

Implementace jazyka TIL-Script

Implementation of the TIL-Script Language

Bc. Filip Peterek

Diplomová práce

Vedoucí práce: prof. RNDr. Marie Duží, CSc.

Ostrava, 2023

Zadání diplomové práce

Student:

Bc. Filip Peterek

Studijní program:

N2647 Informační a komunikační technologie

Studijní obor:

2612T025 Informatika a výpočetní technika

Téma:

Implementace jazyka TIL-Script
Implementation of the TIL-Script Language

Jazyk vypracování:

čeština

Zásady pro vypracování:

Cílem práce je implementovat funkcionální jazyk TIL-Script. Implementace bude vycházet z gramatiky jazyka, která je v souladu s logickým systémem Transparentní Intensionální Logiky (TIL).

Práce bude obsahovat:

1. Popis systému TIL, tj. jazyk konstrukcí a rozvětvená teorie typů.
2. Definici jazyka TIL-Script.
3. Analýzu, návrh a implementaci jazyka TIL-Script včetně typové kontroly.
4. Dokumentace celého projektu včetně uživatelské příručky.

Implementace bude provedena v jazyce C# nebo Java.

Seznam doporučené odborné literatury:

- [1] Duží M., Materna P. (2012): TIL jako procedurální logika (průvodce zvědavého čtenáře Transparentní intensionální logikou). Aleph Bratislava 2012, ISBN 978-80-89491-08-7
- [2] Duží M., Jespersen B. and Materna P. (2010): Procedural Semantics for Hyperintensional Logic. Foundations and Applications of Transparent Intensional Logic. First edition. Berlin: Springer, series Logic, Epistemology, and the Unity of Science, vol. 17, ISBN 978-90-481-8811-6.
- [3] Ciprich N., Duží, M., Košinár M. (2009): The TIL-Script Language. Frontiers in Artificial Intelligence and Applications, Amsterdam: IOS Press, vol. 190, pp. 166-179.

Formální náležitosti a rozsah diplomové práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

Vedoucí diplomové práce: **prof. RNDr. Marie Duží, CSc.**

Datum zadání: 01.09.2022

Datum odevzdání: 30.04.2023

Garant studijního oboru: prof. RNDr. Václav Snášel, CSc.

V IS EDISON zadáno: 07.11.2022 11:59:22

Abstrakt

Cílem práce je implementovat programovací jazyk TIL-Script. Jazyk TIL-Script slouží jako výpočetní varianta logického kalkulu TIL, jenž umožňuje jednoduchý strojový zápis konstrukcí Transparentní intenzionální logiky, ale také jejich následné provedení. Práce dále řeší praktické problémy s interpretací TIL-Scriptu, a to například definice pojmenovaných funkcí, interakce s databází, apod. Dále se práce snaží navrhnout nadmnožinu jazyka TIL-Script, která umožní konstrukce TILu nejen provádět, ale také analyzovat, vytvářet je, a pracovat s nimi.

Klíčová slova

Transparentní intenzionální logika, TIL-Script, překladač

Abstract

The goal of the thesis is the definition and implementation of the TIL-Script language. TIL-Script is a scripting language which serves the purpose of a computational variant of Transparent intensional logic, a logical calculus based on typed lambda calculi. TIL-Script allows for not just representation, but also execution of TIL constructions. This work also deals with practical problems of TIL-Script implementation, such as definitions of named functions, interaction with databases, and so on. Furthermore, this thesis attempts to define a superset of the TIL-Script language, which allows for not just the execution of constructions, but also for their creation and analysis.

Keywords

Transparent intensional logic, TIL-Script, interpreter

Obsah

Seznam použitých symbolů a zkratk	6
Seznam obrázků	7
Seznam tabulek	8
1 Úvod	9
2 Transparentní intenzionální logika	11
2.1 Objektová báze	12
2.2 Funkce	13
2.3 Konstrukce TIL	13
2.4 Typy 1. řádu	15
2.5 Rozvětvená hierarchie typů	15
2.6 Analytické a empirické výrazy	16
3 TIL-Script	17
3.1 Charakteristické rysy TIL-Scriptu	17
3.2 TIL-Script jako výpočetní varianta TILu	20
3.3 Rozšíření TIL-Scriptu	28
4 Implementace	36
4.1 Zvolené technologie	36
4.2 Architektura projektu	37
4.3 Implementace překladače	38
Přílohy	41

Seznam použitých zkratek a symbolů

TIL	– Transparentní intenzionální logika
JVM	– Java Virtual Machine
JRE	– Java Runtime Environment
JAR	– Java Archive
TCO	– Tail Call Optimization
REPL	– Read-Eval-Print Loop
CLI	– Command Line Interface
AST	– Abstract Syntax Tree

Seznam obrázků

2.1	Schéma procedurální sémantiky TIL	12
4.1	Komponenty projektu	37

Seznam tabulek

2.1	Výchozí báze pro analýzu přirozeného jazyka	12
-----	---	----

Kapitola 1

Úvod

Analýza přirozeného jazyka jako disciplína stále rychleji stoupá na oblibě i důležitosti. Jistě málokomu unikly například n -gramové modely založené na predikci následujícího slova z předcházejících n slov, či vektorové modely jako Word2Vec, umožňující reprezentovat význam slov pomocí vektorů. Poslední dobou se velmi často mluví o jazykovém modelu GPT-3.

Výčet přístupů k analýze přirozeného jazyka však nekončí n -gramy a neuronovými sítěmi. K analýze přirozených jazyků lze přistoupit také pomocí logiky. Mezi průkopníky tohoto přístupu patřil také český logik Pavel Tichý, tvůrce Transparentní intenzionální logiky.

Transparentní intenzionální logika (dále také TIL) je logický systém založený na typovaném lambda kalkulu. TIL umožňuje modelovat a analyzovat výrazy přirozeného jazyka za pomoci logiky. Jako doplněk TILu poté vznikl také funkcionální programovací jazyk TIL-Script, sloužící k interpretaci konstrukcí Transparentní intenzionální logiky. Syntax i sémantika TIL-Scriptu jsou silně inspirovány TILEm, ovšem přirozeně s určitými úpravami tak, aby byl TIL-Script rozumně zapisovatelný a interpretovatelný na počítači.

A právě interpretaci TIL-Scriptu se zabývá tento text. Součástí práce tak je navázání na předcházející vývoj jazyka TIL-Script, rozšíření gramatiky TIL-Scriptu nutné pro akomodaci nových prvků jazyka (např. direktiva *import* sloužící k importu symbolů definovaných v jiném souboru), nebo také samotná interpretace jazyka TIL-Script, společně s kontrolou typové koherence. Dále práce navrhuje způsob, jak volat funkce platformy JVM. Tento přístup otevírá jazyku TIL-Script celý ekosystém platformy JVM. Zatímco tedy práce nijak neřeší a neimplementuje získávání informací z již existující databáze, soustředí se spíše na zjednodušení přístupu k existujícím knihovnám, aby do budoucna nebyl problém rychle a jednoduše naprogramovat přístup k libovolnému zdroji dat.

Nakonec se práce snaží navrhnout určitou nadmnožinu jazyka TIL-Script tak, aby TIL-Script mohl sloužit nejen k interpretaci konstrukcí TIL, ale také k jejich tvorbě a analýze. V současné době pro analýzu a práci s konstrukcemi TIL-Scriptu existují nástroje psané v jazyce Java. Existuje-li však v současné době programovací jazyk vytvořený přímo pro logiky pracující s TILEm, není důvod tento

jazyk nevyužít a nerozšířit také o nástroje pro analýzu, tvorbu i transformaci TIL-Script konstrukcí. Pro pokročilejší práci s TIL-Scriptem by tak nebylo třeba učit se jiný programovací jazyk (Java), nebo využívat externí nástroje, ale naopak by byl k dispozici již familiérní nástroj, jehož základy jsou již příznivům Transparentní intenzionální logiky dobře známy. Cílem této nadmnožiny ovšem není předefinovat TIL, či nějak výrazně měnit jádro TIL-Scriptu, jinak by se ostatně také nemohlo jednat o nadmnožinu. Tyto nové prvky jsou od jádra TIL-Scriptu, které slouží převážně jako výpočetní varianta TIL, jednoduše oddělitelné. Při práci s interpreterem, jehož implementace je součástí této práce, není problém se této nadmnožině vyhnout a nevyužívat ji. Také lze, na základě tohoto textu, implementovat interpreter samotného TIL-Scriptu bez jakýchkoliv rozšíření.

Práce se do značné míry inspirovuje výzkumem amerického profesora Johna McCarthyho a programovacím jazykem Lisp. Jelikož Lisp vychází z lambda kalkulu, pracuje s parciálními funkcemi, a nerozlišuje mezi programem a daty (tedy Lisp programy mohou pracovat s, generovat a transformovat Lisp konstrukce), lze mezi Lispem a TILem nalézt řadu podobností. Primárním odlišovacím znakem Transparentní intenzionální logiky je ovšem rigorózně definovaný typový systém.

Text samotný je rozdělen do čtyř částí. V první části budou představeny základy Transparentní intenzionální logiky, jejichž znalost je nutná k pochopení práce. Druhá část představí programovací jazyk TIL-Script a vyznačí změny a úpravy, které tato práce navrhuje. Třetí část popíše technické řešení a implementaci interpreteru. V poslední části poté bude zdokumentována standardní i matematická knihovna, které jsou součástí implementace. Dále bude poslední část obsahovat návod na použití, návod na implementaci vlastní knihovny, a nakonec také ukázky programů psaných v jazyce TIL-Script.

V terminologii programovacích jazyků a překladačů existuje spousta zavedených anglických názvů, jejichž český ekvivalent je méně zavedený, případně neexistuje vůbec. Proto bude práce ve spoustě případů uvádět také anglický ekvivalent k českému výrazu.

Kapitola 2

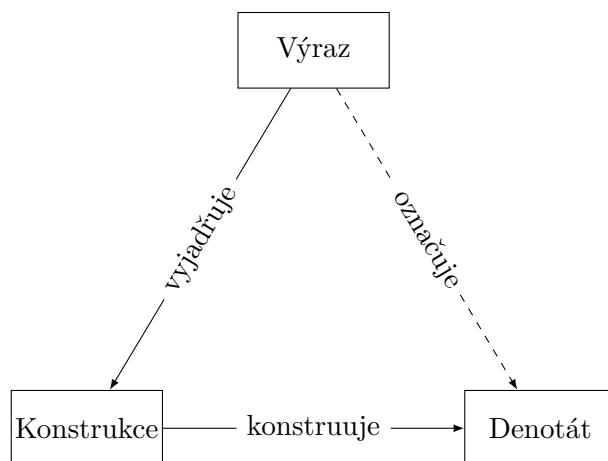
Transparentní intenzionální logika

Transparentní intenzionální logika (TIL) je logický systém založený na typovaném lambda kalkulu. TIL je využíván k logické analýze přirozeného jazyka. Oproti tradičnímu lambda kalkulu, jenž se využívá jako počítačový model, tedy jako pouhý prostředek k dosažení konkrétní hodnoty – výsledku, v Transparentní intenzionální logice hraje konstrukce kalkulu často důležitější roli, než hodnota, kterou by konstrukce po provedení zkonstruovala.

Jako příklad využití lambda kalkulu jako výpočetního modelu lze uvést např. funkcionální programovací jazyk Haskell. Interně je Haskell kompilován do lambda kalkulu (přesněji do jeho supersetu obsahujícího např. čísla nebo logické hodnoty, která jinak v lambda kalkulu musíme kódovat pomocí Churchova kódování – K-kombinátorů, apod.). Ultimátně v Haskellu ovšem lambda kalkul slouží pouze k získání konkrétního výsledku. Nadefinujeme vztah mezi vstupem a výstupem, a program napsaný v Haskellu nám vstup transformuje. Pokud zanedbáme efektivitu programu, nezajímá nás, jakým způsobem program spočítal výsledek, dokud jej spočítal správně.

Naopak Transparentní intenzionální logika je hyperintenzionálním kalkulem, který nám umožňuje vytvářet konstrukce vypovídající o jiných konstrukcích. TIL vychází z myšlenky, že výraz přirozeného jazyka sice označuje denotát – konkrétní objekt (např. individuum, číslo, konstrukci), významem výrazu ovšem není samotný denotát, který ani nemusí nutně existovat. Význam výrazu je abstraktní a lze jej zachytit konstrukcí. Daná konstrukce poté při provedení zkonstruuje denotát výrazu. Jako příklad lze uvést například výraz "francouzský král." V době psaní této práce Francie krále nemá. Výraz nemá žádný denotát, neuvádí žádné konkrétní individuum. Přesto výrazu "francouzský král" rozumíme, výraz má svůj význam, jen v současné době neuvádí žádnou osobu. A budeme-li chtít o významu výrazu "francouzský král" něco vypovědět, například že francouzský král je monarchou v čele Francie, daný monarcha nemusí existovat. Dále lze uvést například rozdíl mezi výrazy "logaritmus 1024 o základě 2" a "5 + 5". Denotátem obou výrazů je 10. Zadáme-li do interpreteru Haskellu výrazy

`logBase 2 1024` a `5 + 5`, získáme v obou případech stejný výsledek. V přirozeném jazyce ovšem chápeme značný rozdíl mezi oběma výrazy, ačkoliv mají stejný denotát. "Logaritmus 1024 o základě



Obrázek 2.1: Schéma procedurální sémantiky TIL

Tabulka 2.1: Výchozí báze pro analýzu přirozeného jazyka

Typ	Popis typu
o	Množina pravdivostních hodnot
ι	Množina individuí (univerzum diskurzu)
τ	Množina časových okamžiků/reálných čísel
ω	Množina logicky možných světů

2"vyjadřuje číslo, kterým musíme umocnit dvojku, abychom získali 1024. Výraz "5 + 5"očividně vyjadřuje úplně jinou matematickou operaci a jeho výsledek spočítáme jiným postupem.

Denotátem výrazu může být nejen objekt z báze, ale i konstrukce nebo funkce.

Jak již bylo zmíněno, Transparentní intenzionální logika vychází z typovaného lambda kalkulu, proto také každý objekt musí mít svůj typ. Pro správné pochopení TILu, a tedy i této práce, je tak nutné znát typovou hierarchii TIL.

2.1 Objektová báze

Objektová báze je kolekce vzájemně disjunktních neprázdných množin, které dohromady vymezují nulární funkce, se kterými budeme pracovat. Bázi volíme dle potřeb konkrétní aplikace a univerza diskurzu. Například používáme-li TIL k logické analýze matematických vět, jako bázi lze zvolit například množinu celých čísel, množinu reálných čísel, a množinu pravdivostních hodnot. Musíme však vzít v potaz, že tato báze neobsahuje čísla komplexní.

Patří-li objekt x do množiny α z báze, říkáme, že se jedná o objekt typu α . K explicitnímu uvedení typu objektu x využíváme zápis x/α . Množinám tvořícím bázi lze přirozeně říkat typy.

Pro analýzu přirozeného jazyka se většinou volí objektová báze skládající se z typů o , ι , τ , ω . Tyto typy jsou podrobněji popsány v tabulce 2.1 – Výchozí báze pro analýzu přirozeného jazyka.

2.2 Funkce

V některých logických systémech, například v predikátové logice, se jako základní molekulární typ využívají relace. Funkce je poté speciální typ relace, která je zprava jednoznačná. V TIL je však základním molekulárním typem funkce. Chceme-li v TIL vyjádřit n -ární relaci nad množinou $\alpha_1 \times \dots \times \alpha_n$, lze tak samozřejmě udělat definicí n -ární funkce z $\alpha_1 \times \dots \times \alpha_n$ do o , která každému prvku z $\alpha_1 \times \dots \times \alpha_n$ přiřadí pravdivostní hodnotu na základě toho, zda prvek do relace patří, nebo ne.

Narozdíl od tradičního lambda kalkulu je Transparentní intenzionální logika kalkulem parciálních funkcí. Z parciality funkcí poté vyplývá další vlastnost TIL – arita funkcí není shora omezená. V lambda kalkulech totálních funkcí lze využít Schönfinkelovu redukci k redukci funkcí n -árních na unární za využití skládání funkcí. Tato redukce však není ekvivalentní, pracujeme-li s funkcemi parciálními.

2.2.1 Intenze a extenze

V TIL dále rozlišujeme funkce na tzv. *intenze* a *extenze*. Intenze jsou funkce z možných světů. Extenze jsou funkce, jejichž doménou množina možných světů není, a tudíž jejich funkční hodnota nezávisí na stavu světa.

Intenze jsou obecně funkce typu $(\alpha\omega)$ pro libovolný typ α . Nejčastěji se však jedná o funkce typu $((\alpha\tau)\omega)$, tedy funkce zobrazující možné světy do chronologií objektů typu α .

2.3 Konstrukce TIL

Konstrukce v Transparentní intenzionální logice jsou abstraktní procedury. Tyto procedury jsou strukturované – nejedná se o množiny, mají pevně danou strukturu, a na uspořádání případných podprocedur záleží. Tyto konstrukce lze podle definovaných pravidel provést. Provedením konstrukce získáme výstup, případně nezískáme nic. Konstrukce, které nekonstruují žádný výstup, nazýváme *nevlastní* (anglicky *improper*). V TIL pracujeme s šesti druhy konstrukcí.

Jak již bylo zmíněno, konstrukce můžou v TIL operovat nejen nad objekty, které nejsou konstrukcemi, tedy nad objekty z báze a funkcemi, ale také nad jinými konstrukcemi. Konstrukce však může operovat pouze nad konstrukcemi nižšího řádu, než je konstrukce samotná, viz 2.5 – Rozvětvená hierarchie typů. Každou podkonstrukci, kterou musíme provést při provedení konstrukce, nazýváme *konstituentem*. V TIL existuje šest různých druhů konstrukcí. Dvě atomické – mají pouze jeden konstituent, a to sebe samotné, a čtyři molekulární. Atomickými konstrukcemi jsou *Trivializace* a *proměnné*. Mezi molekulární konstrukce poté řadíme *Kompozice*, *Uzávěry*, *Provedení* a *Dvojí Provedení*.

Proměnné jsou konstrukce, které na základě valuace v v -konstruují objekty. Skutečnost, že proměnná x v -konstruuje hodnotu typu α značíme $x \rightarrow_v \alpha$.

Trivializace pro libovolný objekt X konstruuje samotný objekt X . Konstrukce je v Transparentní intenzionální logice potřebná, neboť výchozím módem pro konstrukce je provedení. Samotná konstrukce X by tak byla automaticky provedena, a místo konstrukce X bychom dostali pouze její denotát. Pokud bychom chtěli zkonstruovat konstrukci X , musíme ji trivializovat. Tím se provede pouze konstrukce Trivializace. A protože Trivializace nemá jiný konstituent, než sebe samotnou, konstrukce X se tak neprovede. V literatuře se Trivializace X tradičně značí 0X . Alternativně se používá také zápis $'X$. Tento zápis je poté využit i v jazyce TIL-Script. Trivializaci lze považovat za ekvivalent funkce `QUOTE` z jazyka Lisp. Trivializace taktéž bývá využívána ke konstruování hodnot, které nelze provést (objekty z báze, funkce) a tudíž je nelze zmínit netrivializované.¹

Kompozice je procedura aplikace funkce na argumenty. Kompozice $[XY_1...Y_m]$ značí aplikaci funkce konstruované konstrukcí X na argumenty zkonstruované konstrukcemi $Y_1, ..., Y_m$. Pokud konstrukce X v -konstruuje funkci f , všechny podkonstrukce $Y_1, ..., Y_m$ v -konstruují hodnotu, a je-li funkce f na daných argumentech definovaná, kompozice v -konstruuje funkční hodnotu na těchto argumentech. V opačném případě je kompozice v -nevlastní.

Uzávěr $\lambda x_1...x_m Y$ je konstrukce v -konstruující funkci. $x_1, ..., x_m$ musí být navzájem různé proměnné, Y musí být konstrukcí. Konstrukce uzávěru je velmi podobná abstrakci v lambda kalkulu. Narozdíl od lambda kalkulu však v TILu může uzávěr konstruovat funkce s aritou vyšší než jedna. Uzávěr nemůže být nikdy nevlastní, může však konstruovat tzv. *degenerovanou funkci*, tedy funkci, která je nedefinovaná na celém definičním oboru.

Provedení 1X je konstrukce v -konstruující objekt konstruovaný konstrukcí X . Pokud je konstrukce X v -nevlastní, je provedení 1X také v -nevlastní. Jelikož je však provedení výchozím módem pro objekty, většinou se neuvádí. Provést lze pouze konstrukce. Objekty z báze (tedy čísla, individua, apod...) či funkce nelze provést, jejich provedení nekonstruuje nic. Proto je potřeba tyto objekty vždy trivializovat.

Dvojití provedení 2X je poslední z výčtu konstrukcí. Je-li X konstrukcí v -konstruující konstrukci Y , a v -konstruuje-li konstrukce Y objekt Z , pak 2X v -konstruuje Z . V opačném případě je dvojití provedení 2X v -nevlastní.

Jiné konstrukce v Transparentní intenzionální logice neexistují.

2.3.1 Princip kompozicionality

Princip kompozicionality je důležitým rysem Transparentní intenzionální logiky. Z principu kompozicionality vyplývá, že je-li libovolný konstituent konstrukce X v -nevlastní a v dané valuaci nekonstruuje žádnou hodnotu, pak je v -nevlastní i konstrukce X .

¹V jazyce Lisp čísla konstruují sebe samotné, tedy provedením čísla získáme zpět prováděné číslo. 1 a '1 jsou tedy v Lispu ekvivalentní výrazy. V TIL však není možné, aby objekt konstruoval sám sebe, viz rozvětvená hierarchie typů 2.5 – Rozvětvená hierarchie typů.

2.4 Typy 1. řádu

Definice je skoro slovo od slova převzata z knihy *TIL jako procedurální logika – Průvodce zvědavého čtenáře Transparentní intenzionální logikou*. Tato sekce slouží jako krátké vysvětlení základů Transparentní intenzionální logiky se čtenář může obrátit například na tuto knihu.

Nechť B je báze. Pak:

- i) Každá množina z báze B je atomický typ řádu 1 nad B .
- ii) Nechť $\alpha, \beta_1, \dots, \beta_m (m > 0)$ jsou typy řádu 1 nad B . Pak soubor všech m -árních parciálních funkcí $(\alpha\beta_1\dots\beta_m)$, tedy zobrazení z $\beta_1 \times \dots \times \beta_m$ do α , je molekulární typ řádu 1 nad B .
- iii) Nic jiného není typem řádu 1 nad bází B .

2.5 Rozvětvená hierarchie typů

Definice je opět skoro slovo od slova převzata z Průvodce.

Nechť B je báze. Pak:

2.5.1 T_1 (typy řádu 1)

Viz sekce 2.4 – Typy 1. řádu.

2.5.2 C_n (konstrukce řádu n)

- i) Nechť x je proměnná v -konstruující objekt typu řádu n . Pak x je *konstrukce řádu n nad B* .
- ii) Nechť X je prvek typu řádu n . Pak trivializace 0X , provedení 1X a dvojí provedení 2X jsou *konstrukcemi řádu n nad B* .
- iii) Nechť X, Y_1, \dots, Y_m jsou konstrukce řádu n nad B . Pak kompozice $[XY_1\dots Y_m]$ je *konstrukce řádu n nad B* .
- iv) Nechť x_1, \dots, x_m jsou vzájemně různé proměnné a X je konstrukce řádu n nad B . Pak uzávěr $[\lambda x_1\dots x_m X]$ je *konstrukce řádu n nad B* .
- v) Nic jiného není konstrukcí řádu n nad bází B než dle i)-v).

2.5.3 T_{n+1} (typy řádu $n+1$)

Nechť $*_n$ je kolekce všech konstrukcí řádu n nad B .

- i) $*_n$ a každý typ řádu n jsou *typy řádu $n+1$ nad B* .

- ii) Jsou-li $\alpha, \beta_1, \dots, \beta_m$ typy řádu $n + 1$ nad B , pak $(\alpha\beta_1\dots\beta_m)$, tedy kolekce parciálních funkcí, je *typy řádu $n+1$ nad B* .
- iii) Nic jiného není typ řádu $n + 1$ nad B než dle i) a ii).

2.6 Analytické a empirické výrazy

Výrazy přirozeného jazyka lze dělit na dva typy výrazů, a to empirické a analytické.

Analytické výrazy jsou výrazy takové, které označují extenze nebo konstantní intenze. Jedná se například o matematické věty nebo věty vyjadřující relaci rekvizity mezi vlastnostmi (např. věta "Všechny velryby jsou savci" je analytická a nutně pravdivá nezávisle na stavu světa, neboť existuje-li individuum, které je velrybou, pak bude vždy také savcem.)

Empirické výrazy naopak označují intenze, jejichž hodnota na stavu světa závisí. Abychom určili hodnotu dané intenze, musíme empiricky zkoumat stav světa v daném časovém okamžiku. Empirické zkoumání světa ovšem již není záležitost logiky.

Kapitola 3

TIL-Script

Nyní konečně přišel čas představit TIL-Script. TIL-Script je interpretovaný funkcionální programovací jazyk, do znatelné míry inspirovaný jazyky jako Haskell nebo Lisp. Syntax TIL-Scriptu by měla co nejvíce připomínat syntaktické prvky Transparentní intenzionální logiky, aby pouhá znalost TILu stačila k okamžitému pochopení TIL-Scriptu. Sémantika TIL-Script konstrukcí poté musí odpovídat sémantice TIL.

Tato kapitola je rozdělena do tří sekcí. V první sekci jsou popsány důležité základní rysy TIL-Scriptu. Druhá sekce popisuje již existující prvky TIL-Scriptu, případně také dokumentuje změny oproti předchozím verzím TIL-Scriptu. Poslední sekce navrhuje rozšíření jazyka TIL-Script.

3.1 Charakteristické rysy TIL-Scriptu

Tato sekce popisuje charakteristické rysy TIL-Scriptu v takové podobě, jakou nabývá v této práci. Pokud se v některém bodě TIL-Script neshoduje s TILem či předchozími verzemi TIL-Scriptu, je rozdíl náležitě popsán a vysvětlen.

3.1.1 Lambda kalkul parciálních funkcí

3.1.1.1 Shora neomezená arita funkcí

Narozdíl od lambda kalkulu ve své tradiční podobě, nebo například jazyka Haskell, v Transparentní intenzionální logice není arita funkce shora omezená, viz 2.2 – Funkce. TIL-Script musí tento fakt reflektovat. Proto tento jazyk umožňuje definici i aplikaci funkcí libovolné (samozřejmě kladné) arity. Nevyužíváme zde tedy rozvíjení funkcí (anglicky *currying*). Zatímco např. v Haskellu jsou funkce arity dvě nebo vyšší automaticky rozvinuty na sérii několika unárních funkcí, jejichž oborem hodnot jsou unární nebo nulární funkce, v TIL-Scriptu není arita nijak omezená.

3.1.1.2 Parciální funkce a respektování principu kompozicionality

Jelikož v TIL můžou být funkce parciální, musí i TIL-Script počítat s parcialitou funkcí. Dále musí TIL-Script respektovat princip kompozicionality, základní rys Transparentní intenzionální logiky. Jedním z důsledků principu kompozicionality je, že konstrukce, jejíž přinejmenším jeden konstituent je *v*-nevlastní, bude také nutně *v*-nevlastní. Reprezentaci stavu, kdy je parciální funkce aplikována na argumenty, na kterých není definována, se věnuje podsekcce 3.3.7 – Hodnota *Nil* této kapitoly.

Jedinou výjimkou je funkce `IsNil`, jež vrací pravdivostní hodnotu `True`, pokud je její jediný argument `Nil`, v opačném případě je jejím výsledkem `False`. Tato speciální sémantika funkce `IsNil`, ačkoliv porušuje princip kompozicionality a vyžaduje aplikaci unární funkce na *nic*, je zvolena jako doplněk k funkci *Improper*/($o*_n$) definované v Průvodci čtenáře, a jako kompromis mezi dodržáním principů TIL a umožněním zpracování chybových stavů.

3.1.2 Okamžité vyhodnocování

Ačkoliv je TIL-Script funkcionální jazyk, vyhodnocování výrazů probíhá okamžitě (tzv. *Eager evaluation*). Okamžité vyhodnocování je potřeba k zajištění respektování principu kompozicionality. Pokud by vyhodnocování bylo líné, uživatel by mohl jako argument funkce uvést *v*-nevlastní konstrukci, která by nikdy nebyla vyhodnocena, a došlo by tak k aplikaci funkce na chybějící argument. Tato situace by mohla nastat například v případě aplikace funkce, jež některé své argumenty využívá pouze pokud je splněna určitá podmínka.

3.1.3 Neměnnost argumentů funkcí a objektů (*Immutability*)

Jelikož je TIL-Script funkcionální jazyk, jsou objekty a argumenty funkcí neměnné (*immutable*). Zatímco v imperativních jazycích, jako je třeba jazyk C++, není problém v těle funkce modifikovat argument, který funkce obdržela, v TIL-Script samozřejmě nic takového provést nemůžeme. Dále například nemůžeme modifikovat existující seznam. Pokud chceme transformovat již existující seznam, musíme vytvořit nový seznam a požadované hodnoty uložit v novém seznamu. Podobně nemůžeme změnit hodnoty *n*-tice, můžeme však vytvořit novou *n*-tici.

Hodnotu volné proměnné změnit lze, abychom mohli zkoumat konstrukce při různých valuacích *v*. Typ proměnné však změnit nelze.

Obdobně nelze redefinovat funkce nebo změnit typ symbolické hodnoty (viz 3.2.15 – Symbolické hodnoty).

3.1.4 Definice a deklarace symbolů

TIL-Script nově rozlišuje mezi deklaracemi a definicemi proměnných a funkcí. Deklarace pouze uvědomí překladač o existenci proměnné nebo funkce, nijak ale nedefinuje valuaci proměnné nebo sémantiku funkce. Deklarace umožňuje funkci či proměnnou zmínit (např. v trivializaci, v uzávěru),

neumožňuje nám však proměnnou provést nebo funkci aplikovat – jak také, když neznáme hodnotu proměnné, případně sémantiku dané funkce. Provedení deklarované, avšak nedefinované proměnné je chybou, při které interpreter ohlásí chybu a běh programu je ukončen. Deklarovat jeden symbol lze vícekrát, deklarace však nesmí být konfliktní a lišit se typy.

```
$ java -jar interpreter/build/libs/tilscript.jar examples/undef-var.tils
** Error **
(4, 1): myVar.
    ~~~ ^ ~~~
        Variable 'myVar' is declared but undefined
$ java -jar interpreter/build/libs/tilscript.jar examples/undef-fn.tils
** Error **
(2, 1): MyFn/(Int Int Int).
    ~~~ ^ ~~~
        Function MyFn is declared but undefined, application is impossible
```

Ukázka 3.1: Hlášení chyby při chybějící definici

Definice přiřadí proměnné valuaci, funkci sémantiku. Proměnné s řádnou definicí lze provést a můžou tak být konstituentem prováděné konstrukce. Funkce s řádnou definicí lze aplikovat. Funkce i proměnné lze definovat pouze jednou. Opakovaná definice je chybou a vyústí v předčasné ukončení programu.

Symbolické hodnoty, viz 3.2.15 – Symbolické hodnoty, lze pouze deklarovat.

Deklarace jsou automaticky odvozeny z definic. Proto, pokud je známa definice, není třeba dodávat také deklaraci. K interpretaci deklarací automaticky dochází před interpretací definic, aby byla umožněna např. definice vzájemně rekurzivních funkcí. Definice jsou interpretovány v takovém pořadí, v jakém jsou uvedené ve zdrojovém kódu.

Deklarace bez řádných definic na první pohled můžou působit zbytečně. K čemu může sloužit funkce, kterou nelze aplikovat? Nesmíme však zapomenout, že konstrukce TIL samy vyjadřují význam, a nemusí nutně sloužit pouze k provedení. Provedením konstrukce sice dostaneme její denotát, ten nás ale ne vždy zajímá. Představme si tedy případ, kdy provádíme analýzu výrazu přirozeného jazyka. Výraz analyzujeme pomocí Transparentní intenzionální logiky a získáme konstrukci. S danou konstrukcí chceme dále pracovat a chceme ji strojově zpracovat. Její denotát nás ovšem nezajímá, zajímá nás pouze význam konstrukce – procedura. Současně daná konstrukce obsahuje funkci, jejíž definici neznáme, známe, ale nejsme schopni ji strojově vyjádřit, nebo nás pouze nezajímá. Jeli-kož víme, že konstrukci nebudeme provádět, a tedy nebudeme ani aplikovat funkci v ní zmíněnou, nepotřebujeme znát její přesnou definici. Stačí nám znát pouze její název a typ.

Jako příklad nevyjádřitelné funkce lze uvést například všeobecný kvantifikátor. Ačkoliv všeobecný kvantifikátor existuje jako funkce v Transparentní intenzionální logice, nelze jej korektně definovat tak, aby byl vždy strojově vyhodnotitelný.

Názvosloví *deklarace*, *definice* je převzáno z programovacího jazyka C, kde deklarace pouze uvědomí překladač o existenci symbolu, definice poté přiřadí symbolu konkrétní hodnotu. Počet deklarací je shora neomezený, naopak definice může existovat nanejvýš jedna. Deklarace nedefinovaného symbolu není chybou, ovšem snaha nedefinovaný symbol využít (např. volání funkce, přístup k proměnné) vyústí v chybu při procesu linkování.

3.2 TIL-Script jako výpočetní varianta TILu

Tato sekce popisuje základní výrazy a konstrukce TIL-Scriptu, které existovaly již v předchozích verzích jazyka. Pokud práce tyto výrazy nějakým způsobem upravuje, je úprava náležitě popsána a zdůvodněna.

3.2.1 Věty TIL-Scriptu

V TIL-Scriptu za věty (*sentence*) považujeme výrazy na nejvyšší úrovni v programu. Větou je tedy například konstrukce taková, že není podkonstrukcí jiné konstrukce než sebe samotné, ale také definice funkce, proměnné, typu, apod. Každá věta musí být ukončena terminátorem. Roli terminátoru v TIL-Scriptu zastává znak . (tedy ASCII tečka).

3.2.2 Atomické datové typy

Atomické datové typy v TIL-Scriptu vycházejí z výchozí báze využívané v Transparentní intenzionální logice k analýze přirozeného jazyka, tedy z množin o, ι, τ, ω . TIL-Script ovšem rozlišuje mezi časy a reálnými čísly, a pro tyto hodnoty definuje dva nekompatibilní typy, mezi kterými neexistuje implicitní konverze. Dále TIL-Script využívá typ ν představující celá čísla. Nakonec TIL-Script pro názvy typů nevyužívá řecká písmena, která nelze prakticky a jednoduše zapisovat na většině rozložení klávesnic, ale anglická slova nebo zkratky. Názvy typů vždy začínají velkým písmenem.

Typ o představující pravdivostní hodnoty TIL-Script pojmenovává **Bool**. Hodnotami typu **Bool** jsou poté hodnoty **True** a **False**.

Typ ι , v TIL-Scriptu **Indiv**, reprezentuje množinu individuí. Individua v Transparentní intenzionální logice považujeme za *holá* – žádnou netriviální vlastnost nemají nutně. Všechny netriviální vlastnosti individuí jsou určeny stavem světa. Individuum samotné nemá žádnou inherentní valuaci. Slouží pouze jako unikátní identifikátor. Obdobně hodnoty **Indiv** v jazyce TIL-Script nemají žádnou konkrétní reprezentaci. Typ **Indiv** je využíván v konjunkci se symbolickými hodnotami, viz 3.2.15 – Symbolické hodnoty. Tímto TIL-Script umožňuje uživateli referovat na konkrétní individuum pouze pomocí symbolického identifikátoru, aniž by individuum muselo být přiřazeno arbitrárně zvolené konkrétní hodnoty.

Reálná čísla TIL-Script reprezentuje typem **Real**. V implementaci překladače vytvořeném v rámci této práce jsou reálná čísla interně reprezentována typem **double**. TIL-Script samotný žádné

omezení na reprezentaci reálných čísel nestanovuje, prakticky však reálná čísla v současné implementaci reprezentujeme pomocí 64bitové reprezentace dle IEEE 754.

Celá čísla TIL-Script reprezentuje typem `Int`. Obdobně jako u typu `Real` neexistuje omezení pro reprezentaci celých čísel. Interně je využíván datový typ `long`, jedná se tedy o 64bitové znaménkové číslo reprezentované dvojkovým doplňkem.

Množinu možných časů modelujeme typem `Time`. Pro interní reprezentaci časových okamžiků byl v této práci zvolen datový typ `long`. Uživatel se sám může rozhodnout, jak bude tyto hodnoty interpretovat. Ve standardní knihovně lze však nalézt např. funkci `Now`, jejíž aplikací získáme počet milisekund uplynulých od 1. ledna 1970.

Typ konstrukcí, v TIL denotován `*`, byl v TIL-Scriptu přejmenován na `Construction`. V dřívějších verzích TIL-Scriptu se pro typ konstrukcí také využíval znak `*`. Změna na `Construction` byla provedena z důvodu zachování konzistence s ostatními typy, které v TIL-Scriptu taktéž pojmenováváme anglickými slovy, a také z důvodu omezení ambiguity, aby se typ konstrukce nepletl s funkcí násobení. Zatímco typograficky korektní reprezentací násobení je využití znaku \times , na počítači násobení značíme `*`. Ačkoliv lze mezi funkcí násobení a typem konstrukcí odlišit na základě kontextu, v jakém se znak objevuje, odlišení již v názvu umožňuje uživateli jednoduše rozeznat typ konstrukce od násobení čísel okamžitě a bez bližšího zkoumání kontextu.

Dále byl TIL-Script rozšířen o atomický typ sloužící k reprezentaci textu. Tento typ je podrobněji popsán v sekci o rozšířeních TIL-Scriptu, viz 3.3.3 – Typ *Text*.

3.2.3 Generický typ *Any*

V Transparentní intenzionální logice není neobvyklé definovat typově polymorfní funkce. Zvykem je označovat předem neznámé typy řeckým písmenem α . Obdobně TIL-Script umožňuje definici typově polymorfních funkcí. Generické typy v TIL-Scriptu značíme slovem `Any`, okamžitě následovaným indexem polymorfního typu. Index je libovolné číslo z rozsahu $\langle 0; 2^{32} - 1 \rangle$. Nenulové indexy nesmí začínat číslem 0.

Generické typy lze použít pouze v typech funkcí. Má-li více argumentů funkce stejný generický typ, tedy typ `Any` se stejným indexem, interpreter při procesu typové kontroly zajistí, že argumenty, na něž je funkce za běhu aplikována, jsou stejného typu.

3.2.4 Výrazy *TypeDef*

Výrazy `TypeDef` umožňují přiřadit již existujícímu typu alternativní jméno. `TypeDef` nevytváří nový typ, jedná se pouze o alternativní název. V průběhu typové kontroly jsou tato jména rekurzivně přepisována, dokud překladač nezjistí původní typ. Teprve poté porovnává původní typy, nezávisle na jejich alternativních názvech. Typové aliasy můžeme využívat například pro zkrácení zápisu komplikovaných molekulárních typů. Obdobné zkratky v TIL využíváme běžně.

Korektní syntaktický zápis `TypeDef` výrazu začíná právě klíčovým slovem `TypeDef`, následovaným novým názvem, operátorem přiřazení `:=`, a nakonec původním typem. Název musí začínat velkým písmenem.

```
TypeDef Float := Real.  
TypeDef Property := (((Bool Indiv) Time) World).
```

Ukázka 3.2: Výraz `TypeDef`

3.2.5 Funkce

V matematice známe funkce jako jednoznačné zobrazení z množiny možných argumentů (definiční obor) do množiny možných obrazů (obor hodnot). V programovacích jazycích konstrukce nazývané funkcemi často nemusí být nejen jednoznačné (výstup nemusí odpovídat pouze argumentům), ale dokonce nemusí být ani zobrazením (nevrací žádnou hodnotu, v takovém případě většinou modifikují stav světa, jako příklad lze uvést funkce s návratovou hodnotou `void` v jazyce C).

V jazyce TIL-Script, obdobně jako v TIL, jsou funkce vždy zobrazením do určitého oboru hodnot. Díky parcialitě může být funkce degenerovaná, a v takovém případě nebude vracet hodnotu pro žádnou kombinaci argumentů. V takovém případě se však jedná o určitou formu chybového stavu, spíše než záměr, jak by tomu bylo v případě např. jazyka C. V čem se ovšem TIL-Script od Transparentní intenzionální logiky liší, je možnost modifikovat stav světa. TIL-Script je funkcionální jazyk, proto je na nejlepším uvážení uživatele, aby takovéto funkce nedefinoval, a jejich využívání omezil na nutné minimum. Příkladem funkcí s vedlejšími efekty (tedy modifikujícími stav světa) můžou být například funkce `Print`, `Println` ze standardní knihovny, funkce pro zápis do databáze, apod. Aritmetické funkce musí být vždy alespoň jedna.

Pro zápis typu funkce využíváme podobnou notaci jako v Transparentní intenzionální logice. Typy funkcí denotujeme kulatými závorkami. Uvnitř závorek uvedeme nejprve obor hodnot funkce, poté uvádíme postupně typy argumentů. Jediný rozdíl oproti TIL spočívá v nutnosti zapsat mezeru mezi názvy jednotlivých typů. Tedy ekvivalentem k typu $(\sigma\tau)$ by v TIL-Scriptu byl typ `(Bool Int Real)`, případně `(Bool Int Time)`.

Funkce lze deklarovat uvedením názvu funkce následovaným lomítkem a jejím typem. Deklarujeme-li více funkcí stejného typu, můžeme uvést více názvů oddělených čárkami. Pro definici funkce byla přidána nová syntax popsaná v sekci 3.3.2 – Definice pojmenovaných funkcí.

```
Add, Sub, Mult, Div/(Int Int Int).
```

Ukázka 3.3: Deklarace funkcí

3.2.6 Literály

Názvosloví *literál* je přejato z jiných programovacích jazyků. V TIL-Scriptu literály myslíme ne-funkce, tedy členy množin tvořících bázi. Pomocí literálů lze uvádět celá i reálná čísla, pravdivostní hodnoty, a také text (viz oddíl 3.3.3 – Typ *Text* věnující se typu `Text`). Individua zmiňujeme pouze za využití symbolických hodnot, viz Symbolické hodnoty 3.2.15 – Symbolické hodnoty. Literály nikdy nesmíme zapomenout trivializovat.

3.2.7 Trivializace

Trivializace v TIL-Scriptu slouží ke stejnému účelu jako v Transparentní intezionální logice. Narozdíl od TIL ovšem Trivializaci denotujeme apostrofem, namísto nuly zapsané jako levý horní index.

```
'1          -- Trivializace konstanty typu Int
'3.14159    -- Trivializace konstanty typu Real
'['+ '1 '2] -- Trivializace kompozice
```

Ukázka 3.4: Příklad Trivializace.

3.2.8 Proměnné

Proměnné v TIL-Scriptu opět slouží stejném účelu jako proměnné v TIL, tedy *v*-konstruuji hodnotu v závislosti na valuaci *v*.

Volné proměnné lze deklarovat bez přiřazení valuace, ale lze je také definovat a přiřadit jim konkrétní hodnotu. Proměnné deklarujeme, pokud nás nezajímá konkrétní valuace. Přiřazení hodnoty využijeme například v případě, kdy si chceme uložit výsledek drahé operace (např. čtení z databáze), nebo si třeba jen chceme zkrátit zápis dlouhé konstrukce, a chceme pracovat s konkrétní valuací *v*. Vždy je třeba uvést typ hodnoty, kterou proměnná konstruuje.

Syntax pro deklaraci proměnné již existovala. Syntax pro definici proměnné (přiřazení hodnoty) je nově zavedená. Deklarovat můžeme více proměnných najednou, za předpokladu, že se jedná o proměnné stejného typu, stačí názvy jednotlivých proměnných oddělit čárkou. Přiřazujeme-li však proměnné valuaci, můžeme uvést vždy pouze jednu proměnnou.

Pro změnu hodnoty volné proměnné použijeme stejnou syntax jako při úvodním přiřazení hodnoty. Přiřazení hodnoty proměnné však není konstrukce, ale metavýraz, proto lze hodnotu proměnné vždy změnit pouze na nejvyšší úrovni v programu. Typ proměnné změnit nelze.

Deklarace proměnných jsou vždy provedeny hned po spuštění programu, nezávisle na tom, kde v programu jsou uvedeny. Přiřazení hodnot proměnným není prováděno přednostně, aby byla umožněna korektní redefinice proměnné. Hodnota proměnné je tehdy změněna pouze a přesně na místě, kde je tak uvedeno v programu. Deklarace jsou však z definic vždy dedukovány automaticky – proměnná je přednostně deklarována při spuštění programu, i když její deklarace není uvedena explicitně.

Název proměnné musí začínat malým písmenem.

```
x -> Int.           -- Deklarace proměnné v-konstruuující hodnotu typu Int
y, z -> Int.        -- Deklarace více proměnných najednou
pi -> Real := '3.1415. -- Definice proměnné pi aproximující hodnotu pi
['* pi '2].         -- Využití proměnné pi jako konstituent konstrukce
long_varName123 -> Int.
```

Ukázka 3.5: Příklad využití proměnných

3.2.9 Provedení

Provedení se sémantikou opět nijak neliší od svého ekvivalentu v Transparentní intenzionální logice. Syntax ovšem musela být kvůli praktičnosti upravena, a proto bylo upuštěno od pravých horních indexů. Provedení denotujeme ~ 1 . Pro Dvojí Provedení poté využíváme ~ 2 . Dřívejší verze TIL-Scriptu definovaly i trojí až deváté Provedení. Protože však trojí a vícenásobné provedení není v praxi skoro vůbec potřeba (dle Průvodce TIL taková potřeba nenastala), a protože limit devátého Provedení byl poněkud arbitrární (proč ne například desetinásobné Provedení), je tato práce konzervativní a drží se definice Provedení z Průvodce.

Jelikož jsou všechny konstrukce standardně v módu Provedení, není třeba explicitní Provedení využívat příliš často. Většinou tedy budeme používat Dvojí Provedení.

```
 $\sim 1$  x.
 $\sim 2$ ['GetComposition argument1 argument2].
```

Ukázka 3.6: Příklad využití Provedení

3.2.10 Kompozice

Kompozice umožňuje aplikovat funkce na argumenty. Kompozice využívají stejnou syntax jako v TIL. Protože arita funkce musí být alespoň jedna, musí i Kompozice obsahovat alespoň dvě podkonstrukce – funkci samotnou a alespoň jeden její argument. Počet argumentů, na něž funkci aplikujeme, musí odpovídat počtu argumentů funkce. Dále nesmí být žádný argument *v*-nevlastní (s výjimkou funkcí `If` a `IsNil`). V opačném případě k aplikaci funkce vůbec nedojde, neboť nemáme argumenty, na které bychom funkci aplikovali, a Kompozice je tak *v*-nevlastní.

```
['* '2 '6].
```

Ukázka 3.7: Příklad využití Kompozice

3.2.11 Uzávěry

Sémantika Uzávěrů je opět stejná jako v Transparentní intenzionální logice, syntax ovšem byla upravena. Zatímco v TIL často vypouštíme hranaté závorky, v TIL-Scriptu musíme závorky zapsat vždy. Řecké písmeno lambda nahrazujeme v TIL-Scriptu znakem zpětného lomítka.

Zpětné lomítko následuje seznam argumentů funkce konstruované Uzávěrem. V dřívějších verzích TIL-Scriptu bylo možné typ argumentu v některých případech vynechat – existovala-li volná proměnná se stejným názvem jako λ -vázaná proměnná v Uzávěru, a nebyl-li typ λ -vázané proměnné uveden explicitně, byl typ λ -vázané proměnné automaticky dedukován podle typu stejnojmenné volné proměnné. Tato práce však od této automatické dedukce upouští. Jelikož byly do TIL-Scriptu přidány výrazy `Import` umožňující importovat volné proměnné z jiného souboru, uživatel TIL-Scriptu by se mohl dostat do situace, kdy musí procházet několik importovaných souborů, jen aby zjistil typ λ -vázané proměnné. V případě, kdy máme více λ -vázaných proměnných, použijeme čárku pro jejich oddělení.

Za seznamem argumentů může nově následovat explicitní specifikace oboru hodnot konstruované funkce. Specifikaci oboru hodnot denotujeme symbolem \rightarrow , který následuje název existujícího typu. Explicitní specifikace je nepovinná, může však sloužit k zdůraznění úmyslu uživatele, aby čtenář zdrojového kódu na první pohled znal typ funkce. Dále explicitní specifikace typu může pomoci při typové kontrole.

Nakonec je třeba uvést konstrukci, nad kterou abstrakci provádíme.

```
[ \x: Int -> Int ['+ x '2] ].  
[ \x: Int, y: Int -> Int ['+ x y] ].  
[ \x: Int, y: Int ['+ x y] ].  
  
[[ \x: Int, y: Int -> Int ['+ x y] ] '2 '3 ].
```

Ukázka 3.8: Příklad využití Uzávěrů

3.2.12 Zkrácený zápis typu intenzí a extenzionalizace

V TIL často využíváme zkráceného zápisu jak pro typy intenzí, tak pro jejich extenzionalizaci. Pro hodnoty typu α závislé na světamihu zkracujeme zápis $((\alpha\tau)\omega)$ na $(\alpha_{\tau}\omega)$. Obdobně pro extenzionalizaci intenze a využíváme zkráceného zápisu a_{wt} ekvivalentnímu konstrukci $[[aw]t]$, kde a je konstrukce konstruuující funkci (většinou se jedná o Trivializaci), a $w \rightarrow_v \omega$, $t \rightarrow_v \tau$. Jedná se však pouze o notační zkratku – o dohodu, ne součást TIL.

Ekvivalentní zkratky můžeme využívat také v TIL-Scriptu. Chceme-li specifikovat typ intenze, můžeme využít notační zkratku `@tw`. Zkratku `@wt` využijeme k extenzionalizaci intenze. Zkratka `@wt` vždy extenzionalizuje za využití proměnných `w` a `t`. Proměnná `w -> World` je součástí standardní

knihovny, proměnnou `t` -> `Time` musí uživatel definovat sám. Zkrácenou notaci nelze využít s proměnnými jiného názvu.

```
(Bool Indiv)@tw -- intenze typu (((Bool Indiv) Time) World)
['Rektor@wt 'VSB] -- extenzionalizace funkce Rektor/(((Indiv Indiv) Time) World)
-- a nasledna aplikace na argument VSB.
```

Ukázka 3.9: Příklad využití zkrácené notace

3.2.13 Seznamy

Seznamy se v Transparentní intenzionální logice příliš neobjevují. Při strojové analýze a zpracování je však vhodné mít k dispozici způsob vyjádření kolekce dat, proto TIL-Script seznamy obsahuje. Seznam představuje homogenní seřazenou kolekci potenciálně neomezené délky. Seznamy mohou obsahovat duplicity.

V této práci jsou všechny seznamy neměnné. Seznamy jsou definovány induktivně. Seznam (list) je buď prázdný list, nebo *cons cell* skládající se z hlavičky (známé jako *head* nebo *CAR* listu z jiných jazyků) – prvního prvku v seznamu, a z podseznamu reprezentujícího zbytek listu (*tail*, *CDR*). Z definice tedy vyplývá, že vytvoření nového seznamu vložním prvku na začátek lze provést v konstantním čase. Vytvoření nového seznamu přidáním prvku na konec jiného listu naopak bude lineární vzhledem k velikosti původního kolekce. Tato implementace listu je volena zejména proto, že umožňuje snadnou iteraci pomocí rekursivních funkcí. Stejnou implementaci listů využívají například jazyky Lisp, Haskell, a další.

Typ seznamu denotujeme slovem `List` následovaným kulatými závorkami, v nichž je uveden typ prvku ukládaného v seznamu.

Standardní knihovna obsahuje řadu funkcí pro práci s listy. Tyto funkce jsou v této kapitole pouze nastíněny při ilustraci využití seznamů, podrobněji jsou však zdokumentovány v kapitole popisující standardní knihovnu.

Dále interpreter implementuje syntaktický cukr pro jednodušší vytváření listů. Pro vytvoření seznamu stačí v kompozici aplikovat funkci `ListOf` na nenulový, avšak shora neomezený počet argumentů. Během parsování bude tato kompozice korektně přepsána na sérii kompozic využívajících funkci `Cons` k vytvoření *cons cells*. K žádné aplikaci funkce na libovolný počet argumentů tedy nedochází. Při vytváření seznamu pomocí `ListOf` musíme dodat alespoň jeden prvek listu, aby bylo možné správně dedukovat typ prvků ukládaných v seznamu. `ListOf` je pouze syntaktický cukr překládaný při procesu parsování. Pokud bychom chtěli generovat konstrukce konstruuující listy dynamicky za běhu programu, museli bychom zřetězit aplikace `Cons`.

```
['ListOf '1 '2 '3]. -- vytvoreni listu
['Cons '1 ['Cons '2 ['ListOfOne '3]]]. -- ekvivalent predchoziho radku a priklad
prepisu ListOf
```

```
['Head ['ListOf '1 '2 '3]]      -- 1
['Tail ['ListOf '1 '2 '3]]      -- ['ListOf '2 '3]
```

Ukázka 3.10: Příklad využití seznamů

3.2.14 N-tice

N-tice doznaly oproti předchozím verzím TIL-Scriptu změny. Dříve byly n-tice homogenní kolekce nespecifikované, avšak konečné délky. Roli homogenní kolekce však již zastává List, a v TIL-Scriptu neexistoval způsob, jak seskupit více hodnot jiného typu dohromady.

V matematice využíváme n-tice například k reprezentaci prvků kartézského součinu. N-tice lze definovat více způsoby, vždy jsou však uspořádané, konečné, a mohou být heterogenní (například pokud provádíme kartézský součin nad neidentickými množinami). Obdobnou roli plní n-tice i v TIL-Scriptu. Délka n-tice je pevně daná, stejně jako typy hodnot v n-tici. Celý typ n-tice je tedy dán uspořádaným výčtem všech typů všech hodnot obsažených v této n-tici. Syntaktický zápis typu n-tice je podobný typu seznamu. Typ n-tice denotujeme slovem **Tuple**, následovaným kulatými závorkami. V závorkách ovšem uvádíme výčet typů. Jednotlivé typy oddělujeme čárkami. Délka n-tice je určena počtem uvedených typů.

Práce s n-ticemi je do jisté míry podobná práci s listy. N-tici vytvoříme aplikací variadické funkce **MkTuple**. Stejně jako u funkce **ListOf** se však jedná pouze o syntaktický cukr. Funkce **MkTuple** skutečně existuje, je to ale funkce binární, nikoliv variadická, a vytváří dvojice z dodaných argumentů. Pokud ovšem funkci **MkTuple** dodáme více než dva argumenty, parser aplikaci **MkTuple** rozepíše na jednu aplikaci **MkTuple** a následné řetězení funkce **PrependToTuple**. Funkce pro práci s n-ticemi jsou podrobněji popsány v uživatelské dokumentaci.

```
['MkTuple '1 '2.0 'True].      -- vytvoreni n-tice
['PrependToTuple '1 ['MkTuple '2.0 'True]]. -- ekvivalent predchoziho radku

x -> Tuple(Int, Real, Int) := ['MkTuple '1 '3.14159 '10].
```

Ukázka 3.11: Příklad využití n-tic

3.2.15 Symbolické hodnoty

Název *symbolické hodnoty* je v kontextu TIL-Scriptu nový, jedná se však pouze o praktický způsob implementace entit, které již v TIL-Scriptu existovaly. Entitu lze specifikovat uvedením jejího názvu následovaného lomítkem a typem entity. Tímto byl názvu přiřazen typ, ale ne konkrétní hodnota. Entity v předchozích verzích TIL-Scriptu ovšem mohly reprezentovat jak funkce, tak konstanty. V nové verzi TIL-Scriptu jsou entity děleny na *symbolické hodnoty* a funkce na úrovni implementace.

Funkce tedy není *symbolickou hodnotou*. Název *symbolická hodnota* je opět inspirován jazykem Lisp, kde podobný koncept symbolů, tedy jmen bez hodnoty, již dlouhou dobu existuje.

Symbolické hodnoty využíváme, když potřebujeme o daném objektu referovat jménem, ale ne hodnotou. Jako příklad lze uvést například individua. Individua nemají žádnou inherentní hodnotu. Objekt typu ι je pouze identifikátorem, unikátním jménem. Vlastnosti jsou individuům přiřazeny intenzemi a závisí na stavu světa v konkrétním časovém okamžiku, případně se jedná o triviální vlastnosti bez vypovídací hodnoty. Dále můžeme zmínit například číslo π . Číslo π sice bezpochyby hodnotu má, jedná se o reálné číslo. Jelikož je však π číslo iracionální, tedy číslo s nekonečným desetinným rozvojem, bohužel jej nemůžeme s absolutní přesností reprezentovat v počítači. V praktické implementaci se tak musíme spokojit pouze s aproximací čísla π , nebo právě se symbolickou reprezentací.

Symbolické hodnoty nelze provést. Chceme-li je zmínit, musíme využít Trivializaci. Dále nad nimi nelze provádět všechny operace, které jdou provádět s nesymbolickými konkrétními hodnotami. Například jak bychom mohli přičíst k jinému číslu číslo π , když neznáme přesnou hodnotu čísla π ? Kde to však dává smysl, můžeme naprogramovat dodatečnou podporu pro symbolické hodnoty. Funkce `Sin`, `Cos` v matematické knihovně mají zabudovanou kontrolu, zda obdržely jako svůj argument symbol `Pi/Real`. Pokud ano, funkce vrátí korektní hodnotu. Obdobně by například výsledkem aplikace přirozeného logaritmu na symbol `E/Real` mohlo být číslo 1.

`Pi/Real`.

```
[ 'Sin' 'Pi' ]. -- 0
```

```
[ 'Cos' 'Pi' ]. -- -1
```

```
[ '+ '1 'Pi' ]. -- Nil - vysledek existuje, na pocitaci jej vsak nelze spocitat
```

Ukázka 3.12: Příklad využití symbolických hodnot

3.3 Rozšíření TIL-Scriptu

Tento oddíl nastiňuje nadstandardní rozšíření TIL-Scriptu. Cílem těchto rozšíření je udělat z TIL-Scriptu nástroj pro analýzu a práci s konstrukcemi Transparentní intenzionální logiky. Každá změna a rozšíření je zdůvodněna.

3.3.1 Komentáře

Komentáře v TIL-Scriptu plní stejnou roli jako v jiných programovacích jazycích. Jedná se tedy o text, který je interpreterem ignorován. Komentáře tak můžeme využít například k popisu kódu. Komentáře existují pouze v jednořádkové podobě a denotujeme je, podobně jako v jazycích Haskell nebo SQL, dvěma pomlčkami (`--`). Všechny znaky za pomlčkami až po první znak odřádkování jsou překladačem automaticky ignorovány.

```
-- Následující konstrukce spočítá a vypíše kosinus pi
['Println ['Cos 'Pi]]. -- Prints -1
```

Ukázka 3.13: Příklad využití komentářů

3.3.2 Definice pojmenovaných funkcí

Syntax TIL-Scriptu byla rozšířena o možnost definice pojmenovaných funkcí. Jedná se o způsob, jak funkci přiřadit jméno a následně na danou funkci odkazovat jménem, namísto uvádění stejného uzávěru vícekrát ve zdrojovém kódu. Definice pojmenovaných funkcí byla v TIL-Scriptu již dříve přidána ale následně byla zase odstraněna. Dokud TIL-Script sloužil pouze jako notace pro zápis TIL konstrukcí, byl přínos této funkcionality diskutabilní. Pokud však chceme experimentovat s možnostmi využití TIL-Scriptu také jako nástroje pro analýzu TIL konstrukcí, či importovat funkce z jiných souborů, definice pojmenovaných funkcí je nezbytná.

Definici funkce denotujeme klíčovým slovem **Defn**. Následuje hlavička funkce, symbol přiřazení ($:=$), a nakonec konstrukce předepisující sémantiku funkce. Syntax je odlišná od notace pro zápis uzávěrů. Důvodem pro tento rozdíl je grafické odlišení hlavičky funkce specifikující název, argumenty a typ výsledku od těla funkce (konstrukce). Dále tato notace omezuje počet vnořených závorek a speciálních znaků v těle funkce. Vazba argumentů funkce je samozřejmě ekvivalentní λ -vázání

Název funkce musí začínat velkým písmenem. Dále může obsahovat malá i velká písmena, čísla a podtržítka. Povolena jsou všechna písmena české abecedy (tedy například české *č* je povoleno, avšak katalánské *ç* povoleno není).

```
-- Definice binární funkce
Defn Add(x: Int, y: Int) -> Int := ['+ x y].

-- Příklad definice rekurzivní funkce
Defn Sum(list: List(Int)) -> Int :=
  ['If ['IsEmpty list]
    '0
    ['+ ['Head list] ['Sum ['Tail list]]]].
```

Ukázka 3.14: Příklad definice funkcí

3.3.3 Typ *Text*

Typ **Text** slouží k reprezentaci textu. Opět se jedná o typ, který není příliš potřeba pro analýzu přirozeného jazyka, neboť v přirozené řeči často vypovídáme o individuích, málokdy však o samotných korpusech textu. V programovacím jazyku je však forma reprezentace textu nezbytná. Typ **Text**

využíváme jak k reprezentaci delších textů, tak k reprezentaci jednotlivých znaků. TIL-Script tedy nezná koncept typu pro samostatné znaky (jako např. `Char` v jazyce Haskell), a `Text` není `List` znaků. Jedná se o atomický typ. Pomocí funkcí standardní knihovny však lze z textů extrahovat podřetězce (včetně jednotlivých znaků), konkatenovat řetězce, apod.

Literály typu `Text` uvozujeme dvojími uvozovkami. Hodnoty typu `Text` musíme trivializovat. V literálech textů můžeme používat tzv. *escape sekvence*. Texty podporují standard Unicode. Interně jsou reprezentovány objekty typu `java.lang.String` a tudíž využívají kódování UTF-16 a jsou internovány.

```
text -> Text := "Můj text".
text2 -> Text := "Text na\ndva řádky".
['Println "Výpis textu"].
```

Ukázka 3.15: Příklad využití typu `Text`

3.3.4 Výrazy `Import`

Výrazy `Import` umožňují importovat symboly definované v jiném souboru nebo Java archivu (`.jar` souboru). Výrazy `Import` jsou vždy interpretovány jako první při interpretaci souboru. Každý soubor můžeme importovat nanejvýš jednou v rámci jednoho souboru. Obsahuje-li soubor se zdrojovým kódem TIL-Scriptu více importů stejné závislosti, je daná závislost importována pouze jednou, následující importy jsou ignorovány.

Každý soubor je interpretován pouze jednou. Pokud tedy více souborů importuje jeden soubor, daná sdílená závislost bude interpretována pouze při první importu. Při první interpretaci si překladač uloží všechny symboly definované v daném souboru (symbolické hodnoty, funkce, proměnné). Při následných importech stejného souboru jsou poté pouze importovány tyto symboly. Tedy hodnoty proměnných jsou spočítány pouze jednou, poté dané zůstává stále stejná proměnná. Obsahuje-li importovaný skript konstrukce na nejvyšší úrovni v programu (tedy takové, že nejsou podkonstrukcí jiné konstrukce než sebe samotné), jsou dané konstrukce provedeny pouze při prvním importu programu. Při importu skriptu je nejprve interpretován celý skript, teprve poté jsou symboly definované v daném souboru importovány do kontextu skriptu, který na daném souboru závisí.

Pro import souboru uvedeme nejprve klíčové slovo `Import`, následované cestou k požadovanému souboru. Cesta může být absolutní nebo relativní k umístění aktuálně překládaného souboru. Cesta musí být uvozena uvozovkami. Importy se mohou nacházet pouze na nejvyšší úrovni v programu, neboť se jedná o metavýrazy, ne konstrukce.

Symboly můžeme importovat také z Java archivu. V takovém případě musí být daný Java archiv načten již při spouštění Java prostředí společně s kódem překladače. Načítání `.jar` souborů za běhu je z bezpečnostních důvodů velmi komplikované. Dále musí daný Java archiv obsahovat tzv.

registrátor symbolů konformující vůči předdefinovanému rozhraní. Pro import souborů za využití registrátoru stačí místo cesty k TIL-Script souboru uvést plně kvalifikovaný název třídy registrátoru s předponou `class://`. Právě tato předpona značí, že se jedná o import z *.jar* souboru. Překladač poté vytvoří instanci určeného registrátoru a naimportuje soubory definované tímto registrátorem. Rozhraní registrátorů je popsáno v uživatelské dokumentaci.

Importy nejsou tranzitivní. Tedy importujeme-li soubor, který sám importuje jiné soubory, budou importovány pouze symboly definované v daném souboru, ne však symboly definované v jeho závislostech. Importy nesmí obsahovat vzájemně konfliktní definice.

```
Import "dependency.tils".
Import "relative/path.tils".
Import "/usr/local/lib/tilscript/path.tils".
Import "class://org.fpeterek.tilscript.math.Registrar".
```

Ukázka 3.16: Příklad využití výrazů Import

3.3.5 Typ *Type*

Typ *Type* slouží k reprezentaci typů a je primárně metaprogramovacím prvkem. V tradiční Transparentní intenzionální logice o typech vypovídat nemůžeme. Typově polymorfní TIL není momentálně předmětem výzkumu. Pro účely metaprogramování je však možnost reprezentace typů nezbytná. Typ *Type* umožňuje pracovat s typy jako s validními hodnotami jazyka TIL-Script. Typové reference, tedy hodnoty typu *Type*, využijeme například k implementaci statické typové kontroly (překladač TIL-Scriptu provádí typovou kontrolu pouze za běhu), ke generování validních TIL-Script programů, apod.

Pokud by TIL-Script typ *Type* neobsahoval, musel by být za účelem analýzy TIL-Scriptu implementován parser TIL-Scriptu v TIL-Scriptu, nebo by uživatelé museli využívat jiné programovací jazyky. Proto tato práce navrhuje rozšíření TIL-Scriptu o metaprogramovací prvky za účelem unifikace a zjednodušení nástrojů pro práci s Transparentní intenzionální logikou.

S typovými referencemi pracujeme obdobně jako s objekty z bazových množin. Zodpovědností uživatele však je, aby typové reference využíval opatrně a vyhýbal se jim, bude-li TIL-Script používat pouze pro interpretaci TIL konstrukcí.

```
intType -> Type := 'Int'.
typeOf5 -> Type := ['TypeOf '5].
```

Ukázka 3.17: Příklad využití typových referencí

3.3.6 Typ *DeviceState*

Typ *DeviceState* využíváme jako obor hodnot funkcí závislých na stavu zařízení, na němž běží překladač. Standardní knihovna definuje proměnnou *deviceState* -> *DeviceState*, kterou uživatel může využívat kdekoliv, kde je hodnota typu *DeviceState* vyžadována. Hodnoty typu *DeviceState* nemají žádnou inherentní hodnotu, obdobně jako objekty typu *World*. Jako příklad funkcí závislých na stavu zařízení lze uvést například funkci *Now* vracující aktuální čas, nebo funkci *Random*, jejíž výsledkem je náhodné číslo (v aktuální implementaci číslo pseudonáhodné, nic však nebrání využití např. hardwarových generátorů náhodných čísel nebo */dev/random*, pokud je to na dané platformě možné).

```
[ 'Now deviceState ].
[ 'Random deviceState ].
RandFromDevRandom/(Int DeviceState ).
```

Ukázka 3.18: Příklad funkcí závislých na stavu zařízení

3.3.7 Hodnota *Nil*

Hodnota *Nil* slouží k reprezentaci stavu, kdy konstrukce nekonstruuje žádnou hodnotu. Je-li jedním z argumentů funkce *Nil*, je výsledek automaticky opět *Nil*. Pro vyvolání chybového stavu je možné *Nil* zmínit přímo v TIL-Script konstrukci. Trivializace *Nil* však není povolena gramatikou, neboť Trivializace nemůžou nikdy být nevlastní.

Specifikem implementace je, že hodnota *Nil* obsahuje metadata popisující kde a proč došlo k selhání a konstrukce nezkonstruovala žádnou hodnotu. Pokud konstrukce na nejvyšší úrovni v programu nezkonstruuje hodnotu a výsledkem jejího provedení je *Nil*, program skončí neúspěchem a uživateli je nahlášena chyba.

```
[ 'Cond
  [ '= a '0 ] Nil
  [ '= b '0 ] Nil
  'True    [ '/ [ 'Log a ] b ] ].
```

Ukázka 3.19: Příklad využití Nil

```
(1, 17): [ '* [ '+ '1 '2 ] [ '/ '6 '0 ] ].
          ~~~ ^ ~~~
          Division by zero
** Error **
(1, 17): [ '* [ '+ '1 '2 ] [ '/ '6 '0 ] ].
          ~~~ ^ ~~~
```

Ukázka 3.20: Příklad hlášení chyby

3.3.8 Struktury

Struktury, obdobně jako n-tice, umožňují shlukovat více hodnot různorodých typů. Narozdíl od n-tic, jejichž členy indexujeme číselně, atributy struktur jsou pojmenované a přistupujeme k nim jménem. Dále můžou být struktury definovány rekurzivně. Tato charakteristika umožňuje například reprezentovat grafy. Atributu struktury nemůže být přiřazena hodnota `Nil`.

Během typové kontroly jsou porovnávány nejen typy atributů struktur, ale také názvy porovnávaných struktur. Liší-li se struktury názvem, ačkoliv jsou ve všech ostatních ohledech stejné, jsou považovány za rozdílné typy. Struktury nám tedy, narozdíl od n-tic, umožňují lépe využívat typového systému TIL.

Pro představu uveďme následující příklad. Potřebujeme v TIL-Scriptu reprezentovat souřadnice. Dále pracujeme s dvěma typy souřadnic – s kartézskými a s polárními souřadnicemi. Využijeme-li typ `Tuple(Real Real)` k reprezentaci souřadnic, může jednoduše dojít k záměně kartézských a polárních souřadnic. Pokud ale definujeme za účelem struktury, viz ukázka 3.21, nechtěnné záměně souřadnic dokáže překladač zabránit během typové kontroly.

```
Struct Cart {  
    x: Real,  
    y: Real,  
}.  
  
Struct Polar {  
    distance: Real,  
    angle: Real,  
}.
```

Ukázka 3.21: Definice struktur reprezentujících souřadnice

3.3.8.1 Definice struktur

Strukturu definujeme klíčovým slovem `Struct`. Dále uvedeme jméno struktury, a do složených závorek seznam atributů. Nejprve vždy uvedeme název atributu, za dvojtečku pak jeho typ. Jednotlivé atributy oddělujeme čárkami. Název struktury musí být validním názvem typu – musí začínat velkým písmenem. Při pojmenovávání atributů se pak řídíme pravidly pro jména proměnných a začínáme vždy malým písmenem. Za posledním atributem může ale nemusí být čárka – tzv. *trailing comma*.

```
Struct Tree {  
    value: Int,  
    subtrees: List(Tree),  
}.
```

Ukázka 3.22: Příklad definice struktury

3.3.8.2 Konstruktory struktur

Konstruktory struktur umožňují vytvářet objekty uživatelem definovaného typu – struktury. Pro konstruktory byla vytvořena nová syntax. Do složených závorek uvedeme nejprve název struktury, následně pak uvádíme konstrukce, které budou provedeny jako konstituenty konstruktoru. Výsledky provedení poté budou přiřazeny jednotlivým atributům struktury. Argumenty konstruktoru musíme uvést v pořadí odpovídající definici struktury. Konstruktory nevyužívají typové reference, název konstruovaného typu tedy netrivializujeme, jedná se pouze o syntaktický prvek. Jinak jsou konstruktory ekvivalentní kompozicím – pokud konstruktor neobdrží alespoň jeden z požadovaných argumentů, je konstruktor *v*-nevlastní, jinak konstruktor konstruuje hodnotu specifikovaného typu.

Konstruktory můžeme využít také ke konstrukci prázdných listů. Konstrukce listů pomocí `ListOf` využívá typovou inferenci pro určení typu seznamu. Prázdný list však pomocí kompozice nelze vytvořit. Dále využití konstruktorů pro konstrukci prázdných seznamů umožňuje vyhnout se využívání typových referencí. Neprázdný list pomocí konstruktoru vytvořit nelze.

```
{List(Tree)}. -- Konstrukce prazdneho listu  
{Tree '0 {List(Tree) }}. -- Konstrukce objektu typu Tree  
  
myTree -> Tree := {Tree  
    '1  
    ['ListOf  
        {Tree '2 {List(Tree) }}  
        {Tree '3 ['ListOf {Tree '4 {List(Tree) } }]]}}.
```

Ukázka 3.23: Příklad definice struktury

3.3.8.3 Přístup k atributům

Za účelem přístupu k atributům byl přidán syntaktický konstrukt `::`. Za název objektu, k jehož atributu přistupujeme, zapíšeme `::` následované názvem požadovaného atributu. Přístupy k atributům lze řetězit, potřebujeme-li přistoupit k atributu atributu. Přístup k atributu má vyšší prioritu než trivializace – tedy trivializací trivializujeme celý přístup k atributu, nepřistupujeme k atributu trivializace.

Reálně je však atribut pouze proměnná, ačkoliv se jedná o proměnnou ve strukturovaném shluku. Proto i sémantika provedení atributu je stejná jako sémantika provedení proměnné. Přístup k neexistujícímu atributu je chybou. Provedení existujícího atributu (proměnné) vždy zkonstruuje hodnotu, neboť atributy nemůžou být *v*-nevlastní.

```
tree::subtrees.
```

```
['Println struct::nested::attributes'].
```

Ukázka 3.24: Příklad přístupu k atributům objektu

Kapitola 4

Implementace

V této kapitole nastíníme implementační detaily interpreteru. Nejprve zmíníme využití technologie, poté popíšeme architekturu projektu, nakonec pak uvedeme zajímavější problémy, které se objevily při implementaci překladače, a jejich řešení.

4.1 Zvolené technologie

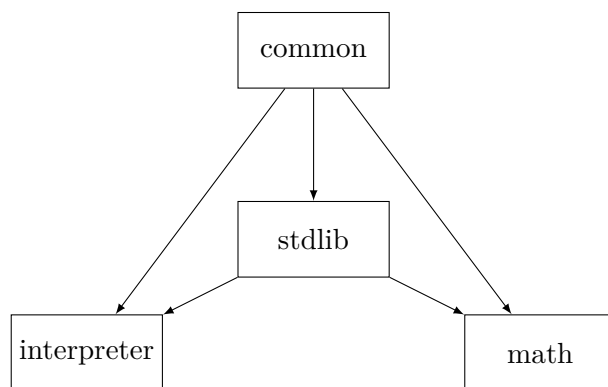
Celý projekt je implementován v jazyce **Kotlin**¹. Jazyk Kotlin je staticky typovaný, multiparadigmatický jazyk kompilovaný do **JVM** bytekódu. Kotlin je vytvářen jako alternativa k jazyku Java, a nabízí plnou kompatibilitu s Javou. Využití Kotlinu umožňuje využívat veškeré výhody Java ekosystému, včetně knihoven psaných v Javě, ale také psát expresivnější kód, než by bylo možné v Javě. Null-safety a jazyková podpora pro algebraické datové typy pak umožňují psát bezpečnější kód, než je možné v jazyce Java. Pro spuštění překladače je samozřejmě nutné mít na počítači nainstalované JRE.

Jako build systém byl zvolen projekt **Gradle**. Důvodem této volby je relativní jednoduchost konfigurace i využití systému Gradle, ale také přístup k Maven repozitáři s Java knihovnami. Dále využíváme několik Gradle pluginů nutných k sestavení projektu.

Parser generujeme za pomoci technologie **Antlr** ve verzi 4. Antlr je open-source generátor parserů podporující tvorbu parserů v řadě jazycích. V našem případě využíváme jako cílový jazyk Java.

Interpreter využívá knihovnu `org.apache.commons:commons-text` pro zpracování escape sekvencí. Tím výčet využitých technologií končí.

¹Přestože je projekt psaný v Kotlinu, v práci často zmiňujeme např. *Java knihovny*, *Java objekty*, apod. Většinou tím myslíme Javu jako platformu. V takových případech poté nezáleží, zda používáme jazyk Java, Kotlin, nebo libovolný jiný jazyk kompilovaný pro platformu JVM.



Obrázek 4.1: Komponenty projektu

4.2 Architektura projektu

Celý projekt je rozdělen do čtyř komponent – společné knihovny (`common`), standardní knihovny (`stdlib`), překladače (`interpreter`) a matematické knihovny (`math`). Schéma projektu je znázorněno v obrázku 4.1, který vyjadřuje, jak na sobě jednotlivé komponenty projektu závisí.

4.2.1 Společná knihovna `common`

Knihovna `common` obsahuje kód společný pro zbytek projektu. Jedná se například o implementace tříd reprezentujících TIL konstrukce, definice společných rozhraní, reprezentaci typů, nebo drobné utility. Knihovna neobsahuje definice TIL-Script objektů, slouží ke sdílení kódu napříč jednotlivými komponentami. Využít ji tak může například programátor implementující novou TIL-Script knihovnu, konečného uživatele se však existence `common` nijak netýká.

Knihovna nemá žádné externí závislosti.

4.2.2 Standardní knihovna `stdlib`

Knihovna `stdlib` obsahuje implementaci standardní knihovny. `stdlib` nekonformuje vůči rozhraní, kterému musí konformovat TIL-Script knihovny implementované jako Java knihovny a distribuované jako Java archivy. Standardní knihovnu překladač automaticky importuje v každém souboru. Není tedy třeba ji importovat explicitně.

Standardní knihovna je nezávislá na použitém překladači TIL-Scriptu. Interpreter, jenž je součástí projektu, můžeme klidně nahradit novým interpreter (za předpokladu, že daný interpreter implementuje potřebná rozhraní, např. `InterpreterInterface` definované v knihovně `common`). Interpreter samotný však na standardní knihovně závisí. Kvůli syntaktickému cukru (funkce `Cond`, `ListOf`, atd.), funkci `If`, jež musí být vyhodnocována líně, apod., musí překladač obsahovat speciální podporu pro standardní knihovnu.

Standardní knihovna definuje základní množinu funkcí, hodnot, typů a proměnných potřebnou pro práci s TIL-Scriptem.

4.2.3 Matematická knihovna `math`

Matematická knihovna `math` slouží jako ukázková implementace TIL-Script knihovny v Kotlinu, případně v Javě. Dále je využívána k testování funkčnosti importování Java knihoven. Narozdíl od standardní knihovny, překladač je naprosto nezávislý na knihovně `math`. `math` je třeba importovat explicitně pomocí výrazu `Import`.

Knihovnu nelze označit za extenzivní, obsahuje pouze malé množství funkcí, definice symbolických hodnot `E`, `Pi` a proměnných `e`, `pi` aproximujících eulerovo číslo a číslo π .

4.2.4 Interpreter

Narozdíl od předchozích komponent, které byly čistě Java knihovnami, Interpreter je spustitelný Java program (tedy program s korektně definovanou funkcí `main`). Jedná se o referenční implementaci překladače jazyka TIL-Script. Překladač podporuje TIL-Script v takové podobě, v jaké je definován touto prací. Obsahuje základní nástroje pro hlášení chyb, aby ulehčil práci s TIL-Scriptem. Typovou kontrolu provádí pouze za běhu programu.

4.3 Implementace překladače

V této sekci nejprve nastíníme fungování překladače. Poté popíšeme zajímavé problémy týkající se implementace překladače. Problémy mohou být zajímavé jak z hlediska řešeného problému, tak z hlediska budoucího rozvoje.

4.3.1 Vysokoúrovňový pohled na interpreter

Interpreter funguje čistě jako neinteraktivní textová aplikace. Programy, které potřebujeme interpretovat, předáváme překladači jako CLI argumenty. Pro překladač momentálně neexistuje žádné grafické rozhraní, ani REPL, který by umožnil interaktivní překlad TIL-Script výrazů. Pokud využíváme funkci `main` definovanou v JAR překladače, spustí se programy sekvenčně v takovém pořadí, v jakém je uživatel specifikoval. Pokud ale běh jednoho z programu skončí chybou, další programy se již nespouští. Pro každý program je však vytvořena nová instance překladače. Běhy jednotlivých programů jsou tedy na sobě nezávislé.

Pro strojovou práci se zdrojovým kódem musíme nejdříve daný kód převést ze sekvenční textové podoby do reprezentace, se kterou počítač umí pracovat lépe. Proto programy převádíme do stromové reprezentace nazývané abstraktní syntaktický strom (také AST – Abstract Syntax Tree). Tomuto převodu se říká *parsování*, případně anglicky *parsing*. Parsing (včetně tokenizace a lexikální analýzy) je v našem interpreteru delegován parseru automaticky vygenerovaném technologií Antlr.

Pokud lexer nebo parser narazí na syntaktické chyby v programu, jsou dané chyby ohlášeny uživateli a překlad automaticky končí neúspěchem. Výsledkem vygenerovaného parseru je abstraktní syntaktický strom.

Výsledné AST je ovšem pro naše potřeby příliš generické. Antlr je nástroj pro všeobecné použití, proto i pomocí Antleru vytvořené AST bývají všeobecné. Antlr však kromě parseru umí vygenerovat také abstraktní třídu `Visitor` sloužící k průchodu AST pomocí návrhového vzoru *visitor*. Pomocí visitoru převedeme výsledek parsování na dočasnou reprezentaci – mezivýsledek. Tento výsledek je, díky omezením automaticky vygenerované třídy, stále nedostatečný. V průběhu tohoto průchodu interpreter nehledá chyby v programu – k tomu chybí překladači v současném momentě dostatečný kontext.

Proto proces parsingu následuje ještě jeden průchod stromem a jeho konečný převod na přívětivou strukturu. Během tohoto průchodu překladač převede reprezentaci zdrojového kódu na objekty tříd definovaných v knihovně `common` – tedy na standardní objekty očekávané napříč projektem (překladačem, funkcemi, apod.). Dále v tomto průchodu překladač provede expanzi funkcí `Cond`, `ListOf`, `TupleOf`, `Progn` z variadických funkcí, jež v TILu neexistují, na sekvenci binárních funkcí. Během tohoto průchodu může interpreter opět zachytávat chyby – většinou se jedná o chybné využití syntaktického cukru

Následně již dochází k interpretaci programu. Jednotlivé výrazy ovšem nejsou interpretovány přesně v takovém pořadí, v jakém jsou v programu uvedeny. Při překladu jsou nejprve provedeny výrazy `Import`. Pokud překladač narazí na výraz `Import`, je interpretace aktuálního souboru pozastavena a překladač přeloží importovaný soubor. Po přeložení importované závislosti jsou do kontextu současného souboru naimportovány symboly definované závislostí ².

Dále dochází k přednostní deklaraci struktur, kterou následuje interpretace výrazů `TypeDef`. Teprve po interpretaci `TypeDef` dojde ke korektní definici struktur a jejich atributů. Využitím tohoto postupu umožníme, aby výrazy `TypeDef` mohly zmiňovat struktury, a aby v definicích struktur bylo možné zmiňovat typové aliasy.

Poté je potřeba provést deklarace proměnných a symbolických hodnot (viz 3.2.15), a současně deklarace a definice funkcí. V tomto kroku jsou deklarace proměnných automaticky vyvozeny z definic. Pokud program obsahuje definici proměnné, není nutné dodávat deklaraci. Dodání nekonfliktní deklarace ovšem není chybou.

Nakonec jsou interpretovány top-level konstrukce a přiřazení proměnným. Pokud je konstrukce na nejvyšší úrovni *v*-nevlastní, program končí chybou. Pokud však uživatel počítá s tím, že konstrukce může být *v*-nevlastní, může její výsledek přiřadit proměnné, nebo jej zpracovat pomocí funkce `IsNil`.

²Symboly zde myslíme názvy – jména funkcí, proměnných, i symbolických hodnot. V daném kontextu slovo symbol nesouvisí se symbolickými hodnotami, využíváme jej stejně jako jej například využívají linkery kompilovaných programovacích jazyků.

Vždy, když je vyhodnocována určitá skupina výrazů, ať už při přednostním vyhodnocování, nebo při vyhodnocování konstrukcí, pořadí výrazů v konkrétní skupině vždy odpovídá jejich relativnímu pořadí ve zdrojovém programu.

4.3.2 Detaily implementace

Tato podsekce popisuje detaily implementace, které mohou být důležité nebo zajímavé. Cílem je vyzdvihnout řešení, které jsou důležité z hlediska budoucího rozvoje, řešení zajímavých problémů, ale také řešení, které jsou suboptimální a zasloužily by si další vývoj.

4.3.2.1 Interpretace konstrukcí a operace nad AST

Velkým designovým rozhodnutím, a zároveň nedostatkem, je rozhodnutí operovat přímo nad abstraktním syntaktickým stromem a interpretovat přímo TIL-Script konstrukce reprezentované Java objekty. Kód není překládán do bytekódu, který by byl následně interpretován.

Tento přístup k interpretaci má velkou výhodu – nízkou náročnost na implementaci. Překlad do bytekódu by s sebou nesl značnou přidanou komplexitu. V první řadě by bylo potřeba implementovat vlastní mechanismus zásobníku. V současné době TIL-Script sdílí zásobník s interpreterem. Aplikace funkcí je řešena rekurzivně. Při aplikaci TIL-Script funkce volá interpreter interně vlastní metodu, která aplikaci provede. S voláním metody na JVM je samozřejmě vytvořen nový rámec na zásobníku. Na tomto rámci se poté uloží kontext aplikované TIL-Script funkce (argumenty funkce, apod.). Po dokončení interpretace TIL-Script funkce se interpreter vrací z metody zpět na místo, odkud byla metoda zavolána. S návratem je přirozeně spojena destrukce rámce na zásobníku – destrukce, která v současné implementaci obsahuje také zapomenutí kontextu funkce. Ten je poté uvolněn garbage collectorem během reklamačního cyklu.

Právě garbage collector je druhou výhodou současného přístupu. Interpretace AST a využití mechanismu zásobníku z JVM nám umožňuje využít také garbage collector z JVM pro automatickou správu paměti. Pokud bychom implementovali vlastní zásobník nezávislý na JVM, museli bychom naprogramovat také vlastní garbage collector.

Ve své podstatě však mechanismus stacku ani jednoduchý mark-and-sweep algoritmus³ není složité naprogramovat.

Podstatně složitějším problémem by byla definice bytekódu, a také řešení oboustranného překladu mezi bytekódem a JVM objekty.

4.3.2.2 Nezávislost knihoven na překladači

Projekt je koncipován tak, aby byly TIL-Script knihovny nezávislé na interpreteru, který uživatel použije. Implementovaný překladač je plně funkční, díky časovým omezením ovšem překladači chybí

³Myšleno mark-and-sweep algoritmus ve své nejjednodušší podobě. Mark-and-sweep algoritmus se používal již v padesátých letech minulého století, a používá se dodnes, v moderních jazycích je však silně optimalizován.

některé nutné prvky, jako třeba optimalizace koncového volání (TCO), jež je ve funkcionálních jazycích nezbytná, či překlad do bytekódu umožňující efektivnější překlad. V budoucnu se může stát, že bude potřeba nahradit současný překladač efektivnější implementací. V takovém případě ovšem není žádoucí, aby nastala potřeba přepsat nebo upravit také všechny již existující TIL-Script knihovny, standardní knihovnu, apod. Pro implementaci TIL-Script knihovny je však potřeba, aby např. TIL-Script funkce implementované v Javě měly určitý přístup k interpreteru, nebo alespoň kontextu, ve kterém jsou volány. Jinak by tyto funkce nemohly přistupovat např. k proměnným, k jiným funkcím, apod.

Za účelem odstranění závislosti na konkrétním překladači definuje knihovna `common` rozhraní `InterpreterInterface`. Toto rozhraní specifikuje základní funkce, které překladač TIL-Scriptu musí implementovat. Při volání TIL-Script funkcí definovaných pomocí Java objektů je danému objektu předána instance překladače jako potomek `InterpreterInterface`.

Tento princip je v softwarovém inženýrství velmi známý, je však nutné, aby na modularitu systému bylo myšleno již od počátku a aby byl dobře zdokumentován. Tato práce umožňuje, aby bylo v budoucnu možno vytvořit drop-in náhradu za konkrétní překladač bez jakékoliv úpravy existujícího kódu.