

Univerzita Jana Evangelisty Purkyně
v Ústí nad Labem
Přírodovědecká fakulta



Tacit programming - návrh doménově
specifického jazyka a implementace jeho
interpretu

BAKALÁŘSKÁ PRÁCE

Vypracoval: Oleg Musijenko

Vedoucí práce: Mgr. Jiří Fišer, Ph.D.

Studijní program: Aplikovaná informatika

Studijní obor: Informační systémy

ÚSTÍ NAD LABEM 2022

Zde bude vloženo zadání bakalářské práce!!

Zadání bakalářské je možné získat až v okamžiku, kdy je práce schválena ve STAGu vedoucím práce, vedoucím katedry a garantem oboru a nelze tak učinit elektronicky.

Požádejte o něj vedoucího katedry informatiky. Zadání vám bude doporučeno v podobě elektronicky podepsaného PDF, které vložíte na místo toho listu (dvojstránky) některým z nástrojů pro práci s PDF dokumenty.

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně a použil jen pramenů, které cituji a uvádím v přiloženém seznamu literatury.

Byl jsem seznámen s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., ve znění zákona č. 81/2005 Sb., autorský zákon, zejména se skutečností, že Univerzita Jana Evangelisty Purkyně v Ústí nad Labem má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Jana Evangelisty Purkyně v Ústí nad Labem oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

V Ústí nad Labem dne 9. dubna 2022

Podpis:

Děkuji vedoucímu práce Mgr. Jiřímu Fišerovi, Ph.D.
za neocenitelné rady a pomoc při tvorbě bakalářské práce.

Abstrakt

TACIT PROGRAMMING - NÁVRH DOMÉNOVĚ SPECIFICKÉHO JAZYKA A IMPLEMENTACE JEHO INTERPRETU

Abstrakt shrnuje základní motivaci práce (kontext), hlavní cíl a následně jednotlivé autorské kroky k jeho splnění (co bylo uděláno od úvodních rešerší, přes návrh, implementaci k případnému nasazení. Minimální rozsah je 800 znaků (maximální půl strany).

Klíčová slova

seznam klíčových slov (obecných termínů vystihujících téma práce) v počtu dva až deset

Abstract

TACIT PROGRAMMING - DESIGN OF A DOMAIN SPECIFIC LANGUAGE AND IMPLEMENTATION OF IT'S INTERPRETER

Translation of Czech abstract.

Key words

Translation of czech key words.

Obsah

Úvod	13
1 Tacit programming	15
1.1 Principy a odlišnosti od klasického procedurálního paradigmatu	16
1.2 Rešerše existujících implementací	18
2 DSL - principy a využití	19
3 Návrh vlastního DSL	21
4 Implementace interpretu navrženého DSL	23
5 Ověření použitelnosti (testování funkčnosti, praktické příklady využití)	25
6 Závěr	27
7 Citace	29
Bibliografie	31

Úvod

Jsou různá programovací paradigmatata s kterými se může programátor setkat. Mezi nejznámější paradigmatata patří strukturované, kde počátek tohoto paradigmatu se datuje v 1967 s Dijkstrovo článkem "Goto statement considered harmful". Dijkstra díky 'goto' výrazům nemohl určit správnost programu, proto se 'goto' nahradil za struktury typu 'if else', 'while', atd... Poté se můžeme setkat s objektově orientovaným paradigmatem, který je v dnešní době využíván jak pro menší aplikace, tak pro enterprise systémy. Toto paradigma dává programátorům možnosti využití polymorfismu, není zapotřebí využívat ukazatele na funkce a pomocí dnešních nástrojů je jednodušší se zorientovat v zdrojovém kódu. Jako finální paradigma je zde funkcionální. Funkcionální jazyky omezují primárně omezují mutaci existujících proměnných a zároveň se zaměřují na kompozici funkcí.

Existují samozřejmě i další paradigmatata, ale mnohdy je na ně pohlíženo jako na kuriozity. Třeba mezi takové patří deklarativní jazyk Prolog, který ověřuje pravdivost výroku v závislosti na předchozích relacích. Hlavně se budeme zaměřovat na tacit - "beztečkové" paradigma. Do tohoto paradigmatu spadá jazyky APL rodiny. Ukážeme si, že i jazyky, které nebyly navrženy jako "beztečkové" umožňují v tomto stylu psát.

Tato bakalářská práce předpokládá, že čtenář zná základy funkcionálních jazyků a obzvlášť Haskellu, protože návrh je vytvořen v Haskellu pomocí knihoven Parsec a LLVM.

1 Tacit programming

Tacit programming je styl programování, kde nevyužíváme parametry funkcí, místo toho funkce řetězíme, nebo kompozicujeme. Na následujících příkladech v programovacím jazyce javascript si tento princip ukážeme.

```
fetch("APIURL")
  .then(x => fancyFunction(x))
  .then(x => console.log(x))
  .catch(e => throw Error(e))
```

Po získání dat provedeme transformaci dat (definice fancyFunction není v tomto kontextu důležitá, stačí pouze vědět, že vrací Promise/Slib) a poté zapíšeme výsledek do konzole. Toto je jeden ze způsobů, jak můžeme na sebe řetězit funkce zpětného volání ("Callbacks").

Toto je zcela běžná praxe javascript programátorů, ale bohužel má jednu malou nevýhodu. Tvoří se zde zbytečná anonymní funkce ("arrow function nebo-li šipková") a pokud bychom prohlubovali čím dál víc zásobník volání, mohou nám tyto anonymní funkce zabírat paměť a během debugingu nám tento styl zápisu "znečišťuje" zásobník volání.

```
fetch("APIURL")
  .then(fancyFunction)
  .then(console.log)
  .catch(throw Error)
```

Zde se nachází kód, který dělá stejné instrukce, jako ten předchozí. Rozdíl je ten, že je zapsán jako *beztečkový* - "point free". Takto programátor předchází potřebě dělat obklopující lambda funkce.

Na následujícím příkladě si ukážeme, jak funguje **currying** a proč souvisí s tacit programováním.

```
const curry = (f) => a => b => f(a,b);

const sayHello = (a, b) = `Hello ${a} from ${b}`;

const applyToFunctionArray = (input,...args) => args.map(a => a(input))

const partiallyAppliedData = ["A", "B", "C"].map(curry(sayHello));
// [(b) => "Hello A from ${b}",
```

```
// (b) => "Hello B from ${b}",  
// (b) => "Hello C from ${b}"]  
  
const partiallyAppliedData2 = ["A", "B", "C"].map(curry(sayHello)(1));  
// ["Hello A from 1",  
//  "Hello B from 1",  
//  "Hello C from 1"]
```

Curry funkce transformuje existující funkci tak, že máme pro každý argument vlastní vracející funkci. Z funkce $f(a,b,c,d)$ vzniká funkce $f(a)(b)(c)(d)$ [1]. V čem je toto výhodné? Například je zde uvedené pole, které se skládá z částečně aplikovaných funkcí. Takto může programátor naiterovat odpověď ze serveru do objektu z předchozí ukázky, které je závislé na třeba na uživatelském vstupu.

Zajímavější část je u *partiallyAppliedData2*. Curryovaná funkce vrací funkci, jež očekává vstupní parametr, aby byla vyhodnocena. Tento princip je důležitý pro lenivé vyhodnocení, který využívá Haskell.

Může zde padnout argument, že v našem případě se curryování nachází pouze pro funkci, která přijímá dva argumenty. Zde je definice funkce, která převádí jakoukoliv funkci na curryovanou.

```
const curry = (f) => (...args) => args.length >= f.length ?  
  f.apply(this, args) : (...args2) => curry.apply(this, args.concat(args2));
```

1.1 Principy a odlišnosti od klasického procedurálního paradigmatu

Procedurální paradigma se zaměřuje na psaní procedurálních instrukcí. Typickým příkladem tohoto paradigmatu je programovací jazyk C, protože se jedná o standard, tak v následujících příkladech budu porovnávat jazyk C s jazykem Haskell. Haskell je primárně funkcionální jazyk, tento jazyk nám umožňuje psát funkce v "beztečkovém" stylu.

Ukažme si následující příklad sumace.

Haskell

```
sumCustom :: (Traversable t, Num a) => t a -> a  
sumCustom = foldr (+) 0
```


C

```
int sum(int* arr, size_t numElements)
{
    int acc = 0;

    for(int i = 0; i < numElements; i++)
    {
        acc += *(arr + i);
    }

    return acc;
}
```

Na příkladu jde vidět, že beztečkový styl zápisu je opravdu kompaktní. V Haskellu vůbec neinteragujeme s parametry funkce. Tento příklad je založen na podstatě tacit programmingu. Co se týče algoritmizace, tacit programming je známý pro vytváření algoritmických řešení pomocí pouze jednoho řádku kódu.

Na dalším příkladě si ukážeme fibonnacciho posloupnost.

Haskell

```
-- Využijeme lenivosti haskellu a vytvoříme nekonečnou
-- posloupnost fibonacciho čísel

fibonacci:: Num a => Int -> [a]
fibonacci = (flip take) fibonacciInfinite
    where
        fibonacciInfinite:: Num a => [a]
        fibonacciInfinite = scanl (+) 0 (1:fibonacciInfinite)
```

C

```
void fibonacci(int* arr, size_t numElements)
{
    if(numElements > 0)
    {
        arr[0] = 0;
    }
    if(numElements > 1)
    {
        arr[1] = 1;
    }
    for(int i = 2; i < numElements; i++)
    {
        arr[i] = arr[i - 1] + arr[i - 2];
    }
}
```

Z pohledu imperativního programátora implementace v C je zcela jasná. Funkce přijímá ukazatel na pole a modifikuje toto pole. Zatímco v Haskellu tato implementace může být matoucí. Funkce `scanl` je velice podobná funkci `foldl`, jen místo vracení akumulátoru, tak vrací průběžně vypočtené hodnoty.

1.2 Rešerše existujících implementací

2 DSL - principy a využití

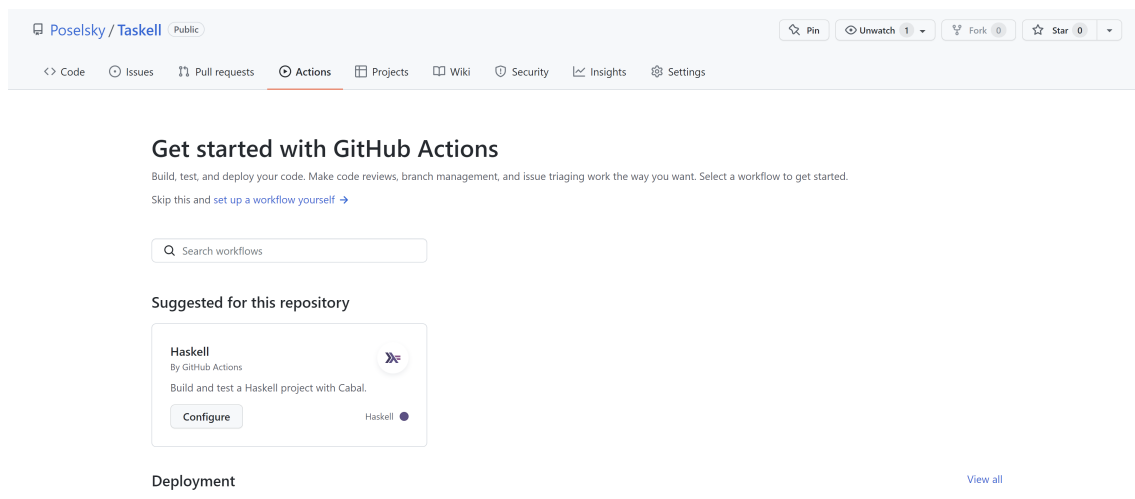
DSL (Domain Specific Language) jsou jazyky, které se zaměřují na specifickou doménu problematiky. Obecně DSL jazyky jsou mnohem jednodušší než jejich plnohodnotné protějšky. Výhodou je, že křivka učení je o mnohem nižší než u GPL (General Purpose Language). Zároveň pokud potřebujete na danou práci experty z oboru, jako jsou například doktoři nebo architekti, tak ty nepotřebují znát detaily implementace algoritmů, ale místo toho pokud budou mít přístup rovnou k DSL - šikmost stěny, hodnota cukrů v krvi pacienta, tak mohou plnit svoji práci o mnohem efektivněji. [2]

Jedním z takových typických jazyků je **HTML a CSS**. HTML se zaměřuje na vytvoření rámce pro zobrazení textu, zatímco CSS se zaměřuje na stylizaci webu pomocí DOM selectorů. Pravdou je, že pro CSS se nenachází žádný protocol a proto v různých webových enginech, můžete dostat různé výsledky. Příkladem z praxe je zpracování fontů.

Těž existují jazyky DSL, které jsou specifické pouze pro jednu danou enterprise aplikaci, kde její implementace často spočívá na bázi XML nebo podobného formátu jako je např YAML. Zde DSL slouží například pro zjednodušení UI nebo business logiky. Třeba pro porovnání **XAML** pro .NET platformu zjednodušuje logiku, stylizuje UI a zároveň zbavuje potřeby tvoření "glue" kódu.

Další jazyk který je velice využíván v hardwarovém prostředí je **VHDL** nebo **Verilog**. Tyto DSL jsou zaměřené na simulaci obvodů pomocí FPGA (hradlových polí). Pro building C/C++ projektů existuje **makefile**. Jedním ze zajímavějších DSL je DSL pro "continuous integration and deployment". Různé firmy co nabízejí online repositáře se v tomto budou trochu lišit, ale většina z nich poskytují jakousi formu automatizaci deploymentu. Toto poskytují firmy jako je GitHub, GitLab nebo Azure Dev Ops. Na GitHubu pomocí YAMLu můžete sepsat instrukce na testování a deployment aplikace.

Obrázek 2.1: Výstřížek z GitHub Actions



3 Návrh vlastního DSL

4 Implementace interpretu navrženého DSL

5 Ověření použitelnosti (testování funkčnosti, praktické příklady využití)

6 Závěr

7 Citace

[3] [4]

Bibliografie

- Currying partials* (2021). Unknown, Open source. URL: <https://javascript.info/currying-partial>.
- Federico, Tomasetti (2021). *Domain Specific Languages*. Itálie, Strumenta S.R.L. URL: <https://tomassetti.me/domain-specific-languages/>.
- Katuščák, Dušan, Barbora Drobíková a Richard Papík (c2008). *Jak psát závěrečné a kvalifikační práce. jak psát bakalářské práce, diplomové práce, dizertační práce, specializační práce, habilitační práce, seminární a ročníkové práce, práce studentské vědecké a odborné činnosti, jak vytvořit bibliografické citace a odkazy a citovat tradiční a elektronické dokumenty*. Nitra: Enigma. ISBN: 978-80-89132-70-6.
- Lattner, Chris (2008). *Intro to LLVM*. Erice, Sicily: Chris Lattner. URL: <https://llvm.org/pubs/2008-10-04-ACAT-LLVM-Intro.pdf>.