



Atherton, Pauline A.

Putting Knowledge to Work; An American View of Ranganathan's Five Laws of Library Science
(Sarada Ranganathan Endowment for Library Science, Series 5).

VIKAS Publishing House, 1970.

Copyright ©Sarada Ranganathan Endowment for Library Science (SRELS)

Digitization: Susan Ditch, SIRLS, University of Arizona

Quality Control: Megan Plesea, SIRLS, University of Arizona

Digitization Training: Han Yan, Information Systems, University of Arizona & dLIST Editor

Project Coordinator: Cheryl K. Malone, Assoc. Prof. SIRLS, University of Arizona & dLIST Editor

Digitized: Spring 2007

Acknowledgments: SRELS (A. Neelameghan, K.N. Prasad, K.S. Raghavan, DRTC) and dLIST
Advisory Board Member, S. Arunachalam (MS Swaminathan Research Foundation)

dLIST Classics (Book) Editor: Barbara Hutchinson, University of Arizona

dLIST Editor-in-chief: Anita Coleman, University of Arizona

dLIST Classics

Titles by S.R.Ranganathan

Five Laws of Library Science, Ed. 1 (1931)

Philosophy of Library Classification (1973)

Prologemena to Library Classification, Ed. 3 (1967)

Classification and Communication (1951)

Documentation Genesis and Development (1973)

Documentation and its Facets (1963)

Library Book Selection, Ed. 2 (1966)

New education and school library: Experience of half a century (1973)

Reference Service, Ed. 2 (1961)

Other titles

S.R. Ranganathan's Postulates and Normative Principles: Applications in Specialized Databases
Design, Indexing, and Retrieval, 1997. Compiled by A. Neelameghan.

Memorabilia Ranganathan: A compilation of useful quotations of S.R. Ranganathan from his various
works, 1994.

Putting Knowledge to Work: An American View of the Five Laws of Library Science, 1970. By
Pauline Atherton.

Read the dLIST Classics online!

<http://dlist.sir.arizona.edu/>



CHAPTER G

LAW 3: EVERY BOOK ITS READER — HOW EVERY BOOK BE FOUND BY ITS READER

1. Catalogue and Readers' Requirements

11 STRUCTURE OF CATALOGUE

Cataloguers have a view of library service which in many ways is most distorted and at the same time most central. Without their efforts at bibliographic control where would library service be? Without some form of listing the items in a library, how could anyone ever find anything? Nevertheless, some people are now asking — even with the help of the cataloguer's effort, we do not seem to find what is available? Has the time come for us to carefully analyse what happens when a cataloguer performs the present ritual of describing books for readers? Many of us have embarked upon such a study only recently because the availability of MARC (Machine Readable Cataloguing) has given us access to current catalogue information in a form easy to manipulate for study. The accessibility of all the information on a catalogue card as access points instead of only three or four elements (main entry, title, subject, call number) makes it pos-

sible to match the individual user's words used in a subject search with any portion of the catalogue record allows us to study our present cataloguing rules for main and added entry and test their potency compared to other elements.

Some of us (including students in library schools, such as Syracuse) are asking:

1 Is the main entry form we establish for corporate bodies really useful to the reader in search of a particular book?

2 Is our rule for main and added entries for works by more than three authors valid if we are trying to describe a book accurately?

3 Are there other items of information on the catalogue card that are more descriptive of the book's content than the main and added entries we have established for it? If yes, shouldn't these be more accessible for the user in search of material on a given subject?

4 Is our strict control over the names of subjects we use in cataloguing essential if we maintain an integrated bibliographic retrieval system, with catalogues, name of subject authority files, and biographical reference sources available at one and the same time in the computer?

12 VIEW AS A WHOLE

I am not sure that we can completely uproot all

of our present practices and rules for descriptive and subject cataloguing. Nor am I sure that we can indict today's cataloguers as persons with a myopic view, that is, do they see service to users for the mountain of books still needing cataloguing. It is safe to say that their work does necessitate their view of each and every book as an entity to be handled separately. Only rarely do they get a chance to view the whole scene of which their individual effort is but a small part. Many of them view their work as assembly, or piece work. Most libraries are not equipped with aids or devices which would facilitate the work of cataloguers so that they could see the true impact on the library catalogue as they catalogue individual works. How many cataloguers know how close their description of a book comes to matching the user's clue words when he is in search of that book? The two studies I discussed in the previous lecture provide some useful information about book descriptions in card catalogues and their use by university scholars. This brings us a little closer to an awareness of this important relationship—the book and the reader—but we still have a long way to go.

2 Investigation

In this lecture I am quite proud to report on some work a few students at Syracuse have done to enlarge our understanding of this subject. With

the availability of MARC tapes from the Library of Congress and some computerized retrieval programs from IBM, we were able to study cataloguing records very carefully, and compare the retrieval of books from a catalogue having traditional access points with retrieval from a catalogue with unlimited access points.

21 MARC FORMAT

Before I begin to discuss our findings, perhaps an explanation of the MARC format is necessary. It is not important for this discussion to be very detailed in this explanation, nor for that matter, entirely up-to-date. For example, we began our work using MARC I records. This was during the MARC Pilot Project at the Library of Congress. Since March 1969, the Library of Congress has been distributing MARC II records. The MARC II format is much more elaborate in identifying and tagging fields and elements of information. Our work with MARC I records could also be done with MARC II records, if we chose. For the purposes of our study, however, which format we used was not as important as the information content. It was the content of a Library of Congress catalogue card, created by cataloguers following the rules for main entry, descriptive cataloguing, subject cataloguing and classification that we were scrutinizing. What they created was now accessible in ways they never

**MARC Pilot Project
Bibliographic Listing**

0) LC Card No. 66-018912

Filed Field Information :

1) <i>Type of Entry</i>	2) <i>Form of Work</i>	3) <i>Biblio</i>	4) <i>Illus</i>	5) <i>Maps</i>	6) <i>Number</i>	7) <i>Conference</i>	8) <i>Work No</i>
Personal Author	Monograph	Yes	None	None	None	No	No
9) <i>Language Data:</i>				10) <i>Publication Data:</i>			
<i>Class</i>	<i>Lang 1</i>	<i>Lang 2</i>	<i>Key Date 1</i>	<i>Date 2</i>	<i>Place</i>	<i>Name</i>	<i>Height</i>
Translation	Eng	Unk	Single	1966	NYNY	PGR	21CM

<i>Name of Variable Field</i>	<i>Tag</i>	
Main Entry	10	Mao, Tse-Tung, 1893-1
Title Statement	20	Basic Tactics. Translated and with an Introd. By Stuart R Schram. Foreword by Samuel B Griffith II
Imprint Statement	30	New York, PRAEGER 1966
Collation Statement	40	VIII, 149 P. 21 CM
Notes	60	Bibliographical Footnotes.
Subject Tracing	70	Guerrilla Warfare.
Pers Author Tracing	71	Schram, Stuart R, Ed and Tr
Title Tracing	74	T
L C Call Number	90	U240.M28 1966
Dewey Class Number	91	355.425

Mao, Tse-tung, 1893—
Basic tactics. Translated and with an introd by Stuart R Schram. Foreword by Samuel. B Griffith, II. New York, Praeger [1966]-

viii, 149p. 21 cm.

Bibliographical footnotes.

1 Guerrilla warfare I. Schram, Stuart R. and tr. II. Title
U240.M28 355.425 66-1891
Library of Congress

imagined, and we were curious to find out how useful their descriptions of books were for machine searching.

The figure of a MARC I record explained and set beside the printed catalogue card for Mao Tse-tung's book entitled *Basic tactics*, shows how every field of information on the catalogue card is not only analysed but explicitly tagged.

In the actual machine readable record on magnetic tape or punched cards, the field tag 10 would appear directly before the main entry line, field tag 20 would appear before the title statement, tag 70 would appear before the subject entry, tag 90 before the Library of Congress call number, and tag 92 before the Dewey class number. Additional information about the work in hand is described in coded fashion, in what are called fixed-length fields. For example, in character position 21 on the MARC I record, a code would appear which indicated the type of main entry. For example, Mao Tse-tung is a personal main entry. Therefore, the code P would appear in character position 21. If the book's main entry were a government body, a code letter G would appear in character position 21, S for Society, R for Religious body, T for Title main entry, and so on. If the notes section of the catalogue card made reference to a bibliography, there

would be an X in character position 23. On this figure all of these fixed field codes have been translated for ease of use of the reader. For example, the listing reads "yes" under *Biblio* in the Fixed Field portion. This is a translation of the coded information on the MARC record, indicating that there is an "X" in character position 23.

211 *Reformatted Marc Record*

When we reformatted the MARC records for use with the IBM Document Processing System, we knew that every group of characters on the MARC record bounded on either side by spaces would be considered a keyword, accessible for searching. That meant that the Dewey Class Number was a Keyword, by definition. So was every word in the title statement including *and*, *with*, *by* (unless we instructed the computer to ignore them as common words). This also meant that every word in a subject heading was now a keyword, not only the first word as is true in our catalogues. As if through a glass darkly, I saw the first glimmerings of something like a classified catalogue approach available to American Library users! Is this too big a concept to grasp over so many miles? The workings of the computer program for accessing the information on the MARC records is fairly complicated, but the use of the MARCS/DPS system as we

called it was not — luckily. All that the user needed to do was to structure his searches in English-like sentences, combining keywords with logical operators (such as *and*, *or*), availing himself of some useful modifiers or delimiters. He could for example specify that two keywords had to be side-by-side in a MARC record or he should not want to see the record. He could specify which field (or sentence) to search for his particular keyword. This was very useful when searching the file by using Dewey class numbers as keywords. A three-digit Dewey number looks like any other three-digit number to the computer, and we knew that the collation contained the number of pages in a book, and numbers appeared elsewhere. By phrasing his search in the following way, he could exclude any three-digit numbers other than those in the Dewey Class number field:

L1 355 (\$);

L2 L1 AND T92(SEN);

We knew that the field tag numbers would be confounded with other two-digit numbers when the MARC record was processed by the Document Processing System, so we instructed the computer programmer to insert the letter T before every field tag number. This gave us a group of unique keywords representing the MARC field tags.

3 Library Education Experimental Project

31 FACILITY

With such a laboratory facility at our disposal, we proceeded to study catalogue records and test the finding of books by new search strategies. We called this laboratory LEEP (Library Education Experimental Project). During its first year-and-a-half, it was supported by the U S Office of Education. Now it is maintained by Syracuse University as an integral part of the educational facilities available to students and faculty in the School of Library Science.

32 SAMPLE CATALOGUE

For the studies reported here in detail we used a sample catalogue of approximately 10,000 books in the social sciences, representing titles published primarily in 1967-68. It was a stratified sample of the larger MARC Pilot Project file of approximately 48,000 titles. We were able to create such a sample by computer manipulation of the file because we had analysed the content of the entire file during the earlier stages of LEEP.

33 MAJOR OBJECTIVE

Most of the projects the library school students developed centred around the third question I asked at the outset of this lecture: Are there other fields of information on the catalogue card that are more

descriptive of the book's content than the main and added entries we have established for it?

If the answer was found to be yes, we intended to tell the world of librarianship so they could begin to consider making such data more accessible. This new information, we felt, would help justify library automation and MARC records not on the basis of efficiency of library operations but on the basis of the third law of library science, namely *every book its reader*.

34 SEARCHES

Many students, more than two hundred in fact, performed searches using MARCS/DPS. They performed these searches while they were students in reference and bibliography classes, cataloguing and technical services classes, or in advanced courses in information systems. Several students went beyond one or two searches, and two of them organised their work into a small research study and wrote down their findings in a fashion suitable for publication. One report is in the process of being published by *Library resources and technical services* and the second is almost ready for submission to the *Journal of the American Society for Information Science*. My report of their work will be the first public report. I have their permission to report their work to you and I am indeed proud to do so.

4 Study 1

41 OBJECTIVE

One student, Mrs Judith Hudson was interested in finding all the relevant information in the MARCS/DPS file for *area studies* bibliographies. This is not a simple matter in a traditional card catalogue and she was not sure which would be the most efficient way to find documents in a machine-based catalogue, such a MARCS/DPS. She set out to:

- 1 Discover the optimum search strategy to find at least 95% of the relevant documents in our computer-based catalogue on an area of the world;

- 2 Discover what other methods to use if one search strategy is not optimum; and

- 3 Determine why non-relevant document references would be found.

If possible, she hoped that her study would allow her to recommend search strategies which would minimize "false drops" and maximize "hits."

42 DEFINITION

The first step in constructing the searches was to begin defining terms and geographic areas. Area studies or area research are considered to be "the interdisciplinary study of a given geographic region,

country or group of countries (considered as a unit), which takes account of the socially significant data about the area, including the history of the people, their economic and political development and their use of natural resources" (24). "Area studies . . . may treat the topography, ethnography, natural resources, religion, sociology, economy, law, government, history, culture (or any combination of these) for an area, large or small" (23).

The area to be studied was the Middle East. Later on in the study, it was decided to examine another area as a check on the results of the Middle East Study. Latin America was chosen.

It was decided that "the Arab countries, inclusive of North Africa, together with Israel, Turkey, Iran and Afghanistan — often called collectively the Near East or Middle East . . . constitute a regional unit" (4). Latin America was defined as "that part of the Western hemisphere that lies south of the border of the United States and of the Florida Strait. It includes Middle America (Mexico, Central America and the West Indies) and South America" (3).

43. SEARCH STRATEGY

In order to find all the relevant MARC records from the file, four different types of searches were written:

1 Searches for keywords—English language words and phrases—for example, Middle East, Palestine, etc.

2 Searches for *Library of Congress* (LC) *Classification* numbers on the area—for example, DS36, F1869, etc.

3 Searches for *Dewey Classification* numbers on the area—for example, 956, 972, etc.

4 “*If-scan*” searches—The “If-scan” function enables the user to scan a specified field of the Record portion of the MARC/DPS record for a given character or character string. In this case, we specified that the computer scan the field called DDCNO, which is the Dewey Classification number, for an area code. Such a search for the area code 56 (Middle East) would yield such Dewey numbers as 560, 956, 321.00956, etc.

44 PREPARATION FOR SEARCH

441 *Search by Keyword*

The first step in preparing the keyword searches was to compile a list of words and phrases which would yield records relevant to an area study. We examined the Dewey and LC Classification Schedules for the two areas and listed all of the relevant words and phrases we could find. To this list were added such words as geographical names, names of

religions, and famous people representative of the area, etc. Each word and phrase was checked in our index list of keywords in MARCS/DPS to see if the word appeared in the file of 8,000 MARC records. If a word was not listed in this index, it was discarded.

The words which remained formed the *keyword searches*. They were combined according to DPS vocabulary and syntax rules in such a way as to find every MARC record which contained any one or more of the specified words or phrases regardless of where it might appear in the text portion.

442 *Search by Class Number*

The LC and Dewey Classification searches were prepared by examining the LC and Dewey schedules to determine the appropriate class number on the area. These class numbers were checked in the 'index list'. If they appeared in this index, the DPS search was written to find every MARC record which contained the specified LC or Dewey Classification numbers.

443 *"If-scan" Search*

The "If-scan" search was prepared for the applicable area notations from the Dewey Classification Area Table (2). All MARC records having a Dewey class number built with these area notations were to be retrieved.

45 EXAMPLES

Searches were run on MARCS/DPS. The Middle East searches were as follows:

451 *Keyword Searches*

Search Number 206

- L1 DEAD & SEA(+1);
- L2 MOUNT & CARMEL(+1);
- L3 TEL & AVIV(+1);
- L4 IBN & SAUD(+1);
- L5 BENI & KURT(+1);
- L6 PALESTINE, ISRAEL, JORDAN,
ARAB(\$);
- L7 ANTIOCH, PALMYRA, SMYRNA,
SUMERIANS, TYRE, UR;
- L8 JEW(\$), JERUSALEM, BETHLEHEM,
EMMAUS, MEGIDDO, NAZARETH;
- L9 NEGEV, NEGEB, CAPPADOCIA,
HEROD, HUSSEIN, JUDEA, AFGHA-
NISTAN;
- L10 L6 & T10 (NOT SEN);
- L11 PERGAMON & T30 (NOT SEN);
- L12 MEDINA, DARIUS, BAHRAMPOUR,
ISMAIL, MAHMUD;
- L13 L1, L2, L3, L4, L5, L6, L7, L8, L9,
L10, L11, L12;

LIST OFFLINE, TEXT, SUBJECTS

Search Number 207

- L1 MIDDLE & EAST(+1);
- L2 NEAR & EAST(+1);
- L3 UNITED & ARAB(+1) & REPUBLIC
(+1);
- L4 TURKEY, TURKS, CYPRUS, SYRIA,
LEBANON;
- L5 IRAQ, IRAN, PERSIAN, MESOPO-
TAMIA, EGYPT(\$);
- L6 ASIA & MINOR(+1);
- L7 MOROCCO, ALGERIA(\$), TUNIS(\$),
LIBYA, SUDAN;
- L8 ARMENIA, ISLAM(\$), BEDOUIN;
- L9 EUPHRATES, ANATOLIA, ANKARA,
ARARAT, ADEN, MOSQUE, MUSLIM
(\$);
- L10 BABYLON(\$), KURDISH, KISH,
BEIRUT, LEVANT;
- L11 L1, L2, L3, L4, L5, L6, L7, L8, L9, L10;

LIST OFFLINE, TEXT, SUBJECTS;

452 *LC Searches*

Search Number 208

- L1 DS1(\$), DS2(\$), DS313(\$), DS326(\$),
DS38(\$), DS43.S5, DS44(\$), DS48. F53,
DS49(\$);



Atherton, Pauline A.

Putting Knowledge to Work; An American View of Ranganathan's Five Laws of Library Science
(Sarada Ranganathan Endowment for Library Science, Series 5).

VIKAS Publishing House, 1970.

Copyright ©Sarada Ranganathan Endowment for Library Science (SRELS)

Digitization: Susan Ditch, SIRLS, University of Arizona

Quality Control: Megan Plesea, SIRLS, University of Arizona

Digitization Training: Han Yan, Information Systems, University of Arizona & dLIST Editor

Project Coordinator: Cheryl K. Malone, Assoc. Prof. SIRLS, University of Arizona & dLIST Editor

Digitized: Spring 2007

Acknowledgments: SRELS (A. Neelameghan, K.N. Prasad, K.S. Raghavan, DRTC) and dLIST
Advisory Board Member, S. Arunachalam (MS Swaminathan Research Foundation)

dLIST Classics (Book) Editor: Barbara Hutchinson, University of Arizona

dLIST Editor-in-chief: Anita Coleman, University of Arizona

dLIST Classics

Titles by S.R.Ranganathan

Five Laws of Library Science, Ed. 1 (1931)

Philosophy of Library Classification (1973)

Prologemena to Library Classification, Ed. 3 (1967)

Classification and Communication (1951)

Documentation Genesis and Development (1973)

Documentation and its Facets (1963)

Library Book Selection, Ed. 2 (1966)

New education and school library: Experience of half a century (1973)

Reference Service, Ed. 2 (1961)

Other titles

S.R. Ranganathan's Postulates and Normative Principles: Applications in Specialized Databases
Design, Indexing, and Retrieval, 1997. Compiled by A. Neelameghan.

Memorabilia Ranganathan: A compilation of useful quotations of S.R. Ranganathan from his various
works, 1994.

Putting Knowledge to Work: An American View of the Five Laws of Library Science, 1970. By
Pauline Atherton.

Read the dLIST Classics online!

<http://dlist.sir.arizona.edu/>



G452 PUTTING KNOWLEDGE TO WORK

DS53(\$), DS62(\$), DS70.6(\$),
DS71.07(\$), DS80.9(\$), DS84(\$),
DS89.T(\$);

L3 DS99(\$);

L4 L1, L2, L3;

LIST OFFLINE, TEXT, LCCALL:

453 *Dewey Searches*

Search Number 209

L1 915.6(\$), 955(\$), 956(\$);

L2 L1 & T92(SEN);

LIST OFFLINE, TEXT, DDCNO;

46 RECORD OF SEARCH RESULT

The method used for recording the information gathered in the searches was straightforward: Once run, each of the searches resulted in a print-out, with statistics and document references which met the search criteria. Each MARC record in the MARC/DPS file has a unique number and that document number was used for identification of individual MARC records.

Each MARC record selected was examined and the following information was recorded:

- 1 Document number;
- 2 Is it relevant (R), questionable (Q), or not relevant (NR)?;

- 3 Search number;
- 4 Kind of search; and
- 5 Keyword(s) which caused the record to be selected.

For keyword searches other than LC or Dewey class numbers, the location (by MARC tag number) where the keyword appeared was recorded. For example, here's one document record:

Doc No 4469		RXQ NR
Search No	Kind	KW
208	LC	DS 80.9
209	DC	956.92
210	IFSc	956.92
207	KW	Lebanon

Side 1 (For side 2 see next page) :

47 FINDINGS

The findings of Mrs Hudson's study are a little startling for any of us who hoped to prove that searching the file by class number would be more

Doc No 4469					
KW	T10	T20	T60	T70	Other
Lebanon		X		X	

Side 2

useful than any other keyword approach. The following table points up how few relevant documents were found using either the LC or DC numbers, when compared to keywords found anywhere in the MARC record.

AREA : MIDDLE EAST

Search Strategy	Relevant Hits	Precision ratio	Recall Ratio
		$\frac{R \text{ ret}}{T \text{ ret}}$	$\frac{R \text{ ret}}{\text{Total R}}$
KW	101	40.6%	97.1%
LC	36	14.4	34.6
DC	26	10.4	25.0
IF Sc	21	8.8	20.1
KW & LC	103	41.3	99
KW & DC	103	41.3	99
KW & IF SC	102	40.9	98
KW & LC & DC	104	41.7	100

AREA : LATIN AMERICA

Search Strategy	Relevant Hits	Precision ratio	Recall ratio
		$\frac{R \text{ found}}{T \text{ found}}$	$\frac{R \text{ found}}{\text{Total } R}$
KW	143	33.4%	98.6%
LC	60	14.1	41.3
DC	47	11.0	32.4
IF SC	36	8.4	24.8
KW & LC	145	33.9	100
KW & DC	144	33.7	99.4
KW & IF SC	143	33.4	98.6
KW & LC & DC	145	33.9	100

471 *Annotation*

Total R is 104 (Middle East), 145 (Latin America)

T found for any single search strategy is the sum of documents found using it alone or in any combination with other strategies.

48 ANALYSIS OF FINDINGS

The analysis of the placement of the keywords in the MARC record is at the heart of the problem. Where do the keyword hits and false drops come from? The table on page 106 tells an interesting story which looks consistent for area studies searches

KEYWORDS IN DATA FIELD	MIDDLE EAST			LATIN AMERICA		
	RELE- VANT	QUES- TION- ABLE	NON-RE- LEVANT	RELE- VANT	QUES- TION- ABLE	NON-RE- LEVANT
<i>Single field hits:</i>						
T10 (author)	0	0	9	0	0	11
T20 (title)	8	1	12	9	5	15
T30 (imprint)	0	3	11	0	1	10
T60 (notes)	2	2	3	3	2	16
T70 (subject heading)	19	1	7	16	5	5
<i>Hits in more than one field:</i>						
T20/T70 (title and subject heading)	70	0	7	49	4	15
T20/T60/T70 (title, noter and subject heading)	9	0	0	7	0	1

Table 4: Keyword Hits by MARC Data Field

for two parts of the world. The fields for author and imprint often yielded non-relevant reference, or false drops. These same fields were never the only reason for the retrieval of a relevant MARC record. It appeared that keyword searches restricted to fields for the title, series notes, notes, and subject headings eliminated many false drops without sacrificing the finding of any relevant records. The table also shows that the fields for title *and* subject heading (T20 and T70 in the MARC I records respectively) are the most effective for finding relevant documents. Keywords in the title field can be both useful and a nuisance, according to these findings.

To find documents on an area of the world, then, we concluded that the keyword search seems optimum, with the use of LC numbers and then Dewey class numbers following for exhaustivity. The homonym problem caused a great many false drops during such searches and seems unavoidable with an "uncontrolled" vocabulary of keyword searching such as that done in MARC/DPS. For example, Israel (the country) and Israel (the surname) cannot be distinguished.

491 *Conclusion*

Mrs Hudson concluded her papers as follows: The fact that keyword searches were better than

searches using a classification system was not surprising. Material on a given geographic area is not easily found on the shelf in any library. It is scattered because other aspects of the material are highlighted and used for primary subject access. The subdivisions of subject headings, more than the geographical codes in D C, appear to provide access to some of this material. To access the subdivisions, however, a catalogue record must be in machine readable form and the programming system used for searching the records must allow access to any word in the subject field, not just the first word. MARC/DPS is such a system. Our results show that two MARC I fields (title and subject) provide access to the relevant documents for an area study search. Neither is sufficient alone.

We are not certain that our findings hold true for other kinds of subject searches (for example, biographical, political, historical, literary, etc). Much more work along the lines described here will have to be done before the class numbers (either L C or D C) on MARC records should be discarded as subject access points in a mechanized document finding system. Before we can say that the title and subject heading fields should be considered the primary routes of subject access to library materials, someone will have to test the entire system more rigorously than we have here.

5 Study 2

51 OBJECTIVES

Michael Ubaldini, a recent graduate of the School of Library Science at Syracuse University, delved a little more deeply into MARC searching than Mrs Hudson. He refined some of the objectives of her work and pinpointed for very close scrutiny the placement of keywords in various MARC fields. He had eight objectives when he undertook his research:

1 To analyse *keyword searches* (phase 1) to determine if there were any well-defined *relationships* between finding relevant documents and specifying which tagged MARC fields should be searched.

2 If such relationships existed, to determine if an increase in the *precision ratio* could be effected by such field specification.

3 If so, to determine the subsequent effect on the *total estimated recall*.

4 If the effect on the total estimated recall was favourable, to determine the sequence in which the tagged fields should be searched to insure high precision and equally high recall.

5 To determine the effect of restricting searches to *single fields*; to compare these results with *multiple-field* searching; and to conclude which method,

if either, is better for obtaining high precision and recall.

6 To determine what part, if any, the following *search characteristics* played in the question of tagged field specification: subject area searched, search author, length and degree of specificity of the search.

7 To construct keyword, L C, and D C searches (phase 2) on one topic to *test* the results of phase 1 (steps 1-6 above).

8 To compare the results concerning tagged field specification of Judy Hudson's report on *area studies* with the results of his work and to determine if her recommendations could be extended to all *keyword* searches.

The concept of single-field versus *multiple-field* searches is an important one. Some document finding systems allow selection only by single-field keyword occurrences. The KWIC indexing of article *title words* (analogous to field T20 in the MARC I records) is an example. It allows access to the contents of these articles solely by the indexing of keywords in the titles of the articles indexed. This may, in fact, be desirable. Allowing specified keywords to be indexed and accessed from multiple parts of an article (for example, abstract, affiliation) may cause the finding of many irrelevant drops,

TABLE 8: *Unique* documents found by *single-field* keyword occurrences (Analogous to KWIC indexing)

FIELD	Total Docs found	Relevant	Non-rel.	Quest.	Precision Ratio %	T Estimated Recall %
T10 (main entry)	4	3	0	1	75	3
T20 (title)	36	30	5	1	83	27
T30 (imprint)	4	0	4	0	0	0
T40 (collation)	0	-	-	-	-	-
T51 (series)	3	0	3	0	0	0
T60 (notes)	9	6	3	0	67	5
T70 (subject headings)	32	23	6	3	72	21
T7+ (corporate author)	6	1	5	0	16	1
T80 (copy information)	0	-	-	-	-	-
TOTAL RESULTS	79	51R	23N	5Q	65	48

TABLE 9: Documents found by *multiple-field* keyword occurrences (Phase I results)

FIELD	Total Docs found	Relevant	Non-rel.	Quest.	Precision Ratio %	T Estimated Recall %
T10 (main entry)	27	13	10	4	48	12
T20 (title)	122	78	30	14	64	70
T30 (imprint)	20	9	8	3	45	8
T40 (collation)	2	2	0	0	100	2
T51 (series)	25	5	18	2	20	5
T60 (notes)	59	30	20	9	51	27
T70 (subject headings)	107	76	19	12	71	69
T7+ (corporate author)	28	10	17	1	36	9
T80 (copy information)	1	1	0	0	100	1
TOTAL RESULTS*	162	95	44	23	59	86

* These total results do not represent the arithmetic sums of their respective columns because one keyword may have appeared in many fields of a document retrieved by the keyword strategy.

TABLE 10: Documents found by *multiple-field* keyword occurrence (Phase 2 results)

	Total	Documents Retrieved			Precision Ratio %	Total Est. Recall %
		Relevant	Non-rel.	Quest.		
Documents retrieved by <i>all</i> strategies	43	14	20	9	33	100
Documents retrieved by <i>keywords</i>	19	12	1	6	63	86
Documents retrieved by <i>LC</i> class nos	16	6	8	2	38	43
Documents retrieved by <i>DC</i> class nos	23	10	12	1	44	72
<i>Field (s) specified or keyword search</i>						
T10 (main entry)	0	-	-	-	-	-
T20 (title)	6	5	0	1	83	36
T30 (imprint)	0	-	-	-	-	-
T40 (collation)	0	-	-	-	-	-
T51 (series)	0	-	-	-	-	-
T60 (notes)	3	1	1	1	33	7
T70 (subject headings)	15	11	0	4	73	79
T80 (copy information)	0	-	-	-	-	-
T20 or T70	16	11	0	5	69	79
T20 or T60 or T70	19	12	1	6	63	86
T10 or T20 or T60 or T70						
T20 or T30 or T60 or T70	19	12	1	6	63	86

but may insure a substantially higher number of additional relevant documents to be selected.

52 FINDINGS

Ubalдини's findings show interesting differences if single-field searches are compared with multiple-field searches.

53 CONCLUSION

He summarised his results as follows:

1 Analysis of the table of single-field keyword occurrences indicates that the most useful fields for finding *relevant* MARC records may be ranked as follows:

- T20 (title)
- T70 (subject headings)
- T60 (notes)
- T10 (main entry)

The ranking would be different if the precision ratio were used as the sole criterion because the number of non-relevant retrievals in these fields varies.

2 Analysis of the table of multiple-field keyword occurrences supports the validity of this sequence. The table which records the results of the test keyword search (phase 2), shows that all 12 relevant documents were retrieved by specifying only fields T20, T60 and T70 to be searched.

3 Single-field keyword searching was not as effective as multiple-field searching. 79 of a possible 162 documents were retrieved by the single-field keyword strategy. Of these 79 documents, 51 were assessed as relevant (a precision ratio of 65 per cent). By allowing keyword occurrence in all fields, 95 relevant documents were retrieved (a precision ratio of 59 per cent). Multiple-field retrievals gave a slightly lower precision ratio, but the increase in the total estimated recall was critical, an increase from 48 per cent to 86 per cent. This would mean, for example, that a KWIC indexing of all the fields of the MARC record would furnish less than half of the relevant documents in the data base while multiple-field searching would furnish over 6/7 of the relevant documents. From another viewpoint, allowing multiple-field searching nearly doubled the total number of relevant documents retrieved, with little cost to the overall precision.

4 Using MARCS/DPS, relevant documents were selected by the occurrence of one or more of a number of specified keywords in the *titles* of these documents (field T20). Manually searching an alphabetically arranged card catalogue by the title of a document would have selected only those documents which had a specified keyword as the *first word* of its title. Analysis of the first indexable word of a sample of 20 relevant documents

selected by searches analysed in this report indicates that only 40 per cent of these documents would be retrieved by a manual search of a card catalogue which contained entries for these documents. That is, for the searches studied here, approximately only 28 per cent of the total number of relevant documents in the card catalogue would be selected by a manual search compared to the retrieval of 70 per cent by use of computer access.

5 Using MARCS/DPS allowed the keyword searching of the *notes* portion of a catalogue entry (field T60). Some of the documents selected in this study could not have been found by a manual search of a card catalogue under title or subject heading. One example is the case of a compilation of articles in which the article titles are given in the notes. A few of the documents retrieved by the search on Shakespearean drama, for example, were retrieved solely by keyword occurrence in the notes. In the light of the results of this report, the ability to search this part of a catalogue entry for specified keywords may be a very valuable facility.

6 In summary, when one considers the title, notes, and subject entries, the occurrence of specified keywords in these fields taken together rather than in any *one* of these fields alone, the total number of relevant documents found is increased by

26 per cent while the precision ratio is decreased by only 3 per cent.

6 Comparison of Findings

The following statement of Mrs Hudson is corroborated by the results of Mr Ubaldini's report: "The fields for *title* and *subject heading* (T20 and T70 respectively) are the most effective for retrieving relevant MARC records." However, her statement that the "author, imprint, and title fields (T10, T30, and T20 respectively) account for most of the non-relevant keyword hits," may not be extended to all keyword searches as is evidenced by the analyses in Ubaldini's report.

In her report, Mrs Hudson recommends that the following fields be searched to obtain high precision and recall in an area study: "filing or conventional title (T15), title statement (T20), series, traced (T50), series, not traced (T51), notes (T60), and subject headings (T70)." The results of Ubaldini's report do not confirm the wholesale extension of her recommendations for area studies to all keyword searches. They do confirm the searching of fields title, notes, and subject. However, since *no* keyword occurrences in his study were noted in fields for conventional title or series (traced), it is not expedient to recommend that they be searched in all keyword searches. In his report, field

T51 (series, not traced) was responsible for retrieving 3 documents (all non-relevant) by single-field keyword occurrence and 25 documents (18 non-relevant) by multiple-field keyword occurrence. In multiple-field searches, T51 was responsible for selecting 5 relevant documents, all of which would have been selected anyway if field T51 had been excluded from being searched. Therefore, the recommendation that field T51 be searched in an area study does not appear to be easily extended to all keyword searches.

7 Remarks

Ubalдини's conclusions have great impact on the future of cataloguing as we know it today. If we are to continue to obey the third law, it is possible that we will have to revise our practices quite dramatically. We cannot ignore much longer such statements as the following: The most productive fields for subject searching, as determined by the results of this report, are title, notes and subject headings. Specifying additional MARC fields for searching will either result in (a) a greater decrease in precision with little increase in recall, or (b) the selection of *no* additional MARC records.

Requiring that all the keywords (specified in a search) be found in a single field *critically reduces* the number of relevant MARC records selected by

the search. The resulting recall ratio is likely to be lower than is desired by the users of the system.

The use of LC subject headings as a means of access to catalogue entries of monographs may not be as useful as is generally accepted. It appears, from the results of this report, that a KWIC index of the monograph titles for specified keywords would have selected a greater number of relevant MARC records than searching the subject headings for these keywords, without appreciably lowering the precision ratio of the search. The necessity for LC subject headings in the MARC record is put in serious doubt. Nevertheless, the results of this report also point to the conclusion that searching the subject heading field *in addition to* the title and notes fields did select many more relevant records than searching the titles and subject headings alone.



Atherton, Pauline A.

Putting Knowledge to Work; An American View of Ranganathan's Five Laws of Library Science
(Sarada Ranganathan Endowment for Library Science, Series 5).

VIKAS Publishing House, 1970.

Copyright ©Sarada Ranganathan Endowment for Library Science (SRELS)

Digitization: Susan Ditch, SIRLS, University of Arizona

Quality Control: Megan Plesea, SIRLS, University of Arizona

Digitization Training: Han Yan, Information Systems, University of Arizona & dLIST Editor

Project Coordinator: Cheryl K. Malone, Assoc. Prof. SIRLS, University of Arizona & dLIST Editor

Digitized: Spring 2007

Acknowledgments: SRELS (A. Neelameghan, K.N. Prasad, K.S. Raghavan, DRTC) and dLIST
Advisory Board Member, S. Arunachalam (MS Swaminathan Research Foundation)

dLIST Classics (Book) Editor: Barbara Hutchinson, University of Arizona

dLIST Editor-in-chief: Anita Coleman, University of Arizona

dLIST Classics

Titles by S.R.Ranganathan

Five Laws of Library Science, Ed. 1 (1931)

Philosophy of Library Classification (1973)

Prologemena to Library Classification, Ed. 3 (1967)

Classification and Communication (1951)

Documentation Genesis and Development (1973)

Documentation and its Facets (1963)

Library Book Selection, Ed. 2 (1966)

New education and school library: Experience of half a century (1973)

Reference Service, Ed. 2 (1961)

Other titles

S.R. Ranganathan's Postulates and Normative Principles: Applications in Specialized Databases
Design, Indexing, and Retrieval, 1997. Compiled by A. Neelameghan.

Memorabilia Ranganathan: A compilation of useful quotations of S.R. Ranganathan from his various
works, 1994.

Putting Knowledge to Work: An American View of the Five Laws of Library Science, 1970. By
Pauline Atherton.

Read the dLIST Classics online!

<http://dlist.sir.arizona.edu/>

