INFORMATION RETRIEVAL

C. J. van RIJSBERGEN B.Sc., Ph.D., M.B.C.S.

Department of Computing Science

University of Glasgow

PREFACE TO THE SECOND EDITION

The major change in the second edition of this book is the addition of a new chapter on probabilistic retrieval. This chapter has been included because I think this is one of the most interesting and active areas of research in information retrieval. There are still many problems to be solved so I hope that this particular chapter will be of some help to those who want to advance the state of knowledge in this area. All the other chapters have been updated by including some of the more recent work on the topics covered. In preparing this new edition I have benefited from discussions with Bruce Croft, David Harper, Stephen Robertson and Karen Sparck Jones. I am grateful to the University of Cambridge Computer Laboratory for providing me with the facilities for carrying out the work. Finally, I am indebted to the Royal Society for supporting me on their Scientific Information Research Fellowship.

PREFACE TO THE FIRST EDITION

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*.

---

* This is Chapter 7 in the second edition.

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 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 Kings 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