

Поиск соседних регионов России и их классификация по численности населения с помощью языка запросов SPARQL

Andrew Krizhanovsky^a, Maria Khudorozhnikova^b

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

^b*Petrozavodsk State University*

Abstract

Аннотация

Статья посвящена исследованию наиболее важных свойств объекта Викиданных "Субъекты России". С помощью SPARQL-запросов были получены данные о количестве экземпляров объекта "Области России", о количестве всех существующих на данный момент времени субъектов Российской Федерации (рассматривались области России, республики, города федерального значения, края, автономные области, автономные округа, бывшие административно-территориальные единицы), построен граф соседних субъектов РФ и стран, а также нарисована карта, на которой отмечена численность населения отдельных субъектов Российской Федерации. Более того, рассматривалась задача, в которой исследовалась заполненность свойства "shares border with" (граничит с) у каждого экземпляра рассматриваемых объектов. По ходу работы пополнялись поля со свойствами в Викиданных. Читатель познакомится с компьютерной обработкой Викиданных и визуализацией информации о регионах России.

1. 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., Худорожникова М. Ю.

2. Экземпляры объекта области России

- Объект: области России (Q835714).
- Свойство: экземпляры (P31).

Построим список всех областей России. Построим список всех банков. См. листинг 1.

```

1 #List of 'instances of' "oblast of Russia"
2 SELECT ?oblast ?oblastLabel
3 WHERE
4 {
5   ?oblast wdt:P31 wd:Q835714.
6   SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
7 }

```

1: Список областей

SPARQL-запрос, 48 записей.

”+”: Наиболее полными и проработанными областями России на Викиданных являются: Московская область, Тюменская область, Курская область.

”-”: Почти пустыми и малоинформативными областями России оказались: Читинская область, Костромская область, Оренбургская область.

3. Субъекты Российской Федерации

- Объект: области России (Q835714).
- Объект: республики России (Q41162).
- Объект: города федерального значения России (Q183342).
- Объект: края России (Q831740).
- Объект: автономные области России (Q309166).
- Объект: автономные округа России (Q184122).
- Объект: бывшая административно-территориальная единица (Q19953632).
- Свойство: экземпляр (P31).

Построим список всех субъектов Российской Федерации: республики, края, области, города федерального значения, автономные области и автономные округа. См. листинг 2.

```

1 #List of 'instances of' "subjects of Russia"
2 SELECT ?subject ?subjectLabel
3 WHERE
4 {
5   { ?subject wdt:P31 wd:Q835714 } UNION # Oblast of Russia
6   { ?subject wdt:P31 wd:Q41162 } UNION # Republic of Russia
7   { ?subject wdt:P31 wd:Q183342 } UNION # Federal city of Russia
8   { ?subject wdt:P31 wd:Q831740 } UNION # Krai of Russia
9   { ?subject wdt:P31 wd:Q309166 } UNION # Autonomus oblast of Russia
10  { ?subject wdt:P31 wd:Q184122 } # Autonomus okrug of Russia
11
12 FILTER NOT EXISTS {?subject wdt:P31 wd:Q19953632} # Former administrative
   territorial entity

```

```

13 SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
14 }

```

2: Список субъектов

SPARQL-запрос, 85 записей.

4. Соседние субъекты

- Объект: области России (Q835714).
- Объект: республики России (Q41162).
- Объект: города федерального значения России (Q183342).
- Объект: края России (Q831740).
- Объект: автономные области России (Q309166).
- Объект: автономные округа России (Q184122).
- Объект: бывшая административно-территориальная единица (Q19953632).
- Свойство: экземпляр (P31).
- Свойство: граничит с (P47).

Построим граф соседних субъектов РФ по свойству "shares border with" (граничит с). См. листинг 3.

```

1 #Graph of "subjects of Russia" 'shares border with'.
2 #defaultView: Graph
3 SELECT ?subject ?subjectLabel ?sharesBorderWith ?sharesBorderWithLabel
4 WHERE
5 {
6   { ?subject wdt:P31 wd:Q835714 } UNION # Oblast of Russia
7   { ?subject wdt:P31 wd:Q41162 } UNION # Republic of Russia
8   { ?subject wdt:P31 wd:Q183342 } UNION # Federal city of Russia
9   { ?subject wdt:P31 wd:Q831740 } UNION # Krai of Russia
10  { ?subject wdt:P31 wd:Q309166 } UNION # Autonomus oblast of Russia
11  { ?subject wdt:P31 wd:Q184122 } # Autonomus okrug of Russia
12
13  FILTER NOT EXISTS {?subject wdt:P31 wd:Q19953632} # Former administrative
14  territorial entity
15  SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
16  ?subject wdt:P47 ?sharesBorderWith
17 }

```

3: Соседние субъекты

SPARQL-запрос, 467 записей.

Полученное число формируется путем сложения количества соседних территорий для всех субъектов России.

Результат работы скрипта - граф, отображающий соседние субъекты, представлен на рис. 1

На нем отчетливо видно изолированную компоненту, которая является Калининградской областью.

Фрагмент получившегося графа, на котором изображены соседствующие территории республики Карелия и Вологодской области, представлен на рис. 2

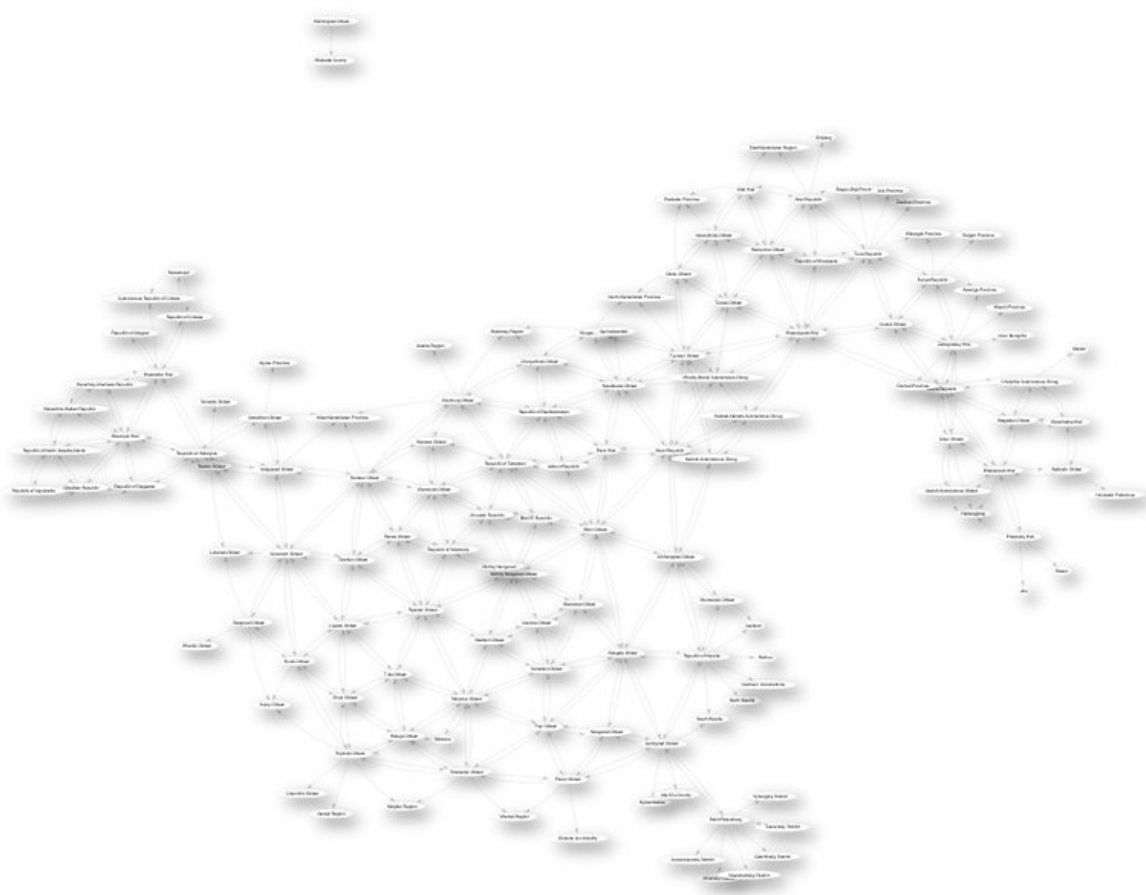


Figure 1: Результат работы скрипта - граф, отображающий соседние субъекты. Вершинами графа являются наименования субъектов и стран. Два любых соседствующих субъекта соединяются ребром

5. Полнота викиданных

Построим список субъектов РФ с пустым свойством "shares border with" (граничит с).

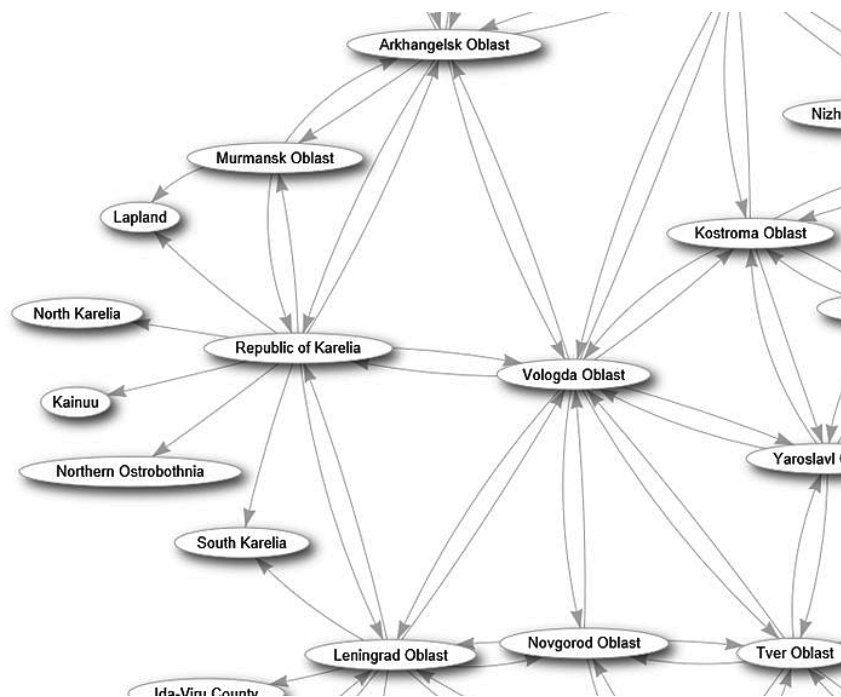


Figure 2: Фрагмент полученного графа, на котором изображены соседствующие территории республики Карелия и Вологодской области

- Объект: области России (Q835714).
- Объект: республики России (Q41162).
- Объект: города федерального значения России (Q183342).
- Объект: края России (Q831740).
- Объект: автономные области России (Q309166).
- Объект: автономные округа России (Q184122).
- Объект: бывшая административно-территориальная единица (Q19953632).
- Свойство: экземпляр (P31).
- Свойство: граничит с (P47).

См. листинг 4.

```

1 #List of "subjects of Russia" without 'shares border with'
2 SELECT ?subject ?subjectLabel ?sharesBorderWith ?sharesBorderWithLabel
3 WHERE
4 {
5   { ?subject wdt:P31 wd:Q835714 } UNION # Oblast of Russia

```

```

6 { ?subject wdt:P31 wd:Q41162 } UNION # Republic of Russia
7 { ?subject wdt:P31 wd:Q183342 } UNION # Federal city of Russia
8 { ?subject wdt:P31 wd:Q831740 } UNION # Krai of Russia
9 { ?subject wdt:P31 wd:Q309166 } UNION # Autonomus oblast of Russia
10 { ?subject wdt:P31 wd:Q184122 } # Autonomus okrug of Russia
11
12 FILTER NOT EXISTS {?subject wdt:P31 wd:Q19953632} # Former administrative
    territorial entity
13 MINUS { ?subject wdt:P47 [] } . #Shares border with
14 SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
15
16 }

```

4: Полнота

SPARQL-запрос, 0 записей.

Таким образом, на 8 марта 2017 года на Викиданных нет изолированных субъектов РФ, что соответствует действительности.

Информация, необходимая для решения задачи:

- По данным Конституции Российской Федерации Россия состоит из 85 субъектов: республик, краёв, областей, городов федерального значения, автономной области, автономных округов.
- В этой задаче не учитываются субъекты, которые на текущий момент времени не входят в состав РФ (например: Читинская область (Q182902)), поскольку они не являются экземплярами объектов "oblast of Russia", "republic of Russia", "federal city of Russia", "krai of Russia", "autonomus okrug of Russia", "autonomus oblast of Russia", а относятся к объекту "former administrative territorial entity" (бывшая административно-территориальная единица).
- В Русской Википедии существует 85 субъектов РФ по данным категории "Субъекты Российской Федерации".
- В Английской Википедии так же существует 85 субъектов РФ по данным категории "Federal subjects of Russia".

Численность населения отдельных субъектов Российской Федерации

Обозначим на карте субъекты Российской Федерации, разделив их на 6 групп по количеству населения. Субъекты, принадлежащие одной группе, будут отображаться на карте одним цветом, а именно:

- субъекты с количеством населения менее 500000 обозначаются синим цветом
- субъекты с количеством населения более 500000, но менее 1000000 обозначаются оранжевым цветом
- субъекты с количеством населения более 1000000, но менее 3000000 обозначаются зеленым цветом

- субъекты с количеством населения более 300000, но менее 8000000 обозначаются красным цветом
- субъекты с количеством населения более 800000, но менее 10000000 найдены не были
- субъекты с количеством населения более 1000000 обозначаются фиолетовым цветом

Используются:

- Объект: области России (Q835714).
- Объект: республики России (Q41162).
- Объект: города федерального значения России (Q183342).
- Объект: края России (Q831740).
- Объект: автономные области России (Q309166).
- Объект: автономные округа России (Q184122).
- Объект: бывшая административно-территориальная единица (Q19953632).
- Свойство: экземпляр (P31).
- Свойство: координаты (P625).
- Свойство: население (P1082).

См. листинг 5.

```

1 #Map of 'population' "subject of Russia"
2 #defaultView:Map
3 SELECT DISTINCT ?subject ?subjectLabel (SAMPLE(?location) AS ?location) (MAX
  (?population) AS ?population) (SAMPLE(?layer) AS ?layer)
4 WHERE
5 {
6   {
7     { ?subject wdt:P31 wd:Q835714 } UNION # Oblast of Russia
8     { ?subject wdt:P31 wd:Q41162 } UNION # Republic of Russia
9     { ?subject wdt:P31 wd:Q183342 } UNION # Federal city of Russia
10    { ?subject wdt:P31 wd:Q831740 } UNION # Krai of Russia
11    { ?subject wdt:P31 wd:Q309166 } UNION # Autonomus oblast of Russia
12    { ?subject wdt:P31 wd:Q184122 } # Autonomus okrug of Russia
13  }
14  ?subject wdt:P625 ?location .
15  ?subject wdt:P1082 ?population .
16  FILTER NOT EXISTS {?subject wdt:P31 wd:Q19953632} # former administrative
  territorial entity

```

```

17 BIND(
18   IF(?population < 500000, ">500000",
19   IF(?population < 1000000, "500000-1000000",
20   IF(?population < 3000000, "1000000-3000000",
21   IF(?population < 8000000, "3000000-8000000",
22   IF(?population < 10000000, "8000000-10000000",
23   "<10000000"))))
24   AS ?layer).
25 SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
26 }
27 GROUP BY ?subject ?subjectLabel
28 ORDER BY ?population

```

5: Численность населения

SPARQL-запрос, 85 записей.

Результат работы скрипта представлен на рис. 3

На карте видно, что восточная часть России является самой малонаселенной, в то время как в западной части России число проживающих жителей является самым высоким по все России.

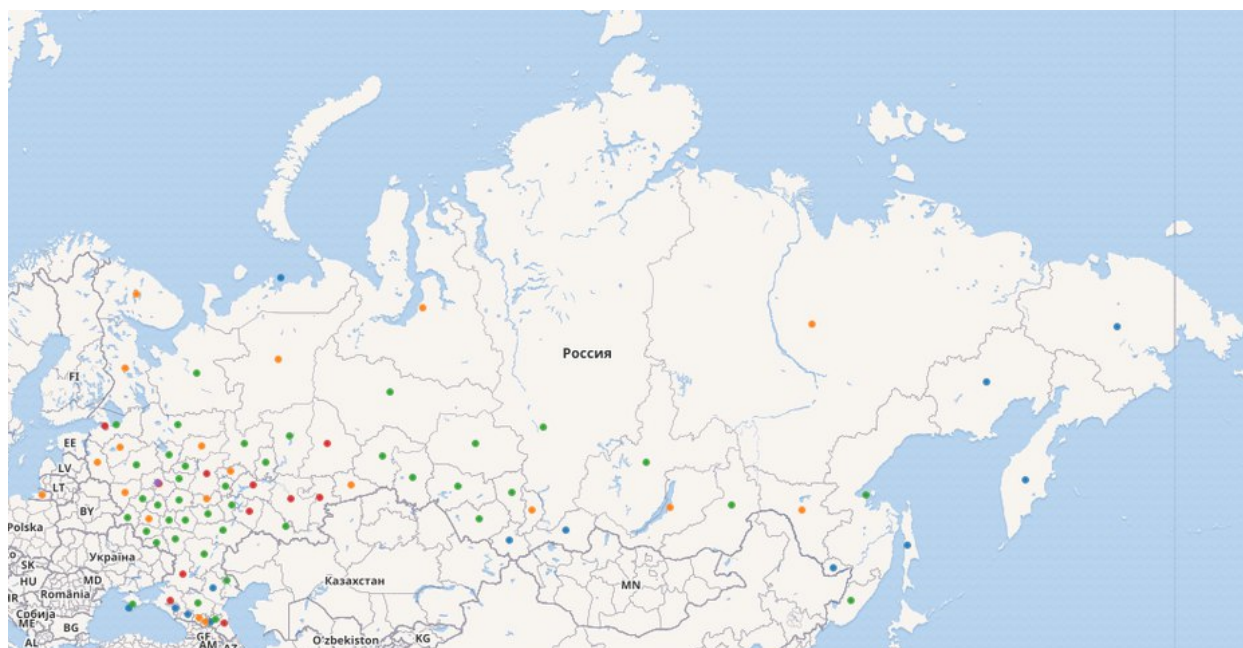


Figure 3: Карта субъектов Российской Федерации, разделенных на 6 групп по количеству населения и отмеченных разными цветами в зависимости от группы, в которую субъект входит

[1] Oblast of Russia. 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%9E%D0%B1%D0%BB%D0%B0%D1%81%D1%82%D0%B8_%D0%A0%D0%BE%D1%81%D1%81%D0%B8%D0%B8, accessed 12 Mar 2017 (2017).