INTELLIGENT DATABASES

Object-Oriented, Deductive Hypermedia Technologies

Kamran Parsaye
Mark Chignell
Setrag Khoshafian
Harry Wong

INTELLIGENT DATABASES

Object-Oriented, Deductive Hypermedia Technologies

Kamran Parsaye

Mark Chignell

Setrag Khoshafian

Harry Wong

WILEY

John Wiley & Sons, Inc.

New York • Chichester • Brisbane • Toronto • Singapore

Library of Congress Cataloging-in-Publication Data:

Intelligent databases.

Bibliography: p.
1. Data base management. 2. Expert systems (Computer science) I. Parsaye, Kamran.
QA76.9.D31548 1989 005.74 88-33948
ISBN 0-471-50346-0
ISBN 0-471-50345-2 (pbk.)

PREFACE

Anthropologists have argued over whether humans became smart because they walked upright, because they had to kill other animals without having the powerful jaws and claws needed for the job, or because they learned how to communicate using a complex form of articulation and vocalization. However, lurking behind our definitions of humans as tool users and language users is the prior concept of humans as information users. Information and its use allow us to build and use tools and to express ourselves in language.

It is often remarked that we are living in the information age, yet we are struck by the fact that many existing technologies (1) are under-utilized and (2) tend to be treated in isolation. In the past few years, major advances have been made in several fields of critical relevance to information technology. These advances have been significant, but fragmented. Bringing these separate technical leaps into a unified framework will benefit people involved in all of the contributing disciplines by facilitating cross-disciplinary activity and by pointing out new avenues for both research and application.

This book is about the emerging intelligent database technology that will have a dramatic impact on the way we think and work. Intelligent databases will greatly expand our capabilities as information users. The phenomenon of the shrinking world is reaching its natural conclusion (i.e., McLuhan's global village), and intelligent databases are the means by which we may store and retrieve the information and expertise of this global village.

We may have to discard some familiar intellectual habits and presuppositions along the way. For instance, we are accustomed to thinking about information as passive,

residing in books and documents. Similarly, we often use libraries because we want to get at the information contained in their books, and we use databases as retrieval systems to find "answers." These "answers" are not always predefined in the system and may have to be created "on the fly," depending on the particular characteristics of the user and his or her current interests. Intelligent databases will change how we do research, how we look for ideas, how we make decisions, and how knowledge is transmitted.

In writing this book, we began with the premise that the database is at the heart of any information system and then sought to combine databases with a variety of tools that would make them more intelligent. In so doing, we have developed the model for intelligent databases that is described in this book.

The intelligent database model is based on five information technologies:

- Databases
- Object-Oriented Programming
- Expert Systems
- Hypermedia
- Text Management

Each of the four authors of this book has expertise in one or more of these areas. The book represents a balance between the need to provide sufficient background in each of the five technologies while also including new material on intelligent databases.

Databases and expert systems have been extensively discussed elsewhere, but the material in Chapters 2 and 4 distills the key concepts, thus allowing this book to be a self-contained treatment of intelligent databases. Although many texts on databases and expert systems are available, none relates them to the other technologies discussed here.

We also felt the need to provide clear expositions of object-oriented programming, hypertext, and text management. In addition to covering the major points made elsewhere in a number of scattered texts and journals, our treatment of these topics extends them where necessary so as to provide the tools for constructing intelligent databases.

The eight chapters of the book provide a complete and self-contained exposition of intelligent databases. Chapter 1 begins by defining intelligent databases and places them in the context of evolving information technology. The next five chapters present the key concepts for each of the five contributing technologies. In its discussion of databases, Chapter 2 focuses on relational and semantic data models in particular. Chapter 3 discusses the major issues of object orientation, including encapsulation,

inheritance, and object identity; it also reviews existing object-oriented languages and databases. Chapter 4 introduces the expert-system building blocks of knowledge representation and inference, and describes the process of building expert systems. The chapter concludes with a discussion of how expert systems and databases may be combined to form deductive retrieval systems.

Chapter 5 provides a thorough introduction to hypertext, relating it to the knowledge representation methods described in Chapter 4 and showing how hypertext networks may be linked to conventional databases. Chapter 6 reviews the extensive topic of text management, focusing on these elements most relevant to the construction of intelligent databases. Chapter 7 elaborates on the themes developed in Chapters 1 through 6. A formal model for intelligent databases is described that incorporates the five technologies described earlier. The main points of the earlier chapters are summarized in Chapter 8.

While writing this book, we constructed a prototype of the Fortune Finder intelligent database described in Chapter 7, thus demonstrating the viability of our intelligent database model. The material presented in this book represents the fruits of research and development that have been carried out by universities and industry over the past few years. In particular, we would like to acknowledge the contribution of Project Jefferson at the University of Southern California in demonstrating how hypermedia can be used for information management.

We write the book with several types of readers in mind. Those who want to master the topic of intelligent databases in its entirety should read all the chapters of the book in the order they are written. Database specialists who want to update their technical knowledge should skim Chapter 2 before focusing on Chapters 3, 4, 5, and 7. Administrators and managers should begin by reading Chapters 1 and 6, followed by Chapter 7. They should then return to the more technical material provided in such chapters as 4 and 5. Readers who have a good grasp on topics such as object orientation and expert systems, but who want to know what intelligent databases are, should concentrate, at least initially, on Chapters 1 and 7. Conversely, the book can also be used to review individual topics such as expert systems or hypertext. We see it being used in this way, for instance, in graduate-level college seminars on information technology or expert system applications.

All of the techniques described in this book can be implemented on a microcomputer such as a Macintosh II or a 286 or 386 class machine, and most of the techniques have already been implemented. In many ways, this books describes what is needed to build a shell for intelligent databases, much like the shells that are used to construct expert systems. In the same way that knowledge engineering of expert systems can be a difficult task, even when one has the expert system shell, so our experience has shown us that intelligent database engineering is a difficult task. However, we have no doubt that the benefits of intelligent databases will lead to them becoming a firmly established information technology.

We would like to thank those people who have helped us in preparing this book: Jenny Ghielmetti, Sandra Chignell, Silva Khoshafian, Linda Wong, Bob Blum, Bob Fraser, Joe Goguen, Lee Jaffe, Rei-Chi Lee, Diana Lin, Phil Smith, David Thompson and Gio Wiederhold. In preparing this book, we also benefited from discussion and interactions with a wide variety of colleagues, including the members of the Project Jefferson Team at USC.

KAMRAN PARSAYE
MARK CHIGNELL
SETRAG KHOSHAFIAN
HARRY WONG

Los Angeles and Walnut Creek, California
January, 1989

CONTENTS

5.
HYPERMEDIA 223

1 INTRODUCTION

1.1 INTRODUCTION

This book is about intelligent databases, which can be loosely defined as *databases that manage information in a natural way, making that information easy to store, access, and use.* We use the term *information* rather than *data* here since the intelligent databases discussed in this book deal not only with traditional applications such as inventory management, but also with knowledge bases, automatic discovery systems, textual data, images, and so on.

The concepts of "ease" and "naturalness" are so well accepted that it is hard to argue with their worth. However, while the definition above is rather general, in Section 1.5 we provide detailed criteria for what makes a database intelligent. These criteria establish that by using intelligent databases users can perform tasks involving large amounts of information that otherwise could not possibly be performed. This relies on defining three levels of database intelligence.

i. Intelligence of the high-level tools provided by the database.
ii. Intelligence at the user-interface level.
iii. Intelligence of the underlying database engine.

These are illustrated in Figure 1.1.

Intelligent databases represent a new technology for information management that has evolved as a result of the integration of traditional approaches to databases with more recent fields such as

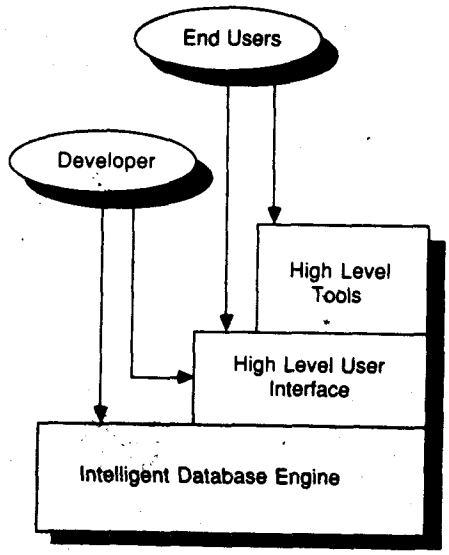

Figure 1.1 Three levels of database intelligence.

- Object-oriented programming
- Expert systems
- Hypermedia
- Online information retrieval

The merger of these technologies is illustrated in Figure 1.2.

Intelligent databases thus represent the evolution of a number of distinct paths of technological development. As Figure 1.3 suggests, traditional database systems were developed in the mainframe environments of the early 1960s. They were typically applied to numeric and record-based data (e.g., employee records or parts inventories) and soon began to spread throughout industry. In the early 1970s, online information systems became available as a specialization of database technology, and they began dealing with textual databases. In the mid-1970s the technology of expert systems developed and had become prominent by the late 1970s. Similarly, in the late 1970s and early 1980s, techniques relating to *object-oriented programming* grew out of work in software engineering, user interfaces, and high-level programming languages. Finally, in the mid-1980s practical hypermedia systems, which dealt with information in a variety of forms such as text, images, and sound, became available.

Until recently, these technologies were treated in isolation, with each technology being only *weakly linked* to others. For instance, in the early days expert systems relied on little more than file-transfer protocols to gather data from databases. This was partly