

Working with vagueness: A pragmatic incremental approach to ontology development in the Japanese Visual Media Graph project

Zoltan Kacsuk

Stuttgart Media University, Germany

kacsuk@hdm-stuttgart.de

 04 December 2023

Abstract

The Japanese Visual Media Graph (JVMG) project – funded by the German Research Foundation’s (Deutsche Forschungsgemeinschaft, DFG) e-Research Technologies program – is developing a knowledge graph (available at <https://mediagraph.link/>) for researchers working with popular Japanese visual media such as anime, manga, video games and so on [1]. The project is working from the same premise the Databased Infrastructure for Global Games Culture Research (diggr) project has already demonstrated to be very productive for media studies research [2][3], namely the fact that there exist large amounts of high quality, granular, well-structured metadata online that are compiled by various fan or enthusiast communities [4][5]. By working with such online communities towards integrating their descriptive metadata resources into a single knowledge graph – this type of data aggregation approach has also been explored by [6] – the JVMG project aims to enable new forms of quantitative research for researchers working with these domains of interest, and at the same time provide a template for building similar resources in other areas of inquiry.

The ontology development process for the JVMG knowledge graph faces three distinct sources of challenges: 1) challenges stemming from the domain, 2) challenges arising from the data sources, and finally, 3) challenges from the goals of the project itself.

Starting with domain specific challenges, first, 1a) work on modeling areas such as manga is still not as developed as that of other more established publication types [6][7]. Second, 1b) media mixes or multimedia franchises are both common in and a defining element of popular Japanese visual media. In this domain data is often also gathered on the level of the media mix, franchise or ‘superwork’ entity, which needs to be adequately handled and connected to the other levels of the ontology (such as the abstract work or actual release level) [8][9][10]. And third, 1c) the centrality of fictional characters [11] is also reflected in the data models of the communities [12][13][14] as well as the research interests of the researchers working on this domain [15][16]. Thus, it is very important to be able to develop an ontology that can handle information on characters and their relationship to various levels of the ontology describing, for

example, the works they appear in adequately.

Next up, the challenges stemming from the data sources employed in the JVMG project again fall into three large categories. 2a) The challenges arising from the heterogeneity of the data sources, are one of the main reasons the JVMG knowledge graph is developed as a linked data resource using RDF technology [4][17]. RDF technology further enables the project to accommodate 2b) changes in the structure of the source data in a more flexible manner [4][17]. Last, and most importantly, 2c) the source data itself is not always as strictly structured and defined as an ideal version of a data model for the domain could be.

Finally, the third group of challenges arising from the goals of the JVMG project itself, most prominently the need to serve the needs of the researchers working on the domain, place further practical constraints on the ontology development process. 3a) The JVMG project needs to be able to provide researchers with material to work with as soon as possible, even at the potential cost of keeping the data model potentially simpler. In the experience of the pilot research questions that have already been explored in the framework of the project 3b) data completeness (or lack thereof) can be an issue for a range of questions that researchers would like to explore. Furthermore, due to the expanding nature of the knowledge graph and the ongoing work on incorporating further data sources, 3c) the JVMG ontology needs to be both flexible and extensible.

The presentation will introduce our “two phase – two layer” approach to ontology development in the JVMG project, which aims to answer the above listed challenges in a pragmatic and incremental way, with a special focus on “bridge characters.” This entity type is unique to the JVMG ontology, and it was introduced to provide a solution for the vagueness found in some of our source data, and to enable the generation of “good enough” data for actual research use cases.

References

- [1] Pfeffer, Magnus and Martin Roth, 2020. Japanese Visual Media Graph: Providing researchers with data from enthusiast communities. In: *2019 Proceedings of the International Conference on Dublin Core and Metadata Applications*, eds. Koraljka Golub et al. Seoul: Dublin Core Metadata Initiative (DCMI), 136-41.
- [2] Roth, Martin, 2023. Was Wird Denn Hier Gespielt? Genrebezogene Unterschiede Der Digitalen Spielelandschaft in Deutschland Und Japan. In: *Spielzeichen IV: Genres – Systematiken, Kontexte, Entwicklungen*, eds. Martin Hennig and Hans Krahl. Glückstadt: Verlag Werner Hülsbusch, 207-30.

- [3] Roth, Martin, 2021. Terebigēmu bunka no kūkanteki tenkai: FromSoftware gēmu no seianteki-ryūtsuteki tenkai to sore ni okeru sofuto na bunkateki kyōkaisen. *Replaying Japan* 3, 127-41, <https://doi.org/10.34382/00014541>.
- [4] Kiryakos, Senan, and Magnus Pfeffer, 2021. Exploring the research utility of fan-created data in the Japanese visual media domain. *International Conference on Asian Digital Libraries*. Cham: Springer International Publishing, 210-218.
- [5] Price, Ludi and Lyn Robinson, 2017. ‘Being in a knowledge space’: Information behaviour of cult media fan communities. *Journal of Information Science* 43(5), 649-64, <https://doi.org/10.1177/0165551516658821>.
- [6] Kiryakos, Senan, Shigeo Sugimoto, Mitsuharu Nagamori and Tetsuya Mihara, 2017. Aggregating metadata from heterogeneous pop culture resources on the Web. In: *International Conference on Dublin Core and Metadata Applications 2016*, 65-74.
- [7] Morozumi, Ayako, Satomi Nomura, Mitsuharu Nagamori and Shigeo Sugimoto, 2009. Metadata framework for Manga: A multi-paradigm metadata description framework for digital comics. In: *International Conference on Dublin Core and Metadata Applications*, 61-70.
- [8] Kiryakos, Senan and Shigeo Sugimoto, 2018. The representation of a multimedia franchise as a single entity: Contrasting existing bibliographic entities with web-based superwork portrayals. *LIBRES: Library and Information Science Research Electronic Journal* 28(2), 4057.
- [9] Zevio, Stella, Tetsuya Mihara and Shigeo Sugimoto, 2018. Identifying Transmedia Works from User-Generated Knowledge Bases: Japanese Pop Culture Study Case. *The Fifth International Workshop on Practical Application of Ontology for Semantic Data Engineering (PAOS2018)*.
- [10] Lee, Jin Ha et al., 2018. Reconceptualizing superwork for improved access to popular cultural objects. *Proceedings of the Association for Information Science and Technology* 55(1), 27481.
- [11] Azuma, Hiroki, 2009 [2001]. *Otaku: Japan's Database Animals*. Introduction and translation by Jonathan E. Abel and Shion Kono. Minneapolis: University of Minnesota Press.
- [12] Kiryakos, Senan and Magnus Pfeffer, 2021. Japanese Visual Media Graph – Visual

Novel Database Ontology. *Zenodo*, 10.5281/zenodo.5036040.

- [13] Kiryakos, Senan and Magnus Pfeffer, 2021. Japanese Visual Media Graph – Anime Characters Database Ontology. *Zenodo*, 10.5281/zenodo.5710959.
- [14] Kiryakos, Senan and Magnus Pfeffer, 2021. Japanese Visual Media Graph – AnimeClick Ontology. *Zenodo*, 10.5281/zenodo.5508683.
- [15] Bruno, Luca, 2021. A Glimpse of the Imaginative Environment: Exploring the Potential of Data-driven Examinations of Visual Novel Characters. In: *Japan's Contemporary Media Culture between Local and Global: Content, Practice and Theory*, eds. Martin Roth et al. Heidelberg: CrossAsia, 143-70, <https://doi.org/10.11588/crossasia.971.c12882>.
- [16] Kacsuk, Zoltan, 2021. Using Fan-Compiled Metadata for Anime, Manga and Video Game Research: Revisiting Azuma's "Otaku: Japan's Database Animals" Twenty Years On. In: *Japan's Contemporary Media Culture between Local and Global: Content, Practice and Theory*, eds. Martin Roth et al. Heidelberg: CrossAsia, 117-42, <https://doi.org/10.11588/crossasia.971.c12881>.
- [17] Kiryakos, Senan and Magnus Pfeffer, 2021. The Benefits of RDF and External Ontologies for Heterogeneous Data: A Case Study Using the Japanese Visual Media Graph. In: *Information between Data and Knowledge, Proceedings of the 16th International Symposium of Information Science (ISI 2021)*. Glückstadt: Verlag Werner Hülsbusch, 308-320, <https://doi.org/10.5283/epub.44950>.