



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

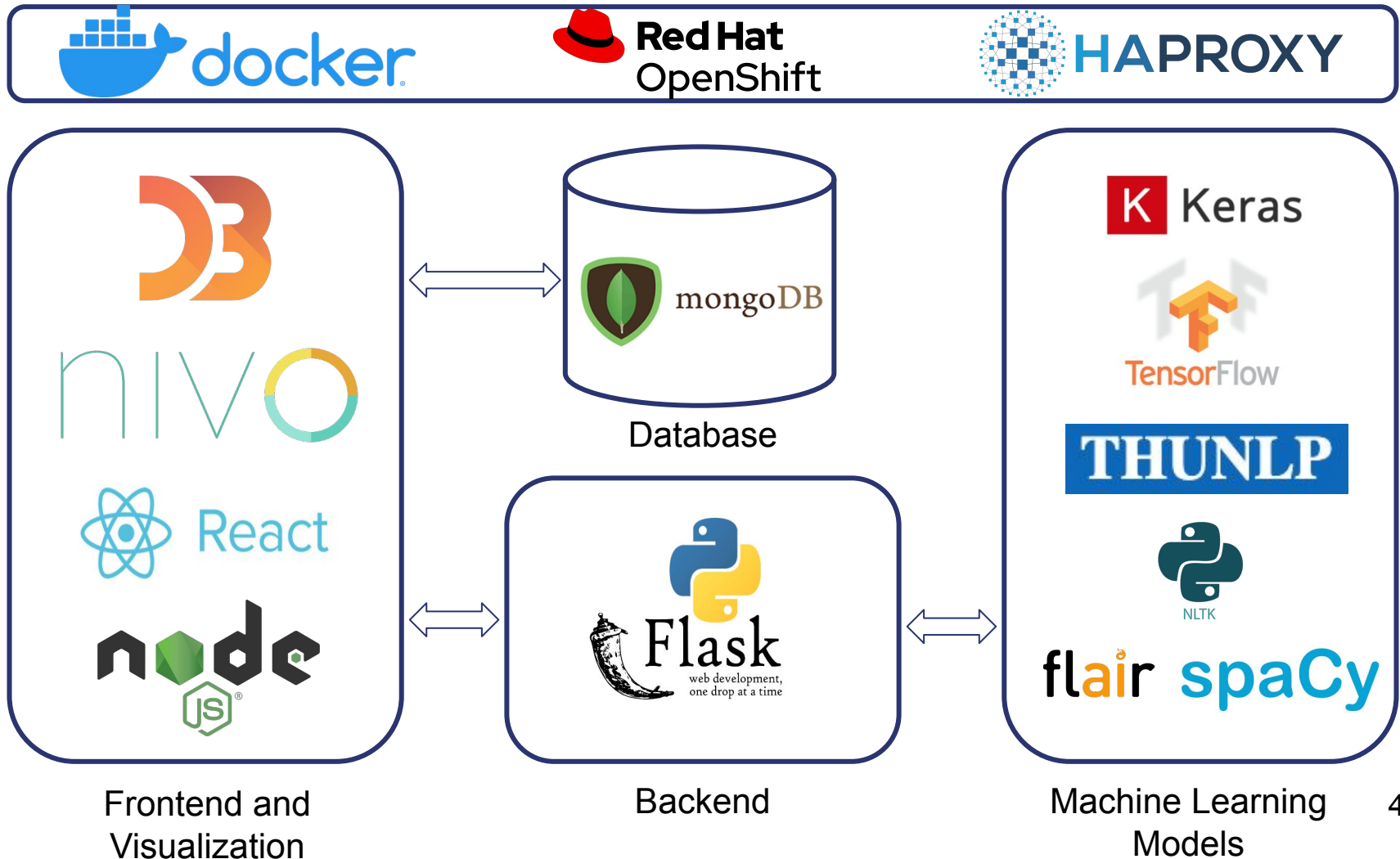


Background



- 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



[About](#) [Login](#) [Upload](#)



Drop some files here!
Only .pdf and .txt files accepted

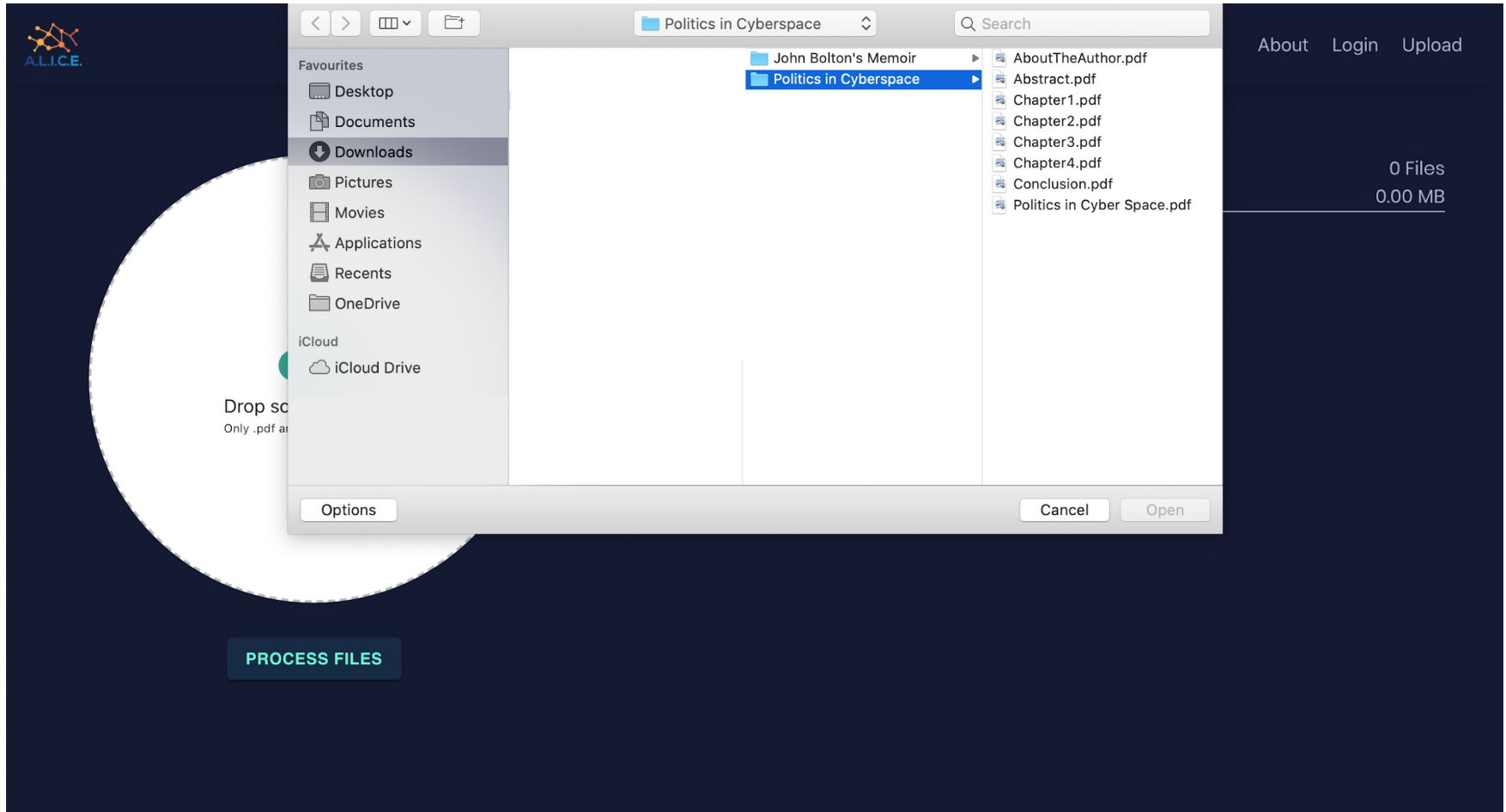
PROCESS FILES

Uploaded Files

0 Files
0.00 MB




Walkthrough of Use Case

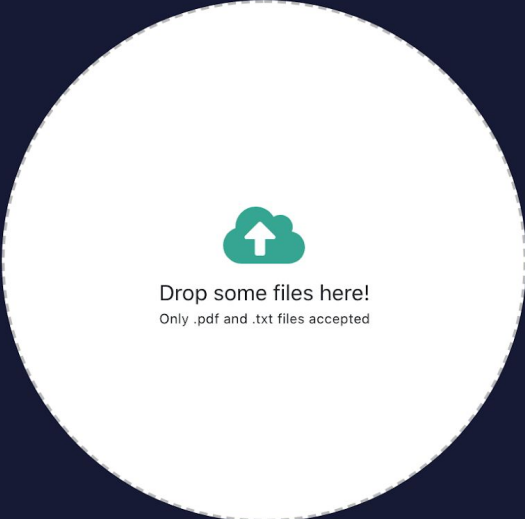




Walkthrough of Use Case



[About](#) [Login](#) [Upload](#)













Drop some files here!
Only .pdf and .txt files accepted

PROCESS FILES

Uploaded Files

15 Files
4.03 MB

	File Name: Chapter 1.pdf File Size: 0.30MB	
	File Name: Chapter 2.pdf File Size: 0.22MB	
	File Name: Chapter 3.pdf File Size: 0.21MB	
	File Name: Chapter 4.pdf File Size: 0.32MB	
	File Name: Chapter 5.pdf File Size: 0.27MB	

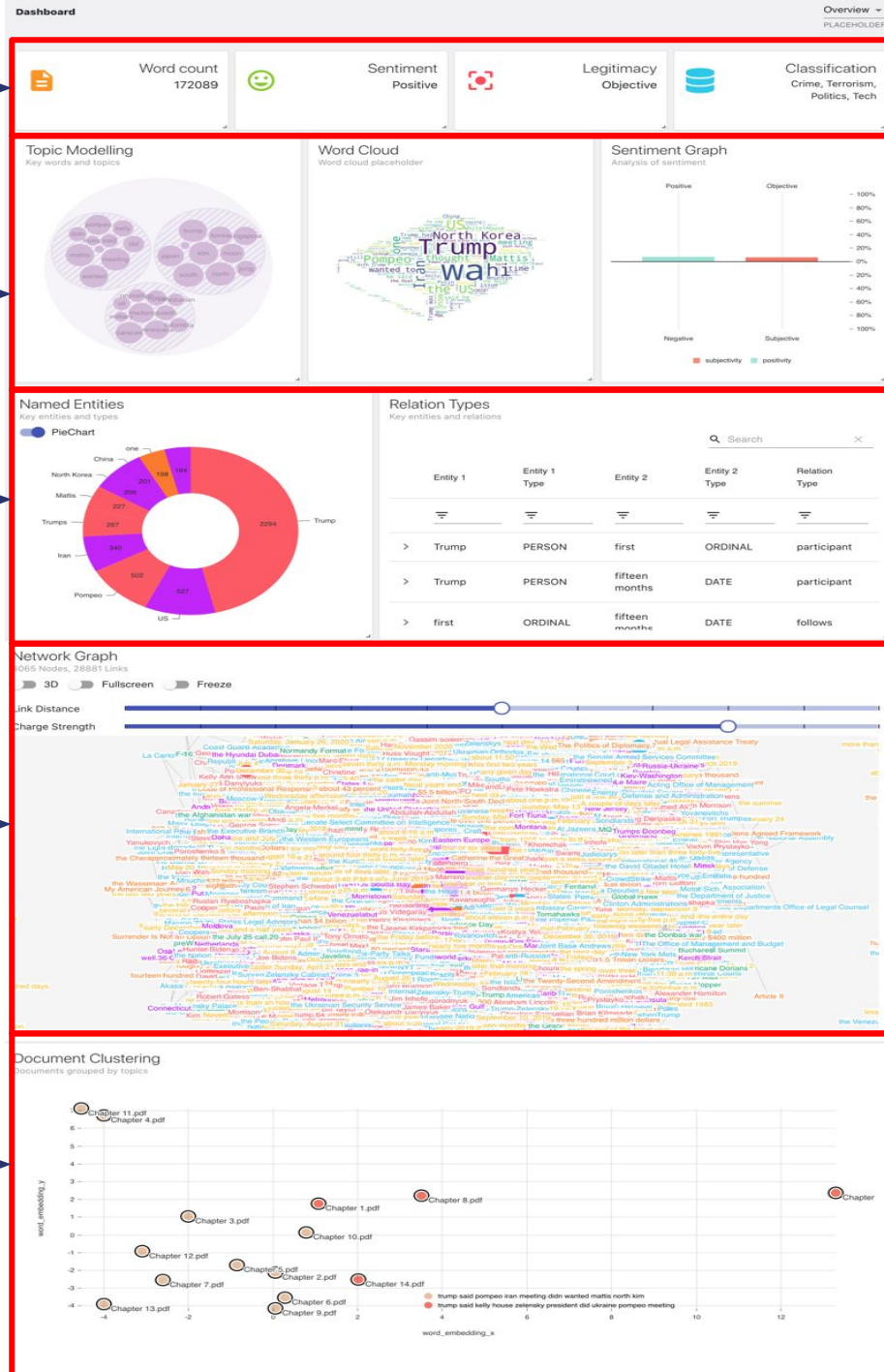
Key Data

Topic Modelling, Word Cloud, Sentiment Graph

Named Entity Recognition, Relation Extraction

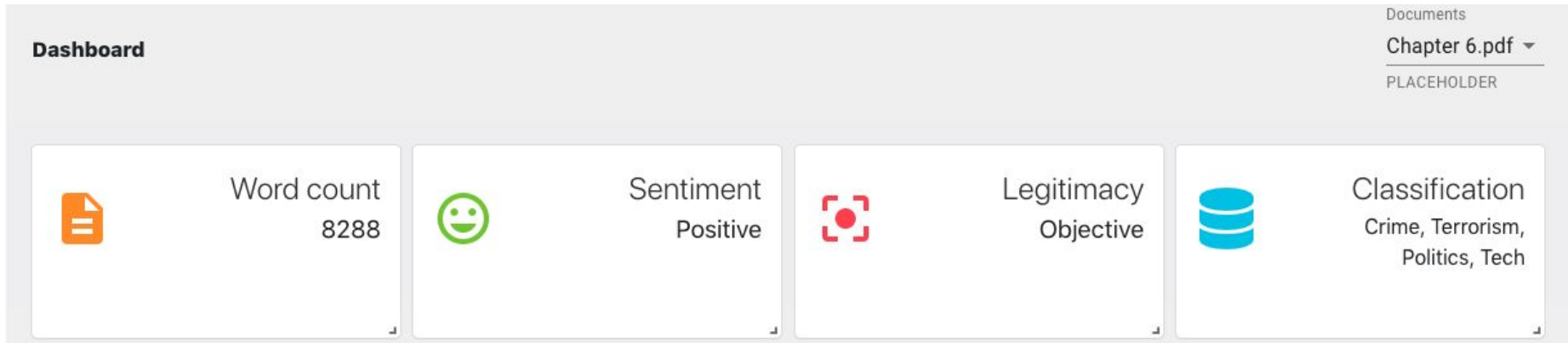
Network Graph

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



Summary of document

PLACEHOLDER

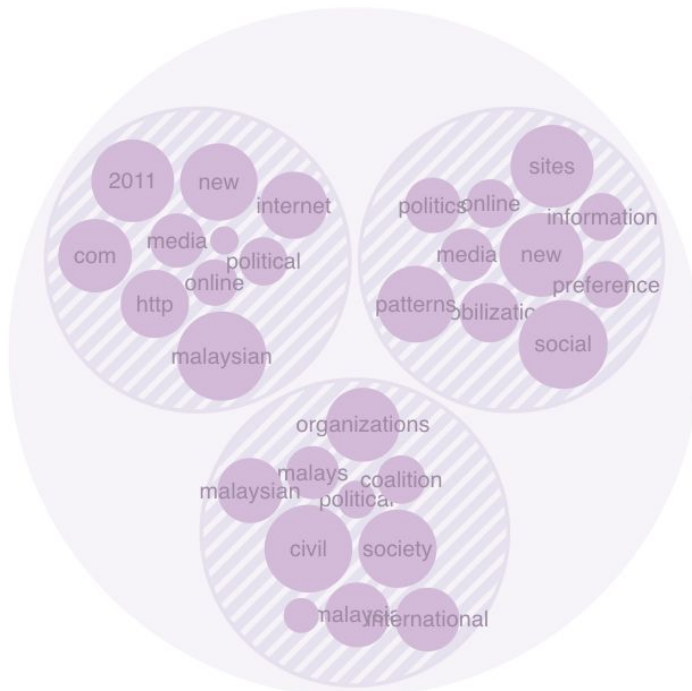
Even so, the presence and activity of a wide and growing range of civil society organizations significantly widens the public sphere in Malaysia, enhancing the democratic character of the polity and expanding possibilities for political change, whether in the direction of Islam, noncommunalism, or something else. Issues related to Islam and racial preferences continue to raise hackles, however, extending to threats (rarely fulfilled) or symbolic acts of violence or aggression.⁵ Both civil society organizations and political parties have made serious efforts to bridge those divisions, building coalitions around common issues of concern (for instance, the noncommunal Women's Agenda for Change initiative launched in the late 1990s or Article 11 Coalition for religious freedom in the 2000s, 2007 and 2011 electoral reform initiatives described below, or the Pakatan Rakyat itself).⁸ Meanwhile, Malaysia's Centre for Independent Journalism, launched at the same time (and an associate member of SEAPA as well as a member of the International Freedom of Expression Exchange, IFEX), includes regional and international campaigns and alerts, despite its primarily domestic focus.⁹ Journalists also formed a Foreign Correspondents Club Malaysia in July 2011 (officially launched by the prime minister in March 2012), which regularly hosts events and talks, including on sensitive political issues.¹⁰ Violations of press freedom in Malaysia do meet with opprobrium overseas, not least due to these connections—although again, such condemnation seems to have little effect.

- 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

Topic Modelling

Key words and topics



- 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



Word Cloud

Word cloud placeholder



- 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.

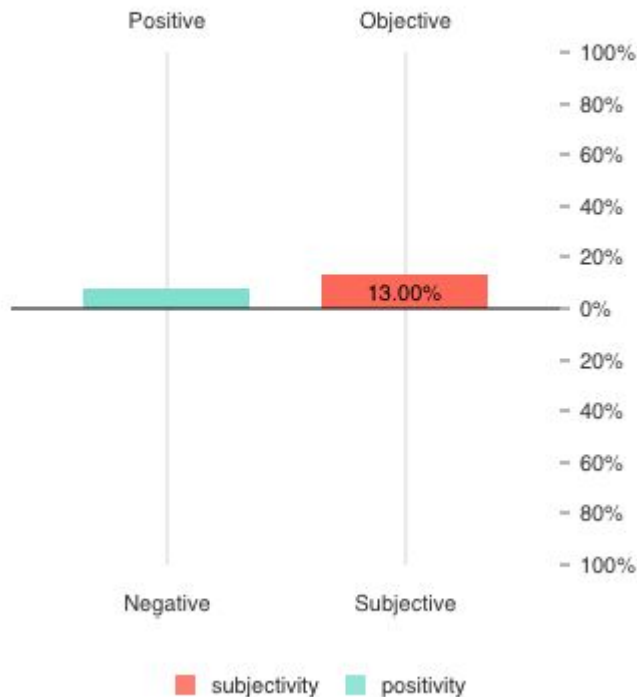


Walkthrough of Use Case



Sentiment Graph

Analysis of sentiment



- 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

Named Entity Recognition

Tagged entities

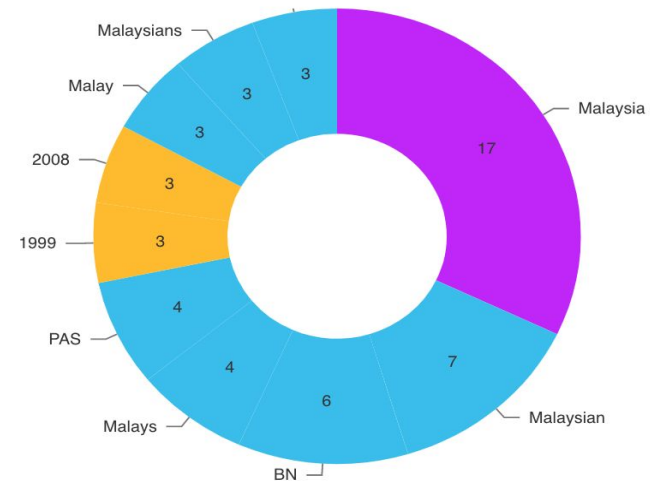
Malaysia's Political Situation Overview: Communalism and Control With a population of just over 28 million **CARDINAL**, Malaysia **GPE** is comparatively small, yet positions itself as a leading force in Southeast Asia **LOC**. The polity is distinctive for several reasons.

First **ORDINAL**, it has a markedly stable competitive electoral authoritarian regime, with a federal structure in which state-level governments do have real authority in certain areas. (Local elections, though, were eliminated in the 1970s **DATE**.) Heading that regime is the Barisan Nasional **ORG** (National Front **ORG**, BN **ORG**) coalition, comprised of the United Malays National Organisation **ORG** (UMNO **ORG**, the dominant partner), Malaysian Chinese Association **ORG** (MCA **ORG**), Malaysian Indian Congress **ORG** (MIC **ORG**), and a cluster of smaller parties. The BN **ORG** bases its political legitimacy upon a mixture of economic performance and popular sovereignty, even though multiparty elections are far from fully free or fair. Second **ORDINAL**, Malaysian **NORP** politics is

Named Entities

Key entities and types

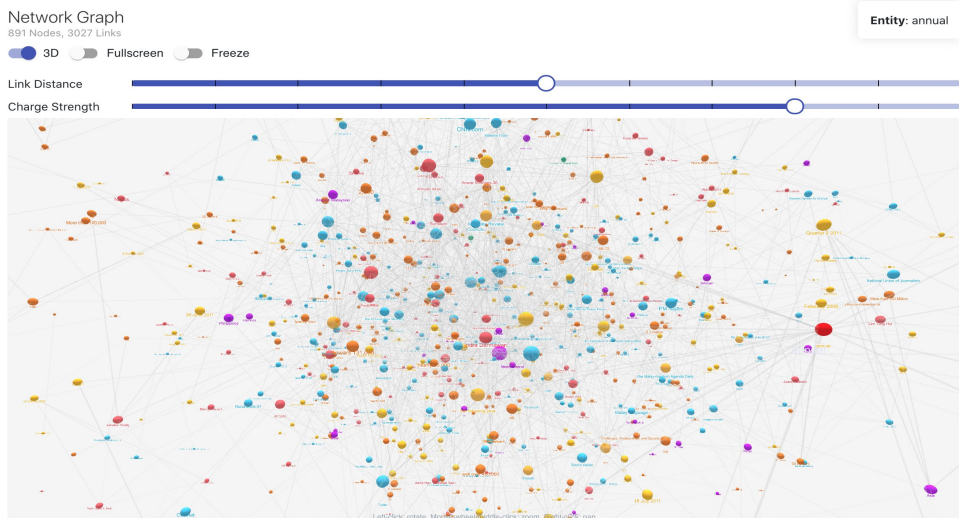
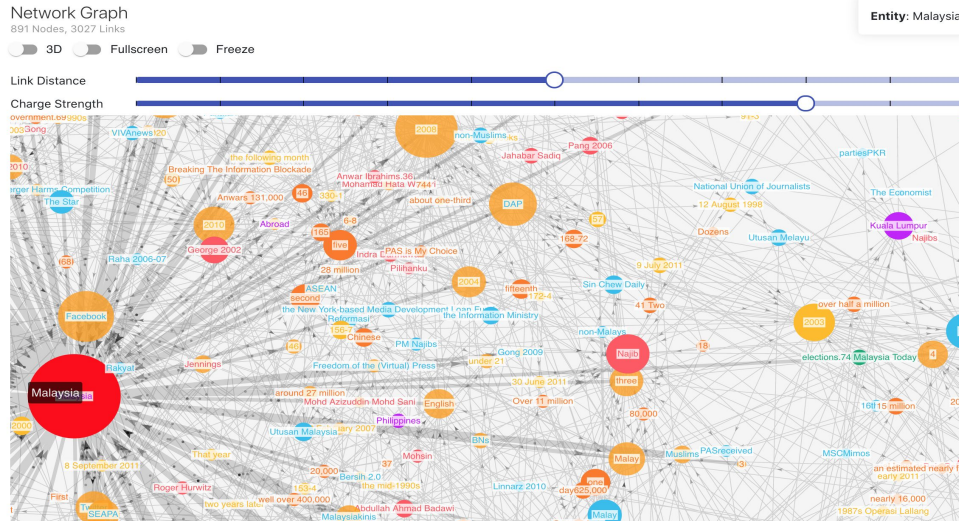
☒ PieChart



- 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

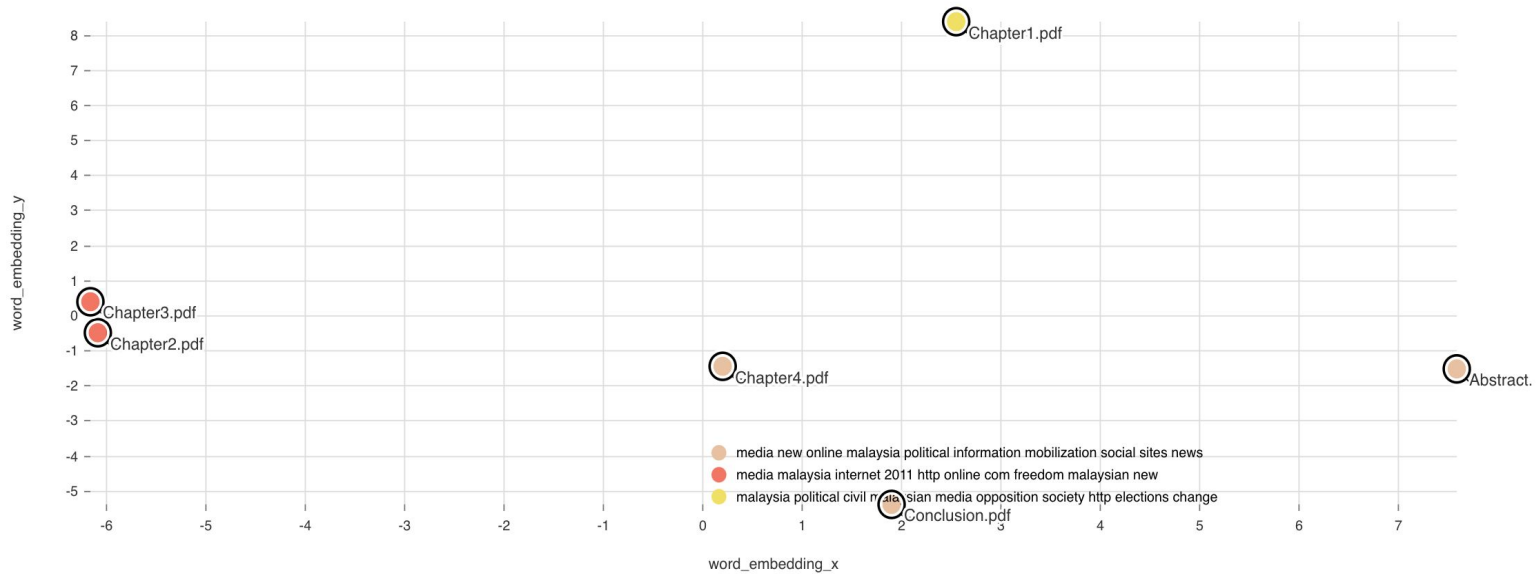


- The network graph displays all the entities and relations via nodes and links respectively.
- The network graph has several options available to the user - 3D mode, fullscreen mode, freezing the graph so the nodes do not move, zooming in/out, and changing the link distance and charge strength between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

Walkthrough of Use Case

Document Clustering

Documents grouped by topics



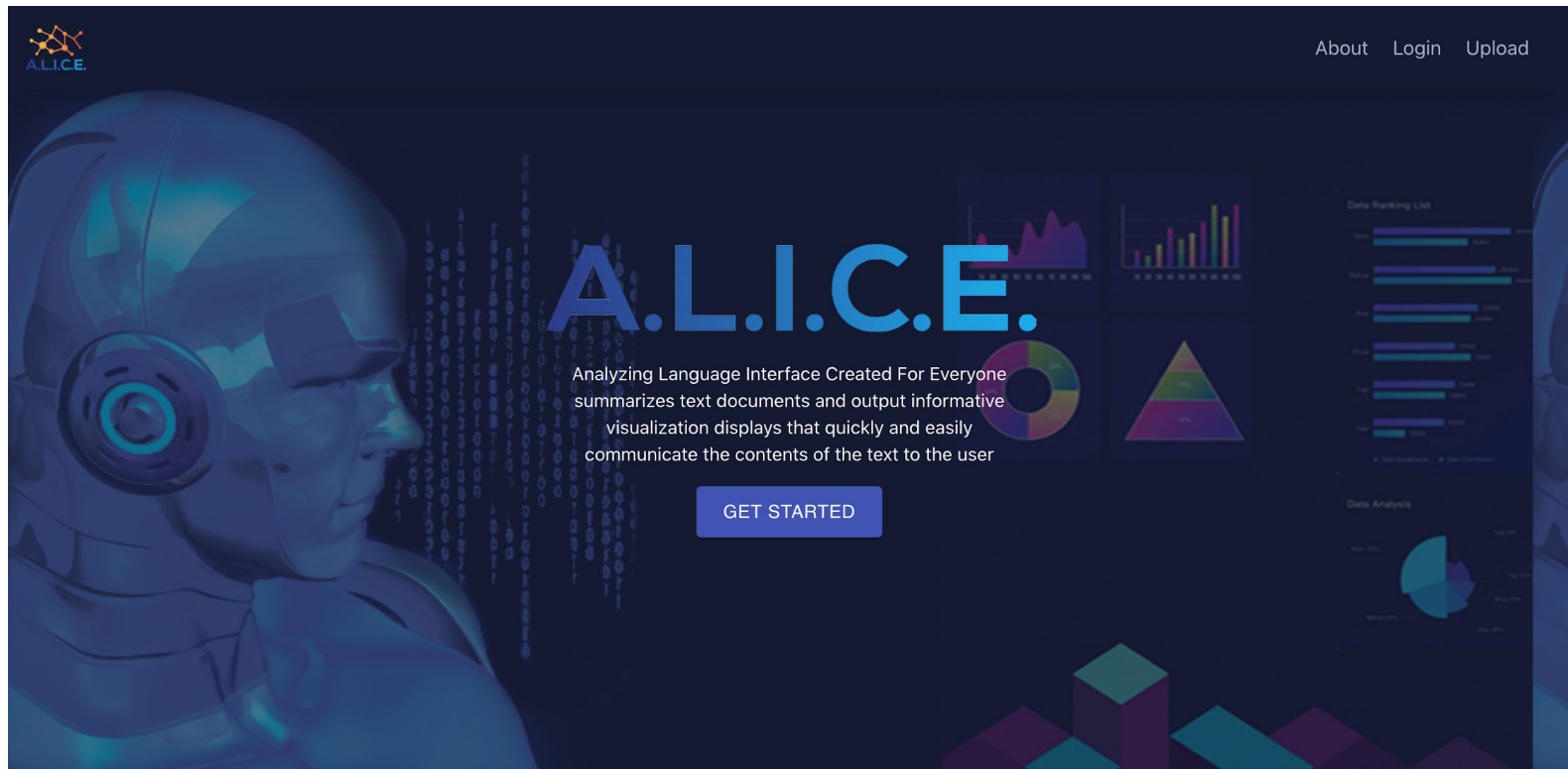
- Shows how closely related each document is to each other. The closer the distance between the document node in the graph, the more similar the documents are.
- Chapter 2 and 3 are similar in topic, where both are talking about media, internet in Malaysia and possibility its impact on freedom



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!