

# Географический анализ школ России и их ВЫПУСКНИКОВ

Andrew Krizhanovsky

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

Arina Demidova

*Petrozavodsk State University*

24 декабря 2017 г.

## Аннотация

### Аннотация

**Школа** — это учебное заведение для получения общего образования. Статья посвящена исследованию школ на основе Викиданных. Для изучения школ используются SPARQL-запросы, работающие с объектами Викиданных типа «школа». Получен список всех ныне существующих школ, описанных в Викиданных. В статье присутствует анализ полноты базы Викиданных, содержащей информацию о школах, а также по Викиданным построены карта расположения российских школ и линейная диаграмма, на которой отображено количество известных выпускников каждой школы. По этой карте можно убедиться, что основные скопления школ находятся в Москве и Санкт-Петербурге, а по линейной диаграмме видно, что в основном Викиданным известны по два выпускника из каждой школы.

## 1 Экземпляры объекта “Школа”

- Объект: школа (Q3914).
- Свойство: экземпляр (P31).

Построим список всех школ (Листинг 1).

```

1 #List of schools
2 SELECT ?school ?schoolLabel
3 WHERE
4 {
5   ?school wdt:P31 wd:Q3914.
6   SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
7 }

```

Листинг 1: Скрипт для построения списка всех школ

SPARQL-запрос, 18615 объектов.

Примеры наиболее полно проработанных экземпляров объекта "Школа":

1. Консерватория искусств и ремёсел
2. Коннектикут-колледж
3. Гринвиллский колледж

Примеры плохо заполненных ранее экземпляров объекта "Школа":

1. Державинский лицей
2. Школа имени Горчакова
3. Классическая гимназия №1 имени В. Г. Белинского

## 2 Полнота Викиданных

По данным Федеральной службы государственной статистики РФ[1], 2016 году в России насчитывалось 42,6 тыс. школ. По данным SPARQL-запроса, школ, находящихся на территории России, насчитывается всего 82. Имеет место быть огромная разница между цифрами Росстата и Викиданными.

То же самое можно наблюдать в отношении любой страны. Например, по данным ещё одного SPARQL-запроса, школ в США насчитывается 20, что, само собой разумеется, не соответствует действительности (по статистике — 28 220 школ в 2008 году [2]).

Трудности в оценке полноты Викиданных также вызывает то, что для огромного количества школ (4629 объектов), представленных в Викиданных, не указана страна (SPARQL-запрос). Такие школы составляют около 25% от общего числа школ в Викиданных. Это не позволяет отнести школу к какой-либо стране и сравнить официальные данные страны с Викиданными.

## 3 Заполнение Викиданных

Решено заполнить свойство "ученики (P802)" у российских школ.

Это свойство показывает список известных выпускников школы, людей, так или иначе отмеченных в истории. Рассмотрим в качестве примера Державинский лицей. Известной выпускницей лицея является Г. И. Ширшина, российский политический и общественный деятель.

Определим, сколько российских школ до момента выполнения задания не имели свойства "ученики (P802)" (Листинг 2).

```
1 SELECT ?school WHERE {
2   ?school wdt:P31 wd:Q3914. # school
3   ?school wdt:P17 wd:Q159. # Russian schools
4   FILTER NOT EXISTS { ?school wdt:P802 [] } # empty property "
5     student"
6 }
```

Листинг 2: Скрипт для построения списка российских школ без свойства "ученики"

По данным SPARQL-запроса (Листинг 2), школ без известных учеников оказалось 82.

Также было обнаружено 7 российских школ, имеющих известных учеников (по данным Википедии), но с незаполненным свойством "страна (P17)" на Викиданных. Поэтому в начале было заполнено свойство "страна (P17)" у этих объектов, таким образом, количество российских школ для заполнения известных учеников увеличилось на 7 и стало равняться 89.

В ходе заполнения свойства было обнаружено, что пара школ имели ложное свойство "страна (P17)" Россия, и потому это было исправлено, отчего количество российских школ слегка уменьшилось.

Также в ходе работы была выполнена дополнительная работа по заполнению Label на английском для известных выпускников, связанных свойством "ученики (P802)" с исследуемыми школами.

Итог работы таков: удалось заполнить свойство "ученики (P802)" у 45 российских школ. Остальные школы либо не имели известных выпускников, либо не были описаны в Википедии. Список отечественных школ с заполненным свойством "ученики (P802)" получим с помощью следующего скрипта:

```
1 SELECT ?school ?schoolLabel (count(*) as ?countStudents) WHERE {
2   #the object is the school
3   ?school wdt:P31 wd:Q3914.
4   { ?school wdt:P17 wd:Q34266 } UNION # Russian Empire
5   { ?school wdt:P17 wd:Q15180 } UNION # Soviet Union
6   { ?school wdt:P17 wd:Q159 }. # Russia
7   ?school wdt:P802 ?student.
8   SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
9 }
10
11 GROUP BY ?school ?schoolLabel
```

Листинг 3: Скрипт для построения списка отечественных школ с заполненным свойством "ученики"

SPARQL-запрос, 43 объекта, то есть 43 отечественных школы с известными учениками.



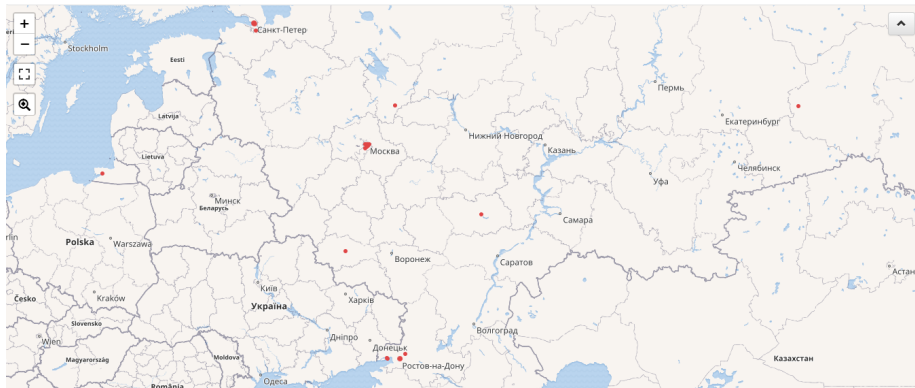


Рис. 2: Карта расположения российских школ с заполненным свойством “известные выпускники”

```

3 ?school wdt:P31 wd:Q3914. #school
4 ?school wdt:P17 wd:Q159. #russian schools
5 FILTER NOT EXISTS { ?school wdt:P802 []} #empty property "
6 student"
7 SERVICE wikibase:label { bd:serviceParam wikibase:language "ru" }

```

Листинг 5: Скрипт для построения списка отечественных школ без известных учеников

SPARQL-запрос, 41 объект.

## Список литературы

- [1] Number of state and municipal general education organizations. [http://www.gks.ru/free\\_doc/new\\_site/population/obraz/o-obr1.htm](http://www.gks.ru/free_doc/new_site/population/obraz/o-obr1.htm), 2016.
- [2] Thomas D. Snyder. *Digest of Education Statistics*. u.s. department of education, june 2011.