IV-30-2010

Abstract – Regular Paper – iEvoBio Conference 2010

Title: Biological taxonomy and ontology development: scope and limitations

Authors: Nico M. Franz[1] and David Thau[2]

[1] Presenter; Department of Biology, Call Box 9000, University of Puerto Rico, Mayagüez, PR 00681-9000 U.S.A., nico.franz@upr.edu

[2] Department of Computer Science, University of California at Davis, Kemper Hall, One Shields Avenue, Davis, CA 9561 U.S.A., thau@learningsite.com

Abstract. The prospects of integrating full-blown biological taxonomies into an ontological reasoning framework are reviewed. We contrast the common usage of a *static* 'snapshot' hierarchy in ontological representations of taxonomy with a more realistic situation that involves *dynamic,* piece-meal revisions of particular taxonomic groups and requires alignment with relevant preceding perspectives. Taxonomic practice is characterized by a range of phenomena that are orthogonal to the logical semantic background from which ontological entities and relationships originate, and therefore pose special challenges to ontological representation and reasoning. Among these phenomena are: (1) the notion that there is a single phylogenetic hierarchy in nature which taxonomy can only gradually approximate; (2) the evolvability of taxa which means that taxon-defining features may be lost in subordinate members or independently gained across multiple sections of the tree of life; (3) the hybrid approach of defining taxa both in reference to properties (intensional) and members (ostensive) which undermines the individual/class dichotomy sustaining conventional ontologies; (4) the idiosyncratic yet inferentially valuable usage of Linnaean ranks; (5) the indelible and semantically complex 250-year legacy of nomenclatural and taxonomic changes that characterizes the current system; (6) the insufficient taxonomic exploration of large portions of the tree of life; and the need to use a sophisticated terminology for aligning taxonomic entities in order to integrate both (7) single and (8) multiple hierarchies. We briefly such how such integration may proceed based on an initial expert alignment of concept relationship and subsequent use of first-order logic algorithms to maximize consistency, reveal implied relationships, and ultimately merge taxonomies.

In light of the aforementioned obstacles, we suggest that research along the taxonomy/ontology interface should focus on either strictly nomenclatural entities or specialize in ontology-driven methods for producing alignments between multiple taxonomies. We furthermore suggest that the prospects of developing successful ontologies for taxonomy will largely depend on the ability of the taxonomic expert community to present their phylogenies and classifications in a way that is more compatible with ontological reasoning than concurrent practice. Minimally, this means (1) adopting rigorous standards for linking new core taxonomies to relevant peripheral

taxonomies through comprehensive alignments so that their ontological/taxonomic connections are transparent; (2) using lineage-specific ontological standards for phenotype-based accounts of taxa while taking into account the phylogenetic contextuality of phenotypic descriptors; (3) presenting all nomenclatural and taxonomic novelties in an explicit, ontology-compatible format, including intensional and ostensive definitions; and (4) offering comprehensive intensional/ostensive alignments to entities in relevant preceding taxonomies.

Select references

Franz, N.M. 2005. On the lack of good scientific reasons for the growing phylogeny/ classification gap. Cladistics 21: 495-500.

Franz, N.M., R.K. Peet & A.S. Weakley. 2008. On the use of taxonomic concepts in support of biodiversity research and taxonomy; pp. 63-86. In Wheeler, Q.D. (Ed.): The New Taxonomy, Systematics Association Special Volume Series 74. Taylor & Francis, Boca Raton, FL.

Franz, N.M. & R.K. Peet. 2009. Towards a language for mapping relationships among taxonomic concepts. Systematics and Biodiversity 7: 5-20.

Franz, N.M. & D. Thau. 2010. Biological taxonomy and ontology development: scope and limitations. Biodiversity Informatics 7. (accepted pending revision)

Schulz, S., H. Stenzhorn & M. Boeber. 2008. The ontology of biological taxa. Bioinformatics 24: i313–i321.

Thau, D. & B. Ludäscher. 2007. Reasoning about taxonomies in first-order logic. Ecological Informatics 2: 195-209.

Thau, D. 2008. Reasoning about taxonomies and articulations; pp. 11-19. In Ph.D. '08: Proceedings of the 2008 EDBT Ph.D. Workshop. ACM, New York, NY.

Thau, D., S. Bowers & B. Ludäscher. 2008. Merging taxonomies under RCC-5 algebraic articulations; pp. 47-54. In Conference on Information and Knowledge Management. Proceeding of the 2nd International Workshop on Ontologies and Information systems for the Semantic Web. ACM.

Thau, D., S. Bowers & B. Ludäscher. 2009. Merging sets of taxonomically organized data using concept mappings under uncertainty; pp. 1103-1120. In Proceedings of the 8th International Conference on Ontologies, Databases, and the Applications of Semantics (ODBASE 2009). Lecture Notes in Computer Science. OTM 2009.