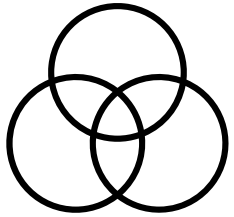
Semantic Similarity Demonstrator

Stefan Velev

Graph Databases, 0MI3400521

Big Data Technologies, GATE Institute

Sofia University “St. Kliment Ohridski”

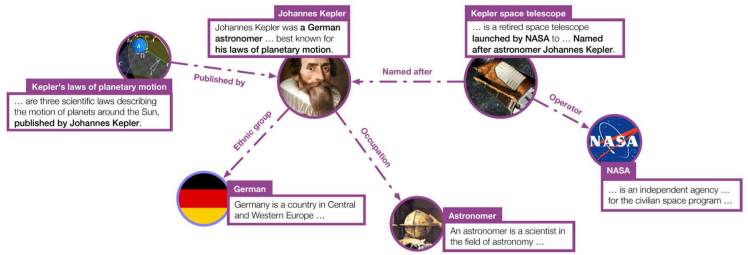


CONTENTS

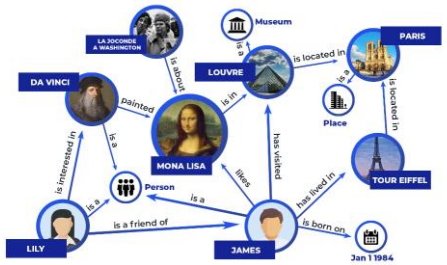
- **Web-based demonstrator** for computing **semantic similarity** between notable **individuals** using shared properties, such as place of birth, occupation, nationality, etc.
- **Sematch** – an integrated framework for the development, evaluation, and application of semantic similarity for knowledge graphs
- **Two methods** to measure similarity in **Sematch**:
  - **Entity similarity** – how much two entities are alike in meaning

    E.g.: entity_sim.similarity('http://dbpedia.org/resource/Apple_Inc.', 'http://dbpedia.org/resource/Steve_Jobs')
  - **Entity relatedness** – how much two entities are connected or associated in some way

    E.g.: entity_sim.relatedness('http://dbpedia.org/resource/Apple_Inc.', 'http://dbpedia.org/resource/Steve_Jobs')



Wikidata5m



human subgraph

Somatch

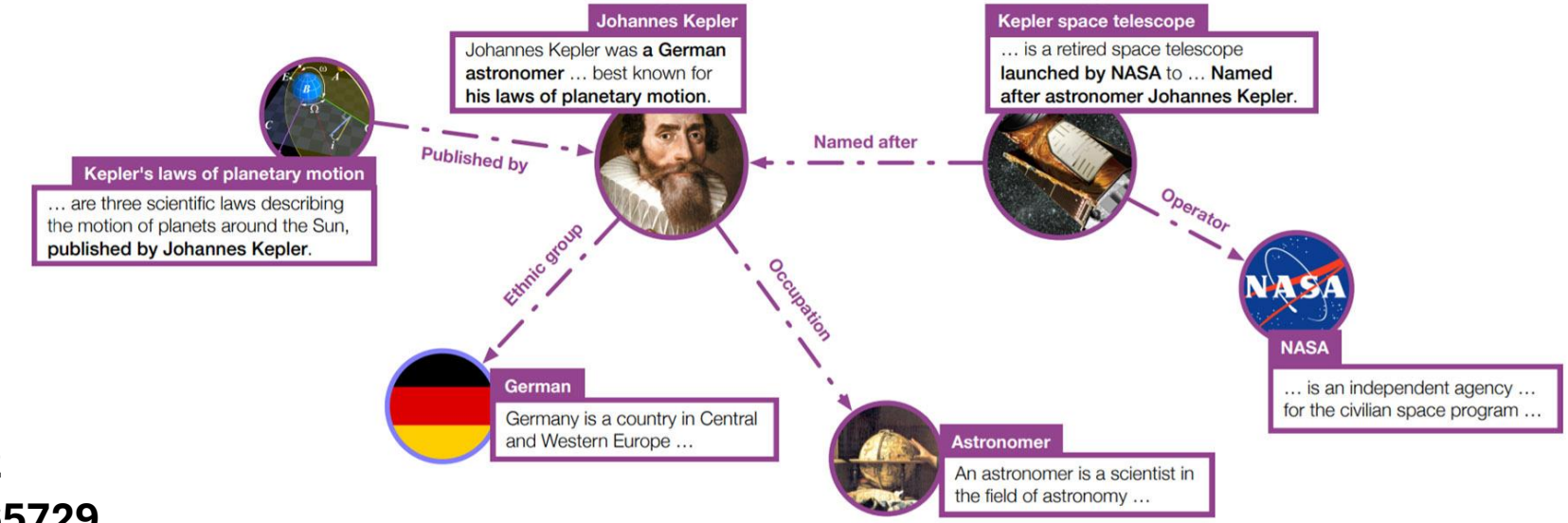
Streamlit

WIKIDATA

DBpedia

GraphDB

Wikidata5m: Transductive Split



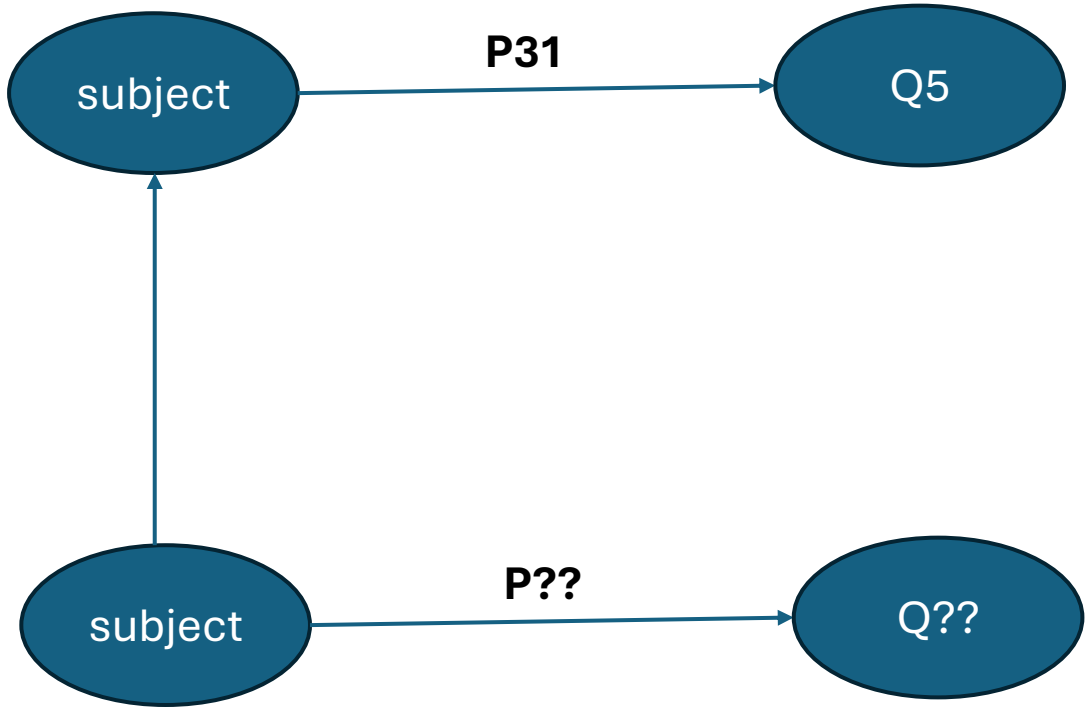
Q29387131	P31	Q5
Q326660	P1412	Q652
Q7339549	P57	Q1365729
Q554335	P27	Q29999
Q20641639	P54	Q80955
Q14946683	P31	Q5
Q4221140	P27	Q399
Q6925786	P131	Q488653
Q4890993	P19	Q931116
Q3198638	P156	Q2859200
Q24905727	P161	Q88139

1. Extract of the subgraph

```
human_entities = set()

subject, predicate, object = line.strip().split("\t")
if predicate ==  “P31” and object == “Q5”:
    human_entities.add(subject)
```

```
if subject in human_entities:
    fout.write(line)
    line_counter += 1
```



2. Convert to RDF Turtle format

```
def to_uri(qid, is_property=False):
    if is_property:
        return f"<http://www.wikidata.org/prop/direct/{qid}>"
    else:
        return f"<http://www.wikidata.org/entity/{qid}>"
```

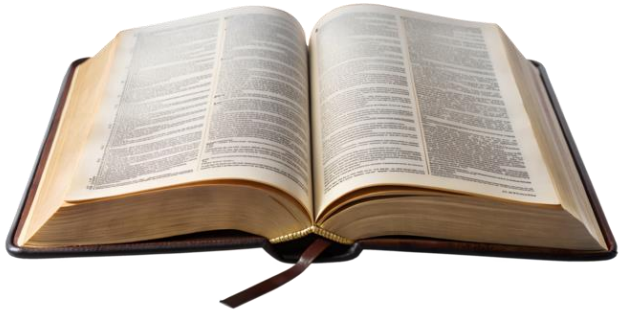
Q29387131	P31	Q5
Q326660	P1412	Q652

```
@prefix wd: <http://www.wikidata.org/entity/> .
@prefix wdt: <http://www.wikidata.org/prop/direct/> .

<http://www.wikidata.org/entity/Q29387131> <http://www.wikidata.org/prop/direct/P31> <http://www.wikidata.org/entity/Q5> .
<http://www.wikidata.org/entity/Q326660> <http://www.wikidata.org/prop/direct/P1412> <http://www.wikidata.org/entity/Q652> .
```

3. Enrich human subgraph with labels from Wikidata

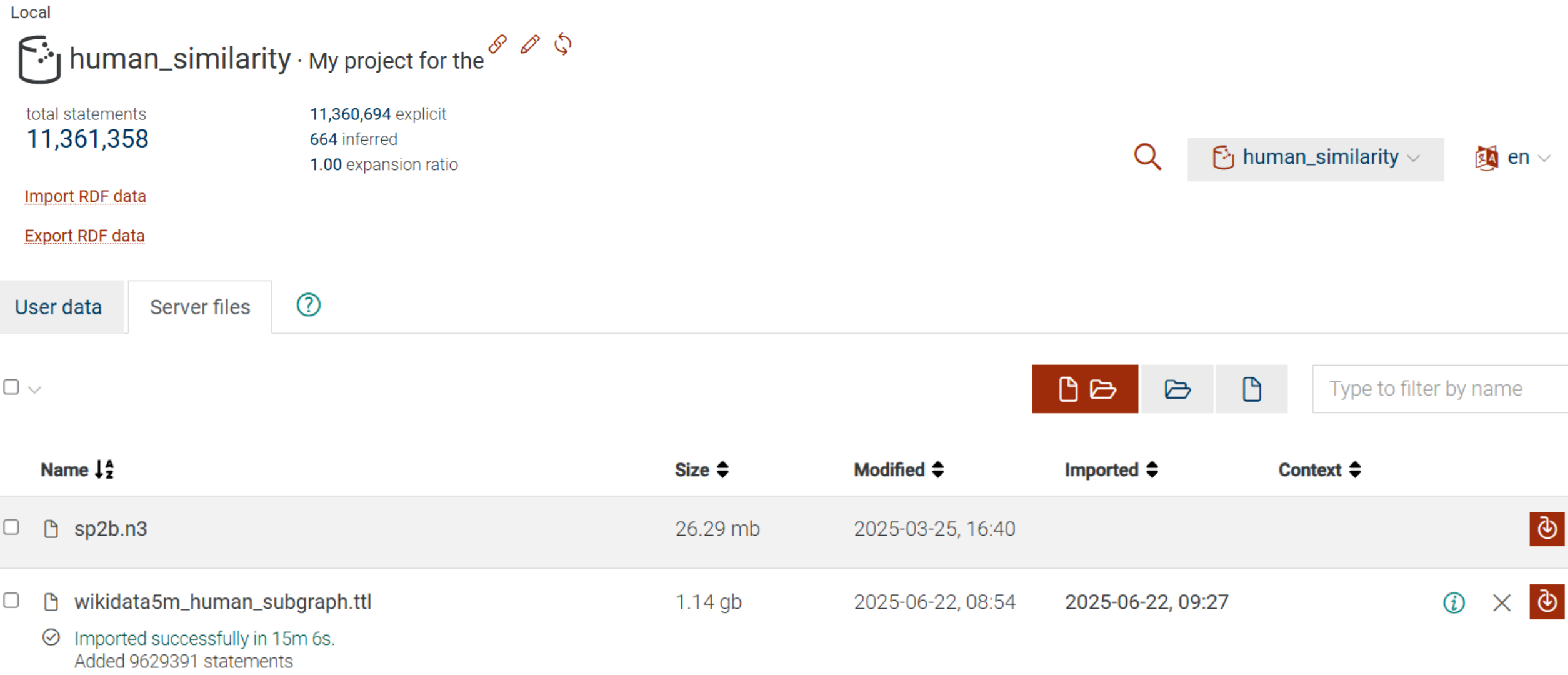
```
batch = qids[i:i+BATCH_SIZE]
ids = "|".join(batch)
url = "https://www.wikidata.org/w/api.php"
params = {
   "action": "wbgetentities",
   "ids": ids,
   "format": "json",
   "props": "labels",
   "languages": "en"
}
```



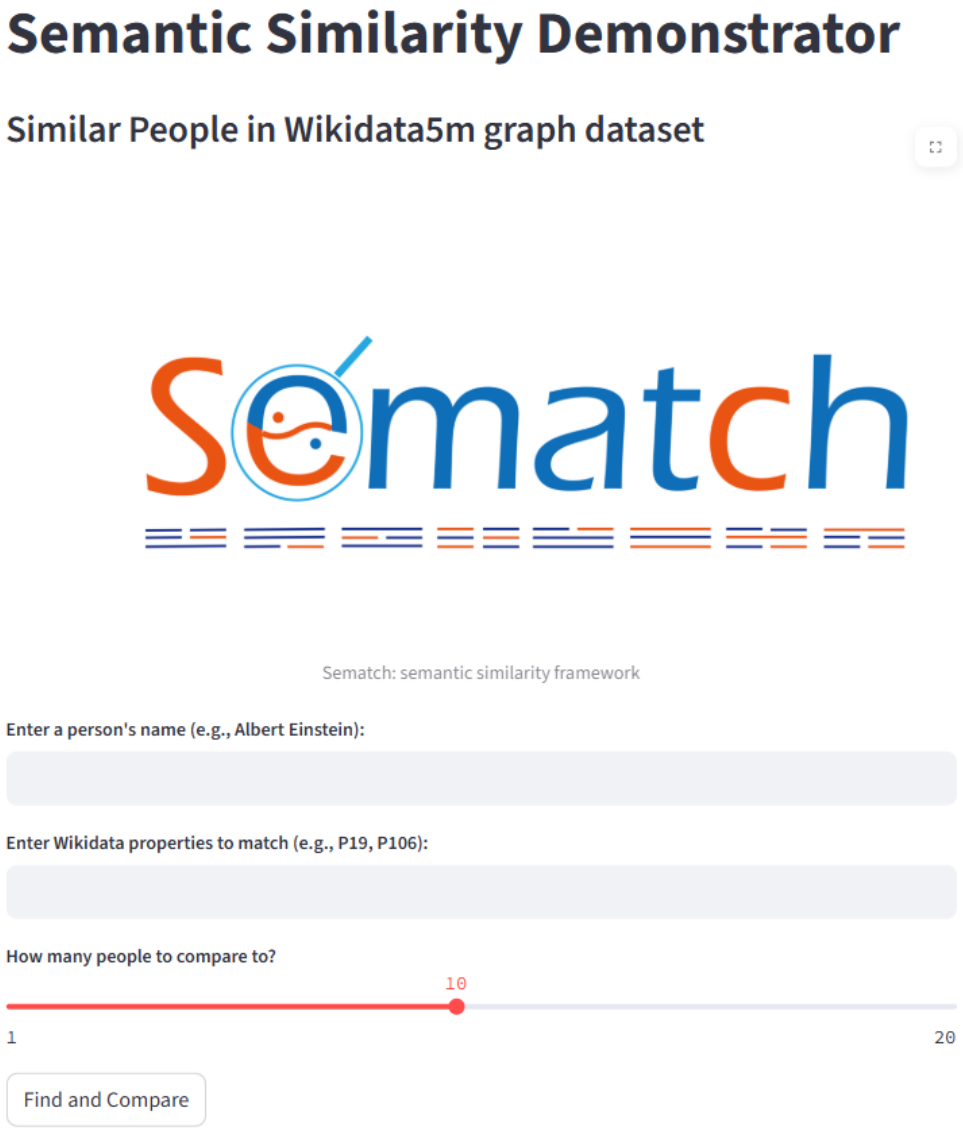
```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix wd: <http://www.wikidata.org/entity/> .

<http://www.wikidata.org/entity/Q100> rdfs:label "Boston"@en .
<http://www.wikidata.org/entity/Q1000> rdfs:label "Gabon"@en .
<http://www.wikidata.org/entity/Q100000> rdfs:label "Cadier en Keer"@en .
<http://www.wikidata.org/entity/Q1000009> rdfs:label "Neuville-Saint-Vaast"@en .
<http://www.wikidata.org/entity/Q1000020> rdfs:label "Melrose RFC"@en .
```

4. Create GraphDB repository and load the TTL files



5. Semantic Similarity Demonstrator – Streamlit




```python
def get_person_qid_by_label(label):
    query = f"""
    PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
    SELECT ?person WHERE {{
      ?person rdfs:label "{label}"@en .
    }} LIMIT 1
    """
    results = sparql_query(query)
    bindings = results["results"]["bindings"]
    if bindings:
        return bindings[0]["person"]["value"].split('/')[-1]
    return None
```

```python
dbpedia_uri = f"http://dbpedia.org/resource/{urllib.parse.quote(label.replace(' ', '_'))}"
target_dbpedia = f"http://dbpedia.org/resource/{urllib.parse.quote(person_label.replace(' ', '_'))}"
```

Top 10 similar people to Albert Einstein

1. Johann Josef Loschmidt

Similarity: 0.5157

2. Harald Lesch

Similarity: 0.4955

3. Johann Wolfgang von Goethe

Similarity: 0.4640

4. Pascual Jordan

Similarity: 0.4562

5. Immanuel Kant

Similarity: 0.4330

6. Gerhard Herzberg

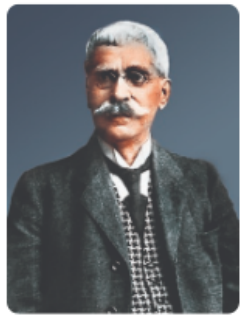
Similarity: 0.4248

7. Paul Ehrenfest

Similarity: 0.3104

8. Angela Merkel

Similarity: 0.3071



Ivan Vazov

Enter Wikidata properties to match (e.g., P19, P106):

P106, P136, P27

How many people to compare to?

3

1

Find and Compare

Top 3 similar people to Ivan Vazov

1. Stanislav Stratiev

Similarity: 0.6590

2. Anton Strashimirov

Similarity: 0.5908

3. Aleksandar Hadzhihristov

Similarity: 0.5698

Top 10 similar people to Dimitar Berbatov

1. Yanko Valkanov

Similarity: 0.8552

2. Georgi Yordanov

Similarity: 0.8552

3. Dimitar Makriev

Similarity: 0.8552

4. Radostin Kishishev

Similarity: 0.8552

5. Ivan Karamanov

Similarity: 0.8552

6. Vladislav Romanov

Similarity: 0.8552

7. Dimitar Nakov

Similarity: 0.8552

8. Dimitar Rangelov

Similarity: 0.8552