Preface

About This Book

The work presented in this book came about after we recognized that ill-designed semistructured databases can lead to update anomalies, and there is a strong need for algorithms and tools to help users design storage structures for semistructured data. We have been publishing papers in the design of databases for semistructured data since 1999, and believe that after a number of attempts we have defined a data model that captures the necessary semantics for representing the semantics that are necessary in the design of good semistructured databases.

This book describes a process that initially takes a hardline approach against redundant data, and then relaxes the approach for gains in query performance. The book is suited to both researchers and practitioners in the field of semistructured database design.

Some of the material in this book has been published at international conferences. The material in Chapter 5 was originally based on work presented in [Wu et al., 2001a] and Chapter 6 was originally based on [Chen et al., 2002]. The material in Chapter 3 was published as a technical report at the National University of Singapore [Dobbie et al., 2000].

Use of the Book

The target audience of this book is practitioners who design semistructured data file organizations or semistructured databases, researchers who work in the area of semistructured data organization, and students with an interest in the design of storage organizations for semistructured data. The material is as relevant for file organizations as it is for databases since inconsistencies can also exist in data files.