Internship 2020

A.L.I.C.E.

Cornelius Yap, Tiffany Goh, Samuel Lye

Agenda

1. **Overview of A.L.I.C.E.**
   - Background - Analyzing Language Interface Created for Everyone (A.L.I.C.E.)

2. **Key Components**
   - Frontend & Visualization Tools
   - Machine Learning Algorithms
   - DevOps (Docker & OpenShift)

3. **Live Demo**
   - Walkthrough of Use Case
   - Live Demo

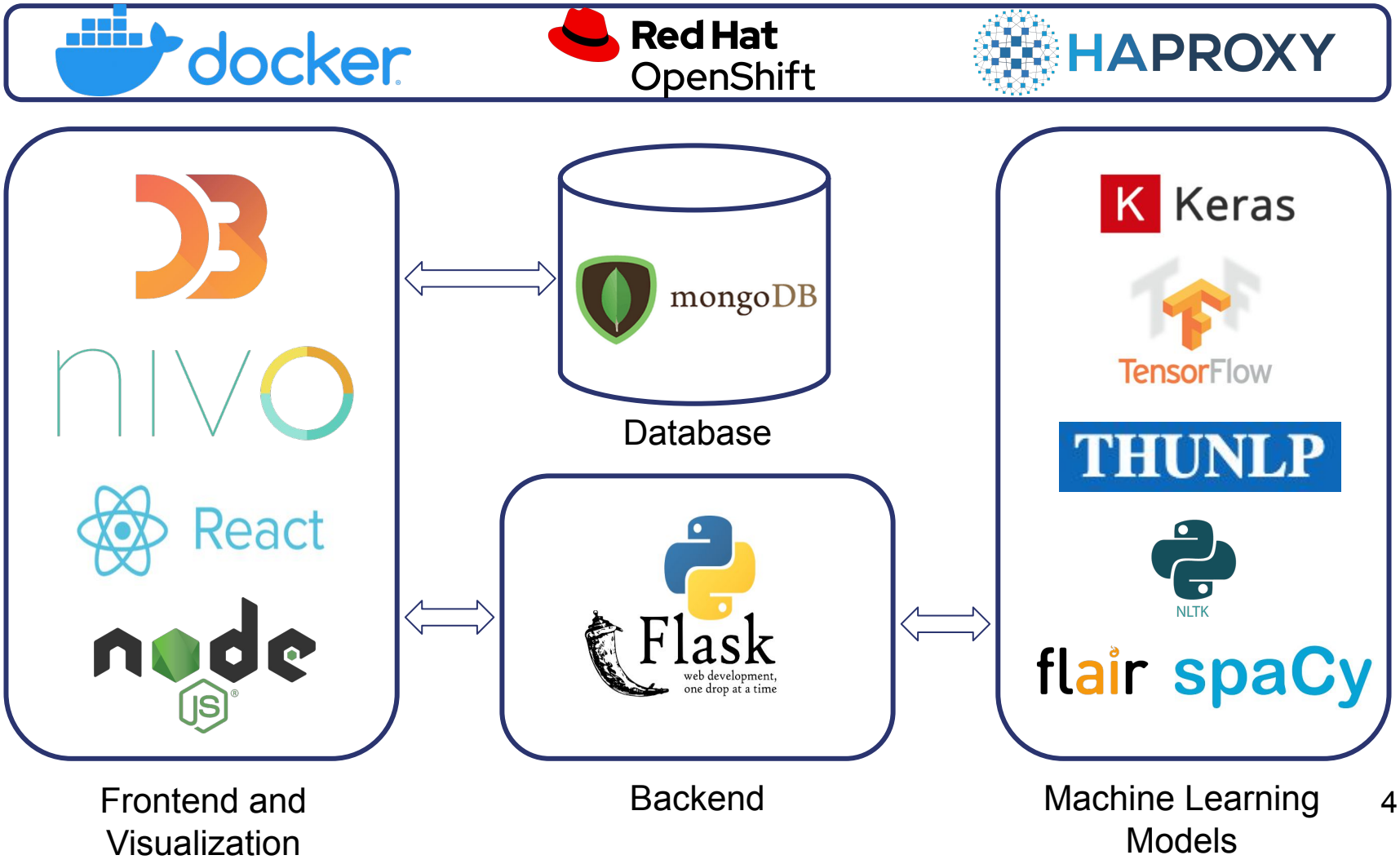
Background

- Automatic text summarization methods are greatly needed to address the ever-growing amount of text data available online to both better help discover relevant information and to consume relevant information faster.

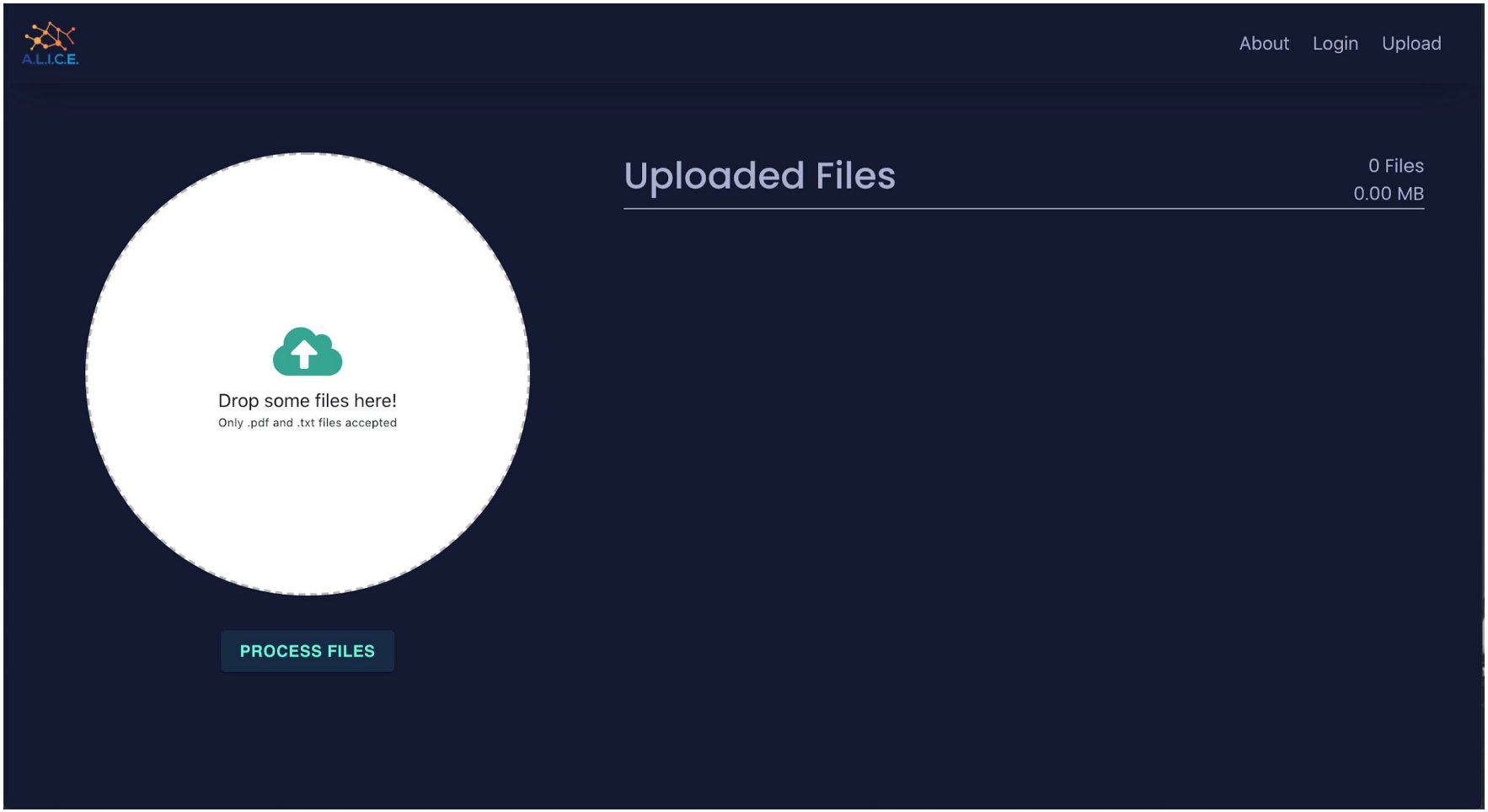
- Analyzing Language Interface Created for Everyone (A.L.I.C.E.) is purposed to summarize text documents and output informative visualization displays that quickly and easily communicate the contents of the text to the user.

- The key components of A.L.I.C.E. are the Frontend (React, D3), Backend (Flask, Machine Learning models), as well as DevOps (Docker, Openshift).

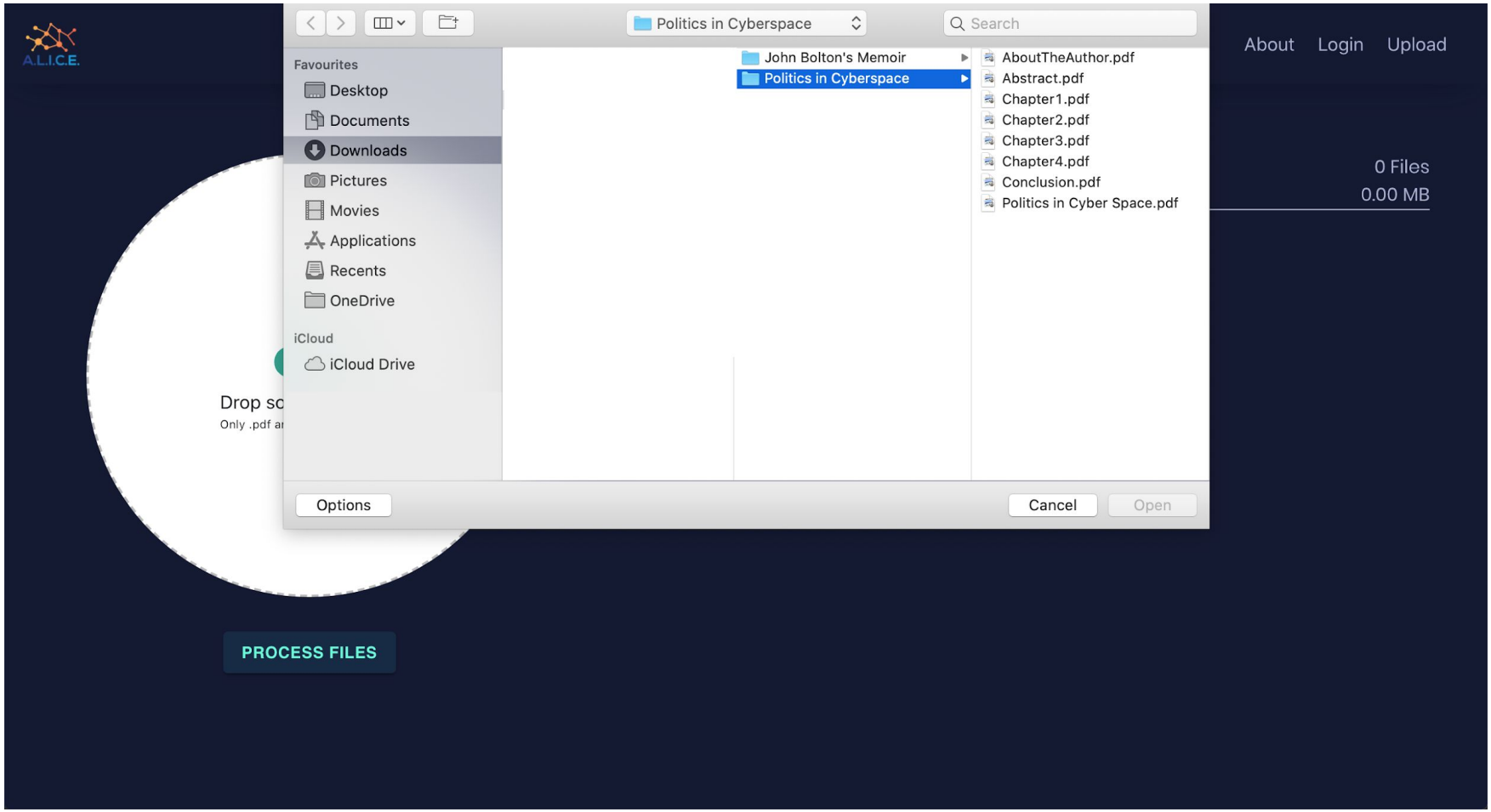
Key Components



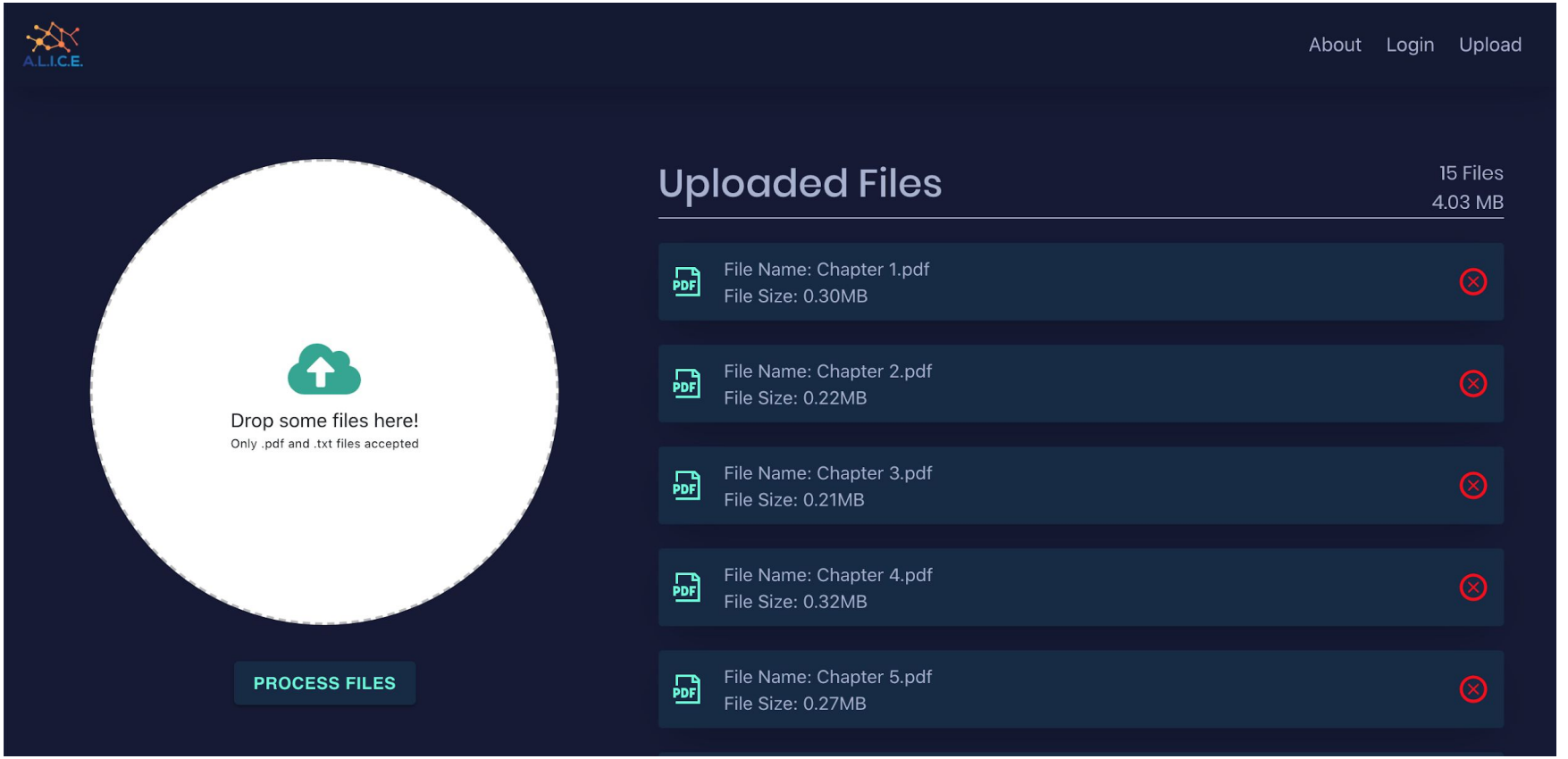
Walkthrough of Use Case



Walkthrough of Use Case



Walkthrough of Use Case



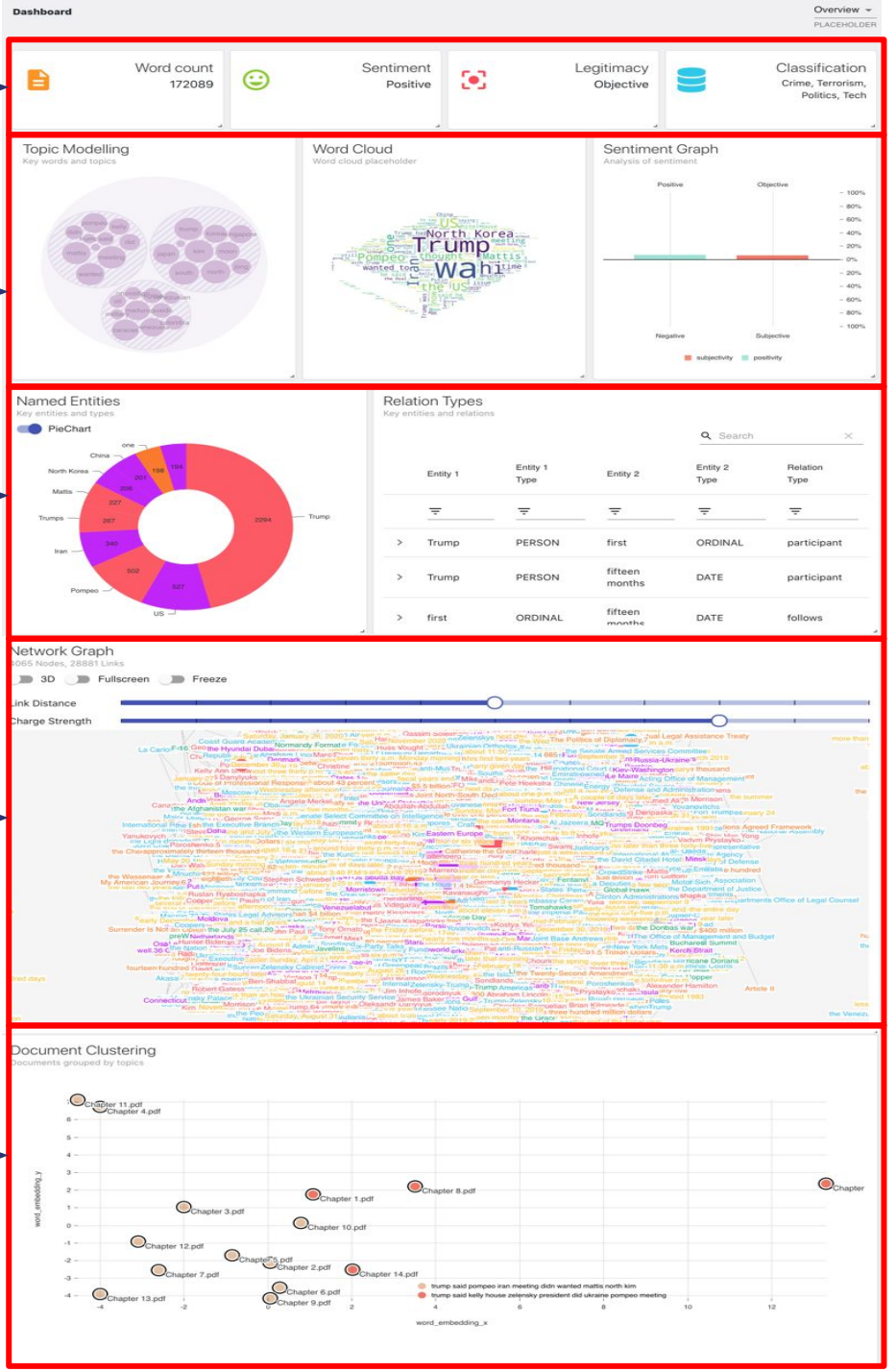
Key Data

Topic Modelling, Word Cloud, Sentiment Graph

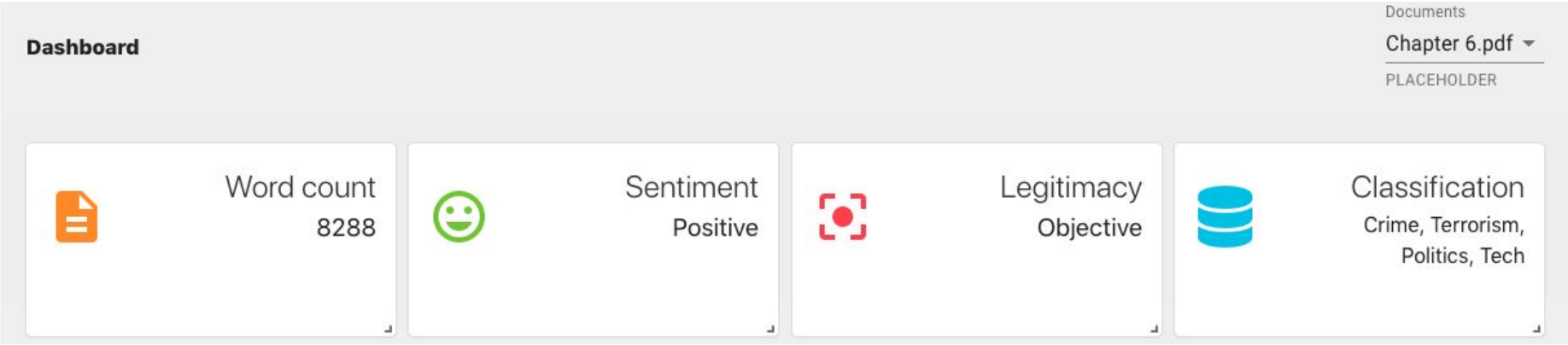
Named Entity Recognition, Relation Extraction

Network Graph

Document Clustering



Walkthrough of Use Case



- The key data provides a simple summary of the entire document or chapter.

- For example, Chapter 6 of the document has 8288 words, has a positive sentiment, is written objectively, and has the classification of crime, terrorism, politics, and technology.

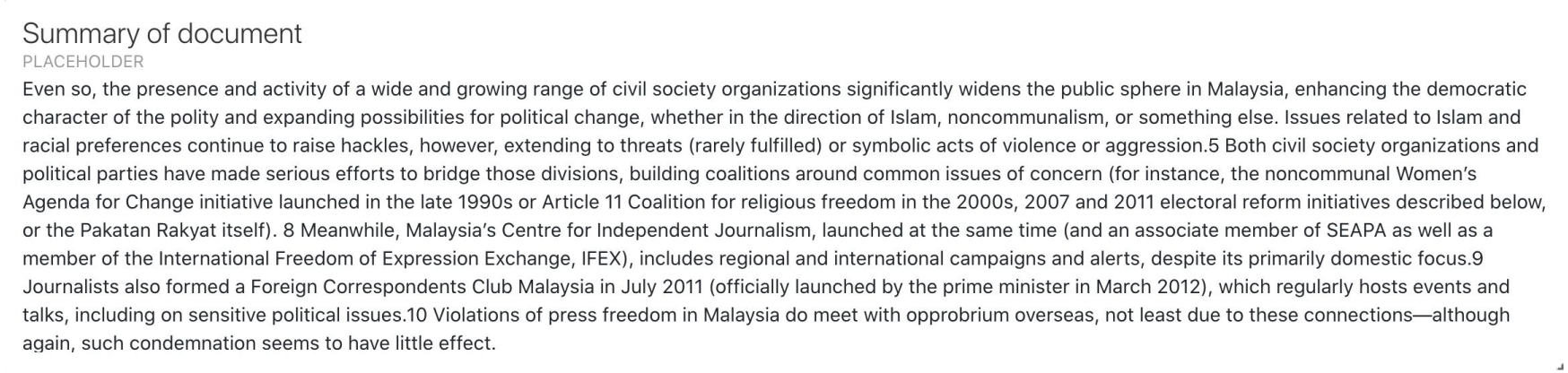
Walkthrough of Use Case

Classification

crime, terrorism,
finance, politics,
tech

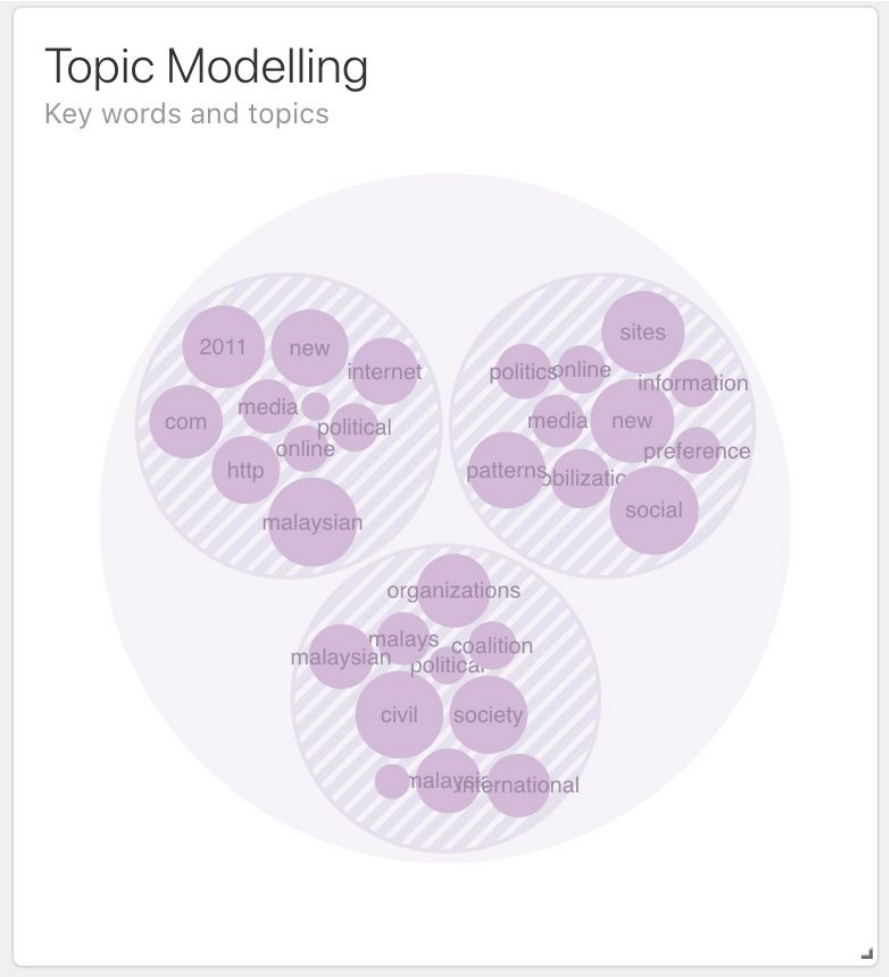
- Classifies the document into any of the following six categories: Health, Crime, Terrorism, Finance, Politics and Tech

Walkthrough of Use Case



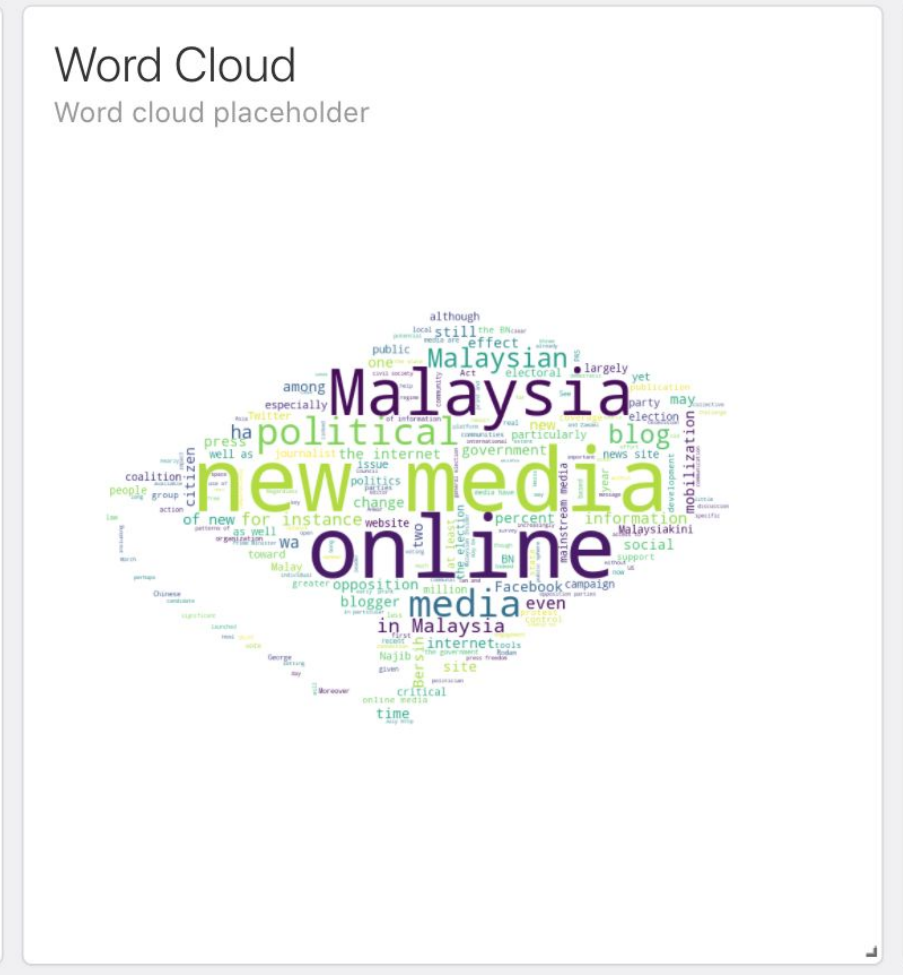
- The text summarizer takes the most important sentences of the document as the summary. Currently the default settings displays 4-5 lines as the summary

Walkthrough of Use Case



- The larger bubble represents a topic and the smaller bubbles within it represents the words that make up the topic
- The user has to infer what the topic is based on the words that are given

Walkthrough of Use Case



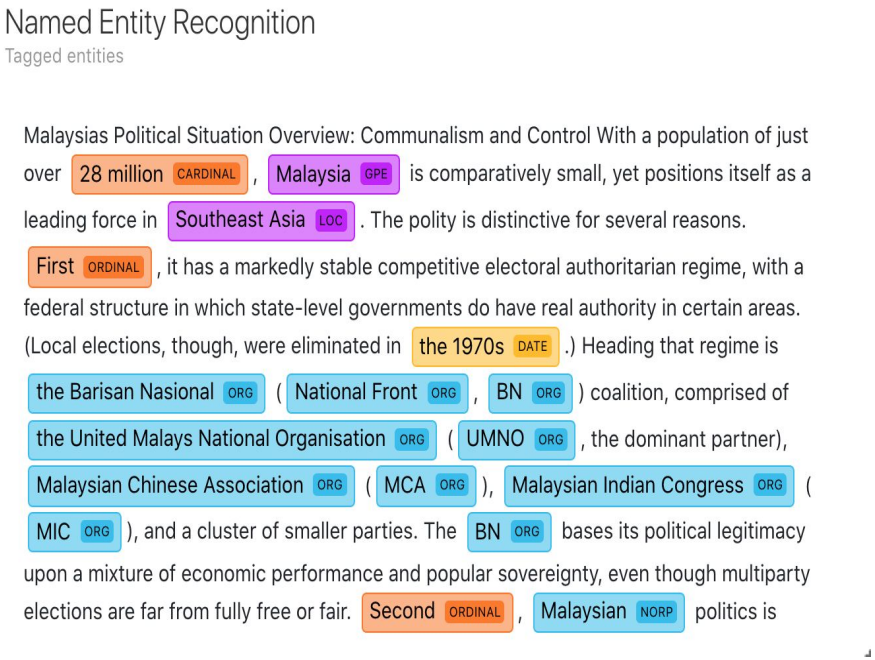
- The word cloud is an image that is made up of the most frequent words in the document.

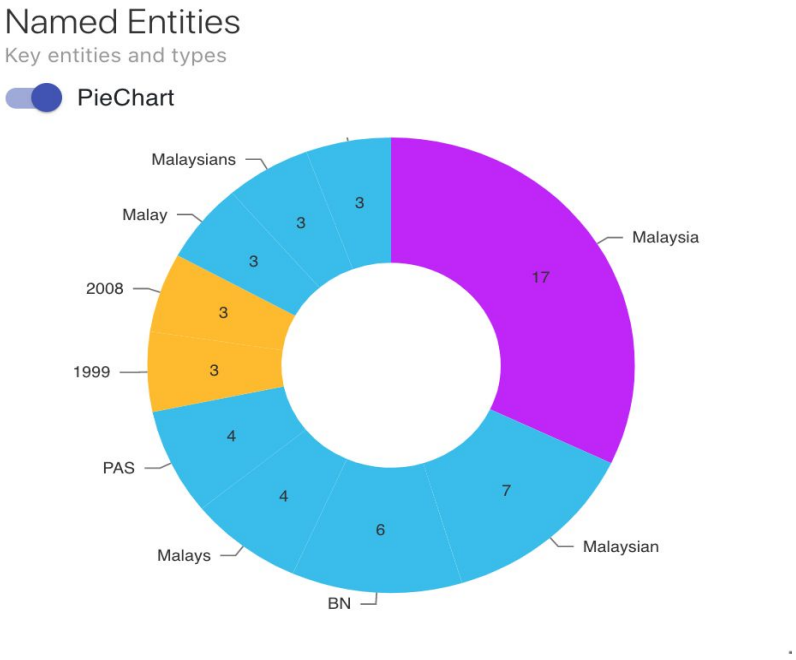
- The larger the words, the higher the frequency count in the document.

Walkthrough of Use Case

- The sentiment graph provides more details about the positivity and subjectivity of the document.
- The left bar represents the positivity.
- The right bar represents the subjectivity.

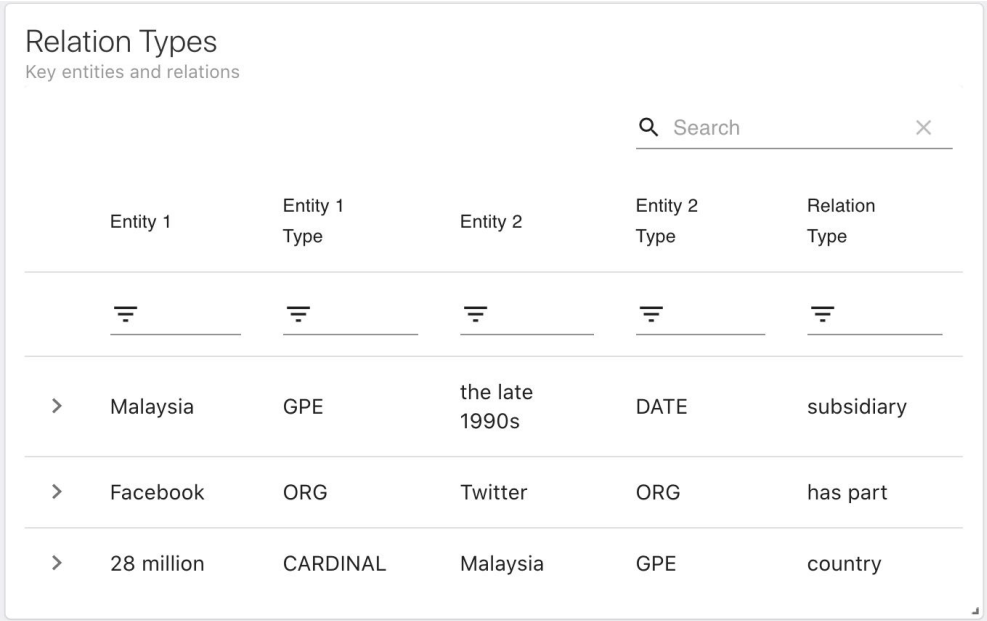
Walkthrough of Use Case





- The entity visualizer highlights named entities and their labels in a text.
- The entity pie chart displays the top 10 entities within the document, so that users can focus on the most frequent entities.

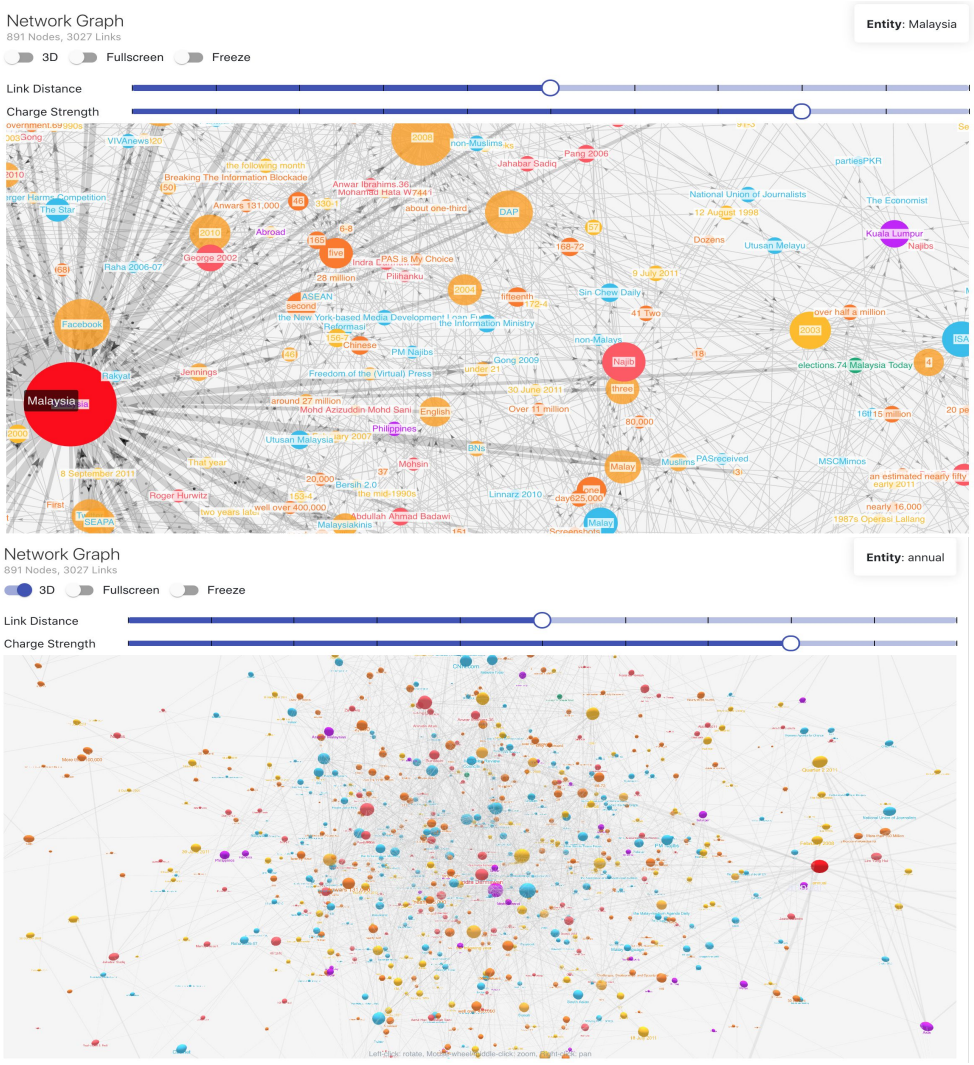
Walkthrough of Use Case



- In the Overview dashboard, the relation table allows the user to know which document the relation appears in.

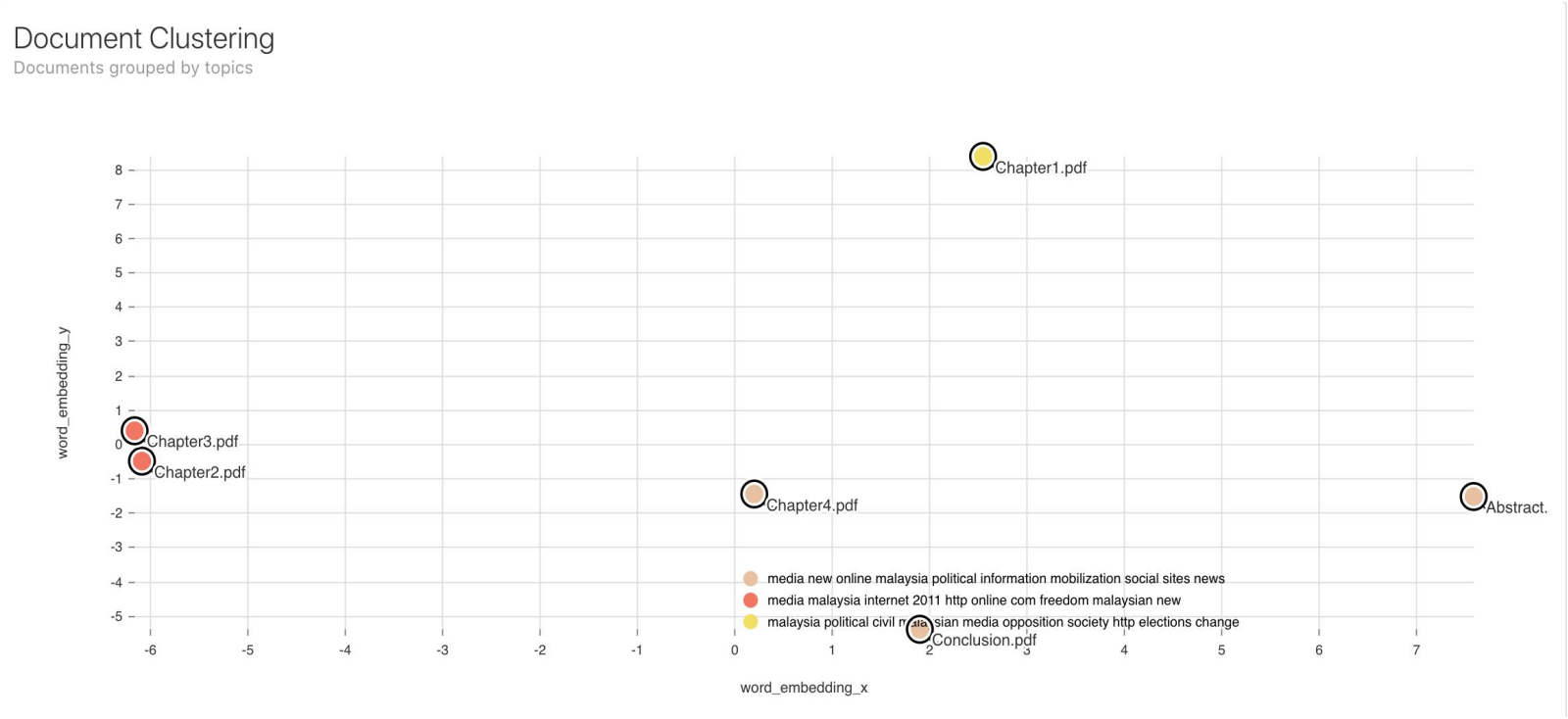
- In the individual document dashboard, the filtering of entities in the network graph is done by the relation table because of the smaller scale of relations in the individual documents. Also, the relation table allows the user to delete the relation if it is irrelevant.

Walkthrough of Use Case



- The network graph displays all the entities and relations via nodes and links respectively.
- The network graph has several options available to the user - 3D mode, fullscreen mode, freezing the graph so the nodes do not move, zooming in/out, and changing the link distance and charge strength between nodes.

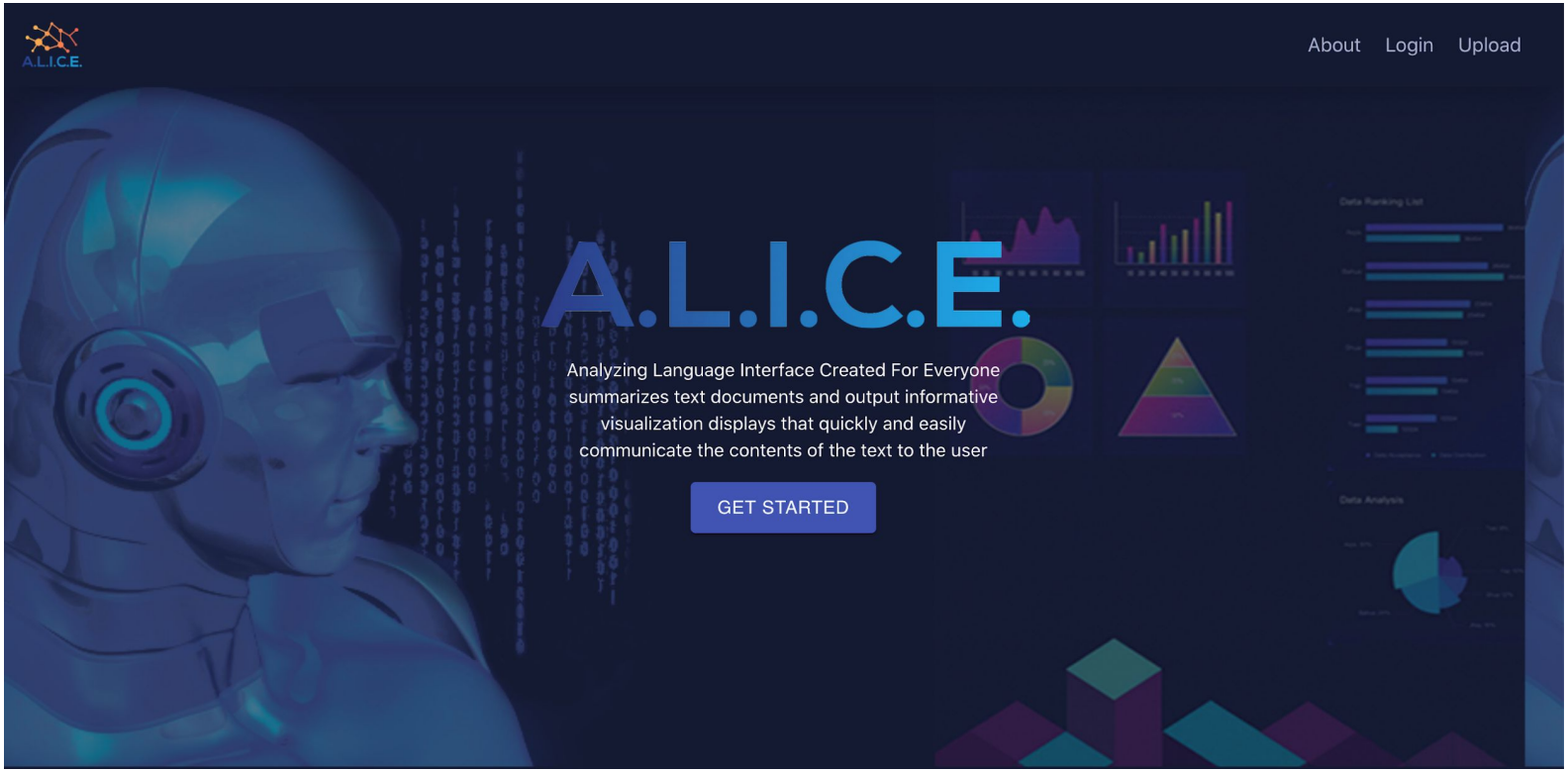
Walkthrough of Use Case



- Shows how closely related each document is to each other. The closer the distance between the document node in the graph, the more similar the documents are.
- Chapter 2 and 3 are similar in topic, where both are talking about media, internet in Malaysia and possibility its impact on freedom

Live Demo

1. **Politics in Cyberspace: New Media in Malaysia**
2. **The Room Where It Happened (John Bolton's Memoir)**



Internship 2020

A.L.I.C.E.

Thank you!