

Spracovanie parametrov

Skript spracováva parametre pomocou funkcie `getopt`, ktorej výstup porovná s počtom parametrov `$argv` čo zamedzí ich redundancii. V prípade, že sú korektne zadané sa uložia do globálnych premenných.

Vstupné a výstupné dáta

Primárnym vstupom a výstupom sú štandardný vstup `STDIN` a štandardný výstup `STDOUT`. Pokiaľ je parametrom zadaný vstupný súbor, otvorí sa a formátuje jeho obsah. Rovnako môže byť zadaný výstupný súbor.

Vstupný súbor obsahujúci informácie o formátovaní je implicitne nastavený ako prázdny. Pokiaľ je špecifikovaný parametrom, súbor sa otvorí a spracuje.

Spracovanie formátovacieho súboru

V prípade nezadania formátovacieho súboru je volaná funkcia `noFormatOutput` ktorá vytlačí vstup na výstup, prípadne pridá element `
` na koniec každého riadku pôvodného textu.

V prípade, že formátovací súbor zadaný bol, dá sa čítať a je neprázdny, funkciou `fgets` prebehne prečítanie súboru po riadkoch. Prázdne riadky sú preskočené, neprázdne sú spracované funkciou `createNewFormatExpression`. Funkcia rozdelí reťazec na časť regulárneho výrazu a zoznamu formátovacích príkazov. Vyhľadanie vo vstupnom súbore je implementované pomocou vstavanej funkcie `preg_match_all`. Prevod vstupného regulárneho výrazu znak po znaku na regulárny výraz regex je implementovaný konečným automatom.

Konečný automat vždy najskôr skontroluje ascii hodnotu platného znaku, ktorá musí byť väčšia ako 32. Každý stav automatu vymedzuje platné a neplatné znaky. Začína v stave štart, ktorý obmedzuje platné znaky na tie, ktoré môžu stáť na začiatku výrazu. Na základe aktuálneho znaku má možnosť prejsť do stavu negácie, znaku, špeciálneho znaku alebo chyby. Stav negácie hneď vyhodnotí každý normálny znak, v prípade špeciálnych znakov nastaví znak negácie ktorý je vyhodnotený po každom prechode automatu. Stav špeciálneho znaku očakáva iba špeciálne znaky, stav znak zas všetky znaky ktoré sa môžu vyskytovať ako následník. Automat taktiež kontroluje znaky, ktoré sú špeciálne pre regex ktoré zmení za normálne znaky s pôvodným významom.

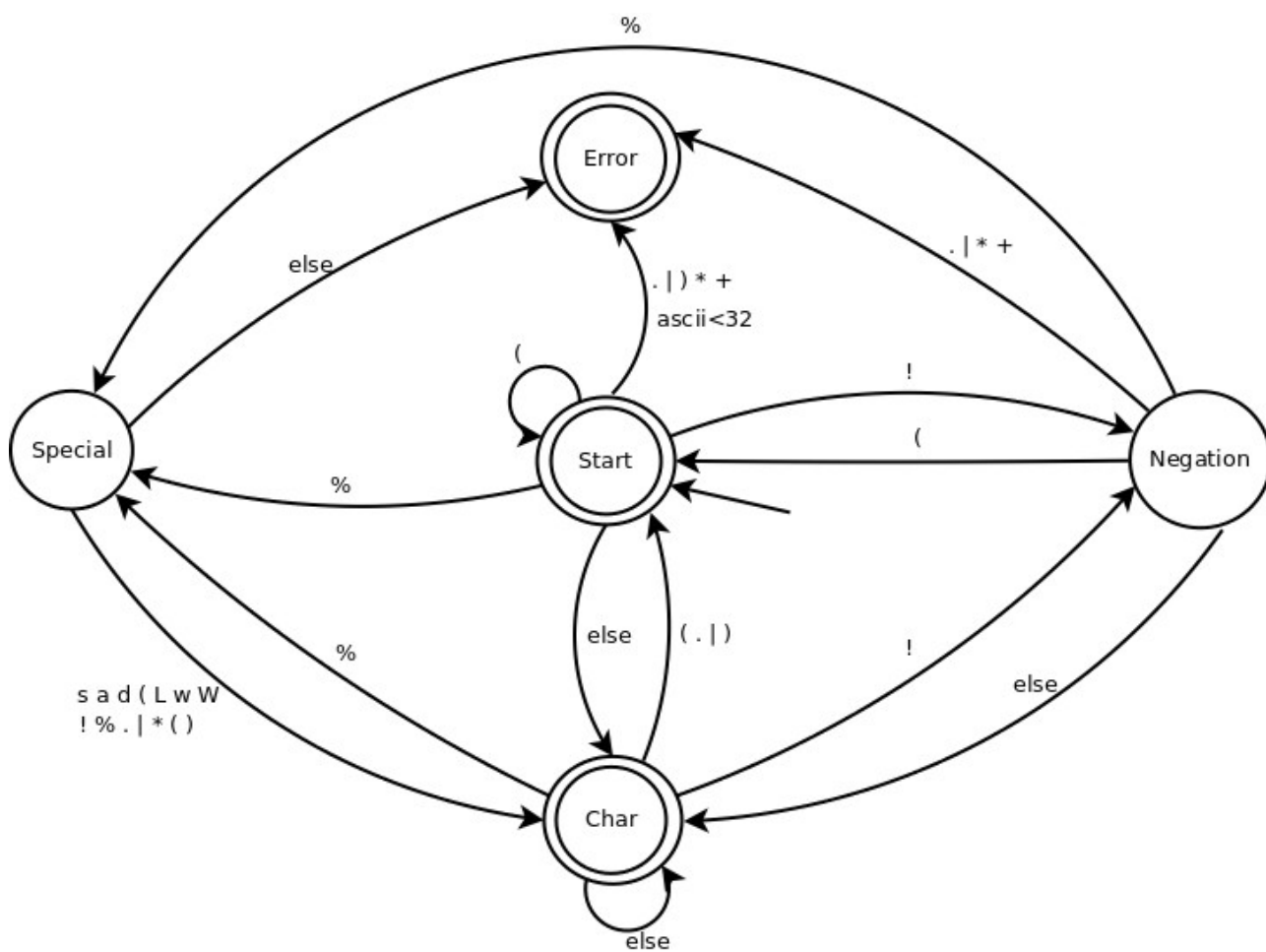
Pre kontrolu spätosti uzatvárania a poradia zátvoriek skript obsahuje zásobník ktorý na konci spracovania výrazu vyhodnotí a určí ich korektnosť.

Časť reťazca, ktorý obsahuje zoznam formátovacích príkazov je najprv zbavená nadbytočných tabulátorov na začiatku a rozdelená na dve časti oddelené čiarkou. Následne je sú odstávané biele znaky zo začiatku a konca reťazca. Na základe porovnania príkazu s zoznamom povolených príkazov je v prípade úspechu volaná funkcia `Format`. Tento proces sa opakuje v cykle pokiaľ nieje pôvodný zoznam prázdny.

Funkcia `Format` aplikuje regex výraz na vstupný súbor. Pre každú zhodu v texte si uloží jej začiatkový a koncový index. V cykle prejde všetky zhody, pre každý jej index sa pozrie do poľa `array_of_tags`. Pokiaľ pole ešte neobsahuje žiadnu položku s daným indexom, vytvorí ju a vloží doňho daný formátovací príkaz. Ak už pole položku malo, príkaz naň nadpojí na koniec.

Vytvorenie finálneho textu

Pri tvorbe konečného výstupného textu sa prechádza po indexe každého vstupného znaku. Najprv sa skontroluje pomocné pole `array_of_tags` s daným indexom, pokiaľ je neprázdne, vloží obsahujúci tag do výstupu. Následne sa skontroluje, či je zadaný parameter `br` a aktuálny znak je koniec riadku, vloží element `
` pred tento znak.



Obrázok 1: schéma konečného automatu