

# KNOWLEDGE GRAPH FOR NETFLIX DATASET

by Aman Roy

Step 1 :-

Analyzing the Netflix.csv data and creating the ontology.

	A	B	C	D	E	F
1	show_id	type	title	director	cast	country
2	81145628	Movie	Norm of the North: King Sized Adventure	"Richard Finn	Tim Maltby"	"Alan Marriott
3	80117401	Movie	Jandino: Whatever it Takes		Jandino Asporaat	United Kingdom
4	70234439	TV Show	Transformers Prime		"Peter Cullen	Sumalee Montano
5	80058654	TV Show	Transformers: Robots in Disguise		"Will Friedle	Darren Criss
6	80125979	Movie	#realityhigh	Fernando Lebrija	"Nesta Cooper	Kate Walsh
7	80163890	TV Show	Apaches		"Alberto Ammann	Eloy Azorín
8	70304989	Movie	Automata	Gabe Ibáñez	"Antonio Banderas	Dylan McDermott
9	80164077	Movie	Fabrizio Copano: Solo pienso en mi	"Rodrigo Toro	Francisco Schultz"	Fabrizio Copano
10	80117902	TV Show	Fire Chasers			United States
11	70304990	Movie	Good People	Henrik Ruben Genz	"James Franco	Kate Hudson
12	80169755	Movie	Joaquin Reyes: Una y no más	José Miguel Contreras	Joaquin Reyes	
13	70299204	Movie	Kidnapping Mr. Heineken	Daniel Alfredson	"Jim Sturgess	Sam Worthington
14	80182480	Movie	Krish Trish and Baltiboy		"Damandeep Singh Baggan	Smita Malhotra
15	80182483	Movie	Krish Trish and Baltiboy: Battle of Wits	"Munjal Shroff	Tilak Shetty"	"Damandeep Singh Baggan
16	80182596	Movie	Krish Trish and Baltiboy: Best Friends Forever	"Munjal Shroff	Tilak Shetty"	"Damandeep Singh Baggan
17	80182482	Movie	Krish Trish and Baltiboy: Comics of India	Tilak Shetty	"Damandeep Singh Baggan	Smita Malhotra
18	80182597	Movie	Krish Trish and Baltiboy: Oversmartness Never Pays	Tilak Shetty	"Rishi Gambhir	Smita Malhotra
19	80182481	Movie	Krish Trish and Baltiboy: Part II		"Damandeep Singh Baggan	Smita Malhotra
20	80182621	Movie	Krish Trish and Baltiboy: The Greatest Trick	"Munjal Shroff	Tilak Shetty"	"Damandeep Singh Baggan
21	80057969	Movie	Love	Gaspar Noé	"Karl Glusman	Klara Kristin
22	80060297	Movie	Manhattan Romance	Tom O'Brien	"Tom O'Brien	Katherine Waterston
23	80046728	Movie	Moonwalkers	Antoine Bardou-Jacquet	"Ron Perlman	Rupert Grint
24	80046727	Movie	Rolling Papers	Mitch Dickman		"United States
25	70304988	Movie	Stonehearst Asylum	Brad Anderson	"Kate Beckinsale	Jim Sturgess
26	80057700	Movie	The Runner	Austin Stark	"Nicolas Cage	Sarah Paulson
27	80045922	Movie	6 Years	Hannah Fidell	"Taissa Farmiga	Ben Rosenfield
28	80244601	TV Show	Castle of Stars		"Chaiyapol Pupart	Jintanutda Lummakanon
29	80203094	Movie	City of Joy	Madeleine Gavin		"United States
30	80190843	TV Show	First and Last			
31	70241607	Movie	Laddaland	Sopon Sukdapisit	"Saharat Sangkapreecha	Pok Piyatida Woramusik
32	80988892	Movie	Next Gen	"Kevin R. Adams	Joe Ksander"	"John Krasinski
33	80239639	Movie	Sierra Burgess Is A Loser	Ian Samuels	"Shannon Purser	Kristine Froseth
34	80159586	Movie	The Most Assassinated Woman in the World	Franck Ribière	"Anna Mougialis	Niels Schneider
35	80152447	Movie	Cézanne et moi	Danièle Thompson	"Guillaume Canet	Guillaume Gallienne
36	80221550	TV Show	Archibald's Next Big Thing		"Tony Hale	Rosamund Pike
37	81154455	Movie	Article 15	Anubhav Sinha	"Ayushmann Khurrana	Nassar
38	81113928	Movie	Care of Kancharapalem	Maha Venkatesh	"Subba Rao Vepada	Radha Bessy
39	81052275	Movie	Ee Nagarainiki Emaindi	Tharun Bhascker	"Vishwaksen Naidu	Sushanth Reddy
40	81132437	Movie	Kill Me If You Dare	Şenol Şönmez	"Murat Boz	Seda Bakan
41	80178151	TV Show	The Spy		"Sacha Baron Cohen	Noah Emmerich
42	80058026	Movie	Heli and Back	"Tom Ganas	Ross R. Shuman"	"Nick Swardson
43	70303496	Movie	PK	Rajkumar Hirani	"Aamir Khan	Anuska Sharma
44	80162141	Movie	Hard Tide	"Robert Osman	Nathanael Wiseman"	"Nathanael Wiseman
45	80095641	Movie	Elstree 1976	Jon Spira	"Paul Blake	Jeremy Bulloch
46	81176188	Movie	American Factory: A Conversation with the Obamas		"President Barack Obama	Michelle Obama
47	80159880	Movie	ATM	Mez Tharatorn	"Chantavit Dhanasevi	Preechaya Pongthanankorn
48	81016044	Movie	Bangkok Traffic (Love) Story	Adisorn Tresirikasem	"Theeradej Wongpuapan	Sirin Horwang

Ontology Design

- Id
  - movieId
  - showId
- Movie
- Show
- NetflixContent
- Title
- Director
- Cast
- Country

- Date
- releaseYear
- Rating
- Duration
  - timeDuration
  - showDuration
- genre
- Description

**Note :-** All in **red** leads to Data Property while in black are Concepts

**Classes :-**

	<b>Rigid</b>	<b>Unity</b>	<b>Identity</b>
<b>Movie</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>
<b>Show</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>
<b>NetflixContent</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>
<b>Director</b>	<b>No</b>	<b>Yes</b>	<b>Yes</b>
<b>Cast</b>	<b>No</b>	<b>Yes</b>	<b>Yes</b>
<b>Country</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>
<b>Rating</b>	<b>Yes</b>	<b>Yes</b>	<b>No</b>
<b>Genre</b>	<b>Yes</b>	<b>Yes</b>	<b>No</b>

**Properties :-**

Property	Characteristics(Inverse Relation)	Domain	Range	Rigidity
<b>hasId</b>	-	NetflixContent	String	
<b>hasMovieId</b>	-	Movie	String	
<b>hasShowId</b>	-	Show	String	
<b>hasDirectorRole</b>	isDirectorRoleOf	NetflixContent	DirectorRole	

hasCastRole	isCastRoleOf	NetflixContent	CastRole	
isReleasedInCountry	isReleasedCountryOf	NetflixContent	Country	
hasDuration	-	NetflixContent	String/Int	
hasTimeDuration	-	Movie	Int	
hasShowDuration	-	Show	String	
hasRated	isRatingOf	NetflixContent	Rating	
hasDate	-	NetflixContent	DateTime	
hasGenre	isGenreOf	NetflixContent	genre	
hasDescription	-	NetflixContent	String	
hasCasted	isCastedBy	Cast	CastRole	
hasDirected	isDirectedBy	Director	DirectorRole	
hasName	-	Cast/Director	String	

Note :- All in red are data property while in black are object property.

TBoxes Statements :-

1. A director is someone who hasDirected some DirectorRole.
2. A cast is someone who hasCasted some DirectorRole.
3. Every movie has one MovieId.
4. Every show has one ShowId.
5. A NetflixContent is something that has atleast one DirectorRole and atleast one CastRole
6. Movie and Show are disjoint classes.
7. Every NetflixContent has atleast one genre.
8. DirectorRole and CastRole are disjoint.
9. Every NetflixContent has some country where it is released.
10. Every NetflixContent has only one rating.

ODPs Used

Two type of ODPs are used namely :-

1. Agent-Role ODP

- a. Agents are director and cast which performs role of director role and cast role respectively.
2. Event-ODP
  - a. Agent Role ODP is combined with Event ODP where Event is movie and show which has both location and duration information.

#### Hierarchy Explanation :-

1. Every netflix content will always be a netflix content and is either a movie or a show. Also all movies and shows will always sustain to be movies and shows. This is the reason why movies and shows are subclasses of netflix content.
2. Every movie or show has data properties which are listed in terms of various data properties as show in red in the table.
3. Apart from that every other class has no subclasses and also provides information regarding movies and shows.
4. The ontology has primarily designed in order to answer the below competency questions which it fulfills :-

#### Competence Questions :-

1. List all the DirectorRole of a Director ?
2. List all the movies directed by a director ?
3. List all the movies or shows in which a cast was casted?
4. List all the movies or shows which was directed by a director ?
5. List all the movies released in a country ?

Also below is the validity of ontology after running through Hermit Reasoner.

```
INFO 19:08:38 ----- Running Reasoner -----
INFO 19:08:38 Pre-computing inferences:
INFO 19:08:38   - class hierarchy
INFO 19:08:38   - object property hierarchy
INFO 19:08:38   - data property hierarchy
INFO 19:08:38   - class assertions
INFO 19:08:38   - object property assertions
INFO 19:08:38   - same individuals
INFO 19:08:38 Ontologies processed in 54 ms by Hermit
INFO 19:08:38
```

#### The Explanation Diagram



"http://www.semanticweb.org/amanroy/ontologies/2020/3/Movie#India" ,  
"https://www.geonames.org/countries/IN/india.html"

"http://www.semanticweb.org/amanroy/ontologies/2020/3/Movie#Indonesia" ,  
"https://www.geonames.org/countries/ID/indonesia.html"

"http://www.semanticweb.org/amanroy/ontologies/2020/3/Movie#Ireland" ,  
"https://www.geonames.org/countries/IE/ireland.html"

"http://www.semanticweb.org/amanroy/ontologies/2020/3/Movie#Pakistan" ,  
"https://www.geonames.org/countries/PK/pakistan.html"

"http://www.semanticweb.org/amanroy/ontologies/2020/3/Movie#United\_Kingdom" ,  
"https://www.geonames.org/countries/GB/united-kingdom.html"

"http://www.semanticweb.org/amanroy/ontologies/2020/3/Movie#United\_States" ,  
"https://www.geonames.org/countries/US/united-states.html".

Linking can be done using protege by below methods :-

1. **owl:sameAs (adding extra dbpedia or geonames IRI for every individual)**
2. **By changing the IRI of individuals to dbpedia or geonames IRI.**

**The above KG is a 5 star linked dataset as it has :-**




1. **Sparql endpoints.**
2. **Is in RDF format.**
3. **Linked with at least two external datasets i.e. dbpedia and geonames.**

b) will generate the ontology file named ontology.ttl.

#### **Step 4 :- Setup Of Fuseki Server for SparQL query**

**Run Fuseki bash script and upload knowledge graphs to run SparQL queries.**  
**The named graphs and the fuseki server running on hostname.**

**Here is the screenshot of the same**

QUERY RESULTS	
	Table
	Raw Response
	
Showing 1 to 11 of 11 entries	
Search:	<input type="text"/>
Show	50 entries
a	b
1	<http://dbpedia.org/resource/Aamir_Khan>
2	base:Aamir_Khan
3	<http://dbpedia.org/resource/Adam_Conover>
4	base:Adam_Conover
5	<http://dbpedia.org/resource/Anushka_Sharma>
6	base:Anushka_Sharma
7	<http://dbpedia.org/resource/Ayushmann_Khurrana>
8	base:Ayushmann_Khurrana
9	base:Denmark
10	<https://www.geonames.org/countries/DK/denmark.html>
11	base:India
12	<https://www.geonames.org/countries/IN/india.html>
13	base:Indonesia
14	<https://www.geonames.org/countries/ID/indonesia.html>
15	base:Ireland
16	<https://www.geonames.org/countries/IE/ireland.html>
17	base:Pakistan
18	<https://www.geonames.org/countries/PK/pakistan.html>
19	base:United_Kingdom
20	<https://www.geonames.org/countries/GB/united-kingdom.html>
21	base:United_States
22	<https://www.geonames.org/countries/US/united-states.html>
Showing 1 to 11 of 11 entries	

Activities Google Chrome Thu 20:49

Apache Jena Fuseki - Inspect dataset - Google Chrome

Not secure | amanroy-asuspro-p5440uf3030/dataset.html?tab=upload&ds=Netflix\_2

Apps My company Mining of Mas... README | Out... (106) Hadoop... ISRO CS previ... GATE CS Topic... ISRO CSE Sylla... ISRO CS: Eligib... ISRO Recruitm... Math 561 - The...

Apache Jena Fuseki dataset manage datasets help Server status

Dataset: /Netflix

query upload files edit info

SPARQL query

To try out some SPARQL queries against the selected dataset, enter your query here.

EXAMPLE QUERIES

Selection of triples Selection of classes

PRESERVES

rdf rdfs owl xsd

SPARQL ENDPOINT /Netflix/query CONTENT TYPE (SELECT) JSON CONTENT TYPE (GRAPH) Turtle

```

1 prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
2 prefix owl: <http://www.w3.org/2002/07/owl#>
3 prefix base: <http://www.semanticweb.org/amanroy/ontologies/2020/3/Movie#>
4 SELECT ?Movie ?ab
5 FROM <http://iiitd.ac.in/web/2016011/oldmoviesgraph>
6 FROM <http://iiitd.ac.in/web/2016011/newmoviesgraph>
7 FROM <http://iiitd.ac.in/web/2016011/remainingmoviesgraph>
8 WHERE {
9   ?Movie base:hasDirected ?DirectorRole.
10  ?Director base:hasDirected ?DirectorRole.
11  Filter regex(str(?Director), "(?!)(?:shetty)$")
12  ?ab base:hasDirected ?DirectorRole
13 }
14

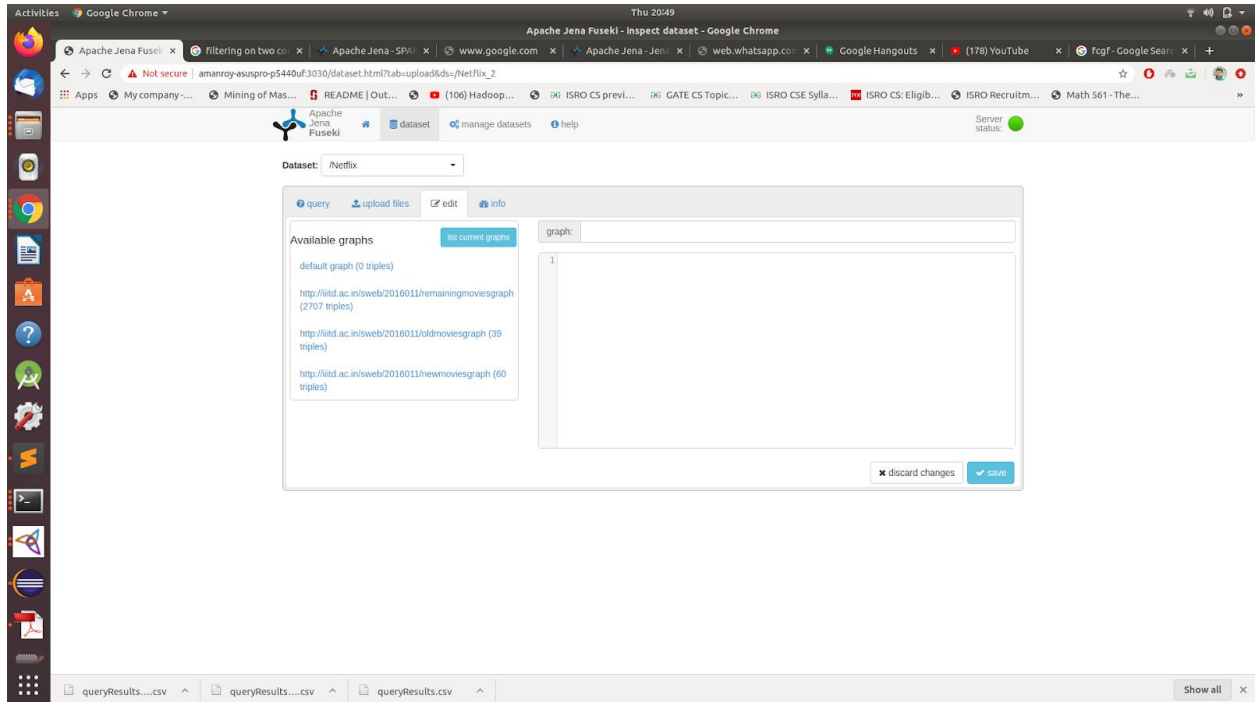
```

QUERY RESULTS

Table Raw Response

queryResults....csv queryResults....csv queryResults.csv Show all

## Step 5 :- Splitting Graph into Named Graphs



I have created three named graphs i.e.

- i) [<http://iiitd.ac.in/sweb/2016011/oldmoviesgraph>](http://iiitd.ac.in/sweb/2016011/oldmoviesgraph)
- ii) [<http://iiitd.ac.in/sweb/2016011/newmoviesgraph>](http://iiitd.ac.in/sweb/2016011/newmoviesgraph)
- iii) [<http://iiitd.ac.in/sweb/2016011/remainingmoviesgraph>](http://iiitd.ac.in/sweb/2016011/remainingmoviesgraph)

The first two graphs contain triples as defined in the question and the remaining graphs contain all the triples which are not present in the first two graphs. Also after moving triples to these named graphs, default graphs become empty at the end.

**Step 6 :- Give knowledge Graph a front end using Pubby**



Activities Google Chrome Thu 17:44 PK | Netmovie - Google Chrome

PK | Netmovie  
http://localhost:8080/pubby/resource/PK

Property	Value
? hasCastRole	• <http://localhost:8080/pubby/resource/cast_role_of_PK>
? hasDate	• September_6_2018
? hasDescription	• Aamir_Khan_teams_with_director_Rajkumar_Hirani_to_play_a_social_crusader_in_a_political_satire_on_the_state_of_corruption_in_India.
? hasDirectorRole	• <http://localhost:8080/pubby/resource/director_role_of_PK>
? hasGenre	• <http://localhost:8080/pubby/resource/Comedies> • <http://localhost:8080/pubby/resource/Dramas> • <http://localhost:8080/pubby/resource/International_Movies>
? hasMovieId	• 70303496
? hasRated	• <http://localhost:8080/pubby/resource/TV-14>
? hasReleaseYear	• 2014
? hasShowDuration	• 146_min
? isReleasedInCountry	• <http://localhost:8080/pubby/resource/India>
rdftype	• <http://localhost:8080/pubby/resource/Movie> • owl:NamedIndividual

Metadata

Anon\_0

rdftype prv:DataItem

rdftype <http://www.w3.org/2004/03/trix/tdg:1/Graph>

foaf:primaryTopic <http://localhost:8080/pubby/resource/PK>

foaf:topic Anon\_0

is:realizes <http://localhost:8080/pubby/data/PK>

prv:createdBy Anon\_1 (more)

expand all

This page shows information obtained from the SPARQL endpoint at http://localhost:3030/Netflix\_2spargl.  
As N3 | As RDF/XML | Browse in Disco | Browse in Tabulator | Browse in OpenLink Browser

IsreleasedInCountry **India** is linked to Geonames.org link

Activities Google Chrome Thu 20:25 India | Netmovie - Google Chrome

India | Netmovie  
http://localhost:8080/pubby/resource/India

Property	Value
owl:sameAs	• <https://www.geonames.org/countries/IN/India.html>
rdftype	• <http://localhost:8080/pubby/resource/Country> • owl:NamedIndividual

Metadata

Anon\_0

rdftype prv:DataItem

rdftype <http://www.w3.org/2004/03/trix/tdg:1/Graph>

foaf:primaryTopic <http://localhost:8080/pubby/resource/India>

foaf:topic Anon\_0

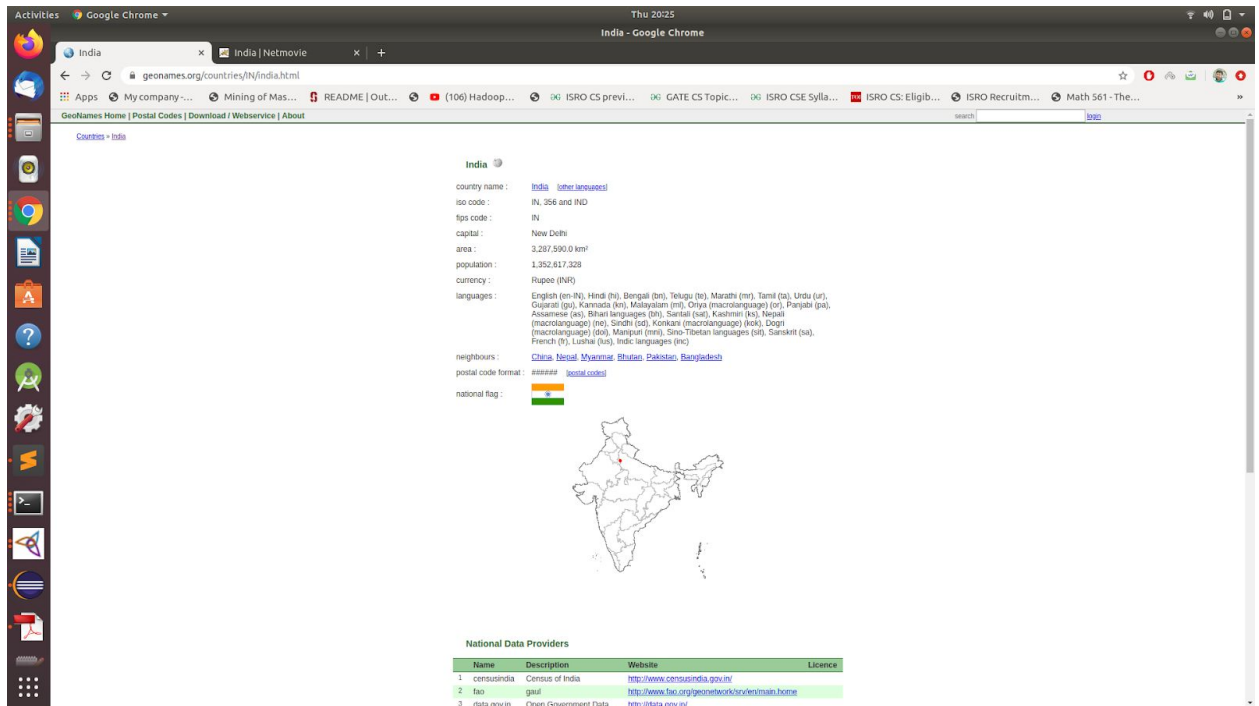
is:realizes <http://localhost:8080/pubby/data/India>

prv:createdBy Anon\_1 (more)

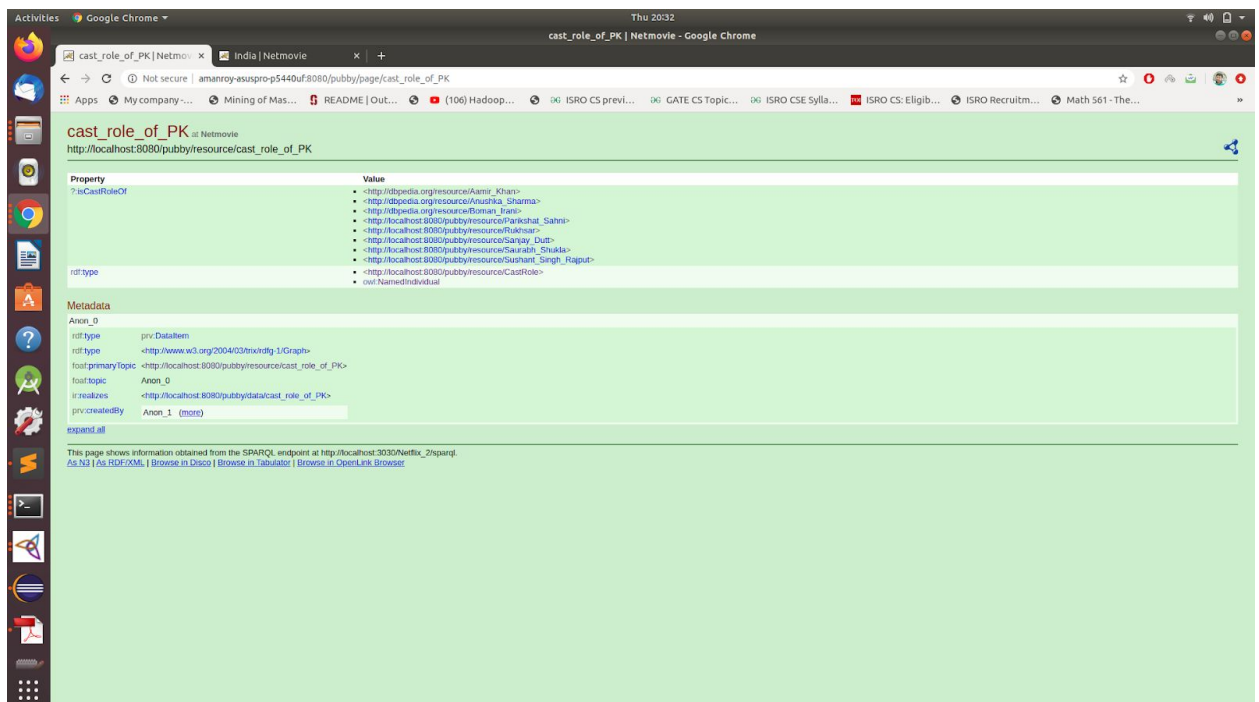
expand all

This page shows information obtained from the SPARQL endpoint at http://localhost:3030/Netflix\_2spargl.  
As N3 | As RDF/XML | Browse in Disco | Browse in Tabulator | Browse in OpenLink Browser

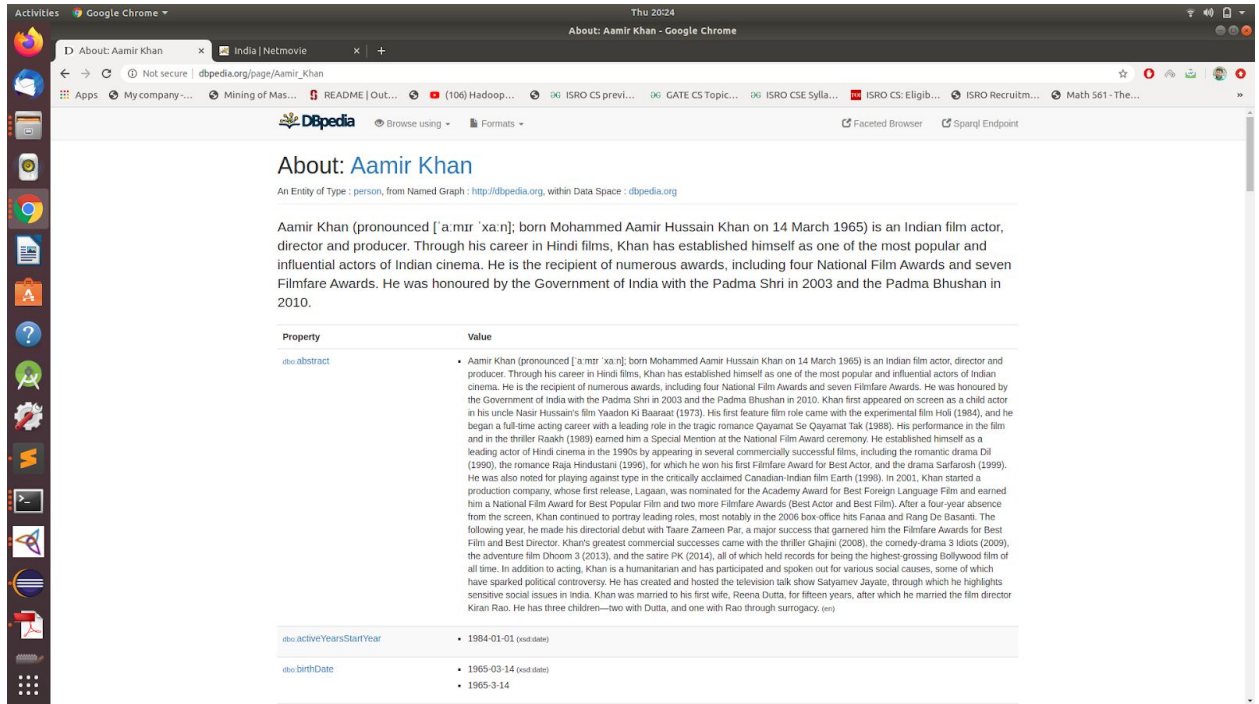
Geoname link takes to Geoname page.



hasCastRole is linked to CastRole which has dbpedia entries for different actors



Clicking on Aamir Khan link opens dbpedia page of Aamir Khan



These images show that front-end is working successfully.

## Bonus Part

Some queries to run on the above graphs on Fusceki Server

a)

Copy the query in `query.rq` file and run the command in fusceki's bin folder.

Run the command :-

```
sudo ruby ./s-query --service=http://amanroy-asuspro-p5440uf:3030/Netflix --query=query.rq >> Q2-a.txt
```

Query

```
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
```

```
prefix owl: <http://www.w3.org/2002/07/owl#>
```

```
prefix base: <http://www.semanticweb.org/amanroy/ontologies/2020/3/Movie#>
```

```
SELECT ?Movie ?ab
```

```
FROM <http://iiitd.ac.in/sweb/2016011/oldmoviesgraph>
```

```
FROM <http://iiitd.ac.in/sweb/2016011/newmoviesgraph>
```

```

FROM <http://iiitd.ac.in/sweb/2016011/remainingmoviesgraph>
WHERE {
  ?Movie base:hasDirectorRole ?DirectorRole.
  ?Director base:hasDirected ?DirectorRole.
  Filter regex(str(?Director), "(?i)(?:shetty)$")
  ?ab base:hasDirected ?DirectorRole
}

```

**b)**

**Copy the query in query.rq file and run the command in fusceki's bin folder.**

Run the command :-

```

sudo ruby ./s-query --service=http://amanroy-asuspro-p5440uf:3030/Netflix --query=query.rq >>
Q2-b.txt

```

Query

```

prefix owl: <http://www.w3.org/2002/07/owl#>
prefix base: <http://www.semanticweb.org/amanroy/ontologies/2020/3/Movie#>
SELECT distinct ?Movie ?Genre
FROM <http://iiitd.ac.in/sweb/2016011/oldmoviesgraph>
FROM <http://iiitd.ac.in/sweb/2016011/newmoviesgraph>
FROM <http://iiitd.ac.in/sweb/2016011/remainingmoviesgraph>
WHERE {
  ?Movie base:hasGenre ?Genre.
  Filter regex(str(?Genre), "(?i)(?:Comedies|comedy|drama)")
}

```

**c)**

**Copy the query in query.rq file and run the command in fusceki's bin folder.**

Run the command :-

```

sudo ruby ./s-query --service=http://amanroy-asuspro-p5440uf:3030/Netflix --query=query.rq >>
Q2-c.txt

```

Query

```

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
prefix owl: <http://www.w3.org/2002/07/owl#>
prefix base: <http://www.semanticweb.org/amanroy/ontologies/2020/3/Movie#>

```

```

SELECT ?Movie ?Year ?country
FROM <http://iiitd.ac.in/sweb/2016011/oldmoviesgraph>
FROM <http://iiitd.ac.in/sweb/2016011/newmoviesgraph>
FROM <http://iiitd.ac.in/sweb/2016011/remainingmoviesgraph>
WHERE {
    ?Movie base:hasReleaseYear ?Year.
    ?Movie base:isReleasedInCountry ?country.
    ?Movie base:hasDescription ?description.
    Filter (xsd:int(?Year) > 2010 && xsd:int(?Year) < 2020 && regex(str(?description),
"(?i)(?:couple)") && regex(str(?country), "(?i)(?:united_state)") )
}

```

d)

**Copy the query in query.rq file and run the command in fusccki's bin folder.**

Run the command :-

```

sudo ruby ./s-query --service=http://amanroy-asuspro-p5440uf:3030/Netflix --query=query.rq >>
Q2-d.txt

```

Query

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

prefix owl: <http://www.w3.org/2002/07/owl#>

prefix base: <http://www.semanticweb.org/amanroy/ontologies/2020/3/Movie#>

```

SELECT ?Movie ?Movie1 ?Director ?Director1 ?time ?time1
FROM <http://iiitd.ac.in/sweb/2016011/oldmoviesgraph>
FROM <http://iiitd.ac.in/sweb/2016011/newmoviesgraph>
FROM <http://iiitd.ac.in/sweb/2016011/remainingmoviesgraph>
WHERE {
    ?Movie base:hasDirectorRole ?DirectorRole.
    ?Director base:hasDirected ?DirectorRole.
    ?Movie1 base:hasDirectorRole ?DirectorRole1.
    ?Director1 base:hasDirected ?DirectorRole1.
    ?Movie base:hasTimeDuration ?time.
    ?Movie1 base:hasTimeDuration ?time1
    Filter (xsd:int(?time) >= 60 && xsd:int(?time1) >= 60 &&
str(?DirectorRole)!=str(?DirectorRole1) && str(?Movie) > str(?Movie1) && str(?Director) =
str(?Director1))
}

```

e)

**Copy the query in query.rq file and run the command in fusceki's bin folder.**

Run the command :-

```
sudo ruby ./s-query --service=http://amanroy-asuspro-p5440uf:3030/Netflix --query=query.rq >>
Q2-e.txt
```

Query

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
prefix owl: <http://www.w3.org/2002/07/owl#>
prefix base: <http://www.semanticweb.org/amanroy/ontologies/2020/3/Movie#>
SELECT distinct ?Movie ?CastRole ?DirectorRole ?Country ?Genre
FROM <http://iiitd.ac.in/sweb/2016011/oldmoviesgraph>
FROM <http://iiitd.ac.in/sweb/2016011/newmoviesgraph>
FROM <http://iiitd.ac.in/sweb/2016011/remainingmoviesgraph>
WHERE {
  ?Movie base:hasCastRole ?CastRole.
  ?Movie base:hasDirectorRole ?DirectorRole.
  ?Movie base:isReleasedInCountry ?Country.
  ?Movie base:hasGenre ?Genre.
}
```

This is the diagram for the star pattern.

