

## 1 Zadanie

Úlohou projektu bolo vytvoriť skript v programovacom jazyku PHP verzie 5 pre zvýrazňovanie zadaných častí textu. Skript pracuje s formátovacím súborom, ktorý definuje aké časti textu sa budú zvýrazňovať a ako sa budú zvýrazňovať. Jeden riadok formátovacieho súboru je jedno pravidlo pre zvýraznenie. Toto pravidlo sa rozdeľuje na dve časti, prvou časťou je regulárny výraz vo formáte predmetu IFJ a druhou časťou je zoznam príkazov podľa ktorých budú vkladané html značky. Tie musia byť vkladané do textu podľa poradia v akom sú zapísané vo formátovacom súbore.

## 2 Návrh

Návrh programu je možné rozdeliť na dve hlavné časti. Prvou časťou je serializácia formátovacích údajov a druhou je správne vkladanie značiek do vstupného textu.

### 2.1 Serializácia formátovacích údajov

Regulárne výrazy vo formátovacom súbore majú špeciálnu syntax, preto je potrebné zaručenie ich správneho čítania napríklad tak, že sa budú transformovať na regulárne výrazy odpovedajúce syntaxe PHP. Ďalej je potrebná kontrola formátovacích značiek a ich parametrov a správne priradenie k odpovedajúcemu regulárnemu výrazu. K jednému regulárnemu výrazu podľa zadania môže prislúchať viac formátovacích značiek, preto je potrebné nájsť optimálne a jednotné ukladanie týchto údajov do pamäte napríklad pomocou asociatívnych polí.

### 2.2 Vkladanie html značiek do vstupného textu

Dôležitou časťou návrhu je aj samotné vkladanie html značiek do vstupného textu. Je potrebné nájsť spôsob, v akom poradí budú vkladané html značky. Ak bude vkladanie priebehať okamžite pri nájdení časti textu ktorému vyhovuje regulárny výraz, môže dôjsť k problému v prípade, že html značka znemožní opätovné nájdenie časti textu, ktorému regulárny výraz odpovedá. Preto je nutné prejsť celým vstupným textom a nájsť začiatkové a koncové pozície, na ktoré budú vkladané odpovedajúce html značky. Neskôr nasleduje prechádzanie vstupného textu reverzne a vkladaniu samotných html značiek. Reverzný postup je dôležitý preto, aby nedošlo k zmenám pozícií html značiek.

## 3 Implementácia

Pri implementácii bolo použitých 5 nasledujúcich tried:

- Help - Zobrazenie nápovedy
- File - Parsovanie argumentov, čítanie a zápis do súborov
- Parser - Parsovanie formátovacieho súboru
- Syntax - Vkladanie html značiek do vstupného textu.

Po spustení skriptu je vytvorený objekt triedy `Files`, ktorý sa pokúsi a parsovanie argumentov skriptu. V prípade, že je úspešný, je možné pomocou tohto objektu čítať z súborov input a formát zapisovať do súboru output. Ak nie je zadaný súbor output alebo input je použitý štandardný vstup resp. štandardný výstup. Počas parsovania môže dôjsť k chybám, ktoré vyvolajú výnimku a ukončia skript s odpovedajúcim chybovým kódom.

Nasleduje cyklus, ktorý pomocou metód v triede `Files` číta po riadku formátovací súbor a na každý riadok je zavolaná metóda triedy `Parser`, ktorá tento riadok naformátuje do jednotnej podoby. Súčasťou tohto formátovania je aj transformácia regulárnych výrazov so špeciálnou syntaxou na syntax regulárnych výrazov jazyka PHP.

Formátovacie príkazy sú skontrolované spolu s prípadnými parametrami. Upravený regulárny výraz spolu so všetkými

formátovacími príkazmi je uložený do asociatívneho poľa štýlom: regulérny výraz => pole formátovacích príkazov. Ak dôjde k chybe počas parsovania, je vyvolaná výnimka, ktorá ukončí program s správnym chybovým kódom.

Potom je vytvorený objekt triedy `Syntax` s parametrami formátovacia databáza, vstupný text a príznak nového riadku. Je zavolaná trieda aplikovania vloženia html značiek do vstupného textu. Táto metóda začne čítať formátovaciu databázu, z ktorej je vybratý regulárny výraz. Skript sa pokúsi o nájdenie odpovedajúcej časti vo vstupnom texte a uloží si počiatočnú a koncovú pozíciu spolu so všetkými html značkami, ktoré sa majú vložiť do poľa. V prípade, že pozícia už je obsiahnutá v poli, nevytvorí sa nový element ale je len pridaná html značka už do existujúceho elementu k danej pozícii. Ak dôjde k aplikovaniu všetkých regulárnych výrazov, je pole usporiadané podľa kľúča, teda podľa pozície. Následne je toto pole iterované reverzne a do vstupného textu sú vkladane všetky html značky, ktoré prislúchajú danej pozícii. Ak je nastavený aj príznak nového riadku, po vložení formátovacích značiek je vložená aj html značka nového riadku pred každý znak nového riadku.

Po dokončení je vstupný text spolu s formátovacími značkami zapísaný do výstupného súboru alebo na štandardný výstup a sú uvoľnené všetky popisovače použitých súborov. Skript je ukončený odpovedajúcim chybovým kódom.