A Movie Recommendation System

Based on Knowledge Graph and Collaborative Filtering

Team ACT

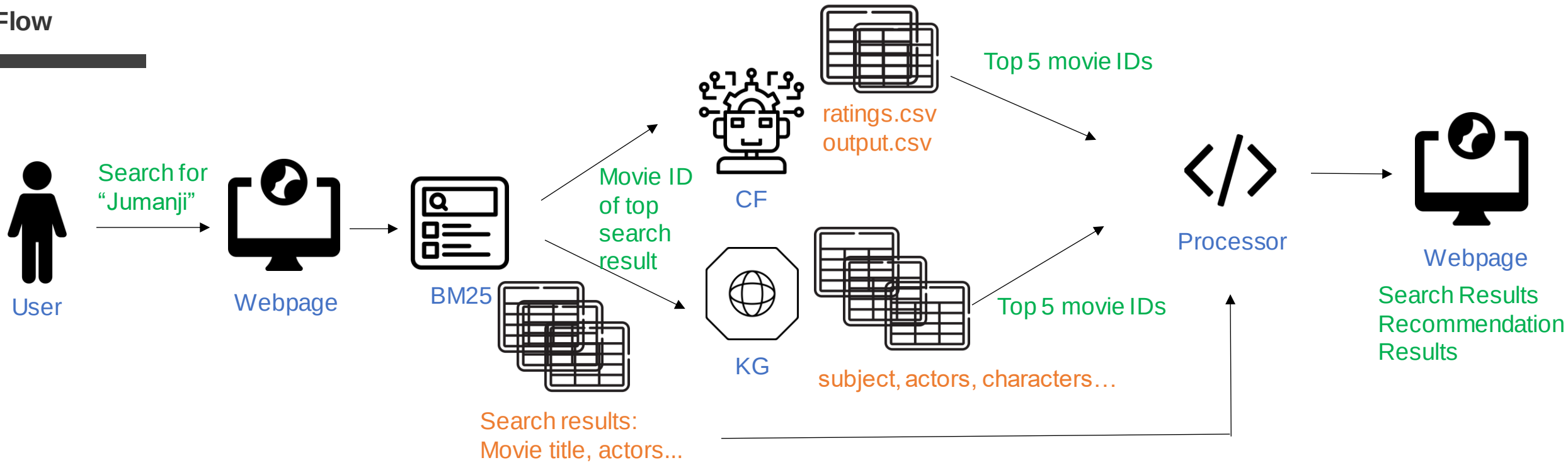
Hyeonjae Cho	hc53@illinois.edu
Xi Chen	xi12@illinois.edu
Sean Kan	clkan2@illinois.edu
Danwoo Park	danwoop2@illinois.edu
Kihyuk Song	kihyuks2@illinois.edu

Overview

A movie search and recommendation system

- Text retrieval model : Okapi BM25
- Recommendation models : WikiData Knowledge Graph, Collaborative Filtering

Flow



Installation & Setup

GitHub Repository Link: https://github.com/myevertime/CourseProject

1. Git clone the main branch of git@github.com:myevertime/CourseProject.git

2. Create a virtual environment in the current directory:

```
python3.7 -m venv {{virtual_env_name}}
```

3. Activate the virtual environment:

```
source {{virtual_env_name}}/bin/activate
```

4. Install packages using *pip*:

```
cd CourseProject
```

```
pip install -r requirements.txt
```

5. Execute the app *main.py*

```
cd ../Flask\ App
```

```
python main.py
```

6. Run the app on http://127.0.0.1:5000/ in your browser.

Notes:

1. Your environment must be connected to the Internet. Our application connects to a remote search server and KG server.
2. The app is tested to work on Visual Studio, Python 3.7.15. We do not guarantee it to work on Python 3.7 on other environments, like Mac Terminal.

Installation & Setup

Package Installation

Link: https://github.com/myevertime/CourseProject/blob/main/requirements.txt

```
pip install -r requirements.txt
```

```
bottle==0.12.23
certifi==2022.9.24
charset-normalizer==2.1.1
click==8.1.3
Flask==2.2.2
idna==3.4
importlib-metadata==5.1.0
itsdangerous==2.1.2
Jinja2==3.1.2
joblib==1.2.0
MarkupSafe==2.1.1
metapy==0.2.13
mwparserfromhell==0.6.4
numpy==1.21.6
pandas==1.3.5
python-dateutil==2.8.2
pytz==2022.6
pywikibot==7.7.2
requests==2.28.1
scikit-learn==1.0.2
scipy==1.7.3
six==1.16.0
threadpoolctl==3.1.0
typing_extensions==4.4.0
urllib3==1.26.13
Werkzeug==2.2.2
zipp==3.11.0
```

Collaborative Filtering(CF)

- Item-based Collaborative Filtering

- User-based Collaborative Filtering

	title	imdbId	tmdbId	genres	tag
2	Jumanji (1995)	113497	8844	Adventure|Children|Fantasy	Robin Will
3	Waiting to Exhale (1995)	114885	31357	Comedy|Drama|Romance	based on
4	Father of the Bride Part II (1995)	113041	11862	Comedy	aging|bab
5	Heat (1995)	113277	949	Action|Crime|Thriller	imdb top
6	Sabrina (1995)	114319	11860	Comedy|Romance	remake|ch
7	Tom and Huck (1995)	112302	45325	Adventure|Children	based on
8	Sudden Death (1995)	114576	9091	Action	explosive|I
9	GoldenEye (1995)	113189	710	Action|Adventure|Thriller	007|Bond|
10	American President, The (1995)	112346	9087	Comedy|Drama|Romance	Romance|v
11	Dracula: Dead and Loving It (1995)	112896	12110	Comedy|Horror	dracula|sp
12	Balto (1995)	112453	21032	Adventure|Animation|Children	Ei muistaj
13	Nixon (1995)	113987	10858	Drama	biography
14	Cutthroat Island (1995)	112760	1408	Action|Adventure|Romance	exotic isla
15	Casino (1995)	112641	524	Crime|Drama	Mafia|Maf
16	Sense and Sensibility (1995)	114388	4584	Drama|Romance	chick flick|
17	Four Rooms (1995)	113101	5	Comedy	anthology
18	Ace Ventura: When Nature Calls (1995)	112281	9273	Comedy	detectiv
19	Money Train (1995)	113845	11517	Action|Comedy|Crime|Drama|Thriller	new york

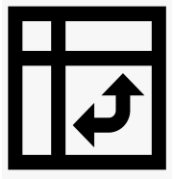
output.csv

rating groupby &
Merge
on 'movieId'

	userId	movieId	rating
2	1	306	3.5
3	1	307	5
4	1	899	3.5
5	1	1088	4
6	1	1175	3.5
7	1	1217	3.5
8	1	1237	5
9	1	1250	4
10	1	1260	3.5
11	1	2011	2.5
12	1	2012	2.5
13	1	2161	3.5
14	1	2351	4.5

ratings.csv

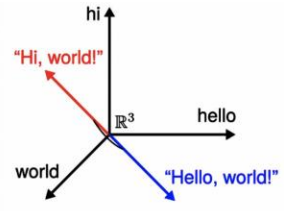
'movieId', 'userId', 'rating'
as a pivot input
(made as pickle file for Github upload)



$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

pivot & normalization

$$\begin{pmatrix} 10 & 0 & 0 & 12 & 0 \\ (0,0) & & & (0,3) & \\ 0 & 0 & 11 & 0 & 13 \\ & & (1,2) & & (1,4) \\ 0 & 16 & 0 & 0 & 0 \\ & (2,1) & & & \\ 0 & 0 & 11 & 0 & 13 \\ & & (3,2) & & (3,4) \end{pmatrix}$$



csr_matrix & cosine_similarity * cosine_similarity(csr_matrix.T)
= User-based CF

"Item-based CF"

similar_5movies()

Result : Top 5 movie IDs
[1380, 597, 1101, 1035, 1307]

Processor