ONLINE
INFORMATION RETRIEVAL SYSTEMS

an introductory manual to
principles and practice

Second Edition

ONLINE INFORMATION RETRIEVAL SYSTEMS

an introductory manual to principles and practice

Second Edition

BERNARD HOUGHTON MPhil MA FLA
Senior Lecturer in Information Work
Liverpool Polytechnic
School of Librarianship and Information Studies

and

JOHN CONVEY BEd ALA
Senior Librarian
Lancashire Library

CLIVE BINGLEY  LONDON

Published by Clive Bingley Limited, The Library Association, 7 Ridgmount Street, London W1, and printed in England by Redwood Burn Limited, Trowbridge, Wiltshire.

First published 1977
This second, revised edition published 1984

British Library Cataloguing in Publication Data

Houghton, Bernard, *1935* –
Online information retrieval systems.
— 2nd ed., rev.
1. On-line bibliographic searching
2. Interactive computer systems
3. Information storage and retrieval systems
I. Title II. Convey, John
025.5'2 Z699.3

ISBN 0 85157 363 0

Typeset by Allset in 10/12 Press Roman and Univers

12345688685848382

CONTENTS

ONLINE INFORMATION RETRIEVAL SYSTEMS

an introductory manual to principles and practice

Second Edition

BERNARD HOUGHTON MPhil MA FLA
Senior Lecturer in Information Work
Liverpool Polytechnic
School of Librarianship and Information Studies

and

JOHN CONVEY BEd ALA
Senior Librarian
Lancashire Library

CLIVE BINGLEY  LONDON

Published by Clive Bingley Limited, The Library Association, 7 Ridgmount Street, London W1, and printed in England by Redwood Burn Limited, Trowbridge, Wiltshire.

First published 1977
This second, revised edition published 1984

British Library Cataloguing in Publication Data

Houghton, Bernard, *1935* –
Online information retrieval systems.
— 2nd ed., rev.
1. On-line bibliographic searching
2. Interactive computer systems
3. Information storage and retrieval systems
I. Title II. Convey, John
025.5'2 Z699.3

ISBN 0 85157 363 0

Typeset by Allset in 10/12 Press Roman and Univers

12345688685848382

CONTENTS

ILLUSTRATIONS

INTRODUCTION

This work is designed for use by students of librarianship and information science, by practising information workers and researchers who wish to familiarize themselves with the principles and techniques of online access to bibliographic databases, and in addition by the increasing numbers of end-users who are now using online information retrieval systems to fulfil their own information needs. It is based on the experience the authors have accumulated while developing teaching programmes for students and practising information workers at the School of Librarianship and Information Studies at Liverpool Polytechnic and it also incorporates the expertise they have gained while operating online information services at Liverpool Polytechnic and the Lancashire Library.

This second edition, while maintaining the basic layout of the first edition, updates and considerably enlarges the text published in 1977 to reflect current practice. Part One of the work seeks to introduce the reader to the principles of online access by concentrating on essentials and consciously omitting extended technical detail which could be considered extraneous by information workers. Part Two is a practical exposition of the principles outlined in Part One demonstrating the techniques and conventions of online interaction with six of the major systems currently accessible in Europe and the United States. It is arranged to form part of a training programme for users new to online systems whether they be intermediaries (ie librarians or information workers) or end-users.

Liverpool, May 1983.

PART ONE

Introduction to online information retrieval principles and systems

1: Development of online information retrieval; the information retrieval industry; databases

The development of online information retrieval

The general availability of online access to bibliographical and non-bibliographical databases is the inevitable technological consequence of applying computers to the processing of bibliographical and referral information. The computer revolution has proceeded rapidly and its successive stages are now clearly discernible.

Initially, computers were employed to produce printed abstracting and indexing services more efficiently. As a by-product of the generation of the machine-readable databases necessary for the production of the printed indexes, the machines were used to process both current-awareness and retrospective literature searches in batch mode; ie a number of searches or user-profiles were run at one time 'offline' at an information centre to which the user addressed his request for information. Now remote terminals are used by librarians, information scientists ('intermediaries') acting on behalf of their users and, increasingly, by the end-users (ie those who will ultimately use the information from the search) themselves, to interact conversationally with the machine-readable databases in online searching. The information retrieval industry is now expanding rapidly and its structure is still evolving. The electronic database, with no corresponding hard copy version, has now emerged and as the costs of paper and print escalate and the cost of online access to databases over telecommunication networks diminishes, the continuing existence of hard copy abstracting, indexing, and in many cases referral publications, is uncertain.

The computer was first publicly applied to the processing of bibliographical information in 1961, when the *Chemical abstracts service (CAS)* produced *Chemical titles* (*CT*), a machine-generated alphabetical subject index to the 600 most influential journals covered by the parent journal *Chemical abstracts*. The keyword-in-context (KWIC) format devised for *CT* manipulated actual keywords taken from the titles of the papers published in the journals, so dispensing with the intellectual processes involved in concept indexing. The KWIC indexing principle thus represents a crude form of the natural language and full-text searching facility now available in some measure in all online retrieval systems. The KWIC index could be produced quickly and cheaply, and was readily adopted by other abstracting and indexing

services to alleviate the problems imposed by the phenomenal growth of the literature.

By the end of the 1960s a whole range of new *CAS* publications had been produced from a machine-readable database. In this system the abstracting and indexing information was culled from the source journals in a single process of intellectual analysis before it was input to the database in a single keyboarding operation. Computerization was steadily adopted by other services, and by 1970 the first phase of the computer revolution was completed. The new processing methods gave the secondary services (ie indexing and abstracting services) the increased flexibility necessary to adapt to the changing patterns of research. They were now able to produce not only a single abstracting publication but also to repackage the references held in the database to meet the varying needs of groups and individual users.

The result of this first phase of the revolution was greatly improved secondary publications; they were more up to date, more comprehensive in their coverage of the literature, and they were equipped with better indexes which cumulated more frequently. In addition, users were given the facility of running 'profiles' against the database to keep them informed of recent papers covering their spheres of interest, and they could also search the database retrospectively to retrieve all papers which satisfied a search strategy designed to circumscribe their subject demands.

These current-awareness and retrospective searches were run in 'batch mode' in information centres, each of which dealt with a specific discipline or subject field, eg the UK MEDLARS service based at the British Library Lending Division at Boston Spa or the UK Chemical Information Service (UKCIS) at Nottingham University. These centres acquired the machine-readable tapes from the database producers, loaded them into their own computers and produced the necessary software to enable the batch processing of current-awareness and retrospective searches. Batch processing of searches was economical in terms of machine time but it had two serious disadvantages. Firstly, it necessitated the user experiencing delays of maybe three weeks before he received his search product from the information centre, as it was necessary to accumulate a sufficient number of searches to make batch processing economical; secondly, the user was forced to search 'blind'; he could not sample the relevance of the references retrieved by the search strategy formulated on his behalf by a search analyst at the information centre, defining the question he was putting to the service.

Interactive online retrieval systems, which enabled the user to communicate directly (in a conversational mode) with the database being searched were being developed by the System Development Corporation (SDC) as long ago as 1965. The following year, the Lockheed Space and Missile Corporation, now DIALOG Information Services Inc., introduced its DIALOG system, an upgraded version of which became operational in 1969. After a developmental period which extended into the early seventies, the use of online bibliographical systems became widespread in the United States and more recently in Europe. Batch processing of searches is now virtually obsolete, and even theses searches can be entered online. Even if the search for the end-user is not carried out in his own library or information centre by an intermediary or, increasingly, by himself, an external information centre will process the search online over its own terminal against a remote database and mail the search product to the end-user rather than batch process a series of searches against an in-house database on its own computer.

MEDLINE (Medical Analysis and Retrieval System On-line) was one of the first online bibliographic databases to be developed and made generally available, and subsequently it has become one of the most heavily used of the wide range of databases which can now be accessed online. The US National Library of Medicine (NLM), following the example of the American Chemical Society, started using a computerized system to produce its monthly subject index to the world's medical literature *Index medicus* in 1963, and shortly after this date on-demand, one-off, retrospective searches were being run against the database in batch mode. In the autumn of 1967 the NLM began to experiment with online access to its database, when a contract was signed with SDC, who installed and evaluated their ORBIT (On-line Retrieval of Bibliographical Information Time-shared) system at the NLM, using a small database in the field of neurology. A pilot online experimental service offering access to a subset of the *Index medicus* database became operational in June 1970 to about 90 medical institutions in the United States, giving access to *Abridged index medicus (AIM)* which covered 100 of the most important journals. This project utilized SDC's IBM 360/67 computer linked by the Teletypewriter Exchange Network (TWX) to remote terminals in medical institutions throughout the country. The response from users was enthusiastic, and a fully developed MEDLINE system giving access to a database containing references from 1,200 of the journals covered by *Index medicus* in one online file (ie 'database') was introduced in October 1971. The current MEDLINE

files which cover over 2,500 source journals are now marketed throughout the world by a variety of suppliers including BLAISE-LINK in the United Kingdom, DIALOG from the United States and Data-Star from Europe.

In Europe the European Space Research Organization (ESRO) Space Documentation Service (SDS) established in 1965 was one of the first centres to provide online searching facilities. The service gave access to a number of databases of relevance to space research and technology, eg the NASA database, *Chemical abstracts*, *Biological abstracts* etc from a host computer located at Frascati, near Rome, using the RECON software developed under contract by Lockheed. In the United Kingdom the Department of Industry's DIALTECH service began to make dial-up access available to the twelve RECON databases in October 1970 via a minicomputer at Orpington, Kent. The European Space Agency was founded by twelve member states in 1975 to supersede ESRO, and the RECON software was replaced by the more sophisticated QUEST programs in the Information Retrieval Service (IRS) which replaced SDS late in 1979.

The planning of the European online network and retrieval service EURONET/DIANE began in 1971 and the service became operational in 1980. EURONET is the data transmission facility developed by the post, telegraph and telephone authorities (PTTs) of the EEC countries to provide users inside the Community with the means of access to bibliographical and referral databases while DIANE (Direct Information Access Network for Europe) is the ensemble of the service suppliers available within the Community. Among the first suppliers (or 'hosts') in the new network were ESA/IRS and the British Library's BLAISE system. Currently about 370 databases and databanks are available through EURONET via about 40 hosts.[1]

The British Library's BLAISE (British Library Automated Information Service) was introduced in April 1977, based on the ELHILL software developed for the US National Library of Medicine's MEDLINE service by SDC. The initial databases to be made available were in the biomedical field – MEDLINE, CANCERLINE, TOXLINE etc – and these were soon followed by the British and American MARC databases and later by others including *Conference proceedings index* and *British education index*. In 1982, BLAISE divided their service into BLAISE-LINK – a link with the National Library of Medicine for searching the biomedical and related databases, and BLAISE-LINE, which continues to offer the non-biomedical databases from Harlow, United Kingdom.

Significant European services now include Data-Star and Pergamon