

April 08, 2014

Manuscript for submission to *PLoS ONE*

Reasoning over Taxonomic Change: Exploring Alignments for the *Perelleschus* Use Case

Nico M. Franz^{1*}, Mingmin Chen², Shizhuo Yu², Shawn Bowers³, Bertram Ludäscher²

1 School of Life Sciences, Arizona State University, Tempe, Arizona, United States of America

2 Department of Computer Science, University of California at Davis, Davis, California, United States of America

3 Department of Computer Science, Gonzaga University, Spokane, Washington, United States of America

* E-mail: nico.franz@asu.edu

Abstract

We explore a novel approach for aligning multiple classifications and phylogenies based on the use of taxonomic concepts, Region Connection Calculus (RCC-5) articulations, and Answer Set Programming (ASP) reasoners. The *Perelleschus* use case of Franz & Cardona-Duque (2013) includes six related taxonomies (ranging from 1936 to 2013), 54 taxonomic concepts, and 75 relevant articulations asserted in that publication. The Euler/X toolkit is used to analyze 13 configurations of six pairwise alignments for this use case under heterogeneous taxonomic constraints and interpretations. The Open Source toolkit ingests two input taxonomies, an initial set of expert-made articulations, and additional taxonomic constraints. The reasoning approach utilizes ASP to iteratively optimize the logical consistency and expressiveness of the input and infer the set of maximally informative relations, leading to visual representations of single or multiple possible world merge taxonomies. In this small use case with 6-53 input concepts per alignment, the information gained through the reasoning process is on average one order of magnitude greater than in the input. The approach offers solutions for longstanding taxonomic provenance and integration challenges; including compatibility with contemporary Linnaean nomenclature; many-to-many name/circumscription relationships across multiple input taxonomies; traditional, ranked taxonomies versus fully bifurcated, informally named phylogenies; monotypic concept lineages with multiple ranks yet congruent extensions; insufficiently specified higher-level entities; undersampled outgroup entities; differentially sampled ingroup entities; taxonomically overlapping entities and merge concepts; strictly ostensive (member-centered) versus intensional (property-centered), or hybrid readings of concept articulations; and localized resolution versus ambiguity in alignments. We review the challenges for a wider implementation of RCC-5/ASP multi-taxonomy alignments with emphasis on human, computational, and resolution constraints. The focus of concept taxonomy is to build sound provenance chains that are amenable to computational representation and reasoning, irrespective of the historical or future stability of name/circumscription relationships for perceived organismal lineages.