

Ontology-based Tools to Enhance the Curation Workflow

Patricia L. Whetzel, Nigam H. Shah, Natasha F. Noy, Clement Jonquet, Cherie Youn, Paul Alexander, Michael Dorf, and Mark A. Musen

Stanford University, Stanford, CA, USA

Introduction

In order to effectively search, retrieve, and analyze data oftentimes it is curated and tagged with ontology terms. However, the amount of effort to curate the existing set of data resources is beyond the limits of purely manual curation. We present three ontology-based tools developed by the National Center for Biomedical Ontology to enhance the curation workflow: Ontology Widgets, Notes, and the Annotator. The Ontology Widgets provide a mechanism to use ontologies in Web-based forms without the need to locally parse and store the ontology. The Ontology Widgets provide a variety of functionality including term autocompletion and ontology visualization. The Ontology Widgets are implemented for all BioPortal ontologies, including those from the OBO Foundry and Unified Medical Language System. The Notes feature of BioPortal allows structured term proposals to be submitted in order to request the addition or modification of a term in an ontology. The term proposals can be added directly via the BioPortal Web interface or programmatically via the Notes Web service. Notification of new Notes and replies are both RSS- and Email-enabled. Once the term curation process is complete, the OWL class or OBO stanza can be generated via the Notes Web service. Finally, the Annotator can be used to automatically process textual metadata to identify ontology terms found within the text. The Annotator can be accessed programmatically via the Annotator Web service and can be used with all BioPortal ontologies. In summary, the Ontology Widgets, Notes, and Annotator provide mechanisms to enhance curation by helping collect annotated data upon data submission, by facilitating ontology term curation, and by tagging unstructured textual data with ontology terms.

BioPortal Web Services

- Prefix for BioPortal Web Services http://rest.bioontology.org/bioportal/
- Ontology Web services
 - List metadata for all ontologies
 - List all versions of an ontology
 - Download ontologies
 - List all ontology categories
 - List all ontology groups
- Search Web service
 - Search for terms from all BioPortal Ontologies
 - Search for term within a subtree of an ontology
- Term Web services

- Hierarchy Web services
 - Get parents/children of a term
 - Get path to roots/leaves of a term
 - Get siblings of a term
- RDF Generation Web service
 Generate a N3 RDF file
- Notes Web service
- Get comments for a term
 - Get all comments for an ontology
- Mapping Web service
 - Get mappings for a term
 - Get all mappings for an ontology Mapping Web service

BioPortal Ontology Library

- Open ontology repository for publishing biomedical ontologies
 Features
 - Search across and within all ontologies
 - Browse ontologies
 - Provide feedback Notes and Reviews

•Statistics

- Total number of ontologies: 219
- Number of classes/types: 1,438,792

•Supports many ontology formats: OWL, OBO format, Rich Release Format, Protégé frames



- Get term by id
- Get all root terms for an ontology
- Get all terms from an ontology

Annotator Web Service

- Prefix for Annotator Web Service http://rest.bioontology.org/obs/annotator/
- Description
 - The NCBO Annotator annotates or "tags" free-text data with terms from BioPortal and UMLS ontologies. The current entity recognizer used by the Annotator is Mgrep, developed by the National Center for Integrative Biomedical Informatics. Plans are underway to include an additional entity recognizer, MetaMap from the National Library of Medicine. The annotation is performed in two steps; direct annotations via matching the input text with the preferred name or synonyms of ontology terms and expanded annotations taking into account the ontology hierarchy and mappings.
- Documentation
 - For more information and sample clients, see:

http://www.bioontology.org/wiki/index.php/Annotator_Web_service



Enhancing the Curation Workflow

Data Submission

- Ontology Widgets
 Ontology Web Services
 Ontology Enrichment
 BioPortal Notes
- Data Annotation
 - NCBO Annotator

Ontology Widgets

- Ontology widgets provide ready-to-use code to access your ontology of interest from your Web site
- For more information, see: http://www.bioontology.org/wiki/index.php/Ontology_Widget



Web Service Access

Ontology Development Lifecycle

- Involves both Subject Matter Experts (SMEs) and Ontology Developers
 Consists of:
 - Discussion between SMEs and Ontology

Developers

- Ontology Editing
- Publishing Ontology to the Web
- Collect Feedback on Published Ontology

BioPortal Notes

Biomedical Resource Ontology 🛛 🛞 NCI The	saurus										
medical Resource Ontolo	gy 🔝 Subscribe	to updates									
bmit New Version Edit Ontology Inform	mation										
adata Views (6) Projects (4) Rev	iews (1) Notes (66)	Metrics Ontolog	y Widgets								
ow 50 🗘 entries				Search:	Biomedical Resource Ontology Version	n 3.1 Physic	ology Facility Link Here	Subscribe			
SUBJECT	AUTHOR	TYPE	TARGET		View Ontology Summary	Details Visualization	Notes (3) Mappings (0) F	lesource index			
New Term Proposal: Exercise Study Facility	Mette	New Term Proposal	Physiology Facility (Class)		Jump To: (Go)	Show 50 ¢ entries		Search:			
New Term Proposal: Sleep Study Facility	Mette	New Term Proposal	Physiology Facility (Class)		Area of Research Biostemaps Information Model core:Collection core:coreset	SUBJECT New Term Proposal: Ex	AUTHOR	TYPE New Term Proposal	CREATED		
Deprecate	whetzel	Comment	Symbolic and Analytic Mode	<u>el</u> (Class)	eore:ConceptScheme	Study Facility		- Feelike			
e-activate this term	whetzel	Comment	Web Service (Class)		Deprecated Area of Research Deprecated Resource	New Term Proposal submitted by Mette 3 months ago on <u>Physiology Facility</u> in <u>Biomedical Resource Ontology</u>					
Move in hierarchy	whetzel	Comment	Material Analysis Service (C	Class)	Resource Funding Resource	PREFERRED NA	ME Exercise PROPOSED Study	D PARENT http://bioor /Biomedical	ntology.org/ontologies ResourceOntology.owl#Physiology_Facil		
Proposal	whetzel	Comment	People Resource (Class)		Information Resource Material Resource	REASON FOR CHAN	Facility GE Physiology STATE	IS CONTACT INFO			
Add Education Service as a Synonym	whetzel	Comment	Training Service (Class)		Biomaterial Supply Resource Biomedical Supply Resource	facility child					
Synonym proposal	whetzel	Comment	Robotic Arrayer (Class)		Facility Core Biosafety Level Facility	SYNONY	MS	to everyise studies			
Term name	whetzel	Comment	Molecular Modeling and Classification (Class)		Cell Biology Facility Fabrication Facility	MEETINET OF THINKING IN SAME ASTRONOM IN SAMELING SUBJECT					
Re-structure branch	whetzel	Comment	People Resource (Class)		Imaging Facility Molecular Biology Facility	Responses hide al	1 show all				
Discuss this branch with collaborating groups, e.g. eagle-i, OBI, NIF	g whetzel Comment Instrument (Class)		Physiology Facility Research Animals Facility Tissue Organ Facility		RE: New Term Proposal: Exercise Study Facility by whetzel 3 months ago						
Review definition	whetzel	Comment	Physiology Facility (Class)		Laboratory Supply Resource Medical Device	Can you expan	d on what an exercise study is				
Refine definition	whetzel	Comment	Molecular Biology Facility (Class)		Reagent Resource Resource Software	Exercise study facility definition by Mette 3 months ago A facility or core that provides services and equipment related to exercise physiology					
Add Image to Data_Resource branch	whetzel	Comment									
Clarify action item	whetzel	Comment	Programmatic Access (Class	s)	Training Resource	reply					
		1									
						Add Reply					

Resource Index Web Service

- Prefix for Resource Index Web Service http://rest.bioontology.org/resource_index/
- Description
 - Provides an ontology-based index of publicly available biomedical resources.

- Prefix for NCBO Web Services: http://rest.bioontology.org/{application}
- List of Web Services http://bioontology.org/wiki/index.php/NCBO_REST_servic es
- Access via Code (Perl, Java, Flash, Ruby, etc.), User Interface, Browser, or Widget



BioPortal Brows	se Search	n Pro	Projects		All Map	All Mappings		All Resources	
Biomedical Resource Ontology	Gene Ontology Extension	ension 📀 Expe	erimental Factor	Ontology	Pathway onto	.ogy 📀	Ontology for	r Bior	
Search all ontologies						Searc	h time: 280 r	ns –	
melanoma		# Search	Categories	All Categories			•		
			Groups	All Groups					
Include attributes in search		Help	Filter t	ype filter text					
 Contains Exact Match 		× Clear	Ontologies		Select All Sele	ect None			
			ABA Adul	t Mouse Brain (A	BA)				
Most popular searches (all users)	:			Event Ontelany (450)			•	
melanoma diabetes heart cancer	cell gene Influenza p	atient brain 🕩	Adverse	Event Ontology (AEO)		_		
Recent searches (all users):			African T	raditional Medicir	ne (ATMO)				
MRI amnion schwann cell macroglia	a C0025202 melanoma	generalized exa 🕨	AIR (AIR)					
Selected Ontologies (199):			Amino Ac	id (amino-acid)					
All Ontologies		Amphibian gross anatomy (AAO)							
Matching Terms								lts –	
Filter type filter text	Exact Matches Only	Ontology Filter	43 ontologies	•			Columns	•	
Term Name	Term ID	Ontology		Version ID	Ontology ID	Details	Visualize		
Melanoma	Melanoma	Galen		4525	1055				
Melanoma	lanoma DOID:1909		,	39927	1245				
Melanoma	anoma Melanoma		NCI Thesaurus		1032		4	ш	
lelanoma D008545		Descritores en C	iencias de la Sa	42236	1420		-		
nelanoma 11223:68		Health Level Sev	<u>en</u>	42545	1343			ш	
Melanoma PMA_1619		PMA 2010		42837	1497				
Melanoma 10053571		MedDRA		42280	1422		4	ш	
Melanoma	MedlinePlus Healt	th Topics	40397	1347					
delanoma C4116		National Drug File	<u>e</u>	40402	1352			ш	
Melanoma	MTHU000127							ш	
melanoma	CDR0000038833		se le	ato	ef o c			ш	
melanoma	EV:0400102	USE		ne		,e	-		
melanoma	MPATH:359								
melanoma 2020-2434		CRISP Thesaurus	s, 2006	44432	1526		- * *		
melanoma	efo:EFO_0000756	efo:EFO_0000756 Experimental Fac		44524	1136		<u>.</u>		
melanoma DOID:1909		Human disease		44594	1009				
Mallana at an ala ana a	2092003	SNOMED Clinical	Terms	42789	1353				
Malignant melanoma									

- Currently indexed resources include: Array Express, ARRS Goldminer, Biositemaps, caNanoLab, ClinicalTrials.gov, Conserved Domain Database, DrugBank, Gene Expression Omnibus, Molecular Imaging and Contrast Agent Database, NextBio, OMIM, Pathway Commons, PharmGKB, PubChem, Reactome, Research Crossroads, Stanford Microarray Database, UniProt, and WikiPathways
- For more information see:

http://www.bioontology.org/wiki/index.php/Resource_Index



Acknowledgements

BioPortal is developed by the National Center for Biomedical Ontology (NCBO), a National Center for Biomedical Computing under the NIH Roadmap. BioPortal is developed in conjunction with partners at the University of Victoria and the Mayo Clinic. For more information on NCBO see <u>http://www.bioontology.org</u>.

Contact

BioPortal Support: <u>support@bioontology.org</u> Protégé/WebProtege: <u>protege-</u> <u>discussion@lists.stanford.edu</u>