Modeling data relationships with named sets

Mark Burgin[1] and H. Paul Zellweger[2]

[1] UCLA, Los Angeles, California, USA
[2] ArborWay Labs, Rochester, Minnesota, USA

Data are representations and containers (carriers) of information while data with their relationships form data structures. That is why modeling data relationships is important for organization and optimization of information processes.

Data have different structures and data processing in general and data mining in particular depends on these structures. The basic data structures include: Boolean values, characters, integers, fixed-precision number values, floating-point number values, arrays, records, lists, streams, sets, multisets, stacks, queues, and graphs to mention just the most important of them. Here in addition to these data structures, we consider named sets and chains of named sets as the fundamental data structures.

So, the question is: Why named sets are really essential and what is so specific about named sets?

First, it is proved that any mathematical structure is a named set or is built of named sets (Burgin, 2011). For instance, functions, relations, variables, graphs, multigraphs, and morphisms (arrows) in categories are special cases of named sets. Ordinary sets are also specific named sets, namely, they are singlenamed sets because all elements in a set with the name, say Q, have the common name “an element of the set Q” (Burgin, 2011). In essence, named set is the most fundamental structure in mathematics (Burgin, 2011).

Second, named sets have been extensively utilized in databases and knowledge bases. As relations are a special case of named sets, all relational databases store named sets and work with them (Date, 2004). Named set chains are key structures in temporal databases (Snodgrass and Jensen, 1999; Burgin, 2008). Named sets have been constructively utilized for data visualization and information retrieval in databases (Burgin and Zellweger, 2005; Zellweger, 2017) and database management (Ivanova, 2015).

Third, it is proved that named set (also called fundamental triad) is the most basic structure in nature (Burgin, 2011). As a consequence, named sets become ubiquitous in modeling natural systems.

So, what is a named set?

A *basic named set*, also called a *fundamental triad*, is a triad **D** = (*X*, *f*, *Z*), in which the components *X* and *Z* are two objects and *f* is a correspondence (e.g., a binary relation) between *X* and *Z*. A *bidirectional named set* has the same structure but in it, the correspondence (e.g., a binary relation) *f* goes in two directions.

An important peculiarity of utilization of named sets in databases is that algorithms in general and software systems, in particular, for operation with data are also specific named sets and systems of named sets. Namely, they are algorithmic named sets and their systems, i.e., such named sets in which the relation *f* is an algorithm or a program.

Manipulation with data demands various operations and in the case of using named sets for data representation, a variety of operations, such as mappings of different kinds, union, intersection, difference, renaming, naming, interpreting, and reinterpreting, is provided by the theory of named sets (Burgin, 2011).

We consider these operations and their application in database management, data modeling, data organization, data mining, data transformation, and information search.

Burgin, M. (2008) Structural Organization of Temporal Databases. In: *Proceedings of the 17th International Conference on Software Engineering and Data Engineering* (SEDE-2008), ISCA, Los Angeles, California, pp. 68-73

Burgin, M. *Theory of Named Sets*, Mathematics Research Developments, Nova Science, New York, 2011

Burgin, M. and Zellweger, P. (2005) A Unified Approach to Data Representation. In: *Proceedings of the* 2005 *International Conference on Foundations of Computer Science*, CSREA Press, Las Vegas, pp. 3-9

Date, C. J. *An Introduction to Database Systems*, Addison Wesley, Boston/San Francisco/New York, 2004

Ivanova, K. (2015) Storing and processing accounting information based on collect/report paradigm, International Journal *Information Content and Processing*, **2**(2), 103-114

Snodgrass, R.T. and Jensen, C.S. *Developing Time-Oriented Database Applications in SQL*, Morgan Kaufmann, 1999

Zellweger, H.P. (2017) A Decision Tree Interface Based on Predicate Calculus, 21st *International Conference on Information Visualisation* (*IV* 2017), pp. 188-193