Atlantis

Problem Statement

An application that extracts the title, authors, and abstract from the pdf of a research paper.

Milestones

Understanding and analyzing the problem

Finding Title and Author

Creating Android Application for extraction

| 9:00 | 11:00 | 13:30 | 17:00 | 19:30 | 21:00 | 02:00 | 05:00 | 08:00 | 11:00 | 12:00 | 13:30 |

PDF to text conversion

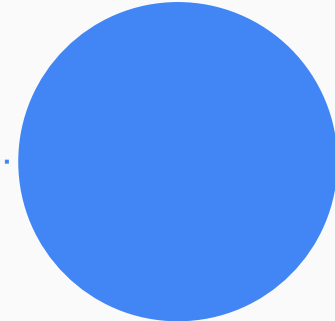Finding or creating abstract
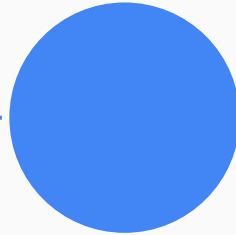
How it works

PDF 2 TXT

We need to convert .pdf to .txt to extract the information

Finding Title, Author and Abstract

Using ML concepts we find these info

Uploading the .JSON file and reading data

Output is extracted in JSON format.

- Converted PDF to Text (.txt) using pdfminer library.
  - Any kind of research paper pdf can be converted.
- Find the "Title" using pattern matching.
- Find the "Authors" using POS - tagging.
- Find the "Abstract" using text summarizer.
- Find the "References" using pattern matching again.
- Android application using android studio.

Title, Authors, Abstract and References

- The title from a PDF is one-liner heading which also gives the idea of a problem. So using this knowledge, we make a pattern matcher and find the title.
- The authors on the other hand are proper noun(NNP). Using POS-tagging we can find all the proper noun and find the authors
- The abstract can be given or we have to make one. So we search for the given abstract first. If found we use it, else text summarizer makes abstract.

- The references are given at the end and so we search for it if it given. If so we print it, else 'reference not found'.

Android Application

- The file is to be selected using a button that browses through the user's phone.
- The pdf selected is stored and uploaded and path is used to extract the information and the output, i.e. in the form of .json, is read and the output is presented.

Conclusion

- The pdf research paper when selected gives directly the title, abstract and references so that the user can get idea of the paper.

The Team

1. Chirag Bajaj
2. Shivani Patel
3. Vaidehi Desai
4. Kandarp Kakkad

Thank You!!!