

The GOLEM-Knowledge Graph and Search Interface: Perspectives into Narrative and Fiction

Franziska Pannach^{1,*}, Luotong Cheng¹ and Federico Pianzola¹

¹Centre for Language and Cognition, University of Groningen, The Netherlands

Abstract

This contribution presents the GOLEM Knowledge Graph and interface, offering different perspectives into content-related data and metadata from the domain of fanfiction narratives. The Knowledge Graph is aligned with common ontologies and vocabularies from the domains of narrative and cultural heritage. In this short paper, we outline how narrative organization and characters' features are modelled in the GOLEM knowledge graph. The GOLEM UI is also presented, a user-friendly access point to the data that allows to browse the knowledge graph even without knowledge of SPARQL.

Keywords

narrative structure, knowledge graphs, literature and fiction, semantic web technology, fanfiction

1. Introduction

One of the main aims of the GOLEM (Graph Ontologies for Literary Evolution Models) project is to build an ontology¹ that can be used to model narratives independent of their domain of application (e.g. fiction or news), their association with a literary tradition, or their geographical and cultural contexts. The formal semantic model is also designed to be language-agnostic and independent of the format of the stories. Such a model should be able to express how narrative elements (e.g. events), characters, and their individual representations, as well as readers' engagement and literary evolution, are related to each other. The theoretical framework guiding the creation of the model is grounded in literary theory, narratology, and best practices of formal ontology design.


The ontology is used to model the data of the GOLEM triple store [10], which contains over eight million stories, into a knowledge graph (KG). The KG contains a subset of ca. 19.000 stories from the original triple store, and will be extended continuously in the near future. At the moment, the modelled data are fanfiction stories from the popular online-platform *Archive of Our Own* (AO3) [3]². This particular genre of stories holds immense potential, not only as a case study for modeling the literary domain, but also for in-depth study of user-produced narratives, reader response, semantic and narrative modeling approaches, and for the development of natural language processing (NLP) tools. Within the communities associated with

CHR 2024: Computational Humanities Research Conference, December 4–6, 2024, Aarhus, Denmark

*Corresponding author.

✉ f.a.pannach@rug.nl (F. Pannach); skylarcheng585@gmail.com (L. Cheng); f.pianzola@rug.nl (F. Pianzola)

🆔 0000-0003-4216-8410 (F. Pannach); 0009-0002-6567-8923 (L. Cheng); 0000-0001-6634-121X (F. Pianzola)

 © 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

¹<https://ontology.golemlab.eu/>

²<https://archiveofourown.org>

a specific fandom, fanfiction works are polyvocal interpretations of shared stories, internet folklore that is produced at a growing pace [11]. The works are polyvocal in the sense that, through a work, reader and writer are in dialogue with a canonical narrative universe that inspired it, but they are also uniquely engaged with each other. Not only in the traditional cycle of production-reception, but also through active and immediate engagement via comments or other interactions (e.g. kudos). This engagement can be measured and subsequently used to study the evolution of (cultural) traits of literary works, e.g. the appearance (and disappearance) of certain character traits, or the change in character roles (villain-to-hero, enemy-to-lover, secondary-to-main character).

In order to make these valuable data more easily accessible to interested researchers and other stakeholders, we created the GOLEM-UI, an easy-to-use interface based on the SAMPO framework [7, 6].³ This is an example of good practice also adopted by other Digital Humanities projects [4].

2. Domain Modelling

In the knowledge graph, a subset of the triple store data has been modelled. Fanfiction stories and canonical works are represented as instances of *lrm:F1_Work* [12]. Characters have two different class representations: *gc:G1_Character* (a *crm:E89_Propositional_Object*) for instances that appear in a specific story, and *gc:G0_Character-Staff* [17, 16] (a *crm:E28_Conceptual_Object*) that refers to all the possible variations (*Staff*) of a character. This allows modelling the relationship between an instance of a character in a specific version of a narrative material, e.g. Harry Potter in the novel *Harry Potter and the Philosopher's Stone*, and the general idea of the character that appears in different books and is thus a set of various physical or biographical features, and personality traits.

The second main aspect to be modelled is that of social relations. In fanfiction, romantic or sexual pairings are expressed through “Character/Character”-relationships (so-called “slashing”). This category is of special interest, because it allows users to investigate the recurrent use of gender-specific features and character pairings, as well as their influence on the popularity of a story both within a specific fandom and across fandoms in comparison with the canonical relationships [13].

Additional metadata are modelled with *DCMI Metadata Terms*⁴, e.g. *dct:title* or *dct:creator*. At the moment, domain-specific terms, such as the number of kudos – a form of user-interaction similar to ‘likes’ – are modelled using project-specific categories, e.g. *golem:numberOfKudos*. This will be updated in the future once the KG makes full use of the GOLEM ontology. For copyright reasons, the KG (as well as the triple store) does not provide access to the fulltexts of the fanfiction stories. This is consistent with the goal of a large scale study of fiction using derived features [9, 5, 14].

³The interface is available at <http://search.golemlab.eu:3006/>.

⁴<https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>

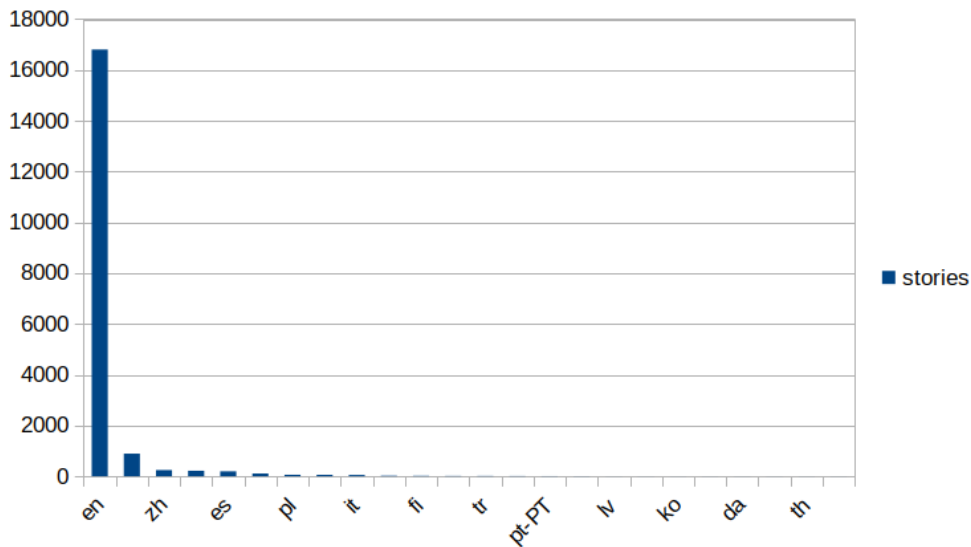


Figure 1: Distribution of stories per language

3. Search Perspectives

SAMPO offers a framework to create different focused perspectives on the data. The queries are predefined. Therefore, end-users do not need to be proficient in SPARQL. Table 1 gives an overview of the current state of the KG, while Figure 1 provides a breakdown of stories by language.

Table 1

Current statistics of the knowledge graph (2024-10-15)

Class	Number of Instances
Stories	18,710
Fandoms	1,698
Languages	27
Fanfiction Characters (total, distinct)	6,637
Potterverse Characters (distinct)	523
Potterverse Character Stoffe (distinct)	253

The GOLEM-UI search perspectives include four distinct views into the KG:

1. Metadata view: Overview of the stories' metadata,
2. Fandoms view: Statistics on fandoms and associated stories,
3. Characters view: A closer look into characters and their various romantic pairings, for now with a special focus on the *Harry Potter* fandom (*Potterverse*),
4. Literary quality view: Statistics on readability and literary quality of the stories.

3.1. Metadata View

The metadata view gives an overview over the common properties of the stories, such as the word count, the language of the work, or the associated fandom (Figure 3). Facets on the left hand side of the search perspective allow users to filter data by language and fandom.

3.2. Fandoms View

The fandoms view gives statistical overview over the representation of specific narrative universes, e.g. how many stories are associated with the respective fandom (Figure 4). Fandoms are represented by their *skos:prefLabel*, which in some cases can be a translation of the title. Sub-fandoms are connected to their more general fandom (e.g. *Harry Potter and the Philosopher's Stone* to the fandom *Harry Potter - J.K. Rowling*) by *skos:broader/skos:narrower* relations.

3.3. Characters View

The characters view shows a character's aliases (e.g. *Voldemort*, *Lord Voldemort*, *Tom Riddle* and so on), keywords that are associated to a certain character (i.e. all the possible variations a character can have), and the number of stories a specific version of the character appears in. In order to link between them all the versions of a character, we introduce the class *G0_Character-Stuff* a *crm:E28_Conceptual_Object*. This represents the general idea of a character, e.g. *Voldemort*. Specific realisations of a character, e.g. *Voldemort* in *Harry Potter and the Half Blood Prince*, are represented as individuals of *G1_Character* a *crm:E89_Propositional_Object*.

The GUI shows characters and their associated synonyms from the *Potterverse* fandom, as well as the number of social relationships in which a character is involved. This statistic gives us an insight into which characters are popular love-interests in a narrative universe or with

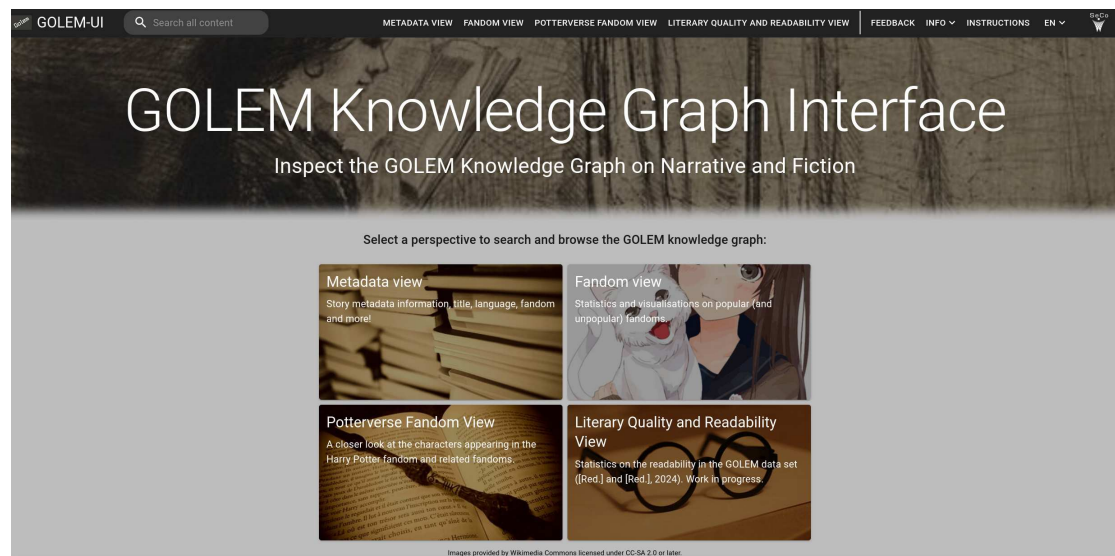


Figure 2: Landing page of the GOLEM-UI giving access to the four individual data perspectives

GOLEM-UI Search all content

METADATA VIEW FANDOM VIEW POTTERVERSE FANDOM VIEW LITERARY QUALITY AND READABILITY VIEW FEEDBACK INFO INSTRUCTIONS EN

Metadata view 18710 Fanfiction stories

Narrow down by:

Language

- en [16804]
- de [890]
- zh [243]
- ru [215]
- 其他 [1504]

Fandom

- Harry Potter - J. K. Rowling [9805]
- My Hero Academia [1395]
- Boku no Hero Academia [1395]
- 僕のヒーローアカデミア [1395]
- BTS [1391]
- Bangtan Boys [1391]
- 방탄소년단 [1391]
- Supernatural [1195]
- Marvel Cinematic Universe [694]
- MCU [694]
- Teen Wolf (TV) [589]
- Yuri!!! on Ice (Anime) [555]

URI	Title	Word Count	Kudos Count	Comment Count	Language	Publication date	Fandom
https://data.golemlab.eu/entity/W00017327	I've been a bad girl, daddy	1551	592	4	en	2018-07-14	One Direction (Band)
https://data.golemlab.eu/entity/W00017328	Jikook AU #2 (ABO)	5012	490	19	en	2018-07-12	BTS
https://data.golemlab.eu/entity/W00017329	sweet honey past your lips	2016	1085	32	en	2018-07-12	Yuri!!! on Ice (Anime)
https://data.golemlab.eu/entity/W00017330	Potential	727	93	4	en	2018-07-12	One Direction (Band)
https://data.golemlab.eu/entity/W00017331	The Care and Feeding of a Pack's Omega	8690	537	31	en	2018-07-11	My Hero Academia
https://data.golemlab.eu/entity/W00017332	Heartache (I Miss You)	5425	750	33	en	2018-07-11	<ul style="list-style-type: none"> Captain America (Movies) Marvel Cinematic Universe
https://data.golemlab.eu/entity/W00017333	The Wolf and The Lamb	8688	279	36	en	2018-07-10	<ul style="list-style-type: none"> Hannibal (TV) Hannibal Lecter Series - All Media Types
https://data.golemlab.eu/entity/W00017334	School of the Chosen	9749	249	57	en	2018-07-10	Yuri!!! on Ice (Anime)
https://data.golemlab.eu/entity/W00017335	I Don't Want To Share	1520	731	16	en	2018-07-09	BTS
https://data.golemlab.eu/entity/W00017336	What Have I Become	4077	52	4	en	2018-07-09	BTS

Figure 3: Story Metadata Perspective

URI	Label	Count
https://data.golemlab.eu/entity/F000000058	Iron Man (Movies)	5
https://data.golemlab.eu/entity/F000000031	<ul style="list-style-type: none"> Prince of Tennis Tennis no Oujisama 奮斗吧少年! 	2
https://data.golemlab.eu/entity/F000000036	Transformers - All Media Types	1
https://data.golemlab.eu/entity/F000000055	Star Trek	7
https://data.golemlab.eu/entity/F000000056	Stargate SG-1	6
https://data.golemlab.eu/entity/F000000057	Star Wars ...	6
https://data.golemlab.eu/entity/F000000068	Star Trek: The Next Generation	4
https://data.golemlab.eu/entity/F000000080	Return of the Seven (1966)	3
https://data.golemlab.eu/entity/F000000069	Star Trek: The Original Series	4
https://data.golemlab.eu/entity/F000000060	Thor (Movies)	5

Figure 4: Fandoms Perspective

whom they interact more frequently. For instance, Figure 5 shows that currently⁵, the character of *Albus Severus Potter* has several alternate names used by authors to refer to him, he is involved in 30 slash (romantic and/or erotic) relationships, and most commonly (17 times) with the character *Scorpius Malfoy*.

⁵Date of the submission: 2024-10-15.

Rows per page 10 1-10 of 254 < < > >				
Name ⓘ	Alternate Names ⓘ	No. /-Relationships ⓘ	Top /-Partner ⓘ	No. Top /-Partner ⓘ
∨ Alastor "Mad-Eye" Moody	Alastor "Mad-Eye" Moody ...	-	-	-
∧ Albus Dumbledore	<ul style="list-style-type: none"> Albus Dumbledore (Mentioned) Albus Dumbledore (minor) Dumbledore Abius Dumbledore Afterlife Albus Dumbledore Alburns Dumbledore Alburn Dumbledore Albus (too many middle names) Dumbledore 	10	Minerva Mc Gonagall	4
∧ Albus Severus Potter	<ul style="list-style-type: none"> Albus Severus Potter Al mentioned Al Potter Alburns Severus Potter Albus Albus Orion Black Albus Poter Albus Potter Albus Potter (mentioned) Albus Potter Malfoy 	30	Scorpius Malfoy	17
∨ Alicia Spinnet	Alicia Spinnet ...	5	Lee Jordan	5

Figure 5: Characters Perspective

3.4. Literary Quality View

Lastly, we calculated scores modelling the literary quality (and readability) of the texts associated to the stories in the knowledge graph.⁶ These measures include: Flesch Reading Ease, Flesch-Kincaid Grade Level, SMOG Readability Formula, Automated Readability Index, New Dale-Chall Readability Formula, for readability; and sentence-length, type-token ratio, and compressibility, for stylistic complexity (adapted from [2]). Figure 7 shows an overview of the available data fields in the literary quality view. In Figure 6, we illustrate how this extracted features can be combined with other information for data exploration, e.g. looking at the association of individual measures of literary quality with reader response. Kudos is the reader's appreciation measure used on AO3 and it seems to be positively correlated with the average word entropy of stories [cfr. 8].

⁶We adopt the term *literary quality* from [1], recognizing that the proposed measures are not all-encompassing to describe literary quality.

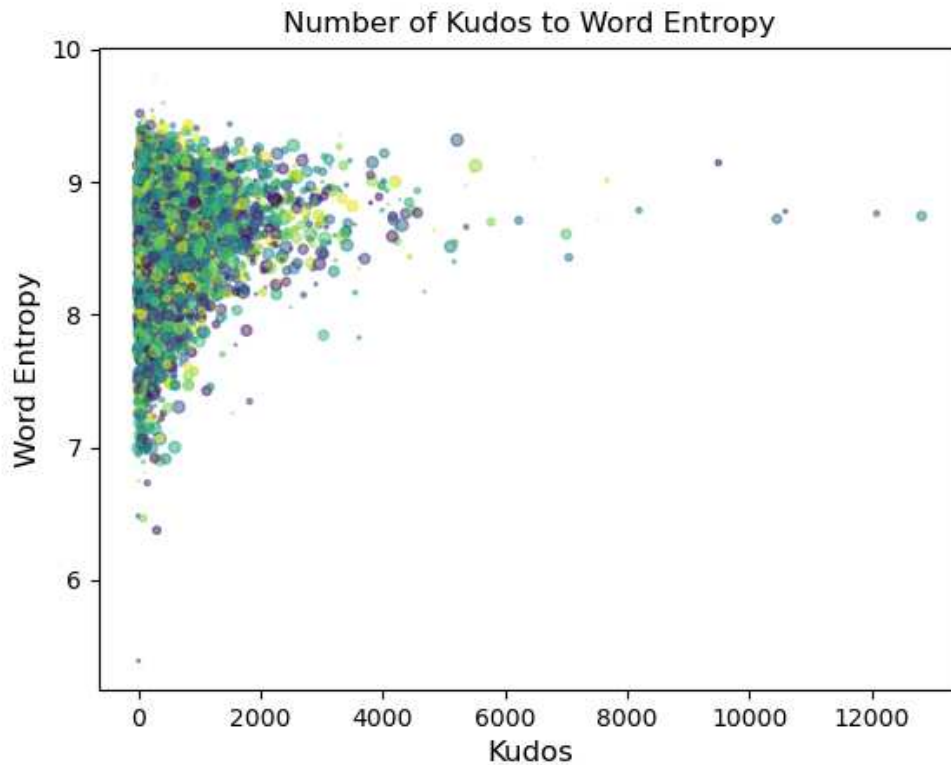


Figure 6: Average word entropy plotted against the number of kudos (dot size represents the number of stories for each value on the x-axis, colors are added only for aesthetic purpose).

4. Querying the Data

While the triple store [10] remains available at <http://graph.golemlab.eu:8890/sparql>, the knowledge graph can be queried via the same endpoint and the designated graph using *WITH GRAPH* <<http://golemlab.eu/graph/>>.

For example, the following query will yield all the stories in the knowledge graph with their respective authors (anonymised) and the romantic category (e.g. F/M-relationships):

```
prefix golem: <http://golemlab.eu/graph/>
```

```
WITH GRAPH <http://golemlab.eu/graph/>
```

```
SELECT * WHERE
```

```
{
  ?s dcterms:title ?title .
  ?s dcterms:creator ?author .
  ?s golem:romanticCategory ?category .
}
```

URI	Title	Kudos	fleschGrade	fleschEase	word entropy	MSTR	SMOG	ARI	daleChallNew	nominal ratio	inverse nominal ratio	average word length
https://data.golemlab.eu/entity/W00016233	Just (Room) Mates	1217	5.2	82.75	8.3326631773	0.7007407407	7.9	6.7	5.7	1.0575342466	0.9455958549	3.5241608263
https://data.golemlab.eu/entity/W00016234	Playing Pretend	330	4.7	83.86	8.870012604	0.7146808511	8.2	6.6	6.07	1.4006309148	0.713963964	3.7624415142
https://data.golemlab.eu/entity/W00016235	Red Carpet	461	5.4	82.04	8.3278901692	0.6804166667	8.5	8.6	6.21	1.2841823056	0.7787056367	3.7642113691
https://data.golemlab.eu/entity/W00016236	I'm knot stuck	2602	5.5	81.93	8.3871842077	0.694375	8.2	7.3	5.78	1.223880597	0.8170731707	3.6728224069
https://data.golemlab.eu/entity/W00016237	Izuku and The Sex Forest	875	5.8	81.02	8.5444274456	0.6797058824	8.4	8.3	6.07	1.5066964286	0.6637037037	3.7278554779
https://data.golemlab.eu/entity/W00016238	Found	328	3.8	91.82	7.8205957348	0.6763636364	5.9	4.8	5.28	1.2694610778	0.7877358491	3.537366548
https://data.golemlab.eu/entity/W00016239	Make a Deal With the Bad Wolf	61	3	93.74	8.1728388996	0.695	6.2	5.2	5.57	1.6035502959	0.6236162362	3.4889349931
https://data.golemlab.eu/entity/W00016240	come a little bit closer	187	4	91.31	8.7616674423	0.6774418605	6.5	7.1	5.34	0.9491039427	1.0536253776	3.5546470114
https://data.golemlab.eu/entity/W00016241	Movie Theater Shenanigans	1047	5.3	82.34	7.8731156248	0.69	8.1	7.1	6.15	1.7011494253	0.5878378378	3.6686434396
https://data.golemlab.eu/entity/W00016242	The Runaway Pack	87	6.9	78.18	8.5239913176	0.7285365854	8.1	8.7	5.74	1.5229202037	0.6566332219	4.0051107325

Figure 7: Literary Quality Perspective

The result of this query can be found at this link.

5. Discussion

The GOLEM-UI is an intuitive interface that allows different insights into the GOLEM knowledge graph. Four different search perspectives allow users who are not familiar with SPARQL queries to gain insights into the data modelled from the domain of fanfiction stories. Four predefined perspective are currently available: Metadata, Fandoms, Characters, and Literary Quality. This interface, together with the possibility of exporting the results of queries in CSV format will allow researchers to easily create corpora that they can use for their analyses of the online production and reception of fiction.

The presented user interface is one step towards presenting the full-fledged GOLEM ontology, which will include many more features than those presented here. For instance, it will allow modeling events and event chains according to different narrative theories, e.g. the hylistic approach [17, 16], more character relationships [15], and character traits.

Acknowledgements

The authors would like to thank Heikki Rantala for his helpful comments and support in creating the GOLEM-UI. This work is part of the *Graphs and Ontologies for Literary Evolution Models (GOLEM)* project funded by the European Commission.

References

- [1] Y. Bizzoni, I. M. Lassen, T. Peura, M. R. Thomsen, and K. Nielbo. “Predicting Literary Quality How Perspectivist Should We Be?” In: *Proceedings of the 1st Workshop on Perspectivist Approaches to NLP LREC2022*. Ed. by G. Abercrombie, V. Basile, S. Tonelli, V. Rieser, and A. Uma. Marseille, France: European Language Resources Association, 2022, pp. 20–25. URL: <https://aclanthology.org/2022.nlperspectives-1.3>.
- [2] P. Feldkamp, Y. Bizzoni, I. M. S. Lassen, M. Rosendahl Thomsen, and K. Nielbo. “Readability and Complexity: Diachronic Evolution of Literary Language Across 9000 Novels”. In: *Proceedings of the Joint 3rd International Conference on Natural Language Processing for Digital Humanities and 8th International Workshop on Computational Linguistics for Uralic Languages*. Ed. by M. Härmäläinen, E. Öhman, F. Pirinen, K. Alnajjar, S. Miyagawa, Y. Bizzoni, N. Partanen, and J. Rueter. Tokyo, Japan: Association for Computational Linguistics, 2023, pp. 235–247. URL: <https://aclanthology.org/2023.nlp4dh-1.27>.
- [3] C. Fiesler, S. Morrison, and A. S. Bruckman. “An Archive of Their Own: A Case Study of Feminist HCI and Values in Design”. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. Chi ’16. San Jose, California, USA: Association for Computing Machinery, 2016, pp. 2574–2585. DOI: 10.1145/2858036.2858409.
- [4] F. Fischer, I. Börner, M. Göbel, A. Hechtel, C. Kittel, C. Milling, and P. Trilcke. “Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama”. In: *Proceedings of DH2019: “Complexities”, Utrecht, July 9–12, 2019*. Utrecht University, 2019, pp. 1–6. DOI: 10.5281/zenodo.4284002.
- [5] Htrc. *HTRC Derived Datasets - Documentation - HTRC Docs*. 2023. URL: <https://wiki.htrc.illinois.edu/display/COM/HTRC+Derived+Datasets>.
- [6] E. Hyvönen. “Digital Humanities on the Semantic Web: Sampo Model and Portal Series”. In: *Semantic Web – Interoperability, Usability, Applicability* 14.4 (2023), pp. 729–744. DOI: 10.3233/sw-223034.
- [7] E. Ikkala, E. Hyvönen, H. Rantala, and M. Koho. “Sampo-UI: A Full Stack JavaScript Framework for Developing Semantic Portal User Interfaces”. In: *Semantic Web – Interoperability, Usability, Applicability* 13.1 (2022), pp. 69–84. DOI: 10.3233/sw-210428.
- [8] M. Jacobsen, Y. Bizzoni, P. Feldkamp, and K. Nielbo. “Patterns of Quality: Comparing Reader Reception Across Fanfiction and Published Literature”. In: *Proceedings of the Computational Humanities Research 2024*. 2024, pp. X–x.
- [9] Oecd. *Derived data element*. 2005. URL: <https://stats.oecd.org/glossary/detail.asp?ID=5130>.
- [10] F. Pannach, X. Yang, N. V. Solissa, Z. Yu, A. Van Cranenburgh, M. Van Der Ree, and F. Pianzola. “The GOLEM Triple Store: A Graph-based Representation of Narrative and Fiction”. In: *Joint Proceedings of the ESWC 2024 Workshops and Tutorials, ESWC-JP 2024*. CEUR Workshop Proceedings (CEUR-WS. org). 2024, pp. 1–9.

- [11] F. Pianzola, A. Acerbi, and S. Rebora. “Cultural accumulation and improvement in online fan fiction”. In: *CHR 2020: Workshop on Computational Humanities Research, November 18–20, 2020, Amsterdam, The Netherlands*. Vol. 2723. CEUR Workshop Proceedings, 2020, pp. 2–11. URL: <http://ceur-ws.org/Vol-2723/short8.pdf>.
- [12] P. Riva, M. Žumer, and T. Aalberg. *LRMoo, a high-level model in an object-oriented framework*. <https://repository.ifla.org/handle/20.500.14598/2217>. 2022.
- [13] K. Schneider. *A Study on the Relevance of Gender within the Shipping Phenomenon in the Worlds of Fanfiction. Analysis of Relationship Patterns in Comparison to Canon Books with Digital Humanities Methods*. Master’s thesis. Mainz, 2024.
- [14] C. Schöch, M. Hinzmann, J. Röttgermann, K. Dietz, and A. Klee. “Smart Modelling for Literary History”. In: *International Journal of Humanities and Arts Computing* 16.1 (2022), pp. 78–93. DOI: 10.3366/ijhac.2022.0278.
- [15] X. Yang and F. Pianzola. “Exploring the Evolution of Gender Power Difference through the Omegaverse Trope on AO3 Fanfiction”. In: *Proceedings of the Computational Humanities Research 2024*. 2024, pp. X–x.
- [16] C. Zgoll. “Myths as Polymorphous and Polystratic Erzählstoffe”. In: *Mythische Sphärenwechsel: Methodisch neue Zugänge zu antiken Mythen in Orient und Okzident*. Berlin, Boston: De Gruyter, 2020, pp. 9–82. DOI: 10.1515/9783110652543-002.
- [17] C. Zgoll. *Tractatus mythologicus: Theorie und Methodik zur Erforschung von Mythen als Grundlegung einer allgemeinen, transmedialen und komparatistischen Stoffwissenschaft*. Berlin, Boston: De Gruyter, 2019. DOI: 10.1515/9783110541588.