Valerie Darling
Summer 2020
INFO 5200 Denton
Concept Briefing

Controlled Vocabulary

Abstract

Controlled vocabularies (CVs) are important tools in the cataloging and retrieving of information. There are four types of CVs: *lists*, *synonym rings*, *taxonomies*, and *thesauri.* The purpose of CVs is to organize information so that similar and relevant content can be indexed according to a set standard, which in turn makes the content easier to find during retrieval. CVs help in limiting and defining information so that indexers can consistently input data. Created before digital content, CVs face challenges that emerged with advancing technology. A major challenge of CVs involves users who search for content by keyword. Without an understanding of the methodology behind CVs, users may abandon a search that did not yield the expected results. The timeliness of additions and corrections to CVs is also a challenge when users expect correct and accurate data updates promptly.

Introduction

Controlled vocabularies (CVs) are important components in the cataloging and retrieving of information. CVs are used to organize information for both the indexer and the user so that relevant information may be found. With four different types, CVs are multipurpose tools that create consistent, organized, and understandable word relationships within information systems so that relevant indexed information can be found.

Definition

CVs are sets of words or phrases used in metadata elements that organize and standardize concepts to ensure consistency in indexing, cataloging, and that helps in retrieving information.

Purpose and implications

The Four Types of Controlled Vocabularies

There are four main types of CVs: *lists*, *synonym rings*, *taxonomies*, and *thesauri* (NISO, 2005). *Lists* provide a limited set of terms that can be used in a given field with a definite set of options (see Example 1). *Synonym rings* are a set of synonymous terms. By defining and cataloging synonymous terms, the information system can more accurately produce relevant results (see Example 2). *Taxonomies* are a list of terms that are related in a hierarchal way (see Example 3). *Thesauri* are the most complex type of CVs and consists of standardized sets of terms that show relationships between entries (see Example 4).

The Purpose of Controlled Vocabularies

There are five main purposes of CVs: translation, consistency, indication of relationships, label and browse, and retrieval (NISO, 2005). Overall, CVs are used to help limit and organize information so that indexers can be consistent in the tagging of content. CVs also help users find relevant content by showing the relationship between words they may be searching for and the indexed words used. One of the most common uses for CVs is subject headings. One of the most popular subject headings standards is the Library of Congress Subject Headings which aims to define, organize, and standardize all informational knowledge so that it can be easily indexed and searchable.

Challenges Associated with Controlled Vocabularies

A major challenge of CVs is that many users are unfamiliar with interacting with them. CVs were created before digital content. Subject headings were used in card catalogs so that users would be able to find (and cross-reference if needed) similar content based on a certain, predetermined term. Due to the advances of technology, many users do not know how CVs work to organize information, and subsequently, retrieve it. This challenge emerges when users approach a system and search for content by keyword. Trained by years of internet search engines, systems with limited search capability make finding relevant content more difficult for users with low systems knowledge. The user may abandon their search when the system returns inaccurate or incomplete results. Though relevant content is indexed and may fulfill their request, the user may refuse to search further.

Another challenge of CVs is the process in which changes occur. With rules about who can add, remove, or alter entries, the process of making changes to a CV can become lengthy. This protracted approach can mean that the CV is deemed obsolete relatively quickly (McCutcheon, 2009). Even if the majority of the CV is relevant, users may assume that the CV has low authority and may discontinue using it.

Examples

Example 1 – Lists

Lists are the least complex of the CV types. A *list* limits the options the indexer can use for a given field. The purpose of a *list* is to help in consistency and retrieval. In the example below, the indexer may only choose one option. This predetermined *list* forces consistency across indexers. For a photograph, all indexers will choose image instead of typing in any synonymous words such as photo, photography, or

still image. This consistency helps the user in retrieving information as well. When searching for an image, this *list* ensures that all relevant content is indexed correctly and will be found.

Format: (Choose One)
        Movie
        Image
        Audio
        Text

Example 2 – Synonym Rings

Synonym rings help users find relevant information by ensuring that words or phrases that are similar are searched and retrieved during the retrieval process. The purpose of a *synonym ring* is to indicate relationships. In a *synonym ring*, the words are not based on hierarchy but are equal in relation to each other and can thought of as on a single plane.

Book ←→ Novel ←→ Hardback ←→ Paperback ←→ Textbook ←→ Text

Example 3 – Taxonomies

The main feature of *taxonomies* is the way words and their relationships are displayed in a hierarchy. The purpose is to aid users in labeling and browsing and retrieving information. In *taxonomies*, it is easy to think of words as a parent/child relationship.

- Psychology
  - Abnormal
  - Cognitive
    - Memory
      - Short Term
      - Long Term
      -

Example 4 – Thesauri

Thesauri are the most complex of the four CVs. It "contains all the features of a hierarchy, plus the associative relationship" of the entries (Leise, 2008). The purpose of *thesauri* is for translation, consistency, label and browsing, and retrieval. Many include a definition, broader and narrower terms, and related terms. The example below is from the online ERIC Thesaurus (1966).

Reading Skills
Complex behaviors developed through practice in order to read proficiently.

Broader Terms
Language Skills
Reading Ability

Narrower Terms
Reading Comprehension
Reading Fluency
Reading Rate

Related Terms
Adult Literacy
Basic Skills
Cloze Procedure
Decoding (Reading)
Functional Literacy
Inferences
Literacy
Literacy Education
Phonics

Conclusion

CVs are an important part of information systems. Though keyword searching has developed as a common search method for users, CVs are still an important part of information retrieval. Even when keyword searching, many keyword search results are found due to subject headings, one of the most common uses of CVs (Gross & Taylor, 2005). The creation and use of the various types of CVs ensure that indexed content is consistent, and that relevant content will be found. CVs create an order to indexed

information and presents it to users so that they may understand the relationships between similar content.

References

Gross, T., & Taylor, A. G. (2005). What have we got to lose? The effect of controlled vocabulary on keyword searching results. *College & Research Libraries, 66*(3), 212-230. doi:10.5860/crl.66.3.212

Leise, F. (2008). Controlled vocabularies: An introduction. *The Indexer, 26*(3), 121-126. doi:10.3828/indexer.2008.37

McCutcheon, S. (2009). Keyword vs controlled vocabulary searching: The one with the most tools wins. *The Indexer, 27*(2), 62-65. doi:10.3828/indexer.2009.21

National Information Standards Organization, (U.S.), & American National Standards Institute. (2005). *Guidelines for the construction, format, and management of monolingual controlled vocabularies: An American national standard / developed by the national information standards organization.* Retrieved from http://parlinfo.aph.gov.au/parlInfo/search/summary/summary.w3p;query=Id:%22library/lcatalog/00155874%22

Reading Skills. (1966). In ERIC Thesaurus. Retrieved from https://eric.ed.gov/?ti=Reading+Skills