BL R&D Report 5481

STATISTICAL BASES OF RELEVANCE ASSESSMENT FOR THE 'IDEAL' INFORMATION RETRIEVAL TEST COLLECTION

H. Gilbert
K. Sparck Jones

Computer Laboratory
University of Cambridge
Corn Exchange Street
Cambridge, England

(C)

March 1979

Acknowledgement

This work was supported by British Library Research and Development Department under Grant SI/G/267.

We are grateful to Dr. P.M.E. Altham of the University Statistical Laboratory, and to Dr. C.J. van Rijsbergen and Dr. S.E. Robertson for their assistance.

Contents

		page
Summary		S1
Section A :	The 'ideal' test collection and obtaining relevance assessments for it	A1
A1	Introduction: the 'ideal' information retrieval test collection	A1
A2	The 'ideal' collection specification	A2
A3	Relevance data	A3
A4	Constraints on statistical methods	A4
Section B :	Statistical methods of determining relevance assessment requirements	B1
B1	Preface: the research proposed and its conduct	B1
1.1	General requirement	B3
B2	The Pool method	B4
2.1	Non-statistical summary	B4
2.2	Statistical presentation	B10
2.2.1	Scrutiny of design study argument	B10
2.2.2	Improved sampling rationale: the modified Pool method	B15
2.2.3	Accuracy of estimators	B16
2.2.4	Weakening the assumptions	B18
2.2.5	Conclusion on the Pool method	B18
B3	Other methods: dead ends	B19
3.1	Logistic model	B20
3.2	Wilcoxon's signed ranks test	B23
3.3	Likelihood ratio test	B25
B4	The Squares method	B26
4.1	Non-statistical summary	B26
4.2	Statistical presentation	B32
B5	Comparison between the Pool and Squares method	B37
B6	Multi-strategy comparisons	B41
Section C :	Practical implications of the statistical methods for building and using the 'ideal' collection	C1
C1	Summary of the properties of the statistical methods	C1
C2	Application of the methods	C3
2.1	Testing to obtain statistical parameter values	C3
2.2	Practical implications of the methods	C3
2.3	Assessment of the methods	C7
References		R1
Appendix 1 :	Tables and figures	A1
	Table 0	A2
	Table 1	A3
	Table 2	A4
	Table 3	A6
	Table 4	A8
	Table 5	A12
	Table 6	A13
	Table 7	A14
	Figures 1 and 2	A19
Appendix 2 :	Computing	A23