

Два с половиной SPARQL-запроса и воздушные суда

Andrew Krizhanovsky¹, Rybin Egor²

¹ Institute of Applied Mathematical Research of the Karelian Research Centre of the Russian Academy of Sciences

¹ Petrozavodsk State University

Аннотация

Исследуем воздушные суда на основе базы знаний международного проекта Викиданные. С помощью SPARQL-запросов, вычисляемых на объектах типа "Воздушные суда" в Викиданных, решены такие задачи: выведен список всех воздушных судов, построена диаграмма соотношения количества производителей воздушных судов по странам. Выполнена оценка полноты Викиданных по этой теме. Осуществлено заполнение данных при помощи скрипта.

Примечание

Статья распространяется по лицензии Creative Commons Attribution-ShareAlike. Материалы этой статьи использованы в главе курса Викиверситета "Воздушные суда" [1]. Иллюстрации загружены на Викисклад. Над статьёй в 2017 году работали Крижановский А. А., Рыбин Е. И..

Экземпляры объекта "Воздушные суда"

- Объект: Воздушное судно (Q11436).
- Свойство: Экзмпляр (P31)

Построим список всех экземпляров объекта "Воздушные суда".

```

1 SELECT ?item ?itemLabel
2 WHERE
3 {
4     ?item wdt:P31 wd:Q11436.
5     SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
6 }
```

SPARQL-запрос, 1564 записи.

SPARQL-запрос, 153 записи имеют название на русском языке(label).

Заполнены не очень плохо: МиГ-3, Як-36, Mitsubishi A5M.

Заполнены не очень хорошо: МиГ-1, Су-6, Ил-103.

Производители воздушных судов

Построим список производителей воздушных судов

```

1 # Count aircraft having property manufacture
2 # Group by manufacture
3 SELECT ?manufactureLabel (COUNT(?item) AS ?count)
4 WHERE {
5     ?item wdt:P31 wd:Q11436.      # instance of aircraft
6     ?item wdt:P176 ?manufacture. # show manufacture
7     SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
8 }
9 GROUP BY ?manufacture ?manufactureLabel # group by manufacture
```

SPARQL-запрос, 300 результатов.

В результате запроса было получено 300 предприятий, занимающихся производством воздушных судов в разное время и в разных странах .

Страны происхождения производителей воздушных судов

```

1 # Count manufacture having property country
2 # group by country
3 SELECT ?countryLabel (count(?item) as ?count)
4 WHERE
```

```

5 {
6   ?item wdt:P31 wd:Q936518. # instance of manufacture
7   ?item wdt:P17 ?country. # shw country
8   SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
9 }
10 GROUP BY ?country ?countryLabel # group by country

```

SPARQL-запрос 39 записей.

Пузырьковая диаграмма

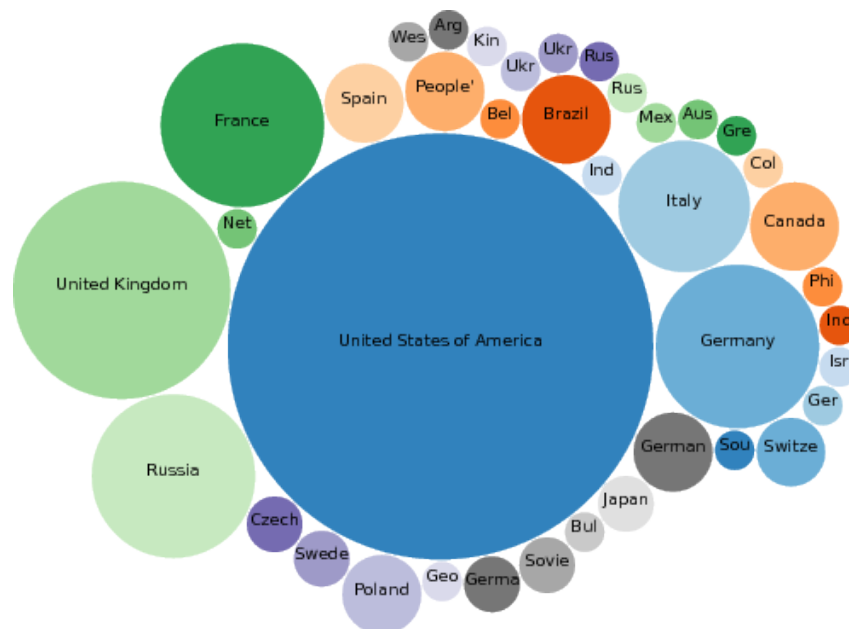


Figure 1. Соотношение количества производителей воздушных судов по странам

Как видно из запроса страны происхождения производителей воздушных судов заполнены куда хуже, чем могли бы быть, и предоставляют мало информации. Больше всего производителей указано у США (115), Великобритания (30), Германия (17), Россия (17).

Полнота Викиданных

Согласно сайту aviationfanatic.com существует около 1700 производителей воздушных судов [2], но SPARQL-запрос вернул всего 300 записей. Из этого можно сделать вывод о неполноте Викиданных. Скорее всего, оставшиеся полторы тысячи производителей сделали слишком мало самолетов или не сделали их вовсе, поэтому из-за недостатка информации они не были включены в Викиданные.

В категории Авиастроительные компании России указано наличие в стране 58 Авиастроительных компаний, но в то же время на сайте aviationfanatic.com указано наличие 61 производителя [3], например такие компании как Иркут, МиГ, Туполев.

Заполнение викиданных

Для заполнения были выбраны поля label и description для объектов, перечисленных в категории Авиастроительные компании России. Так как объектов там много, было решено автоматизировать заполнение, для чего была написана соответствующая программа [4]. Сначала был создан JSON-файл с объектами из этой категории и пустыми полями для заполнения:

```

    {
    "121 авиационный ремонтный завод": {
    "description": "",
    "descriptionen": "",
    "nameen": "",
    "qid": "Q4028573"
    },
    ...
    }

```

В первой части программы считывалась информация из Викиданных и записывалась в JSON-файл. После чего было необходимо заполнить оставшиеся пустыми поля, которых оказалось достаточно много. В итоге, в JSON-файле не осталось пустых полей:

```

    {
    "121 авиационный ремонтный завод": {
    "description": "авиаремонтное предприятие, расположенное посёлке Старый Городок",
    "descriptionen": "aircraft repair facility, located in the village Stary Gorodok",
    "nameen": "121 aircraft repair plant",
    "qid": "Q4028573"
    },
    ...
    }

```

Во второй части программы записывались данные из JSON-файла в Викиданные.

С помощью этой программы удалось упростить работу с Викиданными, так как не приходилось самостоятельно заходить на страницы объектов и вносить изменения, если существующие данные не удовлетворяют ожиданиям, то есть поле в Викиданных отличается от локального.

References

- [1] (2017). Aircraft. Research in programming Wikidata [Programmirovaniye Vikidannykh] // Russian Wikiversity. https://ru.wikiversity.org/wiki/%D0%9F%D1%80%D0%BE%D0%B3%D1%80%D0%B0%D0%BC%D0%BC%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5_%D0%92%D0%B8%D0%BA%D0%B8%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85/%D0%92%D0%BE%D0%B7%D0%B4%D1%83%D1%88%D0%BD%D1%8B%D0%B5_%D1%81%D1%83%D0%B4%D0%B0. Accessed 14 Mar 2017.
- [2] List of all Manufacturers. http://www.aviationfanatic.com/ent_list.php?ent=3.
- [3] List of all Russian Manufacturers. http://www.aviationfanatic.com/ent_list.php?ent=3&MAN_Country=RU.
- [4] Code of the Program. https://github.com/notautorobot/wiki_category_aviation_manufacturers.