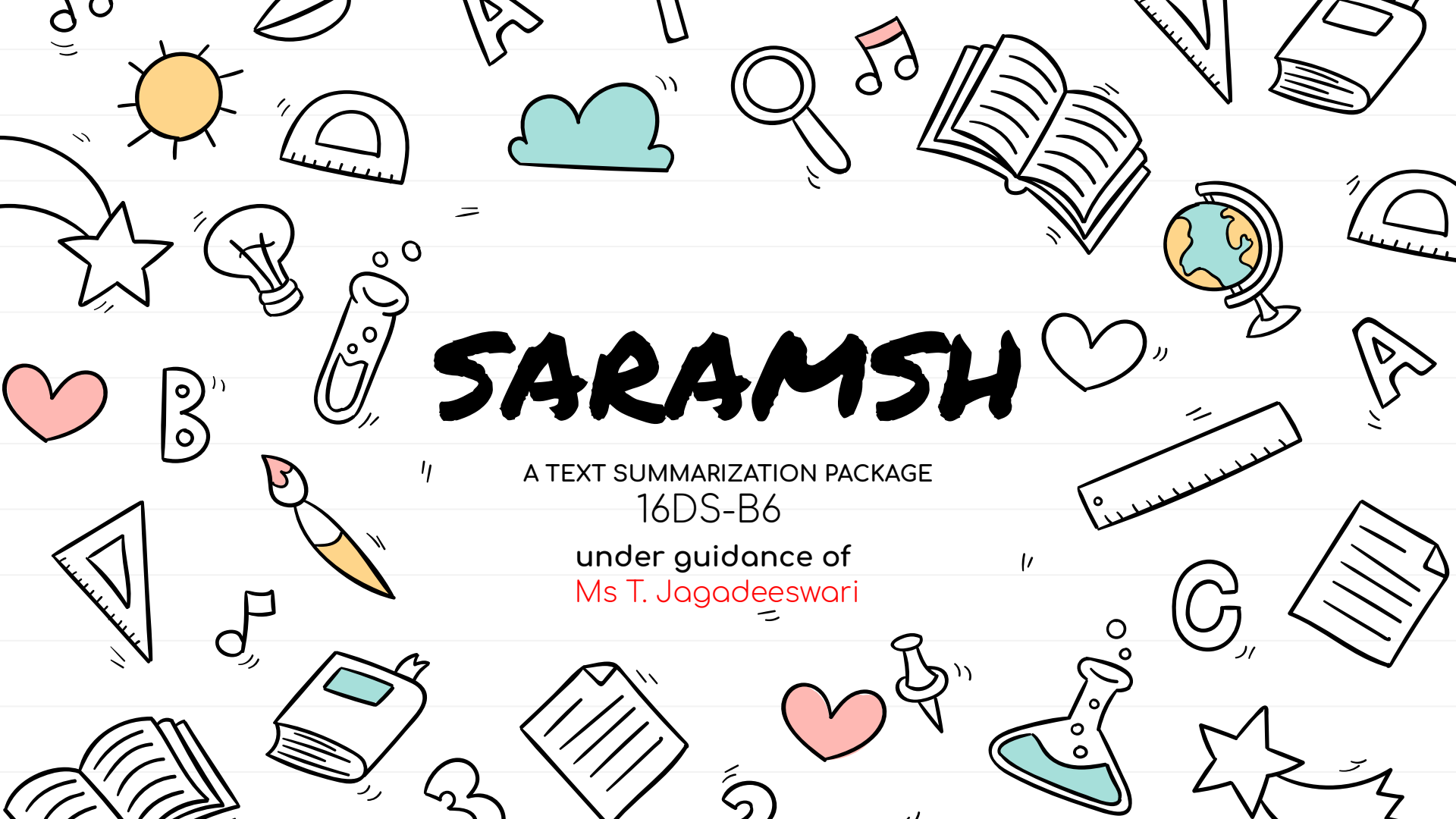

SARAMSH
A TEXT SUMMARIZATION PACKAGE
16DS-B6
under guidance of
Ms T. Jagadeeswari

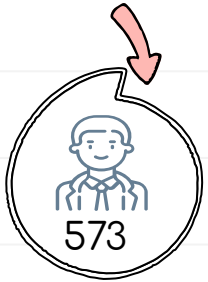
Team Members

KARTHIK

Developing the TF-IDF from scratch

PRUDHVI

Building the final Summarizer package



SATHVIK

Testing and Debugging

VIVEK

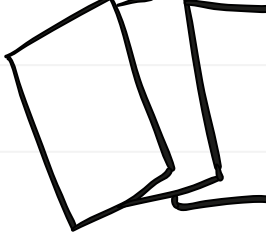
Writing documentation for the code

TABLE OF CONTENTS

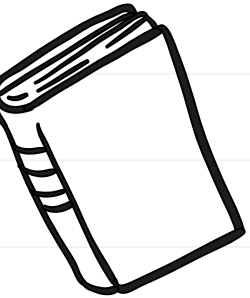
How we arrived at this idea ?

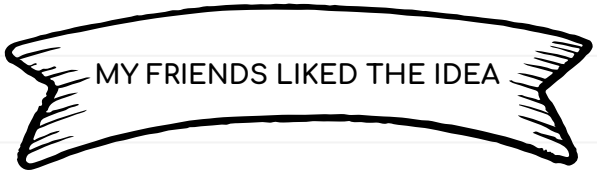
IT ALL STARTED WITH AN EMAIL

Once I got an email which contained a huge chunk of text. It was forwarded to me by a person very dear. It took me 20 mins to completely read the email and understand it. It would have been easy, if there was a summary of that chunk of information!...



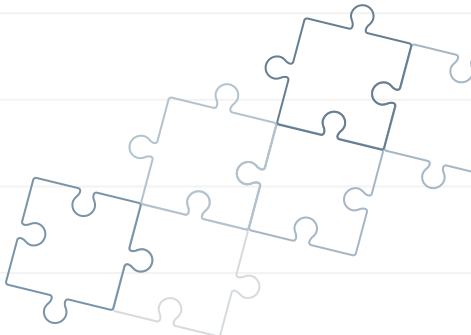
" Then I thought why not build a program that could take any chunk of text and generate a summary! "





THINGS THAT WE FOUND OUT

- **Text Summarization is not a new idea. And it is already a challenging and open problem in NLP.**
- Developing this as a package/API would be more useful.
- There are many advanced tools and packages that are available for doing text summarization. But most of them either have many dependencies or take much time and resources to run.

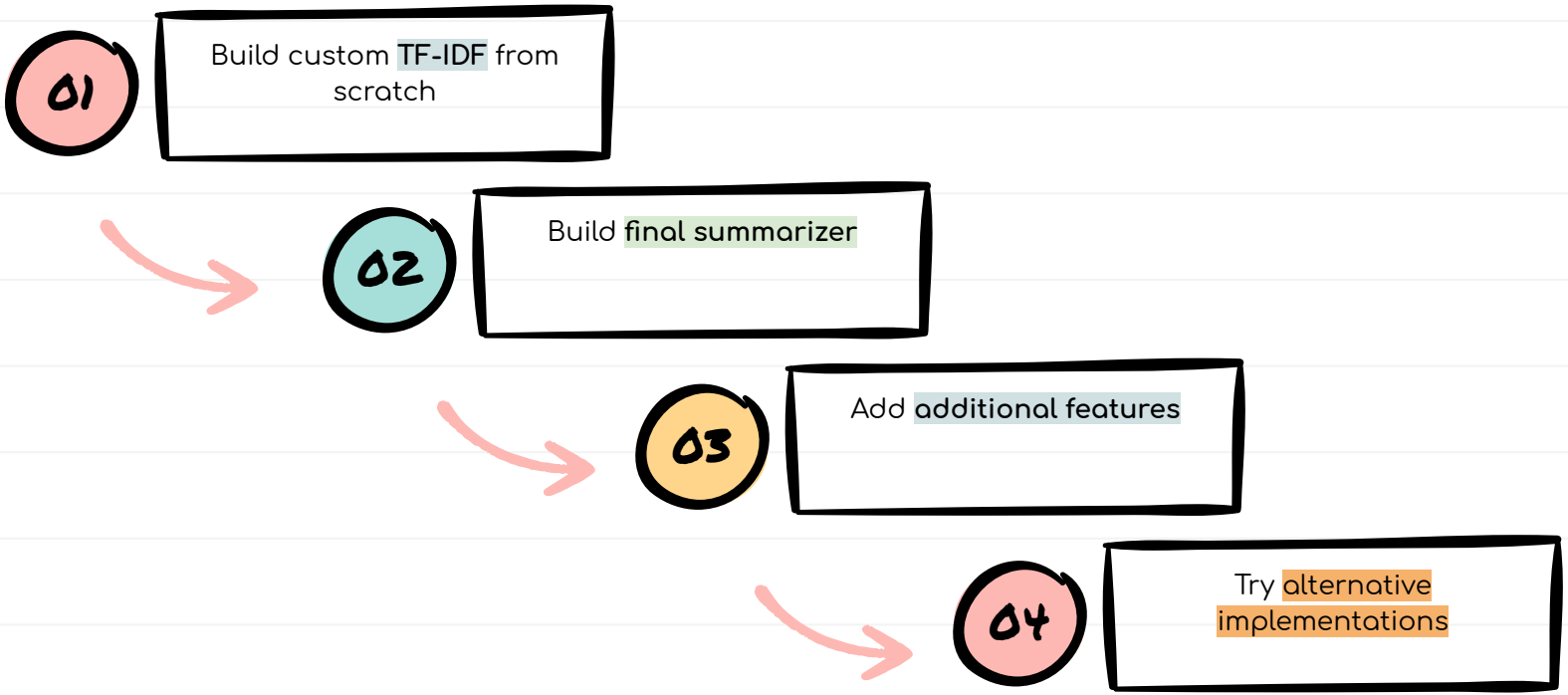


OUR GOALS!

- To develop the text summarizer as a **python package**.
- To enable the summarization of large amounts of text in a very **less amount of time**.
- To write the code from **scratch**.
- To make this project **open-source**.
- To continue the development of this package for as long as possible **incorporating new advancements** if needed.

What are we gonna do ?

OUR PLAN


01
Build custom TF-IDF from scratch
02
Build final summarizer
03
Add additional features
04
Try alternative implementations

DEFINITION OF CONCEPTS

TF-IDF

It is the heart of the project. It is basically a scoring mechanism which indicate how important is a word in a document.

FINAL SUMMARIZER

This module uses TF-IDF output to score the sentences and generate final summary.

ADDITIONAL FEATURES

Things like generating summary of text extracted from wikipedia about a given term.

ALTERNATE IMPLEMENTATIONS

Using other tools instead of TF-IDF and measuring their **rouge score.**

Why things should be the way they are planned ?

REASONS TO JUSTIFY OUR APPROACH

UNSUPERVISED LEARNING

Our approach do not need seperate training, it works directly on the given text and generates summary. Supervised learning needs lot of training data.

EXTRACTIVE IS BETTER

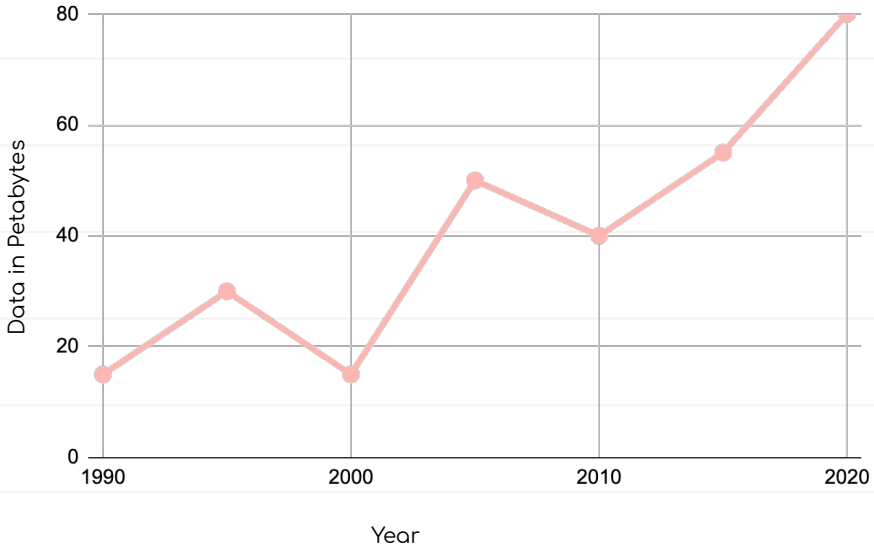
Abstractive summarization is generally not preferred. Our approach follows an extractive text summarization technique.

THERE IS A LONG WAY TO GO

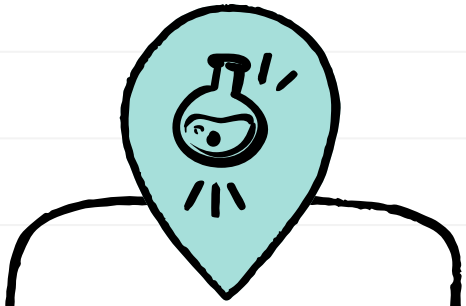
Text summarization is an open problem and there is a long way to go in this area. And the problems that we can solve using this approach are huge in number.

Where it can be used?

INCREASING DATA

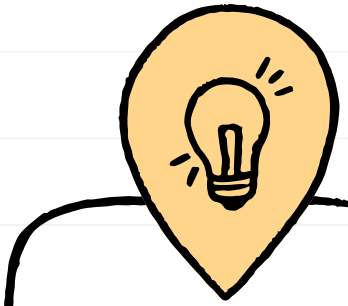


There is going to be a lot of textual information generated in the near future.



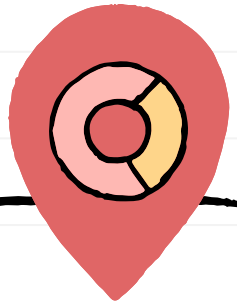
Querying Summaries

We can store the huge chunks of text in our database and query the summary for a particular index. This is a possibility in the near future.



Generating Reports

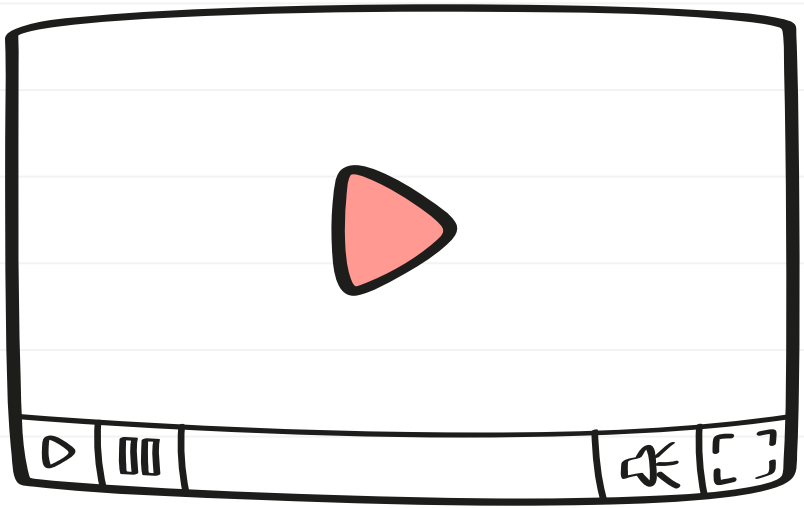
We can use text summarization to generate reports which could help in financial research.



UNDERSTANDING LITERATURE

With the help of text summarization we could understand lots of books easily. And also these summaries can be used as input to advanced AI systems.

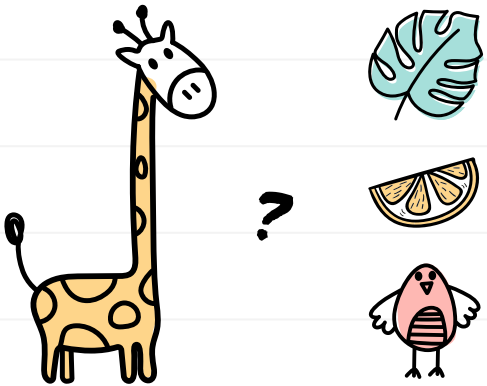
We can also summarize a video by generating a summary out of the video`s transcript.




THANK YOU!...
ANY QUESTIONS ?

IN WHAT SENSE DOES IT COME UNDER THE THEME OF SMART CITY ?

We believe that our tool would help in development of (ICT) Information and communication technologies by broadly improving the **operational efficiency.**



" It need not to be a conventional project, but it can be also a tool that promote the development of various other projects under the same domain or area."

HOW ?

HELP DESKS

COMPLAINTS GENERALIZATION

CROSS LANGUAGE SUMMARIZATION