



Movies Dataset

Julián Ferreira
Tiago Antunes



To be remembered... Prospective Search Tasks

- Search for a movie by its Title
- Search for a movie by its Gender
- Search for a movie by its Origin/Ethnicity
- Search for a movie by its Release Year (with ranges)
- Search for a movie by its Director
- Search for a movie by its Cast (Actors)
- Search for a movie by keywords in the Plot

Information Retrieval Tool



A screenshot of the Solr Admin UI. The left sidebar shows navigation links: Dashboard, Logging, Cloud, Collections, Java Properties, Thread Dump, and a dropdown menu for "films" containing Overview, Analysis, Dataimport, Documents, Files, Query (selected), and Schema. Below this is a "Core Selector" dropdown. The main panel is titled "Request-Handler (qt) /select" and contains various configuration fields: "common" (empty), "q" (set to "genre:Fantasy"), "fq" (empty), "sort" (empty), "start, rows" (0 to 10), "fl" (empty), "df" (empty), "Raw Query Parameters" (key1=val1&key2=val2), "wt" (set to "json"), and checkboxes for "indent" (checked), "debugQuery", "dismax", and "edismax". On the right, the JSON response is displayed for the URL "http://127.0.1.1:7574/solr/films/select?indent=on&q=genre:Fantasy&wt=json". The response includes a "responseHeader" with status, QTime, and params, and a "response" object containing "numFound": 82, "start": 0, and a list of document objects. Two documents are shown, each with fields for id, directed_by, initial_release_date, and genre.



Collection of data

- Build the docker image: 1 core only
- collection.sh script
 - Posts schema with the indexes
 - Populates the core with the content of the csv file



Indexing I

Field	Type	Solr class	MultiValued
Title	synonym_text	TextField	false
Origin	porter_text	TextField	false
Plot	plot_text	TextField	false
Release Year	number	IntPointField	false
Genre	synonym	TextField	true
Director	char_text	TextField	true
Cast	char_text	TextField	true

Table 2: Description of schema fields



Indexing II

- synonym_text
 - ASCII Folding filter
 - Lower Case Filter
 - Synonym Graph Filter
 - Porter Stem Filter
- porter_text
 - ASCII Folding filter
 - Lower Case Filter
 - Porter Stem Filter
- plot_test
 - ASCII Folding filter
 - Lower Case Filter
 - Synonym Graph Filter
 - Porter Stem Filter
 - English Possessive Filter
 - Hyphenated Words Filter
 - Stop Filter
- char_test
 - ASCII Folding filter
 - Lower Case Filter
 - Mapping Char filter



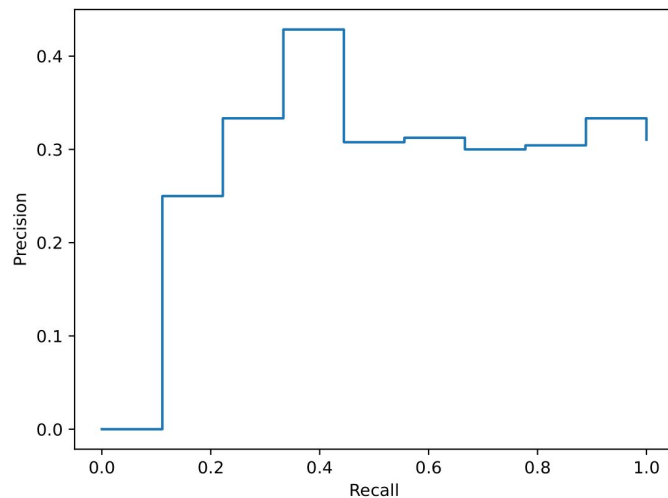
Retrieval: Query 1

Drama movies with Bollywood origin released in 2015 or after

Option	Value
q	Genre:drama AND Origin_Ethnicity:bollywood AND Released_Year:[2015 TO *]
qf (boost)	Genre^5 Origin_Ethnicity^3 Released_Year

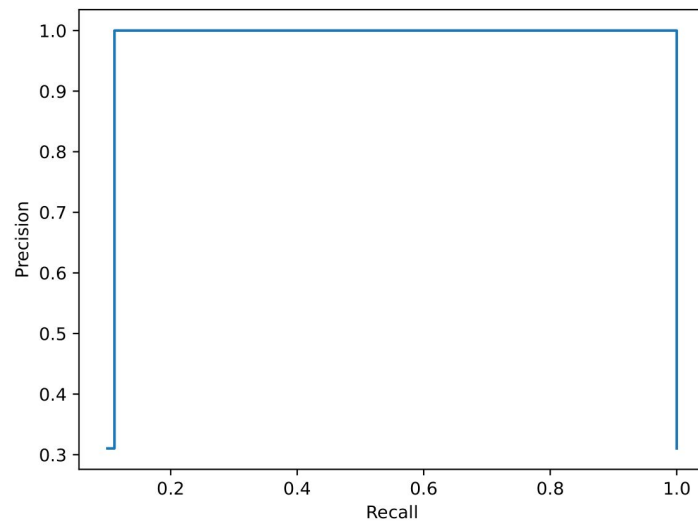
Table 3: Arguments of query 1

Evaluation: Query 1



Metric	Value	Value with boost
AvP	0.34	0.67
P@10	0.4	0.9

Table 4: Precision metrics for query 1





Retrieval: Query 2

Movies starring Robert De Niro and are of italian mafia

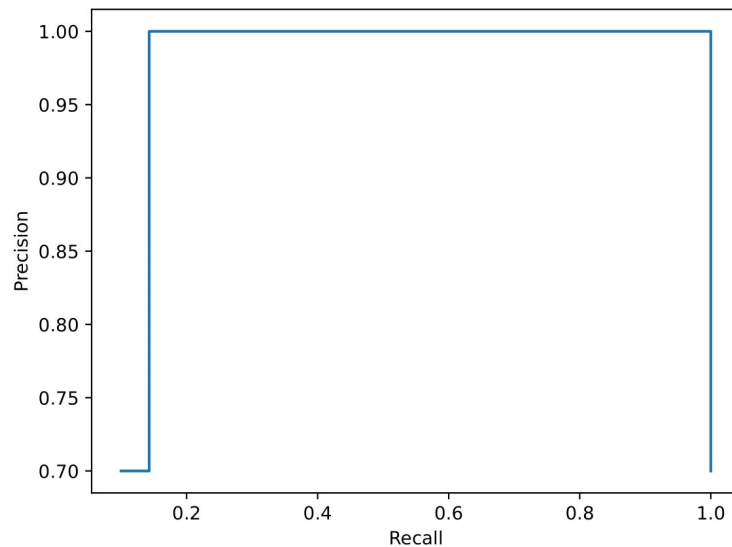
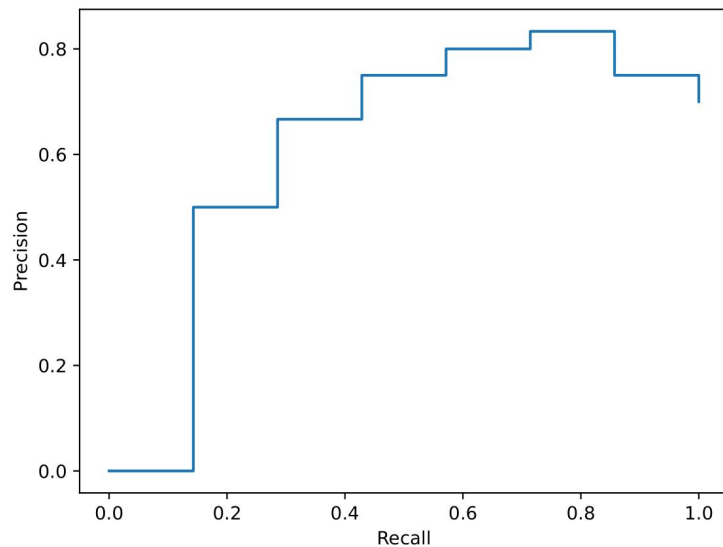
Option	Value
q	Cast:"Robert de Niro" AND (Plot:mafia OR Plot:Italian)
qf (boost)	Cast^5 Plot:mafia^3 Plot:italian

Table 5: Arguments of query 2

Evaluation: Query 2

Metric	Value	Value with boost
AvP	0.66	0.96
P@10	0.7	0.7

Table 6: Precision metrics for query 2





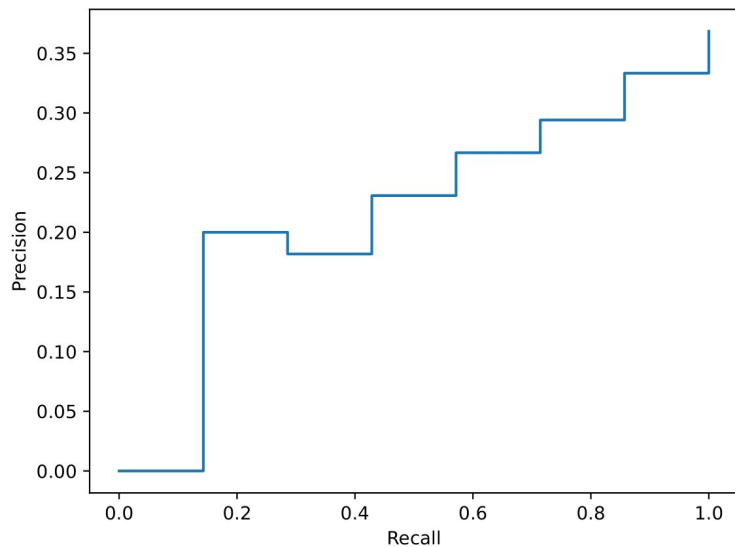
Retrieval: Query 3

Movies directed by Steven Spielberg about murder

Option	Value
q	Director:"Steven Spielberg" AND (Plot:murder OR Plot:kill)
qf (boost)	Director^5 Plot:murder^3 Plot:kill

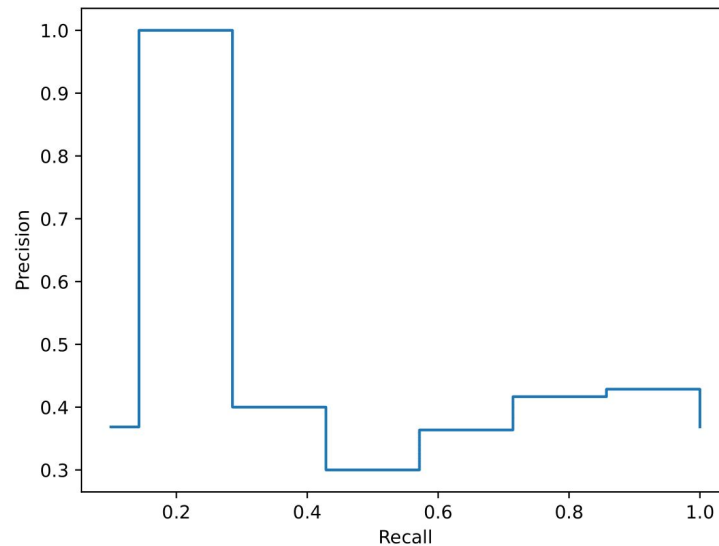
Table 7: Arguments of query 3

Evaluation: Query 3



Metric	Value	Value with boost
AvP	0.26	0.49
P@10	0.2	0.3

Table 8: Precision metrics for query 3





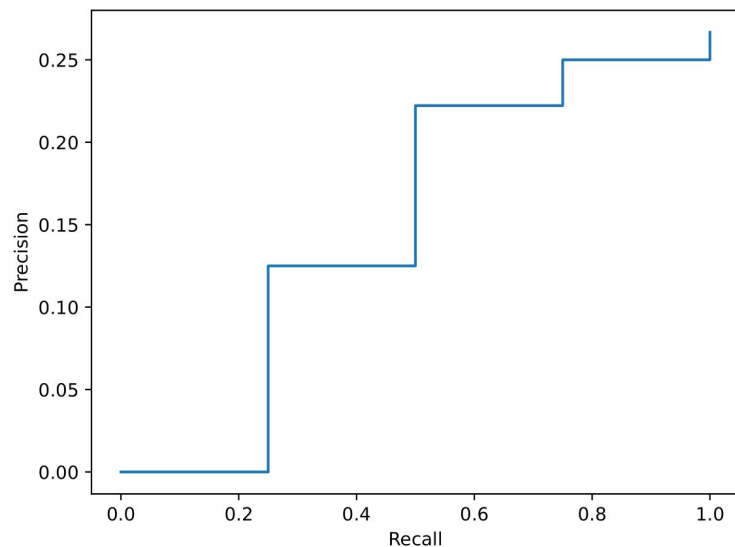
Retrieval: Query 4

Police action movies from the year 2000

Option	Value
q	Plot:police AND Genre:action AND Release_Year:2000
qf (boost)	Plot:police ⁵ Genre:action ³ Release _{year}

Table 9: Arguments of query 4

Evaluation: Query 2



Metric	Value	Value with boost
AvP	0.18	0.6
P@10	0.3	0.4

Table 10: Precision metrics for query 4

