Task documentation CST: C Stats in PHP for IPP 2012/2013
Name and surname: Marek Marušic
Login: xmarus05

1 Purpose of script

Main task of this script written in PHP is to read, analyze and output statistics of files written in C language based on supplied arguments.

2 Processing of arguments

First thing done by this script is processing and check rightness of arguments. It is done by function parse() which uses php function getopt(). Getopt is function which will return an array of options. Those options are processed whit parse and then array params[] is returned and used for !!further run !!.

3 Files and directory listing

Every input directory and subdirectory is scanned by function findFiles(). Found files with .c or .h extensions are stored in array with or without real paths to files depending on whether -p parameter is set or not. This array is then used to iterate over files for further processing.

4 File handling

Every file is checked whether it is readable followed by reading file line by line. For each line there is executed extraction of relevant data and elimination of irrelevant data, such as comments macros and strings in some cases. For this extraction was used few regular expressions. During the extraction every relevant data are processed and counted according to its type and added to counter of that type. Information about file path and obtained count of searched data type is appended to array $Output, after all data from file are processed.

4.1 Comments extraction

When comments are being counted there is while loop used for every line to extract comments, one by one till no comments left. We don't have to worry about comments in macros or strings, because those were eliminated before.

4.2 Keywords and identifiers extraction

There was used regular expression for this search to find every word starting with _ or alphabetic character followed by _ or alphanumeric characters. Every found word was checked by php function in_array() whether it exists in a predefined array of keywords. The word is then counted as a keyword or as an identifier depending on the result from comparison.

4.3 Pattern extraction

When -w parameter is used it has to be followed by a pattern. This pattern is then used to count occurrences of supplied pattern. For this is used php function substr_count().

4.4 Operators extraction

There are used few regular expressions to remove characters which are not operators but could cause misleading results. Like pointers, parentheses, strings, comments etc. Afterward regular expression to find operators in every line is used and every result is counted.

5 Output formatting

Output is generated from previously populated array with file paths and statistics. when -p parameter is set the array is sorted by name of file. Next we calculate indent from the longest file path or file name and from current file which one is being processed. After all files and their counts are printed, we add one more line with overall count of specified data type.