Abstract: The article aims at establishing a logical approach to class-based data modeling. After a discussion on class-based formalisms for data modeling, we introduce a family of logics, called Description Logics, which stem from research on Knowledge Representation in Artificial Intelligence. The logics of this family are particularly well suited for specifying data classes and relationships among classes, and are equipped with both formal semantics and inference mechanisms. We demonstrate that several popular data modeling formalisms, including the Entity-Relationship Model, and the most common variants of object-oriented data models, can be expressed in terms of specific logics of the family. For this purpose we use a unifying Description Logic, which incorporates all the features needed for the logical reformulation of the data models used in the various contexts. We also discuss the problem of devising reasoning procedures for the unifying formalism, and show that they provide valuable supports for several important data modeling activities.
8.1 INTRODUCTION

Data modeling is the activity of specifying the structure of the data to be managed within an application. In the last two decades, data modeling has been the subject of a large body of work in several areas, including Databases, Information Systems, Software Engineering, and Knowledge Representation. In particular, recent approaches to conceptual data modeling advocate the use of abstract formalisms for describing data, mostly based on the notion of class [HK87]. In this paper, we concentrate on such class-based formalisms for data modeling, with the aim of demonstrating that they can be profitably reconstructed within a logical framework. We argue that the reasoning techniques available in the logical framework provide valuable support for the data modeling activity.

Generally speaking, a class denotes a subset of the domain of discourse, and a class-based representation formalism allows one to express several kinds of relationships and constraints (e.g. subclass constraints) holding among classes [MM92]. Moreover, class-based formalisms aim at taking advantage of the class structure in order to provide various information, such as whether an element belongs to a class, whether a class is a subclass of another class, and more generally, whether a given constraint holds among a given set of classes.

Two main families of class-based formalisms for data modeling are addressed in this paper. The first one originates in the field of databases and in particular from the work on semantic data models (see e.g. [HK87]). The second one arises from the work on types in programming languages and object-oriented systems (see e.g. [KL89]).

In the past, there have been several attempts to establish relationships among class-based formalisms used in knowledge representation (e.g. semantic networks and frames [Leh92; Sow91]) and the above two families of class-based formalisms. One significant aspect of this work is the identification of reasoning problems, where one can take advantage of techniques for reasoning on hierarchical structures that have been developed in different areas.

The relationship between frame-based languages and types has been addressed in [BHR90; LNS91; Bor92], while in [BS92; PSS92; Bor95; ACS96] frame-based languages are used to enrich the deductive capabilities of data models. The analysis of the above cited works makes it clear that, although a number of steps have been accomplished, several difficulties arise in identifying a common framework, which is expressive enough to capture the essential features of various class-based formalisms, while still providing techniques for the associated reasoning problems. Other formalisms have been recently proposed with the aim of integrating the object-oriented and the logic programming paradigms. A notable example of this effort is F-Logic [KLW95], which pro-