The potential of using sets of specimens to handle species concepts: initial thoughts

Ed Baker

November 2015

Contents

1 Problem

Recent increases in the rate of specimen digitisation in natural history museums, combined with persistent identifiers for these specimens allows for robust species concepts defined by sets of specimens. This paper uses set notation to

define operations on groups of specimens that can be considered equivalent to taxonomic and nomenclatural acts.

At the present time insufficient numbers of specimens have been assigned unique identifiers for this solution to be generally practical. It is presented here as an example potential use of unique identifiers, and to encourage discussion as to whether this approach has any merit (it may not).

The mathematics of relational databases is understood in terms of manipulations (relations) of sets. The expression of taxonomic and nomenclatural acts as functions on sets may aid the design of systems that record and track species concepts, which will by necessity need to be stored in databases. This approach is the reverse of Ytow *et al.* (2001) who developed an object-orientated model and later showed it could be made into a relational database.

1.1 Reduction of problem

Species concepts are generally considered to comprise all organisms that are defined by that concept, that is to say that all wild living organisms that match the description and primary types are included. Placing the entirety of a population in to set is impractical, so here sets (and their concepts) are confined to specimens. Specimens are increasingly citable 'objects' and are here considered to include physical specimens, nucleotide sequences, etc. These concept sets may be considered to be a representative subset of the population if philosophically desirable.

This work deals solely with the species concept. Similarly methods (nested sets) could be applied to genus- and family- group concepts if so desired, with modification for the differing method of type-designation at these levels.

This paper does not deal with publications for the purpose of clarity. The association of publications of new species with their type specimens is straightforward, if time consuming for the historical literature. It is hoped that any real world system would have this functionality.

1.2 Nomenclature

This paper does not deal predominantly with nomenclature. The assignment of scientific names to species concepts is a great aid in communicating about organisms. This approach does not change anything relating to the naming of concepts. If anything it may help with algorithmically determining the name of a species concept based on specimens.

2   Introduction

We consider x to be a collection (set) of specimens $(s_1, s_2, s_3, ...)$ including a number of primary type specimens $(t_1, t_2, t_3, ...)$.

$$x = \{s_1, s_2, s_3, ..., t_1, t_2, t_3, ...\}$$

A competent taxonomist takes this set of specimens, and sorts them into piles they consider to represent species. If the group has been recently well studied then the piles may each contain a single primary type.

$$x_1 = \{s_1, ..., t_1\}$$
$$x_2 = \{s_2, ..., t_2\}$$
$$x_3 = \{s_3, ..., t_3\}$$

where $x_n \subset x; x = x_1 \cup x_2 \cup x_3 \cup ...$

Each x_n is a species concept, typified by the specimen t_n when there is a single primary type. No specimen appears in more than one subset.

2.1   Synonymy

If the taxonomist selected set contains multiple primary types this is an indication of synonymy.

$$x_n = \{s_a, s_b, s_c, ..., t_x, t_y\}$$

Define *type cardinality* as the number of valid types in a set.

$$typec(x) = |\{t | t \in x\}|$$

Assuming that t_n are valid types then typification can be resolved by the appropriate nomenclatural code.

1. When $typec(x) = 1$ then the concept can be named by the primary type.

2. When $typec(x) > 1$ then a selection of primary type is needed following the appropriate rules of nomenclature, following the concept of priority.

3. When $typec(x) = 0$ then the concept does not contain a type. The set should be expanded to include an appropriate primary type, or if no appropriate type is available then a specimen from the set should be described as the primary type.

2.1.1 Precedence

Of the valid types the oldest is the one used to formalise the species concept.

Define
$$type(x) = min(\{t_{[date]} | t \in x\})$$

3 Comparison of species concepts

3.1 Identity of species concepts

The identity of species concepts in this scheme occurs when the concepts are sets containing the same specimens. Identity between concepts is potentially not the most useful way of determining if two or more sets are compatible (see Consistency). The mathematical identity of two sets is equivalent to each set being a subset of the other.

$$x = y \iff x \in y \land y \in x$$

3.2 Consistency of species concepts

Test to see if two, non-identical, species concepts are compatible.

Author A: $x_A = \{x_1, x_2, x_3\}$
Author B: $x_B = \{x_1, x_2, x_3, x_4\}$

The species concepts x_A and x_B can be considered to be consistent. An example would be where Author A creates their concept before Author B. Author B later expands Author A's concept with the addition of a new specimen. As Author A has not placed x_4 in any other species concept these two concepts can be considered to be compatible: it is only the fact that Author A was not aware of x_4 that they did not include it in x_A.

Define species concepts are *consistent* when the only specimens not in the intersection of the two concepts are not placed in another concept.

$consitent(x_a, x_B) = \forall x_n \notin x_A \cap x_B$ not in other concepts by author

4 Expanding scope of a species concept

4.1 Identification of expanded scope

We can test that x_B is an expansion of the scope of x_A by checking that x_A is a subset of x_B and that x_b has a larger cardinality than x_A.

$$x_A \subset x_B \land |x_A| < |x_B|$$

4.2 Expansion of scope

4.2.1 Addition of a specimen to a species concept

Adding a specimen to an existing species concept can be achieved through the union of that concept with the new specimen.

$$x_B = x_A \cup x_4$$

4.2.2 Synonymy of two previous species concepts

If two concepts, both previously considered to be valid, are found to be synonyms of each other then a new concept can be created that is the union of these two.

$$x_{new} = x_A \cup x_B$$

5 Reducing scope of a species concept

5.1 Identification of reduced scope

5.2 Reduction in scope

5.2.1 Removal of specimens from a species concept

5.2.2 Splitting of a species concept

6 Remarks

To create a usable system of species concepts based on specimens a robust citation system for concepts needs to be created, with persistent identifiers. Persistent identifiers would allow the easy expansion and changing of concepts as new specimens are collected or digitised. Life Sciences Identifiers (LSIDs) or Digital Object Identifiers (DOIs) could be used for this purpose.