

The Literary Theme Ontology for Media Annotation and Information Retrieval

Paul SHERIDAN^{a,1}, Mikael ONSJÖ^b and Janna HASTINGS^c

^a*Tupac Bio, Inc., USA*

^b*Independent Researcher, UK*

^c*Department of Clinical, Educational and Health Psychology, University College London, UK*

Abstract. Literary theme identification and interpretation is a focal point of literary studies scholarship. Classical forms of literary scholarship, such as close reading, have flourished with scarcely any need for commonly defined literary themes. However, the rise in popularity of collaborative and algorithmic analyses of literary themes in works of fiction, together with a requirement for computational searching and indexing facilities for large corpora, creates the need for a collection of shared literary themes to ensure common terminology and definitions. To address this need, we here introduce a first draft of the Literary Theme Ontology. Inspired by a traditional framing from literary theory, the ontology comprises literary themes drawn from the authors own analyses, reference books, and online sources. The ontology is available at <https://github.com/theme-ontology/theming> under a Creative Commons Attribution 4.0 International license (CC BY 4.0).

Keywords. ontology, tagset, information retrieval, ontology-based information system, knowledge graph, standardization, literary theme

1. Introduction

Literary themes are the notable topics that structure and give focus to works of fiction, by relating art to life around such themes as “romantic love”, “coping with death”, and “the lust for power”, or to the science fictional and fantastical in themes ranging from “astronomical-scale engineering” projects to “the occult”. They are a natural organising hub for classification and retrieval of works of fiction.

Despite many informal textual collections of literary themes, no comprehensive *ontology* has yet been developed to underpin thematic analysis in digital literary studies. Ontologies are computationally formalised representations of entities from a given domain. By capturing domain knowledge in a standardised computable form, they enable sophisticated applications such as clustering and intelligent information retrieval. Based on best practices such as the use of the Basic Formal Ontology (BFO) [1] as upper level and the Web Ontology Language (OWL2) [2] as encoding formalism, we here introduce, and present a first draft of, the Literary Theme Ontology (LTO).

¹Corresponding Author: Chief Science Officer, Tupac Bio, Inc., 717 Market St., San Francisco, CA, 94103; E-mail: paul.sheridan.stats@gmail.com. Copyright © 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

The LTO has been developed to be a general-purpose literary theme ontology able to serve multiple applications. However, in this initial presentation we focus in the main on science fiction themes, largely derived from a case study of the analysis of the *Star Trek* television series episodes, and the application scenarios of classification and information retrieval. The LTO is available for download at <https://github.com/theme-ontology/theming> under a Creative Commons Attribution 4.0 International license (CC BY 4.0). For user-friendly access to the current developmental version of the ontology along with a searchable database of annotated stories visit the related Theme Ontology project website (<https://themeontology.org>) [3].

2. Materials and Methods

2.1. Literary Themes

It is customary among literary theorists to distinguish between subject and literary theme [4,5]. According to this view, a subject of a literary work is a topic that is explored within that work. A literary theme, on the other hand, is an opinion that a work conveys about one of its subjects. Kelley Griffith illustrates this distinction by noting that “love” is a subject of Shakespeare’s *Sonnet 116* and “love remains constant even when assaulted by tempestuous events or by time” a literary theme [4]. To take another example, in Jules Verne’s *Twenty Thousand Leagues Under the Sea* (1870), “the obsessive pursuit of retribution inevitably leads to wreck and ruin” is arguably a literary theme associated with the subject “vengeance”.

For the purposes of the LTO, a literary theme of a work of fiction is considered to be a topic treated by that work or an opinion conveyed by the work about a topic treated therein. In other words, we use the term “literary theme” to encompass both the subjects and literary themes of classical literary theory. Several practical considerations motivate this choice. In the first place, the above-drawn distinction between literary theme and subject is far from universally observed. To single out one example, C.E. Preston and J.A. Cuddon write of “jealousy” (a subject) being a literary theme of Shakespeare’s *Othello* [6]. Second, in practice, subjects are more susceptible to objective characterization, than the comparatively subjective classical literary themes associated with them. That said, we do wish to accommodate the identification of classical literary themes whenever feasible. Third, subjects sometimes evoke an implicit theme to begin with. For example, the subject “Martian extraterrestrial” evokes the associated theme “the planet Mars is home to sentient beings” without too much stretching of the imagination. From the perspective of classification and information retrieval, little is gained by distinguishing between topics and the opinions a work conveys about them in such cases. In the authors’ judgement, the more profitable classical literary themes to include in the ontology amount to aphorisms, such as “appearances can be deceiving”, “a danger shared can bring people together”, and “be careful of what you wish for” to name but a few.

2.2. Literary Theme Acquisition and Curation

The LTO terms were primarily collected by watching science fiction video productions and recording themes as the authors understood them. We refer readers to [7,8] for a

discussion of the methodology we used to populate the ontology with themes. Table 1 catalogs the lion’s share of works covered in the ontology to date. The heavy focus on *Star Trek* is explained by its being a long-running, culturally significant franchise with a serviceable balance of themes relating to the human condition, society, and futuristic technologies. Although it will also be admitted that the authors are particular to the show. In order to cover the pre-*Star Trek* era, we watched and recorded themes for roughly 500 classic science fiction films, dating from 1895 to 1973, drawn from Wikipedia lists of science fiction films by decade [9]. Note that *Star Trek* first aired in 1966. We performed the same analysis for about 100 selected science fiction films from 1974 through to the present day. On top of that we watched and recorded themes for the *Black Mirror* television series and the animated series *Futurama*, in an effort to capture some of the more modern-day science fiction themes. Lastly, we are at present in the process of synthesizing and semantically structuring existing compiled lists of heterogeneous themes from *The Encyclopedia of Science Fiction* [10], *Dictionary of Literary Themes and Motifs* [11], *Encyclopedia of Themes in Literature* [12], *Encyclopedia of Film Themes, Settings and Series* [13], and scattered online sources [14,15].

Table 1. Science fiction television series and films used to populate LTO with terms.

Story Collection	Time Period	Story Count
Classic science fiction films	1895 – 1973	491
Selected science fiction films	1974 – 2018	105
Star Trek: The Original Series	1966 – 1969	79
Star Trek: The Animated Series	1973 – 1974	22
Star Trek: The Next Generation	1987 – 1994	178
Star Trek: Deep Space Nine	1993 – 1999	177
Babylon 5	1994 – 1998	110
Star Trek: Voyager	1995 – 2001	172
Futurama	1999 – 2013	129
Star Trek: Enterprise	2001 – 2005	99
Black Mirror	2011 – 2019	23

The scope of the LTO is the representation of operationally verifiable themes of some significance that can be expected to arise in multiple stories. The term *story* is used in the present work to refer to any work of fiction, be it a work of written literature, a film, a television drama, comic, video game plot, and so on.

Theme identification is a subjective enterprise. The LTO has been designed with an understanding that some important themes will fall necessarily outside of its scope on account of their ambiguous or controversial nature. Indeed, people may reasonably disagree because people with different perspectives are liable understand stories differently. Instead, the LTO’s focus is squarely on themes which are amenable to being defined in precise terms. Take the timeless theme “the desire for vengeance” (alias “vengeance”) as an example. It is defined as “A character seeks retribution over a perceived injury or wrong.”. The classic science fiction theme “flying car” constitutes another example. It is defined as “There is a hypothetical air vehicle that looks like a car.”. Both themes are relatively straightforward to identify in stories. We minimize this subjective element of story theme identification by insisting on maximally unambiguous definitions.

2.3. *Literary Theme Ontology Construction*

In compiling the LTO from the above mentioned collection of themes, we followed BFO best practices to ensure interoperability with existing and future humanities related ontologies. The ontology was constructed using Protégé version 5.5.0 [16] in the Web Ontology Language (OWL2) [2]. We captured themes as classes in OWL, related using *subClassOf*, and annotated with the annotation property *label* from the RDF Schema. In addition, we used the Open Biological and Biomedical Ontologies (OBO) [17] metadata standard annotation properties *definition*, *alternative term*, and *reference*. The Fast Classification of Terminologies FaCT++ [18] logic reasoner shows that the LTO is formally consistent and contains no cycles.

2.4. *Release and Quality Control*

Current and archival releases of LTO are made available for download at the project’s GitHub repository (<https://github.com/theme-ontology/theming>) under a Creative Commons Attribution 4.0 International license (CC BY 4.0). In the future, we aim to recruit a community of domain experts (i.e. digital literary scholars) to quality check all themes in the ontology.

2.5. *Theme Requests*

Technical support and new theme requests are presently handled at the LTO GitHub repository (<https://github.com/theme-ontology/theming>).

3. Results

3.1. *LTO Overview*

The LTO version 0.3.1 contains 2,875 classes (i.e. literary themes). The upper-level organization is loosely based around a literary theme classification scheme proposed by William Henry Hudson in 1913 [19]. He identifies five sweeping categories, which we quote:

1. the personal experiences of the individual as individual — the things which make up the sum-total of his private life, outer and inner;
2. the experiences of man as man. . . which transcend the limits of the personal lot, and belong to the race as a whole;
3. the relations of the individual with his fellows, or the entire social world, with all its activities and problems;
4. the external world of nature, and our relations with this; and
5. man’s own efforts to create and express under various forms of literature and art.

Figure 1A shows an overview of the LTO upper level themes. The root class “literary thematic entity” has three children: “the human condition thematic entity”, “knowledge and belief thematic entity”, and “speculative fiction thematic entity”. As these class names suggest, theme names in the LTO follow the convention “such and such a theme thematic entity”. However, the “thematic entity” phrasal appendage is dropped in the exposition

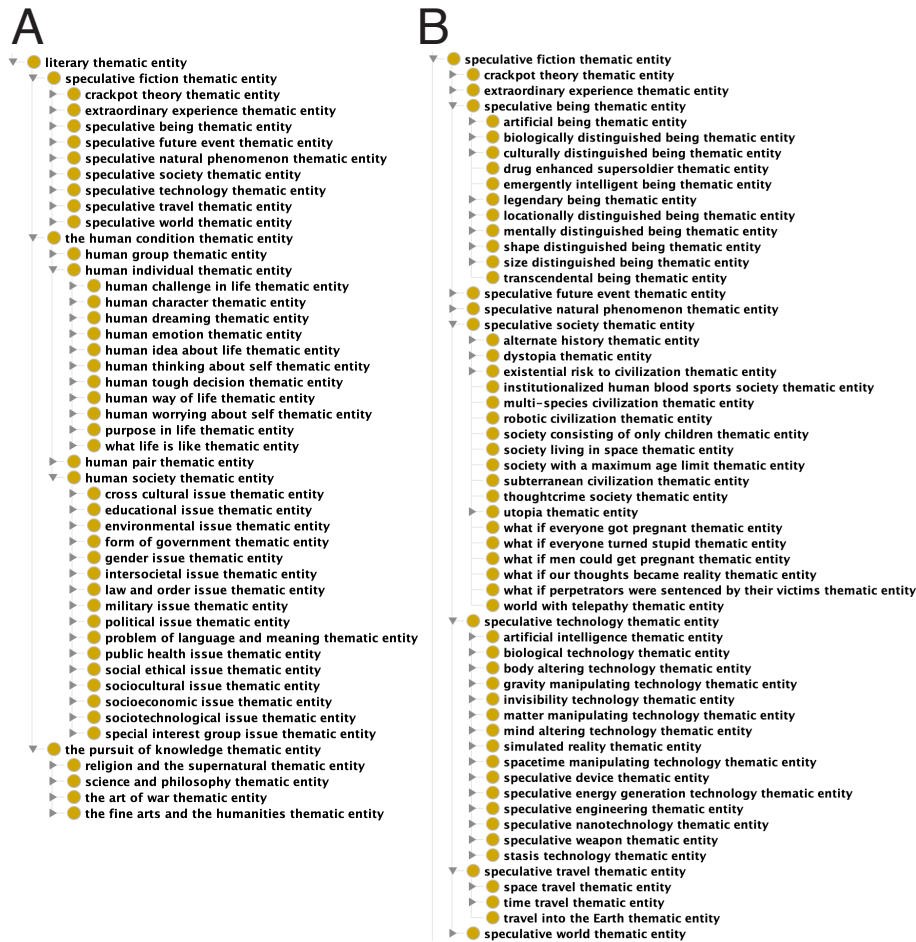


Figure 1. A) LTO bird's eye view. B) LTO with focus on speculative fiction themes.

below for brevity's sake. Each term is accompanied by a definition, and a reference when possible.

The “the human condition” theme covers the first and third of Hudson’s five categories. In keeping with Hudson, the LTO covers human experience at expanding levels of interaction. First, “human individual” covers themes about characteristic key events, and situations which compose the essentials of human experience, such as birth and death, growth, emotionality, aspiration, conflict, and life decisions. Second “human pair” themes examine the respective interactions and attitudes of two people vis-à-vis each other. This theme is specialized in such descendants as “mother and daughter”, “husband and wife”, and “friendship” (i.e. “friend and friend”). Third, “human group” covers themes pertaining to small groups of people in which everyone effectively knows everyone else (e.g., family, friends, and the workplace). And last, “human society” covers themes about how groups of people in the same geographical territory subject to the same political authority and cultural expectations organize themselves and interact with other

such entities. In the future, these levels of human interaction can be extended according to the “levels of analysis” approach used in sociology [20].

The upper level class “the pursuit of knowledge” functions as a generalization of Hudson’s three remaining theme categories. In practice, this is a home for themes pertaining to the expression of a view about how the world operates, and how humans fit in relation to it. Put another way, these are themes about scientific, religious, philosophical, artistic, and humanist views on the nature of reality.

The last upper level LTO class, which is unaccounted for in Hudson’s classification scheme, is “speculative fiction thematic entity”. Figure 1B shows an overview. The class covers subject matter either falling outside of reality as it is presently understood or to human experiences of such an extraordinary nature that no reasonable person would expect to experience them in the course of their lifetime. Themes of these sorts are prevalent in, but are not exclusive to, the genres of science fiction, fantasy, adventure fiction, and horror. The theme “Dyson sphere”, defined as “There is a truly enormous, hollow sphere constructed around a star to capture its entire energy output.”, constitutes a typical example of a theme-worthy entity that does not presently exist in our universe so far as anybody knows. Meanwhile, the theme “what if I were stranded on a deserted island” serves as a stock example of a situation that historical figures, notably Alexander Selkirk [21], actually experienced, but is so extraordinary that one would be utterly astonished to find oneself in the same unenviable predicament.

3.2. Alignment with BFO

To ensure interoperability with other ontologies, including other literature-related ontologies that might be developed in the future, the LTO was designed to fit within the BFO class hierarchy [22]. While BFO has mainly been developed to provide an upper level ontology for *scientific* domains, and has not specifically been developed to categorise fictional entities or themes, we contend that it is nevertheless not incompatible with our aims. BFO provides a broad metaphysical framework that gives structure and definitions for many of the general distinctions between types of entity that cut across domains and subjects, based on different ontological modes of existence. For example, BFO distinguishes between entities that exist as stand-alone entities and those that are dependent on other entities for their existence. An example of the former is an apple, while the red colour of an apple is an example of the latter, since the colour cannot exist without the apple of which it is the colour.

We take as our precedent for considering literary themes as BFO entities, the *information content entities* that are the subject matter of the Information Artifact Ontology (IAO) [23]. Information content entities are *dependent* entities in BFO, in that they cannot exist without some bearer. The bearer of an information content entity is its *concretization*. For example, the content may be concretized by printing it on paper (e.g. a book) or it may be concretized in a file on a computer (e.g. a PDF file). The information content entity is not identical with its concretizations, but is rather multiply realized by them (in BFO terms, it is *generically dependent*) insofar as the same (for example) book *content* can be reproduced in multiple copies.

Information content entities are intended to describe the realm of data and scientific research reports, rather than fiction. They are defined by a relation of *is about*, or representation: information content entities are *about* some entity in the real world. A single

data item in a database may be about an individual entity in the world, but more typically information content entities are about complex aggregates of entities and interrelations, which may be called *portions of reality*.

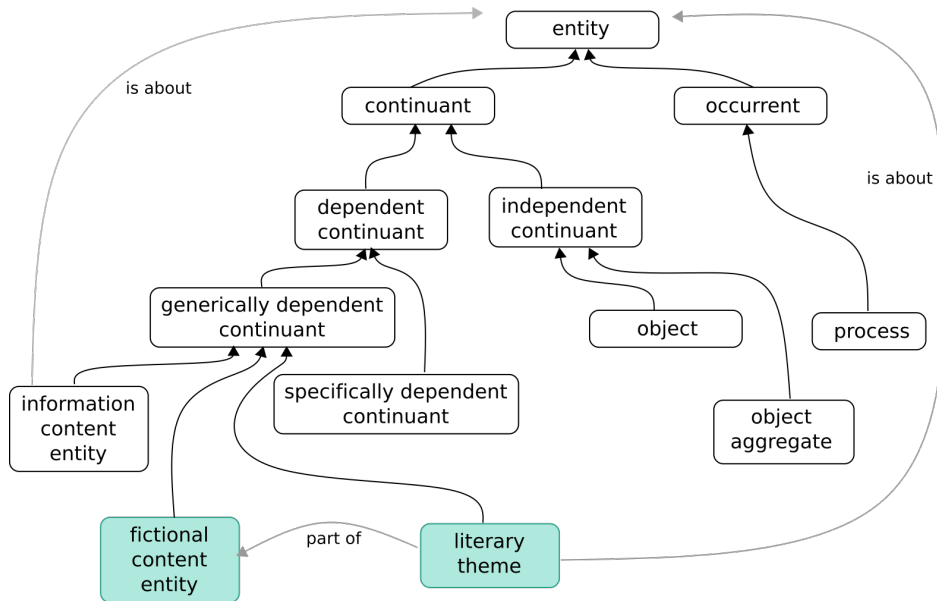


Figure 2. Literature themes in the context of the BFO class hierarchy. Black boxes indicate BFO classes, while coloured boxes indicate proposed classes relevant to literature and literary themes. Black arrows represent hierarchical (is a) relations, while labelled grey arrows represent other types of relation.

Works of literature are not exactly information content entities, since they are typically not intended to carry information about portions of reality in the same way. (Although, of course, in some genres, such as historical fiction, the boundaries may be blurred.) Nevertheless, it should be clear that there are many similarities: works of fiction are also *content* entities, generically dependent on their bearers in which they are concretized, able to be copied and reproduced in different concrete forms. However, they are not *information* content entities, but rather *fictional* content entities. The distinction amounts to a difference in the nature of the type of representation involved: while information content entities are *about* portions of reality quite straightforwardly, works of literature only *appear to be* about portions of reality. Fictional entities have a peculiarly dual ontological aspect: an apparent nature (within the fictional work) and a real nature (as a fictional work), as discussed at length, for example, in [24]. Thus, we categorise works of fiction as *dependent continuants* in BFO, alongside information content entities. Figure 2 illustrates this proposal.

As discussed above, we define literary themes pragmatically as subjects, or opinions or judgements about subjects, expressed in works of literature. In this broad sense, literary themes can be understood as *part of* the broader literary content entity, and are often *about* (in the IAO sense) the world in which the work of literature is situated. For example, a literary theme such as “war”, or “romantic love”, is straightforwardly *about* the class war or romantic love as understood as a universal in the world. However, not

all literary themes are about entities in the world and thus this relation is not an essential one.

3.3. Comparisons with Related Knowledge Curation Systems

The LTO constitutes, to the authors' knowledge, the first comprehensive ontology of literary themes. Traditional scholarship in literary studies occasioned numerous reference works containing lists of themes [11,12,13]. Themes lists also abound on the web. The Encyclopedia of Science Fiction website [10], which features an invaluable list of science fiction themes, is one example of immediate relevance to the LTO. Moreover, there is nowadays a growing number of wikis focused on the curation of terms that overlap in scope with the LTO, including, TV Tropes [25], the Alien Species Wiki [26], and Non-alien Creatures Wiki [27]. The terms on wikis are typically organized into categories, rather than fully fleshed out hierarchies.

Ontologies have been and continue to be used to model information in the humanities. Relevant to literary and film studies is the *Noctua literaria* ontology, which is an ontology of literary characters [28]. Bartalesi and Meghini [29] developed a semantic network-style ontology in RDF for representing knowledge contained in literary texts with emphasis on the works of Dante Alighieri. The ontology features 60 thematic areas (e.g. astronomy, occultism, and theology). More recently, Fabio Ciotti has advanced a more general ontological framework of narrative that includes characters as a special case [30]. OntoMedia is a long-established ontology that includes a hierarchical classification scheme quite relevant to setting [31]. And most recently, Damiano et al. have published The Ontology of Drama [32], dubbed Drammar, with a view toward formally encoding the domain of dramatic media. Lastly, Khan et al. [33] reorganized an existing taxonomy of literary themes and motifs to fit within the framework of the DOLCE-lite upper level ontology. The ontology is implemented in OWL. The underlying literary themes and motifs were compiled by experts for the purpose of annotating Ancient Greek and Latin texts. There is overlap with the LTO in such timeless themes as "love", "death", and "war". But there are interesting differences also. For example, a cursory inspection of their ontology reveals the LTO to be conspicuously lacking in the different types of physical metamorphoses. Khan et al. record "metamorphosis in to animals", "metamorphosis in to flowers", "metamorphosis in to fruit", and "metamorphosis in to trees". By contrast, the LTO contains the "what if I underwent a physical metamorphosis" themes: "what if I grew in size", "what if I shrank in size", and "what if I were gradually turning into stone". In the future, the LTO stands to be improved by incorporating a variety of these themes.

3.4. LTO Applications

The LTO covers a broad range of themes that can be expected to recur in works of literature, films, television and online programs, and other media. Consequently, we envision a number of knowledge curation and information retrieval applications.

Concerning knowledge curation, LTO functions as an ontology-based vocabulary that can be used to tag stories with themes in a manner that exploits its hierarchical organization. Organizing tags into different levels of abstraction helps facilitate tagging by compartmentalizing ambiguity. Consider the subject of aliens, to take one example.

In the classic sci-fi film *The Day the Earth Stood Still* (1951), a nondescript, humanoid alien comes to Earth to deliver an important message to humanity concerning the dangers of atomic weapons. The theme “extraterrestrial being” (a child of “locationally distinguished being”) applies, since there is nothing particularly noteworthy about him beyond the fact that he comes from outer space. However, it is possible to use a more specialized theme when appropriate. For example, “Venusian extraterrestrial” in stories featuring beings from the planet Venus, or something more exotic, like “mysterious maker alien race” should the occasion demand it. Both themes descend from “extraterrestrial being”. The lesser known film *Last Woman on Earth* (1960) presents another illustrative case study. In the film, all animal life on Earth, save for three people who were diving under the sea, appeared to have perished after the oxygen suddenly disappeared from the atmosphere. What caused the de-oxygenation event was never explained. But plants soon thereafter began to replenish the atmosphere with oxygen so that the three lucky survivors could breathe again. We deemed it appropriate to tag the story with the high level theme “existential risk to civilization”, since the process by which Earth’s oxygen disappeared is left open to speculation. A final general example is the “international issue” of the use of “weapons of mass destruction” in war. LTO presently features “biological weapons”, “chemical weapons”, and “nuclear weapons” as “weapons of mass destruction” themes. These can be used to tag stories at different levels of granularity as required.

There is a host of information retrieval applications. A statistical test making use of an earlier version of the LTO was developed to identify over-represented themes in stories of interest relative to a background list of stories [7]. An ontology-enhanced fiction content recommender system based on another earlier version of the LTO has also been proposed [8]. Future applications include LTO ontology-enhanced document clustering [34] methods for clustering stories together into subtypes based on themes they share in common, and time-course analysis methods to identify themes whose usages differ significantly between two group of stories over time (e.g. author gender or nationality). Yet another potential application is to use LTO themes as a gold standard for document classification tasks in conjunction with inter-rater concordance to assess the classification quality. The larger our corpus of thematically annotated stories grows over time, the more powerful these sorts of statistical methods will become for systematically exploring story data, testing existing hypotheses, and generating new hypotheses with the potential to offer novel insights into the real world.

Lastly, automated methods have been developed for extracting themes from texts. Topic modeling techniques [35] as implemented in such software packages as MALLET [36], the Python module gensim [37], topicmodels [38], and the R package **tm** [39] have been used successfully to identify literary themes in large textual corpora [40,41,42,43]. An interesting challenge awaits in adapt the current methods for automatic topic labeling [44,45,46] to the problem of mapping identified topics to themes in the LTO.

4. Discussion

We have presented a first draft of an ontology of literary themes. The LTO, as we call it, is a natural extension of traditional literary theme lists found in reference books [11,12,13] and newly emerging wiki-style, collaborative trope curation websites [25,26,27]. The

LTO is the first comprehensive ontology of literary themes. It is designed to be BFO compliant, implemented in the OWL2 ontology language, and features upper classes based on a traditional literary classification scheme advanced by William Henry Hudson [19]. The LTO version 0.3.1 is comprised of 2,875 literary themes that can be expected to cover stories across a wide range of different genres and media types.

The version of the LTO presented in this paper uses OWL2 as a formalism to express a taxonomy of literary themes. An important reason for developing the LTO as an ontology, instead of just as a taxonomy or vocabulary, is to capitalize on the established role ontologies play in the development of community-wide shared standards. While taxonomies can and often do serve as community standards, ontologies are explicitly designed for that purpose and there are communities, such as the OBO Foundry [17], with a wealth of expertise on best practices for the development of ontologies to serve as standards. On top of that we plan to extend the LTO beyond a taxonomy via the incorporation of additional relations. For example, “what if I shrank in size” (an “extraordinary experience”) and “miniturization technology” (a “speculative technology”) could both be linked to a non-thematic appropriately defined processual entity “shrinking” by the *is-about* relation, to facilitate querying the ontology for all themes related to things becoming smaller in size. We furthermore aspire to add full logical definitions for themes to the LTO whenever possible.

It is our aim to improve upon the LTO to the point where it can function as a serviceable community standard for researchers involved in the digital annotation of story corpora. A most urgent future work, therefore, is for the LTO to be critiqued by digital literary scholars, and revised accordingly. On the subject of annotation, we compiled a modest collection of annotated science fiction media as a proof of concept. Future work will begin with the drafting of a workable protocol for the collaborative tagging of stories with LTO themes. Aligning the LTO with other media ontologies is another future work of some importance. Already existing ontologies, such as OntoMedia [31], are designed with a view toward annotating media objects according to characters, settings, and other qualities.

LTO development is currently managed at the project’s GitHub repository (<https://github.com/theme-ontology/theming>). We encourage anyone interested in contributing to the development of LTO to submit an outside collaborator request to the Theme Ontology GitHub organization. Project collaborator contributions (e.g. proposals for new themes, suggested updates to existing themes, annotated stories) will be subject to approval by a repository manager according to the standard GitHub pull request workflow. In the long term, we aim to transform the Theme Ontology website (<https://themeontology.org>) into a thriving online community platform by following the non-profit model of Wikipedia. At the start we envision that contributors will be required to register with either an academic email address or by referral from an existing member. Wiki policies and guidelines remain to be drafted and a collaborative story annotation protocol fully formulated. But community members will ultimately be able to annotate stories with LTO themes, propose new themes, update or delete old ones, and improve on the ontology structure—all based on a wiki-style consensus model. The more the Theme Ontology database grows with LTO thematically annotated stories, the more the sorts of information retrieval methods outlined above can be harnessed to test hypotheses about human culture and society.

To close on a more general note, the biological and medical sciences enjoy stable platforms for ontology management [17,47]. However, no such system enjoys wide adoption in the humanities, to the authors knowledge. There is a pressing need for the creation of a platform with the goal of organizing interoperable reference ontologies in the digital literary studies domain.

References

- [1] B. Smith and T. Bittner, Basic Formal Ontology for Bioinformatics, *Journal of Information Systems* (2005), 1–16.
- [2] P. Hitzler, M. Krötzsch, B. Parsia, P.F. Patel-Schneider and S. Rudolph (eds), *OWL 2 Web Ontology Language: Primer*, W3C Recommendation, 2009.
- [3] The Theme Ontology contributors, Theme Ontology, 2019, [Online; accessed 30-April-2019].
- [4] K. Griffith, *Writing Essays About Literature*, Cengage Learning, 2010.
- [5] R.C. Murfin and S.M. Ray, *The Bedford glossary of critical and literary terms*, Bedford/St. Martins, 2009.
- [6] C.E. Preston and J.A. Cuddon, *A Dictionary of Literary Terms and Literary Theory*, The Language Library, Wiley, 1998.
- [7] M. Onsjö and P. Sheridan, Theme Enrichment Analysis: A Statistical Test for Identifying Significantly Enriched Themes in a List of Stories with an Application to the Star Trek Television Franchise, *Digital Studies/Le champ numérique* (2019, Forthcoming).
- [8] P. Sheridan, M. Onsjö, C. Becerra, S. Jimenez and G. Dueñas, An Ontology-based Recommender System with an Application to the Star Trek Television Franchise, 2019.
- [9] Wikipedia contributors, Lists of Science Fiction Films — Wikipedia, The Free Encyclopedia, 2019, [Online; accessed 30-April-2019]. https://en.wikipedia.org/wiki/Lists_of_science_fiction_films.
- [10] The Encyclopedia of Science Fiction contributors, The Encyclopedia of Science Fiction, 2019, [Online; accessed 30-April-2019].
- [11] J.C. Seigneuret, *Dictionary of Literary Themes and Motifs*, Dictionary of Literary Themes and Motifs Vol. v. 1, Greenwood Press, 1988.
- [12] J. McClinton-Temple, *Encyclopedia of Themes in Literature*, Encyclopedia of Themes in Literature Vol. v. 1, Facts On File, 2010.
- [13] R.B. Armstrong and M.W. Armstrong, *Encyclopedia of Film Themes, Settings and Series*, McFarland, 2001.
- [14] Wikipedia contributors, List of Science Fiction Themes — Wikipedia, The Free Encyclopedia, 2019, [Online; accessed 30-April-2019]. https://en.wikipedia.org/wiki/List_of_science_fiction_themes.
- [15] LiteraryDevices Editors, LiteraryDevices.net, 2019, [Online; accessed 30-April-2019].
- [16] M.A. Musen, The Protégé Project: A Look Back and a Look Forward, *AI Matters* 1(4) (2015), 4–12.
- [17] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L.J. Goldberg, K. Eilbeck, A. Ireland, C.J. Mungall, N. Leontis, P. Rocca-Serra, A. Ruttenberg, S.-A. Sansone, R.H. Scheuermann, N. Shah, P.L. Whetzel and S. Lewis, The OBO Foundry: Coordinated Evolution of Ontologies to Support Biomedical Data Integration, *Nature Biotechnology* 25(11) (2007), 1251–5.
- [18] D. Tsarkov and I. Horrocks, FaCT++ Description Logic Reasoner: System Description, in: *Automated Reasoning*, U. Furbach and N. Shankar, eds, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 292–297.
- [19] W.H. Hudson, *In Introduction to the Study of Literature*, George G. Harrap & Company, 1913.
- [20] R. Babbie, *The practice of social research*, Practice of Social Research, Thomson/Wadsworth, 2004.
- [21] J. Howell and A. Selkirk, *The Life and Adventures of Alexander Selkirk*, Oliver & Boyd, Tweeddale-Court; and Geo. B. Whittaker, London, 1829.
- [22] R. Arp, B. Smith and A.D. Spear, *Building Ontologies with Basic Formal Ontology*, MIT Press, 2015.
- [23] W. Ceusters, An Information Artifact Ontology Perspective on Data Collections and Associated Representational Artifacts, *Studies in Health Technology and Informatics* (2012), 68–72.

- [24] R. Ingarden, *The Literary Work of Art: An Investigation on the Borderlines of Ontology, Logic, and Theory of Literature*, Northwestern University Press, 1973.
- [25] TV Trope contributors, TV Tropes — The All Devouring Pop-culture Wiki, 2019, [Online; accessed 30-April-2019]. <https://tvtropes.org>.
- [26] Alien Species contributors, Alien Species Wiki, 2019, [Online; accessed 30-April-2019]. https://aliens.fandom.com/wiki/Alien_Species_Wiki.
- [27] Non-alien Creatures Wiki contributors, Non-alien Creatures Wiki, 2019, [Online; accessed 30-April-2019]. https://non-aliencreatures.fandom.com/wiki/Non-alien_Creatures_Wiki.
- [28] A. Zöllner-Weber, Ontologies and Logic Reasoning as Tools in Humanities?, *Digital Humanities Quarterly* **3**(4) (2009).
- [29] V. Bartalesi and C. Meghini, Using an Ontology for Representing the Knowledge on Literary Texts: The Dante Alighieri Case Study, *Semantic Web* **8**(3) (2017), 385–394.
- [30] F. Ciotti, Toward a Formal Ontology for Narrative, *Matlit: Revista do Programa de Doutorado em Materialidades da Literatura* **4**(1) (2016), 29–44.
- [31] M.O. Jewell, OntoMedia: An Ontology for the Representation of Heterogeneous Media, 2005.
- [32] R. Damiano, V. Lombardo and A. Pizzo, The Ontology of Drama, *Applied Ontology* **14**(1) (2019), 79–118.
- [33] F. Khan, S. Arrigoni, F. Boschetti and F. Frontini, Restructuring a Taxonomy of Literary Themes and Motifs for More Efficient Querying, *MATLIT: Materialities of Literature* **4**(2) (2016), 11–27.
- [34] S.A. Fahad and W.M. Yafooz, Review on Semantic Document Clustering, *International Journal of Contemporary Computer Research* **1**(1) (2017), 14–30.
- [35] D.M. Blei, Probabilistic Topic Models, *Communications of the ACM* **55**(4) (2012), 77–84.
- [36] A.K. McCallum, MALLET: A Machine Learning for Language Toolkit, 2002, <http://mallet.cs.umass.edu>.
- [37] R. Řehůřek and P. Sojka, Software Framework for Topic Modelling with Large Corpora, in: *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, ELRA, Valletta, Malta, 2010, pp. 45–50.
- [38] B. Grün and K. Hornik, topicmodels: An R Package for Fitting Topic Models, *Journal of Statistical Software* **40**(13) (2011), 1–30.
- [39] I. Feinerer, K. Hornik and D. Meyer, Text Mining Infrastructure in R, *Journal of Statistical Software* **25**(5) (2008), 1–54.
- [40] M.L. Jockers, *Macroanalysis: Digital Methods and Literary History*, 1st edn, University of Illinois Press, Champaign, IL, USA, 2013.
- [41] M.L. Jockers and D. Mimno, Significant Themes in 19th-century Literature, *Poetics* **41**(6) (2013), 750–769.
- [42] A. Goldstone and T. Underwood, The Quiet Transformations of Literary Studies: What Thirteen Thousand Scholars Could Tell Us, *New Literary History* **45**(3) (2014), 359–384.
- [43] J.L. Boyd-Graber, Y. Hu and D.M. Mimno, Applications of Topic Models, *Foundations and Trends in Information Retrieval* **11**(2–3) (2017), 143–296.
- [44] J.H. Lau, K. Grieser, D. Newman and T. Baldwin, Automatic Labelling of Topic Models, in: *The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, 19-24 June, 2011, Portland, Oregon, USA*, 2011, pp. 1536–1545.
- [45] A.E.C. Basave, Y. He and R. Xu, Automatic Labelling of Topic Models Learned from Twitter by Summarisation, in: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, June 22-27, 2014, Baltimore, MD, USA, Volume 2: Short Papers*, 2014, pp. 618–624.
- [46] S. Bhatia, J.H. Lau and T. Baldwin, An Automatic Approach for Document-level Topic Model Evaluation, in: *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017), Vancouver, Canada, August 3-4, 2017*, 2017, pp. 206–215.
- [47] M. Musen, N. Noy, N. Shah, P. L. Whetzel, C. Chute, M.-A. Story and B. Smith, The National Center for Biomedical Ontology, *Journal of the American Medical Informatics Association : JAMIA* **19** (2011), 190–5.