

### ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE Fakulta stavební Katedra geomatiky, předmět Free Software GIS Thákurova 7, 166 29 Praha 6

# TECHNICKÁ DOKUMENTACE

Název: ČTI OS

**Účel:** Skript posílá požadavek na službu ČTI OS na základě pole pseudonymizovaných oprávněných

subjektů (POSIdentů) a ukládá odpověď do SQLITE databáze.

Datum: červenec 2019

Copyright: (C) 2019 Linda Kladivová

Email: <a href="mailto:l.kladivova@seznam.cz">l.kladivova@seznam.cz</a>

#### Úvod

Tento skript byl zpracován v rámci předmětu 155yfsg 2019 na fakultě stavební ČVUT a slouží k deanonymizaci VFK souboru. V budoucnu bude na jeho základě vytvořena knihovna, která bude zakomponována do VFK pluginu v softwaru QGIS. V případě dotazů, nejasností či návrzích na vylepšení se obracejte na uvedenou emailovou adresu.

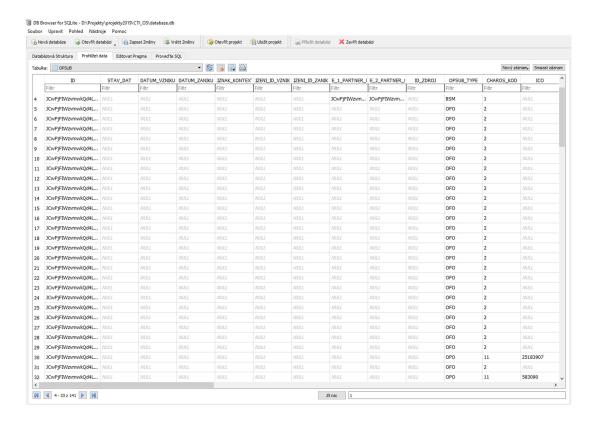
#### Návod ke skriptu

Pro správný chod skriptu je nutno si připravit několik souborů. Zaprvé textový soubor s pseudonymizovanými oprávněnými subjekty s názvem posidents.txt. Zadruhé soubor request.xml s přístupovým jménem a heslem do služby ČTI OS. Zatřetí databázi Export\_vse.db. Vzorové ukázky souboru request.xml a posidents.txt jsou součástí projektu zde na GitHubu, databáze s koncovkou .db se z VFK souboru získá následujícím postupem:

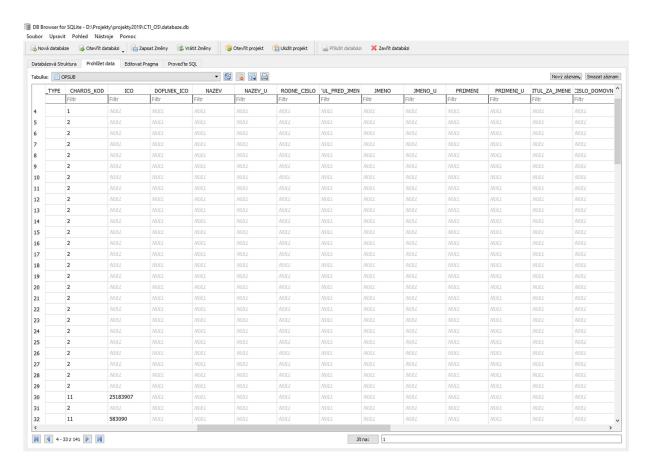
Je třeba mít k dispozici free software QGIS a to konkrétně jednu z Standalone verzí. Možnosti stažení jsou na této webové stránce: https://qgis.org/en/site/forusers/download.html. Společně se stažením QGISu se stáhne i GRASS a SAGA. Co je ale v tuto chvíli nejpodstatnější, je, že se stáhne také příkazová řádka OSGeo4W Shell. Pomocí této příkazové řádky získáme v dalším kroku z VFK souboru databázový soubor s koncovkou db. Práce v této příkazové řádce je znázorněna na **Obrázku 1**. Je nutné se přesunout do adresáře s VFK souborem a pak pustit příkaz ogrinfo Export\_vse.vfk OPSUB (se správným názvem souboru). V tuto chvíli se ve stejném adresáři vytvoří nový soubor, v mém s koncovkou db, který je pro správný běh skriptu nutné si přejmenovat na Export\_vse.db. Tento databázový soubor si můžete prohlédnout nejlépe v softwaru DB Browser (SQLite), který je volně ke stažení na této webové adrese: https://sqlitebrowser.org/dl/. Z celé databáze nás pro deanonymizaci údajů zajímá pouze tabulka OPSUB, která v nyní neobsahuje žádné osobní údaje, viz **Obrázek 2, 3**.

Pokud se chcete o VFK souboru dozvědět další informace, je možné použít výše zmíněný příkaz s dalšími přepínači, jako například: ogrinfo Export\_vse.vfk OPSUB –so (vypíše souhrnné informace o souřadnicovém systému, počtu prvků, rozsahu apod.), ogrinfo Export\_vse.vfk OPSUB -fid 1 (vypíše informace o prvním prvku), ogrinfo Export\_vse.vfk OPSUB –layer (vypíše informace o vrstvě).

Obrázek 1: Vytvoření databáze ze souboru VFK pomocí příkazové řádky OSGeo4W Shell

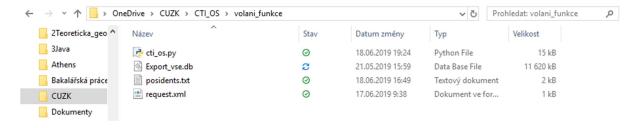


Obrázek 2: Tabulka OPSUB v databázovém prostředí DB Browser for SQLite před aktualizací ČTI OS skriptem



Obrázek 3: Tabulka OPSUB v databázovém prostředí DB Browser for SQLite před aktualizací ČTI OS skriptem

Aby skript fungoval správně, je nutno mít všechny 3 výše vyjmenované soubory ve stejném adresáři jako skript cti\_os.py. Příklad správného sestavení vstupních dat je uveden na následujícím **Obrázku 4**.



Obrázek 4: Struktura adresáře před vykonáním skriptu

## Funkcionalita skriptu

Nejprve se vytvoří LOG soubor (funkce create\_log\_file), do kterého se budou později vypisovat chybné pseudonymizované oprávněné subjekty (dále používejme označení *posidenty*) a statistiky o počtu správně zpracovaných posidentů a o počtu konkrétních chyb: NEPLATNY IDENTIFIKATOR, EXPIROVANY IDENTIFIKATOR, OPRAVNENY SUBJEKT NEEXISTUJE. LOG soubor se vytvoří ve stejném adresáři, jako jsou všechny výše jmenované soubory.

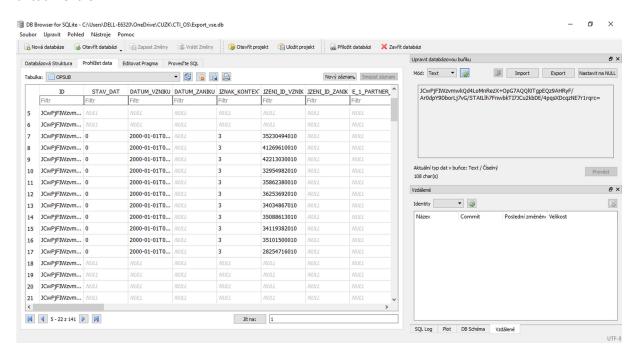
Dále se otevře soubor s polem posidentů a zkontroluje se, zda nemá žádné duplicity, pokud ano odstraní se (funkce remove\_duplicates). Pokud bude v poli více než 100 posidentů, je dotaz na službu rozdělen do více dalších dotazů. Pokud je počet posidentů menší než 100, nic se nerozděluje.

Vzhledem k tomu, ze jména atributů v XML dokumentu v několika případech není možné jednoduše převést na jmena sloupců v databázi, bylo nutné pro tyto speciální případy definovat převodní slovník.

Podle pole posidentů se sestaví XML a to ve funkci draw\_up\_xml\_request (maximálně pro 100 posidentů). Vstupem do teto funkce je cesta k souboru request.xml, ve kterém se může změnit jméno a heslo, ale žádné posidenty se nevyplňují, ty jsou obsahem souboru posidents.txt. Další funkce call\_service zavolá službu CTI OS a služba vrátí odpověď, která bude poměrně rozsáhlá - velký XML soubor, který bude obsahovat atributy k jednotlivým posidentům. Je nutné, aby HTTP status kód byl 2xx, což je ve funkci také kontrolováno.

Dále byly vytvořeny dvě funkce transform\_names a transform\_names\_dict na konverzi jmen atributů v XML dokumentu na názvy sloupců v databázi. První funkce se zabývá jednodušším případem konverze, druhá využívá speciální převodní slovník. V další funkci save\_attributes\_to\_db se postupně prochází všechny podatributy v atributu os a za pomoci funkce na konverzi jmen se vkládají postupně všechny tyto podatributy do databáze. Hned na začátku for cyklu je atribut os kontrolován, jestli neobsahuje podatribut chybaPOSIdent, který může obsahovat hesla: NEPLATNY IDENTIFIKATOR, EXPIROVANY IDENTIFIKATOR, OPRAVNENY SUBJEKT NEEXISTUJE. Pokud se chyba najde, tak se vypíše chyba (podle situace) do log souboru a zbytek for cyklu se přeskočí a jde se na další posident. V databázi se ještě vytvoří sloupeček OS\_ID, do kterého se vloží podatribut osld. Takto se for cyklus vykoná postupně pro každý atribut os v XML souboru (tedy tolikrát, kolik je posidentu v zadaném seznamu).

Na následujícím **obrázku 5** je databázová struktura po vykonání skriptu. Několik posidentů se aktualizovalo.



Obrázek 5: Databáze po aktualizaci v databázovém prostředí DB Browser for SQLite

Příklad log souboru je vidět na následujícím **obrázku 6**.

```
ð
   logfile_07_53_05_19_06_2019.log.txt – Poznámkový blok
 Soubor Úpravy Formát Zobrazení Nápověda
96-19 07:53 INFO Puvodní pocet
                                                                                                                                          Cobrazem Napoveda
Posidentu v seznamu: 12
Delka seznamu POSIdentu po odstraneni duplicitnich zaznamu: 12
Pocet duplicitnich zaznamu: 0
Maximalni pocet POSIdentu v ramci jednoho pozadavku na sluzbu: 100
STAVOVY KOD HTTP 2xx: SUCCESSI
Radky v databazi u POSIdentu JCwPjFlWzvmwkQd4LoMnRezX+OpG7AQQ10TgpEQz9AHRyF/Ar0dpY9DborLj7vG/STAtLlh7FnwbkTI7JCu2kbDE/4pqsXDcqzNE7r1rqrc=
   06-19 07:53 INFO
   06-19 07:53 INFO
 06-19 07:53 INFO
06-19 07:53 INFO
06-19 07:53 INFO
06-19 07:53 INFO
aktualizovany
06-19 07:53 INFO
 nadky v databazi u POSIdentu

JCwPjFINzvmwkQd4LoMnRezX+OpG7AQQ10TgpEQ29AHRyF/Ar0dpY9DborLj7vG/Y6Tz5cyLC+k2K2sRnLgxhtqWYar0DFKSnqkKCuUEsrg= aktualizovany

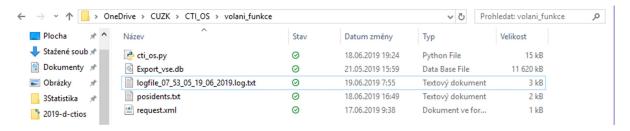
06-19 07:53 INFO Radky v databazi u POSIdentu
JAMFJFIAWZVMWKQd4LOMNKeZX+OpG7AQQ18TgpEQ:9AHRyF/Ar0dpY9DborLj7vG/Y6Tz5cyLC+k2K2sRnLgxhtqWYar0DFKSnqkKCuUEsrg= aktualizovany
06-19 07:53 INFO Radky v databazi u POSIdentu
JCwPjFIWzvmwkQd4LOMnRezX+OpG7AQQ18TgpEQ:9AHRyF/Ar0dpY9DborLj7vG/sBkU4/Re91u/W+d1ev6I7dc4/ZzqEBFYvako7xmUu4U= aktualizovany
06-19 07:53 INFO Radky v databazi u POSIdentu
JCwPjFIWzvmwkQd4LOMnRezX+OpG7AQQ18TgpEQ:9AHRyF/Ar0dpY9DborLj7vG/ztdfKCaIntTWQNhdRJ/6itaxruQMc7df7eUQoskngC4= aktualizovany
06-19 07:53 INFO Radky v databazi u POSIdentu
JCwPjFIWzvmkQd4LoMnRezX+OpG7AQQ18TgpEQ:9AHRyF/Ar0dpY9DborLj7vG/RIE8id0peyYvmoNi15SoUjfnGR0L6dntbREQOPSoBvw= aktualizovany
06-19 07:53 INFO Radky v databazi u POSIdentu
JCwPjFIWzvmkQd4LoMnRezX+OpG7AQQ18TgpEQ:9AHRyF/Ar0dpY9DborLj7vG/RIE8id0peyYvmoNi15SoUjfnGR0L6dntbREQOPSoBvw= aktualizovany
06-19 07:53 INFO Radky v databazi u POSIdentu
06-19 07:53 INFO
CROWNER OF A CONTROLL OF A 
 06-19 07:53 INFO Radky v databazi u POSIdentu POSIdentu Rozavany 06-19 07:53 INFO Radky v databazi u POSIdentu Rozavany 06-19 07:53 INFO Radky v databazi u POSIdentu Rozavany 06-19 07:53 INFO Radky v databazi u POSIdentu Rozavany 06-19 07:53 INFO Radky v databazi u POSIdentu Rozavany 06-19 07:53 INFO Radky v databazi u POSIdentu Rozavany 06-19 07:53 INFO Radky v databazi u POSIdentu Rozavany 06-19 07:53 INFO Rozavany v ramci 1 pozadavku Pozavany 06-19 07:53 INFO Rozavany v ramci 1 pozadavku Pozavany 06-19 07:53 INFO Rozavany v ramci 1 pozadavku Pozavany 07:00 Rozavany 07:00 Roza
                                                                                                                                              Zpracovano v ramci 1 pozadavku
Pocet uspesne stazenych posidentu: 12
   06-19 07:53 INFO
                                                                                                                                              Neplatny identifikator: 0x.

Opravneny subjekt neexistuje: 0x.

Expirovany identifikator: 0x.
   06-19 07:53 INFO
   06-19 07:53 INFO
   06-19 07:53 INFO
```

Obrázek 6: Log soubor informacemi a statistikami o průběhu skriptu

Struktura adresáře po vykonání skriptu je zřejmá z obrázku 7:



Obrázek 7: Struktura adresáře po vykonání skriptu