KNOWLEDGE ACQUISITION FOR EXPERT SYSTEMS

A PRACTICAL HANDBOOK

Edited by
ALISON L. KIDD

Knowledge Acquisition for Expert Systems

A Practical Handbook

Edited by

ALISON L. KIDD

Hewlett Packard Laboratories
Bristol, England

Plenum Press • New York and London

Library of Congress Cataloging in Publication Data

Knowledge acquisition for expert systems.

Bibliography: p.
Includes index.
1. Expert systems (Computer science). 2. System design. I. Kidd, Alison L.
QA76.76.E95K554 1987 006.3'3 87-12274
ISBN 0-306-42454-1

A Division of Plenum Publishing Corporation
233 Spring Street, New York, N.Y. 10013

Printed in the United States of America

Knowledge Acquisition for Expert Systems

A Practical Handbook

Contributors

Joost Breuker, Department of Social Science Informatics, University of Amsterdam, 1016 BS Amsterdam, The Netherlands

John Fox, Biomedical Computing Unit, Imperial Cancer Research Fund, London WC2A 3PX, England

Brian R. Gaines, Department of Computer Science, University of Calgary, Calgary, Alberta, Canada T2N 1N4

John G. Gammack, MRC Applied Psychology Unit, Cambridge CB2 2EF, England

Melvyn F. Greaves, The Leukaemia Research Fund Centre, Institute for Cancer Research, London SW3 6JB, England

Anna Hart, Lancashire Polytechnic, Preston PR1 2TQ, England

Leslie Johnson, Department of Computer Science, Brunel University, Uxbridge, Middlesex UB8 3PH, England

Nancy E. Johnson, Department of Computer Science, Brunel University, Uxbridge, Middlesex UB8 3PH, England

Jerome P. Kassirer, Department of Medicine, Tufts University School of Medicine, Boston, Massachusetts 02111

Alison L. Kidd, Hewlett Packard Laboratories, Stoke Gifford, Bristol BS12 6QZ, England

Benjamin Kuipers, Department of Computer Sciences, University of Texas at Austin, Austin, Texas 78712

Christopher D. Myers, Biomedical Computing Unit, Imperial Cancer Research Fund, London WC2A 3PX, England, and Department of Microbiology, University of Texas, Dallas, Texas 75235

Susan Pegram, The Leukaemia Research Fund Centre, Institute for Cancer Research, London SW3 6JB, England

Mildred L. G. Shaw, Department of Computer Science, University of Calgary, Calgary, Alberta, Canada T2N 1N4

Bob Wielinga, Department of Social Science Informatics, University of Amsterdam, 1016 BS Amsterdam, The Netherlands

Preface

Building an expert system involves eliciting, analyzing, and interpreting the knowledge that a human expert uses when solving problems. Experience has shown that this process of "knowledge acquisition" is both difficult and time consuming and is often a major bottleneck in the production of expert systems. Unfortunately, an adequate theoretical basis for knowledge acquisition has not yet been established. This requires a classification of knowledge domains and problem-solving tasks and an improved understanding of the relationship between knowledge structures in human and machine.

In the meantime, expert system builders need access to information about the techniques currently being employed and their effectiveness in different applications. The aim of this book, therefore, is to draw on the experience of AI scientists, cognitive psychologists, and knowledge engineers in discussing particular acquisition techniques and providing practical advice on their application. Each chapter provides a detailed description of a particular technique or methodology applied within a selected task domain. The relative strengths and weaknesses of the technique are summarized at the end of each chapter with some suggested guidelines for its use.

We hope that this book will not only serve as a practical handbook for expert system builders, but also be of interest to AI and cognitive scientists who are seeking to develop a theory of knowledge acquisition for expert systems.

Alison L. Kidd

Bristol, England

Contents

CHAPTER 1

KNOWLEDGE ACQUISITION—AN INTRODUCTORY FRAMEWORK

Alison L. Kidd

1. Introduction 1
2. The Knowledge Acquisition Process: Mining Is a Misguided Analogy 2
3. Knowledge Domains: Capturing or Creating a Language 3
4. Tasks: Defining the Problem That the Expert System is Designed to Solve 5
5. System Modality: Defining the Set of Tasks 8
6. Users: Acquiring Their Knowledge 9
7. An Alternative Approach: Identifying Weaknesses in Human Reasoning 10
8. Concluding Guidelines for Knowledge Engineers 11
9. Brief Overview of Chapters 12
10. References 15

CHAPTER 2

USE OF MODELS IN THE INTERPRETATION OF VERBAL DATA

Joost Breuker and Bob Wielinga

1. Problems and Solutions in Building Expert Systems 17
2. A Methodology for Knowledge Acquisition 20
3. Interpretation of Data on Expertise 24

3.1. Domains and Tasks 24
3.2. Levels of Interpretation 26
3.3. Primitives of the Epistemological Level 28
3.4. Interpretation Models 30
3.5. Use of Interpretation Models 34
3.6. Supporting Knowledge Acquisition: KADS and ROGET 38
4. Concluding Remarks 39
5. Guidelines Summary 40
6. References 42

CHAPTER 3

KNOWLEDGE ACQUISITION BY ANALYSIS OF VERBATIM PROTOCOLS

Benjamin Kuipers and Jerome P. Kassirer

1. Introduction 45
2. Design of the Experiment 47
3. The Nephrotic Syndrome 50
4. Analysis of the Transcript 51
5. The Domain Model—Structural Description 55
6. Qualitative Simulation in the Explanation 58
7. The Domain Model—Qualitative Description of State 61
8. The Domain Model—Qualitative Simulation 63
9. Conclusion 65
10. Postscript 68
    10.1. Computer Implementation 68
    10.2. Types of Analysis 68
    10.3. Guidelines Summary 69
11. References 70

CHAPTER 4

A SYSTEMATIC STUDY OF KNOWLEDGE BASE REFINEMENT IN THE DIAGNOSIS OF LEUKEMIA

John Fox, Christopher D. Myers, Melvyn F. Greaves, and Susan Pegram

1. Introduction 73
2. Leukemia Diagnosis 74
3. Knowledge Elicitation and Expert System Development 75
    3.1. Expert System Package 76
    3.2. System Development 78
    3.3. Knowledge Base Development 78

4. System Performance 80
5. Analysis of Disagreements between System and Expert 83
6. Discussion 86
7. Guidelines Summary 87
7.1. Limits of the Methods Proposed 88
8. References 89

CHAPTER 5

KNOWLEDGE ELICITATION INVOLVING TEACHBACK INTERVIEWING

Leslie Johnson and Nancy E. Johnson

1. The Knowledge Elicitation Process 91
1.1. Theoretical Stance 92
1.2. Conversation Theory 93
2. Case Studies 95
2.1. Why Use Teachback Interviewing? 95
2.2. Arithmetic Study 96
2.3. VLSI Design Study 99
3. Mediating Representation—SGN 101
4. Discussion 103
4.1. Teachback as a Complete Methodology 104
4.2. Teachback Interviewing as a Viable Technique 105
5. Guidelines Summary 106
5.1. Strengths 106
5.2. Weaknesses 107
5.3. Rules of Thumb 107
6. References 108

CHAPTER 6

AN INTERACTIVE KNOWLEDGE-ELICITATION TECHNIQUE USING PERSONAL CONSTRUCT TECHNOLOGY

Mildred L. G. Shaw and Brian R. Gaines

1. Knowledge Engineering 109
2. Personal Construct Psychology 111
3. What Is a Repertory Grid? 112
3.1. Eliciting Constructs 113
4. Techniques for Repertory Grid Elicitation and Analysis 115
4.1. Repertory Grid Analysis 116
4.2. Analysis of a Single Grid 116

4.3. Analysis of a Pair of Grids ... 117
4.4. Analysis of a Group of Grids ... 118
5. Soft Systems Analysis ... 118
5.1. The Significance of Different Perspectives ... 119
5.2. Techniques of Soft Systems Analysis ... 119
6. PLANET: A Computer-Based Knowledge-Engineering System ... 122
7. PEGASUS in Action ... 123
8. ENTAIL in Action ... 126
9. Validation ... 130
10. Guidelines Summary ... 133
11. References ... 134

CHAPTER 7

DIFFERENT TECHNIQUES AND DIFFERENT ASPECTS ON DECLARATIVE KNOWLEDGE

John G. Gammack

1. Introduction ... 137
2. Methods ... 138
2.1. Concept Elicitation ... 139
2.2. Structure Elicitation ... 144
2.3. Structure Representation ... 151
2.4. Developing the Representation ... 154
3. Using the Knowledge Base ... 156
4. Future Research ... 158
5. Guidelines Summary ... 159
6. References ... 162

CHAPTER 8

ROLE OF INDUCTION IN KNOWLEDGE ELICITATION

Anna Hart

1. Introduction ... 165
2. Induction ... 166
2.1. General Principles ... 166
2.2. The ID3 Algorithm ... 167
3. A Case Study ... 173
3.1. Background and Rationale ... 173
3.2. The Knowledge Domain ... 173
3.3. Procedures ... 173
3.4. Summary of Findings ... 178

3.5. Interviewing the Expert . . . 179
3.6. Comments on the Interviews . . . 182
4. Conclusion . . . 183
4.1. Issues in Induction . . . 183
4.2. Guidelines Summary . . . 187
5. References . . . 188

INDEX . . . 191

1

Knowledge Acquisition

An Introductory Framework

ALISON L. KIDD

1. INTRODUCTION

Knowledge acquisition (KA) is a crucial stage in the development of an expert system. As a process, it involves eliciting, analyzing, and interpreting the knowledge that a human expert uses when solving a particular problem and then transforming this knowledge into a suitable machine representation. KA is critical since the power and utility of the resulting expert system depend on the quality of the underlying representation of expert knowledge. The aim of this book is to provide the builders of expert systems with some practical advice on what KA involves and some of the methodologies and techniques that have been developed to aid its effectiveness.

The fact that KA has proved such a major stumbling block in the production of expert systems is not surprising. To solve the problem of KA in any real sense would entail answering some fundamental questions, not only in AI, but also in psychology and the philosophy of science. For example:

- What is the relationship between knowledge and language?
- How can we characterise different domains of knowledge?
- What constitutes a theory of human problem solving?

ALISON L. KIDD • Hewlett Packard Laboratories, Stoke Gifford, Bristol BS12 6QZ, England.

The fact that KA as a process now has its own name and has received so much recent attention in AI does not make these underlying questions any easier to answer. Until we have some answers, there is no hope that AI research can produce one "magical" technique that solves the KA problem. Any claims that such a technique exists (e.g., Michie and Johnston, 1985) are fundamentally misguided.

This book was designed to fill a gap in the existing literature on KA between, on the one hand, AI research papers that describe only scantily the techniques employed in acquiring the knowledge for any expert system and, on the other hand, the large body of literature describing techniques developed within cognitive and behavioral psychology. The latter, although seemingly relevant, have rarely been applied to expert system applications or even described in terms that system builders can easily interpret and exploit.

Each chapter of the book provides a detailed description of a particular KA technique or methodology, applied within a selected task domain. The relative strengths and weaknesses of the technique are summarized at the end of each chapter with some suggested guidelines for its use. Some of the chapters are written from the standpoint of the AI scientist, interested primarily in building a powerful working system to simulate some expert task (e.g., Kuipers and Kassirer, Fox *et al.*). Others are written from the standpoint of the psychologist applying some well-tried psychological technique to an expert system style application (e.g., Shaw and Gaines, Gammack).

By way of introduction, this opening chapter makes a series of points that clarify the purpose of the KA process and outline what it involves. The aim is, first, to uncover some common misconceptions in the way KA is currently described and practiced and, second, to provide a theoretical framework that will help readers to interpret and assess the value of the diverse set of techniques described in the following chapters, in relation to their own KA requirements. Many of the theoretical issues raised in this chapter are discussed in greater depth by individual authors. However, all the opinions expressed in this chapter are not necessarily shared by my fellow authors.

2. THE KNOWLEDGE ACQUISITION PROCESS: MINING IS A MISGUIDED ANALOGY

A popular view of KA has been to consider experts' heads as being filled with bits of knowledge and the problem of KA as being one of "mining those jewels of knowledge out of their heads one by one"

(Feigenbaum and McCorduck, 1983). The underlying assumption is that some magical one-to-one correspondence exists between the verbal comments made by an expert during an elicitation session and real items of knowledge that are buried deep within his head (Kidd and Welbank, 1984). Once these comments are transcribed onto a piece of paper, they are considered to be nuggets of the truth about that domain. The analogy is misleading; there is no *truth* in that sense. Rather, it is the case that KA involves the following:

1. Employing a technique to elicit data (usually verbal) from the expert.
2. Interpreting these verbal data (more or less skillfully) in order to infer what might be the expert's underlying knowledge and reasoning processes.
3. Using this interpretation to guide the construction of some model or language that describes (more or less accurately) the expert's knowledge and performance. Interpretation of further data is guided in turn by this evolving model (Johnson, 1985).

Knowledge engineers need to recognize that this is the basic process in which they are engaging and that the techniques described in this book are ways of helping them carry out this process more accurately and effectively.

3. KNOWLEDGE DOMAINS: CAPTURING OR CREATING A LANGUAGE

In the past 10 years, expert systems have been constructed in a wide variety of domains. Even within this book, the list includes domains as dissimilar as leukemia diagnosis, air-conditioning design, children's arithmetic, and the selection of college students. Unfortunately, AI currently has no formal (or even informal) characterization of different knowledge domains. Only as we develop such a characterization, will it be possible to identify meaningfully and to classify KA techniques in relation to different domains and tasks. An approach to increasing our present understanding of this classification is proposed in section 4.

In the meantime, there are two fundamental questions that knowledge engineers can ask about any domain now, the answers to which are critical in determining (1) the suitability of that domain for developing a successful expert system and (2) the type of KA that is necessary for developing such a system.

What language do humans possess for representing and reasoning within this domain? (Is it formally defined? If not, how stable and well-agreed is it?)

Example Class 1: Mathematics, Geometry, Programming Languages. In this class of domains, humans have already developed a formal language for representation and reasoning; i.e., axioms, semantics, and syntax have been defined. KA for such domains should therefore be straightforward, involving the capture of some subset of this language and mapping it into an appropriate machine equivalent. The only difficult part of the KA process is discovering and describing how the expert maps this formal language onto any real-world problem. This activity is no longer formally defined, and here the knowledge engineer will have to create a language to describe the process.

Example Class 2: Medicine, Chemistry, Electronic Equipment Design and Diagnosis. Humans have not yet developed a formal language for domains in this class. At the deepest level, functionality is described in terms of physical laws, but these are hypotheses, not axioms: They may have a basis in deeper theory and be subject to error (Clancey, 1986). However, humans have developed a fairly stable, well-agreed language for representing and reasoning about such domains. KA therefore involves capturing the essence of this language and then systematizing it for machine purposes.

Example Class 3: Applications Software, Management, Marketing. Domains in this class do not currently have a stable, well-agreed language for representation or reasoning. They lack a coherent underlying theory. For example, when an applications programmer is required to develop a new electronic mail system, he first has to develop his own theory of electronic mail before writing the software to realize this theory in his own idiosyncratic way. As a result, such domains present the worst problems for expert system development because KA must involve the creation of a "new" language (and therefore an implicit theory) that will adequately support reasoning within the domain. This new language is likely to be the subject of continuous negotiation with both experts and users even when the expert system is developed. Surprisingly, however, this class of domain is proving increasingly popular for expert system development, e.g., electronic mail advice systems. Presumably this is because such domains present the biggest problems to human reasoning simply because they lack an agreed language or any underlying theory. Knowledge engineers need to be very aware of these problems and of what KA in such a domain must involve.

Is there an existing machine representation that can adequately support the human's domain language?

Example Class 4: Spatial Reasoning Tasks, e.g., Stick-and-Ball Models of Molecules. In some domains, humans have developed a fairly formal and powerful language for reasoning about a class of difficult problems, but such a language is currently unsupportable with state-of-the-art machine representation techniques. The main thrust of the KA process in such a domain is to develop a new machine language that can adequately capture the power of the existing human language.*

4. TASKS: DEFINING THE PROBLEM THAT THE EXPERT SYSTEM IS DESIGNED TO SOLVE

The aim of an expert system is not merely to capture a static representation of some knowledge domain but to simulate a particular problem-solving task (or tasks) carried out within that domain. A generic set of such tasks has been identified by Hayes-Roth *et al.* (1984) and is now widely referenced. This list comprises the following:

INTERPRETATION Inferring situation descriptions from sensor data.
PREDICTION Inferring likely consequences of a given situation.
DIAGNOSIS Inferring system malfunctions from observables.
DESIGN Configuring objects under constraints.
PLANNING Designing actions.
MONITORING Comparing observations to plan vulnerabilities.
DEBUGGING Prescribing remedies for malfunctions.
REPAIR Executing a plan to administer a prescribed remedy.
INSTRUCTION Diagnosing, debugging, and repairing student behavior.
CONTROL Interpreting, predicting, repairing, and monitoring system behaviors.

The utility, and indeed validity, of these categories is questionable for three reasons: (1) What appears to be an inherent property of a real-world problem may rather be a property of how AI programs currently

*I would like to acknowledge the contribution of my colleague, Bill Sharpe, to many of the ideas discussed in this section.