



METHODOLOGICAL FOUNDATIONS OF AUTOMATED QUESTION GENERATION (ASY) FOR LEARNING MANAGEMENT SYSTEMS USING ARTIFICIAL INTELLIGENCE (SI) AND NATURAL LANGUAGE PROCESSING (TQI)

Yuldashev N. A

Andijon mashinasozlik instituti,

“Texnologik jarayonlar va ishlab chiqarishlarni avtomatlashtirish va boshqarish” ixtisosligi bo'yicha mustaqil izlanuvchi

Dr. Danish Ather

Ilmiy-tadqiqot bo'limi boshlig'i, dotsent.

Toshkent shahridagi Amiti universiteti

Tayanch so'zlar: avtomatik ravishda savol yaratish, ta'lim sektori, matni qayta ishlash, tokenizatsiya, lemmatizatsiya, nutq qismlarini (POS) teglash, stemming, tabiiy tilni qayta ishlash (NLP), mavzuni modellashtirish, faktik savollar.

Ключевые слова: автоматическое создание вопросов, образовательный сектор, обработка текста, токенизация, лемматизация, маркировка частей речи (POS), стемминг, обработка естественного языка (NLP), моделирование тем, фактические вопросы.

Key words: automatically question generation, educational sector, text processing, tokenization, lemmatization, Parts-of-Speech (POS) tagging, stemming, natural language processing (NLP), topic modeling, factual questions.

Introduction. Using Natural Language Processing (NLP), an Automated Question Generator can produce precise and pertinent questions from various input formats like text, structured databases, or knowledge bases. NLP, a subset of Artificial Intelligence (AI), concentrates on facilitating human-computer interaction by understanding and deciphering spoken and written language. The central objective of NLP is to comprehend and extract meaning from written or textual content in natural languages. This initiative employs NLP to generate questions for computer-based exams, educators, and students gearing up for competitive assessments. The goal is to streamline the process of crafting and adjusting Multiple-Choice Questions (MCQs) while establishing a valuable



repository of questions for educational purposes. Consequently, the project ensures that the MCQs encompass fitting questions and choices aligned with the learning objectives and the relevance of the covered subjects. The utility of an automated question generator spans across diverse domains, encompassing Massive Open Online Courses (MOOCs), objective question formulation, search engines, automated assistance systems, customer interaction chatbots, and mental health evaluation in healthcare. Nonetheless, it's worth noting that the manual creation of meaningful and pertinent questions still demands time and dedication.

Developing questions for educators and students preparing for competitive exams is a challenging endeavor. The current method involves manual question creation, which is time-intensive and demands considerable human labor. Consequently, there's a growing demand for a system capable of efficiently and swiftly generating questions with minimal human involvement.

The process of manually crafting test papers and quizzes consumes significant time for educators, professors, and tutors alike. Similarly, students invest considerable time in assessing their own comprehension and knowledge. [1]

Literature Review. To summarize the articles related to the proposed work on applying NLP and machine learning models in the education sector, a literature survey was conducted. Here are the key points from each article:

In a paper by Ankita, K. A. et al., the complexity of Part-of-Speech (POS) tagging is discussed, particularly the computational levels involved in determining POS tags. The study emphasizes simplifying POS labeling complexity through the utilization of Hidden Markov Models (HMM) and proposes integrating Named Entity Recognition (NER). Despite the abundance of POS taggers, researchers continue to seek methods that are less labor-intensive and less susceptible to complexities. Unlike single-word-based methods, HMM-based taggers ascertain tags on a sentence level.

Aleena et al. implemented a question generation system that focuses on natural language understanding and processing. Their system conducts data preprocessing, extracts crucial phrases, and employs natural language processing methods. It can produce question papers from text, documents, or PDF files through processes like stop word removal, key phrase extraction using the TF-IDF algorithm, verifying term presence on Wikipedia, generating question triplets via WordNet, and ensuring input clarity.

In a paper by Onur et al. proposed a rule-based methodology for automating question generation. Their approach considers both syntactic and semantic structures of sentences to create thorough questions based on word semantic roles, the system employs layouts or rules grounded in dependency, named-

entity recognition (NER), and semantic role marking (SRL). Chunking is used to determine whether to generate «who» or «what» questions.

Priti et al. The discussion focused on generating questions according to Bloom’s taxonomy. Their method encompasses data preprocessing, involving feature extraction and Stanford POS tagging. It includes syntactic and semantic analyses to examine structures and identify named entities. Appropriate «WH» question words are mapped to the pre-processed text, and questions are generated accordingly.

Overall, these articles propose different methodologies and techniques for automating question generation in the education sector using NLP and machine learning approaches.

The researchers have focused on developing NLP applications associated with the generation of question papers and multiple-choice questions. They propose the development of a model that considers object-based learning and Bloom’s taxonomy to generate questions. Additionally, they have developed a web-based application that can generate multiple-choice questions based on the input of a text or paragraph. The application also assesses the test scores, provides results, and displays the correct answers to the participants.

Model Design. Feature engineering refers to the process of extracting features from raw data and converting them into formats suitable for machine learning (ML) models. A common technique for feature engineering in natural language processing (NLP) is the bag-of-words approach, which simplifies the representation of a text document based on word count statistics. A typical framework used in supervised classification problems is depicted in the provided figure. This framework involves two phases: training and prediction. In the training phase, feature sets are generated by a feature extractor. These feature sets, along with corresponding labels, are utilized by the machine learning algorithm to build the model. During prediction, unseen inputs are processed by the same feature extractor to generate feature sets, which are then fed into the model for label prediction.

NYAI also includes a syntactic dependency parser that allows for navigating the tree structure of parsed sentences. This parser is capable of identifying sentence frontiers and iterating through noun phrases or chunks. Named Entity Recognition (NER) is also a significant task in

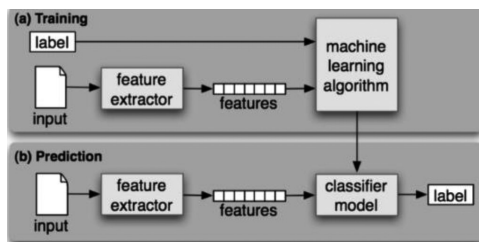
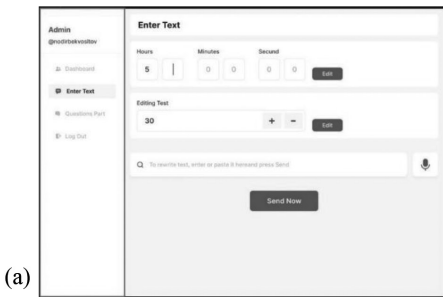


Figure: The framework for supervised classification. Image is sourced from.

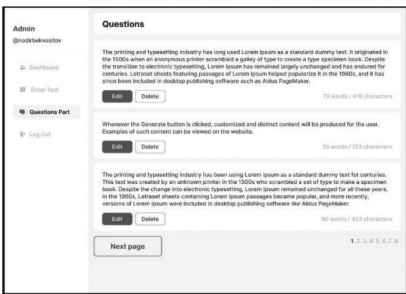
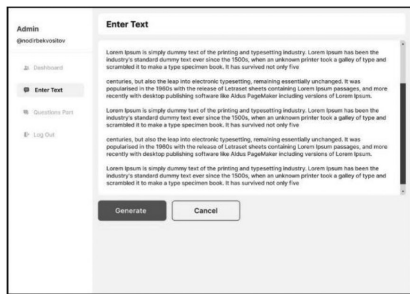


NLP, and project (NYAI) offers features to recognize and categorize named entities, including organizations, individuals, locations, temporal expressions, and monetary figures. The library categorizes named entities based on their characteristics and assigns them to different classes.

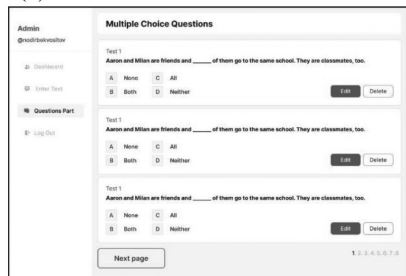
Evaluation and Performance analysis. Upon inputting the text paragraph and clicking on the «Generate question» button, the code executes in the command prompt, as illustrated in the provided figure. For instance, if a brief paragraph is entered, it will be divided into sentences to produce questions for each sentence.



(a)



(b)



Two types of questions are generated: multiple-choice and fill-in-the-blank questions. Sentence segmentation is facilitated using the NLTK package, it appears there might be a typo in query. The package commonly used for natural language processing tasks is NLTK (Natural Language Toolkit). To generate valid distractors along with the correct answer for the generated questions, the Sense2vector package is employed. It extracts keywords from each sentence and identifies related words to generate distractors solely from that paragraph.

The questions are making constructed based upon the keywords identified within the paragraph.

Results and discussion. It is described that the educator can use a dropdown menu to select a subject for which they want to generate questions. Users have the option to input paragraph text, with the flexibility to include multiple paragraphs if necessary. Upon selecting the «Create (Generate) questions» button, the system produces questions derived from the input and subsequently stores them in the database.



After generating the questions, the educator is redirected to the question paper list page, where they can demonstrate the automated questions generate. The file includes information such as the subject and the date and time when the questions

were generated. Different question types are available, including multiple-choice questions (MCQs), fill in the blanks.

The question column in the question paper list displays the generated questions, and the options column includes the answer choices, consisting of one true answer and three obstructions. The correct answer column indicates the correct answer for each question.

The photos given above are provided to illustrate the process of question generation, although the details of these figures are not mentioned in the text.

Conclusion and Future Scope. Generating questions from text has demonstrated its value across different fields, yet the manual approach is both time-consuming and costly. To tackle this challenge, a system has been created, accessible via web and desktop platforms, offering functionalities tailored for generating multiple-choice questions (MCQs) and fill-in-the-blank questions. The main goal of this application is to furnish an automated question generator according to NLP techniques.

The approach not only facilitates analysis itself for students by generating questions but also finds utility in the educational field for generating assessment test papers. The generated questions are presented to students along with their corresponding answers, serving as practice tests for assessment. It can be particularly beneficial for students preparing for competitive exams, as it generates questions for practice and preparation.

This work can be expanded further to include the generation of paraphrasing questions, descriptive questions, and the overall generation of question papers encompassing different types of questions. Additionally, there is scope for exploring the incorporation of automated answer validation systems to authenticate responses and furnish results to students, thereby enriching the functionality and practicality of the tool.

Test 30
A: I have got two dictionaries, but _____ of them include the meaning of that word in it.
B: You can look it up in an online dictionary then.

A None (True) C Neither
 B Both D All (False)

Test savollari

Full_Name	Correct	Incorrect	Result
Abrorbek Ibrokhimov	20	10	30 ball

Done



Literature:

1. N.A. Yuldashev “The role and utilizing of automated question generation into web based online examination systems” Journal of Engineering and Technology (JET) ISSN(P): 2250-2394; ISSN(E): Applied Vol. 13, Issue 2, Dec 2023, 157-164 © TJPRC Pvt. Ltd.
2. O. KEKLIK, “Automatic question generation using natural language processing techniques,” 2018. Accessed: Nov. 20, 2022. [Online]. Available: <http://openaccess.iyte.edu.tr/bitstream/handle/11147/6938/T001801.pdf?sequence=1&isAllowed=y>
3. M. R. S. M. S. MPhil and G. K, “Automatic Question Paper Generator System,” Int. J. Trend Sci. Res. Dev., vol. Volume-3, no. Issue-3, pp. 138–139, 2019, doi:10.31142/ijtsrd21646.
4. P. 'Gumaste, S. 'Joshi, S. 'Khadpekar, and S. 'Mali, “Automated Question Generator System Using NLP,” Int. Res. J. Eng. Technol., vol. 7, no. 6, pp. 4568– 4572, 2020, [Online]. Available: <https://www.irjet.net/archives/V7/i6/IRJETV7I6848.pdf>
5. D. R. Ch and S. K. Saha, “Automatic Multiple Choice Question Generation from Text: A Survey,” IEEE Transactions on Learning Technologies, vol. 13, no. 1. Institute of Electrical and Electronics Engineers, pp. 14–25, Jan. 01, 2020. doi: 10.1109/TLT.2018.2889100.

РЕЗИОМЕ

Avtomatik savol yaratuvchining maqsadi berilgan matndan tabiiy tovushli, semantik jihatdan to'g'ri va grammatik jihatdan izchil bo'lgan yangi savollar yaratishdir. Xulosa va parafraz kabi boshqa til yaratish vazifalaridan farqli o'laroq, bu tizim javob talab qiladigan savollarni yaratishga qaratilgan. U bir nechta tanlovli savollarni yaratish uchun yuqori sifatli chalg'ituvchilardan va samarali savollarni yaratish usullaridan foydalanadi. O'qituvchilar ushbu tizimdan to'g'ri javoblar va chalg'ituvchi vositalar bilan ko'p tanlovli baholash savollarini tuzib, o'quvchining fan bo'yicha tushunchasini tezkor baholash uchun foydalanishlari mumkin. Ushbu yondashuv, shuningdek, o'quvchilarga mavzu bo'yicha o'z tushunchalarini baholashga imkon beradi. Model, ayniqsa, ta'lim sohasida test ishlarini yaratish uchun foydalidir. O'qituvchilar tizimga bir yoki bir nechta paragraflarni nusxalash va joylashtirish orqali osongina savollar yaratishi mumkin. Python dasturlarini tabiiy til asboblari to'plami yordamida yozish mumkin, u matnni qayta ishlash kutubxonalarini tokenizatsiya, tahlil qilish, lemmatizatsiya, qismlarga ajratish, Nutq qismlari belgilarini belgilash va stemming kabi vazifalar uchun funktsiyalar bilan ta'minlaydi. Matn tabiiy tilni qayta ishlashning turli usullarida, jumladan, muhim ma'lumotlarni aniqlashga yordam beradigan mavzuni modellashtirishda hal qiluvchi rol o'ynaydi. Olingan muhim yoki ko'rsatma matni asosida savollar ro'yxati tuziladi. Turli xil savollarni yaratish usullari odatda kim, qachon, qaerda, nima uchun va nima bilan bog'liq faktik savollarni beradi. Tabiiy tilni qayta ishlash dasturlari mashinalar tomonidan til va og'zaki muloqotni tushunishga yordam beradi. Jarayon tarkibini tarkibiy qismlarga ajratish, tilning ma'nosini izohlash, tegishli harakatlarni tanlash va kontentni foydalanuvchilar tushunadigan tarzda taqdim etishni o'z ichiga oladi.

РЕЗИОМЕ

Целью автоматического генератора вопросов является создание новых вопросов из заданного текста, которые будут звучать естественно, семантически точны и грамматически связны. В отличие от других задач генерации языка, таких как обобщение и перефразирование, эта система фокусируется на создании вопросов, требующих ответов. Он использует высококачественные отвлекающие факторы и эффективные методы построения вопросов для создания вопросов с несколькими вариантами ответов. Преподаватели могут использовать эту систему, чтобы быстро оценить понимание учащимся предмета, создавая оценочные вопросы с несколькими вариантами ответов, правильными ответами и отвлекающими факторами. Этот подход также позволяет учащимся оценить собственное понимание предмета. Модель особенно полезна для создания контрольных работ в образовательном секторе. Учителя могут легко создавать вопросы, копируя и вставляя один или несколько абзацев в систему. Программы на Python можно писать с помощью набора инструментов Natural Language Toolkit (NLTK), который предоставляет библиотеки обработки текста с функциями для таких задач, как токенизация, синтаксический анализ, лемматизация, разбиение на фрагменты, тегирование Части речи и формирование корней. Текст играет решающую роль в



различных методах обработки естественного языка, включая тематическое моделирование, которое помогает идентифицировать важную информацию. На основе извлеченного значимого или инструктивного текста создается список вопросов. Различные методы генерации вопросов обычно дают фактические вопросы, связанные с тем, кто, когда, где, почему и что. Программы обработки естественного языка помогают машинам понимать язык и разговорную речь. Этот процесс включает в себя разбиение контента на составные части, интерпретацию значения языка, выбор соответствующих действий и представление контента таким образом, чтобы пользователи могли его понять.

SUMMARY

The primary objective of the Automatically Question Making system is to craft authentic, semantically precise, and grammatically correct questions from provided text. Unlike other language generation tasks such as summarization or paraphrasing, this system focuses specifically on generating questions that necessitate answers. It employs well-crafted distractors and effective question construction techniques to formulate multiple-choice questions. This tool serves educators by enabling swift assessment of students' comprehension through the creation of multiple-choice assessment questions with both correct answers and distractors. This approach empowers students to evaluate their own grasp of the subject matter. Particularly beneficial for educators in crafting test papers, this system allows for easy question generation by simply inputting one or more paragraphs into the system. Implementation can be achieved through Python programming utilizing the Natural Language Toolkit (NLTK), which provides comprehensive text processing libraries capable of tasks such as tokenization, parsing, lemmatization, chunking, Parts-of-Speech (POS) tagging, and stemming. Text is pivotal in numerous natural language processing techniques, including topic modeling, which aids in identifying essential information. Drawing from extracted significant or instructional text, a series of questions can be generated. Various methods of question generation typically result in factual questions revolving around who, when, where, why, and what. Natural language processing programs facilitate machine understanding of language and verbal communication by disassembling content into its constituent elements, interpreting language meaning, selecting relevant actions, and presenting content in a comprehensible manner to users.