Subject representation

6.1.2. Subject description processes

Overview

Four major processes or areas of practice for representing subjects are classification, subject cataloging, indexing, and abstracting. The processes vary in their origins, purposes and results, not all of which are directly related to libraries. For this reason, they also differ in specific techniques, conventions, and (unfortunately) terminology.

The library and information science (LIS) literature tends to focus on only one or two of the processes at a time, which can be confusing for the reader. Nowhere does there seem to be balanced coverage of all four processes or an explicit explanation of their differences. This module is intended to make those distinctions.

Functions of subject representations

LIS views subject representations as secondary information objects that describe the intellectual content (aboutness) of primary information objects (documents and queries).

Although queries certainly have subjects, the emphasis here is on documents: the objects represented in information retrieval system records. The term document is used below instead of object to avoid the awkwardness of saying "the subject of the object."

Subject descriptions serve to . . .

- Support subject access to documents of all kinds
- Inform searchers about intellectual content of the documents
- Provide consistency of representations

They also serve to . . .

- Organize shelving for physical browsing and retrieval
- Assist in collection development and acquisitions
- Assist in collection maintenance
The objectives of subject descriptions are to . . .

- Provide access by subject to all relevant material
- Provide subject access to materials through all suitable principles of subject organization
- Show various kinds of subject concepts, based on similarities of matter studied, method, point of view, or application of knowledge
- Provide entry to any subject field at any conceptual level from the most general to the most specific
- Bring together references to materials that treat substantially the same subject regardless of disparities in terminology
- Provide entry through the vocabulary understood by particular groups of users
- Provide a formal description of the subject content of any document in the most accurate and precise terms possible, whether the description is in the form of a word or brief phrase or in the form of a classification number or symbol
- Provide a means for the user to select from among all documents in any particular category, according to any chosen set of criteria, such as most thorough, most current, or most elementary.

Processes and products of subject description

Each of the four processes—classification, subject cataloging, indexing, and abstracting—differs from the others in several ways:

- **Purpose**: The type of intellectual access provided and whether physical access is also provided
- **Entity level**: The level of document content described, from whole document to parts of document
- **Surrogate**: The product or result of the process, consisting of one or more terms or codes that serve as access points and surrogates for the document. Surrogates vary in the
  - Number of terms and symbols used
  - Precision and specificity of terms and symbols
  - Internal organization of terms and symbols
- **Language**: The type of vocabulary or code used to create the surrogate
- **Source**: The source of the language for creating the surrogate. (Major examples only are shown below: many more sources of terms and codes exist.)

The differences in the processes are detailed below and summarized in a table at the end of this module. Note that these processes originated in a variety of information settings, each with its own terminology. If you understand the distinctions below, you will find it much easier to understand the LIS literature!
Classification

Classification is probably the oldest formal method for organizing information objects, dating back to the earliest known library collections of clay tablets. Its original—and continuing—purpose is for physical access. Its product, a code called notation, is part of the familiar library call number. This very compact surrogate is what the user scans to find a document on a shelf. It also collocates documents on similar subjects and therefore facilitates intellectual access through browsing. The limitation of classification is that a document can be shelved in only one place, even if it has more than one subject. The concept of physical access is evolving with the need to classify electronic information objects.

- **Purpose**: Physical access, for locating documents on the shelf; and intellectual access through collocation of documents on the same subject, relative to rest of collection
- **Entity level**: Subject of whole document; one subject only for shelf location
- **Surrogate**: Notation (one only), a combination of class codes; usually first part of document call number
- **Language**: Classification codes
- **Source**: Library of Congress Classification (LCC), Dewey Decimal Classification (DDC), others

Subject Cataloging

Subject cataloging, also based in library tradition, grew from the desire for a representation containing more subject information than a classification code. Catalogs began as printed lists of scrolls or books and, as publishing increased and libraries began to grow at a faster rate, evolved into card catalogs with far greater flexibility. A catalog provides intellectual access as an alternative to physical access: it identifies subjects that are not evident from physical location. Online catalogs provide even more access because the user can search subject description and physical description fields (e.g., subject and author) simultaneously.

- **Purpose**: Intellectual access; identifies multiple subjects and relationships among subjects
- **Entity level**: Subject(s) of whole document
- **Surrogate**: Subject heading; usually several terms combined in prescribed order; about one to three subject headings per catalog record
- **Language**: Controlled vocabulary
- **Source**: Library of Congress Subject Headings (LCSH), Sears List of Subject Headings, others

Indexing

Indexing is a more recent development than cataloging. It grew, in part, out of the desire of the scholarly, scientific, and technical communities to describe individual journal articles instead of the whole-journal entity level of catalogs. Indexes in their many forms, including back-of-the-book indexes, use multiple terms to describe a variety of topics and subtopics.

- **Purpose**: Intellectual access; identifies multiple subjects and relationships among subjects
- **Entity level**: Subject(s) of whole document and/or internal subtopics
Abstracting

Abstracting developed along with indexing. Next to full text, abstracts are the most complete representation of intellectual content and often the most helpful to the user. Because they are written in natural language, they can indicate overall purpose and direction of a work as well as complex relationships among subject concepts.

- **Purpose:** Intellectual access; identifies complex internal concepts and relationships
- **Entity level:** Subject(s) of whole document plus internal subtopics
- **Surrogate:** Content synopsis; usually one at beginning of document and/or in an index to literature in a discipline
- **Language:** Natural language
- **Source:** Text of document, although thesaurus may be helpful in finding appropriate terms for the discipline

All four processes are now used to describe information objects in libraries and other settings, such as the World Wide Web. In many settings, the processes are also used partially or exclusively for physical description. They are presented here as subject description, however, because subject description is particularly challenging and a major basis for their inception.

### Relevance to the course

This course covers classification, subject cataloging, and indexing but does not cover abstracting. In library practice, classification is commonly subsumed under cataloging, where cataloging consists of three tasks: descriptive cataloging (physical description), subject cataloging, and classification. Libraries typically do cataloging and classification, publishers do indexing and cataloging, and bibliographic utilities do both.

The vocabularies (sets of terms) used for subject cataloging and indexing are generically known as indexing languages, as described in other modules. Classification is addressed last in a separate set of modules, which is appropriate because it is often the last subject description task to be completed.

**Project Alert!** For the IOP, you practice two of the processes above: indexing and classification. The IOP results in a catalog (the Inmagic database), but the approach to controlled vocabulary is taken from indexing practice (multiple discrete subject terms) instead of subject cataloging practice (combined terms in subject headings). This is not unusual and may in fact be a trend, especially with information objects in electronic formats.
# SLIS 5200

## Subject Description Processes and Products

<table>
<thead>
<tr>
<th>PROCESS</th>
<th>PURPOSE (kind of access provided)</th>
<th>ENTITY LEVEL (level of document described)</th>
<th>SURROGATE (form of document representation)</th>
<th>LANGUAGE (type of word or code used for surrogate)</th>
<th>SOURCE (example of content resource)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLASSIFICATION</td>
<td>Intellectual access; traditionally also physical access, for collocation (shelving) of same-subject documents</td>
<td>Subject of whole document; one subject only for shelf location</td>
<td>Notation: combination of class codes, often first part of document call number; one notation only per document</td>
<td>Class codes</td>
<td>Dewey Decimal Classification (DDC) or Library of Congress Classification (LCC) scheme</td>
</tr>
<tr>
<td>SUBJECT CATALOGING</td>
<td>Intellectual access, as alternative to physical access; identifies subjects not evident in shelving</td>
<td>Subject of whole document; multiple subjects for interdisciplinary documents</td>
<td>Subject heading: one or more terms; often several subject headings per document in catalog record</td>
<td>Controlled vocabulary</td>
<td>Library of Congress Subject Headings (LCSH) list</td>
</tr>
<tr>
<td>INDEXING</td>
<td>Intellectual access, as alternative to physical access; identifies conceptual relationships among subjects</td>
<td>Subject(s) of whole document and/or parts of document (subtopics)</td>
<td>Index term or descriptor: one or more terms; usually many descriptors per document in subject-area index and/or back-of-the-book index</td>
<td>Controlled vocabulary or natural language</td>
<td>ERIC Thesaurus (controlled vocabulary) or text of document (natural language)</td>
</tr>
<tr>
<td>ABSTRACTING</td>
<td>Intellectual access only; identifies complex concepts and relationships</td>
<td>Subject(s) of whole document and parts of document (subtopics)</td>
<td>Content synopsis: sentences or phrases; one abstract per document in subject-area index and/or beginning of document</td>
<td>Natural language</td>
<td>Text of document; thesaurus may also be helpful</td>
</tr>
</tbody>
</table>

**Note:** Document includes any format of information object (e.g., text, image, sound). Term is one or more words describing a single subject.