

Тематический и жанровый анализ газет мира и России

Andrew Krizhanovsky

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

Anastasia Bodryakova

Petrozavodsk State University

18 января 2018 г.

Аннотация

Газета – это печатное периодическое издание. Статья посвящена исследованию объекта Викиданных "газета". С помощью SPARQL-запросов, вычисляемых на объектах типа "газета" Викиданных, решены следующие задачи: выведен список всех газет мира, получен перечень самых популярных жанров газет в виде пузырьковой диаграммы, определено, что в мире есть 106 газет с географической привязкой в Викиданных (свойств "coordinate location") и больше всего газет, обладающие данным свойством имеют города в Европе и Америке. Так же был получен результат о самых многочисленных жанрах в газетах, а именно: сатира, информация, аналитика, научный журнал и всеведение.

1 Экземпляры объекта "Газеты"

- Объект: <https://www.wikidata.org/wiki/Q11032>газета (Q11032)
- Свойство: <https://www.wikidata.org/wiki/Property:P31>экземпляр (P31)

Получим список газет во всем мире с помощью следующего скрипта: См. листинг 6.

```
1 SELECT ?newspaper ?newspaperLabel
2 WHERE
3 {
4   ?newspaper wdt:P31 wd:Q11032 .
```

```

5 SERVICE wikibase:label { bd:serviceParam wikibase:language "en".
6   }
7 }

```

Листинг 1: Список газет во всем мире

<https://query.wikidata.org/>

Список газет, у которых есть метка (label) на русском и английском языках: См. листинг 6.

```

1 SELECT ?newspaper ?label_en ?label_ru WHERE {
2   ?newspaper wdt:P31 wd:Q11032 .
3   ?newspaper rdfs:label ?label_en .
4   ?newspaper rdfs:label ?label_ru .
5   FILTER((LANG(?label_en)) = "en")
6   FILTER((LANG(?label_ru)) = "ru")
7 }

```

Листинг 2: Список газет с меткой label

<https://query.wikidata.org/SELECT>

Во втором скрипте получилось меньше записей, чем у первого, так как не все газеты имеют метку (label) на русском и английском языках одновременно.

Наиболее полные и проработанные газеты на Викиданных:

- <https://www.wikidata.org/wiki/Q264003>Зери и популлит,
- <https://www.wikidata.org/wiki/Q336197>Наблюдатель,
- <https://www.wikidata.org/wiki/Q753932>Известия,
- <https://www.wikidata.org/wiki/Q50008>Таймс.

Малоинформативными газетами на Викиданных были:

- <https://www.wikidata.org/wiki/Q4242920>Крымская правда,
- <https://www.wikidata.org/wiki/Q4400770>Русский инвалид,
- <https://www.wikidata.org/wiki/Q985193>Кодима,
- <https://www.wikidata.org/wiki/Q28647608>Армия и флот рабочей и крестьянской России,
- <https://www.wikidata.org/wiki/Q16933408>Меркуриус Политикус,
- <https://www.wikidata.org/wiki/Q18342176>Павлово-Посадские известия,
- <https://www.wikidata.org/wiki/Q20036766>Дик Мануш,
- <https://www.wikidata.org/wiki/Q21210293>Вестник Маньчжурии,
- <https://www.wikidata.org/wiki/Q21099945>Бухорои Шариф,
- <https://www.wikidata.org/wiki/Q1983882>Кыым.

2 Полнота Викиданных

Проанализируем полноту Викиданных.

- По данным учебного пособия Радченко М.В. http://www.unizd.hr/portals/35/2011-12/jezik_medi_ja_p_rirucnik.doc, к 2009 году в Российской Федерации было зарегистрировано более 50 тысяч печатных СМИ, в том числе 27425 газет и еженедельников и 20433 журнала.
- По данным категории https://en.wikipedia.org/wiki/List_of_newspapers_in_Russia Английской Википедии есть 16 ежедневно выпускающихся газет в России, а так же 9 газет, которые издаются с периодичностью от одного до четырех выпусков в неделю, что составляет 0,09% от зарегистрированных газет, согласно учебному пособию Радченко М.В.
- По данным категории https://en.wikipedia.org/wiki/List_of_national_newspapers Английской Википедии газета рассматривается, как национальная газета, то есть она должна распространяться по всей стране, в отличие от местной газеты, которая выходит в определенном городе или регионе. В данной категории определены 87 национальных газет (0,3%), включающие в себя столичные газеты.
- По данным категории <https://ru.wikipedia.org/wiki/Категория:Газеты> Русской Википедии насчитывается 115 газет (0,4%), которые печатаются в России. Многие газеты имеют не только печатное издание, но и <https://ru.wikipedia.org/wiki/Сайтвеб-сайт>, например <https://www.rbth.com> сайт газеты Russia Beyond the Headlines.

От общего числа зарегистрированных газет (27425) всего 0,8% газет являются заполненными, по данным всех вышеперечисленных категорий. Это говорит о низкой заполненности Викиданных.

2.1 Жанр газет

Газетные материалы должны обладать определенной направленностью – тщательным учетом всех специфических черт, свойственных аудитории той или иной страны или группы стран, для которых предназначается публикация.

Существует три основных жанра в газетах:

1. информационный
2. аналитический
3. художественно-публицистический

К информационному жанру относятся: заметки, репортажи, интервью, отчеты. Непосредственно этот жанр передает аудитории все минувшие анонсы.

Аналитические жанры – корреспонденция, комментарий, статья, рецензия, обзор печати, письмо, обозрение – имеют более широкие временные границы, в них содержится изучение и анализ системы фактов, ситуаций, обобщения и выводы.

Художественно-публицистические жанры – очерк, фельетон, памфлет – имеют большую эмоциональную силу, содержат образно-выразительные средства.

Построим пузырьковую диаграмму распределения газет по жанрам.

```
1 #basic form of government ranking
2 #defaultView:BubbleChart
3 SELECT ?genre ?form (count(*) as ?count)
4 WHERE
5 {
6     ?newspaper wdt:P31 wd:Q11032. # instance of newspaper
7     ?newspaper wdt:P136 ?genre . # genre of newspaper
8     OPTIONAL {
9     ?genre rdfs:label ?form
10    filter (lang(?form) = "ru")
11    }
12 }
13 GROUP BY ?genre ?form
14 ORDER BY DESC(?count) ASC(?form)
```

Листинг 3: Газеты со свойством "main subject"(основная тема) во всём мире

<https://query.wikidata.org/>

Самыми многочисленными оказались жанры: сатира, информация, аналитика, научный журнал, всеведение.

2.2 Газеты на карте

Выведем газеты, которые обладают свойством "coordinate location" на карте мира.

```
1 #Newspapers on the world map
2 #defaultView:Map
3 SELECT ?newspaper ?location ?newspaperLabel WHERE {
4     ?newspaper wdt:P31 wd:Q11032. # instance of newspaper
5     ?newspaper wdt:P625 ?location. # location of newspaper
6 }
```

Листинг 4: Газеты на карте мира со свойством "coordinate location"

<https://query.wikidata.org/>

С помощью данного скрипта можно заметить, что газеты со свойством "coordinate location" имеют географические координаты, в большинстве случаев, соответствующие городам в Европе и Америке.

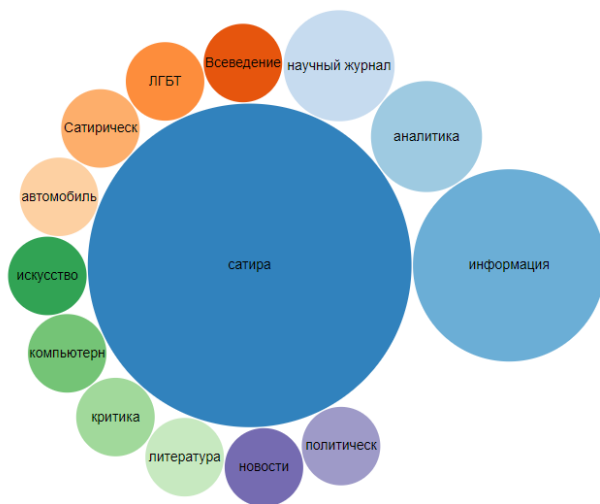


Рис. 1: Пузырьковая диаграмма распределения газет по жанрам (свойство “genre”)



Рис. 2: Газеты со свойством “coordinate location” (географические координаты) на карте мира

3 Заполнение Викиданных

Свойство "genre" (жанр) подразумевает способ и форму передачи информации в газетах. Например, возьмём газету "Труд". В ней свойство "genre" (жанр) имеет значение "information" и это означает, что эта газета относится к информационному жанру.

Построим список газет без заполненного свойства <https://www.wikidata.org/wiki/Q483394genre> (Q483394) и [https://www.wikidata.org/wiki/Property:P921main subject \(P921\)](https://www.wikidata.org/wiki/Property:P921main subject (P921)),

чтобы узнать к каким газетам следует добавить "genre" и "main subject".

```
1 SELECT ?newspaper ?label_ru
2 WHERE
3 {
4   ?newspaper wdt:P31 wd:Q11032.           # instance of newspaper
5
6   { ?newspaper wdt:P17 wd:Q34266 } UNION # Russian Empire
7   { ?newspaper wdt:P17 wd:Q15180 } UNION # Soviet Union
8   { ?newspaper wdt:P17 wd:Q159 }        # Russia
9   MINUS { ?newspaper wdt:P136 [] }.     # newspaper without a genre
10  MINUS { ?newspaper wdt:P921 [] }.     # newspaper without a main
11 }                                     subject
```

Листинг 5: Газеты без заполненного свойства "genre" и "main subject"

<https://query.wikidata.org/>

В ходе работы были заполнены свойства genre и main subject в 100 объектах газеты.

С помощью скрипта удалось получить список из 26 газет, не имеющие свойств genre и main subject. Данные свойства были заполнены. Так же были рассмотрены 74 объекта в категории <https://ru.wikipedia.org/wiki/Категория:Газеты.genreremainssubject>.

Итого, заполнены свойства genre и main subject в 100 объектах (газеты).

Так как тем может быть много, а жанр у газеты один, чтобы продемонстрировать заполнение Викиданных, строка нахождения свойства main subject закомментирована (Рис. 3). С помощью следующего скрипта построим список газет, которые имеют свойство <https://www.wikidata.org/wiki/Q483394genre> (Q483394).

```
1 SELECT ?newspaper ?label_ru
2 WHERE
3 {
4   ?newspaper wdt:P31 wd:Q11032.           # instance of newspaper
5
6   { ?newspaper wdt:P17 wd:Q34266 } UNION # Russian Empire
7   { ?newspaper wdt:P17 wd:Q15180 } UNION # Soviet Union
8   { ?newspaper wdt:P17 wd:Q159 }        # Russia
9   { ?newspaper wdt:P136 [] }.           # genre of newspaper
10  # { ?newspaper wdt:P921 [] }.         # main subject of newspaper
11 }
```

Листинг 6: Газеты со свойством genre

<https://query.wikidata.org/>

Построим пузырьковую диаграмму по свойству "основная тема" по газетам всего мира на Викиданных:

```
1 #defaultView:BubbleChart
2 SELECT ?subject ?form (COUNT(*) AS ?count) WHERE {
3   ?newspaper wdt:P31 wd:Q11032.           # instance of newspaper
4   ?newspaper wdt:P921 ?subject.         # main subject of newspaper
5   OPTIONAL {
6     ?subject rdfs:label ?form.
7     FILTER((LANG(?form)) = "ru")
8   }
}
```

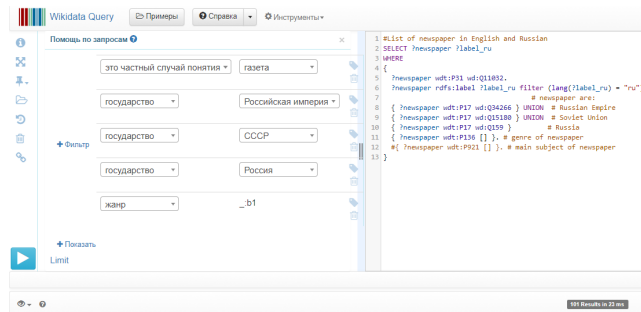


Рис. 3: Газеты с жанром с Российской Империи, Советском Союзе и в России

```

9 }
10 GROUP BY ?subject ?form
11 ORDER BY DESC(?count) ?form

```

Листинг 7: Газеты со свойством "main subject" (основная тема) во всём мире

<https://query.wikidata.org/>

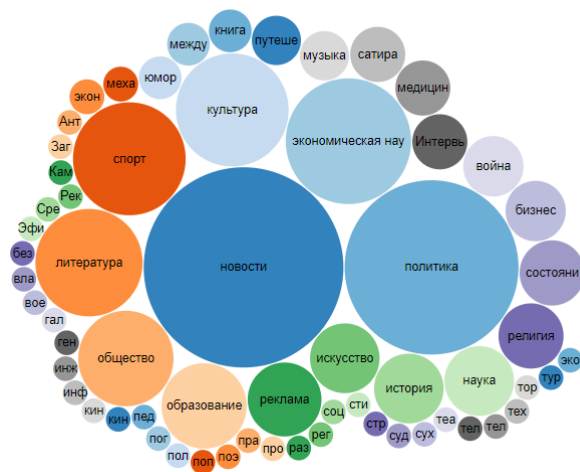


Рис. 4: Пузырьковая диаграмма распределения газет со свойством “main subject” (основная тема) во всём мире

Данный скрипт показал, что наиболее популярными темами в газетах являются:

- новости (66 газет),

- политика (50 газет),
- экономическая наука (26 газет),
- культура (21 газета),
- спорт (21 газета).

Построим пузырьковую диаграмму по свойству "жанр" по газетам всего мира на Викиданных.

<https://query.wikidata.org/>

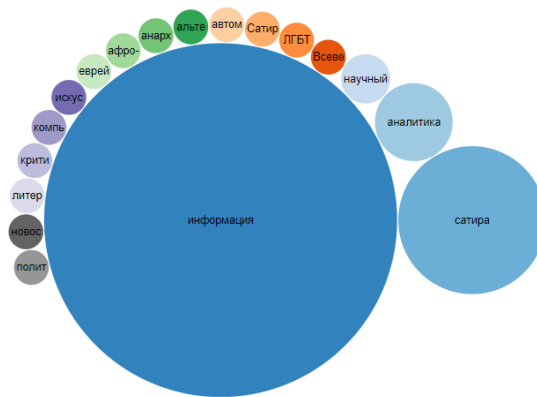


Рис. 5: Пузырьковая диаграмма распределения газет со свойством “genre” (жанр) во всем мире

Основными газетными жанрами являются:

- информация (103),
- сатира (18),
- аналитика (5).

Жанров намного меньше, чем основных тем, так как газета может обладать только одним жанром, а тем может быть много.