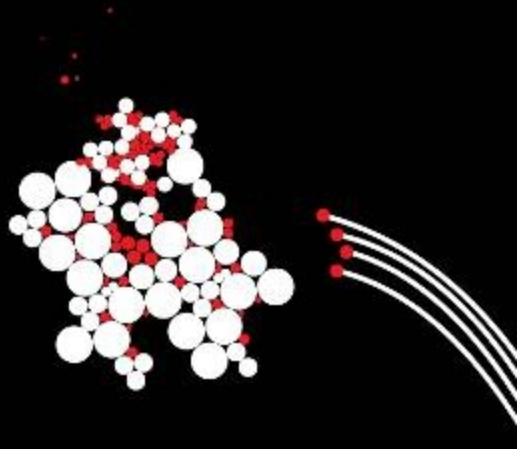
UNIVERSITEIT TWENTE.



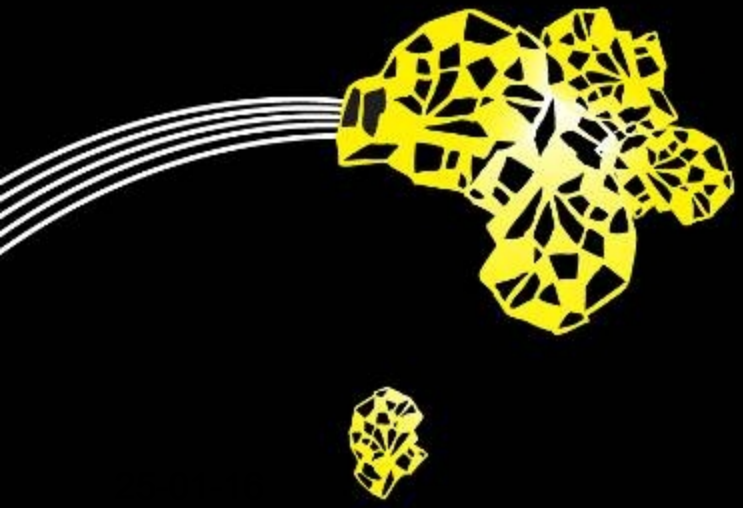
ENRICHING WOLFF BIOSCOPEN WEBSITE USING LINKED DATA FROM DBPEDIA

GROUP 4

MUHAMMAD ARIF WICAKSANA,

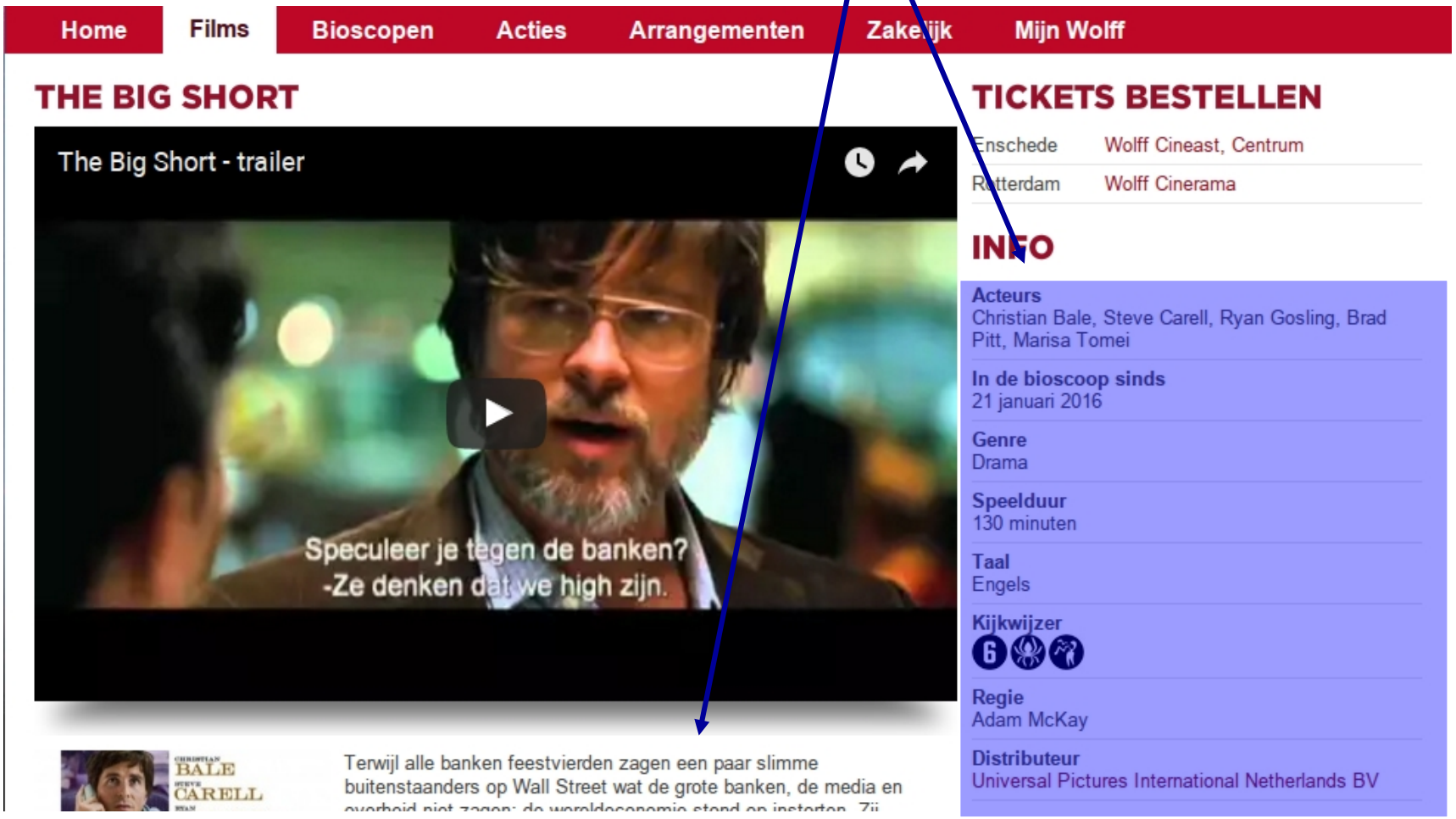
IWAN TIMMER,

KHRISHNA KT CHANDRASEKAR



INTRODUCTION

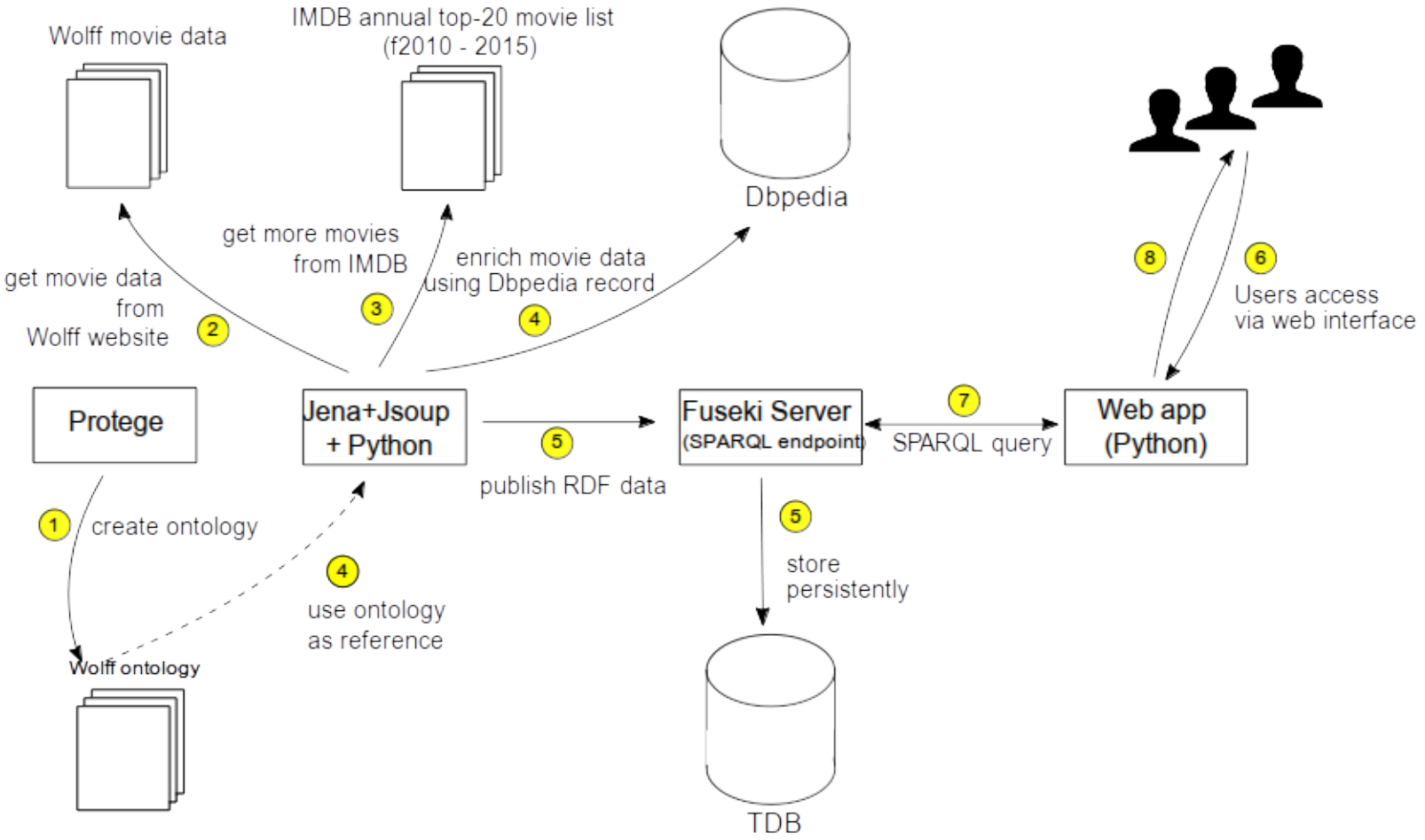
Limited information



GOAL

To enrich movie data from Wolff Bioscopen with linked data from Dbpedia

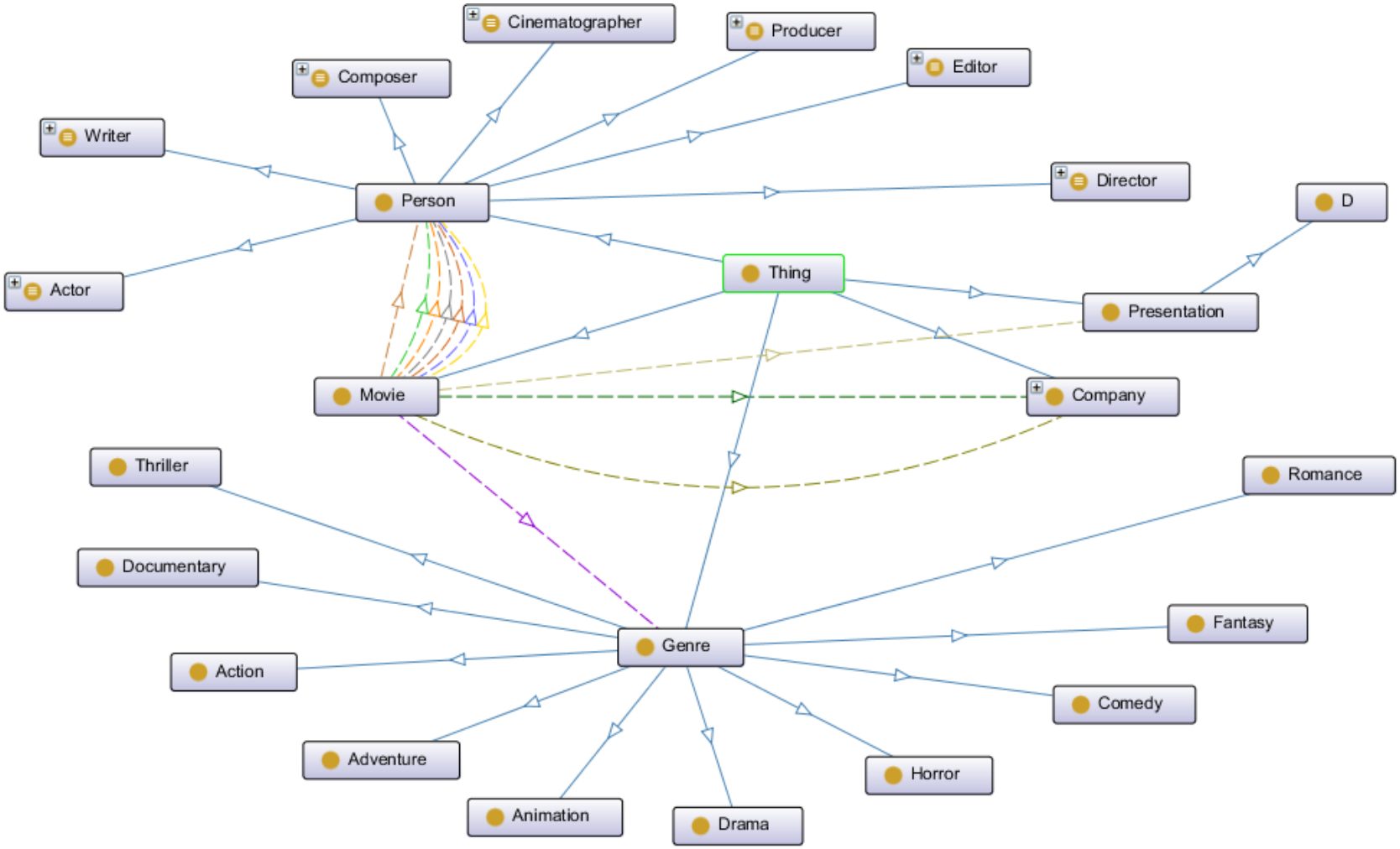
SYSTEM DESCRIPTION



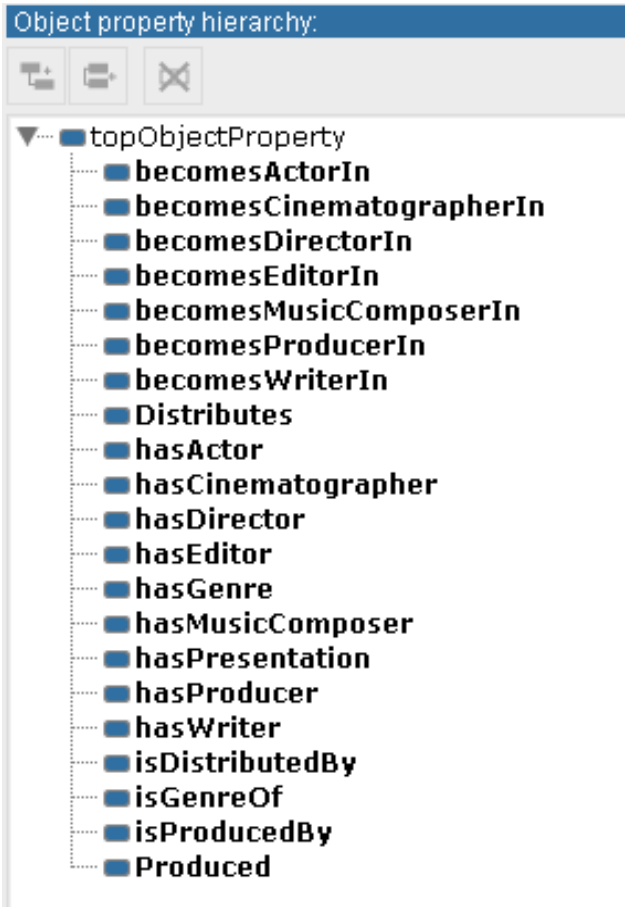
ONTOLOGY

- Uses OWL due to its expressivity
- Describes classes and properties related to film
- 5 main classes (Company, Genre, Movie, Person, Presentation)
- 20 sub-classes in total

ONTOLOGY


Cinematographer
Producer
Composer
Editor
Writer
Director
Person
D
Actor
Thing
Presentation
Movie
Company
Thriller
Romance
Documentary
Fantasy
Genre
Action
Comedy
Adventure
Horror
Animation
Drama

ONTOLOGY: PROPERTIES



ONTOLOGY

$$\text{becomesActorIn} \equiv hasActor^{-}$$
$$\text{becomesCinematographerIn} \equiv hasCinematographer^{-}$$
$$\text{becomesDirectorIn} \equiv hasDirector^{-}$$
$$\text{becomesEditorIn} \equiv hasEditor^{-}$$
$$\text{becomesMusicComposerIn} \equiv hasMusicComposer^{-}$$
$$\text{becomesProducerIn} \equiv hasProducer^{-}$$
$$\text{becomesWriterIn} \equiv hasWriter^{-}$$

$$\text{Actor} \equiv \text{Person} \sqcap \exists \text{becomesActorIn.Movie}$$
$$\text{Cinematographer} \equiv \text{Person} \sqcap \exists \text{becomesCinematographerIn.Movie}$$
$$\text{Composer} \equiv \text{Person} \sqcap \exists \text{becomesMusicComposerIn.Movie}$$
$$\text{Director} \equiv \text{Person} \sqcap \exists \text{becomesDirectorIn.Movie}$$
$$\text{Producer} \equiv \text{Person} \sqcap \exists \text{becomesProducerIn.Movie}$$
$$\text{Writer} \equiv \text{Person} \sqcap \exists \text{becomesWriterIn.Movie}$$

ENRICHING ORIGINAL DATASET

- Original data: movie title, actors, directors, release date, duration
- Use only additional data from Dbpedia
  - Linked movie database (linkedmdb) has outdated database
  - Connection to linkedmdb server is unreliable
- Additional data from Dbpedia:
  - Movie-related: producers, composers, editors, budget, abstract
  - Person-related: birth-date
- RDF individuals are given owl:sameAs towards its corresponding Dbpedia resources, if available

EXAMPLE QUERIES:
People with whom the subject frequently works with

```
SELECT DISTINCT ?name (GROUP_CONCAT(DISTINCT ?movietitle; SEPARATOR = "; ") AS ?movielist)
    (COUNT(DISTINCT ?movietitle) AS ?movietotal) WHERE {
  wolff-p:Quentin_Tarantino owl:sameAs ?dbplink .
  SERVICE <http://dbpedia.org/sparql> {
    ?movie a dbo:Film .
    ?movie ?p ?dbplink .
    ?movie rdfs:label ?movietitle .
    ?movie ?p2 ?actor .
    ?actor a dbo:Person .
    ?actor rdfs:label ?name .
    FILTER (langMatches(lang(?movietitle),"en"))
    FILTER (langMatches(lang(?name),"en"))
    FILTER (?actor != ?dbplink)
  }
} GROUP BY ?name  ORDER BY DESC(?movietotal)  LIMIT 15
```

EXAMPLE QUERIES:
Oscar Winners with whom the subject frequently works with

```
SELECT DISTINCT ?strippedname ?strippedtitle WHERE {
  wolff-p:Quentin_Tarantino owl:sameAs ?p_link .
  SERVICE <http://dbpedia.org/sparql> {
    ?movie a dbo:Film .  ?movie ?pre ?p_link .    ?movie dbo:starring ?actor .
    { ?actor dct:subject <http://dbpedia.org/resource/Category:Best_Actor_Academy_Award_winners> .}
    UNION { ?actor dct:subject <http://dbpedia.org/resource/Category:Best_Actress_Academy_Award_winners> . }
    UNION { ?actor dct:subject <http://dbpedia.org/resource/Category:Best_Supporting_Actor_Academy_Award_winners> }
    UNION { ?actor dct:subject <http://dbpedia.org/resource/Category:Best_Supporting_Actress_Academy_Award_winners> }
    ?actor rdfs:label ?name .
    ?movie rdfs:label ?title .
    FILTER(?p_link != ?actor)    FILTER (langMatches(lang(?name),"en"))
    FILTER (langMatches(lang(?title),"en"))
    BIND(str(?name) AS ?strippedname)    BIND(str(?title) AS ?strippedtitle)
  }
}
```

CONCLUSIONS

- Linked data allows us to reuse information provided by others automatically
- Linking data could be complicated
  - Ambiguity
  - Different ontologies describing same thing
- Self-developed ontology vs reuse existing ontology
- Difficulties to find good linked data provider for movie subject
  - Linkedmdb problems: outdated datasets, server reliability, more often does not provide more complete information than Dbpedia
  - Despite its status as the largest linked datasets, Dbpedia does not provide complete information for movies. But at least, it is the most up-to-date