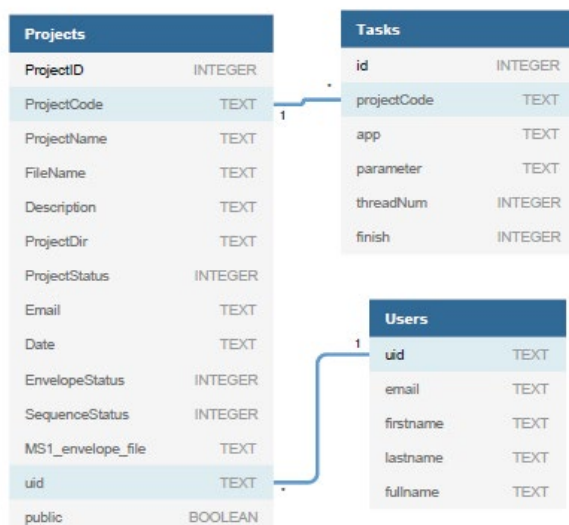


Database Design

1. Project Database



Each Topview node.js app will have one SQLite database for storing projects information, registered users information and tasks information in task scheduler. The SQLite database file should be under top_view_app/db directory. This database contains three tables, Projects, Tasks and Users.

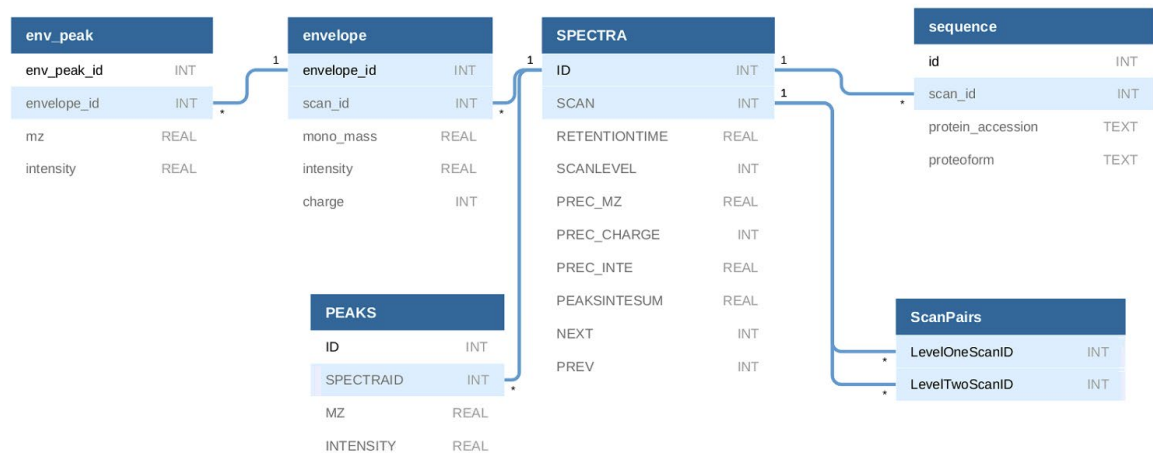
Projects table contains all the information of each projects. ProjectID is a primary key of Projects table, which is usually for internal use. ProjectCode is also unique for each project and it is always for external reference instead of ProjectID. ProjectName is a name assigned by users who create the project, which can be changed anytime in project management page. FileName is usually the file name of mzML file uploaded by user. Description is a short paragraph for users to describe the corresponding project, which can also be changed anytime in project management page. ProjectDir is a directory to store all files of corresponding project, which is important for topview app to locate files. ProjectStatus is a integer code to indicate project status, code 0 means processing, code 1 means success, code 2 means removed, code 3 means failed, code 4 means waiting. Email is an email address from upload user, which is used to send notifications of project status change. Date is timestamp in UTC time zone, which is created when project uploaded by user at the first time. EnvelopeStatus is also a status code, which indicates whether this project contains envelopes information. Code 0 means no envelope exists, code 1 means it contains envelope information. Envelope information can be either uploaded by users or generated by TopFD task automatically. SequencesStatus is a status code which indicates whether the project contains sequence information. Code 0 means it doesn't contain sequence information, and code 1 means it contains sequence information. Sequence information can be uploaded by users or generated by Toppic task automatically. MS1_envelope_file is a optional file during user creating a project. Users can choose to use env file to import envelope information instead of uploading msalign files later. Uid is a foreign key

which refers to `Users.uid`, it helps `topview` to connect projects with users. `Public` is a Boolean value to indicate if the project is public or not.

Tasks table contains all tasks in topview task scheduler. Id is the primary key of Tasks table. ProjectCode is a foreign key which refers to Projects.ProjectCode. It helps task scheduler find files which the task will need. App is a text value to indicate which app the task will run, for example, it can be node or cpp app. Parameter is also a text value for running task app, it contains how user configures task. ThreadNum is a integer value which tells how many thread the task will use, it can help task scheduler assign available resources to tasks.

Users Table contains all registered users on topview app. Users information will be inserted into this table when they first log into topview. Users table helps topview give users access to their own private projects and give guests access to public projects without privilege to modify data.

2. Dataset Database



Each dataset from user will generate a SQLite database. This SQLite database file should be located inside each project folder, for example, /data/XXX/XXX.db. There are three tables that all datasets will have, SPECTRA, PEAKS and ScanPairs. Env_peak and envelope tables will only appear after users upload envelope information files or run TopFD on server. Sequence table will only appear after users upload sequence information files or run Toppic on server.

SPECTRA table shows the SPECTRA data from mzML file. In SPECTRA table, ID is the table's primary key. SCAN, RETENTIONTIME, SCANLEVEL, PREC_MZ, PREC_CHARGE, PREC_INTE, PEAKSINTESUM are directly read from original mzML file. SCAN is a unique number for each scan, which is usually the same as SPECTRA ID. NEXT and PREV is generated by mzML process app, which can tell spectra's next or previous spectra's ID. NEXT and PREV are usually used for navigating on frontend.

PEAKs table shows the peaks information from mzML file. It has ID as its own primary key. SPECTRAID is a foreign key which refers to SPECTRA.ID. MZ and INTENSITY are values read directly from original mzML file.

ScanPairs table shows the relationship between level one scan and level two scan. It helps user to know corresponding scan level two and scan level one information. For example, if I want to know all scan level two of specific scan, I can loop up this table to get the information.

Sequence table shows sequence data of one spectra. It is an optional table and it only shows up after importing sequence data from users. It uses id as its own primary key. Scan_id is a foreign key referring to SPECTRA.ID. Protein_accession and proteoform are directly read from sequence files, which is always a tsv file.

Envelope table shows envelope data of spectra. It is an optional table, and it only shows up after importing envelope data from users. It has envelope_id as its primary key and scan_id is a foreign key referring to SPECTRA.ID. Mono_mass, intensity and charge are values read from envelope files which may be uploaded or generated by TopFD.

Env_peak table stores envelope peaks data. It is an optional table and it only appear after importing envelope data from users. Envelope_id is a foreign key referring to envelope.envelope_id. Mz and intensity are calculated based on envelope information such as mono_mass, charge, and intensity, also with a corresponding peaks list.