7.3. Hierarchical classification tutorial

Introduction

This is one of two alternative tutorials on how to develop a classification scheme. Follow only the tutorial for the approach you choose, faceted or hierarchical. Before you start developing a scheme, read the project assignment and assigned materials on classification, then read through the entire tutorial.

The first step is to determine your approach to classification.

1. Determine classification approach.

Choose a faceted approach if you
- have a variety of different facets
- have few kinds of subjects or no subjects
- have broad nonhierarchical subjects
- are highly likely to add facets or classes

Choose a hierarchical approach if you
- have primarily subject-based organization
- have a variety of kinds of subjects
- have detailed hierarchical subjects
- are less likely to add facets or classes

Most Information Organization Projects use a faceted approach. A faceted approach is recommended unless the organization is primarily subject-based and the subjects are hierarchical. If you are unsure about the best approach for your project, ask the instructor.

This tutorial is for a simple hierarchical classification scheme being used as a first facet in a faceted scheme. If you take a faceted-only approach, use the other tutorial.

This tutorial makes three assumptions: (1) that organization is primarily subject-based, (2) that the hierarchical portion of the scheme involves a subject facet, and (3) that information objects are arranged on shelves, in cabinets, etc. for physical access. However, there may be exceptions to any of these assumptions. If in doubt, ask the instructor.

The following steps do not necessarily have to be done in order.
2. Choose facets.

This is probably the most challenging step. Think about the purpose of classification. The classification process results in a unique code (call number) that identifies an individual object for physical access (shelving). **Limit the number of facets.** The code should provide only enough information for physical access, not the level of detail contained in all fields in the record.

Choose about two to four facets that are appropriate for your users and objects. Possible facets are subject (theme, genre), author, period (historical), form (literary), format (physical), grade level (audience), purpose (use), region (geographic), language, and date (publication).

The facets may or may not be based on fields already in the record structure. For example, you may choose a subject facet that is not a field (e.g., theme, genre, discipline) in order to combine broad forms of several kinds of subjects. (Remember that the user searches subject field(s), not the classification field, for detailed intellectual access.)

Give the facets simple but descriptive names. These may be the same names as elements or fields if the meanings are comparable.

**Example**

The sample collection is books and videos in a cooking school. These objects have many kinds of subjects. The facets are overall subject theme, format, author, and publication date.

3. Decide citation order (facet combination order)

Each facet has classes within it. The codes for each class within a facet are combined in a predetermined citation order to create notation. Traditionally the primary (first) facet determines the major approach to physical arrangement; whether objects are collocated by subject, author, date, etc.

For your project, decide on a citation order that relates directly to physical arrangement of the objects. Visualize the location in which the objects are stored. How do users want or expect them to be physically organized? Is it important for the objects to be collocated so users can browse the shelves? Remember that collocation not only places similar items together, but it also separates items. For example, if the primary facet is subject and the secondary facet is format, the shelf can be browsed by subject but contains mixed formats. Conversely, if the primary facet is format, the formats are separated and subjects can be browsed only within each format.

**Example**

The citation order is Theme, Author, Date, Format. This order determines the order of shelving. Placing Format last collocates books and videos on the same theme by the same author.

4. Choose classes.

Choose classes one facet at a time. For each facet:

- **Limit the number of classes.** You must have enough classes to represent all ten objects in your sample collection. This may be fewer than ten if a class represents more than one object. You may add a few classes to fill out or balance the scheme, but don't try to cover all possibilities: this is only a sample.

- **Make classes mutually exclusive.** Each object can be placed in only one physical location, so the classes must not overlap each other.
• **Do not list classes in open facets.** Some facets, such as author and year, are open: they have an indefinite or unpredictable number of classes. For open facets, do not try to list every class code, but rather explain the code in the rule (see example below and step 6).

• **Decide how to deal with exceptions.** Two kinds of exceptions frequently occur:

  1. A certain facet may not apply to an object. For instance, you have an Author facet but the object has no author. In such a case, you can (a) omit that facet in notation or (b) add a generic class, such as Anon, for that facet. One facet you cannot omit is the primary facet because this is the major basis for physical arrangement—you must have a generic class to guide the shelving. A generic class commonly used for a subject facet is General.

  2. There may be no class listed for an object. For instance, the object is a CD but the only two format classes are Book and Video. In such a case, you can (a) add a specific new class, such as CD, or (b) add a generic class called Other. The latter solution is appropriate for rare exceptions that do not justify adding a specific class.

   Explain how to deal with any of these exceptions in the notation rules (see step 6).

**Example** (hierarchical Theme facet)

- Regional/ethnic (American, Chinese, French, Jewish)
- Ingredient (meat, dairy, pastry)
- Event/meal (traditional holiday, family occasion, meal/course)
- Technique (cooking method, equipment/tools, presentation/display)
- General (mixed recipes, diet/health, quantity)

5. **Order classes and determine notation codes.**

These tasks take place almost simultaneously. In the hierarchical facet, the order of classes is predetermined, but in other facets the order may or may not be important, depending on the notation code. Here are ideas for codes:

- **Letters:** may designate main classes if there are fewer than 26 (A-Z) of these. You can skip letters to leave room to add other facets later (i.e., instead of A, B, C use A, C, E).

- **Numbers:** may designate main classes and subclasses because the number of classes is potentially unlimited. If classes are numbered, their order cannot change.

- **Mnemonics:** letters or abbreviations that aid identification of facets or classes. It may be necessary to use two to four characters to distinguish between classes. With mnemonic codes, the order of classes can be adjusted later. Possible ways of ordering are alphabetical, chronological (historical), spatial (geographic), or by importance, complexity (simple to complex), or proportion of collection (largest to smallest).

- **Punctuation:** typically used to separate codes in citation order. Common separators are periods, slashes, colons, spaces. Punctuation can be used instead of letters to distinguish facet codes.

Creating the notation code is probably the most fun! There is no right or wrong code. The code should:

- Be related to physical arrangement
- Clearly distinguish facets from each other
- Cover all objects in the sample collection
- Allow for facet and class exceptions
- Be reasonably short in length
Appendix E is only a sample: it does not contain all parts of a traditional scheme. To explain this, write a brief paragraph at the top that lists all the facets in citation order, clarifies which facet is hierarchical, and states that the index to the hierarchy is not included.

**Examples** (hierarchical Theme facet)

The scheme consists of four facets: Theme, Author, Date, Format. The Theme facet is hierarchical (see hierarchy below). The index to classes in the hierarchy is not included in this appendix.

1. General
   1.1. Diet/health
   1.2. Quantity

2. Regional/ethnic
   2.1. American
      2.1.1. Southern
      2.1.2. Cajun
   2.2. Chinese
   2.3. French
   2.4. Jewish

3. Ingredient
   3.1. Meat
   3.2. Dairy

4. Event/meal
   4.1. Traditional holiday
   4.2. Family occasion
   4.3. Meal/course

5. Technique
   5.1. Cooking method
   5.2. Equipment/tools
   5.3. Presentation/display

Here the General class is first because it is the broad generic class and should be easy to find if other classes are added below Technique. Depending on the rules, notation might look like this:


**Another example:** If parts of the code are based on abbreviations and more letters . . .

Reg. Regional/ethnic
   1. American
      1.1. Southern
      1.2. Cajun
   2. Chinese

. . . the code might look like this:

- *Hot stuff Cajun cooking* by André Blanc, 1985, book Reg1.2/Bla.1985b/
6. Write rules for notation.

Rules help ensure that the classifier synthesizes (combines) notation accurately and consistently. Write a separate rule for each facet in citation order, following the assigned rule format.

Rules must be specific, including:

- Chief source of information
- How to choose a class from the hierarchy
- Spelling and capitalization
- Punctuation between facets
- When to omit a facet
- What to add a class
- When and how to apply a generic class

Example (based on last example in step 5)

**Facet name:** Theme  
**Chief source of information:** Title page, or if title is insufficient, table of contents or text  
**Notation rules:** Classify under the most specific class available for the dominant subject.

- When two or more subjects are applicable, place under the dominant subject.  
- If the dominant subject cannot be determined, classify as General.  
- When two or more subclasses are applicable, place under the main class.  
- When a clear subject exists but there is no class, find the closest subject and add a subclass.

Use class code exactly as shown in the hierarchy. End with a slash.

**Facet name:** Author  
**Chief source of information:** Title page or verso of title page  
**Notation rules:** Use first three letters of author’s last name. Capitalize first letter. End with a period. If there are multiple authors, use first author only. If there is no author, omit this facet.

**Facet name:** Date  
**Chief source of information:** Verso of title page  
**Notation rules:** Use four-digit publication year. Do not punctuate end. If there is no year, omit this facet.

**Facet name:** Format  
**Chief source of information:** Physical package  
**Notation rules:** Use lowercase b for book or v for video. End with a slash. If the object is a format other than book or video, add a class for that format.

**Hint:** Some facets, such as date and format, are also record fields that contain *only one data value*. In this case, the chief source of information can be the *specific field in the record*. This saves the classifier’s time because he/she does not have to re-examine the object.

7. Choose unique identifier.

This is a final code added to notation to create a unique identifying code (call number) for the object. An easy solution is to use the unique RecordID number from the Inmagic record. Another solution is to use a number in/on the object. Make sure the number is *in/on every object*, that it is *truly unique*, and that it is *relatively short*. 

8. Write rule for unique identifier.

This rule goes under a separate heading in Appendix E. It should cover punctuation even if there is none.

**Example**

Append RecordID number from database record to end of notation. Do not punctuate.

9. Provide example of classification code.

Provide example of complete classification code (call number), including unique identifier, for one sample object. Explain every part of the code in order.

**Example**

The classification code **Reg1.2/Bla.1985b/1** represents *Hot stuff Cajun cooking*, which has the regional theme of American Cajun (Reg1.2), was written by André Blanc (Bla) and was published in 1985 (1985). This is a book (b) that is represented in the first database record created (1).

10. Finalize classification sections.

Be sure Appendix E contains four parts under these assigned headings:

1. **Scheme** [hierarchy for hierarchical facet]
2. **Notation rules** [rules for synthesizing notation, with three-part format for each facet]
3. **Rule for unique number** [separate rule]
4. **Example** [complete classification code]

Test your rules by asking someone else to follow them.

Write narrative for section 4.3. This should contain only enough information to respond to assignment, not all details of decisions you made for Appendix E. The example of the classification code, however, can be exactly the same as in Appendix E.

In the narrative, it is sometimes difficult to explain the effectiveness of classification for physical organization. This can be done by describing the relationship of notation to physical arrangement clearly and completely. Here is one way this might be stated:

**Example**

The facets are Theme, Author, Date, and Format. Objects are shelved in citation order: alphabetically by theme, alphabetically by author, chronologically by date, and alphabetically by format. Because some authors, such as Julia Child, create both books and videos, this order collocates books and videos on the same theme by the same author.

---

Final checkpoints

Classification schemes can vary widely, so yours may look quite different from some else’s. As a final check on this part of the project, here are reasons for which grade points are most often deducted:

- Narrative 4.3 and/or Appendix E is incomplete
- Narrative 4.3 and Appendix E are inconsistent
- Appendix E does not follow assigned format
- Notation rules are confusing or incomplete
- Classification rules are not applied in Appendix G records

All INFO 5200/4200 course materials are copyrighted and may not be copied, revised, or distributed in any form or venue, beyond their use by students for purposes of fulfilling course requirements, without prior permission of the authors or the University of North Texas.