Zadání projektu z předmětu IPP 2022/2023

Zbyněk Křivka
E-mail: krivka@fit.vut.cz

1 Základní charakteristika projektu

Navrhněte, implementujte a dokumentujte sadu skriptů pro interpretaci nestrukturovaného imperativního jazyka IPPcode23. K implementaci vytvořte odpovídající stručnou programovou dokumentaci. Projekt se skládá ze dvou úloh a je individuální.

První úloha se skládá ze skriptu `parse.php` v jazyce PHP 8.1 (viz sekce 3) a dokumentace k tomuto skriptu (viz sekce 2.1). Druhá úloha se skládá ze skriptu `interpret.py` v jazyce Python 3.10 (viz sekce 4) a dokumentace k tomuto skriptu (viz sekce 2.1 a upřesnění v sekci 4).

2 Požadavky a organizační informace

Kromě implementace skriptů a vytvoření dokumentace je třeba dodržet také řadu následujících formálních požadavků. Pokud některý nebude dodržen, může být projekt hodnocen nula body! Pro kontrolu alespoň některých formálních požadavků lze využít skript `is_it_ok.sh` dostupný v *Moodle* předmětu IPP.

Termíny:

připomínky[1] k zadání projektu do 13. února 2023;

fixace zadání projektu od 14. února 2023;

odevzdání první úlohy v pondělí **13. března** 2023 do 23:59:59;

odevzdání druhé úlohy v úterý **18. dubna** 2023 do 23:59:59.

Dodatečné informace a konzultace k projektu z IPP:

- *E-Learning* (Moodle) předmětu IPP včetně Často kladených otázek (*FAQ*) a Souborů k projektu.
- *Fórum* v *E-Learning* (Moodle) IPP pro ak. rok 2022/2023, témata `Projekt.*`.
- Zbyněk Křivka (garant projektu): dle konzultačních slotů (viz webová stránka) nebo po dohodě e-mailem (uvádějte předmět začínající `"IPP:"`), viz `http://www.fit.vut.cz/person/krivka`. Další cvičící najdete na kartě předmětu.
- Dušan Kolář (garant předmětu; jen v závažných případech): po dohodě e-mailem (uvádějte předmět začínající `"IPP:"`), viz `http://www.fit.vut.cz/person/kolar`.

[1] Naleznete-li v zadání nějakou chybu či nejasnost, dejte prosím vědět na *Fóru* předmětu nebo e-mailem na krivka@fit.vut.cz. Validní připomínky a upozornění na chyby budou oceněny i bonusovými body.

Pokud máte jakékoliv dotazy, problémy či nejasnosti ohledně tohoto projektu, tak po přečtení *FAQ* a *Fóra* (využívejte i možnosti hledání na *Fóru*), neváhejte napsat na *Fórum* (u obecného problému, jenž se potenciálně týká i vašich kolegů) či kontaktovat garanta projektu, cvičícího (v případě individuálního problému), nebo nouzově garanta předmětu, kdy do předmětu vždy uveďte na začátek řetězec `"IPP:"`. Na problémy zjištěné v řádu hodin až jednotek dní před termínem odevzdání některé části projektu nebude brán zřetel. Začněte proto projekt řešit s dostatečným předstihem.

Forma a způsob odevzdání: Každá úloha se odevzdává individuálně prostřednictvím StudIS do předmětu IPP (odevzdání e-mailem je bodově postihováno) a odevzdáním stvrzujete výhradní autorství skriptů i dokumentace.

V aktivitě „Projekt - 1. úloha v PHP 8.1“ odevzdáte archív pro první úlohu v jazyce PHP 8.1 (skript `parse.php`; včetně dokumentace tohoto skriptu). Po termínu odevzdání 1. úlohy bude otevřeno odevzdání archívu v aktivitě „Projekt - 2. úloha v Pythonu 3.10“ pro druhou úlohu (skript `interpret.py`; včetně dokumentace tohoto skriptu). Součástí archívu první úlohy může být i nedokončený skript a dokumentace odevzdávané k hodnocení až v druhé úloze a naopak v archívu druhé úlohy se smí vyskytovat skript z první úlohy a adresář s vašimi testy.

Každá úloha bude odevzdána ve zvláštním archívu, kde budou soubory k dané úloze zkomprimovány programem ZIP, TAR+GZIP či TAR+BZIP do jediného archívu pojmenovaného `xlogin99.zip`, `xlogin99.tgz`, nebo `xlogin99.tbz`, kde `xlogin99` je váš login. Velikost každého archívu bude omezena informačním systémem (pravděpodobně na 2 MB). Archív nesmí obsahovat speciální či binární spustitelné soubory. Názvy všech souborů mohou obsahovat pouze písmena anglické abecedy, číslice, tečku, pomlčku a podtržítko. **Skripty budou umístěny v kořenovém adresáři odevzdaného archívu**. Po rozbalení archívu na serveru *Merlin* bude možné skript(y) spustit. Archív smí obsahovat v rozumné míře pomocné adresáře (typicky pro vaše vlastní testy, vaše knihovny a pomocné skripty nebo povolené knihovny, které nejsou nainstalovány na serveru *Merlin*).

Hodnocení: Výše základního bodového hodnocení projektu v předmětu IPP je **maximálně 20 bodů**. Navíc lze získat maximálně 5 bonusových bodů za kvalitní a podařené řešení některého z rozšíření nebo kvalitativně nadprůměrnou účast na *Fóru* projektu apod.

Hodnocení jednotlivých skriptů: `parse.php` až 8 bodů (z toho až 1 bod za dokumentaci a úpravu zdrojových kódů skriptu `parse.php`); `interpret.py` až 12 bodů (z toho dokumentace a úprava zdrojových textů skriptů až **3 body**). Dokumentace je však hodnocena maximálně 30 % ze sumy hodnocení funčnosti skriptu dané úlohy (tedy v případě neodevzdání žádného funkčního skriptu v dané úloze bude samotná dokumentace hodnocena 0 body). Body za každou úlohu včetně její dokumentace se před vložením do StudIS zaokrouhlují na desetiny bodu.

Skripty budou spouštěny na serveru *Merlin* příkazem: *interpret skript parametry*, kde *interpret* bude `php8.1` nebo `python3.10`, *skript* a *parametry* závisí na dané úloze. Hodnocení většiny funkčnosti bude zajišťovat automatizovaný nástroj. Kvalitu dokumentace, komentářů, návrh a strukturu zdrojového kódu budou hodnotit cvičící.

Vaše skripty budou hodnoceny nezávisle na sobě, takže je možné odevzdat například jen skript `interpret.py` do 2. úlohy bez odevzdání 1. úlohy.

Podmínky pro opakující studenty ohledně případného uznání hodnocení loňského projektu najdete v *FAQ* na *Moodle* předmětu.

Registrovaná rozšíření: V případě implementace některých registrovaných rozšíření za bonusové body bude odevzdaný archív obsahovat soubor `rozsireni`, ve kterém uvedete na každém řádku

identifikátor jednoho implementovaného rozšíření[2] (řádky jsou ukončeny unixovým koncem řádku, tj. znak s dekadickou ASCII hodnotou `10`). V průběhu řešení mohou být zaregistrována nová rozšíření úlohy za bonusové body (viz *Fórum*). Nejpozději do termínu pokusného odevzdání dané úlohy můžete na *Fórum* zasílat návrhy na nová netriviální rozšíření, která byste chtěli navíc implementovat. Cvičící rozhodne o přijetí/nepřijetí rozšíření a hodnocení rozšíření dle jeho náročnosti včetně přiřazení unikátního identifikátoru. Implementovaná rozšíření neidentifikovaná v souboru `rozsireni` nebudou hodnocena.

Pokusné odevzdání: Pro zvýšení motivace studentů pro včasné vypracování úloh nabízíme koncept nepovinného pokusného odevzdání. Výměnou za pokusné odevzdání do uvedeného termínu (cca týden před finálním termínem) dostanete zpětnou vazbu v podobě zařazení do některého z pěti rozmezí hodnocení (0–10 %, 11–30 %, 31–50 %, 51–80 %, 81–100 %). Bude-li vaše pokusné odevzdání v prvním rozmezí hodnocení, máte možnost osobně konzultovat důvod, pokud jej neodhalíte sami. U ostatních rozmezí nebudou detailnější informace poskytovány.

Pokusné odevzdání bude relativně rychle vyhodnoceno automatickými testy a studentům zaslána orientační informace o správnosti pokusně odevzdané úlohy z hlediska části automatických testů (tj. nebude se jednat o finální hodnocení; proto nebudou sdělovány ani body či přesnější procentuální hodnocení). Využití pokusného termínu není povinné, ale jeho nevyužití může být negativně vzato v úvahu v případě reklamace hodnocení projektu.

Formální požadavky na pokusné odevzdání jsou totožné s požadavky na finální termín a odevzdání se bude provádět do speciálních aktivit „Projekt - Pokusné odevzdání 1. úlohy“ do **6. března 2023** a „Projekt - Pokusné odevzdání 2. úlohy“ do **11. dubna 2023**. Není nutné zahrnout dokumentaci, která spolu s rozšířeními pokusně vyhodnocena nebude.

2.1 Dokumentace

Implementační dokumentace (dále jen dokumentace) musí být stručným a uceleným průvodcem **vašeho způsobu řešení** skriptů 1., resp. 2. úlohy. Bude vytvořena ve formátu **PDF** nebo **Markdown** (viz [4]). Jakékoliv jiné formáty dokumentace než PDF či Markdown[3] (přípona `md`) budou ignorovány, což povede ke ztrátě bodů za dokumentaci. Dokumentaci je možné psát buď česky, slovensky (s diakritikou, formálně čistě), nebo anglicky (formálně čistě).

Dokumentace bude popisovat celkovou filozofii návrhu, interní reprezentaci, způsob a váš specifický postup řešení (např. řešení sporných případů nedostatečně upřesněných zadáním, konkrétní řešení rozšíření, případné využití návrhových vzorů, implementované/nedokončené vlastnosti). Dokumentaci je vhodné doplnit např. o UML diagram tříd (především u 2. úlohy), navržený konečný automat, pravidla vámi vytvořené gramatiky nebo popis jiných formalismů a algoritmů. Nicméně **nesmí obsahovat ani částečnou kopii zadání**.

Vysázená dokumentace 1. úlohy popisující skript `parse.php` by neměla přesáhnout 1 stranu A4 (tj. rozsah přibližně 1 až 2 normostrany). U 2. úlohy (popisující skript `interpret.py`) pak nepřesáhněte vysázené 2 strany A4 (nepočítaje vhodně velký UML diagram). Doporučení pro sazbu: 10bodové písmo Times New Roman pro text a Courier pro identifikátory a skutečně krátké úryvky zajímavého kódu (či zpětné apostrofy v Markdown); nevkládejte žádnou zvláštní úvodní stranu, obsah ani závěr. V rozumné míře je vhodné používat nadpisy první a případně druhé úrovně (12bodové a 11bodové písmo Times New Roman či ## a ### v Markdown) pro vytvoření logické struktury dokumentace.

[2] Identifikátory rozšíření jsou uvedeny u konkrétního rozšíření tučně.

[3] Bude-li přítomna dokumentace v Markdown i PDF, bude hodnocena verze v PDF, kde je jistější sazba.

Nadpis a hlavička dokumentace[4] bude na prvních třech řádcích obsahovat:

```
Implementační dokumentace k %cislo%. úloze do IPP 2022/2023
Jméno a příjmení: %jmeno_prijmeni%
Login: %xlogin99%
```

kde `%jmeno_prijmeni%` je vaše jméno a příjmení, `%xlogin99%` váš login a `%cislo%` je číslo dokumentované úlohy.

Dokumentace bude v kořenovém adresáři odevzdaného archívu a pojmenována `readme1.pdf` nebo `readme1.md` pro první úlohu a `readme2.pdf` nebo `readme2.md` pro druhou úlohu.

V rámci dokumentace bude hodnocena i funkční/objektová dekompozice a komentování zdrojových kódů (minimálně každá funkce a modul (třída) by měly mít komentář o svém účelu a parametrech; u složitějších funkcí okomentujte i omezení na vstupy či výstupy).

2.2 Programová část

Zadání projektu vyžaduje implementaci dvou skriptů[5], které mají parametry příkazové řádky a je definováno, jakým způsobem manipulují se vstupy a výstupy. Skript nesmí spouštět žádné další procesy či příkazy operačního systému. Veškerá chybová hlášení, varování a ladicí výpisy směřujte pouze na standardní chybový výstup, jinak pravděpodobně nedodržíte zadání kvůli modifikaci definovaných výstupů (ať již do externích souborů, nebo do standardního výstupu). Jestliže proběhne činnost skriptu bez chyb, vrací se návratová hodnota `0` (nula). Jestliže dojde k nějaké chybě, vrací se chybová návratová hodnota větší jak nula. Chyby mají závazné chybové návratové hodnoty:

- 10 - chybějící parametr skriptu (je-li třeba) nebo použití zakázané kombinace parametrů;
- 11 - chyba při otevírání vstupních souborů (např. neexistence, nedostatečné oprávnění);
- 12 - chyba při otevření výstupních souborů pro zápis (např. nedostatečné oprávnění, chyba při zápisu);
- 20 – 69 - návratové kódy chyb specifických pro jednotlivé skripty;
- 99 - interní chyba (neovlivněná vstupními soubory či parametry příkazové řádky; např. chyba alokace paměti).

Pokud zadání níže nestanoví jinak, veškeré vstupy a výstupy jsou v kódování UTF-8. Pro účely projektu z IPP musí být na serveru *Merlin* ponecháno implicitní nastavení `locale`[6], tj. `LC_ALL=cs_CZ.UTF-8`.

Jména hlavních skriptů jsou dána zadáním. Pomocné skripty nebo knihovny budou mít příponu dle zvyklostí v daném programovacím jazyce (`.php` pro PHP 8 a `.py` pro Python 3). Vyhodnocení skriptů bude prováděno na serveru *Merlin* s aktuálními verzemi interpretů (dne 24. 1. 2023 bylo na tomto serveru nainstalováno `php8.1`[7] verze 8.1.12 a `python3.10`[8] verze 3.10.6).

[4] Anglické znění nadpisu a hlavičky dokumentace najdete v *FAQ* na *Moodle* předmětu

[5] Tyto skripty jsou aplikace příkazové řádky neboli konzolové aplikace.

[6] Správné nastavení prostředí je nezbytné, aby bylo možné používat a správně zpracovávat parametry příkazové řádky v UTF-8. Pro správnou funkčnost je třeba mít na UTF-8 nastaveno i kódování znakové sady konzole (např. u programu PuTTY v kategorii *Window.Translation* nastavíte *Remote character set* na UTF-8). Pro změnu ovlivňující aktuální sezení lze využít unixový příkaz `export LC_ALL=cs_CZ.UTF-8`.

[7] Upozornění: Na serveru *Merlin* je třeba dodržet testování příkazem `php8.1`, protože pouhým `php` se spouští verze, která má omezen přístup k souborovému systému!

[8] Upozornění: Na serveru *Merlin* je třeba dodržet testování příkazem `python3.10`, protože pouhým `python` se spouští stará nekompatibilní verze! Python 3.x není zpětně kompatibilní s verzí 2.x!

K řešení lze využít standardně předinstalované knihovny obou jazykových prostředí na serveru *Merlin*. Případné využití jiné knihovny kromě knihovny podporující načítání/ukládání formátu XML, zpracování parametrů příkazové řádky a zpracování regulárních výrazů je třeba konzultovat s garantem projektu (především z důvodu, aby se řešení projektu použitím vhodné knihovny nestalo zcela triviálním). Seznam povolených a zakázaných knihoven je udržován aktuální na *Moodle*. Ve skriptech v jazyce PHP jsou některé funkce z bezpečnostních důvodů zakázány (např. `header`, `mail`, `popen`, `curl_exec`, `socket_*`; úplný seznam je u povolených/zakázaných knihoven na *Moodle*).

Každý skript bude pracovat s jedním společným parametrem:

- `--help` vypíše na standardní výstup nápovědu skriptu (nenačítá žádný vstup), kterou lze převzít ze zadání (lze odstranit diakritiku, případně přeložit do angličtiny dle zvoleného jazyka dokumentace), a vrací návratovou hodnotu 0. Tento parametr nelze kombinovat s žádným dalším parametrem, jinak skript ukončete s chybou 10.

Kombinovatelné parametry skriptů jsou odděleny alespoň jedním bílým znakem a mohou být uváděny v libovolném pořadí, pokud nebude řečeno jinak. U skriptů je možné implementovat i vaše vlastní nekolizní parametry (doporučujeme konzultaci na *Fóru* nebo u garanta projektu).

Není-li řečeno jinak, tak dle konvencí unixových systémů lze uvažovat zástupné zkrácené (s jednou pomlčkou) i dlouhé parametry (se dvěma pomlčkami), které lze se zachováním sémantiky zaměňovat (tzv. alias parametry), ale testovány budou vždy dlouhé verze.

Je-li součástí parametru i soubor (např. `--source=`*`file`* nebo `--source="`*`file`*`"`) či cesta, může být tento soubor/cesta zadán/a relativní cestou[9] nebo absolutní cestou; výskyt znaku uvozovek a rovnítka ve *file* neuvažujte. Cesty/jména souborů mohou obsahovat i Unicode znaky v UTF-8.

Doporučení: Jelikož je velká část hodnocení odvozena automatickými testy, doporučujeme i pro vývoj skriptů využívat vlastní automatické testování. K tomu je potřeba sada testů, které si sestavíte na základě správného pochopení zadání. Lze se inspirovat několika zveřejněnými testy. Kromě návratového kódu je vhodné testovat i správnost výstupu v případě úspěšného dokončení skriptu. Mezi studenty je možné testy i sdílet.

3 Analyzátor kódu v IPPcode23 (`parse.php`)

Skript typu filtr (`parse.php` v jazyce PHP 8.1) načte ze standardního vstupu zdrojový kód v IPPcode23 (viz sekce 5), zkontroluje lexikální a syntaktickou správnost kódu a vypíše na standardní výstup XML reprezentaci programu dle specifikace v sekci 3.1.

Tento skript bude pracovat s těmito parametry:

- `--help` viz společný parametr všech skriptů v sekci 2.2.

Chybové návratové kódy specifické pro analyzátor:

- 21 - chybná nebo chybějící hlavička ve zdrojovém kódu zapsaném v IPPcode23;
- 22 - neznámý nebo chybný operační kód ve zdrojovém kódu zapsaném v IPPcode23;
- 23 - jiná lexikální nebo syntaktická chyba zdrojového kódu zapsaného v IPPcode23.

[9] Relativní cesta nebude obsahovat zástupný symbol ~ (vlnka).

3.1 Popis výstupního XML formátu

Za povinnou XML hlavičkou[10] následuje kořenový element `program` (s povinným textovým atributem `language` s hodnotou IPPcode23), který obsahuje pro instrukce elementy `instruction`. Každý element `instruction` obsahuje povinný atribut `order` s pořadím instrukce. Při generování elementů je pořadí číslováno od 1 v souvislé posloupnosti. Dále element obsahuje povinný atribut `opcode` (hodnota operačního kódu je ve výstupním XML vždy velkými písmeny) a elementy pro odpovídající počet operandů/argumentů: `arg1` pro případný první argument instrukce, `arg2` pro případný druhý argument a `arg3` pro případný třetí argument instrukce. Element pro argument má povinný atribut `type` s možnými hodnotami `int`, `bool`, `string`, `nil`, `label`, `type`, `var` podle toho, zda se jedná o literál, návěští, typ nebo proměnnou, a obsahuje textový element.

Tento textový element potom nese buď hodnotu literálu (již bez určení typu a bez znaku @), jméno návěští, typ, nebo identifikátor proměnné (včetně určení rámce a @). U proměnných vypisujte označení rámce vždy velkými písmeny, jak by mělo být již na vstupu. Velikosti písmen samotného jména proměnné ponechejte beze změny. Formát celých čísel je dekadický, oktalový nebo hexadecimální dle zvyklostí PHP (viz funkce `intval`), nicméně na výstup tato čísla vypisujte přesně ve formátu, v jakém byla načtena ze zdrojového kódu (např. zůstanou kladná znaménka čísel nebo počáteční přebytečné nuly). U literálů typu `string` při zápisu do XML nepřevádějte původní escape sekvence, ale pouze pro problematické znaky v XML (např. <, >, &) využijte odpovídající XML entity (např. `<`, `>`, `&`). Podobně převádějte problematické znaky vyskytující se v identifikátorech proměnných. Literály typu `bool` vždy zapisujte malými písmeny jako `false` nebo `true`.

Doporučení: Všimněte si, že analýza IPPcode23 je tzv. kontextově závislá (viz přednášky), kdy například můžete mít klíčové slovo použito jako návěští a z kontextu je třeba rozpoznat, zda jde o návěští, nebo ne. Při tvorbě analyzátoru doporučujeme kombinovat konečně-stavové řízení a regulární výrazy a pro generování výstupního XML využít vhodnou knihovnu.

Výstupní XML bude porovnáváno s referenčními výsledky pomocí nástroje A7Soft JExamXML[11], viz [2]. Pozor, v PHP příkaz `return` neslouží pro návrat chybového kódu, použijte funkci `exit`. Pro výpis varování na standardní chybový výstup uveďte na začátku skriptu příkaz

```
ini_set('display_errors', 'stderr');
```

3.2 Bonusová rozšíření

STATP Sbírání statistik zpracovaného zdrojového kódu v IPPcode23. Skript bude podporovat parametr `--stats=`*file* pro zadání souboru *file*, kam se budou vypisovat statistiky v parametrech umístěných za tímto `--stats`. Statistiky se do souboru vypisují po řádcích dle pořadí v parametrech s možností jejich opakování; na každý řádek nevypisujte nic kromě požadovaného číselného výstupu a odřádkování; případně existující soubor je přepsán. Pro sběr skupin statistik do různých souborů se použije další výskyt parametru `--stats` s jiným jménem souboru a následovaný dalšími parametry nové skupiny statistik. Parametr `--loc` vypíše do statistik počet řádků s instrukcemi (nepočítají se prázdné řádky ani řádky obsahující pouze komentář ani úvodní řádek). Parametr `--comments` vypíše do statistik počet řádků, na kterých se vyskytoval komentář. Parametr `--labels` vypíše do statistik počet definovaných návěští (tj. unikátních možných cílů skoku). Parametr `--jumps` vypíše do statistik počet všech instrukcí návratů z volání a instrukcí pro *skoky* (souhrnně podmíněné/nepodmíněné skoky a volání),

[10] Tradiční XML hlavička včetně verze a kódování je `<?xml version="1.0" encoding="UTF-8"?>`

[11] Nastavení A7Soft JExamXML pro porovnávání XML (soubor `options`) je na Moodle.

`--fwjumps` počet dopředných skoků, `--backjumps` zpětných skoků a `--badjumps` počet skoků na neexistující návěští. Je-li uveden pouze parametr `--stats` bez upřesnění statistik k výpisu, bude výstupem prázdný soubor. Parametr `--frequent` vypíše do statistik jména operačních kódů (velkými písmeny, čárkou oddělené, bez mezer), které jsou ve zdrojovém kódu nejčastější dle počtu statických výskytů. Parametr `--print=`*`string`* vypíše do statistik řetězec *string* a parametr `--eol` vytiskne do statistik odřádkování. Chybí-li při zadání parametrů statistik `--loc`, `--comments` a dalších před nimi parametr `--stats`, jedná se o chybu 10. Pokus o zápis více skupin statistik do stejného souboru během jednoho spuštění skriptu vede na chybu 12. [1,5 b]

NVP Při návrhu a implementaci skriptu `parse.php` a pomocných skriptů bude aplikováno objektově orientované programování a využit alespoň jeden vhodný standardní návrhový vzor (viz [5] a tipy *na Moodle*), což bude řádně zdokumentováno (proč, kde, jak, popis omezení). [1 b]

4 Interpret XML reprezentace kódu (`interpret.py`)

Skript (`interpret.py` v jazyce Python 3.10) načte XML reprezentaci programu a tento program s využitím vstupu dle parametrů příkazové řádky interpretuje a generuje výstup. Vstupní XML reprezentace je definována trochu volněji než XML generované skriptem `parse.php`, ale lze předpokládat, že vstupní XML reprezentace již nebude obsahovat chyby, jež měl ze zadání za úkol detekovat `parse.php`. Interpret navíc oproti sekci 3.1 podporuje existenci volitelných dokumentačních textových atributů `name` a `description` v kořenovém elementu `program` a různě naformátované značky (např. volitelné využití zkráceného zápisu značek, pokud neobsahuje značka podelement). Sémantika jednotlivých instrukcí IPPcode23 je popsána v sekci 5. Interpretace instrukcí probíhá dle atributu `order` vzestupně (nicméně, sekvence nemusí být souvislá/seřazená na rozdíl od sekce 3.1).

Tento skript bude pracovat s těmito parametry:

- `--help` viz společný parametr všech skriptů v sekci 2.2;
- `--source=`*`file`* vstupní soubor s XML reprezentací zdrojového kódu dle definice ze sekce 3.1 a doplnění v úvodu sekce 4;
- `--input=`*`file`* soubor se vstupy[12] pro samotnou interpretaci zadaného zdrojového kódu.

Alespoň jeden z parametrů (`--source` nebo `--input`) musí být vždy zadán. Pokud jeden z nich chybí, jsou chybějící data načítána ze standardního vstupu.

Chybové návratové kódy specifické pro interpret:

- 31 - chybný XML formát ve vstupním souboru (soubor není tzv. dobře formátovaný, angl. *well-formed*, viz [1]);
- 32 - neočekávaná struktura XML (např. element pro argument mimo element pro instrukci, instrukce s duplicitním pořadím nebo záporným pořadím);

Chybové návratové kódy interpretu v případě chyby během interpretace jsou uvedeny v popisu jazyka IPPcode23 (viz sekce 5.1).

[12] Vstup/vstupní soubor může být prázdný; např. neinterpretuje-li se žádná instrukce READ.

Dokumentace a objektově-orientovaný (OO) návrh: Kód skriptu bude navržen objektově, což znamená, že je třeba se seznámit s terminologií OOP (min. v rozsahu přednášek), zamyslet se nad možnostmi využití OO v Python 3 pro tuto úlohu, návrh stručně a jasně (správnou terminologií) popsat v dokumentaci (až 1 bod z hodnocení dokumentace) a samozřejmě také implementovat. Vhodným způsobem promítněte OO návrh do kódu (např. volitelné otypování signatur metod a funkcí) i do komentářů ve zdrojovém kódu (např. uvádění účelu a návrhové myšlenky u definice třídy apod.). I když nebudete implementovat rozšíření, tak bude váš návrh hodnocen i z hlediska intuitivnosti pro rozšíření implementace o některé bonusové rozšíření, což popište i v dokumentaci spolu s určením, zda takový doplňující návrh byl také implementován, či nikoli (až 1 bod z hodnocení dokumentace a úpravy zdrojového kódu).

Dokumentace bude tedy povinně obsahovat popis OO návrhu včetně vhodného UML diagramu s potřebným popisem. Neimplementované části UML diagramu vhodně odlište (např. sazbou šedou barvou). Zvažte též využití návrhových vzorů (více viz *Poznámky k projektu na Moodle*).

Doporučení: Doporučujeme použít knihovnu pro načítání XML. V případě nekompletní implementace se zaměřte na funkčnost globálního rámce, práce s proměnnými typu `int`, instrukce `WRITE` a instrukce pro řízení toku programu.

4.1 Bonusová rozšíření

FLOAT Podpora typu `float` v IPPcode23 (dekadický i **hexadecimální zápis** v analyzátoru i v interpretu včetně načítání ze standardního vstupu; např. `float@0x1.2000000000000p+0` reprezentuje 1,125; viz funkce `float.fromhex()` a `float.hex()` v Python 3). Podporujte instrukce pro práci s tímto typem: INT2FLOAT, FLOAT2INT, aritmetické instrukce (včetně DIV), atd. (viz [3]). Podpora v `parse.php` je možná, ale nebude testována. [1 b]

STACK Podpora zásobníkových variant instrukcí (přípona S; viz [3]): CLEARS, ADDS/SUBS/MULS/IDIVS, LTS/GTS/EQS, ANDS/ORS/NOTS, INT2CHARS/STRI2INTS a JUMPIFEQS/JUMPIFNEQS. Zásobníkové verze instrukcí z datového zásobníku vybírají operandy se vstupními hodnotami dle popisu tříadresné instrukce od konce (tj. typicky nejprve $\langle symb_2 \rangle$ a poté $\langle symb_1 \rangle$). Podpora v `parse.php` je možná, ale nebude testována. [1 b]

STATI Sbírání statistik interpretace kódu. Skript bude podporovat parametr `--stats=`*file* pro zadání souboru *file*, kam se agregované statistiky budou vypisovat (po řádcích dle pořadí v dalších (případně opakovaných) parametrech; na každý řádek nevypisujte nic kromě požadovaného číselného či řetězcového výstupu). Podpora parametru `--insts` pro výpis počtu tzv. vykonaných instrukcí (tj. kromě ladicích instrukcí a speciální instrukce `LABEL`) během interpretace do statistik. Parametr `--hot` vypíše do statistik hodnotu atributu `order` u vykonané instrukce, která byla provedena nejvícekrát a má hodnotu atributu `order` nejmenší. Podpora parametru `--vars` pro výpis maximálního počtu inicializovaných proměnných přítomných v jeden okamžik ve všech platných rámcích během interpretace zadaného programu do statistik. Parametr `--frequent` vypíše do statistik jména operačních kódů (velkými písmeny, čárkou oddělené, bez mezer), které jsou ve zdrojovém kódu nejčastější. Parametr `--print=`*string* vypíše do statistik řetězec *string* a parametr `--eol` vytiskne do statistik odřádkování. Chybí-li při zadání některého z akčních parametrů úvodní parametr `--stats`, jedná se o chybu 10. [1 b]

5 Popis jazyka IPPcode23

Nestrukturovaný imperativní jazyk IPPcode23 vznikl úpravou jazyka IFJcode22 (jazyk pro mezikód překladače jazyka IFJ22, viz [3]), který zahrnuje instrukce tříadresné (typicky se třemi argumenty) a případně zásobníkové (typicky méně parametrů a pracující s hodnotami na datovém zásobníku). Každá instrukce se skládá z operačního kódu (klíčové slovo s názvem instrukce), u kterého nezáleží na velikosti písmen (tj. case insensitive). Zbytek instrukcí tvoří operandy, u kterých na velikosti písmen záleží (tzv. case sensitive). Operandy oddělujeme libovolným nenulovým počtem mezer či tabulátorů. Také před operačním kódem a za posledním operandem se může vyskytnout libovolný počet mezer či tabulátorů. Odřádkování slouží pro oddělení jednotlivých instrukcí, takže na každém řádku je maximálně jedna instrukce a není povoleno jednu instrukci zapisovat na více řádků. Každý operand je tvořen proměnnou, konstantou, typem nebo návěštím. V IPPcode23 jsou podporovány jednořádkové komentáře začínající mřížkou (#). Vyjma prázdných[13] či zakomentovaných řádků obsahuje kód v jazyce IPPcode23 na začátku místo instrukce jen identifikátor jazyka (tečka následovaná jménem jazyka, kdy nezáleží na velikosti písmen):

```
.IPPcode23
```

5.1 Návratové hodnoty interpretu

Proběhne-li interpretace bez chyb, vrací se návratová hodnota `0` (nula). Chybovým případům odpovídají následující návratové hodnoty:

- 52 - chyba při sémantických kontrolách vstupního kódu v IPPcode23 (např. použití nedefinovaného návěští, redefinice proměnné);
- 53 - běhová chyba interpretace – špatné typy operandů;
- 54 - běhová chyba interpretace – přístup k neexistující proměnné (rámec existuje);
- 55 - běhová chyba interpretace – rámec neexistuje (např. čtení z prázdného zásobníku rámců);
- 56 - běhová chyba interpretace – chybějící hodnota (v proměnné, na datovém zásobníku nebo v zásobníku volání);
- 57 - běhová chyba interpretace – špatná hodnota operandu (např. dělení nulou, špatná návratová hodnota instrukce EXIT);
- 58 - běhová chyba interpretace – chybná práce s řetězcem.

5.2 Paměťový model

Hodnoty během interpretace nejčastěji ukládáme do pojmenovaných proměnných, které jsou sdružovány do tzv. rámců, což jsou v podstatě slovníky proměnných s jejich hodnotami. IPPcode23 nabízí tři druhy rámců:

- globální, značíme `GF` (Global Frame), který je na začátku interpretace automaticky inicializován jako prázdný; slouží pro ukládání globálních proměnných;
- lokální, značíme `LF` (Local Frame), který je na začátku nedefinován a odkazuje na vrcholový/aktuální rámec na zásobníku rámců; slouží pro ukládání lokálních proměnných funkcí (zásobník rámců lze s výhodou využít při zanořeném či rekurzivním volání funkcí);

[13] Řádek obsahující pouze bílé znaky je považován za prázdný.

- dočasný, značíme `TF` (Temporary Frame), který slouží pro chystání nového nebo úklid starého rámce (např. při volání nebo dokončování funkce), jenž může být přesunut na zásobník rámců a stát se aktuálním lokálním rámcem. Na začátku interpretace je dočasný rámec nedefinovaný.

K překrytým (dříve vloženým) lokálním rámcům v zásobníku rámců nelze přistoupit dříve, než vyjmeme později přidané rámce.

Další možností pro ukládání nepojmenovaných hodnot je datový zásobník využívaný zásobníkovými instrukcemi.

5.3 Datové typy

IPPcode23 pracuje s typy operandů dynamicky, takže je typ proměnné (resp. paměťového místa) dán obsaženou hodnotou. Není-li řečeno jinak, jsou implicitní konverze zakázány. Interpret podporuje speciální hodnotu/typ nil a tři základní datové typy (int, bool a string), jejichž rozsahy i přesnosti jsou kompatibilní s jazykem Python 3.

Zápis každé konstanty v IPPcode23 se skládá ze dvou částí oddělených zavináčem (znak `@`; bez bílých znaků), označení typu konstanty (int, bool, string, nil) a samotné konstanty (číslo, literál, nil). Např. `bool@true`, `nil@nil` nebo `int@-5`.

Typ int reprezentuje celé číslo (přetečení/podtečení neřešte). Typ bool reprezentuje pravdivostní hodnotu (`false` nebo `true`). Literál pro typ string je v případě konstanty zapsán jako sekvence tisknutelných znaků v kódování UTF-8 (vyjma bílých znaků, mřížky (#) a zpětného lomítka (\)) a escape sekvencí, takže není ohraničen uvozovkami. Escape sekvence, která je nezbytná pro znaky s dekadickým kódem 000–032, 035 a 092, je tvaru $\backslash xyz$, kde xyz je dekadické číslo v rozmezí 000–999 složené právě ze tří číslic[14]; např. konstanta

`string@řetězec\032s\032lomítkem\032\092\032a\010novým\035řádkem`

reprezentuje řetězec

```
řetězec s lomítkem \ a
novým#řádkem
```

Pokus o práci s neexistující proměnnou (čtení nebo zápis) vede na chybu 54. Pokus o čtení hodnoty neinicializované proměnné vede na chybu 56. Pokus o interpretaci instrukce s operandy nevhodných typů dle popisu dané instrukce vede na chybu 53.

5.4 Instrukční sada

U popisu instrukcí sázíme operační kód tučně a operandy zapisujeme pomocí neterminálních symbolů (případně číslovaných) v úhlových závorkách. Neterminál ⟨*var*⟩ značí proměnnou, ⟨*symb*⟩ konstantu nebo proměnnou, ⟨*label*⟩ značí návěští. Identifikátor proměnné se skládá ze dvou částí oddělených zavináčem (znak `@`; bez bílých znaků), označení rámce `LF`, `TF` nebo `GF` a samotného jména proměnné (sekvence libovolných alfanumerických a speciálních znaků bez bílých znaků začínající písmenem nebo speciálním znakem, kde speciální znaky jsou: `_`, `-`, `$`, `&`, `%`, `*`, `!`, `?`). Např. `GF@_x` značí proměnnou `_x` uloženou v globálním rámci.

Na zápis návěští se vztahují stejná pravidla jako na jméno proměnné (tj. část identifikátoru za zavináčem).

Příklad jednoduchého programu v IPPcode23:

[14] Zápis znaků s kódem Unicode větším jak 126 pomocí těchto escape sekvencí nebudeme testovat.

```
.IPPcode23
DEFVAR GF@counter
MOVE GF@counter string@  #Inicializace proměnné na prázdný řetězec
#Jednoduchá iterace, dokud nebude splněna zadaná podmínka
LABEL while
JUMPIFEQ end GF@counter string@aaa
WRITE string@Proměnná\032GF@counter\032obsahuje\032
WRITE GF@counter
WRITE string@\010
CONCAT GF@counter GF@counter string@a
JUMP while
LABEL end
```

Instrukční sada nabízí instrukce pro práci s proměnnými v rámcích, různé skoky, operace s datovým zásobníkem, aritmetické, logické a relační operace, dále také konverzní, vstupně/výstupní a ladicí instrukce.

5.4.1 Práce s rámci, volání funkcí

MOVE ⟨*var*⟩ ⟨*symb*⟩ — Přiřazení hodnoty do proměnné

Zkopíruje hodnotu ⟨*symb*⟩ do ⟨*var*⟩. Např. `MOVE LF@par GF@var` provede zkopírování hodnoty proměnné `var` v globálním rámci do proměnné `par` v lokálním rámci.

CREATEFRAME — Vytvoř nový dočasný rámec

Vytvoří nový dočasný rámec a zahodí případný obsah původního dočasného rámce.

PUSHFRAME — Přesun dočasného rámce na zásobník rámců

Přesuň `TF` na zásobník rámců. Rámec bude k dispozici přes `LF` a překryje původní rámce na zásobníku rámců. `TF` bude po provedení instrukce nedefinován a je třeba jej před dalším použitím vytvořit pomocí CREATEFRAME. Pokus o přístup k nedefinovanému rámci vede na chybu 55.

POPFRAME — Přesun aktuálního rámce do dočasného

Přesuň vrcholový rámec `LF` ze zásobníku rámců do `TF`. Pokud žádný rámec v `LF` není k dispozici, dojde k chybě 55.

DEFVAR ⟨*var*⟩ — Definuj novou proměnnou v rámci

Definuje proměnnou v určeném rámci dle ⟨*var*⟩. Tato proměnná je zatím neinicializovaná a bez určení typu, který bude určen až přiřazením nějaké hodnoty. Opakovaná definice proměnné již existující v daném rámci vede na chybu 52.

CALL ⟨*label*⟩ — Skok na návěští s podporou návratu

Uloží inkrementovanou aktuální pozici z interního čítače instrukcí do zásobníku volání a provede skok na zadané návěští (případnou přípravu rámce musí zajistit jiné instrukce).

RETURN — Návrat na pozici uloženou instrukcí CALL

Vyjme pozici ze zásobníku volání a skočí na tuto pozici nastavením interního čítače instrukcí (úklid lokálních rámců musí zajistit jiné instrukce). Provedení instrukce při prázdném zásobníku volání vede na chybu 56.

5.4.2 Práce s datovým zásobníkem

Operační kód zásobníkových instrukcí je zakončen písmenem „S“. Zásobníkové instrukce případně načítají chybějící operandy z datového zásobníku a výslednou hodnotu operace případně ukládají zpět na datový zásobník.

PUSHS ⟨*symb*⟩ — Vlož hodnotu na vrchol datového zásobníku

Uloží hodnotu ⟨*symb*⟩ na datový zásobník.

POPS ⟨*var*⟩ Vyjmi hodnotu z vrcholu datového zásobníku

Není-li zásobník prázdný, vyjme z něj hodnotu a uloží ji do proměnné ⟨*var*⟩, jinak dojde k chybě 56.

5.4.3 Aritmetické, relační, booleovské a konverzní instrukce

V této sekci jsou popsány tříadresné instrukce pro klasické operace pro výpočet výrazu. Přetečení nebo podtečení číselného výsledku neřešte.

ADD ⟨*var*⟩ ⟨$symb_1$⟩ ⟨$symb_2$⟩ Součet dvou číselných hodnot

Sečte ⟨$symb_1$⟩ a ⟨$symb_2$⟩ (musí být typu int) a výslednou hodnotu téhož typu uloží do proměnné ⟨*var*⟩.

SUB ⟨*var*⟩ ⟨$symb_1$⟩ ⟨$symb_2$⟩ Odečítání dvou číselných hodnot

Odečte ⟨$symb_2$⟩ od ⟨$symb_1$⟩ (musí být typu int) a výslednou hodnotu téhož typu uloží do proměnné ⟨*var*⟩.

MUL ⟨*var*⟩ ⟨$symb_1$⟩ ⟨$symb_2$⟩ Násobení dvou číselných hodnot

Vynásobí ⟨$symb_1$⟩ a ⟨$symb_2$⟩ (musí být typu int) a výslednou hodnotu téhož typu uloží do proměnné ⟨*var*⟩.

IDIV ⟨*var*⟩ ⟨$symb_1$⟩ ⟨$symb_2$⟩ Dělení dvou celočíselných hodnot

Celočíselně podělí celočíselnou hodnotu ze ⟨$symb_1$⟩ druhou celočíselnou hodnotou ze ⟨$symb_2$⟩ (musí být oba typu int) a výsledek typu int přiřadí do proměnné ⟨*var*⟩. Dělení nulou způsobí chybu 57.

LT/GT/EQ ⟨*var*⟩ ⟨$symb_1$⟩ ⟨$symb_2$⟩ Relační operátory menší, větší, rovno

Instrukce vyhodnotí relační operátor mezi ⟨$symb_1$⟩ a ⟨$symb_2$⟩ (stejného typu; int, bool nebo string) a do ⟨*var*⟩ zapíše výsledek typu bool (`false` při neplatnosti nebo `true` v případě platnosti odpovídající relace). Řetězce jsou porovnávány lexikograficky a `false` je menší než `true`. Pro výpočet neostrých nerovností lze použít AND/OR/NOT. S operandem typu nil (další zdrojový operand je libovolného typu) lze porovnávat pouze instrukcí EQ, jinak chyba 53.

AND/OR/NOT ⟨*var*⟩ ⟨$symb_1$⟩ ⟨$symb_2$⟩ Základní booleovské operátory

Aplikuje konjunkci (logické A)/disjunkci (logické NEBO) na operandy typu bool ⟨$symb_1$⟩ a ⟨$symb_2$⟩ nebo negaci na ⟨$symb_1$⟩ (NOT má pouze 2 operandy) a výsledek typu bool zapíše do ⟨*var*⟩.

INT2CHAR ⟨*var*⟩ ⟨*symb*⟩ Převod celého čísla na znak

Číselná hodnota ⟨*symb*⟩ je dle Unicode převedena na znak, který tvoří jednoznakový řetězec přiřazený do ⟨*var*⟩. Není-li ⟨*symb*⟩ validní ordinální hodnota znaku v Unicode (viz funkce `chr` v Python 3), dojde k chybě 58.

STRI2INT ⟨*var*⟩ ⟨$symb_1$⟩ ⟨$symb_2$⟩ Ordinální hodnota znaku

Do ⟨*var*⟩ uloží ordinální hodnotu znaku (dle Unicode) v řetězci ⟨$symb_1$⟩ na pozici ⟨$symb_2$⟩ (indexováno od nuly). Indexace mimo daný řetězec vede na chybu 58. Viz funkce `ord` v Python 3.

5.4.4 Vstupně-výstupní instrukce

READ ⟨*var*⟩ ⟨*type*⟩ Načtení hodnoty ze standardního vstupu

Načte jednu hodnotu dle zadaného typu ⟨*type*⟩ ∈ {int, string, bool} a uloží tuto hodnotu do proměnné ⟨*var*⟩. Načtení proveďte vestavěnou funkcí `input()` (či analogickou) jazyka Python 3, pak proveďte konverzi na specifikovaný typ ⟨*type*⟩. Při převodu vstupu na typ bool nezáleží na velikosti písmen a řetězec „true“ se převádí na `bool@true`, vše ostatní na `bool@false`. V případě chybného nebo chybějícího vstupu bude do proměnné ⟨*var*⟩ uložena hodnota `nil@nil`.

WRITE ⟨*symb*⟩ Výpis hodnoty na standardní výstup

Vypíše hodnotu ⟨*symb*⟩ na standardní výstup. Až na typ bool a hodnotu `nil@nil` je formát výpisu kompatibilní s příkazem `print` jazyka Python 3 s doplňujícím parametrem `end=''` (zamezí dodatečnému odřádkování). Pravdivostní hodnota se vypíše jako `true` a nepravda jako `false`. Hodnota `nil@nil` se vypíše jako prázdný řetězec.

5.4.5 Práce s řetězci

CONCAT ⟨*var*⟩ ⟨*symb*$_1$⟩ ⟨*symb*$_2$⟩ Konkatenace dvou řetězců

Do proměnné ⟨*var*⟩ uloží řetězec vzniklý konkatenací dvou řetězcových operandů ⟨*symb*$_1$⟩ a ⟨*symb*$_2$⟩ (jiné typy nejsou povoleny).

STRLEN ⟨*var*⟩ ⟨*symb*⟩ Zjisti délku řetězce

Zjistí počet znaků (délku) řetězce v ⟨*symb*⟩ a tato délka je uložena jako celé číslo do ⟨*var*⟩.

GETCHAR ⟨*var*⟩ ⟨*symb*$_1$⟩ ⟨*symb*$_2$⟩ Vrať znak řetězce

Do ⟨*var*⟩ uloží řetězec z jednoho znaku v řetězci ⟨*symb*$_1$⟩ na pozici ⟨*symb*$_2$⟩ (indexováno celým číslem od nuly). Indexace mimo daný řetězec vede na chybu 58.

SETCHAR ⟨*var*⟩ ⟨*symb*$_1$⟩ ⟨*symb*$_2$⟩ Změň znak řetězce

Zmodifikuje znak řetězce uloženého v proměnné ⟨*var*⟩ na pozici ⟨*symb*$_1$⟩ (indexováno celočíselně od nuly) na znak v řetězci ⟨*symb*$_2$⟩ (první znak, pokud obsahuje ⟨*symb*$_2$⟩ více znaků). Výsledný řetězec je opět uložen do ⟨*var*⟩. Při indexaci mimo řetězec ⟨*var*⟩ nebo v případě prázdného řetězce v ⟨*symb*$_2$⟩ dojde k chybě 58.

5.4.6 Práce s typy

TYPE ⟨*var*⟩ ⟨*symb*⟩ Zjisti typ daného symbolu

Dynamicky zjistí typ symbolu ⟨*symb*⟩ a do ⟨*var*⟩ zapíše řetězec značící tento typ (int, bool, string nebo nil). Je-li ⟨*symb*⟩ neinicializovaná proměnná, označí její typ prázdným řetězcem.

5.4.7 Instrukce pro řízení toku programu

Neterminál ⟨*label*⟩ označuje návěští, které slouží pro označení pozice v kódu IPPcode23. V případě skoku na neexistující návěští dojde k chybě 52.

LABEL ⟨*label*⟩ Definice návěští

Speciální instrukce označující pomocí návěští ⟨*label*⟩ důležitou pozici v kódu jako potenciální cíl libovolné skokové instrukce. Pokus o vytvoření dvou stejně pojmenovaných návěští na různých místech programu je chybou 52.

JUMP ⟨*label*⟩ Nepodmíněný skok na návěští

Provede nepodmíněný skok na zadané návěští ⟨*label*⟩.

JUMPIFEQ ⟨*label*⟩ ⟨*symb*$_1$⟩ ⟨*symb*$_2$⟩ Podmíněný skok na návěští při rovnosti

Pokud jsou ⟨*symb*$_1$⟩ a ⟨*symb*$_2$⟩ stejného typu nebo je některý operand nil (jinak chyba 53) a zároveň se jejich hodnoty rovnají, tak provede skok na návěští ⟨*label*⟩.

JUMPIFNEQ ⟨*label*⟩ ⟨*symb*$_1$⟩ ⟨*symb*$_2$⟩ Podmíněný skok na návěští při nerovnosti

Jsou-li ⟨*symb*$_1$⟩ a ⟨*symb*$_2$⟩ stejného typu nebo je některý operand nil (jinak chyba 53), tak v případě různých hodnot provede skok na návěští ⟨*label*⟩.

EXIT ⟨*symb*⟩ Ukončení interpretace s návratovým kódem

Ukončí vykonávání programu, případně vypíše statistiky a ukončí interpret s návratovým kódem ⟨*symb*⟩, kde ⟨*symb*⟩ je celé číslo v intervalu 0 až 49 (včetně). Nevalidní celočíselná hodnota ⟨*symb*⟩ vede na chybu 57.

5.4.8 Ladicí instrukce

Následující ladicí instrukce (DPRINT a BREAK) nesmí ovlivňovat standardní výstup. Jejich skutečnou funkcionalitu nebudeme testovat, ale mohou se v testech objevit.

DPRINT ⟨*symb*⟩	Výpis hodnoty na `stderr`
Předpokládá se, že vypíše zadanou hodnotu ⟨*symb*⟩ na standardní chybový výstup (stderr).	

BREAK	Výpis stavu interpretu na `stderr`
Předpokládá se, že na standardní chybový výstup (stderr) vypíše stav interpretu (např. pozice v kódu, obsah rámců, počet vykonaných instrukcí) v danou chvíli (tj. během vykonávání této instrukce).	

Reference

[1] Extensible Markup Language (XML) 1.0. W3C. World Wide Web Consortium [online]. 5. vydání. 26. 11. 2008 [cit. 2020-02-03]. Dostupné z: `https://www.w3.org/TR/xml/`

[2] A7Soft JExamXML is a java based command line XML diff tool for comparing and merging XML documents. c2018 [cit. 2020-02-03]. Dostupné z: `https://www.a7soft.com/jexamxml.html`

[3] Křivka, Z., a kol.: Zadání projektu z předmětu IFJ a IAL. c2022 [cit. 2023-01-24]. Dostupné z: `https://www.fit.vutbr.cz/study/courses/IFJ/private/projekt/ifj2022.pdf`

[4] The text/markdown Media Type. Internet Engineering Task Force (IETF). 2016 [cit. 2020-02-03]. Dostupné z: `https://tools.ietf.org/html/rfc7763`

[5] Gamma, E., a kol.: Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley, 1994.