A WEB APP AND DATABASE SCHEMA FOR MANAGING AND DISPLAYING TAXONOMIC CONCEPT RELATIONSHIPS

Campbell O. Webb [1,*]
Kimberly J. Cook [2]
Steffi Ickert-Bond [1]

1: University of Alaska Museum of the North
*<cowebb@alaska.edu>   2: Indiana University

Goal

While interest in sharing data about taxonomic concepts is increasing, the process of generating such data is still labor intensive, and generally depends on careful human reading of and inference from taxonomic literature. Some of that work can be reduced by using an **efficient means of recording, managing and visualizing the taxonomic concepts and their relationships**. During work on taxonomic concept mapping for a new Flora of Alaska (see related oral presentation by Cook et al.), we have developed a lightweight, easy-to-install web app and related database schema that tracks publications, names, taxonomic concepts and their relationships. Within the app, the data can be visualized as a graph of relationships (edges) and concepts (nodes); we have found this greatly assists in the human process of logical inference from taxonomic literature.

Taxonomic Concepts

Taxonomic names alone are not sufficient to communicate specific circumscriptions of taxonomic groups of organisms, because the same name may be used in different ways by different taxonomists. The name is anchored only to a single type specimen, and not to an understanding or taxonomic concept (TC) of the taxon. However, one may more precisely specify the particular usage of a name, with a reference to the publication using it (an 'according to'). Our goal for the New Flora of Alaska (https://alaskaflora.org/) is to list not just names, but also the names' various TCs and their interrelationships; because TCs represent *sets* of specimens they can be related to each other using set relationship terms: 'is congruent with', 'includes', 'overlaps with', 'intersects' and 'is disjunct from' (Franz et al., 2008).

Previous tools

Recording taxonomic concept relationships in a database is not a novel idea, and various tools have been created for this purpose. TaxLink (Gradstein et al., 2001) was a MS Windows app developed to store Berlin Model (Berendsohn and al., 2003) taxonomic data. ConceptMapper (Liu et al., 2007) was a more recent Java tool developed out of NCEAS. However neither is currently available.

Database schema

The heart of the app is an open-source relational database (`MariaDB`), with a schema of four key tables: publications, names, taxonomic concepts, and relationships (Fig. 1). Names, TCs and relationships all have a foreign key in the publications table. Data integration is maximized by recording DOIs for each taxonomic treatment where possible, and recording the URIs of names in external names authorities (e.g., IPNI). A Boolean field in the relationships table indicates whether the relationship is to be graphed.
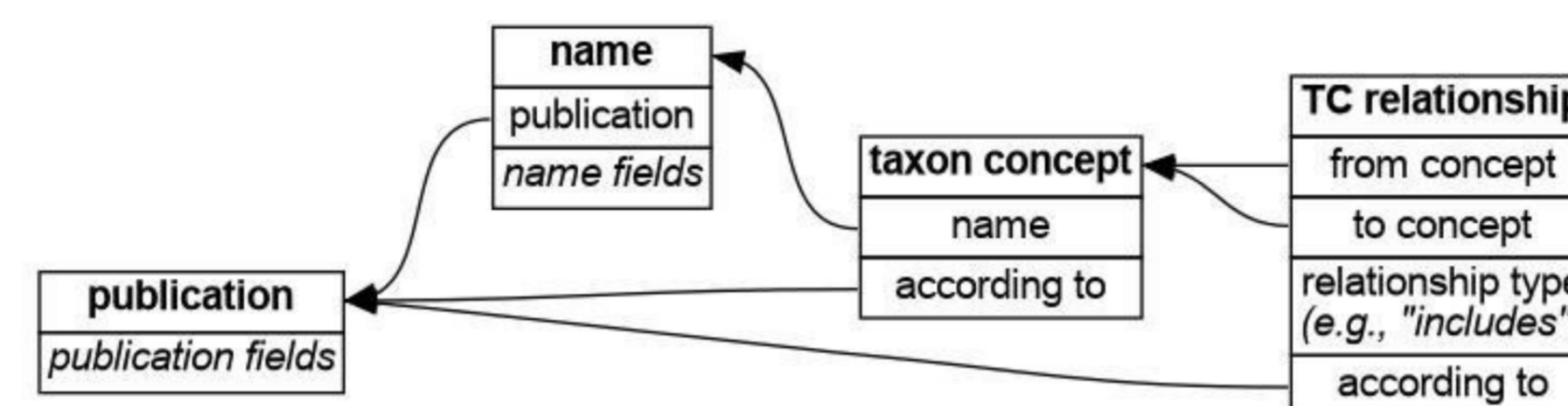

Figure 1: Simplified database schema

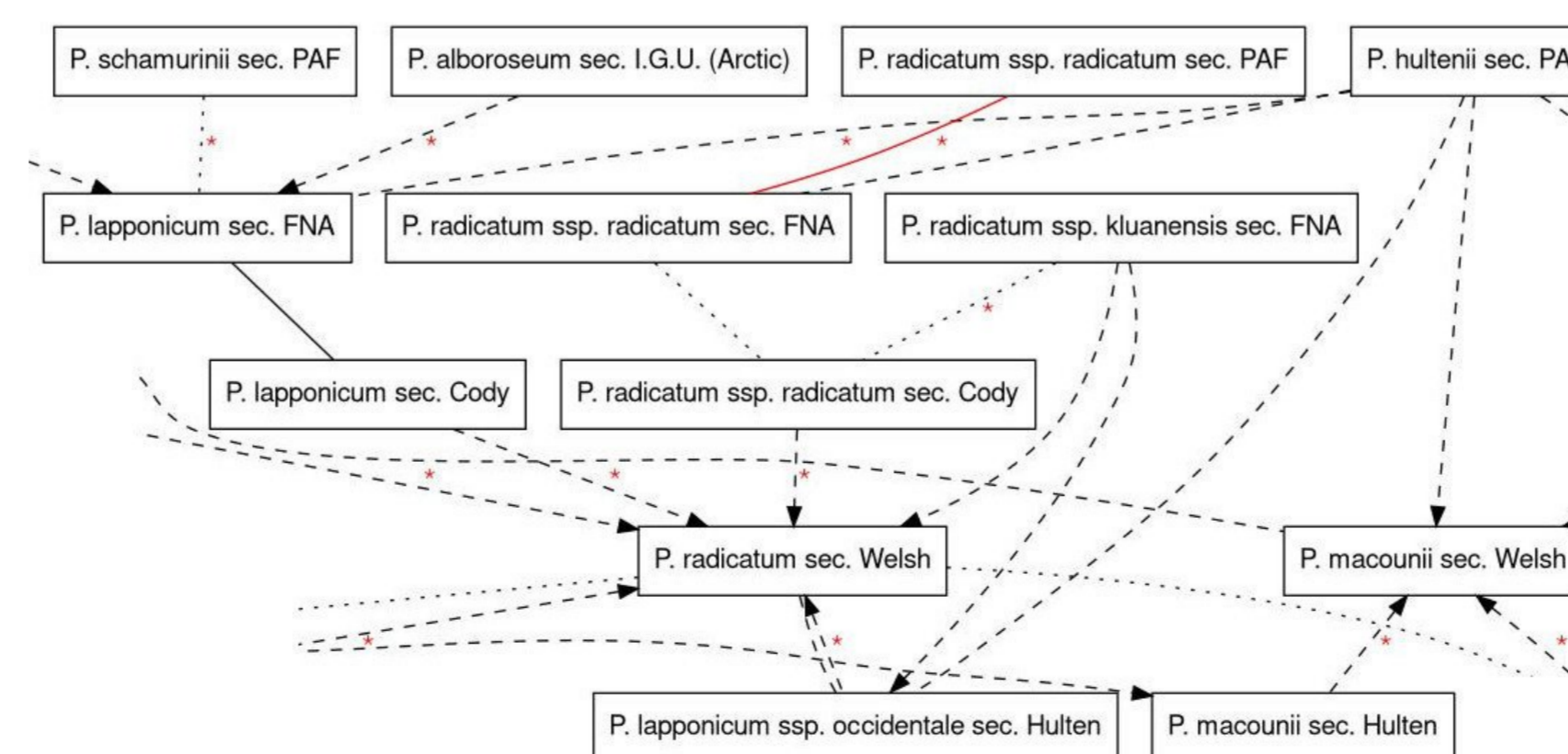**Figure 2:** Web app view of table of TC relationships

Figure 3: In-app Graphviz representation of TC relationships

TCM web app

The web app itself is a single CGI script (called `tcm` and written in Awk) running on an Apache web server, reading from and writing to the database, and generating HTML tables (e.g., Fig 2) and HTML forms. Only a few minutes are needed to install the script on a local Linux laptop or remote web hosting platform.

The basic user workflow is:

1. Enter publication details for a name, a name usage, or as a reference for an inference,
2. Enter a name, referencing its original publication,
3. Create a taxon concept, linking a name with the publication it is used in, and
4. Create a relationship record between two concepts, indicating the type of relationship (e.g., A is 'included in' B).

TC relationships can be viewed as a node-and-edge graph. This visualization helps a user greatly in making logical inferences and developing chains of relationships. This app uses Graphviz (https://www.graphviz.org/) to generate an image file. Different styles of line (edges) indicate the type of TC relationship: solid black for 'congruent with' (RCC-5 symbol: =), dashed black for 'overlaps' (><), dotted black for 'intersects' (∩), dashed black with arrow for 'is included in' (<), and solid red for 'is disjunct from' (|). If newer TCs are always placed before older TCs as relationships are entered then the resultant graph can be read as having a vertical time axis (Fig. 3).

Finally, the taxon concepts and TC relationships can be exported as RDF, using TDWG TCS terms.

Berendsohn, W., and al., 2003. The Berlin Model: a concept-based taxonomic information model. Pages 15–42 in W. G. Berendsohn, editor. MoReTax - Handling Factual Information Linked to Taxonomic Concepts in Biology. Landwirtschaftsverlag, Bonn-Bad Godesberg.

Franz, N., R. K. Peet, and A. Weakley, 2008. On the use of taxonomic concepts in support of biodiversity research and taxonomy. Pages 61–84 in Q. D. Wheeler, editor. The New Taxonomy. Chapman & Hall, London.

Gradstein, S. R., M. Sauer, W. Braun, M. Koperski, and G. Ludwig. 2001. TaxLink, a program for computer-assisted documentation of different circumscriptions of biological taxa. Taxon 50:1075–1084.

Liu, X., R. K. Peet, N. Franz, and L. L. Downey, 2007. ConceptMapper: a tool to link concept relationships across taxonomic classifications. in 5th International Conference on Ecological Informatics, Santa Barbara, CA.

https://github.com/akflora/tcm