

Онтологии и представление знаний


Прохоров Даниил Р42301



Предметная область

Велосипедные компоненты

Велосипедные компоненты имеют большое количество особенностей, стандартов, размеров, производителей, стран изготовителей и других данных, которые можно удобно представить в формате графа



Данные для анализа

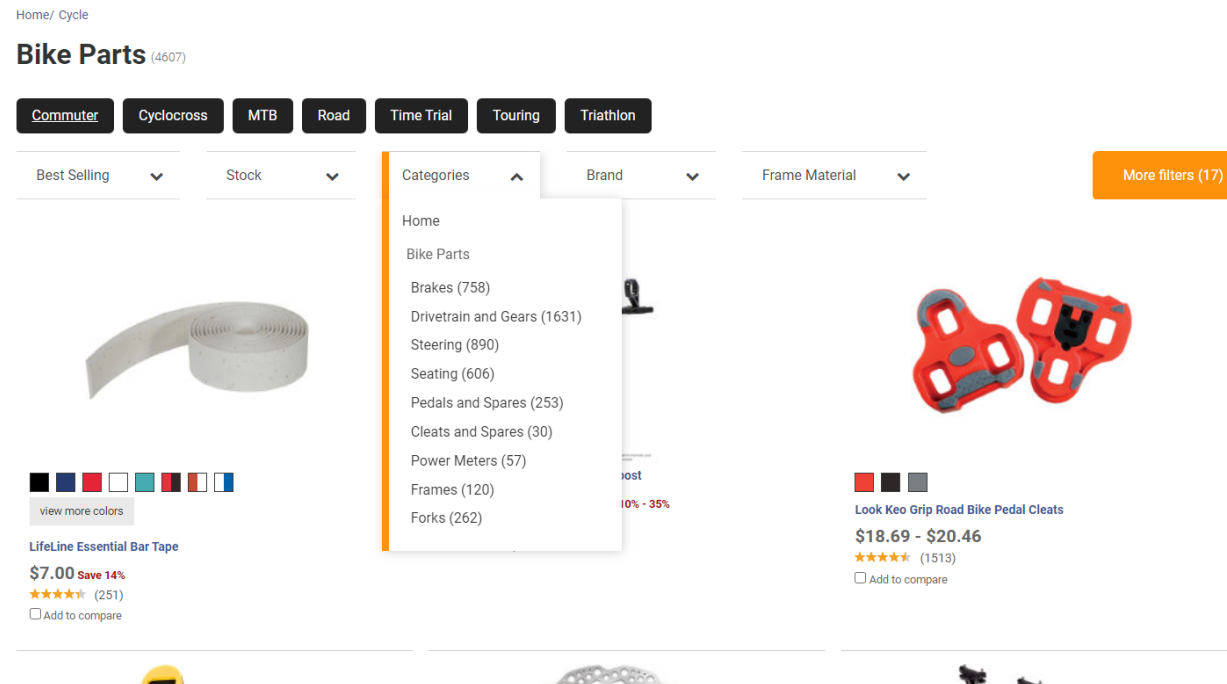
Открытых API с данными не существует (возможно только использовать API Ebay, однако там только бывшие в употреблении запчасти)

Сайты для парсинга:

- www.bike24.com
- www.wiggle.com

Первые сложности

Bike24 не дает парсить собственные страницы, поэтому было принято решение использовать только данные сайта Wiggle.com, к тому же это стало и так стало комплексной задачей из-за сложности организации доступа к конкретной запчасти (требовалось пройти много ссылок вложенности, чтобы сформировать структуру сайта, с помощью которой потом получить ссылки на отдельные запчасти и загрузить данные с их страниц)



Иерархия расположения запчастей на сайте имеет представленную в примере структуру

```
{
  "bike-parts": {
    "seating": {
      "saddles": {},
      "seat-posts": {},
      "dropper-seats-posts": {},
      "seat-post-clamps": {},
      "seat-post-spares": {},
      "dropper-remotes": {},
      "saddle-care-kits": {}
    },
    "cleats-and-spares": {
      "cleats": {},
      "cleat-covers": {}
    },
  },
}
```


```
"power-meters": {
  "power-meter-chainsets": {},
  "power-meter-cranksets": {},
  "power-meter-pedals": {}
},
"frames": {
  "road-bike-frames": {},
  "mountain-bike-frames": {
    "hard-tail-mountain-bike-frames": {},
    "full-sus-mountain-bike-frames": {}
  },
  "cyclocross-bike-frames": {},
  "adventure-bike-frames": {},
  "time-trial-bike-frames": {},
  "bmx-bike-frames": {},
  "electric-bike-frames": {},
  "kids-bike-frames": {}
},
}
```



Сложности 2

- 1) Я никогда не писал парсеры сайтов (что-то похожее было только на тестировании)
- 2) Я решил использовать ЯП Clojure, потому что он всегда мне казался интересным

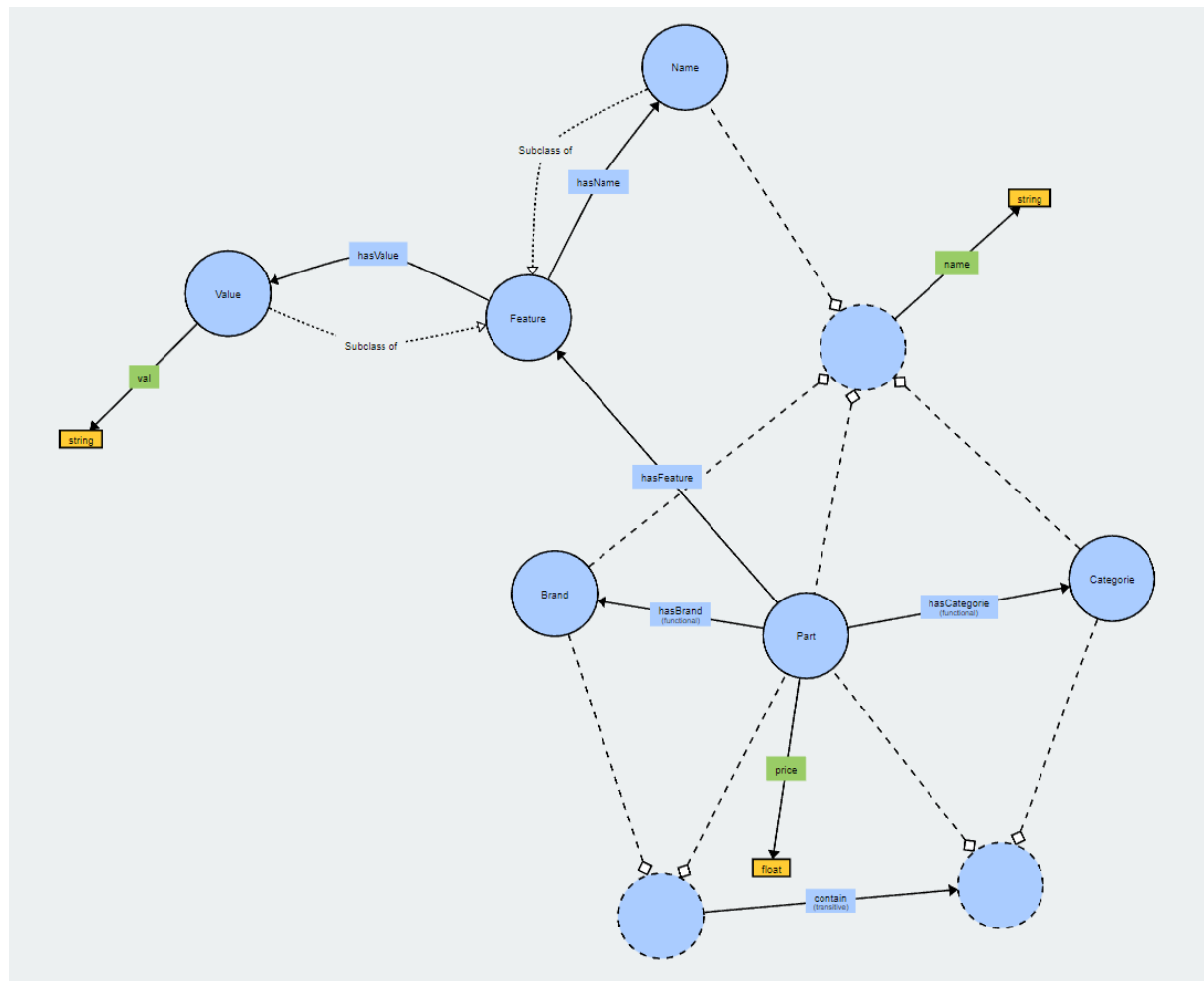
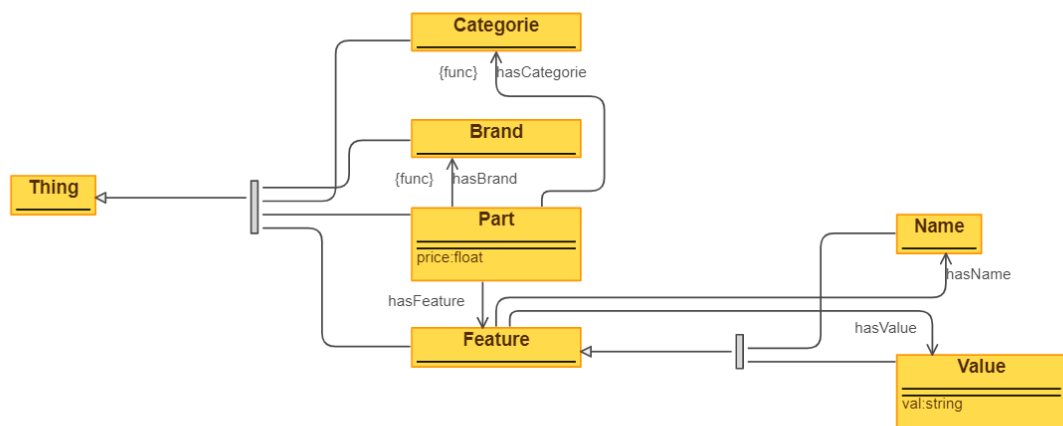
Из-за этих двух пунктов время которое я потратил на парсинг данных многократно увеличилось, но все получилось!



Пример получаемой структуры данных

```
{
  "title": "Selle Italia SLR Kit Carbonio Flow Saddle with Carbon Rails",
  "topFeatures": [
    "Cut-out for pressure relief and greater comfort",
    "Lightweight and stylish",
    "CarboKeramic rails and Fibra-Tek cover",
    "Carbon inserts at the back",
    "Approximate Weight: 130g (+/-5g)",
    "Rails: CarboKeramic 7x9mm",
    "Use: Road/Offroad",
    "Dimensions:",
    "S2: 135x275mm (WxL)",
    "L2: 145x275mm (WxL)",
    "Rail Material: Carbon",
    "Road: Yes",
    "MTB: Yes"
  ],
  "price": "£99.99"
```

Онтологическая модель



Компетентностные вопросы

- Все бренды карбоновых рулей
- Все седла для шоссейных велосипедов с покрытием из кожи и ценой до 100 фунтов
- Соответствие цепей кассетам по бренду и ширине цепи (ширина зависит от количества скоростей)
- Все подседельные штыри длиной 400мм



Все бренды карбоновых рулей

```
PREFIX base: <urn:bike#>
SELECT ?brand_name ?part_name
WHERE {
  ?Part a base:Part.
  ?Part base:price ?price.
  ?Part base:name ?part_name.
  ?Part base:hasFeature ?Feature.
  ?Part base:hasCategorie ?Categorie.
  ?Part base:hasBrand ?Brand.
  ?Categorie base:name ?categorie_name.
  ?Brand base:name ?brand_name.
  ?Feature base:hasName ?Name.
  ?Feature base:hasValue ?Value.
  ?Name base:name ?feature_name.
  ?Value base:val ?feature_value.
  FILTER(?categorie_name = 'handlebar' || ?categorie_name = 'drop_handlebar')
  FILTER CONTAINS(?feature_name, 'material')
  FILTER CONTAINS(?feature_value, 'carbon')
}
GROUP BY ?brand_name
```

cinelli
e.thirteen
easton
fizik
ion
pro
protaper
race face
renthal
ritchey
met
zipp



Все седла для шоссейных велосипедов с покрытием из кожи и ценой до 100 фунтов

```
PREFIX base: <urn:bike#>
SELECT ?part_name ?price ?feature_name1 ?feature_value1 ?feature_name2 ?feature_value2
WHERE {
    ?Part a base:Part.
    ?Part base:price ?price.
    ?Part base:name ?part_name.
    ?Part base:hasCategorie ?Categorie.
    ?Categorie base:name ?categorie_name.
    FILTER(?categorie_name = 'saddle')
    FILTER(?price <= 100 )
    {
        ?Part base:hasFeature ?Feature1.
        ?Feature1 base:hasName ?Name1.
        ?Feature1 base:hasValue ?Value1.
        ?Name1 base:name ?feature_name1.
        ?Value1 base:val ?feature_value1.
        FILTER CONTAINS(?feature_name1, 'road')
        FILTER CONTAINS(?feature_value1, 'yes')
    }
    {
        ?Part base:hasFeature ?Feature2.
        ?Feature2 base:hasName ?Name2.
        ?Feature2 base:hasValue ?Value2.
        ?Name2 base:name ?feature_name2.
        ?Value2 base:val ?feature_value2.
        FILTER CONTAINS(?feature_name2, 'cover')
        FILTER CONTAINS(?feature_value2, 'leather')
    }
}
```

```
Selle Italia C2 Gelflow Racing Saddle;
Selle Italia Diva Gel Flow Bike Saddle;
Selle Italia Flite 1990 Bike Saddle;
Selle Italia SLR Superflow Bike Saddle;
Selle Italia Turbo Bullitt Saddle;
```

Соответствие цепей кассетам по бренду и ширине цепи (ширина зависит от количества скоростей)

```
PREFIX base: <urn:bike#>
SELECT ?part_name1 ?part_name2 ?feature_name1 ?feature_value1
WHERE {
  {
    ?Part1 a base:Part.
    ?Part1 base:name ?part_name1.
    ?Part1 base:hasCategorie ?Categorie1.
    ?Part1 base:hasBrand ?Brand1.
    ?Brand1 base:name ?brand_name1.
    ?Categorie1 base:name ?categorie_name1.
    FILTER(?categorie_name1 = 'cassette')

    ?Part1 base:hasFeature ?Feature1.
    ?Feature1 base:hasName ?Name1.
    ?Feature1 base:hasValue ?Value1.
    ?Name1 base:name ?feature_name1.
    ?Value1 base:val ?feature_value1.
    FILTER CONTAINS(?feature_name1, 'speed')
    FILTER NOT EXISTS {
      FILTER CONTAINS(?feature_value1, '-')
    }
  }
}
```

```
{
  ?Part2 a base:Part.
  ?Part2 base:name ?part_name2.
  ?Part2 base:hasCategorie ?Categorie2.
  ?Part2 base:hasBrand ?Brand2.
  ?Brand2 base:name ?brand_name2.
  ?Categorie2 base:name ?categorie_name2.
  FILTER(?categorie_name2 = 'chain')

  ?Part2 base:hasFeature ?Feature2.
  ?Feature2 base:hasName ?Name2.
  ?Feature2 base:hasValue ?Value2.
  ?Name2 base:name ?feature_name2.
  ?Value2 base:val ?feature_value2.
  FILTER CONTAINS(?feature_name2, 'speed')
  FILTER NOT EXISTS {
    FILTER CONTAINS(?feature_value2, '-')
  }
}
FILTER (?feature_value2 = ?feature_value1)
FILTER (?brand_name2 = ?brand_name1)
```

```
Campagnolo Centaur 11 Speed Cassette Campagnolo Potenza 11 Speed Chain speed: 11 speed;
Campagnolo Centaur 11 Speed Cassette Campagnolo Record 11 Speed Chain speed: 11 speed;
Campagnolo Chorus 11 Speed Cassette Campagnolo Potenza 11 Speed Chain speed: 11 speed;
Campagnolo Chorus 11 Speed Cassette Campagnolo Record 11 Speed Chain speed: 11 speed;
Campagnolo Chorus 11 Speed Road Cassette Campagnolo Potenza 11 Speed Chain speed: 11 speed;
Campagnolo Chorus 11 Speed Road Cassette Campagnolo Record 11 Speed Chain speed: 11 speed;
Campagnolo Veloce 10 Speed Cassette Campagnolo Record 10 Speed Chain speed: 10 speed;
Campagnolo Veloce 10 Speed Cassette Campagnolo Veloce 10 Speed Chain speed: 10 speed;
Campagnolo Veloce 10 Speed Cassette Campagnolo Veloce 10 Speed Chain speed: 10 speed;
Campagnolo Veloce UltraDrive 9 Speed Cassette Campagnolo Record 9 Speed Chain speed: 9 speed;
Miche Light Primato 11x Shimano Cassette CeramicSpeed UFO Racing Shimano Chain speed: 11 speed;
Miche Light Primato 11x Shimano Cassette Shimano Dura-Ace/XTR HG901 11sp SIL-TEC Chain speed: 11 speed;
Miche Light Primato 11x Shimano Cassette Shimano HG601 105 5800 11 Speed Chain speed: 11 speed;
Miche Light Primato 11x Shimano Cassette Shimano HG601Q 105 5800 11 Speed Chain speed: 11 speed;
Miche Light Primato 11x Shimano Cassette Shimano HG701 Ultegra and XT 11 Speed Chain speed: 11 speed;
```


Все подседельные штыри длиной 400мм

```
PREFIX base: <urn:bike#>
SELECT ?brand_name ?part_name ?feature_name ?feature_value
WHERE {
  ?Part a base:Part.
  ?Part base:price ?price.
  ?Part base:name ?part_name.
  ?Part base:hasFeature ?Feature.
  ?Part base:hasCategorie ?Categorie.
  ?Part base:hasBrand ?Brand.
  ?Categorie base:name ?categorie_name.
  ?Brand base:name ?brand_name.
  ?Feature base:hasName ?Name.
  ?Feature base:hasValue ?Value.
  ?Name base:name ?feature_name.
  ?Value base:val ?feature_value.
  FILTER CONTAINS(?categorie_name, 'seat post')
  FILTER CONTAINS(?feature_name, 'length')
  FILTER CONTAINS(?feature_value, '400')
}
```

Fizik Cyrano R3 Seat Post length: 270/400mm;
Fizik Cyrano R5 Seat Post length: 270/400mm;
Ritchey WCS Carbon Flexlogic Seat Post lengths: 350, 400mm;
Ritchey WCS Link Road Seat Post 2017 lengths: 350, 400mm;
ULTIMATE USE Vybe Suspension Seat Post overall length: all options are 400mm long;



VoID



classes **12** ;
distinctObjects **17359** ;
distinctSubjects **11652** ;
properties **12** ;
triples **46706** .

Embedding

