A Synopsis

on

NLP BASED INTELLIGENCE

in partial fulfillment of the requirement for the degree

of

Bachelor of Technology

In

COMPUTER SCIENCE AND ENGINEERING

Submitted by

Sivasish Koch(1513310214)

Saif Ali(1513321165)

Rahul Bilra(1513310161)

Akash Kandpal(1513310027)

under the supervision of

Dr. Deepti Gupta

(Associate Professor Computer Science Dept.)

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY

GREATER NOIDA

Index

Sr.No.	Topics	Page No.
1	Abstract	1
2	Existing System(Literature Survey)	2
3	Problem Statement	3
4	Objective of the project and Features	4
5	Resource Requirements	5
6	Conclusion	6
7	References	7

Supervisor Sign:

(Sign of supervisor on synopsis is mandatory while submission)

ABSTRACT

Today people are very busy they want to get results in lesser time. So we thought of reducing the time taken by a human being in going through a book, be it a student's class book or a professor’s PHD thesis. We want to create an NLP backed Deep Learning Model which can answer all your queries after going through the data (book/ notes fed into the web app). It can perform much better than searching queries in Google as the results would be made available from the data presented by you, also as you know how the intention of question changes from person to person. For eg, bank means river bank in a certain context and it also means a financial institution that accepts deposits from the public. Thus we came up with this unique idea.

MODIFICATION AND IMPROVEMENT OVER THE EXISTING IMPLEMENTATION

Present State:

- Software present currently are limited in scope as they only deals with user providing a query and in response it scrapes the most relevant data from wikipedia and gets the response.
- Currently the systems mainly uses Google API's for covering wide area and hence unable to give answers specific to users.
- Non-ability to detect handwritten notes and query on them.
- Summarization of text has been not been implemented as of yet .

After implementation of project:

- Easy to use GUI.
- Answer to queries will be very much relevant and specific to the user.
- It can summarize text for input data (fed in the form of pdf/txt/doc) page-wise/chapter-wise/whole-book.
- Handwritten notes can also be fed in the form of pdf (which comes under Computer Vision part) which will be much beneficial for students in getting quick responses for their questions, especially during exam preparations .

PROBLEM STATEMENT

Currently, we as a human want to get results faster and we have scarcity of time. So , when we (as a student) have loads of data provided by our professors which is to read and we don't have that much time to go through all of it, wither we skip it or resort to google for the information which gives results on the basis of page ranks and popularity of links. We wanted data which is relevant to the topic or subject we were going through in that scenario google doesn't helps.

Consider a case, where we are given an unseen passage which is not even popular thus google won't help. So where to find answers to questions which are not popular and one has data for that, one need to get someone else to go through the data and give them the data which is relevant to them as the user don't have time.

In cases, where we have handwritten data, we don't have softwares which can even detect it properly properly. So, loads of information which could have been beneficial cannot be automated.

OBJECTIVE

The following are the objectives of the A.I. :

1. ML model will be able to read and understand the book completely and it will provide an answer from the book itself thus saving you lots of time from reading the whole book.

2. It can answer questions on most popular books/ holi-books/novels /comics with much accuracy and score.

3. It can summarize text for input data (fed in the form of pdf/txt/doc) page-wise/chapter-wise/whole-book.

4. The performance will be better than searching queries on Google as it would be domain specific.

5. Handwritten notes can also be fed in the form of pdf (which comes under Computer Vision part) which will be much beneficial for students in getting quick responses for their questions, especially during exam preparations .

6. One can get answers to their questions from a book without reading the book, all they have to do is to upload the book into the webapp.

New features can be added as per requirements.

This software package can be readily used by non-programming personal avoiding human handled chance of error. This project is mainly used by two types of users :

i. Students.

ii. Researchers

TOOLS AND TECHNOLOGY USED

Tool and Libraries:

- Gensim : https://radimrehurek.com/gensim/
- Fasttext : https://fasttext.cc/
- NLTK : https://www.nltk.org/

Technology:

- Django : Web App
- REST API's : For hitting calls between ML model and app.
- Tensorflow : As a base for Keras and more optimization.
- Keras : For making Deep Learning Models
- Pandas : For cleaning the data
- Numpy : For mathematical purposes.
- Plotly : For visualising graphs.
- Scikit : For Shallow learning Algorithms.
- SHLDA : For Topic Modelling purposes.
- Open CV : For computer vision part.
- Convnet : For Handwritten Notes detection.
- Heroku : For Deployment of Web App.

CONCLUSION

Our effort has been to make it easier for students, researchers and other people who have lots of data and don't get time to read it. They can easily get relevant information from their data. This has never been approached before as previously people have tried to get the most relevant information from Google API's and thus ignoring the individuality of every person.

Also, the approach involves further advancements like usage of Deep NLP and Computer Vision for detecting the words and getting relevant information out of it. Often a hybrid approach, judiciously blending apparently different techniques, provides improved results in the form of faster speed, increased relevancy, and higher precision and recall measures.

REFERENCES

Research Papers involved are :

- Green BF, Wolf AK, Chomsky C, and Laughery K. Baseball: An automatic question answerer.
- Weizenbaum J. ELIZA - a computer program for the study of natural language communication between man and machine.
- Woods W. Progress in Natural Language Understanding - An Application to Lunar Geology.
- Bobrow DG, Kaplan RM, Kay M, Norman DA, Thompson H, and Winograd T. Gus, a frame-driven dialog system.
- Katz B. Annotating the World Wide Web using natural language.
- Clark P, Thompson J, and Porter B. A knowledge-based approach to question answering.
- Riloff E and Thelen M. A Rule-based Question Answering System for Reading Comprehension Tests.
- Reading Comprehension Tests as Evaluation for Computer-Based Language Understanding Systems, Vol. 6, 2000, pp. 13-19.
- Ittycheriah A, Franz M, Zhu WJ, Ratnaparkhi A and Mammone RJ. IBM's statistical question answering system.