Fact Validation Ensemble Learning
Summer Semester 2022

Michael Röder

DICE
Data Science Group
Paderborn University

February 7, 2022
Knowledge Graphs

Alice
http://example.org/alice#me

BOB
http://example.org/bob#me

Person
foaf:Person

"1990-07-04"^^xsd:date

foaf:knows

Leonardo Da Vinci
http://dbpedia.org/resource/Leonardo_da_Vinci

The Mona Lisa
http://www.wikidata.org/entity/Q12418

"Mona Lisa"

dcterms:creator

dcterms:title

dcterms:subject

La Joconde à Washington
http://data.europeana.eu/item/04802/243FA861638F4117025Fi17A8B813C5F9AA4D619
Fact Validation

Motivation

KnowGraphs

- Geonames
- DBpedia
- MusicBrainz
- Foaf

M. Röder (DICE): Fact Validation
Fact Validation

Motivation

KnowGraphs

Motivation

Geonames

DBpedia

Music

Brainz

Foaf

Fancy

algorithm 42

M. Röder (DICE): Fact Validation
Fact Validation

Motivation

Extraction Pipeline

KnowGraphs

DBpedia
Fact Validation

Motivation

Link set

Linking algorithm

DBpedia, GeoNames, MusicBrainz
Fact Validation

Motivation

M. Röder (DICE): Fact Validation
Fact Validation

Motivation

M. Röder (DICE): Fact Validation
Fact Validation

Motivation

Fact Validation

\( f(t) \rightarrow [-1,1] \)
Fact Validation

Motivation

:BarackObama :birthPlace :Kamerun

Fact Validation

\( f(t) \rightarrow [-1,1] \)

-0.64
Fact Validation

Motivation

:BarackObama :birthPlace :Kamerun -0.64

f(t) → [-1,1]

:BarackObama :birthPlace :USA 0.97
Fact Validation

Motivation

Fact Validation

\[ f(t) \rightarrow [-1, 1] \]

-0.64

:BarackObama :birthPlace :Kamerun

0.97

:BarackObama :birthPlace :USA
Fact Validation
Approaches

- COPAAL
- ESTHER
- Knowledge Streams
- Knowledge Linker
- Relational Knowledge Linker
- Predicate Path Mining
- Katz
- PathEnt
- ...

M. Röder (DICE): Fact Validation
### Project Group

#### Goals

<table>
<thead>
<tr>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>- <strong>Question</strong>: Does a combination of approaches lead to a better performance?</td>
</tr>
</tbody>
</table>
Summary

▶ **Question:** Does a combination of approaches lead to a better performance?

▶ **Solution:** Train and evaluate a meta classifier on the system’s results.
Summary

- **Question**: Does a combination of approaches lead to a better performance?
- **Solution**: Train and evaluate a meta classifier on the system’s results.

- Get and set up implementations of Fact Validation systems
- Gather and create relevant benchmark datasets
- Implement a meta classifier for Fact Validation
- Run a detailed evaluation
GERBIL
Further Information

Publications:

Github projects:
- [https://github.com/dice-group/copaal](https://github.com/dice-group/copaal)
- [https://github.com/dice-group/esther](https://github.com/dice-group/esther)
- [https://github.com/shiralkarprashant/knowledgestream](https://github.com/shiralkarprashant/knowledgestream)

Technologies:
- Java, Maven, Python (for existing systems)
- Knowledge about graph theory, RDF and machine learning is helpful