

Jméno a příjmení: Martin Gráca

Login: xgraca00

## 1. ZADÁNÍ

Zadáním 2. projektu do předmětu IPP bylo vytvořit skript v jazyce python, který vytvoří sql skript pro tvorbu tabulek, které by mohli odpovídat vstupnímu xml souboru. Další možností je vytvoření xml souboru s možnými tabulkami a relacemi mezi nimi.

## 2. ŘEŠENÍ

### 2.1. Zpracování parametrů

Pro zpracování parametrů slouží funkce *makeargs*. Tato funkce používá třídu *argparse*, která slouží pro práci s parametry programu. Funkce za pomocí této třídy uloží potřebné informace o parametrech, jména souborů, hlavičku nebo počty argumentů k identifikaci chyby.

*Makeargs* se taky stará o detekci chyb, jako jsou chybně zadané parametry, duplicitně zadané parametry, více zadaných vstupních/výstupních souborů nebo nepovolená kombinace parametrů. Při detekci vypíše chybové hlášení na standardní chybový výstup a ukončí skript s návratovou hodnotou podle zadání.

### 2.2. Zpracování vstupního textu

Zpracování vstupního textu provádí funkce *parsexml* za pomocí třídy *etree*, která je určena pro parsování xml souborů. Funkce začíná od kořene a prochází postupně jednotlivé potomky. Nejprve ukládá do slovníku jména tabulek a vytváří primární klíč. Poté, pokud element obsahuje atributy nebo pokud nebyl zadán parametr a, ukládá do slovníku jména atributů a příslušný datový typ, který zjišťuje funkce *gettype*. Pokud již je daný atribut ve slovníku, dochází k porovnání typů právě procházeného atributu a atributu uloženého ve slovníku. Váha jednotlivých typů je zjišťována pomocí funkce *getweight*. Pokud je váha větší, uloží se k atributu nový datový typ. Obdobný postup je prováděn při ukládání hodnoty elementu, s tím rozdílem, že u ukládání hodnoty je použit jiný datový typ pro textové řetězce.

Ukládání cizích klíčů závisí na zadaných parametrech. Při nezadání parametru etc, který určuje maximální počet sloupců vzniklých ze podelementů stejného jména, jsou vytvářeny všechny cizí klíče s číslem podelementu. Při zadání parametru etc, je třeba navíc zjistit maximální počet výskytů daných podelementů stejného jména, ke správnému určení cizích klíčů. Počty výskytů zjišťuje funkce *countFK*. Při generování cizích klíčů dochází také ke kontrole kolizí jmen atributů a textového obsahu a případně je program ukončen s chybovým hlášením a návratovým kódem 90.

Prohledání všech elementů je zajištěno rekurzivním sestupem.

### 2.3. Vytváření relací

Při zadání parametru -g nedochází k vytváření sql příkazů ale ke generování xml souborů obsahují relace mezi tabulkami. Tyto relace vytváří funkce *make\_relations*, která prochází slovník vytvořený funkcí *parsexml* a ukládá základní vztahy do slovníku *relations*. Poté dochází k opakovanému průchodu toho slovníku a přidávání tranzitivních vztahů tak dlouho, dokud nedojde k žádné změně.

### 2.4. Formátování výstupu

Pokud se má generovat sql soubor, volá se funkce *printddl*, která prochází slovník s uloženými informacemi a vytváří jednotlivé tabulky. V případě, že se má generovat xml soubor s relacemi, volá se funkce *printxml*, která vytváří validní xml soubor.