Elin K. Jacob  
Indiana University, Bloomington, Indiana, USA

The structure of context: implications of structure for the creation of context in information systems

Abstract: Structural differences between systems of classification and categorization lead to meaningful differences in the contexts within which information can be apprehended. Examination of systemic properties, forms of interaction and functional capabilities that distinguish classification and categorization can identify the fundamental features that influence the conceptual context of an information system and contribute to its utility as a meaningful information environment.

1. Introduction

Classification and categorization are used to group related phenomena and organize the environment. However, there are fundamental differences between these two processes that influence how that organization is effected. The process of classification divides a universe of entities into an arbitrary system of classes arranged according to a set of established principles. The resulting structure is both rigid and rigorous: all classes are mutually exclusive and non-overlapping and each entity can be a member of only one class. The fact that neither the composition of these classes nor their structural context varies is the basis for the stability of reference provided by a system of classification. In contrast, categorization is a flexible and creative process that draws non-binding associations between entities. Categorization organizes the world of experience as potentially overlapping categories based on the simple recognition of those similarities which distinguish a set of entities. The fact that the membership of these categories may vary across different contexts accounts for both the flexibility and the power of categorization (Jacob, 1992).

Analysis of the structural properties and forms of interaction that distinguish systems of classification from systems of categorization reveals fundamental differences in the configuration of classes or categories that influence the functional activities of an information system and contribute to its constitution as an interactive environment. More importantly, fundamental syntactic distinctions between the structure of classification and the structure of categorization lead to meaningful differences in the contexts generated by each type of representational system. These differences are significant because context not only mediates the individual's interaction with such a system but also influences the availability of semantic information "about" reality (Floridi, 2004). Analysis of the intrinsic characteristics of classification and categorization systems furthers understanding of how the functional structure of an information system contributes to context; how different forms of organization mediate communication between the searcher and the system; and how concrete organizational structures and specific types of relationships contribute to the production of meaningful information environments (Jacob, 2004).

2. Classification and categorization

Traditionally, a classification scheme consists of a set of mutually exclusive and non-overlapping classes arranged within a hierarchical structure according to a unitary and preconceived ordering of reality. Ideally, the process of creating a classification scheme is
universal, lawful and systematic: universal because it legitimises a single paradigm or world view that presumes to reflect ontological consensus (Olson, 1994); lawful because it is carried out within the framework of an established set of principles that governs the structure of classes and class relationships; and systematic because it mandates definition and arrangement of the scheme's constituent classes through consistent application of these governing principles. It is important to recognize that the scheme itself is artificial precisely because it is a tool or an artifact designed for the purpose of creating meaningful organization; but it is also arbitrary because the criteria used to define classes reflect a unitary perspective of the domain. Furthermore, because classification mandates that an entity can be a member of only one class, it ensures stability of reference and supports communication of meaningful information through the aegis of unique class labels that link all members of a given class to the formal definition of that class.

While traditional classification is both systematic and rigorous, the process of categorization is flexible and creative. Classification divides a universe of entities into an arbitrary system of mutually exclusive and non-overlapping classes according to a set of established principles. In contrast, categorization draws non-binding associations between entities based on the simple recognition of the similarities that exist across a set of entities. And, while classification provides for stability of semantic reference precisely because neither the definition of its classes nor their composition varies as a function of time or context, categorization divides the world of experience into potentially dynamic groups (or categories) whose members bear some immediate similarity within a given context. That this context may vary and with it the description and composition of the category is the basis for the flexibility and conceptual power of categorization (Jacob, 1992). Thus, in contrast to the formal character of classes, categories are conceptual tools used to organize the diversity of everyday experience. Membership in a category indicates that, within the immediate context, the entity is perceived to share some common bond with other members of that category. And the category label not only points to this shared bond that links category members but also serves as a place-holder for information associated with all category members.

While it is obvious that there are similarities between systems of classification and systems of categorization, failure to distinguish between these two approaches to organization holds significant implications for the design of information environments. This confusion appears to stem from the misconception that "classification" and "categorization" are synonymous terms referring to the same process – a misconception apparently arising from the fact that both are mechanisms for organizing information. Although the literature of library and information science contains many examples of the indiscriminate use of "classification" and "categorization", there appears to be growing appreciation for the differences in structure that characterize these two systems of organization.

There are six dimensions that can be used to analyse and compare systems of classification and categorization: process, boundaries, membership, criteria for assignment, typicality and structure (Jacob, 2004). Of these six dimensions, structure is the most discriminating because it encompasses distinctions based on the other dimensions of process, boundaries, membership and criteria for assignment. Thus the structure of a classification system is generally hierarchical, consisting of well-defined, mutually exclusive and non-overlapping classes nested in a series of superordinate-subordinate or genus-species relationships. The structure of a classification system is a powerful cognitive tool – an "external scaffolding" (Clark, 1997; Jacob 2001, 2002) – that minimizes the cognitive effort required of the individual by embedding information about reality within the systematic organization of classes. For example, because an entity either is or is not a member of a particular class in a system of classification, the nesting of well-defined classes provides for determination of class membership as a relatively simple pattern-matching or pattern-completing activity. At a more complex level, the structure of the classification system forms
information-bearing relationships between classes and serves as a medium for the accumulation, storage and communication of information associated with each class in the structure. Thus the vertical relationships between superordinate and subordinate classes, which are subject to the mechanism of inheritance, serve to reduce the load on individual memory by minimizing the amount of information that must be stored with each class.

In contrast, a categorization system consists of variable clusters of entities that may or may not be organized within a hierarchical structure. Because categories need not be mutually exclusive, membership in one category does not prohibit membership in any other category. More importantly, however, the plasticity and flexibility that give categorization its creative power may actually prohibit its use as a persistent information structure because the potentially transitory and overlapping nature of categories provides that any relationships established between categories will be mutable. Thus a system of categorization creates a conceptual framework whose meaning may be short-lived and ephemeral – a framework that cannot serve as cognitive scaffolding and whose ability to provide a medium for the accumulation, storage and communication of information is limited.

3. Systems of organization

A system for ordering (Jacob & Loehrlein, 2003) provides access to resources by arranging them in some recognizable order. Typically, such a system supports access to known items through an alphanumeric or chronological sequence precisely because these patterns are familiar to most individuals. While such a system may appear to create groupings of like resources (e.g., individuals with the surname Jones), sequential order as a syntactic device cannot create meaningful relationships either between individual entities or between groups of entities. In contrast, a system of organization (Jacob & Loehrlein, 2003) establishes a network of relationships among the classes or categories that make up the system. This network of relationships consists of the conceptual linkages among concepts. Relationships established in a system of organization are both meaningful and information-bearing because they specify principled connections between two or more groupings within the same system.

A traditional classification scheme exemplifies a system of organization because it provides for the conceptual arrangement of a set of mutually exclusive and non-overlapping classes within a systematic structure of hierarchical, genus-species relationships. But a classification scheme is not a system for ordering: because distinctions between classes are conceptual, they do not conform to a recognizable, syntactic pattern of arrangement and thus require an auxiliary mechanism to support access.

Categorization may or may not produce a system of organization. Although a categorization system groups entities on the basis of similarity, identifying a set of categories without establishing meaningful, information-bearing relationships among them does not constitute a system of organization. If a process of categorization does not impose a systematic organization of member categories, if it does not establish meaningful relationships among categories, then it is simply a mechanism for grouping. For example, dividing items on a shopping list into categories based upon point of purchase (e.g., grocery store, gas station) simplifies the individual’s interaction with her environment but neither specifies meaningful relationships nor imposes a recognizable order.

4. The nature of context

Zerubavel contends that the individual finds order and meaning in her environment by splitting and lumping – by imposing boundaries on the objects of experience in order to mark categories as distinct “islands of meaning” (Zerubavel, 1991, 5). How an entity is categorized creates a context that both provides information about it and shapes one’s interaction with it.
There is no universally accepted understanding of what it is to be "context." Dourish (2004) points out that the positivist position sees context as a problem of representation, an approach that accords with the "engineering point of view" that understands context as "the collection of relevant conditions and surrounding influences that make a situation unique and comprehensible" (Brézillon & Pomerol, 2001). But Dourish argues that the positivist position does not capture "the kind of thing that context is." He contends that it is more appropriate to understand context as a problem of interaction supported by four fundamental assumptions: 1) context is not a given but is related to a particular object or activity; 2) the features of context are dynamic; and 3) relevant to the immediate situation; and 4) context is generated by and maintained within the immediate situation or activity. In this way, Dourish lays the theoretical groundwork for the understanding of context espoused by Nardi: "Context is both internal to people – involving specific objects and goals – and, at the same time, external to people, involving artifacts, other people, specific settings. A context cannot be reduced to an enumeration of people and artifacts; rather the specific transformative relationship between people and artifacts ... is at the heart of any definition of context" (Nardi, 1996, 76). Elaborating on Nardi's observation of the "transformative" aspect of context, Brézillon & Pomerol (2001) propose that, by focusing on a particular task or activity, context is transformed into functional knowledge that the individual uses to anticipate the outcome of an activity.

Context may be understood as a conceptual framework that influences activity - an abstraction of the physical, temporal, social or affective conditions that emerge from a particular situation and mediate or "transform" the individual's understanding of and interaction with her immediate environment. Barsalou (1989) observes that the concept associated with a category on any given occasion is comprised of: 1) context-independent information that is present for all instances of a category; 2) context-dependent information associated with, or cued by, the immediate interaction with a category (object); and 3) goal-related or experiential context-dependent information brought to the interaction by the individual. Stability of reference across contexts is provided by context-independent information that is intrinsic to any interaction with a given category; it frequently consists of category features that are available to or shared across a population of individuals. The stability that exists across any two interactions with a particular category, whether observed across individuals or within the same individual across interactions, will depend on the amount of overlap between the context-independent and context-dependent information that form the conceptual basis for each interaction. If there are no known features that determine category membership, if the membership of a category is not fixed by a set of context-independent criteria but is wholly dependent upon immediate context and/or individual goals and experience, stability of reference is eroded and the possibility of effective communication between the individual and the system is undermined.

5. Structure as context

The functional role of context within a system of organization can be investigated by analyzing the structure of four information environments: classification, postcoordination, precoordination and free-text searching. Of these, free-text searching is the most rudimentary. Although it shares with systems of classification the creation of mutually exclusive, non-overlapping and rigidly bounded groups whose membership is constrained by an explicit criterion of assignment (i.e., the alphanumeric search string used to query the system), free-text searching lacks an established set of principles that govern the structure of classes and class relationships. It is a system of categorization in the very broadest sense; but it is, at best, an elementary mechanism for grouping. Because similarity is measured by simple matching of alphanumeric strings, members of a free-text category frequently share only a superficial,
syntactic similarity without deeper semantic connections. And, because free-text searching exemplifies the simplest of structures (i.e., two antonymous classes, one comprised of resources that match the query string and the other of those that do not), it fails to provide an expressive context that can mediate interaction between searcher and system.

Unlike free-text searching, classification, postcoordination and precoordination are indexing systems that assign to each resource one or more descriptors representing intellectual content. Although classification is the most highly constrained, each of these three systems relies on a controlled vocabulary both to create an indexical, one-for-one correspondence between each descriptor and the concept to which it points and to support effective communication between the system and the searcher by specifying those terms which can be used to pose queries to the system.

Theoretically, a classification scheme epitomizes a system of organization because it creates a principled structure of well-defined classes that are linked by a system of hierarchical, genus-species relationships. Classification is the most rigid of organizational systems because its structure of mutually exclusive and non-overlapping classes mandates an absolute relationship between resource and class: each resource may be assigned to only one class in the structure. And, because it is governed by the application of accepted principles, classification provides the searcher with a persistent and meaningful context that emanates from a structure of explicit, conceptual relationships. However, communication with a classification system is inherently one-way - from system to searcher. In order to interact with the system in an effective manner, the searcher is frequently forced to rely on her own knowledge of the conceptual context to mediate the interpretation of structural relationships.

A classification system is precoordinated in that the classes to which a resource may be assigned are predetermined. A subject heading system is also precoordinated but it is less constrained — and less constraining — than a classification because it does not require its categories to be mutually exclusive. Because a subject heading system allows the assignment of multiple descriptors to a single resource, categories formed by the assignment of subject headings are not rigidly bounded but frequently overlap, providing multiple access points for each entity rather than the single access point prescribed by classification. However, unlike the systematic and principled structure of a classification system, structural relationships between subject headings are frequently unprincipled, unsystematic and ambiguously polyhierarchical. The relationships subject heading systems create between headings are generally descriptive, occasionally idiosyncratic and, too often, potentially meaningless. Thus the absence of principled and meaningful relationships between subject headings undermines the ability of the system to establish a context that can contribute to effective interaction between the searcher and the information environment. Unlike the explicit and meaningful context provided by a structure of well-defined and mutually exclusive classes, conceptual context afforded within a subject heading system cannot be blindly assumed to be either meaningful or information-bearing, rendering its utility negligible at best.

Precoordination constrains interaction between the individual and the system through the establishment of a finite collection of class labels or subject headings that delimits the complete set of possible search queries and thus predetermines the composition of retrieval sets. In contrast, postcoordination allows the individual to build her own categories at the time of retrieval. Because these searcher-constructed categories are limited only by the set of individual terms that comprise the indexing language, postcoordination affords the searcher greater interaction with the system. Unfortunately, the advantage that might accrue to the individual who creates her own categories is offset by the absence of a meaningful structure of relationships. The individual is empowered to create unique and potentially idiosyncratic search categories precisely because the system itself does not establish any but the simplest categories — those defined by the individual descriptors assigned by an indexer. Thus, like free-text searching, posing a query to a postcoordinate system simply divides the collection
into two groups: the set of resources that match the search query and those that do not. Obviously, postcoordination is a mechanism for grouping, not a system of organization. Unlike free-text systems, however, the basis for grouping in a precoordinate system is semantic, rather than syntactic. But, because postcoordination fails to establish a principled framework that supports information-bearing relationships between categories, a postcoordinate system can neither provide nor support a persistent structure of conceptual relationships that contributes to the establishment of meaningful context.

References