Università degli Studi di Verona
DIPARTIMENTO DI INFORMATICA
Linguaggi di programmazione
Riassunto dei principali argomenti
Autore: Davide Bianchi

Indice

1	Intr	oduzione	2
2	Eseı	mpio di linguaggio basilare	2
	2.1	Semantica big-step	2
		2.1.1 Esempio	2
	2.2	Semantica small-step	3
3	Ling	guaggio imperativo	3
	3.1	Memoria	4
	3.2	Sistemi di transizione	4
	3.3	Semantica small-step su un linguaggio imperativo	5
	3.4	Esecuzione di programmi e proprietà	5
	3.5	Funzione di valutazione della semantica	5
	3.6	Possibili varianti del linguaggio	6
		3.6.1 Inversione dell'ordine di valutazione	6
		3.6.2 Regole di assegnamento	6
		3.6.3 Inizializzazione della memoria	6
		3.6.4 Valori memorizzabili	6
	3.7	Type systems	6
		3.7.1 Regole di tipaggio	7
		3.7.2 Proprietà di tipaggio	7
4	Fori	me di induzione	8
	4.1	Induzione matematica	8
	4.2	Induzione strutturale	8
		4.2.1 Induzione strutturale su numeri naturali	8
		4.2.2 Induzione strutturale su strutture complesse	8
	4.3	Rule induction	9
5	Asp	etti funzionali	9
	5.1	Variabili free e bound	9
	5.2	Alpha conversion	10
	5.3	Sostituzioni	10
		5.3.1 Sostituzioni singole	10
		5.3.2 Sostituzioni simultanee	10
6	Dati	i e memoria variabile	11
7	Sott	o-tipaggio	11

1 Introduzione

Un linguaggio di programmazione è composto da:

- Sintassi: insieme di regole di scrittura del linguaggio;
- Semantica: descrizione del comportamento del programma a tempo di esecuzione;
- Pragmatica: descrizione delle caratteristiche del linguaggio, delle sue funzionalità ecc.

Gli stili per dare la semantica di un linguaggio sono 3:

- *Operazionale*: la semantica è data con sistemi di transizione, fornendo i passi della computazione passo passo;
- Denotazionale: il significato di un programma è dato dalla struttura di un insieme;
- Assiomatica: il significato è dato attraverso regole assiomatiche o qualche tipo di logica.

2 Esempio di linguaggio basilare

La semantica operazionale di un linguaggio è data attraverso un sistema di regole di inferenza, date come segue:

$$(Assioma) \; \frac{-}{(Conclusione)} \qquad (Regola) \; \frac{(Hyp_1) \; (Hyp_2) \; ... \; (Hyp_n)}{(Conclusione)}$$

Introduciamo la sintassi di un linguaggio basato solo su espressioni aritmetiche:

$$E := n | E | E + E | E * E \dots$$

La valutazione di programmi generati con questa sintassi può procedere in due modi:

- Small step: la semantica fornisce il sistema per procedere nell'esecuzione, passo dopo passo;
- Big step: la semantica va subito al risultato finale.

2.1 Semantica big-step

Forniamo la semantica big-step per il linguaggio dato sopra:

$$\operatorname{B-Num} \frac{-}{n \Downarrow n} \qquad \operatorname{B-Add} \frac{E_1 \Downarrow n_1 \ E_2 \Downarrow n_2}{E_1 + E_2 \Downarrow n_3} \ n_3 = add(n_1, n_2)$$

La semantica big-step fornisce immediatamente il risultato, dando subito il valore finale dell'espressione che si sta valutando.

2.1.1 Esempio

Teorema 2.1 (Determinatezza per semantica big-step). $E \Downarrow m \ e \ E \Downarrow n \ implica \ m = n$.

2.2 Semantica small-step

Indichiamo con $E_1 \rightarrow E_2$ lo svolgimento di un solo passo di semantica.

$$\begin{array}{c} E_1 \rightarrow E_1' \\ \hline E_1 + E_2 \rightarrow E_1' + E_2 \\ \hline \text{S-N.Right} & E_2 \rightarrow E_2' \\ \hline n_1 + E_2 \rightarrow n_1 + E_2' \\ \hline \text{S-Add} & - \\ \hline n_1 + n_2 \rightarrow n_3 \\ \hline \end{array} n_3 = add(n_1, n_2)$$

Con queste regole l'ordine di valutazione degli statement è fisso, procede sempre da sinistra verso destra. Diamo un'alternativa:

S-Left
$$\frac{E_1 \rightarrow_{ch} E_2}{E_1 + E_2 \rightarrow_{ch} E_1' + E_2}$$
S-Right
$$\frac{E_2 \rightarrow_{ch} E_2'}{E_1 + E_2 \rightarrow_{ch} E_1 + E_2'}$$
S-Add
$$\frac{-}{n_1 + n_2 \rightarrow_{ch} n_3} n_3 = add(n_1, n_2)$$

In questo caso l'ordine di valutazione è arbitrario. La notazione utilizzata in generale è la seguente:

- la relazione \rightarrow^k , con $k \in \mathbb{N}$, indica una sequenza di n passi applicando la semantica small-step;
- la relazione →*, indica una sequenza di derivazione lunga un certo numero di passi. Questa relazione è riflessiva ed è la chiusura transitiva di →.

Teorema 2.2 (Determinatezza per semantica small-step). Definiamo:

- strong determinacy: $E \to F$ e $E \to G$ implies F = G;
- weak determinacy: $E \rightarrow^* m$ e $E \rightarrow^* n$ implies m = n;

3 Linguaggio imperativo

Definiamo la sintassi di un semplice linguaggio imperativo:

$$\begin{split} b &:= true \mid false \\ n &:= \{...-1, 0, 1, 2, ...\} \\ l &:= \{l_0, l_1, ...\} \\ op &:= + \mid \geq \\ e &:= n \mid b \mid e \ op \ e \mid \text{if} \ e \ \text{then} \ e \ \text{else} \ e \mid l := e \mid !l \mid skip \mid e; e \mid \text{while} \ e \ \text{do} \ e \end{split}$$

Nota: lo statement !l indica l'intero memorizzato al momento alla locazione l. Inoltre il linguaggio non è tipato, quindi sono ammesse le sintassi come $2 \ge true$.

3.1 Memoria

La memoria è necessaria per poter valutare gli statement di lettura da una locazione. In particolare definiamo

$$dom(f) = \{a \in A \mid \exists b \in B \ s.t. \ f(a) = b\}$$

 $ran(f) = \{b \in B \mid \exists a \in A \ s.t. \ f(a) = b\}$

Lo store del linguaggio imperativo in questione è un insieme di funzioni parziali che vanno dalle locazioni di memoria nei numeri interi:

$$s: \mathbb{L} \to \mathbb{Z}$$

L'aggiornamento della memoria funziona come segue:

$$s[l \to n](l') = \begin{cases} n & \text{if } l = l' \\ s(l') & \text{altrimenti} \end{cases}$$

3.2 Sistemi di transizione

Le semantiche operazionali sono date attraverso sistemi di transizione, ovvero strutture composte da:

- un insieme Config di configurazioni;
- una relazione binaria $\Rightarrow \subseteq Config \times Config$;

Per indicare un generale passo di semantica si usa la notazione

$$\langle e, s \rangle \rightarrow \langle e', s' \rangle$$

che rappresenta una trasformazione di un programma e con una memoria s in un programma e' con memoria associata s'. I singoli passi di computazione sono singole applicazioni di regole della semantica.

3.3 Semantica small-step su un linguaggio imperativo

$$(\operatorname{op+}) \frac{-}{\langle n_1 + n_2, s \rangle \to \langle n, s \rangle} \quad n = \operatorname{add}(n_1, n_2) \qquad (\operatorname{op-geq}^1) \frac{-}{\langle n_1 \geq n_2, s \rangle \to \langle b, s \rangle} \quad b = \operatorname{geq}(n_1, n_2)$$

$$(\operatorname{op1}) \frac{\langle e_1, s \rangle \to \langle e'_1, s' \rangle}{\langle e_1 \text{ op } e_2, s \rangle \to \langle e'_1 \text{ op } e_2, s' \rangle} \qquad (\operatorname{op2}) \frac{\langle e_2, s \rangle \to \langle e'_2, s' \rangle}{\langle v \text{ op } e_2, s \rangle \to \langle v \text{ op } e'_2, s' \rangle}$$

$$(\operatorname{deref}) \frac{-}{\langle !l, s \rangle \to \langle n, s \rangle} \quad \text{if } l \in \operatorname{dom}(s) \text{ and } s(l) = n \qquad (\operatorname{assign1}) \frac{-}{\langle l := n, s \rangle \to \langle skip, s[l \to n] \rangle} \quad \text{if } l \in \operatorname{dom}(s)$$

$$(\operatorname{assign2}) \frac{\langle e, s \rangle \to \langle e', s' \rangle}{\langle l := e, s \rangle \to \langle l := e', s' \rangle} \qquad (\operatorname{if-tt}) \frac{-}{\langle \text{if true then } e_1 \text{ else } e_2, s \rangle \to \langle e_1, s \rangle}$$

$$(\operatorname{if-ff}) \frac{-}{\langle \text{if false then } e_1 \text{ else } e_2, s \rangle \to \langle e_1, s \rangle} \quad (\operatorname{if-ten}) \frac{\langle e, s \rangle \to \langle e', s' \rangle}{\langle \text{if } e \text{ then } e_1 \text{ else } e_2, s \rangle \to \langle e_1, s \rangle}$$

$$(\operatorname{seq.skip}) \frac{-}{\langle skip; e_2, s \rangle \to \langle e_2, s \rangle}$$

3.4 Esecuzione di programmi e proprietà

L'esecuzione di programmi con questa semantica consiste nel trovare una memoria s^\prime tale per cui valga che

$$\langle P, s \rangle \rightarrow^* \langle v, s' \rangle$$

ovvero che si raggiunga una configurazione terminale in un certo numero di passi.

(while) $\frac{-}{\langle \text{while } e \text{ do } e_1, s \rangle \rightarrow \langle \text{if } e \text{ then } (e_1; \text{while } e \text{ do } e_1) \text{ else } skip, s \rangle}$

Illustriamo inoltre due importanti proprietà:

Teorema 3.1 (Strong normalization). Per ogni memoria s e ogni programma P esiste una qualche memoria s' tale che

$$\langle P, s \rangle \rightarrow^* \langle v, s' \rangle$$

Teorema 3.2 (Determinatezza). $Se \langle e, s \rangle \rightarrow \langle e_1, s_1 \rangle e \langle e, s \rangle \rightarrow \langle e_2, s_2 \rangle$ allora $\langle e_1, s_1 \rangle = \langle e_2, s_2 \rangle$.

3.5 Funzione di valutazione della semantica

Date le regole nella sezione 3.3, possiamo dire che in generale, per valutare una porzione di programma, viene applicata la regola

$$\llbracket - \rrbracket : Exp \rightarrow (Store \rightharpoonup Store)$$

¹Problemi con La implicano l'uso di un nome diverso.

dove, data una generica espressione e, la funzione $[\cdot]$ prende una memoria e ne ritorna una aggiornata dopo la valutazione di e.

$$\llbracket e \rrbracket(s) = \begin{cases} s' & \text{se } \langle e, s \rangle \rightarrow \langle e', s' \rangle \\ undefined & \text{altrimenti} \end{cases}$$

3.6 Possibili varianti del linguaggio

Nel linguaggio illustrato possono essere introdotte anche diverse varianti.

3.6.1 Inversione dell'ordine di valutazione

È possibile ad esempio introdurre un ordine di valutazione right-to-left, ossia:

$$(op1b) \frac{\langle e_2, s \rangle \to \langle e'_2, s' \rangle}{\langle e_1 + e_2, s \rangle \to \langle e_1 + e'_2, s' \rangle} \qquad (op2b) \frac{\langle e_1, s \rangle \to \langle e'_1, s' \rangle}{\langle e_1 + v, s \rangle \to \langle e'_1 + v, s' \rangle}$$

Aggiungendo queste due regole alla semantica ovviamente salta la regola della determinatezza.

3.6.2 Regole di assegnamento

Una piccola variante alla regola dell'assegnamento:

$$(\text{assign1b}) \frac{-}{\langle l := n, s \rangle \Rightarrow \langle n, s[l \to n] \rangle} \text{ if } l \in dom(s) \quad (\text{seq.skip.b}) \frac{-}{\langle v; e_2, s \rangle \Rightarrow \langle e_2, s \rangle}$$

3.6.3 Inizializzazione della memoria

Possibili varianti a livello di inizializzazione della memoria potrebbero essere:

- inizializzare implicitamente tutte le locazioni a 0;
- permettere assegnamenti ad una locazione l tale che $l \notin dom(s)$ per inizializzare quella locazione.

3.6.4 Valori memorizzabili

Altre estensioni relative alla memoria (qui definita staticamente, ovvero l'insieme delle locazioni possibili è fisso) possono includere:

- la possibilità di memorizzare anche altri tipi di dato (non solo interi come in questo caso);
- la possibilità di avere una memoria definita dinamicamente, quindi dare la possiblità di avere sempre nuove locazioni disponibili oltre a quelle già in uso.

3.7 Type systems

Un type system è una struttura i cui usi principali sono:

- descrivere quando i programmi sono sensati;
- prevenire certi tipi di errore;
- strutturare i programmi;
- dare delle linee guida per la progettazione del linguaggio;

- dare informazioni utili per la fase di ottimizzazione da parte del compilatore;
- rinforzare alcune proprietà di sicurezza del programma.

Definiamo la funzione

$$\Gamma \vdash e : T$$

che sostanzialmente assegna il tipo T all'espressione e, per qualche tipo T del linguaggio. Aggiungiamo al linguaggio i tipi delle espressioni T e i tipi delle locazioni T_{loc} :

$$T ::= int \mid bool \mid unit$$

$$T_{loc} ::= intref$$

3.7.1 Regole di tipaggio

$$(int) \frac{-}{\Gamma \vdash n : int} \text{ per } n \in \mathbb{Z}$$

$$(bool) \frac{-}{\Gamma \vdash b : bool} \text{ per } b \in \{true, false\}$$

$$(op+) \frac{\Gamma \vdash e_1 : int}{\Gamma \vdash e_1 + e_2 : int}$$

$$(op-geq) \frac{\Gamma \vdash e_1 : int}{\Gamma \vdash e_1 : int} \frac{\Gamma \vdash e_2 : int}{\Gamma \vdash e_1 \ge e_2 : bool}$$

$$(if) \frac{\Gamma \vdash e_1 : bool}{\Gamma \vdash e_1 : bool} \frac{\Gamma \vdash e_2 : T}{\Gamma \vdash if e_1 \text{ then } e_2 \text{ else } e_3 : T}$$

$$(assign) \frac{\Gamma \vdash e : int}{\Gamma \vdash l := e : unit} \text{ se } \Gamma(l) = intref$$

$$(skip) \frac{-}{\Gamma \vdash skip : unit}$$

$$(seq) \frac{\Gamma \vdash e_1 : unit}{\Gamma \vdash e_1 : e_2 : T}$$

$$(while) \frac{\Gamma \vdash e_1 : bool}{\Gamma \vdash e_1 : bool} \frac{\Gamma \vdash e_2 : unit}{\Gamma \vdash while e_1 \text{ do } e_2 : unit}$$

Nota: le regole di tipaggio sono *syntax-directed*, ovvero per ogni regola della sintassi astratta si ha una regola di tipaggio.

3.7.2 Proprietà di tipaggio

Teorema 3.3 (Progress). Se $\Gamma \vdash e : T \ e \ dom(\Gamma) \subseteq dom(s)$ allora $e \ \grave{e}$ un valore oppure esiste una coppia $\langle e', s' \rangle$ tale che

$$\langle e, s \rangle \rightarrow \langle e', s' \rangle$$

Teorema 3.4 (Type preservation). Se $\Gamma \vdash e : T \ e \ dom(\Gamma) \subseteq dom(s) \ e \ \langle e, s \rangle \rightarrow \langle e', s' \rangle$ allora si ha che $\Gamma \vdash e' : T \ e \ dom(\Gamma) \subseteq dom(s')$

Mettendo insieme le due proprietà sopra, si ottiene una nuova proprietà, esplicativa del fatto che programmi ben tipati non vanno mai in deadlock.

Teorema 3.5 (Safety). Se $\Gamma \vdash e : T$, $dom(\Gamma) \subseteq dom(s)$ e $\langle e, s \rangle \rightarrow^* \langle e', s' \rangle$ allora e' è un valore oppure esiste una coppia $\langle e'', s'' \rangle$ tale che $\langle e', s' \rangle \rightarrow \langle e'', s'' \rangle$

Teorema 3.6 (Type inference). Dati Γ , e, può essere trovato il tipo T tale che $\Gamma \vdash e : T$ oppure può essere provato che T non esiste.

Teorema 3.7 (Decidibilità del type-checking). Dati Γ , e, T, è decidibile $\Gamma \vdash e : T$

Teorema 3.8 (Unicità del tipaggio). Se vale che $\Gamma \vdash e : T$ e $\Gamma \vdash e : T'$ allora T = T'.

4 Forme di induzione

L'induzione è una tecnica formale che consente di provare delle proprietà su determinate categorie di oggetti, sfruttando la natura di questi oggetti. Esistono 3 tipi di induzione:

- · matematica:
- strutturale:
- rule induction².

4.1 Induzione matematica

È la forma di induzione più semplice, consiste infatti nel dimostrare una proprietà P(-) su numeri naturali procedendo nel modo seguente:

1. Caso base: provare che P(0) è vera, usando qualche procedimento matematico;

2. Caso induttivo:

- (a) assumere che l'ipotesi induttiva valga, ovvero che valga P(k);
- (b) dall'ipotesi induttiva dimmostrare che vale P(k+1), usando qualche procedimento matematico.

Se i punti precedenti sono veri, allora P(n) è vera per ogni numero naturale.

4.2 Induzione strutturale

4.2.1 Induzione strutturale su numeri naturali

Per dimostrare una proprietà P su numeri naturali basta applicare il seguente metodo:

- Caso base: dimostrare che vale P(0);
- Caso induttivo: dimostrare che è vera P(succ(K)) assumendo come ipotesi induttiva che valga P(K) per qualche $K \in \mathbb{N}$.

L'induzione strutturale consiste quindi nell'assumere che l'ipotesi induttiva valga per la sottostruttura di succ(K).

4.2.2 Induzione strutturale su strutture complesse

Prendiamo come esempio la costruzione di alberi binari. Diamo la seguente grammatica per costruire gli alberi:

$$T ::= leaf \mid tree(T, T)$$

In tal caso partiamo col presupposto che:

- Caso base: una foglia sia un albero binario;
- Caso induttivo: se L e R sono alberi binari, allora lo è anche tree(L,R).

²Appena avrò una traduzione valida la metterò.

4.3 Rule induction

L'idea di base della rule induction consiste nell'ignorare la struttura di ciò che si deriva per fare induzione sulla dimensione dell'albero di derivazione.

Per provare formalmente una proprietà P(D) su una derivazione D, si procede come segue:

- 1. **Caso base:** dimostrare che P(A) è vera, per ogni assioma A;
- 2. Caso induttivo: per ogni regola della forma

(regola)
$$h_1 h_2 \dots h_n$$

si dimostra che ogni derivazione che termina con l'utilizzo di questa regola soddisfa la proprietà. Questa derivazione ha sottoderivazioni $D_1, D_2, ..., D_n$ che terminano con le ipotesi $h_1, h_2, ..., h_n$. Per ipotesi induttiva si assume che valga $P(D_i)$ con $1 \le i \le n$.

5 Aspetti funzionali

Estendiamo la sintassi data in 3 con:

Variabili
$$x \in \mathbb{X}$$
, con $x = \{x, y, z, ...\}$
Espressioni $e ::= ... \mid \text{fn } x : T \Rightarrow e \mid ee \mid x$

e i tipi con:

$$T \; ::= \; \dots \mid T \to T$$

Assumiamo che:

- l'applicazione di funzioni ee è associativa a sinistra;
- i tipi delle funzioni sono associativi a destra;
- il corpo di fn si estende a sinistra quanto più è possibile;
- $fn \ x : unit \Rightarrow fn \ y : int \Rightarrow x; y \ e \ di \ tipo \ unit \rightarrow int \rightarrow int.$

5.1 Variabili free e bound

Intuitivamente, diciamo che una variabile è *free* in e se x non occorre in nessun termine della forma fn $x:T \Rightarrow ...$

Più formalmente definiamo le variabili free come una funzione:

$$fv(): Exp \to 2^{\mathbb{X}}$$

dove \mathbb{X} è l'insieme delle variabili. La funzione fv è definita come:

$$fv(x) \triangleq \{x\}$$

$$fv(\texttt{fn } \texttt{x} : \texttt{T} \Rightarrow e) \triangleq fv(e) \setminus \{x\}$$

$$fv(e_1e_2) = fv(e_1; e_2) \triangleq fv(e_1) \cup fv(e_2)$$

$$fv(n) = fv(b) = fv(!l) = fv(\texttt{skip}) \triangleq \emptyset$$

$$fv(e_1 \texttt{op} e_2) = fv(\texttt{while } e_1 \texttt{ do } e_2) \triangleq fv(e_1) \cup fv(e_2)$$

$$fv(l : =e) \triangleq fv(e)$$

$$fv(\texttt{if } e_1 \texttt{ then } e_2 \texttt{ else } e_3) \triangleq fv(e_1) \cup fv(e_2) \cup fv(e_3)$$

Definizione 5.1. Un'espressione e si dice **chiusa** se $fv(e) = \emptyset$.

5.2 Alpha conversion

Nell'espressione fin x : $T \Rightarrow e$ la variabile x è bound in e. Diciamo che:

- x è il parametro formale della funzione, quindi ogni occorrenza di x che non è in una funzione annidata alla funzione attuale indica la stessa cosa;
- al di fuori della definizione della funzione, la variabile x non ha significato;

Basandoci su quanto detto nei due punti precedenti, possiamo concludere che il nome del parametro formale nella funzione **non modifica** il comportamento della funzione. Assumiamo quindi di poter rimpiazzare il vincolo su x e tutte le occorrenze di x in una certa espressione e con una variabile fresh che non occorra da nessun 'altra parte.

5.3 Sostituzioni

5.3.1 Sostituzioni singole

Un perno della semantica delle funzioni è dato dalle sostituzioni, ovvero il rimpiazzo a runtime di un parametro attuale con un parametro formale. Una sostituzione è indicata con

$$e_2\{e_1/x\}$$

e indica la sostituzione di x con e_2 nell'espressione e_1 , per tutte le occorrenze libere di x. Le sostituzioni funzionano come segue:

$$n\{e/x\} \triangleq n$$
 $b\{e/x\} \triangleq b$
 $\mathrm{skip}\{e/x\} \triangleq \mathrm{skip}$
 $x\{e/x\} \triangleq e$
 $y\{e/x\} \triangleq y$

$$(\mathrm{fn} \ \mathbf{x}: \mathbf{T} \Rightarrow e_1) \{e/z\} \triangleq (\mathrm{fn} \ \mathbf{x}: \mathbf{T} \Rightarrow e_1 \{e/z\}) \mathrm{se} \ x \notin fv(e)$$
 $(\mathrm{fn} \ \mathbf{x}: \mathbf{T} \Rightarrow e_1) \{e/z\} \triangleq (\mathrm{fn} \ \mathbf{x}: \mathbf{T} \Rightarrow (e_1 \{e/z\}) \{e/z\}) \mathrm{se} \ x \in fv(e) \land y \ \mathrm{\acute{e}} \ \mathrm{fresh}^3$
 $(\mathrm{fn} \ \mathbf{x}: \mathbf{T} \Rightarrow e_1) \{e/x\} \triangleq (\mathrm{fn} \ \mathbf{x}: \mathbf{T} \Rightarrow e_1)$
 $(e_1; e_2) \{e/x\} \triangleq (e_1 \{e/x\} e_2 \{e/x\})$

Le altre espressioni dopo una sostituzione rimangono invariate per il semplice motivo che non sono funzioni, quindi non va sostituito alcun parametro.

5.3.2 Sostituzioni simultanee

Le sostituzioni possono essere implementate in modo da poter essere anche simultanee, ovvero con lo scopo di poter sostituire più variabili contemporaneamente. In generale, una sostituzione simultanea è una funzione parziale $\sigma: \mathbb{X} \to Exp$.

Sintatticamente viene indicato con $e\sigma$ l'espressione risultante dalla sostituzione simultanea di ogni $x \in dom(\sigma)$ con la corrispondente espressione $\sigma(x)$. La sostituzione dei parametri viene indicata con

$$\{e_1/x_1,...,e_k/x_k\}$$

³Operando con alpha-conversion.

- 6 Dati e memoria variabile
- 7 Sotto-tipaggio