Linked-Potter: an example of ontology for the study of the evolution of literature and readers communities

Federico Pianzola

University of Milano-Bicocca, Italy and Sogang University, South Korea

In the case of libraries and other book archives, ontologies and linked data are mostly used for metadata describing the materiality of books or paratextual information. I present here an example of how an ontology can be used for the distant reading of literary texts, to study literary history, the cultural evolution of fiction, or as a selection mechanism to identify themes of interest. I created a knowledge base using the tags of the fanfiction website Archive of Our Own (AO3) (Organization for Transformative Works, 2009), which has implemented an excellent system of tags management (Dalton, 2012; McCulloch, 2019).

When publishing on AO3, authors can specify tags for characters, relationships, and additional freeform tags for any use they may think of. Autocompleting typing suggest canonical forms for the tags, so that uniformity is guaranteed across the all archive. Moreover, specialized volunteers, called "wranglers," aggregate synonym tags: e.g. "harrypotter" and "Harry Potter" (AO3 Admin, 2012). The goal of AO3 is to help readers find exactly the kind of stories they are looking for, but researcher can exploit the well-maintained and accurate tags database to draw insights about the history and evolution of a specific genre of literature (fanfiction) and its readership. In particular, freeform tags offer authors the possibility to make explicit in the metadata any relevant aspect of the story, like a psychological trait of a character (e.g. "Morally grey Harry Potter"), a narrative strategy (e.g. "point of view of Draco"), a setting (e.g. "Diagon Alley"), a timeframe (e.g. "post first war with Voldemort"), etc. A distant reading of fanfiction through the lens of tags has benefits that go beyond the understanding of a widespread – and growing – cultural phenomenon. Data driven insights from research on AO3 can be used to formulate better hypotheses regarding the evolution of other cultural systems – like literary classics or genre fiction – and to more strategically plan labour-intensive and time-consuming tasks like manual annotation of textual corpora.

A current limitation of AO3 is that, when tags are aggregated, the "canonical tag" is linked to the synonym tag in the backend database but in the frontend only the user-generated synonym tag is displayed. Therefore, readers can retrieve all stories linked to a canonical tag thanks to the built-in search engine, but the link is lot when content is scraped from the website. In order to benefit from AO3 tags aggregation, but also to improve it for research purposes, I used the software Protégé (Musen, 2015) and the Web Ontology Language (OWL) to replicate AO3-generated ontology for the Harry Potter fandom, adding further subclasses and properties. These are the steps I followed:

1. create classes for the four main categories (FandomTag, CharacterTag, RelationshipTag, and FreeformTag) and relevant subclasses: LoveTag, FriendshipTag, FreeformCharacterTag, FreeformRelationshipTag, FreeformPlotTag, FreeformPlaceTag, FreeformTimeTag;
2. copy all the tags from the main page of the fandom tag "Harry Potter - J. K. Rowling" (Anon, n.d.), create objects of the type owl:NamedIndividual for each of them, and assign them to the respective classes;
3. define which tags are considered canonical in the AO3 database;

4. copy the synonyms of every canonical tag from the tag's page, e.g. for "Hermione Granger" (Anon, n.d.); create objects of the type owl:NamedIndividual for each of them; link them to the respective canonical tag through the property owl:SameAs;
5. link CharacterTags to the RelationshipTags through the property "participatesIn" and define "hasParticipant" as the inverse property;
6. link FreeformTags to the CharacterTags and RelationshipTags through the property "isTagOf" and define "isTaggedAs" as the inverse property;
7. to complete the whole knowledge base, run the reasoner (Sirin et al., 2007) to resolve coreferences and infer axioms.

An example of how data are linked is shown in Fig. 1.
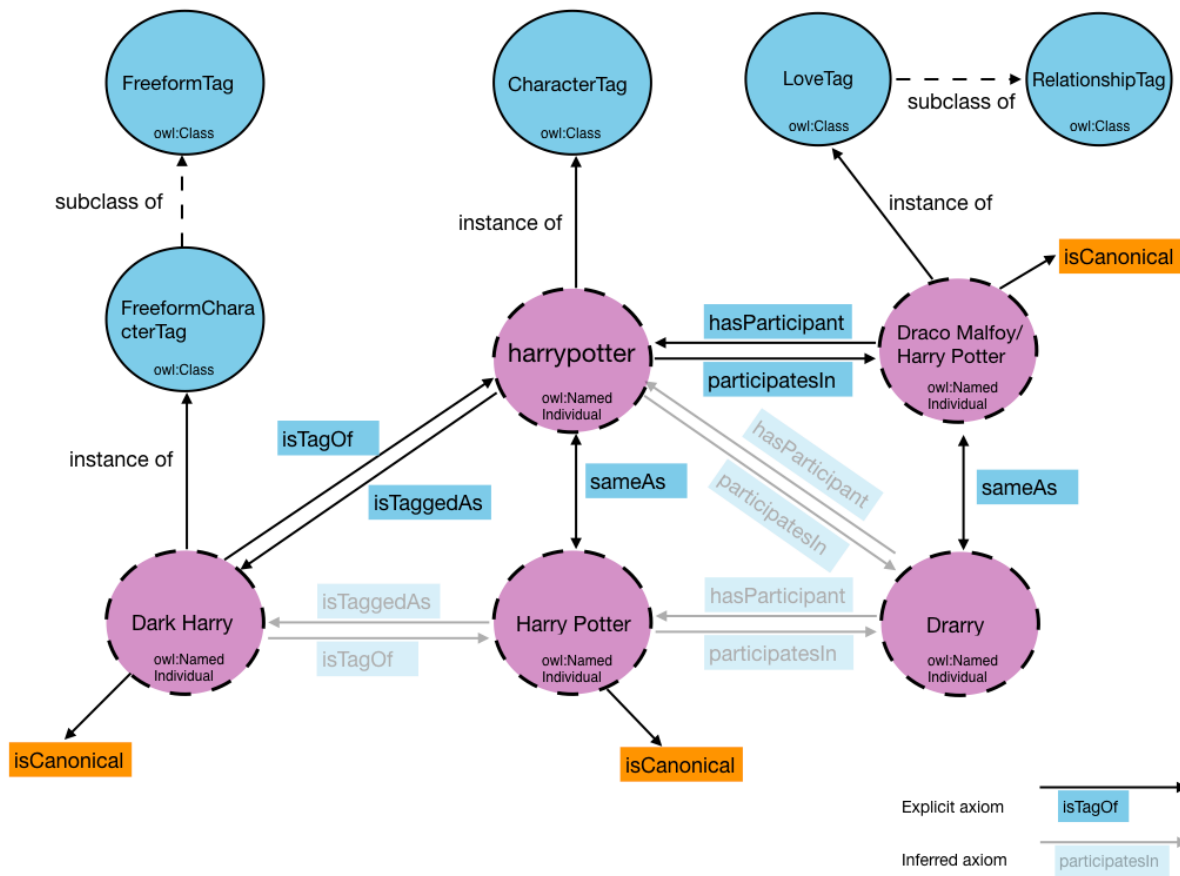
Fig. 1 Example from the knowledge base of the Harry Potter fandom on AO3

The resulting knowledge base has around 33,000 individuals (tags) linked between them. It is complete with respect to all the data retrievable scraping AO3 frontend website for the Harry Potter fandom. Using the knowledge base with a dataset retrieved in March 2020 (217,772 stories), the number of synonyms tags found for each class are: 247 CharacterTags (31,785 occurrences, 3.3% of the total occurrences for this class), 317 FreeformTags (54,517 occurrences, 4.6% of the total), 48 for RelationshipTags (13,833 occurrences, 4.5% of the total). Using aggregated tags allows to perform more complete analysis on the selected dataset.

Moreover, specific information regarding various topics can be extracted from the knowledge base and used to analyze the metadata of the stories published on AO3 or to find subset of stories with specific features. For instance, we can aggregate all synonyms of the

FreeformTags associated to the CharacterTag "Harry Potter" and count the number of stories written every year about a certain character version of the young wizard. The themes that attracts the most interest are female versions of Harry, his profession after graduating (Auror), the last stage of mastery of wizardry reached in the official novels (Master of Death), and a "dark" version which drastically changes the plot of the official novels (Slytherin Harry Potter) (Fig. 2).
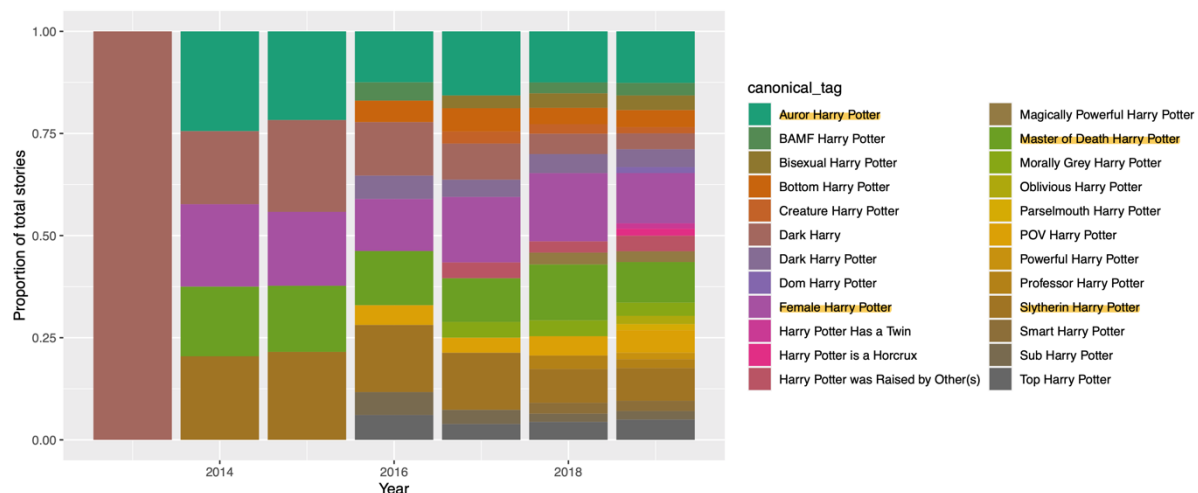
Fig. 2 FreeformTags associated to the character "Harry Potter" in different years. The four most frequent tags are highlighted in yellow

A broader study taking advantage of the Harry Potter ontology is Pianzola et al. (2020), which studied the cultural evolution and growing diversification of Harry Potter fanfiction.

Compared to other archives of online fiction, like fanfiction.net and wattpad.com, AO3 has metadata that can be very useful for the study of literature in a digital age. However, since their database structure is not publicly available, knowledge bases like Linked-Potter have to be created, with the advantage that additional information can be classified to explore specific aspect of plot or style.

References

AO3 Admin. (2012). *The Past, Present, and Hopeful Future for Tags and Tag Wrangling on the AO3*. *Archive of Our Own*. https://archiveofourown.org/admin_posts/267 (accessed 25 February 2020).

Dalton, K. L. (2012). *Searching the Archive of Our Own: The Usefulness of the Tagging Structure*. University of Wisconsin-Milwaukee.http://dc.uwm.edu/etd/26/ (accessed 24 June 2013).

Harry Potter - J. K. Rowling. *Archive of Our Own*. https://archiveofourown.org/tags/Harry%20Potter%20-%20J*d*%20K*d*%20Rowling (accessed 5 May 2020a).

Hermione Granger. *Archive of Our Own*. https://archiveofourown.org/tags/Hermione%20Granger (accessed 5 May 2020b).

McCulloch, G. (2019). *Fans Are Better Than Tech at Organizing Information Online | WIRED*. *Wired*. https://www.wired.com/story/archive-of-our-own-fans-better-than-tech-organizing-information/ (accessed 25 February 2020).

Musen, M. A. (2015). The Protégé Project: A Look Back and a Look Forward. *AI Matters*, **1**(4), pp. 4–12. 10.1145/2757001.2757003.

Organization for Transformative Works. (2009). *Archive of Our Own*.https://archiveofourown.org/ (accessed 5 May 2020).

Pianzola, F., Acerbi, A. and Rebora, S. (2020). Cultural Accumulation and Improvement in Online Fan Fiction. *OSF Preprint*. 10.31219/osf.io/4wjnm.

Sirin, E. et al. (2007). Pellet: A Practical OWL-DL Reasoner. *Journal of Web Semantics*, **5**(2), pp. 51–3. 10.1016/j.websem.2007.03.004.