

Applying NLP to extract entities from queries to
answer relevant questions and surface search
results from an IMDB movie dataset

CS410 Final Project Usage Tutorial

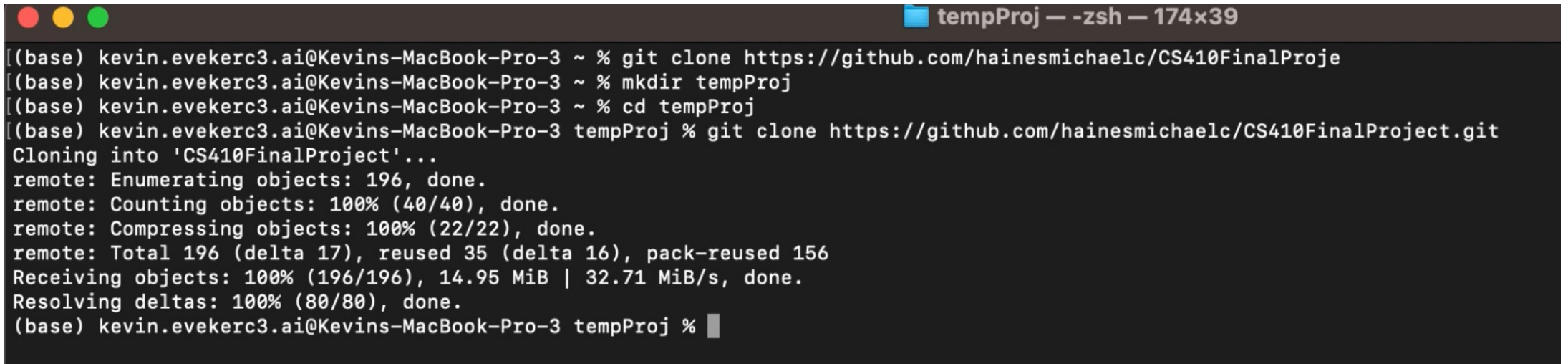
Team-MK

Captain: Michael Haines (mhaines2)

Kevin Eveker (keveker2)

Usage Guide: Solution Setup (1 of 5)

- We have a video that walks through these steps that can be found in our repo in the docs folder called: TeamMK_environment_setup_demo_720p.mov. It can also be found on github at this [link](https://github.com/hainesmichaelc/CS410FinalProject.git).
- Start by opening a terminal, creating a directory to work in, and cloning our repo:
 - git clone <https://github.com/hainesmichaelc/CS410FinalProject.git>



```
tempProj — -zsh — 174x39
((base) kevin.eveker3.ai@Kevins-MacBook-Pro-3 ~ % git clone https://github.com/hainesmichaelc/CS410FinalProje
((base) kevin.eveker3.ai@Kevins-MacBook-Pro-3 ~ % mkdir tempProj
((base) kevin.eveker3.ai@Kevins-MacBook-Pro-3 ~ % cd tempProj
((base) kevin.eveker3.ai@Kevins-MacBook-Pro-3 tempProj % git clone https://github.com/hainesmichaelc/CS410FinalProject.git
Cloning into 'CS410FinalProject'...
remote: Enumerating objects: 196, done.
remote: Counting objects: 100% (40/40), done.
remote: Compressing objects: 100% (22/22), done.
remote: Total 196 (delta 17), reused 35 (delta 16), pack-reused 156
Receiving objects: 100% (196/196), 14.95 MiB | 32.71 MiB/s, done.
Resolving deltas: 100% (80/80), done.
(base) kevin.eveker3.ai@Kevins-MacBook-Pro-3 tempProj %
```

Usage Guide: Solution Setup (2 of 5)

- Next step is to create a conda environment that has the right version of python:
 - `conda create --name envTeam-MK python=3.9.6`

```
(base) kevin.eveker3.ai@Kevins-MacBook-Pro-3 tempProj % conda create --name envTeam-MK python=3.9.6
Collecting package metadata (current_repodata.json): done
Solving environment: failed with repodata from current_repodata.json, will retry with next repodata source.
Collecting package metadata (repodata.json): done
Solving environment: done
```

```
==> WARNING: A newer version of conda exists. <==
current version: 4.14.0
latest version: 22.11.0
```

Please update conda by running

```
$ conda update -n base -c defaults conda
```

Package Plan

environment location: /opt/anaconda3/envs/envTeam-MK

added / updated specs:
- python=3.9.6

The following NEW packages will be INSTALLED:

ca-certificates	pkgs/main/osx-64::ca-certificates-2022.10.11-hecd8cb5_0
certifi	pkgs/main/osx-64::certifi-2022.9.24-py39hecd8cb5_0
libcxx	pkgs/main/osx-64::libcxx-14.0.6-h9765a3e_0
libffi	pkgs/main/osx-64::libffi-3.3-hb1e8313_2
ncurses	pkgs/main/osx-64::ncurses-6.3-hca72f7f_3
openssl	pkgs/main/osx-64::openssl-1.1.1s-hca72f7f_0
pip	pkgs/main/osx-64::pip-22.2.2-py39hecd8cb5_0
python	pkgs/main/osx-64::python-3.9.6-h88f2d9e_1
readline	pkgs/main/osx-64::readline-8.2-hca72f7f_0
setuptools	pkgs/main/osx-64::setuptools-65.5.0-py39hecd8cb5_0
sqlite	pkgs/main/osx-64::sqlite-3.40.0-h880c91c_0
tk	pkgs/main/osx-64::tk-8.6.12-h5d9f67b_0
tzdata	pkgs/main/noarch::tzdata-2022f-h04d1e81_0
wheel	pkgs/main/noarch::wheel-0.37.1-pyhd3eb1b0_0
xz	pkgs/main/osx-64::xz-5.2.8-h6c40b1e_0
zlib	pkgs/main/osx-64::zlib-1.2.13-h4dc903c_0

Proceed ([y]/n)? y

```
Preparing transaction: done
Verifying transaction: -
```

```
Preparing transaction: done
Verifying transaction: done
Executing transaction: done
#
# To activate this environment, use
#
#     $ conda activate envTeam-MK
#
# To deactivate an active environment, use
#
#     $ conda deactivate
```

```
Retrieving notices: ...working... done
(base) kevin.eveker3.ai@Kevins-MacBook-Pro-3 tempProj %
```

Usage Guide: Solution Setup (3 of 5)

- Then activate the conda environment and install dependencies:
 - conda activate envTeam-MK
 - pip install flask google urllib3 bs4 pandas datasloth

```
[(base) kevin.eveker3.ai@Kevins-MacBook-Pro-3 tempProj % conda activate envTeam-MK
(envTeam-MK) kevin.eveker3.ai@Kevins-MacBook-Pro-3 tempProj % pip install flask google urllib3 bs4 pandas datasloth cinemagoer
Collecting flask
  Using cached Flask-2.2.2-py3-none-any.whl (101 kB)
Collecting google
  Using cached google-3.0.0-py2.py3-none-any.whl (45 kB)
Collecting urllib3
  Using cached urllib3-1.26.13-py2.py3-none-any.whl (140 kB)
Collecting bs4
  Using cached bs4-0.0.1-py3-none-any.whl
Collecting pandas
  Using cached pandas-1.5.2-cp39-cp39-macosx_10_9_x86_64.whl (12.1 MB)
Collecting datasloth
  Using cached datasloth-0.4-py3-none-any.whl
Collecting cinemagoer
  Using cached cinemagoer-2022.12.4-py3-none-any.whl (299 kB)
Collecting Jinja2>=3.0
  Using cached Jinja2-3.1.2-py3-none-any.whl (133 kB)
Collecting importlib-metadata>=3.6.0
  Using cached importlib_metadata-5.1.0-py3-none-any.whl (21 kB)
Collecting itsdangerous>=2.0
  Using cached itsdangerous-2.1.2-py3-none-any.whl (15 kB)
Collecting click>=8.0
  Using cached click-8.1.3-py3-none-any.whl (96 kB)
Collecting Werkzeug>=2.2.2
  Using cached Werkzeug-2.2.2-py3-none-any.whl (232 kB)
Collecting beautifulsoup4
  Using cached beautifulsoup4-4.11.1-py3-none-any.whl (118 kB)
Collecting python-dateutil>=2.8.1
  Using cached python_dateutil-2.8.2-py2.py3-none-any.whl (247 kB)
Collecting pytz>=2020.1
  Using cached pytz-2022.6-py2.py3-none-any.whl (498 kB)
Collecting numpy>=1.20.3
  Using cached numpy-1.23.5-cp39-cp39-macosx_10_9_x86_64.whl (18.1 MB)
Collecting pandasql
  Using cached pandasql-0.7.3-py3-none-any.whl
Collecting openai
  Using cached openai-0.25.0-py3-none-any.whl
```

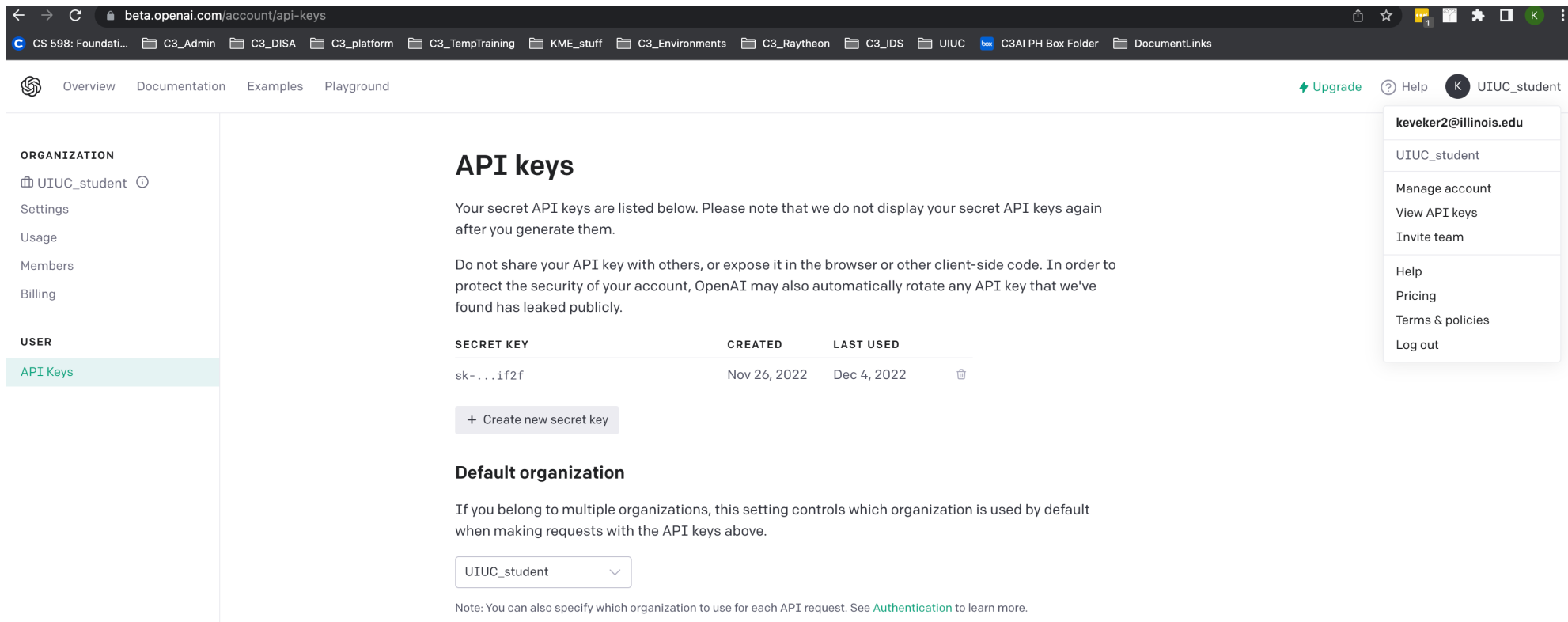
```
Collecting SQLAlchemy
  Using cached SQLAlchemy-1.4.44-cp39-cp39-macosx_10_15_x86_64.whl (1.6 MB)
Collecting zipp>=0.5
  Using cached zipp-3.11.0-py3-none-any.whl (6.6 kB)
Collecting MarkupSafe>=2.0
  Using cached MarkupSafe-2.1.1-cp39-cp39-macosx_10_9_x86_64.whl (13 kB)
Collecting six>=1.5
  Using cached six-1.16.0-py2.py3-none-any.whl (11 kB)
Collecting soupsieve>1.2
  Using cached soupsieve-2.3.2.post1-py3-none-any.whl (37 kB)
Collecting typing-extensions
  Using cached typing_extensions-4.4.0-py3-none-any.whl (26 kB)
Collecting requests>=2.20
  Using cached requests-2.28.1-py3-none-any.whl (62 kB)
Collecting pandas-stubs>=1.1.0.11
  Using cached pandas_stubs-1.5.2.221124-py3-none-any.whl (146 kB)
Collecting tqdm
  Using cached tqdm-4.64.1-py2.py3-none-any.whl (78 kB)
Collecting openpyxl>=3.0.7
  Using cached openpyxl-3.0.10-py2.py3-none-any.whl (242 kB)
Collecting greenlet!=0.4.17
  Using cached greenlet-2.0.1-cp39-cp39-macosx_10_15_x86_64.whl (203 kB)
Collecting et-xmlfile
  Using cached et_xmlfile-1.1.0-py3-none-any.whl (4.7 kB)
Collecting types-pytz>=2022.1.1
  Using cached types_pytz-2022.6.0.1-py3-none-any.whl (4.7 kB)
Collecting charset-normalizer<3,>=2
  Using cached charset_normalizer-2.1.1-py3-none-any.whl (39 kB)
Collecting idna<4,>=2.5
  Using cached idna-3.4-py3-none-any.whl (61 kB)
Requirement already satisfied: certifi>=2017.4.17 in /opt/anaconda3/envs/envTeam-MK/lib/python3.9/site-packages (from requests>=2.20->openai->datasloth) (2022.9.24)
Installing collected packages: types-pytz, pytz, zipp, urllib3, typing-extensions, tqdm, soupsieve, six, pandas-stubs, numpy, MarkupSafe, lxml, itsdangerous, idna, greenlet, et-xmlfile, click, charset-normalizer, Werkzeug, SQLAlchemy, requests, python-dateutil, openpyxl, Jinja2, importlib-metadata, beautifulsoup4, pandas, google, flask, cinemagoer, bs4, pandasql, openai, datasloth
Successfully installed Jinja2-3.1.2 MarkupSafe-2.1.1 SQLAlchemy-1.4.44 Werkzeug-2.2.2 beautifulsoup4-4.11.1 bs4-0.0.1 charset-normalizer-2.1.1 cinemagoer-2022.12.4 click-8.1.3 datasloth-0.4 et-xmlfile-1.1.0 flask-2.2.2 google-3.0.0 greenlet-2.0.1 idna-3.4 importlib-metadata-5.1.0 itsdangerous-2.1.2 lxml-4.9.1 numpy-1.23.5 openai-0.25.0 openpyxl-3.0.10 pandas-1.5.2 pandas-stubs-1.5.2.221124 pandasql-0.7.3 python-dateutil-2.8.2 pytz-2022.6 requests-2.28.1 six-1.16.0 soupsieve-2.3.2.post1 tqdm-4.64.1 types-pytz-2022.6.0.1 typing-extensions-4.4.0 urllib3-1.26.13 zipp-3.11.0
(envTeam-MK) kevin.eveker3.ai@Kevins-MacBook-Pro-3 tempProj %
```

Usage Guide: Solution Setup (4 of 5)

- Add your OpenAI API key to config.py:
 - vi config.py

```
CS410FinalProject — vi config.py — 174x39
openai_api_key = "sk-secret"
browser = 'Mozilla/5.0'
~
~
```

- Replace sk-secret with the API key you get from <https://openai.com/api/>



The screenshot shows a web browser window with the URL `beta.openai.com/account/api-keys`. The browser's address bar and tabs are visible at the top. The page content is divided into a left sidebar and a main area. The sidebar has sections for 'ORGANIZATION' (with links to Overview, Documentation, Examples, and Playground) and 'USER' (with links to Settings, Usage, Members, and Billing). The main area is titled 'API keys' and contains a warning about not sharing API keys. Below the warning is a table with columns 'SECRET KEY', 'CREATED', and 'LAST USED'. The table contains one row with the value 'sk-...if2f', 'Nov 26, 2022', and 'Dec 4, 2022'. Below the table is a button '+ Create new secret key'. At the bottom, there is a 'Default organization' section with a dropdown menu showing 'UIUC_student'. A note at the very bottom states: 'Note: You can also specify which organization to use for each API request. See [Authentication](#) to learn more.'

API keys

Your secret API keys are listed below. Please note that we do not display your secret API keys again after you generate them.

Do not share your API key with others, or expose it in the browser or other client-side code. In order to protect the security of your account, OpenAI may also automatically rotate any API key that we've found has leaked publicly.

SECRET KEY	CREATED	LAST USED
sk-...if2f	Nov 26, 2022	Dec 4, 2022

[+ Create new secret key](#)

Default organization

If you belong to multiple organizations, this setting controls which organization is used by default when making requests with the API keys above.

UIUC_student

Note: You can also specify which organization to use for each API request. See [Authentication](#) to learn more.

Usage Guide: Solution Setup (5 of 5)

- Initialize your database:
 - cd scripts
 - python ./init_db.py

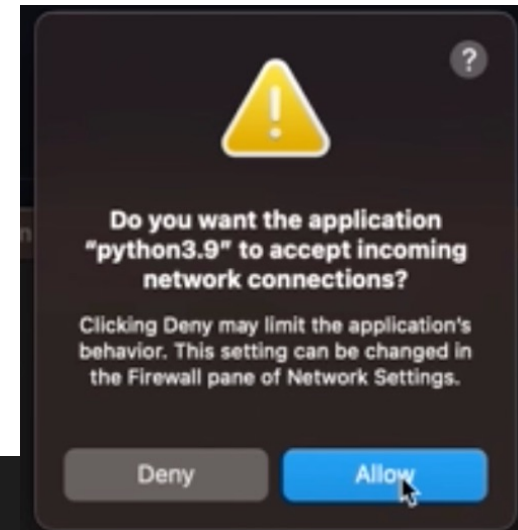
```
(envTeam-MK) kevin.eveker3.ai@Kevins-MacBook-Pro-3 CS410FinalProject % vi config.py  
(envTeam-MK) kevin.eveker3.ai@Kevins-MacBook-Pro-3 CS410FinalProject % cd scripts  
(envTeam-MK) kevin.eveker3.ai@Kevins-MacBook-Pro-3 scripts % python ./init_db.py
```

this step can take 30secs to 1 minute to complete, please stand by

Usage Guide: Solution Usage (1 of 4)

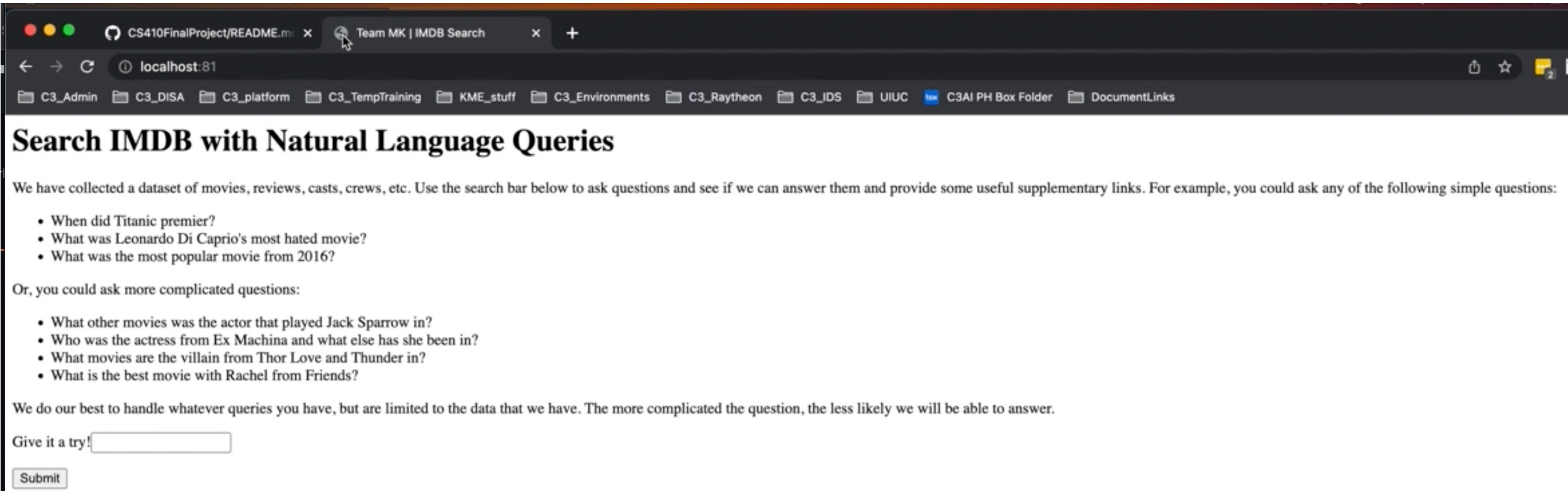
- We have a video that walks through these steps that can be found in our repo in the docs folder called: TeamMK_software_usage_demo_720p.mov. It can also be found on github at this [link](#).
- Move up to the main directory and run the start_app script
 - `cd ..`
 - `./scripts/start_app.sh`
- Click on allow so that web app will work correctly

```
(envTeam-MK) kevin.eveker3.ai@Kevins-MacBook-Pro-3 scripts % python ./init_db.py
(envTeam-MK) kevin.eveker3.ai@Kevins-MacBook-Pro-3 scripts % cd ..
(envTeam-MK) kevin.eveker3.ai@Kevins-MacBook-Pro-3 CS410FinalProject % ./scripts/start_app.sh
Password:
* Serving Flask app 'app'
* Debug mode: off
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
* Running on all addresses (0.0.0.0)
* Running on http://127.0.0.1:81
* Running on http://192.168.4.107:81
Press CTRL+C to quit
□
```



Usage Guide: Solution Usage (2 of 4)

- Navigate in browser to:
 - localhost:81
- This give you access to our software to run searches against out GPT-3 based engine, IMdB, and Google to compare results



Usage Guide: Solution Usage (3 of 4)

- A simple search for “Ex Machina” yields results from all 3 sources

Search Results Team-MK GPT-3 Based Result

Question: Ex Machina

Answer:

	title	release_date	vote_average	vote_count	revenue	budget	runtime	status	original_language	popularity	id
0	Ex Machina	2015-01-21	7.6	4737	36869414	15000000	108.0	Released	en	95.130041	264660

Generated Database Query:

```
SELECT m.title, m.release_date, m.vote_average, m.vote_count, m.revenue, m.budget, m.runtime, m.status, m.original_language, m.popularity, m.id
FROM m
WHERE m.title = 'Ex Machina'
```

Google Search Results

1. [https://en.wikipedia.org/wiki/Ex_Machina_\(film\)](https://en.wikipedia.org/wiki/Ex_Machina_(film))
2. https://en.wikipedia.org/wiki/Tiffany_Pisani
3. https://en.wikipedia.org/wiki/Knowledge_argument
4. <https://en.wikipedia.org/wiki/Valldalen>
5. [https://en.wikipedia.org/wiki/The_Machine_\(2013_film\)](https://en.wikipedia.org/wiki/The_Machine_(2013_film))
6. [https://en.wikipedia.org/wiki/Ex_Machina_\(comics\)](https://en.wikipedia.org/wiki/Ex_Machina_(comics))
7. <https://www.imdb.com/title/tt0470752/>
8. https://www.rottentomatoes.com/m/ex_machina
9. <https://www.youtube.com/watch?v=hggUmgeMCdc>
10. <https://www.youtube.com/watch?v=EoQuVnKhxaM>

IMDB Search Results

Movies:

1. [Ex Machina](#)
2. [Appleseed: Ex Machina](#)
3. [Deus Ex Machina](#)
4. [I Am Mother](#)
5. [Deus ex Machina](#)

People:

1. [Ex Machina](#)
2. [Deus Ex Machina](#)
3. [Deus Ex Vapore Machina Devm](#)
4. [Cimex ex Machina](#)
5. [Jeffrey Katzenberg](#)

Usage Guide: Solution Usage (4 of 4)

- A complex search for “Who was the actress from Ex Machina and what else has she been in” results in:

Search Results Team-MK GPT-3 Based Result

Question: Who was the actress from Ex Machina and what else has she been in

Answer:

	actress_name	movie_title
0	Alicia Vikander	Jason Bourne
1	Alicia Vikander	Seventh Son
2	Alicia Vikander	The Man from U.N.C.L.E.
3	Alicia Vikander	Anna Karenina
4	Alicia Vikander	The Fifth Estate
5	Alicia Vikander	Burnt
6	Alicia Vikander	Ex Machina

Generated Database Query:

```
SELECT ca.name AS actress_name, m.title AS movie_title
FROM ca
JOIN m ON ca.movie_id = m.id
WHERE ca.name = 'Alicia Vikander'
```

Google Search Results

1. <https://www.imdb.com/name/nm2539953/>
2. https://en.wikipedia.org/wiki/Alicia_Vikander
3. https://en.wikipedia.org/wiki/Alicia_Vikander#Early_life
4. https://en.wikipedia.org/wiki/Alicia_Vikander#Career
5. https://en.wikipedia.org/wiki/Alicia_Vikander#Personal_life
6. https://en.wikipedia.org/wiki/Alicia_Vikander#Filmography
7. [https://en.wikipedia.org/wiki/Ex_Machina_\(film\)](https://en.wikipedia.org/wiki/Ex_Machina_(film))
8. <https://www.wmagazine.com/story/alicia-vikander-ex-machina-actress>
9. <https://www.gradesaver.com/ex-machina-film/study-guide/cast-list>
10. <https://www.nylon.com/articles/alicia-vikander-interview>

IMDB Search Results

Movies:

No results found for "who was the actress from ex machina and what else has she been in"

People:

No results found for "who was the actress from ex machina and what else has she been in"

Please try your own queries to
see how we compare to IMdB
and Google