

1 Zadání

Úkolem bylo vytvořit skript v jazyce Python, který na základě vstupního XML souboru vytvoří sadu SQL příkazů pro vkládání tabulek do databáze. Program obsahuje několik podprogramů a používá dostupné knihovny `argparse` a `XML.dom.minidom`.

2 Implementace

Celý program je rozdělený do menších celků, které budou popsány níže.

2.1 Zpracování parametrů

Zpracování argumentů příkazové řádky zajišťuje funkce `parseArgs()` za pomoci knihovny `argparse`. Ve funkci jsou rovněž ošetřeny všechny chybové stavy, které mohou při zadávání argumentů nastat.

2.2 Práce se soubory

Manipulace se vstupem a výstupem je zajištěna přímo ve funkci `main()`, zároveň bylo třeba ošetřit práci s případným vstupním souborem pro rozšíření, to se děje ve funkci `validation()`, která bude rozebrána dále.

2.3 Zpracování XML

Zpracování vstupu probíhá ve funkci `parse_xml()` prostřednictvím knihovny `XML.dom.minidom`. Ta prochází vstup po jednotlivých uzlech a ukládá získanou strukturu elementů, podelementů a atributů do slovníků `all_tables` a `all_attrs`. V prvním zmíněném jsou zaznamenány názvy tabulek a v nich počty stejnojmenných podelementů. Ve druhém slovníku jsou uloženy atributy a textové podelementy včetně svých datových typů. V případě, že je zadán přepínač `-a`, do slovníku `all_attrs` ukládáme pouze textové podelementy a jejich datové typy a zpracování atributů ignorujeme.

2.4 Přepínač `--etc`

Pokud je zadán přepínač `--etc=n`, hned po vytvoření struktury vstupních dat dochází k volání funkce `my_etc()`. Tato funkce má za úkol kontrolovat, zda číslo zadané u parametru je nižší než počet podelementů se stejným jménem ve struktuře. Pokud se tak stane, musí dojít k záměně cizího klíče – tedy, pokud bylo v tabulce *A* několik odkazů (více, než je zadáno parametrem `--etc=n`) do tabulky *B*, nyní bude v tabulce *B* 1 odkaz do tabulky *A*. Přepínač `--etc=n` nesmí být zadán zároveň s přepínačem `-b`, neboť `-b` z více podelementů stejného názvu vytváří pouze 1 sloupec tabulky.

2.5 Kontrola názvů

Kontrolu konfliktů v názvech sloupců tabulky zajišťuje funkce `conflict()`. Kontroluje konflikty primárních klíčů, cizích klíčů a atributů. Pokud je v tabulce více stejnojmenných podelementů a není zadán přepínač `-b`, ještě před kontrolou data projdou funkcí `set_counter()`, která tyto podelementy očíslovuje.

2.6 Vztahy mezi tabulkami

Generování vztahů mezi jednotlivými tabulkami obstarává funkce `generate_rel()`. Tato funkce nejdříve uloží všechny vztahy zjištěné z cizích klíčů do pomocného slovníku `tables_g` a následně vyhledává vztahy se všemi ostatními tabulkami. Vztahy mezi tabulkami jsou tranzitivní a symetrické, tedy pokud

jsou propojeny tabulky A a B i tabulky B a C , existuje vztah i mezi tabulkami A a C . V případě, že mezi A a B je vztah $1:N$ a mezi B a C totéž, platí stejný vztah i pro A a C . To stejné platí pro příklad se vztahem $N:1$. Každá tabulka má relaci sama na sebe s kardinalitou $1:1$. Všechny ostatní případy mají kardinalitu $N:M$. Podle těchto pravidel se vytvoří množina vztahů (pomocí funkce `return_rel()`), tisk těchto výsledků pak zajišťuje funkce `print_rel()`.

2.7 Výstup

Tisk výsledných tabulek na výstup je zajištěn funkcí `print_table()`. Data jsou ve struktuře již připravená přímo pro tisk, veškeré úpravy a kontroly se provádí v jiných funkcích před voláním tisku.

2.8 Rozšíření

Součástí programu je rovněž implementace rozšíření VAL. Jedná se o validaci, zda můžeme do již vytvořené struktury tabulek vložit další data zadaná parametrem `--isvalid=filename`. Ověření probíhá ve funkci `validation()`. Nejdříve dojde k otevření souboru a zpracování vstupu funkcí `parse_xml()`. Následně jsou data upravena podle zadaných přepínačů, a nakonec funkce porovná, zda je možné přidat elementy a atributy do již stávající struktury dat. Pokud je to možné, data jsou přidána, v opačném případě program končí s chybovým kódem.

3 Závěr

Skript byl řádně otestován sadou testů přiloženou k zadání projektu a několika vlastními testy. K porovnání výstupů přepínače `-g` s referenčními výstupy byl použit program JExamXML. Testování proběhlo na operačním systému Linux Ubuntu 10.04 a na školním serveru Merlin s operačním systémem CentOS. Všechny tyto testy dopadly úspěšně.