Anime List - Semantic Web

Andreia Rodrigues
up201404691@fe.up.pt
Faculty of Engineering
of University of Porto

Francisco Queirós
up201404326@fe.up.pt
Faculty of Engineering
of University of Porto
Porto

Miriam Gonçalves up201403441@fe.up.pt Faculty of Engineering of University of Porto

9th of January of 2019

Abstract

Anime refers to "animation" and it represents all animation produced by Japan. It has a distinct look-and-feel compared to western animations and, over the last forty years, it has become an international phenomenon, attracting millions of fans and being translated into many languages. With the increasing number of shows being released every year, large collections of documents need to be indexed and ranked so that information retrieval tools can deliver relevant documents according to the user's information need. As an alternative to this method, semantic web ontologies can be built, representing knowledge by favouring the interoperability of knowledge obtained from different sources. The aim of this paper is to focus on the creation of an ontology for animes and identify appropriate queries to explore the created ontology.

1 Introduction

Anime is an abbreviation of the word "animation" and it is used by the Western culture to describe a Japanese-style animated film or tv show [1]. These animations are characterized by the distinct look of its characters, with huge eyes, bright colored hair and exaggerated emotional expressions and gestures [2].

Japan began producing animation in 1917, but *animes* only started to become famous after the 60s due to the creation of television, that had a crucial role in making these Japanese characteristic films an increasing trend [3] [4].

Nowadays, with the growth of *anime* popularity and the number of animated films being produced every year, *anime* fans gather in online platforms where information about *anime* and

its reviews are collected and can be accessed, allowing users to interact with each other, share what animations they are interested in and keep track of what they have watched, are watching and want to watch in the future.

These online platforms gather their own data separately and there isn't a centralized place where users can have an overview of an *anime* and the rating it is given in each of them. To counter this, we previously assembled and refined the chosen information in a big sample of data stored in a single *CSV* file.

We begin this paper with the analysis of the data domain and its main concepts, along with the existing vocabularies or ontologies of the domain. This is followed by the description of the tool used to build the ontology and the queries created to explore the ontology.

2 Ontology

On the current iteration of this project, the goal is to analyze, design and explore an ontology for the data domain of *animes*, based on the previously acquired and processed data from the previous iterations of the project.

2.1 Analysis of the data domain and its main concepts

In the domain to explore, an *anime* is an animation originated from Japan. It can be distributed as a TV show, as a movie or as a less common type like an *OVA* (Original Video Animation) or *ONA* (Original Net Animation), usually special episodes delivered for television or for the web. Since most *animes* are distributed for television they can easily be compared to

a TV show as both types have producers, licensors, a defined genre(s), a studio where it was developed, the dates from when the show started and finished airing, rating, type, synopsis, episode amount, duration of episodes, and status.

With the current growth of *anime* popularity, *anime* fans gather in online platforms to evaluate these animations so we also included the rating and some other information of two of the most known platforms as a characteristic of them.

The concepts mentioned are used to describe an *anime* and to relate it to other entities. For example, producers, licensors, studios are entities that can also be defined in our ontology. Genres, ratings, the current status of an *anime*, it's source and type can also be used to relate *anime*'s as we may want to know all the *animes* with the "Adventure" genre, for example.

These values are predefined in order to proceed to a consistent classification of *anime* characteristics.

To relate an *anime* with other *animes*:

- hasAlternativeSetting relates an anime to other anime which tells an alternative setting of the story told by the first one;
- hasFullStory relates an *anime* to other *anime* which tells the full story of the story told by the first one;
- hasParentStory relates an anime to other anime which tells the origin story of the story told by the first one;
- hasPrequel relates an anime to its prequel;
- hasSequel relates an anime to its sequel;
- hasSideStory relates an anime to other anime which tells a side story of the story told by the first one;
- hasSpinOff relates an anime to its spinoff;
- hasSummary relates an *anime* to other *anime* which is a resumed version of the first one.

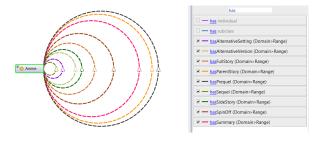


Figure 1: Properties between entities of the anime class.

To relate an anime with other entities:

- hasGenre relates an anime to one of its genres (usually an anime has more than one genre);
- hasLicensor relates an anime to its licensor (usually an anime only has one licensor);
- hasProducer -relates an anime to one of its producers (usually an anime has more than one producer);
- hasRating relates an anime to the age rating it is classified;
- hasSource relates an anime to its main source (the story told can be originated from a book, for example);
- hasStatus relates an *anime* to its current status (airing or finished);
- hasStudio relates an *anime* to the studio where it is developed;
- hasType relates an *anime* to its type (tv, movie, etc).

To classify an *anime*, which is the main class in this domain, some subclasses were created, regarding the current status, type or score attributed to that *anime*. In regard to the current status, the subclasses "AiringAnime" and "FinishedAnime" made sense to create so that we can identify easily what *animes* were still being released. The classes "TV_Anime", "Movie_Anime", "Music_Anime", "ONA_Anime", are subclasses that separate the different existing types of *anime* and are dependent on the value attributed to each *anime* regarding the hasType

relation. The type attributed will determine what subclass of *anime* types it should be categorized as. Finally, "HighRatedAnime", "MediumRatedAnime" and "LowRatedAnime" seemed interesting to include as we can categorize an *anime* based on the score it has been given by the viewers.

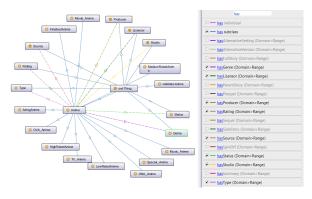


Figure 2: Properties that relate Animes to other classes.

2.2 Analysis of existing vocabularies or ontologies

After doing some research, 2 different ontologies related to *animes* were found. Some other ontologies regarding TV shows or movies were analyzed but none was fitting for the domain of this project.

The first ontology found is from DBpedia[5] where an *anime* is categorized as a subclass of "Cartoon". The second ontology found is from Wikidata[6] where an *anime* is categorized as a subclass of "Animated Film of Japan". It is also part of the "Anime and Manga" class and says to be an instance of "Film Genre". Both these ontologies refer each other in the equivalent class section.

Although it was hard to find a suitable ontology for this project, it doesn't mean it does not exist and perhaps with extensive research, it may be possible to find one in the future.

3 Building an ontology using *Protégé*

For the creation of the ontology, we used *Protégé*, an open-source framework and knowledge

management software, that provides a graphic user interface to define ontologies [7]. It supports a lot of plug-ins that can help in the creation of complex ontologies.

The ontology was created using the *OWL* language, which eases the interpretation of knowledge by machines. OWL has a rich set of operators, which makes possible to create complex languages out of simple concepts. This language can also verify the consistency and correctness of the concepts and architecture of the hierarchy [8].

The first step on the creation of the ontology was to define its International Resource Identifier (IRI). Then, the classes *Anime*, *Genre*, *Licensor*, *Producer*, *Rating*, *Source*, *Status*, *Studio* and *Type* were created, as well as the *Anime* subclasses already mentioned that will classify the different *anime* individuals included in our ontology through the reasoner. Both existing ontologies mentioned in the previous section were included in our ontology so we could classify our own *Anime* class as equivalent to those from DBpedia and Wikidata.

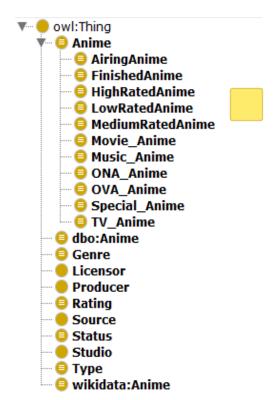


Figure 3: Classes defined for the Anime ontology.

The object properties represent the relation

between classes or instances of classes. These properties were already described in the previous section and can be seen in the figure below. Regarding their characteristics each of the properties was analyzed and the ones found most suitable were attributed, as well as the domains and ranges for each property, as it can be seen in figure 15.



Figure 4: Object Properties.

The data properties describe relations between the individuals and their data values. Regarding these properties, most of them were defined as functional, since they are only defined once for each individual. The properties defined are to represent the data related to each anime or just to give a name to the individuals of the other classes (represented by the property with the same name as the class). The properties attributed can be seen in figure 14, as well as the domain and range for each one.

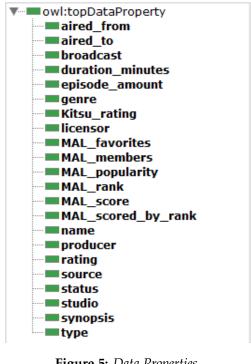


Figure 5: Data Properties.

Populating the ontology

Due to the large size of the dataset, more than 6900 animes, it was not feasible to populate the ontology with all the examples of the dataset, so only a small part was translated into *RDF*.

After some unsuccessful tries into populating the ontology directly with the data from the files collected on the previous iterations, the examples used were added manually. The feature to import data was not very intuitive and there wasn't a lot of time to explore other possible ways to import data faster. The small sample imported to the ontology was first analyzed to cover the most cases possible of the ontology.

3.2 Queries

The queries below were created using the SPARQL Query and Snap SPARQL Query plugins on Protégé.

1 - Retrieve the japanese name of an anime by its english name.

```
SELECT ?name WHERE {
```

Figure 6: Result of query 1.

2 - Retrieve all *animes* with a medium rated score on my *anime* list, ordered from higher to lower.

```
SELECT ?name ?rating WHERE {
?anime a ont:Anime;
  ont:name ?name;
  ont:MAL_score ?rating
  FILTER(langMatches(lang(?name), "en")
  &&
    (?rating >= 5) && (?rating < 8)).
} ORDER BY DESC(?rating)</pre>
```

name	rating
"Tokyo Ghoul:re"@en	"7.89"^^ <http: 2001="" td="" www.w3.org="" xmlschema#float<=""></http:>
"Sword Art Online"@en	"7.64"^^ <http: 2001="" td="" www.w3.org="" xmlschema#float<=""></http:>
"Tokyo Ghoul √A"@en	"7.45"^^ <http: 2001="" td="" www.w3.org="" xmlschema#float<=""></http:>
"Sword Art Online Alternative: Gun Gale Online"@en	"7.3"^^ <http: 2001="" www.w3.org="" xmlschema#float=""></http:>
"Sword Art Online II"@en	"7.2"^^ <http: 2001="" www.w3.org="" xmlschema#float=""></http:>
"Space Battleship Tiramisu"@en	"6.08"^^ <http: 2001="" td="" www.w3.org="" xmlschema#float<=""></http:>
"Regalia: The Three Sacred Stars"@en	"6.03"^^ <http: 2001="" td="" www.w3.org="" xmlschema#float<=""></http:>

Figure 7: Result of query 2.

3 - Retrieve all *animes* with the genre Sci-fi that are sequels of other *animes*.

```
SELECT DISTINCT ?name WHERE {
?z a ont:Genre; ont:genre "Sci-Fi".
?y a ont:Anime.
?x a ont:Anime; ont:name ?name;
ont:hasGenre ?z.
?y ont:hasSequel ?x
}
```

```
?name
Steins;Gate: Oukoubakko no Poriomania@jp
Steins;Gate: Egoistic Poriomania@en
Steins;Gate 0@en
```

Figure 8: Result of query 3.

4 - Retrieve the *animes* previous to 2017 that have a good rank (< 100) on my *anime* list, ordered by the obtained score on that platform.

```
SELECT ?name ?rank WHERE {
?x rdf:type ont:Anime;
ont:name ?name ;
ont:MAL_rank ?rank;
ont:aired_to ?endDate.

FILTER (langMatches(lang(?name), "en")
&& ?endDate <
"2017-01-01T00:00:00"^^xsd:dateTime
&& ?rank < 100).
} ORDER BY DESC(?rank)
```

Figure 9: Result of query 4.

5 - Retrieve the total amount of *animes* for each genre registered in the ontology, ordered by higher to lower.

```
ActionGenre
                         12"^^<http://www.w3.org/2001/XMLSchema#integer
                        "7"^^<http://www.w3.org/2001/XMLSchema#integer>
Sci-FiGenre
                        "6"^^<http://www.w3.org/2001/XMLSchema#integer>
FantasyGenre
                        "5"^^<http://www.w3.org/2001/XMLSchema#integer>
GameGenre
DramaGenre
                        "5"^^<http://www.w3.org/2001/XMLSchema#integer>
MysteryGenre
                        "5"^^<http://www.w3.org/2001/XMLSchema#integer>
SeinenGenre
                        "5"^^<http://www.w3.org/2001/XMLSchema#integer>
PsychologicalGenre
                        "4"^^<http://www.w3.org/2001/XMLSchema#integer>
                        "3"^^<http://www.w3.org/2001/XMLSchema#integer>
ThrillerGenre
RomanceGenre
                        "3"^^<http://www.w3.org/2001/XMLSchema#integer>
```

Figure 10: *Result of query 5.*

6 - Retrieve all the *animes* that contain the word "human" in the synopsis and its genres

```
SELECT distinct ?name
(group_concat(?genreName;separator=",_")
as ?genres) ?synopsis WHERE {
?x ont:name
?name; ont:hasGenre ?genre;
ont:synopsis ?synopsis.
?genre ont:genre ?genreName
FILTER( regex(?synopsis, "human"))
} group by ?name ?synopsis
```

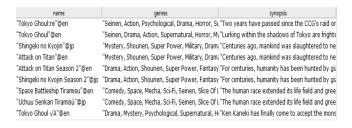


Figure 11: Result of query 6.

4 Conclusion

With this work, it was possible to learn how knowledge is represented and linked in the web, so that when a user searches for certain data, related connected data is retrieved as well.

The main goal of this milestone was successfully achieved, not only an ontology about *animes* was created and explored through queries, but also the domain of the project was clear and the main concepts about semantic web were understood.

During the whole process of creation of the ontology, we realized the importance and the power of representing knowledge and linking data in an ontology, that can be available for and analyzed by everyone in the web. *OWL* showed to be a really powerful and rich language that can easily represent knowledge about things and relations between things, leading to complex representations of concepts. Besides learning a new language, it was really advantageous to use *Protégé* and get to know some of its features.

References

- [1] Anime Wikipedia, https://en.wikipedia.org/wiki/Anime, 2018. [Online, accessed 13-October-2018]
- [2] What is Anime? A Brief History of Anime Genres, Culture and Evolution The Daily Dot https://www.dailydot.com/parsec/what-is-anime/, 2018. [Online, accessed 14-October-2018].
- [3] A Brief History of Anime Thought Co, https://www.thoughtco.com/

- brief-history-of-anime-144979, 2018. [Online, accessed 14-October-2018].
- [4] History of Anime Wikipedia, https://en.wikipedia.org/wiki/History_of_anime, 2018. [Online, accessed 13-October-2018].
- [5] Ontoloy DBpedia, https://wiki.dbpedia.org/services-resources/ontology, 2018. [Online, accessed 20-December-2018].
- [6] Wikidate, https://www.wikidata.org/wiki/Wikidata:Main_Page, 2018. [Online, accessed 20-December-2018].
- [7] Protégé, https://protege.stanford.edu/, 2018. [Online, accessed 22-December-2018]
- [8] W3C, OWL 2 Web Ontology Language Primer (Second Edition), W3C Recommendation 11 December 2012, http://www.w3.org/TR/owl2-primer/, 2018. [Online, accessed at 23-December-2018].

5 Appendix

5.1 Ontology

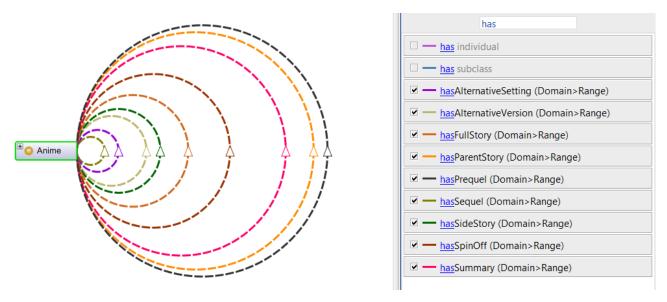


Figure 12: *Properties that relate entities of the Anime class.*

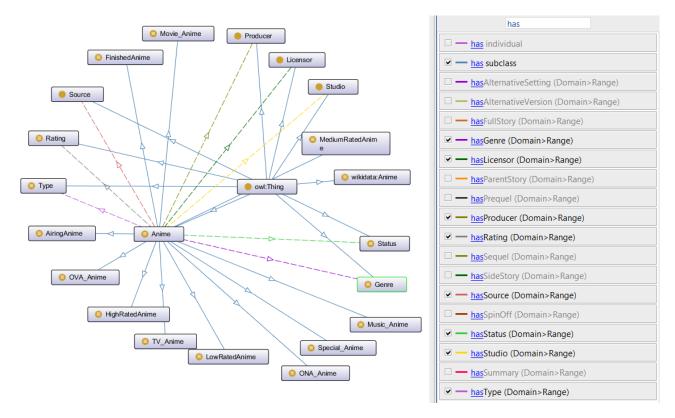


Figure 13: Properties that relate entities of the Anime class with other classes.

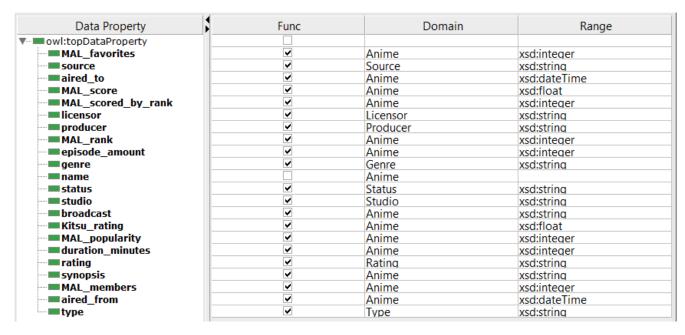


Figure 14: Data Properties Matrix.

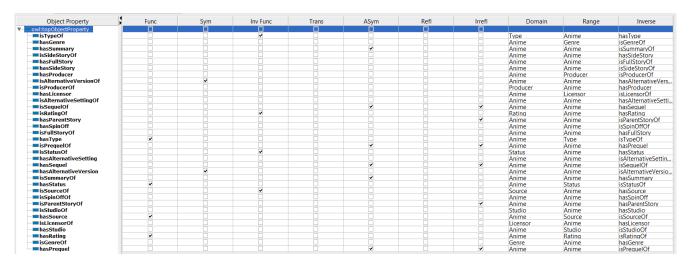


Figure 15: Object Properties Matrix.

5.2 Queries Results

Query 1



Query 2

name	rating
"Tokyo Ghoul:re"@en	"7.89"^^ <http: 2001="" www.w3.org="" xmlschema#float=""></http:>
"Sword Art Online"@en	"7.64"^^ <http: 2001="" www.w3.org="" xmlschema#float=""></http:>
"Tokyo Ghoul √A"@en	"7.45"^^ <http: 2001="" www.w3.org="" xmlschema#float=""></http:>
"Sword Art Online Alternative: Gun Gale Online"@en	"7.3"^^ <http: 2001="" www.w3.org="" xmlschema#float=""></http:>
"Sword Art Online II"@en	"7.2"^^ <http: 2001="" www.w3.org="" xmlschema#float=""></http:>
"Space Battleship Tiramisu"@en	"6.08"^^ <http: 2001="" www.w3.org="" xmlschema#float=""></http:>
"Regalia: The Three Sacred Stars"@en	"6.03"^^ <http: 2001="" www.w3.org="" xmlschema#float=""></http:>

Query 3

	?name
Steins;Gate: Oukoubakko no Poriomania@jp	
Steins;Gate: Egoistic Poriomania@en	
Steins;Gate 0@en	

Query 4

name	rank
"Steins;Gate"@en	"5"^^ <http: 2001="" www.w3.org="" xmlschema#integer=""></http:>

Query 5

animes	
"12"^^ <http: 2001="" www.w3.org="" xmlschema#integer=""></http:>	
"7"^^ <http: 2001="" www.w3.org="" xmlschema#integer=""></http:>	
"6"^^ <http: 2001="" www.w3.org="" xmlschema#integer=""></http:>	
"3"^^ <http: 2001="" www.w3.org="" xmlschema#integer=""></http:>	
	"12"^^ <http: 2001="" www.w3.org="" xmlschema#integer=""> "7"^^<http: 2001="" www.w3.org="" xmlschema#integer=""> "6"^^<http: 2001="" www.w3.org="" xmlschema#integer=""> "5"^^<http: 2001="" www.w3.org="" xmlschema#integer=""> "4"^^<http: 2001="" www.w3.org="" xmlschema#integer=""> "3"^^<http: 2001="" www.w3.org="" xmlschema#integer=""></http:></http:></http:></http:></http:></http:></http:></http:></http:></http:>

Query 6

name	genres	synopsis
"Tokyo Ghoul:re"@en	"Seinen, Action, Psychological, Drama, Horror, St	."Two years have passed since the CCG's raid or
"Tokyo Ghoul"@en	"Seinen, Drama, Action, Supernatural, Horror, My	"Lurking within the shadows of Tokyo are frighte
"Shingeki no Kyojin"@jp	"Mystery, Shounen, Super Power, Military, Drama	"Centuries ago, mankind was slaughtered to ne
"Attack on Titan"@en	"Mystery, Shounen, Super Power, Military, Drama	"Centuries ago, mankind was slaughtered to ne
"Attack on Titan Season 2"@en	"Drama, Action, Shounen, Super Power, Fantasy	"For centuries, humanity has been hunted by giz
"Shingeki no Kyojin Season 2"@jp	"Drama, Action, Shounen, Super Power, Fantasy	"For centuries, humanity has been hunted by giz
"Space Battleship Tiramisu"@en	"Comedy, Space, Mecha, Sci-Fi, Seinen, Slice Of L	"The human race extended its life field and gree
"Uchuu Senkan Tiramisù"@jp	"Comedy, Space, Mecha, Sci-Fi, Seinen, Slice Of L	"The human race extended its life field and gree
"Tokyo Ghoul √A"@en	"Drama, Mystery, Psychological, Supernatural, H	"Ken Kaneki has finally come to accept the mons