Children stories Ontologies
M2 TAL
University of Lorraine

Md Zobaer Hossain, Averie Ho Zoen So, Silviya Silwal, H. Andres Gonzalez Gongora

1 Introduction

The goal of this project was to create a children stories ontology that could help researchers analyze and explore large corpora of children stories. The initial step in defining an ontology is to first find examples and conceptualize the domain. A such out first task was to come up with a various examples of children stories that incorporated diverse themes, across different time and culture.

As a first step we discussed stories from different backgrounds, themes, and media:

- Little Mermaid
- Hansel and Gretel
- Goldilocks and the Three Bears
- El Encanto
- Kiki's delivery service
- Mama Built a Little Nest
- Tom and Jerry
- Conifers
- Jungle Book
- Yankee Doodle Went To Town

Some of these examples such as the Jungle Book are cross-cultural because it was written by a British Indian author. Whereas other examples like the Little Mermaid have changed over time to adjust to the sensibilities of a time period. E.g. the little mermaid dies by self-inflicted harm in the original 1837 story, but in the Disney adaptation she survives. Additionally, stories like Kiki's delivery service have received various adaptations and others like Goldilocks and the Three Bears have been interpreted in different ways depending on the culture.

Despite their stark cultural differences, a common trait between them is the *moral* or *lesson* of the story. For example, Conifers (loosely translated from Bangladesh as "Needle Trees") is a story about an older sibling who mistreats his younger brother and is overcome with greed. However, as the plot develops, he learns the importance of being kind to others. Similar lessons are imparted on the reader in these other stories. Facing your fears in the Jungle Book, addressing self-doubt in Kiki's delivery service, finding true happiness in one's real form in the Little Mermaid, and never trust strangers in both Hansel and Gretel as well as Goldilocks and the Three Bears.

Other lessons or takeaways may not be moral, but rather educational as seen in Mama Built a Little Nest where children learn about bird nests.

Note that it is not necessary for a story to have been adapted to another medium to be a story and that some may be excluded to one form of medium such as El Encanto, which is a movie.

Based on the examples explored we developed informal definitions for stories and children stories:

- What is a story?
  - A story is an account or narrative of a sequence of events. Its medium can be lexical, visual or audio.
- What is a children story?
  - A children story is a story that is meant for children. It means that the story has a target audience where certain elements may be filtered out. Stories have a target audience and children's stories would be related to a specific target audience.

Based on the definition, we came up with various concepts that may be important to further elaborate on the definition of stories and children stories. Some of these concepts are: origin, medium, audience, characters, purpose, etc. Since, children stories is a subclass of stories we know any object property (concept) of story is therefore also applied to a children story.

Naturally we had to make one clear distinction between a story and a children story. The basis of this was the concept of children which defines a story's target audience based on age. For the purpose of our ontology, we defined three age ranges: toddlers (0 to 3), child (4 to 12), and teenager (13 to 18). Injecting a target audience means that certain themes from more mature stories are filtered out.

However, it is important to acknowledge that certain themes are not only dependent on a target audience, but also on culture. For instance, Kiki's delivery service lesson of balancing her passions (hobbies) and the need of growing to be more responsible is intrinsically Asian. Alternatively while the themes may be targeted to a younger audience, the previous example titled "Conifers" conveys its message in a more graphic manner. In our ontology, this definition does not that take into account such nuances to facilitate our implementation.

2 Competency Questions

We came up with ten competency questions before building the ontology that we want to be able to answer by interrogating the final ontology. The competency questions were created using the examples we discussed as well as some of the topics a researcher might want to investigate about children's stories.

1. Which stories are published in 19th Century Europe?
2. Which stories are books?
3. Which media are usually published by organizations?
4. Which stories are fictional?
5. Which stories are non-lexical?
6. Which stories have more than one female protagonists?
7. Which stories have animal characters?
8. Which stories have more than one version?
9. Which stories are not suitable for a 5 year-old to read?
10. Which stories belong to the fantasy genre?

Based on the definitions, concepts, and competency questions we made a first draft skeleton ontology of children stories as seen on Figure 1. The skeleton ontology revealed improvements that were necessary in defining our concepts. As such we added more concepts such as genres, publication, purpose, plot, etc.

An example of this was how we had to introduce subclasses of media to find non-lexical stories. In other words, this meant that the "concept" story had to contain the concept "medium" which can correspond to an instance like "book".

Considering our definition of story, the children stories concept must be a subclass of story that is defined by a target audience attribute. Since stories in general may have themes that do not target children, this definition is more suitable while not restricting what a story entails.

3 Surveying existing ontologies

To properly build an ontology capable of answering our competency questions, we explored three main ontologies.

Time ontology[1]**:** From which we did not reuse concepts due to the reasons provided below.

Comic ontology[2]**:** we reuse the Genre concept and their instances by import since it is helpful to competency question 10. We also specialize this concept by adding sub-concepts Fiction and Non-Fiction for competency question 4. We additionally include few instances of genres (eg. Proverbs, Religious) to cover a broader range of stories that have come up during our discussion. We also reuse the Agent concept from this ontology since it includes sub-concepts such as Characters which are relevant to competency question 6 and 7.

Although the descriptions of some properties such as *story* and *interval* seemed promising to address questions 1, 3, 7, and 9, they were either empty or did not answer our questions. Notably the time ontology's time interval was not specific to a time period. This level of sparsity and lack of suitable concepts in this specific ontology let us to a limited usage of its concepts, as well as, relations.

We have also surveyed a few other open-source book ontologies but found that many of the their structures do not align with ours (eg. their concepts were equivalent to our instances). Because of this we defined several concepts in our ontology, their respective properties, and instances.

[1] https://www.w3.org/TR/owl-time/
[2] https://comicmeta.org/

4 Children Stories Ontology

Our ontology was defined using logical relationships to find instances given our queries. In other words our ontology addresses the following logical descriptor $\exists Ra$, examples of what it answers can be seen in table 1 and it adheres to our skeleton from figure 1.

Not only does our ontology yield instances of stories, but it also contains information from publishers. For example, *The Little Mermaid* has different instances belonging to certain time periods where certain information such as publisher will differ. In fact, in the most recent animated iteration from this story, the publisher is Disney whereas in the older version we return the name of the original author.

In order to answer some of our competency questions, we have created new concepts and subconcepts after surveying existing ontologies.

Object Properties of Characters: Nature, Gender To target competency questions 6, we create a concept Gender which have instances Female, Male and Others. To target competency question 7, we create a concept Nature which includes subconcepts Human, Non-Human and Others. In both concepts, while it may be logically accurate to have two disjoint classes (eg. female and male, human and non-human), we realise that there are cases where characters could have unknown or partial Nature and Genders. For example, mermaid is neither human nor non-human.

Object Properties of Story: Medium During our initial discussion, we would like to include some unusual stories which may be considered unusual due to their medium, hence competency questions 2 and 5. For example, stories in the form of songs or non-lexical stories such as comics. We introduce this concept and the related subconcepts (lexical, audio, visual) as a way to comprehensively categories all stories, such that each story should at least be of one of these media.

Object Properties of Publication: Location, Parties To provide some context of which each story is published, we added concepts of Location and Parties. This allows us to answer competency questions 1 and 3. Their respective sub-concepts are intended to allow some common queries that we think researchers are interested in. For example, researchers might be interested in the broader concept of Continents of the publication rather than Country for historical and cultural reasons. For Parties, we also include sub-concepts Author and Organization.

Object Properties of Story: hasOtherVersions This property is designed to address competency question 8. This object property is transitive and symmetric which makes it convenient to add new versions of the same story and adding all other versions by defining just one object property.

Other than object properties, we define some other data properties in order to allow for numerical comparisons across different instances of stories.

Data Property of Publication: hasYear We define year of publication as a data property in order to allow for queries that searches for an arbitrarily defined ranges of time. From our skeleton, we had intended to include pre-defined historical time periods (eg. renaissance) but could not find an open-source existing ontology for that. We find that adapting this goal to a time and location combination to be a flexible alternative.

Data Property of Story: minAge To relate instances of stories to the concept of "Children's stories", we define a minAge data property which captures the minimum age of audience for which a story is suitable. Other than defining what is or isn't a children's story, this would also allow for potential queries which search for stories that are suitable above a specific age (competency question 9).

5 Evaluation and Discussion

Reasoner To assess how robust our reasoner is, we passed queries to answer our 10 competency questions where we leveraged the logic relationships. An example of a much more complex query that answers question 6 is the following:

> *Story and hasCharacter min 2 (Characters cyanand hasGender value Female cyanand hasRole value Protagonist)*

The answer to this query is Frozen because we defined two main characters as object properties from this instance: Elsa and Anna. As observed in both table 1 and the example above, the reasoner is also capable of yielding instances from cardinalities. However, instances of $\forall$ or **ONLY**

Query	Instance
hasMedium **VALUE** song	*Baby Shark*
hasYear **SOME** xsd:integer[$>$ 1801, $<$ 1900]	*Goldilocks and the three bears*
hasCharacter **MIN** 2 (Characters and hasGender value Female and hasRole **VALUE** Protagonist)	*Frozen*
Medium and inverse hasMedium **SOME** (Story and byParties **SOME** Organization)	*video*

Table 1: Logic-relationships in our ontology. Note that the first 3 queries start with *Story and* and that some of the instance results have been omitted

for concepts such as fiction stories are not properly handled as we did not consider exclusivity in our skeleton or in the competency questions.

Conclusion Although our ontology answers our competency questions, adding more universal restrictions to provide more skewed answers is an appropriate next step. Furthermore, expanding the ontology with concepts from others that may be suitable to our definition of "stories" and "children stories" will yield a more robust as well as reusable ontology.

A Appendix

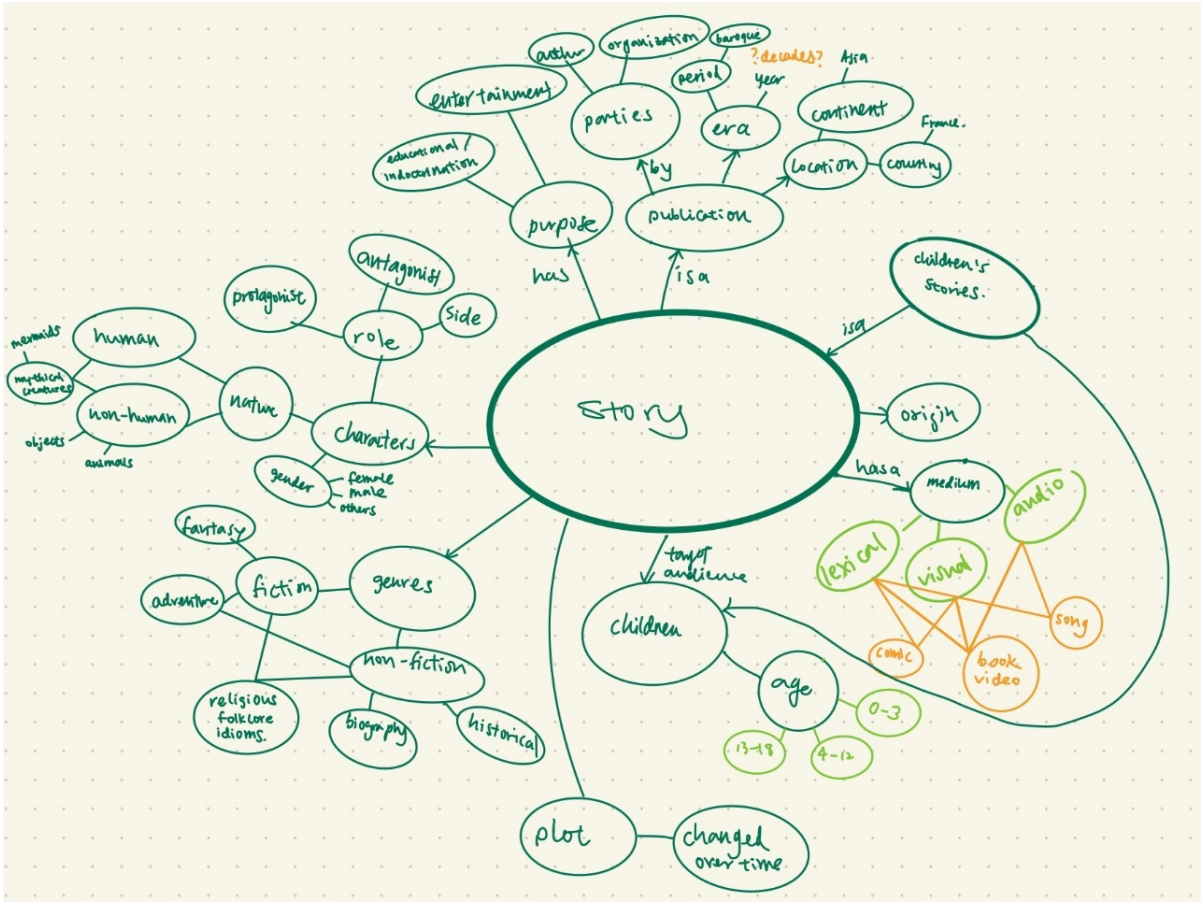

Figure 1: Skeleton ontology of children stories

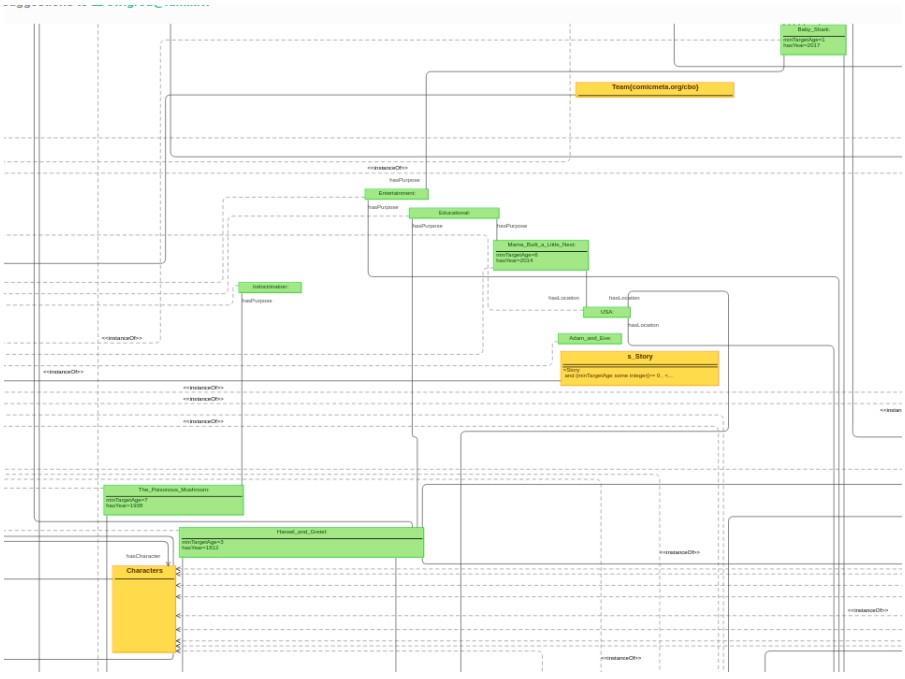

Figure 2: OwlGred Visualization of our Ontology