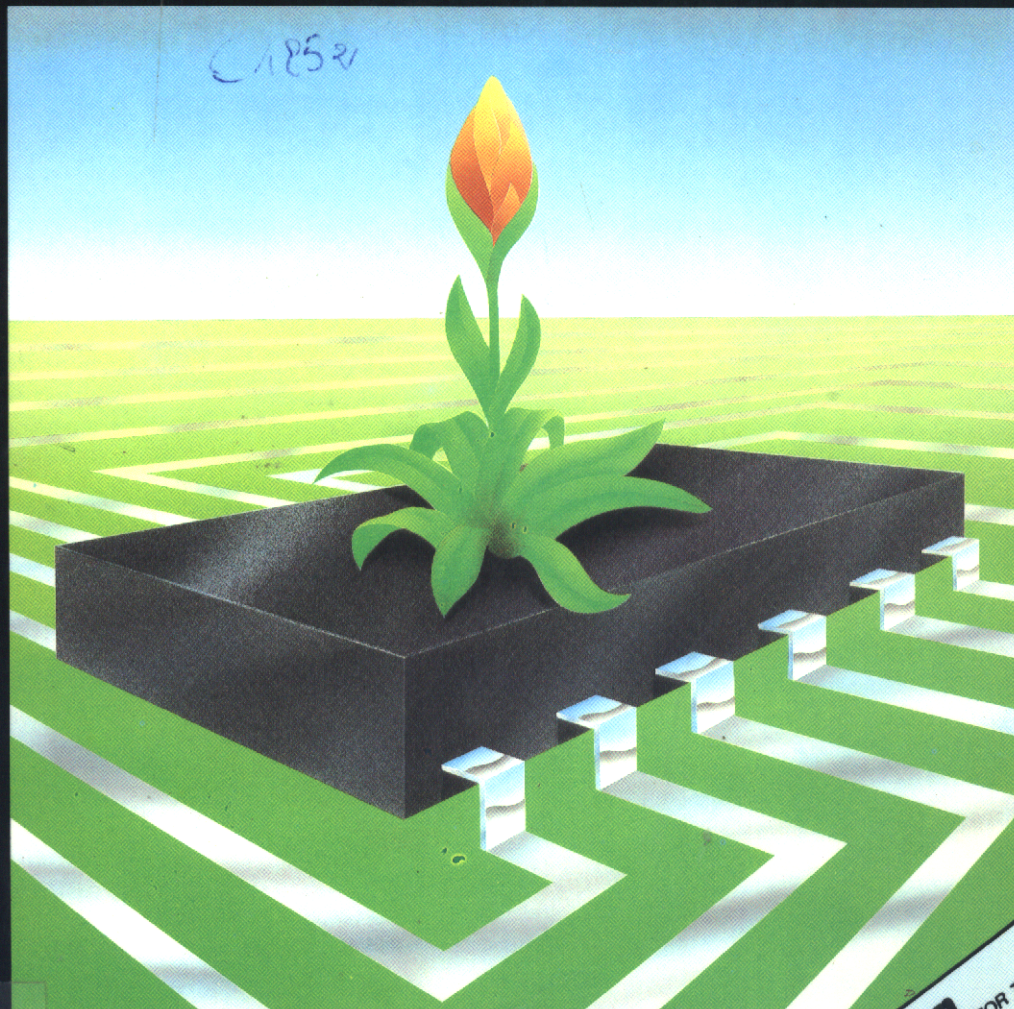


BUILD YOUR OWN

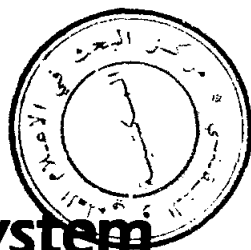
Expert System

CHRIS NAYLOR



SIGMA
PRESS

2nd. Edition FOR THE
IBM-PC and AMSTRAD PC 1512



Build Your Own Expert System

2nd Edition
for the IBM PC and Compatibles

Chris Naylor



Publishers. Wilmslow



HALSTED PRESS a division of JOHN WILEY & SONS
New York · Chichester · Brisbane · Toronto

Copyright © Chris Naylor, 1987.

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior written permission.

ISBN 1-85058-071-5 (Sigma Press)

ISBN 0-470-20946-1 (Halsted Press)

First published in 1987 by

Sigma Press

98a Water Lane, Wilmslow, SK9 5BB, England.

Printed in Malta by Interprint Limited

British Library Cataloguing in Publication Data

Naylor, Chris

Build your own PC expert system.

1. Expert systems (Computer science)
2. Microcomputers
3. System design

I. Title

006.3'3

QA76.76.E95

Library of Congress Cataloguing in Publication Data

Naylor, Chris, 1947-

Build your own expert system.

Includes index.

1. Expert systems (Computer science)
 2. Microcomputers
- Programming. I. Title.

QA76.76.E95N38 1987

006.3'3

87-15006

Distributed in UK and Europe by

John Wiley & Sons Ltd., Baffins Lane, Chichester, West Sussex, England.

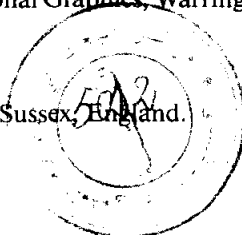
Acknowledgments

IBM PC and PC-DOS are trademarks and IBM is a registered trademark of International Business Machines Corporation. PC1512 is a trademark of Amstrad Consumer Electronics PLC. BASIC 2 is a trademark of Locomotive Software Ltd.

Cover graphics and illustration by David Collins, Professional Graphics, Warrington.

Distributed by

John Wiley & Sons Ltd., Baffins Lane, Chichester, West Sussex, England.



Preface

A potential reader is standing in a bookshop with a copy of a possible purchase in his or her hand. Having looked at the front cover, and looked at the back cover, he or she now reads the Preface to determine whether or not the book should be purchased. The Preface, says the Theory, clinches the sale.

So much for theory. The real reason you should buy this book is because, for a book on computers, it is relatively cheap. It also contains working examples in BASIC for the IBM PC and close compatibles so you get some 'free' programs thrown in for your money. Actually though, the BASIC used is so basic that it could be pretty readily converted to most other computers if you don't happen to have a PC.

It tells you a moderate amount about Expert Systems, but, frankly, to disclose exactly what it does tell about Expert Systems would rather annihilate the reason for buying it. After all, you could just stand there, in the bookshop, reading the Preface and hang onto your money. But it will (just as a sort of appetiser) enable you to build your own expert system for medical diagnosis. Or an expert system to work out why your car won't start in the morning. Or it will enable you to build a learn-by-example expert system which can be taught expertise in a wide range of areas.

It will also teach you a fair amount about statistics and inferencing systems but, despite that, you should still shell out the necessary and buy a copy.

I know money is tight these days and you could usefully spend it on something else, like drink, which would give *you* greater pleasure but I have to make a living too you know and the cost of printer ribbons alone was pretty enormous when it came to bashing this lot out.

Well, after all that, maybe you've had the heart to fork out on a copy and you're actually planning to read the thing now. What you do is start at the beginning and carry on until you get to the Technical Overview. Then think of something you'd like to try out – such as a machine learning system or a bit of diagnosis – and dig out the relevant parts using the Contents pages, the Index, and the Technical Overview to tie it together.

Alternatively, you could have your computer switched on as you read the book and key in the examples as you go along, that way seeing how they work as an aid to understanding. If you do this it could take you ages to get to the end, of course.

The big point to note though is that this book is not arranged like a normal text book. The chapters are not isolated entities. One way or another you *do* have to churn right through from front to back to get the ideas in the proper sequence. A few people have commented that it reads more like a novel than a text book in the way it's arranged – which is a fair point, but it's probably as well to warn you that it's like that.

Prior to publication this book was read by Graham Beech, Phil Bradley, Bill Hudspith and Phil Manchester (in alphabetical order).

They all chipped in with comments of one sort or another and they each have caused some improvements to be made to the final version. Which was nice of them.

If anyone, having read the book, has any comments which might lead to useful alterations, then drop the publisher a line and let him know.

A Note on the Programs

All of the programs in this book have been written in Basic to run on the IBM PC and close compatibles.

They have been tested under IBM's Advanced Basic (Basica), GW Basic and Locomotive Basic 2 (as supplied with the Amstrad PC 1512).

However, they have all been written using a very simple sub-set of Basic which should enable the average Basic programmer to readily convert the programs to some other dialect to suit their particular machine or individual preferences.

Where examples of actual run-time output are given these were obtained using an IBM PC running Basica. Readers using other machines or other Basic interpreters can expect to obtain similar results at run-time to those shown but minor variations may arise. This is usually due to such things as the different methods used by other machine/interpreter combinations when rounding numeric variables and the differing behaviour of their random number generators and other built-in functions. Any variations in run-time performance due to these factors are usually of little practical consequence.

All of the program code shown in this book has been printed in upper case text except in the case of string constants where a combination of upper and lower case text is sometimes used. Depending on the Basic interpreter you use you may find that the program code displayed on your screen is not in the same combination of upper and lower case. For instance, users of the Amstrad PC 1512 running Locomotive Basic 2 will notice that their screen only displays Basic 2 key words in upper case whilst variable names are displayed in lower case. This is quite normal and is not a program fault.

Also, where a program line extends over more than one line on the printed page the line has been broken, indented, and continued in such a way as to try to make the overall program text as legible as possible to the reader. The result is that the lines as displayed on your screen may not be broken and indented in exactly the same way as shown in this book. For example, users of the IBM PC running Basica will notice that any lines extending over more than 80 characters are displayed on their screens in such a way as to continue on the next line without any indentation. Again, this is quite normal and is not a program fault.

In general, when entering programs into your machine try to maintain the sense of the code and do not try to make the physical spacing, layout or appearance exactly match that shown in the book.

Reader Convenience Disk

A 5¼" PC-compatible disk containing all the programs in this book—including the medical knowledge base is available from Sigma Press for £11.50 including post, packing and VAT. This price is valid until September 1988 and is applicable to UK and Europe only. Other purchasers should remit the equivalent of £15.00 to include airmail despatch.

Sigma Press, 98a Water Lane, Wilmslow, Cheshire, SK9 5BB. England.

Contents

Chapter 1: Why Expert Systems?	1
1.1 What do you want an Expert System for?	2
1.2 What do Other People Want an Expert System for?	3
1.3 What is an Expert System?	4
1.4 What do you want your Expert System to do?	7
1.5 Expert Systems: Some Untrue Things	8
Chapter 2: A Statistical Scheme	11
2.1 Setting up a Matrix	11
2.2 Probabilities	16
2.3 More Probabilities	18
2.4 More Variables	21
2.5 Bayes' Theorem	27
Chapter 3: Avoiding Probabilities	29
3.1 How to Make the Computer do the Hard Work	30
3.2 The Learning System	30
3.3 Other Types of Data	35
3.4 The Judgement Rule	36
3.5 Building a Rule	39
3.6 Prior Probabilities	43
3.7 Expanding your Options	43
3.8 Can it make a Mistake?	47
Chapter 4: Improving your Expert	51
4.1 Parallel and Sequential Decisions	51
4.2 Adding some Commonsense	57
4.3 A Trial Run of our New Expert	62
Chapter 5: A Real World Expert	71
5.1 The Weather Again	71
5.2 A Chi-Squared Program	78
5.3 Exercising your Expert	78
5.4 Direct Estimation	83
Chapter 6: Running for Real	87
6.1 Using your Expert	87
6.2 Reserved Judgement	91
6.3 The Problem of Distance	92
6.4 Understanding your Problem	98
Chapter 7: An Expert on Everything in the Entire Known World	99
7.1 Nodes	99
7.2 The Variables so Far	104
7.3 Going Through the Nodes	108
7.4 Tailor Made Nodes	112
7.5 Specific Code	115

7.6 Saving your Expert	116
7.7 The Multi-Node Code	116
7.8 Some Examples	123
7.8.1 A single node example.....	123
7.8.2 A two-node example	127
7.8.3 And now for something useful	133
Chapter 8: How can you use your Expert?	139
8.1 Choosing a Problem	139
8.2 Analysing the Problem	141
Chapter 9: Large-Scale Expert Systems	145
9.1 MYCIN—Medical Diagnosis	145
9.2 PUFF—Breathing Disorders	151
9.3 DENDRAL—Chemical Structures	157
9.4 PROSPECTOR—Searching for Minerals	160
9.5 Some Other Examples	166
Chapter 10: A Rule Based BASIC Expert	171
10.1 A System that Works Backwards	171
10.2 The BASIC Program	174
10.3 A Medical Knowledge Base	180
Chapter 11: The Tower of Babel	189
Chapter 12: Summary and Technical Overview	197
12.1 Events	197
12.2 Probabilities	197
12.2.1 Bayes' Theorem	198
12.2.2 Prior and posterior probabilities	199
12.2.3 Odds	199
12.2.4 Approximations	199
12.2.5 Combinations	200
12.2.6 Descriptive statistics	200
12.2.7 Normal distribution	201
12.2.8 Discrete and continuous variables	201
12.3 Surfaces	201
12.4 Discrimination	202
12.5 The Learning Algorithm	202
12.6 Parallel and Sequential Procedures	203
12.7 Minimum and Maximum Values	203
12.8 Processing Strategies	204
12.8.1 Goal-driven strategies	204
12.8.2 Data-driven strategies	204
12.8.3 Selecting the next variable	204
12.9 Intermediate Conclusions	205
12.10 Linear Interpolation of Responses	206
12.11 Data Formats	206
Chapter 13: Select Readings	209