

Using Ontologies in Biological Research

Instructors

Janet Kelso
Nicola Mulder
Johann Visagie
Winston Hide

Winston Hide (PhD)

Winston Hide has 14 years of experience in the analysis of sequence diversity and in large and small scale sequence analysis with an established publication record in molecular evolution and genome analysis systems development. His work centres around the understanding of gene expression, development of integrative systems, and HIV bioinformatics.

Janet Kelso (PhD)

Janet is a postdoctoral researcher at the South African National Bioinformatics Institute (SANBI). During her PhD, Janet worked on the development and application of ontologies for human gene expression data. Janet has been involved in the establishment of bioinformatics training, services and research in South Africa through her involvement in the formation of the South African National Bioinformatics Network (NBN), and as chair of the NBN education committee. She has acted as the South African representative to the EMBnet board, and as secretary of the EMBnet Education and Training committee. In addition to a period as a lecturer in human biology, Janet has taught on a number of local and international bioinformatics courses and contributed towards the development of curricula for bioinformatics programs.

Nicola Mulder (PhD)

Nicola has a PhD in Medical Microbiology and spent the next four and a half years at the European Bioinformatics Institute, the last three of which in the role of manager of the InterPro database. Her main interests are in genome annotation and comparative genomics using ontologies. She is responsible for the mapping of InterPro entries to Gene Ontology (GO) terms for large-scale mapping of proteins to GO terms. She has taught at a number of bioinformatics courses in Brazil, the UK and Europe, particularly in the field of protein classification.

Johann Visagie (B.Eng)

Johann has extensive experience in object oriented software design and development. During his time at software vendor Electric Genetics he researched ontology design and laid the groundwork for a commercial ontology system for gene expression information. He has presented lectures on subjects as diverse as internet usage and open source software development.

Expected goals, objectives and motivation of the tutorial

Title

Using ontologies in biological research

Objectives

In this tutorial we describe and explain ontologies and controlled vocabularies with a view to showing those interested how these can be used in biological research. We provide a background on the field of ontologies, showing the value of implementing ontologies for

providing standardised descriptions and querying tools for biologists. We then move to a survey of the available ontology resources and describe how these can be used. Finally, we demonstrate the available tools for biologists wishing to create or edit ontologies.

Motivation

Biologists require the ability to use biological information from a variety of sources, and to be able to integrate this information in order to make biologically meaningful discoveries. The rapidly increasing amounts of many types of biological data, stored in numerous and disparate biological databases using different accessions and annotated with inconsistent terminology have made it difficult for the average biologist to identify and coherently query biological data.

Recent years have seen a growing trend towards the adoption of ontologies for the management of biological knowledge. Ontologies represent a powerful means to analyse and integrate biological data. Their successful utilisation is dependent upon broad use by the community.

Ontologies provide a formal specification of how to represent the objects, concepts and other entities that are assumed to exist in some area of interest, and the relationships that hold among them. Various ontologies and controlled vocabularies have grown out of the effort to provide a shared language for communicating biological information. Biological ontologies aim to overcome the semantic heterogeneity commonly encountered in molecular biology databases, and to provide a common terminology for the description of a focussed aspect of biology. However, while these resources are widely developed and used in bioinformatics research community, they have as yet to be as widely accepted and used by lab biologists. This is in part due to a lack of understanding of what these resources offer, and how they can be used. We aim to provide an overview of the available ontologies and an introduction to how these can be used in biological research.

Intended audience – indication of level at which the tutorial will be taught, and required background of participants

This tutorial is aimed at biologists with no prior experience with using or developing ontologies. It will:

- Guide participants to develop an understanding of the concept of ontologies
- Survey major approaches to building ontologies
- Describe widely used ontologies and discuss potential uses for the ontologies and any data annotated using the ontology.

Length

Half day

Detailed outline of the presentation

1. Introduction

The introduction will provide a definition and brief overview of ontologies including their background, application, and the essential elements of an ontology. The difference between term lists, controlled vocabularies and ontologies will be discussed with a view to providing context for the tutorial topics to be covered. The pure computer scientist understanding of an ontology will be compared to the definitions applied by biologists.

An outline of the broad categories or types of ontologies will be provided.

Some examples of the application on ontologies outside of biological sciences will be described.

2. Applications and Types of Ontologies in Biology

This section will deal with the ways in which ontologies can contribute to the biological sciences. Specifically, the types of ontologies useful for biological data and their applications to biological data entry, storage, integration and querying will be addressed.

3. Survey of Biological Ontologies

This chapter will provide an in-depth survey of some of the biological ontologies available and relevant to molecular biology. Ontologies chosen will be those included among the Open Biological Ontologies (OBO). These include general ontologies such as the Gene Ontology (GO), the MGED Ontology, the NCI Thesaurus, eVOC, the Plant ontology consortium. The TAMBIS ontology (TaO) and EcoCyc will also be discussed.

The ontologies surveyed will be discussed in terms of:

- The scope/domain of the ontology
- Concepts and relationships
- The size of the ontology
- Commonly used data/databases to which the ontology has been applied, or in which the ontology is implemented
- Software for using, or modifying the ontology
- Availability

A brief overview of other related ontologies will be included. These include organism specific ontologies such as those included in FlyBase, and and the Edinburgh Human developmental anatomy.

4. Ontology Manipulation Tools

A selected set of the software available for querying and editing biological ontologies will be presented. Many of the ontology projects have developed software for querying and editing their own ontologies. These will be described, and their specific strengths and weaknesses discussed.

Other applications which are useful in the analysis of laboratory data including microarray and other high throughput expression data will be presented with a view to guiding users as to which tools are appropriate for their needs.

Applications to be covered are:

- 4.1. Ontology Editors**
 - 4.1.1. DAG EDIT**
 - 4.1.2. COBrA**
- 4.2. Ontology Browsers**
 - 4.2.1. AmiGo**
 - 4.2.2. GOBrowser**
 - 4.2.3. PANDORA**
- 4.3. Other Useful Tools**
 - 4.3.1. GOSurfer**
 - 4.3.2. Onto-Miner**

5. Building Ontologies

Users will be guided through the process of setting up a basic ontology. Using specific General principles of setting up an ontology, including what should be included, and rules which need to be followed will be discussed.

The tools available for setting up ontologies will be used to demonstrate the basic principles.

6. Ontology Consortia

There are numerous efforts to establish ontologies in various areas of biology. These will be listed and briefly described as they are a valuable resource for biologists wishing to utilise them. Some are listed below.

- 6.1. The Open Biological Ontologies (OBO)**
- 6.2. Microarray Gene Expression Database Ontology Working Group (MGED OWG)**
- 6.3. Gene Ontology (GO)**
- 6.4. Plant Ontology Consortium (POC)**