



International Journal for Innovative Engineering and Management Research

A Peer Reviewed Open Access International Journal

www.ijiemr.org

COPY RIGHT



ELSEVIER
SSRN

2022 IJIEMR. Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper, all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 25th Jun 2022. Link

[:http://www.ijiemr.org/downloads.php?vol=Volume-11&issue= Spl Issue 05](http://www.ijiemr.org/downloads.php?vol=Volume-11&issue= Spl Issue 05)

DOI: 10.48047/IJIEMR/V11/SPL ISSUE 05/20

Title **AUTOMATIC QUESTION AND ANSWER GENERATION USING NLP**

Volume 11, SPL ISSUE 05, Pages: 130-135

Paper Authors

Mr.O.Shiv Bhagwan,Sanku Tanuja, RavuriGopi , Rejeti Sandhya ,Tadikamalla Praveen



USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per **UGC Guidelines** We Are Providing A Electronic Bar Code

AUTOMATIC QUESTION AND ANSWER GENERATION USING NLP

Mr.O.Shiv Bhagwan¹,Sanku Tanuja², RavuriGopi³, Rejeti Sandhya⁴,Tadikamalla Praveen⁵

¹AssociateProfessor,Dept.ofCSE,²18ME1A0598,³18ME1A0590,⁴18ME1A0591,⁵18ME1A05A9
Ramachandra College of Engineering, A.P., India
cseplacements@rcee.ac.in, tanujasanku490@gmail.com, chinagopireddy@gmail.com,
sandhya.rejeti2@gmail.com, praveentadikamalla143@gmail.com

Abstract

Potential open doors for question-answer age have been proposed. The need of inquiries and answers is provoked for different reason, for example self-study, scholastic evaluation, and coursework. Be that as it may, the regular method for making question-answer matches has been both dreary and tedious. In this task we propose a programmed question age for sentences from message entries in understanding cognizance. We present a standard based programmed question age for the undertaking, as well as execute measurable sentence choice and different setups of named substance acknowledgment. Various sorts of inquiries like "what", "Who", and "Where" will be produced by our framework utilizing NLP. The framework performs well on producing inquiries from straightforward sentences. And afterward the responses are separated which are applicable to those inquiry and show them.

Introduction

People are interested in nature. We pose inquiries to acquire information and adjust the information in our day to day routines. Questions are essential wellspring of learning. Enormous counts of students are terrible at presenting requests. Dunlosky and Graesser states that understudies generally disapprove of recognizing their own understanding about the information they have. Consequently, they will more often than not pose less inquiries. Mechanization of inquiry age frameworks can help students to find the levels of their aptitude and increment their insight by making a difference them in explaining their questions. It can likewise be generally utilized in different fields like for placement exercises. It very well may be utilized by researchers for producing questions. Instructive foundations can utilize this framework making the

assessment arrangement of understudies more straightforward and simple. FAQs can be created for different applications, courses an for documentations.

Robotization of inquiry age frameworks help mechanized question addressing frameworks, for example, IBM Watson to perform self-preparing (IBM Watson Biological system, 2014).[3] As opposed to contingent upon human experts to characterize ground physically truth replies from the inquiries, programmed question age frameworks automatize this cycle. Savvy coaching frameworks can likewise use this benefit. Depending on human experts to physically separate inquiries from concentrate on materials takes time and is on occasion drawn-out; rather each end client can characterize its

own coaching system. Finally, robotized frameworks like this help the improvement of commented on datasets for question addressing and perusing cognizance. Question addressing (QA) is a PC science discipline inside the fields of data recovery also, normal language handling (NLP), which is concerned with building frameworks that consequently answer questions presented by people in a characteristic language.

There are three primary parts in a QA system that is Question handling, entry determination and reply extraction. There is a huge assortment of text information accessible in the question responding to framework as are the issues looked by the analyst and "question getting it" is one of the issues of the QA framework. Grammatical feature labeling is first level in any NLP application. The higher the amount of computational advances expected for choosing the pieces of discourse tag of a sentence, the higher is its unpredictability. Named Substance Acknowledgment (NER) is a fundamental component in NLP. Utilizing Grammatical feature tagger we track down the most ideal tag for each word in a sentence and comprehend it appropriately possibly it is thing, action word, and descriptor, etc. Out of numerous NLP libraries, NLTK is one of most remarkable library, it contains bundles which causes machine to figure out human language what's more, can give fitting reaction. Spacy is one of NLPs canny libraries, we use it for tokenization, lemmatization.

Related Work

There are several approaches to

accomplish question-answer generation. Du et al. introduced the usage of end-to-end fashion and deep sequence-to-sequence learning to generate questions for reading comprehension [4]. They implemented an attention-based sequence learning with two encodings, namely sentence encoding and paragraph-level information. Their system, Neural Question Generation (NQG), is able to generate "what", "when", "where", "who", "why", and "how" questions (without answers). There is also an ongoing recent work about question generation by Sarvaiya, A., whose framework we adopt and develop into our system [5]. The strategy used in this study is selecting important sentences, finding candidate gaps in sentences, then forming the interrogative sentences.

Problem Statement

1. Existing System

Programmed age of inquiries from text assumes a necessary part in different spaces. A standard strategy for robotized question age centers around the assessment of both semantic and syntactic organization of a sentence.

Different libraries, for example, Spacy, NLTK are utilized for tokenization, stemming, lemmatization, accentuation. Utilizing those libraries the framework performs sections of text as info and creates an inquiries as result. The inquiries will be produced in view of the given text.

The framework have modules like inquiry passage, question the board, paper age, paper the executives. The framework can be utilized by schools, universities, instructing establishments for producing Question paper.

Disadvantages of existing System:

1. In this system three type of WH-questions are generated i.e., "What", "Where", "Who".
2. This process may not yield exact accurate results.

2. Proposed System

In proposed framework, here we are going to creating responsive matches. The inquiries are produced and the responses are separated from sentences in a given message.

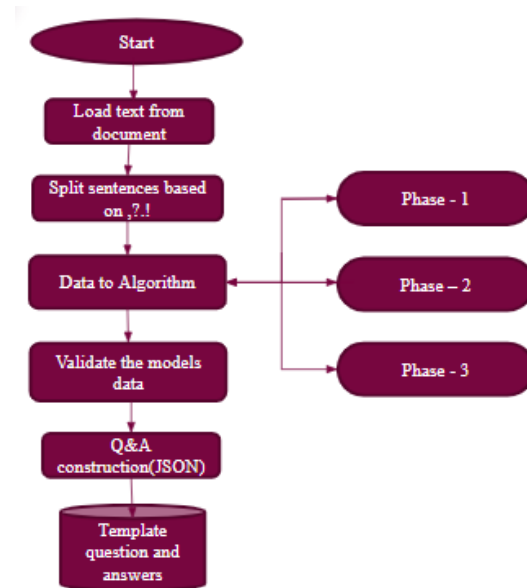
Utilizing NLP libraries, the proposed framework performs actually using passages of text as info and produces question-answer pair as result. In this we are adding a module that a code is produced alongside age of inquiry answer matches, the inquiry paper is simply noticeable to the client who are taking an appraisal by entering a specific code.

The framework characterizes a few modules, for example, client organization, subject choice, detail, question-answer passage, question-answer the executives, paper age, paper the board. This framework can be utilized by schools, universities, training establishments for producing Question paper and furthermore by understudies for self-evaluation. The framework can likewise be utilized by colleges to make question paper for understudies.

Advantages:

1. The system mainly focuses on Pre-processing of Data.
2. The accuracy obtained in this model is nearly 93.1%.

System Architecture



Methodologies

Techniques used in this system

1. SPACY

At the point when we are working with an enormous measure of text, we will ultimately need to find out about the text. Questions like: What does the words mean in the sentence, How would they act together to give a significant sentence? , Which texts are like one another, etc. Spacy is explicitly worked to cycle and assist us with seeing enormous volumes of text. Spacy system is written in Cython, and is a very quick library. It gives admittance to its strategies and capacities which are educated by artificial intelligence/AI models. In its bundle spaCy contains various models which contain the data about vocabularies, prepared vectors, punctuations and elements. It gives elements to numerous regular language

assignments, can be utilized to construct data extraction frameworks. These models are to be stacked into our code to get to them. Following is an instance of stacking the default bundle "english-center web":

```
>>>import spacy
>>>nlp=spacy.load("en")
```

2. NLTK

Natural Language Processing is manipulation of information, literary substance or discourse through any program or gadget. An relationship is that individuals have connection, comprehend each other point of view, and answer with a fitting arrangement. In NLP, this exchange, understanding, the response is made by means of a PC rather than a human. NLTK represents

Natural Language Tool Kit. This tool kit is one of the most striking NLP libraries which integrate projects to make machines see human language and answer it with a reasonable response. It gives various libraries to message handling, arrangement, tokenization, stemming, and labeling, marking, parsing, and semantic thinking. Tokenization in NLP is more hearty. It comprises of breaking the given sentence into tokens and accentuations prior to handling the information. Tokenization is done as displayed:

```
>>>text="I was absent yesterday!"
>>>tokens=nltk.word_tokenize(text)
>>>tokens
['I', 'was', 'absent', 'yesterday', '!']
```

3. POS TAGGING

POS labeling is the initial phase in any NLP based application. Labeling is a kind of class that might be portrayed as the electronic task of portrayal to the tokens. Here the descriptor is known as tag, which can likewise address one of the part of- talks, semantic measurements, etc. Presently, in the event that we discuss Grammatical feature (POS) labeling, it might be deciphered as a technique for allotting grammatical features to the given word. It is for the most part called POS labeling. In simple phrases, we can say that POS labeling is an endeavor of naming each expression in a sentence with its suitable name of discourse. We definitely know that components of discourse incorporate things, action word, modifiers, descriptors, pronouns, combination and their sub-classifications. Taggers utilize different information of the structure: word references, vocabularies, rules, etc. [6] The greater part of the POS labeling falls under Rule Base POS labeling, Stochastic POS labeling and Change based labeling. Label set is an assortment of labels utilized for a specific task. Each tagger will be given a standard label set. The tag set might be coarse like NN (Thing), VB (Action word), JJ (Modifier), RB (Intensifier), IN (Relational word), and CC (Combination, etc. [6] coming up next is a model on how tokenization happens, where words are

tokenized in view of assuming they are formal person, place or thing/individual thing, and so forth.

```
>>>import nltk
>>> nltk.pos_tag (nltk.word_tokenize("Hey, how are you
doing?"))
[('Hey', 'NNP'), ('.', '.'), ('how', 'WRB'), ('are', 'VBP'),
('you', 'PRP'), ('doing', 'VBG'), ('?', '.')]

```

Result

Figure 2 shows test question-answer matches created by our proposed framework, as well as its setting sentence. The initial five inquiry answer matches show effective inquiry age; the last pair, in any case, presents a fizzled question age.

Questions created utilizing our framework effectively changed over straightforward revelatory explanations into questions, however for additional mind boggling sentences, our concocted rules show challenges in switching more mind boggling proclamations over completely to questions. This is because of an execution blunder — the standards to switch an assertion over completely to an inquiry are made in-house, rather than executing a generally settled rule-based question-age framework, e.g., Heilman and Smith [7].

One more significant impediment of our framework is its restricted pool of conceivable inquiry words, i.e., just WH-questions are produced, and just inquiries which don't need complex semantic examination (for example "Why", "How") can be created.

Sentence: The building was on fire, and it wasn't my fault.
Question: What was The building on?
Answer: fire

Sentence: Distant streetlights provided the only light in the dusty hall, and left huge swaths of blackness crouching in the old classroom doors.

Question: What provided the only light in the dusty hall, and left huge swaths of blackness crouching in the old classroom doors?

Answer: Distant streetlights

Sentence: I sprinted around a corner and toward the exit doors to the abandoned school building on the southwest edge of Chicagoland.

Question: Where did I sprint around a corner and toward the exit doors to the abandoned school building on the southwest edge of?

Answer: Chicagoland

Sentence: The movie of my life must be really low-budget.

Question: What must be really low-budget?

Answer: The movie of my life

Sentence: There is, I think, humor here which does not translate well from English into sanity.

Question: What does does not do?

Answer: sanity

Fig 2. Sample generated question-answer pairs

Conclusion

We have introduced a standard based programmed question age framework for understanding perception. We use numerous strategies to choose important sentences in a passage, then utilize named substance acknowledgment and constituent parsing to create conceivable inquiry reply matches. The sentence is then changed to an inquisitive structure in view of a bunch of rules and potential responses. The framework performs well on straightforward sentences, yet flounders on more intricate sentences. Our inquiry answer age frame work can deliver WH-questions.

Future Work

Future upgrades to the framework incorporate executing a more powerful, currently settled rule-basedquestion0.25-age framework, e.g., Heilman and Smith

[7].Another choice is to execute a profound brain organization engineering for the inquiry age itself, e.g., utilizing succession to-arrangement.

References

[1] Holy Lovenia, Felix Limanta, AgusGunawan -*Automatic Question-Answer Pairs Generation from Text*. November 2018

[2] PritiGumaste, Shreya Joshi, SrushteeKhadpekar, Shubhangi Mali-*AUTOMATED QUESTION GENERATOR SYSTEM USING NLP LIBRARIES*. June 2020

[3] Onur KEKL – *Automatic Question Generation Using Natural Language Processing Techniques*. July 2018.

[4] X. Du, J. Shao, and C. Cardie, “Learning to Ask: Neural Question Generation for Reading Comprehension,” in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2017.

[5] A. Sarvaiya- “Using Natural Language Processing for Smart Question Generation,” Intel AI Academy, Jul. 2018.

[6] AmrutaUmardand, Ashwini – *A survey on Automatic Question Paper Generation System*, International Advanced Research Journal in Science, Engineering and Technology (IARJSET), January 2017.

[7]M. Heilman and N. A. Smith - *Question Generation via Overgenerating Transformations and Ranking*,2009.