



HAL
open science

Ontology-based Web Services for Data Annotation and Integration

Patricia L. Whetzel, Nigam H. Shah, Natalya Noy, Clement Jonquet, Adrien Coulet, Cherie Youn, Michael Dorf, Mark A. Musen

► **To cite this version:**

Patricia L. Whetzel, Nigam H. Shah, Natalya Noy, Clement Jonquet, Adrien Coulet, et al.. Ontology-based Web Services for Data Annotation and Integration. 18th International Conference on Intelligent Systems for Molecular Biology (ISMB 2010), Technology track, Jul 2010, Boston, United States. , TT33, 2010. lirmm-01605892

HAL Id: lirmm-01605892

<https://hal-lirmm.ccsd.cnrs.fr/lirmm-01605892>

Submitted on 2 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Ontology-based Web Services for Data Annotation and Integration

Patricia L. Whetzel, Nigam H. Shah, Natalya F. Noy, Clement Jonquet, Adrien Coulet, Cherie Youn,
Michael Dorf and Mark A. Musen
Stanford Center for Biomedical Informatics Research, Stanford University
Stanford, California 94305-5479

Researchers have turned to the Semantic Web to annotate and integrate disparate knowledge. Ontologies provide the domain knowledge to drive these processes and the successful creation of semantic applications in the health and life sciences require Web services that provide access to bioontologies. The National Center for Biomedical Ontology (NCBO), one of the seven National Centers for Biomedical Computing created under the NIH Roadmap, has developed BioPortal, which provides access to one of the largest repositories of biomedical ontologies via both Web services and Web browsers.¹ NCBO has developed a set of Web services that access and use BioPortal ontologies in applications for data annotation and integration. These applications include the *Ontology Recommender*, which suggests ontologies to use for annotation based on textual metadata or keyword input, the *NCBO Annotator*, which “tags” text automatically with terms from BioPortal ontologies, and the *NCBO Resource Index*, which provides an ontology-based index of publicly available, online data resources.

The *Ontology Recommender* Web service uses textual metadata or keywords to suggest BioPortal ontologies that are best suited to annotate data based on three criteria; coverage, connectivity, and ontology size. Coverage refers to the number of terms in the ontology that match the input, connectivity relates to how often an ontology is mapped to another other ontology, and size is the overall number of terms in the ontology. The *Ontology Recommender* processes the textual input and generates a ranked list of ontologies that are best suited for annotation of the data.

The *NCBO Annotator* Web service processes text to recognize relevant biomedical ontology terms. Users can customize the Web service to limit results to a particular ontology (e.g. SNOMED CT) or to a certain UMLS semantic type (e.g. T017 for ‘Anatomical Structure’) as well as many other parameters documented at: http://www.bioontology.org/wiki/index.php/Annotator_Web_service. The entity recognition engine, MGREP, was developed by the National Center for Integrative Bioinformatics and is combined with the BioPortal *Ontology* Web services to create the *NCBO Annotator* Web

service. The availability of this functionality makes the task of creating ontology-based annotations accessible for any biomedical researcher by using the Web service directly through programmatic access or by using Addins such as the NCBO Annotator Excel Addin.

The *NCBO Annotator* Web service has also been used by NCBO to generate an ontology-based index of several online biomedical data repositories (e.g., GEO, ClinicalTrials.gov, dbGaP, DrugBank, PharmGKB, and Reactome) resulting in the *NCBO Resource Index*. These indexes are created using the textual metadata of the database record as input to the *NCBO Annotator* and the database records are linked together via shared ontology annotations. These linkages take advantage of the semantic relationships within the ontology, including subsumption relationships among ontology entities and mappings between entities in different ontologies. The *NCBO Resource Index* Web service provides biomedical investigators a mechanism for programmatic search of the index using ontology terms. For example, one can search for all experiments and clinical trials corresponding to ‘malignant melanoma’ from GEO and ClinicalTrials.gov.

In conclusion, the ontology-based Web services provided by NCBO provide a “1-stop shop” to access ontologies as well as to annotate and integrate data. More information on these Web services can be found at www.bioontology.org/wiki/index.php/NCBO_REST_services.

REFERENCES

- Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith N, Jonquet C, Rubin DL, Storey MA, Chute CG, Musen MA. (2009) BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Res.*, **37**, W170–3.
- Shah NH, Jonquet C, Chiang AP, Butte AJ, Chen R, Musen MA. (2009) Ontology-driven indexing of public datasets for translational bioinformatics. *BMC Bioinformatics*, **10**, Suppl 2:S1.

¹<http://biportal.bioontology.org>