

Question and Answer Extraction Using NLP

Akash Shekar

UG Scholar, Department of Computer Science and Engineering
Velammal Engineering College, Chennai.

Jeevanantham.P

UG Scholar, Department of Computer Science and Engineering
Velammal Engineering College, Chennai

-----**ABSTRACT**-----

A website which is used for the extraction of question and answer using NLP. This website is used to extract information about a given topic or word from the Wikipedia and generates questions and answers for the given topic. The challenge is to make use of ML and NLP for question and answer generation from a topic or paragraph. Systematic reviews require expert reviewers to manually screen citations in order to identify all relevant articles. Our proposed method uses a question extraction based on neural network vector space model. With the paragraph vector method -discover latent topics & explains content of each topic by meaningful and comprehensive text labels. In the existing system, the NLP is used to search for only keywords and characters but in the current system it generates questions and answers for given topic. The website uses the languages Python, Java, PHP, HTML. The database is done with SQL Database. This website gives a simple UI Environment.

Keywords- Question Extraction, Answer Generation, Wikitrivia, topic, Python, NLP

-----**I. INTRODUCTION**

There are many Pages such as Wikipedia for obtaining information about a topic, but the usage of those data and information is utilized by many people for generating questions and answers. So the basic need of those data and information is needed by teachers for creating question papers and generation of answers for those questions.

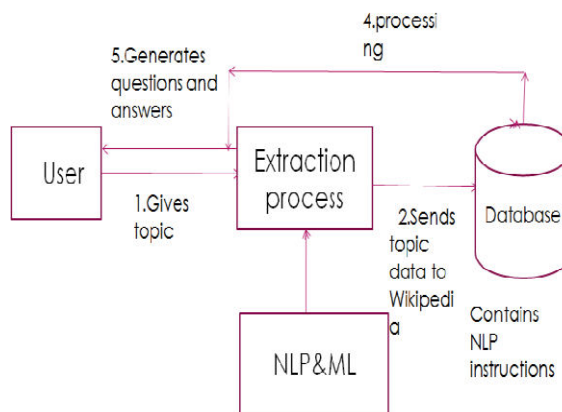
II. RELATED WORK

Technologies are developed and there is a very high rise in use of smart devices. Thus current generation students are taking online assessments and tests to evaluate themselves and queries they have are all being posted and answered in social network. Because of the rise of internet many organizations have come forward in development of online assessments and test sites to improve students and professionals skills. This website is available for students and people to practice the questions generated from a particular topic and answering the questions to know if it is correct or wrong.

III. PROPOSED SYSTEM

To develop a generalisation model that uses NLP and ML to extract questions based on the content. Also it gets the users answer and tells if it is correct or wrong based on the content using the question and input text. In the existing system, the NLP is used to search for only keywords and characters but in the current system it generates questions and answers for given topic. This system uses the extraction of questions in a comprehensive manner with options given to that question.

JS: program the behavior of web pages. JavaScript enables interactive web pages and is an essential part of web applications.



Python: Python is a general purpose and high level programming language. You can use Python for developing desktop. GUI applications, websites and web applications. Also, Python, as a high level programming language, allows you to focus on core functionality of the application by taking care of common programming tasks.

NLP: Natural language processing (NLP) is the ability of a computer program to understand human language as it is spoken. ... Human speech, however, is not always precise - it is often ambiguous and the linguistic structure can depend on many complex variables.

HTML: Hypertext Markup Language, a standardized system for tagging text files to achieve font, colour, graphic, and hyperlink effects on World Wide Web pages.

CSS: Cascading style sheets are used to format the layout of Web pages. They can be used to define text styles, table sizes, and other aspects of Web pages that are **defined** in a page's HTML.

IV. IMPLEMENTATION

The proposed system is implemented using Python in the back end and uses HTML,CSS in the front end. This website uses Paragraph vector model.With the paragraph vector method -discover latent topics & explains content of each topic by meaningful and comprehensive text labels. When the user enters the website,the user searches for a topic in the topic search engine. The questions about the particular topic is generated from the Wikipedia page of that topic. The options of answers to that topic is also generated. When the user selects the particular option it displays whether the given option is correct or wrong. It searches for the particular topic intensively. The website uses the NLP application for processing the topic and generating the data from the Wikipedia page. The questions are extracted the Wikipedia database and then displayed on the website for the user.

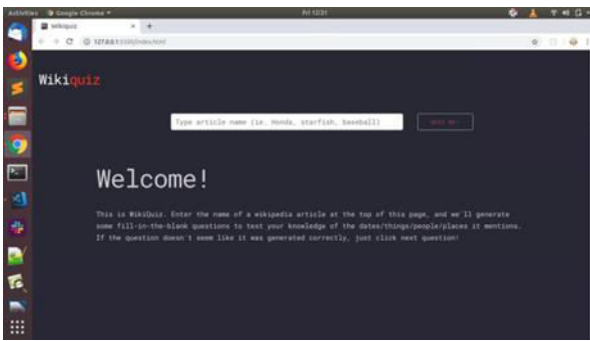


Fig 1. Welcome Page.

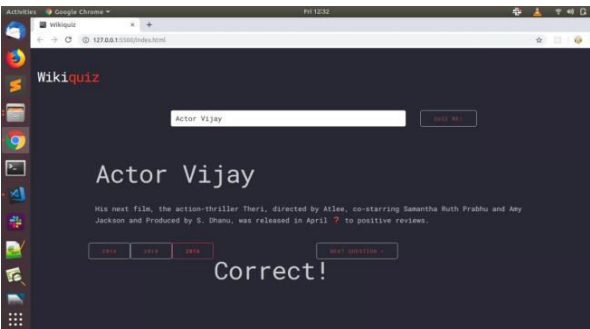
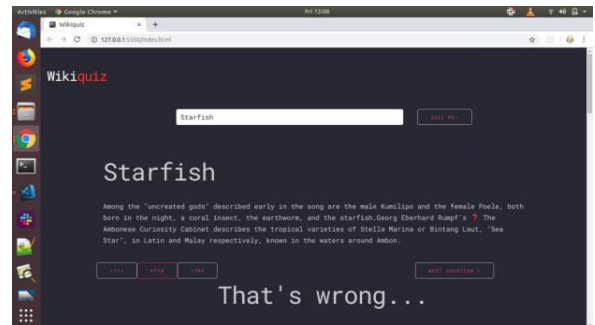
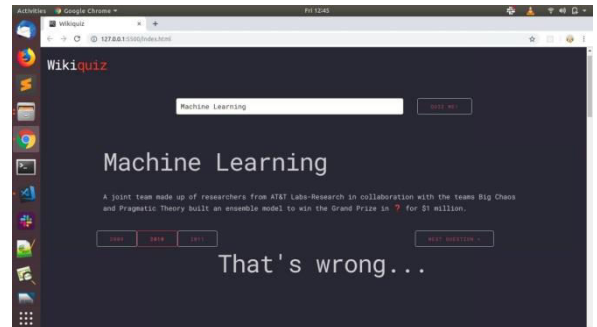
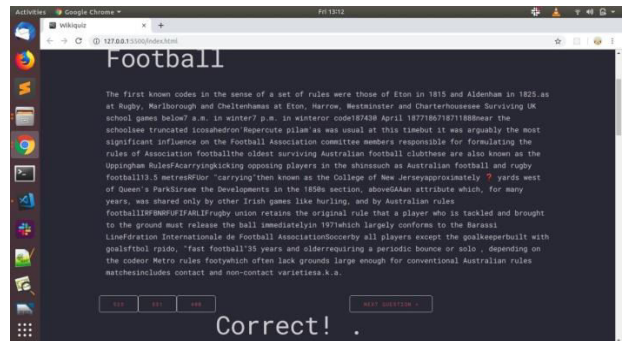
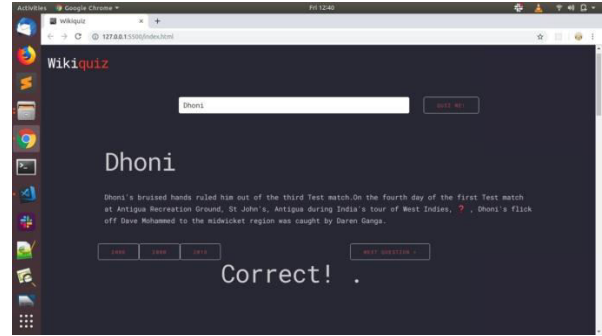
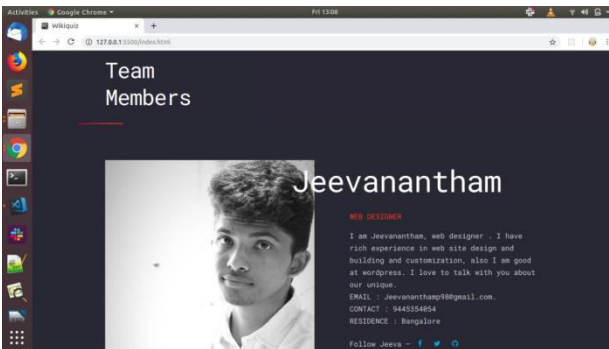
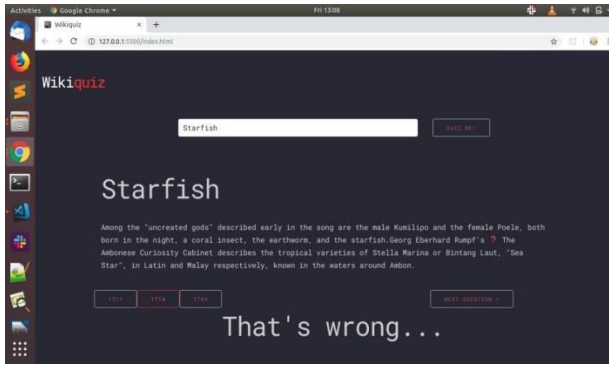


Fig 2.Question Page.





V. CONCLUSION

The website helps mainly for the student community and teacher's community to generate questions with answers as options to check if the answer is right or wrong. Thus this website enables us to analyse a given topic in a comprehensional manner. It also helps the teachers in

creating questions from a given topic and the students to study and revise from the topic.

VI. FUTURE ENHANCEMENTS

- **Choosing more appropriate multiple choice options, especially for numbers:** Ignoring the less text heavy parts of a Wikipedia page. Creating more interesting grammar for the text chunker, which would lead to more interesting question types.
- **Interpret the context of sentence in question itself:** Some of the questions presented currently lack context about what the question is referring to. A further version of the project would attempt to interpret the context of the sentence in question and include that in the question.

REFERENCES

- [1] Using Wikipedia and Conceptual Graph Structures to Generate Questions for Academic Writing Support. <https://ieeexplore.ieee.org/document/4740489>.
- [2] Extracting Concept Hierarchy Knowledge from the Web Based on Property Inheritance and Aggregation. <https://ieeexplore.ieee.org/document/4740489>
- [3] Recent Trends in Deep Learning Based Natural Language Processing. <https://ieeexplore.ieee.org/abstract/document/8416973>
- [4] Answering Natural Language Questions by Subgraph Matching over Knowledge Graphs <https://ieeexplore.ieee.org/abstract/document/8085196>
- [5] Visual Question Answering with Compressively Sensed Images. <https://ieeexplore.ieee.org/document/8451445>