24.11.2011

SQL Lab: Assignment 2
(due to 13. / 14.12.2011)

Task

In the previous assignment each group modeled the PubMed digital library. Out of these models, we choose one which will be starting point for all of you and can be downloaded from our website (http://www.ifis.cs.tu-bs.de/teaching/ws-1112/sql). This solution can still have logical errors: Just ignore them! The task for this assignment is to integrate this model into the model of a chemical digital library (given below).

The aim of the integration is a global model for both libraries. In principle the integration consists of four steps (see RDB1, Lecture 4):
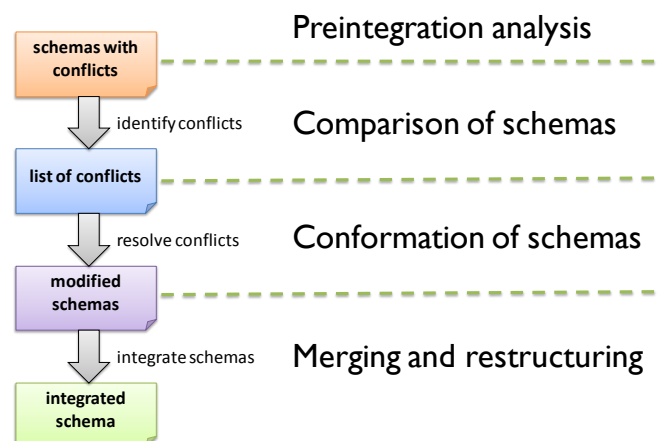
One of the most important steps is the merging and restructuring. It is very important to resolve all conflicts in the global model, e.g., naming conflicts. Of course, the resulting model should be well-arranged and comprehensible, e.g., think about using hierarchical structures with packages. Furthermore, it is important to explain all your design decisions and to document all your changes, e.g., what is renamed, merged classes, new classes, new attributes… Based on the bad experiences from last week, we advise you to **document your solution in a way that you tutor can understand it and that each of the four steps are considered**. The documentation of the given starting model is a good example. **If you do not document your solution you will fail this assignment!**

For your solution you have to use a modeling tool (for example Visual Paradigm from the last time). Handwritten models are not allowed. You have to send your solution **to your**

tutor via email before your tutorial group meeting. In addition, you must also hand in a **printed version to your tutor in the respective tutorial group**.

Models for integration

The first model can be downloaded from the website (http://www.ifis.cs.tu-bs.de/teaching/ws-1112/sql). This model is very well documented and there should not be a problem to understand it. If there are still questions, please contact your tutor! Close to reality, the following second model is not that well documented and structured. And again, if you have questions, please contact your tutor. A scalable SVG version of the second model can also be downloaded from the website.

Schema

In general: an **empty diamond** indicates that this attribute can be null. A **filled diamond** indicates that an attribute cannot be null. A turquoise diamond indicates a *normal* attribute whereas a **red diamond** indicates a foreign key for this attribute. And of course a **key** indicates a primary key.
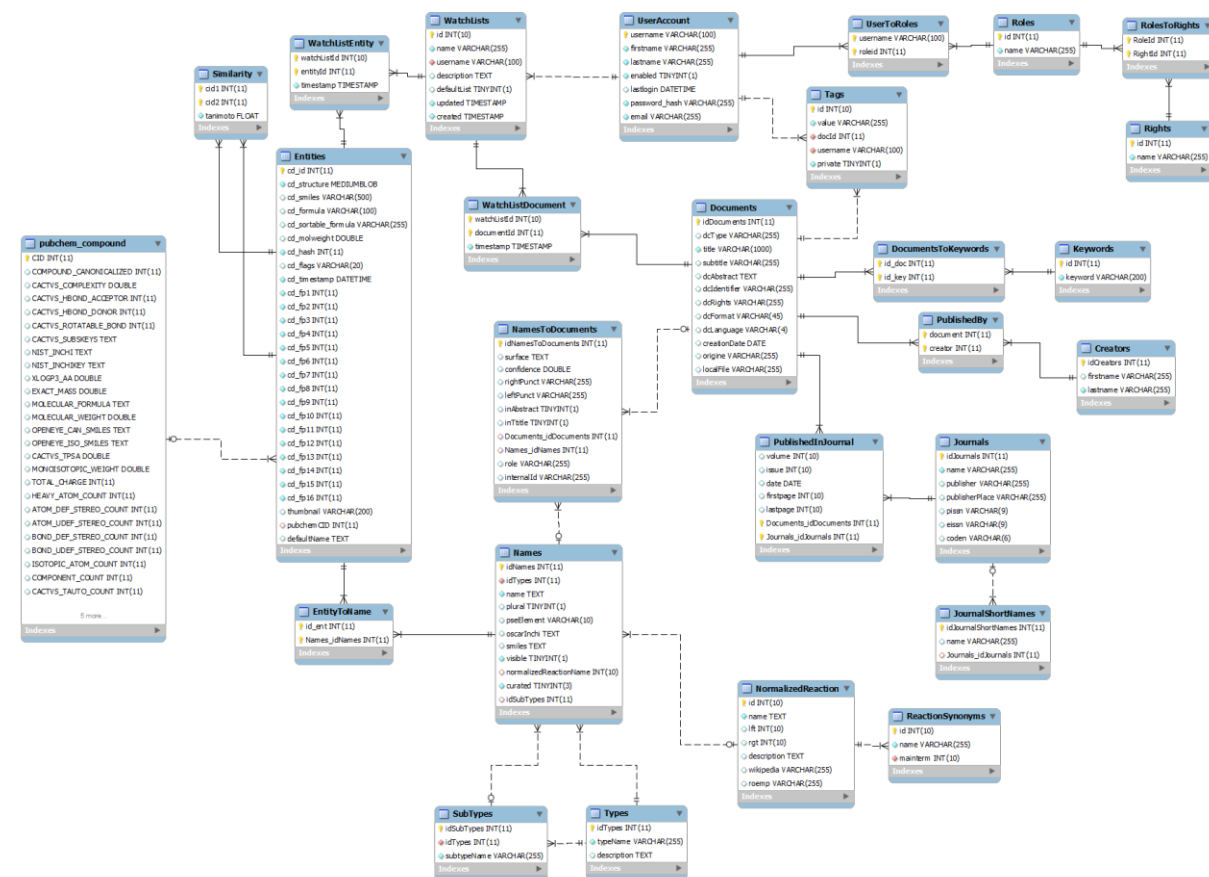
Documentation

- In the center of this schema we have our chemical **Documents.** In the attribute list, **dc** is always an acronym for Dublin Core.

- Each document can have several **Keywords** assigned.

- Each document is authored by some **Creators** and published in a **Journal**, which may have short names (**JournalShortNames**).

- Chemical documents always contain chemical entities, such as chemical reactions or substances. All of these **Names** are extracted in an automated way. The used algorithm calculates a **confidence** value for each extracted Name within the document and highlights if the name was found in the abstract or in the title of the document.

- If a names object is a reaction it may have a normalized reaction name (**NormalizedReaction**) and thus **ReactionSynonyms**.

- In addition a names object may contain to a type (**Types**) and respective to a subtype (**SubTypes**). Both are a predefined list, which is comparable to an enumeration in Java.

- If a names object is a chemical entity it is connected to the **Entitites** table. Each of the *cd_fp** attribute represents a part of a fingerprint representing this entity. In addition for each entity we may have a thumbnail and a default name.

- If an entity is listed in the PubChem database, it is a **pubchem_compound**, with a big list of additional attributes.

- Two entities are similar (**Similarity**) to each other with a specific tanimoto coefficient.

- A user may have a **UserAccount**. Each user can tag documents (**Tags**) and add documents to their watch lists. In addition, he / she can also add chemical entities to their watch lists.

- In respect to the **role** and the assigned **rights** a user may access different areas and methods within the chemical digital library.