INTRODUCTION

This article deals with the process of cutter classification, i.e., assignment of cutter numbers for specific monographs. In other words, the focus is placed on the process of finding an appropriate alpha-numeric representation for the monograph’s author or its title, in the event that an author is missing.

The importance of the research which formed the basis for this article, lies in the fact that an old classification problem was solved using a new approach. The solution reached provided higher classification accuracy and resulted in saving classifiers time and energy. The higher accuracy and better time and energy use were the initial goals for striving toward an automatic cutter classification system. Once developed, tested and installed, it allowed classifiers to concentrate more on other aspects of book classification.

This article includes background on practical explanations and a short historical review of cutter numbers, but primarily concentrates on an analysis of the cuttering process and on the way an automatic solution was developed. A thorough description of the CUTT-x, an Expert System for Automatic Assignment of Cutter Numbers is

Mr. Savić lives in Montreal, Quebec, Canada.

The author would like to acknowledge the generous support given to the development of this Expert System for Automatic Assignment of Cutter Numbers by Mrs. Ghislaine Giroux, ICAO Library Assistant.

Cataloging & Classification Quarterly, Vol. 22(2) 1996
© 1996 by The Haworth Press, Inc. All rights reserved.
provided. Special attention is devoted to three basic elements of the expert system: the knowledge base, inference engine, and user interface.

The actual system was designed, tested and implemented in the Library of the International Civil Aviation Organization (ICAO). The application was developed using Microsoft ACCESS relational database management system in the MS-Windows personal computer based environment.

**BOOK LABELS AND CUTTER NUMBERS**

A call number represents "a combination of characters assigned to a library book, to indicate its place on the shelf relative to other books" (*Webster's*, 1976, p. 319). As shown in Figure 1, at the ICAO Library they usually consist of two lines, or two parts: classification number and cutter number. This number may be typed or written on labels glued to book spines. ICAO follows the Dewey Decimal Classification (DDC).

There are different ways to obtain the classification number. The DDC classification number usually has to be assigned by the classifier using classification tables, corresponding classification rules, and existing library practice. Another way to obtain the classification number is for it to be assigned by some external agency, such as numbers found on cataloguing-in-publication (CIP) present in the monograph. The important element is that at ICAO Library the classification number appears on the label as the first line and becomes a part of the call number.

The second line is the cutter number which is the alpha-numeri-
The general purpose of call numbers is to establish a unique identifier for various books and documents enabling their proper shelving, location, and retrieval. Besides this general purpose, the call number serves an additional double-fold function. Firstly, the classification number, from Dewey Decimal Classification (DDC), groups related materials together. Secondly, the cutter number "uniquely identifies different works in the same class" (Wynar, 1992, p. 373), making them, when joined together, an accurate identifier.

The call number practices followed by some libraries also include the assignment of a "work mark," a lower-case letter appended to the cutter number. "The work mark is used only in those cases where the library acquires two or more books by the same author which are assigned the same classification number, making it necessary to distinguish between the different works" (Lehnus, 1980, p. 81). One rule is that the first book by an author added to the library collection is the only one which is not assigned the work mark, while all the others are assigned the first letter of the first word in the title, except if the first word is an article. In that case the first letter of the second word is used.

There are typically four exceptions to this rule: biographies, family histories, bibliographies for individuals, and criticisms. There, "the cutter number stands for a subject, [and] the work mark represents the initial of the author’s surname" (Comaromi, 1981, p. 69). In principle, the work mark should not be used if there is no author. To make the job more interesting for classifiers, work marks can
consist of two letters, if there are two or more books whose title starts with the same letter and are written by the same author. If the entire word is the same, then some other word, such as a keyword, is used. Dictionaries, bibliographies, concordance letters, criticisms, commentaries, and similar works require an additional lower-case letter ‘z’ appended to the work mark. This type of work mark needs an additional cutter number to identify the actual book or work discussed. There are also some additional rules for translations where some specific combinations of letters and numbers are used to indicate a specific language. More details regarding the rules governing the creation of work marks can be found in *Book Numbers* by Lehnus (1980).

It should be noted that the work mark, as a part of the call number used on a book label, is not used by too many libraries. The Library of the International Civil Aviation Organization (ICAO), on whose practice the CUTT-x was based, does not make use of work marks, so the CUTT-x does not support it. The work mark is also not used with the Library of Congress classification scheme. However, from the CUTT-x design point of view, since Lehnus has described distinct rules governing the creation of the work mark, it would not be too difficult to incorporate those rules to cover assignment of work marks.

**HISTORY OF CUTTER NUMBERS**

The cutter number system was designed by Charles Ammi Cutter. Born in 1837, Cutter is best known for designing Expansive Classification. This classification scheme is no longer in use, but it provided a plan on which the Library of Congress classification was developed.

One part of the Expansive Classification was the alphabetical-order tables, also known as Cutter tables. These tables assigned a capital letter plus a two digit number (therefore the name ‘two-figure’ tables), to the author’s last name or the other word from the title if the author was missing. Cutter’s initial table covered all the consonants (except ‘S’), followed by vowels and finally by the letter ‘S’.

Between 1892 and 1895, in order to provide more coverage and
differentiation between names which were of special importance to larger libraries, Kate E. Sanborn developed an expanded version of the tables. She added another digit, so the ‘two-figure’ tables became ‘three-figure’ tables, also known as Cutter-Sanborn tables.

Since Sanborn’s expansion did not include the initial Cutter’s ‘two-figure’ tables, Cutter worked between 1899 and 1901 on expanding his own ‘two-figure’ to ‘three-figure’ tables. However, his ‘three-figure’ tables did not undermine the popularity of previously mentioned Cutter-Sanborn tables, which became the most popular version and are still used by many libraries in conjunction with the DDC classification.

**THE PROCESS OF CUTTERING**

According to some dictionaries like Webster’s (1976), it is allowed to use the word ‘cutter’ as a verb—‘to cutter’ or ‘cuttering’ for example. At the same time, it must be admitted that most of the modern dictionaries do not mention this possibility. Whether grammatically proper or not, library science and librarians use this term to indicate the process of assigning a cutter number to a particular book or document. *Encyclopedia of Library and Information Science* mentions that “the use of any numerical interpolation of the alphabet is called ‘cutter numbers’ and the act of using such numbers is called ‘cuttering’ ” (*Encyclopedia, 1968, p. 383*).

When discussing the process of cuttering it is important to make the following distinction. Author number and cutter number are sometimes used interchangeably, although as noted previously, it is more precise to use the term cutter number. The cutter number is a part of the call number or the book number but it is not its synonym. Therefore, when some authors (Godden, 1984) define assignment of book numbers as the process of cuttering or shelflisting, they are giving cuttering a wider role than it should have. Since the book number may include, besides the cutter number, also the classification number and the work mark, cuttering covers only one aspect. The work mark is not the same as the cutter number, but the process of cuttering may include work mark assignment. During the process of cuttering, the author or the other main bibliographic entry is
replaced by the number from the Cutter-Sanborn tables without going into the area of work marks.

The actual procedure for cuttering is defined in the Explanation of the Cutter-Sanborn Author-Marks as an activity carried out generally in two subsequent steps:

1. Find the first few letters of the author’s name in the table; the figures following added to the initial are the mark.
2. If the first letters of the name do not occur in the table take the letters next previous in the alphabetical order (Cutter, 1969, pp. 3-4).

**CUTT-x DESIGN: AN EXPERT SYSTEM APPROACH**

Following a need for a sophisticated, flexible, yet inexpensive system for information and documentation processing within the ICAO Library, an application was developed using available resources: Microsoft ACCESS relational database management system (Version 2.0), personal computers and MS-Windows. The application developed is called Automated Library System (ATLAS). It is installed on the Banyan-Vines local area network (LAN). MS-ACCESS was used for the creation of all of the ATLAS and CUTT-x components. CUTT-x is incorporated into the classification function, and a corresponding screen (a ‘form’ in MS-ACCESS terminology) is used for processing monographs. The actual program procedure (a ‘module’ in MS-ACCESS terminology) is written using the MS-ACCESS dialect of the computer language, BASIC.

For the user, the CUTT-x is centered around the Monograph Entry Form. It is the place from where the CUTT-x is initiated by double-clicking on the ‘cutter field’. The entry form is also the place from where CUTT-x takes necessary information for its input processing (author and main title), and it is the place where again in the ‘cutter field’ it displays the result of its work. The Monograph Entry Form, among other features, offers a Book Label printing function. It combines the DDC classification number with the cutter number, and using a special Seiko Smart Label Printer PRO, prints the actual book-spine label.

The CUTT-x is developed using expert system methodology with
its basic principles and available application guidelines. It follows a classical expert system structure: user interface; inference engine; knowledge base. Figure 2 shows this structure and it also shows the flow of data that takes place inside CUTT-x. It starts with information input, goes through input identification, normalization, demarcation, matching process, assigns the cutter number and the initial letter, and at the final stage displays the result—the complete cutter number.

**Knowledge Base**

The knowledge base is the foundation of any expert system. It is the founding block which determines all the other elements and the whole structure of inferencing and user interfacing. The knowledge base represents the system of “facts and rules that embody the expert’s knowledge” (Harmon & King, 1985, p. 34). The crucial element that distinguishes various expert system methodologies is the methodology used for knowledge representation. Knowledge representation “implies some systematic means of encoding what
an expert knows about a knowledge domain in an appropriate medium . . . it involves mapping knowledge at the psychological level into a representation of knowledge at the computational level” (Beynon-Davis, 1993, p. 27).

In order to represent knowledge, the expert system designer (sometimes also known as knowledge engineer) has to go through the process of knowledge acquisition. Knowledge has to be firstly acquired through various means in order to be later represented in such a way that it could be used by the expert system. Depending on the source, there are two general ways or two approaches for knowledge acquisition (Savic, 1994) by the knowledge engineer:

1. written knowledge which requires rigorous document analysis and translation of its contents into facts and rules
2. knowledge possessed by a certain individual (domain expert) which requires the process of elicitation through interviews and questionnaires.

In practice, the combination of these two sources is usually the way to get the best results. This combined approach was used in designing the CUTT-x. The Cutter-Sanborn Three-Figure Author Table (1969) and its accompanying Explanations (Cutter, 1969), as well as some other related literature were carefully analyzed, and the necessary rules were extracted making the CUTT-x knowledge base. The Cutter-Sanborn Table was translated and transformed into MS-ACCESS table enabling the expert system to use it for the automatic assignment of cutter numbers.

As shown in the Figure 2, there are rules regulating the use of author or main title fields as text inputs for cuttering; rules regulating the cases when the title starts with an article or with a numeral or rules for when the author’s last name starts with Mc (Mac), St(Saint), Ste(Sainte) or some other related letter combination which needs to be normalized. This process of normalization allows standard use of only apparently different last names and therefore, always provides a proper assignment of the cutter number.

Inference Engine

Inference engine is ‘the brain’ or ‘the logic’ of any expert system. Its role is to guide the expert system in using the facts and the rules
stored in the knowledge base. Besides the role of using facts and rules the inference engine also has an important role to play in deciding the order or the sequence in which the rules will be tested or ‘fired’.

Harmon and King (1985), in their book on expert systems discuss the ways of drawing inferences and the inference engine structure. According to them, there are two different components of an inference engine with its corresponding types:

**Inference Strategy**

*Modus ponens.* This strategy uses the logic where if A is true and the rule is “If A then B,” then the conclusion is that B is also true.

*Reasoning about uncertainty.* Uncertainty is handled by attaching a number to a fact determining its certainty which is used later for the calculation of final certainty degree.

*Resolution.* This strategy is based on translating IF-THEN statements to OR statements, and then combining those OR statements.

**Control Mechanism**

*Backward and forward chaining.* Backward chaining starting position is the goal which has to be known and is to be reached. That is the reason the expert systems based on the backward chaining are also called “goal-directed systems.” This type of reasoning strategy is only possible with a small number of possible solutions or outcomes. Forward chaining is an approach where possible outcomes are numerous so the system has to go from one rule to another in order to gather as much information as possible to finally select an answer. Because of this data dependence, the forward chaining expert systems are also known as “data-driven systems.”

*Depth-first and breadth-first search.* Most of the expert systems use depth-first search when the system follows a meaningful interdependent chain of questions aimed at a particular direction.
breadth-first based expert system usually examines a whole array of rules through corresponding questions, or by other means. Often this line of questioning seems to be random and not connected at all. Only after examining all the questions at one level can the expert system proceed to the following level.

Monotonic and nonmonotonic reasoning. The main characteristic of a monotonic system is that one's confirmed values or attributes remain unchanged until the end of the particular case. A nonmonotonic system, however, allows the values to be changed at a later stage due to a change of some other attribute. Nonmonotonic expert systems are not as common as monotonic ones and they require a careful system for following and restructuring the line of reasoning.

The inference engine built in CUTT-x, like those built in some other systems, uses some of the concepts mentioned here. With regard to inference engine strategy, Modus ponens is the general approach utilized. The following example illustrates the use of Modus ponens strategy as applied in the CUTT-x:

If A is true
rule: if A then B
then B is also true

if <Amstrong is the author> is true
then <the cutter number is 735> is true.

In other words, every time the system locates the author in the Cutter-Sanborn table it will rightfully assume that its corresponding cutter number is the one related to that name.

Another example of the Modus ponens in action is the elimination of articles as the starting words in the input. If the system finds that the title (in the absence of the author) starts with an article, it eliminates the article and assumes that the remaining title is the part to be matched against the Cutter-Sanborn Table.

Regarding the control mechanism of the inference engine, CUTT-x is a data driven expert system and therefore, it uses forward-chaining. The inference engine moves forward from the intermediate results temporarily stored in the working memory, matching them against the rules and facts, towards the ultimate goal. The extensive number of possible options (over 12,000 authors listed
with the Cutter-Sanborn tables) was the main reason for adapting the forward-chaining approach.

Figure 2 illustrates the structure and the data flow through CUTT-x. There are six processing stages, or functional parts of the inference engine: Identification; Normalization; Demarcation; Matching; Number assignment; Letter assignment. The actual processing sequentially follows the stages until it reaches its goal, cutter number assignment, and displays it through the interface.

The process of identification aims at locating or identifying the text sequence which will be used for later processing. To do this, the system has to consult the rule on use of either the author or the title field from the input form. Whenever the system locates the presence of the author, according to the rule, the author input becomes the text string for further processing. If the author is missing the main title field becomes the text sequence. In the process of identification the system also has to do other accompanying tasks. The first one is to make sure that only the first author is used as a text sequence, and secondly to correct the author format used for the data entry making it adequate for CUTT-x processing. Data entry obeys the following sequence: `<last name><comma><space><first name/or initial><period to follow initial>`. CUTT-x format does not require punctuation marks, therefore commas and periods are eliminated during this stage.

Once identified, the text sequence undergoes the process of normalization. Here, the system checks the starting characters of the input in order to determine if the starting character is in fact an article or a numeral (Arabic number). If the answer is negative, the system proceeds to the following stage. Otherwise it goes through a set of IF-THEN rules eliminating articles and transliterating the numerals. The system also checks if the author’s name starts with ‘St’, ‘St.’, ‘St-’, ‘Ste’, ‘Ste.’, ‘Ste-’ normalizing it to ‘Saint’, or ‘Sainte’ respectively. A similar processing is done with ‘Mc’ or ‘M’ where the input is normalized to ‘Mac’ and therefore always assigned a proper cutter number.

Demarcation is the stage which helps with speeding up the system during subsequent processing. The analysis done on the Cutter-Sanborn Three-Figure Author Table showed that the longest author’s name in that table has a total of 13 characters (including
spaces). This finding enabled the demarcation of the input text sequence to the same amount of 13 characters. So, instead of matching and comparing a whole main title field, which could be long, the system uses a demarcated string of 13 characters maximum.

Matching is the heart of the whole inference engine and the CUTT-x system. A special algorithm was developed which is capable of matching the input text sequence with the Cutter-Sanborn Table. Besides this simple task of physical matching, the algorithm also compares the input to entries in the Cutter-Sanborn Table in order to accommodate the rule where “if the first letters of the name do not occur in the Table take the letters next previous in the alphabetical order” (Cutter, 1969). For example:

\[
\begin{array}{l}
\text{Cutter-Sanborn Table} \\
\text{Anderson} & 545 \\
\text{Anderson D} & 546 \\
\text{Anderson J} & 547 \\
\text{Anderson M} & 548 \\
\end{array}
\]

Direct matching for authors Anderson, Clark and Anderson, George, for example, would produce the same cutter number—545. In the case of Anderson, George that cutter number is incorrect. The correct one would be 546. In other words, the span of letters between D and J, as in this example, should be classified under letter D.

It was possible to solve this challenge by expanding the actual Cutter-Sanborn Table in order to include all the missing letters in all the gaps, but this approach was not found to be very practical. Firstly, the number of table entries would probably become 4 to 5 times larger, while the actual process of data entry would open possibilities for mistakes. Secondly, the performance speed would definitely deteriorate with such an increased number of entries, making the system less attractive for use. Instead, an algorithm was developed which can do the same task without using an extended table.

Number assignment and letter assignment are the last stages of input processing related to the inference engine. They are also the result, or the outcome, of the matching stage. Once the match is successfully established, using Modus ponens principle, the system assumes that the corresponding cutter number is the right one. CUTT-x assigns the initial cutter letter, using the rule that this letter
is always the author’s last name initial. In the case that there is no author, it is the starting letter from the title, while obeying the rules on articles and numerals. Then the system supplies the numeric portion of the cutter number from the table, according to the matching step above.

**User Interface**

The third element of any expert system is the interface which is “the means by which users interact with the expert system” (Pedersen, 1989, p. 40). Every expert system has some type of an interface, because, as Mishkoff in his book on artificial intelligence noted, “even the most sophisticated expert system is worthless if the intended user cannot communicate with it” (Mishkoff, 1986, pp. 3-5).

Different expert systems use different approaches, among which the most common are queries: lines of questions demanding answers through typed inputs or through multiple choices. They are increasingly combined with windows technology and graphics features available today on almost any personal computer. Two main functions of the user interface are to provide input and output points. They are also often used for editing the knowledge base, both rules and facts, and for displaying the explanations as to why a particular solution or answer was selected. In essence, the explanations offered are usually displays of the expert system’s actual lines of reasoning. The user interface designed for CUTT-x provides input and output points, as well as the triggering mechanism. Any editing of the knowledge base itself, for whatever reason, could be done through the regular program, known in MS-ACCESS as *module editing*. The Cutter-Sanborn Table is accessible through standard database functions. CUTT-x does not provide an explanation to the user of its inference or reasoning process primarily because it was deemed unnecessary.

CUTT-x user interface is incorporated in a larger window-based user interface designed for monograph cataloguing, connected to the cataloguing screen. That screen, known as the “Monograph Entry Form,” is shown in Figure 3. The main approach adopted for CUTT-x with regard to the actual interface design, had to fulfil three requirements:
• to be simple to use
• to require no additional data entry
• to be efficient.

The requirement to develop a simple interface for use was important because the Monograph Data Entry Form was already reaching the maximum acceptable level of complexity. Any additional buttons, fields, selection or combo boxes, and similar form objects had to be avoided. So the decision was made to use the available Cutter number field as a triggering mechanism, and the actual output. A double-click on the Cutter Number Field initiates the expert system whose work ends with the cutter number being displayed. Existing information in the author and the main title fields are used as inputs.

The requirement that there should be no additional data entry was also sensible, since the information was already present. Retyping something already entered would only introduce the chance of error in rekeying information. The CUTT-x system simply takes the data already available on the screen, after deciding to use either author or

FIGURE 3. Monograph Entry Form
main title field based on the built-in rules, and processes that information as an input text for cutting. The system also has a safety mechanism built in to warn the user of cases when both author and title fields are empty.

The element of required efficiency is understood since the process of cutting is straightforward, always bringing a successful result—a correct cutter number. There is no need to double check the cutter number found, nor to verify the rules applied during the reasoning process. Since it was assumed and later proved, through extensive testing, that the CUTT-x is a reliable system, it was not necessary to develop and build in any checking or verification procedures to be performed manually by the classifier or by the cataloguer.

CONCLUSIONS

Before drawing any conclusions from the development of the CUTT-x automatic classification system it is worth acknowledging that the system has performed well for the ICAO Library from the very beginning. The perfect accuracy of CUTT-x, as well as the time and effort saved through its use, fully justify the decision to experimentally develop the system. Benefits of its use are obvious and it is believed that this article gives enough detail regarding its structure and functioning to encourage other librarians to develop their own CUTT-x clones. However, it should be noted that larger libraries require more complex cutting and therefore, they would require a more complex CUTT-x system.

Besides this practical side, there are at least two other conclusions that could be drawn here. At a more general level, it demonstrates again that libraries can benefit from using modern methodologies, such as expert systems. It is not only ‘the technical world’ that should make use of modern methodological approaches and technologies. It is up to librarians to look for new ways of improving existing practices and increasing efficiency by making use of already developed and available technologies.

On a more specific level, it could be concluded that, just as CUTT-x has demonstrated an application of technology to the area of cutter number assignment, new technologies could also be ap-
plied to the area of classification. Library classification is known to be a very time consuming and technical activity where any enhancement, efficiency improving, and time saving device would be very much appreciated. The actual experimental applications do not have to offer fully automatic classification. Instead, as an intermediate stage, they can offer solutions in the form of automatic assistance classification, which reduces unnecessary manual work while retaining the important role played by human classifiers. Expert systems technology offers possibilities for improving the classifier's work, for example, through offering look-up lists; offering a list of probable solutions; comparing the found classification term/number with previously classified items, or it could be of help in some other appropriate way. Whatever the chosen application may be, expert systems are to play a greater and greater role in library classification.

NOTE

1. "A table is simply a collection of information that's neatly organized into rows and columns . . . in database management terminology, we refer to each row in a table as a record, and each column as a field" (Simpson, 1993, p. 7).

REFERENCES


Cutter-Sanborn Three-Figure Author Table. Swanson-Swift revision, 1969.


---

for faculty/professionals with journal subscription recommendation authority for their institutional library . . .

If you have read a reprint or photocopy of this article, would you like to make sure that your library also subscribes to this journal? If you have the authority to recommend subscriptions to your library, we will send you a free sample copy for review with your librarian. Just fill out the form below—and make sure that you type or write out clearly both the name of the journal and your own name and address.

( ) Yes, please send me a complimentary sample copy of this journal:

________________________________________

(please write in complete journal title here—do not leave blank)

I will show this journal to our institutional or agency library for a possible subscription.
The name of my institutional/agency library is:

________________________________________

NAME:_____________________________________
INSTITUTION: _______________________________
ADDRESS:___________________________________
CITY:_____________ STATE:_____________ ZIP:________

Return to: Sample Copy Department, The Haworth Press, Inc.,
10 Alice Street, Binghamton, NY 13904-1580
CATALOGING & CLASSIFICATION QUARTERLY

Editor:
Ruth C. Carter, MA, MS
Assistant Director for Automated and Technical Services, University of Pittsburgh Libraries, Pennsylvania

"Readable, strongly researched articles are made more useful by EXTENSIVE USE OF EXAMPLES AND BIBLIOGRAPHIC SUPPORT."
Information Retrieval & Library Automation

"Allows a refreshing view of both 'the forest and the trees'... Should indeed acquire significance for catalogers, administrators, public service librarians, and library science students, as a thoughtful forum concerning how and why bibliographic organization is one of the more dynamic aspects of librarianship."
Bulletin of the Medical Library Association

Sample copies of journals are available free of charge for libraries and faculty with library recommendation authorization. Please send your request to us on institutional letterhead to Sample Copy Department. Allow 6-8 weeks for delivery.

Contents

Volume 22, No. 2: The Nexus of Subject Analysis and Bibliographic Description
Cataloging Southeast Asian Language Materials
Changing Cataloging Rules in Relation to Changing Patterns of Publication
Queens Borough Public Library's Guidelines for Cataloging Community Information
CUTT-x: An Expert System for Automatic Assignment of Cutter Numbers
Cataloging News

Volume 21, No. 1—Fall 1995.
Volume 22, No. 1—Spring 1996.
Quarterly (4 issues).
Subscription rates (per volume):
Special Discounted Rate for Individuals: $48/
Institutions: $150/
Libraries: $150

We will send you a subscription to the current published volume.
Orders outside US/Canada/Mexico must be prepaid by check or credit card.
Outside US and Canada: please add 40% to journal price.
In Canada: Add 30% to journal price.
Canadian residents: please add 7% GST to journal total.
If paying in Canadian funds, please use the current exchange rate to convert total to your currency.
Payment in UNESCO coupons welcome.
Discounts are good only in US and Canada.
Discounts are not good in conjunction with any other offer.
Discounts are not good to subscription agencies.
Prices subject to change without notice.

The Haworth Press, Inc.
10 Alice Street, Binghamton, New York 13904-1580 USA