

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Toni Kocjan Turk

**Vgradnja objektno usmerjenih
gradnikov v programski jezik PINS**

DIPLOMSKO DELO

VISOKOŠOLSKI STROKOVNI ŠTUDIJSKI PROGRAM
PRVE STOPNJE
RAČUNALNIŠTVO IN INFORMATIKA

MENTOR: doc. dr. Bištjan Slivnik

Ljubljana, 2017

COPYRIGHT. Rezultati diplomske naloge so intelektualna lastnina avtorja in Fakultete za računalništvo in informatiko Univerze v Ljubljani. Za objavo in koriščenje rezultatov diplomske naloge je potrebno pisno privoljenje avtorja, Fakultete za računalništvo in informatiko ter mentorja.

Besedilo je oblikovano z urejevalnikom besedil \LaTeX .

Fakulteta za računalništvo in informatiko izdaja naslednjo nalogo:

Tematika naloge:

Besedilo teme diplomskega dela študent prepíše iz študijskega informacijskega sistema, kamor ga je vnesel mentor. V nekaj stavkih bo opisal, kaj pričakuje od kandidatovega diplomskega dela. Kaj so cilji, kakšne metode uporabiti, morda bo zapisal tudi ključno literaturo.

Na tem mestu zapišite, komu se zahvaljujete za izdelavo diplomske naloge. Pazite, da ne boste koga pozabili. Utegnil vam bo zameriti. Temu se da izogniti tako, da celotno zahvalo izpustite.

Svoji dragi Alenčici.

Kazalo

Povzetek

Abstract

1	Uvod	1
2	Prevajalniki	3
2.1	Uvod v prevajalnike in programske jezike	3
2.2	Zgradba prevajalnika	4
3	Programski jezik PINS	13
3.1	Leksikalna pravila	13
3.2	Sintaksna pravila	13
3.3	Semantična pravila	15
4	Programski jezik Atheris	17
4.1	Sintaksa	18
4.2	Funkcije	20
4.3	Enumeracije	22
4.4	Terke	24
4.5	Razredi	24
4.6	Dedovanje in polimorfizem	29
4.7	Instance of operator	31
4.8	Pretvarjanje tipov	32
4.9	Vmesniki	33
4.10	Razširitve	33
4.11	Type inference	34
5	Meritve in testiranje	37
5.1	Bubble sort	37

5.2	Quick sort	38
5.3	Rekurzivni fibonacci	38
5.4	Quicksort nad razredi	38
6	Sklepne ugotovitve	39
	Literatura	40

Seznam uporabljenih kratic

kratica	angleško	slovensko
CA	classification accuracy	klasifikacijska točnost
DBMS	database management system	sistem za upravljanje podatkov- nih baz
SVM	support vector machine	metoda podpornih vektorjev

Povzetek

Naslov: Vgradnja objektno usmerjenih gradnikov v programski jezik PINS

Avtor: Toni Kocjan Turk

V diplomskem delu bom predstavil programski jezik Atheris, ki je nastal kot nadgradnja programskega jezika PINS. Prog. jezik PINS, oz. prevajalnik zanj, je bil zgrajen tekom semestra pri predmetu prevajalniki in navidezni stroji. Ker mi je bilo delo na prevajalniku izjemno zanimivo, sem se odločil, da ustvarim svoj programski jezik in sam določim pravila zanj.

V diplomskem delu na kratko predstavim prevajalnike in programske jezike, kaj sploh so in kaj je njihov namen. Opišem kakšne so sodobne prakse pri ravoju prevajalnikov, s kakšnimi probleme se prevajalnik sooča ter kako je zgrajen.

Podrobneje bom obrazložil nadgradnje jezika PINS, s kakšnimi problemi sem se tekom razvoja soočal ter kako sem jih reševal. Pokazal bom nekaj primerov programov v mojem jeziku ter primerjal z drugimi jeziki.

Ključne besede: prevajalnik, programski jezik, sintaksa, semantika, Java, Swift.

Abstract

Title: Diploma thesis sample

Author: Toni Kocjan Turk

This sample document presents an approach to typesetting your BSc thesis using \LaTeX . A proper abstract should contain around 100 words which makes this one way too short.

Keywords: compiler, programming language, syntax, semantics, Java, Swift.

Poglavje 1

Uvod

Razvoj prevajalnikov, ter s tem tudi programskih jezikov, je, po mojem mnenju, izjemno pomembna panoga v računalništvu. Programski jezik je medij, preko katerega komuniciramo z računalnikom. Prevajalniki razvijalcem omogočajo, da se med razvojem programske opreme ne rabijo osredotočati na nizkovojske detaile, ampak se lahko posvetijo reševanju praktičnih problemov. Naloga prevajalnika je, da pretvori človeku berljivo kodo v računalniku razumljivo zaporedje strojnih ukazov.

Dandanes lahko za razvoj programske opreme izbiramo med veliko programskih jezikov. Trenutno eni izmed najbolj popularnih so JavaScript, Java, Python in C++, popularnost pa dobivajo tudi novejši jeziki, kot so GoLang, Swift, Kotlin in podobni. [1]

S prevajalniki sem se začel ukvarjati pri predmetu Prevajalniki in Navidezni Stroji (PINS) v drugem letniku na Fakulteti za Računalništvo in Informatiko. Tekom diplomske naloge bom predstavil programski jezik Atheris ter prevajalnik zanj. Osredotočil se bom predvsem na vgradnjo objektno usmerjenih gradnikov v programski jezik.

Poglavje 2

Prevajalniki

2.1 Uvod v prevajalnike in programske jezike

Programski jezik je poseben jezik, ki se uporablja za razvoj programske opreme. Programski sistemi, ki poskrbijo, da se izvede pretvorba kode, napisane v programskem jeziku, v računalniku razumljivo obliko, se imenujejo *prevajalniki*. Nekaj definicij:

1. **Računski model** (angl. **computational model**): zbirka vrednosti in računskih operacij
2. **Izračun** (angl. **computation**): zaporedje operacij nad vrednostjo (ali več vrednosti), ki vrne nek rezultat
3. **Program**: specifikacija izračuna
4. **Programski jezik**: zapis (notacija) za pisanje programov

[1]

Program lahko predstavimo kot funkcijo, pri kateri je rezultat (angl. *output*) funkcija vhodnih parametrov (angl. *input*):

```
rezultat = program(vhodni parametri)
```

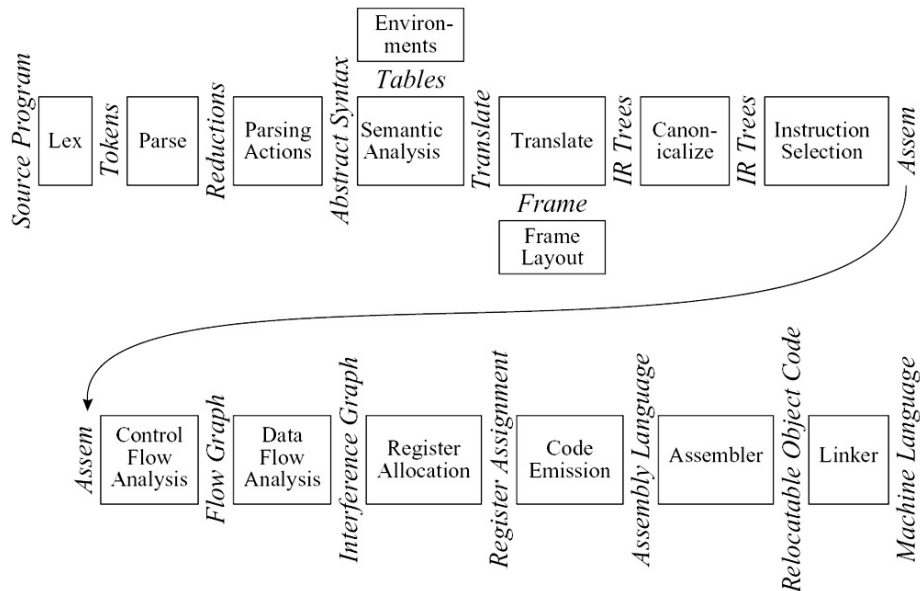
Iz drugega zornega kota si lahko program predstavljamo tudi kot model problemske domene, kjer je instanca izvedbe programa simulacija problema:

```
program = model problemske domene  
izvedba programa = simulacija problema
```

[1]

2.2 Zgradba prevajalnika

Sodobni prevajalniki so pogosto organizirani v več posameznih faz, vsaka izmed njih pa operira na različnem nivoju abstrakcije jezika. [2]



Slika 2.1: Faze prevajalnika ter vmesniki, ki jih povezujejo med seboj.

Zato da lahko prevajalnik program prevede iz ene oblike v drugo, ga mora najprej analizirati, razumeti njegovo strukturo ter pomen, šele nato ga pretvori nazaj v drugačno obliko.

Analizo programa običajno delimo v naslednje korake:

1. **Leksikalna analiza** (angl. *lexical analysis*)
2. **Sintaksna analiza** (angl. *syntax analysis*)
3. **Semantična analiza** (angl. *semantic analysis*)

[2]

2.2.1 Leksikalna analiza

Tako imenovani leksikalni analizator kot vhod prejme tok znakov (angl. *stream of characters*), kot izhod pa vrne tok v naprej definiranih žetonov (angl. *stream of tokens*). Žeton je običajno zgrajen iz imena, vrednosti (t.i. *lexeme*) ter lokacije v izvorni datoteki. [2]

vrsta žetona	angl.	primeri
ime	identifier	x foo bar thisIsAnIdentifier
rezervirana beseda	keyword	while for if public override
operator	operator	, . && ==
niz znakov	string	"this is a string"
znak	character	'a' 'x' '@'
celo št.	integer	10 125 082
decimalno št.	real	201.5 3.14 1.2e10

Tabela 2.1: Primeri žetonov v programskem jeziku Java

Leksikalni žetoni:

Žeton (ali simbol) je zaporedje znakov, ki ga interpretiramo kot samostojno enoto v slovnici programskega jezika. [2]

Tabela 2.1 prikazuje nekaj vrst simbolov ter primere.

```

let x: Int
let y: Int
x * y

```

Listing 2.1: Primer programa v programskem jeziku Atheris

Rezultat:

```

[1:1-1:4]      LET:let
[1:5-1:6]      IDENTIFIER:x
[1:6-1:7]      COLON::
[1:8-1:11]     IDENTIFIER:Int
[1:11-1:12]    NEWLINE:\n
[2:1-2:4]      LET:let
[2:5-2:6]      IDENTIFIER:y
[2:6-2:7]      COLON::
[2:8-2:11]     IDENTIFIER:Int
[2:11-2:12]    NEWLINE:\n
[3:1-3:2]      IDENTIFIER:x
[3:3-3:4]      MUL:*
[3:5-3:6]      IDENTIFIER:y
EOF:$

```

Listing 2.2: Rezultat leksikalne analize za program 2.1

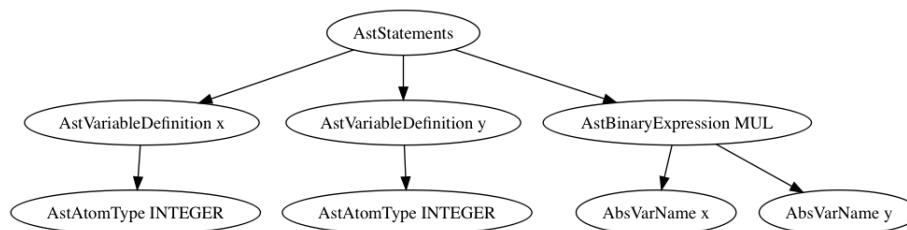
2.2.2 Sintaksna analiza

Druga faza prevajanja je sintaksna analiza (angl. *syntax analysis* ali *parsing*). Naloga te faze je, da zagotovi, da je napisan program slovnično pravilen in v skladu s sintaksnimi pravili. Sintaksni analizator prejme kot vhod tok žetonov, ki ga zgenerira prejšnja faza, rezultat pa je abstraktno sintaksno drevo.

Abstraktno sintaksno drevo (AST) je drevesna podatkovna struktura, ki predstavlja slovnično strukturo programa. Vsako vozlišče drevesa ponazarja konstrukt v programski kodi.

Iz slike 4.1 lahko razberemo, da gre za dve definiciji spremenljivk in množenje.

Abstraktno sintaksno drevo je bistvenega pomena, saj nadaljne faze operirajo izključno nad njim.



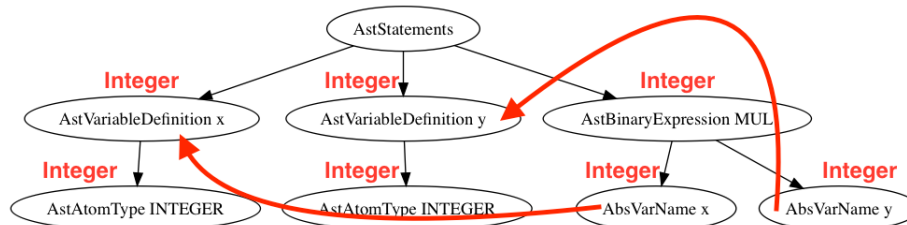
Slika 2.2: Abstraktno sintaksno drevo za program 2.1.

2.2.3 Semantična analiza

Semantična analiza poveže definicije spremenljivk z njihovimi uporabi ter preveri, ali so vsi izrazi pravih podatkovnih tipov. [2]

Običajno semantično analizo razdelimo na dve pod-fazi:

1. **Razreševanje imen:** zagotovi, da za vsako uporabo imena obstaja znotraj trenutnega območja vidnosti definicija z istim imenom, ter uporabo poveže z definicijo
2. **Preverjanje tipov:** vsakemu vozlišču v AST določi podatkovni tip, ter na podlagi postavljenih semantičnih pravil zagotovi, da so vsi izrazi pravih tipov



Slika 2.3: Rezultat semantične analize za program 2.1

Pri implementaciji semantične analize nam pomaga *simbolna tabela*.

Simbolna tabela

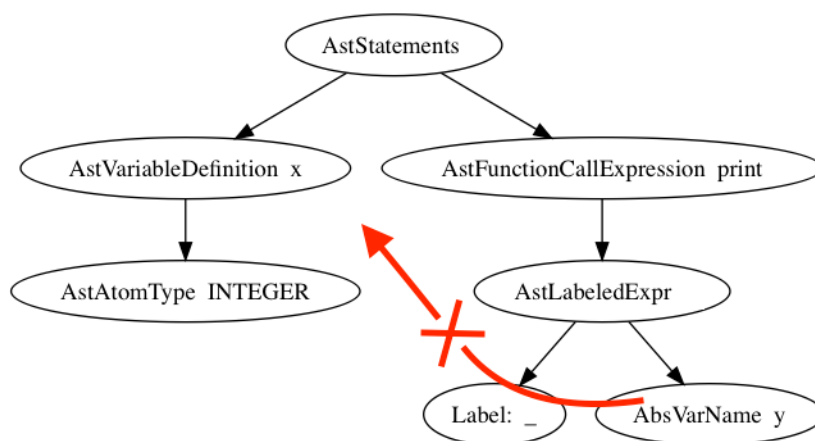
Simbolna tabela je podatkovna struktura, ki mapira imena v njihove definicije in podatkovne tipe. [2] Ker običajno programi vsebujejo več tisoč unikatnih definicij imen, mora podatkovna struktura omogočati učinkovito poizvedovanje. Iz slike 2.3 lahko razberemo, kaj se med izvajanjem semantične analize zgodi

v ozadju: puščice predstavljajo povezave med definicijami in uporabami, evaluacija podatkovnih tipov pa je pri vseh vozliščih *Integer*, razen pri korenu, ki nima tipa oz. je tipa *Void*.

Kot sem omenil, semantična analiza zagotovi, da za vsako uporabo imena obstaja njena definicija, in da so podatkovni tipi pravilni. Sledita dva primera, kjer to ne drži:

```
let x: Int
print(y)
```

Listing 2.3: Primer programa, kjer spremenljivka *y* ni definirana



Slika 2.4: Napaka v programu 2.3. Spremenljivka *y* ni definirana.

```
let x: Int
let s: String
x + s
```

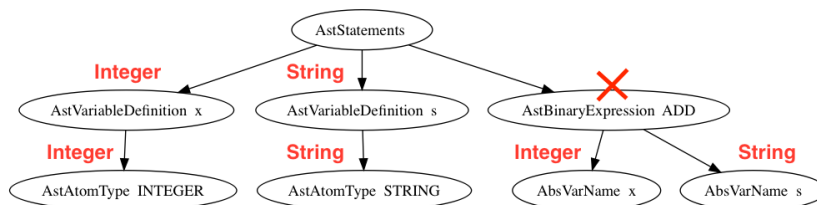
Listing 2.4: Primer programa, kjer je napaka v podatkovnih tipih

2.2.4 Klicni zapisi

V skoraj vsakem modernem programskem jeziku ima lahko funkcija *lokalne* spremenljivke, ki so kreirane ob vstopu v funkcijo. Hkrati lahko naenkrat obstaja več zapisov iste funkcije, zato je pomembno, da ima vsak zapis lastne instance lokalnih spremenljivk. [2]

V funkciji

```
todo
```

Slika 2.5: Napaka v programu 2.4. Seštevanje med podatkovnima tipoma *Integer* in *String* ni dovoljeno.

se za vsak njen klic ustvari nova instanca x , ki jo inicializira klicalec funkcije. Ker je funkcija rekurzivna, živi v pomnilniku naenkrat veliko x -ev. Podobno se ob vsakem vstopu v jedro funkcije ustvari tudi nova instanca y . [2]

Sklad

Klicni zapisi funkcij se shranjujejo na sklad. Sklad je v pomnilniku predstavljen kot velika tabela s posebnim registrom imenovanim *stack pointer*, ki kaže na konec sklada. Ob vsakem klicu funkcije se sklad poveča za velikost klicnega zapisa klicane funkcije. Podobno se ob vrnitvi funkcije zmanjša za enako vrednost. Prostor v klicnem zapisu namenjen hrambi vhodnih parametrov, lokalnih spremenljivk in ostalih registrov se imenuje *activation record*. [2]

Kazalec na klicni zapis

Predpostavimo da funkcija g kliče funkcijo f . Ob vstopu v f kazalec na sklad (SP) kaže na prvi vhodni argument funkciji f . Nov prostor na skladu je rezerviran tako, da se od SP odšteje velikost klicnega zapisa f . Tako SP sedaj kaže na konec sklada. Stara vrednost SP postane nova vrednost kazalca na klicni zapis (*angl. frame pointer*). V programskih jezikih, kjer je velikost klicnega zapisa za posamezno funkcijo konstantna, je vrednost kazalca na klicni (FP) zapis vedno izračunljiva, zato si je ni potrebno posebej shranjevati na sklad. $FP = SP + \text{velikost sklada}$. [2]

Prenos parametrov

Standarden način klicanja funkcij je, da klicoča funkcija rezervira prostor na skladu za prenosih njenih izhodnih parametrov (oz. vhodnih parametrov za klicano funkcijo). [2]

Prostor za njih se običajno nahaja pred SP. Klicana funkcija tako naslov njenega i -tega parametra izračuna z enačbo

```
argumentAddress(i) = FP + (4 * i)
```

Lokalne spremenljivke

Prevajalnik mora na skladu alocirati dovolj prostora za vse lokalne spremenljivke. todo

Static link

V jeziki, ki podpirajo gnezdenje funkcij, lahko gnezdene funkcije dostopajo do spremenljivk, ki se nahajajo v zunanjih funkcijah. Da lahko gnezdena funkcija dostopa do spremenljivk, ki niso na njenem klicnem zapisu, ji ob klicu poleg ostalih parametrov posredujemo FP funkcije, ki jo neposredno definira. Temu kazalcu rečemo *static link*.

```
func f() {
    var x: Int = 10
    func e() {
        var z: Int = 100
    }
    func g() {
        var y: Int = 20
        func h() {
            print(y, x)
        }
    }
}
```

Listing 2.5: Primer gnezdenih funkcij

Primer 2.5 vsebuje dve gnezdeni funkciji *g* in *h*. Funkcija *h* lahko dostopa do spremenljivk definiranih v *g* in *f*, ne pa tudi tistih v *e*.

2.2.5 Generiranje vmesne kode

Vmesna koda

Vmesna koda (*angl. intermediate representation*) je abstraktna predstavitev strojnega jezika in predstavlja ukaze, brez da bi poznali arhitekturo ciljne naprave. Poleg tega je vmesna koda neodvisna od izvirnega jezika. [2]

Dobra predstavitev vmesne kode ima naslednje lastnosti:

1. Njegovo generiranje mora biti priročno
2. Generiranje strojne kode v dejansko strojno kodo mora biti priročno za vse ciljne arhitekture

3. Vsak gradnik mora imeti jasen pomen, da so lahko optimizacijske transformacije enostavno implementirane

[2]

Posamezni deli abstraktnega sintaksnega drevesa so lahko kompleksne stvari, na primer zanke, klici funkcij, itd., ki jih ne moremo neposredno mapirati v strojne ukaze. Zato morajo gradniki vmesne kode predstavljati le enostavne operacije, kot so npr. LOAD (preberi vrednost iz pomnilnika), STORE (shrani vrednost v pomnilnik), ADD (seštej dve vrednosti), itd. Tako lahko vsak posamezen del AST prevedemo v ravno pravo zaporedje ukazov abstraktne vmesne kode. [2]

Poglavje 3

Programski jezik PINS

Programski jezik PINS je učni programski jezik, zato je tudi dokaj preprost. Prevajalnik zanj smo implementiral v sklopu domačih nalog pri predmetu Prevajalniki in navidezni stroji.

3.1 Leksikalna pravila

Programski jezik PINS podpira tri atomarne podatkovne tipe: *integer*, *logical*, *string*, za katere so rezervirane istoimenske besede. Celoštevilske konstante so poljubno predznačeno zaporedje števk, logične konstante so ali *true* ali *false*, znakovne konstante pa so definirane kot poljubno (lahko prazno) zaporedje znakov z ASCII kodami med vključno 32 in 126, ki je obdano z enojnima navednicama (ASCII koda 39); izjema je en sam enojni narekovaj, ki je podvojen.

Imena so definirana kot poljubno zaporedje črtk, številčk in podčrtajev, ki se ne začne s številko in ni rezervirana beseda ali kakšna od prej naštetih konstant.

Belo besedilo (*angl. whitespace*) so presledki (ASCII 32), tabulatorji (ASCII 9) in znaka za konec vrstice (ASCII 10 in 13).

Komentarji se začnejo z '#' (ASCII 35) in se raztezajo do konca vrstice.

3.2 Sintaksna pravila

Celotna izvorna koda je sestavljena iz seznama definicij. Vsaka definicija je lahko:

1. Definicija tipa (oz. sklic na tip - *typealias*)
2. Definicija spremenljivke

3. Definicija funkcije

Kot sem že omenil, podpira PINS tri osnovne podatkovne tipe, vendar sintaksa omogoča tudi definicijo tabel (*angl. array*) s fiksno velikostjo.

Definicija funkcije je sestavljena iz imena, seznama parametrov, tipa ki ga funkcija vrača, ter *izraza* oz. jedra funkcije. Zanimivo pri PINSu je to, da izven jedra funkcij ne moremo početi ničesar drugega, kot ustvarjati definicije (podobno kot pri Javi).

Izraz (*angl. expression*) je lahko:

1. Logični izraz
2. Primerjalni izraz
3. Seštevalni izraz
4. Multiplikativni izraz
5. Seštevalni izraz
6. Prefiksni izraz
7. Postfiksni izraz
8. Atomarni izraz

Atomarni izraz se še naprej deli in je lahko:

1. Logična konstanta
2. Celoštevilska konstanta
3. Znakovna konstanta
4. Ime
5. Klic funkcije
6. If stavek
7. If else stavek
8. While stavek
9. Zaporedje izrazov

Vsakemu izrazu lahko sledijo tudi definicije gnezdene znotraj zavitih oklepajev.

3.3 Semantična pravila

Območja vidnosti

Imena so vidna v celotnem območju vidnosti, ne glede na mesto definicije. Izraz *expression* { *WHERE definitions* } ustvari novo vgnezdено območje vidnosti. To pomeni, da definicije znotraj zavrtih oklepajev niso vidne navzven. Tudi definicija funkcije ustvari novo vgnezdено območje vidnosti, ki se začne za imenom in se razteza do konca funkcije..

Tipiziranost

1. *integer*, *logical*, *string* opisujejo podatkovne tipe INTEGER, LOGICAL in STRING, zaporedoma
2. izraz

`arr [n] type`

opisuje podatkovni tip ARR(*n*, *type*), kjer je *n* celoštevilska konstanta

Deklaracije

1. Deklaracija tipa

typ identifier : type

ustvari sklic na podatkovni tip *type* z imenom *identifier*

2. Deklaracija funkcije

fun identifier(identifier₁ : type₁, ..., identifier_n : type_n) : type = expression

določa funkcijo, ki je tipa

*type₁ *, ..., * type_n → type*

3. Deklaracija spremenljivke

var identifier : type

določa spremenljivko tipa *type*

4. Deklaracija parametra

identifier : type

določa parameter tipa *type*

Poglavje 4

Programski jezik Atheris

Kljub temu, da prevajalnik za Atheris izhaja iz prevajalnika za PINS, sta si jezika med seboj zelo različna. Razlikujeta se predvsem v sintaksi, ki je v programskem jeziku Atheris zelo podobna tisti, ki jo ima programski jezik Swift.



Slika 4.1: Atheris Squamigera. [<http://rebrn.com/>]

Med drugim nova sintaksa omogoča definiranje kompleksnejših podatkovnih tipov z uporabo razredov, vmesnikov, terk in enumeracij.

Posamezni stavki so med seboj ločeni z novimi vrsticami (Python, Swift), vendar pa prevajalnik omogoča njihovo ločevanje tudi s podpičjem ';' (C++, Java).

Komentarji se lahko začnejo bodisi z '#' in se raztezajo do konca vrstice (Python, PINS), bodisi z '/' in končajo z '*/' (Java, C++).

Podprt je tudi *switch* kontrolni stavek s sintakso identične tej od Swifta.

Definicije funkcij ter klici funkcij so spremenjeni; parametri funkcije so sedaj sestavljeni iz labela ter imena parametra (tako kot v Swiftu). Pri klicu funkcije se uprabi labela parametra, znotraj jedra funkcije pa se uporablja ime.

4.0.1 Podatkovni tipi

Podprti atomarni podatkovni tipi so: *integer*, *double*, *string*, *char* in *bool*. Poleg atomarnih tipov pa so podprti tudi sestavljeni podatkovni tipi:

1. Razredi
2. Enumeracije
3. Terke

Poleg naštetih podatkovnih tipov obstaja še podatkovni tip, ki ga ni mogoče eksplicitno uporabljati. To je kazalec (*angl. pointer*). Ena izmed prednosti kazalcev je v tem, da lahko do kompleksnejših objektov, ki v pomnilniku zasedejo veliko prostora, dostopamo preko njihovega naslova s kazalcem, namesto kopiranja celotne vsebine (na primer pri pošiljanju objekta kot argument funkciji). Spremeljivke, ki jim priredimo objekt, terko ali tabelo, se implicitno smatrajo kot kazalci.

4.1 Sintaksa

Celoten program v jeziku PINS je sestavljen iz seznama definicij, definicija pa je lahko definicija spremenljivke, definicija tipa ali definicija funkcije. Sintaksa programskega jezika Atheris se razlikuje v tem, da je program sestavljen iz zaporedja stavkov, stavek pa je lahko *izraz* ali *definicija*. Definicije so predstavljene z naslednjimi kontekstno neodvisnimi gramatikami:

1. Spremenljivke

```
var_def -> visibility var identifier
var_def -> visibility var identifier = expression
var_def -> visibility var identifier: type
var_def -> visibility var identifier: type =
        expression
```

```
visibility -> public
visibility -> private
visibility -> $
```

2. Funkcije

```
func_def -> func identifier ( parameters ) {
    statements }
func_def -> func identifier ( parameters )
type { statements }

parameters -> $
parameters -> paramater
parameters -> paramaters, paramater

parameter -> identifier : type
parameter -> identifier identifier : type
```

3. Enumeracije

```
enum_def -> enum { enum_defs }

enum_defs -> $
enum_defs -> enum_member_def
enum_defs -> enum_defs, enum_member_def

enum_member_def -> case identifier
enum_member_def -> case identifier = expression (
    literal)
enum_member_def -> enum_member_def,
    enum_member_def // todo
```

4. Terke

```
todo
```

5. Razredi

```
class_def -> class { definitions }
```

Sintaksa izrazov je skoraj identična PINSu, razen izrazov za nadzor toka (angl. *control flow*):

1. If stavek

```
if_expression { statements } if_expression '  
if_expression' -> $  
if_expression' -> else if { statements }  
    if_expression '  
if_expression' -> else { statements }
```

2. Switch stavek

```
switch_expression { cases }  
  
cases -> case expressions : statements cases '  
cases' -> $  
cases' -> cases  
cases' -> default : statements
```

3. While stavek

```
while_expression -> while expression { statements  
    }
```

4. For stavek

```
for_expression -> for identifier in expression  
{ statements }
```

Kot sem že omenil, omogoča sintaksa programskega jezika Atheris, da so posamezni stavki ločeni bodisi s ';' bodisi z novo vrstico. Za ta namen je dodan nov tip žetona, ki predstavlja novo vrstico v izvorni kodi (pred tem se je znak za novo vrstico štel kot belo besedilo). V primeru, da je v izvorni kodi več zaporednih novih vrstic, jih leksikalni analizator *požre* in vedno vrne samo en zaporedni *newline* simbol. To naredi sintaksno analizo enostavnejšo.

4.2 Funkcije

Programski jezik Atheris zahteva, da ob klicu funkcij poleg imena funkcije navedemo tudi imena parametrov. Podobno kot to zahteva tudi Swift.

Ob deklaraciji funkcije lahko opcijsko definiramo poljubno število poimenovanih ter tipiziranih vrednosti, ki jim skupaj rečemo parametri funkcije. Poleg tega lahko opcijsko navedemo še podatkovni tip vrednosti, ki jo bo funkcija vračala (privzeto funkcije ne vračajo ničesar - *Void*.)

Imena parametrov so sestavljena iz dveh delov: labela argumenta ter imena parametra. Labela argumentov se uporabljajo ob klicu funkcije, medtem ko se imena parametrov uporabljajo znotraj jedra funkcije. Labela so privzeto identične imenom.

```
func someFunction(firstParameterName: Int,
    secondParameterName: Int) {
    /* znotraj telesa funkcije, sta firstParameterName
       in secondParameterName referenci na vrednosti
       argumentov za prvi in drugi parameter. */
}
someFunction(firstParameterName: 1,
    secondParameterName: 2)
```

Listing 4.1: Primer definicije in klica funkcije

4.2.1 Določanje label argumentov

Če želimo, da sta labela in ime različna, navedemo labelo argumenta pred imenom:

```
func someFunction(argumentLabel parameterName: Int) {
    /* znotraj telesa funkcije, se parameterName sklicuje
       na vrednost argumenta funkciji */
}
someFunction(argumentLabel: 1)
```

Listing 4.2: Labela argumenta in ime parametra se razlikujeta

Podprta je tudi možnost, da se izognemo navajanju imen argumentov pri klicanju funkcij. To storimo tako, da označimo labelo argumenta z `_`.

```
func someFunction(_ firstParameterName: Int, _
    secondParameterName: Int) {
    ///
}
someFunction(1, 2)
```

V ozadju so opisane funkcionalnosti implementirane tako, da se v simbolno tabelo ne shrani samo *ime* funkcije, ampak se celotna definicija pretvori v posebno znakovno predstavitev, ki je sestavljena iz imena ter label argumentov. Funkcije z istim imenom, a različnimi imeni parametrov, lahko tako shranimo v simbolno tabelo brez težav. Znakovne predstavitve zgornjih funkcij izgledajo tako:

```
someFunction(firstParameterName:secondParameterName)
someFunction(argumentLabel)
someFunction(_:_)
```

Listing 4.3: Znakovne predstavitve funkcij

T.i. *overloadanje* funkcij sedaj ni problem, saj čeprav imajo funkcije ista imena, vsako izmed njih predstavimo z drugim imenom, zato jih lahko v simbolni tabeli ustrezno poiščemo. Seveda je potrebno v podobni obliki predstaviti tudi klice funkcij.

4.3 Enumeracije

V primeru, da enumeracija nima surovih vrednosti, je semantična analiza dokaj preprosta. Kompleksnejša postane kadar so surove vrednosti prisotne, saj je potrebno zagotoviti, da so ustreznega podatkovnega tipa. V primeru, da surove vrednosti niso eksplicitno navedene, jih prevajalnik skuša določiti sam. To lahko stori samo v primeru, če so surove vrednosti tipa **Int** ali **String**.

Privzeto so surove vrednosti za **Int** zaporedna števila začenši z 0. V primeru, da je vrednost eksplicitno navedena, je vrednost naslednje surove vrednosti naslednje zaporedno število od prejšnje vrednosti.

```
enum Days: Int {
    case Monday // vrednost = 0
    case Tuesday = 10 // vrednost = 10
    case Wednesday // vrednost = 11
    case Thursday = 100 // vrednost = 100
    case Friday // vrednost = 101
}
```

Listing 4.4: Enumeracija s surovimi vrednostmi tipa Int

Za **String** so privzete surove vrednosti imena članov enumeracije.

```
enum Fruit: String {
    case Apple // privzeta vrednost = "Apple"
```

```
        case Orange = "Annoying Orange"
        case Strawberry // privzeta vrednost = "
Strawberry"
    }
```

Listing 4.5: Enumeracija s surovimi vrednostmi tipa String

Do posameznih elementov enumeracije dostopamo z *DOT* operatorjem (*.*), kjer je na levi strani ime enumeracije, na desni pa ime elementa. Podobno dostopamo tudi do surovih vrednosti z uporabo imena *rawValue*. Operator *DOT* bom podrobneje opisal na razdelku razredov, za zdaj lahko povem da je njegova naloga to, da zagotovi, da v sestavljenem podaktovnem tipu na levi strani obstaja član z imenom na desni strani.

```
print(Fruit.Apple.rawValue) // 'Apple'
print(Fruit.Orange.rawValue) // 'Annoying
Orange'
print(Fruit.Strawberry.rawValue) // '
Strawberry'
```

Listing 4.6: Primer dostopa do elementov enumeracije 4.5

Člane enumeracije lahko prirejamo spremenljivkam ter nad njimi izvajamo logični operaciji primerjanja vrednosti.

```
let x: Languages = Languages.Cpp
if x == Languages.Java {
    print("Java")
}
else {
    print("Some other language")
}
```

Da lahko realiziramo *runtime* operacije nad člani enumeracije, prevajalnik vsak član enumeracije nadomesti z njegovim indeksom v definiciji enumeracije. Za zgornji primer bo tako spremenljivka *x* vsebovala vrednost 0. V *if* stavku primerjamo vrednosti 0 in 1, iz česar sledi, da se bo izvedel *else* blok.

4.4 Terke

Terka je podatkovna struktura sestavljena iz poljubnega števila elementov, ki so lahko različnih podatkovnih tipov. Terke poznajo jeziki, kot sta Python in Swift, medtem ko jih Java in C++ ne poznata.

Terka je definirana kot zaporedje izrazov znotraj oklepajev. Do posameznih elementov terke dostopamo podobno kot dostopamo do elementov v razredu, s to razliko, da je ime elementa njegov indeks v terki. Vsakemu izrazu znotraj terke lahko opcijsko določimo tudi ime. V primeru, da ime ni eksplicitno določeno, se za ime uporabi indeks izraza v terki (začenši z 0).

Do elementov prav tako dostopamo z uporabo DOT operatorja, ki zagotovi, da terka vsebuje izraz z željenim imenom.

V pomnilniku so terke predstavljene skoraj identično kot tabele, s to razliko, da se podatkovni tipi elementov med sabo lahko razlikujejo. Ravno zaradi tega ne moremo odmika izračunati na enak način kot pri tabelah. Pri terkah odmik posameznega elementa od začetnega naslova terke izračunamo tako, da seštejemo velikosti podatkovnih tipov vseh elementov pred željenim elementom:

$$offset(element_i) = sum(size(element_0), size(element_1), \dots, size(element_{n-1}))$$

4.5 Razredi

Razred je sestavljena podatkovna struktura, ki, podobno kot terka, v sebi hrani spremenljivke različnih podatkovnih tipov. Vsak razred lahko vsebuje poljubno število atributov (spremenljivk) ter poljubno število metod - to so člani razreda.

Atributi in metode so lahko statični, kar pomeni, da niso del posamezne instance razreda, ampak živijo v statični instanci, ki je kreirana avtomatsko. Poleg tega so lahko člani razreda *privatni* ali *javni*. Razlika je v tem, da do privatnih članov ni možno dostopati izven razreda.

Metodam lahko dodamo tudi modifikator *final*, kar prepreči, da bi bila funkcija re-implementirana v dedujočem se razredu. Če želimo, da metoda re-implementira metodo v starševkem razredu, ji moramo dodati modifikator *overriding*.

Poleg atributov in metod lahko razred vsebuje tudi definicije gnezdenih razredov, enumeracij ter vmesnikov.

V pomnilniku je instanca razreda (objekt) predstavljena podobno kot terka, t.j. kot tabela z vrednostmi različnih podatkovnih tipov. Zato tudi odmik elementov izračunamo podobno kot pri terkah.

4.5.1 Dostop do članov in nadzor dostopa

Do članov razreda dostopamo preko DOT operatorja. DOT operator je binarni operator in je sestavljen iz dveh izrazov, ki ju povezuje pika '.'.

Običajno se izvaja povezovanje uporab spremenljivk z definicijami v fazi razreševanja imen, vendar to pri sestavljenih podatkovnih strukturah (vključno z enumeracijami in terkami) ni mogoče.

Zato se s tem ne ukvarjamo v razreševanju imen, ampak problem prepustimo razreševanju tipov. Razreševanje tipov zato delimo na dve pod-fazi (dva sprehoda po AST). V prvem sprehodu nas zanimajo samo podatkovni tipi definicij, izraze zaenkrat pustimo pri miru. Ko zberemo vse potrebne informacije o podatkovnih tipih članov razreda, lahko zgradimo razredni tip, ki predstavlja razred in njegove definicije.

Sedaj imamo dovolj informacij, da razrešimo imena znotraj DOT operatorja, kar storimo v drugem sprehodu.

Nadzor dostopa

Nadzor dostopa omejuje dostop do članov razreda izven definicije, če to eksplicitno navedemo z uporabo modifikatorja *private*. Privzeto so vsi člani *public*. Programski jezik Atheris trenutno pozna samo public in private modifikatorja; public modifikator omogoča dostop do in spreminjanje vrednosti člana razreda kodi, ki se ne nahaja v razredu, za razliko od modifikatorja private, ko to prepoveduje.

4.5.2 Instančne funkcije

Razred ima lahko v sebi definirane funkcije, ki se jim v OOP žargonu pogosto reče *metode*. Metode se od funkcij, ki niso definirane v razredu, razlikuje v tem, da vsebujejo impliciten parameter, ki je referenca na objekt, ki metodo kliče. Ta parameter prevajalnik v metodo vstavi avtomatsko.

V C++ in Javi se ta parameter imenuje *this*, v Pythonu, Swiftu in v Atherisu pa *self*. V Pythonu je ta razlika, da je parameter potrebno eksplicitno navesti, če ga ne, se funkcija / metoda smatra kot statična (več o statičnih metodah v nadaljevanju).

Prevajalnik mora implicitni *self* parameter avtomatsko vstaviti v vse metode razreda. Delno to stori tekom razreševanja imen, ko definiciji funkcije vstavi nov parameter, delno pa tekom razreševanja tipov, ko parametru nastavi tip razreda.

4.5.3 Konstruktorji

Konstruktorji so posebne vrste metod, njihova naloga pa je kreacija objektov ter inicializacija atributov. Podobno kot ostalim metodam razreda tudi konstruktorju prevajalnik implicitno vstavi *self* parameter.

Naloga konstruktorja je, da na kopici, t.j. delu pomnilnika, kamor shranjujemo dinamično alicirane objekte (za razliko od sklada, kamor shranjujemo statične), alocira prostor, kjer bo shranjen objekt. Za dinamično alociranje pomnilnika se uporabi gradnik vmesne kode **ImcMALLOC**. Ta gradnik na kopici rezervira željeno velikost pomnilnika ter vrne naslov.

Pri generiranju vmesne kode prevajalnik vstavi na začetek kode funkcije MALLOC gradnik, ki bo rezerviral prostor za objekt. Nato bo prevajalnik vrnjen naslov shranil na sklad v parameter *self*, zato da lahko znotraj konstruktorja nastavimo vrednosti ostalih atributov razreda. Na koncu bo ta naslov konstruktor tudi vrnil. Vrnjen naslov je referenca na objekt in ga uporabljamo za nadaljne operacije nad objektom.

Razred ima lahko poljubno število konstruktorjev, ki jih definiramo z uporabo ključne besede *init*.

```
class Atheris {  
    var x: Int  
    init() {  
        self.x = 0  
    }  
    init(x: Int) {  
        self.x = x  
    }  
}
```

Listing 4.7: Konstruktorji

Poleg eksplicitnih konstruktorjev prevajalnik zgenerira tudi impliciten *privzeti* konstruktor. Vanj je vstavljena koda za inicializacijo atributov, ki so inicializirani ob definiciji.

```
class Atheris {  
    var x: Int = 10  
    var y: Int = 20  
}  
  
print(Atheris().x) // 10  
print(Atheris().y) // 20
```

Listing 4.8: Implicitni privzeti konstruktor

V primeru, da je v razredu privzeti konstruktor tudi eksplicitno definiran, bo poleg svoje kode vseboval tudi kodo implicitnega.

```
class Atheris {  
    var x: Int = 10  
    var y: Int = 20  
    init() { // privzeti konstruktor  
        y = 100  
    }  
}  
  
print(Atheris().x) // 10  
print(Atheris().y) // 100
```

Listing 4.9: Eksplicitni privzeti konstruktor

4.5.4 Statični člani razreda

Statični člani razreda so tisti člani, ki ne živijo znotraj posameznih instanc, ampak živijo v globalni statični instanci razreda. Definiramo jih z uporabo modifikatorja **static**. Statična instanca razreda je naložena v pomnilnik avtomatsko ob zagonu programa. Do statičnih članov dostopamo preko imena razreda. Ker statične funkcije niso del instance razreda, ne vsebujejo *self* parametra.

```
class Static {
    static func foo() {
        print("I am a static function")
    }
}
Static.foo()
```

Listing 4.10: Klicanje statične funkcije

V času razreševanja tipov so statične definicije razreda shranjene v ločeno podatkovno strukturo od instančnih. Vozlišču AST, ki definira razred (*AstClassDefinition*), v simbolno tabelo ne pripišemo razredni podatkovni tip, ampak *statični / kanonični* tip, ki hrani podatke o statičnih definicijah razreda ter referenco na dejanski instanci podatkovni tip. Ko se sklicujemo na člane statične instance preko DOT operatorja, prevajalnik preveri, ali statični tip vsebuje iskan član.

Razredi lahko vsebujejo tudi enumeracije in gnezdene razrede. Definicije enumeracij in razredov se smatrajo privzeto kot statične, zato do njih dostopamo enako kot do statičnih članov.

```
class Static {
    enum Languages: Int {
        case Cpp, Java
    }
}
let x = Static.Languages.Java
print(x.rawValue) // '1'
```

Listing 4.11: Enumeracija znotraj razreda

Da lahko dostopamo do konstruktorjev znotraj gnezdenih razredov, jih prevajalnik smatra, kot da so to statične metode.

```
class Atheris {
    class NestedClass {
        var x = 20
    }
}
```

```
    }  
}  
  
print(Atheris.NestedClass().x) // '20'
```

Listing 4.12: Gnezden razred

4.6 Dedovanje in polimorfizem

Dedovanje in polimorfizem sta, poleg enkapsulacije in abstrakcije, najpomembnejša koncepta OOP programiranja. Dedovanje je mehanizem, s katerim objekt pridobi (podeduje) nakatere ali vse lastnosti drugega objekta, medtem ko je polimorfizem sposobnost predstaviti isto funkcionalnost z različnimi implementacijami. Programski jezik Atheris omogoča enkratno dedovanje, tako kot Java in Swift.

V pomnilniku se atributi dedovanega razreda nahajajo pred atributi dedujočega se razreda. Enačba za izračun velikosti razreda (koliko prostora porabi objekt v pomnilniku) postane kompleksnejša, saj je potrebno rekurzivno prišteti velikosti nadrazradov. Prav tako postane kompleksnejši izračun odmkov atributov od začetnega naslova.

4.6.1 Dynamic Dispatch

Dynamic dispatch je proces izbire, katera implementacija polimorfne operacije naj se izvede v času **izvajanja programa**. Običajno lahko pri klicih funkcij že v času prevajanja izberemo, katera *implementacija* funkcije naj se kliče, saj obstaja samo ena. Pri razredih nastane problem, kadar dovolimo, da imajo lahko dedujoči se razredi svoje implementacije funkcij (oz. metod), kot je razvidno v spodnjem primeru:

```
class Fruit {  
    func kind() {  
        print("I am a Fruit!")  
    }  
}  
  
class Apple: Fruit {  
    override func kind() {  
        print("I am an Apple!")  
    }  
}  
  
class Orange: Fruit {
```

```

        override func kind() {
            print("I am an Orange!")
        }
    }
}

```

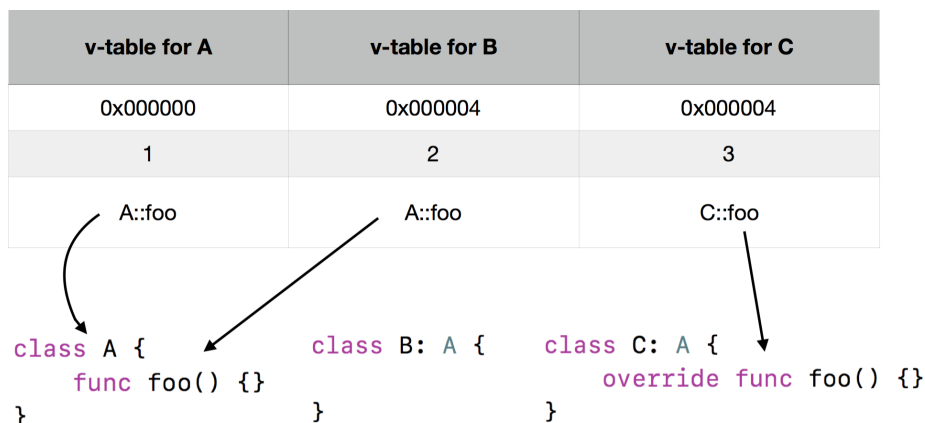
Listing 4.13: Več implementacij za isto funkcionalnost

Problem je možno rešiti na več načinov. Rešitev prevajalnika za Atheris temelji na rešitvi, ki jo ima tipično C++, t.j. z uporabo podatkovne strukture *v-table*.

Virtualna tabela

Virtualna tabela je podatkovna struktura, ki vsebuje kazalce na funkcije implementirane v danem razredu. Prevajalnik za vsak razred zgenerira in shrani v pomnilnik unikatno virtualno tabelo.

Vsak objekt v pomnilniku, poleg svojih atributov, hrani še kazalec na lastno virtualno tabelo.



Slika 4.2: Virtualne tabele za dane razrede. (todo popravi shemo da uporabiš razrede iz zgornjega primera)

Virtualna tabela vsebuje kazalce na naslove, kjer se nahajajo vse ne-statične metode, ki so implementirane v razredu. Vsaka tabela vsebuje še kazalec na virtualno tabelo nadrazreda (ali *null*, če ga nima), ter unikatni identifikator tipa. Identifikator je pozitivno celo število, ki ga prevajalnik dodeli vsakemu tipu in je unikatno za vsak tip.

4.6.2 Realizacija

Problem rešimo tako, da kazalec na metodo shranimo v virtualno tabelo na tisti indeks, na katerem je metoda definirana v razredu. Prevajalnik za klic funkcije zgenerira vmesno kodo, ki izračuna naslov metode z naslednjo enačbo:

$$\text{address}(\text{method}_i) = \text{address}(\text{vtable}) + (i * \text{sizeof}(\text{void*}) + 8)$$

kjer je

$$\text{address}(\text{vtable})$$

lokacija tabele v pomnilniku,

$$(i * \text{sizeof}(\text{void*}) + 8)$$

indeks pomnožen z velikostjo kazalca (običajno 4 byti) plus odmik od začetka tabele zaradi prej opisanih dodatnih podatkov v tabeli in

$$\text{address}(\text{method}_i)$$

naslov metode na indeksu i .

4.7 Instance of operator

Instance of operator je operator, ki v času izvajanja programa preveri, ali je objekt instance danega razreda. To naredi z uporabo virtualne tabele. Kot sem omenil, vsaka tabela vsebuje unikatni deskriptor njenega razreda. Preko deskriptorja lahko preverimo, ali se dva podatkovna tipa ujemata.

V programskem jeziku Atheris se operator imenuje *is*.

```
var object: A
object = B()
print(object is A) // 'true'
print(object is B) // 'true'
print(object is C) // 'false'
```

Listing 4.14: Uporaba operatorja *is* za razrede iz sheme 4.2

Dejanski algoritem za preverjanje, ali je objekt instance danega razreda, je kompleksnejši kot sem opisal zgoraj, saj je lahko razredna hierarhija obsežna in je zato potrebno preverjanje izvesti rekurzivno po celotni hierarhiji.

todo

Listing 4.15: Algoritem za izračun ali je objekt instance danega razreda

4.8 Pretvarjanje tipov

Mehanizem pretvarjanja iz enega tipa v drug (angl. *type casting*) deluje po zelo podobnem principu kot *as* operator. Razlikujeta se predvsem v njunem rezultatu. Rezultat *as* operatorja je *true* ali *false*, medtem ko je pri neuspešnem pretvarjanju rezultat *null*, pri uspešnem pa objekt, pretvorjen v željen tip. Takemu tipu pretvarjanja rečemo *safe casting*, saj program nadaljuje z izvajanjem tudi če pretvorba ni uspešna. Za razliko od *non-safe castinga*, kjer bi program vrnil *null pointer exception*.

```
var object: A = B()
let casted = object as B
    print("cast failed")
}
else {
    print("cast succeeded") // izvede se 'else
' blok
}
```

Listing 4.16: Pretvorba za razrede iz sheme 4.2

4.9 Vmesniki

Vmesnik je abstraken opis akcij, ki jih lahko izvede objekt. Tekom prevajanja prevajalnik zagotovi, da razred vsebuje vse metode v vmesnikih, ki jih želi implementirati. Ker je vmesnik abstraktna podatkovna struktura, ne moremo kreirati instance neposredno, lahko pa spremenljivki priredimo instanco razreda, če razred implementira vmesnik. Vse metode so klicane dinamično, po postoku opisanem v razdelku 4.6.1.

```
interface I {
    func foo()
}
class A: I {
    func foo() {
        print("A's implementation of foo")
    }
}
class B: I {
    func foo() {
```



```
        print("B's implementation of foo")
    }
}
var x: I
x = A()
x.foo() // 'A's implementation of foo'
x = B()
x.foo() // 'B's implementation of foo'
```

Listing 4.17: Vmesniki

4.10 Razširitve

Razširitev (angl. *extension*) je mehanizem, ki omogoča dodajanje funkcionalnosti obstoječim razredom.

```
extension A {
    func toString() {
        print("ext::toString()")
    }
}
A().toString()
```

Listing 4.18: Razširitev razreda A

Razred lahko razširimo z metodami, enumeracijami in razredi.

4.10.1 Razširitev z vmesnikom

Razredom lahko preko razširitev dodamo tudi implementacijo vmesnikov, čemur rečemo *interface extension*.

```
class Collection {}
interface Iterable {
    func next() Any
}
extension Collection: Iterable {
    func next() Any {
        return nil
    }
}
let iterable: Iterable = Collection()
```

```
print(iterable.next())
```

Listing 4.19: Razširitev z vmesnikom

4.11 Type inference

Type inference omogoča prevajalniku, da avtomatsko prepozna podatkovni tip izraza, na podlagi vrednosti, ki so v izrazu. To je posebej uporabno pri definicijah spremenljivk, ki jim pogosto nastavimo vrednost ob sami definiciji.

V primeru, da definicija spremenljivke nima eksplicitno navedenega podatkovnega tipa, prevajalnik najprej izračuna podatkovni tip izraza na desni strani prirejevalnega operatorja, ter ga nato dodeli definiciji.

```
let meaningOfLife = 42
// meaningOfLife je tipa Int
let pi = 3.14159
// pi je tipa Double
let anotherPi = 3 + 0.14159
// anotherPi je tudi tipa Double
```

Listing 4.20: Avtomatično prepoznavanje podatkovnih tipov

Prepoznavanje tipov postane kompleksnejše pri tabelah, kadar njeni elementi niso istih tipov. Prevajalnik tekom razreševanja tipov vedno poišče najmanjši možen tip, ki je skupen vsem elementom tabele. V primeru, da so v tabeli objekti, ki jim je skupen nadrazred *A*, bo tudi tabela tipa *A*. V primeru, da elementi nimajo nobenega skupnega tipa, prevajalnik tabeli dodeli poseben abstrakten tip ***Any***, ki je definiran v standardni knjižnici. *Any* je v programskem jeziku Atheris to, kar je v Javi *Object*, s to razliko, da je *Object* razred, *Any* pa vmesnik. Vsak podatkovni tip je implicitno tudi tipa *Any*.

```
let array = [C(), B(), C(), C()] //
spremenljivka array = tipa ARR ( A )
```

Listing 4.21: Elementi tabele s skupnim nadrazredom

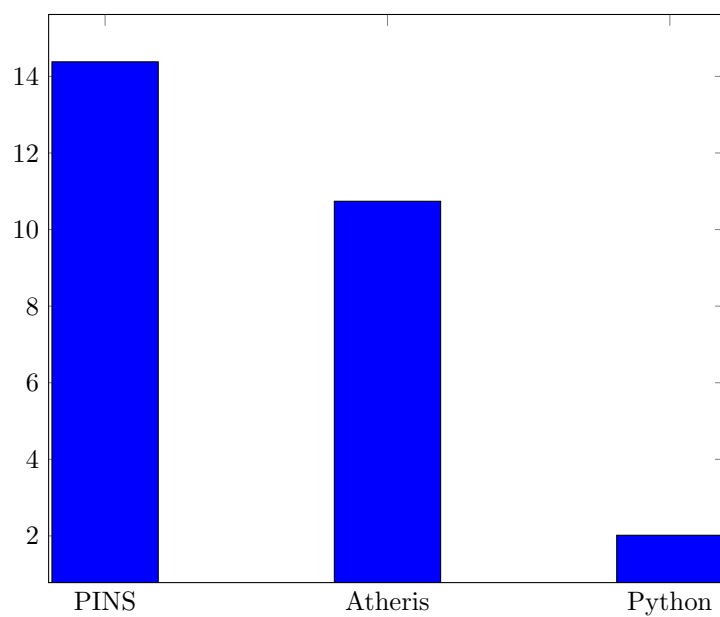
```
let array = [10, "string", A(), 3.14] //
spremenljivka array = tipa ARR ( Any )
```

Listing 4.22: Elementi tabele nimajo skupnega tipa

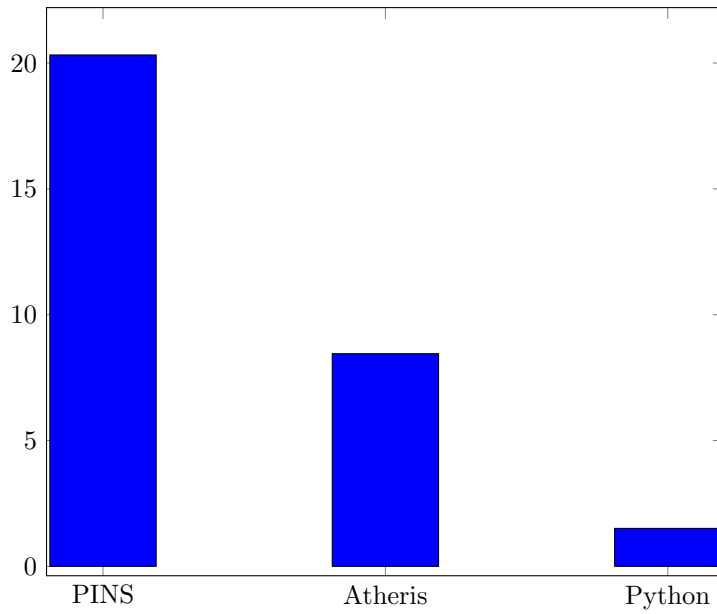
Poglavje 5

Meritve in testiranje

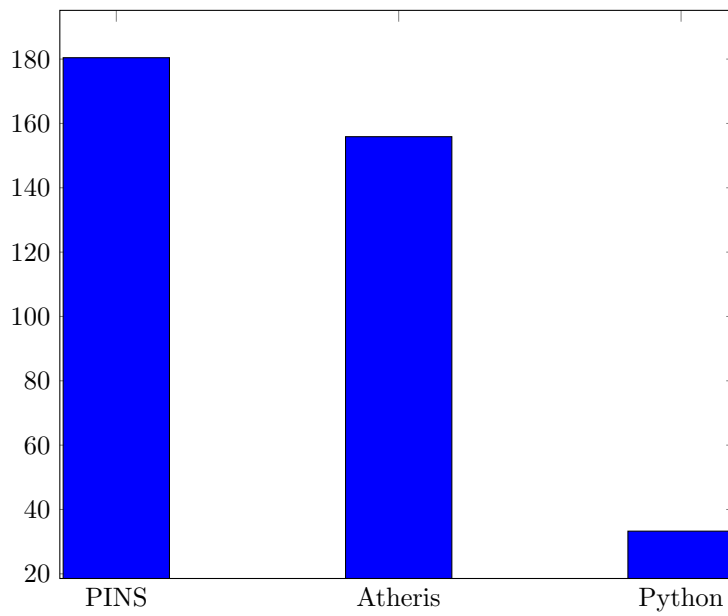
5.1 Bubble sort



5.2 Quick sort



5.3 Rekurzivni fibonacci



5.4 Quicksort nad razredi

37821ms

Poglavje 6

Sklepne ugotovitve

Uporaba \LaTeX in \BibTeX je v okviru Diplomskega seminarja **obvezna**! Iz-bira \LaTeX ali ne \LaTeX pri pisanju dejanske diplomske naloge pa je prepuščena dogovoru med vami in vašim mentorjem.

Res je, da so prvi koraki v \LaTeX u težavni. Ta dokument naj vam služi kot začetna opora pri hoji. Pri kakršnihkoli nadaljnjih vprašanjih ali napakah pa svetujem uporabo Googla, saj je spletnih strani za pomoč pri odpravljanju težav pri uporabi \LaTeX a ogromno.

Literatura

- [1] A complete description of a programming language includes the computational model, the syntax and semantics of programs, and the pragmatic considerations that shape the language. Dosegljivo: <http://www.emu.edu.tr/aelci/Courses/D-318/D-318-Files/plbook/intro.htm>. [Dostopano: 26.11.2017].
- [2] Andrew W. Appel and Jens Palsberg. *Modern Compiler Implementation in Java, Second Edition*. Cambridge University Press, 2002.