# Perspective for Database Preservation

D

**Elvira Immacolata Locuratolo**
*ISTI Consiglio Nazionale delle Ricerche, Italy*

**Jari Juhani Palomäki**
*Department of Information Technology, Tampere University of Technology/Pori, Finland*

## INTRODUCTION

The approaches of database preservation described in (Buneman & Christophides, 2007) do not consider the need to *preserve database concepts.* In *database* models a label denotes a class/a set of objects/ and connotes a corresponding concept; as an example, the label "department" denotes the class "department" and connotes the concept of *department*. Concepts and classes of objects are modeled in a similar way. In *concept theory* (Kauppi, 1997; Palomäki, 1994), concepts are introduced at the *intensional/concept level*, which is the level of human thinking, whereas classes/ set of objects are introduced at the *extensional/ set-theoretical level*, which is the level of *computer science*. A distinction is made between concepts and classes of objects.

At the intensional level, concepts are modelled in terms of information contents of concepts, and concept structures are defined through an *is-in* relation, called *intensional containment relation,* (Palomäki & Kangassalo, 2012). Correspondently, at the extensional level, an *is-a* relationship is exploited to model database specialization hierarchies. Oriented acyclic graphs of classes in *is-a* relationships supported by Semantic Data Models (Cardenas & McLeod, 1990) or Enhanced Entity-Relationships models (Elmasri & Navathe, 2000), called *Conceptual graphs,* are similar to oriented acyclic graphs of classes in $is\text{-}a_o$ relationships supported by object database systems, called *Logical object graphs* (Locuratolo, 2013). The following properties hold:

- **Is-a relationship:** Each object instance can belong to any class of a conceptual graph. This property enhances *flexibility* in modeling changes occurring in the real life.
- **Is-a$_o$ relationship:** Each object instance belongs to one and only one class of the logical object graph. This property enhances *efficiency* in accessing and storing objects.

The difference between an *is-in* relation and an *is-a/* $is\text{-}a_o$ relationship can be explained as follows: in a conceptual model using an *is-a* relationship as its taxonomical link, objects to be modeled are presupposed to exist; vice-versa, in a concept structure using an *is-in* relation, the existence of objects falling under a concept is not presupposed. In a conceptual model using an *is-a* relationship, objects can be added to classes or removed from them while the concepts of the corresponding classes remain unchanged.

Original research based on concept theory has recently been proposed for database preservation (Locuratolo & Palomäki, 2013). By *ontology for database preservation* we will refer a structure where all and only the concepts related with a conceptual graph are retained; the logical coherence among these concepts is maintained and a mapping from the concept structure to a corresponding logical object graph is provided (Locuratolo & Palomäki, 2013). This concept structure is characterized by the following properties:

- Results in leaves which are incompatible concepts.
- Encloses all and only the concepts related to an initial concept structure.
- Encloses all and only the intensional inclusion relations among concepts.
- Is mapped to a logical object graph.

In ontology for database preservation, concepts are mapped to classes of objects, but separated from them, thus, in addition to the database preservation results

DOI: 10.4018/978-1-4666-5888-2.ch179

achieved in (Buneman & Christophides, 2007), the further result of preserving all and only the concepts included in oriented acyclic graphs of object classes is considered.

In this article, the main results achieved in (Locuratolo & Palomäki, 2013) are outlined and, taking into considerations that databases change over time and include schemas and integrity constraints which can be interesting for current and future interpretations, a possible extension of the article is proposed. Ontology for database preservation has been mapped to statically object database schemas; however, it can be effective for dynamic situations. Each class of the logical database model can be understood as a box with a label. Each object of the considered real-world domain belongs to one and only one box. The boxes are potentialities and according to time and real world situation can be either empty or occupied. Empty boxes at a given time can then be occupied and vice versa. The article evidences that logical/object graphs can be extended with basic operations able to add objects to classes or remove objects from the classes.

## BACKGROUND

Concept theory, developed in philosophy, puts into evidence the distinction between the intentional and the extensional aspects of a concept. In Figure 1, the former aspect, represented by a cloud, is referred to the information content of the concept, whereas the latter aspect, represented by an oval, is referred to the set of objects which fall under the concept. It is correct to go from the intensional to the extensional level of a concept, but not vice-versa. This is because a set of objects can falls under many different concepts, as an example a set of dogs can fall under the concepts of *dog*, *mammal*, *and animal.*

An *is-in* relation, called *intensional containment relation,* enables to make concept structures. More formally, let us consider concepts $u$ and $v$ and the respective extensions set ($u$) and set ($v$), also denoted by $U$ and $V$. The intensional containment relation between the concepts $u$ and $v$ is converse to the extensional set-theoretical inclusion between their extensions. Thus, we get,

$$u \geq v$$
$$\downarrow$$
$$< U \subseteq V >$$

As an example, since the information content of the concept *apple* is greater than the information content of the concept *fruit*, the concept *apple* contains intensionally the concept of *fruit*. Formally, from *apple* ≥ *fruit*, it follows set (apples) $\subseteq$ set (fruits).

The term *ontology* has nowadays at least three different meanings: in its original philosophical significance, ontology is part of metaphysics concerning the Universe of Discourse and the building blocks of the world (Palomäki, 2009). In Information systems, ontology may means a conceptual model/a conceptualized knowledge representation or a vocabulary/dictionary containing the basic terms used in a conceptual language. In this article, the term *ontology* is justified by the fact that the concept structure is mapped to database models. The term *preservation* is justified by the fact that all and only the explicit and implicit database concepts, as well as all and only the intensional links for their logical coherence are retained.

*Figure 1. From concept to objects (© Copyright 2013, IGI Global. Used with permission)*