

Bioontologies as a tool for plant research

Anna S. DOBRONOGOVA, Mikhail Yu. CHEREDNICHENKO

Department of Genetics, Biotechnology, Plant Breeding and Seed Science,
Russian Timiryazev State Agrarian University, Russian Federation
adobronogova@gmail.com

Aim of the study: Due to the huge volume of electronic documents, there is an ever-growing need for processing unstructured textual information, improving the quality and efficiency of existing text processing methods. Initially, information retrieval thesauri were developed to adapt information for retrieval and for needs of information and analytical systems, which were supposed to be manually indexed. The modern paradigm of computer resources for information retrieval applications are formal ontologies. Ontologies cover most of the words of a language or domain and simultaneously have an ontological structure that manifests itself in the relations between concepts.

Material and Methods: The most significant bioontologies were observed. GO is a structurally controllable dictionary dedicated to the unification of the terminology of gene and gene product annotations of all biological species. PO is a public structured dictionary of terms for the anatomy and morphology of plants, as well as the stages of their development. ChEBI includes the ontology of natural compounds or synthetic products that affect the processes in living organisms. CL is a structured controlled layer that includes a description of the cellular types of various organism types. KEGG is a series of ontologies that focus on a broad field of molecular biology - from genes and proteins to metabolic and gene networks. SO includes many concepts and controlled dictionaries used to describe the properties and primary annotation of nucleotide or protein sequences, the structural representation of these annotations in genomic databases, mutations in both sequence types and at a higher level. The MGED ontology is used to describe experiments and data on gene expression. MIAME is used to describe expression data. PharmGKB provides information on pharmacogenetics. Cell Cycle Ontology is an extension of existing cell cycle-related ontologies to integrate and manage knowledge about its components and regulatory aspects.

Results: In general, in bioinformatics and system biology, the following tasks can be distinguished, in which the application of ontologies gives a noticeable effect. (·) Interpretation of molecular genetic knowledge, semantic interpretation of methods of data analysis and models in system biology. (·) Prioritization of genes, proteins, biomarkers, etc. (·) Analyzing the similarity and clustering of objects. (·) Support for interoperability and knowledge sharing - unified access to sets of heterogeneous data sources. (·) Creation of new ontologies based on the repeated use of basic canonical ontologies and various types of operations with them, including ontology matching, ontology merging, ontology mapping, ontology alignment. (·) Ensuring consistency and correctness of knowledge representation. (·) Support inductive inference for extracting additional knowledge from a variety of facts and testing hypotheses. (·) Increasing the rationale for bioinformatics methods.

Keywords: plant ontology, bioinformatics, secondary metabolism, plant research.