Bionomina, **6**: 52–56 (2013)
www.mapress.com/bionomina/

Correspondence

ISSN 1179-7649 (print edition)
BIONOMINA
ISSN 1179-7657 (online edition)

http://dx.doi.org/10.11646/bionomina.6.1.4
http://zoobank.org/urn:lsid:zoobank.org:pub:8AC3C682-2767-4A8D-A1D9-9A015D5B5E03

The misleading use of the terms parent, child, ancestor and descendant in databases dealing with biological evolution and taxonomy

Alain DUBOIS & Mohamed BERKANI

Reptiles & Amphibiens, UMR 7205 OSEB, Département de Systématique & Evolution, Muséum national d'Histoire naturelle, CP 30, 25 rue Cuvier, 75005 Paris, France. <adubois@mnhn.fr>, <mohamed.berkani@gmail.com>.

The rapid burst of electronic databases in the recent decades has led to a boom in specialised terminology meant to describe the structure and way of functioning of these databases. Some of this terminology was borrowed from "common language" vocabulary but used in a specialized technical language. This decision to use existing terms in a new sense, rather than coining new terms for the new concepts, was a questionable one as in some cases this terminology may be misleading, at least in certain contexts. This is the case, in our opinion, with the use of terms expressing genealogical relationships to express hierarchical taxonomic relationships in databases.

There are various kinds of hierarchical relationships but a basic distinction is between inclusive or "whole-part" relationships and genealogical relationships. In an inclusive relationship, B is *included* in A, whereas in a genealogical relationship B was *produced* by A, two situations which have little in common.

In graph theory and other computer science contexts, in a directed graph A→B, A was soon designated as the "*parent*" and B as the "*child*". These terms are used in various contexts to indicate hierarchical relationships between items. After several decades of such a practice, many users feel that the meaning of these terms is "obvious" and clear to all readers. However, the indiscriminate use of the terms "parent" and "child" in all hierarchical databases tends to describe both kinds of hierarchical relationships as genealogical.

In hierarchical database models, data are organized following a tree-like structure. In this structure, each entity is related to several subordinate entities, which is known as one-to-many relationships (Kamfonas 1992; Groff & Weinberg 1999; Celko 2004). The descriptive term commonly used for such a structure using *1:N* mappings is "parent-child relationships", according to which each parent can have many children, but each child has only one parent. However, to any biologist, the latter sentence doubtless looks strange, because, in bisexual species at least, each child has two parents! But even more misleading is the fact that parent-child relationships (in the ordinary meaning) are genealogical relationships, which is not the case with hierarchical relationships as expressed in databases, even dealing with phylogeny (genealogy) or taxonomy (a conventional way of expressing genealogy). Besides taxonomic hierarchies, examples that can be mentioned include taxonomies of products, customer segmentations, organizational and location hierarchies (countries, states, departments, etc.), component breakdowns, etc.

A principle of "pseudo-inheritance" is used and well established in object-oriented databases (Garcia-Molina 2008). In such a context, each "piece" of data, or data object, is linked to subordinate data objects through attributes (or characteristics) and this results in a hierarchy of data. The analogy with phylogenetic systematics is obvious here, but it is misleading. It would appear then important in our opinion to use specific terms in databases dealing with taxonomic hierarchies in order to avoid semantic confusion.

Two pairs of terms are often used to express relationships of superordination and subordination in databases: *parent* and *child* refer to immediate such relations, whereas *ancestor* and *descendant* designate more remote relations (Su *et al.* 2006; Singh 2009). Both pairs of terms are misleading in evolutionary and taxonomic databases because they clearly evoke genealogical relationships although they do not refer to such