British Library R & D Report no. 5587

# NEW MODELS IN PROBABILISTIC INFORMATION RETRIEVAL

C.J. van Rijsbergen

S.E. Robertson

M.F. Porter

Computer Laboratory
University of Cambridge
Corn Exchange Street
Cambridge CB2 3QG
England

© 1980

CONTENTS


| | | | |
|---|---|---|---|
| | Introduction | ... | 1 |
| 1 | Establishing the NPL test collection at Cambridge | ... | 4 |
| 2 | The selection of good search terms | ... | 8 |
| 3 | Probabilistic models of indexing and searching | ... | 36 |
| 4 | Loose ends | ... | 63 |
| 5 | Programs for setting up document test collections | ... | 68 |
| 6 | An algorithm for suffix stripping | ... | 98 |
| 7 | The CAMIR system | ... | 107 |