Linked Movie Recommender

Semantic Web Technologies
Fall Semester 2018

Patrick Heinze, Anh Le, Leon Schüller

4th December 2018

University of Mannheim

Agenda

1. Goals and vision for the recommender

2. Semantic movie datasets

3. Similarity measures and recommendation algorithm

4. Example results

5. Outlook

6. Lessons learned

Movie Recommender

Semantic Web Technologies
University of Mannheim

Goals and vision for the recommender

Introduction

- Use semantic web data to provide recommendations
- Provide suggestions based on attributes other than user ratings, e.g.
    - director
    - subject
    - genre
    - actors
- Combine weighted attributes with linear combination rule
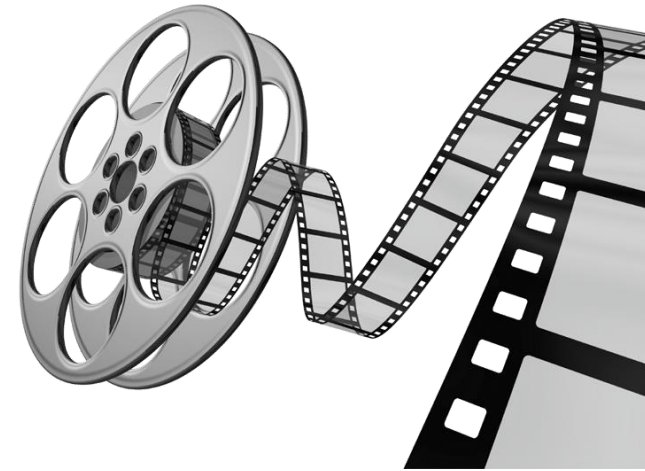
Vision statement

Exciting recommendations that are sometimes main stream and sometimes surprising alternative movies.

Semantic Movie Datasets

About: http://dbpedia.org/resource/Gladiator_(2000_film)

dbo:budget	• 1.03E8
dbo:cinematography	• dbr:John_Mathieson_(cinematographer)
dbo:director	• dbr:Ridley_Scott
dbo:distributor	• dbr:DreamWorks_Pictures • dbr:Universal_Pictures
dbo:editing	• dbr:Pietro_Scalia
dbo:gross	• 4.576E8
dbo:musicComposer	• dbr:Lisa_Gerrard • dbr:Hans_Zimmer
dbo:producer	• dbr:Branko_Lustig • dbr:Douglas_Wick • dbr:David_Franzoni
dbo:productionCompany	• dbr:Scott_Free_Productions • dbr:Red_Wagon_Entertainment
dbo:runtime	• 9300.000000 (xsd:double)
dbo:soundRecording	• dbr:Gladiator_(2000_film)__1
dbo:starring	• dbr:Richard_Harris • dbr:Connie_Nielsen • dbr:Oliver_Reed • dbr:Djimon_Hounsou • dbr:Russell_Crowe • dbr:Derek_Jacobi • dbr:Joaquin_Phoenix

Semantic Movie Datasets

dbo:abstract	
	• Gladiator is a 2000 British-American epic historical drama film directed by Ridley Scott. It stars Russell Crowe, Joaquin Phoenix, Connie Nielsen, Ralf Möller, Oliver Reed (in his final film role before his death), Djimon Hounsou, Derek Jacobi, John Shrapnel, and Richard Harris. Crowe portrays the fictional character, loyal Roman general Maximus Decimus Meridius, who is betrayed when Commodus, the ambitious son of Emperor Marcus Aurelius, murders his father and seizes the throne. Reduced to slavery, Maximus rises through the ranks of the gladiatorial arena to avenge the murders of his family and his emperor.The film was released in the United States on May 5, 2000, and grossed $457.6 million worldwide. The film won multiple awards, including Best Picture, Best Actor for Crowe, and three other Academy Awards at the 73rd Academy Awards. It has also been credited with rekindling interest in the historical epic. (en)

dct:subject	
	• dbc:2000s_drama_films
	• dbc:Films_about_death
	• dbc:2000s_adventure_films
	• dbc:Films_set_in_Rome
	• dbc:Nerva–Antonine_dynasty
	• dbc:American_films
	• dbc:English-language_films
	• dbc:Films_about_revenge
	• dbc:Regicide_in_fiction
	• dbc:Best_Drama_Picture_Golden_Globe_winners
	• dbc:Best_Film_BAFTA_Award_winners
	• dbc:British_films
	• dbc:Films_featuring_a_Best_Actor_Academy_Award-winning_performance
	• dbc:Films_that_won_the_Best_Sound_Mixing_Academy_Award
	• dbc:Films_that_won_the_Best_Visual_Effects_Academy_Award
	• dbc:Incest_in_film
	• dbc:British_epic_films
	• dbc:British_historical_films
	• dbc:Films_shot_in_Morocco
	• dbc:British_action_films

Semantic Movie Datasets – DBPedia to WikiData

About: http://dbpedia.org/resource/Gladiator_(2000_film)

owl:sameAs
- http://sw.cyc.com/concept/Mx4r53a45axiT0q0oCkdEh0bqg
- dbpedia-ja:グラディエーター
- dbpedia-fr:Gladiator_(film,_2000)
- dbpedia-de:Gladiator_(Film)
- dbpedia-es:Gladiator
- dbpedia-it:Il_gladiatore
- dbpedia-cs:Gladiátor_(film)
- dbpedia-el:Μονομάχος_(ταινία_2000)
- dbpedia-eu:Gladiator
- dbpedia-id:Gladiator_(film)
- dbpedia-ko:글래디에이터_(2000년_영화)
- dbpedia-nl:Gladiator_(film)
- dbpedia-pt:Gladiador_(filme)
- dbpedia-wikidata:Q128518
- http://wikidata.org/entity/Q128518
- dbpedia-pl:Gladiator_(film)
- freebase:m.09q5w2
- dbr:Gladiator_(2000_film)

WIKIDATA   Item | Discussion

Gladiator (Q128518)

Semantic Movie Datasets

nominated for	Academy Award for Best Supporting Actor		edit
	statement is subject of	73rd Academy Awards	
	nominee	Joaquin Phoenix	
	point in time	25 March 2001	
	▾ 0 references		
			+ add reference
	Academy Award for Best Actor		edit

award received	Academy Award for Best Picture		edit
	point in time	2001	
	statement is subject of	73rd Academy Awards	
	winner	Douglas Wick	
		David Franzoni	
		Branko Lustig	
	▾ 0 references		
			+ add reference
	Academy Award for Best Actor		edit

FSK film rating	FSK 16		edit
	▸ 1 reference		
			+ add value

Semantic Movie Datasets – Sparql Query Plan

Semantic Movie Datasets

IIa

Query Movie Data DBPedia

I

Query Movie URI DBPedia

Latch

IIb

Query Movie Data WikiData

concurrent requests

limit/offset 10.000
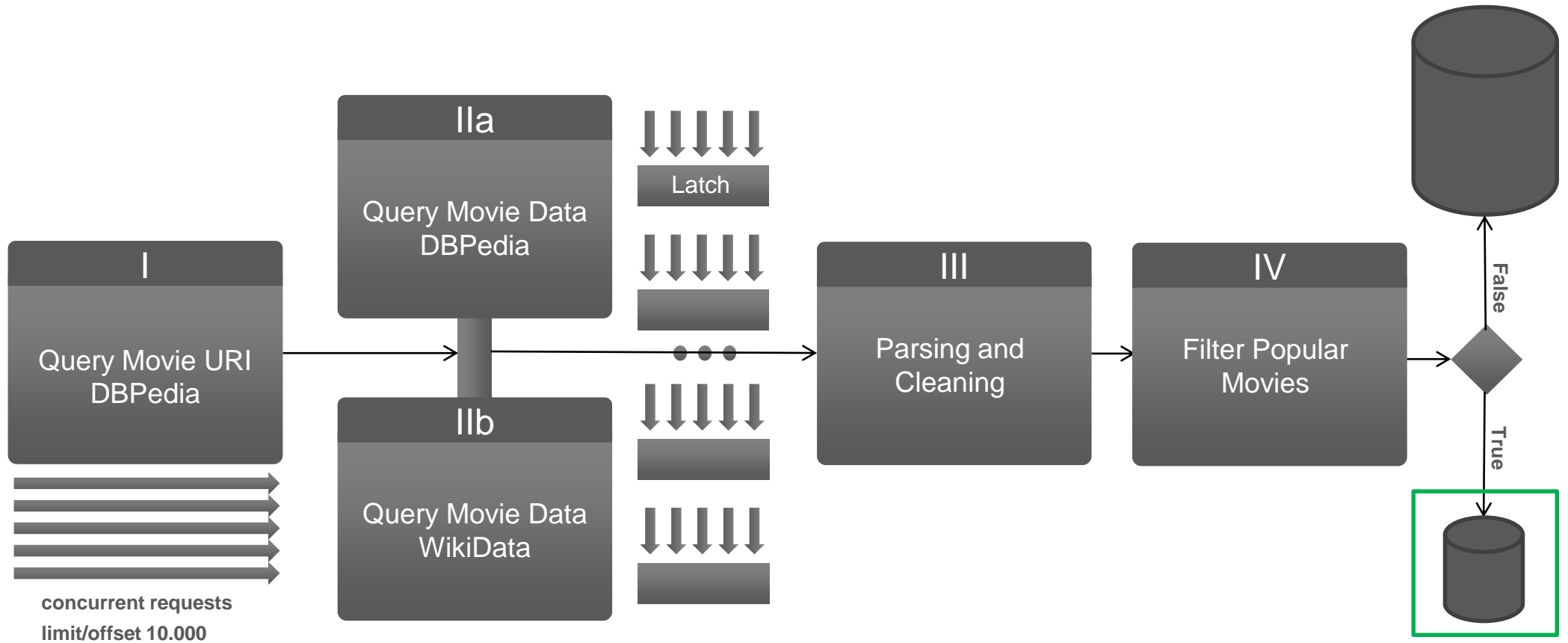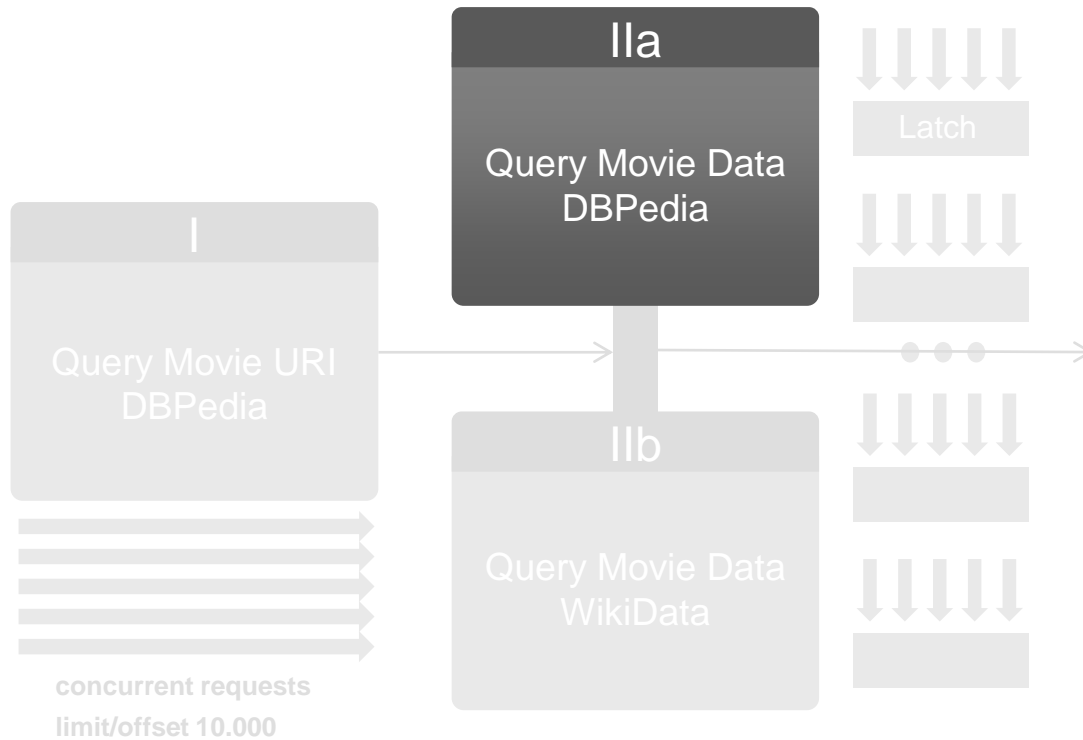
```
SELECT ?budget ?director ?link ?cinematography ?gross
(GROUP_CONCAT(DISTINCT ?name;separator='|') AS ?names)
(GROUP_CONCAT(DISTINCT ?runtime;separator='|') AS ?runtimes)
(GROUP_CONCAT(DISTINCT ?country;separator='|') AS ?countries)
(GROUP_CONCAT(DISTINCT ?writer;separator='|') AS ?writers)
(GROUP_CONCAT(DISTINCT ?starring;separator='|') AS ?actors)
(GROUP_CONCAT(DISTINCT ?language;separator='|') AS ?languages)
(GROUP_CONCAT(DISTINCT ?distributor;separator='|') AS ?distributors)
(GROUP_CONCAT(DISTINCT ?composer;separator='|') AS ?composers)
(GROUP_CONCAT(DISTINCT ?producer;separator='|') AS ?producers)
(GROUP_CONCAT(DISTINCT ?subject;separator='|') AS ?subjects)
(GROUP_CONCAT(DISTINCT ?prodcompany;separator='|') AS ?companies)
(GROUP_CONCAT(DISTINCT ?studio;separator='|') AS ?studios)
WHERE {
  mov: rdf:type schema:Movie.
  OPTIONAL {mov: dbo:budget ?budget.}
  OPTIONAL {mov: dbo:cinematography ?cinematography.}
  OPTIONAL {mov: dbo:director ?director.}
  OPTIONAL {mov: dbo:distributor ?distributor.}
  OPTIONAL {mov: dbp:studio ?studio. FILTER(regex(?studio,'resource'))}
  OPTIONAL {mov: owl:sameAs ?link. FILTER(regex(?link,'wikidata'))}
  OPTIONAL {mov: dbo:gross ?gross.}
  OPTIONAL {mov: dbo:writer ?writer.}
  OPTIONAL {mov: dbp:country ?country.}
  OPTIONAL {mov: dbo:musicComposer ?composer.}
  OPTIONAL {mov: dbp:productionCompanies ?prodcompany.}
  OPTIONAL {mov: dbo:starring ?starring.}
  OPTIONAL {mov: dbp:language ?language.}
  OPTIONAL {mov: dct:subject ?subject.}
  OPTIONAL {mov: dbo:producer ?producer.}
  OPTIONAL {mov: dbo:runtime ?runtime.}
  OPTIONAL {mov: foaf:name ?name. FILTER (lang(?name)= 'en')}
  OPTIONAL {mov: dbo:abstract ?abstract. FILTER (lang(?abstract)= 'en')}
}

GROUP BY ?budget ?link ?director ?cinematography ?gross ?abstract
LIMIT 1;
```

Similarity measures and recommendation algorithm
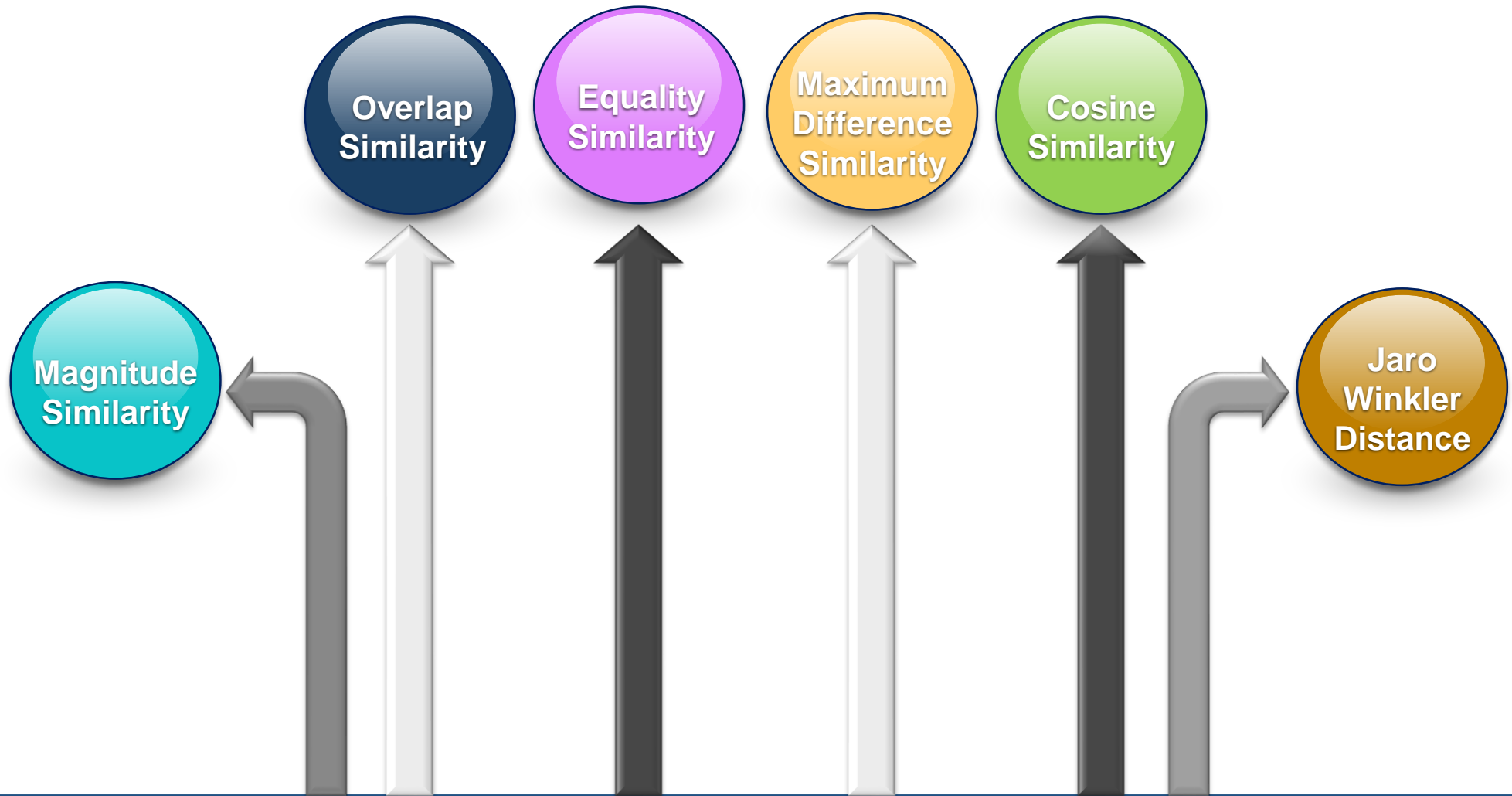
# Similarity measures and recommendation algorithm
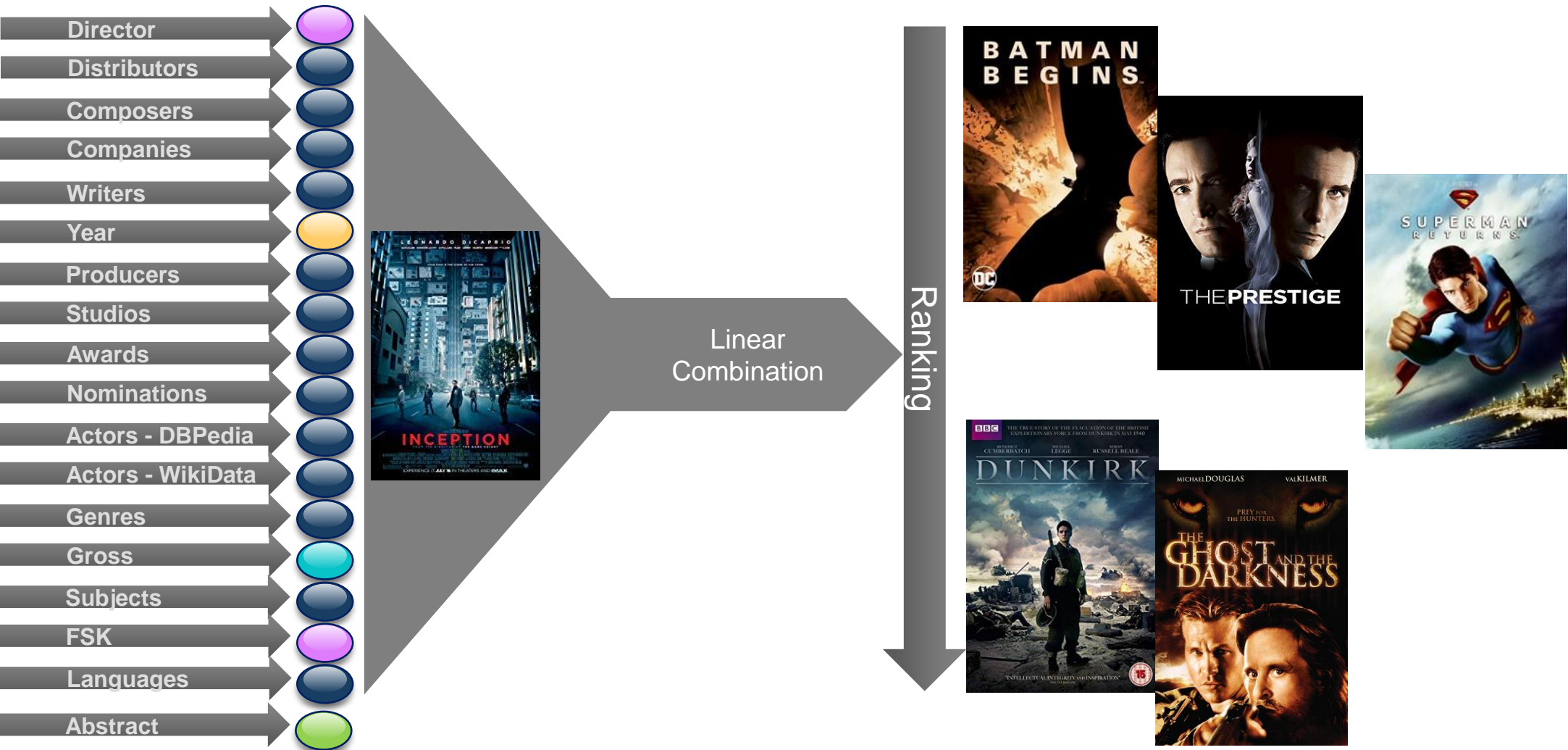
Example results - Application

- Uses precomputed JSON dataset

- User can enter up to 5 movies he wants similar movies for

- Fuzzy search to find movie(s) in the dataset while

  accounting for possible spelling mistakes

e.g. for input "Inception"

1. Batman Begins (0.4115)

2. The Prestige (0.4031)

3. Superman Returns (0.3734)

4. Dunkirk (0.3633)

5. The Ghost and the Darkness (0.3614)

6. The Abyss (0.3603)

7. Mad Max: Fury Road (0.3519)

8. Man of Steel (0.3518)

9. Transformers: Revenge of the Fallen (0.3464)

10. Spider-Man 2 (0.3409)
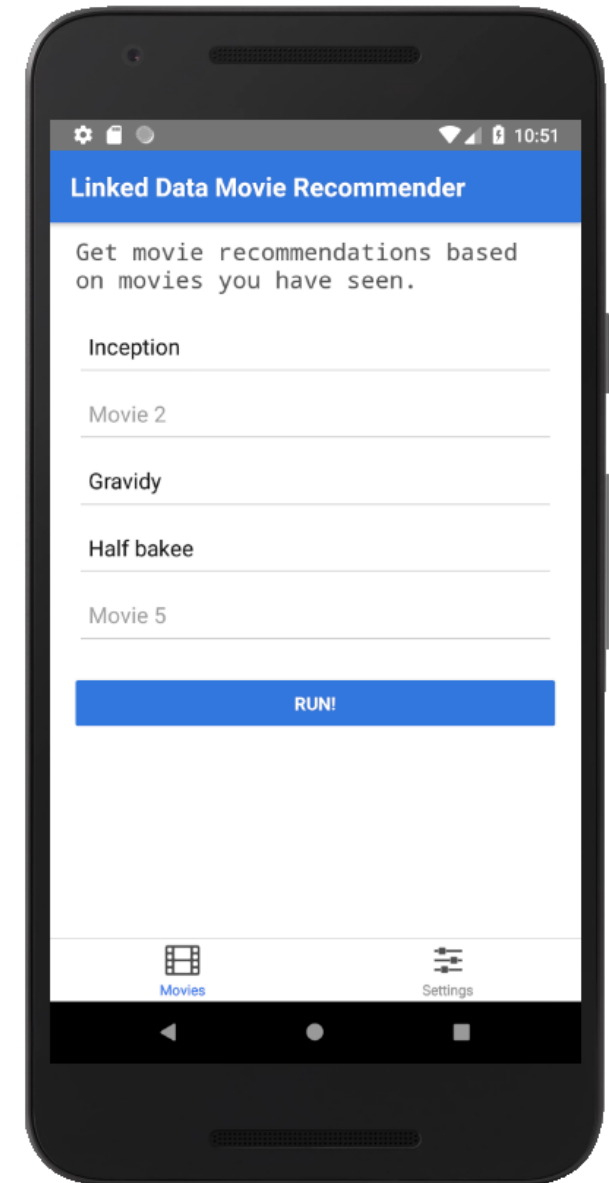
Example results - Application

One entry of dataset

```
{
    "budget": 23000000,
    "gross": 174800000,
    "name": "The Grand Budapest Hotel",
    "cinematography": "Robert_Yeoman",
    "abstractParagraph": "The Grand Budapest Hotel is a 2014 comedy film written and directed by Wes Anderson from a story by Anderson a
    "director": "Wes_Anderson",
    "runtime": 100,
    "fsk": "FSK 12@en",
    "year": 2014,
    "genres": "drama film | comedy film | adventure film | crime film | romance film",
    "actorsWiki": "Ralph Fiennes | Bill Murray | Waris Ahluwalia | Florian Lukas | Claudia Jung | Lisa Kreuzer | Daniel Steiner | Phili
    "actorsDP": "Adrien_Brody | Bill_Murray | Edward_Norton | F._Murray_Abraham | Harvey_Keitel | Jason_Schwartzman | Jeff_Goldblum | Ju
    "producers": "Jeremy_Dawson_(producer) | Scott_Rudin | Steven_M._Rales",
    "distributors": "Fox_Searchlight_Pictures",
    "subjects": "Category:2010s_criminal_comedy_films | Category:2014_films | Category:American_alternate_history_films | Category:Ameri
    "studios": "Scott_Rudin_Productions | Steven_M._Rales | Studio_Babelsberg | TSG_Entertainment",
    "writers": "Wes_Anderson",
    "composers": "Alexandre_Desplat",
    "companies": "",
    "languages": "English",
    "countries": "United States Germany United Kingdom",
    "nominations": "Academy Award for Best Writing Original Screenplay | Academy Award for Best Picture | Academy Award for Best Directo
    "awards": "Academy Award for Best Makeup and Hairstyling | Academy Award for Best Costume Design | Academy Award for Best Production
    "top10Recommendations": "The Life Aquatic with Steve Zissou|Finding Neverland|The Milagro Beanfield War|Me and Earl and the Dying Gi
},
```

Example results - Application

- Movies' Top 10 recommendations are counted and sorted

- Movies that appear most often will be on top of the list

- If count is equal, movies are sorted by score
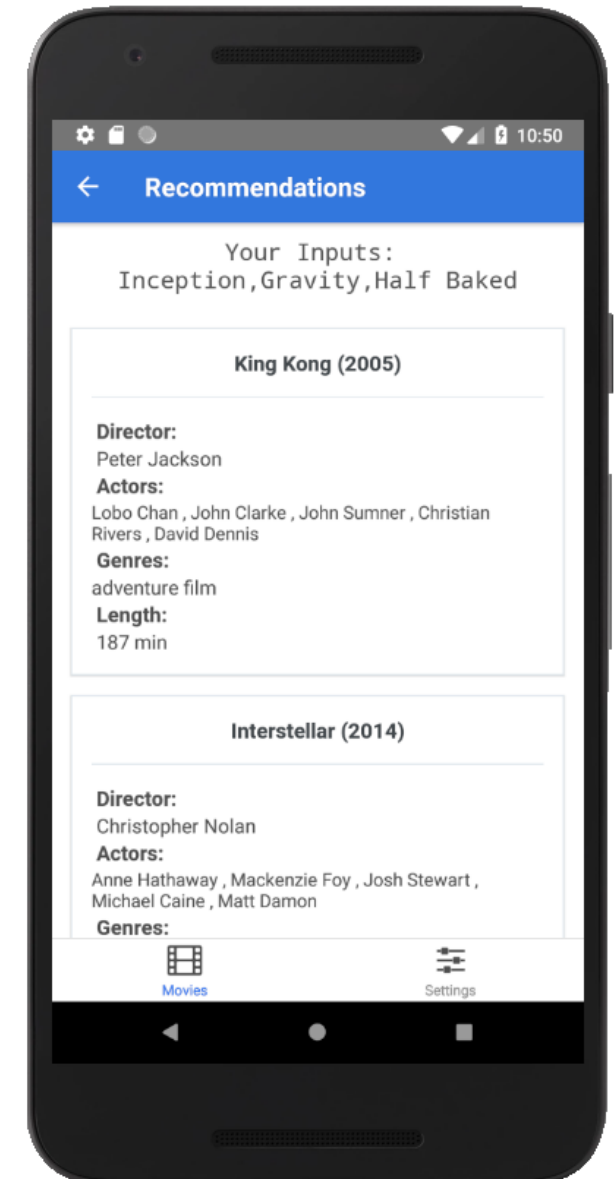
Outlook – Real time retrieval with ElasticSearch

- **ElasticSearch uses BM-25 probabilistic weighting scheme to index attribute fields**
- **We create an index for the movie data and indexed the attribute fields accordingly**
- **We use the Bulk Load API to push to data to the cluster**
- **We use the Search API to retrieve movies within milliseconds**
- **APIs use RestHighLevelClient and are tailored to Java**

Integration with Javascript-based App failed

Outlook

- More movies in the dataset (currently: 10000 highest grossing)

- Filter by Year, Genre, Language …

- Show more information about recommended movies, integration of services or links

  e.g. Wikipedia, Rotten Tomatoes, IMDB …

- Use a machine learning based approach to support further domains

Lessons learned

- Java is a good fit for Jena and SPARQL

  - Computation too costly for real-time results

  - SPARQL endpoint not built to handle ad-hoc queries

- Use python for data science related tasks

  - Reuse of existing frameworks

  - Enable a machine learning based approach with auto-encoders

Questions

?

Patrick Heinze, Anh Le, Leon Schüller

4th December 2018

University of Mannheim