



Contents lists available at ScienceDirect

Web Semantics: Science, Services and Agents on the World Wide Web

journal homepage: www.elsevier.com/locate/websem

Analyzing user interactions with biomedical ontologies: A visual perspective

Maulik R. Kamdar^{*}, Simon Walk, Tania Tudorache, Mark A. Musen

Stanford Center for Biomedical Informatics Research, Stanford University, USA

ARTICLE INFO

Article history:

Received 20 February 2017
 Received in revised form 20 November 2017
 Accepted 11 December 2017
 Available online xxxx

Keywords:

Data visualization
 Ontology reuse
 Ontology exploration
 User behavior
 Knowledge exploration
 Log analysis

ABSTRACT

Biomedical ontologies are large: Several ontologies in the BioPortal repository contain thousands or even hundreds of thousands of entities. The development and maintenance of such large ontologies is difficult. To support ontology authors and repository developers in their work, it is crucial to improve our understanding of how these ontologies are explored, queried, reused, and used in downstream applications by biomedical researchers. We present an exploratory empirical analysis of user activities in the BioPortal ontology repository by analyzing BioPortal interaction logs across different access modes over several years. We investigate how users of BioPortal query and search for ontologies and their classes, how they explore the ontologies, and how they reuse classes from different ontologies. Additionally, through three real-world scenarios, we not only analyze the usage of ontologies for annotation tasks but also compare it to the browsing and querying behaviors of BioPortal users. For our investigation, we use several different visualization techniques. To inspect large amounts of interaction, reuse, and real-world usage data at a glance, we make use of and extend PolyGonto, a visualization method that has been successfully used to analyze reuse of ontologies in previous work. Our results show that exploration, query, reuse, and actual usage behaviors rarely align, suggesting that different users tend to explore, query and use different parts of an ontology. Finally, we highlight and discuss differences and commonalities among users of BioPortal.

Published by Elsevier B.V.

1. Exploring empirical usage of ontologies

Biomedical researchers have adopted ontologies for use in various tasks, such as knowledge management, data annotation, data integration, data exchange, decision support, and reasoning [1,2]. These ontologies, which often serve as standard vocabularies for a specific domain, are exhaustively large. For example, the Chemical Entities of Biological Interest (ChEBI) ontology [3] contains more than 50,000 and SNOMED CT more than 300,000 entities. To support the biomedical community in finding and using ontologies, the National Center for Biomedical Ontology (NCBO) has developed BioPortal [4],¹ an open online repository of biomedical ontologies and terminologies. BioPortal currently hosts more than 500 biomedical ontologies. Users access BioPortal frequently in their work. In the first half of 2016, more than 215,000 unique IP addresses accessed the BioPortal website and submitted more than 2.52 million requests for various ontologies and services [5].

BioPortal provides two modalities through which researchers can access the content of the ontologies: (1) an interactive website

(referred to in the rest of the paper as WebUI) that enables biomedical researchers to explore ontologies using a class hierarchy visualization (see Fig. 1) [6]; and (2) an application programming interface (API) that allows researchers to programmatically query the repository for specific ontologies and entities, which allows them to perform tasks such as search, mapping and annotation [7]. Fig. 2 shows an example of a BioPortal API request.

In the remainder of the paper we will refer to users *exploring* ontologies when users browse the ontologies and their classes using the BioPortal WebUI (Fig. 1). We will refer to users *querying* ontologies when the users make the BioPortal API requests (Fig. 2), usually to retrieve content for their application. Finally, *reuse* will refer to the situation in which an ontology uses a class defined in another ontology (see Section 3.3).

Problem. Despite the success and widespread adoption of BioPortal for biomedical and Semantic Web research during the last 10 years, we do not have a clear understanding of how researchers use BioPortal to explore, query and use ontologies in their own projects. Remedying this missed opportunity by investigating empirical usage data from BioPortal will allow BioPortal developers to better target their efforts to meet the needs of biomedical researchers. Through such an investigation, biomedical researchers can identify frequently used classes in their ontologies, while

^{*} Corresponding author.

E-mail address: maulikrk@stanford.edu (M.R. Kamdar).

¹ <http://bioportal.bioontology.org>.

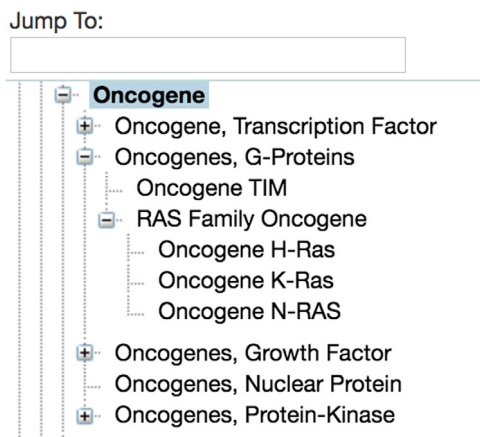


Fig. 1. The Oncogene subtree of the National Cancer Institute Thesaurus (NCIT) displayed in class hierarchy of the BioPortal WebUI. The “JumpTo” field allows users to quickly select a class in the hierarchy.

ontology engineers could concentrate their efforts on improving the content of highly accessed classes in their ontologies. We believe that insights gained by such an investigation can guide the development of new web-based interfaces that embed efficient semantic resource search and exploration methods. It will also enable the development of methods to profile users based on their behavioral characteristics and provide targeted recommendations.

We attempt to answer the following research questions in our work:

1. **RQ1: Do BioPortal WebUI exploration and API querying inform reuse?** Are the classes that users explore and access more often through the WebUI and API also reused more often in other ontologies?
2. **RQ2: Do BioPortal WebUI exploration and API querying correlate?** Do users explore the same classes in the BioPortal WebUI that they query through the API?
3. **RQ3: Do BioPortal WebUI exploration and API querying inform usage?** Are the classes that users explore and access more often through the WebUI and API also used more in downstream applications?

The research questions were formulated, in part, due to our prior research, categorizing exploration behaviors from BioPortal WebUI logs and analyzing ontology reuse across biomedical ontologies [5,8].

Approach. In this paper, we present an empirical investigation to help improve our understanding of how researchers (i) explore, (ii) query, and (iii) reuse biomedical ontologies and terminologies from BioPortal [5,8,9]. We hypothesize that users of BioPortal mainly use the WebUI to explore classes of ontologies to determine their utility for a specific task before programmatically querying (API) and using these classes for ontology reuse or downstream applications. For example, a user might first explore the *Oncogene* class of the NCI Thesaurus [10] in the BioPortal WebUI (Fig. 1), and then, if she deemed it appropriate for her application, she would access the *Oncogene* class through the BioPortal API. Specifically, we apply several visualization techniques to analyze commonalities and differences among information consumption strategies across different interfaces (i.e., WebUI and API) and purposes (i.e., exploration, querying, reuse and downstream applications).

For each ontology, we analyze correlation statistics between WebUI exploration data, API querying data and reuse data. To inspect and compare large amounts of exploration, query and reuse

```
http://data.bioontology.org/search?q=Oncogene
&ontologies=NCIT
```

Fig. 2. An example BioPortal API request to retrieve all classes called “Oncogene” from the NCI Thesaurus.

data, we make use of and extend PolyGonto [8], a visualization method in which an ontology is represented as an abstract geometrical polygon. The PolyGonto visualization method enables a quick comparison of individual usage and information consumption strategies across different interfaces and ontologies. Through three scenarios, we further analyze whether user interactions in BioPortal inform real-world applications of ontologies.

We discuss some of the key findings of our empirical analysis with respect to ontology users’ exploration and querying behavior, and highlight opportunities for further investigations. All results presented in this paper, including PolyGonto and other visualizations we have developed for each ontology, are available online at: <http://onto-apps.stanford.edu/vision>.

The datasets used in this study have also been published as TSV and RDF files at: <http://onto-apps.stanford.edu/bionic> under the Creative Commons CC-BY-NC-SA license [11].

The remainder of this paper is organized as follows: Section 2 describes work related to the analysis of user logs in the context of the Semantic Web, and to ontology visualization. In Section 3, we characterize the datasets used in the different analyses. In Section 4 we outline the methods used to compute and visualize correlation statistics, and delve deeper into the PolyGonto visualization method. We present the results of our empirical investigation in Section 5. Finally, in Section 6, we discuss few key findings regarding the ways in which users explore and query biomedical ontologies.

2. Related work

2.1. Log analysis to characterize user behavior

Several studies have tried to identify ways in which users interact with ontologies and ontology editors in the context of collaborative ontology development. These studies have used the data provided by logs of user activity in collaborative ontology development tools. Strohmaier et al. [12] conducted an empirical investigation using user activity logs to measure the impact of collaboration on ontology-engineering projects. The authors developed several new metrics to quantify different aspects of the hidden social dynamics that take place in these collaborative ontology-engineering projects from the biomedical domain. Falconer et al. [13] investigated and classified users according to different roles in collaborative projects by clustering users according to the types of changes they contributed.

Debruyne et al. [14] used different characteristics (reputation sensors) to measure the reputation of users in collaborative ontology-engineering projects with the goal of identifying “leaders” that drive activity, quality or social interactions. Using a combination of *k*-means and the GOSPL methodology [14], Van Laere et al. [15,16] classified users by clustering interactions that users engage in while engineering an ontology. Vigo et al. [17] analyzed eye-tracking data and event logs from the Protégé ontology editor to identify common user activity patterns, and proposed guidelines for bulk editing and modifications to the Protégé user interface.

In 2013, Wang et al. [18] applied association-rule mining to the change-logs of several different collaborative ontology-engineering projects to extract edit patterns, which were then used to predict the next change actions in the corresponding projects. Similarly, Walk et al. [19–21] used (higher-order) Markov chains

Download English Version:

<https://daneshyari.com/en/article/6950439>

Download Persian Version:

<https://daneshyari.com/article/6950439>

[Daneshyari.com](https://daneshyari.com)