DB Schema

Dashamir Hoxha

10 March 2012

Contents

1 DB Schema

Files, Templates, Locations and Projects are related to the import/export of the PO files.

Projects and Categories can be used to limit the scope of the search (and other operations).

A project contains the translations of a certain application (software). A project can have several template (POT) files. A template file can have several PO files (one for each different language). Each of these PO files has many PO entries, which are stored in the table Locations.

The table Locations stores only the comments, line references, flags, previous strings, etc. of each PO entry.

The msgid (and msgctxt) of the entry is stored on the table Strings. A string can be connected to several locations, since the same string can be used on different projects.

Each string can have several translations (or suggestions) in each language. Each translation can have many votes. Each vote is given by a certain user.

The DB tables and their fields:

Files A PO file that is imported and can be exported from the DB.

    fid : serial Auto-increment internal identifier.

filename : varchar(250) The path and filename of the imported PO file.

hash : char(40) The SHA1() hash of the whole file content.

potid : int Reference to the template (POT) for which this PO file is a translation.

lng : varchar(10) The code of the translation language.

headers : text Headers of the imported PO file, as a long line. Needed mainly for exporting.

comments : text Translator comments of the file (above the header entry). Needed mainly for exporting.

uid : int Id of the user that imported the file.

time : datetime The date and time that the record was registered.

Templates POT files that are imported.

potid : serial Auto-increment internal identifier.

tplname : varchar(50) The name of the POT template (to distinguish it from the other templates of the same project).

filename : varchar(250) The path and name of the imported POT file.

pguid : char(40) Reference to the project to which this PO template belongs. it come from).

uid : int(11) Id of the user that registered the project.

time : datetime The date and time that the template was imported.

Projects A project is the software/application which is translated by the PO files.

pguid : char(40) Project Globally Unique ID, pguid = SHA1(CONCAT(origin,project))

project : varchar(100) Project name (with the release appended if needed).

origin : varchar(100) The origin of the project (where does it come from).

uid : int(11) Id of the user that registered the project.

time : datetime The date and time that the project was registered.

Locations Locations (lines) where a l10n string is found.

> **lid : serial** Internal numeric identifier of a line.
>
> **sguid : char(40)** Reference to the id of the l10n string contained in this line.
>
> **potid : int** Reference to the id of the template (POT) that contains this line.
>
> **translator_comments : varchar(500)** Translator comments in the PO entry (starting with "# ").
>
> **extracted_comments : varchar(500)** Extracted comments in the PO entry (starting with "#. ").
>
> **line_references : varchar(500)** Line numbers where the sting occurs (starting with "#: ").
>
> **flags : varchar(100)** Flags of the PO entry (starting with "#, ").
>
> **previous_msgctxt : varchar(500)** Previous msgctxt in the PO entry (starting with "#| msgctxt ").
>
> **previous_msgid : varchar(500)** Previous msgid in the PO entry (starting with "#| msgid ").
>
> **previous_msgid_plural : varchar(500)** Previous msgid_plural in the PO entry (starting with "#| msgid_plural ").

Strings Translatable strings that are extracted from projects.

> **string : text** The string to be translated: CONCAT(msgid,CHAR(0),msgid_plural)
>
> **context : varchar(500)** The string context (msgctxt of the PO entry).
>
> **sguid : char(40)** Globally Unique ID of the string, as hash of the string and context: SHA1(CONCAT(string,context))
>
> **uid : int** ID of the user that inserted this string on the DB.
>
> **time : datetime** The time that this string was entered on the DB.
>
> **count : int/tiny** How often this string is encountered in all the projects. Can be useful for any heuristics that try to find out which strings should be translated first.
>
> **active : boolean** The active/deleted status of the record.

Translations Translations/suggestions of the l10n strings. For each string there can be translations for different languages, and more than one translation for each language.

sguid : int Reference to the id of the l10n string that is translated.

lng : varchar(5) Language code (en, fr, sq_AL, etc.)

translation : varchar(1000) The (suggested) translation of the string.

tguid : char(40) Globally Unique ID of the translation, defined as the hash: SHA1(CONCAT(translation,lng,sguid))

count : int/tiny Count of votes received so far. This can be counted on the table Votes, but for convenience is stored here as well.

uid : int id of the user that initially suggested/submitted this translation

time : datetime Time that the translation was entered into the database.

active : boolean The active or deleted status of the record.

Votes Votes for each translation/suggestion.

vid : serial Internal numeric identifier for a vote.

tguid : char(40) Reference to the id of the translation which is voted.

uid : int Reference to the id of the user that submitted the vote.

time : datetime Timestamp of the voting time.

active : boolean The active or deleted status of the record.

Users Users that contribute translations/suggestions/votes.

uid : int The numeric identifier of the user.

points : int Number of points rewarded for his activity.

config : varchar(250) Serialized configuration variables.

Snapshots Snapshots are tgz archives of project-lng translation files.

pguid : char(40) Reference to the project.

lng : varchar(10) The language of translation.

snapshot : mediumblob The content of the tgz archive.

uid : int Id of the user that updated the snapshot for the last time.

time : datetime The time of last update.

Diffs Diffs between the current state and the last snapshot.

pguid : char(40) Reference to the project.

lng : varchar(10) The language of translation.

nr : smallint Incremental number of the diffs of a project-language.

diff : text The content of the unified diff (diff -u).

ediff : text The embedded diff (generated with the command poediff of pology).

comment : varchar(200) Comment/description of the diff.

uid : int Id of the user that inserted the diff.

time : datetime The date and time that the diff was saved.
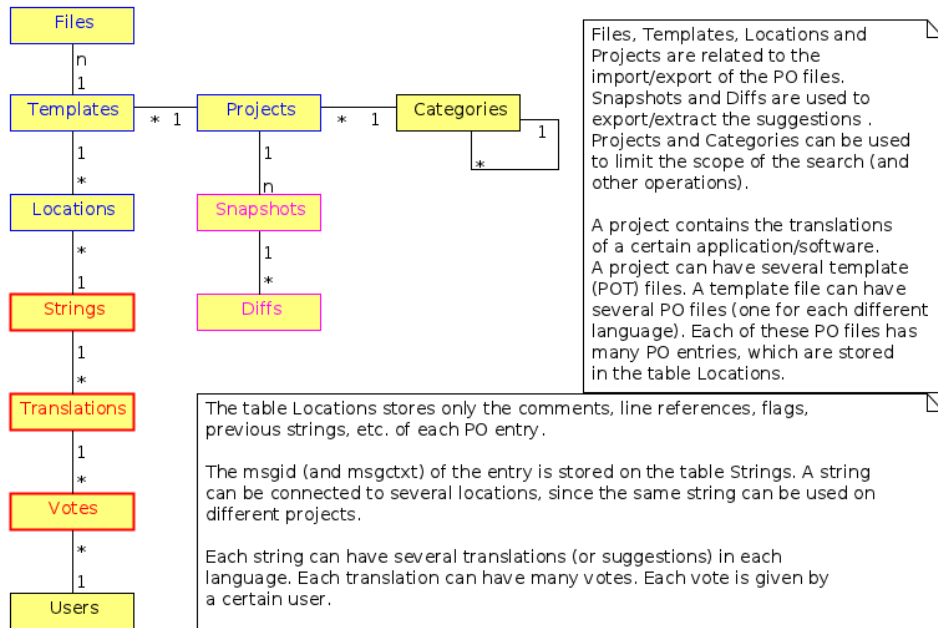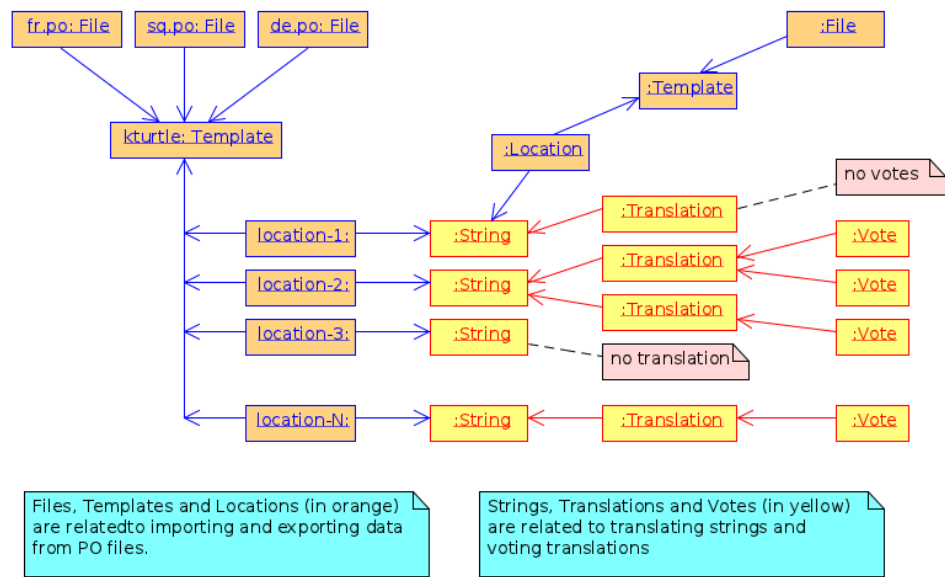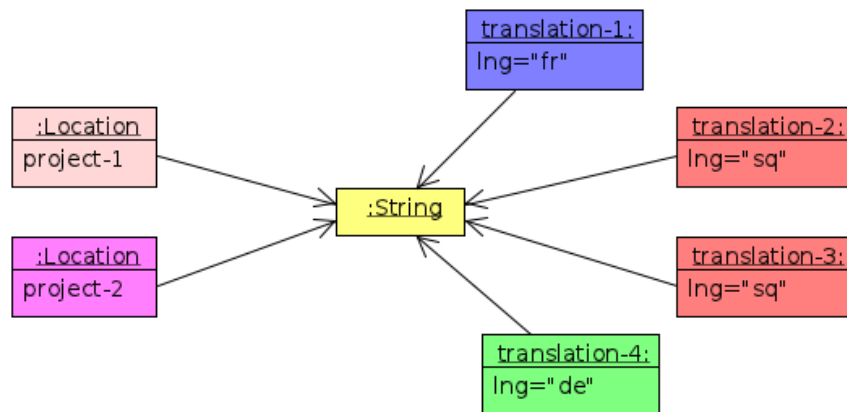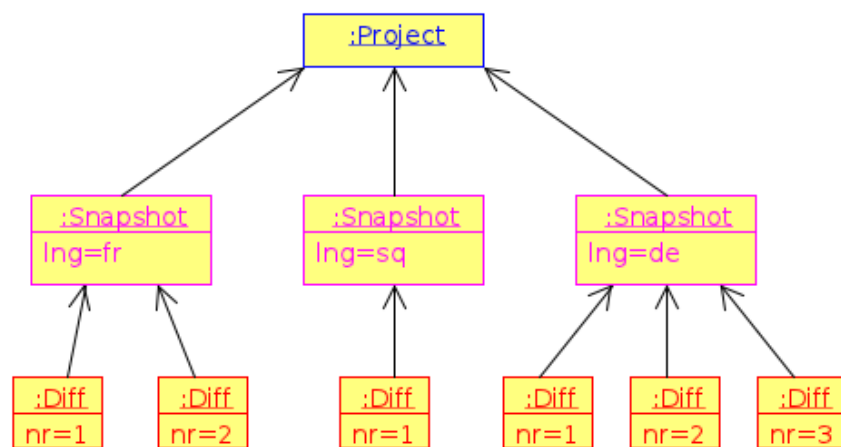

Figure 1: Tables and their relations.

Figure 2: Structure of the DB.

Figure 3: Structure of the DB.

A snapshot is an export of the PO files of a project-language.
A project has a snapshot for each language.
A diff is the difference between a snapshot and the last one.
The diffs are sequentially numbered and they keep the
history of changes.

Figure 4: Structure of the DB.