

Подсчёт количества университетов России и мира, их география

Andrew Krizhanovsky^a, Vladislav Mamaev^b

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

^b*Petrozavodsk State University*

Abstract

Статья посвящена исследованию объекта Викиданных — “Университет”. С помощью SPARQL-запросов к Викиданным, решены следующие задачи: построение списка всех университетов, построение пузырьковой диаграммы, показывающей соотношение университетов в разных странах, отображение на карте университетов, расположенных в России. В ходе работы дополнена информация у объектов Викиданных, соответствующих университетам России, сделаны выводы о полноте данных, представленных в Википедии и в Викиданных.

Примечание

Статья распространяется по лицензии Creative Commons Attribution-ShareAlike. Материалы этой статьи использованы в главе курса Викиверситета "Программирование Викиданных-[1]. Иллюстрации загружены на Викисклад. Над статьёй в 2017 году работали Крижановский А. А., Мамаев В. С.

1. Экземпляры объекта “Университет”

- Объект: университет (Q3918),
- Свойство: экземпляр (P31).

Построим список всех университетов. См. листинг 1.

```
1 #List of 'instances of' "university"
2 SELECT ?university ?universityLabel
3 WHERE
4 {
5     ?university wdt:P31 wd:Q3918. # instance of university
6     SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
7 }
```

Листинг 1: Список университетов

SPARQL-запрос, 13298 записей.

Наиболее полными и проработанными университетами на Викиданных являются: Токийский университет, Массачусетский технологический институт, Московский государственный университет.

Почти пустыми и малоинформативными университетами оказались: Московский государственный университет пищевых производств, Технический университет УГМК, Саратовский государственный социально-экономический университет.

1.1. Полнота Викиданных по университетам мира

По данным международной системы рейтинга университетов Webometrics Ranking of World Universities на Земле существует более чем 19 тысяч университетов. В основной список [2] входит около 12 тысяч из них. Некоммерческий рейтинг 4 International Colleges & Universities [3] включает информацию о 12 тысячах университетов и колледжей, имеющих веб-сайт.

По данным категории “Университеты по алфавиту” Русской Википедии существует более двух с половиной тысяч университетов. Единого списка всех университетов в Английской Википедии не существует, но есть списки по странам, категория “Universities and colleges by country”.

Данные, которые можно получить, просматривая эти категории, отличаются от тех, что используются в международных рейтингах университетов. Это связано с тем, что многие страницы университетов неправильно заполнены и в Википедии, и в Викиданных, имеют недостаточно информации и не относятся к рассмотренным категориям. Например, статья в Английской Википедии “Lincoln University of Business and Management” имеет лишь категорию “Companies based in Sharjah”, которая никак не связана с университетами.

1.2. Полнота Викиданных по университетам России

По данным сайта “Статистика Российского образования” [4] на 2004 год в России существовало 4157 университетов. Построим SPARQL-запрос, чтобы узнать, о скольких российских университетах есть информация в Викиданных. См. листинг 2.

```
1 #added 2017-03
2 #Number of universities in Russia
3 SELECT ?university ?universityLabel ?country
4 WHERE
5 {
6 ?university wdt:P31 wd:Q3918; # instance of university
7           wdt:P17 wd:Q159;      # with country Russia
8           rdfs:label ?item_label.
9
10     FILTER (LANG(?item_label) = "ru") .
11 SERVICE wikibase:label {
12 bd:serviceParam wikibase:language "ru".
13 }
14 }
```

Листинг 2: Список российских университетов

SPARQL-запрос, 436 записей.

В категории “Университеты России” Русской Википедии есть страница “Проект:Образование/Список университетов России”, которая содержит список из более чем 350 университетов. Категория “Universities in Russia” Английской Википедии содержит информацию о 77 университетах России.

Таким образом, в Википедии представлена информация лишь о десятой части университетов России.

2. Университеты в разных странах

- Объект: университет (Q3918),
- Свойства: государство (P17), экземпляр (P31).

Построим пузырьковую диаграмму, показывающую соотношение количества университетов разных стран мира. См. листинг 3.

```
1 #added 2017-03
2 #Universities in different countries
3 #defaultView:BubbleChart
4 SELECT ?university ?country (count(*) as ?count)
5 WHERE
6 {
7   ?univer wdt:P31 wd:Q3918 ; # instance of university
8     wdt:P17 ?university. # show country
9   OPTIONAL {
10    ?university rdfs:label ?country
11    filter (lang(?country) = "en")
12  }
13 }
14 GROUP BY ?university ?country
15 ORDER BY DESC(?count)
```

Листинг 3: Соотношение количества университетов разных стран мира

SPARQL-запрос, 199 записей.

Результат представлен на рис. 1. Итого, из 250 стран у двухсот есть университеты. Лидерами по количеству университетов являются: Соединённые Штаты Америки - 1608 учебных заведений, Индия - 930, Япония - 836. Россия находится на пятом месте и имеет 453 университета.

3. The map of Russian universities

- Объект: университет (Q3918),
- Свойства: государство (P17), географические координаты (P625), официальный сайт (P856), расположение штаб-квартиры (P159), экземпляр (P31).

Отобразим на карте университеты, расположенные в России (свойство "country"). Если у университета есть веб-сайт, добавим его во всплывающую подсказку (свойство "official website"). См. листинг 4.

```

1 #added 2017-02
2 #Locations of universities in Russia
3 #defaultView:Map
4 SELECT ?universityLabel ?universityDescription ?website ?coord
5 WHERE {
6 ?university wdt:P31 wd:Q3918; # instance of university
7 wdt:P17 wd:Q159;           # with country Russia
8 wdt:P625 ?coord.         # show coordinate location
9 OPTIONAL {
10 ?university wdt:P856 ?website # show official website
11 }
12 SERVICE wikibase:label {
13 bd:serviceParam wikibase:language "en,ru".
14 }
15 }

```

Листинг 4: Университеты на карте России

SPARQL-запрос, 205 записей.

3.1. *Incorrect filling*

ПетрГУ

На карте в результате выполнения скрипта не будет отображен Петрозаводский государственный университет (ПетрГУ). Вызвано это тем, что у объекта на Викиданных, соответствующего ПетрГУ, отсутствует свойство "coordinate location".

ТГУ

Как видно на рис. 2, Тверской государственный университет (ТГУ) отображается на карте в Атлантическом океане. Это также связано с некорректным заполнением свойства "coordinate location".

3.2. *100 objects*

Добавим данные о расположении ста университетов в Викиданные. Проверим успешность заполнения, еще раз выполнив SPARQL-запрос, представленный выше. Теперь он вернет 308 записей, результат на рис. ???. Из рисунка видим, что большая часть университетов расположена в европейской части России.

- [1] University. 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%A3%D0%BD%D0%B8%D0%B2%D0%B5%D1%80%D1%81%D0%B8%D1%82%D0%B5%D1%82, accessed 23 Feb 2017 (2017).
- [2] Webometrics Ranking of World Universities // Ranking of World Universities, <http://www.webometrics.info/en/world>, accessed 23 Feb 2017 (2017).
- [3] A-Z list of Universities and Colleges // 4 International Colleges & Universities (4ICU), <http://www.4icu.org/reviews/index2.htm>, accessed 23 Feb 2017 (2017).
- [4] Education statistics // Статистика Российского образования, http://stat.edu.ru/scr/db.cgi?act=listDB&t=2_6_2&ttype=2&Field=All, accessed 23 Feb 2017 (2004).

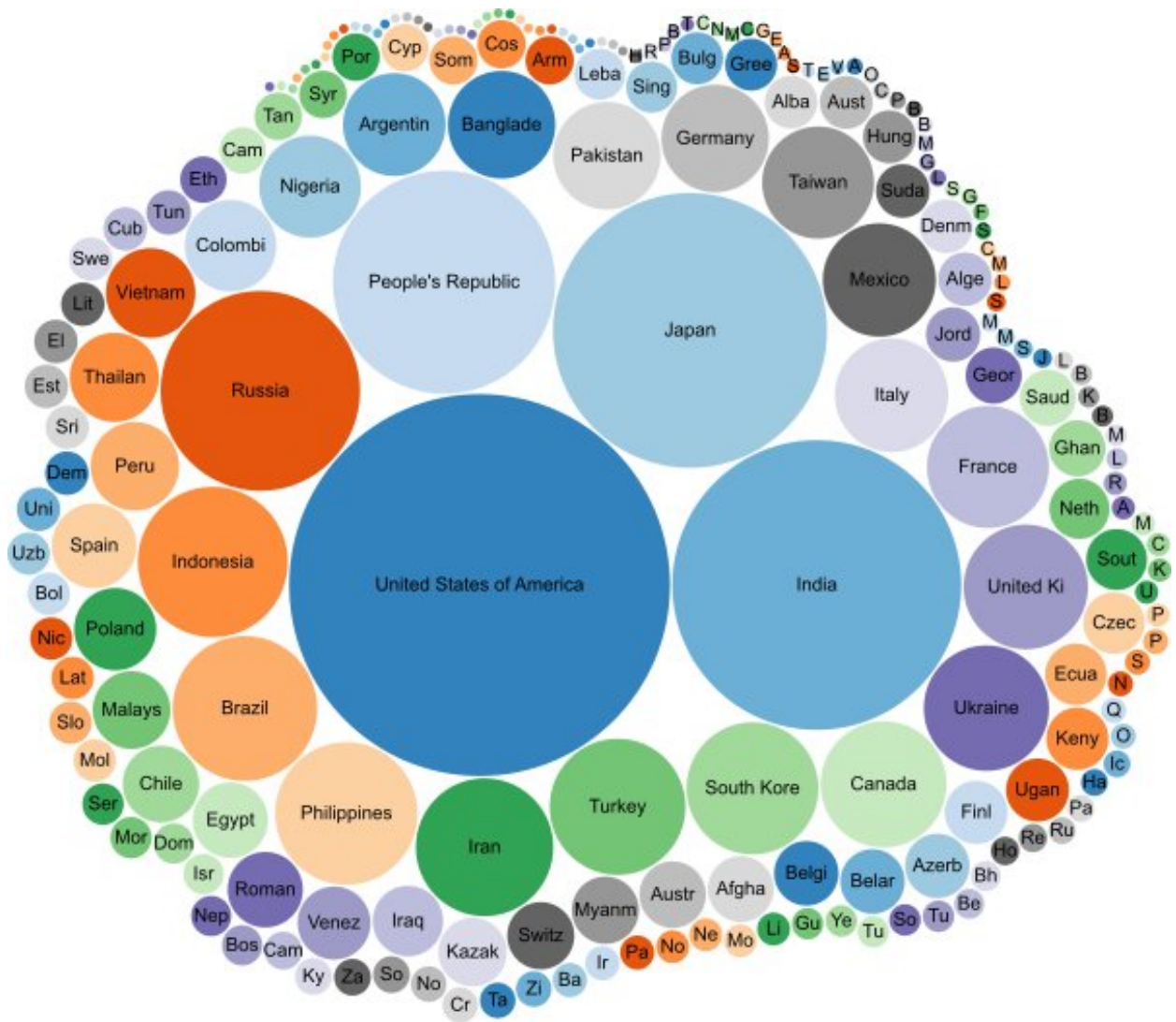


Рис. 1: Пузырьковая диаграмма, показывающая соотношение количества университетов разных стран мира.



Рис. 2: Карта расположения университетов России; некорректное заполнение свойства “coordinate location” у объекта Викиданных, соответствующего ТГУ, привело к тому, что он отображается в Атлантическом океане, эта ошибка была исправлена