# Tests with Mondial and IMDb

(Companion tests for the paper "RDF Keyword-based Query Technology Meets a Real-World Dataset", by Grettel M. García, Yenier T. Izquierdo, Elisa S. Menendez, Frederic Dartayre, and Marco A. Casanova)

## 1. Introduction

We tested the keyword-search tool described in [1] against

- a triplified version of the full Mondial dataset[1]
- a triplified version of the Mondial dataset adopted in Coffman's benchmark[2] [2]
- a triplified version of the full IMDb dataset[3]

Contrasting with the versions adopted in Coffman's benchmark, the full versions of Mundial and IMDb feature conceptual schemas with a complexity closer to the schema of the target industrial dataset (see Table 1) used in [1]. We ran all queries in the original list of the Coffman's benchmark against each of these datasets and compared the results returned with the expected results.

An analysis of the failed queries reveals that: some may not be classified as full failures, since they returned the desired result, plus additional results; some failures can be blamed to the lack of keyword semantics; and some failures can be credited to the lack of accuracy of the keywords. These results actually indicate that the list of queries and query results in Coffman's benchmark should be reassessed.

**Table 1. RDF Statistics**

| Triple Type | #Triples | | |
|---|---|---|---|
| | **Full Version of Mondial** | **Coffman's Benchmark Mondial** | **Full Version of IMDb** |
| Class declarations | 40 | 26 | 21 |
| Object property declarations | 62 | 33 | 24 |
| Datatype property declarations | 130 | 79 | 24 |
| Indexed properties | 71 | 39 | 34 |
| Distinct indexed datatype property instances | 11.094 | 6.091 | 14.259.846 |
| Class instances | 43.869 | 16.621 | 72.973.275 |
| Object property instances | 63.652 | 19.706 | 184.818.637 |
| **Total number of triples** | **235.387** | **109.362** | **395.394.424** |

[1] https://www.dbis.informatik.uni-goettingen.de/Mondial/
[2] https://www.cs.virginia.edu/~jmc7tp/resources.php
[3] https://sites.google.com/site/ontopiswc13/home/imdb-mo

## 2. Tests with Mondial

### 2.1 Keyword Queries in Coffman's Benchmark for Mondial[2]

1-5: countries

Relevant results should contain a single tuple that is the specified country's tuple from the country relation.

6-10: cities

Relevant results should contain a single tuple that is the specified city's tuple from the city relation.

11-15: geographical

Relevant results should contain a single tuple that is the specified geographical entity's tuple from the appropriate relation. Refer to the qrels files for additional details (e.g., which relation the tuple is from).

16-20: organization

Relevant results should contain a single tuple that is the specified organization's tuple from the organization relation.

21-25: border between countries

Relevant results should contain 3 tuples (2 from the country relation and 1 from the borders relation) that identify the shared border between two countries. The relevant result answers the question "What is the length of the border between X and Y?" where X and Y are the respective countries.

26-35: geopolitical or demographic information

Relevant results should contain 2 tuples (1 from the country relation and 1 from another specified relation -- see the qrels files for additional details). These relevant results address questions related to political or demographic information for the specified country.

36-45: member organizations between two countries

Relevant results contain 5 tuples (2 from the country relation, 2 from the is_member relation, and 1 from the organization relation) and identify all organizations which the 2 specified countries are members of.

46-50: [misc]

In general, these topics relate arbitrary information within the database. The questions used to create the topics are provided below.

46: Which African countries contain the Hutu ethnic group?

47: Which European countries contain the Serb ethnic group?

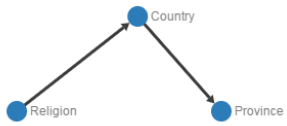
48: Which Asian countries contain the Uzbek ethnic group?

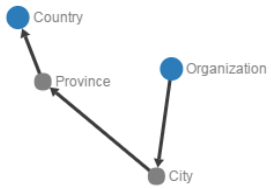
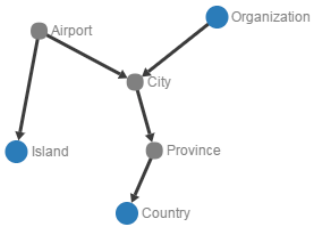
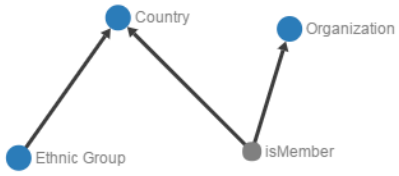
49: Which German provinces does the Rhein River flow through?

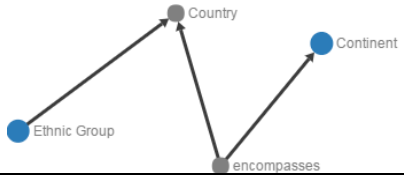
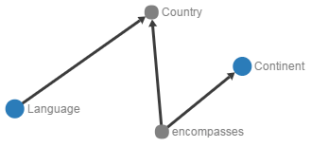
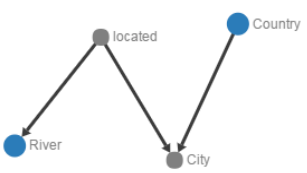
50: Which Egyptian provinces does the Nile River flow through?

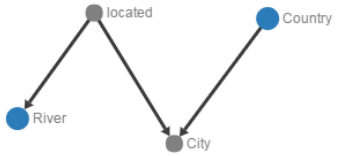
## 2.2 Results Obtained for the Full Mondial Dataset

| # | Keywords | Expected Answer | Application Answer | Observation |
|---|---|---|---|---|
| 1. | thailand | country | Thailand (Country) | √ |
| 2. | netherlands | country | Netherlands (Country) | √ |
| 3. | georgia | country | Georgia (Country) | √ |
| 4. | country china | country | China (Country) | √ |
| 5. | bangladesh | country | Bangladesh (Country) | √ |
| 6. | alexandria | Alexandria, USA | Alexandria (Romania)<br>Alexandria (USA) | There are 2 cities named "Alexandria" |
| 7. | sonsonate | Sonsonate, ES | - | The version of Mondial adopted does not have a city called "Sonsonate" |
| 8. | xiaogan | Xiaogan, TJ | Xiaogan (China) | √<br>The value of Code for "China" in class Country is **TJ.** |
| 9. | city glendale | Glendale, USA | Glendale, Arizona (USA)<br>Glendale, California (USA) | √ |
| 10. | city granada | Granada, E | Granada, Andalucía (ES)<br>Granada, Nicaragua (NIC) | √ |
| 11. | Lake Kariba | Lake Kariba | Lake Kariba | √ |
| 12. | Niger | Niger river | Niger (Country) | "Niger" is both a country and a river |
| 13. | Arabian Sea | Arabian Sea | Arabian Sea | √ |
| 14. | Asauad | Asauad desert | - | The version of Mondial adopted does not have a desert called "Asauad" |
| 15. | Sardegna | Sardegna (island) | Sardegna (Island) | √ |
| 16. | arab cooperation council | Arab Cooperation Council | Answer with 75 instances of class Organization | Queries were not correctly answered since the expected values were not listed in class Organization (in the version of Mondial used). |
| 17. | world labor | World Confederation of Labor | Answer with 9 instances of class Organization | |
| 18. | islamic conference | Organization of the Islamic Conference | • Islamic Development Bank<br>• Organization of Islamic Cooperation<br>• United Nations Conference on Trade and Development<br>• Conference of Interaction and Confidence-Building Measures in Asia<br>Class Organization | |
| 19. | 30 group | Group of 30 | Instances of class EthnicGroup | The expected value does not exist in property Name of class Organization. Since the numeric value is not indexed, the match was with the label of class EthnicGroup. |
| 20. | caribbean economic | Economic Commission for Latin America and the Caribbean | • Organization for Economic Cooperation and Development<br>• Economic and Monetary Union<br>Class Organization | The expected value does not exist for property Name of class Organization. |

| | | | | |
|---|---|---|---|---|
| 21. | slovakia hungary | borders | • Hungary<br>• Slovakia<br>Classe Country | Keywords match the labels of two instances of class Country; but the keywords are not sufficient to infer that the question is about the borders between countries and, thus, were not correctly answered. |
| 22. | mongolia china | borders | • China<br>• Mongolia<br>Class Country | |
| 23. | niger algeria | borders | • Niger<br>• Algeria<br>Class Country | |
| 24. | kuwait saudi arabia | borders | • Saudi Arabia<br>• Kuwait<br>Class Country | |
| 25. | lebanon syria | borders | • Syria<br>• Lebanon<br>Class **Country** | |
| 26. | cameroon economy | Cameroon | Cameroon \| ECONOMY OF CAM<br>Country ← Economy | √ |
| 27. | nigeria gdp | Nigeria | Nigeria \| ECONOMY OF WAN \| 502000<br>Country ← Economy | √ |
| 28. | mongolia republic | Mongolia | • Mongolia<br>• Czech Republic<br>• Dominican Republic<br>• Central African Republic<br>Class Country | As keywords match with labels of class Country |
| 29. | kiribati politics | Kiribati | Kiribati \| POLITICS OF KIR<br>Country ← Politics | √ |
| 30. | poland language | Poland | Poland \| Polish<br>Country ← Language | √ |
| 31. | spain galician | Spain | Spain \| Galician<br>Country ← Language | √ |
| 32. | uzbekistan eastern orthodox | Uzbekistan | -<br>  <br> | "eastern orthodox" is not a value of property Name of class Religion. |
| 33. | haiti religion | Haiti | Haiti \| Protestant<br>Haiti \| Roman Catholic<br>Country ← Religion | √ |
| 34. | suriname ethnic group | Suriname | Suriname \| African<br>Suriname \| Javanese<br>Suriname \| Creole<br>Suriname \| Amerindian<br>Suriname \| Hindustani<br>Suriname \| Chinese<br>Suriname \| Europeans<br>Country ← Ethnic Group | √ |
| 35. | slovakia german | Slovakia | Slovakia \| German<br>Country ← Ethnic Group | √ |

| | | | | |
|---|---|---|---|---|
| 36. | poland cape verde organization | • Nonaligned Movement<br>• Customs Cooperation Council<br>• Food and Agriculture Organization<br>Others… | -<br>  <br> | The expected answer is the list of organizations that the countries belong to; however, the translation algorithm did not identify the IS_MEMBER class when generating the nucleuses. |
| 37. | saint kitts cambodia | • International Bank for Reconstruction and Development<br>• International Civil Aviation Organization<br>• International Criminal Police Organization<br>Others… | • Saint Kitts and Nevis<br>• Cambodia<br>• Saint Lucia<br>• Saint Martin<br>• Saint Helena<br>• Saint Barthelemy<br>• Saint Pierre and Miquelon<br>• Saint Vincent and the Grenadines<br>Class Country | |
| 38. | marshall islands grenadines organization | • Food and Agriculture Organization<br>• Group of 77<br>• International Bank for Reconstruction and Development<br>Others… | -<br>  <br> | |
| 39. | czech republic cote divoire organization | • Customs Cooperation Council<br>• Food and Agriculture Organization<br>• International Atomic Energy Agency<br>Others… |  <br> | Only the “czech republic” is returned and appears in the nucleus with class EthnicGroup |
| 40. | panama oman | • Nonaligned Movement<br>• Food and Agriculture Organization<br>• Group of 77<br>Others… | • Oman<br>• Panama<br>Class Country | The expected answer is the list of organizations that the countries belong to; however, this is difficult to infer from the keywords. |
| 41. | iceland mali | • Customs Cooperation Council<br>• Food and Agriculture Organization<br>• International Atomic Energy Agency<br>Others... | • Iceland<br>• Mali<br>Class Country | |
| 42. | guyana sierra leone | • Commonwealth<br>• Nonaligned Movement<br>• African, Caribbean, and | • Iceland<br>• Mali<br>Class Country | |

<table>
<tr><td></td><td></td><td>Pacific Countries<br>Others…</td><td></td><td rowspan="4"></td></tr>
<tr><td>43.</td><td>mauritius india</td><td>• Group of 77<br>• International Atomic Energy Agency<br>• International Bank for Reconstruction and Development<br>Others…</td><td>• Mauritius<br>• India<br>Class Country</td></tr>
<tr><td>44.</td><td>vanuatu afghanistan</td><td>• Nonaligned Movement<br>• Asian Development Bank<br>• Group of 77<br>Others…</td><td>• Vanuatu<br>• Afghanistan<br>Class Country</td></tr>
<tr><td>45.</td><td>libya australia</td><td>• Nonaligned Movement<br>• Customs Cooperation Council<br>• Food and Agriculture Organization<br>Others…</td><td>• Libya<br>• Australia<br>Class Country</td></tr>
<tr><td>46.</td><td>hutu africa</td><td>Rwanda</td><td><table><tr><td>Africa</td><td>Hutu</td></tr></table>  </td><td>If we added the keyword “Country”, then the answer would be:<table><tr><td>Rwanda</td><td>Africa</td><td>Hutu</td></tr></table></td></tr>
<tr><td>47.</td><td>serb europe</td><td>Slovenia<br>Bosnia and Herzegovina<br>Croatia<br>Romania<br>Macedonia<br>Hungary</td><td><table><tr><td>Europe</td><td>Serb</td></tr><tr><td>Europe</td><td>Serb</td></tr><tr><td>Europe</td><td>Serb</td></tr><tr><td>Europe</td><td>Serb</td></tr><tr><td>Europe</td><td>Serb</td></tr><tr><td>Europe</td><td>Serb</td></tr></table></td><td><table><tr><td>Serbia</td><td>Europe</td><td>Serb</td></tr><tr><td>Slovenia</td><td>Europe</td><td>Serb</td></tr><tr><td>Croatia</td><td>Europe</td><td>Serb</td></tr><tr><td>Hungary</td><td>Europe</td><td>Serb</td></tr><tr><td>Montenegro</td><td>Europe</td><td>Serb</td></tr><tr><td>Macedonia</td><td>Europe</td><td>Serb</td></tr></table></td></tr>
<tr><td>48.</td><td>uzbek asia</td><td>Afghanistan<br>Tajikistan<br>Turkmenistan<br>Uzbekistan<br>Kazakstan<br>Kyrgyzstan</td><td><table><tr><td>Asia</td><td>Uzbek</td></tr><tr><td>Asia</td><td>Uzbek</td></tr><tr><td>Asia</td><td>Uzbek</td></tr></table>  </td><td>The expected answer should contain all countries with the uzbek ethny, but the keywords do not reflect this information.</td></tr>
<tr><td>49.</td><td>rhein germany</td><td>Baden Wurttemberg<br>Hessen<br>Nordrhein Westfalen<br>Rheinland Pfalz</td><td>-<br>  </td><td>The expected answer should be the provinces, but the keywords do not reflect this information.</td></tr>
</table>

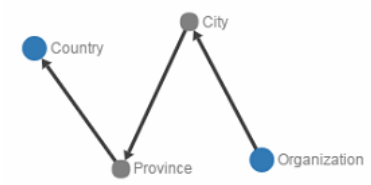
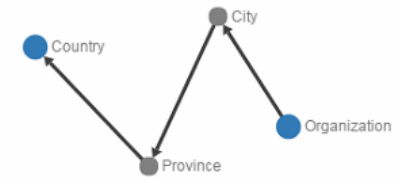
<table>
<tr>
<td>50.</td>
<td>egypt nile</td>
<td>Asyut<br>Beni Suef<br>El Giza<br>El Minya<br>El Qahira (munic.)</td>
<td>
<table>
<tr><td>Egypt</td><td>Nile</td></tr>
</table>


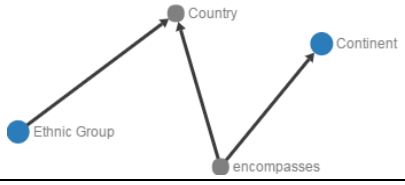
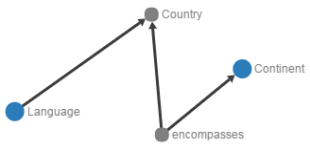
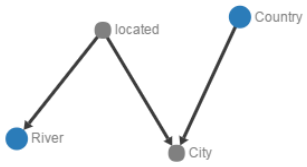
</td>
<td>If we added the keyword “City”, then the answer would be:
<table>
<tr><td>Egypt</td><td>Al Minya</td><td>Nile</td></tr>
<tr><td>Egypt</td><td>Al Qahirah</td><td>Nile</td></tr>
<tr><td>Egypt</td><td>Al Jizah</td><td>Nile</td></tr>
<tr><td>Egypt</td><td>Bani Suwayf</td><td>Nile</td></tr>
<tr><td>Egypt</td><td>Asyut</td><td>Nile</td></tr>
</table>
</td>
</tr>
</table>

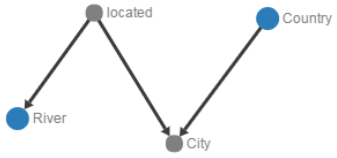
## 2.3 Results Obtained for the Version of the Mondial Dataset used in Coffman’s Benchmark

| # | Keywords | Expected Answer | Application Answer | Observation |
|---|---|---|---|---|
| 1. | thailand | country | Thailand (Country) | √ |
| 2. | netherlands | country | Netherlands (Country) | √ |
| 3. | georgia | country | Georgia (Country) | √ |
| 4. | country china | country | China (Country) | √ |
| 5. | bangladesh | country | Bangladesh (Country) | √ |
| 6. | alexandria | Alexandria, USA | Alexandria (Romania)<br>Alexandria (USA)<br>Class Country | √ |
| 7. | sonsonate | Sonsonate, ES | Sonsonate (City) | √ |
| 8. | xiaogan | Xiaogan, TJ | Xiaogan (City) | √ |
| 9. | city glendale | Glendale, USA | Glendale, Arizona (USA)<br>Glendale, California (USA)<br>Class Country | √ |
| 10. | city granada | Granada, E | Granada (**City**) | √ |
| 11. | Lake Kariba | Lake Kariba | Lake Kariba (**Lake**) | √ |
| 12. | Niger | Niger river | Niger (**Country**) | If the keyword “river” were added, then the expected answer would be returned. |
| 13. | Arabian Sea | Arabian Sea | Arabian Sea (**Sea**) | √ |
| 14. | Asauad | Asauad desert | Asauad (**Desert**) | √ |
| 15. | Sardegna | Sardegna (island) | Sardegna (**Island**) | √ |
| 16. | arab cooperation council | Arab Cooperation Council | • Council of Arab Economic Unity<br>• North Atlantic Cooperation Council<br>• Arab League<br>Others…<br>Class Organization | The expected value does not exist for property Name of class Organization. |
| 17. | world labor | World Confederation of Labor | • World Confederation of Labor<br>• World Food Program<br>Others...<br>Class Organization | Returns the expected value as well as other organizations. |
| 18. | islamic conference | Organization of the Islamic Conference | - | The version of Mondial adopted does not have any match for “Islamic” |
| 19. | 30 group | Group of 30 | Instâncias da classe **Ethnic Group** | Since the numeric value is not indexed, the match was with the label of class EthnicGroup.<br>If the user entered “Group of 30”, the expected answer would be returned. |
| 20. | caribbean economic | Economic Commission for Latin America and the Caribbean | • Economic and Social Council<br>• Economic Commission for Europe<br>• Council of Arab Economic Unity<br>Others…<br>Class Organization | The expected value does not exist for property Name of class Organization. |

| | | | | |
|---|---|---|---|---|
| 21. | slovakia hungary | borders | • Hungary<br>• Slovakia<br>Class Country | Keywords match the labels of two instances of class Country; but the keywords are not sufficient to infer that the question is about the borders between countries and, thus, were not correctly answered. |
| 22. | mongolia china | borders | • China<br>• Mongolia<br>class Country | |
| 23. | niger algeria | borders | • Niger<br>• Algeria<br>class Country | |
| 24. | kuwait saudi arabia | borders | • Saudi Arabia<br>• Kuwait<br>class Country | |
| 25. | lebanon syria | borders | • Syria<br>• Lebanon<br>class Country | |
| 26. | cameroon economy | Cameroon | Cameroon \| ECONOMY OF CAM<br>Country ← Economy | √ |
| 27. | nigeria gdp | Nigeria | Nigeria \| ECONOMY OF WAN \| 135900<br>Country ← Economy | √<br>Observe that the value in the previous table is different. |
| 28. | mongolia republic | Mongolia | • Mongolia<br>• Czech Republic<br>• Dominican Republic<br>• Central African Republic<br>class Country | The keywords match with labels of class Country |
| 29. | kiribati politics | Kiribati | KIR republic \| Kiribati<br>**Politics → Country** | √ |
| 30. | poland language | Poland | Poland \| Polish<br>**Country ← Language** | √ |
| 31. | spain galician | Spain | Galician \| Spain<br>**Language → Country** | √ |
| 32. | uzbekistan eastern orthodox | Uzbekistan | Eastern Orthodox \| Uzbekistan<br>**Religion → Country** | √ |
| 33. | haiti religion | Haiti | Protestant \| Haiti<br>Roman Catholic \| Haiti<br>**Religion → Country** | √ |
| 34. | suriname ethnic group | Suriname | Suriname \| African<br>Suriname \| Javanese<br>Suriname \| Creole<br>Suriname \| Amerindian<br>Suriname \| Hindustani<br>Suriname \| Chinese<br>Suriname \| Europeans<br>**Country ← Ethnic Group** | √ |
| 35. | slovakia german | Slovakia | -<br>**Country ← Language** | There is no value that relates "Slovakia" with "german" (Classes Language and EthnicGroup) |

| | | | | |
|---|---|---|---|---|
| 36. | poland cape verde organization | • Nonaligned Movement<br>• Customs Cooperation Council<br>• Food and Agriculture Organization<br>Others… | -<br>  <br> | The expected answer is the list of organizations that the countries belong to; however, the translation algorithm did not identify the IS_MEMBER class when generating the nucleuses. |
| 37. | saint kitts cambodia | • International Bank for Reconstruction and Development<br>• International Civil Aviation Organization<br>• International Criminal Police Organization<br>Others… | • Saint Kitts and Nevis<br>• Cambodia<br>• Saint Lucia<br>• Saint Vincent and the Grenadines<br>Classe **Country** | |
| 38. | marshall islands grenadines organization | • Food and Agriculture Organization<br>• Group of 77<br>• International Bank for Reconstruction and Development<br>Others… | -<br>  <br> | |
| 39. | czech republic cote divoire organization | • Customs Cooperation Council<br>• Food and Agriculture Organization<br>• International Atomic Energy Agency<br>Others… | -<br>  <br> | |
| 40. | panama oman | • Nonaligned Movement<br>• Food and Agriculture Organization<br>• Group of 77<br>Others… | • Oman<br>• Panama<br>Classe **Country** | |
| 41. | iceland mali | • Customs Cooperation Council<br>• Food and Agriculture Organization<br>• International Atomic Energy Agency<br>Others... | • Iceland<br>• Mali<br>Classe **Country** | |
| 42. | guyana sierra leone | • Commonwealth<br>• Nonaligned Movement<br>• African, Caribbean, and Pacific Countries<br>Others… | • Iceland<br>• Mali<br>Classe **Country** | |

<table>
<tr><td>43.</td><td>mauritius india</td><td>• Group of 77<br>• International Atomic Energy Agency<br>• International Bank for Reconstruction and Development<br>Others…</td><td>• Mauritius<br>• India<br>Classe Country</td><td rowspan="3"></td></tr>
<tr><td>44.</td><td>vanuatu afghanistan</td><td>• Nonaligned Movement<br>• Asian Development Bank<br>• Group of 77<br>Others…</td><td>• Vanuatu<br>• Afghanistan<br>Classe Country</td></tr>
<tr><td>45.</td><td>libya australia</td><td>• Nonaligned Movement<br>• Customs Cooperation Council<br>• Food and Agriculture Organization<br>Others…</td><td>• Libya<br>• Australia<br>Classe Country</td></tr>
<tr><td>46.</td><td>hutu africa</td><td>Rwanda</td><td><table><tr><td>Africa</td><td>Hutu</td></tr></table> 
</td><td>If we added the keyword “Country”, then the answer would be:<br>√<table><tr><td>Rwanda</td><td>Africa</td><td>Hutu</td></tr></table></td></tr>
<tr><td>47.</td><td>serb europe</td><td>Slovenia<br>Bosnia and Herzegovina<br>Croatia<br>Romania<br>Macedonia<br>Hungary</td><td><table><tr><td>Europe</td><td>Serb</td></tr><tr><td>Europe</td><td>Serb</td></tr><tr><td>Europe</td><td>Serb</td></tr><tr><td>Europe</td><td>Serb</td></tr><tr><td>Europe</td><td>Serb</td></tr><tr><td>Europe</td><td>Serb</td></tr></table></td><td>√<table><tr><td>Bosnia and Herzegovina</td><td>Europe</td><td>Serb</td></tr><tr><td>Slovenia</td><td>Europe</td><td>Serb</td></tr><tr><td>Croatia</td><td>Europe</td><td>Serb</td></tr><tr><td>Romania</td><td>Europe</td><td>Serb</td></tr><tr><td>Hungia</td><td>Europe</td><td>Serb</td></tr><tr><td>Macedonia</td><td>Europe</td><td>Serb</td></tr></table></td></tr>
<tr><td>48.</td><td>uzbek asia</td><td>Afghanistan<br>Tajikistan<br>Turkmenistan<br>Uzbekistan<br>Kazakstan<br>Kyrgyzstan</td><td><table><tr><td>Asia</td><td>Uzbek</td></tr><tr><td>Asia</td><td>Uzbek</td></tr></table> 
</td><td>The expected answer should contain all countries with the uzbek ethny, but the keywords do not reflect this information. If the keyword “country”, then the answer would be:<table><tr><td>Asia</td><td>Uzbek</td><td>Turkmenistan</td></tr><tr><td>Asia</td><td>Uzbek</td><td>Uzbekistan</td></tr></table></td></tr>
<tr><td>49.</td><td>rhein germany</td><td>Baden Wurttemberg<br>Hessen<br>Nordrhein Westfalen<br>Rheinland Pfalz</td><td>- 
</td><td>The expected answer should be the provinces, but the keywords do not reflect this information.</td></tr>
</table>

<table>
<tr>
<td>50.</td>
<td>egypt nile</td>
<td>Asyut<br>Beni Suef<br>El Giza<br>El Minya<br>El Qahira (munic.)</td>
<td>
<table>
<tr><td>Egypt</td><td>Nile</td></tr>
</table>


</td>
<td>If we added the keyword “City”, then the answer would be:
<table>
<tr><td>Egypt</td><td>Al Minya</td><td>Nile</td></tr>
<tr><td>Egypt</td><td>Al Qahirah</td><td>Nile</td></tr>
<tr><td>Egypt</td><td>Al Jizah</td><td>Nile</td></tr>
<tr><td>Egypt</td><td>Bani Suwayf</td><td>Nile</td></tr>
<tr><td>Egypt</td><td>Asyut</td><td>Nile</td></tr>
</table>
</td>
</tr>
</table>

## 3. Tests with the Full IMDb

### 3.1 Keyword Queries in Coffman’s Benchmark for IMDb

We run the queries in Coffman’s Benchmark for IMDb made available from a different source[4].

### 3.2 Results Obtained for the Full IMDb Dataset

**Queries 1-10**: consist of the name of movie stars. Relevant results contain a single tuple from the person relation that is the tuple of the specified individual.

1. denzel washington
2. clint eastwood
3. john wayne
4. will smith
5. harrison ford
6. julia Roberts
7. tom hanks
8. johnny depp
9. angelina jolie
10. morgan freeman

**Accuracy: 10 of 10**. The result contained more than one tuple, if the movie star’s name matched one of the keywords, but the top result was the expected actor.

**Queries 11-20**: consist of the name of movies. Relevant results contain a single tuple from the title relation that is the tuple of the specified film.

11. gone with the wind
12. star wars
13. casablanca
14. lord of the rings
15. the sound of music
16. wizard of oz
17. the notebook
18. forrest gump
19. the princess bride

[4] https://github.com/periclesoliveira/cnrank/tree/master/queryset/IMDb
https://github.com/periclesoliveira/cnrank/blob/master/queryset/IMDb/topics.txt

20. the godfather

**Accuracy: 9 of 10**. Again, the result contained more than one tuple, if the movie name matched one of the keywords, but the top result was the expected movie.

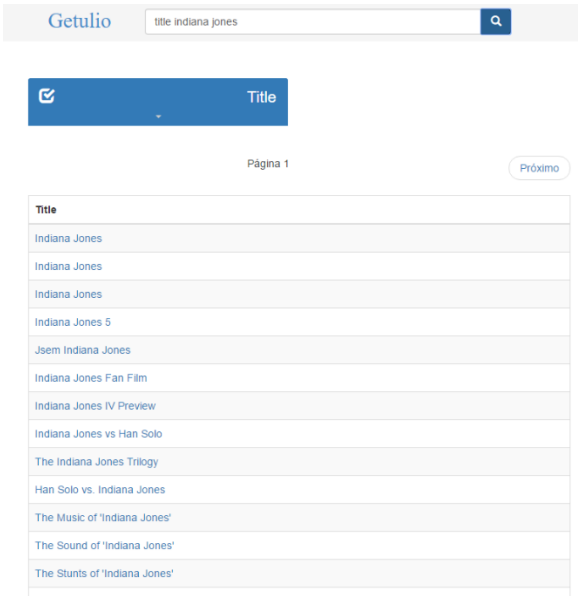
**Error in Query 13** – “casablanca”. “casablanca” is the name of a movie and of an actor; the score for both values was the same, but the algorithm returned the name of the actor, since the Actor class had a higher score than the Movies class. The movie name was the second generated query.

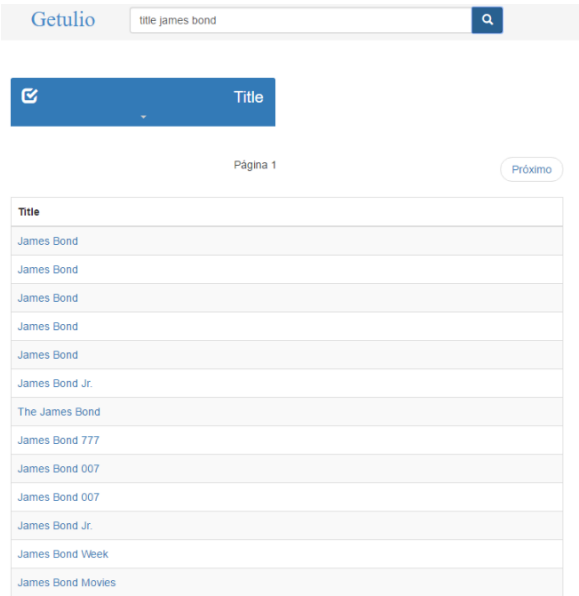
**Queries 21-30**: consist of the keyword “"title"” plus the name of film characters. Relevant results contain 3 tuples (1 from the char_name relation, 1 from the cast_info relation, and 1 from the title relation) that link the character to the film(s) in which s/he appears. (The keyword “title” is intentionally added to differentiate this group of topics from topics 1-20)

21. title atticus finch
22. title indiana jones
23. title james bond
24. title rick blaine
25. title will kane
26. title dr. hannibal lecter
27. title norman bates
28. title darth vader
29. title the wicked witch of the west
30. title nurse ratched

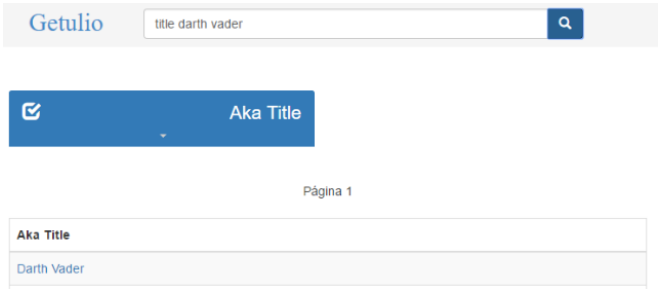
**Accuracy: 7 of 10**. Again, the result contained more than one tuple.

**Error in Queries 22, 23**. The name of the character is part of the name of some title. The nucleus with class Title contained all keywords and had the best score. The answers of the algorithm were the titles with the character names.



**Error in Query 28.** In this case, the class AKA_TITLE has “darth vader” in one of its values. This nucleus was the best scored because the label of the class had the keyword “title” and “darth vader” as a value. Class Title only matched the keyword “title” and class char_name only matched “darth vader”.




**Queries 31-35**: consist of the keyword “'title'” plus a film quote. Relevant results contain 2 tuples (1 from the movie_info relation and 1 from the title relation) that link the movie quote to the film in which it appears. (The keyword “title” is intentionally added so that relevant results answer the question "In which film does this quote appear?".) Note that a quote may appear in multiple films.

31. title frankly my dear i don't give a damn
32. title i'm going to make him an offer he can't refuse
33. title you don't understand i coulda had class i could been a contender i coulda been somebody instead of a bum which is what i am
34. title toto i've a feeling we're not in kansas any more
35. title here's looking at you kid

**Accuracy: 4 of 5**. The result was not a single tuple, as in previous blocks.

**Error in Query 32**: The quotes were not in the dataset used for the tests.

**Query 36.** Relevant results must denote the films in which the actor Mark Hamill plays the character Luke Skywalker.

36. mark hamill luke skywalker

**Accuracy: 1 of 1**

**Query 37**: Relevant results contain 3 tuples (name <- cast_info -> title) that must denote all films in which the actor Tom Hanks appeared in the year 2004.

37. tom hanks 2004

**Accuracy: 1 of 1**

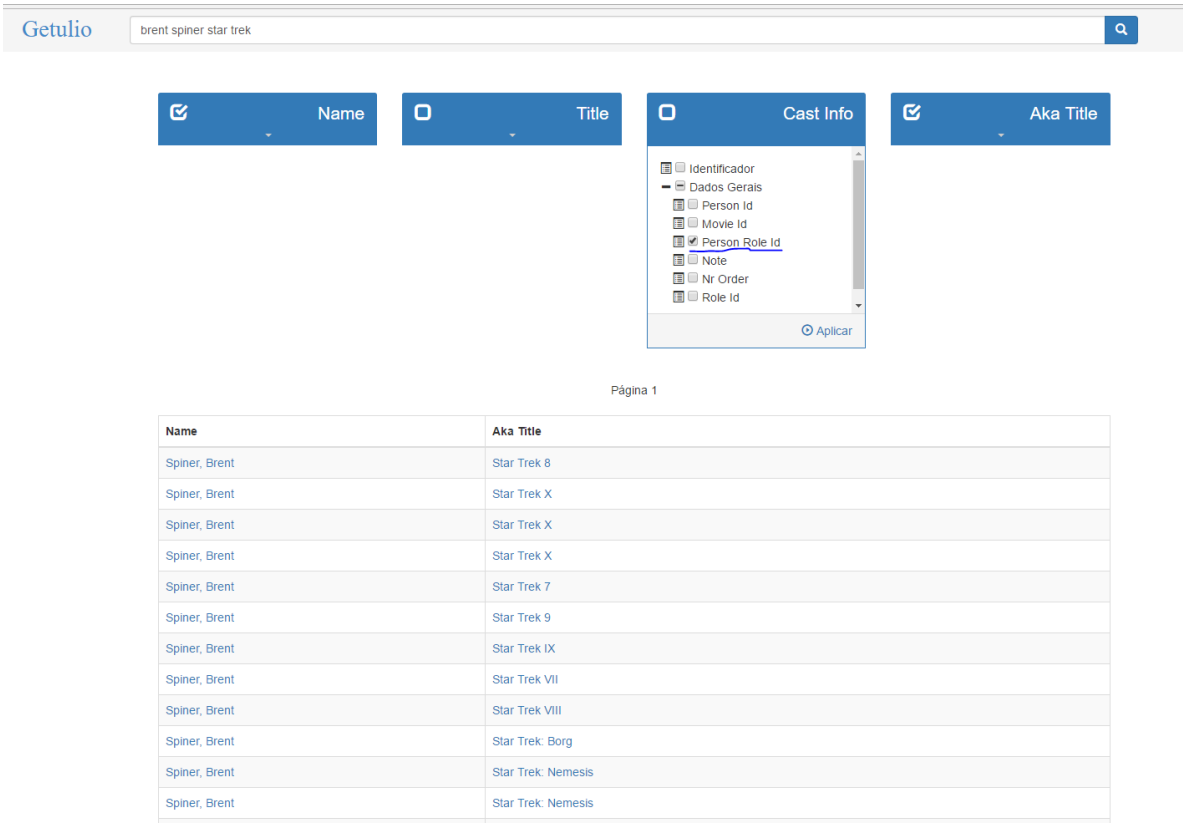
**Queries 38-40**: Relevant results must denote the character that an actor plays in a film.

38. henry fonda yours mine ours char_name
39. russell crowe gladiator char_name
40. brent spiner star trek

**Accuracy: 1 of 3**

**Error in Queries 38 and 39:** There are values in char_name that match “Henry Fonda” and “Russell Crowe”. The algorithm assumed that the query was about these character names and tested with the movie name.

Note: The query 30 return the actor, the movie and the cast_info, with the cast info we can access to the char_name.




**Query 41**: Relevant results contain 3 tuples (name <- cast_info -> title) that must denote all films in which the actor Audry Hepburn appeared in the year 1951.

41. audrey hepburn 1951

**Accuracy: 0 of 1**

**Error**: The nucleus with Title covered all three keywords since there is a film whose name matches “Audrey Hepburn” and whose production year matches 1951.

**Queries 42-43**: A relevant result must identify an actor who plays Jacques Clouseau in a movie.

42. name jacques clouseau
43. name jack ryan

**Accuracy: 0 of 2**

**Error:** The algorithm found only the nucleus with class char_name, the character name matched with property name, and the keyword “name” matched with the label of the nucleus.

**Query 44**: Relevant results must denote a film in which Sylvester Stallone plays the character Rocky. Note that because of limitations of existing systems, relevant results are *not* required to include the appropriate tuple from the title relation (which would prevent any system from identifying a single relevant result).

44. rocky stallone

**Accuracy: 0 of 1**

**Error:** the keywords are very ambiguous. The algorithm found both keywords in a PERSON_INFO#INFO value.

**Query 45**: A relevant result must identify an actor who plays "The Terminator".

45. name terminator

**Accuracy: 0 of 1**

**Error:** same as for Queries 42-43.

**Queries 46-49**: Relevant results identify relationships (through the title relation) between an actor and another class, such as “harrison ford george lucas”.

46. harrison ford george lucas
47. sean connery ian fleming
48. keanu reeves wachowski
49. dean jones herbie

**Accuracy: 3 of 4**

**Error in Query 48**: “wachowski” only had matches in the AKA_NAME class.

**Query 50**: Relevant results identify cast members in common between the films "Raiders of the Lost Ark" and "Indiana Jones and the Last Crusade."

50. indiana jones last crusade lost ark

**Accuracy: 0 of 1**

**Error**: The algorithm did not return the actors that both movies had in common, but returned the movies themselves.

## References


[1] García, G., Izquierdo, Y., Menendez, E., Dartayre, F., and Casanova, M. RDF Keyword-based Query Technology Meets a Real-World Dataset” (submitted for publication).

[2] Coffman, J. and Weaver, A. 1999. An empirical performance evaluation of relational keyword search techniques. TKDE 1999.