

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.

Примечание

Статья распространяется по лицензии Creative Commons Attribution-ShareAlike. Материалы этой статьи использованы в главе курса Викиверситета «Программирование Викиданных» [2]. Иллюстрации загружены на Викисклад. Научный руководитель к.т.н. Крижановский А.А.

Экземпляры объекта "Коммерческие организации"

Используются:

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

Список всех коммерческих организаций (листинг 1).

```
1 #added 2017-02
2 #List of 'instances of' "business enterprise"
3 SELECT ?lang ?langLabel
4 WHERE
5 {
6   ?lang wdt:P31 wd:Q4830453 .
7   SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
8 }
```

Листинг 1: Список всех коммерческих организаций

SPARQL-запрос, 109383 записей.

Наиболее полными и проработанными по теме коммерческие организации на Викиданных являются: Google, Apple, Microsoft.

Почти пустыми и малоинформативными организациями 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 организаций, у которых поле «label» будет непустым. (Листинг 2)

```
1 #List of 'instances of' "business enterprise" only with a label.
2 SELECT ?item ?item_label
3 WHERE
4 {
5   ?item wdt:P31 wd:Q4830453
6   ; rdfs:label ?item_label .
7
8   FILTER (LANG(?item_label) = "en") .
9 }
```

Листинг 2: Список организаций, у которых указано имя в Викиданных на английском языке

SPARQL-запрос, 74556 записей.

Распределение организаций по отраслям

Каждая организация специализируется на какой-либо отрасли. Для того чтобы понять, какая отрасль является самой популярной на данный момент вычислим число организаций, работающих в этой отрасли (листинг 3).

Результаты представлены в виде пузырьковой диаграммы (рис. 3).

Используются:

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

```
1 #enterprise industry ranking
2 #defaultView:BubbleChart
3 SELECT ?industry ?company (count(*) as ?count)
4 WHERE
5 {
6   ?org wdt:P31 wd:Q4830453.
7   ?org wdt:P452 ?industry.
8   OPTIONAL {
9     ?industry rdfs:label ?company
10    filter (lang(?company) = "en")
11  }
12 }
13 GROUP BY ?industry ?company
14 ORDER BY DESC(?count) ASC(?company)
```

Листинг 3: Пузырьковая диаграмма распределения организаций по отраслям

SPARQL-запрос, 864 записи.

Проанализировав данную диаграмму (рис. 1), можно сделать вывод о количестве организаций, специализирующихся в той или иной отрасли. На основе (рис. 1) построим таблицу, включающую 5 самых популярных отраслей.

Название отрасли	Количество организаций
Автомобильная промышленность	1149
Розничная торговля	843
Телекоммуникации	648
Видеоигры	633
Обрабатывающая промышленность	506

Таблица 1: Топ-5 самых популярных отраслей

Название отрасли	Количество организаций
Розничная торговля	78
Автомобильная промышленность	13
Оружейная промышленность	10
Аэрокосмическая промышленность	9
Видеоигры	9

Таблица 2: Топ-5 самых популярных отраслей в России

```

1 SELECT ?countryLabel (count(?org) as ?count)
2 WHERE
3 {
4     ?org wdt:P31 wd:Q4830453.
5     ?org wdt:P17 ?country.
6
7     SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
8 }
9 GROUP BY ?country ?countryLabel
10 ORDER BY DESC (?count)

```

Листинг 6: Количество организаций в каждой стране мира

SPARQL-запрос, 198 записей.

В полученном списке Россия находится на 17-ом месте(614 организаций). Лидером является США(14 658 организаций).

Организации и их дочерние организации

Построим граф из существующих организаций, а также их дочерних организаций (листинг 7).

Используются:

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

```

1 #subsidiary graph
2 #defaultView:Graph
3 SELECT ?org ?orgLabel ?subsidiary ?subsidiaryLabel
4 WHERE
5 {
6     ?org wdt:P31 wd:Q22687
7     ; rdfs:label ?item_label.
8
9     SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
10    OPTIONAL { ?org wdt:P355 ?subsidiary. }
11    FILTER (LANG(?item_label) = "en")
12 }

```

Листинг 7: Построение графа родительских и дочерних организаций

SPARQL-запрос, 428 записей(рёбер).

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 (листинг 8).

```

1 #subsidiary graph
2 #defaultView:Graph
3 SELECT ?org ?orgLabel ?subsidiary ?subsidiaryLabel
4 WHERE
5 {
6     ?org wdt:P31 wd:Q22687
7     ; rdfs:label ?item_label.
8     ?org wdt:P355 ?subsidiary.
9
10    SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }

```

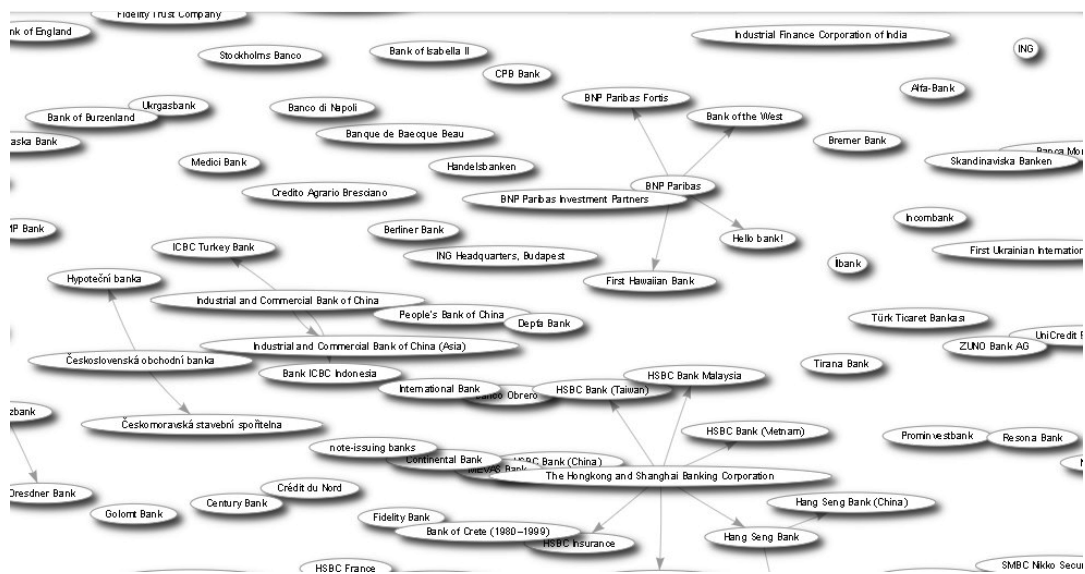


Рис. 3: Диаграмма дочерних организаций мира

```

11 FILTER (LANG(?item_label) = "en")
12 }
13 }

```

Листинг 8: Построение графа родительских и дочерних организаций без висячих вершин

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. В этой категории перечислен рейтинг крупнейших компаний России по объему реализации продукции. Можно сделать вывод, что даже крупные организации (Яндекс, Тинькофф) 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, Единый государственный реестр юридических лиц (ЕГРЮЛ), 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 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.

Оценим полноту информации о коммерческих организациях на Викиданных. Необходимо вспомнить цифру, полученную вначале, об общем количестве организаций на Викиданных (около 110 000). Найдем иллюстрации, соответствующие организациям с помощью запроса из листинга 9.

```

1 #List of organizations with image
2
3 SELECT ?org ?orgLabel ?image
4 WHERE
5 {
6   ?org wdt:P31 wd:Q4830453 . #instance of orgs

```

```

7 ?org wdt:P18 ?image #has image
8
9 SERVICE wikibase:label { bd:serviceParam wikibase:language "en"}
10 }

```

Листинг 9: Организации с изображением

SPARQL-запрос, 2913 записей.

Количество организаций с изображением равно 2 913, что составляет 3

Оценим количество заполненных свойств у коммерческих организаций на Викиданных. Результат в табл. 3. Например inception(Дата создания) указано у 31 организаций, что составляет 31

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

Таблица 3: Свойства объекта «Коммерческая организация» на Викиданных и их заполненность.

Результаты данной таблицы (табл. 3) говорят о том, что количество необходимой информации об организациях очень мало, учитывая их общее количество на Викиданных.

Исследуем полноту представления российских организаций в Викиданных (листинг 10).

```

1 #List of organizations
2
3 SELECT ?org ?orgLabel
4 WHERE
5 {
6   ?org wdt:P31 wd:Q4830453. #instance of organizations
7   ?org wdt:P17 wd:Q159. #Russia country
8
9   SERVICE wikibase:label { bd:serviceParam wikibase:language "en"}
10 }

```

Листинг 10: Организации России

SPARQL-запрос, 577 записей.

В Викиданных представлено 577 отечественных организаций. Отобразим на карте те российские организации, у которых указан географические координаты (листинг 11).

```

1 #Map of organizations
2 #defaultView:Map
3
4 SELECT ?org ?orgLabel ?location
5 WHERE
6 {
7   ?org wdt:P31 wd:Q4830453. #instance of orgs
8   ?org wdt:P17 wd:Q159. #Russia country
9   ?org wdt:P625 ?location #display location
10
11   SERVICE wikibase:label { bd:serviceParam wikibase:language "en"}
12 }

```

Листинг 11: Карта организаций России

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 (листинг 12).

```

1 #List of organizations
2 #defaultView:Map
3
4 SELECT ?org ?orgLabel ?location

```



```

5 WHERE
6 {
7   ?org wdt:P31 wd:Q4830453. #instance of orgs
8   ?org wdt:P625 ?location
9
10  SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
11 }

```

Листинг 12: Карта организаций мира

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. 4: Карта 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. Наличие авторитетных источников, неафилированных и общедоступных крайне мало, а это главный козырь для создания статьи об организации в Википедии и Викиданных.. Но информация даже о таких крупнейших организациях (Apple, Microsoft, Intel) не полна и нуждается в доработке (например, у организации Intel не указан девиз).

Список литературы

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