4

Knowl. Org. 24(1997)No.1
H.B. Irving: Computer-Assisted Indexing Training and Electronic Text Conversion

Computer-Assisted Indexing Training and Electronic Text Conversion at NAL

Holly Berry Irving

National Agricultural Library, U. S. Department of Agriculture

Holly Berry Irving is an Indexer (Technical Information Specialist, Biological Sciences) for the National Agricultural Library, specializing in human nutrition, food science and food technology. She is a Registered Dietitian and Nutritionist, and a member of the District of Columbia Metropolitan Area and American Dietetic Associations.

ABSTRACT: The Indexing Branch of the National Agricultural Library (NAL) has launched a project to move from their traditional one-on-one method of training novice indexers to a computer-based, largely self-directed training program. CAIT, the Computer-Assisted Indexing Tutor, has been designed to standardize, streamline and improve the quality of indexer training. In addition, to support the increased need for all NAL indexers to more fully utilize microcomputers for indexing production, the Branch investigated the "reinvention" of paper versions of frequently used indexing tools as electronic references.

1. General Background

As the nation's primary source for agricultural information, the National Agricultural Library (NAL) has a mission to increase the availability and utilization of agricultural information for researchers, educators, policy makers, and consumers of agricultural products. The Library is one of the world's largest and most accessible agricultural research libraries and plays a vital role in supporting research, education, and applied agriculture.

The Indexing Branch of NAL indexes approximately 1,400 journals and 500 monographs on agriculture and related subjects for both the NAL online public-access catalog, called ISIS and AGRICOLA (AGRICcultural OnLine Access), the Library's bibliographic database. AGRICOLA contains records for agricultural literature citations of journal articles, monographs, theses, patents, computer software, audiovisual materials, and technical reports. Indexing records comprise approximately 85 percent of the AGRICOLA database.

Two types of indexing are required for each indexing record: descriptive indexing and subject indexing. Descriptive indexing provides data on the physical/bibliographic aspects of the work. Subject indexing is the process of assigning descriptors (or terms) from a controlled vocabulary, identifiers which are indexer-defined and subject category codes which adequately express the subject matter content of the work. NAL indexers use the *CAB Thesaurus* (CAB, 1995) as their controlled vocabulary to assign terms. It has over 59,000 terms covering all aspects of agriculture, horticulture, animal production and forestry. The *AGRICOLA Subject Category Codes with Scope Notes* (USDA, 1993) are used for the selection of codes. The "category codes", as they are called by indexers, divide the field of agriculture into 21 broad subject units, each of which is further subdivided. These subdivisions are grouped by subject matter and assigned an alpha-numeric notation consisting of a letter and three numbers. This code, with a title known as a "subject heading", denotes the specific agricultural category (e.g. F200 - Plant Breeding and Genetics, E100 - Agricultural Economics, L500 - Animal Nutrition, Q120 - Microbiology of Food Processing). Both descriptive and subject indexing also involve the assignment of appropriate MARC (MAchine Readable Cataloging) tags. MARC is a standardized system which allows for the creation, storage and retrieval of information between users and libraries.

2. Project Background

When new indexers join the staff of the National Agricultural Library (NAL), they generally spend six

Knowl. Org. 24(1997)No.1
H.B. Irving: Computer-Assisted Indexing Training and Electronic Text Conversion

5

months to one year under the tutelage of an experienced indexer, learning the complex procedures and intellectual skills required to maintain the high standards of the AGRICOLA bibliographic database. This one-on-one educational approach has two major disadvantages: 1) It requires a significant time commitment on the part of an experienced professional, creating a serious reduction of productivity. Trainers report an overall 25 to 30 percent decrease in their indexing production during the usual 12-18 month period before the novice can index independently. This decrease may be as high as 50% during the initial one or two months when a new employee's needs are greatest. 2) Although they are highly qualified to perform other aspects of their jobs, experienced staff members are not always the most effective teachers.

In addition, microcomputer software and hardware developments have made it possible for individuals to create, process, and access extraordinary amounts of widely varied information. Indexers can greatly benefit from these capabilities, using data processing and reference tools that will let them index in a "seamless" manner, using their computer screens as a central access point for information and communications as they work. Converting existing in-house print tools into electronic format, therefore, has become of increasing importance.

3. Goals and Expected Benefits

For the development of a training program, the primary project goals are to provide training for NAL indexers that is comprehensive, consistent, high quality, and largely self-directed and, therefore, less time-intensive. The expected benefits are: reduced time required for novice indexers to achieve journeyman level; reduced time investment of experienced indexers in training of novices; increased productivity; increased consistency and quality of the AGRICOLA database; opportunities for other NAL staff, outside of the Indexing Branch, to more easily learn about indexing; updating the existing NAL Indexing Manual; creation of an "online" indexing manual as a reference tool; simultaneous development of an Indexing Technician training program; and potential use by others outside NAL.

For the conversion of print publications to electronic format, the goals were to: 1) reinvent an existing print working tool as an accessible and accepted electronic reference which supports the established ways in which the print version is used but also encourages new exploration and information seeking; Figures 1 and 2 illustrate the use of a complete electronic work space for instructional purposes and for electronic reference for all NAL staff. The first publication converted to an electronic format was *AGRICOLA Category Codes with Scope Notes,* the reference used to assign the alphanumeric classification(s) mentioned above.

4. Project Design

4.1 CAIT

CAIT (pronounced KATE), the instructional program, is divided into four sections, addressing three major subject areas. Each section has several individual topic chapters providing thorough and comprehensive coverage of each area: *Section One – "Introduction and Orientation"* features information on using the program, along with an explanation of the Indexing Branch and overall NAL function and organization. *Section Two – "Descriptive Indexing"* discusses descriptive bibliographic data. *Section Three – "Subject Indexing"* covers MARC format, category codes, use of the *CAB Thesaurus*, NAL indexing rules and policies, and indexer workflow (Figure 2). *Section Four – "Resources"* offer supplementary reference materials and exercises.

CAIT: 1) simulates actual indexing procedures; 2) presents information in
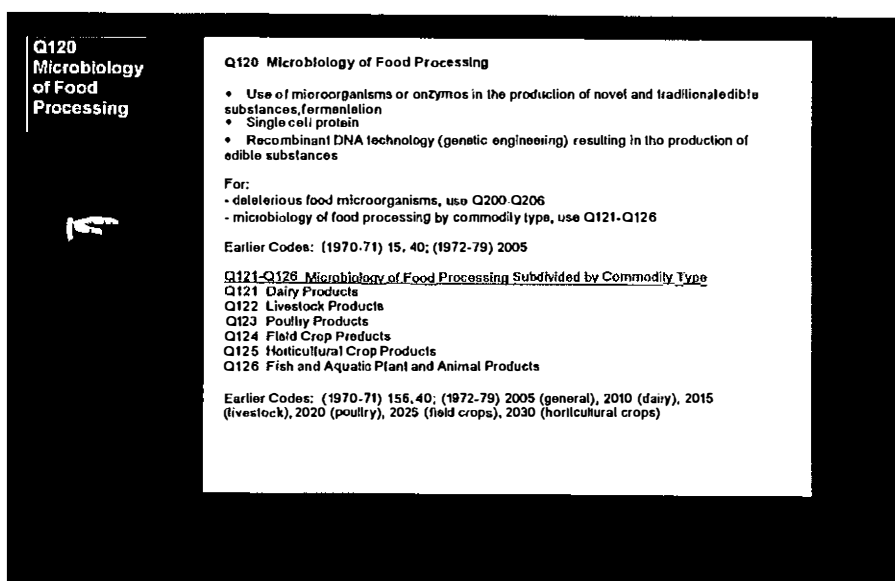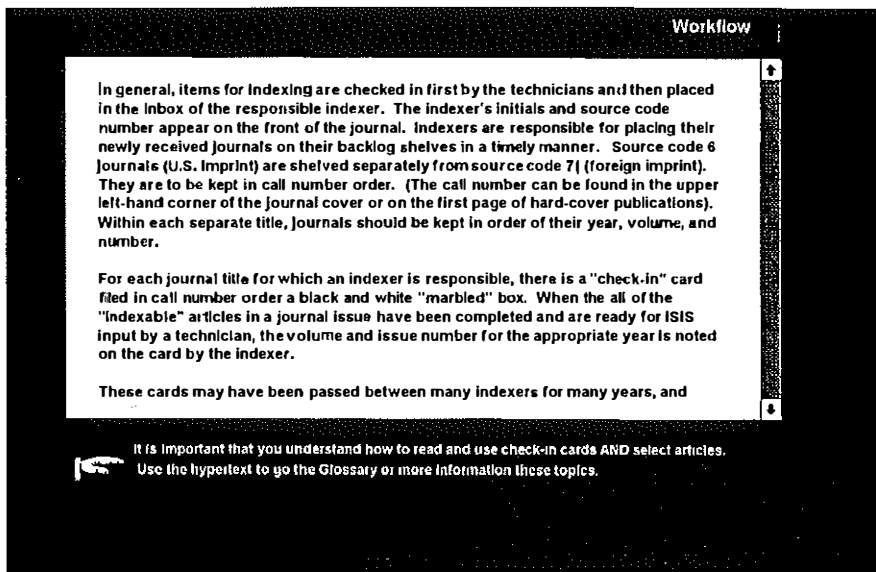

Fig. 1

Fig. 2

Fig. 3

staff or visitors), and an experienced NAL Indexer who serves as a Mentor to provide the needed "human touch" by: making introductions to other staff members; gathering needed materials; making the Trainee comfortable with the computer-based training; and serving as an experienced "second opinion" on more complex subject indexing issues.

4.2 Electronic AGRICOLA Category Codes

The main objective was to integrate an electronic reference tool into the daily work habits of Branch indexers. The electronic publication was designed to:

1) maintain the basic structure and organization of the original AGRICOLA category codes and create a familiar screen format consistent with the printed version; 2) incorporate functionalities which provide an easy-to-use and convenient alternative to the print text, such as: the ability to examine information from two or more codes; extensive utilization of hypertext; and ability to search text or specific code; and 3) account for the varied degrees of computer expertise within the Branch.
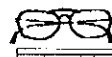
a clear, concise, task-oriented manner that is consistent with the principles of adult education and computer education; 3) provides examples of articles that illustrate various NAL indexing rules and philosophies; 4) utilizes "hypertext" to facilitate the user's access to information; and 5) includes review questions and exercises to reinforce the presented information (Figure 3).

The complete training program includes: 1) the computer-based training program; and 2) printed and/or electronic reference materials and working tools: including the *CAB Thesaurus; AGRICOLA Guide to Subject Indexing* (USDA, 1990).

Participants in the program include a Trainee (new indexer, new technician, student help, interested NAL-

5. Implementation

In-house expertise and resources were used to develop, produce, implement and evaluate this program. ToolBook™ by Asymetrix and Microsoft Windows™ was used to produce the computer-based portion of the training program. Printed materials are produced using standard word-processing microcomputer programs and desktop publishing. Other resources and services such as scanning of text and images, obtaining of documents for training purposes, etc. have been secured from existing NAL resources.

Knowl. Org. 24(1997)No.1
H.B. Irving: Computer-Assisted Indexing Training and Electronic Text Conversion

7

6. Project Evaluation

The computer files were loaded onto the microcomputers of interested indexers in the Branch. Documentation and questionnaires were provided for each evaluator.

6.1 Evaluation of CAIT

Eight members of the Indexing Branch evaluated the initial "CAIT Evaluation Module", developed to demonstrate CAIT's instructional features in an abbreviated format. Most found CAIT's overall appearance and performance to be satisfactory, although three individuals indicated that they felt "lost" or confused about how to proceed through CAIT. There was a great deal of interest and enthusiasm for the use of hypertext. Six of the evaluators had some trouble understanding the directions for the exercises. However, most individuals found that the exercises reinforced the concepts presented and found the level of difficulty to be just right.

Overall, respondents felt that CAIT shows a great deal of promise. The evaluators were satisfied with its basic structure and design. They seem to feel that it has the potential to fulfill its intended purpose as a comprehensive training program for novice indexers, a source of reference information and review for experienced indexers, and an awareness/educational tool for other interested library personnel.

6.2 Evaluation of the Electronic AGRICOLA Category Codes

Nine NAL indexers assessed the efficiency and potential value of the electronic reference, focusing specifically on: 1) navigational functions built into the text, 2) the user friendliness of the interface, 3) help screens within the text, and 4) preferences between print and electronic texts for assigning category codes.

All primary navigational functions (hypertext, text search and searching by code) were given the most favorable comments and highest ratings. The majority of indexers favored the print source when comparing codes. However, they did favor the electronic text for finding codes, and locating referral and relator codes. In assessing the efficiency of the electronic text, five of nine respondents favored the electronic text, and overall, six of the nine indexers favored the electronic text.

7. Summary

The data gathered from the initial evaluation was used to finalize the structure of the basic CAIT instructional program, although its content continues to grow as additional exercise questions are added.

New indexer trainees who have used CAIT as the New indexer trainees who have used CAIT as the first step in their training have responded favorably.

The electronic version of the *AGRICOLA Subject Category Codes* have been updated to include refinements in the interface; improvement in the performance of existing functions; "bookmarks" for rapid access to frequently used codes; and the ability for individuals to create personalized annotations or "sticky notes". Because our data suggests the NAL Indexing Branch is ready to utilize computer technology in daily branch tasks, other indexing reference works will be converted to similar electronic texts[1].

Notes

1. The Branch has made the *AGRICOLA Subject Category Codes* and the *AGRICOLA Guide to Subject Indexing* available on the Internet (http://www.usda.nal.gov/indexing/). A demonstration of CAIT can also be downloaded from this same URL. Free print copies of these references or floppy disks with the CAIT demonstration may be obtained from: National Agricultural Library, Agricultural Research Service, U.S. Department of Agriculture, Indexing Branch, Room 011, 10301 Baltimore Avenue, Beltsville, MD 20705-2351, USA.

References

CAB. (1995). (Center for Agriculture and Biosciences International). *CAB Thesaurus* (1995 ed., Vols. 1-2). UK: CAB International. 1315p.

USDA. (1990). (United States Department of Agriculture). *AGRICOLA Guide to Subject Indexing*. Beltsville, MD: National Agricultural Library. 21p.

USDA. (1993). (United States Department of Agriculture). *AGRICOLA Subject Category Codes with Scope Notes*. (Modified AGRIS), Beltsville, MD: National Agricultural Library. 223p.

Ms. Holly B. Irving, National Agricultural Library, Indexing Branch, USDA, 10301 Baltimore St., Rm. 011, Beltsville, MD 20705 USA