



УНИВЕРСИТЕТ ИТМО

Разработка графовой базы знаний “Гонки Формула-1”

Дисциплина: “Графовые базы знаний”

Группа: Р41142

Автор: Рогаленко Н.А.

Санкт-Петербург
2022

Предметная область

- Разрабатывается граф знаний по автогонкам “Формулы-1”
- Информация о командах, гонщиках, конструкторах, сезонах, чемпионатах, заездах, пит-стопах, болидах и др.
- Обогащение данными о книгах, фильмах и в перспективе других объектов культуры
- Возможно применение в вопросно-ответных системах



Ergast Developer API

- Веб-сервис, предоставляющий доступ к историческим записям автогонок на некоммерческой основе
- Широкий спектр данных (гонщики, результаты, трассы, пит-стопы, конструкторы и др.)
- Извлечение данных посредством запросов к API

```
<MRData xmlns="http://ergast.com/mrd/1.4" series="f1" url="http://ergast.com/api/f1/drivers/alonso" limit="30" offset="0" total="1">
  <DriverTable driverId="alonso">
    <Driver driverId="alonso" code="ALO" url="http://en.wikipedia.org/wiki/Fernando_Alonso">
      <PermanentNumber>14</PermanentNumber>
      <GivenName>Fernando</GivenName>
      <FamilyName>Alonso</FamilyName>
      <DateOfBirth>1981-07-29</DateOfBirth>
      <Nationality>Spanish</Nationality>
    </Driver>
  </DriverTable>
</MRData>
```

f1technical.net

- Интернет-ресурс, предоставляющий доступ к данным о гонках
- В первую очередь интересуют данные о командах и болидах
- Извлечение данных посредством веб-скрапинга

Grand Prix cars since 1950

1950 - 1951 - 1952 - 1953 - 1954 - 1955 - 1956 - 1957 - 1958 - 1959
1960 - 1961 - 1962 - 1963 - 1964 - 1965 - 1966 - 1967 - 1968 - 1969
1970 - 1971 - 1972 - 1973 - 1974 - 1975 - 1976 - 1977 - 1978 - 1979
1980 - 1981 - 1982 - 1983 - 1984 - 1985 - 1986 - 1987 - 1988 - 1989
1990 - 1991 - 1992 - 1993 - 1994 - 1995 - 1996 - 1997 - 1998 - 1999
2000 - 2001 - 2002 - 2003 - 2004 - 2005 - 2006 - 2007 - 2008 - 2009
2010 - 2011 - 2012 - 2013 - 2014 - 2015 - 2016 - 2017 - 2018 - 2019
2020 - 2021 - 2022 -

Formula One cars of 2022

Alfa Romeo C42 Ferrari	 Sauber F1 Team
AlphaTauri AT03 RBPT	 Scuderia Toro Rosso
Alpine A522 Renault	 Alpine F1 Team
Aston Martin AMR22 Mercedes	 Aston Martin
Ferrari F1-75	 Scuderia Ferrari SpA
Haas VF-22 Ferrari	 Haas F1 Team
McLaren MCL36 Mercedes	 McLaren International Ltd
Mercedes F1 W13 E Performance	 Mercedes AMG F1
Red Bull RB18 RBPT	 Red Bull Racing
Williams FW44 Mercedes	 Williams F1

f1-fansite.com

- Интернет-ресурс, предоставляющий доступ к данным о гонках
- В первую очередь интересуют таблицы результатов и команд
- Извлечение данных посредством веб-скрапинга

Classification 1976 Brazilian F1 GP							
P	Nº	DRIVER	TEAM	TIME	LAP	GRID	PT
1	1	 Niki Lauda	 Ferrari	01:45:16.780	40	2	9
2	4	 Patrick Depailler	 Tyrrell	01:45:38.250	40	9	6
3	16	 Tom Pryce	 Shadow	01:45:40.620	40	12	4
4	34	 Hans Joachim Stuck	 March	01:46:44.950	40	14	3
5	3	 Jody Scheckter	 Tyrrell	01:47:13.240	40	13	2

<https://www.f1-fansite.com/f1-result/results-1976-formula-1-grand-prix-of-brazil/>

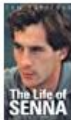



Goodreads

- Данные о книгах, посвященных Формуле-1
- Извлечение данных посредством веб-скрапинга

Best Formula 1 Racing Books

Best ever Formula 1 books.

Vote for the all-time Formula 1 novel.

	All Votes	Add Books To This List
1	 The Life of Senna by Tom Rubython ★★★★★ 3.75 avg rating — 1,683 ratings score: 2,400, and 24 people voted	Want to Read ▾ Rate this book ★★★★★
2	 The Death of Ayrton Senna by Richard Williams ★★★★★ 3.71 avg rating — 1,641 ratings score: 2,174, and 22 people voted	Want to Read ▾ Rate this book ★★★★★
3	 Ayrton Senna: The Whole Story by Christopher Hilton ★★★★★ 3.64 avg rating — 1,412 ratings score: 1,866, and 19 people voted	Want to Read ▾ Rate this book ★★★★★
4	 Burning Rubber: The Extraordinary Story of Formula One by Charles Jennings ★★★★★ 3.83 avg rating — 198 ratings score: 781, and 8 people voted	Want to Read ▾ Rate this book ★★★★★



IMDB

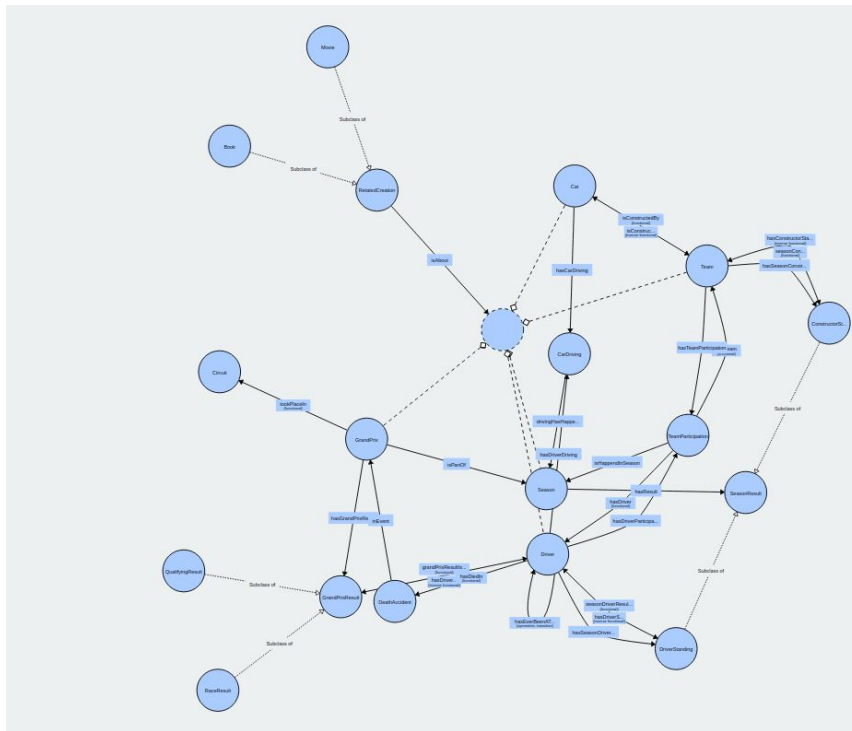
- Данные о фильмах, посвященных Формуле-1
- Извлечение данных посредством запросов к API

```
{
  "id": "tt1979320",
  "image": "https://imdb-api.com/images/original/MV5BOWEwODJmZDItYTNmZC00OGM4LTlhNDktOTQzZjIzMGMQxODA4XkEyXkFqcGdeQXVyNjU0OTQ0OTY0._V1_Ratio0.6837_AL_.jpg",
  "title": "Rush",
  "description": "(I) (2013)",
  "runtimeStr": "123 min",
  "genres": "Action, Biography, Drama",
  "genreList": [
    {
      "key": "Action",
      "value": "Action"
    },
    {
      "key": "Biography",
      "value": "Biography"
    },
    {
      "key": "Drama",
      "value": "Drama"
    }
  ],
  "contentRating": "R",
  "imDbRating": "8.1",
  "imDbRatingVotes": "468414",
  "metacriticRating": "74",
  "plot": "The merciless 1970s rivalry between Formula One rivals James Hunt and Niki Lauda.",
  "stars": "Ron Howard, Daniel Brühl, Chris Hemsworth, Olivia Wilde, Alexandra Maria Lara",
}
```


Сложности этапа

- Отсутствие API у многих источников
- Ограничения различных бесплатных API
- Получение кода 403 без добавления заголовка User-Agent
- Поиск фильмов и книг

Разработанная онтология



Сложности этапа

- Продумывание некоторых типов связей
- Овеществление участия в команде и пилотирования болида

Наполнение графа

Metrics

Axiom	408973
Logical axiom count	408870
Declaration axioms count	103
Class count	19
Object property count	24
Data property count	59
Individual count	46348
Annotation Property count	0

Class axioms

SubClassOf	6
EquivalentClasses	0
DisjointClasses	1
GCI count	0
Hidden GCI Count	0

Object property axioms

SubObjectPropertyOf	0
EquivalentObjectProperties	0
InverseObjectProperties	4
DisjointObjectProperties	0
FunctionalObjectProperty	8
InverseFunctionalObjectProperty	4
TransitiveObjectProperty	1
SymmetricObjectProperty	1

Сложности этапа

- Лишние символы и неточности в результате скраппинга
- Поиск соответствия объекта самому себе в другом источнике
- Обработка описаний фильмов и книг

Компетентностные вопросы

- Каковы характеристики болида Формулы-1, о котором сняли наибольшее число фильмов и написали наибольшее число книг?
- Сколько книг и фильмов создано о погибших гонщиках?
- Какой гонщик оказал наибольшее влияние на культуру, согласно числу снятых фильмов и написанных книг?
- Каково среднее количество кругов на трассах, на которых проходили гонки, о которых написали книги с рейтингом не менее 4 звезд?
- Сколько сезонов выиграла команда, о которой сняли наибольшее число фильмов?

Запрос №1

Ferrari 312B (1 books and movies):

Chassis: Type 001, monocoque body, aluminium panels riveted on a tubular steel structure, partially stress-bearing engine

Front Suspension: Double wishbones, upper rocker arm, lower wishbone, inboard spring/damper units and anti-roll bar

Rear Suspension: Upper arm, reversed lower wishbone, 2 radius arms and anti-roll bar

Brakes: Outboard Girling discs and callipers, separate adjustable hydraulic circuits on the two axles

Steering: Rack and pinion

Fuel Tank: 240 litres, side-mounted in canvas/Pirelli rubber

Bodywork: Aluminium single-seater body with composite elements

Transmission: Rear-wheel drive, Borg & Beck multi-plate clutch, 001, rear longitudinal overhanging 5-speed gearbox + reverse

Gear Ratios: Variable depending on track

Запрос №2

```
Number of creations about drivers died in accidents is: 53
```


Запрос №3

```
Driver Ayrton Senna: 40 related creations
Driver Bruce McLaren: 19 related creations
Driver Jackie Lewis: 18 related creations
Driver Damon Hill: 17 related creations
Driver Alain Prost: 17 related creations
Driver Max Jean: 15 related creations
Driver Niki Lauda: 15 related creations
Driver Michael Schumacher: 15 related creations
Driver Lewis Hamilton: 14 related creations
Driver Jonathan Williams: 13 related creations
```

Запрос №4

```
Average number of laps in grand prix described in books with rating >= 4.00: 68.833333333333333333333333
```



Запрос №5

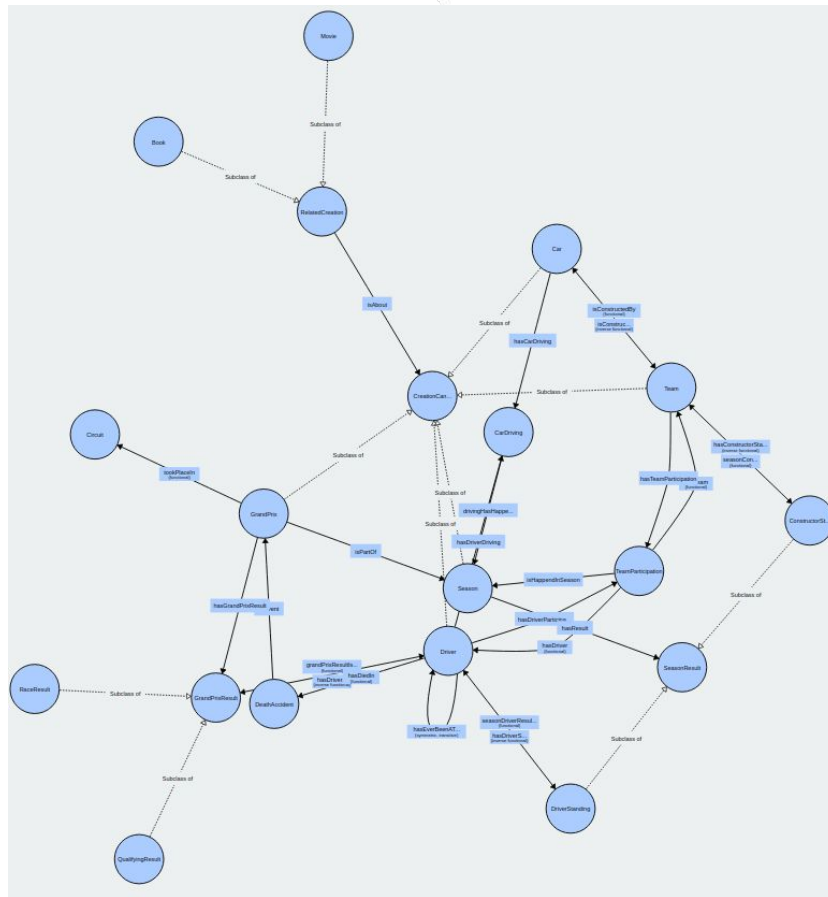
```
Team Ferrari won 17 seasons, has 4 movies
Team Williams won 9 seasons, has 3 movies
Team Mercedes won 8 seasons, has 1 movies
Team McLaren won 8 seasons, has 8 movies
Team Red Bull won 4 seasons, has 2 movies
Team Team Lotus won 4 seasons, has 1 movies
Team Brabham-Repco won 2 seasons, has 0 movies
Team Lotus-Climax won 2 seasons, has 0 movies
Team Renault won 2 seasons, has 1 movies
Team Cooper-Climax won 2 seasons, has 0 movies
```

Сложности этапа

- Невозможность выполнять SPARQL запросы в Protege из-за размеров онтологии

Онтология v2

- Небольшое изменение - отдельный суперкласс для сущностей, о которых могут снять фильм или написать книгу



Reasoner

- Использовался Pellet
- Исправлены неточности и ошибки в datatype properties

```
----- Running Reasoner -----  
Pre-computing inferences:  
- class hierarchy  
- object property hierarchy  
- data property hierarchy  
- class assertions  
- object property assertions  
- same individuals  
Ontologies processed in 63859 ms by Pellet
```

Void

- void:triples - 409097
- void:entities - 46425
- void:classes - 26
- void:properties - 93

Документация (pyLODE)

made by **pyLODE** 3.0.1

Formula 1 Knowledge base

Metadata

IRI

<https://github.com/RogoGit/F1-knowledge-base/f1-ontology>

Title

Formula 1 Knowledge base

Version Info

In this version only movies and books are supported as 'related creation'

Description

Formula One knowledge base, which contains information about drivers, events, results and other entities related to subject area.

Also contains "related creations" data - information about cultural impact of this racing series

Classes

Book^C

IRI

<https://github.com/RogoGit/F1-knowledge-base/f1-ontology#Book>

Description

Entity, which represents book related to Formula One racing

Sub Class Of

[f1.RelatedCreation^C](#)

In Domain Of

[f1.author^{dp}](#)
[f1.goodreadsRating^{dp}](#)
[f1.pages^{dp}](#)
[f1.reviewsNum^{dp}](#)

Table of Contents

Metadata

Classes

[Book](#)
[Car](#)
[Car Driving](#)
[Circuit](#)
[Constructor Standing](#)
[Creation Can Be About](#)
[Death Accident](#)
[Driver](#)
[Driver Standing](#)
[Grand Prix](#)
[Grand Prix Result](#)
[Movie](#)
[Qualifying Result](#)
[Race Result](#)
[Related Creation](#)
[Season](#)
[Season Result](#)
[Team](#)
[Team Participation](#)

Object Properties

[driving_has_happened_in](#)
[season](#)
[grand_prix_result_is_related_to](#)
[to](#)
[has_car_driving](#)
[has_constructor_standing](#)
[result](#)
[has_died_in](#)
[has_driver](#)
[has_driver_driving](#)
[has_driver_grand_prix_result](#)
[has_driver_participation](#)
[has_driver_standing_result](#)
[has_ever_been_a_teammate](#)
[has_grand_prix_result](#)
[has_result](#)
[has_team](#)
[has_team_participation](#)
[in_event](#)
[is_about](#)
[is_constructed_by](#)
[is_constructor_of](#)
[is_happened_in_season](#)
[is_part_of](#)
[season_constructor_result](#)
[is_related_to](#)
[season_driver_result_is_related_to](#)
[took_place_in](#)

Верификация графа (SHACL)

```
def verify_graph(ontology_path, data_format, shapes_path):
    f1_graph = Graph().parse(ontology_path, format=data_format)
    f1_shapes = Graph().parse(shapes_path, format=data_format)
    results = pyshacl.validate(
        data_graph=f1_graph,
        shacl_graph=f1_shapes,
        data_graph_format=data_format,
        shacl_graph_format=data_format,
        inference="rdfs",
        debug=True,
        serialize_report_graph=data_format,
    )
    conforms, report_graph, report_text = results
    print(f"Conforms: {conforms}\nReport graph:\n{report_graph}")
```

```
f1:CircuitShape
  a sh:NodeShape ;
  sh:targetClass f1:Circuit ;
  sh:property [
    sh:path f1:circuitLocationLong ;
    sh:datatype xsd:double ;
    sh:minInclusive -180 ;
    sh:maxInclusive 180 ;
  ] ;
  sh:property [
    sh:path f1:circuitLocationLat ;
    sh:datatype xsd:double ;
    sh:minInclusive -90 ;
    sh:maxInclusive 90 ;
  ] .
```

```
f1:GrandPrixShape
  a sh:NodeShape ;
  sh:targetClass f1:GrandPrix ;
  sh:property [
    sh:path f1:tookPlaceIn;
    sh:node f1:CircuitShape ;
    sh:maxCount 1 ;
  ] ;
  sh:property [
    sh:path f1:grandPrixDate ;
    sh:datatype xsd:date ;
    sh:maxCount 1 ;
  ] ;
```



Графовые эмбединги

- Обучающая выборка: 326225 триплетов
- Тестовая выборка: 36247 триплетов
- Модель: ComplEx

```
model = ComplEx(batches_count=100,  
               epochs=300,  
               k=100,  
               eta=20,  
               optimizer='adam',  
               optimizer_params={'lr':1e-4},  
               loss='multiclass_nll',  
               regularizer='LP',  
               regularizer_params={'p':3, 'lambda':1e-5},  
               seed=0,  
               verbose=True)
```

Пример эмбединга

```
[('f1:team_participation_1968_john_surtees_honda', array([ 0.15058921, 0.
-0.19745371, -0.15330258, 0.19007845, -0.14624396, -0.18978721,
0.01731578, 0.16899933, 0.19650525, -0.16436966, -0.16565551,
0.00911854, -0.01223622, 0.1442946 , 0.07178091, -0.13925822,
-0.17620327, -0.17334235, 0.15280165, -0.19565189, -0.1704562 ,
0.01307552, 0.15865031, -0.17643066, 0.15020834, -0.10737054,
-0.00249887, -0.20082456, -0.10074502, -0.15203606, -0.17853774,
-0.09260228, -0.12076462, -0.07302562, 0.17158657, 0.00034421,
-0.19432108, 0.14239483, 0.06417444, -0.14825055, -0.12346714,
0.1280998 , 0.02698476, 0.15995395, -0.13690637, 0.2305665 ,
-0.23822747, 0.12913081, -0.10044429, -0.05091096, 0.2150847 ,
-0.1700849 , 0.11347755, 0.05339024, 0.01367681, -0.00865733,
0.05937546, -0.1170646 , -0.07530995, -0.15010785, -0.03708335
```

Предсказывание связей (результат гонки)

	statement	rank	score	prob
17	f1:race_result_2022_azerbaijan_grand_pr...	122005	-1.947975	0.124774
12	f1:race_result_2022_azerbaijan_grand_pr...	116362	-1.010131	0.266954
13	f1:race_result_2022_azerbaijan_grand_pr...	107985	-0.665249	0.339561
19	f1:race_result_2022_azerbaijan_grand_pr...	72788	-0.009744	0.497564
11	f1:race_result_2022_azerbaijan_grand_pr...	41536	0.769191	0.683346
18	f1:race_result_2022_azerbaijan_grand_pr...	42165	0.857910	0.702224
16	f1:race_result_2022_azerbaijan_grand_pr...	38889	0.991935	0.729470
2	f1:race_result_2022_azerbaijan_grand_pr...	40258	1.028274	0.736581
7	f1:race_result_2022_azerbaijan_grand_pr...	38289	1.377143	0.798532