

컴퓨터 프로그램의 분류방안과 목록작성 방법의 개발에 관한 연구

Development of computer software classification scheme and cataloging methodology

문기주* · 오현승**

ABSTRACT

The size of software collection at any organization is increasing rapidly with the use of computer becomes popular for various applications. This trend points out the importance of proper 'catalog' or 'directory' for software collections, however, only 'title list' is found in practice. The problem of no proper directory with software collection would be more serious than the problem with book collection since browsing is not possible with softwares. Many of the expensive softwares will be a waste of money to purchase and to hold if no catalog about them is provided.

In this paper the difficulties in applying AACR2 to software cataloging and the most important characteristic of computer software are discussed as well as presenting the necessity of localized information due to the requirement of a machine to be used, the difficulty using subject-based-call-number scheme, an expected way of searching software to design a location indicator, software qualifications to be cataloged, and so on. In addition, an analysis of randomly selected softwares after a cataloging practice is followed.

1. Introduction

Software collections for the use of staff and students have become a feature of many college and university libraries over past few years[9]. It is not surprising at all that the size of collection is increasing rapidly. The variety of ways which softwares can help staff and students from research to lecture show the evidence of this trend.

* Department of Industrial Engineering, Dong-A University

** Department of Industrial Engineering, Han Nam University

The importance of 'directory' or 'catalog' for computer softwares must be pointed out if the size of collection become large. It may not be necessary to have directory or catalog with reasonable number of softwares, but not with a large size of collections. Many of the expensive softwares will be a waste of money to purchase and to hold if no catalog about them is provided.

Any software collection without proper catalog will be exactly like a book collection without catalog. In addition, the problem with software collection will be more serious than the problem with book collection. While a patron can probably browse books to find what he/she wants, but not with softwares. Software titles could be browsed if they are stored on floppy diskettes, but it is not true if they are on the mini/mainframe or available only through remote access.

Cataloging computer software also needs rules as well as cataloging printed material does. A very comprehensive set of rule called the Anglo-American cataloging rules, 2nd edition (AACR2) can be applied to software catalog. Dodd[4] described all the aspects of cataloging machine readable data files(MRDF) with examples in her book published in 1982. Two years later, the British Library published a research report about their experimental software cataloging works[11]. In 1985, Dodd and Sandberg-Fox[5] released a book for cataloging microcomputer files. Also the draft revision of AACR2 Chapter 9 of 1987, describing itself as 'only a preliminary draft of the rules for computer files' is found as an example of AACR2 application[8]. These serial works tell us two things. One is the importance of having computer software catalog and the necessity of cataloging rules. The other is the necessity of continues modifications on the rule for applying AACR2 to computer softwares.

Development of software catalog without a set of stable rules could be risky and painful tasks. However, the project to develop an online catalog for computer software began at Iowa State University(ISU). The initial idea was to create an online catalog of the software available to engineering students at ISU since there was a lot of software around, but nobody knew precisely where it all was or what it did. Through cataloging practice, we found several interesting points for discussion. They are some important characteristics of software affecting to cataloging practice, location indicator problem, software qualifications to be cataloged, and some statistics obtained by analyzing catalog entries what we made.

2. Software cataloging

How different software is from printed material in cataloging point of view? In this section, some interesting characteristics of software will be discussed in comparison of printed material.

2-1. The necessities of local information

A catalog is intended to help someone to find what is available in a particular library. Therefore, all the necessary information to locate a particular item must be included on catalog. Generally, catalog includes title, author, version, publisher, physical description/requirements, series, note, standard number, classification, etc[5,8,11]. However, a catalog for computer software may have to be more localized one than printed materials. One of the reasons could be on the characteristic of software which needs to have a machine to be used. No software can be used alone and not all local libraries or computer laboratories would have the same computer system. Some of the important catalog information supplied by Library of Congress(LC) or Online Computer Library Center (OCLC) must be changed based upon the locally available machines. For example, the catalog entry supplied by OCLC need not to be modified if it is for printed materials, and location information may have to be changed if the software had been modified for local use.

2-2. Documentation, storage medium and software

Computer software is not a single item like a book, but a compound matter which includes storage medium and documentation. The role of documentation is important as well as software itself because it is often impossible to use software without manuals. The difficulties involved with computer software cataloging is to handle the case that cataloging information from documentation or label of storage medium is not same as the one from software itself.

Cataloging a computer software is to catalog a software, not to catalog the documentation. Therefore, the recommended priority listing of 'by running program, from code of program, from documentation, etc.' by Templeton and Witten[11] would be proper. The necessity of priority list also means the possibility of developing different catalogs for the same software. The information which could be found through executing a software or listing code of a program is depends upon the expertise of cataloger's computer skills or availability of special hardware and software. There is a big chance of possibility that a person could find more cataloging information than the other on the same software if no cataloging information is given in the documentation by author or publisher.

Other difficulty is that not all softwares have documents or have only locally developed versions which do not have a complete information like the original one supplied by publisher. The risk of developing incorrect catalog entry is increased if all of the original materials of a software package is not kept properly. No university will be willing to purchase duplicate copies of all the softwares they own only to develop catalog with the original condition.

The conflicts between software itself and supplementary documentation originates from the characteristic of software existing in electronic form. This characteristic makes software easy to

update, but not the whole printed manual. Generally, a few page or a small sized update note to the original manual is all or no update note is provided. This increases the possibility of developing wrong catalog entries.

2-3. Location and call number

Library uses call number to tell patron the physical location of books. This scheme is used by most of the regular libraries. Whenever moving or re-shelving is necessary for a group of items, library can simply change the physical location index on the guide chart. For example, if a group of books which has call number between HA and HF moved from floor 4 to floor 3 of main library, then librarian simply change that information on the guide chart only, not the individual card catalog. The difficulty using this scheme for software is that software is not possible to place based on the classification. Most of the case, location of software is determined only by the location of machine which a software runs on or the real physical location of software does not have to be known at all if only remote access is allowed. The subject based call number scheme may not be a good choice to be used as a preliminary locator to the software.

One possible approach to this problem is that using supplementary printed documentation, so called user guide or manual. The printed documentation can be arranged exactly same way as regular books. Therefore, we may can have user locate documentation first and then let them find information how to access the software or where to loan the software on manual. This could be one of the best workable approaches. The fact that a patron may need to learn how to use the software by reading manual well supported this claim. The expected difficulty with this scheme is that a lot of softwares at university do not include printed documentation and adding all the necessary information to access or locate software in documentation and altering them as necessary would not be a simple task. However, it could be better than hiring people to have them tell patron where it is and how to find it every time.

Another approach is that using codes for a specific group of softwarws. For example, the numerous softwares on CD-ROM are assigned a specific code 'CRD' and a format file switches realistic location information for the code 'CDR' to '131-U Coover Hall.' Whenever the CD-ROM is moved to some other building on campus, the information on the format file only changed rather than the individual catalog one by one. The catalog still has 'CDR', but the format file now will display different location information as changed for 'CDR.' Some softwares need to have a specific access information. In this case, the specific information is given for the software. If the information is necessary to be changed, then individual catalog record must be changed. However, alteration is not so much big trouble as with manual card catalog if it exists in electronic form.

2-4. Electronic form and publication

As discussed before, software exists in electronic form. This is a very convenient feature to develop and to use, but in a sense it creates serious problems for cataloging tasks. Printed materials need to meet a certain qualification to be released to public because of the expenses to make enough copies. However, software in electronic form does not. It is not expensive at all to make copies of it. This feature requires cataloger tough extra work of making judgement whether a software is qualified to be cataloged or not. A lot of softwares are available at ISU through SIG, but surely not all of them are quality software. How do we make decision which one to catalog? We suggest to catalog software only if all of the following criteria are met:

1. An organization or individual must be responsible for a particular record and must make a commitment to actively maintain the catalog by notifying the library of any changes which should be made to the record until it is deleted.
2. The software referred to must be accessible to the organization community. This does imply that use must be unrestricted, only that all members of the organization community have an opportunity to gain access to it.
3. The software referred to must be owned by or under the long term control of the organization. References to software which be summarily withdrawn form circulation by an agency external to the organization should not be included.
4. Software referred to must be reasonably functional, nontrivial and otherwise useful. No guarantee of usefulness is implied by cataloging but too much junk may dilute the database to the point where the database itself is useless.

Item 1 is a fundamental requirement for the inclusion of data. It means the validity of the data must be assured. The requirement of a long term commitment to maintain a catalog record may actually be the only criterion necessary for putting the record in the catalog since we can logically expect that nobody will make a commitment to maintain catalog information pertaining to trivial, ephemeral, or otherwise useless material. Regarding item 2, it seems reasonable to insist that only software which is accessible to a large fraction of the university community be cataloged. Some groups may have easier access to a particular piece of software than other groups, but we must insist that some mechanism exist by which virtually any member of the university community can gain access to the software which is cataloged.

3. Software and cataloging practice

In this section we are going to discuss how do we make catalog, what kind of software do we

have, the information what we could find on software, and so on. In addition, an expected way of searching software will be discussed. Searching a book is usually done with name of author, title, publisher, etc. However, searching a software possibly will be done quite differently. We can easily guess that a software searching would be done with function or system requirements of software rather than the name of programmer or publisher.

3-1. The Softwares at university

The pool of software we have drawn from has been that accessible to students at ISU. We have cataloged most of the software which is in publicly accessible mainframe directories and for which documentation exists in the Computation Center Library. All of the software which is in ISU's Software Clearinghouse, an outlet for VAX academic software, has also been cataloged. Almost all of the microcomputer software we have cataloged is from a CD-ROM collection of about 20,000 programs published by PC-SIG. This collection is available online 24 hours per day on ISU's network and on a dial-up basis.

The cataloged softwares were grouped into three different categories such as public domain softwares, locally developed softwares, and commercial softwares. The total figures of analysis results is not necessarily represent the whole trends of university software collections. Random sampling among all softwares must be made to represent all the softwares what university uses, however, it was not possible due to two reasons. First of all, not all the softwares at ISU are cataloged yet. And also the number of software is increasing too rapidly. Second, no complete lists of softwares were known prior to cataloging for random sampling purpose.

The softwares in different group represent different characteristics on the development of catalog entries. In this study, softwares were classified as Courseware Authoring System(CAS) group, GENERAL group such as the softwares on mini/main frame, PC-SIG software group, and COMMERCIAL software group which runs on microcomputer.

The CAS softwares are developed by instructors at several colleges and universities, and are separated from regular softwares on mini/mainframe. These groups of softwares already has online menu which displays name of software and function to allow browsing for trial except some of them. All of the available softwares are cataloged and analyzed. Some softwares on mini/mainframe are analyzed under the name GENERAL. These are the softwares generally provided for students and staff use. Total number of softwares to be cataloged is not known exactly, but 79 of them are cataloged and analyzed. The softwares under the name PC-SIG are public domain microcomputer softwares. The analyzed 43 softwares are randomly sampled among more than 260 cataloged softwares. However, the 260 cataloged softwares are only the first 5% of the total softwares. The result will represent the total trends better if softwares are sampled randomly from

the beginning to the end, but no seriously different trends are expected. The COMMERCIAL group has commercial softwares available at computer laboratories. Currently the number of analyzed software is relatively very small because university owns small.

3-2. Analysis of software

In this section the results of cataloged software analysis is provided with general comments. A major trend is that the softwares for mainframe usually do not have title screen while the softwares to be used on microcomputer has a good display of title screen. A problem with mainframe software is on documentation. Some of the softwares on mainframe are almost impossible to access without manual. The original manuals supplied with the softwares were not sufficient to be used due to the local modification.

Author

Softwares in CAS group do not follow this, but generally software do not include name of authors on them. This probably true since too many people work together to develop a system in comparison of book such as one to three. Also the development of a software may not be a real creative work. As shown in the Table 1, approximately 40% of the softwares did not have name of authors in GENERAL and PC-SIG group.

Numner of title

Approximately half of the softwares had one title and others had two or more titles. The softwares which has two or more titles usually have an acronym and a descriptive title. Also, it is easy to see manual writer call the software little differently in manual or in online menu.

Source of title

Commercial softwares run on microcomputer and most of the softwares in CAS group have title screen. Therefore, it was possible to have title information by running the software. The softwares in SIG group also show good figures. 42% of the title were found on title screen and 23% from the program codes. Finding title from source code was possible with SIG group softwares because most of them were written in BASIC and the source code was released. GENERAL group only 24%. Therefore, relied more on documentation to find title information.

Version

The information about version is a pretty important due to the frequent update of software. The softwares in CAS and COMMERCIAL group present good rate on this, but GENERAL and PC-SIG group are not shown up good. The rate of known and unknown of software version is approximately 3:7.

Publisher

This information is in trouble only with the GENERAL softwaers on mini/mainframe. This is true with the softwares which do not have any manuals or any written documentation of them or small sized utility programs.

Date

This is about the information for copyright date or published date. The figures indicate similar trend with the information about publisher. CAS and COMMERCIAL group present very nicely, but GENERAL and PC-SIG softwares are half and half for known vs. unknown.

System requirements

Software in CAS or COMMERCIAL group contain these information but the software in SIG does not. Also, the GENERAL group software on mini/mainframe does not. However, this information is not necessary at all to ordinary user if it is available on mini or mainframe computer. All the user need to know is how to access, not the very detailed system requirements as in microcomputer case. System requirements for microcomputer softwares are very important, however, 84% of PC-SIG were not shown. The reason could be the program size is small or it is too obvious to mention if program runs on any IBM or compatible machines.

4. Conclusions and comments

The fundamental difficulty involved with software cataloging comes from the characteristic of exiting in electronic form. This characteristic makes software easy to copy, modify, move, and access on any terminal. Also it makes software real hard to find changes if the information about alteration is not given. Only a management level of decision to set up a policy could solve this problem, but not with cataloging technique.

Another question is that how long a cataloger has to struggle with a software to find cataloging information. Finding necessary information on a software is neither easy, nor simple task. If a manual is not provided, then cataloger does not know anything about the hardware and software requirements to use the software. It is not practical that trying all available hardware and software to find necessary information and to check all aspect of it. The most accruate information can be given only by author or publisher of the software. The cataloging information provided by publisher or programmer could resolve the mess involved in software cataloging works.

TABLE 1. Summary of Cataloging Experience

CLASSI - FICATION	MINI /MAIN				MICRO /PERSONAL				TOTAL	%
	CAS		GENERAL		PC -SIG		CO - MMERCIAL			
AUTHOR										
Unknown	3	6%	30	38%	18	42%	4	100	55	30.7
Known	50	94	49	62	25	58			124	69.3
NO. OF TITLES										
1	19	36	47	59	25	58	2	50	93	52.0
2	33	62	30	38	17	40	2	50	82	45.8
3			2	3	1	2			3	1.7
4	1	2							1	0.6
TITLE SOURCE										
Title screen	40	75	19	24	18	42	4	100	81	45.3
File name			10	13	12	28			22	12.3
Program code			1	1	10	23			11	6.1
Documentation	2	4	49	62	3	7			54	30.2
Online menu	11	21							11	6.1
VERSION										
Unknown	14	26	53	67	32	74			99	55.3
Known	39	74	26	33	11	26	4	100	80	44.7
PUB- LISHER										
Unknown			45	57					45	25.1
Known	53	100	34	43	43	100	4	100	134	74.9
DATE										
Unknown	2	4	44	56	18	42			64	35.8
Known	51	96	35	44	25	58	4	100	115	64.2
REQUIREMENTS										
Unclear	1	2	76	96	36	84			113	63.1
Clear	52	98	3	4	7	16	4	100	66	36.9

REFERENCES

1. Crawford, W., MARC for Library Use: Understanding the USMARC Formats, Knowledge Industry Publication, Inc., White Plains, NY, 1984.
2. Crawford, W. et al., Bibliographic Displays in the Online Catalog, Knowledge Industry Publication, Inc., White Plains, NY, 1986.
3. Demas, S., "Microcomputer Software Collections," *Special Libraries*, Vol. 76, No. 1(1985), pp. 17-23.
4. Dodd, S., Cataloging Machine-Readable Data Files-An Interpretive Manual, American Library Association, Chicago, 1982.
5. Dodd, S. and A. Sandberg-Fox, Cataloging Microcomputer Files-A Manual of Interpretation for AACR2, American Library Association, Chicago, 1985.
6. Faust, J., "Microcomputers as On-Line Catalogs in Special Libraries", *Special Libraries*, Vol. 79, No. 3(1986), pp. 133-139.
7. Gorman, M., "Microcomputers and Online Catalogs," *Drexel Library Quarterly*, Vol. 20, No. 4 (1984), pp. 25-33.
8. Gorman, M.(Editor), *Anglo-American Cataloguing Rules*, second edition. Chapter 9, Computer files (draft revision), The American Library Association, Chicago, 1987.
9. Piele, L. et al., "Teaching Microcomputer Literacy:New Roles for Academic Librarians," *College & Research Libraies*, Vol. 47, No. 4 (1986), pp. 374-377.
10. Stanek, D., "Videotapes, Computer Programs, and the Library," *Information Technology and Libraries*, Vol. 5, No. 1(1986).
11. Templeton R. and A. Witten, *Study of Cataloguing Computer Software : Applying AACR2 to Microcomputer Programs*. Library and Information Research Report 28, The British Library, Wetherby, West Yorkshire, Great Britain, 1984.
12. Troutner, J., *The media specialist, the micro computer, and the curriculum*, Libraries Unlimited, Inc., Littleton, Colorado, 1983.