Implementační dokumentace k 1. úloze do IPP 2023/2024
Jméno a příjmení: Milan Takáč
Login: xtakac09

1. Úvod

Tato dokumentace slouží k popsání mé implementace parseru v jazyku Python. Využíval jsem převážně vestavěných či importovaných knihoven.

2. Popis chování

Prvně byla potřeba kontrola argumentů, v tomto projektu bylo potřeba jen argumentu *help*, k tomu mi posloužila knihovna *argparse*. Dále po spuštění programu příkazem *python3.10 parse.py < input > output* se postupně začne zpracovávat předaný soubor. Jako první se provede kontrola přítomnosti hlavičky *.IPPcode24*, pokud je na prvním místě (vyjma předcházejících bílých znaků, či komentářů), vytvoří se XML element program, a může se pokračovat na ostatní řádky, chybí-li, program se ukončí s chybovou hláškou, a chybovým kódem 21.

Načtení řádku s instrukcí se provede kontrola, zda řádek obsahuje komentáře, odstranění okolních bílých znaků, a rozdělení do listu po slovech. Tento list je poté předán funkci *checkInstructionArgsCount*, kde se provede kontrola jména, a rozdělení strukturou match podle počtu argumentů. List se dál předá do funkce *writeInstructionToXML*, kde se vytvoří element *instruction* s názvem instrukce a pořadím. Argumenty se dále předají funkci *checkArguments*, kde se znovu podle názvu a počtu argumentů řadí do sekce pro kontrolu argumentů. Celkově je 8 možností kombinace argumentů, kde se v každé zavolá lexikální kontrola pro daný argument. Výsledkem volání je list, kterého elementy slouží pro jednotlivé argumenty. Obsahuje-li nějaký element hodnotu *False*, vypíše se chybová hláška na standardní chybový výstup, a program se poté ukončí s chybovým kódem 23, je-li vše v pořádku, element obsahuje další list s připraveným typem a hodnotou pro zápis do XML.

Při kontrole stringů jsem použil funkci *replace* pro nahrazení znaků, u kterých by mohl nastat problém při zápisu do XML (<, >, &, ‘, “), a pro kontrolu escape sekvencí jsem využil regexové výrazy.

Po zpracování souboru se vytvoření XML strom se základní hlavičkou zapíše do Vámi zvoleného souboru *output.*