

The search of music gaps in public domain using Wikidata

Andrew Krizhanovsky and Ilya Vershinin

This article is about research of musical compositions using the knowledge base of international project called Wikidata. With the help of SPARQL queries for items classified as "musical compositions" the following were received: the list of all musical compositions, the list of musical compositions that has a composer, the bubble diagram for composers that shows composers with most compositions. Moreover, the task of searching music gaps in public domain was done and completeness of Wikidata was evaluated.

1 List of musical compositions

- Item: musical composition (Q207628).
- Property: instance of (P31).

Let's build a list of all musical compositions. See the listing 1.

```
1 #List of all musical compositions
2 SELECT ?composition ?compositionLabel
3 WHERE {
4   ?composition wdt:P31 wd:Q207628.
5   SERVICE wikibase:label { bd:serviceParam wikibase:language "ru". }
6 }
```

Listing 1: List of all musical compositions

[SPARQL query](#), 5494 records.

The most complete and well-developed musical compositions on Wikidata are The Magic Flute, Für Elise, Mozart's Requiem, Eine kleine Nachtmusik.

Almost empty and uninformative musical compositions were Flight of the Bumblebee, Romeo and Juliet, Iron Foundry, Binks' Waltz, The Rose-bud March, Leola.

2 Search music gaps in public domain

The task is to find musical works the composers of which passed away more than 70 years ago and audios of which are absent from Wikimedia Commons. The list of compositions must be sorted the in ascending order by publication date. This script can be used to find musical compositions that need to be digitized and then uploaded to Wikimedia Commons. See the listing 2.

```
1 #Search music gaps in public domain
2 SELECT ?composition ?compositionLabel ?publication
3 WHERE {
4   ?composition wdt:P31 wd:Q207628.      # instance of composition
5   ?composition wdt:P86 ?composer.      # composition has a composer
6   ?composition wdt:P577 ?publication.  # composition has a publication date
```

```

7  ?composer wdt:P570 ?death.          # composer has a date of death
8  MINUS {?composition wdt:P51 []}.    # compositions without audio
9  FILTER(?death < "1947-01-01T00:00:00Z"^^xsd:dateTime) # composers that passed away more
    than 70 years ago
10 FILTER(?publication < "1947-01-01T00:00:00Z"^^xsd:dateTime) # compositions that were
    published more than 70 years ago
11 SERVICE wikibase:label { bd:serviceParam wikibase:language "ru". }
12 }
13 ORDER BY ASC(?publication)

```

Listing 2: Search music gaps in public domain

[SPARQL query](#), 140 records.

3 Completeness of Wikidata

Let's analyze the completeness of Wikidata .

According to *Grove Dictionary of Music and Musicians* (2001) there are 20374 composers.

According to the category "[List of composers by name](#)" of Russian Wikipedia there are around 6000 composers.

According to the category "[List of composers by name](#)" of English Wikipedia there are around 5000 composers.

The number of musical compositions with filled property "composer (P86)" equals 3862, which is shown in [SPARQL query](#), and that's if you take into account the fact that one composer could have written several musical compositions. For example Wolfgang Amadeus Mozart is the composer of 95 compositions, which decreases the number of unique composers. The number of 3862 is lower than the amount of composers from both Russian and English Wikipedia and substantially lower than the amount of composers from Grove Dictionary of Music and Musicians which confirms the incompleteness of Wikidata.

[SPARQL query](#) for compositions with filled property "composer (P86)" and property "country of origin (P495)" with value of "Russian Empire (Q34266)", "USSR (Q15180)" or "Russia (Q159)", gave us only 8 compositions, which means that it's impossible to analyze Russian musical compositions due to lack of data.

Let's build the bubble diagram for composers of musical compositions. See the listing 3.

```

1 #composers of musical compositions
2 #defaultView:BubbleChart
3 SELECT ?composer ?form (COUNT(*) AS ?count) WHERE {
4   ?composition wdt:P31 wd:Q207628. # instance of composition
5   ?composition wdt:P86 ?composer. # composition has a composer
6   OPTIONAL {
7     ?composer rdfs:label ?form.
8     FILTER((LANG(?form)) = "en")
9   }
10 }
11 GROUP BY ?composer ?form
12 ORDER BY DESC(?count) ?form

```

Listing 3: Musical compositions grouped by composer

Size of a bubble tells us about the amount of musical compositions. This diagram shows us that some composers have more compositions than the others. Top 5 includes Niels Gade (173 compositions), Johann Sebastian Bach (155 compositions), Christian Sinding (125 compositions), Johan Halvorsen (121 compositions), Alan Hovhaness (108 compositions).

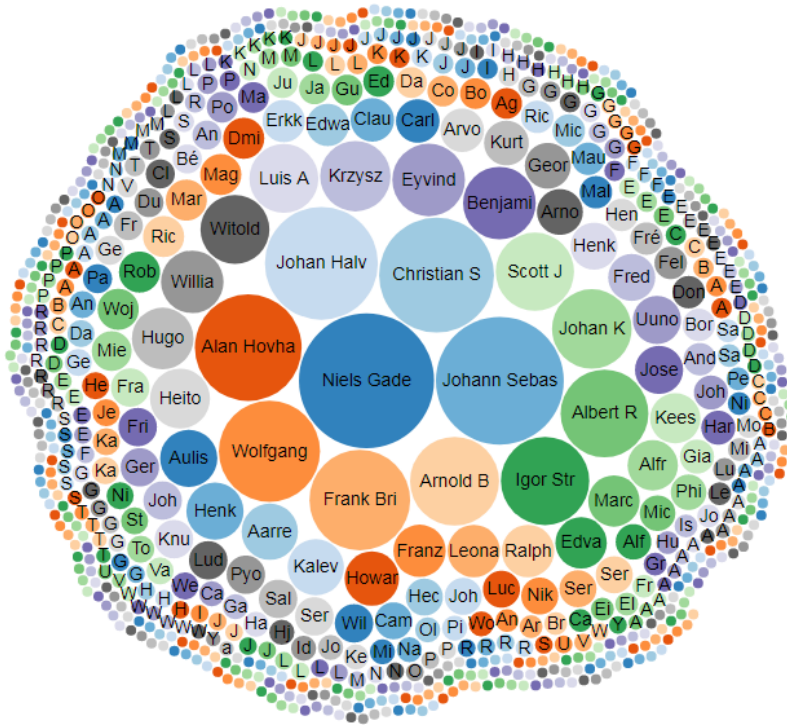


Figure 1: Musical compositions grouped by composer

4 Filling of Wikidata

The decision was made to fill "composer (P86)" property for "musical composition (Q207628)" items to get better results while performing the query for searching music gaps in public domain.

Let's build a list of all musical compositions with filled property "composer (P86)". See the listing 4.

```

1 #Lists of compositions that has a composer in Russian
2 SELECT ?composition ?compositionLabel ?composer ?composerLabel
3 WHERE {
4   ?composition wdt:P31 wd:Q207628. # instance of composition
5   ?composition wdt:P86 ?composer. # composition has a composer
6   SERVICE wikibase:label { bd:serviceParam wikibase:language "ru". }
7 }
8 }

```

Listing 4: Lists of compositions that has a composer in Russian

[SPARQL query](#), 3864 records at 30/10/2017, 10:51.

[SPARQL query](#), 3965 records at 30/10/2017, 12:47.

References

Grove Dictionary of Music and Musicians. (2001). Oxford University Press.

Musical composition. Research in programming Wikidata // English Wikiversity. (2017). https://en.wikiversity.org/wiki/Research_in_programming_Wikidata/Musical_Compositions.