LEARNING BY EXPERIENCE WITH THE SCHOOL OF LIBRARIANSHIP URBAN TRANSPORTATION SYSTEM (SLUTS) (APPRENDRE PAR LA PRATIQUE A L'AIDE DU SCHOOL OF LIBRARIANSHIP URBAN TRANSPORTATION SYSTEM (SLUTS))

Anne B. Piternick and Peter Simmons
School of Librarianship
University of British Columbia
2075 Wesbrook Place
Vancouver, B.C. V6T 1W5

ABSTRACT

Students in the Documentation course at the UBC School of Librarianship learn the procedures and problems involved in creating a document storage and retrieval system through practical experience with the School of Librarianship Urban Transportation System (SLUTS). The course involves the creation of a machine-readable data base from sample document representations in the field of Urban Transportation. Students make the decisions on fields and subfields to be established for discription and searching; agree on standards for bibliographic description, and index the "documents". Searching is done on-line and also by means of KWIC and Uniterm indexes, which are prepared from document titles as by-products of the course. Records of time spent in indexing, consultation and supervision give students an idea of some of the costs involved in system development. Alternative approaches to problems of subject analysis, and design features of existing systems (which are treated in lectures and assignments) provide a context for SLUTS. (A l'Ecole de Biblioth-éeonomie de l'Université de la Colombie Britannique les étudiants du cours en documentation apprennent, par l'expérimentation du School of Librarianship Urban Transportation System (SLUTS), les méthodes et les problèmes qu'entrainent la création d'un système pour le stokage et la recherche de documents. Le cours comprend la création d'une banque de données provennant de l'analyse de documents modèles dans le domaine du transport urbain. Les étudiants choisissent les sujets et les sous-sujets qui seront employés pour fin de description et de recherche; ils adoptent des normes pour la description bibliographique et ils répertorient les "documents". La recherche est effectuée à l'aide de données à accès direct et aussi au moyen des indexes KWIC et UNITERM qui sont préparés d'après le titre du document - produit dérivé du cours. La compilation du temps encouru pour l'indexation, la

consultation et la surveillance donne aux étudiants un aperçu de quelques dépenses nécessaires lors de l'établissement d'un système. Des approaches variées aux problèmes de l'analyse documentaire ainsi que les principales caractéristiques des systèmes existants (qui sont traitées dans les lectures recommandées et les travaux) assurent la matière pour le système SLUTS.

INTRODUCTION [Apologia]

The School of Librarianship at the University of British Columbia offers as a second-year elective in the M.L.S. programme a course entitled "Documentation". Neither the title nor the course description offer much help in determining the content or the methodology of the course. Because we feel, however, that both content and methodology may be of interest to members of CAIS, we have undertaken to describe our system. In so doing, we have no intention of making a "How we do it good at UBC" presentation; indeed, we are conscious that there is much room for improvement in our approach, and we welcome comments and suggestions from members involved in designing or using similar systems in the field.

OBJECTIVES OF THE COURSE

The definitions of "Documentation" are many. The interpretation placed on this term for the purposes of our course was no doubt influenced by the fact that the instructors originally responsible for its design - Piternick, Simmons and Hagler* - had primary interests in the fields of reference and information services, automation, and classification and cataloguing respectively. The methodology of the course was influenced by the fact that almost all the students enrolled have background knowledge of the basic requirements for bibliographic control of specialist subject literatures; of the design of existing information storage and retrieval systems; of the capabilities of automated systems; and of the bibliographic and subject description of books.

As far as they can be formulated, our overall objectives in planning the course could be said to be to start the student thinking in a practical way about approaches to the design of reference retrieval systems for documents other than monographs. The "practical" component was felt to be desirable because of our experience of the demands placed on new graduates, on the UBC library staff, and on the library school faculty themselves, by members of the university faculty and research staff and by local firms, for assistance in creating specialized reference retrieval systems. It was possible, moreover, to consider a practical approach because of the students' previous grounding in the theoretical aspects of the principles involved in the design of such systems.

A FIRST APPROACH

Our first approach to Documentation was made during the Fall Term, 1971. Because of local library and campus interest in Transportation (the University had recently received a four-year, $360,000 grant for study and research in this field) this topic was chosen as the theme for the course. The students were assigned the task of describing an ideal system for the storage and retrieval of document surrogates relating to the field. Working through recommended reading, class discussion, consultation with faculty and with the UBC Library Staff in the

---

*Dr. Ronald Hagler is no longer responsible for teaching the course, although he is still active in a visiting and consultative capacity.

Science, Social Science, Fine Arts (including Planning) and Government Publications Divisions, they worked through the following stages:

a) Survey of the field, and categorization of the types of documents to be covered by the system;
b) Study of users and user preferences (based largely on information supplied by the library staff);
c) Examination of existing systems for the storage and retrieval of document surrogates (including abstracting and indexing services for Transportation, as well as systems such as MEDLARS, ERIC, etc.);
d) Study of problems relating to subject analysis and specification; problems of controlled and uncontrolled vocabularies, etc.;
e) Preparation of specifications for an ideal system.

Although this approach was problem-centred, it was insufficiently tangible to be successful. There were difficulties in imagining the system in action, and the absence of any attempt to assess costs increased the artificiality of the approach. The use of Transportation as a theme was felt to be good, however, since it was represented by a very wide range of document types - government publications; technical reports; theses; periodical articles; conference papers; feasibility studies; promotional brochures; chapters of books - and presented many subject facets - technical; economic; sociological, and so on.

SECOND APPROACH

Our second - and basically our present - approach to the course may be described as that of learning by experience. While retaining the framework of our previous attempt, we decided to confront the students with the problem of creating a system which would accommodate an actual collection of documents. Before term started, we were able to call on part-time help to create a collection of document representations, selected according to the previously determined document categories. These document representations consisted of xerographic copies of portions of documents sufficient to provide data for bibliographic description and subject analysis. For example, a document representation for a periodical article usually comprised the first page including abstract, and the citation; for a book, report, or proceedings, the title and contents pages and foreword if necessary; for a thesis, the title and contents pages and abstract or introduction, and so on. Occasionally, whole articles were copied, or complete duplicate copies of reports or government publications were used. The document representations were given an accession number - or SLUTS Document Number, as it came to be called.

A KWIC index was created to the collection at this time; the keypunchers worked from marked document representations, punching title and document number to provide input.

LEARNING BY EXPERIENCE

The School was fortunate to have access to an on-line information retrieval system, originally created by John Campbell (now of the UBC Library Systems Division) for the Poison Control Centre on campus. A description of the system is given in Appendix A. Although this was by no means a model system for reference retrieval, it nevertheless possessed the necessary flexibility for our purposes.

THE CREATION OF "SLUTS"

The students were now presented with the problem of creating SLUTS - the School of Librarianship Urban Transportation System - using the document representations as their source of input. Working through class discussion they covered the following stages:

1. Bibliographic description

   (i) Determine types of bibliographic elements which are useful for description of document;
   (ii) Determine types of bibliographic elements which are useful for retrieval of document;
   (iii) Categorize these elements according to Field and Subfield;
   (iv) Agree on standards for bibliographic form, citation (for periodical articles, chapters of books and conference papers), abbreviations, etc.

   From these deliberations, the SLUTS Rules for Descriptive Cataloguing and Standards for Forms and Abbreviations (Appendix B) were formulated. Finally, worksheets were prepared for input; this led to much consultation on unforeseen problems, revision and updating of rules, and correction, before the worksheets were finally checked and sent to keypunching.

2. Subject description

   The students were required to examine encyclopedias, guides to the literature, indexes, classification systems, and other sources for background information on the subject field encompassed by Transportation. A "subject profile" was created in class in which all aspects of the subject and their interrelationships were explored. This led to discussion of alternative approaches to classification of the field. The KWIC index was examined; from this an estimate was derived of the proportion of documents with titles useful for indexing in this way. The examination also provided graphic examples of the problems of uncontrolled vocabularies: separation of synonyms and analogous terms; juxtaposition of homographs; presentation of different forms and spellings of

the same word; splitting of compound terms; machine filing problems caused by hyphenation and variable punctuation practices, and so on.

As an adjunct to the KWIC index analysis, the students prepared a Uniterm card index to the collection using "useful" title words; for this they combined different forms of the same word or root, and accepted compound terms.

In a previous year's class we assigned to each student the construction of a "mini-thesaurus" in one area of the subject field, but this proved very time-consuming and difficult to handle in the time available. We abandoned this project last time the course was taught and simply asked the students to index their documents according to an established subject heading list. (Indexing subfields for the SUBJECT field are shown in Appendix B.) The subject indexing was then keypunched and the file updated.

3. Searching

Retrieval procedures for SLUTS are described briefly in Appendix A. For searching, each student was asked to provide three questions which could be answered from documents in his possession, and which could be found by searching the SLUTS fields and subfields, or terms within these subfields (e.g. words in title). These questions were next passed on to other students, who formulated search expressions using subject terms from the subject heading list, and geographic names, identifiers, and title terms from the Uniterm index. Each student spent approximately one hour searching the SLUTS file on-line; searches were then repeated using the Uniterm and KWIC indexes.

For most students, this was a first experience with an on-line system. It provided an excellent object lesson in system design, as problems typical of unsophisticated IR systems - strict punctuation requirements; rather lengthy procedures for combining search terms; lack of truncation - soon became apparent. The difficulties of searching using an inadequate vocabulary, albeit backed up by title terms, were also graphically illustrated.

4. Outputs

Outputs - specifically, the production of printed indexing and abstracting services - formed the final subject for decision-making. The students examined pairs of published abstracting and indexing services in the same subject area

with respect to coverage, time lag, presentation of information, access (subject, author, report number, etc.), type of indexing and ease of use, and cost to the subscriber of each abstract and each journal covered. We were able to compare costs with an estimate for the cost of SLUTS input because all students and faculty had been required to fill out timesheets during indexing. These timesheets were tabulated, and typical wage figures were applied to produce a hypothetical cost for producing the SLUTS data base. Finally, student preferences for format and design of existing indexes were applied to recommendations for a printed SLUTS abstracting service.

PROBLEMS

Although we feel that our approach to teaching Documentation is a good one, we still face several problems:

Subject description is unsatisfactory, partly because there is no useful thesaurus published for this field. We may solve this problem by moving to a new subject field - Water Resources, for example.

The SLUTS collection is very small, making comprehension of the subject difficult, and rendering the searching process rather unsatisfactory as so few documents will ever be retrieved by a search.

It is hard to estimate student load. (At least the timesheets permit some estimate of time spent in indexing.)

The time constraints involved in keypunching input make feedback on student indexing performance difficult.

Student participation in class discussion tends to be uneven. Inevitably, a few students become bored with the whole enterprise.

Student evaluation poses problems. One solution might be to treat this as a "pass-fail" course; another might be to apply contract grading.

ADVANTAGES

To weigh against these problems are the definite advantages of the practical aspects of SLUTS. The use of an actual document collection introduces an immediacy lacking in our previous approach, and allows much opportunity for decision-making. A high level of out-of-class discussion and consultation with instructors testifies to the interest and involvement of many of the students. Certainly the experience gained during the course makes it unlikely that any graduate will in future tackle the problem of creating an information retrieval system

without a full awareness of the work and costs involved, and of the variety of approaches to the problem.

This year has seen the introduction of an advanced course in Documentation which has enabled selected students to further explore problems raised in the basic course; we hope it will also lead to suggestions for improvements in SLUTS.

APPENDIX A
BRIEF SYSTEM DESCRIPTION

Ideally, specifications for an information retrieval package suitable for this course should have been laid out, and programs either found or produced to match these specifications. For reasons of expediency, however, a hurried search for existing programs was confined to the University, where only one suitable system was found.

Called "INFO, An Interactive Information Retrieval Program", the package permits input, modification, and output in either the batch or on-line mode. It permits an unlimited number of variable-length fields and sub-fields, free-form input with minimal special preparation, and a wide variety of optional output formats. Since the package with complete documentation was available, the decision was made to adopt this to our own uses; it has been used in this course since 1971 and despite its limitations, has served us well. It is hoped that in the near future we will be able to use a considerably more powerful data maintenance and report generating system currently being prepared for use in the U.B.C. Library by the Library staff.

FORMAT DESIGN AND INPUT

INFO requires that the names of all fields and subfields be declared before the data base is built. It is necessary, therefore, for the students to engage in the bibliographic format design described in the body of this paper early in the term. Following that, document descriptions are written on lined paper in the pattern: #T#D#, where # is a delimiter, T is the tag for the field, and D is the data in the field. Tags may be any size, and are chosen by the students. In fact, they are usually mnemonic contractions of the field name, e.g., A for author, T for title, PD for publication date. The data is input either in free form, as it appears on the document, or in coded form in the case of certain fields. The decisions regarding coded input are not imposed by the system, but are determined by the students with the sole criterion of facilitating rapid searching and output. Thus authors, for example, are always recorded in as complete a form as is available, while in the field used for form of the work, microforms are input as "MICRO", and phonograph records as "PHONO". Once the students have written (or, preferably, printed) the descriptive data on paper, those sheets are sent to be keypunched for batch input.

DATA MAINTENANCE

The SLUTS document number is input as a separate field and used as the control for the records. This permits the students to complete the bibliographic description early in the course, so that the worksheets can be forwarded to keypunchers for batch input while the students are working on the subject descriptions which occupy much of the term. The subject data is then keypunched and added to the file later, with

last-minute corrections and changes being added via the School of Librarianship's on-line terminal. The ease of on-line access to the file also permits the instructors to manipulate the file experimentally before the students have a chance to do so, a process which often indicates areas that may cause searching problems. For example, in 1973 a few test searches by the instructors revealed a punctuation problem. Though the INFO system ignores commas and periods during searches, it failed to retrieve a number of terms that appeared at the end of a field and were followed by question marks. It became necessary to either eliminate question marks or insert a space before each question mark. Through the use of a simple single-instruction program, the entire data base was changed on-line, solving the problem in a few seconds.

SEARCHING

Search procedures permit the use of the standard Boolean operators IF and AND, but not NOT. Students may search for any word or combination of words in any field or combination of fields; the searcher also has freedom to specify whether or not the terms he provides must appear in the order given, and whether any or all of the provided terms must appear in the field(s) searched. Each search question must constitute a "consider/doit" loop; that is, each search statement must first indicate which file is to be considered in this search and must end with the instruction to "DOIT". The CONSIDER instruction is required because the searcher has the ability to search all records in the SLUTS file, any section of the file he cares to define, or he may create subsets based on various criteria and then search one or more of those subsets. The DOIT permits the search strategy to occupy any number of lines and instructions, since no action will be taken until the final DOIT command.

The pattern of search commands is as follows: *CONSIDER ALL *B X.Y.Z *A *DOIT, where B is one of the Boolean operators, X is the data searched for, Y is the relator, Z is the data specified by the searcher, and A is the action to be taken. In practice, to save keying time CONSIDER is abbreviated to C, and DOIT to D. For example, here is a simple search command: *C ALL * IF AU.IS. SMITH, JOHN *PRINT DN *D. In this case the system is asked to consider all the records. If any records have SMITH, JOHN as one of the authors listed, the document number is to be printed. Another example: *C ALL(1,300) *IF PD. IS.1973 *AND S.CT.MONORAIL MONORAILS *ATTACH X *PRINT ALL *D. In this case, only the first 300 records in the file are to be considered. If any records have a publication date of 1973 and contain the word "monorail" or the word "monorails" in the subject field, those records are to be attached to the subset named X, and all of the data in the file on those records is to be printed. Subsets may be created for further searching or output at a later time; a record may exist in hundreds of subsets simultaneously without being withdrawn from the total file.

In the space available in this paper it is not possible to describe all of the possible relationships that may exist between the data

searched and the search key, nor to describe the system's many special features that are available to users. Those who wish more detailed information should write directly to the authors.

OUTPUT

INFO permits the organization of output into a variety of formats according to the user's specification. One or more fields may be output from any number of records, displayed across a page of any width, with or without headings, with or without spaces or other characters between records, with or without the field name preceding each field, and so on. As with input, output may be produced in batch mode or on-line. In normal student use, output is either displayed on CRT or printed out on a typewriter terminal, depending on which terminal the student is using for searching.

APPENDIX B
RULES FOR DESCRIPTIVE CATALOGUING, 1973

TITLE, T
ØRIGINAL TITLE, ØT
This subfield includes titles, subtitles, and edition statements, separated from each other by the symbol #. Titles in non-Roman characters to be transliterated (see SLUTS Standards).

TRANSLATED TITLE, TT
English translation of foreign titles and subtitles.

CITATION, C
Include here:
- Title, volume, pagination and date for serials (see SLUTS Standards for format). Title abbreviations only as permitted in SLUTS Standards.
- Series title and number.
- Citation and pagination for chapters of books, or for conference papers.

If document is a journal article which was presented at a conference, note this fact here. Include here information about documents which are superseded by the present document.

NAME, NA
AUTHOR, A
Enter all personal and corporate authors here, using AACR for form of entry. Designate editors, compilers and translators by using ED, COMP or TR after the name.

AFFILIATION, AF
Where document was derived from a thesis, include university where thesis was done if it differs from present AF; do this also if work described in document was done at a different place from present AF. Use AACR for form of entry.

LEARNING BY EXPERIENCE

SPONSOR, SP
Include here sponsors of reports and projects, and any other bodies in this category: contractors and granting agencies whose funds supported the work; sponsors of conferences and meetings, etc. Use AACR for form of entry.

PUBLISHER/PLACE, PR
See SLUTS Standards for abbreviations of placenames.

DATE, D
STUDY/SUBMISSION DATE, SD
If both SD and AD are given, prefer SD; give PD also. Give earliest SD if more than one.
ACCEPTANCE DATE, AD
PUBLICATION DATE, PD
If no date given, use "N.D." For all dates, use abbreviations of months as in SLUTS Standards; give year in full.

FORM, F
Acceptable terms and abbreviations (others to be added to this list as established):

MONOGRAPH, MONO	MICROFORM, MICRO
JOURNAL ARTICLE, JL	REVIEW, REV
REPORT, REPT	PHONORECORD, PHONO
THESIS	PAMPHLET, PAM
CONFERENCE PROCEEDINGS, CONF	MANUAL
CONFERENCE PAPER, CONF P	BOOK CHAPTER, CHAP

REPRINT (use only for reprinted monographs, and give PD for both original and reprint. Do not use for reprints of journal articles, etc.)

SPECIAL FEATURES, SF
Acceptable terms and abbreviations (others to be added to this list as established):

MAPS	GRAPHS
PLANS	CASE STUDIES
DIAGRAM(S), DIAGR(S)	ILLUSTRATION(S), ILLUS
TABLES, TABS	PHOTOGRAPH(S), PHOTO(S)
STATISTICS, STATS	INDEX
BIBLIOGRAPHY, BIBLIOG	QUESTIONNAIRE(S), QUEST
GLOSSARY, GLOSS	

NUMBER, NO, N
DOCUMENT NUMBER, DN
This is the SLUTS accession number.
CONTRACT NUMBER, CON
Includes Project Number, Grant Number.
REPORT NUMBER, RN
Use only if a distinctive RN with letters and numbers (e.g. AS-123 456); otherwise include in C as series number.

LEARNING BY EXPERIENCE

CLASS NUMBER, CN
(Note. This was originally intended for SLUTS Class Numbers, if we decided to use a classification scheme; however, this would be better in SUBJECT. Suggest this subfield be used instead for UBC call numbers if we ever decide to give location of documents indexed.)

LANGUAGE, L
See SLUTS Standards for abbreviations.

SUBJECT, S
S1 Level 1 terms (not more than three)
S2 Level 2 terms (unlimited)
S3 Geographic terms
S4 Identifiers (e.g., BART, AMTRAK, C.P.R.)

STANDARDS FOR FORMS AND ABBREVIATIONS, 1973

Ø - Use this form throughout on the worksheets for the letter "O".
ØT - Transliteration of titles: follow INIS Transliteration rules for selected non-Roman characters. (IAEA-INIS-10)
C - Format for periodical citations - follow H.W. Wilson, i.e. Title; volume no.; issue no.; pages; date - e.g. - [TITLE] V1 N2 P3-4 Ø69
- Periodical title abbreviations - do not abbreviate except

J	JOURNAL
Q	QUARTERLY
TRANS	TRANSACTIONS
RES	RESEARCH
SCI	SCIENCE

- Pagination for periodical titles, etc.:
  - give inclusive paging if available: Px-y
  - if starting page only known: Px-
  - if only one page: Px

NA - All corporate names to follow AACR.
- Allowable abbreviations: as above, and

UNIV	UNIVERSITY
DEPT	DEPARTMENT

- Placename abbreviations - country, state and province - as in AACR, except for G.B.
- Other placenames as in Webster's Geographical Dictionary.

L - Abbreviations of names of languages - as in ACCESS (except that we have changed "Engl" to "ENG".)