67th IFLA Council and General Conference
August 16-25, 2001

Teaching of subject access and retrieval at Mexican LIS Schools

Filiberto Felipe Martínez-Arellano

College of Library Science
National Autonomous University of Mexico
Mexico

Abstract:

Subject searching is one of the most important retrieval functions for obtaining useful information from catalogues. Subject searching in catalogues originally took the form of using a controlled language but OPACs have included new techniques like keyword searching. This fact means LIS students must learn about controlled and free-text languages' attributes to be successful in cataloguing and reference work. This paper shows how subject access has changed in online and Web-based catalogues; likewise how teaching of subject access and retrieval has been considered in LIS education, particularly in the curricula of Mexican LIS schools. Likewise, it proposes some actions to improve teaching of subject searching

No longer are library holdings just books, but different kinds of information-bearing entities (visual images, audio recordings, and digital objects, among others) have emerged and have been incorporated into the library collection. Likewise, many organizations and individuals are using the Internet for generating and delivering electronic information. The amount of electronic resources that are available on the Web has been substantially increased in recent years and there is an urgent need to include them in the library collection and, consequently, to include their surrogates in the library catalogue. Additionally, the catalogue's physical form has changed from cards to online and

Web-based catalogues. In the past, the catalogue was only accessible in the library but now it is accessible in many places outside as well. Library users can retrieve information not only about holdings in the local library, but also can examine holdings from other libraries. Moreover, modern catalogues offer a variety of options in terms of their access points. Card catalogues have given way to online and Web-based catalogues that incorporate new search options, particularly subject searches.

Subject access in library catalogues

In card catalogues, the only options for retrieving information about the holdings of a library were by author, title and subject. In comparision, online and Web-based catalogues make searches by the title words or words included in any other field as surrogates possible. In this way, the possibilities of subject access in online and Web-based catalogues are not limited to subject headings, a controlled language, but they are extended to keywords, mainly those from the title, which is the basic constituent of free-text. The use of these alternatives for searching and retrieving information on a particular topic is not something totally new. Indexes and abstracts covering published literature in several disciplines have used controlled and free-text vocabularies for many years. Likewise, many arguments related to the advantages and disadvantages of these two options have appeared in the LIS literature over the years (Lancaster, 1991). Additionally, free-text or natural language searches have been incorporated into search engines and portals in order to perform subject retrieval. With regard to this new retrieval tool, Thomas (2000) comments: "With the Web estimated to be increasing by 10 million pages weekly, the task of indexing Internet resources is clearly gargantuan, and not something that can be accomplished by even the most industrious beehive of catalogers. Instead of relying on the catalog to identify and retrieve relevant Web pages, users have turned instead to Web portals"

Portals have certainly became very popular search and navigation tools and some libraries are developing them. In this context, Thomas (2000) mentions: "Recognizing that some patrons may prefer to connect directly with online resources without being routed through the catalog, some libraries have developed separate gateways to networked resources. These gateways facilitate access to electronic materials selected by the library by providing a single point of entry, by organizing them into categories, and using metadata, often derived from their catalog records, to assist users in locating networked resources"

Portals are being developed by libraries and they have widespread acceptance by library users; however, like search engines, their subject retrieval features are mainly supported by free or natural language which has advantages but also disadvantages. Free language richness and variability lead, ironically, to ambiguity. The use of an unlimited number of words in indexing and searching processes leads to useless effort and a certain degree of error in information retrieval (Cleveland and Cleveland, 1991). Moreover, the use of free-text increases the amount of material recovered but decreases the precision of the retrieved information (Lancaster, 1986, 1991).

Unlike free-text, a controlled vocabulary has more advantages. Gerhan (cited by Peters, 1991) found that catalogue users retrieved more records in fewer attempts making use of the Library of Congress Subject Headings. Additionally, when testing a catalogue that contained a considerable quantity of material in a language different from the one used for the actual indexing, Martínez-Arellano (1999) discovered that a great deal of material was missed. Referring to the importance of a controlled vocabulary, Tillet (2000) points out : "Authority control enables "precision and recall," which are lacking from today's Web searches. Authority control provides precision to

retrieve only those records or items of interest, and the syndetic structure of authority control's cross references assures recall of all the relevant materials, as well as navigation to reach bibliographically related materials. It cannot be stressed enough that this feature of online catalogues adds tremendous value to the user's search and retrieval process. No more wading through tens of thousands of retrieved and computer ranked results for anything close to what we asked for, unless we want to. Let's give users the option for more precise searching, if they want it"

In a networked environment, libraries no longer own most electronic resources; they just have access to them. In this situation, the physical description of the electronic resources plays a secondary role to their content description. A correct description of those topics dealing with an information-bearing entity would allow information users to evaluate surrogates in order to decide which electronic resources are more convenient to access. Library of Congress Subject Headings has been recomended as an important tool that can be used in a networked environment to retrieve electronic resources by subject. Among the reasons for this recommendation, Chan (2000) mentions the following: "(1) LCSH is a rich vocabulary covering all subject areas, easily the largest general indexing vocabulary in the English language; (2) there is synonym and homograph control; (3) it contains rich links (cross references indicating relationships) among terms; (4) it is a pre-coordinate system that ensures precision in retrieval; (5) it facilitates browsing of multiple-concept or multi-faceted subjects; and, (6) having been translated or adapted as a model for developing subject headings systems by many countries around the world, LCSH is a de facto universal controlled vocabulary."

All the above arguments show the importance of controlled vocabularies for subject retrieval in a networked environment. "The reason is that effective cataloguing involves controlled vocabularies and adherence to the standards that have evolved in the past 100 years" (Gorman, 2000). Since controlled vocabularies are important subject retrieval tools, then, it is also important to look at how LIS schools are educating librarians for managing "traditional" cataloguing standards, particularly subject retrieval and controlled languages. Only if LIS students learn the basic principles can they manage electronic resources organization and retrieval adequately.

Information organizacion and subject access teaching

Subject access and retrieval have been integrated into most of the coursework at LIS schools. In a survey carried out among the Masters programs accredited by the American Library Association in the United States and Canada, Cortés-Arriaga found 450 courses dealing with the matter. The titles of those courses included information retrieval, online searching, search strategies, information storage and retrieval, online bibliographic services, organization of information, indexing, information organization, database access, online information retrieval. Many courses dealing with subject access and retrieval were in the service area. Courses related to technical processes or information organization were fewer. Also, there were few courses whose title and content were focused specifically on subject access. Such courses include: Subject analysis and thesaurus construction (The University of Western Ontario), Indexing and surrogation (State University of New York at Buffalo) Indexing (University of Wisconsin-Madison), Text Retrieval, Web Search Engines (Iowa University), Abstracting and Indexing, Thesaurus design and construction (Pratt Institute).

Although subject access has been integrated to some courses at LIS schools, its inclusion has been mainly in those courses dealing with the retrieval process. The importance given to the storage process has declined, cataloguing courses have decreased in the last years. A study by Spillane

(cited by Weiss and Carstens, 2001) showed that in 1998 only 52.1% of library schools required a cataloguing course, compared to 77.1% in 1986. Nevertheless, giving LIS students the basics of cataloguing and subject access will prepare them to help users to find the information they are seeking. A study carried out by Banks (cited by Weiss and Carstens, 2001) showed how learning the rudiments of cataloguing assists users in locating materials in the catalogue. Regarding subject access teaching, Romero (1995) has recommended: "educators should ensure that students not only understand how to use the various subject headings tools, so that headings are valid, but should also emphasize how to analyze the subject content of the work so that appropriate and valid headings can be assigned."

Additionally, this author mentions: "when instructing students in subject cataloging and classification, educators need to stress the overall goal of subject analysis both in application and in access. Educators should ensure that in addition to knowledge of how to use the tools students should also acquire the ability needed to assess the subject content of an item so that retrieval does not become an exercise in second-guessing." All these arguments show the necessity to emphasize teaching the organization of information, particularly those aspects dealing subject content analysis of information-bearing entities to get accurate retrieval.

Information organization and subject access teaching at Mexican LIS schools

At the present time, there are six LIS schools at Mexico that prepare professionals in this discipline. These schools give education at the undergraduate level and their main objective is to train professionals to select, acquire, catalogue, classify and deliver information to all kinds of users. The first Mexican LIS schools were founded approximately fifty years ago, the National School of Library and Archive Sciences in 1945, and the College of Library Science of the National Autonomous University of Mexico in 1956. Both were established in the country's capital. There are four other LIS schools in different Mexican universities: School of Library and Information Science of the Autonomous University of San Luis Potosi, College of Library Science of the Autonomous University of Nuevo Leon, Documental Information Science Program of the Autonomus University of Mexico State and Library Science Program of the Autonomous University of Chiapas (Martínez-Arellano, 2000)

These LIS schools and programmes have their own characteristics and curricula; however, all of them include an area that comprises courses dealing with organization of information (Table 1). Looking at the titles of those courses, it seems that subject access learning is not included. In order to know what courses were related to subject access, a questionnaire about the course objectives covering the organization of information was circulated. Four LIS schools answered this request and the results were as follows:

National School of Library and Archive Sciences

Indexing I To understand the process of documental analysis, foundations of languages and their different types
Indexing II To analyze the concept, origin and functions of subject headings lists and thesaurus , as well as the terminology control.

College of Library Science (National Autonomous University of Mexico)

Cataloguing and classification II-2

To know the content representation theory of documents using controlled vocabularies (subject headings and thesaurus). To use controlled vocabularies for documental organization.

School of Library and Information Science (Autonomous University of San Luis Potosi)

Indexing and languages for information searching.
Students will be able to create or develop an index, as well as apply the accessible search languages and procedures to retrieve information according to established systems. The tools used in information storage and retrieval of a document, a database, or a databank should be part of the student knowledge.

Library Science Program (Autonomous University of Chiapas)

Information analysis and representation I-II (subject content)
At the end of the course, students will be able to know and to analyze the subject content of bibliographic materials. Moreover, they will learn the use, management and application of subject headings lists.
Abstracts: To guide students to understand the nature and diversity of different types of abstracts that can be used in information services
Thesaurus: To guide students to understand the foundations of thesaurus and its use for the content description, as well as to learn the possibilities that these tools can offer information services

Final considerations

The catalogue's physical form has changed and now it is normal to use online and Web-based catalogues. But not only the catalogue's physical form has changed, new catalogues have increased access points and they have incorporated new search options, particularly subject searches. In addition to searches by subject headings or controlled terms, online and Web-based catalogues make searches by the title words or words included in any other field from their surrogates possible. These two ways of searching and retrieving information about a particular topic have advantages and disadvantages. The use of free-text in the searching process leads to the retrieval of a great amount of information but its precision decreases. In addition to printed resources, authors and organizations are using electronic resources for delivering knowledge and information, and new tools for organizing these types of resources have emerged. Portals, gateways and search engines are being used to organize and find electronic resources but the subject retrieval features of these tools are mainly supported by free-text. The result is that sometimes a user gets a great amount of information, but most of it is irrelevant. In order to solve this problem, the use of controlled languages and tools like the Library of Congrees Subject Headings has been suggested.

Although use of free-text is an easy and cheap option for indexing electronic resources, libraries and other information agencies will have to use controlled vocabularies for the storage and retrieval of the precise information that matches user needs. In this way, controlled languages will continue to play an important role in the organization of electronic and networked resources, so knowledge of them will need to be strengthened. LIS schools will have to increase the number of courses whose main focus is subject analysis for the storage and retrieval of information in catalogues and other organization of information tools. Likewise, in these courses, the importance and advantages of controlled languages must play an important role. Traditionally, Mexican LIS schools have given great importance to organization of information courses; however, most of the courses dealing with this area are centred to descriptive cataloging of materials. It is necessary to include in

the LIS curricula more courses whose main objective is to increase knowledge of the subject content and languages used for the storage and retrieval of information in catalogues and other tools. In this way, librarians will be adequately prepared to face the challenges that technology and the new types of information resources impel.

References

Cortés-Arriaga, M. (2000). *Influencia de las tecnologías de la informaci6n en los programas de estudio de las escuelas de bibliotecología de Canadá y Estados Unidos* (Information technologies influence on study programs of Library Science schools from Canada and the United States) (Master in Library Science Dissertation, National Autonomous University of Mexico).

Chan, L. M. (2000). *Exploiting LCSH, LCC, and DDC to retrieve networked resources: issues and challenges.* http://lcweb.loc.gov/catdir/bibcontrol/chan.html

Cleveland, D. B. and Cleveland A. D. (1990). *Introduction to the indexing and abstracting.* Englewood, CO: Libraries Unlimited.

Gorman, M. (2000). *From card catalogues to WebPACS: celebrating cataloguing in the 20th century*. http://lcweb.loc.gov/catdir/bibcontrol/gorman.html

Lancaster, F. W. (1986). *Vocabulary control for information retrieval*. Arlington, Va: Information Resources

Lancaster, F. W. (1991). *Indexing and abstracting in theory and practice.* Champaign, IL: University of Illinois. Graduate School of Library and Information Science.

Martínez-Arellano, F. F. (2000). 'Library Science education in Mexico'. *Journal of Education for Library and Information Science*. 41(2), 158-161.

Martínez-Arellano, F. F. (1999). 'Subject searching in online catalogs including Spanish and English materials'. *Cataloging & Classification Quarterly* 28(2), 45-56.

Peters, A. T. (1991). *The online catalog: a critical examination of public use*. Jefferson, NC: McFarland.

Romero, L. (1996). 'An evaluation of classification and subject cataloging in entry-level cataloging copy: implications for access and instruction'. *Journal of Education for Library and Information Science*. 36(3), 217-229.

Thomas, S. E. (2000). The catalog as portal to the Internet.
http://lcweb.loc.gov/catdir/bibcontrol/thomas.html

Weiss, A. & Carstens, T. V. (2991). 'The year's in cataloging 1999'. *Library Resources & Technical Services*. 45(1), 47-58

TABLE 1. COURSES DEALING WITH ORGANIZATION OF INFORMATION AT MEXICAN LIS SCHOOLS

National School of Library and Archive Sciences

    Area of technical organization
        Foundations of technical organization for documental materials
        Library classification systems
        Cataloging codes I
        Cataloging codes II
        Bibliography
        Indexing I
        Indexing II
        Dewey Classification System
        Library of Congress Classification System
        Catalogs organization

College of Library Science (National Autonomous University of Mexico)

    Area of technical services
        Cataloging and classification I-1 and I-2
        Cataloging and classification II-1 y II-2
        Cataloging and classification III-1 y III-2
        Cataloging and classification IV-1 y IV-2
        Technical services
        Resources selection
        Introduction to data processing 1 and 2 (Elective)
        Statistics applied to education 1 and 2 (Elective)
        Book preservation and restoration 1 and 2 (Elective)

School of Library and Information Science (Autonomous University of San Luis Potosi)

    Information analysis and organization area
        Cataloging I, II, II, IV, V
        Classification I, II, III
        Preservation and restoration workshop
        Indexing and languages for information searching
        Scientific papers writing
        Information processing workshop
        Bibliography I, II
        Mexican bibliography

College of Library Science (Autonous University of Nuevo Leon)

    Bibliographic organization area
        Classification and cataloging I, II, III, IV, V, VI
        Document selection and acquisition
        Serials and periodicals

Introduction to data processing I, II

Documental Information Science Program (Autonomous University of Mexico State)
<u>Information theory</u>
Document carries history
Resources and collections development I
Resources and collections development I
Cataloging I
Cataloging II
Introduction to documentary information science
Documentary information units
Classification systems I
Classification systems II
Classification systems III
Political and administrative system of Mexico
Document legislation
Document administration
Document description I
Document description II

Library Science Program (Autonomus University of Chiapas)

<u>Area of documentary classification systems</u>
Knowledge classification systems
Library classification systems
Information analysis and representation I (description)
Information analysis and representation I II (subject content)
Resources classification
Catalogues classification
Abstracts
Thesaurus