Nanotechnology Perceptions
ISSN 1660-6795
www.nano-*ntp.com*

Automated Question Generation System Using Natural Language Processing to Meet Educational Needs

[1] V.Ravikumar, [2] P. Balakumar, [3] R. Dhanalakshmi

[1,2] Professor , [3,] Associate Professor
[1,2] Prince Shri Venkateshwara Padmavathy Engineering College, Chennai 127
[3] Jeppiaar Engineering College , Chennai 119
Email:[1] vvr_siva@yahoo.co.in, [2] balakumar.p@psvpec.in, [3]dhanalakshmi@gmail.com

Years of educational research have demonstrated the advantage of active learning in enhancing students' learning results when they actively engage with the materials, for example, by responding to questions and solving problems. Most teachers struggle with how to assign questions for various grade levels. Therefore, it is quite challenging for someone to produce the greatest series of questions from greater volumes of data. An essential tool for handling the vast volume of data is natural language processing (NLP). Here, we pre-process the data using a variety of NLP Methodologies like tokenization, stemming, and lemmatization. By identifying the characteristics of every single sentence in the provided paragraph, Named Entity Recognition determines the characteristics of data. The necessary information can be quickly found using NER. The barrier created by the rule-based algorithm prevents the generation of specific queries for a given paragraph. The produced questions come in three difficulty levels: easy, medium, and challenging. To set the various possible questions, the Automatic Question Generator is quite effective for the question setter.

Key Words: Natural Language Processing, Tokenization, Rule-based algorithm, NER, Automatic Question Generator

Introduction

When it comes to developing tests and translating the content into multiple languages, teachers frequently run into difficulties. Along with translation, a helpful and effective technology may be one that automates the process of developing questions and answers depending on content presented as a test. If done right, it can speed up the educational process and raise standards [1]. When test paper creators want to have access to a large collection of questions for regular question generating, they employ the specific and distinctive software known as Question Paper Generator. It can be used in a variety of engineering, medical, and coaching facilities for theory papers, which can quickly produce random test questions. Depending on the storage, capacity, and requirements of the system, an infinite number of units and chapters can be entered [2][21].

A sentence or paragraph is given, and the goal of question generation (QG) is to generate natural questions from it. Creating questionnaire is one important use of question generating in the field of education. As an illustration, displays three hand created questions that assess a user's comprehension of the relevant text segment [3-5]. A new strategy called over generate-and-rank that uses a rule-based approach to generate several questions from an input sentence, ranks them using a supervised learning-based ranker, and outperforms a purely rule-based system. The questions generated frequently overlaps word-for- word with the token in the input text, making them relatively simple to answer, even if the ranking system helps to develop more acceptable questions because it relies largely on a hand-crafted feature set [6][17].

The process of creating questions to aid learning is extremely complicated, imaginative, and knowledge intensive. Typical college professors have years of knowledge in their subject area, making it difficult for one model to fully automate their question creation process. Second, the classroom is a high-stake setting, and teachers frequently have predetermined objectives [7][16]. They might not favour imperfect and vulnerable AI models as a result. To bridge the gap between Natural Language Processing Question Generation and the reality in classrooms, the decision-making work of the instructors should be analysed and the crucial steps to recreate the NLP with high educational value must be identified [8-9].

It would be conceivable to create a system that could generate more complicated questions, but it would probably need encoding a large amount of individual knowledge about the chosen subject and the different kinds of inquiries. Conceptual fulfilment queries and, to a smaller extent, verification questions, are the two categories of questions that this work specifically intends to produce [10][22]. Questions on concept completion elicit specific knowledge that accomplishes a given incomplete concept or proposition. Verification questions ask for a yes-or-no response to confirm the information provided. Other question types include judgmental questions goal-oriented questions and example questions can also be utilized.

Methodology

The suggested system enriched data sets for reading comprehension and question-answering natural-language processing (NLP) research. Computerized systems can complete Natural Language Processing (NLP) tasks in an efficient manner. However, it is first important to appropriately divide the documents into units that are meaningful. Most NLP programmes assume that their input will be divided into sentences with each sentence representing a single token. Since real-world papers don't have the same well-defined structure, some approaches must be used to carry out these duties. The categorization model's information is extracted as a set of rules that are simple to understand and highly expressive. When assessing data with a mix of quantitative and qualitative features, this method works well. This study suggests a rule-based method for automatically generating questions. The suggested method focuses on analysing a sentence's syntactic and semantic structure. Although question generation from sentence is the developed system's major goal, automatic evaluation findings reveal that it also performs admirably on reading comprehension datasets that place more emphasis on querying from paragraphs.

A. Pre-Processing:

At first, Pre-processing steps such as Tokenization, POS, Vectorization and Stemming can be done starts with loading the metadata first, after which it is associated to the data and takes the place of the modified data. The undesired data will then be removed from the list and the data will be divided into the train data and the test data. Tokenization is the process of breaking down a sentence, phrase, essay, or even an open text document into simpler components, like individual words or phrases. Tokens are the name for each of the smaller components. The tokens may take the form of words, integers, or punctuation. Word boundaries are found and used to produce smaller units during tokenization. Parts of Speech Tagging done to recognize entities, themes and to process sentiment.

All possible combinations of letters or words of fixed length n that found in the original text are known as n-grams. Unigrams are ngrams with n=1. Similar structures include bigrams (n=2), trigrams (n=3), and so forth used. Contrary to bigrams and trigrams, unigrams typically don't have as much information. The fundamental idea underlying n-grams would be that they identify the letters or word that will probably come after the supplied word. More the context one must deal with, the lengthier the ngram (greater n). Linguistic normalising process called stemming that strips words of their derived affixes or reduces them to their word roots. When words are lemmatized, they are reduced to their basic words, which are lemmas in the proper sense. With the aid of lexical and morphological analysis, it converts root words.

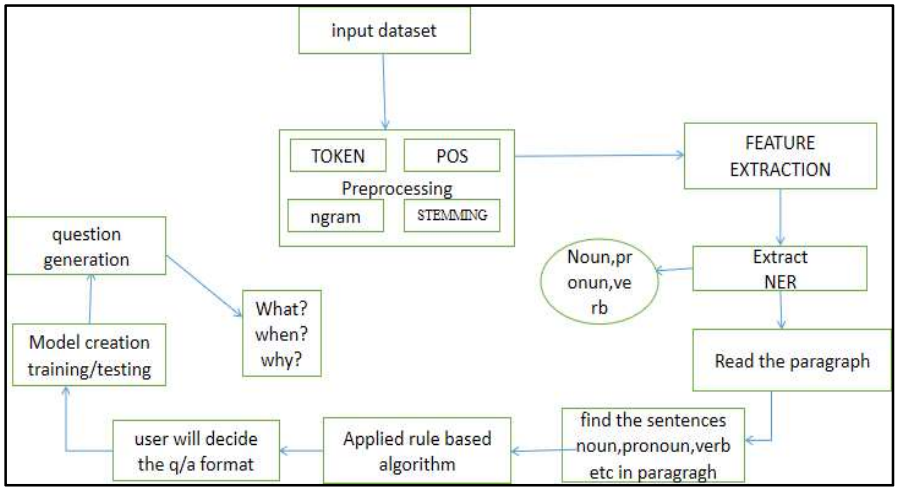

Figure 1 Block Diagram

B. Feature Extraction:

Step 1: From the provided text input, choose the best possible group of sentences from which to create the questions.
Step 2: To determine the sentence's main theme, identify the sentence's subject and context.
Step 3. Determine which type of question can best be derived from that sentence.

At Lexalytics, a significant volume of pre-tagged entities has been used to train supervised machine learning models. With the use of this strategy, accuracy and adaptability can be improved. Additionally, NLP algorithms utilized to identify uncommon items. Noting that Named Entity Recognition models depend on precise PoS tagging from those models is also crucial.

C. Model Creation:

The test data is used to determine how effectively the device can predict new responses based on its training, while the training data and cross-validation data are used to ensure that the technique used to train the model is more accurate and effective.

D. Question Generation:

Each predicted question is evaluated and given a reward by the framework's evaluator model. The total model is built by discovering the variables that maximise reward; it is laborious to identify the solutions and then assess them.

Results and Discussion

The results gathered from the proposed system was assessed using Accuracy, Recall and Validity of the Questions.

The percentage of relevant instances that are retrieved is known as accuracy (A). It's a metric that counts how many accurate, favourable predictions were made during the execution.

$$A = Q_v/(Q_v + Q_i)$$

The percentage of accurately projected predictions made from all potential positive predictions is measured by the metric recall (R).

$$R = Q_v/(Q_v + Q_{ng})$$

The number of queries that the proposed system generates that are legitimate and meaningful is a measure of the system's validity.

Accuracy and Recall of the proposed work in each category are depicted in the graph. The overall accuracy is almost nearly 0.95% and recall is nearly 0.85%.

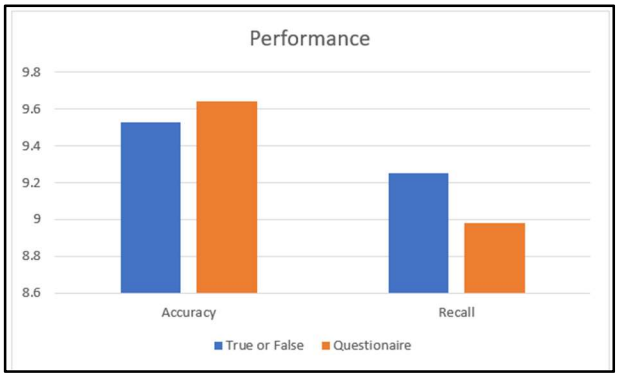

Figure 5 Performance measure of proposed work

Conclusion

The proposed work outlines an automation process that converts the manual process of producing paper into an automated system by granting restricted access to the materials. We have also given randomization's significance in the process of producing papers some thought. Our system features a powerful algorithm that eliminates recurrence of questions in the ensuing question papers, proving it impossible to spot any patterns in the papers. Our system uses tasks to differentiate between administrators and subordinates. As a result, the automation process for question paper creation that was created as a result offers enhancements in terms of safe platform, regulated access to resources, and random question paper generation. In the future, the ideal system will be capable to handle any PDF, Word, or other type of text file independently, evaluate it, and identify key sentences for QG.

References

[1] Beg and A. Beg, Using open technologies for automatically creating question-and-answer sets for engineering MOOCs, Comput. Appl. Eng. Educ. 26 (2018), 617–625.

[2] Michael Heilman and Noah, A Smith. 2009. Question generation via over generating transformations and ranking. Technical Report CMULTI-09-013, Language Technologies Institute, Carnegie Mellon University.

[3] Zhe Liu and B. J. Jansen, Subjective versus objective questions: Perception of question subjectivity in social Q&A, International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction, Springer, 2015, pp. 131– 140.

[4] Swart (2010). Evaluation of Final Examination Papers in Engineering: A Case Study Using Bloom‘s Taxonomy. IEEE Transactions on Education, (May 2010) Vol. 53, No.2 257-264.

[5] G.Deena and K.Raja , ―Sentence Selection Using Latent Semantic Analysis for Automatic Question Generation in E-Learning System‖, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-9, July 2019.

[6] Mannem Prashanth, Rashmi Prasad, and Aravind Joshi. (2010). Question generation from paragraphs at UPenn: QGSTEC system description. In Proceedings of QG2010: The Third Workshop on Question Generation, pages 84–91.

[7] Smith, S. ,Avinesh, P.V.S. and Kilgarriff, Adam (2010) 'Gap-fill Tests for Language Learners: Corpus-Driven Item Generation' Proceedings of ICON-2010: 8th International Conference on Natural Language Processing. held: 8-11 December 2010, Kharagpur, India. India: Macmillan Publishers.

[8] Sheetal Rakangor et al,‖ Computer Aided Environment for Drawing (To Set) True or False Objective Questions From Given Paragraph‖, Int. Journal of Engineering Research and Applications, ISSN : 2248-9622, Vol. 4, Issue 5(Version 3), May 2014, pp.105-108.

[9] lan Mitkov, Le An Ha and NikiforosKaramanis. 2006 A computeraided environment for generating multiple-choice test items, Natural Language Engineering 12(2): 177-194.

[10] Maor, and J. K. Currie, The use of technology in postgraduate supervision pedagogy in two Australian universities, Intl. J. Educ. Technol. Higher Educ. 14 (2017), 1–12.

[11] Karthikeyan S., Mohan B., Kathiresan S., Anbuchezhiyan G.,(2021),"Effect of process parameters on machinability, hemocompatibility and surface integrity of SS 316L using

R-MRAFF", Journal of Materials Research and Technology,Vol.15,no.,pp.2658-2672.doi:10.1016/j.jmrt.2021.09.060

[12] Brearley B.J., Bose K.R., Senthil K., Ayyappan G.,(2022),"KNN APPROACHES BY USING BALL TREE SEARCHING ALGORITHM WITH MINKOWSKI DISTANCE FUNCTION ON SMART GRID DATA",Indian Journal of Computer Science and Engineering,Vol.13,no.4,pp.1210-1226.doi:10.21817/indjcse/2022/v13i4/221304179

[13] Babu, G.N.K.S., Anbu, S., Kapilavani, R.K., Balakumar, P., Senthilkumar, S.R.,(2022),"Development of cyber security and privacy by precision decentralized actionable threat and risk management for mobile communication using Internet of Things (IOT)",AIP Conference Proceedings,vol.2393,pp.-.doi:10.1063/5.0074634

[14] Vijayan D.S., Rose A.L., Arvindan S., Revathy J., Amuthadevi C.,(2020),"Automation systems in smart buildings: a review",Journal of Ambient Intelligence and Humanized Computing,Vol.,no.,pp.-.doi:10.1007/s12652-020-02666-9

[15] Senthilkumar, S.R., Sureshbabu, G.N.K., Reena, R., Kannan, K.N., Balakumar, P.,(2022),"Intelligent therapy aided by advanced computational technology",AIP Conference Proceedings,vol.2393,pp.-.doi:10.1063/5.0074507

[16] Latha A.A., Anbuchezhiyan M., Kanakam C.C., Selvarani K.,(2017),"Synthesis and characterization of γ-glycine - A nonlinear optical single crystal for optoelectronic and photonic applications",Materials Science- Poland,Vol.35,no.1,pp.140-150.doi:10.1515/msp-2017-0031.

[17] Packiaslacer, P., Rajam, P.P., Theresa, M., Jacob, J.J., Benila, A.,(2022),"Iot based automation in the Manhole system",AIP Conference Proceedings,vol.2393,pp.-.doi:10.1063/5.0074220

[18] Amuthadevi C., Vijayan D.S., Ramachandran V.,(2021),"Development of air quality monitoring (AQM) models using different machine learning approaches",Journal of Ambient Intelligence and Humanized Computing,Vol.,no.,pp.-.doi:10.1007/s12652-020-02724-2.

[19] Slacer, P.P., Priyadharshini, R.I., Benila, A., Rajam, P.P., Jacob, J.J.,(2022),"A contemporaneous input vector monitoring Bist architechture using memory",AIP Conference Proceedings,vol.2393,pp.-.doi:10.1063/5.0074185

[20] Mohandass, G., Devi, K.S., Amalorpavaraj,(2010),"Antifungal activity of Vitex agnus castus-in vitro study",Biomedical and Pharmacology Journal,vol.3,pp.187-190.

[21] Rajesh G., Raajini X.M., Sagayam K.M., Dang H.,(2020),"A statistical approach for high order epistasis interaction detection for prediction of diabetic macular edema",Informatics in Medicine Unlocked,Vol.20,no.,pp.-.doi:10.1016/j.imu.2020.100362

[22] Sheeba Rani S., Kamatchi Sundari V., Subha Hency Jose P., Sivaranjani S., Stalin B., Pritima D.,(2020),"Enrichment of material subtraction rate on Eglin steel using electrical discharge machining process through modification of electrical circuits",Materials Today: Proceedings,Vol.33,no.,pp.4428-4430.doi:10.1016/j.matpr.2020.07.670

[23] Shirley D.R.A., Sundari V.K., Sheeba T.B., Rani S.S.,(2021),"Analysis of IoT-Enabled Intelligent Detection and Prevention System for Drunken and Juvenile Drive Classification",EAI/Springer Innovations in Communication and Computing,Vol.,no.,pp.183-200.doi:10.1007/978-3-030-59897-6_10