Analyzing User Interactions with Biomedical Ontologies

A Visual Perspective

Maulik Kamdar, Simon Walk, Tania Tudorache, Mark Musen

Stanford Center for Biomedical Informatics Research
maulikrk@stanford.edu

17th International Semantic Web Conference (ISWC) 2018
11th October 2018

Welcome to BioPortal, the world's most comprehensive repository of biomedical ontologies.

SNOMED CT Details:
- Acronym: SNOMEDCT
- Visibility: Public
- Description: SNOMED Clinical Terms
- Status: Alpha
- Format: UMLS
- Contact: NLM Customer Service, custserv@nih.gov
- Home Page: http://www.snomed.org
- Publications Page: http://www.snomed.org
- Documentation Page: http://www.snomed.org
- Categories: Health
- Groups: Cancer Biomedical Informatics Grid, Unified Medical Language System
- License Information: This ontology is made available via the UMLS. Users of all UMLS ontologies must abide by the terms of the UMLS license, available at https://uts.nlm.nih.gov/license.html

SNOMED CT Metrics:
- Number of Classes: 327128
- Number of Individuals: 0
- Number of Properties: 152
- Maximum Depth: 28
- Maximum Number of Children: 2403
- Average Number of Children: 5
- Classes with a single child: 32114
- Classes with more than 25 children: 2500
- Classes with no definition: 325000

SNOMED CT Visits:
Graph showing visit data with peaks and troughs.

http://bioportal.bioontology.org/
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>
<tr>
<td></td>
<td></td>
<td>1. Using the apikey query string parameter</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Providing an Authorization header:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Authorization: apikey token=your_apikey (replace 'your_apikey' with your actual key)</td>
</tr>
</tbody>
</table>

http://bioportal.bioontology.org/
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
Research Questions

• Do BioPortal WebUI exploration and API querying strategies correlate with each other?

• Do BioPortal WebUI exploration and API querying strategies inform ontology usage?
  – Ontology usage, in this context, means reuse in other ontologies and usage in data annotation.
What this talk is about ...

- **BiOnIC** - Catalog of User Interactions with Biomedical Ontologies
- **VisIOn** application with embedded visualizations.
What this talk is about ...

• **BiOnIC** - Catalog of User Interactions with Biomedical Ontologies

• **VisIoN** application with embedded visualizations.

Analysis of BioPortal WebUI exploration and API querying strategies, and correlation with usage.
What this talk is about ...

- **BiOnIC** - Catalog of User Interactions with Biomedical Ontologies
- **VisIOn** application with embedded visualizations.

Analysis of BioPortal WebUI exploration and API querying strategies, and correlation with usage.
BiOnIC: A Catalog of User Interactions with Biomedical Ontologies

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

Depth of an Ontology

subClassOf

subClassOf
PolygOnto visualizations of interaction sequences

Depth of an Ontology

Breadth of an Ontology
PolygOnto visualizations of interaction sequences

Depth of an Ontology

Breadth of an Ontology
PolygOnto visualizations of interaction sequences

Depth of an Ontology

Breadth of an Ontology
PolygOnto visualizations of interaction sequences

Depth of an Ontology

Breadth of an Ontology
PolygOnto visualizations of interaction sequences

- Depth of an Ontology
- Breadth of an Ontology

User Sequence

Ontology Structure

Single Click
What this talk is about ...

• **BiOnIC** - Catalog of User Interactions with Biomedical Ontologies
• **VisIOn** application with embedded visualizations.

Analysis of BioPortal WebUI exploration and API querying strategies, and correlation with usage.
Research Questions

- Do BioPortal WebUI exploration and API querying strategies correlate with each other?

- Do BioPortal WebUI exploration and API querying strategies inform ontology usage?
  - Ontology usage, in this context, means reuse in other ontologies and usage in data annotation.
Interface influences in browsing and querying

Certain classes browsed or queried significantly more.
Certain classes browsed or queried significantly more.
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)
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.
Exploration and Querying behavioral patterns

SNOMEDCT UI Access

SNOMEDCT API Access

300,000+ classes
200,000+ users

Width of the Ontology (Number of Terms in each Hierarchical Layer)
Do BioPortal WebUI exploration and API querying strategies correlate with each other?

- **Small proportions** of ontological content explored or queried for several ontologies.

- **Minimal Spearman Correlation** and Jaccard Similarity observed between consumption strategies.

- Classes in the **lower layers of the ontological hierarchy** are rarely browsed or queried using the WebUI or API.
Research Questions

• Do BioPortal WebUI exploration and API querying strategies correlate with each other?

• Do BioPortal WebUI exploration and API querying strategies inform ontology usage?
  – Ontology usage, in this context, means reuse in other ontologies and usage in data annotation.
Unified Medical Language System (UMLS)

SNOMEDCT

Diabetes Mellitus

C0011849

Diabetes Mellitus

ICD9CM

Open Biomedical Ontologies (OBO) Foundry

RNA Binding (GO:0003723)

IRI

GO:0003723

xref

Binding to RNA (GRO#BindingToRNA)

Gene Expression Ontology (GEXO)

Gene Ontology (GO)

Gene Regulation Ontology (GEXO)
Ontology usage: Reuse in other ontologies

Distinct class sets reused in other ontologies

Distinct User Polygons
Ontology usage: GWAS Catalog and PubChem annotations

Classes in the higher layers of the ontological hierarchy are very abstract and are never used for data annotations.
Ontology usage: GWAS Catalog and PubChem annotations

Classes in the higher layers of the ontological hierarchy are very abstract and are never used for data annotations.
Do BioPortal WebUI exploration and API querying strategies inform ontology usage?

- **Minimal correlation and similarities** between exploration and querying strategies, and usage.

- **Ontology reuse** occurs from classes located in the **higher layers** of the hierarchy for semantic interoperability between ontologies.

- **Data annotation and data integration** generally uses classes in the **lower layers** of hierarchy.
  - These classes are rarely explored or queried, we need better interfaces!
This talk was about ...

- **BiOnIC** - Catalog of User Interactions with Biomedical Ontologies
- **VisIOn** application with embedded visualizations.

Analysis of BioPortal WebUI exploration and API querying strategies, and correlation with usage.