

WD: Anime

Andrew Krizhanovsky^a, Daria Boollieva^b

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

^b*Affiliation not available*

Abstract

Аннотация

Статья посвящена исследованию объекта Викиданных "аниме". С помощью SPARQL-запросов, вычисляемых на объектах типа "аниме" в Викиданных, решены такие задачи: выведен упорядоченный список сэйю по числу озвученных ими аниме, построена гистограмма по числу сэйю, озвучивших одно и более аниме, построен граф, связывающий сэйю и озвученные ими аниме.

Примечание

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

Экземпляры объекта "Аниме"

Аниме - японская анимация. У каждого аниме есть актеры озвучивания. В дальнейшем мы будем употреблять слово "сэйю"(т.е. японские актеры озвучивания). Актеры озвучивания и сэйю являются синонимами, когда речь заходит о Японии и японской анимации. Слово "тайтл"(название) будет упоминаться при обращении к какому-то аниме (или японской анимации в целом).

- Объект: аниме (Q1107),

Список всех аниме. См. листинг 1.

```
1 #added 2017-06
2 #List of 'instances of' "anime"
3 SELECT ?anime ?animeLabel
4 WHERE
5 {
6   ?anime wdt:P31 wd:Q1107.
7   SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
8 }
```

Листинг 1: Список аниме

SPARQL-запрос, 683 записи.

Наиболее полными и проработанными аниме на Викиданных являются: Гуррен-Лаганн, Space Battleship Yamato, Project A-ko.

Почти пустыми и малоинформативными аниме оказались: Charlotte, Dagashi Kashi, KonoSuba.

Упорядоченный список сэйю по числу озвученных ими аниме

Практически в любом аниме присутствуют несколько актеров озвучивания (сэйю). Большинство сэйю озвучили за свою карьеру несколько тайтлов, а многие даже несколько десятков тайтлов. Талантливых сэйю приглашают озвучивать сразу несколько персонажей в одном аниме.

Построим упорядоченный список сэйю по числу озвученных ими аниме. См. листинг 2.

```
1 #Ordered list of actors (seiyu) according to the quantity of their anime
2 SELECT ?seiyu (SAMPLE(?label) AS ?seiyuLabel) (COUNT(?anime) AS ?count)
3 WHERE
4 {
5   ?anime wdt:P31 wd:Q1107;    # Instance of anime
6           wdt:P725 ?seiyu.    # Instance of seiyu (voice actor)
7   ?seiyu rdfs:label ?label.   # Subclass of label
8   FILTER(LANG(?label) = "en").
9 }
10 GROUP BY ?seiyu    # Group by seiyu
11 ORDER BY DESC(?count) # Order by count of voiced anime
```

Листинг 2: Упорядоченный список сэйю

SPARQL-запрос, 148 записей.

Мы видим список, состоящий из 148 сэйю. Список получится немногочисленным, но и количество аниме на Викиданных было столь же немногочисленным - всего 683 тайтла (речь о полноте будет дальше). В данный список вошли сэйю, озвучившие несколько аниме (например, Aki Toyosaki - 26 аниме).

Гистограмма по числу сэйю, озвучивших одно и более аниме

Было бы интересно построить гистограмму (линейную диаграмму) из сэйю, озвучивших аниме (чем больше аниме озвучил сэйю, тем дальше на диаграмме он будет находиться, "правее" в данном случае). См. листинг 3.

```
1 #added 2017-06
2 #Histogram of the number of seiyu who voiced one or more anime
3 #defaultView:LineChart    # Do line chart as result representation
4 SELECT ?haveseiyu (COUNT(?haveseiyu) AS ?quantity) WHERE { # Count number of
5   seiyu having a voice acting
6   {
7     SELECT (COUNT(?seiyu) AS ?haveseiyu) WHERE { # Count quantity of
8       voice acting
9       ?anime wdt:P31 wd:Q1107;
10              wdt:P725 ?seiyu.
11       SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
12     }
13   }
14 }
```

```

11 GROUP BY ?anime # Group list by quantity of voiced anime
12 ORDER BY DESC(?haveseiyu) # Order by seiyu quantity (descending)
13 }
14 }
15 GROUP BY ?haveseiyu # Group by seiyu quantity
16 ORDER BY DESC(?haveseiyu) # Order by seiyu qty (descending)

```

Листинг 3: Гистограмма по числу сэйю

SPARQL-запрос, 13 записей.

Очевидно, что, чем большее количество аниме берётся в расчёт, тем меньшее количество сэйю участвует в озвучке (рис. 1). Это демонстрирует данная гистограмма. Большинство сэйю, как показано на диаграмме, озвучили только 1 аниме. Это может быть связано с неполнотой Викиданных.

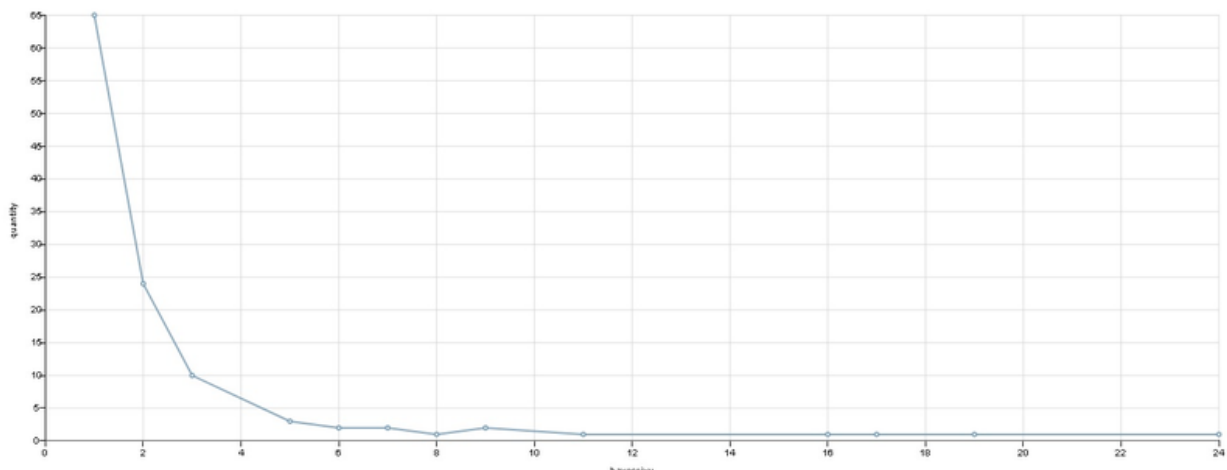


Рис. 1: Гистограмма по числу сэйю, озвучивших одно и более аниме

Граф, связывающий сэйю и озвученные ими аниме

Как было сказано ранее, несколько сэйю могут озвучивать сразу несколько персонажей в одном аниме (встречается редко), а так же озвучивать несколько аниме. Построим граф, связывающий сэйю и озвученные ими аниме, чтобы нагляднее показать это взаимосвязь. См. листинг 4.

```

1 #added 2017-06
2 #Graph contacting seiyu and anime
3 #defaultView:Graph
4 SELECT ?anime ?animeLabel ?seiyu ?seiyuLabel
5 WHERE
6 {
7   ?anime wdt:P31 wd:Q1107
8   ; rdfs:label ?label .
9
10  SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
11  OPTIONAL { ?anime

```

```

12   wdt:P725 ?seiyu . }
13   FILTER (LANG(?label) = "en")
14 }

```

Листинг 4: Граф сэйю и аниме

SPARQL-запрос, 826 записей.

Данный граф (рис. 2) показывает все существующие аниме, а так же сэйю, которые эти аниме озвучивали. Недостаток графа в том, что большое количество аниме остались "без озвучки".



Рис. 2: Граф, связывающий сэйю и озвученные ими аниме

Полнота Викиданных

Список найденных аниме на сайте shikimori. В списке 559 по 20 аниме. Всего 11173 наименований. В то время как количество найденных объектов на Викиданных всего 683. К тому же, стоит учитывать, что скорость выхода новых аниме довольно велика. Даже зимой-летом 2017-го выходит более 600 аниме. После выполнения некоторых запросов, упомянутых ранее, некоторые аниме не находились (Useless Animals, Ryuu no Naisha, Frame Arms Girl). Из этого можно сделать вывод, что Викиданные крайне неполно отражают данные.

Согласно приведённой статистике от того же сайта shikimori industry в разделе "Индустрия аниме" упоминается 16 разделов по японской анимации, в которые входят 10497 тайтлов. Возможно, приведённые ниже статьи и сайты не будут являться авторитетными источниками, но с помощью них можно привести информацию об имеющихся аниме и сделать определённые выводы о неполноте.

На сайте anidub приведён список из 7811 аниме.

На сайте animespirit приведён список из 4905 аниме.

На сайте animelend приведён список из 4751 аниме.

На сайте anistar приведён список из 2971 аниме.

На сайте animevost приведён список из 1881 аниме.

На сайте anidesu приведён список из 763 аниме.

Можно сделать вывод, что различные сайты имеют разную информацию об имеющихся аниме. Какие-то сайты появились позже, какие-то раньше, поэтому количество

аниме может разниться, причём довольно серьёзно. Так же, количество посетителей влияет на появление каких-то аниме на сайте. Если упорядочить все приведённые сайты, данные Русской Википедии, Английской Википедии по количеству аниме, то на последнем месте окажутся Викиданные. На Викиданных нельзя найти все популярные и известные японские анимации мира, что ещё раз говорит о неполноте.

Вспомним ранее упомянутый запрос, в котором говорилось о 148 сэйю на Викиданных. Дело в том, что они искались в подразделе аниме, поэтому результат оказался таким немногочисленным. Если поправить скрипт и попытаться вывести всех актёров озвучивания (то есть, убрать категорию аниме), то результат может измениться. См. листинг 5.

```
1 #added 2017-06
2 #Ordered list of actors according to the quantity of their voiced projects
3 SELECT ?actor (SAMPLE(?label) AS ?actorLabel) (COUNT(?anime) AS ?count)
4 WHERE
5 {
6   ?anime wdt:P725 ?actor.      # Instance of voice actor
7   ?actor rdfs:label ?label.    # Subclass of label
8   FILTER(LANG(?label) = "en").
9 }
10 GROUP BY ?actor # Group by actor
11 ORDER BY DESC(?count) # Order by count of voiced anime
```

Листинг 5: Список актёров озвучивания

SPARQL-запрос, 3965 записей.

Нашлось 3965 актёров озвучивания. Сравним результат по сэйю. Прошлый скрипт вывел у сэйю Аки Toyosaki 26 аниме. Данный скрипт дал результат в 62 аниме. Это говорит о том, что данные заполнялись не совсем корректным образом и о том, что нужно было упоминать категорию аниме.

Заполнение 100 объектов

Любитель японской анимации желает знать в каком году вышло его любимое аниме. Викиданные располагают этой информацией не в полной мере. Напишем скрипт, который бы показывал количество аниме с незаполненным полем "publication date" (дата публикации). См. листинг 6.

```
1 #added 2017-06
2 #List of anime with unfilled publication date
3 SELECT ?anime ?animeLabel
4 WHERE
5 {
6   ?anime wdt:P31 wd:Q1107.
7   FILTER NOT EXISTS { ?anime wdt:P577 [] } #if property publication date is
  unfilled
8   SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
9 }
```

Листинг 6: Аниме с незаполненной датой публикации

SPARQL-запрос, 237 записей.

Видим результат в 237 записи. Это говорит о том, что примерно у трети аниме нет даты публикации.

После заполнения данных и повторного выполнения скрипта выдаётся результат в 134 записи.

[WDAanime] Anime. Research in programming Wikidata [Programmirovaniye Vikidannykh] // Russian Wikiversity. https://ru.wikiversity.org/w/index.php?title=%D0%9F%D1%80%D0%BE%D0%B3%D1%80%D0%B0%D0%BC%D0%BC%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5_%D0%92%D0%B8%D0%BA%D0%B8%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85/%D0%90%D0%BD%D0%B8%D0%BC%D0%B5, 2017. Accessed 14 Mar 2017.