

Исследование национальных парков России и мира по Викиданным

Andrew Krizhanovsky^a, Nikita Bazhenov^b

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

^b*Petrozavodsk State University*

Аннотация

Данная статья посвящена исследованию объекта Викиданных "Национальный парк". С помощью SPARQL-запросов, вычисляемых на объектах типа "национальный парк" в Викиданных, решены такие задачи: выведен список всех ныне существующих национальных парков, список национальных парков, упорядоченных по дате создания, диаграмма парков, упорядоченных по количеству за разные годы и по странам мира, а также карта всех национальных парков, построенная на основе географических координат. Кроме того, сделаны выводы по поводу полноты Викиданных по данной теме.

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., Баженов Н. А.

Экземпляры объекта "Национальный парк"

Национальный парк - территория, где в целях охраны окружающей среды ограничена деятельность человека. В дальнейшем, возможно, будет использовано сокращение слова "национальный парк" до "парк" с целью лучшего понимания текста.

Построим список всех национальных парков. См. листинг 1.

- Свойство: экземпляр (P31),
- Объект: национальный парк (Q46169).

```
#List of 'instances of' "national park"
SELECT ?park ?parkLabel
WHERE
{
    ?park wdt:P31 wd:Q46169.
    SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
```

Листинг 1: Список парков

28 SPARQL-запрос, 1665 записей.

29 Наиболее полными и проработанными национальными парками на Викиданных являются: Тейде, Тингветлир, Национальный парк имени Александра Гумбольдта.

31 Почти пустыми и малоинформативными национальными парками оказались: Припышминские Боры, Смольный национальный парк, Хвалынский национальный парк.

33 **Диаграмма парков, упорядоченных по количеству за разные годы и по** 34 **странам мира**

35 Выведем диаграмму парков, упорядоченных по количеству созданных парков за
36 разные годы и по странам мира (начиная с 1900 года и по сей день). См. листинг 2.

```
37 1 #defaultView:BarChart
38 2 #underscore is for using a variable more than 1 time
39 3 SELECT DISTINCT (SAMPLE(?year) AS ?year) (COUNT(?year) AS ?count) (SAMPLE(?
40    parkLabel) AS ?parkLabel) WHERE {
41 4    ?object (wdt:P31) wd:Q46169. #instance of national park
42 5    BIND(str(YEAR(?inception)) AS ?year) #definition of year
43 6    ?object wdt:P571 ?inception. #definition of inception
44 7    ?object wdt:P17 ?country. #definition of country
45 8    ?country rdfs:label ?parkLabel.
46 9    FILTER((LANG(?parkLabel)) = "ru") #filter in parkLabel = ru
47 0 }
48 1 GROUP BY ?inception ?country
49 2 ORDER BY ?year ?inception
```

Листинг 2: Упорядоченная по количеству за разные годы и по странам мира диаграмма парков

50 SPARQL-запрос, 612 записей.

51 Данный скрипт выводит количество созданных парков в определенные года, а так
52 же страны, в которых они были созданы (рис. 1). Например, столбик голубого цвета
53 за 2006 год означает, что было создано 25 национальных парков в Австралии.

54 Отсортируем столбцы на гистограмме таким образом, чтобы года отображались
55 последовательно (рис. 2).

56 Недостаток полученного скрипта в том, что, возможно, не все национальные парки
57 мира выведутся. Это связано с тем, что у национальных парков мира существует под-
58 класс национальных парков по странам (например, национальные парки Норвегии).
59 Дополним предыдущую диаграмму, включив в нее эти подклассы. См. листинг 3.

```
60 1 #defaultView:BarChart
61 2 SELECT DISTINCT (SAMPLE(?year) AS ?year) (COUNT(?year) AS ?count) (SAMPLE(?
62    parkLabel) AS ?parkLabel) WHERE {
63 3    ?object (wdt:P31/wdt:P279*) wd:Q46169. #instance of national park of .. (
64 4      Russia as example)
65 5    BIND(str(YEAR(?inception)) AS ?year) #definition of year
66 6    ?object wdt:P571 ?inception. #definition of _inception
67 7    ?object wdt:P17 ?country. #definition of _country
68 8    ?country rdfs:label ?parkLabel.
69 9    FILTER((LANG(?parkLabel)) = "ru") #filter in parkLabel = ru
70 0 }
71 1 GROUP BY ?inception ?country
```

```
72 1 ORDER BY ?year ?inception
```

Листинг 3: Упорядоченная диаграмма парков с отсортированным столбцами

73 SPARQL-запрос, 980 записей.

74 Можно заметить по (рис. 3), что количество национальных парков выросло на 378.

75 Полнота Викиданных по паркам

76 Существует много национальных парков мира. Однако, скорее всего, далеко не
77 у всех заполнено поле 'географические координаты' (англ. 'location') Построим спи-
78 сок национальных парков, у которых заполнены географические координаты. См. ли-
79 стинг 4.

```
80 1 #List of parks with filled 'location '  
81 2 #defaultView:Map  
82 3 SELECT ?park ?parkLabel ?location  
83 4 WHERE  
84 5 {  
85 6   ?park wdt:P31 wd:Q46169.  
86 7   ?park wdt:P625 ?location  
87 8   SERVICE wikibase:label { bd:serviceParam wikibase:language "en"  
88 9 }
```

Листинг 4: Список парков с заполненными географическими координатами

89 SPARQL-запрос, 1405 записей.

90 Построим список национальных парков, у которых заполнены географические ко-
91 ординаты, расположенных в России. См. листинг 5.

```
92 1 #List of national parks in Russia  
93 2 #defaultView:Map  
94 3 SELECT ?park ?parkLabel ?location  
95 4 WHERE  
96 5 {  
97 6   ?park wdt:P31 wd:Q46169. #instance of national park  
98 7   ?park wdt:P17 wd:Q159. #country = Russia  
99 8   ?park wdt:P625 ?location #display location  
100 9   SERVICE wikibase:label { bd:serviceParam wikibase:language "en"  
101 0 }
```

Листинг 5: Список парков с заполненными географическими координатами России

102 SPARQL-запрос, 7 записей.

103 Национальных парков по России (которые вывелись данным скриптом) оказалось
104 очень мало, всего лишь семь. Хотя, на самом деле, их шесть, потому что один парк
105 (Национальный парк Бикин) дублируется ввиду повторного заполнения поля 'location'.
106 На самом деле, эти парки являются 'ошибочными', так как они должны входить в
107 подкласс парков конкретной страны.

108 Построим список национальных парков России, используя подкласс (national park
109 in Russia). См. листинг 6.

```
110 1 #List of national parks in Russia  
111 2 #defaultView:Map  
112 3 SELECT ?park ?parkLabel ?location
```

```

113 4 WHERE
114 5 {
115 6   ?park wdt:P31 wd:Q1969226. #instance of national park in Russia
116 7   ?park wdt:P625 ?location #display location
117 8   SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
118 9 }

```

Листинг 6: Список парков с заполненными географическими координатами в России

119 SPARQL-запрос, 44 записи.
 120 Полученный список состоит из 44 парков (рис. 4).
 121 Обратимся к статье на Русской Википедии. Список национальных парков России
 122 включает 49 парков.
 123 В книге "Вокруг Света. Национальные парки России. Поволжье и Северный Кавказ.
 124 Путеводитель" говорится на момент написания (2007 год) о действующих 35 националь-
 125 ных парках России.
 126 Викигид показывает результат в 40 национальных парков в России.
 127 **Заполнение 100 объектов**
 128 Информация на Викиданных говорит о том, что не у всех парков заполнено по-
 129 ле 'географические координаты' (англ.'location'). Напишем скрипт, который выводит
 130 список всех национальных парков с незаполненным полем 'location'.
 131 SPARQL-запрос, 283 записи.

```

132 1 #List of national parks with unfilled property 'location '
133 2 SELECT ?park ?parkLabel ?location
134 3 WHERE
135 4 {
136 5   ?park wdt:P31 wd:Q46169. #instance of national park
137 6   FILTER NOT EXISTS { ?park wdt:P625 [] } #if property location is absent
138 7   SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
139 8 }

```

Листинг 7: Список парков с незаполненными географическими координатами

140 Данный скрипт построил список из 283 национальных парков с незаполненным по-
 141 лем 'location'. См. листинг 7.
 142 Напишем скрипт, который построит список из всех национальных парков, у которых
 143 заполнено поле 'location'.
 144 SPARQL-запрос, 1417 записей.

```

145 1 #List of national parks
146 2 #defaultView:Map
147 3 SELECT ?park ?parkLabel ?location
148 4 WHERE
149 5 {
150 6   ?park wdt:P31 wd:Q46169. #instance of national park
151 7   ?park wdt:P625 ?location #display location
152 8   SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
153 9 }

```

Листинг 8: Список парков с заполненными географическими координатами

154 После выполнения листинга на карте отобразилось 1417 национальных парка. См.
 155 листинг 8. Необходимо заполнить поле 'location', чтобы количество выводимых парков
 156 оказалось больше.

157 Как можно заметить по изображению (рис. 5), количество выведенных парков стало
 158 больше, ввиду заполнения поля. А страной с наибольшим количеством национальных
 159 парков (312) оказалась Австралия.

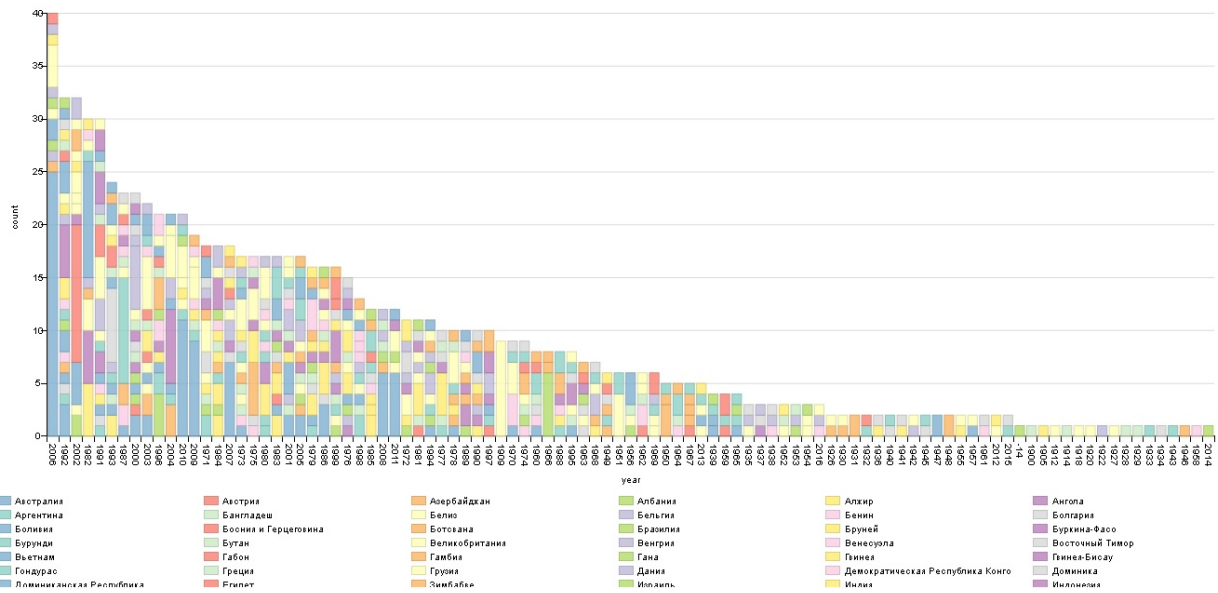
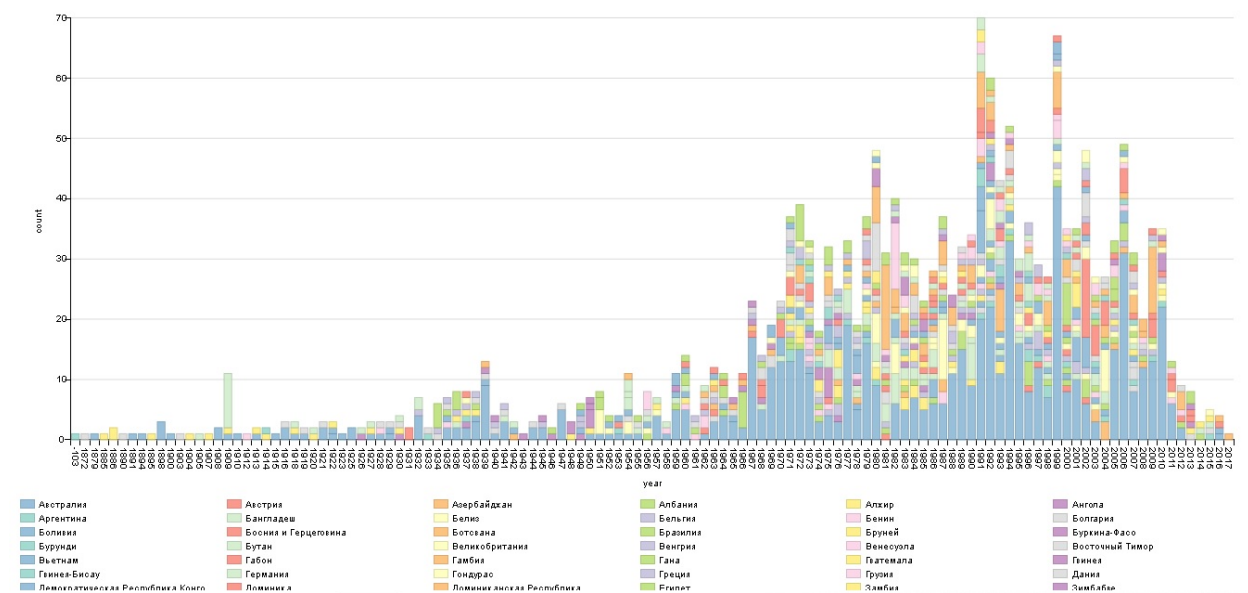
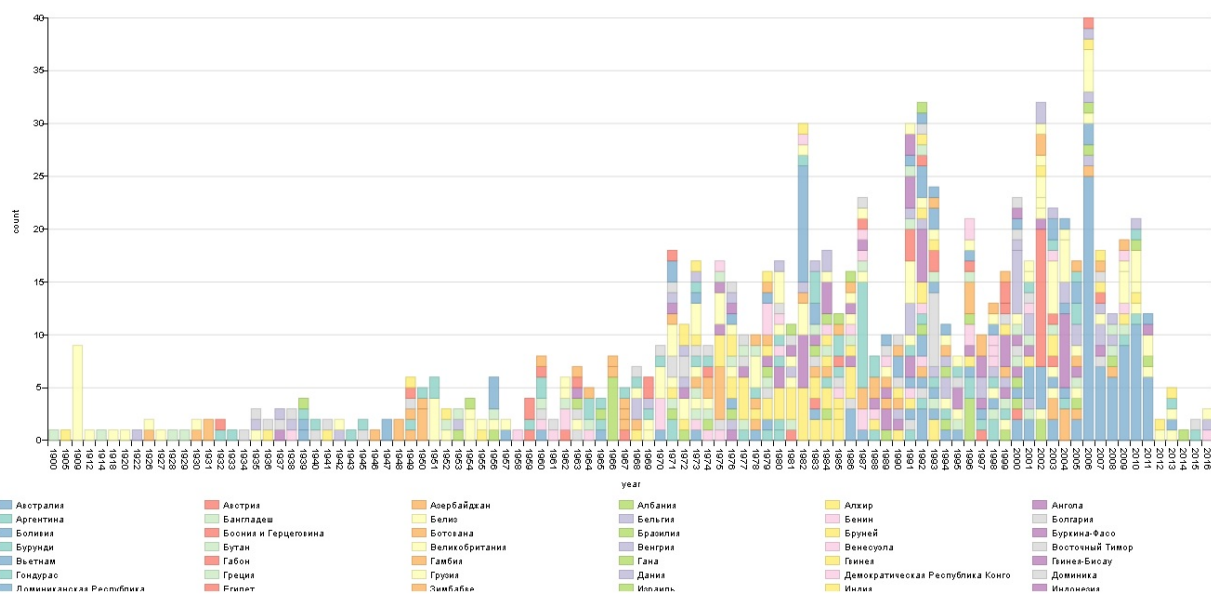


Рис. 1: Гистограмма парков, упорядоченных по количеству за разные года и по странам мира



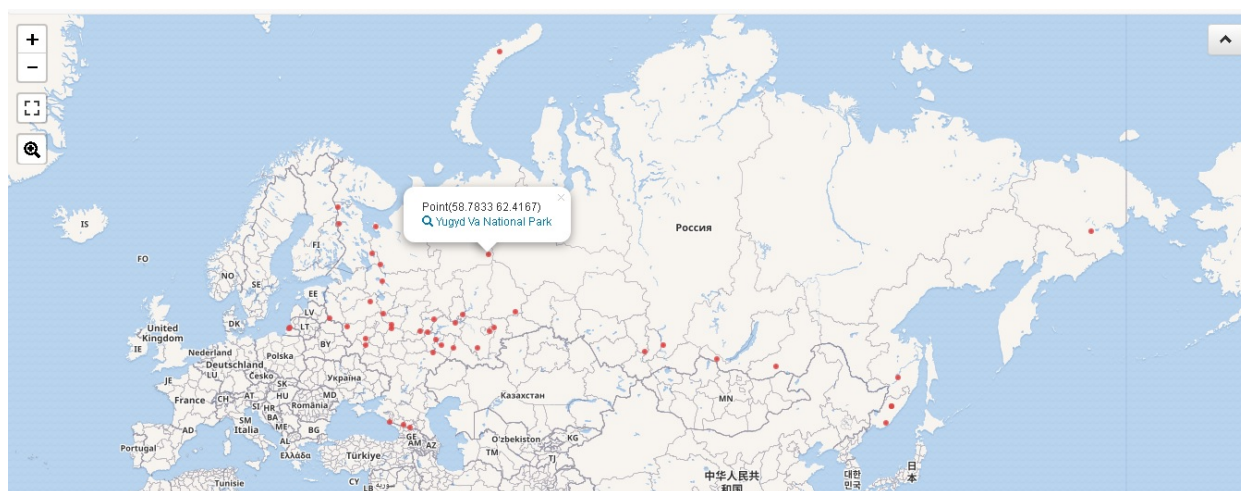


Рис. 4: Карта национальных парков России

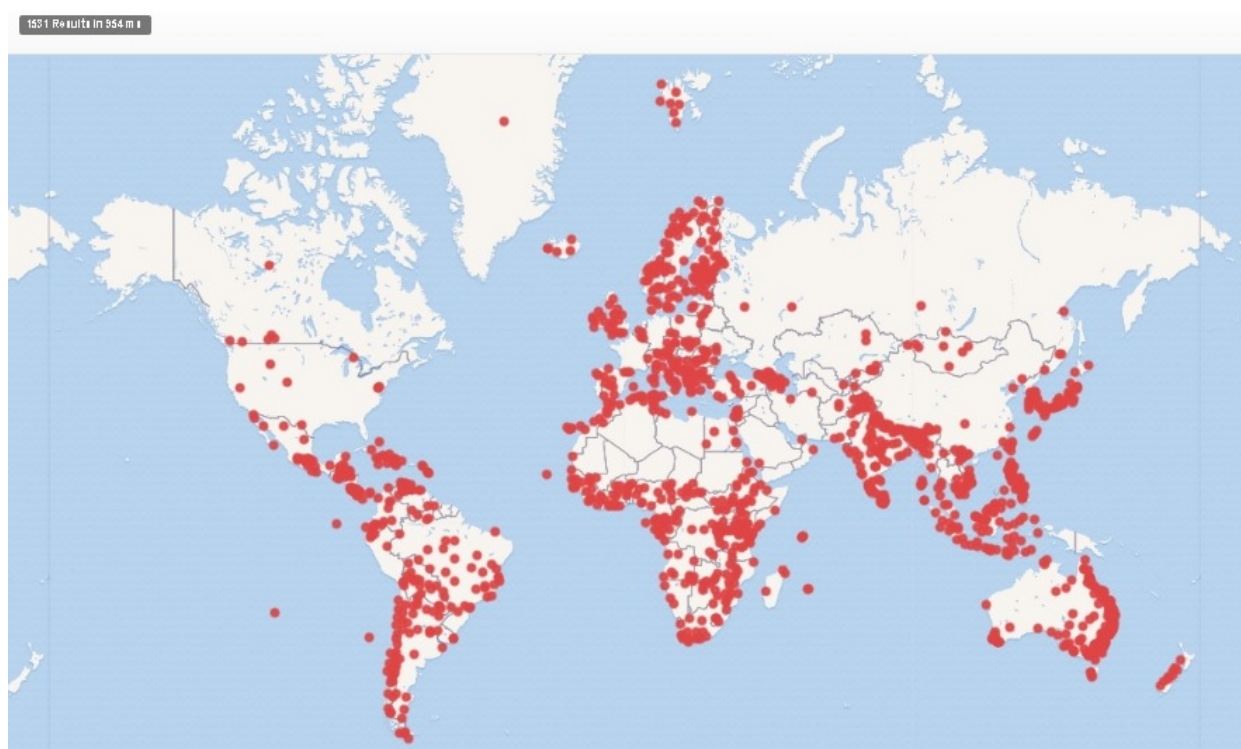


Рис. 5: Карта всех национальных парков, построенная на основе свойства 'географические координаты' (англ. 'location')

160 [1] National Park. Research in programming Wikidata [Programmirovaniye Vikidannykh] // Russian
161 Wikiversity,