

Watching the Watchmen: Pop Culture and the Semantic Web

Ian Custer, Richard Higgins, Lindsay Williams
School of Informatics and Computing
Indiana University
Bloomington, USA
{icuster, willilin, rshiggin}@indiana.edu

Abstract -

The technology of the semantic web promises the ability to elevate disparate data sources into a central repository, where their data can be integrated and transformed into new forms with new uses. This paper discusses one possible application of semantic web technology to comic books and their cultural references. The aim is to create a use case for integrating data from a specialized, small data source into the larger semantic web.

1. Introduction

Data sets have now begun to proliferate across the internet for different cultural entities and domains: MusicBrainz, SoundCloud, Comic Vine, Giant Bomb, Grand Comics Database, and Wikipedia. Yet, the practice of producing links between these data sets is still in its infancy. In this class, our eyes were opened to the power of linking datasets for these domains, many of which we are personally and professionally invested in.

We chose to focus on a popular cultural domain with a large amount of as of yet unlinked data. Comic books have a huge cultural following with a great deal of crowdsourced data existing in standalone databases. We focused on Watchmen, a particular title that inherently contains many cultural references. These references can easily be linked to real world occurrences and figures, making it ripe for connections to

existing linked open data sets. The Watching the Watchmen project aims to harness the power of semantic integration in order to provide a richer appreciation of the cultural context of the Watchmen graphic novel. This could be used in education, furthering the ability of a reader to understand these cultural links and therefore result in deeper analysis and understanding of a given work. This underscores the validity of graphic novels and popular culture as serious cultural artifacts, and not just entertainment. This paper is outlined as follows. Section 1 provides a general introduction. Section 2, Related Work, will discuss the previously existing crowdsourced data about the Watchmen and its potential as linked open data. Section 3, Watching the Watchmen, will go through the procedures of locating this unlinked data, creating ontologies, and linking it to existing semantic data sources. Section 4, Architecture and Implementation, discusses the collection of data from external sources and creation of the ontology in Protege, and finally the methods of accessing the data in a Fuseki server.

Section 5, Conclusion, discusses the implications for further work in this cultural subdomain.

Division of Labor

- Ian - Assembled ontology, acquired data from Comic Vine and Wikia via API calls, wrote report.

- Lindsay - Wrote queries to run against our dataset.
- Richard - Configured Fuseki server, created backend to host dataset, and front end to execute queries, wrote report.

2. Related Work

There are two rich data sources for the Watchmen and other comics and pop culture artifacts that already exist - Comic Vine, and Wikia. Comic Vine is a website that includes a robust, crowd-sourced platform to create wiki-styled entries for comics, characters, artists, writers, and so on. There are thousands of articles that describe every facet of comic books, and any given publication will have a collection of articles, organized and described with a controlled vocabulary and accessible via a rudimentary API. Data can be output in XML format, and from there can be incorporated into web technologies that have the ability to create new ways of interacting with this data. This data is modeled after and inspired by other sources like Wikipedia, DBpedia, Facebook, LinkedIn, and et cetera. Data can be browsed and searched with user-friendly interface, and users can edit and discuss any article. This results in a curated, socially networked experience that enriches one's appreciation of the source material. There are standalone programs that use the Comic Vine API to create new applications, such as the ability to track your own collection of comics.

Wikia functions in a similar manner, but its subject matter encompass nearly everything. There are individual Wikia subdomains that cover a variety of topics: Watchmen is just one of them. Wikia also has its own API which allows for similar

functionality to Comic Vine's. While this data has great potential, search and discovery is still limited. Users and contributors to Wikia link data as a matter of practice, but only in individual subdomains. Search and discovery is inherently limited to one's own interests. If you don't know that a data source exists, you are unlikely to happen upon it. Without semantic web technologies, these subdomains risk being contained in silos, not providing value outside of a specialized group of users.

3. Comic Vine and Wikia -

Our project builds onto the capabilities of Comic Vine and Wikia. Comic Vine is a comic-book focused website with regularly updated news, reviews of existing titles, an active community and forum, and the aforementioned database. Comic Vine attracts a large group of enthusiasts who contribute to the site in several different ways. This critical mass of users enables the database portion of the site to contain massive amounts of rich, detailed information about a cultural subdomain. The site is used by readers, collectors, and artists and writers in the comic industry to preview upcoming work, read reviews, and interact with other users in the forums. The database is the core of the site's knowledge production, taking a previously undocumented set of cultural artifacts and giving them shape and utility for research and appreciation. The database's contributors function similarly to archivists, documenting and preserving valuable cultural data. They use controlled vocabularies to describe titles, characters, publishers, concepts, issues, adaptations into other forms, creators, and locations and real world figures represented in the medium.

The data being produced by Comic Vine is culturally and semantically rich; this isn't necessarily captured in the database itself. Finally, the editorial staff curates a larger cultural discussion surrounding the medium, which encourages further contributions to the database. However, rich cultural discussion is difficult to capture in a database. In addition, while the data being contributed to the site is highly detailed, it comes from a small, self-selecting group of specialized enthusiasts and may miss out on greater cultural contexts existing in a work. Comic Vine's database entries on the Watchmen, for instance, contain little information about its references to other cultural artifacts.

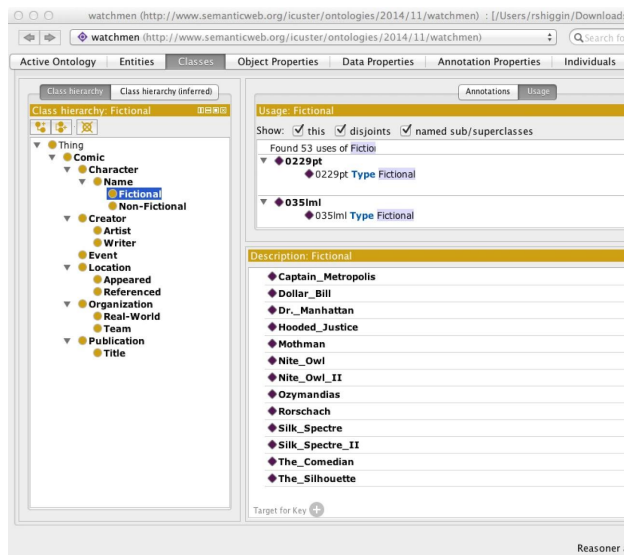


Figure 1 Ontology in Protege

4. Architecture and Implementation -

Without access to a linked open data source for *Watchmen*, or for comic books writ large, we began by creating a dataset and ontology for our project. Our main sources of data are comicvine.com, Grand Comics Database (comics.com), and wikidata.org. These sources provided some support for creating classes and subclasses.

```
rshiggin@ubuntu: /usr/lib/fuseki
%0D%0A&output=text&stylesheet=
11:43:56 INFO [7] Query = PREFIX owl: <http://www.w3.org/2002/07/owl#> PREFIX
rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX xml: <http://www.w3.o
rg/XML/1998/namespace> PREFIX xsd: <http://www.w3.org/2001/XMLSchema#> PREFIX
rdfs: <http://www.w3.org/2000/01/rdf-schema#> SELECT DISTINCT ?Character
?Event WHERE {
?Character <http://www.semanticweb.org/icuster/ontol
ogies/2014/11/comic#isNamed> ?"Rorschach" .
?x <http://www.semantic
eb.org/icuster/ontologies/2014/11/comic#hasAttended> ?Event .
}
11:43:56 INFO [7] exec/select
11:43:56 INFO [7] 200 OK (55 ms)
11:44:12 INFO [8] GET http://localhost:3030/ds/query?query=SELECT+DISTINCT+%3FC
haracter+%3FEvent%0D%0AWHERE%0D%0A%7B%0D%0A+++++++%3FCharacter+%3Chttp%3A%2F%
2Fwww.semanticweb.org%2Ficuster%2Fontologies%2F2014%2F11%2Fcomic%23isNamed%3E+++%
3F%22Rorschach%22+.%0D%0A+++++++%3F%3Chttp%3A%2Fwww.semanticweb.org%2Fic
uster%2Fontologies%2F2014%2F11%2Fcomic%23hasAttended%3E+++%3FEvent+.%0D%0A+++++
++++%09%7D%0D%0A%0D%0A&output=text&stylesheet=
11:44:12 INFO [8] Query = SELECT DISTINCT ?Character ?Event WHERE {
?Character <http://www.semanticweb.org/icuster/ontologies/2014/11/comic#isName
d> ?"Rorschach" .
?x <http://www.semanticweb.org/icuster/ontologies/
2014/11/comic#hasAttended> ?Event .
}
11:44:12 INFO [8] exec/select
11:44:12 INFO [8] 200 OK (18 ms)
```

Figure 2 Fuseki Command Line Interface

The dataset and ontology were produced in Protege.app. The classes for this proof of concept are as follows:

Ontology Classes

- Character
 - Name
 - Fictional
 - Nonfictional
- Creator
 - Artist
 - Writer
- Event
- Location
 - Appeared
 - Referenced
- Organization
 - Real-World
 - Team
- Publication
 - Title

Properties linking these classes included “appearsIn,” “hasAppearance,” “hasAttended,” “hasAuthor.” Once our data entry was complete we ran a built-in reasoner, Hermit 1.3.8, to obtain an ontology. The ontology was output as an

RDF file, which we eventually converted to Terse RDF Triple Language (Turtle) format in order to use with the Jena Project's Fuseki server.

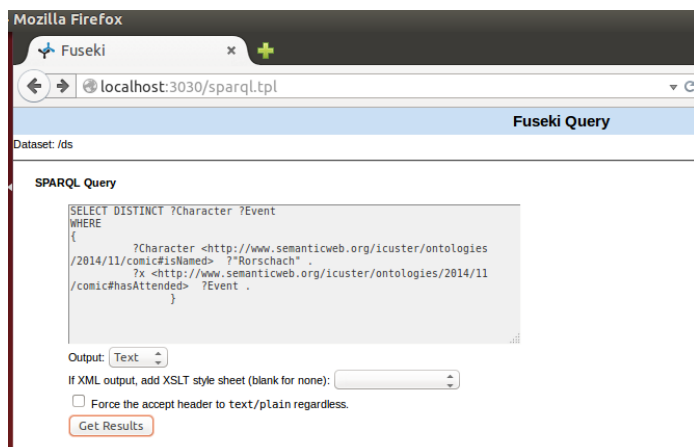


Figure 3 Fuseki Web Interface with Query

Fuseki provided the front-end for SPARQL query service. We set it up on a virtual Linux partition running Ubuntu. Linux eased the work of configuring the server via the command line. In our demo, the front-end web interface accesses the Jena API with a localhost service running on a MacBook Pro. We kept the out-of-the-box fuseki-server interface, which enabled us to load our TTL dataset and provided a text entry box for SPARQL queries. Once the CLASSPATH for Jena was established on our local machine, fuseki-server made the Jena API Java packages fully available. RDF file, which we eventually converted to Terse RDF Triple Language (Turtle) format in order to use with the Jena Project's Fuseki server.

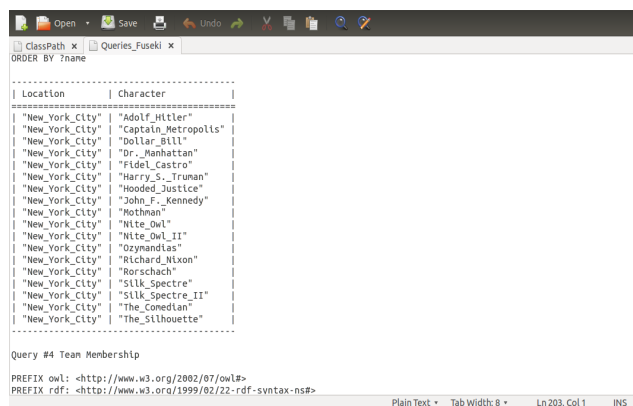


Figure 5 Query Result

5. Conclusion -

This project is a proof of concept for giving some structure to a vast, unstructured data source. By expanding the ontology to encompass the entirety of Comic Vine's data, academic value can be derived and applied to digital library applications, research into comics, and further applications. Much work has been done in applying semantic web technology to scientific data, but the realm of popular culture is still relatively unexplored. This project hopes to start building new bridges between literary analysis and data science.

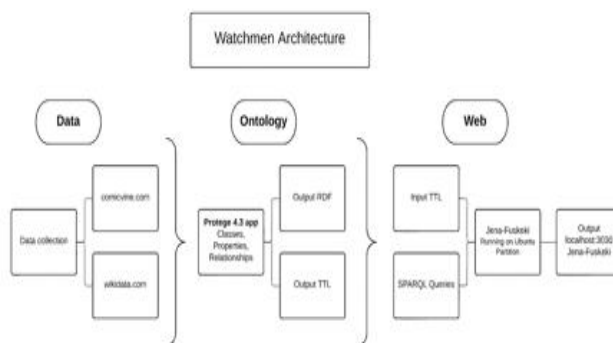


Figure 4 Architecture Flowchart