



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ

FACULTY OF INFORMATION TECHNOLOGY

ÚSTAV INFORMAČNÍCH SYSTÉMŮ

DEPARTMENT OF INFORMATION SYSTEMS

PŘEKLADAČ NOVÉHO MODULÁRNÍHO JAZYKA

COMPILER OF NEW MODULAR LANGUAGE

SEMESTRÁLNÍ PROJEKT

TERM PROJECT

AUTOR PRÁCE

AUTHOR

DANIEL ČEJCHAN

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. ZBYŇEK KŘIVKA, Ph.D.

BRNO 2017

Abstrakt

Do tohoto odstavce bude zapsán výtah (abstrakt) práce v českém (slovenském) jazyce.

Abstract

Do tohoto odstavce bude zapsán výtah (abstrakt) práce v anglickém jazyce.

Klíčová slova

Sem budou zapsána jednotlivá klíčová slova v českém (slovenském) jazyce, oddělená čárkami.

Keywords

Sem budou zapsána jednotlivá klíčová slova v anglickém jazyce, oddělená čárkami.

Citace

ČEJCHAN, Daniel. *Překladač nového modulárního jazyka*. Brno, 2017. Semestrální projekt. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce Křivka Zbyňek.

Překladač nového modulárního jazyka

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením pana Ing. Zbyňka Křivky, Ph. D. ... Další informace mi poskytli... Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

.....
Daniel Čejchan
22. prosince 2016

Poděkování

Velký dík patří i Stefanu Kochovi, který se na internetu vyskytuje pod přezdívkou Uplink-Coder; ve své ochotě se mnou prodebatoval velkou část konceptu jazyka.

V této sekci je možno uvést poděkování vedoucímu práce a těm, kteří poskytli odbornou pomoc (externí zadavatel, konzultant, apod.).

Obsah

1	Úvod	2
1.1	Motivace	2
1.2	Cíle	2
2	Vlastnosti a syntaxe jazyka	3
2.1	Existující prvky programovacích jazyků	3
2.2	Nové a netypické koncepty	9
3	Koncept líhnutí kódu (<i>code hatching</i>)	10
3.1	Dedukce axiomů	10
3.2	Přehled odvozených pravidel	17
3.3	Potenciál a důsledky konceptu	18
3.4	Implementace konceptu	21
	Literatura	22
	Přílohy	23

Kapitola 1

Úvod

Tento dokument popisuje a zdůvodňuje část z mnoha rozhodnutí, která byla vykonána při procesu návrhu programovacího jazyka Beast a vzorového překladače pro něj. Rámcově prozkoumává syntaktické i sémantické prvky moderních programovacích jazyků a popisuje i nové koncepty a prvky.

Programovací jazyk se hodně inspiruje jazyky C++ a D¹. Tyto jazyky budou používány pro srovnávání syntaxe a efektivity psaní kódu.

Příklady kódů z jazyka Beast uvedené v tomto dokumentu čerpají z kompletní specifikace jazyka, kompilátor vytvořený v rámci bakalářské práce ale implementuje pouze její podmnožinu – některé příklady tedy kompilátor není schopný zpracovat.

1.1 Motivace

Jazyk C++ je stále jedním z nejpoužívanějších jazyků².

1.2 Cíle

¹<http://dlang.org/>

²Podle <http://www.tiobe.com/tiobe-index/>

Kapitola 2

Vlastnosti a syntaxe jazyka

Nejdříve musíme určit základní rysy jazyka, které potom budeme dále rozvíjet.

Použití jazyka Naším cílem je navrhnout tzv. *general purpose language*, tedy jazyk nezaměřený na konkrétní případy užití. Cílíme vytvořit „nástupce“ jazyka C++, který by se dal použít ve všech případech, kde se dá využít C++ – tedy i třeba na mikroprocesorových systémech. Složitější struktury se tím pádem budeme snažit řešit spíše vhodnou abstrakcí než zaváděním prvků, které kladou zvýšené nároky na výkon a paměť.

Kompilovaný vs. interpretovaný jazyk Ačkoli interpretované jazyky mají jisté výhody, platí se za ně pomalejším kódem a nutností zavádět interpret. V rámci této práce budeme navrhnout kompilovaný jazyk. Pro zjednodušení práce při psaní překladače nebudeme ale překládat přímo do strojového kódu, ale do jazyka C.

Syntaxe jazyka Abychom maximálně usnadnili přechod případných programátorů k našemu jazyku a zkrátili učební křivku, je vhodné se co nejvíce inspirovat již existujícími jazyky. Vzhledem k tomu, že se jedná o kompilovaný jazyk, je nejrozumnější vycházet se syntaxe rodiny jazyků C, které jsou hojně rozšířené a zažité.

Programovací paradigmaty Jelikož vycházíme z jazyků C++ a D, přejímáme i jejich paradigmaty a základní koncepty. Naš jazyk bude tedy umožňovat strukturované, funkcionální i objektově orientované programování.

2.1 Existující prvky programovacích jazyků

2.1.1 Systém dědičnosti tříd

Jazyky C++ a D mají různě řešené třídni systémy. Zatímco v C++ existuje jen jeden typ objektu, který pracuje s dědičností (*class* a *struct* jsou z pohledu dědění identické), a to s dědičností vícenásobnou a případně i virtuální, D má systém spíše podobný Javě – třídy (*class*) mohou mít maximálně jednoho rodiče, navíc ale existují rozhraní (*interface*), které však nemohou obsahovat proměnné.

Ačkoli vícenásobné dědění není třeba často, občas potřeba je a jen těžko se nahrazuje. Rozhraní nemohou obsahovat proměnné, což omezuje jejich možnosti. D nabízí ještě jedno

řešení – tzv. *template mixins*¹ – které funguje velice podobně jako kopírování bloků kódu přímo do těla třídy. Toto řešení rozbíjí model dědičnosti (na *mixiny* se nelze odkazovat, nefungují jako rozhraní); navíc, protože jsou vloženy funkce prakticky součástí třídy, nefunguje v určitých případech kontrola přepisování (*overridingu*).

Ačkoli je implementace systému vícenásobné dědičnosti tak, jak je například v C++, složitější, její implementace je uskutečnitelná, kód nezpomaluje a fakticky rozšiřuje možnosti jazyka. C++ model dědičnosti umí vše, co umí Javovský model, a ještě víc.

Jazyk Beast tedy bude umožňovat třídní dědičnost, a to způsobem podobným C++. Nicméně implementace systému třídní dědičnosti není primárním cílem projektu, a tak je možné, že v rámci bakalářské práce nebude plně implementován.

Dá se přemýšlet i o zavedení různých direktiv, které umožňují dále nastavovat tento systém – tedy například přidat možnost nastavení zamezení generování tabulky virtuálních metod, atp.

2.1.2 Automatická správa paměti – garbage collector

Spousta moderních jazyků (mezi nimi i D) obsahuje garbage collector (dále GC) v základu. Jeho zavedení nenabízí pouze výhody – programy mohou být pomalejší, GC zvyšuje nároky na CPU a paměť (těžko se zavádí v mikroprocesorech); navíc efektivní implementace GC je velice složitá.

Rozumným kompromisem se jeví být volitelné používání automatické správy paměti. Beast by měl umožňovat efektivní napsání GC přímo v jazyce; GC by tedy mohla být jedna ze základních knihoven jazyka. Vzhledem k už tak velkému rozsahu plánované práce je však GC jen plánem do budoucna.

2.1.3 Implicitní konstantnost proměnných a propagace konstantnosti

Koncept konstantnosti proměnných byl zaveden jednak jako prvek kontroly při psaní kódu, jednak zvětšil potenciál kompilátorů při optimalizování kódu. Označení proměnné za konstantní ale u jazyku C++ (i D, Java, ...) vyžaduje ale napsání dalšího slova (modifikátoru *const* u C, C++ a D, *final* u Javy), a tak tuto praktiku (označovat všechno, co se dá, jako konstantní) spousta programátorů neaplikuje, zčásti kvůli lenosti, zčásti kvůli zapomnětlivosti.

Některé jazyky (například Rust²) přišly s opačným přístupem – všechny proměnné jsou implicitně konstantní a programátor musí použít nějakou syntaktickou konstrukci k tomu, aby to změnil. Beast tento přístup také zavede. V rámci koherence syntaxe jazyka (která je rozvedena v dalších kapitolách tohoto textu), připadají v úvahu dvě možnosti:

1. Vytvoření dekorátoru v kontextu `typeWrapper`³, nejlogičtěji `@mutable` nebo `@mut`
2. Vyčlenění operátoru; nejlepším kandidátem je suffixový operátor `Type!`, protože nemá žádnou standardní sémantiku, je nepoužitý a znak vykřičníku je intuitivně asociován s výstrahou, což je zase asociovatelné s mutabilitou.

Při designu Beastu byla zvolena druhá možnost, především kvůli „upovídánosti“ kódu a ještě kvůli jednomu důvodu, který je popsán níže.

¹<https://dlang.org/spec/template-mixin.html>

²<https://doc.rust-lang.org/nightly/book/mutability.html>

³Viz specifikace jazyka Beast (v příloze), oddíl *Decoration contexts*

Propagace konstantnosti Jazyk D je navržen tak, že je-li ukazatel konstantní (nelze měnit adresu, na kterou ukazuje), je přístup k paměti, na kterou ukazuje, také konstantní. Není tedy možné mít konstantní ukazatel na nekonstantní paměť. Tento mechanismus se nejvíce projevuje při konstantním referencování dat:

```

1 // D
2 class C {
3     int *x;
4 }
5
6 void main() {
7     C c = new C;
8     c.x = new int( 5 );
9
10    const C c2 = c;
11    *c2.x = 6; // Error: cannot modify const expression *c2.x
12 }
```

(2.1)

Toto je zbytečné limitování; tranzitivní konstantnost může být v některých případech užitečná, nicméně existují případy užití, ve kterých je překážkou. V Beastu tedy nebude vynucená propagace konstantnosti; můžeme mít konstantní ukazatel na nekonstantní data.

Chceme-li mít nekonstantní ukazatel na nekonstantní data, musíme tedy specifikovat mutabilitu dvakrát. Syntaktický zápis pro takovou referenci by vypadal takto (syntaxe pro referenci je Typ?, viz 2.1.5; příklady reflektují alternativy syntaktického zápisu referencí, které jsou uvedené výše v tomto oddílu):

1. @mut (@mut Typ)?
2. Typ!?!

Další možností by bylo zcela využít gramatiku C++, nicméně ta je všeobecně považována za velice matoucí – dochází tam ke kombinování suffixových a prefixových modifikátorů podle neintuitivních pravidel. Čistě suffixový zápis je jasný a jednoduchý a značně snižuje potřebu využívat závorek.

Srovnání syntaxe C++, D a Beast

```

1 // C++
2 int a, b; // Mutable integer
3 const int c, d; // Const integer
4 int *e, *f; // Mutable pointer to mutable integer
5 const int *g, *h; // Mutable pointer to const integer
6 int * const i, * const j; // Const pointer to mutable integer
7 const int * const k, * const l; // Const pointer to const integer
```

(2.2)

```

1 // D
2 int a, b; // Mutable integer
3 const int c, d; // Const integer
4 int* e, f; // Mutable pointer to mutable integer
5 const( int )* g, h; // Mutable pointer to const integer
```



```

6 | // const pointer to mutable integer not possible
7 | const int* k, l; // Const pointer to const integer

```

(2.3)

```

1 | // Beast
2 | Int32! a, b; // Mutable integer
3 | Int32 c, d; // Const integer
4 | Int32!?! e, f; // Mutable reference to mutable integer
5 | Int32?! g, h; // Mutable reference to const integer
6 | Int32!? i, j; // Const reference to mutable integer
7 | Int32? k, l; // Const reference to const integer

```

(2.4)

2.1.4 Typová kontrola a konverze typů

Beast má podobný mechanismus typových kontrol jako jazyky C++ a D:

- Typy mohou být implicitně konvertibilní do jiných typů. Proces konverze je definován v jednom z typů (přetížením funkcí `#implicitCastTo` nebo `#implicitCastFrom`)⁴. Implicitní konverze se provádí sama v případě potřeby⁵.
- Typy mohou být explicitně konvertibilní do jiných typů (funkce `#explicitCastTo` a `#explicitCastFrom`). K explicitnímu přetypování se používá výhradně funkce `a.to(Type)` (lze použít i k implicitnímu přetypování).

Důvodem k jiné syntaxi oproti C++ je opět snaha o zjednodušení pravidel a zvýšení přehlednosti kódu.

```

1 | // C++
2 | const int x = *( ( const int* )( voidPtr ) );

```

(2.5)

```

1 | // Beast
2 | Int32 x = voidRef.to( Int32? );

```

(2.6)

2.1.5 Reference, ukazatelé a jejich syntaxe

Typ ukazatel umožňující ukazatelovou aritmetiku je v dnešní době považován za potenciálně nebezpečný prvek, který by se měl užívat jen v nutných případech. Toto se řeší zavedením referencí, které v různých jazycích fungují mírně rozdílně, všeobecně se ale dá říci, že se jedná o ukazatele, které nepodporují ukazatelovou aritmetiku.

⁴Viz specifikace jazyka Beast (v příloze), oddíl *Type casting*

⁵Implicitní konverze ovlivňuje rezoluci volání přetížených funkcí, viz specifikace jazyka Beast (v příloze), oddíl *Overload resolution*

V C++ reference neumožňují měnit adresu odkazované paměti. V praxi ale programátor poměrně často potřebuje měnit hodnotu odkazu a musí se proto uchýlovat k ukazatelům, které podporují ukazatelovou aritmetiku a pro přístup k odkazované hodnotě je třeba buď použít dereferenci (která je implementována prefixovým operátorem `*ptr` a má tendenci znepřehledňovat kód) nebo speciální syntaktickou konstrukci (místo `x.y` `x->y`) pro přístup k prvkům odkazované hodnoty.

D k problematice přistupuje takto:

- Třídy jsou vždy předávány odkazem; adresa odkazované instance třídy se dá měnit, ukazatelová aritmetika není podporována (jako v Javě), přetěžování operátoru `a = b` není povoleno pro třídy jako levé operandy. Kromě tříd existují i struktury, které jsou předávány hodnotou; ty však nepodporují dědičnost.
- Existuje typ ukazatel, který pracuje s ukazatelovou aritmetikou. Pro přístup k odkazovanému prvku je třeba použít dereferenci (`*x`); k prvkům odkazované hodnoty lze použít klasické `x.y` (u C++ je třeba `x->y`). U dvojitého ukazatele je již potřeba použít dereferenci.
- Existuje i reference podobná té v C++, nicméně ta se dá použít pouze v několika málo případech (například v parametrech a návratových typech funkcí)

Vynucené předávání referencí má ale několik nevýhod:

- Programátor musí zajišťovat konstrukci (případně i destrukci) objektů
- Způsobuje více alokací a dealokací
- Indirekce bývá i tam, kde by být nemusela

Řešení Beastu Beast má umožňovat nízkoúrovňové programování, typ ukazatel s ukazatelovou aritmetikou je tedy nutné zavést. Na druhou stranu je ale vhodné nabídnout alternativu pro „běžné“ případy užití, které se většinou týkají vytváření instancí tříd na haldě a manipulace s nimi. Proto náš jazyk nabízí dva ukazatelové typy, kterým říkáme ukazatel a reference.

Způsob deklarace ukazatelů je v C++ i D problematický z hlediska parsování. Výraz `a * b` může totiž znamenat buď výraz násobení `a` krát `b`, stejně tak ale může znamenat deklaraci proměnné `b` typu ukazatel na `a` (obdobně i u reference). Nejjednodušším řešením je použití jiného znaku pro označování ukazatelů.

V Beastu je tímto znakem otazník (`?`). V C++ a D je používán pouze v ternárním operátoru (`cond ? expr1 : expr2`), kteréhož funkčnost Beast obstarává jiným způsobem⁶; tím pádem je znak otazníku „postradatelný“.

Ukazatel funguje skoro stejně jako v C++; podporuje ukazatelovou aritmetiku. Na dereferenci se nepoužívá prefixový operátor `*ptr` kvůli gramatickým konfliktům a protože může způsobovat nepřehlednost kódu a nejasnosti v prioritě operátorů (například u výrazu `*p++`). Místo toho se používá `ptr.data`. Přístup k prvkům odkazované hodnoty je možný pouze pomocí dereference. K získání ukazatele se také, kvůli stejným důvodům jako u dereference, nepoužívá prefixový operátor `&variable`, nýbrž `variable.addr`. Typ ukazatel není deklarován pomocí zvláštní syntaktické konstrukce (suffixový operátor `Typ?`

⁶Viz 2.2.3

přenechává hojněji používaným referencím), jedná se o kompilátorem definovanou třídu `Pointer(Type referencedType)`.

Reference je podobná referencím v C++, navíc ale umožňuje změnu odkazované adresy (může mít i hodnotu `null`). Neumožňuje zanořování – nelze definovat referenci na referenci (ukazatel na referenci ale možný je). Chová se stejně jako odkazovaná hodnota, až na několik výjimek:

- Je přetížen operátor `ref := var` a `ref := null`, který je určen pro změnu odkazované adresy reference.
- Je přetížen operátor `ref is null`, který navrací, zda má reference hodnotu `null`.
- `ref.addr` vrací ukazatel na referenci. `ref.refAddr` vrací ukazatel na odkazovaný objekt.

Všechny typy jsou implicitně konvertibilní na referenci (daného typu i jeho předků) a reference je explicitně konvertibilní na odkazovaný typ. Reference jsou implicitně konvertibilní na reference předků odkazované třídy a explicitně konvertibilní na referenci jakéhokoli typu; při těchto explicitních konverzích probíhá dynamická typová kontrola a jejich výsledek může být `null`.

Reference v Beastu pokrývají naprostou většinu případů užití ukazatelů při programování na vyšších úrovních abstrakce, a to se syntaxí takovou, že programátor nemusí rozlišovat referenci od normální proměnné. K programování na nižší úrovni se dá použít konvenční typ ukazatel `Pointer(T)`.

2.1.6 Dekorátory

2.1.7 Vykonávání funkcí za doby kompilace

2.1.8 Metaprogramování

Deklarace s `if` v D

- 2.1.9 Mixiny
- 2.1.10 Unified function call syntax (UFCS)
- 2.1.11 Systém výjimek
- 2.1.12 Lambda výrazy
- 2.1.13 Rysy (traits)
- 2.1.14 Standardní knihovna
- 2.2 Nové a netypické koncepty
 - 2.2.1 Konstrukce **:ident**
 - 2.2.2 Znak # v identifikátorech
 - 2.2.3 Ternární operátor
 - 2.2.4 Konstrukce **switch**
 - 2.2.5 Vnořovatelné komentáře
 - 2.2.6 Řetězcové literály
 - 2.2.7 Typy IntXX, BinaryXX a Index

Kapitola 3

Koncept líhnutí kódu (*code hatching*)

Toto je nový koncept navržený pro jazyk Beast. Zasahuje do několika již známých konceptů – například šablonové metaprogramování, reflexe jazyka, vykonávání funkcí za doby kompilace; všechny tyto koncepty spojuje do jednoho koherentního celku.

Koncept zavádí jednu jednoduchou myšlenku, ze které pak vyplývá celá řada důsledků. Tou myšlenkou je **zavedení klasifikátoru pro proměnné, jejichž hodnota se dá zjistit bez nutnosti spouštět program**. Tímto klasifikátorem je v jazyce Beast dekorátor `@ctime`.

Vykonávání kódu tak probíhá ve dvou fázích – hodnoty proměnných označených dekorátorem `@ctime` jsou odvozeny již za doby překladu, zbytek je vypočítáván za běhu samotného programu. **Toto se dá připodobnit k líhnutí vajec**, kdy se zárodek vyvíjí za skořápkou, ukryt před světem, a světlo světa spatří až jako vyvinutý jedinec.

3.1 Dedukce axiomů

Máme vyřčenou základní myšlenku – hodnoty proměnných označených dekorátorem `@ctime` musíme být schopni odvodit již během kompilace; nyní tuto myšlenku budeme rozvádět. Začneme jednoduchým příkladem:

```
1| @ctime Int x = 8;
```

(3.1)

Zde je vše jasné. Proměnná je konstantní, takže se nemůže měnit; po celou dobu její existence je její hodnota `osm`.

3.1.1 Datové závislosti

Další jednoduchý příklad:

```
1| @ctime Int x = console.read( Int );
```

(3.2)

Zde je zřejmé, že proměnná `x` nesplňuje naše požadavky. Příkaz `console.readNumber()` čte data z konzole a návratová hodnota této funkce se nedá zjistit bez spuštění programu (mohli bychom požádat o vstup již během kompilace, pro demonstraci konceptu ale toto

nyní neuvažujme). Z tohoto příkladu vyplývá, že `@ctime` proměnná nemůže být datově závislá na volání alespoň některých funkcí.

```
1 | Int add( Int x, Int y ) {
2 |     return x + y;
3 | }
4 |
5 | Void main() {
6 |     @ctime Int a = add( 5, 3 );
7 | }
```

(3.3)

V tomto příkladě lze hodnotu proměnné `a` určit. `@ctime` proměnné tedy mohou být datově závislé na volání funkcí, ale jen některých.

```
1 | Int foo( Int x ) {
2 |     if( x < 3 )
3 |         return console.read( Int );
4 |
5 |     return x + 1;
6 | }
7 |
8 | Void main() {
9 |     @ctime Int a = foo( 5 );
10 |    @ctime Int b = foo( 3 );
11 | }
```

(3.4)

Tento příklad ukazuje, že to, zda funkci lze použít pro výpočet hodnot `@ctime` proměnných, může záležet na předaných parametrech.

Funkce tedy k výpočtu hodnot `@ctime` proměnných všeobecně mohou být použity. To, jestli funkce opravdu lze použít, se zjistí až během vykonávání. Nelze použít žádné funkce, u kterých kompilátor nezná chování nebo závisí na externích datech (soubory, čas, vstup uživatele, ...).

Z tohoto úhlu pohledu nemá funkce smysl označovat dekorátorem `@ctime`; níže v této kapitole je popsáno, že se tento dekorátor na funkce používá, ale s trochu jiným účelem.

Všeobecně se dá říci, že `@ctime` proměnné nemohou být závislé na ne-`@ctime` datech. Hodnoty ne-`@ctime` proměnných mohou sice být odvoditelné bez nutnosti spuštění programu, ale nemusí. Teoreticky by bylo možné klasifikátor vynechat a odvozovat to pouze z kódu, v praxi by toto však značně zpomalilo kompilaci. Předpokládejme tedy, že **hodnoty všech proměnných, které nejsou `@ctime`, se před spuštěním programu nedají odvodit.**

3.1.2 Konstantnost

Nyní zauvažujme nad tím, co se stane, když budeme chtít měnit hodnoty `@ctime` proměnných:

```
1 | @ctime Int! x = 8;
2 | /* code here */
```

```

3 | x += 2;
4 | /* code here */

```

(3.5)

Zde je také všechno v pořádku. Ctime proměnné tedy nemusí být vždy konstantní.

```

1 | @static @ctime Int! x = 5;
2 |
3 | Void foo() {
4 |     @ctime Int! y = 5;
5 |     console.write( x, y, '\n' );
6 |     y += 3;
7 |     x += 3;
8 |     console.write( x, y, '\n' );
9 | }
10 |
11 | Void foo2() {
12 |     x += 2;
13 | }
14 |
15 | Void main() {
16 |     while( true ) {
17 |         if( console.read( Int ) < 2 )
18 |             foo();
19 |         else
20 |             foo2();
21 |     }
22 | }

```

(3.6)

Tady už narážíme na problém. Za doby kompilace nemůžeme zjistit hodnotu proměnné `x`, protože se mění na základě uživatelského vstupu (podle výsledku `console.read(Int)` se totiž volá buď funkce `foo` nebo `foo2`, kde každá manipuluje s proměnnou `x`). Nekonstantní statické proměnné se nedají ohlídat, **@ctime statické proměnné tedy musí být vždy konstantní.**

3.1.3 Podmínky a cykly

Větvení *if-then-else* a cykly sdílejí stejný princip – různé chování na základě hodnoty nějakého výrazu. Pravidla odvozená v tomto oddílu platí pro všechna větvení stejně.

```

1 | Void main() {
2 |     @ctime Int! x = 5;
3 |
4 |     while( x < 6 )
5 |         x += 3;
6 |
7 |     Int y = console.read( Int );
8 |     if( y < 2 )
9 |         x += 8;
10 | }

```

(3.7)

Z řádku 4 je patrné, že větvení v rámci `@ctime` proměnných je možné. Řádek 8 zase ukazuje, že to není možné vždy. Pokud je větvení datově závislé na výrazu, jehož hodnota se dá odvodit za doby kompilace (dále jen `@ctime` výrazu), mohou být v jeho těle přítomny `@ctime` proměnné. Pro usnadnění práce kompilátoru a zpřehlednění kódu Beast vyžaduje, aby `@ctime` větvení byly označeny dekorátorem (jinak se k nim přistupuje jako k `ne-@ctime`, viz dále).

Problematika je ale trochu složitější:

```
1 | Void main() {
2 |     @ctime Int! y = 6;
3 |
4 |     while( console.read( Int ) < 5 ) {
5 |         @ctime Int! z = 3;
6 |
7 |         console.write( y );
8 |         console.write( z );
9 |
10 |        z += 4;
11 |        console.write( z + y );
12 |
13 |        y += 2;
14 |    }
15 | }
```

(3.8)

Zde máme proměnnou `z`, která je definovaná v těle `ne-@ctime` cyklu. Nicméně takovéto použití proměnné naše požadavky nenarušuje, stejně tak čtení `z` proměnné `y` na řádku 7. V `ne-@ctime` větveních tedy mohou být použity `@ctime` proměnné a dokonce i definovány nekonstantní `@ctime` proměnné. Pravidlo, které se z tohoto příkladu dá vyvodit, je že **v tělech `ne-@ctime` větvení nelze měnit hodnoty `@ctime` proměnných deklarovaných mimo něj.**

3.1.4 Tranzitivita `@ctime`

Je zřejmé, že je-li proměnná `@ctime` programátorem definovaného třídního typu, musí být pro danou proměnnou všechny třídní proměnné toho typu také `@ctime`.

```
1 | class C {
2 |     @public Int x, y;
3 | }
4 |
5 | Void main() {
6 |     @ctime C c;
7 |     c.x = 5; // c.x is @ctime
8 | }
```

(3.9)

3.1.5 Reference a dynamické alokace

Pro praktickou demonstraci tohoto problému je již třeba složitější příklad:

```
1  class BinaryTreeNode {
2
3  @public:
4      String key;
5      Int! value;
6      BinaryTreeNode!?! left, right;
7
8  }
9
10 class BinaryTree {
11
12 @public:
13     BinaryTreeNode!?! root;
14
15 @public:
16     Void insert!( String key, Int value ) { ... }
17
18     /// BinaryTree[ "key" ] lookup
19     Int #operator( Operator.brackets, String key ) { ... }
20
21 }
22
23 BinaryTree sampleTree() {
24     BinaryTree result;
25     result.insert( "beast", 3 );
26     result.insert( "best", 5 );
27
28     result.top.data = 10;
29 }
30
31 @static @ctime BinaryTree tree = sampleTree();
32
33 Void main() {
34     console.write( tree["beast"] ); // the tree is @ctime, so the
35                                     compiler is capable of looking up value in the binary tree at
36                                     compile time, making this code extremely fast when running
37
38     tree.root.value = 7; // tree is const, but tree.root is not
39 }
```

(3.10)

Zde jsme si na řádce 31 definovali `@ctime` proměnnou `tree`. Třída `BinaryTree` potřebuje pro svou správnou funkčnost i bloky dynamicky alokované paměti (listy stromu; v kódu to explicitně uvedeno není, nicméně z jeho sémantiky to vyplývá). Ty pro správnou funkčnost stromu také musí splňovat podmínky `@ctime`. Ačkoli by teoreticky bylo možné mít `@ctime` ukazatel na ne-`@ctime` paměť (za doby kompilace by se vědělo, kam ukazatel odkazuje, ale ne, co na té paměti je), výhodnější je udělat `@ctime` tranzitivní i přes ukazatele a reference. Pokud bychom to takto nenastavili, kód příkladu tohoto oddílu by pro funkčnost vyžadoval zvláštní úpravy, nebo by nemohl fungovat vůbec.

Protože `tree` je statická `@ctime` proměnná, musí být konstantní. Dynamicky alokované bloky (listy stromu) vytvořené při její inicializaci tedy také musí být konstantní. Toto ale neplatí zcela – během inicializace proměnné `tree` (během vykonávání funkce `sampleTree` za doby kompilace) s bloky normálně manipulujeme (řádky 25 a 26). Tyto bloky se tedy stávají konstantními až po dokončení inicializace. Zde vyvstává problém – v naší třídě `BinaryTree` je ukazatel na dynamicky alokovaný blok typu `BinaryTreeItem!?!` – odkazovaná paměť je z odkazu mutabilní. Řádek 36 je tedy technicky korektní, i když by podle výše uvedených myšlenek neměl být.

Jak tedy zajistit, aby se po inicializaci proměnné `tree` nedalo k jejím listům přistupovat přes mutabilní reference? Tento problém by se dal vyřešit zavedením vynucené tranzitivity konstantnosti referencí a ukazatelů – tedy že při konstantní referenci/ukazateli by se odkazovaná paměť automaticky brala jako také konstantní. V oddílu 2.1.3 jsme ale rozhodli, že jazyk Beast nebude vynucovat propagaci konstantnosti přes reference – programátor si ji musí zajistit sám tam, kde je to potřeba. U binárního stromu je logické, že je-li konstantní strom, měly by jeho listy být také konstantní; v případě programátorem definovaných typů ale nelze zaručit, že je mutabilita korektně ošetřena.

Při kompilaci se konstantní statické proměnné v některých kompilátorech umísťují do tzv. *.text section* ve výsledném binárním souboru. Stránky paměti načtené z těchto sekcí jsou chráněny operačním systémem proti zápisu a pokus o zapsání do nich vyvolává běhovou chybu (*segfault/access violation*). Beast řeší problém obdobným způsobem – při pokusu o zápis do paměti (během kompilace), která byla alokována během inicializace statické `@ctime` (nebo i jen konstantní) proměnné, mimo dobu inicializace zmíněné proměnné vyvolá kompilátor chybu. Dalším prvkem ochrany je zákaz existence mutabilních referencí a ukazatelů na dané bloky mimo rámec inicializované proměnné.

3.1.6 Konverze na ne-@ctime

Je vhodné objasnit, že dekorátor `@ctime` není modifikátor typu (`typeModifier`¹), ale modifikátor proměnné/parametru (`parameterModifier/variableModifier`). To znamená, že typy proměnných `ref` a `cref` v příkladu 3.11 jsou shodné (`ref.#type == cref.#type == Int!?`). Zákaz přiřazování výrazů závislých na ne-`@ctime` datech do `@ctime` proměnných je řešen jinak než typovou kontrolou.

```

1 | Void main() {
2 |     @ctime Int! x = 3;
3 |
4 |     @ctime Int!? cref := x;
5 |     cref = 9;
6 |
7 |     Int!? ref := cref;
8 |
9 |     if( console.read( Int ) < 5 )
10 |         ref := 8;
11 |
12 |     @ctime Int y = x;
13 | }
```

(3.11)

¹Viz specifikace jazyka Beast (v příloze), oddíl *Decoration contexts*

Výše uvedený příklad neporušuje žádné z pravidel, které jsme si již odvodili, nicméně hodnota proměnné `y` již není odvoditelná během kompilace, protože se hodnota proměnné `x` může změnit na základě uživatelského vstupu. Přiřazení do `cref` na řádku 5 je v pořádku; kontrola nad daty se ztrácí při inicializaci proměnné `ref` na řádku 7. Aby se tomuto předešlo, zavedeme pravidlo, že **na @ctime data nelze odkazovat ne-@ctime referencí/ukazatelem na mutabilní typ**.

Toto pravidlo ale není dostatečné:

```

1 | class C {
2 |     @public Int! x;
3 | }
4 |
5 | class D {
6 |     @public C!?! c = new C;
7 | }
8 |
9 | Void main() {
10 |     @ctime D d;
11 |
12 |     D? dref := d;
13 |     dref.c.x = 5;
14 | }
```

(3.12)

Zde jsme toto pravidlo neporušili, ale přesto se nám opět podařilo porušit axiom pro @ctime. Zavedeme tedy další pravidlo: **odkaz na @ctime data nelze uložit do ne-@ctime reference/ukazatele, pokud všechny třídní proměnné typu reference/ukazatel v odkazovaném typu nejsou konstantní**.

Toto pravidlo se ale také dá obejít:

```

1 | class C {
2 |     @public Int x;
3 | }
4 |
5 | class C2 : C {
6 |     @public Int!? y = new Int;
7 | }
8 |
9 | class D {
10 |     @public C? c = new C2;
11 | }
12 |
13 | Void main() {
14 |     @ctime D d;
15 |
16 |     D? dref := d;
17 |     dref.c.to( C2? ).y = 5;
18 | }
```

(3.13)

Řešení tohoto problému už je značně problematické. Kvůli způsobu implementace by zápis do paměti připadající @ctime datům vedl k nedefinovanému chování aplikace, takže je třeba mu zabránit. Všem těmto komplikacím by se předešlo vynucením tranzitivity kon-

stantnosti přes referenci a vzniká tak dilema, zda tranzitivitu vynutit i za cenu omezení možností programátora, nebo ponechat programátorovi místo, kde by se nechtěně mohl „střelil do nohy“. Na základě názoru, že „jazyk má sloužit programátorovi, ne programátor jazyku“ se autor přiklání spíše k variantě nevynucovat tranzitivitu.

3.1.7 Třídní @ctime proměnné

Zvažme možnost existence nestatických @ctime proměnných ve třídách:

```

1  class C {
2
3  @public:
4      @ctime Int! y = 0;
5
6  @public:
7      Void foo!() {
8          y += 3;
9      }
10
11 }
12
13 Void main() {
14     C! c, c2;
15
16     C!?! cref = select( console.read( Int ) < 4, c, c2 );
17     cref.foo();
18 }
```

(3.14)

Z příkladu lze odvodit, že myšlenka není uskutečnitelná. Volání funkce `foo` je datově závislé na uživatelském vstup a nedá se tomu zabránit jinak, než označením proměnných `c` a `c2` jako @ctime (což by znemožnilo použití `console.read`). I kdybychom zavedli „@ctime“ funkce, které jako parametry přijímají pouze @ctime data, k ničemu by nám to nebylo. Toto se dá chápat i tak, že funkce `foo` obsahuje skrytý parametr – odkaz na instanci třídy. Protože proměnná `cref` není @ctime, právě tento parametr by porušoval axiom @ctime.

3.2 Přehled odvozených pravidel

1. @ctime proměnné nemohou být jakkoli (datově) závislé na proměnných, které nejsou @ctime.
2. Statické @ctime proměnné musí být konstantní.
3. Větvení může být @ctime, pokud je výraz v nich datově závislý pouze na @ctime proměnných.
 - (a) Těla @ctime větvení nejsou nijak omezena.
 - (b) V tělech ne-@ctime větvení se nemůže měnit data @ctime proměnných, které byly deklarovány mimo něj (ale lze je číst).
4. Je-li proměnná @ctime, jsou všechny její třídní proměnné také @ctime.

5. Je-li reference nebo ukazatel `@ctime`, paměť, na kterou odkazuje, je také `@ctime`.
6. Po inicializaci statické `@ctime` proměnné nelze zapisovat do paměti, která byla během její inicializace alokována.
7. Ukládání `@ctime` proměnných do `ne-@ctime` proměnných má další omezení:
 - (a) Na `@ctime` data nelze odkazovat `ne-@ctime` referencí/ukazatelem na mutabilní typ.
 - (b) Odkaz na `@ctime` data nelze uložit do `ne-@ctime` reference/ukazatele, pokud všechny třídní proměnné typu reference/ukazatel v odkazovaném typu nejsou konstantní.
8. Třída (`ne-@ctime`) nemůže obsahovat nestatické `@ctime` proměnné.

3.3 Potenciál a důsledky konceptu

Když už máme odvozená pravidla, můžeme si ukázat, co všechno nám tento koncept umožňuje.

3.3.1 Optimalizace kódu

Pokud kompilátor zná hodnoty `@ctime` proměnných již za doby kompilace, logicky může výpočet jejich hodnot vynechat při běhu. Kód

```
1 | void main() {
2 |     @ctime Bignum myPrime = nthPrimeNumber( 1286345 );
3 |     console.write( myPrime );
4 | }
```

(3.15)

se vyoptimalizuje do

```
1 | void main() {
2 |     console.write( 20264747 );
3 | }
```

(3.16)

Toto samozřejmě umí hodně současných programovacích jazyků. Dekorátor `@ctime` ale dává programátorovi pevnější kontrolu nad optimalizacemi.

3.3.2 Typové proměnné

Díky tomu, že kompilátor zná hodnotu `@ctime` proměnné v každém bodě již během kompilace, může měnit chování `@ctime` proměnné na základě její hodnoty. To umožňuje zavedení mutabilních typových proměnných. Všechny třídy jsou v jazyce Beast instancemi třídy `Type` (i samotná třída `Type` je svou vlastní instancí). Identifikátory tříd jsou ekvivalentní konstantním typovým proměnným.

```

1 | Void main() {
2 |     @ctime Type T = Int16;
3 |
4 |     @ctime if( VERSION > 15 )
5 |         T = Int32;
6 |
7 |     T x = 4;
8 | }

```

(3.17)

```

1 | class C {
2 |     @static Int! x;
3 | }
4 |
5 | class D {
6 |     @static Int! y;
7 | }
8 |
9 | void Main() {
10 |     @ctime Type T = C;
11 |     T.x = 5;
12 |
13 |     T = D;
14 |     D.y = 8;
15 | }

```

(3.18)

Typová proměnná může existovat pouze jako `@ctime`.

3.3.3 Šablonování a unifikace šablonových a klasických parametrů

Koncept `@ctime` se dá jednoduše využít i pro parametry šablon (v parametrech šablon není třeba používat dekorátor `@ctime`).

```

1 | class TreeNode( Type Key, Type Value, Int childrenCount ) {
2 |
3 | @public:
4 |     Key! key;
5 |     Value! value;
6 |     This?!?[ childrenCount ] children;
7 |
8 | }

```

(3.19)

Díky prakticky nulovému rozdílu v syntaxi `@ctime` a `ne-@ctime` proměnných lze šablonové a standardní parametry funkcí zapisovat do stejných závorek a libovolně je míchat:

```

1 | T readFromBuffer( Buffer!?! buf, @ctime Type T ) {
2 |     T? result = buf.ptr.to( T? );
3 |     buf.ptr += T.#size;
4 |     return result;

```

```

5 | }
6 |
7 | Void main() {
8 |     Buffer! buf;
9 |     buf.readFromFile( File( "in.txt", :read ) );
10 |     Int i = readFromBuffer( buf, Int );
11 | }

```

(3.20)

Dokonce lze vytvořit i speciální dekorátor, v Beastu pojmenovaný `@autoctime`, který argument bere jako `@ctime`, pokud to je možné:

```

1 | Index regexFind( String pattern, @autoctime String regex ) {
2 |     // regex implementation
3 | }
4 |
5 | Void main() {
6 |     Index r1 = regexFind( console.read( String ), "Y[a-z]+l" ); //
7 |         Regex argument is known at compile time, so it is considered
8 |         to be @ctime and the regex parsing is done at compile time
9 |
10 |     Index r2 = regexFind( console.read( String ), console.read(
11 |         String ) ); // No optimizations here
12 | }

```

(3.21)

Pomocí klíčového slova **auto** může funkce přijmout hodnotu jakéhokoli typu

```

1 | // C++
2 | template< typename T1, typename T2 >
3 | inline auto max( T1 a, T2 b ) {
4 |     return ( a > b ) ? a : b;
5 | }

```

(3.22)

```

1 | // Beast
2 | @inline auto max( auto a, auto b ) = select( a > b, a, b );

```

(3.23)

3.3.4 `@ctime` funkce a třídy

Pro usnadnění práce programátora se zavádí i podpora dekorátoru `@ctime` pro funkce a třídy.

`@ctime` funkce mají všechny argumenty, návratovou hodnotu, stejně tak i veškerý kód uvnitř automaticky `@ctime` a není tedy třeba všechny prvky takto dekorovat.

Proměnné `@ctime` tříd mohou být pouze `@ctime`. Všechny proměnné a funkce `@ctime` tříd jsou `@ctime`.

3.3.5 Reflexe kódu

Koncept také umožňuje reflexi. Vezmeme-li v úvahu, že každá třída je instancí `@ctime` třídy `Type`, stačí už jen přidat do těchto tříd další `@ctime` proměnné, které je popisují. Jazyk Beast zavádí `@ctime` konstanty jako `Symbol.#identifier`, `Function.#returnType`, třídy `FunctionMetadata`, `ClassMetadata` a funkce jako `Class.#member(String identifier)`, které reflexi zajišťují. Jejich dokumentaci naleznete v referenci jazyka.

3.4 Implementace konceptu

- šablonové parametry - bitové porovnávání
- paměť musí existovat
- sémantická kontrola neprošlých `@ctime` větví se neprovádí

Literatura

Přílohy