



The Noctua Modeling Tool

Seth Carbon*, Heiko Dietze*, Chris Mungall*

17th Bioinformatics Open Source Conference (BOSC) 2016, Orlando, FL, USA

Website: <http://noctua.berkeleybop.org>

Repository: <https://github.com/geneontology/noctua>

License: BSD 3-Clause

We present Noctua, a modern web application and stack for modeling complex biology. Noctua directly models information as a graph, escaping many of the pitfalls of more “tabular” modeling. Noctua also presents a rich, interactive, and collaborative user interface, as well as a complete set of tooling for data extraction and integration.

The Gene Ontology [1] project aims to create a comprehensive and up-to-date model of biological systems based on annotating knowledge graphs (ontologies) with curated and generated information. Over its existence, the project has aimed to capture this information with ever increasing richness and specificity. The traditional storage format for GO annotations, the tabular GAF, has undergone several iterations, but can no longer support the types of annotations required by current use cases. To move forward, the project has adopted LEGO, a graph-base abstraction for modeling biology. The Noctua application and stack was created to annotate information with the LEGO abstraction, but has the built-in flexibility to be used in any number of pathway or workflow models.

The current main end user client application for Noctua presents a real-time collaborative graph editing environment, allowing users to assemble graphs representing biological knowledge, including aspects such as references and evidence. The user environment uses typed inputs combined with a click, drag, and connect interface for easy and intuitive graph editing. As multiple users work on the same model, no matter the client, it updates the common environment in real time, allowing for easy discussion, presentation, and collaboration.

The Noctua stack is composed of three layers: the client, written in JavaScript using the jsPlumb and AmiGO/BBOP libraries; the communication layer, written in JavaScript and providing client-to-client and client-to-server communication; and the graph engine, a Java server that uses OWL to model and OWL universe tools for reasoning. This stack is strongly separated, respecting protocol and common patterns. For example, while the main user interface is a graph editor, it could be easily replaced by a different client that could speak the same wire protocol—the flexibility of the framework allows for the easy creation of alternate clients, such as the form or REPL based ones.

The Noctua stack is being actively used by the Gene Ontology, with the produced annotations finding their way back into the pipelines of several model organism databases. As well, Noctua data produced for the GO is loaded into AmiGO [2], where it is available for exploration and made available to be consumed in clients via our JavaScript API. The data modeling and retrieval systems for the client are available as separate packages, allowing third parties to create their own clients or embed annotation widgets in their own resource pages.

References

- [1] Gene Ontology Consortium: going forward. *Nucl. Acids Res.* 43, D1049D1056.
- [2] AmiGO: online access to ontology and annotation data. *Bioinformatics* 25, 288289.

*Berkeley Bioinformatics Open-source Projects, Lawrence Berkeley National Lab, Berkeley, CA, USA.