BiOnIC: A Catalog of User Interactions with Biomedical Ontologies

MAULIK KAMDAR, SIMON WALK, TANIA TUDORACHE, MARK MUSEN

Stanford Center for Biomedical Informatics Research
maulikrk@stanford.edu

16th International Semantic Web Conference (ISWC)
Vienna, 21st - 25th October 2017
Benefits of analyzing user interactions

- **Ontology Engineers:**
  - Identify exploration and querying patterns
  - Understand ontology usage and reuse
  - Prune unwanted classes and relations

- **Ontology Repository Maintainers:**
  - Categorize user behaviors
  - Develop intelligent interfaces
  - Provide targeted recommendations

- **Biomedical Researchers:**
  - Identify temporal research trends
  - Identify frequently accessed classes
BiOnIC: A Catalog of User Interactions with Biomedical Ontologies

We provide aggregate statistics on the total number of clicks and queries for access to each class in 255 BioPortal ontologies, as well as reuse counts and class characteristics in the class statistics datasets. The access requests are extracted from Apache logs between January 2013 to June 2016. We also provide anonymized user sequences of access for each ontology using WebUI or API modes, under the user interaction sequences datasets. The datasets are available as TSV files and RDF HDT formats. To know more about HDT file formats and converting them to standard N-triples or turtle formats, please visit http://rdfhdt.org

http://onto-apps.stanford.edu/bionic/datasets
Welcome to BioPortal, the world's most comprehensive repository of biomedical ontologies

API Documentation

General Usage

This API is comprised of a set of resources (Ontologies, Classes, etc) and related endpoints (Search, Annotator, Recommender) that are connected together via links, much like webpages. We recommend that you try browsing the API using a web browser (Chrome and Firefox work very well while IE does not) before you start writing code. For more information, please see the documentation on Media Types and Hypermedia Links or view our sample code, available in Java, Python, Ruby and other languages (please email support@bioontology.org if you would like examples in another language).

Common Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Possible Values</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>apikey</td>
<td>(your apikey)</td>
<td>An API Key is required to access any API call. It can be provided in three ways:</td>
</tr>
</tbody>
</table>

1. Using the apikey query string parameter
2. Providing an Authorization header:

   `Authorization: apikey token=your_apikey` (replace 'your_apikey' with your actual key)

http://bioportal.bioontology.org/
BiOnIC datasets creation

Filtering Access Logs → Filtering Ontologies → Computing Class Counts → Computing Sequences → Anonymizing Data

User Interaction Activity Datasets

- **IP addresses** anonymized using unique SHA-224 hash-encoded user identifiers generated from “user_<Random String>_<Random_Integer>”.
  - e.g. 39fd4e6d569a034973fb61bb392a694d4eae1ef2b3e43ee68c4fic86

- **Absolute Time-stamps** converted to relative time-stamps with respect to first interaction with BioPortal repository.
  - e.g. 0, 2757, 2786, 3586, 3618, 3803, 3959, 4047, 5111 (s), ...

Reuse Attributes: from other ontologies.
- Number of ontologies reusing a class

Class Depth ->
- Ontology 1
- Ontology 2

Class Statistics Datasets
- For each class in each ontology:
  - **Access Attributes**:
    - Total IP Requests (WebUI/API)
    - Unique IP Requests (WebUI/API)
  - **Reuse Attributes**:
    - Number of ontologies reusing a class
  - **Structural Attributes**:
    - Number of parent/child/sibling classes
    - Depth from ontology root

User Interaction Sequences Datasets

Anonymization Steps
- IP addresses anonymized using unique SHA-224 hash-encoded user identifiers generated from “user_<Random String>_<Random_Integer>”.
  - e.g. 39fd4e6d569a034973fb61bb392a694d4eae1ef2b3e43ee68c4fic86

- Absolute Time-stamps converted to relative time-stamps with respect to first interaction with BioPortal repository.
  - e.g. 0, 2757, 2786, 3586, 3618, 3803, 3959, 4047, 5111 (s), ...

Anonymizing Data
BiOnIC schema to model statistics and sequences data

SKOS, PROV and DCAT standards are reused in the BiOnIC schema.
BiOnIC datasets

http://www.rdfhdt.org/

http://onto-apps.stanford.edu/bionic/datasets
Characteristics of the BiOnIC Catalog

- **WebUI Access:** 5.4M class requests, 1M unique agents
- **API Access:** 67.2M class requests, 205K unique agents
- 255 biomedical ontologies
VisIOn (Visualizing Ontology Interactions) Web Application

http://onto-apps.stanford.edu/vision
PolygOnto visualizations of interaction sequences

Applications of BiOnIC and VisIOn
Temporal influences in browsing and querying

**Fisher’s exact test with FDR:** Certain classes (e.g. *Ebolavirus*) or sets of classes are browsed or queried significantly more, when compared between different time periods.
Interface influences in browsing and querying

Certain classes browsed or queried significantly more.

- Dysmorphic Syndrome
- Night blindness
- Female Reproductive System
- Dermis
- Number of Unique API Users (Log Scale)
- Number of Unique WebUI Users (Log Scale)

- 1000
- 100
- 10
- 1

- YKL40 measurement
- ovarian cancer cell lines
- Columbia ecotype
- lung cancer cell line
- squamous cell carcinoma

- metastasis free survival
- experimental factor
- cell line
- disease
- cancer
- information entity
- cardiovascular disease
- protocol
- sampling site
- cardiomyopathy
- MSTO-211H
- GM06990
- Homo sapiens
Exploration and Querying behavioral patterns

- Certain classes in the **lower levels of the ontological hierarchy** are rarely browsed and queried – this may be an artifact of the indented tree visualization.
- More **triangular polygons** (1 parent -> 2 children classes, or 2 parents -> 1 child class) observed in WebUI Access polygon due to indented tree visualization.
Modeling user behaviors through Markov Chains

Novel research directions may be enabled through the BiOnIC and VisIOn resources

• Categorize user browsing behaviors by incorporating the structural features of the ontology classes.

• Develop personalized user interfaces for ontology navigation, which take into account the user type and the predictions of the next class that a user is likely to access.

• Develop advanced methods for ontology summarization and modularization, using BiOnIC datasets as features.

...
Acknowledgments

Musen Lab, Stanford
BMI PhD Program, Stanford

US NIH Grants
U54-HG004028
GM086587

maulikrk@stanford.edu
http://onto-apps.stanford.edu/bionic
http://onto-apps.stanford.edu/vision