

1. ORGANIZING AND DESCRIBING (REQUIRED QUESTION)

a)

With respect to digital resources (websites) knowing the number of resource descriptors for a resource helps us know how important that resource is and it is ranked accordingly by search engines and further, organizing the information accordingly giving priority to highly resources data Vs resources that are not referenced much . Eg. Google's page rank algorithm calculates the number of bibliographic references for a resource and ranks it accordingly in the search. With respect to physical resources, having a highly comprehensive resource description helps us decide which location can it be put in. Also, having a highly descriptive resource-description helps us locate the resource efficiently. E.g Knowing just the aisle number will help us know where a book might be. Knowing its shelf number will help us precisely locate it from the vast set.

By knowing the type of resource descriptors helps us create categories that will help optimize the interactions that can be done with the resource. Since resource be described to have many types, it is best to evaluate the most specific resource type and classify it accordingly. Having a strong classification system leads to the system being well organized, which in turn makes interactions easier.

b)

Alphabetical ordering - Resources can be organized based on their names.

E.g. Dictionary

Chronological ordering - Organizing resources based on their time of creation.

E.g. Museums have items grouped together based on the era they belonged to

Size and shape based - Organizing similarly sized items E.g. Pans in the kitchen go in one place and spoons in a separate bin.

c)

In organizing systems with digital resources (information systems), the principles are likely to be more formal. These resources are much likely to be computationally processed. And such processes require highly robust system of resource description (like xml) as opposed to narrative resources that are processed by humans.

Digital resources need highly formal and explicit descriptions because machines cannot understand nuances or references implied in the textual data. They are programmed to take in transactional data to provide relevant interactions.

Likewise, narrative texts (e.g novels) have least explicit resource descriptions. Because the information in has been accrued over time by a single person/ people and the most common interactions it supports is reading. So the resource description tends to have less structure. E.g The sections of the book can span to any length and can have any number of chapters the writer deems fit.

d)

d1) an intrinsic static property - The number sections in the bookstore doesnt change and is an intrinsic.

d2) an extrinsic static property - The location/address of the bookstore. Another example is the bar codes of the books in the bookstore.

d3) an intrinsic dynamic property - The number of books in the store.

d4) an extrinsic dynamic property - Current owner of the bookstore

2.

a)

The DDC was strongly influenced by the Amherst collection, which had a bias towards Christianity and religion. As such among the hundreds of numbers 88 are dedicated to Christianity and only a handful for other religions. Additionally, several important religions in the world are combined together under the category “other religions”. The simplicity of the system has allowed it continued popularity, however it limits the inclusion of newer areas of knowledge.

BISAC, follows a principle of use warrant. As such its categories are biased toward common language usage and popular culture and are more consumer/commercially focused. As such it sticks to standard classes like Law, Medicine, Music and does not allow a less precise and a more user friendly classification.

b)

Literary warrant holds that a classification must be based only on the specific resources that are being classified. A good example is the Dewey decimal classification itself. This system is heavily influenced by the Amherst Collection, which reflected the focus of the time on Christianity and educating young men about it.

c)

The classification systems can have several shortcomings. For example: Lack of flexibility to evolve, requirement of precision, onerous requirements on agreement from different parties, lack of objectivity due to biases introduced due to legacy issues, semantic imbalance, lack of scalability, and the lack of normativity. Historically libraries adopted systems, which facilitated shelf arrangement, subject analysis, collation of information on related subjects etc. Compare this with the requirements of a web portal (like Amazon or Google Books) which does not face a lot of these hurdles and the requirements are

around browsability, search-ability of the text and so on. As such the stakeholders tend to favor classifications, which are optimized for their usage, and this leads to natural conflict.

d)

The Library of Congress Classification (LCC) supported by the Library of Congress would be my choice for a new public library. It divides all knowledge into 21 basic classes, each identified by a single letter. These classes are further divided into more specific subclasses, identified by a combination of letters. It has several advantages. For example it allows the addition of new classes, subclasses, topics and geographic areas. Further, the LC numbers allow for the assignment of unique numbers to a work using cutter numbers, and the expansion of decimal numbers and the publication date. Lastly, it is widely used in other libraries in the world and has very robust systems and solutions already in place.

3. Design a Faceted Classification

a)

Faceted classification is a type of classification system that takes a set of resource properties and then generates only those categories for combinations that actually occur. For my purpose, I will consider a database that has close to 1000 video resources, which is a mid sized collection.

For an online movie database i would like to have the following facets for classification:

1. Name and Genre - I might want to store the name and its broad genre.
Possible values for genre can be Noir, Comedy, Action, Thriller, Drama, et al. E.g Argo, Drama
2. Country and Language - Information about the language of the movie and the country its produced in. Allowed values would be the name of the country, language. E.g Germany, English
3. Year of release - year value in YYYY format.

By considering each of the above facets independently but together, one can easily find a movie from the database as it narrows down the possible results at each facet. Also, each movie can have just one value for each of the above facet, which makes the classification system efficient for a database with 1000 resources.

b)

I decided to not include the facets of “subgenre” and “mood” because these descriptors are very subjective and are not applicable to all movies. Some movies might have a subgenre and some may not. E.g A documentary movie doesn’t have a particular mood. A plain vanilla comedy movie need not have a sub genre. Some movies may be interpreted as screwball comedy by some and romantic screwball comedy by a different set of people.

By not making these my main facets, I have eliminated the issue of the same resource having redundant resource descriptors (sometimes no resource descriptors) which would have made interactions with my database harder.

c)

Since the size of the database definitely determines the efficiency of my old cataloging system, because if the combination of genre and year of release and

language yields a large number of results, it might make interactions harder if we are searching.

If my friend were to use it too, I will find out what the main interactions would be.

If its a recommendation system that supports search, I will introduce another facet of “sub genre” which will describe the resources in a little more detail.

Subgenres for horror would be ‘gory’, ‘funny’ et al. By deciding on the subgenres for each movie in the database using controlled vocabulary for the subgenres, we can devise a good recommendation engine for movies. Also, search results become a little narrower and more precise.

d)

The main issue would be that of vocabulary and semantic gaps. With my system, I would know the meaning of the vocabulary I use for describing my movies - Like “slapstick” is a subgenre for a certain type of comedy movie, the same can be interpreted as “screwball” by my friend. This would lead to inconsistency and limit the interoperability of our databases. Therefore, it is important to unanimously analyze the different genre/types of movies, our interpretation of them, and the common vocabulary that we plan to use to define them.

6.

a)

Languages offer several ways to express a single thought. The choices of words used by people to describe things/ideas are a result of their life experiences and is highly dependent on the context of the situation.

The instance of a book illustrates this issue very well. A book maybe called a paperback, hardcover, text, cartoon, novel, magazine, etc.

b)

One way to address the vocabulary problem is to establish and use a controlled vocabulary. A controlled vocabulary can be thought of as a fixed or closed dictionary that includes all the terms that can be used in a particular domain. For bibliographic resources important aspects of vocabulary control include determining the authoritative forms for author names, uniform titles of works, and the set of terms by which a particular subject will be known. In library science, the process of creating and maintaining these standard names and terms is known as authority control.

Eg: Classification Systems need to decide how the various choices for the name of an author - J K Rowling, JK Rowling, Joanne Rowling, or JKR.

c)

One simple way to address the vocabulary problem is to establish a standard within the industry. For example ISBN numbers eliminate any uncertainty involved in identifying books which could be called different things by different entities. Such standards will be unique with respect to some established scope or context.

Another partial solution to the vocabulary problem is to use a controlled vocabulary. We can impose rules that standardize the way in which names and labels for resources are assigned in the first place. Alternatively, we can define

mappings from terms used in our natural language to the authoritative or controlled terms.

Another way to solve this is to establish a signature for an object by looking at the fundamental differences in domain, format, agency, and focus. A good example is the google knowledge graph.

d)

In information-intensive domains, conversations, documents, databases, software applications, or other explicit repositories or sources of information are ubiquitous. Value is created through the comparison, compilation, transformation of information in some chain. If the resource description in two systems or entities are not the same this value creation might be severely hampered. A good way to illustrate this idea is a conversation between an industry insider and an outsider. For example during a conversation with a telephone industry executive, he/she may use the term ARPU (average revenue per user) . To a person working as a doctor in a hospital this may be a totally new term.