

Dokumentace k projektu IPP 2016/2017, DKA

Zvýraznění syntaxe v Pythonu 3 do IPP

Analýza problému

Úkolem bylo vytvořit skript, který v zadaném vstupním souboru zvýrazní syntaxi podle pravidel určených formátovacím souborem.

Postup řešení

Zpracování argumentů

Tato část projektu je implementována pomocí cyklu, která prochází jednotlivé argumenty a obsahuje povolené přepínače. Pokud byl argument zadán vícekrát, nebo byl zadán neznámý přepínač – jedná se o chybu. Při zpracovávání argumentu se také testuje správnost kombinace přepínačů. Například použití přepínače `-help` s jinými přepínači.

Zpracovávání formátovacího souboru

Formátovací soubor je načten do paměti, pokud byl zadán. v opačném případě se nejedná o chybu a výstupem je vstupní soubor.

Soubor se zpracovává po řádcích, který je vždy kontrolován pomocí regulárního výrazu, jestli odpovídá zadání. Prázdné řádky jsou přeskakovány.

Následně je řádek rozdělen na regulární výraz a formátovací příkaz. Ty jsou následně vloženy do seznamu. Formátovací příkazy jsou následně přepsány do HTML značek.

Zpracování vstupního souboru

Jestliže je zadána cesta ke vstupnímu souboru, je nejdříve zkontrolováno, zda máme dostatečná práva pro přečtení souboru, případně zda soubor existuje. Pokud není zadán parametr `-input=filename` se data načítají ze STDIN. Pokud soubor nelze otevřít, program se ukončuje s chybovou hláškou a návratovou hodnotou 2.

Validace regulárních příkazů

Převod regulárního výrazu na odpovídající PCRE regex je prováděn pomocí konečného automatu, který čte regulární výraz po znaku. Při každém přečteném znaku. Dočasný seznam je následně zkontrolován, zda obsahuje validní regulární výraz. Pokud je regulární výraz platný, je původní regulární výraz přepsán nově vytvořeným.

Vypsání HTML značek

Hledání pozic pro vkládání HTML značek probíhá pomocí funkce *generatetagposition()*. Nejdříve se inicializuje seznam o velikosti vstupního souboru. Nalezení indexu seznamu je provedena pomocí regulárního výrazu, díky které získáme začátek a konec shody se vstupním souborem. Do seznamu s indexem začátku shody regulárního výrazu jsou vloženy otevírací HTML značky, a naopak do seznamu s indexem konce shody jsou uloženy uzavírací značky. Toto je provedeno pro všechny formátovací značky.

Poté je zavolána funkce *generateoutput()* která provede výpis textu na *STDOUT*, pokud nebyl zadán parametr pro uložení. Podprogram *generateoutput()* prochází vstupní soubor po znacích a zároveň v případě shody indexu znaku s indexem v seznamu jsou vkládány HTML značky.

Jako poslední se kontroluje, zda je spuštěn parametr `-br`. Pokud ano, je při nalezení nového řádku vložena značka `
`.

Rozšíření NQS

Rozšíření je implementováno v průběhu validace regulárních výrazů, kdy při nalezení znaku pro libovolné opakování řetězců („*“), nebo pro nenulový počet opakování řetězců se kontroluje následující kvantifikátor – případně nahrazení jediným operátorem.