

СОВРЕМЕННЫЕ БИО-ОНТОЛОГИИ: МОДЕЛЬ, ИНСТРУМЕНТ ИЛИ СПРАВОЧНИК?

Стриж И.Г.

Кафедра физиологии растений биологического факультета Московского государственного университета им. М.В. Ломоносова, Россия, 119992, Москва, Ленинские горы д.1, кор.12, тел: (495) 939-54-87; irina.strizh@mail.ru,

Принципиальным отличием биологии XXI века является то, что практически любое современное исследование не только приводит к появлению многочисленных данных, но и требует привлечения новых технологий, способствующих решению проблемы хранения, обработки и интеграции получаемых данных, а также позволяющих исследователям вычленив необходимую им информацию. Помочь ученым призваны методы биоинформатики, а именно методы организации молекулярно-биологической информации, широко понимаемые компьютерные методы, методы вычислительной математики и статистики. Одним из подходов, применяемых для интеграции данных и знаний в биологии, является использование онтологий. Созданием онтологий для организации молекулярно-биологических и текстовых данных ученые занялись относительно недавно, поэтому до сих пор нет консенсуса, что является «истинной» онтологией и как она должна быть правильно реализована. Из основ философии известно, что онтология это наука о бытии – о природе вещей и взаимосвязях между ними. В контексте информационных технологий представления знаний, этим термином можно определить механизм или способ, используемый для описания некоторой области знаний, в частности базовых понятий этой области и связей между ними. Т.е. изначально онтологии были задуманы как некоторый инструмент для описания знаний. Важное условие при создании онтологий – они должны быть достаточно удобны и понимаемы и для людей, и для программных систем, поэтому часто говорят, что онтология является «машиночитаемой» моделью некоторой предметной области. Большинство современных био-онтологий представляют собой не столько модели, сколько огромные словари терминов-концептов, формирующих описание определенной области знаний. Наиболее известной используемой био-онтологией является Gene Ontology, состоящая из трех контролируемых словарей терминов, обозначающих молекулярную функцию, биологический процесс и клеточную структуру. Такой способ реализации онтологий значительно отличается от «классического», принятого в информационных технологиях, а их громоздкость сильно затрудняет их использование в качестве модели предметной области. Вместе с тем, современные био-онтологии, являющиеся, по сути дела, контролируемые словари, позволяют функционально аннотировать экспериментальные результаты, являются эффективным способом навигации по базам данных, а также могут быть использованы в качестве общедоступного электронного справочника или учебного пособия в образовании.