

Dokumentace úlohy CST: C Stats v PERL do IPP 2012/2013

Jméno a Příjmení: Martin Krippel

Login: xkripp00

Úvod

Tento dokument opisuje návrh a spôsob implementácie projektu (zadanie C Stats) do predmetu IPP v skriptovacom jazyku Perl.

Návrh

Táto časť sa zaoberá spracovaním súboru. Všetok text je spracovávaný pomocou regulárnych výrazov. Problém analýzy zadaného súboru, prípadne súborov, je rozdelený na dve časti. Prvou je odstránenie nepotrebných reťazcov, konkrétne komentárov, textových reťazcov a literálov a makier. Druhou časťou je samotné spočítanie hľadaných prvkov. V oboch sú využívané regulárne výrazy.

Súbor je čítaný programom po jednotlivých riadkoch. Tým sa komplikuje rozpoznanie blokových komentárov či reťazcov a makier na viacerých riadkoch. Problém sa rieši zapamätaním si stavu, v ktorom sa momentálne nachádza. Teda či je v komentári alebo v reťazci. V návrhu sa rátalo s postupným spracovávaním reťazca (riadku). Postupne od začiatku reťazca sa hľadajú komentáre, reťazce a makrá. Pri pozitívnom výskyte na riadku sa začne s hľadaním odznova z pozície, kde bol posledný výskyt.

Samozrejme po odstránení nepotrebných vecí sa musia na riadku vyhľadať potrebné údaje.

Implementácia

Program začne spracovávaním parametrov. Na ich načítanie je použitá knižnica `Getopt::Long` a funkcia `GetOptions`. Všetky premenné, do ktorých sa ukladá hodnota parametrov sú inicializované na hodnotu -1, len premenná `$par_w` (parameter -w) je inicializovaná na hodnotu 93227. Je to náhodné číslo a preto je menšia pravdepodobnosť, že bude zadané ako vzor a spôsobí kolíziu v programe.

Ďalšie knižnice, ktoré boli použité sú `Cwd` pre vypísanie absolútnej cesty, `File::Find` pre prechod stromovou adresárovou štruktúrou.

Jednotlivé prepínače na funkcionalitu (-k, -i, -o, -w, -c), sú spracované v samostatných funkciách. Až na prepínač -w používajú k počítaniu cieľových prvkov regulárne výrazy, pri prepínači -w je použitá funkcia `index()`. Na začiatku týchto funkcií je zavolaná funkcia `odstrániť()`, ktorá odstraňuje nepotrebné reťazce z načítaného riadka a vráti reťazec bez nich. Táto funkcia odstraňuje komentáre, reťazce, reťazcové literály a makrá.

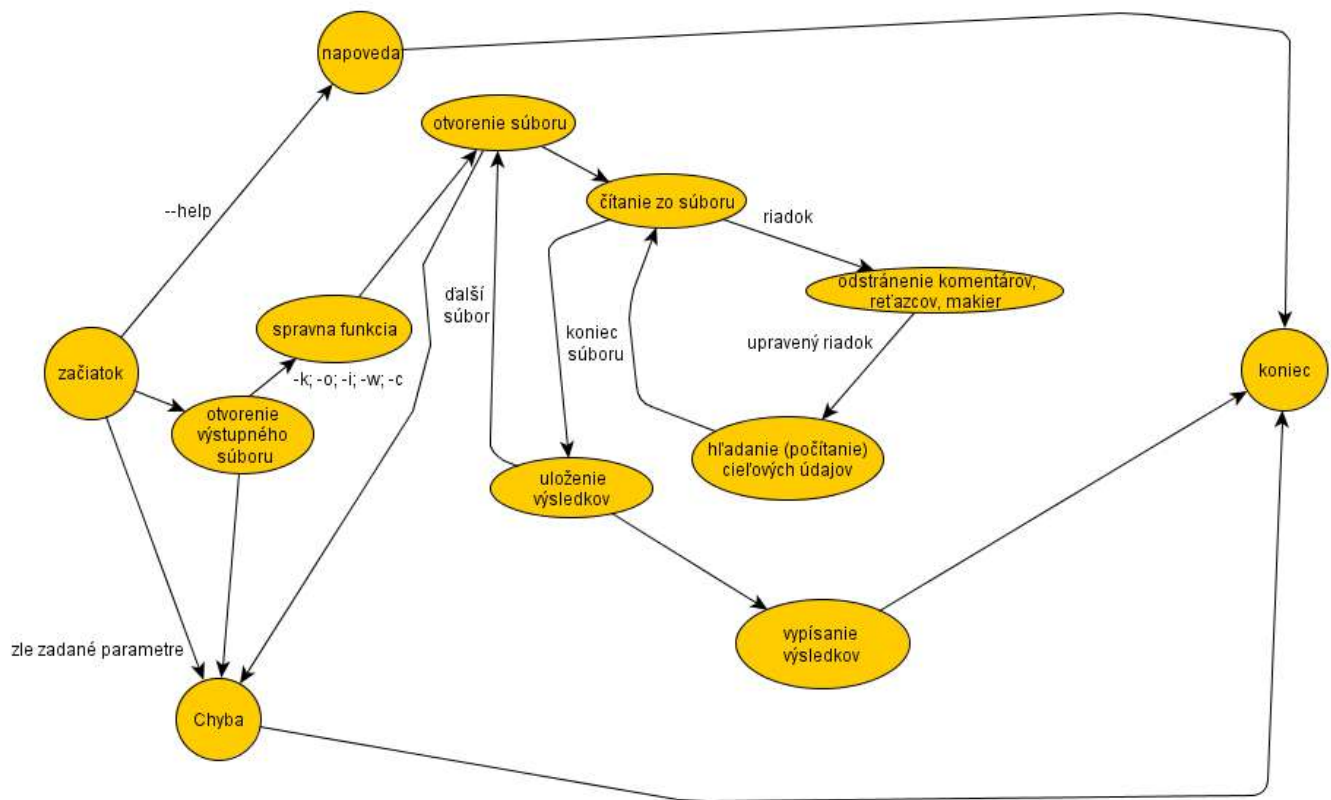
Získané hodnoty sú ukladané do poľa spracované po analyzovaní všetkých súborov. Program nakoniec vypíše získané hodnoty na základe zadania.

Testovanie

Program bol priebežne testovaný na jednoduchých súboroch, ktorými bez problémov prechádzal. Potom bol testovaný aj na zložitejších adresárových štruktúrach a súboroch, kde ručná kontrola bola nemožná. Testovaním sa preukázalo, že pri použití parametrov -k a -i pracuje program dlhšie, rádovo desiatky sekúnd (20 sekúnd bolo nameraných pri analyzovaní dvadsiatich ôsmich súboroch, pričom výskytov bolo rádovo v desať tisícoch). Je to spôsobené prechodom poľa s kľúčovými slovami.

Obrázok

Súčasťou dokumentácie je aj obrázok obr. 1 zobrazujúci jednoduchý diagram celkového priebehu programu.



Obr. 1 – diagram behu programu