INFORMATION RETRIEVAL

INFORMATION RETRIEVAL

C. J. van Rijsbergen B.Sc. Ph.D. M.B.C.S.

Computer Laboratory

Cambridge University

BUTTERWORTHS

London & Boston

THE BUTTERWORTH GROUP

ENGLAND
Butterworth & Co (Publishers) Ltd
London: 88 Kingsway, WC2B 6AB

AUSTRALIA
Butterworths Pty Ltd
Sydney: 586 Pacific Highway, NSW 2067
Melbourne: 343 Little Collins Street, 3000
Brisbane: 240 Queen Street, 4000

CANADA
Butterworth & Co (Canada) Ltd
Toronto: 2265 Midland Avenue, Scarborough, Ontario, M1P 4S1

NEW ZEALAND
Butterworths of New Zealand Ltd
Wellington: 26-28 Waring Taylor Street, 1

SOUTH AFRICA
Butterworth & Co (South Africa) (Pty) Ltd
Durban: 152-154 Gale Street

USA
Butterworth & Co (Publishers) Inc
Reading: 161 Ash Street, Massachusetts, 01867

First published 1975

ISBN 0 408 70717 8

Library of Congress Cataloging in Publication Data

Van Rijsbergen, C J 1943-
Information retrieval.

Bibliography: p.
1. Information storage and retrieval systems.
I. Title.
Z699.V35 029.7 75-33024
ISBN 0-408-70717-8

Printed in England by
The Whitefriars Press Ltd, London and Tonbridge

PREFACE

The material of this book is aimed at advanced undergraduate information (or computer) science students, postgraduate library science students, and research workers in the field of IR. Some of the chapters, particularly Chapter 6, make *simple* use of a little advanced mathematics. However, the necessary mathematical tools can be easily mastered from numerous mathematical texts that now exist and in any case references have been given where the mathematics occur.

I had to face the problem of balancing clarity of exposition with density of references. I was tempted to give large numbers of references but was afraid they would have destroyed the continuity of the text. I have tried to steer a middle course and not compete with the *Annual Review of Information Science and Technology*.

Normally one is encouraged to cite only works that have been published in some readily accessible form such as a book or periodical. Unfortunately much of the interesting work in IR is contained in technical reports and Ph.D. theses. For example most of the work done on the SMART system at Cornell is available only in reports. Luckily many of these are now available through the National Technical Information Service (U.S.) and University Microfilms (U.K.). I have not avoided using these sources although if the same material is accessible more readily in some other form I have given it preference.

I should like to acknowledge my considerable debt to many people and institutions that have helped me. Let me say first that they are responsible for many of the ideas in this book but that only I wish to be held responsible. My greatest debt is to Karen Sparck Jones who taught me to research information retrieval as an experimental science. Nick Jardine and Robin Sibson taught me about the theory of

automatic classification. Cyril Cleverdon is responsible for forcing me to think about evaluation. Mike Keen helped by providing data. Gerry Salton has influenced my thinking about IR considerably, mainly through his published work. Ken Moody had the knack of bailing me out when the going was rough and encouraging me to continue experimenting. Juliet Gundry is responsible for making the text more readable and clear. Bruce Croft, who read the final draft, made many useful comments. Ness Barry takes all the credit for preparing the manuscript. Finally, I am grateful to the Office of Scientific and Technical Information for funding most of the early experimental work on which the book is based; to the King's College Research Centre for providing me with an environment in which I could think, and to the Department of Information Science at Monash University for providing me with the facilities for writing.

C.J.v.R

CONTENTS