

Implementační dokumentace k 1. úloze do IPP 2023/2024

Jméno a příjmení: Thu Tra Phamová

Login: xphamo00

Postup

Základní myšlenka řešení `parse.py` skriptu spočívá v rozdělení vstupních dat na vhodné části, u kterých postupně kontrolujeme jejich správnost. Při kontrole si zároveň potřebné a již korektní informace ukládáme k vygenerování výstupu ve formátu XML.

Programová část

Skript má pouze jeden parametr `-h`, který vypíše správné spouštění. K tomu byla použita `sys` knihovna, díky které si taky zachytáváme vstupní data nebo vypisujeme chyby či xml dokument na výstup.

Zachycený vstup ze `stdin` je uložen v proměnné `file`. Ten se v metodě `parse()`, z třídy `Instructions`, se znalostí `regex` (knihovna `re`) upraví pro usnadnění pozdější manipulaci s výstupem. To zahrnuje odstranění prázdných řádků a komentářů, neboť číslování řádku (označen jako `index`) je využíváno v metodě třídy `Xml`. Po úpravě je text pomocí metod `splitlines()` a `split()` rozdělen do částí, které jsou následně cyklicky kontrolovány.

Nejdřív je kontrolován lexém hlavičky nebo operačního kódu, poté syntaxe operačního kódu, pokud se u něho vyskytuje požadovaný operand. Následuje volání `Xml.instruction()`, která vytváří prvek tzv. `ET.SubElement()` k vytvořenému kořeni `program`. Poté se ověřují lexémy operandů pomocí metod třídy `Instructions`, a to `label_check()`, `var_check()`, `symb_check()` a `type_check()`. Pro detekci chyb je převážně využívána metoda `re.match()`.

Po zjištění správnosti operandů jsou volány metody `attribute()` a `symb_parse()` z třídy `Xml`. Obě slouží pro ukládání argumentů instrukce, přičemž pro typ symbolu je prováděna dodatečná kontrola, zda se jedná o konstantu nebo proměnnou, což vyžadovalo vytvoření metody `Xml.symb_parse()` pro jednodušší zacházení.

K vytvoření XML dokumentu jsou použity následující knihovny: `xml.etree.ElementTree` (jako `ET`), `xml.dom.minidom` (jako `DOM`) a `codecs`. První pro ukládání hierarchie stromu, druhý pro pěkný vzhled (odsazení) a třetí pro dekódování (UTF-8).