Natural Language Processing and Information Retrieval

T e a m   1 0

<u>Subject</u>: How to retrieve a list of movies with a given query (e.g., "love" for movie1, movie2, etc) by searching the movie script dataset from <u>https://imsdb.com/</u>

Adrien Moreau 50231580
Damien Juan 50231589
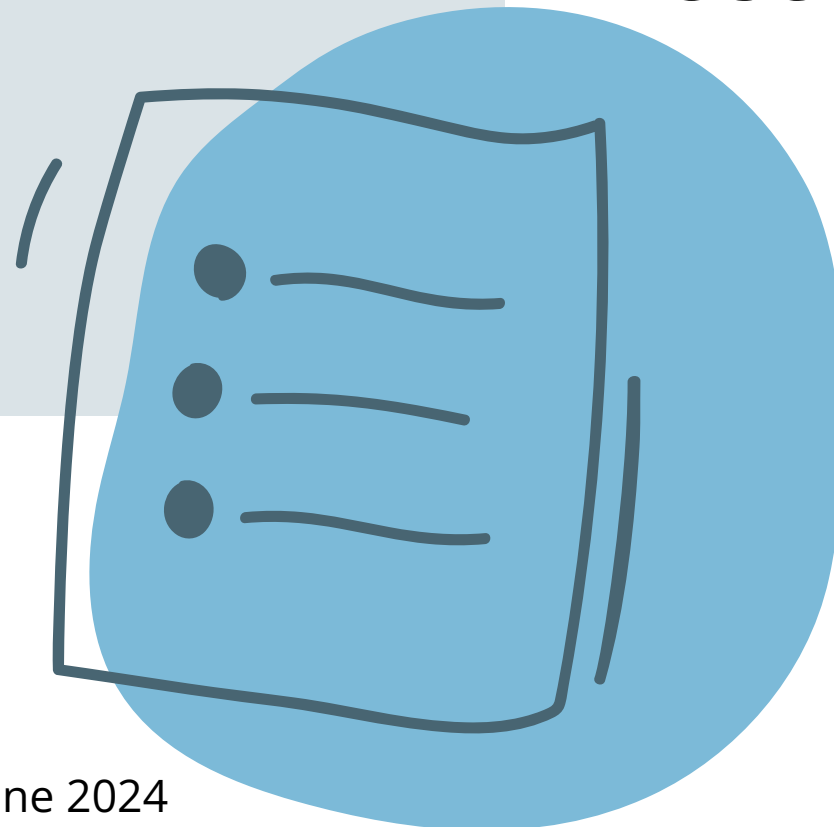Valentin Sauvier 50231622

Summary

- **THE PROJECT & FEATURES**

- **DATA RETRIEVAL**

- **SEARCH ENGINE**

- **EVALUATION**

- **RESULTS INTERPRETATION**

Project & Features

Search engine to retrieve scripts containing a certain keyword(s) in the IMSDb.com.

- Search by Word
- Filter by Genre
- Filter by Writer
- Filter by Date Range
- User Interface

Data Retrieval

- Python
- Web scraping
- Storing data in csv file
  - Title
  - Writers
  - Genres
  - Script date
  - Movie release date
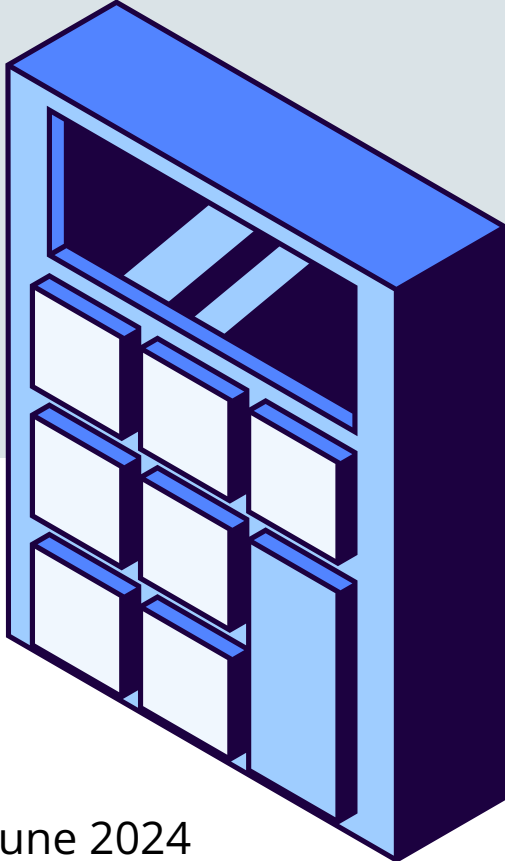 - Script

Search Engine

- Basic keyword search
- Filtering
- Simplified Boolean Model
  - Exact match
  - AND operator
- Narrow down results
  - by genre
  - by writer
  - by date range

More genres than IMSDb filters

Evaluation

- Precision
- Recall
- F1 Score
- Manual evaluation
  - "love" in Western genre
  - obtained 15 results
  - manual check
  - all 15 results contained word "love"

```
$ python3 effectiveness.py
Precision: 1.00
Recall: 1.00
F1 Score: 1.00
```

Results Interpretation

- Exceptionally high precision and recall rates
- Why?
- Simple Keyword Search
- Comprehensive Filtering

Limitations: Incomplete Data
- missing essential information
- not retrieved