1. Consider the web page from AutoTrader.com in Figure 1 and the OSM diagram for car ads in Figure 2. This exercise illustrates several problems and issues in schema integration:

- heterogeneous sources and the necessity to transform sources into a common representation to do schema integration (1a);
- direct (1:1) mappings versus indirect (1:n, n:1, n:m) mappings and their resolution (1b);
- automatic mappings and the difficulty of correctly discovering even 1:1 mappings (1c);
- tool support: synergistic mapping development; diagram merging; diagram clean-up and layout generation (1d).

Be sure to observe these problems and issues as you do the exercise.

(a) Using the Ontology Editor in the Ontology Workbench, create an OSM diagram that captures the information in the table. You should have a non-lexical object set called Vehicle to represent the vehicle, and you should have lexical object sets for each of the four table column headers (Year, Make/Model, Mileage, Price); plus you should have lexical object sets for each of the three subcomponents of the Make/Model column (Color, Details, VIN), where Details should be multi-valued; plus you should have a lexical object set for the “Toll Free” phone number in the unlabeled column. (Take a screen shot of your diagram in the Ontology Workbench to hand in for (a).)

(b) The MapMerge tool in the Ontology Workbench only maps and merges object sets and relationship sets that have a one to one correspondence. Thus, in this next step, modify your OSM diagram for AutoTrader.com so that the object and relationship sets match appropriately with the object and relationship sets in Figure 2. In particular you will need to split up Make/Model appropriately into Make, Model, and Trim, and you will need to add ModelTrim and adjust the relationship sets. (Take a screen shot of your diagram in the Ontology Workbench to hand in for (b).)

(c) Load the OSM model instance in Figure 2. Obtain it at:
   Open the MapMerge tool in the Ontology Workbench (under the tabs menu). Open your modified AutoTrader.com ontology created for (b) and then click on the double-folder icon to open the downloaded Car Ads ontology. Next, click on Preference (under the Edit menu) and set the preferences to Semi-Automatic Mode with the Naive Mapping. Then, click on the mapping icon immediately to the right of the double-folder icon. Follow directions, if any, to complete the naive mapping step. Finally, check the mappings and add/delete any mappings as may be necessary. (Take a screen shot of the diagrams with the mappings in the Ontology Workbench to hand in for (c).)

(d) Use the merge feature in the MapMerge tool in the Ontology Workbench to merge your diagrams. (Click on the merge icon, immediately to the right of the mapping icon.) You need not clean up the resulting diagram, but you may: either adjust it by hand, or click on the automatic layout icon, which looks like a tree diagram and is near the right end of the icon buttons. (Take a screen shot of your merged diagram in the Ontology Workbench to hand in for (d).)

2. Consider the data merge problems for merging your AutoTrader.com ontology with the car ads ontology in Figure 2. (Here, again, be sure to observe the general problem each exercise illustrates.)
(a) **Non-alignment of data values.** The data values under *Make/Model* are a composite of the make, the model, sometimes the trim, and sometimes a feature such as 2WD. What needs to be done to split the *Make/Model*-string into its component parts and place them in their proper object sets? How would you automate this?

(b) **Internal vs. external data-value representation.** What value canonicalization methods do you need for values in the AutoTrader.com page? Why might canonicalization methods be needed for make and model names?

(c) **Value categorization in generalization/specialization hierarchies.** When you merge the information, how could you automatically categorize the *Details* in AutoTrader.com with respect to the *Feature*-specialization object sets in the car ads ontology?

(d) **Record linkage and de-duplication.** Suppose it is possible that the same car is advertised both in the car ads ontology and in the AutoTrader.com site. How could you automatically determine which cars are the same?

(e) **Values left in sources vs. values extracted to global database.** Suppose that instead of
extracting the values into the global schema for querying, we leave the values in the sources. Explain what the system needs to “remember” about how to obtain values so that it does not have to re-align the source schema with the global schema at query-execution time, and explain how the system executes the query. (To be specific, write your explanation in terms of the query: "Find the makes, models, and phone numbers for cars that cost less than $10,000, are newer than a 2007, are red or yellow, and have a sun roof.")