

SuperIDR Code Details and Setup

Code details

There are two compressed files attached with this dissertation. The first is `murthy-u-d-2011-superidr-source-main`. It contains `TabletPC`, which is the SuperIDR source code root directory. `TabletPC` includes the following directories and files:

- `Families` - contains fish images and drawings pertaining to families of fishes
- `Figures` - contains fish images and drawings pertaining to the identification key
- `Descriptors`, `XMLDatabase` - contain files used by the CBISC, including CBISC descriptor information and feature vectors
- `index` - contains the Lucene index
- `app_logs` - contains log files
- `help` - contains the SuperIDR user manual files -

In addition to the `TabletPC` directory, `murthy-u-d-2011-superidr-source-main.zip` contains a database SQL file. The SQL file has database records for fish species information, including fish taxonomical classification information, key information, species descriptions, image information, and subimages and textual descriptions for 269 annotations. Figure 1 shows the database design of SuperIDR. Figure 1 shows the search, annotation, and species description tables and does not include tables that contain taxonomical classification and key information, which is used for browsing. The source code has been tested to compile on a Tablet PC Windows XP Professional operating system, Visual Studio 2008, and MySQL version 5.1.23.

The second compressed file attached with this dissertation is `murthy-u-d-2011-superidr-source-images.zip`. It contains the directory, `Images`, which in turn contains all the images used in SuperIDR. The `Images` directory should be placed in the SuperIDR source code root directory, `TabletPC`.

Source code packages

There are five packages, including classes responsible for specific functions, as described below.

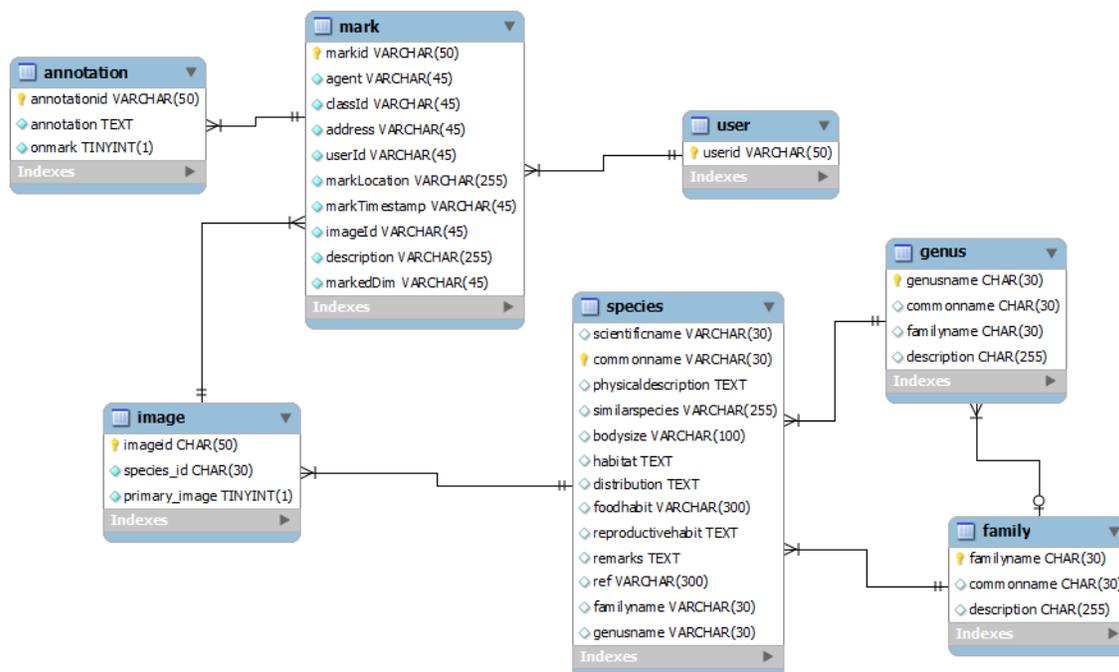


Figure 1: The database design of SuperIDR. Links between tables indicate a foreign key relationship. For example, image id is a foreign key in the mark table.

db

This package is responsible for connecting to the MySQL database. It includes classes that serve as templates for objects stored in the database as well as classes that connect to the MySQL database. It contains four groups of classes:

- Classes that hold information, which might be searched and browsed: These classes hold core information, which might be searched for, using text- and/or content-based image retrieval. Each class has a corresponding DB class to connect to the database. Includes classes for Annotation, ImageO (an image object, which is different from the Image MSDN object), Mark (holds subimage information), and Species.
- Classes that hold information, which might be browsed: These classes hold information, which might only be browsed, using the taxonomical, tree, or the key browser. Each class has a corresponding DB class to connect to the database. Includes classes for Family, Figure, Genus, and Part.
- Administrative classes to connect to the MySQL database: These classes manage server connections. They are DBManage and MySQLDAOFactory.
- Classes holding user information: These classes hold user information, which can neither be searched for or browsed.

dialog

The classes in this package are used for the diary entry interface and function. This functionality was used in the qualitative user study to elicit SuperIDR feedback from participants.

log

The *log* package is responsible for managing the logging of all user interactions with SuperIDR. Classes in the *ui* package invoke methods from this class to log user interactions, where required.

lucene

This package is responsible for full-text and field-wise indexing and search of annotations and species descriptions. The search interfaces invoke methods in classes in the *lucene* package to search for annotations and species descriptions that match the keywords entered by the user. The `FieldKeys` class is responsible for field-wise indexing and search. The `MySQLToLuceneSyncer` class includes methods to synchronize the search-able information (annotations and species descriptions) between the *mysql* database and the Lucene index.

ui

The *ui* package includes all user interface elements of SuperIDR. Also, it handles user interactions and invokes methods in the *db*, *lucene*, and other packages, as required. The `MainMenu` class is responsible for managing the main SuperIDR interface and is the entry point of all other functions. There are multiple forms and controls to handle annotation, species description, text search, image search, combined search, text search results, image search results, and comparison functions and displays.

Code setup

The following steps describe the process of setting up the code.

1. Download and install the following software
 - Visual studio 2008¹
 - MySQL² (setup with user id=root, Database=tablet, and password=si, or change in `App.config` in the source code to reflect the set user id and password)
 - MySQL GUI tool/Workbench for a GUI-based interface to manage the database
2. Restore the database from the SQL file “superidr_v20_images_marks_inkspace 201010250440.sql”.
3. Download and unzip source code file, `superidr-source-v24.zip`, to a directory on your computer, say `PROJECT-HOME`.
4. Make sure that the following `.dll` files are there in the source code:

¹<http://msdn.microsoft.com/en-us/vstudio/aa700830>

²<http://www.mysql.com/>

- Tablet PC SDK
 - Log4net
 - Lucene.Net
 - MySQL.Data (MYSQL Connector)
 - CBISC
 - BIC (in PROJECT-HOME/TabletPC/DLLs/)
5. Open PROJECT-HOME/TabletPC/SuperIDR.sln in Visual Studio 2008.
 6. Open the App.config file and edit the parameters mysqlString, appRootFolder, familyImageRootFolder, figureRootFolder to reflect the correct path on your machine. Make similar changes in the log4net.config file.
 7. Compile and run the code in Visual Studio 2008.

To reindex the images and annotations, delete cbisc.pid and luceneSpecies.pid in PROJECT-HOME/TabletPC/. This process might take a few hours depending on the number of images.

Copyright and licenses

The SuperIDR code, packaged along with the dissertation, has been developed by students at Virginia Tech and by students at UNICAMP, Brazil. The annotation, text search, image management, combined search, comparison, logging functions, and the integration code of SuperIDR have been developed by Uma Murthy at Virginia Tech. The browse function has been developed by Evandro Ramos and Tiago Falcao at UNICAMP, Brazil. The content-based image search function has been jointly developed by all students. This function references the CBISC.dll library, which was developed by students at UNICAMP. The copyright header files in the code indicate the developers of each class.

The source code is shared under the GNU General Public License³. Line drawings and images of live fishes that appear in SuperIDR, have been taken from the book, “Freshwater fishes of Virginia” [2] and are copyright of the book authors and the American Fisheries Society. Images of preserved specimens are shared under the Creative Commons [1] share-alike attribution license⁴.

³<http://www.gnu.org/licenses/gpl.html>

⁴<http://creativecommons.org/licenses/by-sa/3.0/legalcode>

Bibliography

- [1] “Creative commons,” 2011. [Online]. Available: <http://creativecommons.org/>
- [2] R. E. Jenkins and N. M. Burkhead, *Freshwater Fishes of Virginia*. Bethesda, MD: American Fisheries Society, 1994.