

Dokumentace k projektu IPP: XML Query v PHP 5

Úvod

Cílem tohoto projektu bylo vytvoření skriptu v jazyce PHP 5. Tento skript vyhodnotí dotaz nad vstupem ve formátu XML. Dotazovací jazyk je podobný jazyku SQL, ale má zjednodušenou syntaxi. Výstup bude rovněž ve formátu XML. Při řešení jsem použila jednu z knihoven pro práci s XML.

Zpracování parametrů příkazového řádku

Po spuštění programu se nejdříve zpracovávají parametry příkazového řádku. Je nutné kontrolovat, zda soubory zadané parametry `--input` a `--qf` existují a zda je možné z nich číst, a v případě parametru `--output`, daný soubor vytvořit a otevřít pro zápis. Je nezbytné dohlédnout i na to, aby program nebyl spuštěn s několika parametry, které nelze kombinovat mezi sebou.

K této části programu patří i funkce `query_to_array`, která provádí transformaci zadaného dotazu do asociativního pole. Kromě toho uskutečňuje dohled nad správností syntaxe a sémantiky dotazu. Zde je příklad, jak bude vypadat zadaný dotaz jako výstup funkce `query_to_array` (uvozovky u slova Herbert funkce odstání, poté co zkontroluje, zda je to řetězec):

```
Array
(
    [SELECT] => book
    [FROM] => library
    [LIMIT] => 1
    [WHERE] => .name
    [CONTAINS] => Herbert
)
```

Dále se zde definují globální proměnné, protože jejich hodnoty vycházejí právě z dotazu nebo zadaných parametrů.

Postup řešení

Při řešení samotného vyhledávání a vybírání ze vstupu ve formátu XML jsem použila knihovnu SimpleXML zejména pak třídu SimpleXMLIterator. Ta mi umožnila jednoduchou práci se zadaným vstupem a rychlé procházení celou uspořádanou strukturou SimpleXML objektů. Obdobu funkce `main` zde plní několik řádků, kde se volají jednotlivé funkce, z nichž nejdůležitější je `search_array`.

Vstupem funkce `search_array` je samotný dotaz, objekt SimpleXMLIterator a pomocný řetězec. Ten určuje, zda vyhledávám konkrétní element nebo množinu elementů. Funkce pak provádí rekurzivní vyhledávání v struktuře SimpleXML objektů a podle výsledku vrací `true` nebo `false`. Pokud nalezne množinu objektů odpovídající dotazu FROM, volá sebe samu s novými daty a hledá elementy nebo atributy odpovídající skupině SELECT. Počet pozitivních nálezů je monitorován pro případ, že by byl v dotazu zadán limit vyhledávaných elementů. Při vyhledávání je volána funkce `where_check`.

Funkce `where_check` pracuje se zadaným dotazem a daty předanými funkcí `search_array`. Kontroluje, zda byla zadána podmínka WHERE a vyhodnocuje, zda data této podmínce odpovídají a to v attributech i v podelementech. Výstupem je `true` nebo `false`. Pro lepší přehlednost kódu jsem často se opakující podmínky umístila do pomocné funkce `where_check2`, jejíž výstupem je hodnota `true` nebo `false` a je volána právě z funkce `where_check`.

Pro výpis vyhodnocených dat na standardní výstup nebo do souboru byla použita metoda SimpleXMLElement::asXML. Tato metoda je součástí knihovny SimpleXML a upravuje zadaný řetězec do správného tvaru podle standartu XML verze 1.0.

Závěr

Testování skriptu jsem prováděla na referenčním serveru Merlin. Využila jsem sadu testovacích souborů, která byla přiložena k zadání i několik vlastních XML souborů.