

SURVEYING AND MAPPING ABSTRACTS DATABASE AND ITS FUNCTIONS

Wei Qunmin, Xiong Jianqiang, Wang Zhiliang, Li Yong

Wuhan Technical University of Surveying and Mapping
39 Loyu Road Wuhan China

ABSTRACT

At present, there are many large on-line information retrieval systems in the world. These systems include hundreds of specialized database, but there is not any specialized database in the surveying and mapping field among them. In order to ensure that the professionals of surveying and mapping can easily search for literature on surveying and mapping, we created a computerized Surveying and Mapping Abstracts Database (SMADB), based on SIEMENS 7570-C host computer and GOLEM software of information retrieval system. The creation of SMADB and its maintenance are described. The retrieval functions of the database and how to generate the periodical publication of surveying and mapping abstracts and output to a floppy on IBM-PC are also described in this paper.

KEY WORDS: Information Retrieval System, Surveying and Mapping Abstracts Database, On-line information retrieval

INTRODUCTION

There are many on-line information retrieval systems in the world, including thousands of machine-read database. Literature databases cover almost all disciplines and profession: philosophy, politics, economics, laws, business management, mathematics, physics, chemistry, astronomy, geography, medicine, engineering and so on. These vast literature on the above disciplines are data source of retrieval systems. Many international on-line information retrieval systems are formed by means of advanced computer network and satellite communication technology, such as DIALOG, ORBIT in US, ESA/IRS in Europe etc.. Their computer terminals are all over the world. Users can easily search for some scientific and technical literature in which they are interested, only using their specialized data base concerned. However, there is not any specialized database in the surveying and mapping field among those retrieval systems. A vast amount of literature dealing with surveying and mapping disciplines are scattered over many different specialized database. If professionals of surveying and mapping want to find out the literature they are interested in, from international on-line information retrieval system, they must visit many different specialized database. That not only increase the search time and expenditure, but, more importantly, make it very difficult to search for the required literature pertinently and entirely. Therefore, International Society of Photogrammetry and Remote Sensing (ISPRS) 4/VI discussed this problem in 1988, Tokyo. And proposed to establish a computerized information retrieval system of surveying and mapping by international cooperation. Hence, we developed some new subroutines, based on SIEMENS 7570-C host computer and GOLEM software, and created SMADB. Various kinds of scientific and technical literature on each discipline of surveying and mapping and on related discipline are stored in the database. Not only can it provide the professionals with conveniently on-line information retrieval service of literature on surveying and mapping disciplines. but can automatically generate a periodical publication of SMA.

1 FEATURES OF SMADB

SMADB is a friendly integrative system in which the data is accessed in batching and information can be searched efficiently. Managed by the operating system BS2000 of computer, and making use of the GOLEM software and new developed subroutines, the database can be created, expanded and modified. Its features are as follows.

(1) The database covered abstracts and additive information from the main periodicals and international proceedings on surveying and mapping. By means of conversation, it not only can allow several users on the local or remote terminals to access the database at the same time, but can also be modified or expanded by the administrator in the meantime.

(2) It can automatically record the process of search, report how many people have used the data base and how much time they have spent. It is easy to display or print the retrieval results.

(3) It can automatically generate a regular text periodical publication of SMA. Its greatly convenient for manual retrieval.

(4) The data in the database are stored in free flexible format. It is helpful to data transformation and share. There are SMADB thesaurus and the list of key works (free words) in the database, so that users can scan them and administrator manage and modify the thesaurus.

(5) The subroutines for creating, modifying searching and expanding database, and for generating automatically a periodical publication of SMA are stand alone each other. So they can be done easily and flexibly. The database has multilevel password to avoid unauthorized users using the database.

(6) There are two kinds of retrieval commands. One is formed in German and another is CCS (Common Command Set) in English. All commands are complete and Simple for users to grasp.

2 HARDWARE AND SOFTWARE OF SMADB

2.1 Hardware

Host computer	SIEMENS 7570-C CPU speed 3 millions/sec Main memory 16MB cache 32KB
Disks	3024 MB
Tape drivers	2 units
Terminals	50 units
Line printer	2 units
Laser printer	1 units
Paper printer	2 units

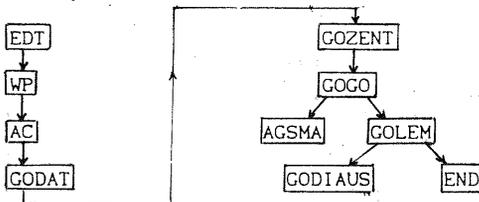
2.2 Software

BS2000 Operating system

EDT	Editing program
WP	Word processing program
GODAT	Program which checks grammar of input documents
GOZENT	Program which creates and manages database
GOGO	Program which loads modifies and expands document
GOLEM	Retrieval program
GODIAUS	Accounting program
AC	Automatically document classify program
AGSMA	Program which automatically generates retrieval reference book(SMA)

3 THE CREATION OF SMADB AND ITS MAINTENANCE

SMADB, in which the data are accessed, expanded and modified in batch processing model, proved stored information on surveying and mapping by means of interactively mode. The flowchart is as below



3.1 The creation of SMADB

3.1.1 creating a file of input documents by EDT
Each document may be divided up into two sections in the database. The format, content and length of the document are decided according to requirements of search users.

The format of a record in the database is as below

```

$
CC:---; (classification Code)
TI:---; (title)
AU:---; (author)
PU:---; (publication)
PD:---,---,---; (publication date, Vol. ... page)
SP:---/---/---; (subject phrases)
---/---/ (free words or key words)
*
AB: --- (abstract)
---
```

A lot of documents may be put in one file

3.1.2 Automatically check the stored data by WP

3.1.3 Creating a command file by EDT as follows

```

$/FILE INPUT=the file name of input document
$/FILE NORM=the infix out put files
$/GODAT
$/NORMAL=ICO
$/END
```

3.1.4 Classifying the stored data according to surveying and mapping disciplines by AC

3.1.5 Reading in and checking the file of input document by following

```

$/SYSFILE SYSDAT=the command file name
/EXEC GODAT
```

At this time, the command file is automatically executed. The documents are read in and checked by the operating system. When the documents have been checked, all of the error messages are put into the output file that called GOLEME. infix. DATFE01, and all of correction documents in the command file are put into another output file that called GOLEM. infix. DATUS01. The documents can only be put into the database after all of error have been corrected.

3.1.6 Creating the database as following

```

/SYSFILE SYSDTA=(PRIMARY)
/EXEC GOZENT
*NEW P=database name, G=password
*INIT ASP=SW
*DASI P=database name, G=password, S=YES
*END
/EXEC GOGO
*ZEIN P=database name, G=password,
I=GOLEM. infix. DATYS01, E=HEUTE
```

If there is no error, the process of creating data base will be finished normally.

3.2 Establishing the SMADB Thesaurus in the database

The SMADB Thesaurus is the main tool for producing indices, accessing and searching for information in the SMADB. The Thesaurus is linked with documents in the database to show the frequency of use each subject phrase so that the Thesaurus will be modified in the future.

3.2.1 Creating the file of relationship between the subject phrase by EDT The format of the file is as below:

```

*SP: Physical geodesy
UF dynamic geodesy
UF geodetic gravimetry
BT Geodesy
RT Geodynamics
... ..
```

3.2.2 Defining the names of the relationship between the subject phrase and their abbreviation

```

/EXEC GOZENT
*UPDATE P=database name, G=password
*BEZ USED FOR, UF
*BEZ NARROW TERM, NT
*BEZ BROAD TERM, BT
*BEZ RELATED TERM, RT
*BEZ USE, us
*END
```

3.2.3 Creating the SMADB Thesaurus in the database

```

/EXEC GOGO
*THAEND P=database name, G=password,
I=the file name of relationship
```

Now, the file of relationships are read in and checked. The Thesaurus can only be put into the data base after all errors have been corrected.

3.3 The extension and modification of the database
After the database is established initially, the daily work would be the extension or modification needed.

3.3.1 Extending the database

```

. see 3.1.1 - 3.1.3
. extending documents into the database as below:
/ EXEC GOGO
*ZEIN P=database name, G=password,
I=GOLEM.infix.DATUS01, E=HEUTE, F=J
```

3.3.2 Deleting or modifying documents

First, create a command file of deleting or modifying by EDT The format of the command file is as below:

```

** 11 - 50 (delete the first to fiftieth)
L (documents in database)
```

Then, use following command to execute.

```

/EXEC GOGO
*ZIAEND P=database name, G=password,
I=the command file name, U=N, H=M
```

3.3.3 Modifying the Thesaurus in database

```

/EXEC GOGO
```

*THAEND P=database name, G=password
 *SP: SPOT image
 *SP: strip adjustment
 *end

The result of the process is the subject phrase 'SPOT image' is added to the Thesaurus and the subject phrase 'strip adjustment' is deleted from the thesaurus in the database.

4 THE FUNCTIONS OF SMADB

4.1 Search terms

The aim of a computerized SMADB is to provide information quickly and accurately for users. According to both factors of the users requirements and the condition of the computer, the search terms or channels of the SMADB are as following:

Subject phrase(SP): A term in the SMADB Thesaurus;
 Free word: A key word outside the SMADB Thesaurus;
 Classification code(CC): A classification code at branches;
 Author's name(AU);
 Publication name(PU);
 Publication data(PD);
 Title(TI);
 Information number.

4.2 Search methods

4.2.1 Subject phrases search Nowadays, the use of subject phrases to search for information is the main method in a computerized information system. An SMADB Thesaurus is compiled according to the pattern of 'INSPEC Thesaurus' and 'List of descriptor for surveying and mapping in Chinese' for the SMADB. It is the main tool for accessing and searching for information, producing a subject index within the SMA. The process of this method is that the subject phrase may be chosen by the user from the Thesaurus in accordance with the content of his researching problem. Then the SMADB is scanned by the subject phrase chosen to search out the particular papers in which the user is interested.

4.2.2 Non-subject phrase search

The method is that the database is scanned through the author's name, publication name, a publication date or a free word which indicates the subject of the researching problem of the user, to enable the user to find any related papers.

4.2.3 Combination search by a logical operator

The searching requirements of a user are often very complicated. A single search term could not indicate the requirements. For this reason, several search terms which are related to the researching problem of the user must be combined to form an expression of search logic. The database is scanned by the expression, which enable the user to acquire more appropriate papers.

The use of the logical operator could combine many search terms in order to constitute an operating formula of search logic. The SMADB provides three kinds of logical operators for users. They are logical 'AND', logical 'OR', and logical negation 'AND NOT'.

Logical AND means that every paper searched for must contain both contents of two terms linked with AND;

Logical OR means that the paper searched for may contain content of one of the terms linked with OR;

Logical negation AND NOT means that the paper searched for must be the content of the first term

linked with AND NOT but not second term.

The subject phrase, free word, author's name, publication name, publication date etc. can be combined by the logical operators and multiply parentheses to make up a complicated formula of search query. This formula can accurately indicate searching requirements of the user.

4.2.4 Masking search The SMADB allows that the characters '&&' are prefixed, appended or both the word fragment by the user, looking like a search word. For example, the 'photo&&' is used as a search word to scan the database for some papers presenting photography, photomosaic, photomap, photogrammetry or photoplan etc.; the '&&sat' for several papers about seasat or landsat etc.; the '&&graph&&' for many papers describing cartography, geography, photography etc..

This search method could expand the user search range and save time on the computer. It not only may be used by the user in the above three cases but it could also be used in the method described in the following passage.

4.2.5 Searching in abstract This is a further search after the papers have been found by the above methods in order to acquire more accurate information, that is to say, the abstract of those papers is scanned by any word or sentence after the user has searched out some papers using the above methods. This search could delete several papers not fully related to the researching problem of user, so that the user finally acquires the most appropriate or important papers.

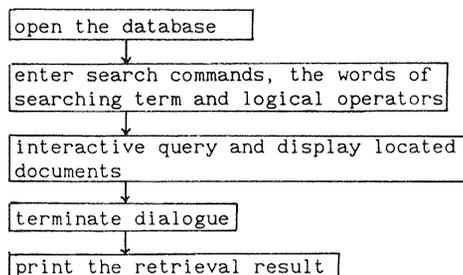
4.2.6 Output of the retrieval results

The retrieval results can be provided in two kinds of form. One is to display the results on screen, e.g. the number of retrieved papers, and their sequential number, and all fields of a record are displayed. If the results are not desirable, users can go back to retrieval mode, and select new related subject phrases on the displayed Thesaurus to extend the domain of searching. Up till users are satisfied with results. Another is to print the results, e.g. first the user write the results on the disk files, and then print out the results with the system commands.

4.2.7 Storing searching strategy The SMADB can store all searching commands of the user in the initialized file to provide selective dissemination of information on-line for the user.

4.3 Searching process

The SMADB provides two languages of searching for the user. One is Germany GOLEM command, another is English CCS commands. They are all easy to operate by the user, and both processes are the same.



5 AUTOMATICALLY GENERATING SURVEYING AND MAPPING ABSTRACTS BY AGSMA

Surveying and Mapping Abstracts (SMA), which be produced automatically by AGSMA subroutine, is a text abstracts compilation. It can provide for readers to search for various kinds of technical papers in the international surveying and mapping field. As a retrieval reference book and an advanced information products, SMA can provide user with excellent service of searching for scientific and technical literature on surveying and mapping disciplines. It would be very useful for many professionals in the course of accomplishing to improve surveying and mapping productivity and industry innovation in 90's.

Each SMA is composed of two parts:

(1) Abstracts section -- Abstracts are arranged according to surveying and mapping disciplines. It contain information number, title, author, source of literature and abstract.

(2) Indices section -- It includes subject index, key word index and author index, which all be related to abstracts section.

Subject index and key word index are arranged in alphabetic order of subject phrases and key word. It is the best approach for user to search for. Users can search for the interested literature on the basis of selecting correct subject phrases or key words, which are related to users' problems. Author index is arranged in alphabetic order of author's name. It contain all authors of each papers which are stored in the database. The information can also be outputed to a floppy disk on IBM-PC.

6 SMADB THESAURUS

The SMADB Thesaurus is compiled according to a pattern of 'INSPEC Thesaurus' and 'List of descriptor for surveying and mapping in Chinese'. It is the main tool for indexing, accessing of literature and searching in the SMADB The Thesaurus now contains approximately 3000, of which some 2500 are preferred terms (i.e. subject phrase) and 500 are cross reference terms (i.e. non-subject phrase). The subject phrase are used in indexing all documents added to the SMADB. The changes and additions to the Thesaurus will arise from the requirements of literature in the branches such as Photogrammetry, Engineering surveying, Cartography and so on covered by SMADB.

Arrangement of the Thesaurus has two parts:

(1) Alphabetical display of terms
This is the main part of the Thesaurus, listing in a single alphabetical order all terms both 'preferred' and 'cross reference'. A capital letter is used to indicate 'subject phrase', and 'non-subject phrase' are shown in small letter e.g.

Basic topographic map.....subject phrase

Close range photography.....subject phrase

UF Nontopographic photography

BT Photography

RT Close range photogrammetry

nontopographic photography...non-subject phrase

US Close range photography

Photogrammetry

NT Aerial photogrammetry

Analytical photogrammetry
Close range photogrammetry
Digital photogrammetry
Dynamic photogrammetry
Ground photogrammetry
Holographic photogrammetry
Space photogrammetry
Underwater photogrammetry

Under each subject phrase are term's relationships. These show other terms which are related to the given term in the Thesaurus.

where US (use) and UF (use for) are replacement relationships.

NT is narrower i.e. more specific term.

BT is broader i.e. more general term.

RT is related term.

Where alternative terms, synonyms or alternative spellings exist for a given term, only one form will be chosen and reference made from the alternative. This is done by the US (use) reference. For each case where a US (use) reference is made, the preferred terms will show any non-preferred terms to which reference is made. This is done by means of the UF (used for).

(2) Category index: All subject phrases in the SMADB thesaurus are divided into seven classifications according to surveying and mapping discipline category, that is, 1. General; 2. Elementary surveying, Topographic surveying; 3. Geodetic survey; 4. Photogrammetry and Remote Seining; 5. Engineering Survey; 6. Cadastral survey and, 7. Cartography. Subject phrases in each classification are arranged in principle according to surveying and mapping disciplines category and two levels extension of the subject phrases subordinate relationship. For example, the subject phrases in photogrammetry and remote sensing classification as below.

Aerotriangulation

NT Analytical aerotriangulation

On line aerotriangulation

Analytical aerotriangulation

NT Strip aerotriangulation

Block aerotriangulation

A subject phrase can appear in more than one branch, but the subject phrase is only located in one main branch. (see-xxxx) is used to indicate the subject phrase of one branch to the main branch, such as subject phrases in 3. Geodetic survey classification.

Control survey

NT Horizontal control survey

Topographic control survey (see-02)

Vertical control survey

Photo control survey (see-04)

Engineering control survey (see-05)

CONCLUSION

From the point of view of the existing information retrieval systems in the world, Surveying and Mapping Abstracts Database (SMADB) is a unique computerized retrieval system of the literature on all disciplines of surveying and mapping at present. The system have been improved continually through past several years of work, furthermore, it is a practical system basically. However, there are some shortcoming in the system as a result of our conditions (financial resource etc.). We hope that international partner would be found to improve the system farther, to enlarge the scopes of surveying and

mapping information service and to share the literature resource of surveying and mapping disciplines.

REFERENCES

- Burette, D., Lata arche, CH.-H., Pouyllau, D., Ten Haken, UJ.H., 1988, Prospective study for an ISPRS Database. In: Int. Arch. Photogramm. Remote sensing., Kyoto-Japan. Vol.27 part.B6, pp. 29-38
- Clark, J.E., 1987, Accessing the resources of the National Technical Information Service. Photogrammetric Engineering & Remote Sensing., LIII/4:395-397
- Hothmer, J., 1988, Information Retrieval for Literature and Factual Data in Photogrammetry and Remote Sensing., Kyoto-Japan, Vol.27 part.B6, pp 85-94
- Hothmer, J., 1984, The information retrieval system ISPRS-IRS for literature and factual data, In: Int. Arch. Photogramm. Remote Sensing., Rio de Janeiro-Brazil, Vol.25-A6, pp. 155-169
- Huang, G. J., 1990, Computer-Aided Quality Control for Bibliographic Database; New Technology of Library and Information Service (China), No. 1:5-8
- Piternick, A. B., 1989, Functions and Capabilities of online Searching Systems., Online Review, 13(6): 460-469
- Wen, S. H., 1991, On the Evaluation and control of indexing Quality in Bibliographic Database., New Technology of Library and Information Service (China), No. 2:36-41
- 1983
INSPEC Thesaurus. IEE