# Interpreter języka W++

# Piotr Gdowski

# $1\quad \mathbf{Sk}\mathbf{\hat{l}}\mathbf{adnia}\ \mathbf{j}\mathbf{\hat{e}}\mathbf{z}\mathbf{y}\mathbf{k}\mathbf{a}\ \mathbf{W}++$

		$Skladnia\ abstrakcyjna$	$Skladnia\ konkretna$	Opis
p	::=	$\operatorname{program}(dt^*, df^*, dv^*, c)$	$dt^* df^* $ vars $dv^* $ in $c$	deklaracja typów, funkcji oraz zmiennych
dt	::=	$\operatorname{newtype}(t, \tau)$	Type $t = \tau$ ;	deklaracja typu
df	::=	function $(f, (x : \tau)^*, dv^*, c, e)$	function $((x : \tau)^*)$ = vars $dv^*$ in $c$ return e;	deklaracja funkcji
dv	::=	newvar(x, e)	x := e	deklaracja zmiennej
		newpointer $(x, e)$	$x:=new\ e$	deklaracja wskaźnika
		$typednewvar(\tau, x, e)$	$\tau \ x := e$	deklaracja zmiennej z własną adno-
		, , ,		tacją typową
		$typednewpointer(\tau, x, e)$	$ au \ x := new \ e$	deklaracja wskaźnika z własną adno-
				tacją typową
c	::=	skip	skip	pusta instrukcja
		$compose(c_1, c_2)$	c; c	złożenie instrukcji
		newvars $(dv^*, c)$	vars $dv^*$ in $c$	deklaracja lokalnych zmiennych i wskaźników
		$if(e, c_1, c_2)$	if $e$ then $c_1$ else $c_2$	instrukcja if
		while(e, c)	while $e$ do $c$	instrukcja while
		varassign(x, e)	x := e	przypisanie wartości do zmiennej
		pointerassign $(x, e)$	*x := e	przypisanie wartości do wartości pod wskaźnikiem
		$\operatorname{callassign}(x, f, \vec{x})$	$x:=\mathrm{f}(ec{x})$	przypisanie wartości zwróconej przez
				funkcję
e	::=	$\underline{n}$	n	liczba
		x	x	wartość zmiennej
		$*_x$	*x	wartość zmiennej wskazywanej przez
				wskaźnik
		$plus(e_1, e_2)$	e + e	dodawanie
		$\operatorname{mult}(e_1,e_2)$	e * e	mnożenie
		$\min(e)$	- <i>e</i>	minus
		$\operatorname{tuple}(\vec{e})$	$\operatorname{tuple}(e_1, \ldots, e_n)$	krotka
		$\operatorname{indexer}(e, \underline{n})$	e[n]	wyjęcie z krotki wskazanej wartości
		injection(l, e)	l.e	włożenie
		$case(e, (l.x \to e)^+)$	case $e$ { $l_1.x \rightarrow e_1,, l_n.x \rightarrow e_n,$ }	case
au	::=	int	int	typ liczb całkowitych
		$\mathrm{TTuple}(ec{ au})$	$Tuple(\tau_1,, \tau_n)$	typ krotki
		$TSum((l \to \tau)^+)$	$Sum(l_1 \to \tau_1,, l_n \to \tau_n)$	typ sum rozłącznych
		$\mathrm{TPtr}( au)$	$\operatorname{Ptr}( au)$	typ wskaźnikowy
		TAbs(t)	t	typ abstrakcyjny

## 2 System typów języka W++

- $\bullet$   $\Gamma$  jest listą zawierającą zmienne wraz z ich typami
- F jest listą zawierającą funkcje wraz z ich definicjami i zwracanymi wyrażeniami
- $\bullet~T$ jest listą zawierającą nazwy predefiniowanych typów wraz z jego prawdziwymi typami

## 2.1 Typowanie wyrażeń

$$\begin{array}{c} \overline{\Gamma,x:\tau\vdash x:\tau} \\ \hline \overline{\Gamma\vdash\underline{n}:int} \\ \hline \Gamma\vdash\underline{e}:TPtr(\tau) \\ \hline \Gamma\vdash *e:\tau \\ \hline \Gamma\vdash e_1:int \quad \Gamma\vdash e_2:int \\ \hline \Gamma\vdash plus(e_1,e_2):int \\ \hline \Gamma\vdash plus(e_1,e_2):int \\ \hline \Gamma\vdash mult(e_1,e_2):int \\ \hline \hline \Gamma\vdash e:int \\ \hline \Gamma\vdash e:int \\ \hline \Gamma\vdash e:int \\ \hline \Gamma\vdash e:int \\ \hline \Gamma\vdash e:tint \\ \hline \Gamma\vdash e:TTuple(\tau_1,...,\tau_n) \quad \underline{i}:int \\ \hline \Gamma\vdash e:TTuple(\tau_1,...,\tau_n) \quad \underline{i}:int \\ \hline \Gamma\vdash e:t \\ \hline \Gamma\vdash e:$$

#### 2.1.1 Relacja $\leq$ na typach (podtypowanie)

Rozważmy instrukcję x := l.5. Przy pomocy powyższych reguł typowania możemy wyinferować typ wyrażenia po lewej stronie - jest to  $Sum(l \to int)$ . Oznacza to, że również x powinien być dokładnie takiego typu, aby instrukcja mogła się poprawnie otypować. Wówczas każdy typ sumy miałby dokładnie jeden konstruktor, nie byłoby bowiem możliwości 'zgadnięcia' pozostałych konstruktorów. Z tego powodu dodałem możliwość zdefiniowania w adnotacji typowej wszystkich konstruktorów. Tym sposobem, mamy możliwość przypisywania do zmiennych 'większych' typów wyrażenia 'mniejszych' typów. Intuicyjnie,  $T \vdash Sum(l \to int) \preceq Sum(l \to int|r \to int)$ , gdzie w T trzymamy predefiniowane typy. Pełna definicja relacji:

#### 2.2 Poprawność sformowania instrukcji

$$\overline{T,F,\Gamma\vdash skip\ cwf}}$$

$$\frac{T,F,\Gamma\vdash c_1\ cwf\ T,F,\Gamma\vdash c_1\ cwf}{T,F,\Gamma\vdash c\ cwf}$$

$$\overline{T,F,\Gamma\vdash e:\tau} \quad T,F,\Gamma,x:\tau\vdash newvars([],c)\ cwf$$

$$\overline{T,F,\Gamma\vdash e:\tau} \quad T,F,\Gamma,x:\tau\vdash newvars(declariations,c)\ cwf$$

$$\overline{T,F,\Gamma\vdash e:\tau} \quad T,F,\Gamma,x:Ptr(\tau)\vdash newvars(declariations,c)\ cwf$$

$$\overline{T,F,\Gamma\vdash e:\tau} \quad T,F,\Gamma,x:Ptr(\tau)\vdash newvars(declariations,c)\ cwf$$

$$\overline{T,F,\Gamma\vdash e:\tau'} \quad T\vdash \tau' \preceq \tau \quad T,F,\Gamma,x:\tau\vdash newvars(declariations,c)\ cwf$$

$$\overline{T,F,\Gamma\vdash e:\tau'} \quad T\vdash \tau' \preceq \tau \quad T,F,\Gamma,x:Ptr(\tau)\vdash newvars(declariations,c)\ cwf$$

$$\overline{T,F,\Gamma\vdash e:\tau'} \quad T\vdash \tau' \preceq \quad T,F,\Gamma,x:Ptr(\tau)\vdash newvars(declariations,c)\ cwf$$

$$\overline{T,F,\Gamma\vdash e:\tau'} \quad T\vdash \tau' \preceq \quad T,F,\Gamma,x:Ptr(\tau)\vdash newvars(declariations,c)\ cwf$$

$$\overline{T,F,\Gamma\vdash e:\tau'} \quad T\vdash \tau' \preceq \quad T,F,\Gamma\vdash c_1\ cwf$$

$$\overline{T,F,\Gamma\vdash e:\tau'} \quad T\vdash c_1\ cwf$$

$$\overline{T,F,\Gamma\vdash b:tf(e,c_1,c_2)\ cwf}$$

$$\overline{T,F,\Gamma\vdash while(e,c)\ cwf}$$

$$\overline{T,F,\Gamma\vdash while(e,c)\ cwf}$$

$$\overline{T,F,\Gamma\vdash while(e,c)\ cwf}$$

$$\overline{T,F,\Gamma\vdash while(e,c)\ cwf}$$

$$\overline{T,F,\Gamma\vdash b:\tau'} \quad T\vdash x:TPtr(\tau)$$

$$\overline{T,F,\Gamma\vdash pointerassign(x,e)\ cwf}$$

$$\Gamma\vdash x:\tau \quad \Gamma\vdash a:\tau_1\ ...\ \Gamma\vdash e_n:\tau_n$$

$$\overline{T,F,(f,x_1:\tau_1,...,x_n:\tau_n,e:\tau),\Gamma\vdash callassign(x,f,e_1,...,e_n)\ cwf}$$

### 2.3 Poprawność sformowania programu

#### 2.3.1 Poprawność sformowania typów

Na potrzeby typów definiowanych przez użytkownika, potrzebujemy nowego jugdementu sprawdzającego poprawność sformowania typu przy pewnym zbiorze typów T:  $T \vdash \tau \ twf$ . Reguły są następujące:

#### 2.3.2 Poprawność sformowania funkcji

Deklaracja funkcji  $function(f, x_1 : \tau_1, ..., x_n : \tau_n, dvs', c, e)$  zawiera nazwę funkcji, nazwy argumentów wraz z ich typami, deklarację zmiennych globalnych w tej funkcji, ciało oraz wyrażenie zwracane. Zwracane wyrażenie ma dostęp do zmiennych globalnych oraz argumentów funkcji. Z ciała funkcji można wywołać samego siebie, bądź funkcję zdefiniowaną wyżej w kodzie. Reguły poprawnego sformowania funkcji:

#### 2.3.3 Właściwe reguły

3

$$\frac{\emptyset,\emptyset,\emptyset \vdash program(dts,dfs,dvs,c) \ pwf}{program(dts,dfs,dvs,c) \ pwf}$$
 
$$\frac{T \vdash \tau \ twf \quad T,(t:\tau),F,\Gamma \vdash program(dts,dfs,dvs,c) \ pwf}{T,F,\Gamma \vdash program(newtype(t,\tau) :: dts,dfs,dvs,c) \ pwf}$$
 
$$\frac{T,F \vdash function(f,x_1:\tau_1,...,x_n:\tau_n,dvs',c,e) \ fwf \quad T,F,(f,x_1:\tau_1,...,x_n:\tau_n,e:\tau),\Gamma \vdash program([],dfs,dvs,c) \ pwf}{T,F,\Gamma \vdash program([],function(f,x_1:\tau_1,...,x_n:\tau_n,dvs',c,e) :: dfs,dvs,c) \ pwf}$$
 
$$T,F,\Gamma \vdash newvars(dvs,c) \ cwf$$

## Semantyka operacyjna języka W++

## 3.1 Definicje wartości i postaci końcowych

$$\frac{\underline{n} \ val}{EPtr(e) \ val}$$

$$\frac{e_1 \ val \ \dots \ e_n \ val}{ETuple(e_1, \dots, e_n) \ val}$$

$$\frac{e \ val}{l.e \ val}$$

 $\overline{T, F, \Gamma \vdash program([], [], dvs, c)} pwf$ 

EPtr(e) jest sztucznym wyrażeniem, służy jako wartość dla zmiennej typu wskaźnikowego

$$\frac{skip\ cfinal}{Program(skip)\ pfinal}$$

## 3.2 Semantyka małych kroków

- M jest pamięcia programu, przechowującą wartości zmiennych.
- H jest stertą, pamiętającą wartości wskazywane przez wskaźniki.
- F jest listą zdefiniowanych funkcji.

#### 3.2.1 Wyrażenia

$$\overline{H,(M,x\hookrightarrow v)\ x\longmapsto H,(M,x\hookrightarrow v)\ v}$$

$$\frac{H,M\ e_1\longmapsto H,M\ e'_1}{H,M\ e_1+e_2\longmapsto H,M\ e'_1+e_2}$$

$$\underbrace{e_1\ val\ H,M\ e_2\longmapsto H,M\ e'_2}_{H,M\ e_1+e_2\longmapsto H,M\ e'_1+e'_2}$$

$$\frac{H,M\ e_1\longmapsto H,M\ e'_1}{H,M\ e_1+e_2\longmapsto H,M\ e'_1*e_2}$$

$$\frac{H,M\ e\mapsto H,M\ e'}{H,M\ e\mapsto H,M\ e'}$$

$$\frac{H,M\ e\mapsto H,M\ e'}{H,M\ -e\longmapsto H,M\ -e'}$$

$$\underbrace{H,M\ e\mapsto H,M\ e'_1}_{H,M\ ETuple(e_1,\ldots,e_i,\ldots,e_n)\longmapsto H,M\ ETuple(e_1,\ldots,e'_i,\ldots,e_n)}_{H,M\ t\models H,M\ t'[\underline{m}]}$$

$$\underbrace{H,M\ t\mapsto H,M\ t'_1}_{H,M\ t[\underline{m}]\longmapsto H,M\ t'[\underline{m}]}$$

$$\underbrace{ETuple(e_1,\ldots,e_i,\ldots,e_n)\ val}_{H,M\ ETuple(e_1,\ldots,e_i,\ldots,e_n)}$$

$$\begin{array}{c} H, M \ e \longmapsto H, M \ e' \\ \hline H, M \ l.e \longmapsto H, M \ l.e' \\ \hline H, M \ e \mapsto H, M \ e' \\ \hline H, M \ case \ e \ of \{l_1.x \rightarrow e_1, ..., l_n.x \rightarrow e_n\} \longmapsto H, M \ case \ e' \ of \{l_1.x \rightarrow e_1, ..., l_n.x \rightarrow e_n\} \\ \hline H, M \ case \ l_i.e \ of \{l_1.x \rightarrow e_1, ..., l_i.x \rightarrow e_i, ..., l_n.x \rightarrow e_n\} \longmapsto H, M \ [e/x]e_i \\ \hline H, M \ e \longmapsto H, M \ e' \\ \hline H, M \ *e \longmapsto H, M \ *e' \\ \hline \hline (H, x \hookrightarrow v), M \ *EPtr(x) \longmapsto (H, x \hookrightarrow v), M \ v \\ \hline \end{array}$$

#### 3.3 Instrukcje

$$\begin{array}{c} H, M \ e \longmapsto H, M \ e' \\ \hline F, H, M \ x := e \ e \ \nabla F, H, M \ x := e' \\ e \ val \\ \hline E, H, (M, x \hookrightarrow_{-}) \ x := e \longmapsto_{-} F, H, (M, x \hookrightarrow_{e}) \ skip \\ \hline F, H, (M, x \hookrightarrow_{-}) \ x := e \longmapsto_{-} F, H, (M, x \hookrightarrow_{e}) \ skip \\ \hline F, H, M \ (c_{1} : c_{2}) \longmapsto_{-} F, H', M' \ (c'_{1} : c_{2}) \\ \hline F, H, M \ (c_{1} : c_{2}) \longmapsto_{-} F, H', M \ (c'_{1} : c_{2}) \\ \hline F, H, M \ (skip; c_{2}) \longmapsto_{-} F, H, M \ c_{2} \\ H, M \ e \longmapsto_{-} H, M \ e' \\ \hline F, H, M \ If ThenElse(e, c_{1}, c_{2}) \longmapsto_{-} F, H, M \ If ThenElse(e', c_{1}, c_{2}) \\ e = 0 \\ \hline F, H, M \ If ThenElse(e, c_{1}, c_{2}) \longmapsto_{-} F, H, M \ c_{1} \\ e \neq 0 \\ \hline F, H, M \ If ThenElse(e, c_{1}, c_{2}) \longmapsto_{-} F, H, M \ c_{2} \\ \hline F, H, M \ If ThenElse(e, c_{1}, c_{2}) \longmapsto_{-} F, H, M \ c_{2} \\ \hline F, H, M \ newvar(x, e, c) \longmapsto_{-} F, H, M \ newvar(x, e', c) \\ e \ val \ x \notin_{-} M \ F, H, (M, x \hookrightarrow_{-} e) \longmapsto_{-} F, H, M \ newvar(x, e, e') \ e' \\ F, H, M \ newvar(x, e, c) \longmapsto_{-} F, H, M \ newvar(x, e, e') \\ e \ val \ x \notin_{-} M \ F, H, M \ c \longmapsto_{-} F, H, M \ newvar(x, e, e') \\ \hline e \ val \ x \notin_{-} M \ F, H, M \ newvar(x, e, c) \longmapsto_{-} F, H, M \ newvar(x, e, e) \\ \hline F, H, M \ newvar(x, e, c) \longmapsto_{-} F, H, M \ newvar(x, e, e) \\ \hline E, H, M \ newpointer(x, e, c) \longmapsto_{-} F, H, M \ newpointer(x, e, e') \\ \hline e \ val \ x \notin_{-} M \ F, H, M \ newpointer(x, e, e') \\ \hline e \ val \ x \notin_{-} M \ F, H, M \ newpointer(x, e, e') \\ \hline e \ val \ x \notin_{-} M \ F, H, M \ newpointer(x, e, e') \\ \hline e \ val \ x \notin_{-} M \ F, H, M \ newpointer(x, e, e') \\ \hline e \ val \ x \notin_{-} M \ F, H, M \ newpointer(x, e, e') \\ \hline e \ val \ x \notin_{-} M \ F, H, M \ newpointer(x, e, e') \\ \hline e \ val \ x \notin_{-} M \ F, H, M \ newpointer(x, e, e') \\ \hline e \ val \ x \notin_{-} M \ F, H, M \ newpointer(x, e, e') \\ \hline e \ val \ x \notin_{-} M \ F, H, M \ newpointer(x, e, e') \\ \hline e \ val \ x \notin_{-} M \ F, H, M \ newpointer(x, e, e') \\ \hline e \ val \ x \in_{-} M \ F, H, M \ newpointer(x, e, e') \\ \hline e \ val \ x \in_{-} M \ F, H, M \ newpointer(x, e, e') \\ \hline F, H, M \ newpointer(x, e, e') \mapsto_{-} F, H, M \ newpointer(x, e, e') \\ \hline e \ val \ x \in_{-} M \ F, H, M \ newpointer(x, e, e') \\ \hline F, H, M \ x := f($$

$$e_1 \ val \ \dots \ e_n \ val$$

$$\overline{(F,(f;x_1,...,x_n;c;e)), H, M \ x := f(e_1,...,e_i,...,e_n) \longmapsto F, H, M \ x := [F,H,(x_1 \hookrightarrow e_1,...,x_n \hookrightarrow e_n) \ (c;e)]}$$

$$F,H',M' \ c \longmapsto F,H'',M''c'$$

$$\overline{F,H,M \ x := [F,H',M' \ (c;e)] \longmapsto F,H'',M \ x := [F,H'',M'' \ (c';e)]}$$

$$c \ cfinal \ H,M' \ e \longmapsto H,M' \ e'$$

$$\overline{F,H,M \ x := [F,H,M' \ (c;e)] \longmapsto F,H,M \ x := [F,H,M' \ (c;e')]}$$

$$\frac{c \ cfinal \ e \ val}{\overline{F,H,M \ x := [F,H,M' \ (c;e)] \longmapsto F,H,M \ x := e}}$$

$$\frac{H,M \ e \longmapsto H,M \ e'}{\overline{F,H,M \ x := e \longmapsto F,H,M \ x := e'}}$$

$$e \ val$$

$$\overline{F,(H,y \hookrightarrow \_),(M,x \hookrightarrow EPtr(y)) \ *x := e \longmapsto F,(H,y \hookrightarrow e),(M,x \hookrightarrow EPtr(y)) \ skip}$$

## 3.4 Program

$$\overline{p} \longmapsto \emptyset, \emptyset, \emptyset \overline{p}$$

$$\overline{F, H, M \ newtype(\tau, t, p)} \longmapsto F, H, M \overline{p}$$

$$\overline{F, H, M \ newfunc(f; x_1, ..., x_n; c; e; p)} \longmapsto (F, (f; x_1, ..., x_n; c; e)), H, M \overline{p}$$

$$\overline{F, H, M \ pcommand(c)} \longmapsto F, H, M \overline{c}$$

## 4 Bezpieczeństwo typów

Twierdzenie: Jeśli p pwf i  $p \mapsto p'$  to p' pwf

## 5 Postęp obliczeń

Twierdzenie: Jeśli p pwf to  $\exists p'$   $p \mapsto p'$  lub p pfinal

## 6 Inne uwagi, które nie pasują nigdzie indziej

- Prezentowane reguły typowania i semantyki małych kroków odpowiada dokładnie implementacji.
- Ciąg deklaracji zmiennych i wskaźników (np.  $vars(x := 1, y := 3, r := 3)in\{skip\}$  jest w czasie parsowania zamieniany na jedną, zagnieżdzoną instrukcję, tak jak w języku W (newvar(x, 1, newvar(y, 3, newvar(r, 3, skip))))
- Podobnie rozwiązany jest rozwiązany problem nowych funkcji i typów definiowanych przez użytkownika. W pliku syntax.ml znajdują się wszystkie definicje typów.
- Definicja funkcji wymaga podania typu wszystkich argumentów.
- Stos wywołań funkcji w języku W++ jest symulowany stosem w języku OCaml (metoda processFunctionCall w pliku eval.ml).
- Nazwy zmiennej w ciele instrukcji nigdy nie może zostać przesłonięta.
- Indekser krotki musi zawsze być liczbą naturalną.
- Krotki sa niemutowalne aby zmienić jej jedna współrzędna, należy stworzyć nowa.
- Zmienna typu wskaźnikowego w pamięci trzyma etykietę komórki pamięci na stercie.
- Raz zaalokowana wartość na stercie nie może zostać usunięta, co najwyżej można ją ponownie nadpisać.
- W case'ie wszystkie przypadki musza zostać zdefiniowane.

- Typ abstrakcyjny nie musi być zdefiniowany tylko pod wskaźnikiem.
- $\bullet\,$ Plik Readme.mdzawiera informacje o komplikowaniu i uruchamianiu kodu.
- Funkcja może zwrócić wskaźnik.
- ullet Jedyny sposób, by zadeklarować zmienną typu TSum z więcej niż jednym konstruktorem, to wpisać cały typ w adnotacji typowej.
- Średnik po ostatniej instrukcji powoduje SyntaxError.
- Błędy typowania są w miarę dobrze opisane w ramach wyjątków. Jeśli program się otypował, to powinien się poprawnie wykonać (tj. zakończyć lub zapętlić).
- Katalog examples/ zawiera przykładowe programy.