Representing Literary Characters and their Attributes in an Ontology

Janna HASTINGS [a,1] and Stefan SCHULZ [b]
[a] *Department of Clinical, Educational and Health Psychology, University College London, UK*
[b] *Institute of Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria*

Abstract. The digital humanities is a burgeoning field of research, bringing computational methods to literary investigations. Ontologies are computational structures that contain descriptions of entities and relations in a domain, and as such they form natural hubs for indexing, search and retrieval as well as enable sophisticated automated inferencing applications. Ontologies are widely used in the medical and biological domains, and are beginning to be adopted in other disciplines such as the social sciences. However, there are unique challenges with the adoption of ontologies for the representation of the elements of literary works. Literary meanings are subjective to some extent, and the ontological mode of being of fictional entities is different from that of the entities in the real world that are the subject matter of the sciences. We focus here on elucidating these challenges through the lens of fictional characters (such as, for example, Macbeth) and the social and personal attributes they are described as having (e.g. age and nationality). We develop a detailed representational strategy – a pattern – for representing fictional characters and their attributes in OWL, in applied ontologies.

Keywords. ontologies, literary characters, digital humanities, OWL

1. Introduction

In the digital humanities, complex requirements for indexing, search and retrieval of works of literature combine with new methods of quantitative analysis to necessitate digital representation of literary content. For example, historical works of literature are being digitalised in projects such as Project Gutenberg and the British Library's Historical Collection.

Library science has a long tradition of representing content by means of thesauri, i.e. semantic lexicons, which provide informal relations (e.g. 'broader' or 'narrower') between keywords. In contrast to thesauri, ontologies consist of formal, computable, logically encoded representations of the entities and relationships in a given domain with a rich range of logical constructs allowing more precise definitions. Based on the philosophical discipline of Ontology, the field of Applied Ontology [1] attempts to bridge be-

[1] Corresponding Author, E-mail: j.hastings@ucl.ac.uk.

tween content and computation across multiple applications. Ontologies and ontology-based terminologies and indexing vocabularies are widely used throughout the natural sciences, medical and engineering domains [2], but they have not yet seen adoption to the same extent in the social sciences and humanities. One of the possible reasons for this is that the types of entities and relations the social sciences and humanities deal with are more complex, controversial and open to subjective interpretations.

This is particularly true for the entities that form the subject matter of the study of literature. Literature studies are one of the core elements of the digital humanities, yet ontologies have not yet been widely applied to this field, although some initial efforts are notable (e.g. [3]). Fictional entities are the primary ontological kind that characterises works of literature, particularly fictional characters [4]. Fictional characters and their attributes, broadly speaking, are an important dimension along which it would be very useful to be able to index, cluster and search for works of literature. For example, one might wish to search for works from a particular genre (e.g. science fiction) which feature a leading *female* character *who is in her thirties*. At present, such a search is not possible, but could be supported via ontologies.

However, there are challenges in developing an ontology for fictional characters according to the same methodological practices with which ontologies in the sciences are usually developed. Examples of these best practices are that ontologies should be developed so as to be able to be shared across a wide community of users, to represent as far as possible a consensus in their domains, and to ensure that the entities that are included in such ontologies actually exist – this latter point is the tenet of a methodological approach known as "ontological realism" [5]. However, fictional characters in many cases do not exist (with the exception of some cases of e.g. historical fiction), and yet there is no clear delineation between the attributes of fictional entities and the corresponding social attributes such as age, gender, nationality, and so on, which are entities in the real world.

In this contribution, we attempt to address the question of how best to represent fictional entities and their attributes. We consider it undesirable to create a parallel hierarchy for fictional attributes which mirrors the hierarchy of attributes which real people can have and are the subject of ontologies in e.g. behavioural science (fictional age vs age as an attribute of people), but we also consider it undesirable to accidentally entail bizarre claims that cross over between the fictional and the real worlds.

2. Background: Ontological Foundations

In the widely used ontology language OWL, which is based on description logics, the traditional A-Box / T-Box distinction separates assertions about individuals from axioms about classes, i.e. universally true statements about all members of a class.

In scientific ontologies, individuals are typically not the primary subject matter (although there are of course exceptions), as these ontologies capture general scientific knowledge about types of entities in the world, rather than individual or particular entities. Thus, the class forms the basic representational unit within an ontology, representing a type of thing in a domain of interest. Classes are arranged in a hierarchy via the subsumption, or 'subclass of' (aka 'is-a') relationship, from the general (high in the hierarchy) to the specific (low in the hierarchy). Other relationship types, or object properties,

are used to capture semantic relationships, most of which are not hierarchical. A common example of another relationship type is 'inheres in', connecting qualities, functions, or dispositions with their material bearer.

Formal and philosophical ontology have given to the discipline of applied ontology a richly developed repertoire of foundational distinctions between the general types of entities that are found in the world, detailed in 'upper level' ontologies that are shared across domain ontologies. One such upper level ontology is the Basic Formal Ontology (BFO; [6]), which we will use as the basis of our discussion.

The most fundamental distinction, according to BFO, is that between continuants and occurrents. Occurrents are entities that unfold in time and have temporal parts. Continuants, on the other hand, are those entities that exist in full at all times that they exist, have no temporal parts, and continue to exist self-identically over an extended period of time. Within continuants, BFO further distinguishes between those entities that are independent and those that are dependent. Independent continuants can exist by themselves, while dependent continuants are those sorts of things that need a bearer in order to exist, such as colours and shapes.

A relevant distinction in BFO is that between specific dependence and generic dependence. Specifically dependent entities are dependent entities that depend for their existence on a specific individual, such as a colour that is, for example, the colour of a particular apple. Generically dependent entities (see also [7]), on the other hand, depend for their existence on the existence of at least one instance of the bearer type existing, but not necessarily the same instance, and not necessarily a single specific such instance. A prototypical example of a generically dependent continuant entity is an information content entity, such as a PDF document or a JPG photograph. Information content entities are said to be *concretized* by a bearer, and can often be transferred between different bearers (as, for example, when a computer file is copied to a different device).

Information content entities have another interesting feature aside from their generic dependence: they point to, denote or are *about* other entities in the world aside from themselves. Human beings have the capability to observe and communicate about the world, and this human capability is the primary source of the relation of representation or *aboutness*. Smith and Ceusters [8] propose a hierarchy of representational entities that can be involved in information content entities (ICEs), including non-referring representational units and representational entities in which individual components refer without the overall representational entity referring, by distinguishing between a 'lower' representational level, mapping to entities directly, and a 'higher' representational level, mapping to complex configurations of multiple entities, which they call 'portions of reality'. This approach accommodates the intuitive compositionality of representational units. For example, they write:

> If someone writes on a piece of paper the sentence Barack Obama is President of Russia, then there is an ICE [information content entity] – concretized by this written string and by any copies made thereof – which is generically dependent on the piece of paper and which is about (on the aforementioned lower level) Barack Obama, his being president, and Russia. But this ICE is not about any corresponding configuration, simply because there is no corresponding configuration. It is for this reason that the given sentence, while it is about certain entities in reality, is nonetheless not true of those entities. [8]

This framework forms the starting point for our discussion of ontological representations of fictional characters.

3. Fictional Characters

Fictional characters are those individual persons who are the subjects and actors in works of fiction. Shakespeare's 'Romeo and Juliet', for example, features the fictional characters Romeo and Juliet. The nature of fictional entities including characters has long been the subject of philosophical debate [4]. A fundamental philosophical question is about their *existence*. And if it is accepted that they do exist, another question relates to their *nature*: what kind of thing is a fictional character? The historical debate has featured both fictional realists and anti-realists. An example of a prominent fictional realist is Meinong [9], while Russell [10] is a prominent anti-realist.

For Meinong, fictional objects are just one type of object amongst other types of object, but which have the additional property of not materially existing. On his account, everything that can be conceived of or have properties is an object, but not all such objects exist in the world [11]. Meinongian objects can be seen as intersections or aggregates of their properties. The properties in question – such as Juliet's age – for Meinong, are the same universals as are instantiated in real persons. This can be contrasted with mental representation approaches to fictional characters, which regard the fictional character as being contained entirely within the mind of the author or the reader, in the same way as, say, we might imagine our potential grandchildren, or a human colony on Mars.

Literary characters have additional properties over and above the manifested properties ascribed to them, which relate to the specific literary context in which they appear. For example, they have roles in the context of the plot, such as being a main character or a side character. They have relationships to an author and to a genre. In a sense, works of fiction have much in common with information content entities: they are generically dependent on their bearers, can be copied and – in some sense – they are representational, at least on the 'lower' level highlighted by Smith and Ceusters. Works of fiction have a peculiar and unique dual mode of being: their characters and happenings both appear to exist to the reader, and not to exist, at the same time: The reader is not confused about the real existence of the fictional entity [12]. We can both ascribe attributes to the character (e.g. being 13 years old) from the perspective *within* the story, and from the *outside* perspective (having been written by Shakespeare). But how can these different aspects be reconciled in an ontology?

4. Case study: Representing Literary Characters in OWL

As a running example, we will draw from the tragedy Romeo and Juliet, as an example of a literary work that should be known to most of our readers, but also because it exhibits typical characteristics for works of fiction:

- A real setting (Verona, Italy, 14th century)
- Realistic characters (Romeo, Juliet), i.e. characters that could have existed
- Unrealistic characters (Tybalt's ghost)
- Wrong beliefs (Romeo believes Juliet is dead)

4.1. Fictional Characters as Parallel Entities in a Fiction-Specific Classification Hierarchy

A naïve initial approach to representing fictional characters in an OWL ontology is to assert that fictional characters simply have a completely different parent type to the regular sorts of characters that exist in the world, and correspondingly all of their properties also have a different parent type to the sorts of properties of real-world people. With this approach, fictional characters are represented with the type 'fictional entity', which is a distinct type of entity alongside other types of entity in BFO. Thus, there are fictional continuants (dependent and independent), fictional occurrents and so on.

An example of a classification hierarchy following this approach is:

Entity
  Continuant
  Fictional Continuant
    Fictional Independent Continuant
      instance = Juliet
    Fictional Dependent Continuant
      instance = Juliet's age
  Occurrent
  Fictional Occurrent
    instance = Juliet's life

The fictional division of the ontology would, thus, duplicate the ordinary ontological counterpart. Fictional entities (persons, objects, events etc.) could then be described in the same way as real entities but with a different set of classes from the ontology.

The obvious problem with this approach is the duplication of attributes between the representation of the fictional world and the representation of the real world, such as character age and nationality. For the age attribute, for example, according to the above pattern, there would be a 'fictional age' classified beneath 'fictional dependent continuant,' distinct from 'age,' classified beneath 'dependent continuant'. Moreover, this approach offers no guidance on what to do in cases where a given character bridges between the real world and the fictional world, e.g. in works of historical fiction, or in cases where it is not agreed whether a character really existed or is fictional (e.g. Abraham and other Biblical characters, cf. [13]).

An additional problem is that some of the axioms that hold in the real world could be challenged in the fictional world. For example, the fictional world could permit time travel, or be populated by immaterial agents such as ghosts and deities.

4.2. Fictional Entities as Representational Entities

In this approach, fictional characters and other fictional entities are characterised as a special type of *information content entity*, created by some author, and continuing to exist self-identically as long as there some concretization of the work of literature still exists (e.g. a copy of the book).

This would look as follows in a classification hierarchy:

Entity
  Continuant
    Dependent Continuant
      Generically Dependent Continuant
        Information Content Entity
          Work of Fiction
            instance – Shakespeare's Romeo and Juliet
          Fictional Character
            instance - Juliet

This classification hierarchy could be expanded further to include additional types of fictional entity as needed, for example, fictional occurrents (Juliet's life).

Representing Juliet with the properties that are ascribed to her would then look something like the following:
(classes italics, object properties and individuals bold face)

Juliet Type *'Fictional character'*
'Romeo and Juliet' Type *'Work of Fiction'*
Juliet part_of 'Romeo and Juliet'
'Romeo and Juliet' Type (output_of some (*Writing process*
  and **has_agent** value **'William Shakespeare'**))

This may succinctly reflect the nature of fictional characters, corresponding to the 'view from the outside' of the literary work. However, it does not yet capture any of the attributes which Juliet is represented to have within the fictional world created by the work, such as her age, the focus of our indexing and clustering literature use case.

To achieve this, we need another element to represent Juliet's attributes as described, to make the connection from the fictional world out to the world of attributes that appear elsewhere in the ontology.

4.3. Fictional Character Attributes as Intersections of Properties

As mentioned above, information content entities are generally characterised by being representational, being 'about' some other entity in the world. However, fictional characters do not genuinely represent an entity in the world in the same way that, say, a photograph of a person does. Rather, they merely *appear* to represent an entity in the world. This could be characterised as a different type of representation relation: not **is_about** (which is the paradigmatic 'aboutness' relation of representation for information content entities, [8]), but rather **as_if_about**, a specifically fictional relation of appearing to represent (see also [14]), which links the fictional character – an individual – to the attributes that the fictional character appears to have.

For our ontological representation of Juliet and her attributes, this would be as follows:

Juliet Type **as_if_about** only *X*
where X is defined as:
X equivalentTo (*Human* and *Female* and *13-Year-Old* and
(**spatially_located_in** value **Verona**) and
(**temporally_located_in** value **16thCenturyAD**))

The characteristics that describe Juliet, summarised as *X*, stand for an object in the Meinongean sense, i.e. something one can think about. In an OWL context this would be a logical expression built from symbols that represent classes of real things (universals, according to BFO), but which are assembled in a way that they represent the purely intensional meaning corresponding to the fictional description of Juliet in the context of the work, regardless of whether this compositional class-like expression is instantiated by any real object.

The precise class-like expression is not necessarily specific to the Juliet character, because the same conjunction of properties might be attributed to other fictional characters or indeed to real characters. Of course, the expression *X* can be extended to any level of detail as needed, so as to include as properties all events in which the character participated and all other characters with whom she interacted.

For example, see Table 1 for a candidate longer list of properties that might be ascribed to Juliet.

No.	Property	OWL expression
P1	A 13-years-old girl	*Human* and *Female* and *13-years-old*
P2	Being in Verona	**spatially_located_in** value **Verona**
P3	Being in the 16th century	**temporally_located_in** value **16thCenturyAD**
P4	Marries someone the fictional character "Romeo" is about	**participates_in** some (*Marriage* and **has_participant** some (inv: **as_if_about** value **Romeo**))
P5	Sees someone the fictional character "Tymbalt" is about and who is a ghost (an immaterial human)	**is_agent_in** some (*Seeing* and **has_participant** some ((inv: **as_if_about** value **Tymbalt**) and *Human* and *Immaterial*))
P6	Believes that Romeo is dead	**is_agent_in** some (*Believing* and **is_about** only ((inv: **as_if_about** value **Romeo**) and *DeadPerson*))
P7	Commits suicide with a dagger	**is_agent_in** some (*CommittingSuicide* and **has_participant** some *Dagger*

Table 1. Properties that characterise the intensional content of the character Juliet

Juliet can be characterised by the conjunction of these properties just as she was characterised by the shorter list given before:

Juliet Type *'Fictional Character'* and
as_if_about only (P1 and P2 and P3 and P4 and P5 and P6 and P7)

This expression, a conjunction of properties, corresponds to Meinong's 'so-being'; here interpreted as the properties of the denotation of a fictional character.

By using the OWL constructor 'only' for the association between Juliet and the composite expression, any existential claim ('being' according to Meinong) is avoided. The

OWL logical connective 'only' can be paraphrased as 'not–some–not', i.e. the fictional character Juliet does not represent (is not **as_if_about**) anything that is not classified under the conjunction P1 through P7.

5. Discussion

The OWL language, which is based on a subset of first-order logic, does not permit objects that represent intensional meanings, i.e. the correlate of Meinong's objects – intersections of properties, to be granted the same status as an individual entity. These are class-level objects, but with the peculiarity that they have no instances. And indeed, from a strict ontologically realist stance [5], classes that have no instances are considered irrelevant. However, the situation with works of literature is complex and the question whether the individual class expressions have instances or not is not the one to be asked. Rather, the objective is to do justice to the combinations of attributes that as particular work of literature juxtaposes, appearing to bring into existence individuals who in some cases might just as well have existed, while building all the complexity of a believable and richly detailed literary world around the interactions of characters within the story.

Even within the sciences, the ordinary applicability domain of applied ontologies, there are occasions where it may be necessary to refer to non-existing (or possibly non-existing) entities in the context of specific types of discourse. For example, scientific hypotheses may postulate the existence of an as yet unconfirmed entity, or medical professionals may need to make contingency plans around as yet unperformed procedures, such as 'treatment of myocardial infarction under zero gravity conditions'. Such entities can often be represented as the intersection of properties, which results in composite expressions that make sense and have a clearly defined meaning due to the compositional nature of their individual attributes. Moreover, it may even be the case on certain occasions that a member of such a class emerges at a future point (e.g. it is possible that an astronaut might suffer from a myocardial infarction). A related use case, discussed in [13], addresses the representation of discourse about entities of religious belief, where no consensus can be assumed about their existence in reality.

It remains somewhat of a challenge to connect these complex class expressions to the fictional entities in OWL axioms without entailing any existential claim. The strategy we have proposed above is to use the value restriction constructor ('only' in Manchester Syntax, $\forall$ in traditional description logics notation). Although somewhat inelegant, since it harnesses the 'only' constructor to create an indirect representation of what is normally considered a direct relationship, in a way that is not the typical purpose for the constructor, this approach does circumvent the ordinary entailment associated with the existential relationship modelling pattern. It is the same strategy as was previously proposed in [15], in which the authors discuss entities of dubious or no existence in the field of natural and life sciences, [16], in which the same modelling approach is applied to represent incorrect or uncertain medical diagnoses (e.g. the diagnostic statement of suspected myocardial infarction in a patient *P* does not imply the existence of any instance of myocardial infarction in *P*), and more recently, [13]. In these former works it has been shown that classification works properly using this pattern: the class describing patients who believe they are being chased by pink unicorns was classified correctly beneath the class describing patients who believe they are being chased by unicorns. Applied to our

fictional case study, in practice this would mean that an OWL classifier would classify Juliet as a fictional character who lives in Europe, who sees an immaterial being and who commits suicide using a metallic object.

However, the reasoning capabilities of the OWL language will reach their limits if the fictional expression leads to unsatisfiability of the ontology. For example, if the foundational ontology declares the classes 'person' and 'immaterial' as disjoint, an OWL reasoner will detect that our description of Tymbalt is unsatisfiable, and nothing further can be inferred in an ontology containing such inconsistencies. Future work will explore alternatives to this modelling pattern, for example, whether the use of OWL 'punning' will allow a more sophisticated representational strategy to be developed while preserving reasoning capabilities.

A problematic case could be fictional characters who are blended with historical characters, such as Julius Caesar in the Asterix comic strips. According to our approach, the different Julius Caesar characters as they appear in different literary works by different authors are different literary character individuals. However, our representation – for the time being – would be agnostic about the relationship between the attributes of those fictional character individuals and the attributes of the real historical person. It is reasonable to assume that not all of the historical attributes of historical persons will be accurately represented in fictional depictions.

In our proposed representational pattern, the asserted type of the character is 'fictional character'. There would be no reason not to add additional categories such as 'fictional depiction of historical character' or indeed other types of character such as characters that are described within a system of religious belief and therefore whose nature with respect to being historical or fictional is contested.

However, for the purpose of digitally connecting works of history to works of fiction, one might imagine that adding an appropriate relationship between the fictional and the real individuals would stand to enrich the represented knowledge of both the domains of history and literature studies.

References

[1] Katherine Munn and Barry Smith, editors. *Applied ontology: an introduction*. Walter de Gruyter, 2008.

[2] Janna Hastings. Primer on ontologies. *Methods in molecular biology (Clifton, N.J.)*, 1446:3–13, 2017.

[3] A Zoellner-Weber. *Noctua literaria : a computer-aided approach for the formal description of literary characters using an ontology*. PhD thesis, Bielefeld, 2008. https://pub.uni-bielefeld.de/record/2301795.

[4] Fred Kroon and Alberto Voltolini. Fictional entities. stanford encyclopedia of philosophy. https://plato.stanford.edu/entries/fictional-entities/, 2018.

[5] Barry Smith and Werner Ceusters. Ontological realism: A methodology for the coordinated evolution of scientific ontologies. *Applied Ontology*, 5(3-4):139–188, 2010.

[6] Pierre Grenon and Barry Smith. Snap and span: Towards dynamic spatial ontology. *Spatial Cognition and Computation*, 4:69–104, 2004.

[7] Antony Galton. On generically dependent entities. *Applied Ontology*, 9(2):129–153, 2014.

[8] Barry Smith and Werner Ceusters. Aboutness: Towards foundations for the information artifact ontology. In *Proceedings of the Sixth International Conference on Biomedical Ontology (ICBO)*, CEUR vol. 1515., 2015.

[9] Alexius Meinong. über gegenstandtheorie. In *Untersuchungen zur Gegenstandtheorie und Psychologie*. Leipzig: Barth, 1904.

[10] Bertrand Russell. On denoting. *Mind*, 14(4):473–493, 1905.

[11] Richard Routley. *Exploring Meinong's jungle and beyond: an investigation of noneism and the theory of items*. Canberra: Research School of Social Sciences, Australian National University, philosophy department monograph 3. interim ed. edition, 1980.

[12] Roman Ingarden. *The Literary Work of Art*. Northwestern University Press, 1973.

[13] Stefan Schulz and Ludger Jansen. Towards an ontology of religious and spiritual belief. *FOIS*, pages 253–260., 2018.

[14] Josef Seifert and Barry Smith. The truth about fiction. In *Kunst Und Ontologie: Für Roman Ingarden zum 100. Geburtstag.*, pages 97–118. Amsterdam: Rodopi, 1994.

[15] Stefan Schulz, Mathias Brochhausen, and Robert Hoehndorf. Higgs bosons, mars missions, and unicorn delusions: How to deal with terms of dubious reference in scientific ontologies. In *ICBO*, 2011.

[16] Stefan Schulz, Catalina Martínez-Costa, Daniel Karlsson, Ronald Cornet, Mathias Brochhausen, and Alan L Rector. An ontological analysis of reference in health record statements. In *FOIS*, pages 289–302, 2014.