UMAP: User's guide

1. Overview:

UMAP has five main steps:

1.1. Correspondence generation

In this step the UMAP's schema matcher component is used to assist the user in generating correspondences between the attributes of the source schema to the attributes of the target schema.

1.2. Merging sibling relations

This is the first step of the UMAP's schema mapping component, in this step UMAP uses the query log of the source and target schema to find relations that have an IS-A relationship with a high-level relation and generates merged relations.

1.3. Aggressive chase

In this step UMAP performs the classical chase and the aggressive chase described in the paper in order to generate logical relations.

1.4. Mapping generation

In this step UMAP generates the mapping rules between the two schemas. UMAP also performs conflict resolution when an attribute in the source schema correspondences to more than one attribute in the target schema.

1.5. Mapping rewriting

In this step UMAP generates the mapping rules using the original relations in the source and target schema (no merged relation is considered).

2. Using UMAP:

After downloading the “umap_demo.zip” file, unzip it and then run the jar file using the following command:

java –jar umap_demo.jar

2.1. UMAP’s main interface

As shown in Figure 1 area 1, the user can input the source schema and its corresponding query log, the same applies to area 2 except that it’s for the target schema.

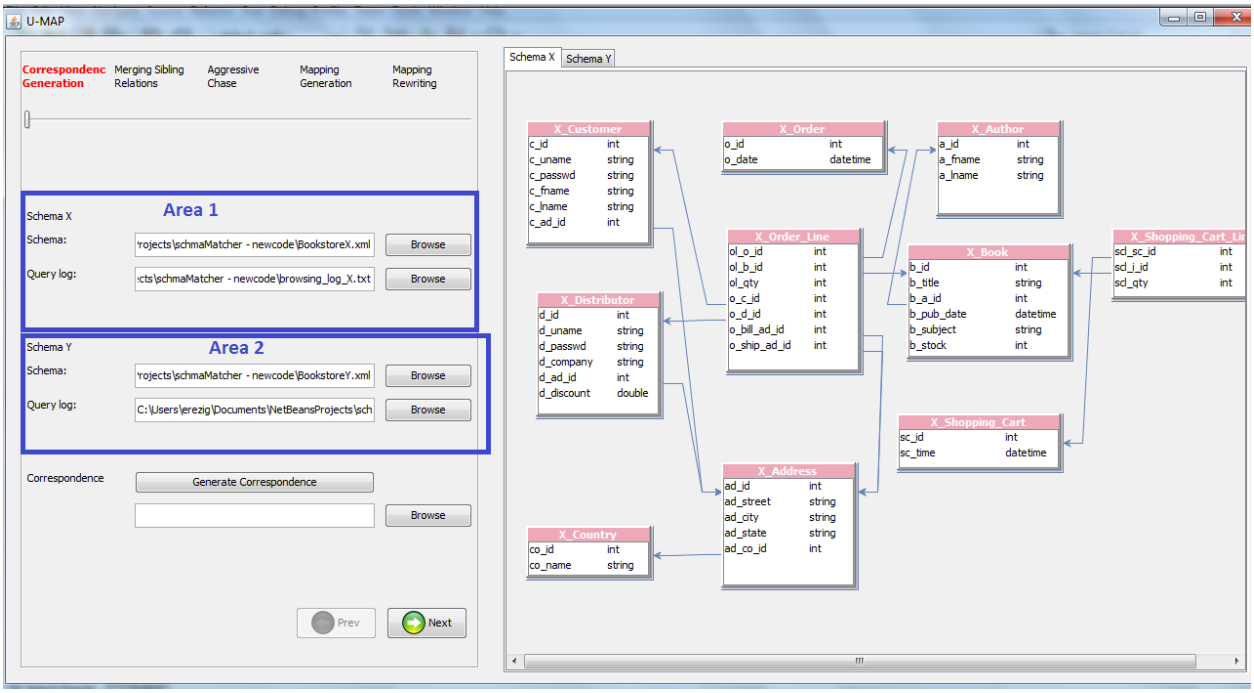

Figure 1. UMAP’s main interface

2.2. Correspondence generation

When the “generate correspondence” button is clicked UMAP pops up the schema matcher interface through which the user can see the correspondence table using different matchers and different levels of names similarity between the source and target schema. The interface also allows comparing the matching results of UMAP with a user-provided ground truth file. Figure 2 shows the schema matcher’s main interface after clicking “match” and comparing the result with a user-provided ground truth file, in order to run the schema matcher the user has to perform the following:

1. From figure 2 area 2, select the matcher and the level of similarity of the attribute names of the two schemas.
2. Click the “match” button

3. In figure 2 area 1, if the user has a ground truth file, then, he can input that file and compare UMAP's results to the ground truth by clicking the "compare to ground truth" button. Then, the correspondence table will show the correct correspondences in green and the wrong ones in pink, the same coloring scheme will appear in the visualization tab as well.
4. The button "correct correspondences" allows the user to correct all the wrong correspondences with respect to the ground truth at once.
5. After the user is done with the schema matcher, he should click "save and proceed" button to proceed to the schema mapping process.

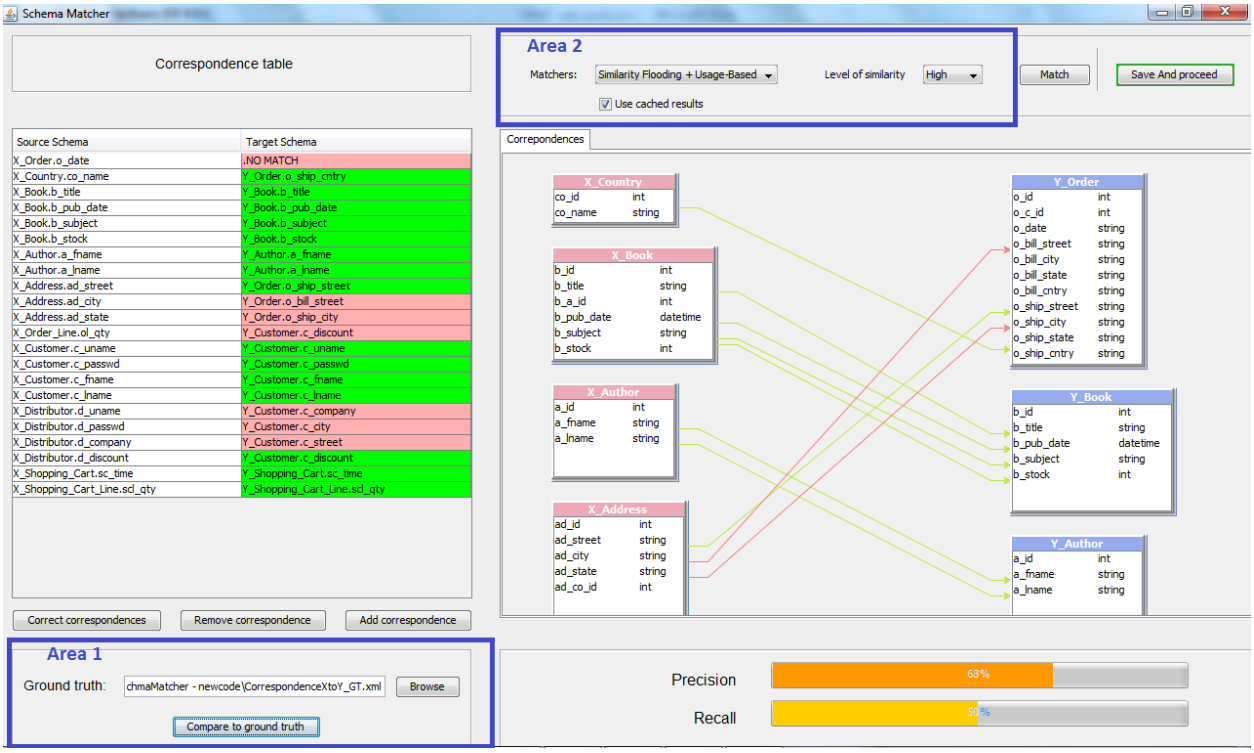

Figure 2. UMAP's schema matcher interface

2.3. Merging sibling relations

Next, the first step of schema mapping is "Merging sibling relations". As shown in figure 3 area 1, X_Customer_Parent is a new merged relation generated by UMAP. This new merged relation is the result of merging X_Customer and X_Distributor relations.

UMAP also shows the queries that supported the merging decision as shown in figure 3 area 2.

A merged relation can also be unmerged by selecting it and clicking the "Unmerge" button shown in Figure 3.

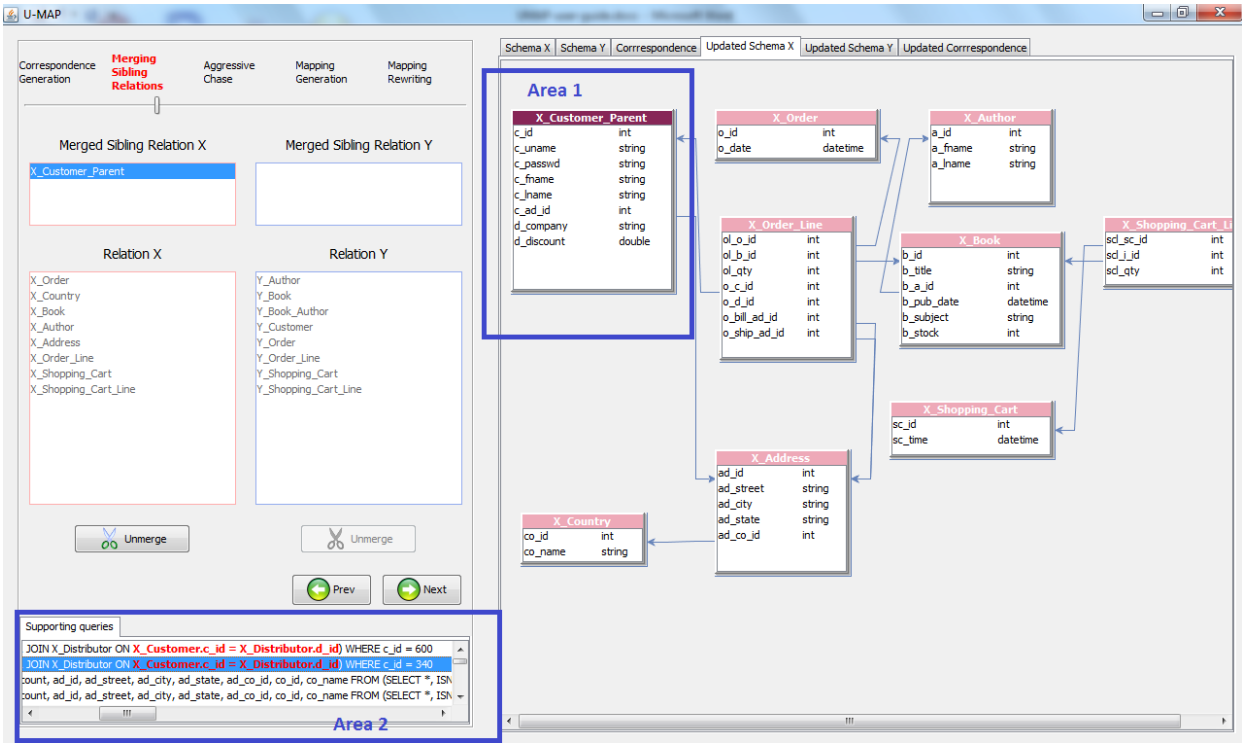

Figure 3. The “Merging sibling relations” step in UMAP.

2.4. Aggressive chase

After clicking Next, we get to the aggressive chase phase of UMAP. As shown in figure 4 area 1, for each generated logical relation in which the aggressive chase has been used UMAP shows the supporting queries and the opposing edges highlighted in pink (figure 4 area 2) between relations in which the aggressive chase has been applied.

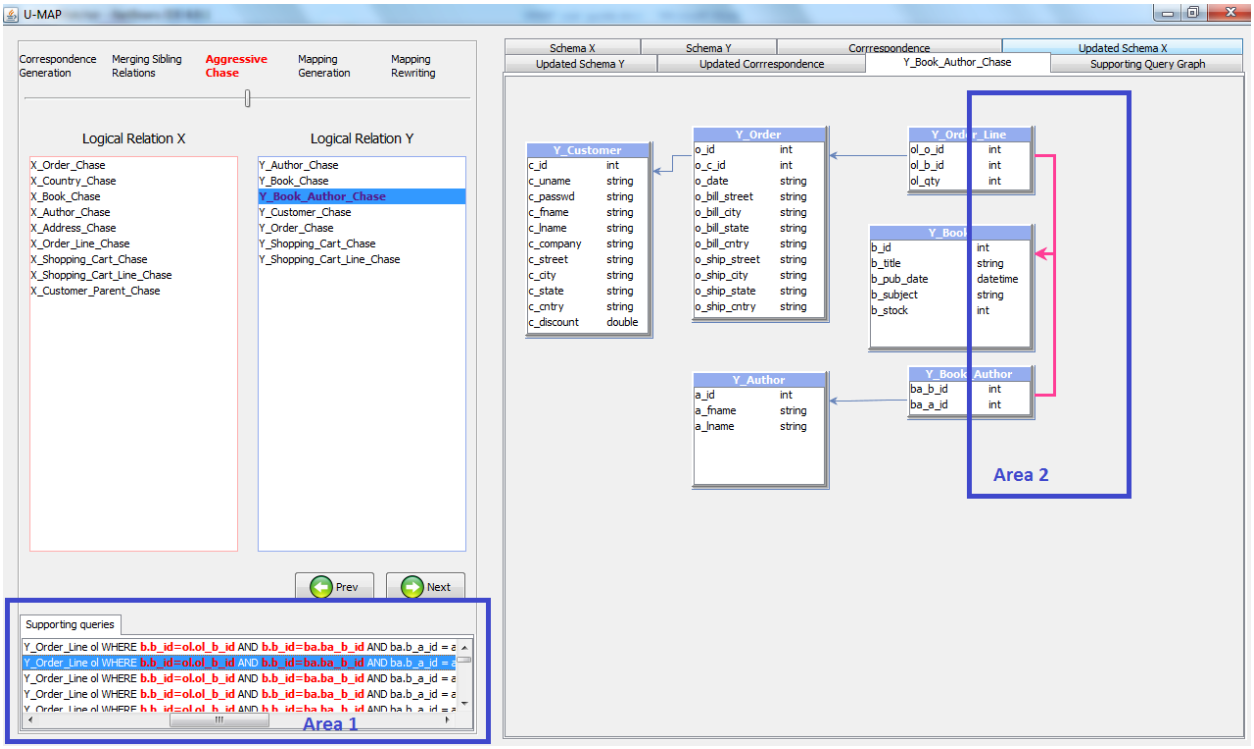

Figure 4. The “aggressive chase” step in UMAP.

2.5. Mapping generation

After clicking Next, we get to the mapping generation step in which mapping rules are generated and conflict resolutions are shown and then resolved. If a mapping rule has conflicts the user can click the “Show conflicts” button and then multiple resolutions for that conflict are shown in different colors in the visualization tab (figure 5 area 1) and the conflict is resolved upon clicking the “resolve conflicts” button.

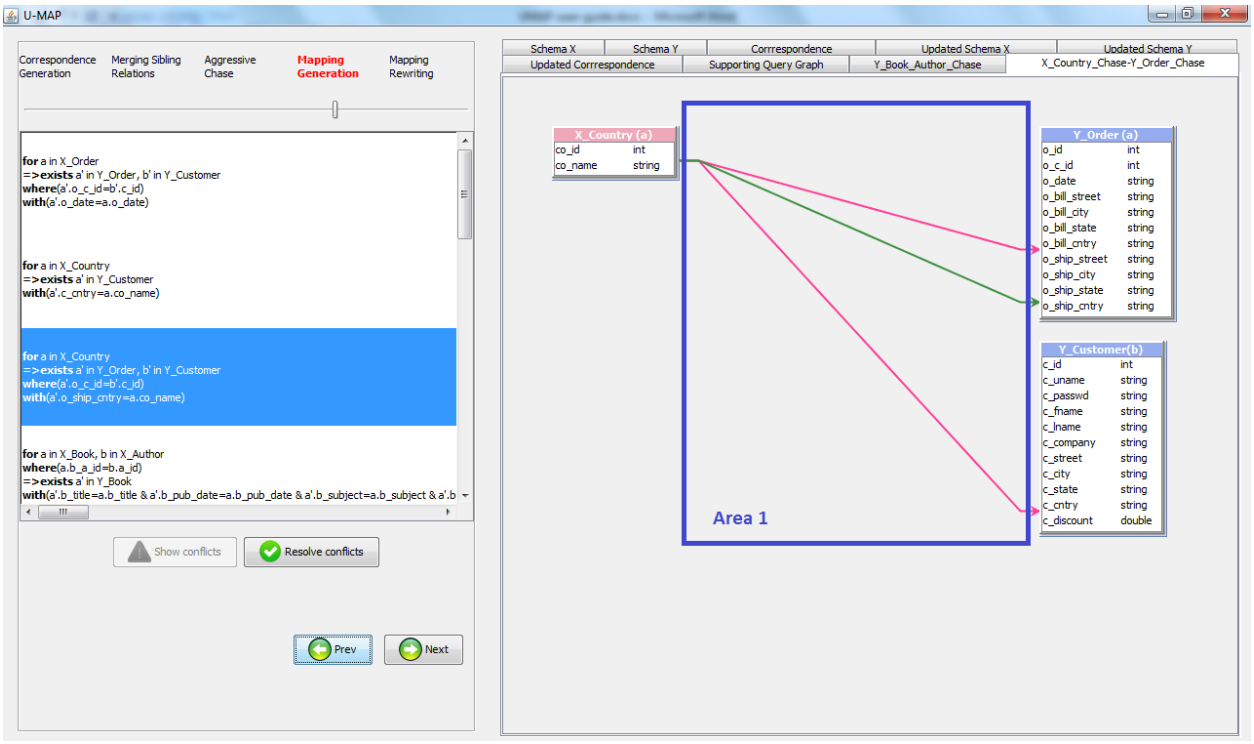

Figure 5. The “Mapping generation” step in UMAP. Multiple conflict resolutions are shown in different colors (area 1).

After clicking Next, we get to the mapping rewriting step in which the mapping rules are written using the original relations from the schemas without considering any merged relation generated by UMAP.