

## Dokumentace pro IPP 2014/2015, Projekt 2 – SYN: Zvýrazňování syntaxe

### 1 Zadání

Cílem projektu bylo vytvořit skript, který v zadaném textu na základě poskytnutého formátovacího souboru zvýrazní syntax pomocí HTML tagů. V případě, že formátovací soubor neexistuje nebo ho nelze otevřít, bude výstupem skriptu původní text.

### 2 Řešení

#### 2.1 Struktura

Celý skript je obsažen v jediném souboru a nevyužívá žádné nestandardní knihovny.

#### 2.2 Zpracování regulérních výrazů

Zpracování regulérních výrazů zadaných ve formátovacím souboru probíhá pomocí konečného automatu. Tento konečný automat má 17 stavů:

START, ESCAPE, SYM, MODIFIED, OR, NEG, NEG\_ESCAPE, NEG\_SYM, GRP\_OPEN, GRP\_ESCAPE, GRP\_SYM, GRP\_OR, GRP\_MODIFIED, GRP\_NEG, GRP\_NEG\_ESCAPE, GRP\_NEG\_SYM a GRP\_CLOSE

Stavy obsahující NEG zpracovávají negace výrazů, stavy obsahující GRP zpracovávají výrazy uvnitř závorek. Zpracování regulérních výrazů probíhá po jednom znaku, který je buď vložen do výstupního výrazu nebo je nahrazen ekvivalentem znaku abecedy regulérního jazyka, který umí zpracovat modul *re*.

Funkce obsahující konečný automat vrací příznak, zda byl vstupní regulérní výraz úspěšně překonvertován nebo došlo k chybě v průběhu zpracování, a výslední regulérní výraz, případně prázdný řetězec jestli došlo k chybě.

#### 2.3 Zpracování formátu

Zpracování formátu probíhá v jednoduché funkci, která porovnáváním vstupního řetězce s validními možnostmi generuje otevírací a zavírací HTML tagy. V případě chybného formátu je vrácen příznak neúspěchu a aplikace ukončená stavovým kódem 4.

#### 2.4 Zvýrazňování syntaxe

Pro zvýrazňování syntaxe je nadefinována pomocná třída *Formatting*, která obsahuje původní text z formátovacího souboru, zkonvertovaný regulérní výraz, HTML tagy, indexy na kterých začíná a končí nalezená slova odpovídající regulérnímu výrazu a počítadlo počtů použití. Indexy jsou vyhledány s pomocí funkce *re.finditer*. Skript poté prochází vstupní text znak po znaku a kontroluje zda na dané pozici nezačíná/končí některý formát. Formátovací tagy jsou aplikovány v pořadí podle, kterého byli definovány ve formátovacím souboru.

V případě, že byl použit parametr *–nooverlap* (rozšíření HTM), jsou po vyhledání indexů ještě vyhledány kolize a následně doplněny indexy, tak aby byl vnořený tag uzavřen před uzavřením „rodičovského“ tagu.

#### 2.5 Rozšíření

Ve skriptu jsou implementovány rozšíření NQS a HTM.

### 3 Závěr

V projektu jsem se pokusil kromě základní funkcionality implementovat i rozšíření NQS a HTM. V rozšíření HTM se mi, ale nepodařilo zabezpečit 100% funkcionality, a tak se občas stává, že jsou tagy uzavřeny v nesprávném pořadí. Dále by bylo možné v konečném automatu zredukovat počet stavů, protože se tam více-méně opakují téměř totožní čtveřice stavů.