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)

# Introduction

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

* a triplified version of the full Mondial dataset[[1]](#footnote-1)
* a triplified version of the Mondial dataset adopted in Coffman’s benchmark[[2]](#footnote-2) [2]
* a triplified version of the full IMDb dataset[[3]](#footnote-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** |

# Tests with Mondial

## 2.1 Keyword Queries in Coffman’s Benchmark for Mondial2

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** |
|  | thailand | country | Thailand (Country) | √ |
|  | netherlands | country | Netherlands (Country) | √ |
|  | georgia | country | Georgia (Country) | √ |
|  | country china | country | China (Country) | √ |
|  | bangladesh | country | Bangladesh (Country) | √ |
|  | alexandria | Alexandria, USA | Alexandria (Romania)  Alexandria (USA) | There are 2 cities named “Alexandria” |
|  | sonsonate | Sonsonate, ES | - | The version of Mondial adopted does not have a city called “Sonsonate” |
|  | xiaogan | Xiaogan, TJ | Xiaogan (China) | √  The value of Code for “China” in class Country is **TJ.** |
|  | city glendale | Glendale, USA | Glendale, Arizona (USA)  Glendale, California (USA) | √ |
|  | city granada | Granada, E | Granada, Andalucía (ES)  Granada, Nicaragua (NIC) | √ |
|  | Lake Kariba | Lake Kariba | Lake Kariba | √ |
|  | Niger | Niger river | Niger (Country) | “Niger” is both a country and a river |
|  | Arabian Sea | Arabian Sea | Arabian Sea | √ |
|  | Asauad | Asauad desert | - | The version of Mondial adopted does not have a desert called “Asauad” |
|  | Sardegna | Sardegna (island) | Sardegna (Island) | √ |
|  | 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). |
|  | world labor | World Confederation of Labor | Answer with 9 instances of class Organization |
|  | islamic conference | Organization of the Islamic Conference | * Islamic Development Bank * Organization of Islamic Cooperation * United Nations Conference on Trade and Development * Conference of Interaction and Confidence-Building Measures in Asia   Class Organization |
|  | 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. |
|  | caribbean economic | Economic Commission for Latin America and the Caribbean | * Organization for Economic Cooperation and Development * Economic and Monetary Union   Class Organization | The expected value does not exist for property Name of class Organization. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | slovakia hungary | borders | * Hungary * Slovakia   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. |
|  | mongolia china | borders | * China * Mongolia   Class Country |
|  | niger algeria | borders | * Niger * Algeria   Class Country |
|  | kuwait saudi arabia | borders | * Saudi Arabia * Kuwait   Class Country |
|  | lebanon syria | borders | * Syria * Lebanon   Class **Country** |
|  | cameroon economy | Cameroon | |  |  | | --- | --- | | Cameroon | ECONOMY OF CAM |   Country ← Economy | √ |
|  | nigeria gdp | Nigeria | |  |  |  | | --- | --- | --- | | Nigeria | ECONOMY OF WAN | 502000 |   Country ← Economy | √ |
|  | mongolia republic | Mongolia | * Mongolia * Czech Republic * Dominican Republic * Central African Republic   Class Country | As keywords match with labels of class Country |
|  | kiribati politics | Kiribati | |  |  | | --- | --- | | Kiribati | POLITICS OF KIR |   Country ← Politics | √ |
|  | poland language | Poland | |  |  | | --- | --- | | Poland | Polish |   Country ← Language | √ |
|  | spain galician | Spain | |  |  | | --- | --- | | Spain | Galician |   Country ← Language | √ |
|  | uzbekistan eastern orthodox | Uzbekistan | - | “eastern orthodox” is not a value of property Name of class Religion. |
|  | haiti religion | Haiti | |  |  | | --- | --- | | Haiti | Protestant | | Haiti | Roman Catholic |   Country ← Religion | √ |
|  | suriname ethnic group | Suriname | |  |  | | --- | --- | | Suriname | African | | Suriname | Javanese | | Suriname | Creole | | Suriname | Amerindian | | Suriname | Hindustani | | Suriname | Chinese | | Suriname | Europeans |   Country ← Ethnic Group | √ |
|  | slovakia german | Slovakia | |  |  | | --- | --- | | Slovakia | German |   Country ← Ethnic Group | √ |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | poland cape verde organization | * Nonaligned Movement * Customs Cooperation Council * Food and Agriculture Organization   Others… | - | 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. |
|  | saint kitts cambodia | * International Bank for Reconstruction and Development * International Civil Aviation Organization * International Criminal Police Organization   Others… | * Saint Kitts and Nevis * Cambodia * Saint Lucia * Saint Martin * Saint Helena * Saint Barthelemy * Saint Pierre and Miquelon * Saint Vincent and the Grenadines   Class Country |
|  | marshall islands grenadines organization | * Food and Agriculture Organization * Group of 77 * International Bank for Reconstruction and Development   Others… | - |
|  | czech republic cote divoire organization | * Customs Cooperation Council * Food and Agriculture Organization * International Atomic Energy Agency   Others… |  | Only the “czech republic” is returned and appears in the nucleus with class EthnicGroup |
|  | panama oman | * Nonaligned Movement * Food and Agriculture Organization * Group of 77   Others… | * Oman * Panama   Class Country | The expected answer is the list of organizations that the countries belong to; however, this is difficult to infer from the keywords. |
|  | iceland mali | * Customs Cooperation Council * Food and Agriculture Organization * International Atomic Energy Agency   Others... | * Iceland * Mali   Class Country |
|  | guyana sierra leone | * Commonwealth * Nonaligned Movement * African, Caribbean, and Pacific Countries   Others… | * Iceland * Mali   Class Country |
|  | mauritius india | * Group of 77 * International Atomic Energy Agency * International Bank for Reconstruction and Development   Others… | * Mauritius * India   Class Country |
|  | vanuatu afghanistan | * Nonaligned Movement * Asian Development Bank * Group of 77   Others… | * Vanuatu * Afghanistan   Class Country |
|  | libya australia | * Nonaligned Movement * Customs Cooperation Council * Food and Agriculture Organization   Others… | * Libya * Australia   Class Country |
|  | hutu africa | Rwanda | |  |  | | --- | --- | | Africa | Hutu | | If we added the keyword “Country”, then the answer would be:   |  |  |  | | --- | --- | --- | | Rwanda | Africa | Hutu | |
|  | serb europe | Slovenia  Bosnia and Herzegovina  Croatia  Romania  Macedonia  Hungary | |  |  | | --- | --- | | Europe | Serb | | Europe | Serb | | Europe | Serb | | Europe | Serb | | Europe | Serb | | Europe | Serb | | |  |  |  | | --- | --- | --- | | Serbia | Europe | Serb | | Slovenia | Europe | Serb | | Croatia | Europe | Serb | | Hungary | Europe | Serb | | Montenegro | Europe | Serb | | Macedonia | Europe | Serb | |
|  | uzbek asia | Afghanistan  Tajikistan  Turkmenistan  Uzbekistan  Kazakstan  Kyrgyzstan | |  |  | | --- | --- | | Asia | Uzbek | | Asia | Uzbek | | Asia | Uzbek | | The expected answer should contain all countries with the uzbekethny, but the keywords do not reflect this information. |
|  | rhein germany | Baden Wurttemberg  Hessen  Nordrhein Westfalen  Rheinland Pfalz | - | The expected answer should be the provinces, but the keywords do not reflect this information. |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | egypt nile | Asyut  Beni Suef  El Giza  El Minya  El Qahira (munic.) | |  |  | | --- | --- | | Egypt | Nile | | If we added the keyword “City”, then the answer would be:   |  |  |  | | --- | --- | --- | | Egypt | Al Minya | Nile | | Egypt | Al Qahirah | Nile | | Egypt | Al Jizah | Nile | | Egypt | Bani Suwayf | Nile | | Egypt | Asyut | Nile | |

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **#** | **Keywords** | **Expected Answer** | **Application Answer** | **Observation** |
|  | thailand | country | Thailand (Country) | √ |
|  | netherlands | country | Netherlands (Country) | √ |
|  | georgia | country | Georgia (Country) | √ |
|  | country china | country | China (Country) | √ |
|  | bangladesh | country | Bangladesh (Country) | √ |
|  | alexandria | Alexandria, USA | Alexandria (Romania)  Alexandria (USA)  Class Country | √ |
|  | sonsonate | Sonsonate, ES | Sonsonate (City) | √ |
|  | xiaogan | Xiaogan, TJ | Xiaogan (City) | √ |
|  | city glendale | Glendale, USA | Glendale, Arizona (USA)  Glendale, California (USA)  Class Country | √ |
|  | city granada | Granada, E | Granada (**City**) | √ |
|  | Lake Kariba | Lake Kariba | Lake Kariba (**Lake**) | √ |
|  | Niger | Niger river | Niger (**Country**) | If the keyword “river” were added, then the expected answer would be returned. |
|  | Arabian Sea | Arabian Sea | Arabian Sea (**Sea**) | √ |
|  | Asauad | Asauad desert | Asauad (**Desert**) | √ |
|  | Sardegna | Sardegna (island) | Sardegna (**Island**) | √ |
|  | arab cooperation council | Arab Cooperation Council | * Council of Arab Economic Unity * North Atlantic Cooperation Council * Arab League   Others…  Class Organization | The expected value does not exist for property Name of class Organization. |
|  | world labor | World Confederation of Labor | * World Confederation of Labor * World Food Program   Others...  Class Organization | Returns the expected value as well as other organizations. |
|  | islamic conference | Organization of the Islamic Conference | - | The version of Mondial adopted does not have any match for “Islamic” |
|  | 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.  If the user entered “Group of 30”, the expected answer would be returned. |
|  | caribbean economic | Economic Commission for Latin America and the Caribbean | * Economic and Social Council * Economic Commission for Europe * Council of Arab Economic Unity   Others…  Class Organization | The expected value does not exist for property Name of class Organization. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | slovakia hungary | borders | * Hungary * Slovakia   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. |
|  | mongolia china | borders | * China * Mongolia   class Country |
|  | niger algeria | borders | * Niger * Algeria   class Country |
|  | kuwait saudi arabia | borders | * Saudi Arabia * Kuwait   class Country |
|  | lebanon syria | borders | * Syria * Lebanon   class Country |
|  | cameroon economy | Cameroon | |  |  | | --- | --- | | Cameroon | ECONOMY OF CAM |   Country ← Economy | √ |
|  | nigeria gdp | Nigeria | |  |  |  | | --- | --- | --- | | Nigeria | ECONOMY OF WAN | 135900 |   Country ← Economy | √  Observe that the value in the previous table is different. |
|  | mongolia republic | Mongolia | * Mongolia * Czech Republic * Dominican Republic * Central African Republic   class Country | The keywords match with labels of class Country |
|  | kiribati politics | Kiribati | |  |  | | --- | --- | | KIR republic | Kiribati |   **Politics** → **Country** | √ |
|  | poland language | Poland | |  |  | | --- | --- | | Poland | Polish |   **Country** ← **Language** | √ |
|  | spain galician | Spain | |  |  | | --- | --- | | Galician | Spain |   **Language** → **Country** | √ |
|  | uzbekistan eastern orthodox | Uzbekistan | |  |  | | --- | --- | | Eastern Orthodox | Uzbekistan |   **Religion** → **Country** | √ |
|  | haiti religion | Haiti | |  |  | | --- | --- | | Protestant | Haiti | | Roman Catholic | Haiti |   **Religion** → **Country** | √ |
|  | suriname ethnic group | Suriname | |  |  | | --- | --- | | Suriname | African | | Suriname | Javanese | | Suriname | Creole | | Suriname | Amerindian | | Suriname | Hindustani | | Suriname | Chinese | | Suriname | Europeans |   **Country** ← **Ethnic** **Group** | √ |
|  | slovakia german | Slovakia | -  **Country** ← **Language** | There is no value that relates “Slovakia” with “german” (Classes Language and EthnicGroup) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | poland cape verde organization | * Nonaligned Movement * Customs Cooperation Council * Food and Agriculture Organization   Others… | - | 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. |
|  | saint kitts cambodia | * International Bank for Reconstruction and Development * International Civil Aviation Organization * International Criminal Police Organization   Others… | * Saint Kitts and Nevis * Cambodia * Saint Lucia * Saint Vincent and the Grenadines   Classe **Country** |
|  | marshall islands grenadines organization | * Food and Agriculture Organization * Group of 77 * International Bank for Reconstruction and Development   Others… | - |
|  | czech republic cote divoire organization | * Customs Cooperation Council * Food and Agriculture Organization * International Atomic Energy Agency   Others… | - |
|  | panama oman | * Nonaligned Movement * Food and Agriculture Organization * Group of 77   Others… | * Oman * Panama   Classe **Country** |
|  | iceland mali | * Customs Cooperation Council * Food and Agriculture Organization * International Atomic Energy Agency   Others... | * Iceland * Mali   Classe **Country** |
|  | guyana sierra leone | * Commonwealth * Nonaligned Movement * African, Caribbean, and Pacific Countries   Others… | * Iceland * Mali   Classe **Country** |
|  | mauritius india | * Group of 77 * International Atomic Energy Agency * International Bank for Reconstruction and Development   Others… | * Mauritius * India   Classe **Country** |
|  | vanuatu afghanistan | * Nonaligned Movement * Asian Development Bank * Group of 77   Others… | * Vanuatu * Afghanistan   Classe **Country** |
|  | libya australia | * Nonaligned Movement * Customs Cooperation Council * Food and Agriculture Organization   Others… | * Libya * Australia   Classe **Country** |
|  | hutu africa | Rwanda | |  |  | | --- | --- | | Africa | Hutu | | If we added the keyword “Country”, then the answer would be:  √   |  |  |  | | --- | --- | --- | | Rwanda | Africa | Hutu | |
|  | serb europe | Slovenia  Bosnia and Herzegovina  Croatia  Romania  Macedonia  Hungary | |  |  | | --- | --- | | Europe | Serb | | Europe | Serb | | Europe | Serb | | Europe | Serb | | Europe | Serb | | Europe | Serb | | √   |  |  |  | | --- | --- | --- | | Bosnia and Herzegovina | Europe | Serb | | Slovenia | Europe | Serb | | Croatia | Europe | Serb | | Romania | Europe | Serb | | Hungia | Europe | Serb | | Macedonia | Europe | Serb | |
|  | uzbek asia | Afghanistan  Tajikistan  Turkmenistan  Uzbekistan  Kazakstan  Kyrgyzstan | |  |  | | --- | --- | | Asia | Uzbek | | Asia | Uzbek | | The expected answer should contain all countries with the uzbekethny, but the keywords do not reflect this information. If the keyword “country”, then the answer would be:   |  |  |  | | --- | --- | --- | | Asia | Uzbek | Turkmenistan | | Asia | Uzbek | Uzbekistan | |
|  | rhein germany | Baden Wurttemberg  Hessen  Nordrhein Westfalen  Rheinland Pfalz | - | The expected answer should be the provinces, but the keywords do not reflect this information. |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | egypt nile | Asyut  Beni Suef  El Giza  El Minya  El Qahira (munic.) | |  |  | | --- | --- | | Egypt | Nile | | If we added the keyword “City”, then the answer would be:   |  |  |  | | --- | --- | --- | | Egypt | Al Minya | Nile | | Egypt | Al Qahirah | Nile | | Egypt | Al Jizah | Nile | | Egypt | Bani Suwayf | Nile | | Egypt | Asyut | Nile | |

# 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]](#footnote-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.

1. gone with the wind
2. star wars
3. casablanca
4. lord of the rings
5. the sound of music
6. wizard of oz
7. the notebook
8. forrest gump
9. the princess bride
10. 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.

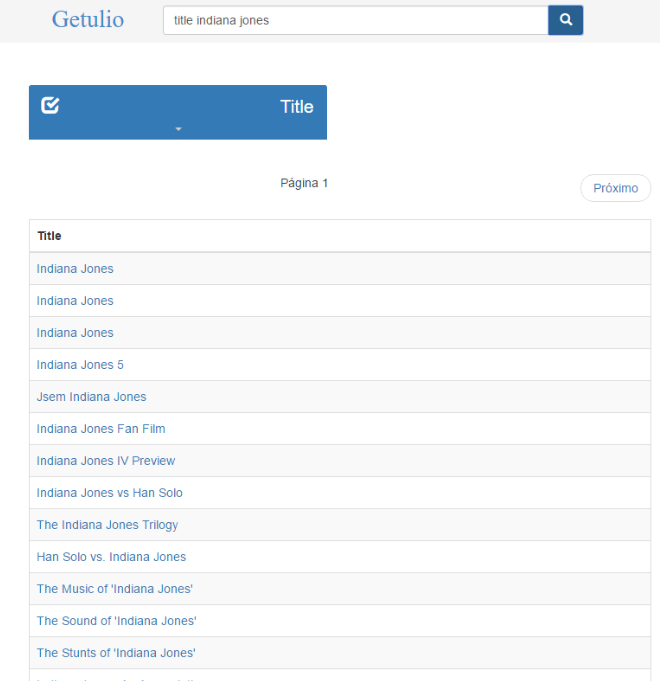
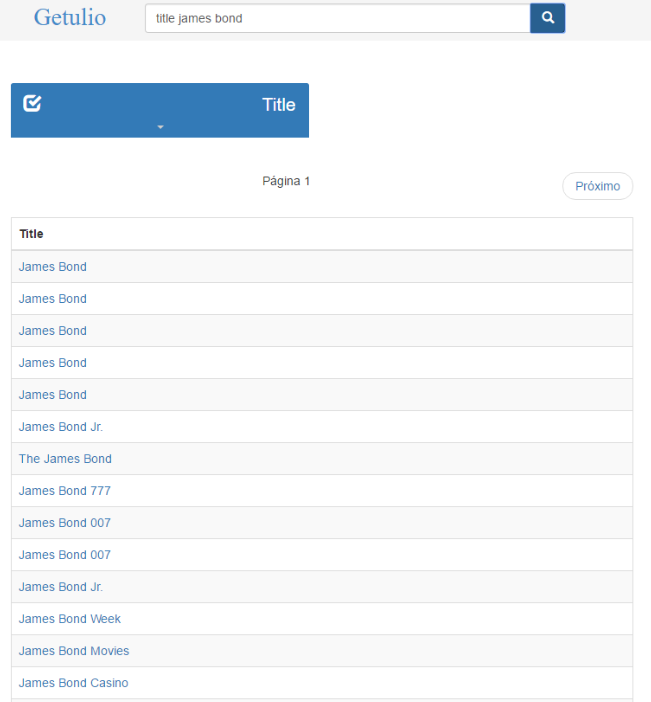
**Error in Query 13** – “casablanca”. “casablanca” is the name of a movie and of an actor; the score for both values was the same, but the algorithm returned the name of the actor, since the Actor class had a higher score than the Movies class. The movie name was the second generated query.

**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)

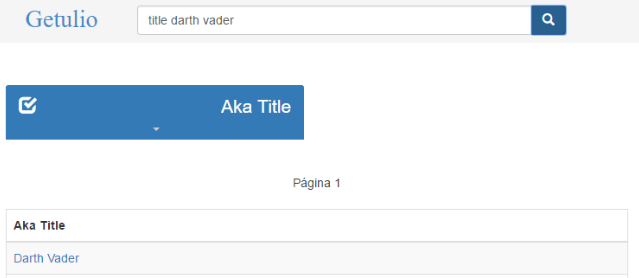
1. title atticus finch
2. title indiana jones
3. title james bond
4. title rick blaine
5. title will kane
6. title dr. hannibal lecter
7. title norman bates
8. title darth vader
9. title the wicked witch of the west
10. title nurse ratched

**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.

1. title frankly my dear i don't give a damn
2. title i'm going to make him an offer he can't refuse
3. 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
4. title toto i've a feeling we're not in kansas any more
5. 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.

1. 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.

1. tom hanks 2004

**Accuracy:** **1 of 1**

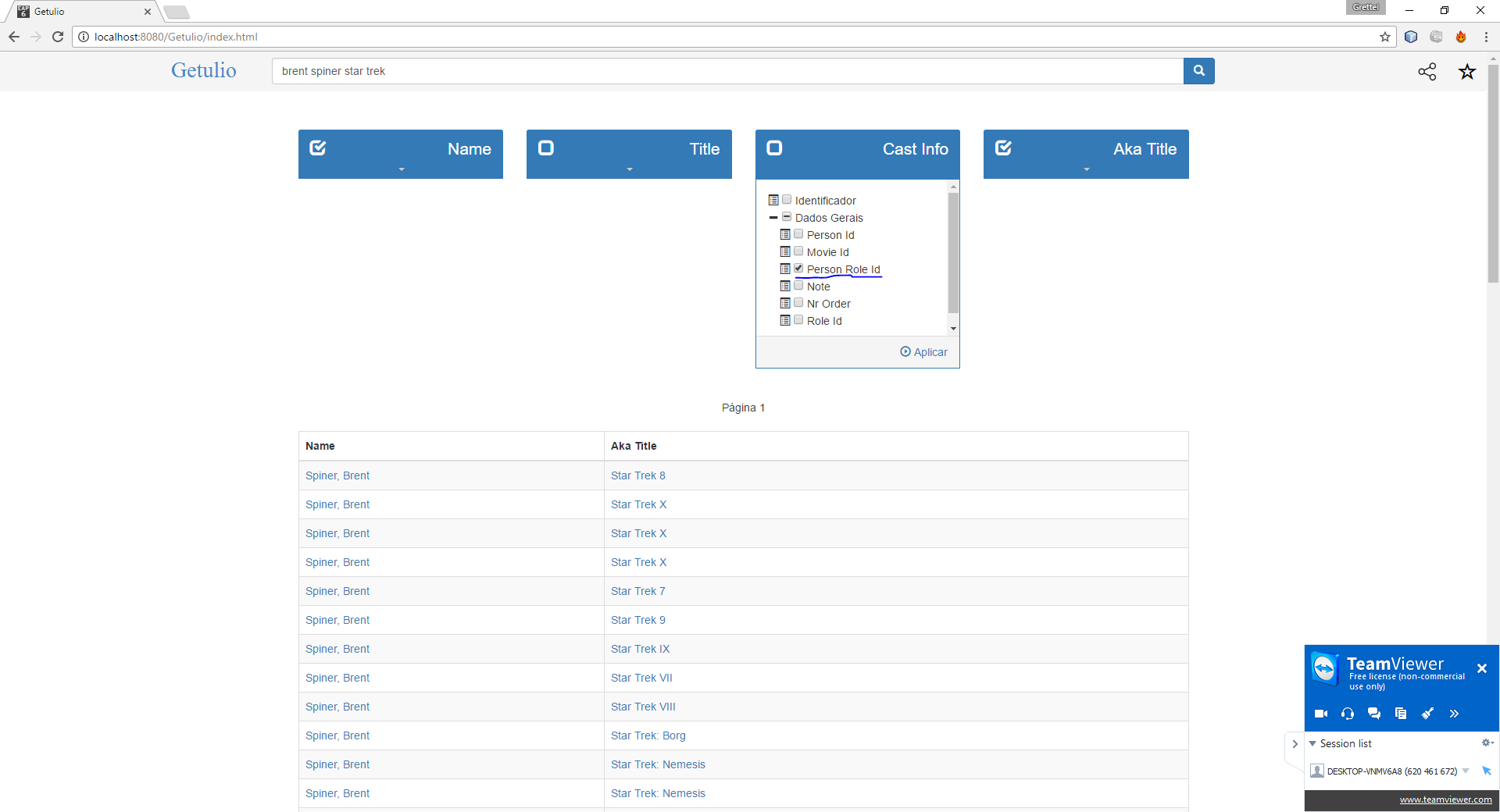
**Queries 38-40**: Relevant results must denote the character that an actor plays in a film.

1. henry fonda yours mine ours char\_name
2. russell crowe gladiator char\_name
3. 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.

1. 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.

1. name jacques clouseau
2. 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).

1. 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".

1. 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”.

1. harrison ford george lucas
2. sean connery ian fleming
3. keanu reeves wachowski
4. 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."

1. 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.

1. https://www.dbis.informatik.uni-goettingen.de/Mondial/ [↑](#footnote-ref-1)
2. https://www.cs.virginia.edu/~jmc7tp/resources.php [↑](#footnote-ref-2)
3. https://sites.google.com/site/ontopiswc13/home/imdb-mo [↑](#footnote-ref-3)
4. https://github.com/periclesoliveira/cnrank/tree/master/queryset/IMDb

   https://github.com/periclesoliveira/cnrank/blob/master/queryset/IMDb/topics.txt [↑](#footnote-ref-4)