Modelling Assignment

Introduction

A prototype ontology for a library is built in Protégé to demonstrate how search capabilities could be improved.

Business Context

A local community library wishes to enhance its search capabilities. Customers can currently browse the library shelves, grouped by genre and organised alphabetically by author's surname, or search for books by title or author.

Challenges include:

- Books can represent multiple genres but only appear on one shelf.

- Customers often don't know what they are looking for, so they browse the shelves until they find a book that looks interesting.

- People expect online search capabilities, but the current library search tool is of limited value unless the customer is looking for a specific book or author.

The library will be organised into an ontology so that relationships between the features of books and customers can be mapped and reasoned. The ontology would provide a superior search system, including books in multiple genres and other searchable features.

An AI-powered recommender system could then be built on the ontology to improve the search capabilities.

Approach

Ideally, knowledge would be solicited from experts through interviews (Kendal & Green, 2019). Without a Librarian to interview, a short literature review was undertaken to examine how library users choose fiction books to read.

Readers' choices are influenced by characteristics of books such as the characters, the setting, the type of ending, the physical size of the book, the genre, and how difficult it is to read Ross (2000).

Saarinen & Vakkari (2013) interviewed library users and established reading motives (figure 1) and criteria (figure 2) for readers categorised into three reader types proposed by Lukin (1994).

Reader types	Type of novels	Approach to novels	Motive to read
Escapists	Genres, entertainment literature	Identification to characters and plot	Relaxation, escapism
Esthetes	High standard novels	Language and aspects of narrative	Self-development, new perspectives
Realists	Realistic novels	Credible and truthful description of everyday reality	Learning

Figure 1. Reader types with book characteristics and reading motivations (Saarinen & Vakkari, 2013: 744)

Reader types	Avid readers	Occasional readers
Escapists	Clear plot A subject to identify with Thrilling Captivating Arousing emotions	Captivating Lots of events
Esthetes	Nuanced language Shaking up story Feeds thought New views Deep characters	
Realists	Social Everyday life description Characters to identify with Captivating Arousing emotions	Credible, realistic Not too many characters or too much dialog Historical

Figure 2. Book criteria by reader type (Saarinen & Vakkari, 2013: 745)

Library users are more influenced by external factors such as family, friends, and media when deciding what to read (Ooi, 2008), so a simple search for a specific book will still be required.

Ontology Output

The seven-step approach by Noy & McGuiness (2001) was broadly followed when designing and implementing the ontology:

1. Determine the domain and scope of the ontology

2. Consider reusing existing ontologies

3. Enumerate important terms in the ontology

4. Define the classes and the class hierarchy

5. Define the properties of classes - slots

6. Define the facets of the slots

7. Create instances

The ontology was built in Protégé by iteratively following steps 4 to 7 using a combination (top-down and bottom-up) class hierarchy approach (Noy & McGuiness, 2001). DeBellis (2021) was used as a reference for the practical implementation in Protégé.

The class hierarchy decomposed publications into books and journals, with books decomposed into fiction and nonfiction, and then adults and children. The decision to decompose book → fiction/nonfiction → adults/children's instead of books → adults/children's → fiction/nonfiction was because there were more properties of fiction/nonfiction than adults/children's so class inheritance would be more efficient.

FictionGenre and NonfictionGenre classes were created with subclasses for the main genres **(**Reedsyblog**,** 2020**)** so each book can have multiple genres. This was preferred to having a genre data property because having a separate class is preferred when the attribute is important to the domain (Noy & McGuiness, 2001), and genre is important to search. Similarly, journal-type classes (Springer, N.D.) were added.

Classes for book difficulty and size were added as suggested by Ross (2000) using the following horn clauses:

- difficulty(easyBook, easy)
- difficulty(regularBook, regular)
- difficulty(difficultBook, difficult)
- small(Book) :- hasNumberOfPages(Book, Pages), Pages < 200
- regular(Book) :- hasNumberOfPages(Book, Pages), Pages >= 200, Pages < 500
- large(Book) :- hasNumberOfPages(Book, Pages), Pages >= 500

Book instances were added from Kaggle (N.D.), which was a laborious process.

The class hierarchy with ActionAndAdventureGenre selected (figure 3) shows the books in that genre while figure 4 shows the complete ontology inferred hierarchy
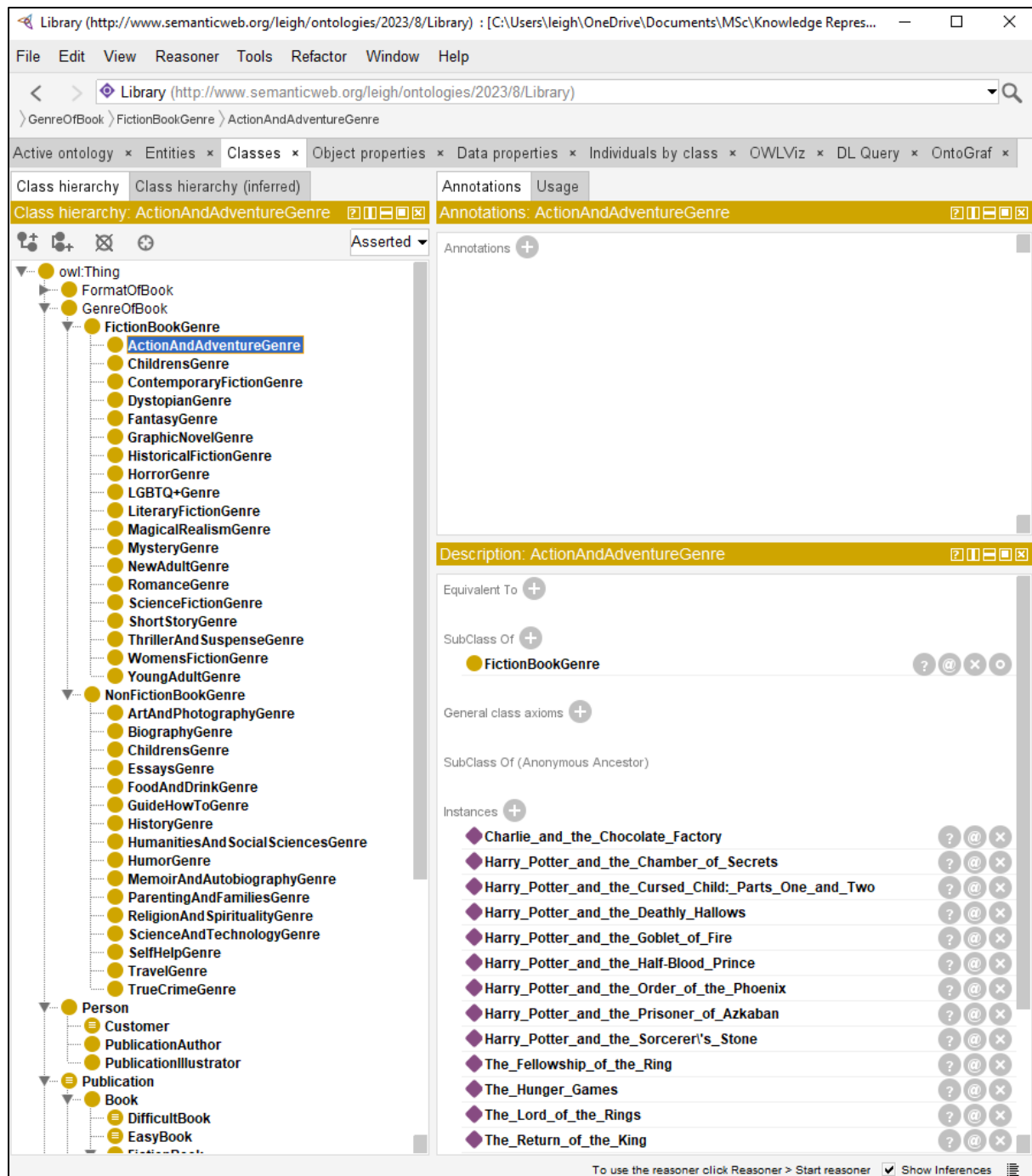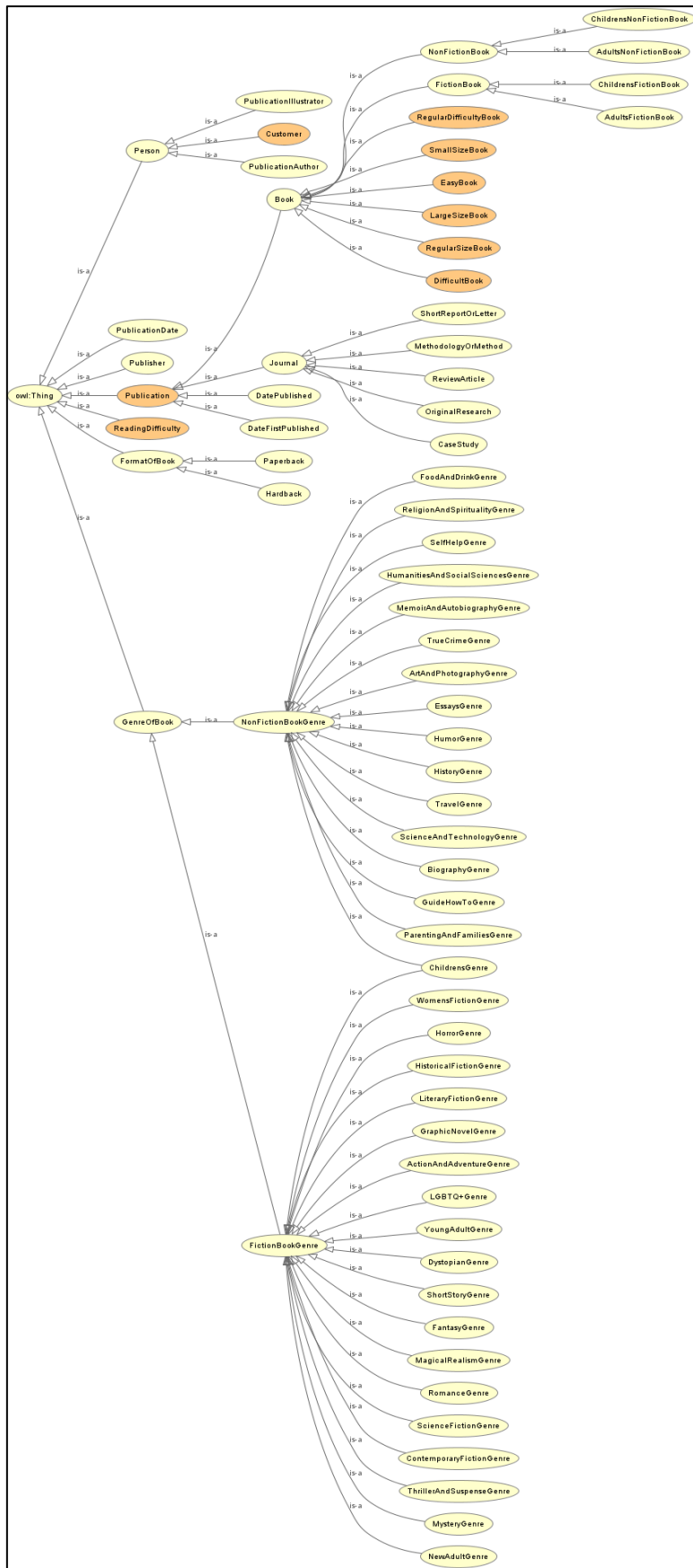

Figure 3. Class hierarchy

Figure 4. Inferred hierarchy

Data properties with hasCharacter selected (figure 5) shows this property has a domain of Books (Journals excluded) and a range of string. This allows characters in books to be searched.
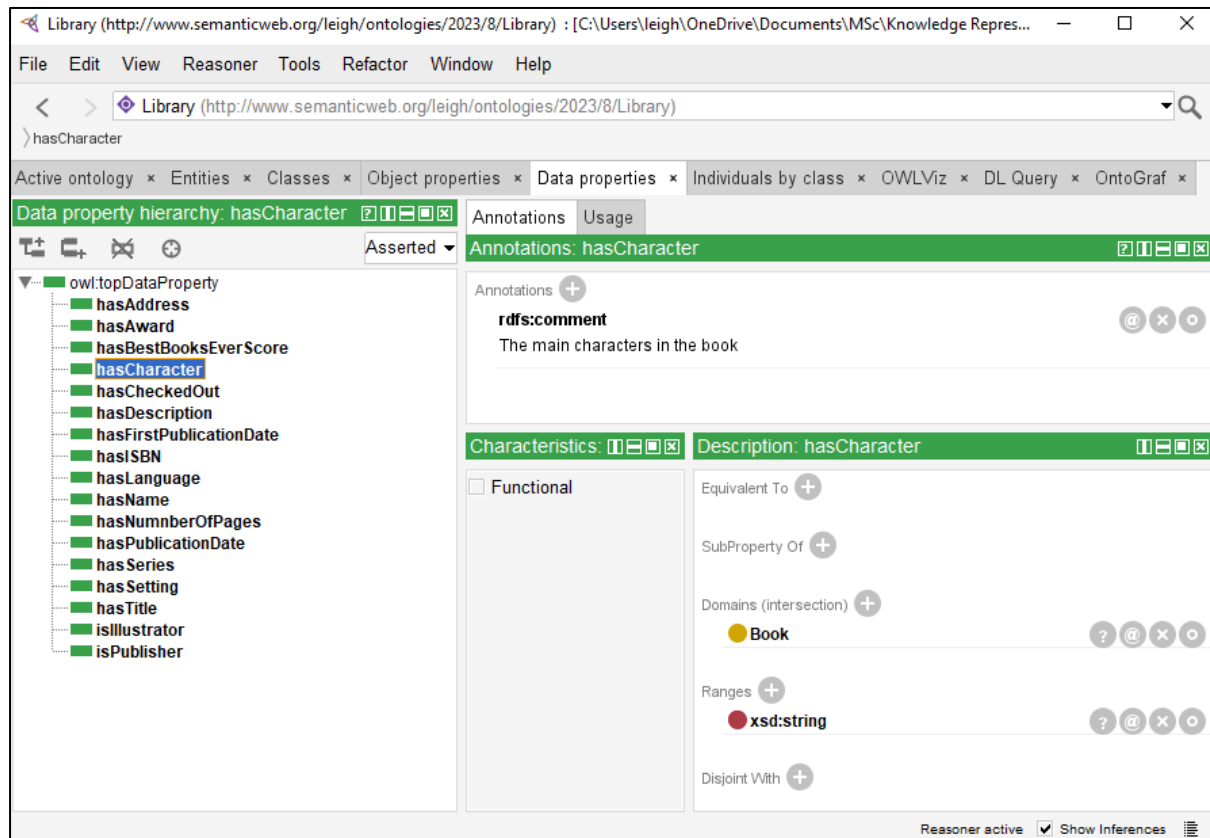

Figure 5. Data properties

Individuals by class with The_Fellowship_Of_The_Ring selected (figure 6) shows some of the property assertions for that book.

Figure 6. Individuals by class 1

Selecting hardback shows only hardback books (figure 7), with The_Way_It_Works selected to show the property assertions.
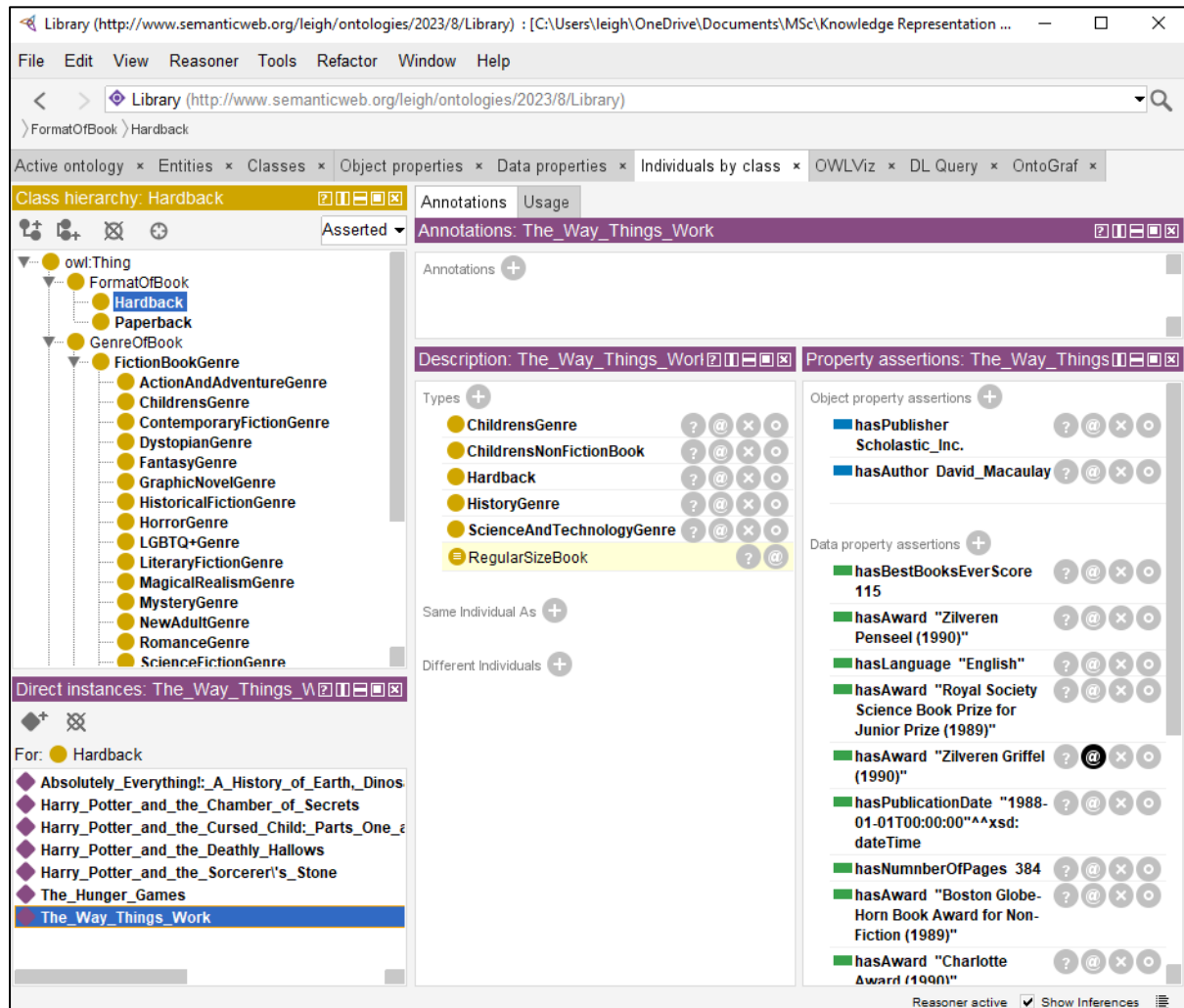

Figure 7. Individuals by class 2

Application to the Business Problem

The ontology can be searched using DL Query. For example, search for books with

the character Ron Weasley (figure 8) with the horn clause:
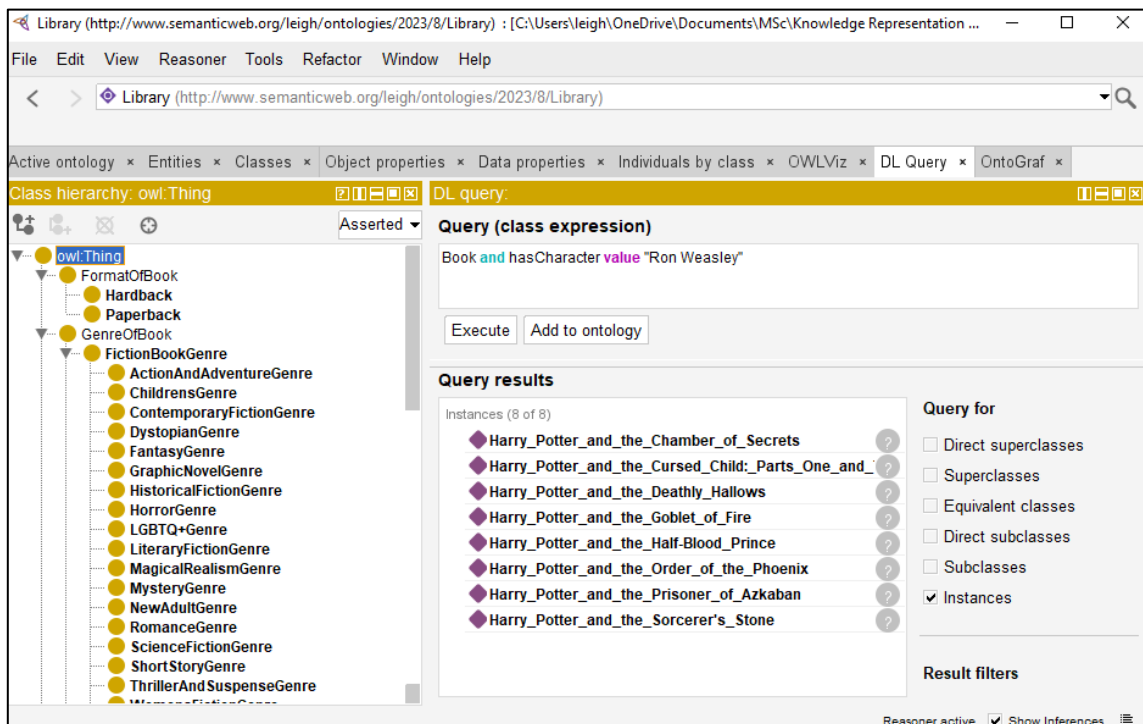
hasCharacter(Book, "Ron Weasley")

Figure 8. Search for books with Ron Weasley

Figure 9 shows a search for AdultFictionBooks with 10 or more awards using the horn clause:

hasAwards(AdultsFictionBook, hasAward), hasAwards >= 10

This could have been searched within a genre, illustrating the power of genre classes.
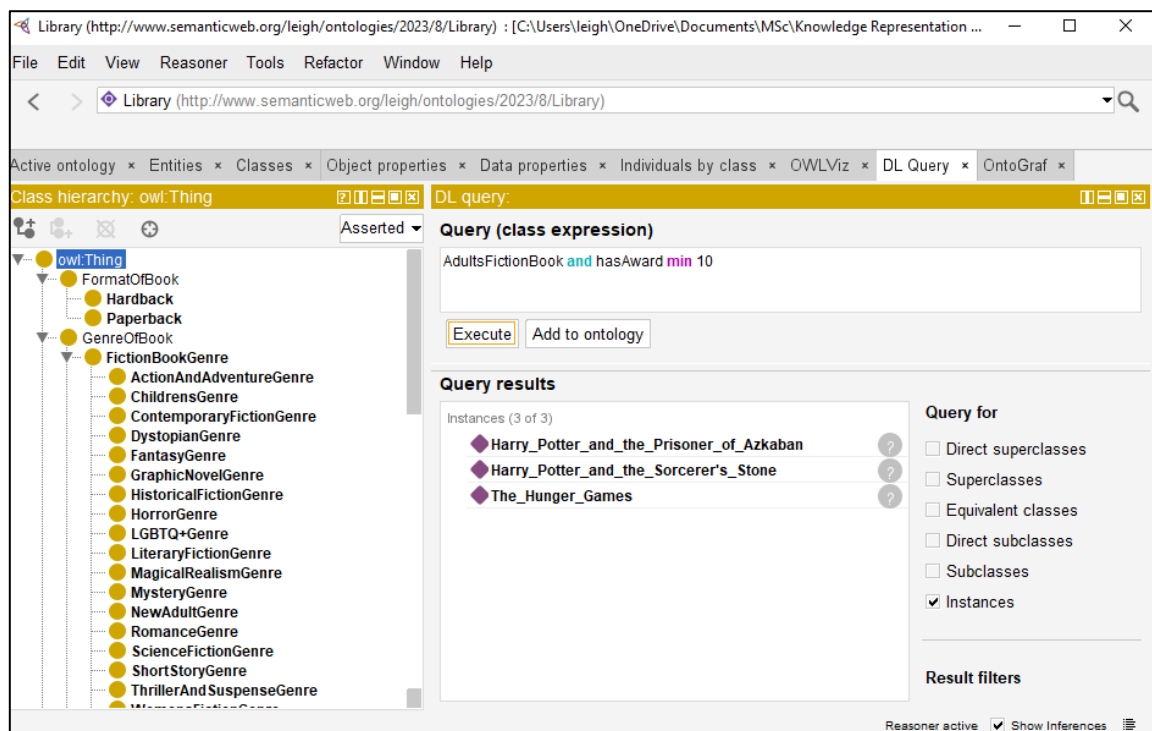

Figure 9. Search for AdultFictionBooks with 10 or more awards

Search for books in the Lord of the Rings series (figure 10) with the horn clause:

hasSeries(Book, "The Lord of the Rings")
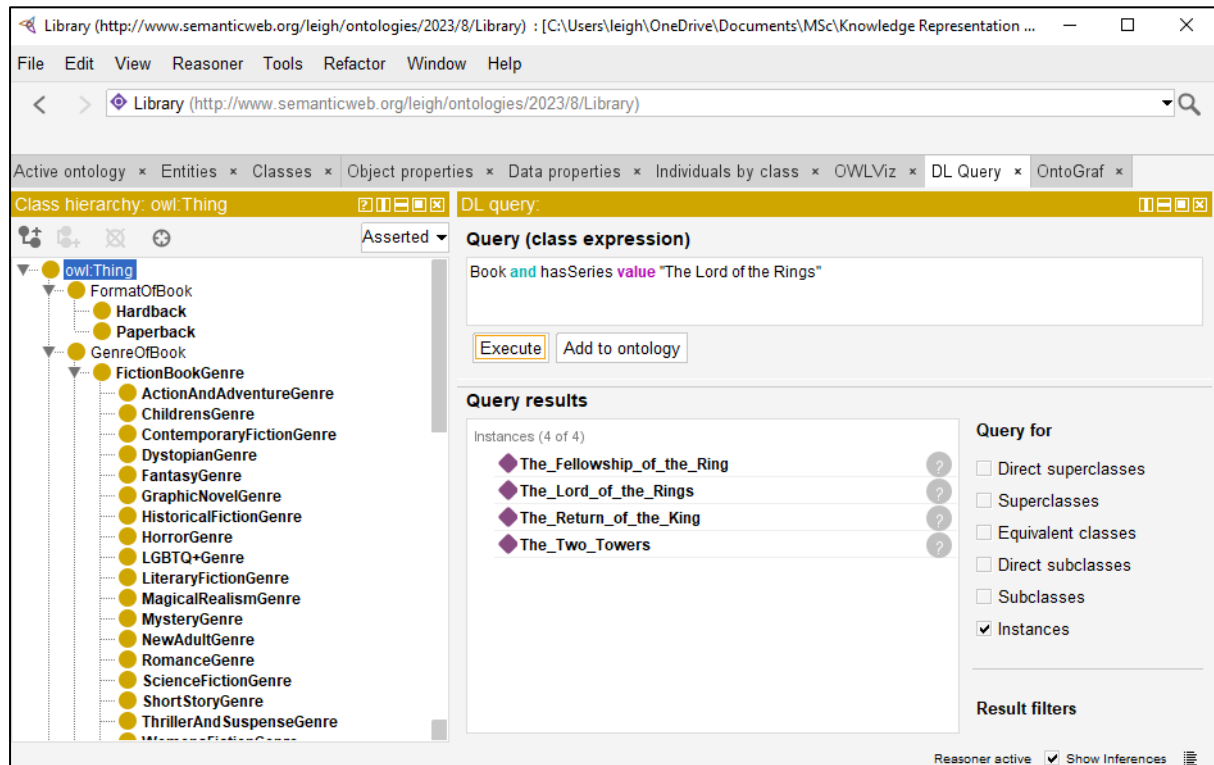
Figure 10. Search for books in The Lord of the Rings series

One of those books is checked out, and figure 11 uses the horn clause

isCheckedOutPublication(Book), hasSeries(Book, "The Lord of the Rings")

to show that the book is TheTwoTowers. Figure 12 uses the horn clause

isCheckedOutBy("The Two Towers", Customer)
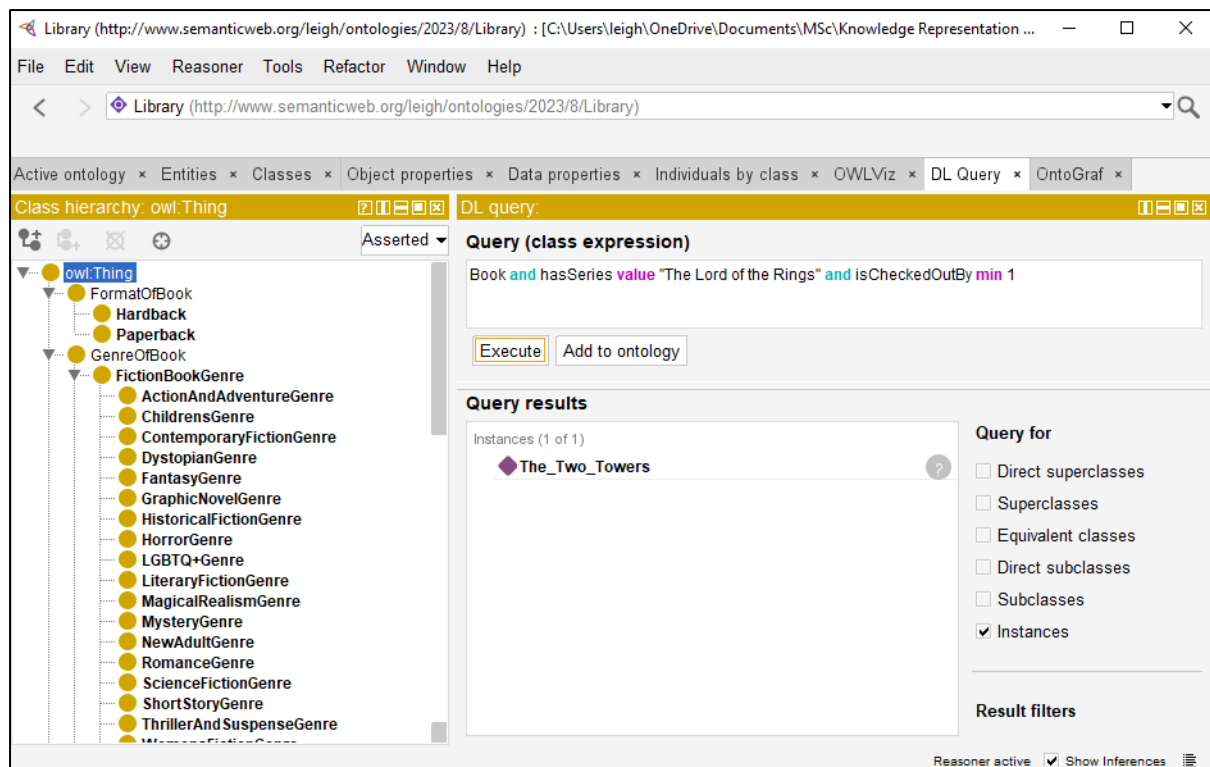
to show that Bob_Dylan has it checked out.
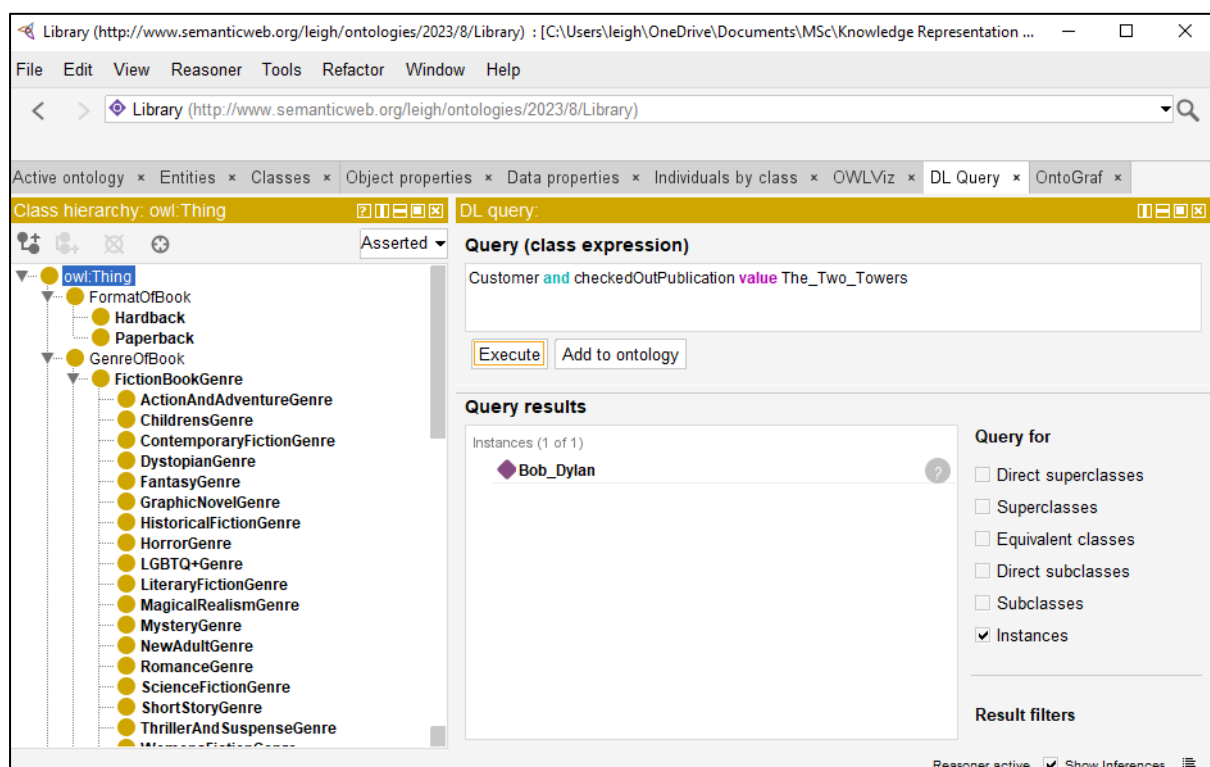
Figure 11. Search for checked-out books

Figure 12. Search for who has a book checked out

Future developments

A recommender system could be built on the ontology to improve and personalise the library customer experience.

A collaborative filtering algorithm could be developed to recommend based on the reading history of similar customers using collaborative filtering and MapReduce (Chen et al, 2015), however collaborative filtering systems suffer from cold-start, and a community library would take a long time to map customer preferences. Ontology-based recommender systems with up-front customer profiles overcome cold-start, with further improvements from machine learning (Obeid et al, 2018), and customer ratings could be added.

The ontology for nonfiction books and journals should be validated and a separate semantic-based recommender for education added, requiring additional features of the customer such as learning style, goals, and subject (Zu et al, 2007).

Conclusions

An ontology of a library, built in Protégé using evidence of fiction search behaviours and loaded with sample data, demonstrated rich search capabilities. However, loading the instances was laborious. A bulk upload tool such as Cellfie should be used when populating the production ontology, which should first be validated with nonfiction search behaviours.

A machine learning recommender system would further improve the customer experience.

Word count: 1,094

References

Chen, L.C., Kuo, P.J. & Liao, I.E. (2015) Ontology-based library recommender system using MapReduce. *Cluster Computing*, 18: 13-121.

DeBellis, M. (2021) *A Practical Guide To Building OWL Ontologies, Using Protégé 5.5 and Plugins.* Available from: https://www.researchgate.net/publication/351037551_A_Practical_Guide_to_Building_OWL_Ontologies_Using_Protege_55_and_Plugins [Accessed 02 October 2023].

Kaggle (N.D.) *Comprehensive Literary Greats Dataset.* Available from: https://www.kaggle.com/datasets/thedevastator/comprehensive-literary-greats-dataset/ [Accessed 30 September 2023].

Kendal, S. & Green, M. (2019) *An Introduction to Knowledge Engineering.* London: Springer.

Lukin, P. (1994) *Where one can find good books.* Master's Theses), University of Tampere.

Noy, N.F. & McGuinness, D.L. (2001) *Ontology Development 101: A Guide to Creating Your First Ontology.* Knowledge Systems Laboratory.

Obeid, C., Lahoud, I., El Khoury, H. & Champin, P.A. (2018) 'Ontology-based recommender system in higher education', *Companion proceedings of the web conference* , Lyon, 23-27 April. Geneva: International World Wide Web Conferences Steering Committee. 1031-1034. https://doi.org/10.1145/3184558.3191533

Ooi, K. (2008) *How adult fiction readers select fiction books in public libraries: a study of information-seeking in context.* Master's dissertation, Victoria University of Wellington. Available from:

https://openaccess.wgtn.ac.nz/articles/thesis/How_Adult_Fiction_Readers_Select_Fiction_Books_in_Public_Libraries_a_Study_of_Information-Seeking_in_Context/16969807 [Accessed 30 September 2023].

Reedsyblog (2020) *The Ultimate List of Book Genres: 35 Popular Genres, Explained.* Available from: https://blog.reedsy.com/book-genres/ [Accessed 30 September 2023].

Ross, C.S. (2000) Making choices: What readers say about choosing books to read for pleasure. The Acquisitions Librarian, 13(25): 5-21.

https://doi.org/10.1300/J101v13n25_02

Saarinen, K. & Vakkari, P. (2013) A sign of a good book: readers' methods of accessing fiction in the public library. *Journal of Documentation*, 69(5): 736-754.

https://doi.org/10.1108/JD-04-2012-0041

Springer (N.D.) Types of journal articles. Available from:

https://www.springer.com/us/authors-editors/authorandreviewertutorials/writing-a-journal-manuscript/types-of-journal-articles/10285504 [Accessed 30 September 2023].

Yu, Z., Nakamura, Y., Jang, S., Kajita, S. & Mase, K. (2007) 'Ontology-based semantic recommendation for context-aware e-learning', *Ubiquitous Intelligence and Computing: 4th International Conference*, Hong Kong, 11-13 July. Heidelberg: Springer. 898-907.