Title

AN INTEGRATED RECORD INPUT SYSTEM (IRIS) FOR THE TOE/TECHNICAL INFORMATION CENTER

Author

Lawrence, J.D.

Publication Date

1983-10-01

Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA

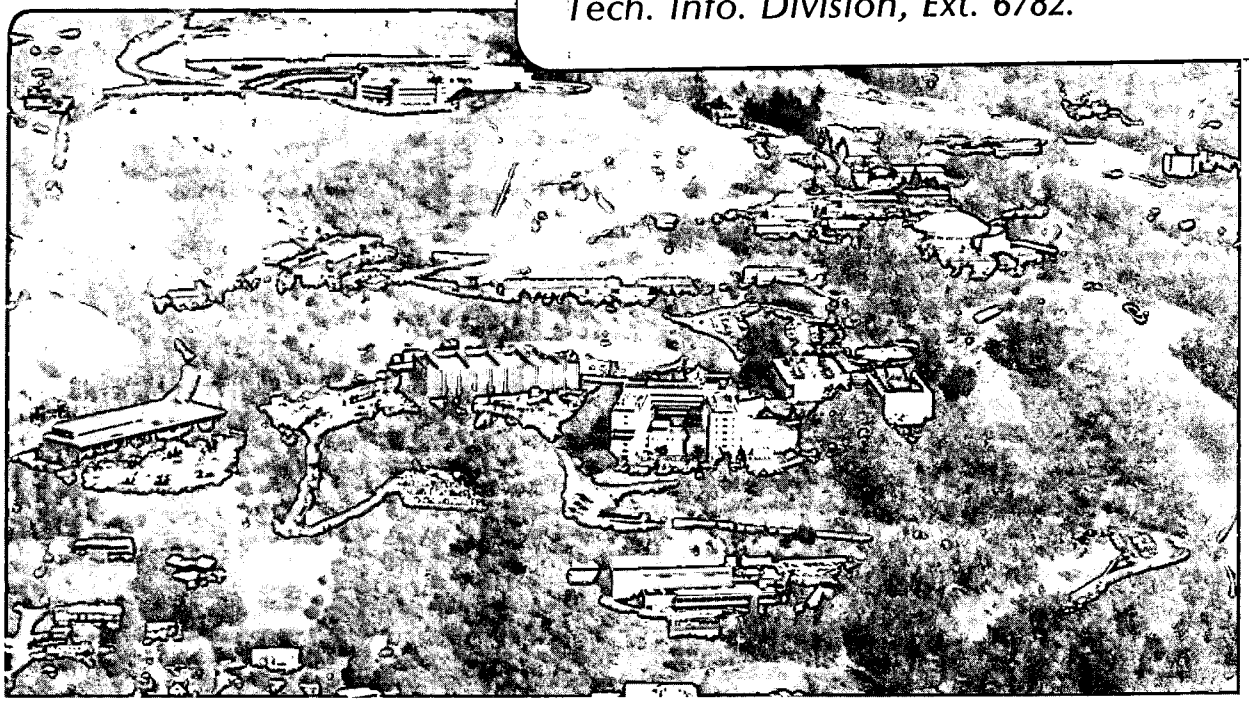
Engineering & Technical Services Division

Presented at the Annual Meeting of the American Society for Information Science, Washington, DC, October 2-6, 1983

AN INTEGRATED RECORD INPUT SYSTEM (IRIS) FOR THE DOE/TECHNICAL INFORMATION CENTER

J.D. Lawrence

October 1983



Prepared for the U.S. Department of Energy under Contract DE-AC03-76SF00098

DISCLAIMER

CONF-8310247--1
LBL-17329

An Integrated Record Input System (IRIS) for the DOE/Technical Information Center

J. Dennis Lawrence

Lawrence Berkeley Laboratory
University of California
Berkeley, California 94720

October 1983

This work was supported by the Director, Office of Energy Research, Office of Basic Energy Sciences, Materials Sciences Division of the U. S. Department of Energy under Contract No. DE-AC03-76SF00098.

An Integrated Record Input System (IRIS) for the DOE/Technical Information System

J. Dennis Lawrence

Lawrence Berkeley Laboratory
University of California
Berkeley, CA

ABSTRACT

A system to provide computer assisted aids to document indexing for the Department of Energy Technical Information Center Database is being developed. One aspect of this system is a special-purpose full-screen text editor. This paper describes the need for a special-purpose editor, experience learned from a prototype, the current status of the project, and implications for TIC's productivity.

1. Introduction

The U. S. Department of Energy/Technical Information Center, Oak Ridge, TN, is responsible for (among other things) preparing document citations for inclusion in the Energy Information Database. The system being developed for this purpose is called IRIS, the Integrated Record Input System. The Lawrence Berkeley Laboratory, Berkeley, CA, as a contractor to DOE, is responsible for developing the computer programs to be used in the new system.

The discussion here is focussed on one small, but vital, aspect of the computer system: an interactive special purpose full screen editor. We first describe the problem, and then briefly discuss a prototype editor developed over the past few years. The major portion of the paper will focus on design features of the new editor, based on lessons learned from the prototype. Finally, we shall give the current status of the project, and implications for TIC's productivity.

The paper is in the nature of a progress report on the project, which was first described at the ASIS midyear meeting in Knoxville in June 1982, by Barbara Cerny [1]. For a more complete background, see references 2 and 3.

2. Statement of the Problem

The Energy Information Database contains about 1 ¼ million citations, giving information about documents related to energy: title, author, publication source, abstract, and subject descriptors. The database grows by about 18,000 citations per month.

The Technical Information Center evaluates 20-30,000 citations per month for possible inclusion in EDB. Citations are edited for grammar and accuracy of description, and subject descriptors (or keywords) are supplied as needed, from the EDB thesaurus of 24,000 indexing terms. All this is done by a staff of 20 subject area specialists, referred to here as indexers. Allowing for vacations, sick leave, memo writing, meetings, and so on, these numbers work out to about 5-6 minutes per citation.

Current processing procedures involve a mixture of batch computer processing and manual editing of computer printouts. One aspect of the IRIS project is to eliminate many of the inefficiencies of this process by providing on-line access to the citations for the indexers.

Project goals have been, and are, as follows: The indexers will have direct access to the citations. This will be through a full screen text editor, whose commands have been tailored to this application. Response is to be rapid — less than ½ to ¾ second for most simple commands. There will be immediate indication of errors, so corrections can be made within seconds of making the error. Most commands will be 1-2 keystrokes. Newly developed terminal capabilities will be incorporated when applicable and desirable.

3. Prototype System

Let us imagine, then, an indexer sitting at a terminal using the prototype system. He has selected a citation to edit; the menu shown in Figure 1 is displayed on the terminal screen [3]. This menu gives a portion of the commands available — the set shown allows the indexer to choose one of the fields for editing. A set of sixteen fields has been selected as interesting to the indexer; those actually present in the record are indicated in the menu. There are, for example, seven fields concerned with document titles and subtitles; generally, only one or two of these are present in a citation.

Suppose the indexer wishes to edit the abstract. He types the 'a' key of the terminal, and sees the display shown in Figure 2. He now has available a general-purpose full screen editor, with the usual editor commands: move cursor; add, delete, or replace characters; move text around; and so forth. When editing is done, a single keystroke returns the screen to the menu.

Now, to edit the keywords. Typing 'k' yields the display shown in Figure 3. The same editor commands are available. A two-level hierarchy of keywords may be created for the citation, termed *main term* and *qualifier*, or M's and Q's. Keywords may be typed, as indicated in the figure.

When editing is finished, and the indexer indicates his desire to return to the menu, extensive checking is done. A variety of error conditions can be detected, and correction is required before control returns to the menu. Terms not in the thesaurus are flagged, main terms without corresponding qualifiers are found, and so on.

The indexer has on-line access to the thesaurus. A portion of the thesaurus centered around any term can be displayed (Figure 4), giving two levels of broader terms, two levels of narrower terms, and the related terms. This can help the indexer to find the precise terms to describe the document.

This prototype was completed under the direction of Barbara Cerny in mid 1982. Testing was carried out in late 1982, to determine if the whole idea be feasible. This testing was informal, and carried out under rather adverse conditions. The indexers received little training and practice before the testing began. The computer being used was (and is) overloaded, so response time was poor. No facilities were available for easily selecting descriptors — all must be keyed in.

Under these conditions, we felt that the system would show promise if, during testing, the indexers were only 2-3 times slower than their manual methods. In fact, editing using the computer system required almost exactly the same amount of time as the manual system. A bonus turned up: a group of people known as descriptive catalogers, who were not included in the original set of users, tried our editor, and found it speeded up their work by a factor of three.

Some problems turned up; problems serious enough to require a complete redesign and rewrite of the editor. The first problem concerns the amount of information available to the indexer at any given time. Consider the actions of an indexer, wanting to generate index terms for a citation.

First, the menu. Then, go to the abstract. Reading this enables the indexer to choose several descriptors. Now, back to the menu, then to the keywords. The first new term is keyed in; 5-10 seconds have elapsed. The indexer then finds he has forgotten the next term he wanted, so back to the menu, abstract, menu, keywords. A lot of time is spent in this back and forth — and it becomes very frustrating for the indexer. The menu, since it doesn't change much, becomes very boring to look at, and an increasing irritation to the indexer.

The second problem concerns keystrokes. Suppose you have a portion of the thesaurus on the screen, and see a term you want to enter — say, "beta-plus decay radioisotopes" — why can't you tell the computer to select this as an index term, rather than having to key it in (30 key strokes, in this case)?

A third problem relates to efficiency. It became apparent to all of us that the prototype was too large and too slow for a production program.

4. The Revised Editor

In February of this year, we began rewriting the editor. The new editor is based on EMACS [4-6], developed at MIT; indeed, most of the features of our editor are taken from EMACS, and modified or extended to fit our special needs. We intend to support multiple windows, allowing (for example) the indexer to have portions of the title and abstract on part of the screen, and subject descriptors on another part. Switching from one display to another will be possible with a few keystrokes. If a thesaurus display be on the screen, index terms can be selected for inclusion in the list of descriptors by simple cursor positioning. We are able to cope with multiple terminal types quite easily, and coding will be flexible and modular enough to add new hardware as it is developed.

The basic change in system philosophy is from a menu-based star-like sequence of operations to a more flexible network-like arrangement. The key to this is multiple windowing, using split screens (Figure 5). There are two ways to split a screen — top and bottom, or right and left. We will allow each indexer to choose which he prefers. Initially, when a citation is made ready for editing, all fields of interest to the indexer except the descriptors will be in one portion; the descriptors will be in the other. A line or two at the bottom will be reserved for error messages, keyed input, and other communication matters.

A window on a terminal screen can be thought of as a view of a portion of a (possibly) longer text (Figure 6). The entire text is called a *frame.* Each window may be scrolled independently to view other portions of that text. By placing several windows on the screen at once, a great deal of control is available to the indexer.

The Citation Record Frame contains all fields thought to be of interest to the indexer (Figure 7). Three digit tag numbers are used instead of the one letter codes given in the menu, since they are more familiar to the indexer and they will no longer be keyed in. The indexer may move about within this frame without affecting any other window that may be visible on the screen. As fields are edited, the window display will change accordingly, and field contents may be verified automatically. Notice the three errors in this figure; they are typical of the types of errors that must be corrected by the indexers.

The Subject Descriptor Frame normally occupies the other window. Initially, the descriptors look like Figure 8. As each descriptor is entered, the program looks it up in the thesaurus [7]; if not there, an error comment is placed in the message area and the error is called to the indexer's attention. This can be by the terminal bell, blinking the descriptor display, etc. The indexer may also add M's and Q's; the display is automatically changed to correspond. The final result will look much like Figure 3 (but without the heading).

The indexer is expected to read through the Citation Record Frame, adding subject descriptors as they occur to him.

Several alternate frames can be invoked when desired, showing the thesaurus in different ways. In addition to the display seen earlier, the indexer may see all terms that begin with a specified stem (Figure 9), or all terms that include a given word (Figure 10). Some degree of stemming will be used in the latter case, to combine singular and plural forms, and other minor lexical variations. If the indexer sees a term in any of these displays that should be in the subject descriptors, he may move the cursor to any portion of that term, and press the proper key; the program will add the term to the descriptor list. If a touch panel be available, a mere touch will accomplish this.

A wide variety of commands is available, including traditional commands available in most full screen editors, and commands required for indexing. Many of the "traditional" commands have been altered, or specialized, somewhat for our purposes.

The cursor may be moved a single step up, down, right, or left. It may also be moved forward or backward in units of words, sentences, and sections. A *section* is meant to be about 2/3 of a window, and will be adjusted somewhat by context. The cursor may also be moved to the top or bottom of the frame. It is possible to search forwards or backwards for text in the frame that matches a typed-in string.

As an example of the specialization I mentioned, consider the citation frame shown in Figure 7. Cursor movement here is limited to the text of the different fields; if the cursor is at the end of a field, and a forward motion command is given, it will move to the beginning of the next field, bypassing the tag on the left. Whenever the cursor is moved out of a field which has been altered, verification of the new field contents will occur, if possible.

A number of commands are available for manipulating frames and windows. In all cases, the cursor is in one window, termed the *working window.* It is possible to scroll the other window up or down — this permits the title/abstract to be examined while entering descriptors,

without changing frames. The indexer may change from split screen to full screen, or vice versa. Full screen allows more information to be displayed, and may be particularly useful for examining the thesaurus. And, of course, the cursor may be moved to the other window, which then becomes the working window.

The usual operations for adding or deleting text are available. However, we prohibit the use of these for adding or deleting fields from the citation. Special operations are available for this purpose; we want to be sure the indexer really wants to do this. Occasionally, the indexer may wish to see all of the fields in the citation; an operation is available for this purpose.

There are about forty commands in all; more may be added if they should prove desirable.

5. New Hardware

New hardware is constantly being developed, and we (ie, both TIC and LBL) continually evaluate it. Two items have been found during the past year. The first of these is the Concept AVT terminal [8], from Human Design Systems, Philadelphia. This terminal includes four screen-loads of memory in the terminal, using direct memory-to-screen mapping. Consequently, several screen-loads of data can be moved directly to the terminal, and it can be used for local examination and editing. This should significantly improve response time for many operations, and we plan to make this our standard terminal.

The other new item is a touch panel [9] from Carroll Touch Technology, Champaign, Illinois. Unlike previous touch panels we have looked at, this one is sensitive enough to allow single characters to be picked out. We will be evaluating this with the goal of reducing much of the tedious part of cursor motion.

6. Status

At this time, the new editor is about 2/3 complete — that is, about 13,000 of the estimated 20,000 lines of code have been written. Much of the completed code was taken over nearly intact from the prototype editor, or from another project at LBL. We plan to have a version ready for testing by January, and the editor finished shortly thereafter.

7. Conclusion

We consider productivity in document indexing to be a combination of effectiveness, efficiency, and consistency. The new editor will, we believe, contribute to all of these aspects. The ease with which an indexer can change the display and the elimination of many of the inefficiencies of the current batch system can increase the amount of time spent on each citation, or decrease the amount of time spent on the indexing activity. Reducing the number of keystrokes contributes to the efficiency of the operation. Finally, computer analysis of indexing has the potential of comparing indexers, and contributing to consistency.

REFERENCES

1. Barbara A. Cerny and J. Dennis Lawrence. An Interactive Indexing-Editing System for DOE Technical Information Center. ASIS Mid-Year Meeting, Knoxville, TN, June 13-16, 1982. Published as LBL Report LBL-14019, Lawrence Berkeley Laboratory, Berkeley, CA.

2. Barbara A. Cerny, J. Dennis Lawrence, Todd Hammond, and Anna M. Okseniuk. Development of an Integrated Record Input System (IRIS) for DOE's Technical Information Center. DOE/TIC-11615, Technical Information Center, U. S. Dept. of Energy, Oak

Ridge, TN, November 1982.

3. J. Dennis Lawrence and Todd Hammond. User's Manual for the DOE Technical Information Center Integrated Record Input System (IRIS). DOE/TIC-11616, Technical Information Center, U. S. Dept. of Energy, Oak Ridge, TN, March 1983.
4. Craig A. Finseth. Theory and Practice of Text Editors, or A Cookbook for an EMACS. MIT/LCS/TM-165, Laboratory for Computer Science, Massachusetts Institute of Technology, Cambridge, MA, May 1980.
5. Richard M. Stallman. EMACS: The Extensible, Customizable, Self-Documenting Display Editor. AI Memo 519a, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, March 1981.
6. Richard M. Stallman. EMACS Manual for Its Users. AI Memo 554, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, October 1981.
7. DOE Energy Information Data Base: Subject Thesaurus. TID-7000, Technical Information Center, U. S. Dept. of Energy, Oak Ridge, TN, 1979.
8. Concept AVT User's Manual. Human Design Systems, Inc., Philadelphia, PA, 1983.
9. Carroll Touch Input System 117-0701-01 for DEC VT100 Terminal. Carroll Touch Technology Corp., Champaign, IL, March 1983.

Title: Use of Sharp Autoionized Resonances for Soft X-Ray Diagnosis

Enter a field/operation

t -----(primary title (A))
f -----(file selected for)
c -----(categories)
k -----(keywords)
a -----(abstract)

>-----(operators — subordinate menu)

here: □

Figure 1. Record Editing Menu

Sharp ionizing resonances (in this paper from He) used in conjunction with a spectrally bright XUV source of moderate power, e.g., laser produced plasma can greatly facilitate the coronal diagnosis of dense, high temperature, inertially confined fusion plasmas.

Figure 2. Example of an Abstract Display

*** Accepted Keywords ***

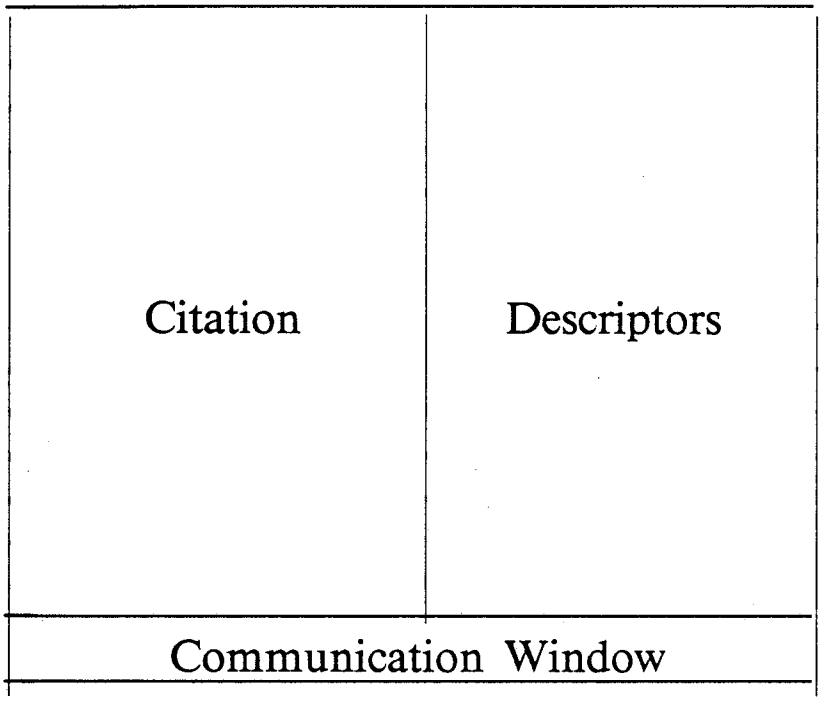
Ev Range 10-100
Inertial Confinement
Isotropy
Plasma Density

Laser-Produced Plasma:m1
    Interferometry:q1
    Soft X Radiation:q1
Hot Plasma:m2
    Interferometry:q2
    Soft X Radiation:q2
Laser Targets:m3
    Blackbody Radiation:q3
    Soft X Radiation:q3

Figure 3. Example of a Keywords Display

BROADER TERMS	RELATED TERMS	NARROWER TERMS
	Antinucleons	
	Brueckner Method	
	Charge Independence	Neutrons
	Effective Range Theory	Cold Neutrons
	Hard-Core Potential	Cosmic Neutrons
	Levinger-Bethe Theory	Epithermal Neutrons
	Massey-Mohr Equation	Fast Neutrons
	Nucleon-Deutron	Fission Neutrons
	Interactions	Intermediate Neutrons
		Photoneutrons
		Pile Neutrons
Fermions	Nucleons	Polyneutrons
Baryons	DA: 120174	Resonance Neutrons
Hadrons		Slow Neutrons
		Solar Neutrons
	Nucleon-Nucleon Potential	Thermal Neutrons
	Ope Potential	Photonucleons
	Pseudovector Coupling	Photoneutrons
	Rosenfeld Force	Protons
	Signell-Marchak Potential	Cosmic Protons
	Stapp Theory	Delayed Protons
		Diprotons

Figure 4. Example of a Thesaurus Display



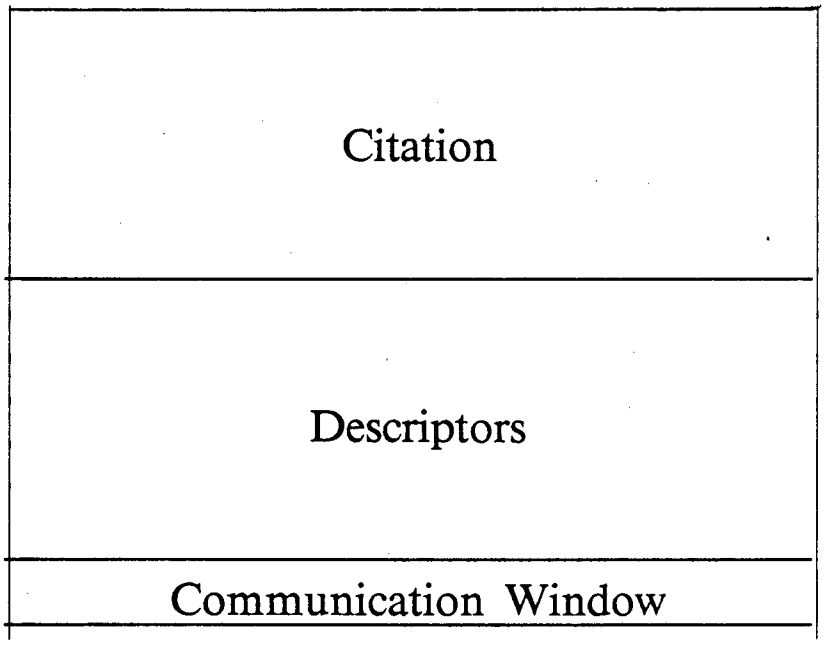


Figure 5. Display Window Formats

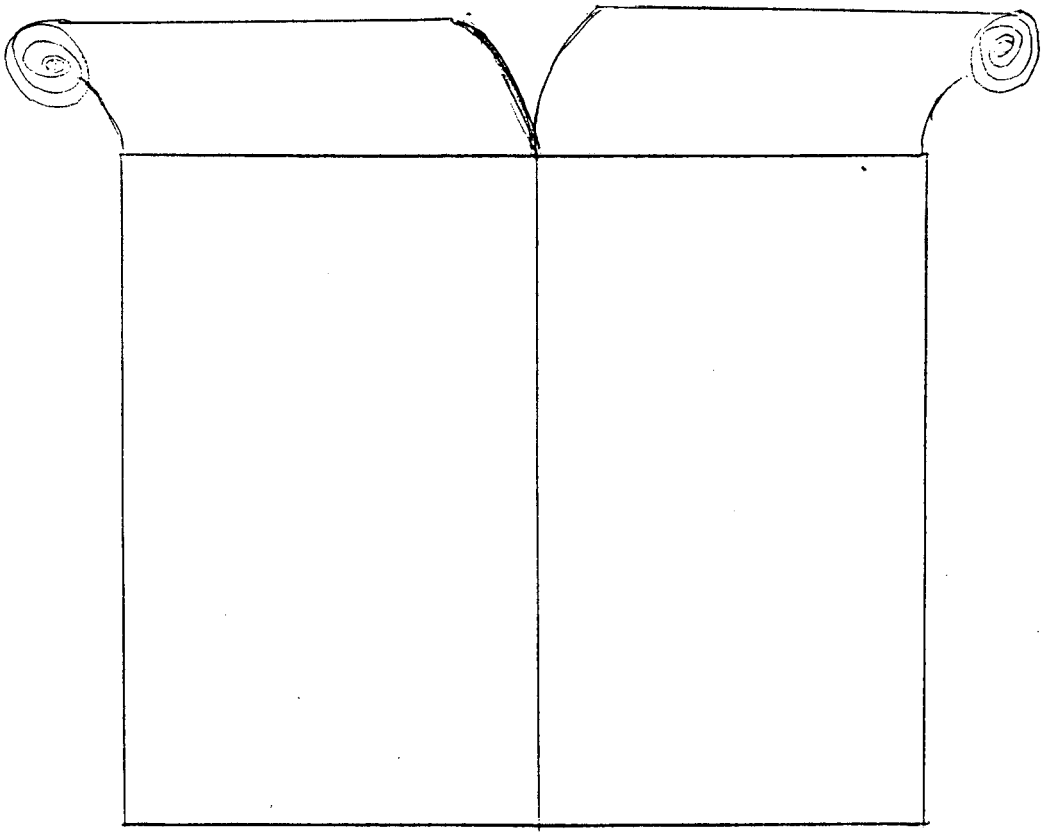


Figure 6. Scrolling

110 Physical bases of radiiology. Diagnosis and therapy.

200 Bases fisicas da radiologia. Diagnostico e terapia.

530 EDB

540 EDB—

950 The following topics are discussed in the book: basic notions of nuclear physics; X-rays; X-or gamma radiation absorption by the matter; ionizing radiation measurement; qualities of X-rays; measurement of the absorbed energy — dose; physical parameters of radiotherapy; isodose curves; irradiation fields; schemes of treatment; teletherapy machines; radiotherapy with radioisotopes; diagnostic diology and ionnizing radiation protection. Tables relative to the matter treated in the text are furnished.

Figure 7. The New Citation Record Display

Biomedical Radiography
Data Compilation
Diagnosis
Isodose Curves
Nuclear Medicine
Radiation Doses
Radiation Protection Laws
Radiology
Radiotherapy
Scintiscanning
X Radiation

Figure 8. An Initial Subject Descriptor Display

Electric Appliances
Electric Arcs
Electric Batteries
Electric Born Model
Electric Bridges
Electric Cables
Electric Charges
Electric Coils
Electric Condensers
Electric Conductivity
Electric Conductors
Electric Contractors
Electric Contacts
Electric Controllers
Electric Currents
Electric Dipole Moments
Electric Dipole Transitions
Electric Dipoles
Electric Discharge Pumping
Electric Discharges
Electric Fields
Electric Filters
Electric Furnaces
Electric Fuses
Electric Generators
Electric Grounds
Electric Heating
Electric Hexadecapole Transitions

Figure 9. Prefix List Display

Batteries (**Electric**)
Bridges (**Electric**)
Cables (**Electric**)
Chugoku **Electric** Power Company
Coils (**Electric**)
Condensers (**Electric**)
Conductivity (**Electric**)
Conductors (**Electric**)
Contacts (**Electric**)
Converters (**Electric**)
Currents (**Electric**)
Discharges (**Electric**)
Electric Appliances
Electric Arcs
Electric Batteries
Electric Born Model
Electric Bridges
Electric Cables
Electric Charges
Electric Coils
Electric Condensers
Electric Conductivity
Electric Conductors
Electric Contractors
Electric Contacts
Electric Controllers
Electric Currents
Electric Dipole Moments

Figure 10. Inverted Index Display

TECHNICAL INFORMATION DEPARTMENT
LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720