

Jméno a příjmení: Milan Gardáš

Login: xgarda04

Postup řešení

Zpracování parametrů

Zpracování probíhá ihned po spuštění skriptu za pomoci proměnných `$argv` a `$argc`, ze kterých zjišťuji, které parametry byly zadány, a následně nastavuji pomocné proměnné. Pokud je zadána nepovolená kombinace parametrů nebo je zadán neznámý parametr, běh skriptu je okamžitě ukončen s odpovídajícím chybovým kódem. Parametry `--input`, `--output` a `-w` jsou rozděleny na pomoci funkce `explode`, za předpokladu svého zadání se správnou syntaxí.

Zpracování souboru

Vstupem může být nejen soubor, ale i složka. V případě zadání pouze souboru je situace jednoduchá, protože ho stačí pouze otevřít, v případě zadání složky je situace o něco složitější, z důvodu jejího prohledání a případného zanořování do podsložek dle zadání parametru `--nosubdir`. V případě zadání tohoto parametru využívám funkci `scandir` a v případě jeho nezadání využívám třídy `RecursiveDirectoryIterator` a `RecursiveIteratorIterator`. Soubory, které takto nalezu postupně uložím do pole, otevřu v režimu pro čtení, jejich obsah si uložím do pole a předám dále funkcím, které tento obsah zpracovávají.

Počet klíčových slov

Tato úloha ke správné funkčnosti vyžaduje ze vstupního souboru odstranění řetězců, maker a komentářů. Na takto upravený soubor se poté aplikuje regulární výraz, ve kterém jsou obsaženy všechna klíčová slova jazyka C, navíc těmto slovům může předcházet a následovat což je v regulárním výrazu také zaznamenáno. Samotné získání počtu klíčových slov je realizováno funkcí `preg_filter`, která je volána v cyklu, protože je možné, aby byla dvě nebo i více klíčových slov za sebou, s čímž by si jedno zavolání funkce neporadilo.

Počet jednoduchých operátorů

Vyžaduje soubor s odstraněnými makry, řetězci, komentáři, klíčovými slovy a identifikátory. Pro získání hodnoty je opět použita funkce `preg_filter`, která je opět volána v cyklu z výše uvedeného důvodu.

Počet identifikátorů

Potřebuje soubor s odstraněnými makry, řetězci, komentáři a klíčovými slovy. K získání správného výsledku je použita funkce `preg_filter`, která postupně počítá a odstraňuje identifikátory ze zadaného souboru. Úlohu trochu znesnadňuje to, že předem nevíme, na rozdíl od úlohy s počítáním klíčových slov, neznáme přesné řetězce, které budeme vyhledávat, ale naštěstí norma jazyka C udává povolené znaky v identifikátorech, takže se máme od čeho odpíchnout, díky tomu je možné sestavit regulární výraz, který je ve výše uvedené funkci využíván.

Vyhledání podřetězce

Nevyžaduje žádné úpravy vstupních souborů, protože zadaný podřetězec chceme vyhledávat v celém souboru včetně komentářů, maker apod. Samotné vyhledání využívá funkci `strpos`, která je volána v cyklu, jenž je ukončen tím, že zadaný řetězec už znovu zadaný podřetězec neobsahuje. Zajištění toho, že nebudeme nacházet stále stejný podřetězec, ale další zajistí nastavení proměnné `$offset` v parametru vyhledávací funkce, přesněji proměnnou `$offset` posuneme o pozici aktuálně nalezeného podřetězce sečteného s délkou vyhledávaného podřetězce.

Počet znaků komentářů

Probíhá pomocí několika řídicích konstrukcí, za pomoci kterých se hledá začátek komentáře a nastavuje se pomocná proměnná, která určuje, zda budeme aktuálně načtený znak ponechávat nebo jej nahradíme mezerou a přičteme jedničku k proměnné s aktuálním počtem znaků.

Formátování výstupu

Formát výstup je zadáním pevně daný a kromě správného odsazení je potřeba výpisy vypsát ve správném pořadí. Statistiku uchovávám ve dvou polích proto je nutné je řadit funkcí **array_multisort**. Po správném seřazení naleznou pomocí funkce **strlen** nejdelší hodnoty v obou polích a formátují výstup. Pole s názvy souborů vložím na index, který je roven nejdelší hodnotě mezeru, čímž se doplní mezery i na ostatní indexy mezi skutečnou délkou a požadovanou délkou. Pole s hodnotami je na formátování obtížnější, protože je nutné jeho obsah zarovnávat k pravému okraji, což je vyřešeno tak, že zkontroluji, zda a případně o kolik je aktuální hodnota v podobě délky menší než nejdelší hodnota a tento rozdíl se poté vloží v počtu mezer před hodnotu.