
Library of Congress Subject Headings

Module 1.3:
What is Subject Analysis?

Policy and Standards Division
Library of Congress
June 2016

Subject Analysis

- “Before we can use any classification scheme or make any index entry for a document, we must be quite clear what the document is about.” – Derek W. Langridge in *Classification: Its Kinds, Systems, Elements, and Applications*

In this unit, we’re discussing subject analysis. I would like to start with a passage from one of my favorite authors in library and information science: D.W. Langridge (of the Classification Research Group). This quote is from his book *Classification: Its Kinds, Systems, Elements, and Applications* (on page 64, if you are interested).


“Before we can use *any* classification scheme or make any index entry for a document, we must be quite clear what the document is about. Unless we start with a correct subject analysis, it is obviously quite impossible to proceed to a correct class number or index entry...

“It is not quite as simple and straightforward to analyze the subject of a document as to classify an object...



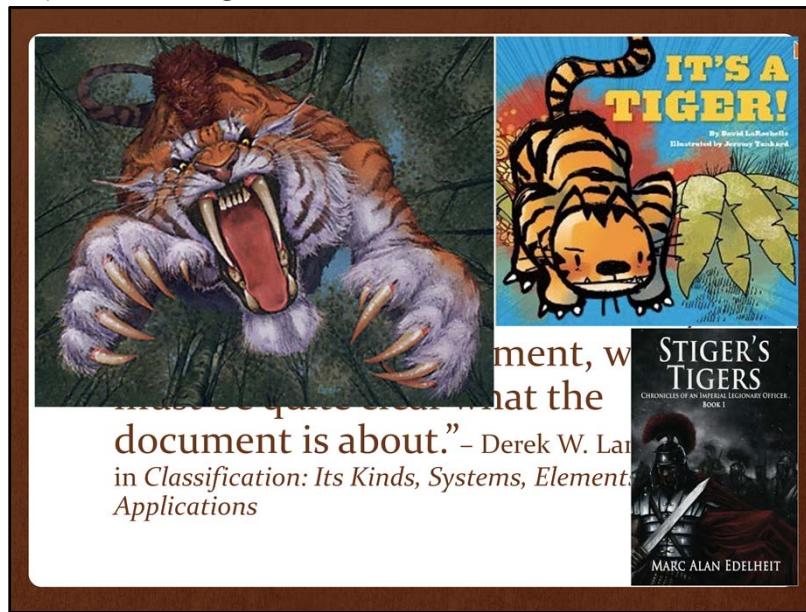
...or make any
...ment, we
...at the
document is about.” – Derek W. Langridge
in *Classification: Its Kinds, Systems, Elements, and
Applications*

“If I come face-to-face with a striped animal in the jungle, I do not have much difficulty in saying,
‘This is a tiger’ ...



ment, we
that the
document is about.” – Derek W. Langridge
in *Classification: Its Kinds, Systems, Elements, and Applications*

“But books on tigers are not striped, and they do not growl or spring at us...



“They consist of several thousand words between two covers with a title that may or may not include the word ‘tiger’. Subject analysis consists in reducing these thousands of words to a very precise statement of the subject scope of the book.”

What is Subject Analysis?

- It is the process of:
 - Examining a resource
 - Determining that resource's *aboutness*
 - Using its *aboutness* to create metadata

So, in order to provide subject access to information materials, we perform an activity known as “subject analysis”.

Subject analysis is the process of examining a resource and figuring out what that resource is, and what it is about.

This is a process of conceptual analysis.

Once we have finished that, we use the “aboutness” to create metadata to reflect the subject matter and genre/form properties of a resource. In effect, we translate the “aboutness” into controlled vocabulary terms and classification notations.

Why Do We Perform Subject Analysis?

- To provide subject access
- To collocate resources
- To provide a logical location for resources
- To save the users' time.

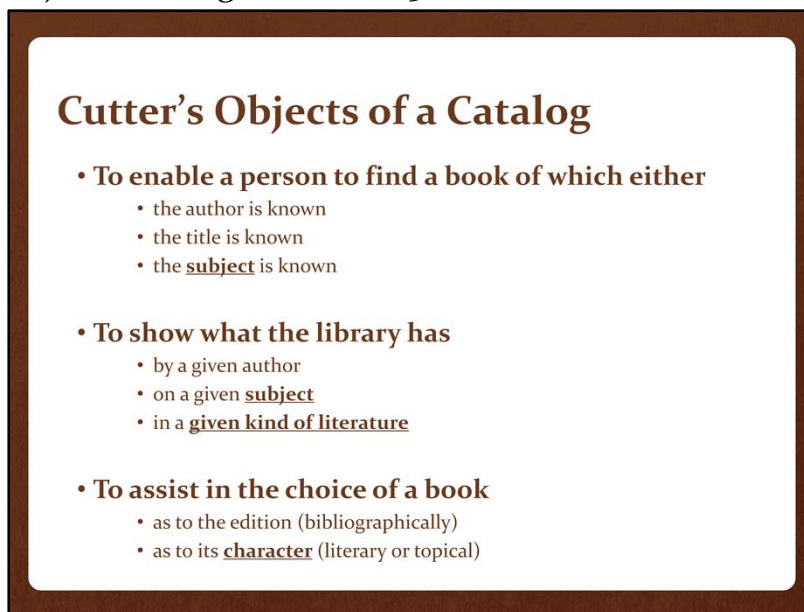
Why do we perform subject analysis?

We do it so that we can identify the subjects of our documents – so that we can say to some degree of certainty, that this is what a particular resource is about.

This allows us to provide subject searching of our resources in our retrieval tools, allowing users to retrieve documents according to their topical contents and their genres and forms. This is important because people do not always know the individual titles or authors of the resources they need, but they probably do know which subjects are of interest to them.

We also do it to bring together materials of a like nature – both on the shelves and in catalogs and indexes and all sorts of databases. We want to relate resources to other like materials and retrieve them in conjunction with those like materials.

Finally, we do it to save the users' time and to meet Cutter's objects of the catalog.



As a reminder, these are Cutter's objects again:

- To enable a person to find a book of which either the author, the title or the subject is known. This is the location or finding function. When it relates to subjects, a controlled vocabulary is an effective way to provide this function.
- To show what the library has by a given author, on a given subject, in a given kind of literature. This is the collocation or gathering function; and again, as it relates to subjects and genre/form (that is, the kind of literature), controlled vocabulary is essential.
- To assist in the choice of a book, as to its edition and as to its character. This is the selection function; and again, the character here relates to the genre *or* the subject, and controlled vocabularies are very important for that as well.

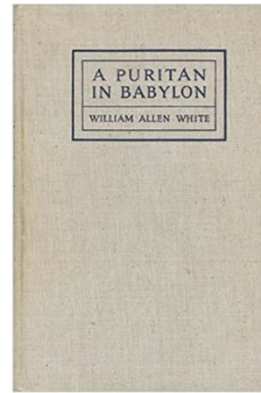
Not only did Cutter want a user to be able to find a particular resource about a subject, but he also wanted a user to be able to find *all* of the resources on that subject.

We see that, over a century ago, subject access was considered by Cutter to be an important function of a catalog. It still is today. Our international cataloging principles and the FRBR model include subject access as a major activity within the bibliographic universe.

What If There Were No Subject Analysis?

- Relying on titles
- Relying on keywords

- A lot of time is wasted
- Infoglut



If we didn't perform subject analysis, then we would have to revert to earlier practices where we relied on title words or words in the table of contents (if available) to understand the resource.

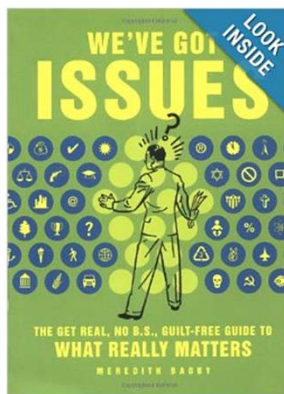
But titles cannot always be relied upon.

For example, what is the *Puritan in Babylon* about?

It is definitely not about Puritans or Babylon. From just this information, one is unlikely to determine that it is actually a biography of one of the Presidents of the United States: Calvin Coolidge.

What If There Were No Subject Analysis?

- Relying on titles
- Relying on keywords
- A lot of time is wasted
- Infoglut



This is another example:

Here's a resource that is currently held in 200 libraries according to OCLC (the Online Computer Library Center in Dublin, Ohio), which is a large bibliographic network that serves libraries and other information institutions across the United States and around the world.

The resource is titled *We've Got Issues: The Get Real, No BS, Guilt-Free Guide to What Really Matters*. In the previous example, we were looking at the title and how it can be unhelpful. This title also may not be particularly clear as to the subject matter. So what if we look at some of the chapter titles? Obviously that's going to help, right?

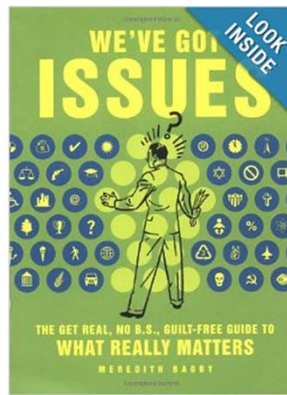
Well, maybe not. If I look at some of the chapter titles, I see,

- "Under the covers"
- "Golden Arches"
- "Yale or Jail"
- "Civil Fights"
- "The Green Generation NOT"
- "You want fries with that, baby?"

From this, it's a little difficult to get a sense of just what this resource is about.

What If There Were No Subject Analysis?

- Relying on titles
- Relying on keywords
- A lot of time is wasted
- Infoglut



With the title proper and the chapter titles as the only indicators of topics, would you know that this item was a handbook for Generation X about 17 major political issues important in the 2000 presidential election in the United States?

A little unlikely.

But, what if it's not just title access or just the table of contents? What if we can search full-text?

With full-text access to the material, thousands and thousands of different phrases will retrieve the text, but then you also deal with another problem: infoglut.

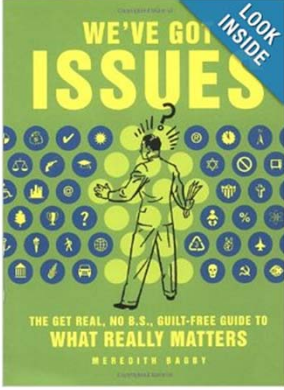
In general, the greater the number of terms available to retrieve the needed information, the higher our recall but the lower our precision. So when we do a search for a term that is used widely among many, many resources, we can quickly and easily get overwhelmed by the sheer number of search results.

And *that* is the idea of infoglut.

What If There Were No Subject Analysis?

- Political issues and Generation X
- Relying on keywords
 - Starbucks
 - Leave it to Beaver
 - Dustin Hoffman
 - Bill Gates

sted



In a full-text system, any one of those 17 major political issues would retrieve the book, along with every other word or phrase that appears in the book.

Searching for “political issues AND Generation X” retrieves the item.

But so will, “Starbucks,” “Leave it to Beaver,” “Dustin Hoffman,” or “Bill Gates.”

And the book is *definitely not* about these things or people.

What We Want From Subject Analysis

- Thoughtful analysis
- Identifying and prioritizing concepts
- Standardized translation

By analyzing resources and identifying the most meaningful subject aspects in them (for instance, saying: this book is about political issues, but it's not about Bill Gates), and then naming the most important topics, and recording those topics in our records, we allow our users to find relevant resources quickly, consistently, and efficiently.

So, subject analysis really does save the users' time.

Our intent, therefore, is thoughtful analysis, focused representation, and standardized terminology to describe the resources.