

INFOMKDE

Student Projects

Mel Chekol

Student Projects

- Goals
 - Develop practical experience with the Semantic Web technologies
 - Become familiar with existing datasets
 - Understand possibilities and limitations of Semantic Web technologies
- Objective
 - Design and implement solutions for real-world problems by leveraging the power of Semantic Web technologies
 - Demonstrate the practical application of Semantic Web principles to enhance data integration, interoperability, and knowledge representation in diverse domains.



Interesting applications

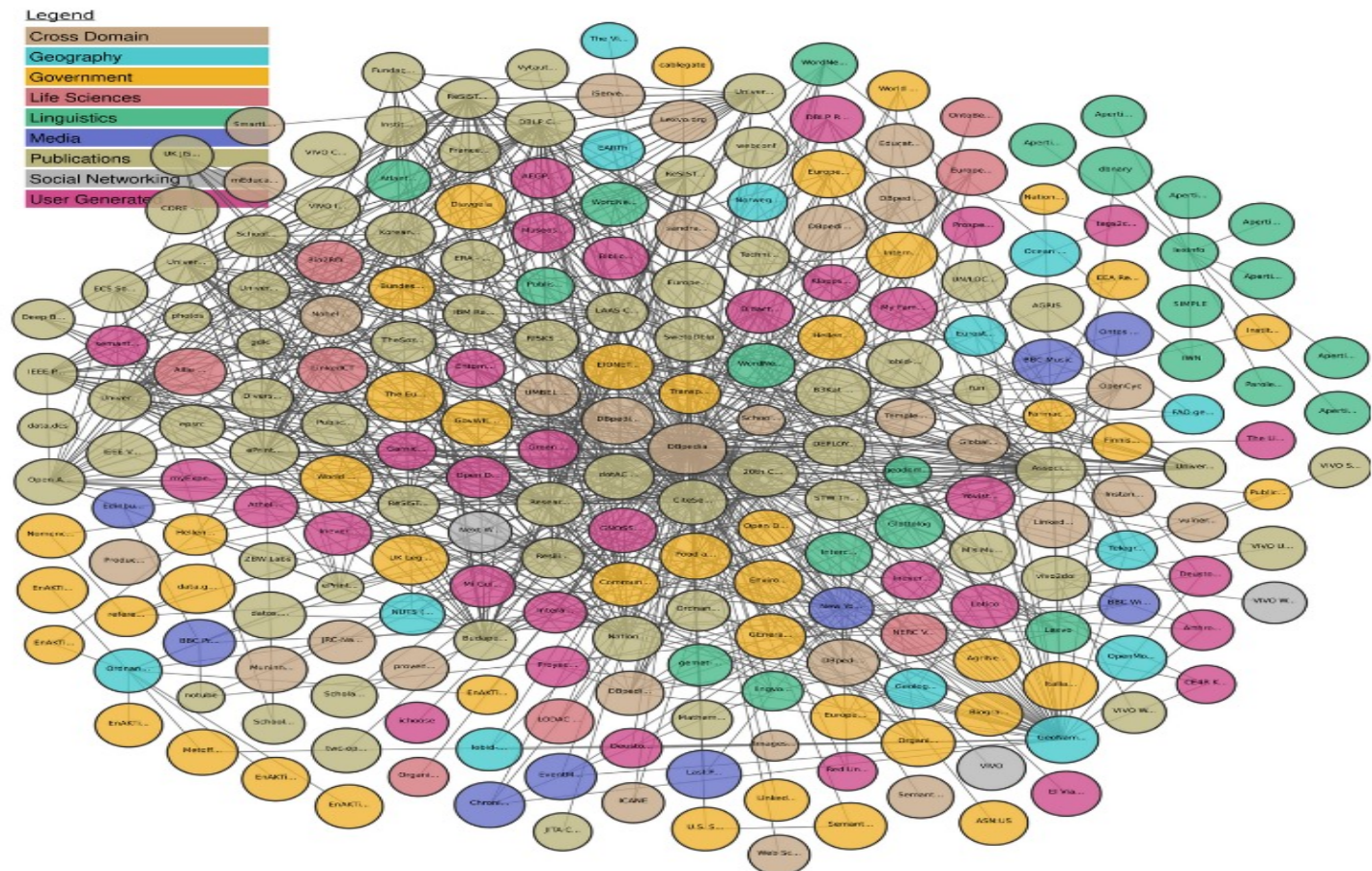
- Automatically transforming structured, unstructured or semi-structured data into RDF, OWL
- Question answering systems
- Data integration
- A simple recommender system
- Geospatial search and retrieval
- Translating SPARQL into Gremlin, cypher
-

Use-case: Datasets



Utrecht University

Wikidata
DBpedia
YAGO



The Linked Open Data Cloud from lod-cloud.net

<https://www.lod-cloud.net/>



Use-case: Applications

- Siri
- Google knowledge graph [1]
- Microsoft's satori [1]
- Facebook's entity graph [1]
- ...

Use-case: Applications



Utrecht University

Bing Infobox

Microsoft Bing

Rafael Bombelli

ALL IMAGES VIDEOS MAPS NEWS SHOPPING

64,200 Results Date Open links in new tab

Rafael Bombelli - Wikipedia
https://en.wikipedia.org/wiki/Rafael_Bombelli

Rafael Bombelli (baptised on 20 January 1526; died 1572) was an **Italian mathematician**. Born in Bologna, he is the author of a treatise on algebra and is a central figure in the understanding of imaginary numbers. He was the one who finally managed to address the problem with imaginary numbers. In his 1572 book, *L'Algebra*, Bombelli solved equations using the method of del ... +

< Life Bombelli's Algebra Reputation Bombelli's method of calc >

Rafael Bombelli was baptised on 20 January 1526 in Bologna, Papal States. He was born to Antonio Mazzoli, a wool merchant, and Diamante Scudieri, a tailor's daughter. The Mazzoli family was once quite powerful in Bologna. When Pope Julius II came to power, in 1506, he exiled the ruling family, the Bentivoglios. The Bentivoglio family attempted to retake Bologna in 1508, but failed. R... +

Wikipedia · Text under CC-BY-SA license

Rafael Bombelli (1526 - 1572) - Biography - MacTutor ...
<https://mathshistory.st-andrews.ac.uk/Biographies/Bombelli>

Biography Rafael Bombelli's father was Antonio Mazzoli but he changed his name from Mazzoli to Bombelli. It is perhaps worth giving a little family background. The Bentivoglio family ruled over Bologna from 1443. Sante Bentivoglio was "signore" (meaning lord) of Bologna from 1443 and he was succeeded by Giovanni II Bentivoglio who improved the city of Bologna, in particular ...

Images of Rafael Bombelli
<bing.com/images>

Rafael Bombelli

Mathematician



Rafael Bombelli was an Italian mathematician. Born in Bologna, he is the author of a treatise on algebra and is a central figure in the understanding of imaginary numbers.

Wikipedia

Born: 20 jan. 1526 · Bologna, Province of Bologna

Died: 1572 · Rome, Italy

Education: University of Bologna

Field of study: Algebra

People also search for

See all (15+)



Gerolamo Cardano



Niccolò Fontana Tartaglia



François Viète



Hero of Alexandria



Leonhard Euler

Data from: Wikipedia · Freebase · Tumblr

Suggest an edit

Use-case: Applications



buys ballot



ALL IMAGES VIDEOS MAPS NEWS SHOPPING

9.390.000 Results

Date ▾

Open links in new tab



Christophorus Buys Ballot - Wikipedia

https://nl.wikipedia.org/wiki/Christophorus_Buys_Ballot ▾

Christophorus Henricus Dedericus Buis Ballot (schrijfwijze op de geboorteakte), beter bekend als Buys Ballot, (Kloetinge, 10 oktober 1817 – Utrecht, 3 februari 1890) was **een Nederlandse meteoroloog, scheikundige en natuurkundige**. Buys Ballot ontdekte een meteorologische wet voor de windrichting die naar hem werd genoemd: de Wet van Buys Ballot. Hij was een va... +



< **Achtergrond** Opleiding en wetenschappel... Buys Ballot-medaille >

Buys Ballot werd in 1817 in Kloetinge geboren als zoon van Antonie Jacobus Buys Ballot, die sinds 1816 predikant was in Kloetinge; daarvoor was hij dominee geweest in Standdaarbuiten. In 1820 vertrok het gezin Buys Ballot al naar de volgende standplaats, Sint-Laurens op Walcheren.

Wikipedia · Text under CC-BY-SA license

Buys Ballotgebouw - Vastgoed & Campus - Universiteit Utrecht

<https://www.uu.nl/buys-ballotgebouw> ▾

Het **Buys Ballotgebouw** heeft zowel werkgroep ruimten als enkele computerleerzalen. De onderwijsruimten bevinden zich hoofdzakelijk op de begane grond, de eerste verdieping en tweede verdieping. Het gebouw heeft geen vrij toegankelijke entree. Toegang zonder toegangspas is mogelijk via de eerste verdieping van het Koningsbergergebouw.

C. H. D. Buys Ballot



Christophorus Henricus Diedericus Buys Ballot ([ˈbœys ˈbalot]; October 10, 1817 – February 3, 1890) was a Dutch chemist and meteorologist after whom Buys Ballot's law and the Buys Ballot table are named. He was first chairman of the International Meteorological Organization, ... +

Wikipedia

Born: October 10, 1817 · Kloetinge, Netherlands

Died: February 3, 1890 (aged 72) · Utrecht, Netherlands

Nationality: Dutch

Alma mater: Utrecht University

Known for: Buys-Ballot's law

Fields: chemistry · meteorology

Life ▾

Rank ▾

Religion ▾

Born ▾

Members of the Royal Netherlands Academy of Arts and Sciences

See all (30+)



Jan Terlouw



Adolphe Quetelet



Frans de Waal




Joseph Dalton Hooker








Justus von Liebig

Use-case: query service

- Rock bands that start with 'K'

 Wikidata Query Service

ExamplesQuery BuilderHelpMore tools



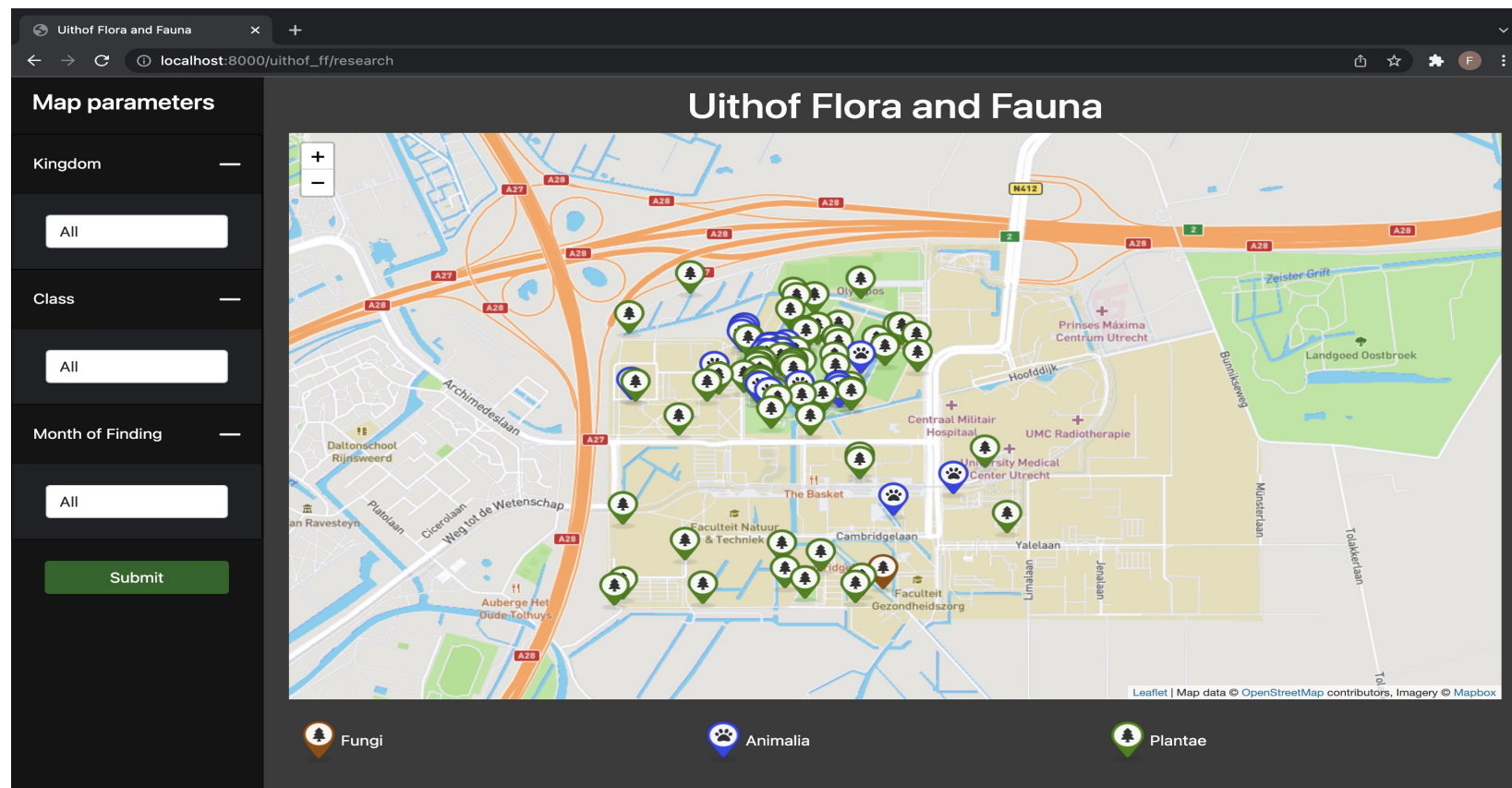
```
1 #Rock bands that start with "K"
2 SELECT DISTINCT ?bandLabel
3 WHERE
4 {
5   ?band wdt:P31 wd:Q5741069 .
6   ?band rdfs:label ?bandLabel .
7   FILTER(lang(?bandLabel)="en")
8   FILTER(STRSTARTS(?bandLabel, 'K') || STRSTARTS(?bandLabel, 'k')) .
9 }
```

bandLabel
Kraan
Kansas
Keimzeit
Keep of Kalessin
Kult
Killing Joke
KISS
Kane
KMFDM



Selected projects from previous years

- A geospatial search application, for flora and fauna at Uithof, that makes use of semantic web technologies [Francois, Hugo and Brent].



Procedure

- Teams of five students
 - realize a semantic web project
 - present the project results to the other students
 - provide a report
 - use Gitlab to track your progress

Upcoming deadlines

- Project report
 - max 5 pages, you need to include in the report the contributions of each member of your group.
 - **January 19, 2024**
- Project presentation (10 minutes)
 - **January 24, 2024 from 09:00 – 12:45, Location: RUPPERT-ROOD**
- Final exam
 - **January 30, 2024, from 17:00 – 20:00, Location: EDUC-BETA**

Modeling Motor Sports using Semantic Web Technologies



Motor sports

- “activities that involve the use of vehicles with a motor” [1].
- activities that involve the use of cars, motorcycles, power boats, etc.
- Categories based on racing
 - Circuit racing
 - Rallying
 - Trials
- Categories based on vehicle type
 - Single-seater
 - Formula 1/2/3/E
 - Stock-car racing
 - NASCAR
- Categories based on surface type
 - Road, off-road, dirt
- Categories based on propulsion method



Source: By Morio - Own work, CC BY-SA 3.0,
<https://commons.wikimedia.org/w/index.php?curid=14930940>

[1] <https://en.wikipedia.org/wiki/Motorsport>



Motor sports – Formula 1

There are:

- Drivers
- Circuits
- Constructors
- Grands Prix
- Cars
- Teams
- Races ...



Source: Source: By Morio - Own work, CC BY-SA 3.0,
<https://commons.wikimedia.org/w/index.php?curid=14930940>



By Morio - This file was derived from: Max Verstappen 2017 Malaysia 3.jpg by Saga City, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=66411705>



Datasets and data sources

- F1 dataset from Kaggle
<https://www.kaggle.com/datasets/rohanrao/formula-1-world-championship-1950-2020>
 - You're responsible for respecting others' rights, including copyright.
 - Please don't post the datasets in forums and other platforms without the proper license.
 - If possible, use the dataset only for the purpose of completing this project.
- F1 fandom Wiki https://f1.fandom.com/wiki/Formula_1_Wiki
- Wikipedia, DBpedia, Wikidata
- F1 official website <https://www.formula1.com/en.html>
- ...



Datasets and data sources

- F1 dataset from Kaggle
<https://www.kaggle.com/datasets/rohanrao/formula-1-world-championship-1950-2020>
 - There are 14 datasets (csv files) containing info about:
 - Circuits, drivers, seasons, constructors, etc.

circ	circuitRef	name	location	country	lat	lng	alt	url
1	albert_park	Albert Park Grand Prix Circuit	Melbourne	Australia	-37.8497	144.968	10	http://en.wikipedia.org/wiki/Melbourne_Grand_Prix_Circuit
2	sepang	Sepang International Circuit	Kuala Lumpur	Malaysia	2.76083	101.738	18	http://en.wikipedia.org/wiki/Sepang_International_Circuit
3	bahrain	Bahrain International Circuit	Sakhir	Bahrain	26.0325	50.5106	7	http://en.wikipedia.org/wiki/Bahrain_International_Circuit
4	catalunya	Circuit de Barcelona-Catalunya	Montmeló	Spain	41.57	2.26111	109	http://en.wikipedia.org/wiki/Circuit_de_Barcelona-Catalunya
5	istanbul	Istanbul Park	Istanbul	Turkey	40.9517	29.405	130	http://en.wikipedia.org/wiki/Istanbul_Park
6	monaco	Circuit de Monaco	Monte-Carlo	Monaco	43.7347	7.42056	7	http://en.wikipedia.org/wiki/Circuit_de_Monaco
7	villeneuve	Circuit Gilles Villeneuve	Montreal	Canada	45.5	-73.5228	13	http://en.wikipedia.org/wiki/Circuit_Gilles_Villeneuve
8	magny_cours	Circuit de Nevers Magny-Cours	Magny Cours	France	46.8642	3.16361	228	http://en.wikipedia.org/wiki/Circuit_de_Nevers_Magny-Cours
9	silverstone	Silverstone Circuit	Silverstone	UK	52.0786	-1.01694	153	http://en.wikipedia.org/wiki/Silverstone_Circuit



Datasets and data sources

- F1 Wiki https://f1.fandom.com/wiki/Formula_1_Wiki
 - Lots of semi-structured data including tables and infoboxes.

Circuit	Type	Location	Country	Race(s)	Season(s)	Races held
A1-Ring	race	Zeltweg	AUT	Austrian GP	1997-2003	7
Adelaide Street Circuit	street	Adelaide	AUS	Australian GP	1985-1995	11
Ain-Diab	road	Casablanca	MAR	Moroccan GP	1958	1
Aintree	race	Aintree	GBR	British GP	1955, 1957, 1959, 1961-1962	5
Albert Park	street	Melbourne	AUS	Australian GP	1996-2019, 2022-present	23
Algarve International Circuit	race	Portomaio	POR	Portuguese GP	2020-2021	2



Tasks

1. Provide a motivation and argue why semantic web technologies are a suitable choice to model motor sports.
2. Transform the data (about motor sports) into RDF
 - i. If it's not possible to use web scrapping tools to extract data from F1 fandom wiki, then you can manually curate triples.
 - ii. When transforming the data into RDF, provide efficient/compact ways to represent it. For instance, how can you represent **temporal data** of the following form:

Hockenheimring	race	Hockenheim	 GER	German GP	1970, 1977-1984, 1986-2006, 2008, 2010, 2012, 2014, 2016, 2018, 2019	37
----------------	------	------------	---	-----------	---	----

- iii. Don't just represent everything, identify those that can be inferred from the data, for instance number of races held (last column in the above table).



Tasks

1. Provide a motivation and argue why semantic web technologies are a suitable choice to model motor sports.
2. Transform the data (about motor sports) into RDF
 - i. If it's not possible to use web scrapping tools to extract data from F1 fandom wiki, then you can manually curate triples.
 - ii. When transforming the data into RDF, provide efficient/compact ways to represent it. For instance, how can you represent **temporal data** of the following form:

Hockenheimring	race	Hockenheim	 GER	German GP	1970, 1977-1984, 1986-2006, 2008, 2010, 2012, 2014, 2016, 2018, 2019	37
----------------	------	------------	---	-----------	---	----

- iii. Don't just represent everything, identify those that can be inferred from the data, for instance number of races held (last column in the above table).
- iv. Provide a method to efficiently represent data that involves reification.
- v. Use a triple store (such as BlazeGraph, Apache Jena Fuseki, OpenLink Virtuoso, etc) to maintain your data



Tasks

1. Provide a motivation and argue why semantic web technologies are an appropriate choice to model motor sports.
2. Transform the data (about motor sports) into RDF
3. Create ontological axioms using RDF schema and OWL
 - i. Axioms to represent different categories of motor sports, drivers, constructors, teams, etc.
 - ii. You need to provide *at least* 20 axioms (10 in RDFS and 10 in OWL)
4. Provide 10 interesting SPARQL queries (the queries need to use constructs such as property paths, negation, aggregates, etc)
 - i. Alternatively, provide a simple interactive interface through which the data can be accessed/searched.



Tasks

5. **Data integration:** link your knowledge base with DBpedia and Wikidata
6. Provide remarks for: what are the challenges for the semantic web? Do you see a future for it in the era of generative AI?
7. Provide dataset statistics, i.e., number of triples, resources, properties, axioms, etc. In addition, show by how much you reduced the size of the data by efficiently modelling temporal facts.
8. **Generate embeddings for the resources and properties** of your dataset using TransE, RotatE or your preferred method.
 - Use the embeddings for a query answering task, for instance, (s, p, ?)
9. Submit a report (max. 5 pages) and the source codes of your system.
10. Present your work.