LIS 571 Guide to Lecture 2.2 Knowledge representation p. 35

You should be able to get this content just by reading. There is no audio recording for this lecture.

This guide is long because it supplements the lecture notes by many examples that I would talk about in a seated class. The guide could be integrated with the lecture notes to make one chapter of a book, but I did not have time to do that.

p. 37 - 42 Introductory exercise
Try first on your own, then you can look at the answers given at the end of this guide, starting on p.4

Some examples for you to work out to discover a simple but very powerful principle. I will send out answers in a few days; try it first on your own.

p. 37 How easy is it for a diner to figure out what the actual differences between the options are? How many things are actually different, how many the same? Present this information so that the diner can see quickly what is the same and what is different, using less text to boot?

p. 39 A list of six bibliographic records in the MARC format. You can guess what the MARC tags (100, 245, etc. mean or you can look at the beginning of the Model catalog where the MARC tags are defined).
Look at these bibliographic records; two of them "do not belong". What about the others? How hard would it be for a user who gets these records returned in a search to figure out what the differences are? Present this information so that the user can see quickly what is the same and what is different, using less text to boot?

p. 40 What are the general types of products (ignoring the manufacturer)? Make a new node for each placed on an invisible horizontal line through the middle of the box. Link each of these nodes to the applicable nodes on the top and the applicable specific products on the bottom.

p. 41 Exactly the same idea in a different representation. What do Food product 1 and Food product 2 have in common? Make a new record that captures these common elements. Then from Food product 1 refer to the new record and omit the information captured already in the new record you created.

p. 43 Just read

p. 44 – 45 A small set of data represented in three different ways for you to have a good look to get a feel for different representations. The entity-relationship representation is the most

basic. The other two representations show exactly the same data, each in a way that is easier to understand for the human reader.

p. 44 Semantic network.
Data common to all birds (ignoring a few exceptions) are listed only for birds; they are understood to be true for every node under Bird. The subordinate nodes **inherit** everything that is said about Bird.

So what can you say about Song Bird?

What can you say about canary?

p. 45 A frame is a record with additional capabilities. The only additional capability used in the example is hierarchical inheritance.
A frame is collection of data item. Each data item is a slot. Here we have a collection of frames that all have the same slot structure. Each data item is from the entity-relationship representation, so the slot names (labels, tags) are the relationship names.

The data in [ ] are inherited. They could be omitted because a frame system would know they are there by inheritance. They are included for the convenience of the reader. A plain record system as used in most library catalogs does not implement hierarchical inheritance so all information about a work must be included in a record, whether on a conceptual level it is inherited or not

The override example illustrates how we can have Penguin under Bird without saying that a Penguin can fly, as would be implied by inheritance.

p. 47 – 49 A larger example with data on foods shown in an entity-relationship representation. The statements are arranged that one can easily see what a frame representation (a frame for each food product) would look like. I intended to list all inherited data in [ ] but discovered that I was not consistent, so I will use this as a teaching opportunity:
Look at FP20. It says FP20 <*isa*> FP5 Soup. Based on that, what can you say about the form of FP20? (Form has values of solid, semi-solid, semi-liquid, liquid).

p. 50 Semantic network. The data represented overlap with the data set from p. 50 – 51. Using spreading activation, what can you say about the risk Fred is exposed to by eating Portion-1.

p. 52 – 53 Extended frame vs. minimal frame

First look at the extended frame on p. 55. Nothing new here, you have seen extended frames in the bird example. An extended frame can be taken apart into many separate statements, most based on a two-way relationship.

The bottom of the page shows the frame for FP22 without all the inherited information. It takes very little space yet gives a lot of information. The <*isa*> FP20 packs a lot of punch: It pulls in all the information given in the frame for FP20.

Another example of hierarchical inheritance: Recipes in a cookbook

The cookbook may have a basic recipe for *basic potato soup*. That recipe may take up two pages, explaining how to select the best potatoes, how long to cook them, how to mash them up, how much water to add, etc. Then there may be 20 recipes for specialty potato soups. Each of these takes just a few lines. For example, the recipe for *leek potato soup* might say simply Use recipe for *basic potato soup.* Cut four medium leeks into small pieces and add to liquid potato soup before cooking. This saves printing the two pages on how to make basic potato soup twenty more times.

Now look on **the minimal frame on p. 52.** This frame gives information on a **buy/sell** event. Note that buy and sell are different names for the same underlying event:

> **buy** is used as the name of the event if the emphasis is on the buyer (the receiver of goods or services
> **sell** is used as the name of the event if the emphasis is on the seller (the source of goods or services

In any event, four pieces of information are needed to give a complete description of the event, as shown in the frame. None of the pieces can be taken out and make sense outside of the frame. In other words, the frame represents a four-way relationship; it **focuses on an event**.

In contrast, t**he extended frame on p. 53 focuses on a food product** and gives many different pieces of information about that food product; each of these pieces of information could be stated outside the frame using a binary relationship.

p. 54 Here is an additional example for activation of all frame elements.

Assume you stand at a street corner, opposite is a store with a little clock tower, a lady in fancy hat walks by. All of a sudden a red sports car appears and hits a pedestrian crossing the street. You store the whole scene in your mind as frame with slots

Frame: Street scene	
General scenery: Building characteristics	Street corner Store with small clock tower
People:	Lady in crazy hat, Person crossing street, many people walking by
Cars	red sports care
Events	red sport car hits pedestrian
. . .	

Two years later you see a red sports car, activating one element in your street scene. The scene you saw two years ago flashes before your eyes; activation of one frame element activates the whole frame, very similar to spreading activation.

Restaurant menu, original version

Third Course		
Fondue Feast	**Fondue Fusion**	**Lobster Indulgence**
Filet Mignon, Sauerbraten NY Strip, Nueske's Applewood Smoked Bratwurst, Hefeweizen Marinated Shrimp, Roasted Garlic Crusted Chicken, Sun Dried Tomato Ravioli and Fresh Vegetables.	Lobster Tail, Filet Mignon, Sauerbraten NY Strip, Nueske's Applewood Smoked Bratwurst, Hefeweizen Marinated Shrimp, Roasted Garlic Crusted Chicken, Sun Dried Tomato Ravioli and Fresh Vegetables.	Lobster Tail(s), Sauerbraten NY Strip, Nueske's Applewood Smoked Bratwurst, Hefeweizen Marinated Shrimp, Roasted Garlic Crusted Chicken, Sun Dried Tomato Ravioli and Fresh Vegetables.

Revised Version

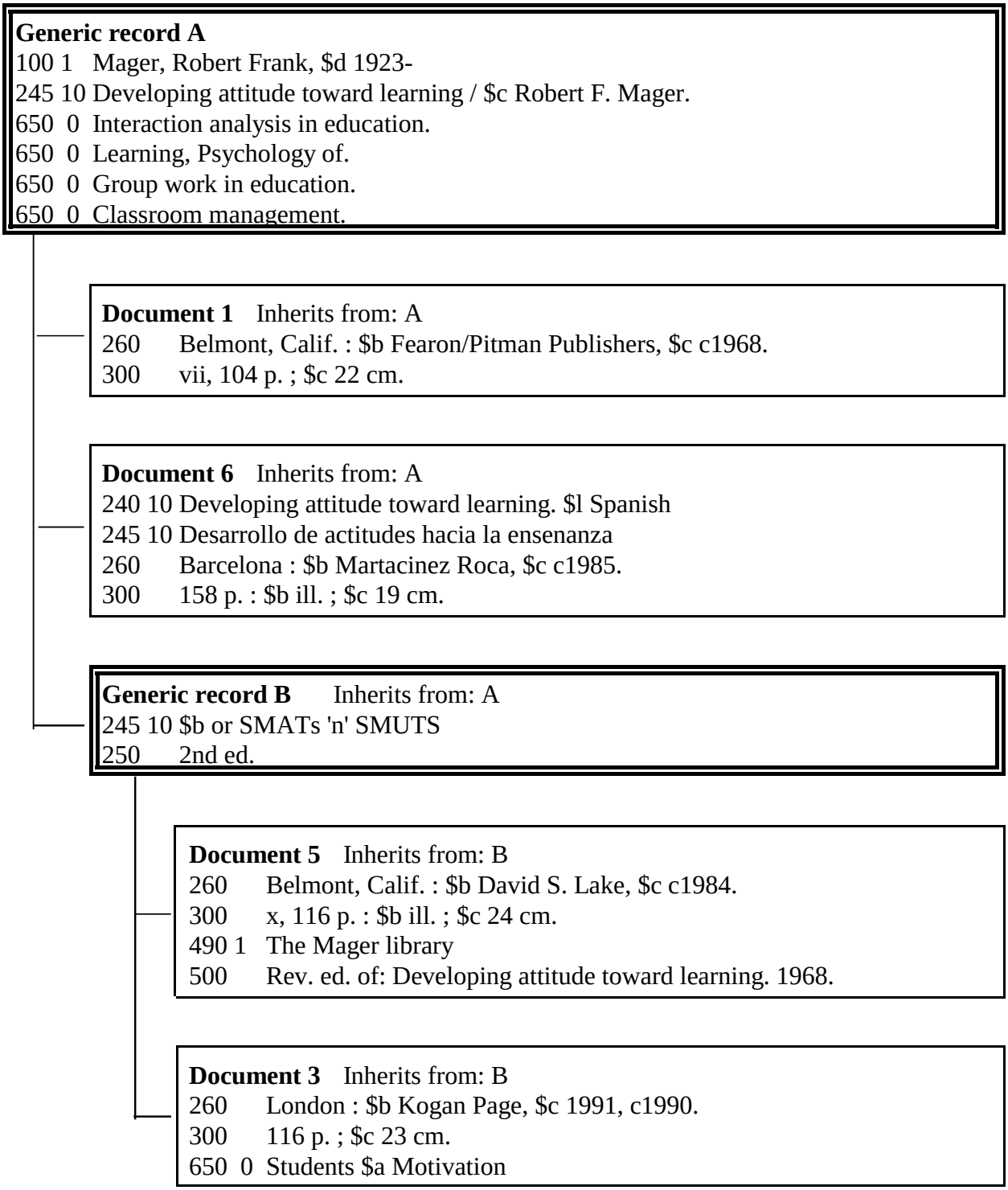
Third Course		
	Sauerbraten NY Strip Nueske's Applewood Smoked Bratwurst Hefeweizen Marinated Shrimp Roasted Garlic Crusted Chicken Sun Dried Tomato Ravioli Fresh Vegetables	
Fondue Feast add	**Fondue Fusion add**	**Lobster Indulgence add**
	Lobster Tail	Lobster Tail(s)
Filet Mignon	Filet Mignon	

To produce the revised version, I pulled out all dishes that are in common to all three choices and listed them first. Then I listed for each choice the additional items.

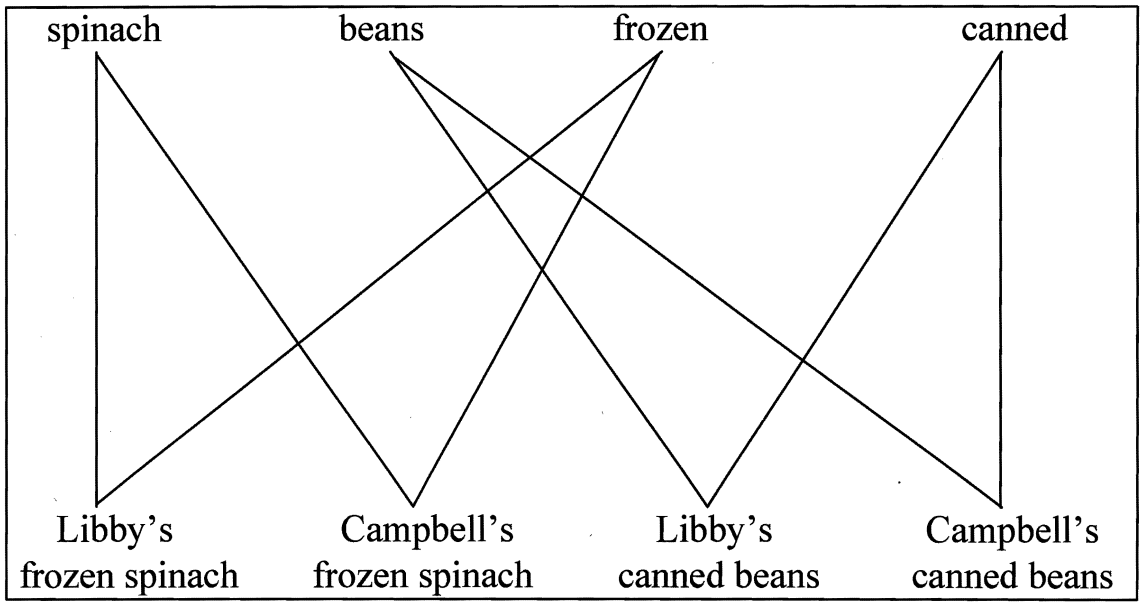
I also put each dish on a separate line to make it easier to read.

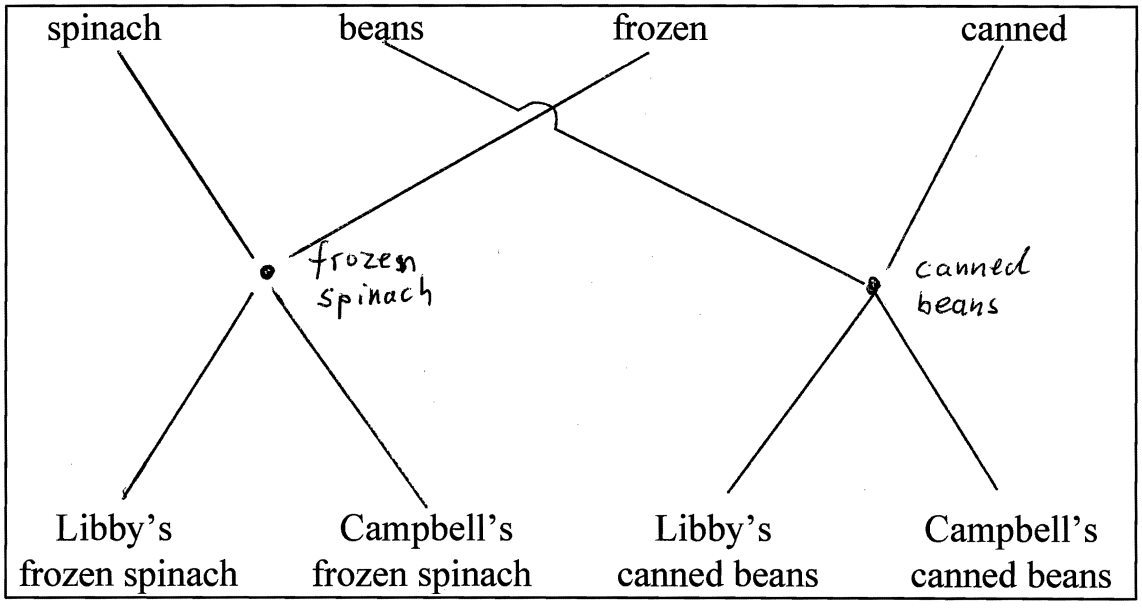
Hierarchical inheritance for efficient storage on bibliographic data

This example uses hierarchical inheritance "with override": If a field from the parent record is repeated in the child record with different information, the child information overrides what would otherwise be inherited from the parent. (Documents 2 and 4 are not related to any other document and thus are omitted here.)



Look at the semantic network below. How can it be restructured for more efficient storage? Complete the second copy of the network to show your restructuring.





- captures what is in common to a group of food products
  if the database includes 50 frozen spinach products this saves many connecting lines

Original database	Restructured database
	Food product A. Frozen spinach *Food*: spinach *Preservation*: frozen
Food product 1. Libby's frozen spinach *Food*: spinach *Preservation*: frozen *Manufacturer*: Libby	**Food product 1. Libby's frozen spinach** *isa:* Food product A *Manufacturer*: Libby
Food product 2. Campbell's frozen spinach *Food*: spinach *Preservation*: frozen *Manufacturer*: Campbell	**Food product 2. Campbell's frozen spinach** *isa*: Food product A *Manufacturer*: Campbell
	Food product B. Canned beans *Food*: beans *Preservation*: canned
Food product 3. Libby's canned beans *Food*: beans *Preservation*: canned *Manufacturer*: Libby	**Food product 3. Libby's canned beans** *isa:* Food product B *Manufacturer*: Libby
Food product 4. Campbell's canned beans *Food*: beans *Preservation*: canned *Manufacturer*: Campbell	**Food product 4. Campbell's canned beans** *isa:* Food product B *Manufacturer*: Campbell

Note: The restructured database has more records, but they are much shorter