant questions from the sentences extracted from the paragraph. The entire input paragraph is scanned and split into individual sentences. This splitting is done based on full stop. Next each of those individual sentences is processed by a Parts Of Speech Tagger (POS Tagger).

Question generation can help an individual to get questions from the given text automatically. It is a process during which given an input text to the system it will create reasonable questions from the input as output. The potential benefits of using automated systems to get questions helps reduce the dependency on humans to get questions and other needs related to systems interacting with natural languages. Question generation can be applied in many fields which can be kind of intelligent tutoring systems, MCQ generation, etc.

These are the part-of-speech tags which are used according to Penn Treebank project:

Number	Tag	Description
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential <i>there</i>
5.	FW	Foreign word
6.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PRP	Personal pronoun
19.	PRPS	Possessive pronoun
20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative
23.	RP	Particle
24.	SYM	Symbol
25.	TO	<i>To</i>
26.	UH	Interjection
27.	VB	Verb, base form

It uses small list of combinations. For eg:

- 1 = ['NNP', 'VBG', 'VBZ', 'IN']
- 2 = ['NNP', 'VBG', 'VBZ']
- 3 = ['PRP', 'VBG', 'VBZ', 'IN']
- 4 = ['PRP', 'VBG', 'VBZ']
- 5 = ['PRP', 'VBG', 'VBD']
- 6 = ['NNP', 'VBG', 'VBD']
- 7 = ['NN', 'VBG', 'VBZ']
- 8 = ['NNP', 'VBZ', 'JJ']
- 9 = ['NNP', 'VBZ', 'NN']
- 10 = ['NNP', 'VBZ']
- 11 = ['PRP', 'VBZ']
- 12 = ['NNP', 'NN', 'IN']
- 13 = ['NN', 'VBZ']

Every sentence is parsed using English grammar rules with the use of conditional statements. A dictionary is created and the part-of-speech tags are added to it.

- **Feature Extraction:** This phase goes through all the individual sentences and extracts a set of features from each of them. Depending on these features it selects all the important sentences on which questions are often generated.
- **Common tokens:** This feature counts the words, only nouns and adjectives that the sentence and the title or the subtitle of the paragraph have in common.
- **Number of Nouns:** This is the count of the number of tokens that are tagged as noun (NN, NNS, NNP, NNPS) by the POS tagger. More number of nouns increases the informational context of the sentence and therefore a sentence with more number of nouns is a good candidate for generating question and is therefore selected for further processing.
- **Number of Pronouns:** This is the count of the number of tokens that are tagged as pronoun (PRP or PRP\$) by the POS tagger. More number of pronouns reduces the informational context of the sentence and therefore a sentence with more number of pronouns (in our case greater than 2) is not a good candidate for generating questions and is thereby not considered for further processing.

Discourse Connective	Wh-Question
Since	When
When	When
Because	Why
As a result	Why

Question generation: We will begin by dividing the selected sentences into simple and complex sentences each of which will be processed separately. The sentences which contains discourse connectives are categorized as complex sentences.

1) Generating questions on simple sentences :

2) So in this phase we have divided the simple sentences into subsections of a English sentence i.e. Subject, Verb, Object. Then Named Entity Recognizer (NER) is being processed over the Subject and Object of the sentence which will identify the coarse class classification of it. The NER will then specify the tagged type of the words such as Person/human, Organization and Location. The coarse class classification is as follows:

- **Human:** This includes the name of a person.
- **Entity:** This includes animals, plant, mountains and any object.
- **Time:** This can be time, date or period such as year, Monday, 9 am, last week, etc.
- **Location:** This will be the words that will represent locations, such as country, city, school, etc.
- **Count:** This class will hold all the counted elements such as 9 mens, 7 workers, measurements like weight and size, etc.
- **Organization:** Organizations which consist of companies, institutes, government, etc.

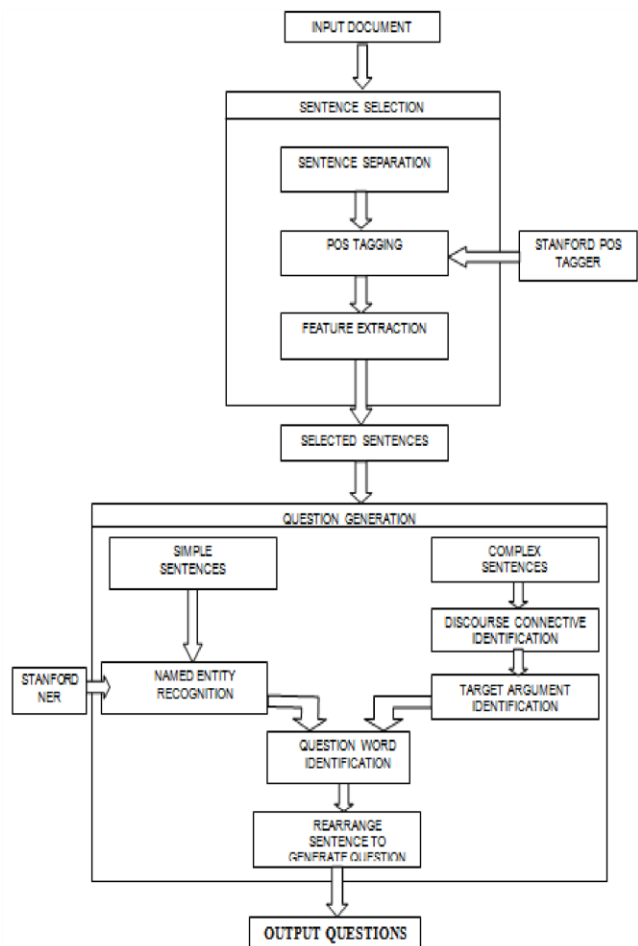


Fig. 3. Question Generation flow

Once the sentence which contain words have been classified to coarse classes, we consider the relationship between the words in the sentence. As an example, if the sentence has the structure “Human Verb Human”, it will classify it as “whom and who” types of question. If it is followed by a preposition that represents location, then we add the “Where” question type to its classification. So then based on the sentence structure and the sentences that are classified based on the various rules. For eg.

- Shivaji Maharaj was born in shivneri fort.
- Geoffrey Hinton is one among the primary researchers who demonstrated the utilization of the generalized back propagation algorithm for training multilayer neural nets.

Output:

- Where was Shivaji Maharaj born?
- Who is Geoffrey Hinton?

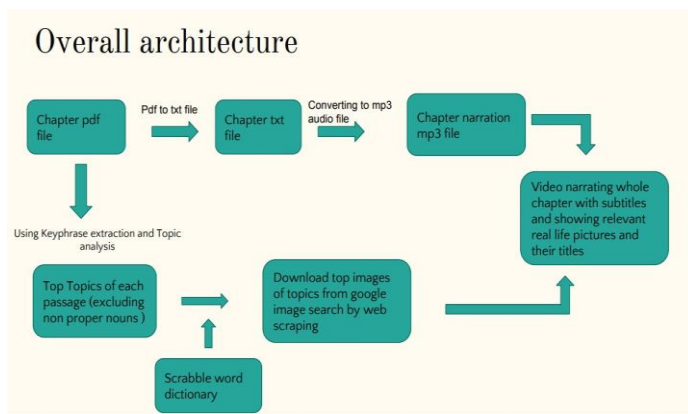


Fig. 4. Text to Narrative Films flow

B] Implementation for generating narrative films: Basically Example (screenshot of film generated) seen in introduction has extracted proper nouns from input text and displayed it along with images of those ,the same text is converted to audio as background sound for video.

- Step1:- Obtaining a chapter as user Input can be a text.
- Step2 :- Extracting proper nouns from the input text using either POS tagger or NER (NER is more better)
- Step3(a):- The text file from step 1 is send to google text to speech API for converting it to audio.
- Step3 (b):- Downloading images from internet by web scraping.
- Step 4:- Merging the audio obtained with slides created using text and relevant images downloaded to form video.

Main challenge in Generating narrative films:

Ambiguity: The main challenge in NLP is to understand and model those elements within a variable context. At times in a language, words are unique but can have different meanings depending on the context in which they are being evaluated.

For eg:- In our case when we searched for Turkey as a country this was the result in google images . So we should be precise in condition like this .



Fig. 5(a). Turkey which is a bird vs



Fig. 5(b) Turkey which is country

1. Solution for this is NER :
 Named Entity Recognition is considered as an important method in order to extract relevant information. For domain specific entities, we have to spend a lot of time on labeling so that we can identify those entities. For general entity such as name, organization and location, we can use pre-trained library which are Stanford NER, spaCy and NLTK to overcome it. In our case we are going to use Spacy for its good accuracy index it properly differentiate between cities, states, companies, agencies, institutions, persons etc.

VI. CONCLUSION

In this paper, we put forward an approach to automatically generate questions given a paragraph. We extract simple as well as complex sentences from the paragraph and generate questions based on subject verb object and prepositions present in the sentence by mapping it to certain predefined rules. The main challenge is to improve the topic analysis and keyphrase extraction part of the process to show more important and relevant images to the current paragraph.

Currently, we are looking at passages individually for topic analysis. It would be better to look at the whole text to gain some insights about the overall theme of the text. Our system can be used in many self-analysis scenarios. For example, It can be helpful for students as it makes learning easier as well as more interactive and interesting. Teachers and professors can use this system to immediately create a quiz. An examination board can utilize this system to create a unique test that will not be known to any professor, discarding the probability of cheating and thereby securing the integrity and privacy of the examination. The main challenge is to improve the topic analysis and key phrase extraction part of the process to show more important and relevant images to the current paragraph. Currently, we are looking at passages individually for topic analysis. It would be better to look at the whole text to gain some information about the overall theme of the text.

REFERENCES

- [1] Learning Text To Image Synthesis With Textual Data Augmentation Hao Dong, Jingqing Zhang, Douglas McIlwraith, Yike Guo Data Science Institute, Imperial College London .
- [2] The Impact Of Using Powerpoint Presentations On Students' Learning And Motivation In Secondary Schools Fateme Samiei Lari 2014.
- [3] Part-Of-Speech Tagging From 97% To 100%: Is It Time For Some Linguistics? Christopher D. Manning Departments Of Linguistics And Computer Science Stanford University.
- [4] Strong Rules Learning Algorithm For Ensemble Text Classification Jin-Hong Liu, Yu-Liang Lu 2007.
- [5] Video Generation From Text Yitong Li, Martin Renqiang Min , Dinghan Shen, David Carlson, Lawrence Carin Duke University, Durham, Nc, United States, 27708.
- [6] Generating Questions And Multiple-Choice Answers Using Semantic Analysis Of Texts Jun Araki, Dheeraj Rajagopal, Sreecharan Sankaranarayanan, Susan Holm Carnegie Mellon University, Pittsburgh, Pa 15213, Usa .
- [7] Computational Intelligence Framework For Automatic Quiz Question Generation Akhil Killawala, Igor Khokhlov, Leon Reznik B. Thomas Golisano College Of Computing And Information Sciences Rochester Institute Of Technology Rochester, Usa.
- [8] Generating Natural-Language Video Descriptions Using Text-Mined Knowledge Niveda Krishnamoorthy UT Austin Niveda .Edu Girish Malkarnenkar * Ut Austin Girish@Cs.Utexas.Edu Raymond Mooney.
- [9] Generating Questions And Multiple-Choice Answers Using Semantic Analysis Of Texts Jun Araki, Dheeraj Rajagopal, Sreecharan Sankaranarayanan, Susan Holm, Yukari Yamakawa, Teruko.
- [10] Strong Rules Learning algorithm For Ensemble Text Classification Jin-Hong Liu, Yu-Liang Lu.
- [11] Enhancing Automatic Ppt Generation Technique Through Nlp For Textual Data Pooja Belote ,Sonal Bidwai , Snehal Jadhav , Pradnya Kapadnis , Nakul Sharma.
- [12] Building An Agent For Factual Question Generation Task Miroslav Blšták, Viera Rozinajová , Institute Of Informatics, Information Systems And Software Engineering, Faculty Of Informatics And Information Technologies.