

Multimedia Project 2

Website scraping into a Elasticsearch database

Jaime Romero Marín - 100346023
Alejandro Abanades Toribio - 100346089

Step 1: Definition of ES structure and mappings.	2
Step 2: Web scrapping of IMDB site.	5
Step 3: Queries	6
Query 1. Films about the World War II produced since 1980.	6
Query 2. Directors with the highest number of action films.	16
Query 3. Films about LGTBQ issues available in the collection and list with number of films per year.	16
Query 4. Films about corrupt politicians in Europe and the United States.	30

Step 1: Definition of ES structure and mappings.

For each document we decided to use an index mapping composed at least by:

- Title of the movie.
- The director or a list of directors.
- The date when the film was released.
 - The day, month and year are also defined in different attributes to be processed in different ways by the system.
- The genre or list of genres where the movie is classified.
- The keywords used by IMDB for classifying the film.
- The description or summary of the film.
- The actor or list of actors participating in the film.

As endpoints for the index we used address 'localhost:9200/film_db/film' and used JSON to store the movies.

The mappings for the index are automatically generated by the ElasticSearch, here is an example for the film 'The Matrix'.

```
{
  "title": "The Matrix",
  "directors": [
    "Lana Wachowski",
    "Lilly Wachowski"
  ],
  "release": "1999-03-31",
  "day": "31",
  "month": "03",
  "year": "1999",
  "genres": [
    "Action",
    "Sci-Fi"
  ],
  "keywords": "simulated reality,artificial reality,w\nar with machines,computer hacker,post apocalypse",
  "description": "The Matr\nix is a movie starring Keanu Reeves, Laurence Fishburne, and Carrie-Anne Mo\nss. A computer hacker learns from mysterious rebels about the true nature o\nf his reality and his role in the war again\nst its controllers.",
  "actors": [
    "Keanu Reeves",
    "Laurence Fishburne",
    "Carrie-Anne Moss",
    "Hugo Weaving"
  ]
}
```

Here is the mapping of index 'film_db' automatically generated from the indexed data.

```
"film_db" : {
  "mappings" : {
    "properties" : {
      "actors" : {
        "type" : "text",
        "fields" : {
          "keyword" : {
            "type" : "keyword",
            "ignore_above" : 256
          }
        }
      },
      "day" : {
        "type" : "text",
        "fields" : {
          "keyword" : {
            "type" : "keyword",
            "ignore_above" : 256
          }
        }
      },
      "description" : {
        "type" : "text",
        "fields" : {
          "keyword" : {
            "type" : "keyword",
            "ignore_above" : 256
          }
        }
      },
      "directors" : {
        "type" : "text",
        "fields" : {
          "keyword" : {
            "type" : "keyword",
            "ignore_above" : 256
          }
        }
      },
      "genres" : {
        "type" : "text",
        "fields" : {
          "keyword" : {
            "type" : "keyword",
            "ignore_above" : 256
          }
        }
      },
      "keywords" : {
        "type" : "text",
        "fields" : {
          "keyword" : {
            "type" : "keyword",
```

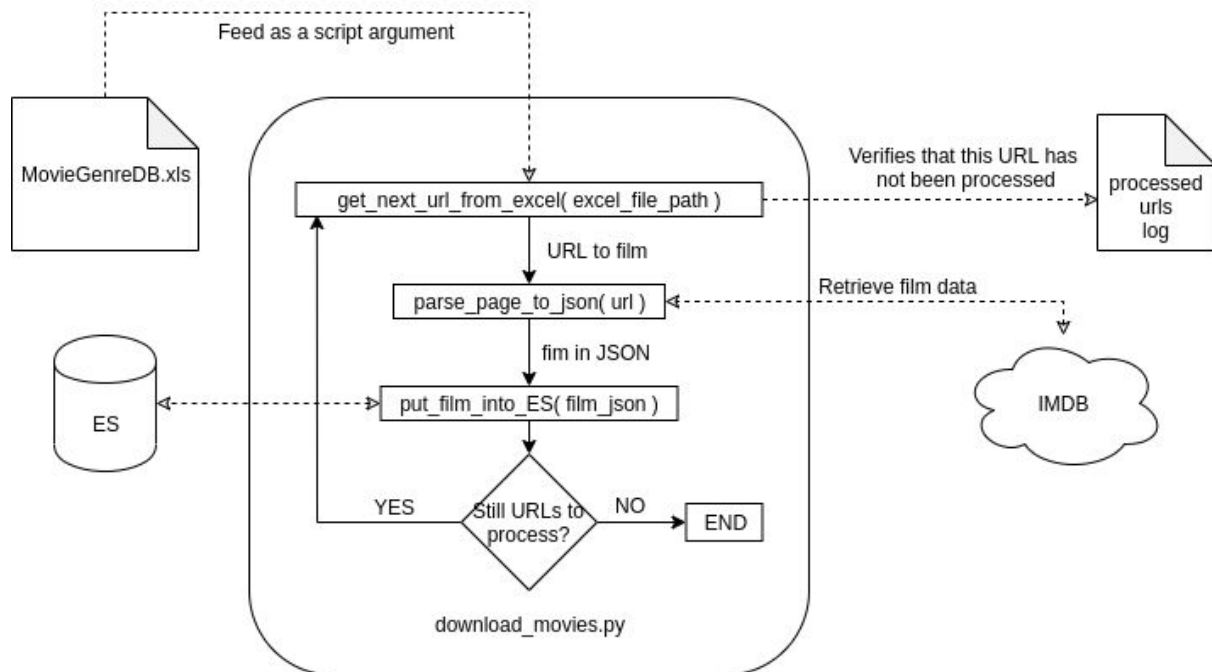
```
        "ignore_above" : 256
      }
    },
    "month" : {
      "type" : "text",
      "fields" : {
        "keyword" : {
          "type" : "keyword",
          "ignore_above" : 256
        }
      }
    },
    "release" : {
      "type" : "date"
    },
    "title" : {
      "type" : "text",
      "fields" : {
        "keyword" : {
          "type" : "keyword",
          "ignore_above" : 256
        }
      }
    },
    "year" : {
      "type" : "text",
      "fields" : {
        "keyword" : {
          "type" : "keyword",
          "ignore_above" : 256
        }
      }
    }
  }
}
```

Step 2: Web scraping of IMDB site.

For the web scraping of the IMDB site we used a Python script using interpreter Pypy3 (python3).

The script consists of a set of functions to transform the data from the excel delivered by the professor containing a list of URLs to movie pages in IMDB.

The basic algorithm is presented in the following graph.



- **MovieGenreDB.xls**: An excel file containing a list of movie movies where one of the attributes is the URL of the film in IMDB.
- **ES**: Our ElasticSearch instance with an endpoint on 'localhost:9200'.
- **Processed URLs Log**: To have a more robust operation, the script stores the already processed URLs in a file, which is reviewed every time the script starts to avoid downloading the same film several times into ES.
- **IMDB**: Is the remote website where the information is scrapped from. At this point, IMDB is located at the domain www.imdb.com.

As an improvement, we added multithread support to speed up the resolution of each of the URLs. However, this feature must be used with great care as its use could be interpreted as an attempted degeneration of service attack to the website and the IP of the current system is subjective to be blocked.

To solve this issue, there is a delay of 500ms for the dispatchment of the next URL processing.

To launch this script, it is recommended to initialize a python virtual environment and install the dependencies contained in the 'dependencies.txt' file using pip or the python package manager of your choice.

Step 3: Queries

Query 1. Films about the World War II produced since 1980.

GET film_db/_search

```
{
  "query": {
    "bool": {
      "must": [
        {
          "range": {
            "year": {"gt": 1980}
          }
        },
        {
          "bool": {
            "should": [
              {
                "match": {
                  "description": "World War II"
                }
              },
              {
                "match": {
                  "description": "World War ii"
                }
              },
              {
                "match": {
                  "description": "World War two"
                }
              },
              {
                "match": {
                  "description": "WWII"
                }
              },
              {
                "match": {
                  "description": "Operation Torch"
                }
              },
              {
                "match": {
                  "description": "Battle of the Atlantic"
                }
              }
            ]
          }
        }
      ]
    }
  }
}
```

```

    }
  },
  {
    "match": {
      "description": "Battle of Britain"
    }
  },
  {
    "match": {
      "description": "Operation Barbarossa"
    }
  },
  {
    "match": {
      "description": "Battle of Stalingrad"
    }
  },
  {
    "match": {
      "description": "Battle of Okinawa"
    }
  },
  {
    "match": {
      "keywords": "World War II"
    }
  },
  {
    "match": {
      "keywords": "World War ii"
    }
  },
  {
    "match": {
      "keywords": "World War two"
    }
  },
  {
    "match": {
      "keywords": "WWII"
    }
  },
  {
    "match": {
      "keywords": "Operation Torch"
    }
  },
  {
    "match": {
      "keywords": "Battle of the Atlantic"
    }
  },
  {

```



```

        "match": {
          "keywords": "Battle of Britain"
        }
      },
      {
        "match": {
          "keywords": "Operation Barbarossa"
        }
      },
      {
        "match": {
          "keywords": "Battle of Stalingrad"
        }
      },
      {
        "match": {
          "keywords": "Battle of Okinawa"
        }
      }
    ]
  }
}
]
}
"size": 10
}

```

Description.

In this query we use a ‘must filter’ and inside of it we check the range of the year in order to have been produce in years higher or equals than 1980. Then also inside the must operation we execute another bool query with should filter in which we compare the description and keywords fields with the next terms:

“World war II”, “World War ii”, “World War two”, “WWII”, “Operation Torch”, “Battle of the Atlantic”, “Battle of Britain”, “Operation Barbarossa”, “ Battle of Stalingrad”, “Battle of Okinawa”,

As you can see we have used in the query the common words for world war II but we have also used words of famous World War II battles.

As it is ask in the results we have added a extra parameter size : 10, for just listing the first 10 results.

Results

```

{
  "took" : 88,
  "timed_out" : false,
  "_shards" : {
    "total" : 1,
    "successful" : 1,
    "skipped" : 0,

```

```

    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 10000,
      "relation" : "gte"
    },
    "max_score" : 104.045364,
    "hits" : [
      {
        "_index" : "film_db",
        "_type" : "film",
        "_id" : "vZUTz24Bs3F6RdQUbj7h",
        "_score" : 104.045364,
        "_source" : {
          "keywords" : "cult director,battle of okinawa,computer,memory,japanese",
          "title" : "Level Five",
          "directors" : "Chris Marker",
          "release" : "1997-02-19",
          "day" : "19",
          "month" : "02",
          "year" : "1997",
          "genres" : [
            "Documentary",
            "Romance",
            "War"
          ],
          "description" : "Level Five is a movie starring Catherine Belkhodja, Kenji Tokitsu, and Nagisa Ôshima. The French computer programmer Laura inherits the task of making a computer game of the Battle of Okinawa during World War II. She searches the...",
          "actors" : [
            "Catherine Belkhodja",
            "Kenji Tokitsu",
            "Nagisa Ôshima",
            "Ju'nishi Ushiyama"
          ]
        }
      },
      {
        "_index" : "film_db",
        "_type" : "film",
        "_id" : "75VDz24Bs3F6RdQUtkjn",
        "_score" : 94.55098,
        "_source" : {
          "keywords" : "propaganda,battle of changsha,sino japanese war,chungking china,shanghai china",
          "title" : "The Battle of China",
          "directors" : [
            "Frank Capra",
            "Anatole Litvak"
          ],
          "release" : "2005-04-06",
          "day" : "06",
          "month" : "04",

```

```

    "year" : "2005",
    "genres" : [
      "Documentary",
      "History",
      "War"
    ],
    "description" : "The Battle of China is a movie starring Claire Chennault, Kai-Shek
Chiang, and Madame Chiang. The Official World War II US Government account of Chinese defense
against Japanese aggression.",
    "actors" : [
      "Claire Chennault",
      "Kai-Shek Chiang",
      "Madame Chiang",
      "Teh Chu"
    ]
  }
},
{
  "_index" : "film_db",
  "_type" : "film",
  "_id" : "wpU5z24Bs3F6RdQUYkY8",
  "_score" : 94.44892,
  "_source" : {
    "keywords" : "soldier,trenches,battle,battle of the somme,war violence",
    "title" : "The Trench",
    "directors" : "William Boyd",
    "release" : "1999-09-17",
    "day" : "17",
    "month" : "09",
    "year" : "1999",
    "genres" : [
      "Drama",
      "History",
      "War"
    ],
    "description" : "The Trench is a movie starring Paul Nicholls, Daniel Craig, and Julian
Rhind-Tutt. A story about a group of soldiers' last days before the battle of the Somme in 1916,
showing the conditions in the trenches during World War I, and...",
    "actors" : [
      "Paul Nicholls",
      "Daniel Craig",
      "Julian Rhind-Tutt",
      "Danny Dyer"
    ]
  }
},
{
  "_index" : "film_db",
  "_type" : "film",
  "_id" : "0ZV80G4Bs3F6RdQUtppP",
  "_score" : 94.44892,
  "_source" : {
    "keywords" : "soldier,trenches,battle,battle of the somme,war violence",
    "title" : "The Trench",

```

```

    "directors" : "William Boyd",
    "release" : "1999-09-17",
    "day" : "17",
    "month" : "09",
    "year" : "1999",
    "genres" : [
        "Drama",
        "History",
        "War"
    ],
    "description" : "The Trench is a movie starring Paul Nicholls, Daniel Craig, and Julian Rhind-Tutt. A story about a group of soldiers' last days before the battle of the Somme in 1916, showing the conditions in the trenches during World War I, and...",
    "actors" : [
        "Paul Nicholls",
        "Daniel Craig",
        "Julian Rhind-Tutt",
        "Danny Dyer"
    ]
}
},
{
    "_index" : "film_db",
    "_type" : "film",
    "_id" : "spVQz24Bs3F6RdQUVkv",
    "_score" : 90.4247,
    "_source" : {
        "keywords" : "sniper,stalingrad,battle,russian,battle of stalingrad",
        "title" : "Enemy at the Gates",
        "directors" : "Jean-Jacques Annaud",
        "release" : "2001-03-14",
        "day" : "14",
        "month" : "03",
        "year" : "2001",
        "genres" : [
            "Drama",
            "History",
            "War"
        ],
        "description" : "Enemy at the Gates is a movie starring Jude Law, Ed Harris, and Joseph Fiennes. A Russian and a German sniper play a game of cat-and-mouse during the Battle of Stalingrad.",
        "actors" : [
            "Jude Law",
            "Ed Harris",
            "Joseph Fiennes",
            "Rachel Weisz"
        ]
    }
}
},
{
    "_index" : "film_db",
    "_type" : "film",
    "_id" : "eJUNz24Bs3F6RdQUxj1U",

```

```

    "_score" : 83.62189,
    "_source" : {
      "keywords" : "falling in love,1940s,post world war ii,aftermath of world war ii,war
aftermath",
      "title" : "A Walk in the Clouds",
      "directors" : "Alfonso Arau",
      "release" : "1995-05-27",
      "day" : "27",
      "month" : "05",
      "year" : "1995",
      "genres" : [
        "Drama",
        "Romance"
      ],
      "description" : "A Walk in the Clouds is a movie starring Keanu Reeves, Aitana
Sánchez-Gijón, and Anthony Quinn. A married soldier returning from World War II poses as a pregnant
woman's husband to save her from her father's anger and honor.",
      "actors" : [
        "Keanu Reeves",
        "Aitana Sánchez-Gijón",
        "Anthony Quinn",
        "Giancarlo Giannini"
      ]
    }
  },
  {
    "_index" : "film_db",
    "_type" : "film",
    "_id" : "wJXuz24Bs3F6RdQUUm3C",
    "_score" : 73.115364,
    "_source" : {
      "keywords" : "deception,character name as title,battle,saboteur,vomiting",
      "title" : "Max Manus",
      "directors" : [
        "Joachim Rønning",
        "Espen Sandberg"
      ],
      "release" : "2008-12-19",
      "day" : "19",
      "month" : "12",
      "year" : "2008",
      "genres" : [
        "Action",
        "Biography",
        "Drama",
        "History",
        "War"
      ],
      "description" : "Max Manus is a movie starring Aksel Hennie, Agnes Kittelsen, and Nicolai
Cleve Broch. The true story about one of the most brilliant saboteurs during World War II and his
battle to overcome his inner demons.",
      "actors" : [
        "Aksel Hennie",
        "Agnes Kittelsen",

```

```

        "Nicolai Cleve Broch",
        "Ken Duken"
    ]
}
},
{
    "_index" : "film_db",
    "_type" : "film",
    "_id" : "TZX1zm4Bs3F6RdQU4TiV",
    "_score" : 72.43471,
    "_source" : {
        "keywords" : "american civil war,battle of gettysburg,hug,gun,robert e lee character",
        "title" : "Gettysburg",
        "directors" : "Ron Maxwell",
        "release" : "1993-10-08",
        "day" : "08",
        "month" : "10",
        "year" : "1993",
        "genres" : [
            "Drama",
            "History",
            "War"
        ],
        "description" : "Gettysburg is a movie starring Tom Berenger, Martin Sheen, and Stephen
Lang. In 1863, the Northern and Southern forces fight at Gettysburg in the decisive battle of the
American Civil War.",
        "actors" : [
            "Tom Berenger",
            "Martin Sheen",
            "Stephen Lang",
            "Richard Jordan"
        ]
    }
},
{
    "_index" : "film_db",
    "_type" : "film",
    "_id" : "bpWPz24Bs3F6RdQUyVkn",
    "_score" : 70.27308,
    "_source" : {
        "keywords" : "world war two,u.s. marine corps,horrors of war,tropical setting,pacific
theater",
        "title" : "The Pacific",
        "release" : "2010-03-14",
        "day" : "14",
        "month" : "03",
        "year" : "2010",
        "genres" : [
            "Action",
            "Adventure",
            "Drama",
            "History",
            "War"
        ],
    },
}

```

```

      "description" : "The Pacific is a TV mini-series starring James Badge Dale, Joseph
Mazzello, and Jon Seda. The Pacific Theatre of World War II, as seen through the eyes of several
young Marines.",
      "actors" : [
        "James Badge Dale",
        "Joseph Mazzello",
        "Jon Seda",
        "Ashton Holmes"
      ]
    }
  },
  {
    "_index" : "film_db",
    "_type" : "film",
    "_id" : "jZWj0G4Bs3F6RdQU4rnB",
    "_score" : 68.709946,
    "_source" : {
      "keywords" : "2010s,world war two,1940s,luftwaffe,british military",
      "title" : "Flight World War II",
      "directors" : "Emile Edwin Smith",
      "release" : "2015-06-02",
      "day" : "02",
      "month" : "06",
      "year" : "2015",
      "genres" : [
        "Action",
        "Adventure",
        "Sci-Fi",
        "Thriller",
        "War"
      ],
      "description" : "Flight World War II is a movie starring Faran Tahir, Robbie Kay, and
Aqueela Zoll. After Flight 42 travels through a storm they find themselves in France, 1940, during
World war II.",
      "actors" : [
        "Faran Tahir",
        "Robbie Kay",
        "Aqueela Zoll",
        "Matias Ponce"
      ]
    }
  }
]
}
}

```

Evaluation

To perform the evaluation as we cannot check the evaluation of the total set of the films we are going to check the precision for the 10 first results.

$$Precision = \frac{\text{Films returned that are relevant}}{\text{Total films}} = \frac{8}{10} = 0.8 = 80\% \text{ of precision}$$

Query 2. Directors with the highest number of action films.

GET film_db/_search

```
{
  "query": {
    "bool": {
      "must": [
        {
          "match": {
            "genres": "action"
          }
        }
      ]
    }
  },
  "aggs": {
    "directores": {
      "terms": {
        "field": "directors.keyword"
        , "size": 10
      }
    }
  }
  , "size": 0
}
```

Description

In this query first we use a must filter to check that the genre of the film is action (or at least one of its genres) and then we use an aggregation of the field “director” that acts like a “group by” in Standard Query Language and group all directors returning the total value of each of them. As it is asked just for the directors we have add a “size : 0” to not skip the collection of films and just return the directors. Inside the aggregation we have added a “size: 10” to list the 10 directors with more action films, although this size is already set to 10 in kibana.

Results

```
{
  "took" : 1,
  "timed_out" : false,
  "_shards" : {
    "total" : 1,
```



```

    "successful" : 1,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 6736,
      "relation" : "eq"
    },
    "max_score" : null,
    "hits" : [ ]
  },
  "aggregations" : {
    "directores" : {
      "doc_count_error_upper_bound" : 0,
      "sum_other_doc_count" : 7088,
      "buckets" : [
        {
          "key" : "John Woo",
          "doc_count" : 28
        },
        {
          "key" : "Sammo Kam-Bo Hung",
          "doc_count" : 25
        },
        {
          "key" : "Albert Pyun",
          "doc_count" : 23
        },
        {
          "key" : "Cheh Chang",
          "doc_count" : 23
        },
        {
          "key" : "Robert N. Bradbury",
          "doc_count" : 22
        },
        {
          "key" : "Chia-Liang Liu",
          "doc_count" : 21
        },
        {
          "key" : "Hark Tsui",
          "doc_count" : 20
        },
        {
          "key" : "Tony Scott",
          "doc_count" : 20
        },
        {
          "key" : "Kenji Misumi",
          "doc_count" : 19
        },
        {

```

```

        "key" : "Daisuke Nishio",
        "doc_count" : 18
    }
]
}
}
}

```

Evaluation:

To perform the evaluation as we cannot check the evaluation of the total set of the films we are going to check the precision for the 10 first results.

$$Precision = \frac{\text{Films returned that are relevant}}{\text{Total films}} = \frac{10}{10} = X = X\% \text{ of precision}$$

Query 3. Films about LGTBQ issues available in the collection and list with number of films per year.

```

GET film_db/_search
{
  "query": {
    "bool": {
      "should": [
        {
          "match": {
            "description": "lesbian"
          }
        },
        {
          "match": {
            "description": "gay"
          }
        },
        {
          "match": {
            "description": "bisexual"
          }
        },
        {
          "match": {
            "description": "transgender"
          }
        },
        {
          "match": {
            "description": "queer"
          }
        }
      ]
    }
  }
}

```

```

        "keywords": "lesbian"
      }
    },
    {
      "match": {
        "keywords": "gay"
      }
    },
    {
      "match": {
        "keywords": "bisexual"
      }
    },
    {
      "match": {
        "keywords": "transgender"
      }
    },
    {
      "match": {
        "keywords": "queer"
      }
    }
  ]
}
},
"aggs": {
  "per_year": {
    "terms": {
      "field": "year.keyword",
      "size": 100,
      "order" : { "_key" : "desc" }
    }
  }
},
"size": 10
}

```

Description

In this query we use a should filter in which we compare the description and keyword fields with the terms of the initials LGTBQ, that is: “lesbian”, “gay”, “transgender”, “bisexual” and “queer”. Then, we use an aggregation to group all the films by its year. For this, we use the “year” field.

We are asked for the films but also for the aggroupation by year so this time both sizes are set to 10 (not just one as in the previous query).

Results

```

{
  "took" : 29,
  "timed_out" : false,
  "_shards" : {

```

```

    "total" : 1,
    "successful" : 1,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 1138,
      "relation" : "eq"
    },
    "max_score" : 24.424006,
    "hits" : [
      {
        "_index" : "film_db",
        "_type" : "film",
        "_id" : "u5Wa0G4Bs3F6RdQUkLAs",
        "_score" : 24.424006,
        "_source" : {
          "keywords" : "queer cinema,gay interest,new york city,lgbt,gay character",
          "title" : "I Do",
          "directors" : "Glenn Gaylord",
          "release" : "2012-10-20",
          "day" : "20",
          "month" : "10",
          "year" : "2012",
          "genres" : [
            "Drama",
            "Romance"
          ],
          "description" : "I Do is a movie starring Jamie-Lynn Sigler, Alicia Witt, and Maurice
Compte. A gay Brit living in New York is deprived of his immigration status, and risks losing his
family and life in the U.S. He marries his lesbian best friend to...",
          "actors" : [
            "Jamie-Lynn Sigler",
            "Alicia Witt",
            "Maurice Compte",
            "David W. Ross"
          ]
        }
      },
      {
        "_index" : "film_db",
        "_type" : "film",
        "_id" : "kJW_zm4Bs3F6RdQUnizy",
        "_score" : 24.401428,
        "_source" : {
          "keywords" : "gay,stonewall riot,lesbian,drag,drag queen",
          "title" : "Before Stonewall",
          "directors" : [
            "Greta Schiller",
            "Robert Rosenberg"
          ],
          "release" : "1985-06-27",
          "day" : "27",

```

```

    "month" : "06",
    "year" : "1985",
    "genres" : [
      "Documentary",
      "History"
    ],
    "description" : "Before Stonewall is a movie starring Rita Mae Brown, Maua Adele Ajanaku,
and Red Jordan Arobateau. The history of the Gay and Lesbian community before the Stonewall riots
began the major gay rights movement.",
    "actors" : [
      "Rita Mae Brown",
      "Maua Adele Ajanaku",
      "Red Jordan Arobateau",
      "Ann Bannon"
    ]
  }
},
{
  "_index" : "film_db",
  "_type" : "film",
  "_id" : "25UZ0G4Bs3F6RdQUhXYK",
  "_score" : 22.207678,
  "_source" : {
    "keywords" : "lesbian,gay,senior citizen,kansas,florida",
    "title" : "Out Late",
    "directors" : [
      "Beatrice Alda",
      "Jennifer Brooke"
    ],
    "release" : "2011-10-25",
    "day" : "25",
    "month" : "10",
    "year" : "2011",
    "genres" : "Documentary",
    "description" : "Out Late is a movie starring Leanna Bradley, Cathy Jambrosic, and Walter
Rothenburger. This feature film looks at five individuals who made a decisive change later in
life-to come out as lesbian, gay, or trans gender, after the age...",
    "actors" : [
      "Leanna Bradley",
      "Cathy Jambrosic",
      "Walter Rothenburger",
      "Ken Rusk"
    ]
  }
},
{
  "_index" : "film_db",
  "_type" : "film",
  "_id" : "qJWS0G4Bs3F6RdQUm6oM",
  "_score" : 22.207678,
  "_source" : {
    "keywords" : "lesbian,gay,senior citizen,kansas,florida",
    "title" : "Out Late",
    "directors" : [

```

```

        "Beatrice Alda",
        "Jennifer Brooke"
    ],
    "release" : "2011-10-25",
    "day" : "25",
    "month" : "10",
    "year" : "2011",
    "genres" : "Documentary",
    "description" : "Out Late is a movie starring Leanna Bradley, Cathy Jambrosic, and Walter
Rothenburger. This feature film looks at five individuals who made a decisive change later in
life-to come out as lesbian, gay, or trans gender, after the age...",
    "actors" : [
        "Leanna Bradley",
        "Cathy Jambrosic",
        "Walter Rothenburger",
        "Ken Rusk"
    ]
}
},
{
    "_index" : "film_db",
    "_type" : "film",
    "_id" : "C5WX0G4Bs3F6RdQU96ye",
    "_score" : 21.958206,
    "_source" : {
        "keywords" : "transgender,ftm,gender reassignment,friendship,lesbian",
        "title" : "Romeos",
        "directors" : "Sabine Bernardi",
        "release" : "2011-12-08",
        "day" : "08",
        "month" : "12",
        "year" : "2011",
        "genres" : [
            "Drama",
            "Romance"
        ],
        "description" : "Romeos is a movie starring Rick Okon, Max Befort, and Liv Lisa Fries. A
drama centered on the relationship between a two young men, as one of them navigates the
difficulties of life as a transgender man.",
        "actors" : [
            "Rick Okon",
            "Max Befort",
            "Liv Lisa Fries",
            "Felix Brocke"
        ]
    }
},
{
    "_index" : "film_db",
    "_type" : "film",
    "_id" : "0pVozm4Bs3F6RdQUBmG",
    "_score" : 19.627552,
    "_source" : {
        "keywords" : "bisexual,lesbianism,menage a trois,lesbian,woman murders a woman",

```

```

    "title" : "Les biches",
    "directors" : "Claude Chabrol",
    "release" : "1968-03-22",
    "day" : "22",
    "month" : "03",
    "year" : "1968",
    "genres" : "Drama",
    "description" : "Les biches is a movie starring Jean-Louis Trintignant, Jacqueline
Sassard, and Stéphane Audran. Architect Paul Thomas insinuates himself into the relationship of two
bisexual women living in a St. Tropez villa with tragic...",
    "actors" : [
        "Jean-Louis Trintignant",
        "Jacqueline Sassard",
        "Stéphane Audran",
        "Nane Germon"
    ]
}
},
{
    "_index" : "film_db",
    "_type" : "film",
    "_id" : "hZWj0G4Bs3F6RdQU3rmH",
    "_score" : 19.586975,
    "_source" : {
        "keywords" : "male to female transgender,female to male transgender,gender
dysphoria,transgender child,transgender girl",
        "title" : "Louis Theroux: Transgender Kids",
        "directors" : "Tom Barrow",
        "release" : "2015-04-05",
        "day" : "05",
        "month" : "04",
        "year" : "2015",
        "genres" : "Documentary",
        "description" : "Louis Theroux: Transgender Kids is a TV movie starring Louis Theroux,
Camille, and Casey. A freewheeling Detroit cop pursuing a murder investigation finds himself dealing
with the very different culture of Beverly Hills.",
        "actors" : [
            "Louis Theroux",
            "Camille",
            "Casey",
            "Eduardo"
        ]
    }
},
{
    "_index" : "film_db",
    "_type" : "film",
    "_id" : "w5Wfzm4Bs3F6RdQUnyXn",
    "_score" : 18.728516,
    "_source" : {
        "keywords" : "gay brother,male rear nudity,closeted homosexual,gay relationship,gay son",
        "title" : "Dimenticare Venezia",
        "directors" : "Franco Brusati",
        "release" : "1979-04-14",

```

```

    "day" : "14",
    "month" : "04",
    "year" : "1979",
    "genres" : "Drama",
    "description" : "Dimenticare Venezia is a movie starring Mariangela Melato, Eleonora
Giorgi, and Erland Josephson. A gay brother visits his childhood home where his lesbian sister still
lives. They reminisce about the better past and their parents.",
    "actors" : [
        "Mariangela Melato",
        "Eleonora Giorgi",
        "Erland Josephson",
        "Nerina Montagnani"
    ]
}
},
{
    "_index" : "film_db",
    "_type" : "film",
    "_id" : "sZUXz24Bs3F6RdQUzD_e",
    "_score" : 18.500101,
    "_source" : {
        "keywords" : "triple f rated,feminist film,new queer cinema,lesbian,f rated",
        "title" : "The Watermelon Woman",
        "directors" : "Cheryl Dunye",
        "release" : "1997-03-05",
        "day" : "05",
        "month" : "03",
        "year" : "1997",
        "genres" : [
            "Comedy",
            "Drama",
            "Romance"
        ],
        "description" : "The Watermelon Woman is a movie starring Cheryl Dunye, Guinevere Turner,
and Valarie Walker. Cheryl is young, Black, and lesbian, working in Philadelphia with her best
friend Tamara and consumed by a film project: to make a video...",
        "actors" : [
            "Cheryl Dunye",
            "Guinevere Turner",
            "Valarie Walker",
            "Lisa Marie Bronson"
        ]
    }
},
{
    "_index" : "film_db",
    "_type" : "film",
    "_id" : "kZWez24Bs3F6RdQUIVwe",
    "_score" : 18.47285,
    "_source" : {
        "keywords" : "road movie,transsexual,sex change,transgender,transgender protagonist",
        "title" : "Transamerica",
        "directors" : "Duncan Tucker",
        "release" : "2006-02-10",

```



```

    "day" : "10",
    "month" : "02",
    "year" : "2006",
    "genres" : [
      "Adventure",
      "Comedy",
      "Drama"
    ],
    "description" : "Transamerica is a movie starring Felicity Huffman, Kevin Zegers, and
Fionnula Flanagan. A preoperative transgender woman takes an unexpected journey when she learns that
she fathered a son, now a teenage runaway hustling on the...",
    "actors" : [
      "Felicity Huffman",
      "Kevin Zegers",
      "Fionnula Flanagan",
      "Andrea James"
    ]
  }
}
],
},
"aggregations" : {
  "per_year" : {
    "doc_count_error_upper_bound" : 0,
    "sum_other_doc_count" : 0,
    "buckets" : [
      {
        "key" : "2017",
        "doc_count" : 2
      },
      {
        "key" : "2016",
        "doc_count" : 20
      },
      {
        "key" : "2015",
        "doc_count" : 23
      },
      {
        "key" : "2014",
        "doc_count" : 15
      },
      {
        "key" : "2013",
        "doc_count" : 28
      },
      {
        "key" : "2012",
        "doc_count" : 30
      },
      {
        "key" : "2011",
        "doc_count" : 30
      }
    ]
  }
}

```

```
{
  "key" : "2010",
  "doc_count" : 25
},
{
  "key" : "2009",
  "doc_count" : 69
},
{
  "key" : "2008",
  "doc_count" : 56
},
{
  "key" : "2007",
  "doc_count" : 63
},
{
  "key" : "2006",
  "doc_count" : 54
},
{
  "key" : "2005",
  "doc_count" : 46
},
{
  "key" : "2004",
  "doc_count" : 46
},
{
  "key" : "2003",
  "doc_count" : 49
},
{
  "key" : "2002",
  "doc_count" : 46
},
{
  "key" : "2001",
  "doc_count" : 26
},
{
  "key" : "2000",
  "doc_count" : 38
},
{
  "key" : "1999",
  "doc_count" : 44
},
{
  "key" : "1998",
  "doc_count" : 48
},
{
  "key" : "1997",
```

```
"doc_count" : 32
},
{
  "key" : "1996",
  "doc_count" : 22
},
{
  "key" : "1995",
  "doc_count" : 23
},
{
  "key" : "1994",
  "doc_count" : 23
},
{
  "key" : "1993",
  "doc_count" : 11
},
{
  "key" : "1992",
  "doc_count" : 6
},
{
  "key" : "1991",
  "doc_count" : 11
},
{
  "key" : "1990",
  "doc_count" : 13
},
{
  "key" : "1989",
  "doc_count" : 9
},
{
  "key" : "1988",
  "doc_count" : 8
},
{
  "key" : "1987",
  "doc_count" : 7
},
{
  "key" : "1986",
  "doc_count" : 7
},
{
  "key" : "1985",
  "doc_count" : 8
},
{
  "key" : "1984",
  "doc_count" : 9
},
},
```

```
{
  "key" : "1983",
  "doc_count" : 10
},
{
  "key" : "1982",
  "doc_count" : 7
},
{
  "key" : "1981",
  "doc_count" : 5
},
{
  "key" : "1980",
  "doc_count" : 6
},
{
  "key" : "1979",
  "doc_count" : 7
},
{
  "key" : "1978",
  "doc_count" : 6
},
{
  "key" : "1977",
  "doc_count" : 6
},
{
  "key" : "1976",
  "doc_count" : 12
},
{
  "key" : "1975",
  "doc_count" : 10
},
{
  "key" : "1974",
  "doc_count" : 7
},
{
  "key" : "1973",
  "doc_count" : 1
},
{
  "key" : "1972",
  "doc_count" : 11
},
{
  "key" : "1971",
  "doc_count" : 6
},
{
  "key" : "1969",
```

```
    "doc_count" : 5
  },
  {
    "key" : "1968",
    "doc_count" : 6
  },
  {
    "key" : "1967",
    "doc_count" : 8
  },
  {
    "key" : "1966",
    "doc_count" : 4
  },
  {
    "key" : "1965",
    "doc_count" : 2
  },
  {
    "key" : "1964",
    "doc_count" : 7
  },
  {
    "key" : "1963",
    "doc_count" : 2
  },
  {
    "key" : "1962",
    "doc_count" : 3
  },
  {
    "key" : "1961",
    "doc_count" : 4
  },
  {
    "key" : "1958",
    "doc_count" : 3
  },
  {
    "key" : "1957",
    "doc_count" : 1
  },
  {
    "key" : "1955",
    "doc_count" : 2
  },
  {
    "key" : "1954",
    "doc_count" : 1
  },
  {
    "key" : "1953",
    "doc_count" : 1
  },
  },
```

```
{
  "key" : "1951",
  "doc_count" : 1
},
{
  "key" : "1950",
  "doc_count" : 1
},
{
  "key" : "1949",
  "doc_count" : 1
},
{
  "key" : "1948",
  "doc_count" : 1
},
{
  "key" : "1947",
  "doc_count" : 5
},
{
  "key" : "1944",
  "doc_count" : 2
},
{
  "key" : "1942",
  "doc_count" : 2
},
{
  "key" : "1936",
  "doc_count" : 4
},
{
  "key" : "1935",
  "doc_count" : 2
},
{
  "key" : "1934",
  "doc_count" : 2
},
{
  "key" : "1933",
  "doc_count" : 4
},
{
  "key" : "1932",
  "doc_count" : 2
},
{
  "key" : "1931",
  "doc_count" : 4
},
{
  "key" : "1930",
```

```

      "doc_count" : 2
    },
    {
      "key" : "1929",
      "doc_count" : 2
    },
    {
      "key" : "1928",
      "doc_count" : 3
    }
  ]
}
}
}
}

```

Evaluation

To perform the evaluation as we cannot check the evaluation of the total set of the films we are going to check the precision for the 10 first results.

$$Precision = \frac{\text{Films returned that are relevant}}{\text{Total films}} = \frac{10}{10} = 1 = 100\% \text{ of precision}$$

Query 4. Films about corrupt politicians in Europe and the United States.

To generate this query we used a script that used a list of European and US states and cities. You can verify this script on the delivery files named as 'create_very_long_query.py'.

Query:

```

GET film_db/_search
{
  "query": {
    "bool": {
      "must": [
        {
          "match": { "keywords": "corrupt politician" }
        }
      ],
      "should": [
        { "match": { "description": "USA" } },
        { "match": { "description": "United States" } },
        { "match": { "description": "EU" } },
        { "match": { "description": "Europe" } },
        { "match": { "description": "Albania" } },
        { "match": { "description": "Andorra" } },
        { "match": { "description": "Austria" } },
        { "match": { "description": "Belarus" } },
        { "match": { "description": "Belgium" } },
        { "match": { "description": "Bosnia and Herzegovina" } },
        { "match": { "description": "Bulgaria" } }
      ]
    }
  }
}

```

```
, { "match": { "description": "Croatia" }}
, { "match": { "description": "Czech Republic" }}
, { "match": { "description": "Denmark" }}
, { "match": { "description": "Estonia" }}
, { "match": { "description": "Faroe Islands" }}
, { "match": { "description": "Finland" }}
, { "match": { "description": "France" }}
, { "match": { "description": "Germany" }}
, { "match": { "description": "Gibraltar" }}
, { "match": { "description": "Greece" }}
, { "match": { "description": "Guernsey" }}
, { "match": { "description": "Hungary" }}
, { "match": { "description": "Iceland" }}
, { "match": { "description": "Ireland" }}
, { "match": { "description": "Isle of Man (British Crown dependency)" }}
, { "match": { "description": "Italy" }}
, { "match": { "description": "Jersey" }}
, { "match": { "description": "Kosovo" }}
, { "match": { "description": "Latvia" }}
, { "match": { "description": "Liechtenstein" }}
, { "match": { "description": "Lithuania" }}
, { "match": { "description": "Luxembourg" }}
, { "match": { "description": "North Macedonia" }}
, { "match": { "description": "Malta" }}
, { "match": { "description": "Moldova" }}
, { "match": { "description": "Monaco" }}
, { "match": { "description": "Montenegro" }}
, { "match": { "description": "Netherlands" }}
, { "match": { "description": "Norway" }}
, { "match": { "description": "Poland" }}
, { "match": { "description": "Portugal" }}
, { "match": { "description": "Romania" }}
, { "match": { "description": "Russia" }}
, { "match": { "description": "San Marino" }}
, { "match": { "description": "Serbia" }}
, { "match": { "description": "Slovakia" }}
, { "match": { "description": "Slovenia" }}
, { "match": { "description": "Spain" }}
, { "match": { "description": "Svalbard" }}
, { "match": { "description": "Sweden" }}
, { "match": { "description": "Switzerland" }}
, { "match": { "description": "Ukraine" }}
, { "match": { "description": "United Kingdom" }}
, { "match": { "description": "Vatican City" }}
, { "match": { "description": "Amsterdam, Noord-Holland" }}
, { "match": { "description": "Andorra la Vella" }}
, { "match": { "description": "Athens, Attiki" }}
, { "match": { "description": "Belgrade, Palilula" }}
, { "match": { "description": "Berlin" }}
, { "match": { "description": "Bern" }}
, { "match": { "description": "Bratislava, Bratislavsky" }}
, { "match": { "description": "Brussels" }}
, { "match": { "description": "Bucharest" }}
, { "match": { "description": "Budapest" }}
```



```

,{ "match": { "description": "Chisinau" }}
,{ "match": { "description": "Copenhagen" }}
,{ "match": { "description": "Douglas" }}
,{ "match": { "description": "Dublin" }}
,{ "match": { "description": "Gibraltar" }}
,{ "match": { "description": "Helsinki, Uusimaa" }}
,{ "match": { "description": "Kyiv, Kiev" }}
,{ "match": { "description": "Lisbon, Lisboa" }}
,{ "match": { "description": "Ljubljana, Osrednjeslovenska" }}
,{ "match": { "description": "London, Greater London" }}
,{ "match": { "description": "Longyearbyen, Svalbard" }}
,{ "match": { "description": "Luxemburg, Luxembourg" }}
,{ "match": { "description": "Madrid" }}
,{ "match": { "description": "Minsk" }}
,{ "match": { "description": "Monaco" }}
,{ "match": { "description": "Nicosia" }}
,{ "match": { "description": "Oslo" }}
,{ "match": { "description": "Paris, Ile-de-France" }}
,{ "match": { "description": "Podgorica" }}
,{ "match": { "description": "Prague, Hlavní město Praha" }}
,{ "match": { "description": "Pristina" }}
,{ "match": { "description": "Reykjavik, Hofudhborgarsvaedhi" }}
,{ "match": { "description": "Riga" }}
,{ "match": { "description": "Rome, Lazio" }}
,{ "match": { "description": "Saint Helier" }}
,{ "match": { "description": "Saint Peter Port" }}
,{ "match": { "description": "San Marino" }}
,{ "match": { "description": "Sarajevo, Federacija Bosna i Hercegovina" }}
,{ "match": { "description": "Skopje" }}
,{ "match": { "description": "Sofia, Sofia grad" }}
,{ "match": { "description": "Stockholm" }}
,{ "match": { "description": "Tallinn, Harju" }}
,{ "match": { "description": "Tirana" }}
,{ "match": { "description": "Torshavn, Strømø" }}
,{ "match": { "description": "Vaduz" }}
,{ "match": { "description": "Valletta" }}
,{ "match": { "description": "Vatican City" }}
,{ "match": { "description": "Vienna" }}
,{ "match": { "description": "Vilnius, Vilniaus" }}
,{ "match": { "description": "Warsaw, Mazowieckie" }}
,{ "match": { "description": "Zagreb, Grad Zagreb" }}
,{ "match": { "description": "Montgomery, Alabama" }}
,{ "match": { "description": "Juneau, Alaska" }}
,{ "match": { "description": "Phoenix, Arizona" }}
,{ "match": { "description": "Little Rock, Arkansas" }}
,{ "match": { "description": "Sacramento, California" }}
,{ "match": { "description": "Denver, Colorado" }}
,{ "match": { "description": "Hartford, Connecticut" }}
,{ "match": { "description": "Dover, Delaware" }}
,{ "match": { "description": "Tallahassee, Florida" }}
,{ "match": { "description": "Atlanta, Georgia" }}
,{ "match": { "description": "Honolulu, Hawaii" }}
,{ "match": { "description": "Boise, Idaho" }}
,{ "match": { "description": "Springfield, Illinois" }}

```

```
, { "match": { "description": "Indianapolis, Indiana" }}
, { "match": { "description": "Des Moines, Iowa" }}
, { "match": { "description": "Topeka, Kansas" }}
, { "match": { "description": "Frankfort, Kentucky" }}
, { "match": { "description": "Baton Rouge, Louisiana" }}
, { "match": { "description": "Augusta, Maine" }}
, { "match": { "description": "Annapolis, Maryland" }}
, { "match": { "description": "Boston, Massachusetts" }}
, { "match": { "description": "Lansing, Michigan" }}
, { "match": { "description": "Saint Paul, Minnesota" }}
, { "match": { "description": "Jackson, Mississippi" }}
, { "match": { "description": "Jefferson City, Missouri" }}
, { "match": { "description": "Helena, Montana" }}
, { "match": { "description": "Lincoln, Nebraska" }}
, { "match": { "description": "Carson City, Nevada" }}
, { "match": { "description": "Concord, New Hampshire" }}
, { "match": { "description": "Trenton, New Jersey" }}
, { "match": { "description": "Santa Fe, New Mexico" }}
, { "match": { "description": "Albany, New York" }}
, { "match": { "description": "Raleigh, North Carolina" }}
, { "match": { "description": "Bismarck, North Dakota" }}
, { "match": { "description": "Columbus, Ohio" }}
, { "match": { "description": "Oklahoma City, Oklahoma" }}
, { "match": { "description": "Salem, Oregon" }}
, { "match": { "description": "Harrisburg, Pennsylvania" }}
, { "match": { "description": "Providence, Rhode Island" }}
, { "match": { "description": "Columbia, South Carolina" }}
, { "match": { "description": "Pierre, South Dakota" }}
, { "match": { "description": "Nashville, Tennessee" }}
, { "match": { "description": "Austin, Texas" }}
, { "match": { "description": "Salt Lake City, Utah" }}
, { "match": { "description": "Montpelier, Vermont" }}
, { "match": { "description": "Richmond, Virginia" }}
, { "match": { "description": "Olympia, Washington" }}
, { "match": { "description": "Charleston, West Virginia" }}
, { "match": { "description": "Madison, Wisconsin" }}
, { "match": { "description": "Cheyenne, Wyoming" }}
, { "match": { "description": "Alabama" }}
, { "match": { "description": "Alaska" }}
, { "match": { "description": "Arizona" }}
, { "match": { "description": "Arkansas" }}
, { "match": { "description": "California" }}
, { "match": { "description": "Colorado" }}
, { "match": { "description": "Connecticut" }}
, { "match": { "description": "Delaware" }}
, { "match": { "description": "District of Columbia" }}
, { "match": { "description": "Florida" }}
, { "match": { "description": "Georgia" }}
, { "match": { "description": "Hawaii" }}
, { "match": { "description": "Idaho" }}
, { "match": { "description": "Illinois" }}
, { "match": { "description": "Indiana" }}
, { "match": { "description": "Iowa" }}
, { "match": { "description": "Kansas" }}
```

```

,{ "match": { "description": "Kentucky" }}
,{ "match": { "description": "Louisiana" }}
,{ "match": { "description": "Maine" }}
,{ "match": { "description": "Maryland" }}
,{ "match": { "description": "Massachusetts" }}
,{ "match": { "description": "Michigan" }}
,{ "match": { "description": "Minnesota" }}
,{ "match": { "description": "Mississippi" }}
,{ "match": { "description": "Missouri" }}
,{ "match": { "description": "Montana" }}
,{ "match": { "description": "Nebraska" }}
,{ "match": { "description": "Nevada" }}
,{ "match": { "description": "New Hampshire" }}
,{ "match": { "description": "New Jersey" }}
,{ "match": { "description": "New Mexico" }}
,{ "match": { "description": "New York" }}
,{ "match": { "description": "North Carolina" }}
,{ "match": { "description": "North Dakota" }}
,{ "match": { "description": "Ohio" }}
,{ "match": { "description": "Oklahoma" }}
,{ "match": { "description": "Oregon" }}
,{ "match": { "description": "Pennsylvania" }}
,{ "match": { "description": "Rhode Island" }}
,{ "match": { "description": "South Carolina" }}
,{ "match": { "description": "South Dakota" }}
,{ "match": { "description": "Tennessee" }}
,{ "match": { "description": "Texas" }}
,{ "match": { "description": "Utah" }}
,{ "match": { "description": "Vermont" }}
,{ "match": { "description": "Virginia" }}
,{ "match": { "description": "Washington" }}
,{ "match": { "description": "West Virginia" }}
,{ "match": { "description": "Wisconsin" }}
,{ "match": { "description": "Wyoming" }}
  ]}},
  "size" : 10
}

```

Description:

In this query we have used a must filter to search in which we have compare the keyword field with the term “corrupt politician” in order to choose the film that are about this topic. Then we have use a should filter comparing the the description and keyword field with each of the states of the United States and also with each of the countries of Europe. We are asked for the 10 first films so we set to 10 the size of the result.

Results:

```

{
  "took" : 81,
  "timed_out" : false,
  "_shards" : {
    "total" : 1,
    "successful" : 1,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 154,
      "relation" : "eq"
    },
    "max_score" : 79.306915,
    "hits" : [
      {
        "_index" : "film_db",
        "_type" : "film",
        "_id" : "n5V-z24Bs3F6RdQUmVUQ",
        "_score" : 79.306915,
        "_source" : {
          "keywords" : "kidnapping,mexico,bodyguard,kidnapped girl,corrupt cop",
          "title" : "Man on Fire",
          "directors" : "Tony Scott",
          "release" : "2004-04-23",
          "day" : "23",
          "month" : "04",
          "year" : "2004",
          "genres" : [
            "Action",
            "Crime",
            "Drama",
            "Thriller"
          ],
          "description" : "Man on Fire is a movie starring Denzel Washington, Christopher Walken,
and Dakota Fanning. In Mexico City, a former CIA operative swears vengeance on those who committed
an unspeakable act against the family he was hired to protect.",
          "actors" : [
            "Denzel Washington",
            "Christopher Walken",
            "Dakota Fanning",
            "Radha Mitchell"
          ]
        }
      },
      {
        "_index" : "film_db",
        "_type" : "film",
        "_id" : "7pWD0G4Bs3F6RdQUbJ5B",
        "_score" : 79.306915,
        "_source" : {
          "keywords" : "kidnapping,mexico,bodyguard,kidnapped girl,corrupt cop",
          "title" : "Man on Fire",

```

```

    "directors" : "Tony Scott",
    "release" : "2004-04-23",
    "day" : "23",
    "month" : "04",
    "year" : "2004",
    "genres" : [
        "Action",
        "Crime",
        "Drama",
        "Thriller"
    ],
    "description" : "Man on Fire is a movie starring Denzel Washington, Christopher Walken,
and Dakota Fanning. In Mexico City, a former CIA operative swears vengeance on those who committed
an unspeakable act against the family he was hired to protect.",
    "actors" : [
        "Denzel Washington",
        "Christopher Walken",
        "Dakota Fanning",
        "Radha Mitchell"
    ]
}
},
{
    "_index" : "film_db",
    "_type" : "film",
    "_id" : "y5WIZm4Bs3F6RdQUYyCc",
    "_score" : 66.25601,
    "_source" : {
        "keywords" : "revenge,corrupt politician,bloody body of child,mafia,1950s",
        "title" : "The Godfather: Part II",
        "directors" : "Francis Ford Coppola",
        "release" : "1974-12-20",
        "day" : "20",
        "month" : "12",
        "year" : "1974",
        "genres" : [
            "Crime",
            "Drama"
        ],
        "description" : "The Godfather: Part II is a movie starring Al Pacino, Robert De Niro, and
Robert Duvall. The early life and career of Vito Corleone in 1920s New York City is portrayed, while
his son, Michael, expands and tightens his grip on the...",
        "actors" : [
            "Al Pacino",
            "Robert De Niro",
            "Robert Duvall",
            "Diane Keaton"
        ]
    }
},
{
    "_index" : "film_db",
    "_type" : "film",
    "_id" : "AJVr0G4Bs3F6RdQubYvV",

```

```

    "_score" : 66.25601,
    "_source" : {
      "keywords" : "revenge,corrupt politician,bloody body of child,mafia,1950s",
      "title" : "The Godfather: Part II",
      "directors" : "Francis Ford Coppola",
      "release" : "1974-12-20",
      "day" : "20",
      "month" : "12",
      "year" : "1974",
      "genres" : [
        "Crime",
        "Drama"
      ],
      "description" : "The Godfather: Part II is a movie starring Al Pacino, Robert De Niro, and
Robert Duvall. The early life and career of Vito Corleone in 1920s New York City is portrayed, while
his son, Michael, expands and tightens his grip on the...",
      "actors" : [
        "Al Pacino",
        "Robert De Niro",
        "Robert Duvall",
        "Diane Keaton"
      ]
    }
  },
  {
    "_index" : "film_db",
    "_type" : "film",
    "_id" : "WJW1zm4Bs3F6RdQUMiqK",
    "_score" : 63.08768,
    "_source" : {
      "keywords" : "bare chested male bondage,gay subtext,poliziottesco,corrupt cop,new york",
      "title" : "Copkiller (l'assassino dei poliziotti)",
      "directors" : "Roberto Faenza",
      "release" : "1983-03-15",
      "day" : "15",
      "month" : "03",
      "year" : "1983",
      "genres" : [
        "Crime",
        "Thriller"
      ],
      "description" : "Copkiller (l'assassino dei poliziotti) is a movie starring Harvey Keitel,
John Lydon, and Nicole Garcia. A pair of corrupt cops spend their illegal cash on an uptown New York
City apartment.",
      "actors" : [
        "Harvey Keitel",
        "John Lydon",
        "Nicole Garcia",
        "Leonard Mann"
      ]
    }
  },
  {
    "_index" : "film_db",

```

```

    "_type" : "film",
    "_id" : "k5XVz24Bs3F6RdQU4WgP",
    "_score" : 49.581413,
    "_source" : {
      "keywords" : "cigar smoking, revenge, corrupt cop, drugs, mexico",
      "title" : "The Shepherd",
      "directors" : "Isaac Florentine",
      "release" : "2008-03-04",
      "day" : "04",
      "month" : "03",
      "year" : "2008",
      "genres" : [
        "Action",
        "Crime",
        "Thriller"
      ],
      "description" : "The Shepherd is a video starring Jean-Claude Van Damme, Stephen Lord, and
Natalie J. Robb. A former New Orleans cop battles drug cartels and ex-Army Special Forces who are
smuggling drugs from Mexico into U.S. through a Texas border...",
      "actors" : [
        "Jean-Claude Van Damme",
        "Stephen Lord",
        "Natalie J. Robb",
        "Gary McDonald"
      ]
    }
  },
  {
    "_index" : "film_db",
    "_type" : "film",
    "_id" : "lJXVz24Bs3F6RdQU5Gih",
    "_score" : 49.581413,
    "_source" : {
      "keywords" : "cigar smoking, revenge, corrupt cop, drugs, mexico",
      "title" : "The Shepherd",
      "directors" : "Isaac Florentine",
      "release" : "2008-03-04",
      "day" : "04",
      "month" : "03",
      "year" : "2008",
      "genres" : [
        "Action",
        "Crime",
        "Thriller"
      ],
      "description" : "The Shepherd is a video starring Jean-Claude Van Damme, Stephen Lord, and
Natalie J. Robb. A former New Orleans cop battles drug cartels and ex-Army Special Forces who are
smuggling drugs from Mexico into U.S. through a Texas border...",
      "actors" : [
        "Jean-Claude Van Damme",
        "Stephen Lord",
        "Natalie J. Robb",
        "Gary McDonald"
      ]
    }
  }
]

```

```

    }
  },
  {
    "_index" : "film_db",
    "_type" : "film",
    "_id" : "iJvYz24Bs3F6RdQU301-",
    "_score" : 43.34479,
    "_source" : {
      "keywords" : "vietnam war veteran,heroin,drugs,drug dealer,corrupt cop",
      "title" : "Exit Wounds",
      "directors" : "Andrzej Bartkowiak",
      "release" : "2001-03-16",
      "day" : "16",
      "month" : "03",
      "year" : "2001",
      "genres" : [
        "Action",
        "Crime",
        "Thriller"
      ],
      "description" : "Exit Wounds is a movie starring Steven Seagal, DMX, and Isaiah Washington. Orin Boyd, a tough cop in an inner-city precinct discovers a web of dirty cops and corruption.",
      "actors" : [
        "Steven Seagal",
        "DMX",
        "Isaiah Washington",
        "Anthony Anderson"
      ]
    }
  },
  {
    "_index" : "film_db",
    "_type" : "film",
    "_id" : "XJW-zm4Bs3F6RdQUkiwe",
    "_score" : 42.760063,
    "_source" : {
      "keywords" : "gay,assassination,gay rights,politician,gay interest",
      "title" : "The Times of Harvey Milk",
      "directors" : "Rob Epstein",
      "release" : "1985-09-20",
      "day" : "20",
      "month" : "09",
      "year" : "1985",
      "genres" : [
        "Documentary",
        "Biography",
        "History"
      ],
      "description" : "The Times of Harvey Milk is a movie starring Harvey Fierstein, Harvey Milk, and Anne Kronenberg. A documentary of the successful career and assassination of San Francisco's first elected gay city supervisor.",
      "actors" : [
        "Harvey Fierstein",

```



```

        "Harvey Milk",
        "Anne Kronenberg",
        "Tory Hartmann"
    ]
}
},
{
    "_index" : "film_db",
    "_type" : "film",
    "_id" : "4pVRz24Bs3F6RdQUL0vf",
    "_score" : 38.16431,
    "_source" : {
        "keywords" : "mini dress,blonde woman,short dress,corrupt cop,pig",
        "title" : "'R Xmas",
        "directors" : "Abel Ferrara",
        "release" : "2001-12-14",
        "day" : "14",
        "month" : "12",
        "year" : "2001",
        "genres" : [
            "Crime",
            "Drama"
        ],
        "description" : "'R Xmas is a movie starring Drea de Matteo, Lillo Brancato, and Lisa Valens. A New York drug dealer is kidnapped, and his wife must try to come up with the money and drugs to free him from his abductors before Christmas.",
        "actors" : [
            "Drea de Matteo",
            "Lillo Brancato",
            "Lisa Valens",
            "Ice-T"
        ]
    }
}
]
}
}

```

Evaluation:

To perform the evaluation as we cannot check the evaluation of the total set of the films we are going to check the precision for the 10 first results.

$$Precision = \frac{\text{Films returned that are relevant}}{\text{Total films}} = \frac{8}{10} = 0.8 = 80\% \text{ of precision}$$