

# Crisis Events Information Extraction

Sully Gregory, Brianna Brown, Will Spies,  
Nishesh Saikrishna, Eitsaam Rabbani

CS 4624 Multimedia and Hypertext, Instructor:  
Mohamed Farag, Virginia Tech, Blacksburg VA 24061

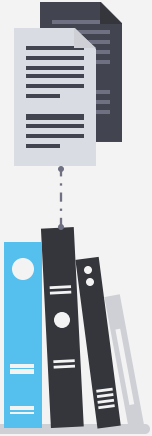
04/22/2024





# Outline

- **Client & Problem**
- **Approach**
- **Architecture**
- **Implementation**
- **Features**
- **Demo**
- **Final Plans**





## Client

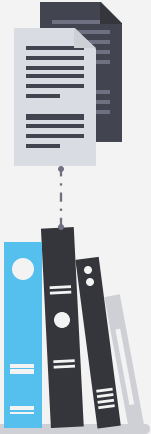


Dr. Mohamed Farag

## Problem

Crisis event news articles can be time-consuming to read and wordy, distracting from the key information.

A straightforward way to extract key details from multiple crisis events articles is needed.

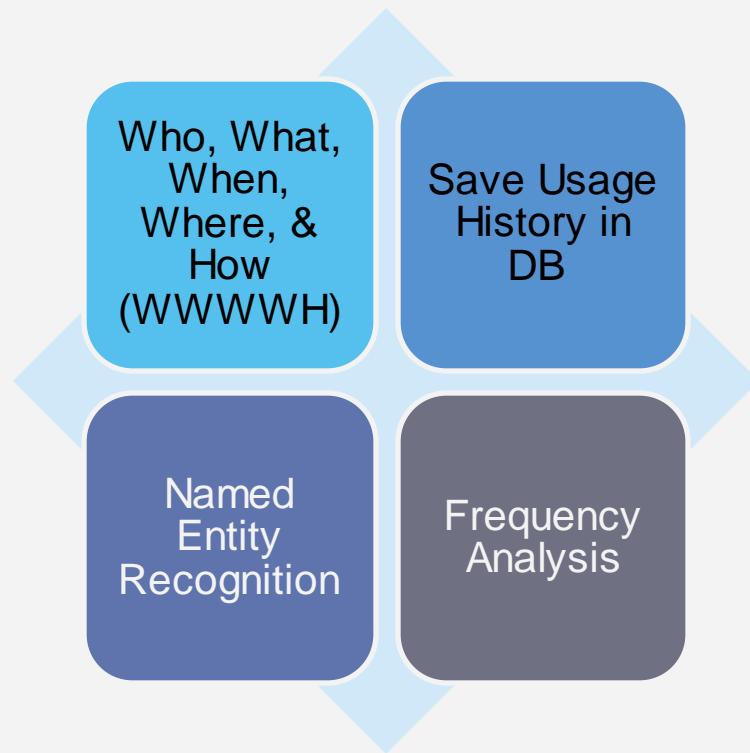




# Approach

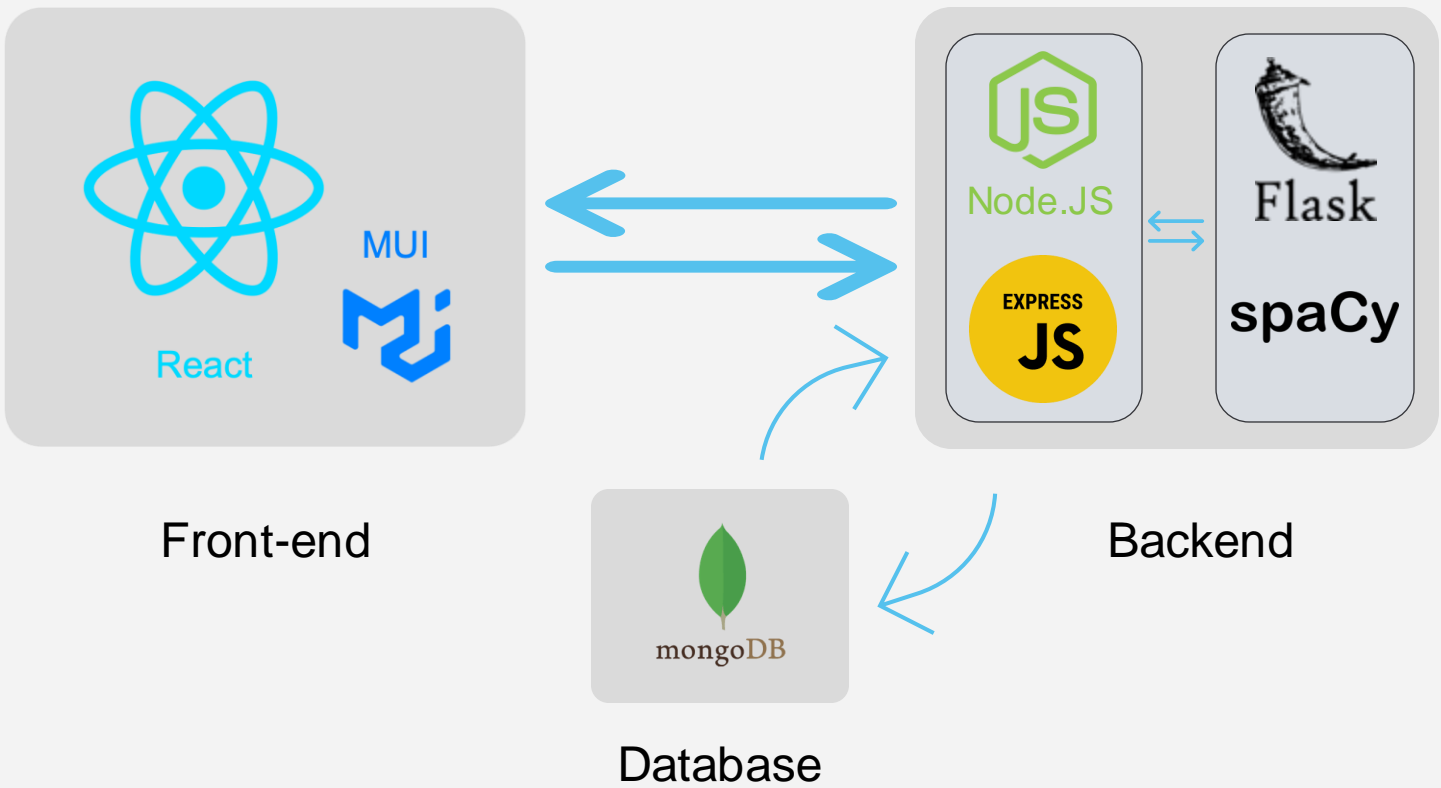
**Design a web application to summarize crisis events with a specific who, what, when, where, and how**

- Accept user input of crisis event webpages (URL/HTML/txt)
- Extract summary data from user input
- Require user login to save input/output collections
- Display collection history for a user, recounting all past input/output collections they have created



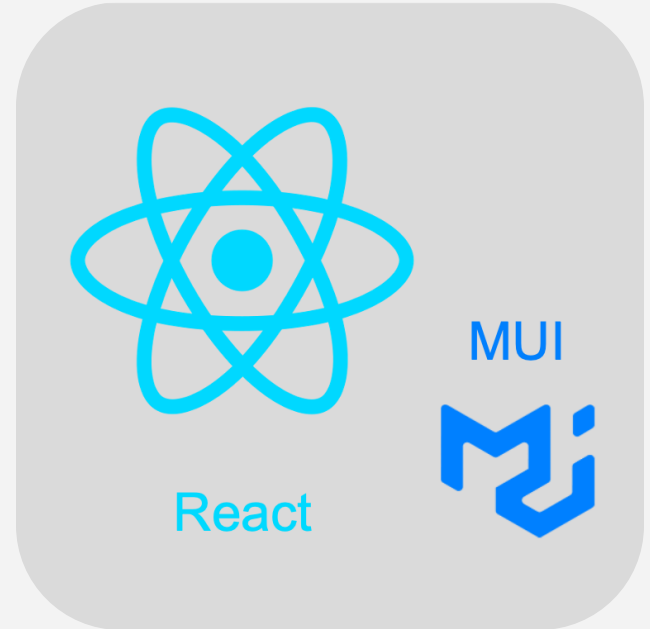


# System Architecture



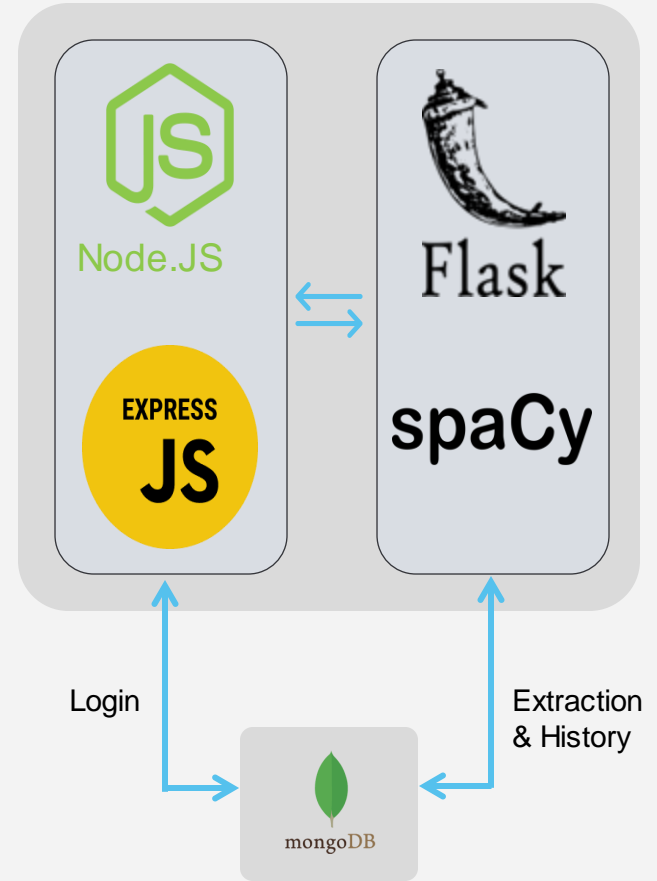
# Front-end Implementation

- Display extracted information in table format
- Extractions can be expanded to show relevant sentences
- MUI component library for consistent design language

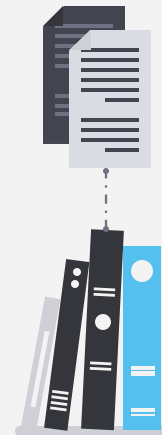
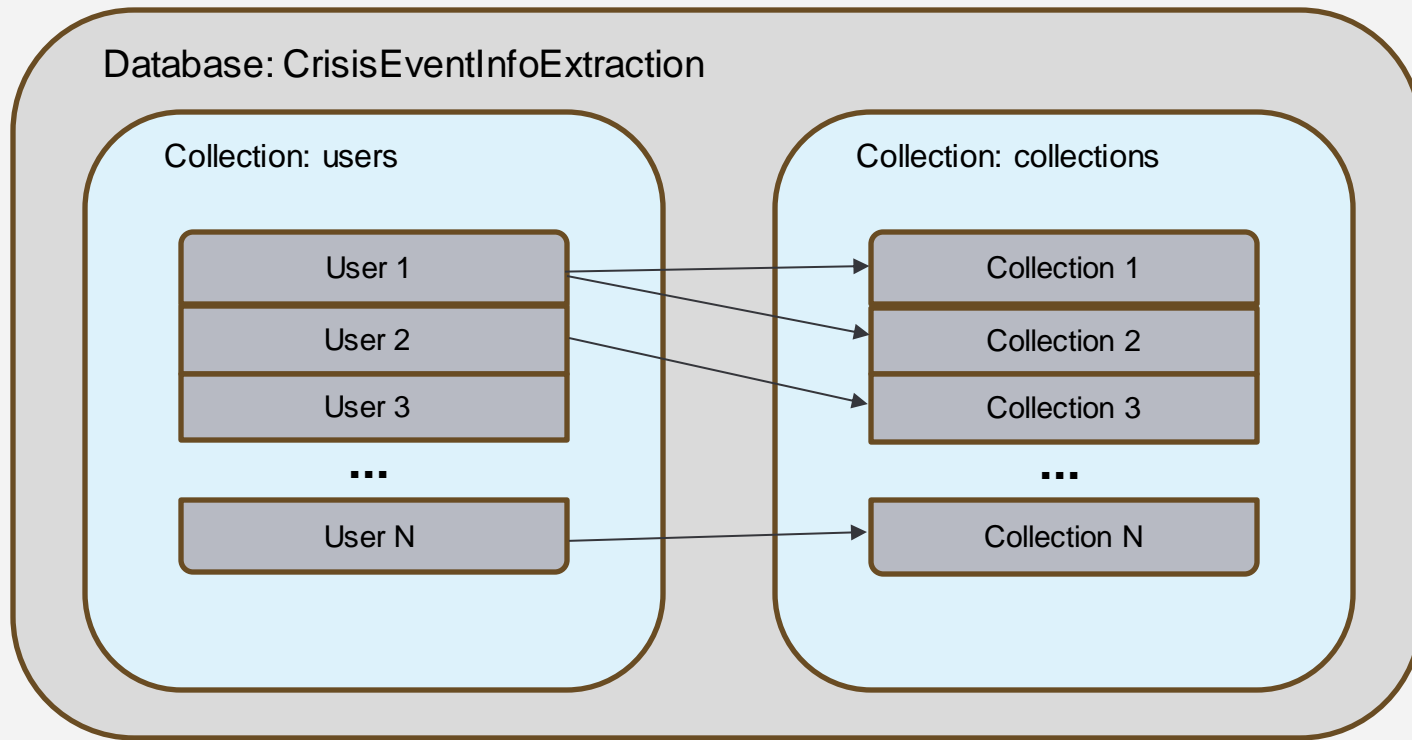


# Back-end Implementation

- Front-end talks exclusively to Node.js / Express server
- Node.js handles authentication natively and calls all NLP via Flask using API requests
- Flask / SpaCy handles all information extraction and NLP and uploads results to MongoDB



# Database Design



# Users Schema

**\_id:** ObjectId

**username:** String

**password:** String

The `_id` field will be used to connect collections to the users who created them.

The `password` field is hashed with `bcrypt` for added database security.



# Collections Schema

```
userId: ObjectId
fileName: String
collectionName: String
inputData: {
  links: [String]
}
outputData: {
  how: [String]
  what: [String]
  when: [String]
  where: [String]
  who: [String]
  howSentences: Object
  whatSentences: Object
  whenSentences: Object
  whereSentences: Object
  whoSentences: Object
}
```

userId – id of user who created the collection

how/what/etc - top K responses for that category

- K value is set in code for client access

\_\_Sentences – object with 1 field for each of the K responses in WWWWH

- Each field's value is an array of the sentences that the given response was extracted from





# Project Features

1. User Authentication
2. Create a new collection
  - File Upload
  - WWWWH Extraction
3. Collection History Table
  - Delete a collection
  - Details pop-up
  - Sentences display



# User Authentication

Login

Crisis Events Information Extraction

FILE UPLOAD

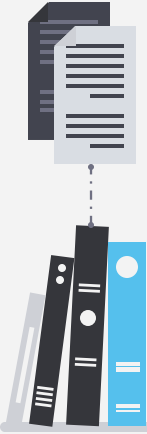
HISTORY

LOGIN

## User Login

LOGIN

SIGN UP

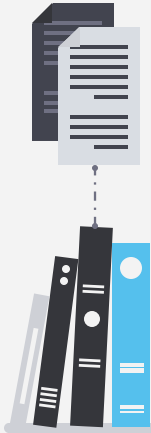
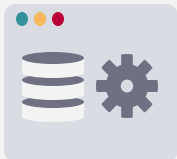


# User Authentication

## Login Success

Crisis Events Information Extraction							FILE UPLOAD	HISTORY	WSTUDENT
File	Collection Name	How	What	When	Where	Who			
desktop.txt	Test2	10	Gunman opens fire near California Lunar New Year festival, killing 10	2023-01-21	Monterey Park	Tran			

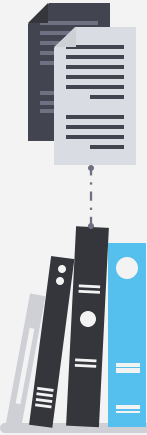
On a successful login, you are redirected to the collection history page and username is displayed



# New Collection

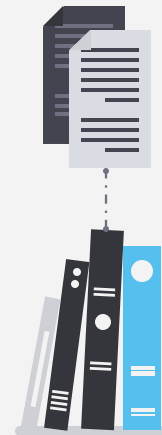
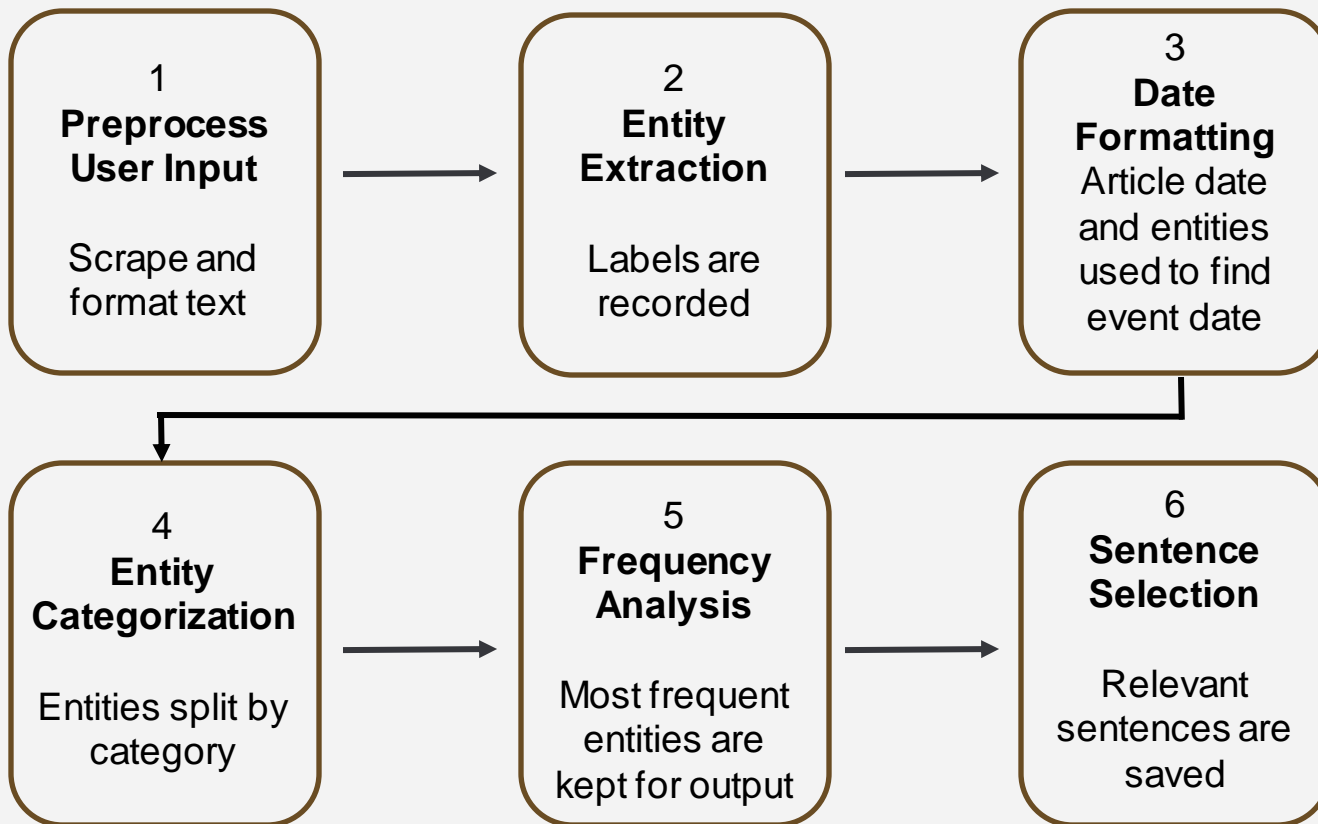
## File Upload

### File Upload

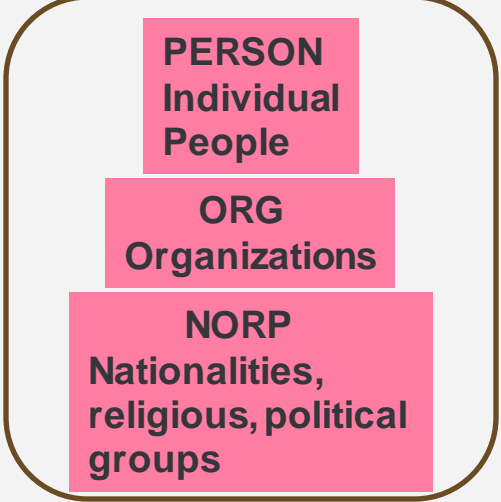


Home page displayed to user, user has the option to upload file and add a collection name. Pressing the submit button will run the extraction.

# Text Processing & Extraction



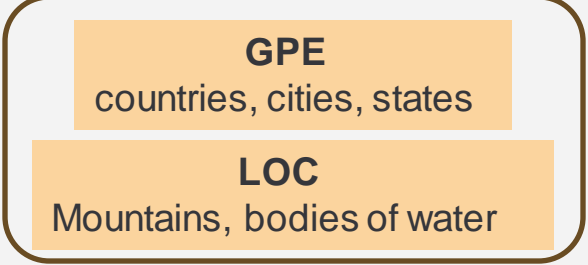
WHO →



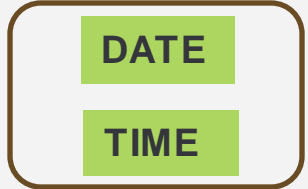
WHAT →



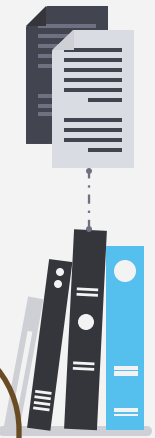
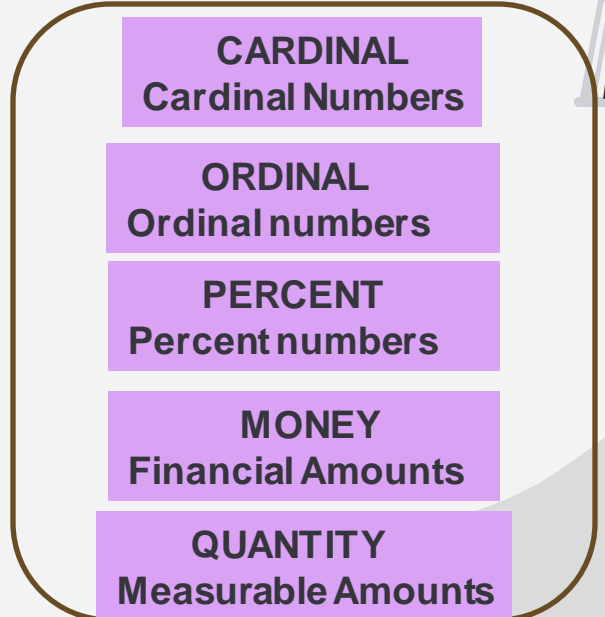
WHERE →



WHEN →



HOW →



WHO

ORG

PERSON

NORP

WHAT

EVENT

WHEN

DATE

TIME

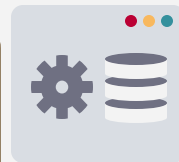
HOW

CARDINAL

ORDINAL

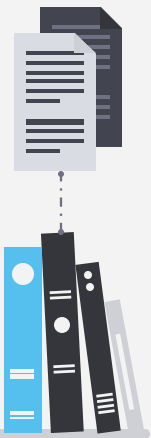
WHERE

GPE



**United States Presidential Election** of 2016, **American presidential election** held on **November 8, 2016**, in which **Republican Donald Trump** lost the popular vote to **Democrat Hillary Clinton** by **more than 2.8 million** votes but won **30** states and the decisive electoral college with **304** electoral votes to Clinton's **227** and thus became the **45th** president of the **United States**.

The tumultuous, abrasive **2016** campaign defied established political norms. **Clinton's** campaign featured superior organization and fund-raising—and almost every election-eve poll pointed to a comfortable victory for her—but **Trump's** anti-Washington appeal to white working-class voters outside major cities in pivotal manufacturing states proved to be the key factor in what several publications called “the most stunning upset in **American** history.” The election of an outsider with no political job experience represented a major repudiation of business as usual by both parties in **Washington, D.C.**





# Date Formatting

1. Get publishing date

## Virginia Tech Studying Solar Eclipse

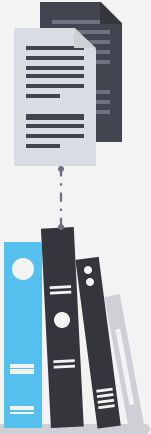
*Published: Apr. 9, 2024 at 6:49 PM EDT*



2. Identify date and time entities

This is the first time in seven years since the U.S. has seen a cosmic event on this scale. However, **Monday's** eclipse is expected to be larger and longer than the one experienced in 2017.

3. Calculate crisis event date





# Frequency Analysis – What

## Solar Eclipse Example

### 1. Join all frequency dictionaries

Where: { Norfolk: 6, Virginia: 4, Earth: 2, VA: 2 }

When: { 2024-04-08: 9 }

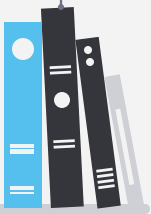
### 2. Select x top entities (ex: x=3)

Top Entities: { 2024-04-08: 9, Norfolk: 6, Virginia: 4 }

### 3. Get article titles containing the most top entities (weighted)

Title 1: Eclipse visible from Norfolk, Virginia – Score = 10

Title 2: Eclipse seen in Virginia on April 8th – Score = 13

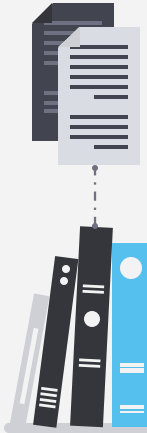


# Collection History

## Display Table

Crisis Events Information Extraction				FILE UPLOAD	HISTORY	WSTUDENT
File	Collection Name	How	What	When	Where	Who
test.txt	Sydney	six	Sydney church bishop stabbed in 'terrorist incident' says he forgives attacker	2024-04-16	Sydney	Reuters
desktop.txt	Collection1	5	Virginia Tech Studying Solar Eclipse	2024-04-08	Virginia	de Nolfo

[CREATE COLLECTION](#) [DELETE COLLECTION](#)



# Collection History

## Delete a Collection

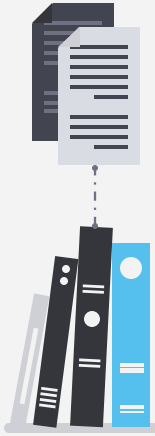
File	Collection Name	How	What	When	Where	Who
test.txt	Sydney	six	Sydney church bishop stabbed in 'terrorist incident' says he forgives attacker	2024-04-16	Sydney	Reuters
desktop.txt	Collection1	5	Virginia Tech Studying Solar Eclipse	2024-04-08	Virginia	de Nolfo

CREATE COLLECTION    DELETE COLLECTION

Select Collection to Delete

Sydney ▾

CANCEL    DELETE



The delete collection button displays a pop-up with a drop-down to select which collection to delete based on the collection name.



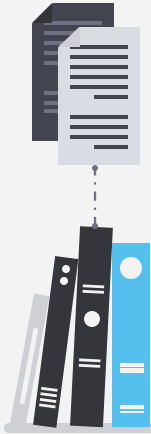
# Collection History

## Details Pop-up

Extraction Details	
<b>Collection Name:</b>	Collection1
<b>How:</b>	5, 3, 2024, 77.9%, 78.8%
<b>What:</b>	<ol style="list-style-type: none"><li>1. Virginia Tech Studying Solar Eclipse</li><li>2. What to expect from the 2024 solar eclipse in Virginia, according to NASA expert</li></ol>
<b>When:</b>	2024-04-08, 2017-04-08
<b>Where:</b>	Virginia, Earth, NORFOLK, Va., the United States
<b>Who:</b>	de Nolfo, NASA, Norfolk, the Lunar Reconnaissance Orbiter, Craig Moeller

[SHOW SENTENCES](#) [CLOSE](#)

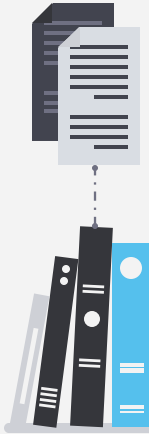
Clicking a collection will display a pop-up with at most the top K results for each category. The K value can be set in the code by the client.



# Collection History

## Sentences Display

Where	Sentences
Virginia	<ol style="list-style-type: none"><li>1. Excitement continues to build ahead of the solar eclipse on Monday that many will be able to experience in the United States, including in Virginia — at least partially.</li><li>2. Meanwhile, Virginia will be in the "penumbra," the larger but partial shadow from the eclipse.</li><li>3. RELATED: 3 Virginia road trips to see the 2024 solar eclipse with the most coverage</li><li>4. "In Virginia, you'll be able to see a partial eclipse, about 78% of the sun will be covered.</li></ol>
Earth	<ol style="list-style-type: none"><li>1. "You basically have the moon coming in between the sun and the Earth, and it'll block out the majority of light coming from the sun.</li><li>2. As a result, it will cast a shadow on the Earth," explained Dr. de Nolfo.</li></ol>



The sentences display shows the top K results for each category and what sentences they were extracted from.



# Final Plans

- Improve “how” extraction
- Improve modularity – extract each WWWWH category separately with minimal added TC
- Testing failures and edge cases





# Acknowledgements



**Dr. Mohamed Farag**  
Professor & Client



# Resources

Slide Template:

- <https://slidesgo.com/theme/multimedia-software-pitch-deck#search-Computer&position-10&results-486>

