Saturday, March 5, 2022

Class 8: Thesauri in Libraries

LIBR 509: Foundations of Resource Description and Knowledge Organization

- **Intro:**

- How controlled vocabulary appears in libraries - focus on semantics (meaning)

- How are thesauri build in lib? => principles, what to include

- Example of LOC, the largest thesaurus

- **Lecture:**

- Thesauri - encounter in print materials and e-books and articles

- "Professional indexers"

- **Aboutness & Evidence**

  - **Aboutness =**

    - Is "the relation between a work and its subject matter"

    - **"**a property of a word or how well the word indicates its subjects matter, the creation between a work and its subject matter"

    - How to make a word stand in for a concept - brevity

    - Sometimes expressed as a *quality* or words - how well they indicate the subject

- **Principles of Aboutness**

  - How do we know how to connect a word to a concept?

  - Evidence, Epistemology (how can I be certain the way I think about a term will be the same you do), Obsolescence

- **Evidence / Warrant**

  - In info org, we use the term "warrant" to describe valid types of evidence (preferred term; a stable tie to a referred concept)

1

- Can help us determine the right word for a concept

- Determine which words are equivalent in meaning

- Help us determine related words

- **Common Types of Warrant**

  • Literary warrant

  • Scientific / consensus warrant

  • User warrant

  • Ethical warrant

- **Literary Warrant system**

  • Most common justification

  • Uses "the literature" as a source to create valid terminology

  • Could refer to only the current collection or the universe of books on a subject

  • Determines not only the term used but also the level of specificity for different subjects

  • Ideally presents the collection in the collection's own voice

- **Literary Warrant - Leads to decisions such as:**

  • "Gold" is not a valid subject heading if there are no books on that element, even if many other elements have subject headings

  • Setting thresholds of how many times a concept must occur in new items before a subject heading is added

  • Whether to re-index existing items to match new items or split records

- **Scientific / Consensus Warrant**

  • Doesn't represent the collection, but the ppl / experts define the valid terminology

  • Might attempt to create a system that describes reality itself

  • Usually make the collection conform to the agreed-upon terms within a community of experts

  • Ideally, creates a system that reflects fields of knowledge accurately

- **Scientific / Consensus Warrant - Leads to decisions such as:**

  - "Gold" is a valid subject heading because it appears on the Periodic Table of Elements (scientific)

  - Which camp in a scholarly community has the authoritative view of the discipline (exp. "library and information studies" or "information science")

  - How to update terminology and term relationships when there are major shifts in scientific paradigm

- **User Warrant system**

  - Using "the users" as a source of valid terminology

  - Lists of possible terms could be derived from passive collection (exp. Search logs)

  - Or could be assembled by direct user research (exp. surveys, interviews)

  - Is most powerful in creating exhaustive lead-in terms

  - Ideally, creates an intuitive system of users' natural language to represent the zeitgeist

- **User Warrant - Leads to decisions such as:**

  - Whether the system should reflect an "average" user or a multiplicity of user types

  - Using novices' terms for the items and concepts they are seeking, rather than the terms in the items themselves

  - Setting thresholds of how many times a new term appears in search logs before you create a subject heading
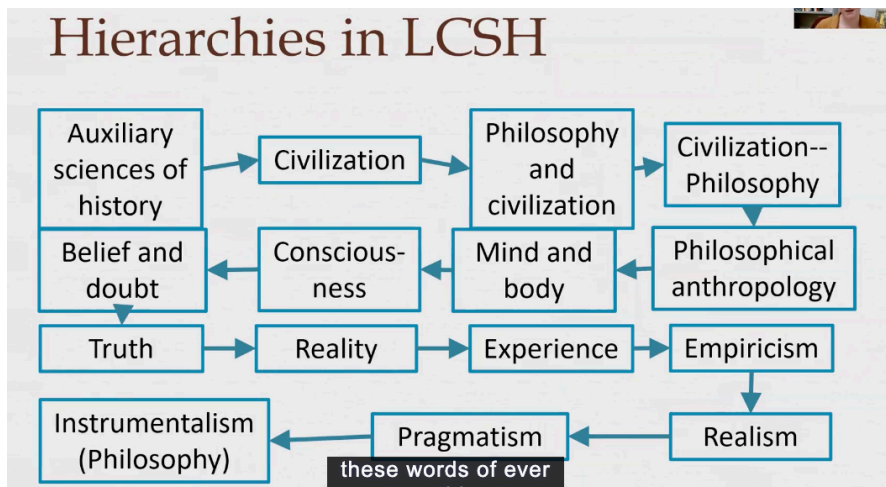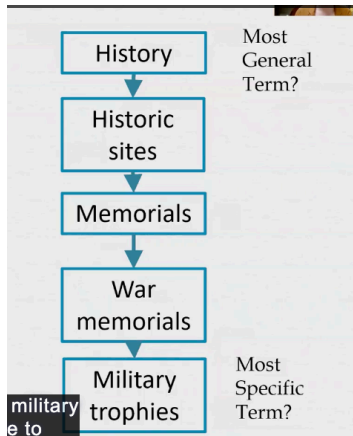
- **Ethical Warrant system**

  - A corrective, final step done on other processes to make sure they're okay

  - Unlike previous forms of warrant, less a source of terms than a test to of their validity for a system

  - Considers the religious, racial, cultural, gender, and language representation of the system (among others)

  - Often the basis for removing slurs or to correct for bias

- **Ethical Warrant - Leads to decisions such as:**

- Revising terms from slurs used by the majority to respectful terms used but the community itself

- Creating symmetry in subject headings to overcome bias (exp. "Male contraception" therefore "female contraception")

- Creating symmetry in subject headings regardless of proportionality of holdings

  - You need more detail among your terms, especially when terms are about ppl, culture, religions to prevent your system from re-enforcing prejudice

- **Library of Congress Subject Headings**

  - **Key features**

    - It's very old

    - It's in nearly every ENG book you'll see

    - It's not quite a thesaurus, do have mechanical features

    - It's got subdivisions

    - There's a set of process for how to propose revisions to terminology

  - **LCSH**

    - Subdivisions into "strings"

    - Subdivision by geography

      - Music – BC

    - Subdivision by chronology

      - France – History - to 1500

    - Subdivision by form

      - Gardens – Pictorial Works

      - Gardens – Poetry

    - Subdivision by other types of topics

      - Greta Britain - Court and courtiers

    - Multiple subdivisions

      - Strikes and lockouts - Petroleum industry - Mexico

- **Hierarchies in LCSH**

  - It can be difficult to organize subject headings in a consistent manner

  - Think about transitive properties

- **Proposing new terms in LCSH**

  - Cataloguing-in-Publication

  - Proposals for term revisions to LCSH

  - Cataloguing Lab

- **"Illegal Aliens" in LCSH**

  - Stop and read the following:

  - What completing principles or warrants seems to be at play here?

- What evidence is sufficient to make a term in LOC vs overturn a pre-existing LOC term and term relationships?

- What is the relevance of the political and institutional context to the design, maintenance, and revision of LCSH?

- **When looking at any knowledge organization system, consider starting from warrant**

  - What is the system's relationship to reality?

  - What is the system's relationship to the sate of knowledge?

  - What is the system's relationship to how we wish the world to be?

  - How is the system meant to serve the user?

  - How is the system meant to serve the authors of the collection

  - How is the system meant to serve society?

- **Assignment:**

  • Analyze an existing controlled vocabulary

  • Should cover:

    - Descriptive Points

      • When was it created / published and by whom?

      • Who maintains it?

      • What resources / concepts does it label and describe?

      • What institutions / collection is it for? Which currently / historically use it?

    - Analytical Points (should be longer than description)

      • What use case is it best for?

      • What are the obvious issues with the system?

      • How are you likely to encounter / implement it?