

УДК 519.689

АНАЛИЗ ТРЁХ АСПЕКТОВ СОВРЕМЕННЫХ СТРАН ПО ВИКИДАНЫМ: ВОЗРАСТ СТРАН, ПОПУЛЯРНЫЕ ФОРМЫ ПРАВЛЕНИЯ И ЭТНОХОРОНИМЫ

Е. М. Смыкова¹, А. А. Крижановский²

¹Смыкова Елизавета Михаловна, студентка Петрозаводского Государственного Университета, Россия, 185910, Петрозаводск, Ленина, 33, esmykova8@gmail.com

²Крижановский Андрей Анатольевич, кандидат технических наук, руководитель Лаборатории информационных компьютерных технологий Института прикладных математических исследований, Россия, 185910, Петрозаводск, Пушкинская, 11, andrew.krizhanovsky@gmail.com

Эта статья посвящена исследованию стран на основе базы знаний международного проекта Викиданные. С помощью SPARQL-запросов, вычисляемых на объектах типа "страна" в Викиданных, получены: выведен список всех ныне существующих стран, перечень стран, упорядоченных по дате создания, список этнохоронимов стран, пузырьковая диаграмма с формами правления стран и граф соседних стран. Кроме того, сделаны относительно полноты Викиданных по данной теме.

Ключевые слова: Викиданные, анализ данных, страны.

1. ЭКЗЕМПЛЯРЫ ОБЪЕКТА СТРАНА

Построим список всех стран.

- Объект: страна (Q6256),
- Свойство: экземпляр (P31).

```
1 #List of 'instances of' "country"
2 SELECT ?lang ?langLabel
3 WHERE
4 {
5     ?lang wdt:P31 wd:Q6256.
6     SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
7 }
```

Листинг 1: Список стран

SPARQL-запрос, 198 записей

"+": Примерами наиболее полных и проработанных стран на Викиданных являются: Соединённые Штаты Америки, Канада, Испания.

"–": Почти пустыми и малоинформативными странами оказались: Сахарская Арабская Демократическая Республика, Приднестровская Молдавская Республика, Косово.

2. ВОЗРАСТ СТРАН

Построим список стран, отсортированных по дате основания страны (первом упоминании о стране).

- Объект: страна (Q6256),
- Свойство: дата основания (P571).

```
1 #List of 'instances of' "countries sorted by inception"
2 SELECT ?country ?countryLabel ?inception
3 WHERE
4 {
5     ?country wdt:P31 wd:Q6256.          #country
6     ?country wdt:P571 ?inception .    #inception of country
7
8     SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
9 }
10 ORDER BY (?inception)
```

Листинг 2: Список стран упорядоченных по дате создания

SPARQL-запрос, 112 записей.

В результате выполнения запроса 2 получен список стран с датами их создания. Например, Абхазия – 1 января 0786, Россия – 1 января 0862, Косово – 17 февраля 2008, Южный Судан – 9 июля 2011.

Годы, в которые было создано наибольшее количество стран – 1991 (17 стран), 1812 (6 стран) и 1918 (5 стран).

3. ПОЛНОТА ВИКИДАННЫХ

Проанализируем полноту Викиданных.

- По данным "Общероссийского классификатора стран мира- [1] за 2016 год на земле существует 251 страна.
- В этой задаче не учитываются древние, уже не существующие государства (например: Ассирия (Q41137)), поскольку они являются экземпляром не объекта "country а объекта "former country"(бывшие страны). Отметим, что количество бывших стран на порядок больше существующих ныне стран (см. SPARQL-запрос, возвращающий более двух тысяч таких стран).
- По данным категории "Алфавитный список стран и территорий" Русской Википедии существует 252 страны. (В "Общероссийском классификаторе стран мира"недостаёт Косово)
- По данным категории "List of sovereign states" Английской Википедии существует 206 стран.

Не всегда можно точно указать дату основания страны по разным причинам: отсутствие, недостаток или противоречие письменных источников. Например, основание Древнерусского государства связывают с призыванием варяжского князя Рюрика в 862 году, но точной даты нет (объект Russia (Q159)). Так же некоторым современным странам предшествовал ряд других и дату образования какого из них считать за дату создания современной страны – это вопрос открытый (например, Монголия (Q711)).

4. СТРАНЫ С НЕЗАПОЛНЕННОЙ ДАТОЙ ОСНОВАНИЯ

Выведем список стран с пустым свойством "дата основания":

- Объект: страна (Q6256),
- Свойство: дата основания (P571).

```
1 #List of 'instances of' "countries without an inception"
2 SELECT ?country ?countryLabel
3 WHERE
4 {
5     ?country wdt:P31 wd:Q6256.           #country
6
7     MINUS { ?country wdt:P571 [] } . #inception of country is empty
8     SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
9 }
```

Листинг 3: Список стран с пустым свойством "дата создания"

SPARQL-запрос, 100 записей.

Итак, на 6 марта 2017 года Викиданные содержат 100 из 198 записей о ныне существующих странах с неизвестным годом основания страны. 3

5. ЭТНОХОРОНИМЫ НА РУССКОМ ЯЗЫКЕ

Этнохороним – название жителей определённой местности соотнесённое с топонимом. Например, Россия – россияне, россиянин, россиянка, Чехия – чехи, чех, чешка.

Помимо географического мотиватора, новых лексемы, используемые для определения происхождения либо принадлежности, происходят так же от этнических, политических, религиозных характеристик людей.

Название жителей может исходить от наименования различных объектов земной поверхности — гор, островов, континентов. Так же обозначение места происхождения людей может зависеть от политико-административного деления. Например, для обозначения гражданства; Украина – украинцы, Канада – канадцы. Внутригосударственное деление так же может породить новые наименования, Крым – крымчане. [2]

Построим список стран у которых есть этнохоронимы на русском языке.

- Объект: страна (Q6256),
- Свойство: этнохороним (P1549).

```

1 #List of countries with demonyms in Russian
2 SELECT ?country ?countryLabel
3 WHERE
4 {
5     ?country wdt:P31 wd:Q6256. #country
6     ?country wdt:P1549 ?demonym . #demonym
7
8     FILTER((LANG(?demonym)) = "ru")
9
10    SERVICE wikibase:label { bd:serviceParam wikibase:language "ru" }
11 }
12 GROUP BY ?country ?countryLabel

```

Листинг 4: Список стран с этнохоронимами на русском языке

SPARQL-запрос, 28 записей.

На 19 марта 2017 года Викиданные содежали 28 из 198 стран с заполненными этнохоронимами на русском языке. 4

6. СПИСОК ЭТНОХОРОНИМОВ

Выведем список всех этнохоронимом на русском языке.

```

1 #List of demonyms in Russian
2 SELECT ?country ?countryLabel ?demonym
3 WHERE
4 {
5     ?country wdt:P31 wd:Q6256. #country
6     ?country wdt:P1549 ?demonym. #demonym
7
8     FILTER((LANG(?demonym)) = "ru")
9
10    SERVICE wikibase:label { bd:serviceParam wikibase:language "ru" }
11 }

```

Листинг 5: Список этнохоронимов на русском языке

SPARQL-запрос, 83 записей.

На 19 марта 2017 года Викиданные содежат 83 заполненных этнохоронима. 5

7. СТРАНЫ С НЕЗАПОЛНЕННЫМИ ЭТНОХОРОНИМАМИ

Построим список стран, у которых нет этнохоронимов на русском языке.

```

1 #List of countries without demonyms in Russian
2 SELECT ?country ?countryLabel
3 WHERE
4 {
5     ?country wdt:P31 wd:Q6256. # country
6
7     MINUS {
8         ?country wdt:P1549 ?demonym. # except with demonyms
9         FILTER((LANG(?demonym)) = "ru") # in Russian
10    }
11    SERVICE wikibase:label { bd:serviceParam wikibase:language "ru" }

```

```
12 }  
13 GROUP BY ?country ?countryLabel
```

Листинг 6: Список стран у которых нет этнохоронимов на русском языке

SPARQL-запрос, 170 записей.

На 19 марта 2017 года Викиданные содержат 170 из 198 стран с незаполненными этнохоронимами. 6

После заполнения авторами статьи этнохоронимов, число стран, у которых их нет на русском языке стало 69, а самих этнохоронимов на русском языке теперь 276 (на 20 марта 2017 года).

8. КОЛИЧЕСТВО ЗАПОЛНЕННЫХ ЭТНОХОРОНИМОВ У СТРАН

Выведем список стран, упорядоченный по количеству заполненных в Викиданных этнохоронимов.

```
1 #Count of demonyms in countries  
2 SELECT ?country ?countryLabel (count(*) as ?count)  
3 WHERE  
4 {  
5     ?country wdt:P31 wd:Q6256. # country  
6     ?country wdt:P1549 ?demonym. # demonyms  
7     SERVICE wikibase:label { bd:serviceParam wikibase:language "ru" }  
8 }  
9 GROUP BY ?country ?countryLabel  
10 ORDER BY DESC(?count)
```

Листинг 7: Страны упорядоченные по количеству заполненных этнохоронимов

SPARQL-запрос, 199 записей.

Наибольшее число этнохоронимов у Соединённых Штатов Америки (41 этнохороним), затем идут Великобритания (40), Германия (40), Канада (36) и Россия (34). 7

9. ФОРМЫ ПРАВЛЕНИЯ СТРАН

Построим пузырьковую диаграмму форм правления стран.

- Объект: страна (Q6256),
- Свойство: форма правления (P122).

```
1 #basic form of government ranking  
2 #defaultView:BubbleChart  
3 SELECT ?bfog ?form (count(*) as ?count)  
4 WHERE  
5 {  
6     ?country wdt:P31 wd:Q6256. #country  
7     ?country wdt:P122 ?bfog . #basic form of government  
8     OPTIONAL {  
9         ?bfog rdfs:label ?form  
10        filter (lang(?form) = "ru")  
11    }
```

```
}
GROUP BY ?bfog ?form
ORDER BY DESC(?count) ASC(?form)
```

Листинг 8: Пузырьковая диаграмма форм правления стран

SPARQL-запрос, 30 записей.

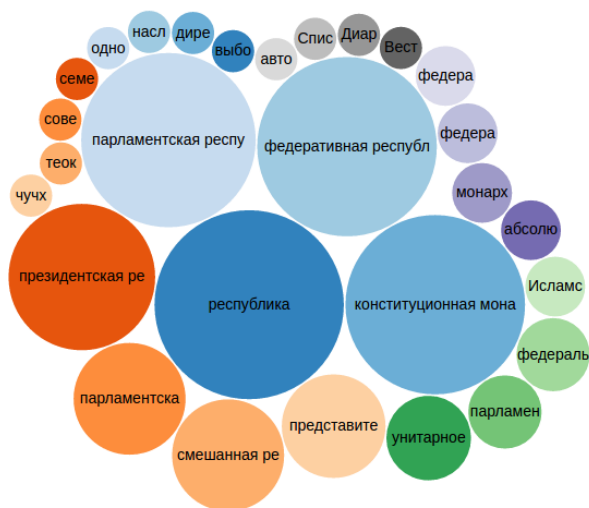


Рис. 1: Пузырьковая диаграмма форм правления стран.

В результате выполнения запроса 8 мы получаем пузырьковую диаграмму 1 с наиболее распространенными формами правления в странах. Основные формы правления стран: республика (в 20 странах), конституционная монархия (в 18 странах), федеративная республика (в 18 странах), парламентская республика (в 17 странах) и президентская республика (в 12 странах).

10. СОСЕДНИЕ СТРАНЫ

Построим граф соседних стран.

- Объект: страна (Q6256),
- Свойство: имеет границы с (P47).

```

1 #neighboring countries graph
2 #defaultView:Graph
3 SELECT ?country ?countryLabel ?sharesBorderWith ?sharesBorderWithLabel
4 WHERE
5 {
6     ?country wdt:P31 wd:Q6256. #countries
7
8     SERVICE wikibase:label { bd:serviceParam wikibase:language "ru" }
9     OPTIONAL { ?country wdt:P47 ?sharesBorderWith . } #shares border with
10 }

```

Листинг 9: Граф соседних стран

Библиографический список

1. *Общероссийский классификатор стран мира* URL: <http://classifikators.ru/oksm> (дата обращения: 11.11.2017).
2. Н.Ю. Журавлева Особенности перевода катойконимов (на материале испанского и русского языков) // Вестник Российского университета дружбы народов. — 2012. — Р. 41–48.

A. A. Krizhanovsky¹, E. M. Smykova²

¹Elizaveta M. Smykova, ORCID: 0000-0002-0426-2046, Petrozavodsk State University, 33, Lenina str., Petrozavodsk, Russia, 185910, esmykova8@gmail.com

²Andrew A. Krizhanovsky, ORCID: 0000-0002-7926-1347, Institute of Applied Mathematical Research of the Karelian Research Centre of the Russian Academy of Sciences, 11, Pushkinskaya str., Petrozavodsk, Russia, 185910, andrew.krizhanovsky@gmail.com

This research is devoted to the study of countries based on the knowledge base of the Wikidata international project. SPARQL queries were used in order to analyse and compare "countries" objects in Wikidata. A list of all currently existing countries, a list of countries ordered by date of creation, a list of demonyms of countries were generated. A bubble chart with the forms of government of countries and a graph of neighboring countries were constructed. In addition, conclusions were drawn regarding the completeness of the Wikidata for this topic.

Key words: Wikidata, data analysis, countries.

References

1. *Russian classification of countries of the world* URL: <http://classifikators.ru/oksm> (Accessed 11, November, 2017).
2. N.Y. Zhuravleva Features of translation katoikonimov (on a material of the Spanish and Russian languages) // Bulletin of the Peoples Friendship University of Russia. — 2012. — P. 41–48.