

### Примечание

Статья распространяется по лицензии Creative Commons Attribution-ShareAlike. Материалы этой статьи использованы в главе курса Викиверситета "Программирование Викиданных"[2]. Иллюстрации загружены на Викисклад. Над статьёй в 2017 году работали Крижановский А. А., Николаев Н. С.

### Аннотация

Статья посвящена исследованию объекта Викиданных "коммерческие организации". С помощью SPARQL-запросов, вычисляемых на объектах типа "коммерческие организации" в Викиданных, решены такие задачи: выведен список с распределением организаций по отраслям в виде пузырьковой диаграммы, построен граф существующих организаций и их дочерних организаций и получена информация о количестве организаций в различных странах. Сделаны выводы по поводу полноты Викиданных по данной теме и построена карта организаций мира.

### Экземпляры объекта "Коммерческие организации"

Используются:

- Объект: business enterprise (Q4830453) (коммерческая организация)

Список всех коммерческих организаций.

```
#added 2017-02
```

```
#List of 'instances of' "business enterprise"
```

```
SELECT ?lang ?langLabel
```

```
WHERE
```

```
{
  ?lang wdt:P31 wd:Q4830453.
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
}
```

SPARQL-query, 109383 записей.

Наиболее полными и проработанными по теме business enterprise на Викиданных являются: Google, Apple, Microsoft.

Почти пустыми и малоинформативными языками оказались: Pininfarina, ANHUI EXPRESSWAY COMPANY LIMITED, Futura et Marge.

### Безымянные объекты

Недостаток полученного списка в том, что ряд объектов получился безымянным на Викиданных (No label defined). Попробуем получить список банков, у которых поле "label" будет непустым.

```
#List of 'instances of' "business enterprise" only with a label.
```

```
SELECT ?item ?item_label
```

```
WHERE
```

```
{
  ?item wdt:P31 wd:Q4830453
  ; rdfs:label ?item_label .

  FILTER (LANG(?item_label) = "en") .
}
```

SPARQL-запрос, 74556 записей.

### Распределение организаций по отраслям

Каждая организация специализируется на какой-либо отрасли. Для того чтобы понять какая отрасль, например, является самой популярной на данный момент (то есть, сколько организаций работают в данной отрасли) можно построить диаграмму.

Тип результата: пузырьковая диаграмма.

Используются:

- Объект: business enterprise (Q4830453) (коммерческая организация)
- Свойство: industry (P542) (отрасль)



Ответим на следующий вопрос: Какие и сколько отраслей существуют в России?

```
#enterprise industry ranking in Russia
#defaultView:BubbleChart
SELECT ?industry ?company (count(*) as ?count)
WHERE
{
    ?org wdt:P31 wd:Q4830453.
    ?org wdt:P452 ?industry.
    ?org wdt:P17 wd:Q159. #Russia country
    OPTIONAL {
        ?industry rdfs:label ?company
        filter (lang(?company) = "en")
    }
}
GROUP BY ?industry ?company
ORDER BY DESC(?count) ASC(?company)
```

SPARQL-запрос, 60 записей.

ТОП5 самых популярных отраслей в России:

Industry name	Quantity of organizations
retail	78
automotive industry	13
arms industry	10
aerospace industry	9
video game industry	9

Отсюда делаем вывод, что такая отрасль как розничная торговля в России преобладает над остальными, причем очень серьезно. Если количество организаций в этой области достигает 78, то в следующей по счету отрасли (автомобильной промышленности) работает только 13 организаций.

Для сравнения можно построить список существующих отраслей какой-нибудь другой страны (например, Норвегии).

```
#enterprise industry ranking in Norway
#defaultView:BubbleChart
SELECT ?industry ?company (count(*) as ?count)
WHERE
{
    ?org wdt:P31 wd:Q4830453.
    ?org wdt:P452 ?industry.
    ?org wdt:P17 wd:Q20. #Norway country
    OPTIONAL {
        ?industry rdfs:label ?company
        filter (lang(?company) = "en")
    }
}
GROUP BY ?industry ?company
ORDER BY DESC(?count) ASC(?company)
```

SPARQL-запрос, 41 запись.

Здесь преобладающей отраслью является manufacturing (производство)

**Количество организаций по странам**

Напишем SPARQL-запрос для получения количества коммерческих организаций в каждой стране мира.

Используются:

- Объект: business enterprise (Q4830453) (коммерческая организация)
- Свойство: country (P17) (страна)

```

SELECT ?countryLabel (count(?org) as ?count)
WHERE
{
    ?org wdt:P31 wd:Q4830453.
    ?org wdt:P17 ?country.

```

```

SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
}
GROUP BY ?country ?countryLabel
ORDER BY DESC (?count)

```

SPARQL-запрос, 198 записей.

### Организации и их дочерние организации

Необходимо построить граф из существующих организаций, а так же их дочерних организаций. Используются:

- Объект: business enterprise (Q4830453) (коммерческая организация)
- Свойство: subsidiary (P355) (дочерняя организации)

```
#subsidiary graph
```

```
#defaultView: Graph
```

```

SELECT ?org ?orgLabel ?subsidiary ?subsidiaryLabel
WHERE
{
    ?org wdt:P31 wd:Q22687
    ; rdfs:label ?item_label.

```

```

SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
OPTIONAL { ?org wdt:P355 ?subsidiary. }
FILTER (LANG(?item_label) = "en")
}

```

SPARQL-запрос, 428 записей.

Полученный граф соседей (рис. 2) состоит из висячих вершин и изолированных. Присутствие изолированных вершин, пожалуй, является недостатком полученного запроса. Необходимо построить такой граф, чтобы в нем отсутствовали эти вершины.

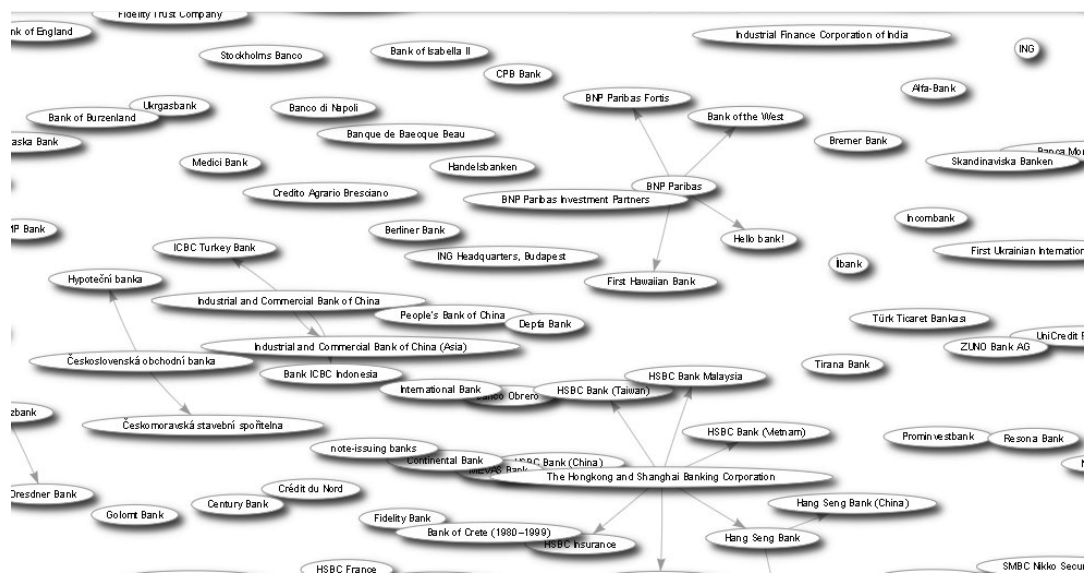


Рис. 2: Диаграмма дочерних организаций мира

```

#neighboring countries graph
#defaultView:Graph
SELECT ?org ?orgLabel ?subsidiary ?subsidiaryLabel
WHERE
{
    ?org wdt:P31 wd:Q22687
    ; rdfs:label ?item_label.
    ?org wdt:P355 ?subsidiary.

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

    FILTER (LANG(?item_label) = "en")
}

```

SPARQL-запрос, 55 записей.

### Полнота Викиданных

По данным категории Компании по алфавиту Русской Википедии существует как минимум 10 272 коммерческие организации. Их количество изменяется с каждым днем (обычно, увеличивается) ввиду появления новых организаций, которые заносятся в данный список.

По данным категории List of companies of Russia Английской Википедии в России существует как минимум 208 коммерческих организаций. Стоит отметить, что в этой категории перечислен рейтинг крупнейших компаний России по объему реализации продукции. Можно сделать вывод, что даже крупные организации не вошли в данный список, не говоря уже про мелкие и средние.

Невозможно получить релевантные данные о количестве коммерческих организаций, так как их количество растёт с каждым днём, а данные о них не хранятся в открытом доступе. Взять, к примеру, ЕГРЮЛ(Единый государственный реестр юридических лиц), который предоставляет данные за плату. [1]

"Количество коммерческих организаций, внесенных в госреестр как вновь созданных, в 2014 году составило 420,5 тыс."свидетельствуют данные на сайте Федеральной налоговой службы (ФНС) России. 30 июня 2015 года вступили в силу приказы Минфина России о том, что данные об имеющихся организациях и информация по ним больше не распространяется в открытом доступе. Данные могут быть предоставлены только органам государственной власти, иным государственным органам, органам местного самоуправления и так далее. Поэтому получить достоверные данные о количестве имеющихся организаций не представляется возможным.

Имеется возможность исследовать полноту с помощью Викиданных. Необходимо вспомнить цифру, полученную вначале, об общем количестве организаций на Викиданных (около 110 000, так как их количество постоянно растет). Обычный пользователь, имеющий общее представление об организациях, возможно, будет заинтересован в том, чтобы посмотреть как выглядит та или иная организация или же в каком месте на карте она расположена.

Чтобы посмотреть, у скольких организаций имеется изображение (то есть, заполнено поле 'image'), необходимо написать следующий скрипт.

```

#List of organizations with image

SELECT ?org ?orgLabel ?image
WHERE
{
    ?org wdt:P31 wd:Q4830453. #instance of orgs
    ?org wdt:P18 ?image #has image

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

```

SPARQL-запрос, 2913 записи.

Можно сделать вывод, что количество организаций с изображением равно 2 913. Это не так уж и много, что говорит о неполноте информации.

Построим таблицу из, возможно, популярных свойств в запросах пользователей по организациям (в зависимости от того, кто в чем будет заинтересован насчет организации). Так же, отсортируем ее по

убыванию найденных результатов.

Таблица запросов на Викиданных:

Имя свойства	Количество результатов
inception (Дата создания)	30995
founded by (Кем основана)	5722
subsidiary (Дочерние организации)	3398
image (Изображение)	2913
location (Географические координаты)	577
motto (Девиз)	2

Результаты данной таблицы говорят о том, что количество необходимой информации об организациях очень мало, учитывая их общее количество на Викиданных.

Исследуем российские организации с помощью Викиданных.

*#List of organizations*

```
SELECT ?org ?orgLabel
WHERE
{
  ?org wdt:P31 wd:Q4830453. #instance of organizations
  ?org wdt:P17 wd:Q159. #Russia country

  SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
}
```

SPARQL-запрос, 577 записей.

Запрос вывел 577 организаций. Например, пользователь захотел посмотреть как эти организации расположены на карте. Стоит написать скрипт.

*#Map of organizations*  
*#defaultView:Map*

```
SELECT ?org ?orgLabel ?location
WHERE
{
  ?org wdt:P31 wd:Q4830453. #instance of orgs
  ?org wdt:P17 wd:Q159. #Russia country
  ?org wdt:P625 ?location #display location

  SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
}
```

SPARQL-запрос, 9 записей.

В результате оказалось очень мало записей с географическими координатами в России. Получить карту организаций не только России, но и всех организаций в мире можно с помощью следующего скрипта.

*#List of organizations*  
*#defaultView:Map*

```
SELECT ?org ?orgLabel ?location
WHERE
{
  ?org wdt:P31 wd:Q4830453. #instance of orgs
  ?org wdt:P625 ?location

  SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
}
```

SPARQL-запрос, 511 записей.

Результат (рис. 3), опять-таки, очень скромный, всего лишь 511 организаций. Количество выведенных организаций с координатами даже меньше, чем общее количество всех организаций в России.



Рис. 3: Диаграмма дочерних организаций мира

Проанализировав полученные данные, можно сделать вывод, что данные об организациях на Викиданных заполнены лишь частично. Не имеется достаточной информации, чтобы делать какие-то определенные выводы насчет организаций и их составляющих. Малое количество информации можно было бы объяснить хаотичным появлением и исчезновением организаций (выживать в условиях конкуренции и существующей экономики непросто). Но информация даже о таких крупнейших организациях (Apple, Microsoft, Intel) неполна и нуждается в доработке (например, у организации Intel не указан девиз).

## Список литературы

- [1] Access to egrul. — <https://www.nalog.ru/rn77/service/egrip2>. — 2017. — Accessed 06 June 2017.
- [2] Business enterprise. research in programming wikidata [kommercheskie organizacii].