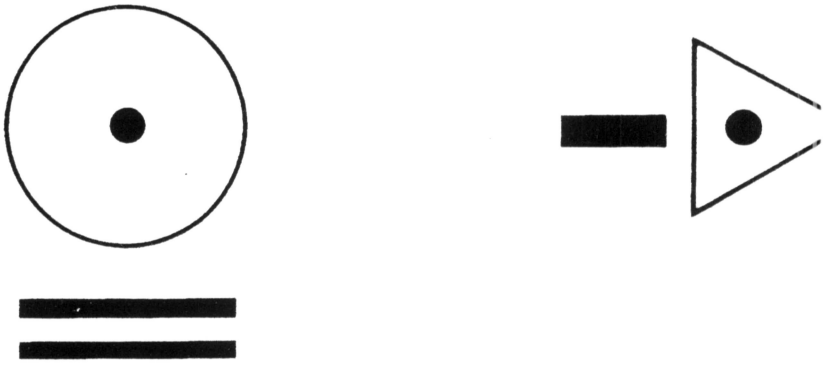
IMPROVING SUBJECT RETRIEVAL IN ONLINE CATALOGUES

Stephen Walker and
Richard M Jones



British Library Research Paper 24

Improving subject retrieval in online catalogues

1. Stemming, automatic spelling correction and cross-reference tables

Stephen Walker and Richard M Jones

with contributions by Nicky Johns

The Polytechnic of Central London

1987

British Library Research Paper 24

British Library Cataloguing in Publication Data

Walker, Stephen
Improving subject retrieval in online catalogues.——(British Library research paper, ISSN 0269-9257; 24).
1 :, Stemming, automatic spelling correction and cross-reference tables
1. Catalogs, On-line
I. Title II. Jones, Richard M. III.Series
025.3'13 Z699

ISBN 0-7123-3129-8

Stephen Walker worked in management services in industry, and as a mathematician, before moving into information science. He has taught information science and computing in several institutions, and worked on the design of information systems. Since 1982 he has worked exclusively in the field of online catalogues, both at the Polytechnic of Central London and as Technical Consultant to Swalcap Library Services Ltd, Bristol.

Richard Jones is a librarian who worked for three years in the Library of the University of London Institute of Education before joining the Okapi projects. He surveyed the operation of the Geac online catalogue at the Polytechnic of the South bank for his Masters' dissertation.

SI/G/720

British Library Research Papers are published by the British Library and distributed by the British Library Publications Sales Unit, Boston Spa, Wetherby, West Yorkshire LS23 7BQ, UK. In the USA and Canada they are distributed by Longwood Publishing Group Inc, 27 South Main Street, Wolfeboro, New Hampshire 03894-2069, USA. In Japan they are distributed by Kinokuniya Co Ltd, PO Box 55, Chitose, Tokyo 156.

Printed in Great Britain by the British Library.

Preface

This report describes research into practical ways of improving subject access in online catalogues. This is part of a continuing programme which began in 1982 with the original Okapi project [1]. This is not a "final" report. The next phase, on the use of relevance feedback, will be published in 1988.

The work described here is concerned both with librarianship and with computing, particularly *linguistic* computing. Some knowledge of online catalogues and of computing and library jargon is assumed. We refer, for example, to "generations" of online catalogues, to "phrase" and "keyword-type" catalogues, subject searching and specific (or known-) item searching, to headings, entries, fields and records. The most important source for background material to this report is [1]. For online catalogues in general, Hildreth's 1983 monograph [2] and his ARIST survey [3] are recommended. The most useful work on subject searching problems is Markey's book [4].

Acknowledgments

This work would not have been possible without the help and support of the British Library Research and Devlopment Department. Particular thanks are due to Maureen Grieves and to Derek Greenwood, who is now our project officer.

We are also most grateful to Neil McLean (Project Head and Head of Library Services, Polytechnic of Central London), to Maura Coghlan (site librarian of our test library) and to Dave Roberts and his staff at the Poytechnic's Computing Services.

Others at PCL who have helped us include Winifred Abbott (Technical Services Librarian), Jacky Conroy (Systems Librarian), Pat Manson (Information Officer, Library Technology Centre) and Penny Pope (Chief Cataloguer). Many other people have helped us with advice, criticism and encouragement. They include Philip Bryant (Director of the Centre for Catalogue Research), Charles Hildreth (READ Ltd), Karen Markey (University of Michigan), Martin Porter (Scott Polar Research Institute) and Peter Willett (Sheffield University). Gill Venner (Plymouth Polytechnic, late of the Okapi team) helped with the rewriting of some of the Okapi programs.

Credits

It would not have been possible to design and program the experimental catalogues used in this project without the foundation provided by the original Okapi system. This was designed and written by Nathalie Mitev, Gill Venner and Stephen Walker. The additional programs were written by Nicky Johns and Stephen Walker. Nicky Johns did much of the detailed design and wrote some complex programs with great speed and accuracy. Richard Jones made substantial contributions to the design and did most of the evaluation work - interviewing users and repeating searches. He also wrote a survey of methods of improving retrieval in subject searching. Chapters 3, 4 and 5 of this report are condensed from this survey, which we hope will be published separately (it is too long to be incorporated in full here). Parts of Chapter 6 were written by Nicky Johns. Most of the rest of the report was written by Stephen Walker, who is responsible for the inevitable inaccuracies. It has not been easy to reconcile the demands of readability and conciseness with the need to include enough detail for other designers to be able to benefit from this work. It is hoped that complete Okapi specifications will be published later this year, together with program source code. Meanwhile, anyone wanting additional detailed information should contact Stephen Walker at PCL.

References

1 MITEV N N, VENNER G M and WALKER S. *Designing an online public access catalogue : Okapi, a catalogue on a local area network* (Library and Information Research Report 39). London : the British Library, 1985.

2 HILDRETH C R. *Online Public Access Catalogs : the user interface.* OCLC Online Computer Library Center, 1982.

3 WILLIAMS M E (editor). *Annual Review of Information Science and Technology. Vol 20.* Knowledge Industry Publications, 1985.

4 MARKEY K. *Subject searching in library catalogs : before and after the introduction of online catalogs.* OCLC Online Computer Library Center, 1984.

Contents

List of figures

List of tables