Chapter 2.
Category-Based Models for Knowledge Representation

Representation is the foundation of information systems. There are many possible representations. Here we will focus on categories and classes as representations, that is what we can call symbolic representations. These also include explicit relationships between the concepts. This approach is often termed "knowledge representation" although it could equally be termed "information representation".

Dog
Perro
Canis Lupus

Figure 2.1: Some representations and descriptions for "dog". Some are for a specific dog; others are for the class of dog.

2.1. Clusters, Categories, and Classes

2.1.1. Categories and Classes are Representations

When we interact with the world we encounter individual objects But, of course, those objects fall into groups. Some of the groups are ad hoc clusters say, all the objects which are on a desk. If the clustering seem important or if there is a similarity among the objects themselves we put them together into category.

Categories and classes usually involve similarity based on several attributes Classes are formalized than categories and are often based on consensus from members of a group. Categories are probably the simplest type of representation. Categories and classes make life easier, people do not have to judge individual situation separately. They can instead, categorize the situation and follow the rules which apply to it. Suppose you are organizing your kitchen. You would probably try to put similar things together — the spices on one shelf, the canned soups on another, and so on. Eventually, the categories help to simplify the complexity of the natural world. Rather than remembering or communicating every detail about a complex situation, the categories provide sufficient detail to allow a person to develop reasonable expectations about that situation. Categorization is the first step in knowledge representation. To create a database, for instance, we must categorize to what entity class each entity belongs (2.3.1). Later, we will consider related topics such as categories in human cognition.... Categories are not always useful; people can over-generalize. Placing items into a category that is not a perfect fit creates representational bias. The properties of the category are assumed to belong to that item. Alternatively, important events which do not fit into any category may simply be ignored (5.3.1).

Every representation leaves something out. Thus, there may be a "representational bias". Some information which does not fit neutrally into the representation. It may not come packaged the way we need it, it may be squeezed into categories into a category where it doesn't fit, or may simply be lost. Instances have specific values for each attribute. The attribute and its associated value are known as attribute-value pairs.

2.1.2. Categories and Classes as Defined by Attributes: Aristotelian Categories

A category is a representation which can simplify complexity by placing entities into a limited number of groups. Movies can be categorized into genres which glosses over many, many details. The simplest type of categories, "Aristotelian categories," are determined solely by attributes or characteristics inherent

Figure 2.2: Plato (left) and Aristotle (right) shown in a detail from *The School of Athens* by Raphael. Plato is pointing upward to signify his belief in prototypes (Platonic Ideals) whereas Aristotle gestures to the ground to indicate his emphasis on empirical attributes. (redraw-K) (check permission)

to the items to be included in the group. These "defining attributes" — those attributes that define whether or not an item can be included in an Aristotelian category — must be universal for the entire category. That is, all the members of an Aristotelian category must share all of the defining attributes that make up that category. This leads us to distinguish between attributes that are required for category membership — defining attributes — and attributes which, though often associated with a category, are not required for membership in that category. These are called "characteristic attributes" (Figure 2.3).

Scientific knowledge is often thought of as identifying attributes and processes so Aristotle is regarded as one of the founders of scientific reasoning (10.2.0). Formally, Aristotelian categories are defined as a conjunction of attributes. Such attributes should be able to be combined and they could be used for logical inference.

Defining Attributes
has two legs
hatched from eggs
warm blooded
has wings and feathers
Characteristic Attributes
sings
found in trees
flies

Figure 2.3: Defining or essential attributes and characteristic attributes of a "bird".

Necessary and sufficient attributes.

While a category system may be very useful for one community or for one application, it may leave out aspects which are crucial for other applications. Not every object fits neatly into a category so sometimes there has to be a forced fit. The need to assign such categories can be biased by representation.

Good representations capture important information in an effective way. Representations can provide information to users within an appropriate context; that is, they can be copied and, in some cases, decomposed and reassembled. Aristotelian categories are symbolic representations. There are other structures for representation such as index representations. In this chapter we focus on symbolic representations but it is also worth remebering that there are also non-symbolic representations such as equations and distributed representations such as some neural networks (A.10.4).

As we have already discussed, classes extend categories by applying a conceptual framework. They are "top-down". A classification could be based on counties of the world. Classification should be differentiated from categorization or clustering which are purely data dependent. Typically, classes are based on a formal classification system while categories are based just on ad hoc similarity [2]. The spirit of this endeavor is clearly Aristotelian.

Classes are sometimes based on superficial attributes. Other classes are based on underlying processes. There has been a shift in botanical classification from external plant structures to classification based on evolution (10.5.1) and genetics. Similarly, there has been a change in the classification of diseases from symptoms to the underlying processes.

2.1.3. Prototypes and Other Approaches to Categories

While models based on Aristotelian categories dominate many information-system applications such as databases, many other models have been proposed for categories although these are not often employed in information systems.

Categories as used by people don't always seem to follow the Aristotelian approach. While we will discuss the implications of this more when we consider human cognition (5.3.0). Is a whale a fish? Although whales are mammals – based on attributes such as feeding milk to their young – many people think of them as fish. People don't seem to use purely attribute-defined categories; rather, they seem to interact with entities as "prototypes". A prototype is an idealized form that captures an essence. This is Plato's approach and unlike Aristotle's approach in which an object is either entirely in or out of a category, there is a degree of similarity or typicality in category membership. That is that some attributes are more typical than others. The distinction has implications across many areas of information systems. While Aristotle's and Plato's approaches are the most widely discussed. there are several other approaches to and issues for categories (Figure 2.4):

Label	Description	Example
Continuous	Some attributes do not have distinct boundaries.	An example is colors. Even, seemingly distinct attributes may be continuous (Figure 2.5).
Abstract	For some categories we cannot define with clear attributes.	Beauty. Many social categories.
Functional	Defined by function rather than by attributes.	Is a tree branch a chair (Figure 2.6)? Are all tree branches chairs?
Radial	Radial categories are extended from a central example or prototype (Figure 2.7). These are the result of analogy and metaphor.	
Family Resemblance	Some categories do not seem well defined by a single set of attributes [8]. These are thought to be show similarity somewhat like the resemblance among members of a family so these are termed "family resemblance" categories. No one attribute is always associated with the categories. That is, these are a disjunction of conjunctions.	The definition of games (Figure 2.8).

Figure 2.4: A variety of other models and issues which have been proposed for categories.

2.1.4. Relationships among Classes

Individual concepts can be part of a larger set of inter-related concepts. there are other concepts and relationships among them. Some common types of relationships can be identified. Indeed, relationships are so important that many of their attributes can be described. From very general to very specific. Related concepts versus named relationships. Binary, n-ary. Relationship among composite objects. Another type of relationship among objects is an ordering.

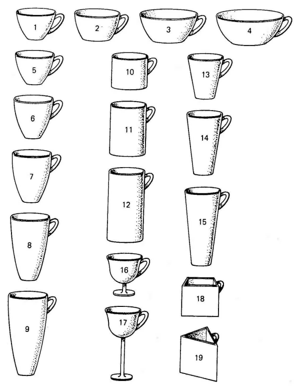

Figure 2.5: At what point does a cup become a glass, a goblet, or a jug? (check permission)(redraw)

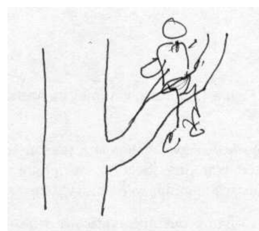

Figure 2.6: Is a tree branch a “chair”? A category may be defined by its function.

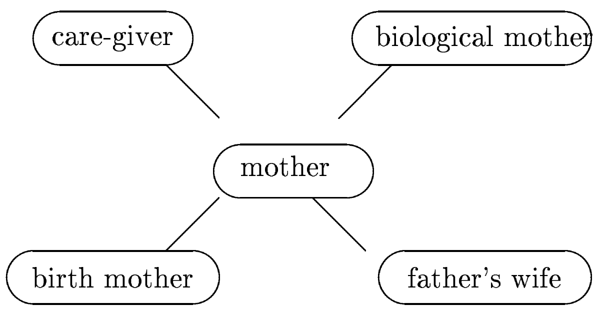

Figure 2.7: Sets of radial categories have a central theme and related concepts, but the related concepts are not differentiated by simple attributes.

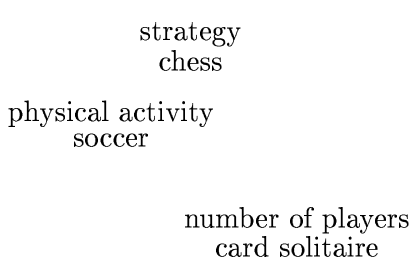

Figure 2.8: No single set of attributes seems to define a game. Rather, there are subsets of attributes which games possess.

Grouping allows intricate objects to be understood and organized more easily by reducing their complexity. Another way to simplify the complexity of the natural world is through grouping. Hierarchy and aggregation are illustrated in Figure 2.9. Hierarchies show "is-a" relationships while aggregations show "has-a" or "part-of" relationships. Aggregation groups together objects that are part of a broader conceptual unit.

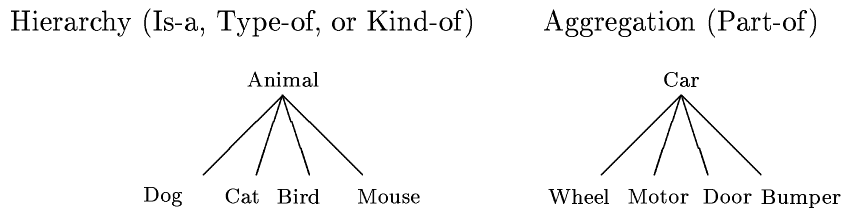

Figure 2.9: Two types of grouping relationships: hierarchy and aggregation.

In hierarchical relationships attributes may be carried, or inherited, from more general classes to more specific ones. An animal is the "parent" of a bird and a bird is the "parent" of a canary. Inheritance is an efficient way to store information because characteristics (such as laying eggs) do not need to be stored with every instance, but only with the parents. By continuing with this logic, we might get even more specific and refer to a particular canary. By doing so, we would move from types (of birds) to tokens (specific examples). Semantic relationships explicitly describe the inter-relatedness of concepts (Figure 2.10).

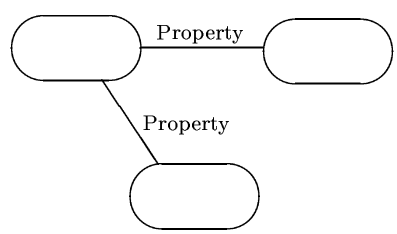

Figure 2.10: Concepts and the relationships between them are identified. This forms a sematic network.

2.2. Knowledge Organizing Systems

Sets of categories and classes are useful for describing things. Descriptions often reflect representations but they should also facilitate access [7]. They need to be tailored to the needs of the people who will be using them. There are many types of descriptions and we consider them at many places in this book. Some descriptions, those we consider in this section, are simply a few words. Descriptions would also include metadata (3.2.0) and abstracts (3.2.7). Descriptions of entire resources versus the contents with semantic annotation. Different ways of describing things.

Epistimology. Frameworks for describing knowledge.

Machine readable names. Descriptions of scientific data sets (10.3.0).

These are useful for bibliographic rcords as described in ((sec:bibliographiccontrol)).

Getting broad acceptance of metadata sets is often very difficult in the digital world. More often, they are used by small specifialist communities.

Systematic classes. Most often this would be part of specifically selected set of terms. There are several ways these models can be structured. Here we consider three approaches: Taxonomies, Thesauri, and Ontologies;

2.2.1. Subject Categories and Controlled Vocabularies

Topic descriptions versus other attributes. Which attributes to select and include in a set of metadata. Systems of metadata (3.2.1). It is useful to have a standard set of descriptive terms — a controlled vocabulary. Although there are differences among concepts, in a controlled vocabulary, these distinctions may be helpful. This process of selecting optimal terms is similar to the process of defining entities. We need to extract terms for a set of documents that are pre-defined as referring to that set. Figure 2.12 shows the stages for such a systematic development, in this case of a thesaurus. Another basis for a developing a controlled vocabulary is by examining the words people use in questions people ask when using the collection.

abode, address, apartment, asylum, bungalow, cabin, castle, cave, commorancy, condo, condominium, cottage, crash pad, diggings, digs, domicile, dormitory, dump, dwelling, farm, fireside, flat, habitation, hangout, haunt, hearth, hideout, home plate, homestead, hospital, house, hut, igloo, illahie, joint, living quarters, manor, mansion, nest, orphanage, pad, palace, parking place, place, residence, resort, roof, rooming house, roost, shanty, shelter, trailer, turf, villa.

Figure 2.11: Terms that may be used to describe a “home” (adapted from Roget). While the variants have different cases, for indexing, it is usually clearer to use just one standard term “home” than all these variants.

	Examples	
	Original Terms	**Final Term**
1. Combine related terms	Aesthetics and Esthetics	Aesthetics
2. Combine related concepts	Aesthetics and Production Values	Production Values

Figure 2.12: Steps in vocabulary reduction for creating controlled vocabulary word lists.

Many concepts are combinations of other concepts. The concept of “doctor” or “nurse” combines the concepts of “person” and “medical treatment”. Each concept is independent, i.e., orthogonal, fom the others. This process of identifying the underlying dimensions is known as “semantic factoring”. Recall that semantics is the study of meaning in language. Furthermore, the concept of “hospital” could be decomposed into “building” and “medical treatment”.

The Semantic Web also addresses many of these issues and it is often used in applications beyond those normally considered by traditional information specialists. Moreover, the Semantic Web emphasizes making the tags machine readable. On the other hand, sometimes the lessons of the traditional approaches are lost in the study of the semantic web. However, the very strengths of controlled vocabularies also suggest limitations. Many nuances are not able to be expressed and there can be drift of meaning across time [6].

2.2.2. Knowledgebases: Describing Domains with Sets of Classes

Concepts do not exist in isolation. Rather, than describing separate descriptions, we need sets of related descriptions. Classification policies. Classification model. Description logic. Conceptual frameworks. These are basic models for networks of concepts. It’s also worth noting that these descriptive system reflect social efforts and help to define the world for members of the social groups.

Hierarchical Classification and Taxonomy

Grouping relationships can be stacked one on another to form a hierarchical classification. Such hierarchical classification is particularly easy to understand and navigate. An obvious example is library classification system which we discuss in the next chapter (3.2.5). Most classification systems are hierarchical. Indeed, the system of biological classification is so strictly hierarchical that we say it is a taxonomy (Figure 2.13).

Kingdom: Animal
Phylum: Chordate
Class: Mammal
Order: Carnivore
Family: Canidae
Genus: Canus
Species: familiarus

Figure 2.13: The zoological taxonomy for dogs.

Inheritance relationships are not always so simple. While it is true that almost all birds can fly, there are exceptions. Penguins and ostriches are birds that cannot fly. A special attribute would be needed to mark such exceptions. such as a subclass of birds that cannot fly, such as penguins and ostriches, be developed? Sometimes, entities may inherit properties from more than one parent (3.2.5).

Thesauri

A thesaurus is a descriptive vocabulary in which the relationships among the terms loosely specified. NTs (Narrower Terms [children]) and BTs (Broader Terms [parents]); define a hierarchical relationship. Typically, a thesaurus also includes RTs (Related Terms) and ST (Synonymous Terms). The familiar *Roget's Thesaurus* lists words which are similar to the given word (7.2.1). Also SN, UF. SKOS.

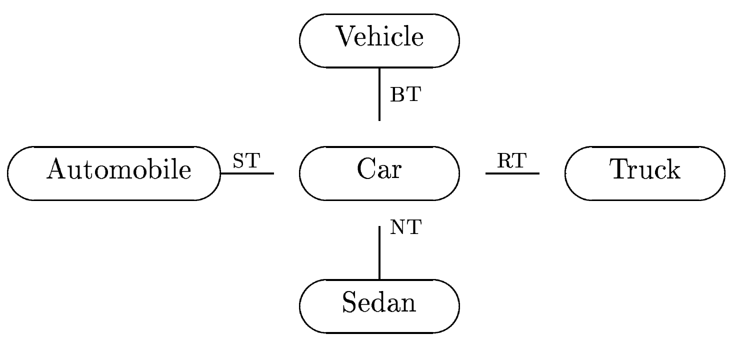

Figure 2.14: Example showing the relationships defined by a thesaurus: BT=broader term, NT=narrower term, RT=related term, ST=synonymous term

Thesauri may also provide a conceptual structure for a domain. Thesauri may facilitate text searches by providing a standard controlled vocabulary for the concepts in that domain (Figure 2.15). Not all concepts can be identified. The appropriate concepts can be selected by examining the questions people use. This is another example of identifying orthogonal, hierarchical concepts and then "composing" them into more complex objects.

Ontologies: Logic-Based Knowledge Representations

While the term ontology is often used loosely, by the formal defintion ontologies extend the semantic network shown in Figure 2.10. Formally, ontologies provide the content for predicate logic (A.6.1), which

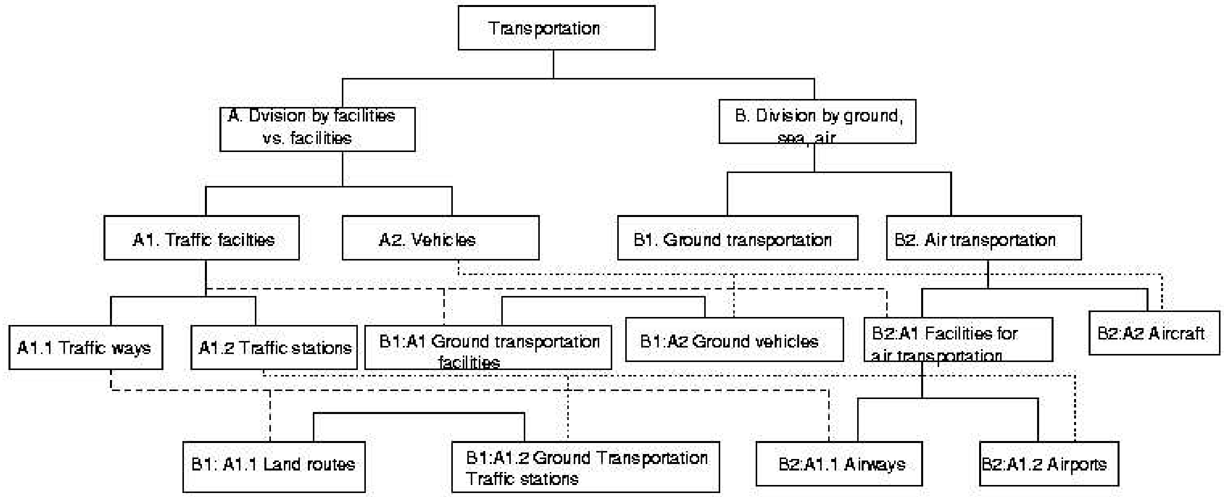

Figure 2.15: A concept hierarchy for aspects of transportation can be used to generate thesaurus terms (from [5]). Some of the resulting concepts can be composed to form complex concepts. “A1.2 B2 Airports” combines “A1.2 Traffic Stations” and “B2 Air Stations”. (check permission)

is the deconstruction of natural language to its actionable elements, thus formalizing and codifying its meaning.

The core of the Semantic Web. Automated inferences from the knowledge on the web. Such named relationships can be useful for logic; indeed, the KR often results in a “knowledge base” which represents the world by the combination of the facts in it and the inference mechanisms which operate on those facts. Several KR languages have been developed. Some of them may be used with natural language processing systems (7.2.3), the Semantic Web, expert systems and logical inference (2.2.2, A.6.0).

There are formalisms for ontologies. Ontologies are discussed further when we introduce XML (3.1.2). Furthermore, they are related to concept maps (5.3.3). Logical inference. Onotologies were originally defined as specifying the entities which are used in reasoning in knowledgebases. More logic in (A.6.0). Ontologies are used to describe the basis of the semantic annotation for multimedia content.

Ontologies, as well as thesauri, are task or domain specific. This is because for a given domain or task, the terms are usually relatively unambiguous. However, coordination across domains can be difficult as are attempts to develop ontologies for general applications, because the terminology can be ambiguous. Furthermore, unlike people for which language is highly fluid, ontologies do not adapt to context or new situations; thus, we say they are brittle. This is less of a problem for thesauri since they do not try to be as exact. Indeed, this many also represent the social uses of language and concepts [4]. Furthermore, there is the difficulty of a combinatoric explosion [1].

Association rules. Market basket analysis attempts to find interesting relationships among concepts.

2.3. Data Models

While taxonomies, thesauri, and ontologies describe domains, data models describe sets of entities and relationships relevant to a task. Data models are systems for specifying information structure. The capture a specific set of attributes which are useful for a given set of activities, tasks, and systems. Databases implement these data models. Statistical data models.

In this section, we briefly consider the Entity-Relationship model and the Relational Model. These implement basic set relationships and entities which are very similar to Aristotelian approaches for categorization described earlier. Later, we consider the Object-Oriented model (4.9.4) which includes grouping and inheritance relationships. Typically, databases do not model general concepts; rather they model specific applications.

2.3.1. The Entity-Relationship Data Model: Entities, Attributes, and Relationships

Conceptual models sketches out the basic entities and relationships among those entitites. The Entity-Relationship model adds some basic details and constraints to that model. The ER model is a semantic data model that employs "entity classes" and relationships to model a complex system. An "entity class" is a group of objects or events which are the basic units in the model. Individual members of an entity class, such as a particular person or object, are known simply as "entities". However, the distinction between entities and entity classes is often ignored and people will speak of an entity when referring to an entity class. These entities are related to categories and classes as we discussed above but they are not quite the same; they are ad hoc constructions for a specific task. For an entity class such as VIDEOS, the specific entity "Gone with the Wind" could have attributes such as Title, Director, Year, and Length (Figure 2.16).

Figure 2.16: An entity class such as VIDEO has several attributes.

field Name	field type	field min	field max	title	order	group	...
Title	text						
Director	text						
Year	integer	1970	2020				
Length	integer	2	20				

Figure 2.17: A highly abbreviated data dictionary.

Defining the influential conceptual units. Entities of one entity class can be related to entities of another class. A STUDIO (from the entity class STUDIOS) may be responsible for a particular entity VIDEO. When groups of data statements, or particular entities and their corresponding attributes, are formed into diagrams, we call these diagrams "entity sets". When constructing a database, we may use entities in many data statements to illustrate the complex relationships that exist between entities of different classes. Figure 2.18 shows a simple Entity-Relationship Diagram (ERD).

Moving from desriptions of entity classes to specific instances. Attribute value pair: Title="North-by-Northwest"

2.3.2. The Relational Data Model: Using Relational Tables to Organize and Merge Attributes

The Relational Model organizes sets of related attributes into tables. Because tables can be computationally efficient, they are often used to implement the ER model. The attributes can be joining entries across tables as needed. Splitting attributed across several tables facilitates efficient storage by minimizing redundancy. Figure 2.19 shows tables with examples of the entity classes in Figure 2.18.

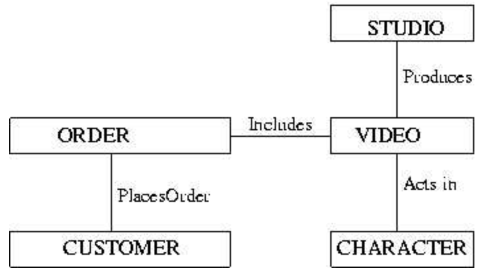

Figure 2.18: A simplified Entity-Relationship Diagram (ERD) for an online video business (attributes are not shown).

To respond to queries, the attributes often have to be re-combined from different tables. A "key" is an attribute of two or more entities or entity classes that forms a link between entities. In Figure 2.19, StudioName is a commonality between the two tables; it is an attribute for both entities. Thus, StudioName is a "key," and links the entities STUDIO and VIDEO, and consequently the tables VIDEO and STUDIO. The key guarantees there will be no ambiguity about which rows of the tables to link. The tables are usually optimized with a processes known as normalization.

VIDEO	**Title**	**Director**	**Year**	**StudioName**
	North-by-Northwest	A. Hitchcock	1959	MGM
	Toy Story	J. Lasseter	1995	Disney
	Crouching-Tiger	A. Lee	2002	Columbia

STUDIO	**StudioName**	**Phone**	**E-mail**
	MGM	800-555-1458	orders@mgm.com
	Disney	800-555-9783	orders@disney.com
	Columbia	800-555-9783	orders@sony.com

Figure 2.19: Relational tables and sample values for the VIDEO and STUDIO entities.

2.3.3. From Data Models to Databases: Databases as Information Systems

While a database program may apply a data model to some data, that is only part of what is needed for the database to be useful. Rather, databases need to be implemented as part of complete database management systems (DBMS). These are complex sets of services which serve human needs. We consider the broader context of information systems in terms of the services they provide (8.0.0).

A database is an information system that usually provides specific answers to specific queries. A database would say what videos are available for rent, what airline flights are available, or determine how much money you have in the bank.

Database Queries and Boolean Logic

Some queries place constraints on complex combinations of attributes. Booleans are generally simple relationships; AND, OR, NOT for combining attributes (Figure 2.20). We can see the formal properties of Boolean logic with "truth tables". Figure 2.21 shows the AND and OR relationships. In the OR relationship, the output is TRUE if either one of the inputs is TRUE (if either x OR y is true, then z is true), while in the AND relationship, output is TRUE only if both of the inputs are TRUE (if x AND y are true, then z is true, but not otherwise). The NOT relationship simply reverses the sense of a relationship so the NOT AND relationship has a TRUE output only when both inputs are off. Boolean logic is used in SQL but more and we will consider it again in (A.6.1). Some Boolean queries are so complex that many users do not readily understand them. For that reason, many users have designed various interactive search interfaces and protocols. Some of these query formats involve visualization and spatializing, or even free-text visual search interfaces (11.7.3).

Year=1959 AND Director='Hitchcock'
(Year>1795 AND Director='Lasseter') NOT (Title='*ToyStory*')

Figure 2.20: Some examples of Boolean queries. The example would match all entries in a movie database where the Year of production was 1959 and the Director was Hitchcock. Parentheses are used to group relationships. So, in the second example the Year and Director must match and from the those some may be deleted.

OR		
Input 1	**Input 2**	**Output**
FALSE	FALSE	FALSE
FALSE	TRUE	TRUE
TRUE	FALSE	TRUE
TRUE	TRUE	TRUE

AND		
Input 1	**Input 2**	**Output**
FALSE	FALSE	FALSE
FALSE	TRUE	FALSE
TRUE	FALSE	FALSE
TRUE	TRUE	TRUE

Figure 2.21: Simple Boolean logic truth tables.

Supporting Database Retrieval

Users need to interact with the database. This generally requires a query language to mediate interaction between the user and the data model. A query language gives the rules by which valid queries are constructed for a given data model. Queries are a useful way for users to interact with information systems. The "query semantics" of a particular information system describe the range of concepts that can be searched in that system. The most widely used query language is the Structured Query Language (SQL), a very common way for database developers to interact with a database. Formal queries must be coordinated with the data model.

Because the attributes in a relational database are organized into tables, responding to SQL often means combining data from different tables. Data from one table may need to be linked with data from another table by means of a key attribute. Figure 2.22 shows an example of using SQL for searching. Figure 2.23 shows the result of the SQL script. In particular, the fields from the tables have been "joined" with the key "StudioName". Despite its name, SQL is more than a query language in the narrow sense. It is like a programming and a system management language. SQL can create tables and control the state of the database.

```
SELECT VIDEO.Title STUDIO.E-mail
  FROM VIDEO STUDIO
    WHERE Title = 'North-by-Northwest' AND
      VIDEO.StudioName=STUDIO.StudioName;
```

Figure 2.22: An example of the SQL instruction for a low-level join operation on a relational database table.

VIDEO.VideoTitle	**STUDIO.E-mail**
North-by-Northwest	orders@mgm.com
Toy Story	orders@disney.com
Crouching Tiger	orders@columbiapictures.com

Figure 2.23: The result of a query on the tables in Figure 2.19. Specifically, there was a "join" of terms from the two tables on the attribute of STUDIO.Name followed by the "selection" of two of the columns.

Transactional Databases

Databases also often used to record transactions. That is, they would allow you to order a video, book an airline ticket, or make a deposit to your bank account. The data model is a framework for representing the complexity of the world. Although it is implicit in most database design and development, databases make sense only for some sets of entities. Part of designing the designing the database, we need to consider what attributes at truly distinctive for a given entity [3].

Database Applications

Many natural data sets are messy. This can occur when the identification of entities is not well defined or when data entry is careless. Many operating databases have duplicate entries. Thus, the data needs to be cleaned. Merging data sets. Processing data sets. The same record appears at several points with small variants. De-duplication of database records. One approach would be name normalization,

As with other information systems, a stable organizational environment for managing a database is essential to their development and maintenance. Information assurance (8.7.3). Data curation (10.3.2).

Exercises

Short Definitions:

Attributes
Attribute-value pair
Data dictionary
Database
Data model
Entities
Epistimology
Inheritance (KR)
Knowledgebase
Multiple inheritance
Ontology
Procedural knowledge
Prototypes
Query language
Representational bias
SQL
Taxonomy
Thesaurus

Review Questions:

1. List some defining and characteristic attributes for an automobile. (2.1.2)
2. Describe the relative advantages of "classification" and "key word" systems. (2.1.2)
3. Give additional examples of the grouping relationships we described. (2.1.4)
4. What are some of the difficulties in a single, simple hierarchical topic classification system. (2.1.2)
5. Compare the process of identifying entities for a database and selecting a controlled vocabulary. (2.2.1)
6. What are some of the advantages and disadvantages of a controlled vocabulary for a given topic? (2.2.1)
7. How is a database different from a knowledgebase? (2.2.2, 2.3.0)
8. In what ways are data models a type of representation. (1.2.1, 2.3.3)
9. Distinguish between "entities" and "entity classes". (2.3.1)
10. Contrast conceptual data models with implementation data models. (2.3.3)
11. List some databases you frequently encounter. What is plausible data model for one of those databases? (2.3.3)
12. Give the "truth value" for the following Booleans (2.3.3) based on the "truth tables" in Figure 2.21:
    a) TRUE AND TRUE
    b) FALSE OR TRUE
    c) (TRUE AND FALSE) OR (TRUE)

Short-Essays and Hand-Worked Problems:

1. What are some of the problems in the standard ("Aristotelian") approach to categorization. (2.1.1)
2. Explain how you would identify the category of "airport". Is an aircraft carrier an airport? (2.1.1)
3. Can you identify any truly unambiguous categories? (2.1.1)
4. What are some examples of prototypes as a model of categorization? (2.1.3)
5. Describe the pros and cons of classification into a single hierarchy versus facets. (2.1.2, 2.2.1)
6. Develop a system for categorizing the food stored in your kitchen (or your parent's kitchen). ((sec:descriptions))
7. Explain the distinction between "types" and "tokens". (3.2.3)
8. Using the approach in Figure 2.12, develop your own controlled vocabulary for either a sport of your choice or for an educational resource used at your university. (2.2.1)
9. Should subjective metadata reflect the creator's view of the material or the user's likely view of that information?((sec:descriptions))
10. Consider the objects around you as you read this. Briefly describe those objects and propose a classification system for them. (2.1.2)
11. Consider the books you own. Make a subject classification system for organizing them. What are the difficulties? (2.1.2)

12. Critique the effectiveness of the library subject classification system used in your university library or in your town's public library. Pick a work from the shelf and explain how it might have classified in a different location. (2.1.2)
13. Give an example of a classification system you have used that is confusing or ambiguous. How could that be improved? (2.1.2)
14. What are the advantages and disadvantages of using subject classification systems as a primary information access technique? (2.1.2)
15. Ask two friends to develop subject classification systems for the same topic independently from each other. For instance, they might make a classification system for games. Compare the results. (2.1.2)
16. Hierarchies are widely used as a navigation structure for hypertext. Describe why it is useful and what are some of the difficulties in using it. (2.1.4)
17. Pick a section of the Dewey Decimal System and attempt to explain why the categories may have been selected. (2.1.2)
18. What makes an effective classification system? (2.1.2)
19. Will search engines replace the need for metadata? (2.1.2, 11.8.0)
20. Select a small domain about which you are very familiar and build an ontology of the concepts for it. (2.2.2)
21. Explain how you might create a thesaurus of (a) your personal photographs and (b) Web objects. (2.2.2)
22. Choose a topic and build a thesaurus for it. The terms should show complete coverage of the area without being redundant. Hint: Use a systematic strategy such as that illustrated in Figure 2.12. (2.2.2)
23. How is a thesaurus different from an ontology? (2.2.2)
24. Some knowledge representation projects have attempted to map all knowledge. What are some of the difficulties of doing this? (2.2.2)
25. What is a "fact"? (2.2.2)
26. Why are people inconsistent about assigning names? (3.2.3, 2.2.2, 7.2.3)
27. A grocery store might use a database for inventory control and marketing. Describe what types of queries these users might use for these applications? ((sec:conceptualschema))
28. Create a sample relational table for the ORDER attribute in Figure 2.18. (2.3.3)
29. Suppose you were designing a database which was the inventory for a book store. What entities would you identify? ((sec:conceptualschema))
30. Draw the truth table for the NOR function which is the negative of the OR function. (2.3.3)
31. Describe the following Boolean query about a book using the Dublin core attributes (2.3.3):
(Title='Ulysses') AND (Date>1900)

Practicum:

1. Build a thesaurus. (2.2.2)
2. Build and E-R Diagram. Implement a Relational Database. (2.3.2)

Going Beyond:

1. Describe how you would create a database. Be sure your account includes the following steps (2.3.0)
2. Are attributes different from entities? (2.1.2, 2.3.1)
3. Explain the difference between "descriptions", "representations", and "models". (1.2.1, (sec:descriptions), 2.3.1)
4. Identify the entities in the following scenario.

- Identify types of users.
- Identify the likely questions each user would need to be answered.
- Develop a conceptual model.
- Identify the data required to answer the users' questions.
- Identify the security needs.
- Identify the computing environment (hardware)

Teaching Notes

Objectives and Skills: Do classification. Create metadata.

Instructor Strategies: This is a short chapter. It could be combined with Chapter 1 or with the first part of Chapter 3. The Data Model section could be dropped or perhaps combined with the other formal models discussed at the end of Chapter 4.

Supplemental Readings

- Bowker, G., and Starr, S. *Sorting Things Out: Classification and its Consequences.* MIT Press, Cambridge MA, 1999

- Brachman, R., and Levesque, H. *Knowledge Representation.* Morgan-Kaufmann, San Francisco, 2004.
- Kent, W. *Data and Reality.* 1stBooks, 2000.
- Lakoff, G. *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind.* University of Chicago Press, Chicago, 1990.
- Sowa, J. *Knowledge Representation: Logical, Philosophical. and Computational Foundations.* Brooks/Cole, Pacific Grove CA, 2000.
- Svenonius, E. *The Intellectual Foundation of Information Organization* MIT Press, Cambridge MA, 2000.
- Taylor, A.G. *The Organization of Information.* Libraries Unlimited, Westport CT, 1999.

References

1. Combinatoric explosion.
2. Jacob, E. K. Categorization and classification: A difference that makes a difference. *Library Trends* (2004).
3. Kent, W. *Data and Reality.* North-Holland, 1978.
4. Shirky, C.
5. Soergel, D. Automatic and semi-automatic methods as an aid in the construction of indexing languages and thesuari. *International Classification 1*, 1 (1974), 34–39.
6. Svenonius, E. Unanswered questions in the design of controlled vocabularies. *Proceedings of the American Society of Information Science and Technology (ASIST)* (2007).
7. Wilson, P. *Two Kinds of Power: An Essay on Bibliographic Control.* U. California Press, Berkeley CA, 1968.
8. Wittgenstein, L. *(title).*