Een natuurlijke semantiek voor prototype oververing en lexicaal bereik

Kelley van Evert & Tim Steenvoorden

29 mei 2012

Inhoudsopgave

1 Inleiding

- Motivatie
- JavaScript
- Lexical scope – vrije/gebeonden variabelen
- Objecten en prototype overerving

2 Notatie en terminologie

In dit hoofdstuk behandelen we zowel een aantal min of meer gebruikelijke wiskundige concepten, als een aantal specifieke notaties en begrippen, die in dit werkstuk vaak zullen terugkeren. Gezien het aard van het onderwerp zullen we bijvoorbeeld vaak over *eindige* functies en verzamelingen willen spreken.

2.1 Functies

In dit werkstuk identificeren we een functie met zijn grafiek, d.w.z. een functie $f : X \to Y$ wordt gedefiniëerd door de verzameling paren $(x, y) \in X \times Y$ waarvoor we beweren dat $f(x) = y$. Uiteraard voldoet zo'n verzameling $f \subseteq X \times Y$ aan de voorwaarde dat:

$$\neg\exists_{x \in X,\, y_1 \in Y,\, y_2 \in Y}\big[\,(x, y_1) \in f \wedge (x, y_2) \in f \wedge y_1 \neq y_2\,\big]$$

De reden voor deze aanpak is zeker niet fundamenteel: het is gewoonweg handig om $\varnothing$ te schrijven voor een nieuwe, "lege", partiële functie.

2.2 Partiële functies

Vrijwel alle functies die we in dit werkstuk behandelen zijn partiële functies. Wanneer een partiële functie $f : X \to Y$ niet gedefiniëerd is op een zeker punt x, d.w.z. $\neg\exists_{y \in Y}\big[\,(x, y) \in f\,\big]$, schrijven we $f(x) = \bot$. Wanneer het omgekeerde het geval is, schrijven we $f(x) \neq \bot$, of kortweg $f(x) = y$ voor de gecombineerde uitspraak dat f wél gedefiniëerd is op x én dat $(x, y) \in f$.

Voor een willekeurige term $\phi = \ldots f(x) \ldots$, waarbij f een partiële functie is die niet gedefiniëerd is op punt x, geldt dat $\phi = \bot$. Op deze manier hoeven we niet iets omslachtigs te schrijven als: "als $f(x) = \bot$, dan $z = \bot$; anders als $f(x) \neq \bot$, dan

$z = \phi$". Deze "verkote schrijfwijze" stelt ons in staat om op een elegantie manier functie definities op te schrijven. Een voorbeeld:

$$f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$$
$$f(0, 0) = 1$$
$$f(n, m + 1) = f(n, m)$$

In dit voorbeeld geldt voor alle $m \in \mathbb{N}$ dat $f(0, m) = 1$, en voor alle $n \in \mathbb{N} \setminus \{0\}$ is $f(n, m)$ niet gedefiniëerd.

2.3 Eindige functies, eindige verzamelingen

De reden dat de meeste behandelde functies partiëel zijn is omdat meeste onderdelen van ons semantisch model eindig van karakter zijn. Functies worden vaak gebruikt om "een verzameling variabelen die een bepaalde waarde bevatten" te representeren, bijvoorbeeld de gedefiniëeerde variabelen in een zekere scope. En het zou ongewoon zijn om in een programmeertaal gebruik te maken van scopes waarin oneindig veel variabelen kunnen bestaan.

Wanneer we het over een eindige functie f hebben, bedoelen we daarmee dat het domein van die functie een eindige verzameling is. Dit kan worden uitgedrukt door te zeggen dat er een zeker getal $n \in \mathbb{N}$ bestaat, zó dat $\underline{n}$ (gedefiniëerd als: $\{m \in \mathbb{N} \mid m \leq n\}$) isomorf is aan het domein van f.

$$f : X \to Y \text{ is "eindig"} \stackrel{\text{def}}{=} \exists_{n \in \mathbb{N}} \big[\, \underline{n} \cong \{x \in X \mid f(x) \neq \bot\} \,\big]$$

We schrijven "eindige" Y^X voor de verzameling functies $\{f : X \to Y \mid f \text{ "eindig"}\}$.

2.4 Lijsten

We zullen meermaals in ons werkstuk gebruik maken van willekeurig grote, maar altijd eindige, *lijsten* van elementen uit een zekere verzameling. Deze lijsten worden gerepresenteerd door eindige (partiële) functies $t : \mathbb{N} \to X$ (met X de verzameling waaruit we de elementen van de lijst nemen), waaraan nog een paar extra voorwaarden worden gesteld. De verzameling van alle lijsten op een zekere verzameling X, genoteerd $X_{\langle\rangle}$, is als volgt gedefiniëerd:

$$X_{\langle\rangle} \stackrel{\text{def}}{=} \{\, t : \mathbb{N} \to X \mid \exists_{N \in \mathbb{N}}\big[\, \forall_{n<N}[t(n) \neq \bot] \wedge \forall_{n \geq N}[t(n) = \bot]\,\big]\,\}$$

We schrijven $\langle\rangle$ (maar ook wel $\varnothing$ aangezien het gewoon een lege functie is zoals beschreven in het voorgaande) voor de lege lijst. Deze is natuurlijk altijd hetzelfde, ongeacht welke invulling wordt gekozen voor X.

Als t een zekere lijst is, dan zeggen we dat hij van *grootte* $N = \min\{n \in \mathbb{N} \mid t(n) = \bot\}$ is.

We schrijven $\langle x_0, x_1, \ldots, x_{N-1}\rangle$ voor de lijst t van grootte N waarvoor geldt dat $\forall_{n<N}[t(n) = x_n]$.

Als t een lijst is uit $X_{\langle\rangle}$, en x een element van X, dan schrijven we $t : x$ voor de lijst $t' = t[N \mapsto x]$, waarbij N de grootte van t is.

2.5 Notationele conventies

Terwille van leesbaarheid en elegantie houden we een aantal gebruikelijke notationele conventies aan.

- Veel wiskundige formules zijn van de vorm $t_1\ R\ t_2$, waarbij R een zeker predikaat is (mogelijk $=$), en t_1 en t_2 termen. Dit soort formules zullen we wel vaker "samenstellen" tot formules als:

  $$6 = 2 * 3 > 2 \geq 42 - 40 \tag{2.1}$$
  $$f(x) = y \in Y \tag{2.2}$$

  De intentie is enkel een elegante schrijfwijze te hanteren die makkelijk en intuïtief leest. Zo zou een college analyse onverdraaglijk zijn als dergelijke verkortingen niet zouden worden gebruikt. Als we bovenstaande formules uitschrijven krijgen we:

  $$6 = 2 * 3 \wedge 2 * 3 > 2 \wedge 2 \geq 42 - 40 \tag{2.3}$$
  $$f(x) = y \wedge y \in Y \tag{2.4}$$

- Bij het definiëren van verzamelingen gebruiken we vaak de volgende voor de hand liggende notatie:

  $$\{a \in A \mid \phi\}$$

…i.p.v. het omslachtigere:

$$\{a \mid a \in A \mid \phi\}$$

3 Taal en syntaxis

In dit hoofdstuk zullen we de taal presenteren waarvoor we een natuurlijk taal construeren. De taal maakt gebruikt van prototype overerving en lexicaal bereik. Eerst zullen we een aantal voorbeeldprogramma's beschouwen, om zo informeel het karakter van de te formaliseren taal over te brengen. Daarna geven we een rigoreuze definitie met behulp van een BNF grammatica. De structuur van de productieregels van grammatica worden in latere hoofdstukken gebruikt om axioma's en deductieregels op te stellen. Daarmee heeft de grammatica in zekere zin een dubbele functie.

Elk voorbeeldprogramma en zijn toelichtingen worden als volg gepresenteerd:

Code fragment 3.1. Het eerste voorbeeldprogramma

```
local f                        — f moet eerst worden gedefiniëerd
f = function (i) returns n
  local n
  n = 2 × (i + 5)
                               — x bestaat niet in deze scope
local x                        — x is ongedefinieerd (maar wel aanwezig)
x = f(42)
```

De toelichtingen moeten als informeel commentaar worden beschouwd, waarmee we aan proberen te geven hoe het programma zich gedraagt. Vaak zijn het uitspraken over de toestand waarin het programma zich bevindt, direct na de linker regel te hebben "uitgevoerd".

3.1 Voorbeeldprogramma's

Een variabele moet gedeclareerd worden, en pas daarna kan er een waarde aan worden toegekend.

```
           — x bestaat niet (in deze scope)
local x    — x is ongedefinieerd (maar wel aanwezig)
x = 5      — x = 5
```

Het concept van declaratie is juist in deze taal, gezien het lexicaal bereik van variabelen, heel belangrijk. Vergelijk het bovenstaande programma fragment bijvoorbeeld met de volgende situatie.

Variabelen hebben geen vaste type. Er zijn drie typen waarden in de taal: getallen, functies en objecten.

```
local x
x = 5                        — de waarde van x is een getal
x = function () { skip }     — de waarde van x is een functie
x object                     — de waarde van x is een object
```

De taal is object georienteerd.

```
local o
o object
                               — o.f is niet gedefinieerd
o.f = function () { skip }     — toekenning waarde aan object attribuut
                               — o.f is wel gedefinieerd
o.n = 5
```

Van de drie typen, zijn getallen en functies *primitief*, en objecten *niet primitief*. Primitieve waarde worden zelf gekopieerd (*by-value*), maar van niet-primitieve waarden worden *referenties* gekopieerd (*by-reference*).

```
local x;  x = 6
local y;  y = x      — x = 6 en y = 6
y = 7                — x = 6 en y = 7

local p;  p.n = 6
local q;  q = p      — p en q verwijzen nu naar hetzelfde object
                     — p.n = 6 en q.n = 6
q.n = 7              — p.n = 7 en q.n = 7
```

3.1.1 Lexical scope

Als in een zekere scope een variabele wordt gereferenceerd (nog) niet is gedefinieerd, wordt in omliggende scopes “gezocht” naar een definitie van deze variabele.

```
local x
local f; f = function (i)
  x = i + 5

f(5)                          — x = 10
```

..maar wanneer deze wel in de huidige scope bestaat, worden omliggende scopes “met rust gelaten”.

```
local x
local f
f = function (i)
  local x
  x = i + 5

f(5)                 — x heeft nog geen waarde
```

Telkens wanneer een functie wordt aangeroepen, wordt een *nieuwe scope* aangemaakt voor lokale variabelen. Variabelen van deze nieuwe scope kunnen later nog gereferenceerd worden, doordat bijvoorbeeld de functie een lokale functie teruggeeft.

```
local f
f = function (n) returns g
  local g
  g = function () returns n
    n = n + 1

local c
c = f(5)                        — c() → 6, 7, 8, . . .
```

Voorbeeld 3.1: Een countervoorbeeld

```
local f
f = function(n) returns g
    local g
    g = function() returns n
        n = n + 1

local c
c = f(5)
c()                              # → 6, 7, 8, . . .
```

maar dan wat beter geschreven, etc...

3.1.2 Prototype overerving

Prototype overerving is een variant van object-geörienteerd programmeren. De kern van object-geörienteerd programmeren is het concept van een *object*, dat ertoe dient een verschijnsel uit de werkelijkheid na te bootsen (een reëel object, een patroon, een abstract idee). Het doel is om meer te kunnen programmeren op een conceptueel niveau. Daarmee wordt bijvoorbeeld zowel creatie als onderhoud van de code makkelijker.

Veel objecten zullen natuurlijk gelijke eigenschappen vertonen, of dezelfde structuur hebben. Verder wilt men concepten als specificering en generalisering toepassen op objecten. Deze problemen kunnen op meerdere manieren worden aangepakt. De bekendste variant is *klasse gebaseerde* object-oriëntatie (ook wel *klassieke object-oriëntatie*) en richt zich op het concept van een *klasse*. Objecten van een bepaalde klasse vertonen de structuur en gedrag van die klasse en heten *instanties*. Van specificering is sprake als een klasse eigenschappen van een andere klasse *overerft*. Klassieke object-oriëntatie vind men in talen als Java en C#.

Een andere aanpak met hetzelfde doel is *prototype gebaseerde* object-oriëntatie. Daarbij wordt geen scheiding gemaakt tussen de concepten klasse, die structuur en gedrag specificeert, en instantie, die enkel deze eigenschappen vertoont. In plaats daarvan wordt gewerkt met een prototype structuur, waarbij elk object naar een bepaald *prototype*-object refereert. Nu zijn objecten zelf de dragers van structuur en gedrag.

Technisch gezien werkt prototype overerving als volgt. Van elk object is een prototype bekend, of het heeft geen prototype. Wanneer men een attribuut opvraagt van een zeker object, kan de op te leveren waarde procedureel als volgt worden opgevat:

1. Bekijk of het attribuut gedefiniëerd is in het object zelf. In dat geval weten we de waarde en leveren deze op.
2. Anders zoeken we het attribuut op in het prototype van het object. Ook dan weten we de waarde en leveren deze op.
3. Wanneer ook het prototype het attribuut niet bevat, herhalen we de zoektocht voor alle volgende prototypen totdat we het attribuut hebben gevonden.

Het grote verschil tussen object-gebaseerde talen en prototype-gebaseerde talen is dus dat de tweede geen onderscheid maakt tussen klassen en instanties. Een prototype heeft beide functies. Neem bijvoorbeeld het object `Deur`:

```
local Deur
Deur object
```

We kunnen `Deur` direct als instantie gebruiken door een attribuut te zetten:

```
Deur.open = 1
```

Een `Deur` is standaard open. We kunnen `Deur` ook als een prototype gebruiken. In prototype-gebaseerde talen heet dit *klonen*:

```
local GeslotenDeur
GeslotenDeur object
GeslotenDeur clones Deur
```

`GeslotenDeur` heeft dan alle attributen van `Deur`:

```
GeslotenDeur.open  — waarde → 1
```

Maar een `GeslotenDeur` moet natuurlijk gesloten zijn. We zetten zijn attribuut `open` op `0`:

```
GeslotenDeur.open = 0
```

Een gewone `Deur` is nog steeds open:

```
Deur.open  — waarde → 1
```

Attributen worden dus per object bewaard. Door `open` op `0` te zetten in `GeslotenDeur` verandert er niks in `Deur`.

We kunnen net zoveel klonen maken van een object als we willen en net zo diep klonen als we willen. Neem een `GlazenDeur`, dit is natuurlijk ook een `Deur`, maar wel doorzichtig:

```
local GlazenDeur
GlazenDeur object
GlazenDeur clones Deur
GlazenDeur.doorzichtig = 1
```

Een gewone `Deur` heeft het attribuut `doorzichtig` niet, en dus een `GeslotenDeur` ook niet:

```
GeslotenDeur.doorzichtig  — fout!
```

Maar we kunnen besluiten dat deuren standaard niet doorzichtig zijn:

```
Deur.doorzichtig = 0
```

Zodat ook onze `GeslotenDeur` niet doorzichtig is:

```
GeslotenDeur.doorzichtig  — waarde → 0
```

Maar er geld nog steeds:

```
GlazenDeur.doorzichtig  — waarde → 1
```

We zien dat we met prototypes een zeer flexibele methode hebben om object-geörienteerd te programmeren. Het is niet nodig om de compiler of parser van te voren uit te leggen dat objecten aan bepaalde “blauwdrukken” moeten voldoen. We creëren objecten “on-the-fly”, alsmede hun attributen en relaties. Deze methode komt terug in talen als JavaScript, IO en Self.

Natuurlijk is het ook mogelijk om *methoden* te definiëren. Dit zijn functie attributen gekoppeld aan een specifiek object. Stel dat we een `GeslotenDeur` graag open willen maken. We definiëren:

```
GeslotenDeur.ontsluit = function (poging)
  if (poging = THIS.code) then
    THIS.open = 1
   else
    THIS.open = 0
```

`this` is hier een expliciete verwijzing naar het huidige object. Op dit moment kunnen we `ontsluit` nog niet aanroepen op `GeslotenDeur`:

```
GeslotenDeur.ontsluit(1234)  — fout!
```

Het attribuut `code` is immers niet gedefinieerd in `GeslotenDeur` noch in zijn prototype `Deur`.

We kunnen natuurlijk een `code` toekennen aan `GeslotenDeur`, maar laten we een specifieke `GeslotenDeur` maken met een `code`:

```
local Kluis
Kluis object
Kluis clones GeslotenDeur
Kluis.code = 4321
```

Wanneer we de methode `ontsluit` aanroepen is deze niet gedefinieerd in `Kluis`, maar wel in zijn prototype `GeslotenDeur`. Die wordt dan uitgevoerd. Een belangrijke observatie is dat `ontsluit` wel wordt aangeroepen op `Kluis`. Dat betekent dat `this` verwijst naar `Kluis` en niet `GeslotenDeur`. Het attribuut `code` wordt dan wel gevonden:

```
Kluis.ontsluit(1234)
Kluis.open                    — waarde → 0
```

Helaas was dat de verkeerde code, we proberen het nog een keer:

```
Kluis.ontsluit(1234)
Kluis.open                    — waarde → 1
```

3.2 Grammatica

Nu volgt een formele definitie van de syntaxis van de taal, aan de hand van een BNF grammatica. Nummers zijn als volgt gedefiniëerd:

$$\mathit{Num} ::= (0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9)^{+}$$

Eigenlijk gebruiken we geen strikte BNF, in deze specifieke gevallen, maar een hele simpele variant, zoals E-BNF, waarbij de ook simpele reguliere expressies toelaten.Bovenstaand voorbeeld maakt dit duidelijk. Voorbeelden van elementen uit *Num* zijn “0”, “1”, “235783” en “0003”. Voorbeelden van elementen die niet in *Num* zitten zijn “”, “-6”, “4.2”.

Identifiers, die gebruikt worden als variabel-namen,zijn op eenzelfde manier als volgt gedefiniëerd:

$$\mathit{Id} ::= (a \mid b \mid c \mid d \mid \ldots)^{+}$$

Hierbij moet je je natuurlijkvoorstellen dat alle letters uit het alfabet in de grammaticaregel staan op de voordehandliggendemanier.

Het is soms ook nodig om meerdere komma-gescheiden namen te gebruiken, of een mogelijk lege lijst, zoals bij functie definities. Vandaar de volgende twee productieregels:

$$\begin{aligned} \mathit{Ids} &::= \mathit{Id} \mid \mathit{Ids}\,,\mathit{Id} \\ \mathit{MaybeIds} &::= \varepsilon \mid \mathit{Ids} \end{aligned}$$

Een *slot*is een opeenvolging door punten (“.”) gescheiden identifiers,en wordt gebruikt om ook naar attributen van objecten te kunnen referenceren:

$$Slot ::= Id \mid Id.Slot$$

Expressies, die ofwel primitieve waarden (getallen en functies), ofwel objecten kunnen weergeven, en *boolse expressies*, die gebruikt worden voor loops en conditionele executie, zijn als volgt gedefiniëerd:

$$\begin{aligned}
Expression &::= Num \mid Slot \mid \varnothing \mid Expression\ (+ \mid - \mid \times \mid /)\ Expression \\
&\quad \mid \texttt{function}\,\texttt{(}MaybeIds\texttt{)}\ [\ \texttt{returns}\ Id\ ]\ \texttt{\{}\ Stm\ \texttt{\}} \\
Expressions &::= Expression \mid Expressions\,,Expression \\
MaybeExpressions &::= \varepsilon \mid Expressions \\
BooleanExpression &::= \textsc{true} \mid \textsc{false} \\
&\quad \mid BooleanExpression\ (\textsc{and} \mid \textsc{or})\ BooleanExpression \\
&\quad \mid \textsc{not}\ BooleanExpression \\
&\quad \mid Expression\ (= \mid < \mid \leq \mid > \mid \geq)\ Expression
\end{aligned}$$

De kern van de hele grammatica draait om de volgende productieregels, die van *statements*. Een statement is een programma van goede vorm. Het betekent niet nodelijkerwijsdat het programma *valide* is, maar alle valide programma’s zitten wel in *Stm*.

$$\begin{aligned}
Stm &::= Stm\,;Stm \\
&\quad \mid \texttt{skip} \\
&\quad \mid \texttt{if}\ BooleanExpression\ \texttt{then}\ Stm\ \texttt{else}\ Stm \\
&\quad \mid \texttt{while}\ BooleanExpression\ \texttt{do}\ Stm \\
&\quad \mid \texttt{local}\ Id \\
&\quad \mid Id\ \texttt{object} \\
&\quad \mid Id\ \texttt{clones}\ Id \\
&\quad \mid Slot = Expression \\
&\quad \mid [Slot =]\ Id\,\texttt{(}MaybeExpressions\texttt{)}
\end{aligned}$$

Vanwege de focus van dit werkstuk definiëren we niet precies wanneer een programma valide is en wanneer niet.

4 Semantisch model

4.1 Bindingen

Aan de basis van ons model ligt het concept van een *binding*. Een binding is een toekenning van een *waarde* aan een variabele (een element uit de syntactische verzameling *Id* komt). Bindingen zijn bijvoorbeeld van belang om de gedefiniëerde variabelen binnen een scope vast te leggen, of de attributen van een bepaald object. Een *groep bindingen* is een eindige functie $b : Id \to \mathbb{V}$. De verzameling van alle groepen van bindingen definiëren we dus als

$$\mathbb{B} \stackrel{\text{def}}{=} \mathbb{V}^{Id}$$

We komen later terug op wat de waarden $\mathbb{V}$ precies zijn in sectie 4.5. Voor nu is het voldoende om te weten dat in ieder geval de natuurlijke getallen $\mathbb{N}$ deel uitmaken van $\mathbb{V}$.

Bindingen komen veelvuldig terug in ons model. In scopes worden *variabelen* gedeclareerd en aan waarden gekoppeld. Bij objecten zijn het de *attributen* die waarden krijgen toegekend.

4.2 Scope en omliggende scopes

In sectie 3.1 is informeel gebleken dat scopes conceptueel goed te zien zijn als een boomstructuur. De evaluatie van een zekere variabele x in scope s zou dan procedureel als volgt kunnen worden uitgelegd:

```
x — We bevinden ons in een zekere scope s.
```

Dan zoeken we x eerst op in de bindingen groep b_s, behorende bij scope s.

$$b_s(x).$$

Zoals ook te zien in **??** hebben we twee mogelijkheden:

1. `x` is gedefinieerd in b_s en we gebruiken de gevonden waarde.
2. `x` is niet gedefinieerd in b_s en we moeten `x` opzoeken in de omliggende scope.

We moeten dus niet alleen de bindingen van de scope zelf bijhouden, maar ook een verwijzing naar zijn *omgevende scope*. Een scope s is definiëren we dus als een paar $(b\,,\pi)$, met in b de bindingen en π een *verwijzing* naar de omgevende scope (ook wel *parent*, of *outer* scope).

We moeten benadrukken dat π een *verwijzing* is, en niet een *kopie* van de bindingen groep van de omgevende scope. Stel dat we het programma in code fragment 4.2 uitvoeren. Op het moment dat we $f()$ aanroepen in regel 7 willen we dat x daarna evalueert naar de waarde 2. Evenzo moet x na regel 8 evalueren naar de waarde 4. De scope s_f van functie f heeft een eigen binding b_f die gedurende de executie van het programma leeg is, x is namelijk niet gedeclareerd als een **`local`** variabele. De omgevende scope π_f van functie f verwijst naar scope s, zodat de variabele x uiteindelijk wel gevonden wordt.

Code fragment 4.2. Lexicale scope: opslaan en vinden van variabelen

```
1 local x
2 x = 1
3 local f
4 f = function ()  — Introductie nieuwe scope
5   x = 2 × x
6                  — Einde nieuwe scope
7 f()              — x = 2
8 f()              — x = 4
```

Stel dat we geen verwijzing in de scope opslaan maar een kopie van de omgevende bindingen. Op het moment dat we f definiëren in regel 4 is scope s_f een paar $(b_f\,,p_f)$ met $b_f\,,p_f \in \mathbb{B}$. Net als hierboven zijn de eigen bindingen b_f leeg. De binding p_f bevat een functie onder naam f en de waarde 1 onder naam x. Wanneer we x aanpassen door de aanroep in regel 7 wordt dit doorgevoerd in de binding p_f maar, omdat dit een kopie is, niet in de binding b_s van de omgevende scope s. We moeten dus wel een verwijzing opslaan willen we het gevraagde gedrag krijgen. Daarnaast wordt het met kopieën erg lastig om een boomstructuur te creëren zodat we een variabele nog hogerop kunnen opzoeken.

Een scope s is dus een element uit de verzameling

$$\mathbb{S} \stackrel{\text{def}}{=} \mathbb{B} \times (\mathbb{L}_s \cup \{\perp\}).$$

Hierbij zijn $\mathbb{B}$ de bindingen zoals besproken in §**??**. $\mathbb{L}_s$ zijn locaties van scopes. Op het begrip locatie komen wij nog terug in §4.6. We moeten er wel rekening mee houden dat er een soort "ultieme" omgevende scope is. Het kan dus zijn dat een scope geen parent heeft. In dat geval zetten we

$$\pi = \perp.$$

We zeggen dat π *niet bestaat* of *niks* is. Vandaar dat we het symbool $\perp$ toevoegen aan $\mathbb{L}_s$.

4.3 Objecten en prototype overerving

In §**??** hebben we een beeld gekregen van prototype overerving. Net als scopes en omgevende scopes, blijken objecten en prototypen te modelleren met een boomstructuur. Geheel in lijn met scopes is een object een paar met daarin zijn eigen bindingen b en een verwijzing naar zijn prototype π. Natuurlijk kan een object ook geen prototype hebben. Dit geven we weer aan met $\perp$. Een object o is dan een element uit

$$\mathbb{O} \stackrel{\text{def}}{=} \mathbb{B} \times (\mathbb{L} \cup \{\perp\}).$$

Hierbij zijn $\mathbb{B}$ weer de bindingen uit §**??** en $\mathbb{L}_o$ zijn locaties van objecten. We maken dus een strikte scheiding tussen locaties van scopes en locaties van objecten.

4.4 Functies

4.5 Waarden: referenties en primitieven

In voorgaande paragrafen spraken we telkens over waardes $\mathbb{V}$ en locaties $\mathbb{L}$ ($\mathbb{L}_s$ voor scopes $\mathbb{L}_o$ voor objecten). We hebben de exacte definities hiervan in het midden gelaten. Een waarde $v \in \mathbb{V}$ is iets wat we toekennen aan een *Id* met behulp van een binding. Bindingen waren immers functies van *Id* naar waardes. In de voorbeelden van §3.1 zijn we verschillende waardes tegengekomen:

```
local n
n = 5                         # Natuurlijke getallen
local f
f = function (x) returns y    # Functies
    y = 2 * x
local A
A object                      # Objecten
```

In onze taal komen dus drie typen waardes voor. Er is echter een verschil in de manier waarop wij ze behandelen. Wat gebeurt er wanneer we bovenstaande waardes opnieuw toekennen aan andere variabelen? Allereerst een natuurlijk getal, deze is het eenvoudigst:

```
local m
m = n
```

Op dit moment hebben zowel `m` als `n` waarde `5`. Wanneer we aan `m` een andere waarde toekennen, gebeurt er niets met `n`. Omgekeerd geld hetzelfde. We kunnen dus simpelweg de waarde van `n` *kopiëren* en in `m` stoppen. Onze waardes $\mathbb{V}$ bevatten dus sowieso $\mathbb{N}$:

$$\mathbb{V} \stackrel{\text{def}}{=} \mathbb{N}$$

Voor functies geld eigenlijk hetzelfde:

```
local g
g = f
```

`g` heeft nu als waarde een kopie van de functie in variabele `f`. Beiden kunnen we aanroepen:

```
f(4)                          # => 8
g(6)                          # => 12
```

Dus $\mathbb{V}$ bevat ook functies:

$$\mathbb{V} \stackrel{\text{def}}{=} \mathbb{N} \cup \mathbb{V}$$

Helaas gaat dit *niet* op voor objecten. Stel dat `A` een attribuut `x` heeft

```
A.x = 7
```

en we koppelen een nieuwe variabele B aan het object in A:

```
local B
B = A
```

Nu heeft `B` sowieso dezelfde attributen als `A`:

```
B.x                         # => 7
```

Maar als we `x` in `B` ophogen

```
B.x = 8
```

moet dit in `A` natuurlijk ook gebeuren:

```
A.x                         # => 8
```

Wanneer we zomaar een kopie van de waarde in `A` aan `B` toekennen, zou het attribuut `A.x` niet gewijzigd worden. Een oplossing is om in plaats van een kopie van een object een *referentie* naar een object aan een variabele te koppelen. Dit zijn dus niet elementen uit $\mathbb{O}$ zelf, maar uit $\mathbb{L}_o$. Dat betekend dat objecten niet op dezelfde manier worden behandeld als getallen en functies, maar de *locaties* van objecten wel. Voor de waardes krijgen we dus

$$\mathbb{V} \overset{\text{def}}{=} \mathbb{L}_o \cup \mathbb{N} \cup \mathbb{V}.$$

Primitieve waardes in onze taal koppelen we *by-value*. Dit zijn natuurlijke getallen en functies. Objecten daarentegen worden *by-reference* gekoppeld. Deze referenties zijn elementen uit $\mathbb{L}_o$. Achter de schermen worden deze gebruikt om objecten door te geven en worden op hun beurt wel *by-reference* gekoppeld.

4.6 Locaties en geheugen

Uit de redenaties over referenties volgt dat we een plaats moeten hebben om alle objecten op te slaan. Dit noemen we het *geheugen voor objecten m_o*. Voor een gegeven locatie $\omega \in \mathbb{L}_o$ kunnen we het bijbehorende object in een geheugen m_o opzoeken met

$$m_o(\omega).$$

Een geheugen is dus een functie van locaties naar objecten. De verzameling van alle geheugens definiëren we als

$$\mathbb{M}_o \overset{\text{def}}{=} \mathbb{O}^{\mathbb{L}_o}.$$

Dan is m_o een element hier uit.

Extra

$\mathbb{L} \stackrel{\text{def}}{=} \{(n, n) \in \mathbb{N}^2\}$	locaties van scopes en objecten
$\mathbb{F} \stackrel{\text{def}}{=} \mathit{Stm} \times \mathit{Id}_{\langle\rangle} \times (\mathit{Id} \cup \{\pm\}) \times \mathbb{L}$	functies
$\mathbb{V} \stackrel{\text{def}}{=} \mathbb{L} \cup \mathbb{N} \cup \mathbb{F}$	waarden
$\mathbb{B} \stackrel{\text{def}}{=}$ "eindige" $\mathbb{V}^{\mathit{Id}}$	binding-verzamelingen
$\mathbb{O} \stackrel{\text{def}}{=} \mathbb{B} \times (\mathbb{L} \cup \{\pm\})$	objecten
$\mathbb{S} \stackrel{\text{def}}{=} \mathbb{B} \times (\mathbb{L} \cup \{\pm\})$	scopes

5 Natuurlijke Semantiek

Deze teksten zijn vooral bedoeld als "tekstvlees" (lorem ipsum's). We zullen axioma's en deductieregels introduceren waarmee we de relatie $(\longrightarrow)$ definiëren, die de volgende signatuur heeft:

$$(\longrightarrow) \subseteq (\mathit{Stm} \times \mathbb{M} \times \mathbb{L} \times \mathbb{L}) \times \mathbb{M}$$

Wanneer we een uitspraak doen van de vorm:

$$\big\langle\, S \,\big\rangle_{m,\sigma,\tau} \longrightarrow m'$$

..dan bedoelen we daarmee dat:

$$\big(\,(S, m, \sigma, \tau), m'\,\big) \in (\longrightarrow)$$

Deze uitspraak moet je lezen als: "In de toestand met geheugen m, scope σ en *this* object τ, termineert het statement S, waarbij het resultaat-geheugen m' is."

Een van deze axioma's [object], heeft betrekking tot de productieregel in de grammatica die de `object` "literal" introduceert.

$$\big\langle\, i\ \texttt{object} \,\big\rangle_{m,\sigma,\tau} \longrightarrow m''$$

$$\begin{array}{rl} \textbf{desda} & \textsc{Find}_m(\sigma, i) = \sigma_{\text{def}} \\ & m' = m\big[\,\sigma_{\text{def}} \mapsto \big(b_{m(\sigma')}[i \mapsto \omega]\,, p_{m(\sigma')}\big)\,\big] \\ & m'' = m'\big[\,\omega \mapsto \big(\varnothing\,, \stackrel{\perp}{-}\big)\,\big] \end{array}$$

Zoals vele axioma's en deductieregels heeft ook dit axioma een aantal voorwaarden waaraan moet worden voldaan. Deze staan eronder genoteerd, elk op een regel.

Wanneer bij een dergelijke opsomming van voorwaarden een nieuwe variabele wordt geïntroduceerd zoals hierboven, met de volgende vorm: **desda** $\square = \theta \ldots$; dan moet deze gelezen worden als: **desda** $\exists_\theta\big[\,\square = \theta \ldots\,\big]$.

$$\big\langle\, i \ \texttt{clones}\ j \,\big\rangle_{m,\sigma,\tau} \longrightarrow m'$$

$$\begin{aligned} \textbf{desda}\quad & [\![\, i \,]\!]_{m,\sigma,\tau} = \omega_i \in \mathbb{L} \\ & [\![\, j \,]\!]_{m,\sigma,\tau} = \omega_j \in \mathbb{L} \\ & m' = m\big[\, \omega_i \mapsto \big(b_{m(\omega_i)}\,, \omega_j\big)\,\big] \end{aligned}$$

bla bla

$$\big\langle\, \texttt{local}\ i \,\big\rangle_{m,\sigma,\tau} \longrightarrow m'$$

$$\textbf{desda}\quad m' = m\big[\, \sigma \mapsto \big(b_{m(\sigma)}[i \mapsto \perp]\,, p_{m(\sigma)}\big)\,\big]$$

bla bla

$$\big\langle\, \textsc{this}.s = e \,\big\rangle_{m,\sigma,\tau} \longrightarrow m'$$

$$\begin{aligned} \textbf{desda}\quad & \textsc{Trav}_m(\tau, s) = (\omega\,, i) \\ & [\![\, e \,]\!]_{m,\sigma,\tau} = v \\ & m' = m\big[\, \omega \mapsto \big(b_{m(\omega)}[i \mapsto v]\,, p_{m(\omega)}\big)\,\big] \end{aligned}$$

bla bla

$$\big\langle\, i = e \,\big\rangle_{m,\sigma,\tau} \longrightarrow m'$$

$$\begin{aligned} \textbf{desda}\quad & \sigma_{\text{def}} = \textsc{Find}_m(\sigma, i) \\ & [\![\, e \,]\!]_{m,\sigma,\tau} = v \\ & m' = m\big[\, \sigma_{\text{def}} \mapsto \big(b_{m(\sigma_{\text{def}})}[i \mapsto v]\,, p_{m(\sigma_{\text{def}})}\big)\,\big] \end{aligned}$$

bla bla

$$\big\langle\, i.s = e \,\big\rangle_{m,\sigma,\tau} \longrightarrow m'$$

$$\begin{aligned} \textbf{desda}\quad & \sigma_{\text{def}} = \textsc{Find}_m(\sigma, i) \\ & b_{m(\sigma_{\text{def}})}(i) = \omega \in \mathbb{L} \\ & \textsc{Trav}_m(\omega\,, s) = (\omega'\,, j) \\ & [\![\, e \,]\!]_{m,\sigma,\tau} = v \\ & m' = m\big[\, \omega' \mapsto \big(b_{m(\omega')}[j \mapsto v]\,, p_{m(\omega')}\big)\,\big] \end{aligned}$$

bla bla

$$\frac{\left\langle\, S_f \,\right\rangle_{m', \sigma_{f\,\text{new}}, \omega'} \longrightarrow m''}{\left\langle\, i.s(e^*) \,\right\rangle_{m,\sigma,\tau} \longrightarrow m''}$$

$$\begin{array}{ll} \textbf{desda} & \sigma_{\text{def}} = \textsc{Find}_m(\sigma, i) \\ & b_{m(\sigma_{\text{def}})}(i) = \omega \in \mathbb{L} \\ & \textsc{Trav}_n(\omega, s) = (\omega', j) \\ & (S_f, I_f, i_f, \sigma_{f\,\text{def}}) = f = b_{m(\omega')}(j) \\ & \sigma_{f\,\text{new}} = \textsc{Next}_{\text{scope}}(m) \\ & m' = m\big[\, \sigma_{f\,\text{new}} \mapsto \big(\llbracket\, e^* \,\rrbracket^*_{m,\sigma,\tau}(I_f), \sigma_{f\,\text{def}}\big) \,\big] \end{array}$$

bla bla

$$\left\langle\, \texttt{skip} \,\right\rangle_{m,\sigma,\tau} \longrightarrow m$$

bla bla

$$\frac{\left\langle\, S_1 \,\right\rangle_{m,\sigma,\tau} \longrightarrow m' \qquad \left\langle\, S_2 \,\right\rangle_{m',\sigma,\tau} \longrightarrow m''}{\left\langle\, S_1 ; S_2 \,\right\rangle_{m,\sigma,\tau} \longrightarrow m''}$$

bla bla

$$\frac{\left\langle\, S_1 \,\right\rangle_{m,\sigma,\tau} \longrightarrow m'}{\left\langle\, \texttt{if}\ (b)\ \texttt{then}\ \{S_1\}\ \texttt{else}\ \{S_2\} \,\right\rangle_{m,\sigma,\tau} \longrightarrow m'}$$

$$\textbf{desda} \quad \llbracket\, b \,\rrbracket^{\mathrm{B}}_{m,\sigma,\tau} = \mathbf{T}$$

bla bla

$$\frac{\left\langle\, S_2 \,\right\rangle_{m,\sigma,\tau} \longrightarrow m'}{\left\langle\, \texttt{if}\ (b)\ \texttt{then}\ \{S_1\}\ \texttt{else}\ \{S_2\} \,\right\rangle_{m,\sigma,\tau} \longrightarrow m'}$$

$$\textbf{desda} \quad \llbracket\, b \,\rrbracket^{\mathrm{B}}_{m,\sigma,\tau} = \mathbf{F}$$

bla bla

$$\frac{\left\langle\, S_1 \,\right\rangle_{m,\sigma,\tau} \longrightarrow m' \qquad \left\langle\, \texttt{while}\ (b)\ \texttt{do}\ \{S_1\} \,\right\rangle_{m',\sigma,\tau} \longrightarrow m''}{\left\langle\, \texttt{while}\ (b)\ \texttt{do}\ \{S_1\} \,\right\rangle_{m,\sigma,\tau} \longrightarrow m''}$$

$$\textbf{desda} \quad \llbracket\, b \,\rrbracket^{\mathrm{B}}_{m,\sigma,\tau} = \mathbf{T}$$

bla bla

$$\left\langle\, \texttt{while}\ (b)\ \texttt{do}\ \{S_1\} \,\right\rangle_{m,\sigma,\tau} \longrightarrow m$$

$$\textbf{desda} \quad \llbracket\, b \,\rrbracket^{\mathrm{B}}_{m,\sigma,\tau} = \mathbf{F}$$

6 Case study: Wiskundige formulering

De vorm van bovenstaand semantisch model geeft een conceptueel sterk beeld van hoe de taal waarschijnlijk geimplementeerd zou worden in een compiler (modulo technische details, etc...). Het is ook vrijwel volgens bovenstaande semantiek dat aan informatica studenten object geörienteerde talen worden uitgelegd (referenties, primitieve waarden, ...). Je kunt echter ook semantisch model maken wat "wiskundiger van aard" is. Daar zal dit hoofdstuk over gaan.

Belangrijke informatie, zoals welk object van welk ander object een prototype is en hoe de scopes elkaar bevatten, maar ook welke informatie binnen een object zit opgeslagen en welke variabelen in een scope zijn gedefiniëerd, worden ditmaal door relaties bevat.

Rest nog identificatie van objecten en scopes, deze zal op eenzelfde manier worden behandeld als eerder de "locaties" binnen het geheugen: ze moeten identiek zijn, maar verder is het van geen belang hoe ze worden gerepresenteerd. We zullen daarom aannemen dat er twee verzamelingen identifiers zijn, $\mathbb{S}$ en $\mathbb{O}$, waarbij bovendien:

1. Voor beide verzamelingen $X \in \{\mathbb{S}, \mathbb{O}\}$, bestaat er een functie $\textsc{Next} : \mathcal{P}(X) \rightarrow X$, zdd $\textsc{Next}(Y) \notin Y$ voor alle $Y \subseteq X$.
2. $\mathbb{S} \cap \mathbb{O} = \varnothing$ (Deze eis is niet een technische benodigdheid, maar geeft wel aan dat de semantiek van deze twee soorten identifiers nu eenmaal anders is.)

De prototype hiërarchie, eveneens de scope hiërarchie, worden natuurlijk heel goed weergegeven door partiële ordeningen. Deze twee relaties zullen zullen we als $\sqsubset^p$ en $\sqsubset^s$, respectievelijk, noteren, en aannemen:

1. $\sqsubset^p$ is een (tweestemmige) partiële ordening op $\mathbb{O}$
2. $\sqsubset^s$ is een (tweestemmige) partiële ordening op $\mathbb{S}$

Vervolgens introduceren we de relatie $attr \subseteq \mathbb{O} \times Id \times (\mathbb{O} \cup \mathbb{P})$. De uitspraak "$p[i] = v$" moet worden gelezen als: $(p, i, v) \in attr$. Deze relatie wordt eveneens gebruikt om de "inhoud" van objecten weer te geven, als de graafstructuur tussen objecten.

Soortgelijk definiëren we de relatie $def \subseteq \mathbb{S} \times Id \times (\mathbb{O} \cup \mathbb{P})$, waarbij de uitspraak "$\sigma[i] = v$" moet worden gelezen als: $(\sigma, i, v) \in def$.

Ten alle tijde moeten de prototype en scope hiërarchieën, de inhoud van objecten en scopes, en de laatste indentificaties van objecten en scopes worden bijgehouden, en dit vormt dan het "geheugen", of de "toestand": $s = (attr, def, \sqsubset^p, \sqsubset^s, p_{next}, \sigma_{next})$.

$$
\begin{aligned}
\langle \texttt{skip}, s, \tau, \sigma \rangle &\longrightarrow s \\
\langle x = e, s, \tau, \sigma \rangle &\longrightarrow s[\ \sigma_d[x] = e\ ] \\
&\quad \textbf{desda}\ \sigma_d = \max\{\ \sigma' \in \mathbb{S} \mid \sigma'[x] \wedge \sigma \sqsubseteq^s \sigma'\ \} \\
\langle x\ \texttt{object}, \tau, \sigma \rangle &\longrightarrow s[\ \sigma_d[x] = p\ ][\ p_{next} \mapsto p + 1\ ] \\
&\quad \textbf{desda}\ \sigma_d = \max\{\ \sigma' \in \mathbb{S} \mid \sigma'[x] \wedge \sigma \sqsubseteq^s \sigma'\ \} \\
&\quad \textbf{en}\ p = p_{next_s}
\end{aligned}
$$