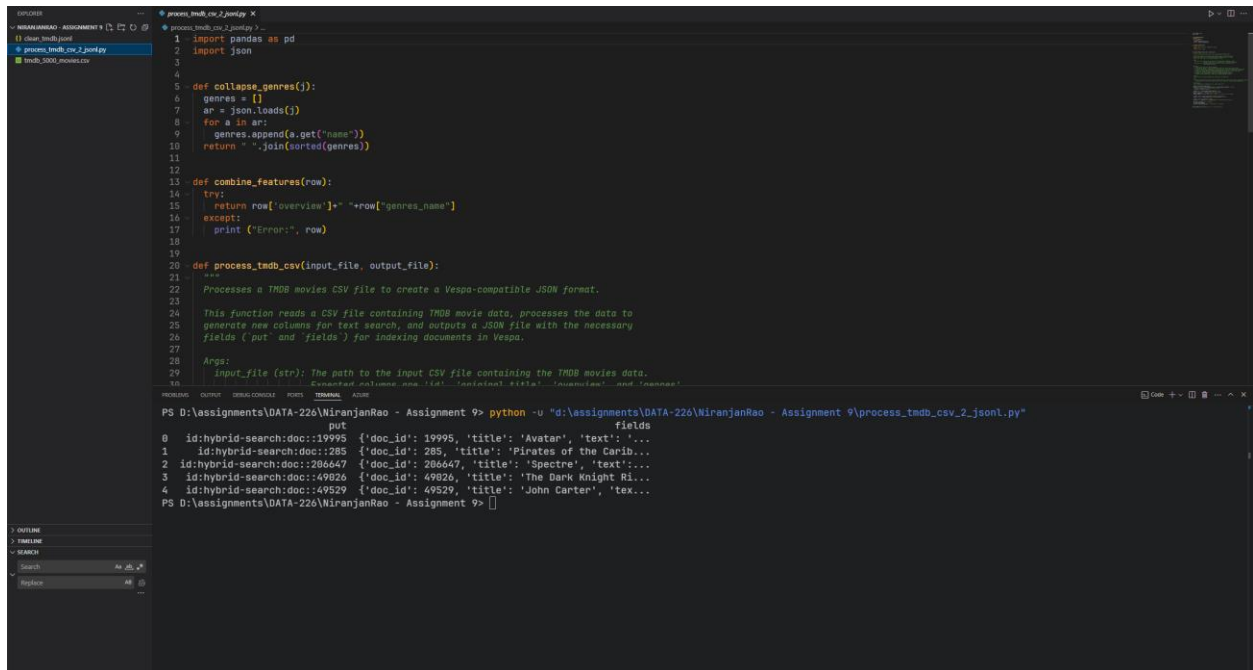


## Data 226 – Assignment 9

### 1. Download, Process and Generate an input file to VESPA (2pt)



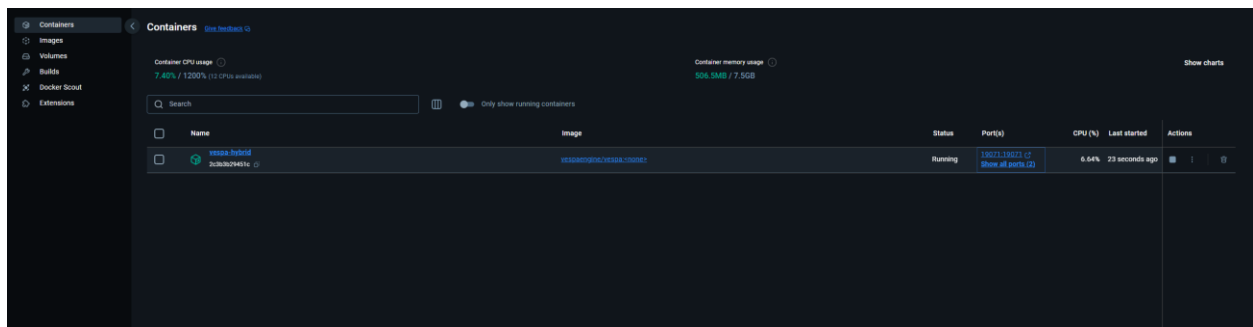
```
1 import pandas as pd
2 import json
3
4
5 def collapse_genres():
6     genres = []
7     ar = json.loads(i)
8     for a in ar:
9         genres.append(a.get("genres"))
10    return ".join(sorted(genres))"
11
12
13 def combine_features(row):
14     try:
15         return row["overview"] + " " + row["genres_name"]
16     except:
17         print("Error:", row)
18
19
20 def process_tmdb_csv(input_file, output_file):
21     """
22     Processes a TMDB movies CSV file to create a Vespa-compatible JSON format.
23
24     This function reads a CSV file containing TMDB movie data, processes the data to
25     generate new columns for text search, and outputs a JSON file with the necessary
26     fields (put and fields) for indexing documents in Vespa.
27
28     Args:
29         input_file (str): The path to the input CSV file containing the TMDB movies data.
30         output_file (str): The path to the output JSON file.
31     """
32     df = pd.read_csv(input_file)
33     df["genres_name"] = df["genres"].apply(collapse_genres)
34     df["text"] = df.apply(combine_features, axis=1)
35     df.to_json(output_file, orient="records", indent=2)
```

PS D:\assignments\DATA-226\NiranjanRao - Assignment 9> python -u "d:\assignments\DATA-226\NiranjanRao - Assignment 9\process\_tmdb\_csv\_2\_jsonl.py"

```
0 id:hybrid-search:doc:19995 {'doc_id': 19995, 'title': 'Avatar', 'text': '...'
1 id:hybrid-search:doc:285 {'doc_id': 285, 'title': 'Pirates of the Carib...
2 id:hybrid-search:doc:206647 {'doc_id': 206647, 'title': 'Spectre', 'text':...
3 id:hybrid-search:doc:49826 {'doc_id': 49826, 'title': 'The Dark Knight Ri...
4 id:hybrid-search:doc:49529 {'doc_id': 49529, 'title': 'John Carter', 'tex...
```

### 2. Run VESPA as a Docker Container (1pt)

```
PS D:\assignments\DATA-226\NiranjanRao - Assignment 9> docker run -d --name vespa-hybrid --hostname vespa-container --publish 8082:8080 --publish 19071:19071 vespaengine/vespa
2c3b3b29451c2f852c7f513a564e64e23886ec43c56c8e93568a43cecb852957
PS D:\assignments\DATA-226\NiranjanRao - Assignment 9>
```



[illegible]

### 3. Configure VESPA and Ingest the file into VESPA (2pt)

```

(vespa_env) PS D:\assignments\DATA-22A\NiranjanRao - Assignment 9> vespa config set target local
+ (vespa_env) PS D:\assignments\DATA-22A\NiranjanRao - Assignment 9> vespa status deploy --wait 300
Waiting up to 5m0s for deploy API...
Deploy API at http://127.0.0.1:19071 is ready
+ (vespa_env) PS D:\assignments\DATA-22A\NiranjanRao - Assignment 9> vespa deploy --wait 300 app
Waiting up to 5m0s for deploy API...
Uploading application package... done

Success: Deployed 'app' with session ID 2

```

```
PS D:\assignments\DATA-226\NiranjanRao - Assignment > vespa feed -t http://localhost:8082 clean_tmdb.jsonl
{
  "feeder.operation.count": 4803,
  "feeder.seconds": 27.479,
  "feeder.ok.count": 4803,
  "feeder.ok.rate": 174.786,
  "feeder.error.count": 0,
  "feeder.inflight.count": 0,
  "http.request.count": 4803,
  "http.request.bytes": 16708990,
  "http.request.Mbps": 0.861,
  "http.exception.count": 0,
  "http.response.count": 4803,
  "http.response.bytes": 429878,
  "http.response.Mbps": 0.016,
  "http.response.error.count": 0,
  "http.response.latency.millis.min": 25,
  "http.response.latency.millis.avg": 113,
  "http.response.latency.millis.max": 533,
  "http.response.code.counts": {
    "200": 4803
  }
}
```

4. Run search against VESPA (3pt)

```

PS D:\assignments\DATA-226\NiranjanRao - Assignment 9> python -u "d:\assignments\DATA-226\NiranjanRao - Assignment 9\pyvesap_script.py"
doc_id      title
0    767    Harry Potter and the Half-Blood Prince
1    671    Harry Potter and the Philosopher's Stone
2    674    Harry Potter and the Goblet of Fire
3    673    Harry Potter and the Prisoner of Azkaban
4    13967   Miss Potter
doc_id      title
0    767    Harry Potter and the Half-Blood Prince
1    675    Harry Potter and the Order of the Phoenix
2    672    Harry Potter and the Chamber of Secrets
3    674    Harry Potter and the Goblet of Fire
4    168795  BloodRayne
doc_id      title
0    767    Harry Potter and the Half-Blood Prince
1    675    Harry Potter and the Order of the Phoenix
2    672    Harry Potter and the Chamber of Secrets
3    671    Harry Potter and the Philosopher's Stone
4    674    Harry Potter and the Goblet of Fire
As Harry begins his sixth year at Hogwarts, he...
Returning for his fifth year of study at Hogsma...
Ignoring threats to his life, Harry returns to...
Harry Potter has lived under the stairs at his...
Harry starts his fourth year at Hogwarts, comp...
PS D:\assignments\DATA-226\NiranjanRao - Assignment 9>

```