14TH CONGRESS OF THE INTERNATIONAL SOCIETY FOR PHOTOGRAMMETRY
HAMBURG 1980

COMMISSION VI

Working Group 4

INVITED PAPER

AUTOMATED STORAGE AND RETRIEVAL SYSTEMS FOR
PHOTOGRAMMETRY AND REMOTE SENSING

by

J.H. ten Haken

Librarian

INTERNATIONAL INSTITUTE FOR AERIAL SURVEY AND EARTH SCIENCES(ITC)

ENSCHEDE, THE NETHERLANDS

Abstract

Because of the enormous flow of publications the conventional bibliogra-
phic storage and retrieval systems have become inadequate to meet the de-
mands.

Automation has become inevitable. Data bases containing bibliographic
references have been created for nearly every branch of science and tech-
nology. These data bases can be consulted on-line by means of a terminal,
thus offering the possibility of retrospective searching in the cumulative
world literature covering several years.

Also literature in the field of photogrammetry and remote sensing has
been included in appropriate data bases. In view of the possible applica-
tion of photogrammetry and above all of remote sensing to fields as earth
sciences, vegetation and agriculture, also the data bases concerning these
fields have to be taken into account. Moreover there are multidisciplinary
data bases which could provide worthful information.

Besides data bases,which only contain bibliographic references, there
are data banks which contain fact information. As far as remote sensing is
concerned, the LEDA-data banks providing information on satellite imagery
are of importance.

1. Introduction

The world stock of scientific and technical publications has reached enormous proportions in the course of the years and is still growing rapidly. The growth is in fact showing an exponential curve; every 15 years the total quantity of published scientific and technical literature is doubled.

An example of the overwhelming flow: in the field of aerospace alone it is estimated that at least one million reports, journal articles, conference papers, etc. are published annually.

At the same time there is an increasing demand for information on behalf of study purposes, research, policy-making, etc.

In order to be able to search in the cumulative world stock of literature and to meet the users' demands, it has become inevitable to involve computers to solve the problems of storage and retrieval.

According to their contents and purpose we distinguish two kinds of storage and retrieval systems:

a. Bibliographic retrieval systems
   Feature: the output is presented as bibliographic references

b. Fact retrieval systems
   Feature: the output is presented as ultimate information or facts

It has to be observed that as for machine-readable bibliographic retrieval systems the term data bases is in use and as for fact retrieval systems the term data banks. Also the term file is used, which in a given case means a part of a data base or data bank containing allied information. With respect to automated storage and retrieval systems there is generally no distinction between files and data bases, resp. data banks.

2. General observations about automated retrieval systems

When a bibliographic reference is stored in a retrieval system, it may be retrieved by means of certain identifications, e.g. author's name, accession number, subject.

Author's name and accession number are easily to be handled, for the author's name is an objective identification and already from the origin connected to a publication and also an accession number, after its addition, is an objective identification and moreover connected to one publication only.

However, most literature is searched after subject. Or, as one of the great information scientists once expressed it: subject retrieval is the artery of our profession. For that purpose the references to be stored are provided with so-called descriptors, i.e. keywords or classification numbers which identify its subject(s). For several reasons there is generally made use of keywords.

2.1 Principles of automated subject retrieval

The principles of automated subject retrieval are based on the rules of mathematical logics developed by the mathematician George Boole. Basically these rules consist of three operators: AND, OR, NOT.

We shall explain this further by means of the diagrams introduced by the mathematician John Venn, the so-called Venn diagrams(figure 1).

In the Venn diagrams the area covered by circle A represents the computer's complete stock of references with one feature. Circle B its stock of references with another feature.
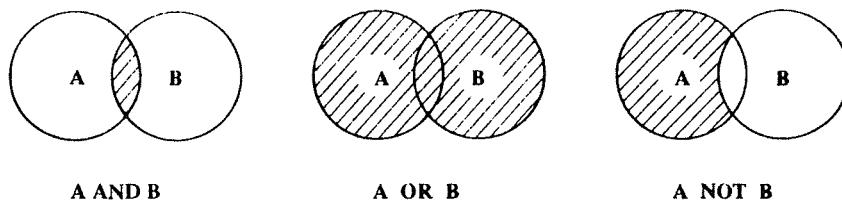
| A AND B | A OR B | A NOT B |

Figure 1. Venn diagrams

For example, we suppose that the feature of circle A is determined by the keyword REMOTE SENSING and that of circle B by FORESTRY.

If literature is wanted on remote sensing we insert the keyword REMOTE SENSING into the computer. The computer will search the data base on this keyword and compose a list of references. This list is corresponding to the area covered by circle A. The same process will take place if literature is desired on forestry. We shall then receive a number of references corresponding to the area covered by circle B.

Both the circles have a common area – the shaded part – representing all the references which use the keywords A AND B, that is, REMOTE SENSING AND FORESTRY. As may be expected, this is a relatively small area compared to that of A and B separately. So, if literature is wanted on remote sensing applied to forestry, we put in the conforming keywords together with the Boolean operator AND. The result will be a list of references corresponding to the common area in the figure.

When using the operator OR you indicate that you wish to find literature on either of the features A OR B; and this will cover a relatively large area.

Through the operator NOT one informs the computer that the desired literature should be on one of the two features only: so if we choose A, it should NOT include B.

2.2 Batch-processing and on-line retrieval

When retrieving literature references by computer this may be done in two ways, that is either by "batch-processing" or "on-line".

When applying batch-processing, the orders for bibliographic retrieval are handed in at the computer department and processed in rotation. The result will be a printed list of appropriate literature references according to the chosen keywords and the possible complementary Boolean operators AND, OR, NOT.

With on-line retrieval you are in contact with the computer during the whole searching process via a terminal. The terminal is composed of a keyboard by means of which the instructions may be given to the computer, a screen showing the preliminary results and a printer. During the whole process of searching there is a possibility to evaluate and to intervene, e.g. to specify a subject further through the operator AND if there have been put out too many references or to insert other related keywords in case there have been put out too few. The final result will be printed out.

It is clear that on-line retrieval offers greater possibilities than batch-processing. However, it should not be forgotten that with on-line retrieval you are in contact with the computer during the whole searching process including the time spent judging the results and thinking out new combinations, thus consuming considerably more expensive computer time.

There is a clear trend in favour of the on-line systems.

3. Review of on-line systems

Since the beginning of the seventies gradually various automated biblio-graphic systems have been developed. Many bibliographic and abstracting services are producing now data bases, often as an extra product of their printed editions of bibliographies and abstract journals, having the result that in the meantime nearly all branches of science and technology are co-vered by data bases. Also a number of national bibliographies are being produced in this way. Further there are large libraries which produce data bases of their collections and combinations of libraries which do so with their cumulative collections.

3.1 The way of access to on-line systems

The exploitation of most of the data bases is channelized through compa-nies which act as hosts or intermediaries between the producers and users of data bases. I mention here as being the most well-known and having the greatest assortment available to the user:

The European Space Agency's   Computer language: QUEST
Information Retrieval Service(ESA/IRS)
ESRIN, Via Galileo Galilei
00044 Frascati(Rome)
Italy

Lockheed Information Systems(LIS)   Computer language: DIALOG
3251 Hanoverstreet
Palo Alto, California 94304
U.S.A.

System Development Corporation(SDC)   Computer language: ORBIT
2500 Colorado Avenue
Santa Monica, California 90406
U.S.A.

It is realized that this enumeration is far from complete the more so as new hosts are still announcing themselves regularly.

Nevertheless as far as photogrammetrists and remote sensing specialists are concerned, all the relevant data bases can be reached through the above hosts.

3.2 Data bases in the field of photogrammetry and remote sensing

We shall consider now which data bases are of importance for the field of photogrammetry and remote sensing. In this connection it should be ob-served that also the conventional photo-interpretation falls under the scope of remote sensing nowadays.

Besides the data bases of direct importance to the field of photogramme-try and remote sensing we shall also have to take into consideration that there is a close link between on the one hand photogrammetry and remote sensing and on the other hand sciences to which they may be applied, such as earth sciences, vegetation and agriculture. So also the data bases covering these fields should be taken into account. Furthermore there are a number of multidisciplinary data bases which also may contain worthful literature for our purpose.

The international most well-known available data bases in the respective fields are listed in the appendix as the parts A, B and C.

A special case is the data base RESORS(Remote sensing on-line retrieval system) originated and exploited by the Technical Information Service of the Canada Centre for Remote Sensing, 2464 Sheffield Road, Ottawa, Ontario K1A OE4. This data base contains remote sensing literature in the broadest sense. However, it is not yet generally available and only operational for "in-house" use. Requests for a search should therefore be applied to the afore-mentioned address.

Furthermore there are good prospects that Geo Abstracts with Part G of which(Remote Sensing, Photogrammetry and Cartography) the former ITC International Bibliography of Photogrammetry has merged also will become available on-line in future.

3.3 Data banks in the field of remote sensing

Contrary to data bases which put out bibliographic references, data banks provide ultimate information on a subject. The data banks of importance for the field of remote sensing are the LEDA(on-Line Earthnet Data Availability) -data banks, which supply direct information on produced imagery.

This information comprises: cycle, track and frame, longitude and latitude, date and hour, name of the satellite, cloud coverage and image quality. For indicating the cloud coverage and the image quality certain parameters have been introduced.

Information can be had by inserting the known data into the computer, often being longitude and latitude. This may further be specified by indicating e.g. the minimum cloud coverage desired and the name of the satellite, using the operators AND, OR, NOT. Finally the full data will be printed out, whereas at the same time there is a possibility to order on-line the desired imagery.

4. Example of an on-line retrospective bibliographic search

We shall now consider how a not too complicated bibliographic search is effectuated.

In the following example, derived from practice, literature was searched about soil mapping by means of aerial or satellite imagery. The most suitable data base for this purpose was supposed to be the STAR + IAA file exploited by the ESA/IRS. With the help of the NASA-thesaurus keywords were chosen covering the subject in question and the following scheme was set up:

Photomapping   )
Reconnaissance   )
Aerial reconnaissance)   to be combined with (Soil mapping
Photo reconnaissance)   (Terrain analysis
LANDSAT 1   )
LANDSAT 2   )

After having been connected to the ESA-computer by the public telephone line, further instructions how to handle are communicated on the screen, among other things which files are accessible. Thereafter you are allowed to insert your first keyword, in this case PHOTOMAPPING, which simultaneously appears on the screen indicating at the same time how many items are corresponding to this subject. Through the command SELECT these items are separated and kept as a set during the search process; in this case SET 1. (See figure 2. Search scheme)

Consecutively the keywords RECONNAISSANCE, AERIAL RECONNAISSANCE, PHOTO RECONNAISSANCE, LANDSAT 1 and LANDSAT 2 are inserted, forming resp. the sets 2 - 6.

```
..............................................
:                                            :
:        E.S.A. DOCUMENTATION SERVICE        :
:                                            :
..............................................
```

USER 265   DATE: 02/28/78   TIME: 16:51:00

SEARCH HISTORY

SET	ITEMS	DESCRIPTION
1	1200	PHOTOMAPPING
2	320	RECONNAISSANCE
3	785	AERIAL RECONNAISSANCE
4	119	PHOTORECONNAISSANCE
5	482	LANDSAT 1
6	42	LANDSAT 2
7	2668	1+2+3+4+5+6
8	421	SOIL MAPPING
9	1411	TERRAIN ANALYSIS
10	1770	8+9
11	267	7x10

SRCH TIME   3.90   PRINT COUNT   267   DESCS.:   8

Figure 2. Search scheme

As we claim that the literature in question should correspond to one of the above subjects, we join the sets 1 - 6 through the command OR, thus resulting in a new set, set no. 7, containing 2668 items.

The next step is to insert the keywords SOIL MAPPING and TERRAIN ANALYSIS, resulting in the sets 8 and 9. Joining these sets through the command OR gives set 10.

The final phase in the search process is that the literature should correspond to both the keywords of set 7 and of set 10. Therefore the sets 7 and 10 are combined through the command AND forming thus set 11 with 267 references. Through the command PRINT the 267 references in question will be printed on a list, of which the first one is shown as an example in figure 3.

PRINT 11/2/1-267   USER 265   PAGE   1   (ITEM   1 OF 267)

78N13516# ISSUE 4 CATEGORY 43 AD-A045871, WES-TR-M-77-2-2 770900
   Acquisition of terrain information using Landsat multispectral data. Report 2, an interactive proceaure for classifying terrain types by spectral characteristics
   STRUVE, H.; GRABAU, W. E.; WEST, H.W.
   Army Engineer Waterways Experiment Station, Vicksburg, Miss. (A0483249)
   142 p. Jpn. 494 HC A07/MF A01
   FLOOD PLAINS / xLANDSAT SATELLITES / OPTICAL DATA PROCESSING / xPHOTOMAPPING / REMOTE SENSORS / SATELLITE-BORNE PHOTOGRAPHY / xSPECTRAL RECONNAISSANCE / xTERRAIN ANALYSIS

Figure 3. Outprint of a reference

LIST OF DATA BASES AND DATA BANKS

Used abbreviations in the list of data bases and data banks

ESA	European Space Agency
ESA/IRS	European Space Agency's Information Retrieval Service
LIS	Lockheed Information Systems
SDC	System Development Corporation
STAR	Scientific and Technical Aerospace Reports
IAA	International Aerospace Abstracts
NASA	National Aeronautics and Space Administration
AIAA	American Institute of Aeronautics and Astronautics
CAB	Commonwealth Agricultural Bureaux
NAL	National Agricultural Library
NTIS	National Technical Information Service

A. Data bases of direct importance to the field of photogrammetry and remote sensing

Data base	Supplier	Field covered	Time span covered	Number of refs. on file	Number of refs. added yearly	Available through
STAR	NASA	The world's unpublished (report) literature in the aerospace field in the broadest sense, including photogrammetry and remote sensing	1962 – present	1,100,000	66,000	ESA/IRS
IAA	AIAA	The world's published (periodicals, conferences) literature in the same field				

Although mentioned as two data bases, they constitute in fact one single data base, indicated as file no. 1 of the ESA/IRS

105.

B. Data bases in the fields earth sciences, vegetation, agriculture

Data base	Supplier	Field covered	Time span covered	Number of refs. on file	Number of refs. added yearly	Available through
OCEANIC	Data Courier	World-wide literature in the field of the oceans in the broadest sense	1964 – present	115,000	6,000	ESA/IRS LIS
CAB ABSTRACTS	CAB	World-wide literature in the field of agriculture in the broadest sense	1973 – present	800,000	150,000	ESA/IRS LIS
AGRICOLA	NAL	The actual holdings of the National Agricultural Library, USA	1970 – present	1,150,000	100,000	LIS
GEOREF	American Geological Institute	World-wide literature in the field of earth sciences in the broadest sense	1967 – present	450,000	48,000	SDC
GEOARCHIVE	Geosystems	World-wide literature in the field of earth sciences in the broadest sense	1969 – present	600,000	100,000	LIS

C. Multidisciplinary data bases

Data base	Supplier	Field covered	Time span covered	Number of refs. on file	Number of refs. added yearly	Available through
NTIS	NTIS	Information released to the public by US Government Agencies and Departments and includes business and economic data as well as scientific and technical report literature; US Government-sponsored translations and some reports written in foreign languages are also included	1964 – present	530,000	50,000	ESA/IRS LIS
PASCAL	Informa-science	World-wide literature of all types normally in its original language with a French translated title and abstract	1973 – present	3,300,000	500,000	ESA/IRS

D. Data banks containing information on imagery

Data bank	Supplier	Field covered	Time span covered	Searchable fields	Available through
LEDA RETROSPECT	ESA	The geographical ground coverage of the LANDSAT satellites roughly encompasses an area bounded by the Faeroe Isles, Namsos(Norway), Pitea(Sweden) and Arkhangel'sk(USSR) in the North; Aaiun (Spanish Sahara), Tamanrasset (Algeria), Araka(Niger) and Dongola(Sudan) in the South; the Canary Isles, Lisbon(Portugal), Gt. Blasket Isle(Irish Rep), and the Faeroe Isles in the West; and Syktyvkar(USSR), Penza(USSR), Tbilisi(USSR), Dihok(Iraq), and Ha'il(Saudi Arabia) in the East. Each scene is covered once every 18 days.	1975 – previous year	Mission Longitude Latitude Cycle	ESA/IRS
LEDA CURRENT	ESA		Current year	Track Frame Date acquisition Time acquisition Cloud coverage Data quality	

The data is currently obtained from LANDSAT 1 and LANDSAT 2 via the Fucino(Italy) ground station. Imagery from other satellites, e.g., HCMM, NIMBUS-G, SEASAT-A and acquired by additional ground stations, e.g., Kiruna(Sweden) will be added as it becomes available.

In the meantime the ESA released the following message:

For operational reasons it will be necessary later this year to merge together the LEDA databanks, presently available as files 12 and 13. File 12, named LEDA RETROSPECT contains all data that are not related to the current year. When the merge occurs, the unique LEDA file will be available as file 13. The merge has no effect whatsoever on the searching. Actual changeover will be announced by a "flash message" at logon time.