

Identifying Transmedia Works from User-Generated Knowledge Bases : Japanese Pop Culture Study Case

Stella Zevio¹, Tetsuya Mihara², and Shigeo Sugimoto²

¹ LIPN - CNRS UMR 7030, Université Paris 13, France
zevio@lipn.univ-paris13.fr

² Faculty of Library, Information and Media Science, University of Tsukuba, Japan
mihara@slis.tsukuba.ac.jp
sugimoto@slis.tsukuba.ac.jp

Abstract. As Japanese pop culture spreads worldwide, digital libraries compiling information about works through representative media (manga, anime, video games) emerge. Some of these works may share the same story, characters or universe, thus being part of a conceptual instance which we call a transmedia work in this paper. Transmedia works are abstract entities composed of works through several media linked together by semantic relationships. Identifying works belonging to the same transmedia work is still a challenge to enhance access, retrieval and organization of media in digital libraries. To overcome this challenge, semantic relationships between works should be identified. As no authority data yet describes semantic relationships between works, we need to find this information in knowledge bases generated by users such as Wikipedia. More precisely, we exploit DBpedia, Wikipedia's Linked Data counterpart, to respect the semantic web standards. In this paper, we present our method and experiment in building work entity datasets of Japanese pop culture (manga, anime and video games) and extracting relationships between these works in order to ease identification of transmedia works from the semantic data structure used in DBpedia. We also extract pertinent information to link works to bibliographic data in the future. We propose an evaluation of our contribution and demonstrate that we can easily and relevantly identify works belonging to the same transmedia work from user-generated knowledge bases.

Keywords: Transmedia work · Semantic Web · Digital Libraries · Linked data · Domain-dependent semantic data analysis

1 Introduction

Several works through different media can sometimes express the same story, take place in the same universe or exploit the same characters. In this case, these works are part of the same transmedia work [17].

Example 1. "Dragon Ball Z: Budokai" video game and "Dragon Ball" anime take place in the same universe and share some common characters, thus they are both part of the same transmedia work.

Identifying transmedia works would be useful for a better access, retrieval and organization of media, in particular for digital libraries. Indeed, identifying semantically linked works is still a challenge and a key issue for recommendation within digital libraries supporting various media formats.

As Japanese pop culture has long been considered as a subculture unworthy of interest, there is no sufficient authorized data of representative media nor reliable knowledge bases describing relations between works. Still, there are emerging digital libraries and databases of Japanese pop culture media, as interest in Japanese pop culture grows worldwide. Media Art Database[6] (MADB) is a database of manga, animation and video games published in Japan, produced by Agency for Cultural Affairs in Japan as the national authority of works through these media. However, MADB compiles information about works through different media but lacks information about relationships between these works. On the other hand, the information is available from knowledge bases generated by users, such as Wikipedia[7].

In this research, our aim is to identify transmedia works of Japanese pop culture, as no authority data describes them. To achieve this goal, we first extract work entity datasets of manga, anime and video games from user-based knowledge bases, then exploit semantic relationships described by users between these works to find works belonging to the same transmedia work. We use DBpedia[1], which is Wikipedia's Linked Open Data dataset, in order to extract relations between works according to the semantic web standards, in an interpretable and interoperable way. Using DBpedia enables us to take advantage of the simplicity, interoperability and interpretability brought by semantic web technologies. We choose to exploit English resources as they are known to be richer than Japanese ones. This method is of course heavily dependent on the data structure used in DBpedia thus we're discussing this issue in section 4. Our contribution lies at the interface between domain-dependent semantic data analysis and knowledge extraction from Linked Data.

This paper presents our method and experiment in building work entity datasets of manga, anime and video games and extracting semantic relationships between them in order to identify works belonging to the same transmedia work. In section 2, we present related work. In section 3 we present our experiment, results we obtained as well as an evaluation. In section 4 we present our conclusions and we discuss about further work.

2 Related Work

Bibliographic information describing semantic relationships between works is useful when it comes to transmedia publications like adaptations for example. The Functional Requirements for Bibliographic Records (FRBR)[13] model, developed by the International Federation of Library Associations and Institutions

(IFLA)[16], defines entities and their relationships for advanced functions of bibliographic records. In the FRBR model, work entity is defined as an abstract one to express distinct intellectual or artistic creation. Different editions or translations of the same creation are semantically connected to each other. In addition, works belonging to the same creation group also have semantic relationships between each other, for example, William Shakespeare’s *Romeo and Juliet* and its namesake film adaptation from Franco Zeffirelli.

The FRBR model is commonly used as a conceptual data model for bibliographic records [16], cataloging rules (Resource Description and Access (RDA)[8] being the most representative one) and even pop-culture databases. McDonough[12] evaluates the usability of the FRBR model to describe relationships between various editions, translations, and adaptations of video games. Jett[10] developed a conceptual model reflecting FRBR for video games and interactive media.

On the other hand, if the FRBR model intends describing relationships between entities, there is actually a lack of datasets or records describing such entity relationships, especially for Japanese pop-culture[11]. OCLC WorldCat Fiction Finder[3] provides data about relationships between different editions of the same work. Unfortunately, records for animation and video game are not well covered by this database.

A method for creating FRBR dataset from existing datasets and conventional bibliographic records is FRBRization[15]. For example, WorldCat Fiction Finder is populated from MARC[5] bibliographic and authority records by using OCLC FRBR Work-Set Algorithm[4]. He *et al.*[9] proposed a method for identifying FRBR Works using Wikipedia, through DBpedia articles for manga. DBpedia is used as a reference authority in order to identify Work level entities of manga in the catalog records of Kyoto Manga Museum which is the largest library for manga in Japan. Takhirov[14] proposed a method for linking a FRBR entity to its corresponding LOD entity and an evaluation using DBpedia and Amazon bookstore’s Web API. Although He and Takhirov focus on the information about books and do not show interest about transmedia works, they suggest that using DBpedia as a source of work entities and their relationships is a viable solution. As DBpedia has many resources about transmedia works including manga, anime and video games, our contribution aims at measuring the quantity and quality of transmedia works and semantic relationships between them that we can extract from DBpedia with simple SPARQL queries.

3 Experiment

3.1 Overview

In order to identify transmedia works, we conduct an experiment consisting in two steps. As few authority datasets are available, our first step described in section 3.2 consists in building our own work entity datasets of manga, anime and video games from DBpedia. The second step described in section 3.3 consists in exploiting semantic relationships described by users to link works through several media together.

We harvest DBpedia SPARQL endpoint as well as DBpedia Live[2] SPARQL endpoint and compare the results obtained with both. DBpedia Live SPARQL endpoint is the most up-to-date one as it is continuously synchronized with Wikipedia, while DBpedia SPARQL endpoint is only updated periodically. In a theoretical setting, we should expect more accurate results with DBpedia Live, assuming that knowledge available on DBpedia is growing larger and more accurate with the Wikipedia users' contributions.

3.2 Datasets

In DBpedia, a concept is described by an article. An article is defined as a RDF resource and additional information such as links to other articles are described as properties of the RDF resource. To determine than an article describes a manga, an anime or a video game, we exploit its *rdf:type* property. Indeed, an article describing a manga would have *rdf:type* property *dbo:Manga*. An article about an anime or a video game would have *rdf:type* property *dbo:Anime* or *dbo:VideoGame* respectively. We are building the manga, anime and video games datasets harvesting DBpedia and DBpedia Live with the SPARQL queries structure described in query 1.1. In table 1 we present the number of results obtained.

```
1 SELECT DISTINCT ?Concept WHERE {
2   ?Concept rdf:type dbo:Manga }
```

Listing 1.1. SPARQL query : Manga

Media	Number of results (DBpedia)	Number of results (DBpedia Live)
Manga	3783	3928
Anime	4271	5014
Video games	28869	20807

Table 1. Datasets of manga, anime and video games obtained by harvesting DBpedia and DBpedia Live on 12-20-2017

3.3 Identification of works belonging to the same transmedia work

In order to identify semantic relationships between works through several media, we exploit semantic links between articles describing works. For an example, an article describing an anime may have a *dct:subject* property which would apply to an article describing a manga. If that so, it would mean that this anime and this manga belong to the same transmedia work. We exploit any direct semantic relationship between articles about works through several media as well as some indirect ones. Queries are all derived from query structure shown in query 1.2. In table 2 we present the number of results obtained.

```
1 SELECT DISTINCT ?Manga ?Anime WHERE {
```

```

2      {?Manga rdf:type dbo:Manga .
3      ?Anime rdf:type dbo:Anime .
4      ?Anime ?p ?Manga}
5  UNION
6      {?Anime rdf:type dbo:Anime .
7      ?Manga rdf:type dbo:Manga .
8      ?Manga ?p ?Anime}
9  UNION
10     {?Anime rdf:type dbo:Anime .
11     ?Anime dct:subject ?Category .
12     ?Category skos:broader dbc:
13         Wikipedia_categories_named_after_anime_and_manga_series
14         .
15     ?Manga rdf:type dbo:Manga .
16     ?Manga dct:subject ?Category}}

```

Listing 1.2. SPARQL query : Manga-Anime belonging to the same transmedia work

Transmedia works	Number of results (DBpedia)	Number of results (DBpedia Live)
Manga - Anime	764	696
Manga - Video games	864	191
Anime - Video games	411	135

Table 2. Couples of works through different media belonging to the same transmedia work obtained by harvesting DBpedia and DBpedia Live on 12-20-2017

3.4 Extraction of links between works and bibliographic data

Digital libraries may compile informations about works through bibliographic data according to the FRBR model [3]. Therefore, it is a key issue to reconcile works to bibliographic data. In order to ease this reconciliation, we exploit semantic links between articles describing manga and list of chapters as well as anime and list of episodes according to the query structure shown in query 1.3. In table 3 we present the number of results obtained.

```

1  SELECT DISTINCT ?Manga ?List WHERE {
2      {?List dct:subject
3      dbc:Lists_of_manga_volumes_and_chapters .
4      ?Manga rdf:type dbo:Manga .
5      ?Manga ?p ?List}
6  UNION
7      {?Manga rdf:type dbo:Manga .
8      ?List dct:subject
9      dbc:Lists_of_manga_volumes_and_chapters .
10     ?List ?p ?Manga}

```

```

11 UNION
12   {?List dct:subject
13     dbc:Lists_of_manga_volumes_and_chapters .
14   ?List dct:subject ?Category .
15   ?Category skos:broader
16     dbc:
17       Wikipedia_categories_named_after_anime_and_manga_series
18       .
19   ?Manga rdf:type dbo:Manga .
20   ?Manga dct:subject ?Category}}

```

Listing 1.3. SPARQL query : Manga - List of chapters

Work - Bibliographic data	Number of results (DBpedia)	Number of results (DBpedia Live)
Manga - List of chapters	244	247
Anime - List of episodes	266	275

Table 3. Couples of works and bibliographic data obtained by harvesting DBpedia and DBpedia Live on 12-20-2017

3.5 Evaluation

As no gold standard is available for data about manga, anime nor video games and semantic links between them as far as we know from the literature, calculate a recall is impossible. It is difficult to judge whether or not our queries have a good coverage of the domain. A possible solution would be to manually collect all works related through several media for a certain number of known works then evaluate the recall of our method according to this restricted gold standard. However, even building a restricted gold standard requires a very high level of expertise and most experts would rely on user-generated knowledge bases at some point. Thus, we don't propose a recall measure.

Still, we can evaluate the accuracy of the queries and detect the relevance of the results returned. To estimate the relevance of our results, we conducted an evaluation consisting in randomly selecting 100 results for each query, asking two external experts of the domain to test the exactitude of each result. In the end, we obtain an accuracy as well as all errors raised. This information is available in tables 4, 5 and 6, along with error types encountered.

We managed to obtain overall precise results. As expected, results obtained with DBpedia Live SPARQL endpoint are more precise than with DBpedia SPARQL endpoint, but with a surprisingly huge gap between them. From the error types and the precision drop with DBpedia SPARQL endpoint concerning the construction of the video games dataset and the identification of transmedia works, we can assert that the results are heavily dependant on the semantic data structure described by the users, which may potentially be inconsistent, as it's human-generated data.

Query	DBpedia	Error types	DBpedia Live	Error types
① Manga	94 %	Something else than a manga (Manga genre, novel, film, wafer silicon, anime)	94 %	Something else than a manga (Novels)
② Anime	91 %	Something else than an anime (Drama, live action, director, studio, magazine, method of animation)	94 %	Something else than an anime (Drama, live action)
③ Video games	26 %	Something else than a video game (Card game, board game, Superbowl, gaming platform)	100 %	

Table 4. Evaluation on query results obtained on 12-20-2017 (Work entity datasets of manga, anime and video games)

Query	DBpedia	Error types	DBpedia Live	Error types
④ Manga - List of chapters	91 %	A list of chapters is associated to something that is not a manga (Volume, mangaka (person))	85 %	<ul style="list-style-type: none"> • A list of chapters is associated to something that is not a manga (Volume) • The list of chapters associated does not correspond to the manga (Chapters of a manga from the same series)
⑤ Anime - List of episodes	80 %	<ul style="list-style-type: none"> • Something else than an anime (DBpedia page about "Anime" in general) • The list of episodes does not correspond to the anime (episodes of an anime from the same series) 	90 %	<ul style="list-style-type: none"> • Something else than an anime (Visual novel) • The list of episodes does not correspond to the anime (episodes of an anime from the same or from different series)

Table 5. Evaluation on query results obtained on 12-20-2017

Query	DBpedia	Error types	DBpedia Live	Error types
⑥ Manga - Anime	8 %	<ul style="list-style-type: none"> • Something else than a manga (Manga genre, company, magazine, mangaka (person)) • Something else than an anime (DBpedia page about "Anime" in general, list of episodes) • The manga and the anime do not belong to the same transmedia work 	89 %	The manga and the anime do not belong to the same transmedia work
⑦ Manga - Video games	41 %	<ul style="list-style-type: none"> • Something else than a manga (Manga series, manga genre, company, magazine, mangaka (person)) • The manga and the video game do not belong to the same transmedia work 	77 %	<ul style="list-style-type: none"> • Something else than a manga (Novels) • The manga and the video game do not belong to the same transmedia work
⑧ Anime - Video games	32 %	<ul style="list-style-type: none"> • Something else than an anime (DBpedia page about "Anime" in general, company, visual novel) • Something else than a video game (Gaming hardware, video game genre, gaming platform) 	95 %	<ul style="list-style-type: none"> • Something else than an anime (Drama) • The anime and the video game do not belong to the same transmedia work

Table 6. Evaluation on query results obtained on 12-20-2017

4 Discussion and conclusion

With the help of very simple SPARQL queries, we managed to build work entity datasets of manga, anime and video games, which is hard to create manually by simple computational method without authority datasets. We prepared a future linkage between works and bibliographic data, by linking manga to their list of chapters and anime to their list of episodes. We also identified semantic relationships between manga, anime and video games, creating a semantic network that enables us to easily identify transmedia works. Although it is difficult to estimate the coverage of the domain as no gold standard is available, we managed to obtain satisfying results in terms of accuracy as well as a solid number of results. As expected, we obtained better results harvesting DBpedia Live SPARQL endpoint, which is the most up-to-date one.

We identified several limitations on this work. First, we use knowledge bases with user-generated content, which are not always exhaustive. Indeed, information may not be available in Wikipedia, or may be available in Wikipedia but not semantically described with accurateness in DBpedia. This limitation is closely related to the lack of authority data in this field, so it is a compromise that has to be made. Then, we obtained disparate results according to the SPARQL endpoint used. Therefore, using an up-to-date endpoint is a key feature. Indeed, consistency of user-generated data is not ensured.

To pursue this work, an interesting research question would be to determine how to link data to records of publications in different countries. A comparison between English and Japanese resources would help us determine if multilingual processes would help us expand our results.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 16H01754.

References

1. DBpedia. <https://dbpedia.org>, [Online; accessed 26-July-2018]
2. DBpedia Live. <http://live.dbpedia.org/sparql>, [Online; accessed 26-July-2018]
3. FictionFinder: A FRBR-based Prototype for Fiction in WorldCat. <https://www.oclc.org/research/activities/fictionfinder.html>, [Online; accessed 15-August-2018]
4. FRBR Work-Set Algorithm. <https://www.oclc.org/research/activities/frbralgorithm.html>, [Online; accessed 15-August-2018]
5. MARC. <http://www.loc.gov/marc/umb/um01to06.html>, [Online; accessed 15-August-2018]
6. Media Art Database. <https://mediaarts-db.bunka.go.jp/>, [Online; accessed 26-July-2018]
7. Wikipedia. <https://en.wikipedia.org>, [Online; accessed 26-July-2018]
8. Steering Committee, T.R.: About RDA. <http://rda-rsc.org/content/about-rda>, [Online; accessed 15-August-2018]

9. He, W., Mihara, T., Nagamori, M., Sugimoto, S.: Identification of Works of Manga Using LOD Resources: An Experimental FRBRization of Bibliographic Data of Comic Books. In: Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries. pp. 253–256. JCDL '13 (2013)
10. Jett, J., Sacchi, S., Lee, J.H., Clarke, R.I.: A Conceptual Model for Video Games and Interactive Media. *J. Assoc. Inf. Sci. Technol.* **67**(3), 505–517 (2016)
11. Kiryakos, S., Sugimoto, S., Nagamori, M., Mihara, T.: Aggregating Metadata from Heterogeneous Pop Culture Resources on the Web. In: International Conference on Dublin Core and Metadata Applications. pp. 65–74 (2017)
12. McDonough, J., Kirschenbaum, M., Reside, D., Fraistat, N., Jerz, D.: Twisty Little Passages Almost All Alike: Applying the FRBR Model to a Classic Computer Game. *Digital Humanities Quarterly* **4**(2) (2010)
13. IFLA Study Group on the Functional Requirements for Bibliographic Records: Functional Requirements for Bibliographic Records. <https://www.ifla.org/publications/functional-requirements-for-bibliographic-records>, [Online; accessed 15-August-2018]
14. Takhirov N., Duchateau F., A.T.: Linking FRBR Entities to LOD through Semantic Matching. *Research and Advanced Technology for Digital Libraries. TPDFL 2011. Lecture Notes in Computer Science* **6966**, 69–76 (2011)
15. Takhirov N., Duchateau F., A.T.: Supporting FRBRization of Web Product Descriptions. *Research and Advanced Technology for Digital Libraries. TPDFL 2011. Lecture Notes in Computer Science* **6966**, 284–295 (2011)
16. Tillett, B.: What is FRBR? A Conceptual Model for the Bibliographic Universe. *Australian Library Journal* **54**(1), 24–30 (2005)
17. Vukadin, A.: Bits and Pieces of Information: Bibliographic Modeling of Transmedia. *Cataloging & Classification Quarterly* **52**(3), 285–302 (2014)