6.1.3. Subject analysis

Overview

This module focuses on the what, why, when, how, and who of subject analysis, or determining what an information object is about. All four subject description processes (classification, subject cataloging, indexing, and abstracting) as well as searching depend on subject analysis. This is a vital skill area for information professionals. Expert subject analysis requires high levels of verbal aptitude and abstract thinking skills.

What is subject analysis?

Subject analysis (also called verbal subject analysis) is the task of determining the intellectual content or aboutness of an information object. But this isn't just the documents in the collection. If you recall the basic information retrieval (IR) model, two kinds of information go into an IR system: representations of information objects (documents) and representations of information need (queries). When these representations match exactly, the system retrieves records containing the data of interest. In order for the concepts and terms to match, subject analysis actually involves both kinds of objects:

- **Document analysis** is studying a document to determine how to represent it in a record, or what indexing terms or codes to enter.
- **Query analysis** is studying an information request to determine how to formulate a search query, or how to choose appropriate search terms.

Why do subject analysis?

At various stages throughout IR, subject analysis serves to:

- Discover or express the subjects of documents and queries
- Clarify and organize subjects of documents and queries
- Express subjects precisely
- Achieve consistency between document and search terms
When is subject analysis done?

Subject analysis is performed . . .

- During production of the primary document:
  - Author’s abstract and/or index
  - Indexing commissioned by publisher
  - Cataloging in publication (CIP) before printing
- Prior to storage of the document for retrieval:
  - Cataloging or indexing by bibliographic utility
  - Cataloging or indexing by individual library
- During information retrieval:
  - Problem statement or question from user
  - Query formulation by intermediary or user

And more—subject analysis takes place any time subject representation is at stake, including in the development of subject authority files.

**Project Alert!** You will be performing subject analysis when you:
- Determine subject elements/fields
- Develop thesaurus
- Develop classification scheme
- Create records

How does one do subject analysis?

Subject analysis involves four steps:

1. **Familiarization:** Becoming acquainted with general content of document and query
2. **Extraction:** Identifying and extracting significant concepts and natural-language terms
3. **Translation:** Converting extracted terms into controlled vocabulary of system
4. **Formalization:** Applying rules for exact format, spelling, punctuation, codes, etc. for input to system

The steps do not necessarily occur in this order: subject analysis requires evaluation and verification at every stage in a continuous, iterative cycle. Taylor (2004, chapter 9) subsumes steps 1 and 2 under **conceptual analysis**, or determining aboutness, and steps 3 and 4 under **subject analysis**, or translating concepts into system terms. Although controlled-vocabulary systems are assumed above, steps 1 and 2 are also applied in natural-language systems. Regardless, the same general considerations come into play:

**Locating subject concepts**

In some cases, subject concepts can be easily found in a title, but in other cases they are buried in text and can be found only by scanning a book in its entirety. The subject analyst develops a method for looking through a book that may begin with the title and progress through locations such as CIP data, table of contents, introduction, and index. But locations of subject concepts vary in books and other types of information objects. Subjects of audio or video objects may be evident from package labels or may require playing the entire piece. Nontextual objects such as artworks may have no obvious clues to subjects or may have a variety of captions, labels, inscriptions, or accompanying materials.
Identifying subject concepts

Subject concepts may be easy or difficult to identify, depending on the questions one asks and the nature of the concepts. Subject analysis is a balancing act between literary warrant and user warrant:

- **Literary warrant** pertains to information objects, both the individual object and the collection as a whole. It comes into play when the indexer/cataloger examines a document to determine what it is about. It addresses questions such as: What are the dominant themes and terms? and What is the author's intent?

- **User warrant** pertains to users' needs and abilities. It requires the subject analyst to understand and predict user behavior. Questions include: What do users want and expect? and What is the extent of their domain knowledge and language skills?

Choosing subject terms

Once concepts and terms are extracted from a document, they must be translated into the vocabulary of the system. Not only must the terms accurately describe the concepts, but they must also be appropriate in conceptual level and number. Questions include: Should the terms be broad or narrow? and Should one term or many terms be assigned?

**Project Alert!**
- For literary warrant, examine objects in your sample collection to determine the locations of subject concepts. State these locations as the chief source of information in your input rules.
- For user warrant, draw on your description of users and their questions in section 1.
- Decisions about the level and number of terms refer to the indexing criteria of specificity and exhaustivity. See the module on thesaurus development.

Who does subject analysis?

The people who perform subject analysis include authors, publishers, catalogers, indexers, abstracters, classifiers, reviewers, and—of course—the users searching for the information. The intellectual characteristics of professional subject analysts include:

- Abstract, holistic thinking
- Flair for analysis and intellectual rigor
- Critical skills and good judgment
- Expert language skills
- Precise, orderly and systematic mind

Does this sound like your line of work?!

Cites & sites