

Бреслав А. А.

**Синтез механизмов композиции в предметно-ориентированных  
языках**

**Оглавление**

<b>Введение</b> . . . . .	2
§ 1. Постановка задачи . . . . .	2
<b>Глава 1. Предварительные сведения</b> . . . . .	3
§ 1. Объектно-ориентированные модели . . . . .	3
1.1. Синтаксис моделей . . . . .	4
1.2. Классы и типы . . . . .	12
1.3. Модели и мета-модели . . . . .	21
§ 2. Языки как множества моделей . . . . .	27
2.1. Нотация для контекстно-свободных грамматик . . . . .	28
2.2. Метамодел, эквивалентные грамматикам . . . . .	30
<b>Глава 2. Расширение предметно-ориентированных языков механизмами композиции</b> . . . . .	33
§ 1. Автоматическое построение языков, поддерживающих ти- пизированные макроопределения . . . . .	33
1.1. Неформальное описание механизма композиции, основан- ного на макроопределениях . . . . .	34
1.2. Шаблоны в языках, порожденных метамоделами . . . . .	37
1.3. Базовый язык шаблонов . . . . .	42
<b>Приложение 1. Нотация для контекстно-свободных грамматик</b>	51
<b>Список литературы</b> . . . . .	53

## Введение

### § 1. Постановка задачи

Целью настоящей работы является **автоматизация разработки механизмов композиции в предметно-ориентированных языках**. Для достижения поставленной цели необходимо решить следующие задачи:

- Разработать автоматизированный метод расширения предметно-ориентированных языков поддержкой типизированных макроопределений.
- Разработать автоматизированный метод расширения предметно-ориентированных языков поддержкой аспектов.
- Формально описать разработанные методы и обосновать их корректность.
- Продемонстрировать применение разработанных методов, применив их к предметно-ориентированному языку, полезному для решения практических задач.

## Глава 1.

### Предварительные сведения

В этой главе...

#### § 1. Объектно-ориентированные модели

Объектно-ориентированные модели (для краткости, ниже мы будем писать просто “модели”) как средство описания предметных областей были предложены в начале 1990-х годов [?, ?] и с течением времени были описаны консорциумом OMG (Object Management Group, [?]) в виде ряда промышленных стандартов. Наибольшую известность среди “модельно-ориентированных” подходов получил унифицированный язык моделирования — UML (Unified Modelling Language, [13]); также очень широко используется библиотека EMF (Eclipse Modelling Framework, [14]), основанная на идеях, заложенных в спецификации MOF (Meta-Object Facility, [54]), являющейся частью UML (Infrastructure). По сути, объектно-ориентированные модели — это структуры, не привязанные к конкретному языку программирования и представляющие данные (и способы их обработки) в объектно-ориентированном стиле.

Следуя работам [?, 83], в последующих разделах мы будем использовать модели как универсальный способ описания предметно-ориентированных языков, однако на этом пути нас ждет небольшое затруднение: промышленные стандарты, упомянутые выше, не предоставляют математической базы для работы с моделями. В литературе описаны различные подходы к формализации этой области [?, ?, ?], в основном нацеленные на описание полной MOF. Использование этих формализаций затруднительно в первую очередь потому, что MOF —

это довольно объемная система, а следовательно, и формальные теории, ее описывающие, получаются достаточно громоздкими. С другой стороны, на практике, при реализации предметно-ориентированных языков, как правило, используется не вся MOF, а лишь небольшая ее часть, “ядро”, которое называется ЕМОФ (Essential MOF, [?]), именно ей соответствует реализация моделей, предоставляемая популярной библиотекой EMF, упомянутой выше. В данном разделе строится формальная теория, описывающая объектно-ориентированные модели, соответствующие ЕМОФ.

### 1.1. Синтаксис моделей

Чтобы строго определить понятие модели, мы вводим ряд вспомогательных определений, начиная со структурного представления и постепенно вводя ограничения.

**Определение 1.1** (Модельный терм). *Модельным термом* (MT) называется формула, построенная по следующим правилам<sup>1</sup>:

$$\begin{aligned}
 \text{MT} &::= \text{object } \text{MT}_{id} : \text{MT}_{cl} \{P_1, \dots, P_n\}, n \geq 0 \\
 &\quad | \text{@MT}_{id} \\
 &\quad | [\text{MT}_1, \dots, \text{MT}_n], n \geq 0 \\
 &\quad | \{\text{MT}_1, \dots, \text{MT}_n\}, n \geq 0 \\
 &\quad | \text{null} \mid \Sigma^* \mid \mathbb{Z} \mid \mathbb{B}, \\
 P &::= \text{MT}_p = \text{MT}_v
 \end{aligned}$$

где  $\mathbb{Z}$  — множество целых чисел,  $\Sigma$  — некоторый конечный алфавит символов, а  $\mathbb{B} = \{\text{true}, \text{false}\}$ . Вспомогательное правило для  $P$  описывает *свойства объектов* (см. ниже).

<sup>1</sup> Для краткости мы пишем  $t_1, \dots, t_n, n \geq 0$  вместо использования рекурсивных продукций вида

$$tList ::= \varepsilon \mid t \mid t, tList$$

Записывая модельные термы, мы, как правило, не будем заключать строки в кавычки, а будем лишь выделять их начертанием, чтобы отличать от *метапеременных*: так, `abc` — это строка из трех символов, а  $x$  — метапеременная, которая, в частности может принимать значение `abc`.

Для различных видов модельных термов мы будем использовать специальные названия. Приведем примеры таких термов и их названия:

- `{abc, 239, null}` — *множество*, состоящее из трех констант: строки, числа и специального значения `null`;
- `[1, 2, 3]` — *список* из трех чисел;
- `@x` — *ссылка* на идентификатор  $x$ .
- `object a : @b {c = false}` — *объект*, где  $a$  — *идентификатор*, `@b` — *ссылка на класс*, а `c = false` — *свойство с маркером c и значением false*.

*Метапеременные* позволяют кратко описать множества термов, обладающих схожей структурой. Если в модельном терме используется метапеременная  $a$ , это означает, что вместо нее может быть подставлен любой подтерм, синтаксически допустимый в данном контексте. Так, например, запись `{a, b}` описывает все термы-множества, состоящие из двух элементов, а `@a` — ссылки на все возможные идентификаторы.

Модельные термы удобны для представления структур абстрактного синтаксиса различных языков [?, 83]. В качестве примера рассмотрим  $\lambda$ -исчисление [?]. Абстрактный синтаксис  $\lambda$ -исчисления

задается тремя конструкторами:

Обозначение	Пример	Название
$LT ::= \text{Abs}(v, LT)$	$\lambda x . t$	<i>абстракция</i>
$  \text{App}(LT_1, LT_2)$	$(f x)$	<i>применение</i>
$  \text{Var}(v)$	$x$	<i>переменная</i>

где  $v$  пробегает алфавит имен переменных. Например,  $\lambda$ -терм  $\lambda x . x x$  в абстрактном синтаксисе записывается как  $\text{Abs}(x, \text{App}(\text{Var}(x), \text{Var}(x)))$ , что соответствует *абстрактному синтаксическому дереву* (Abstract Syntax Tree, AST [?]) для данного терма. То же самое дерево можно представить также в виде объектов; см. Рис. 1.1.

```
object t1 : @Abstraction {
  var = object x : @Variable {name = "x"},
  body = object t2 : @Application {
    function = object t3 : @Usage {var = @x},
    argument = object t4 : @Usage {var = @x}
  }
}
```

Рис. 1.1: Представление AST для терма  $\lambda x . x x$  в виде объектов

Из Рис. 1.1 видно, что объекты модельного терма образуют дерево, соответствующее абстрактному синтаксическому представлению  $\lambda$ -терма, причем роль конструкторов выполняют ссылки на классы. Обратите внимание на использование ссылок в объектах t3 и t4: значение свойства `var` является *ссылкой* на идентификатор объекта, соответствующего определению переменной  $x$ .

Модельный терм можно рассматривать как граф: объекты являются вершинами, а отношение вложенности и ссылки — ребрами. AST является остовным деревом в этом графе — в нем участвуют только ребра вложенности объектов.

### 1.1.1. Графическая нотация для модельных термов

Наравне с нотацией, введенной выше, мы будем использовать для изображения объектов графическую нотацию, основанную на диаграммах объектов языка UML [13]. На Рис. 1.2 в виде такой диаграммы изображены объекты, приведенные на Рис. 1.1.

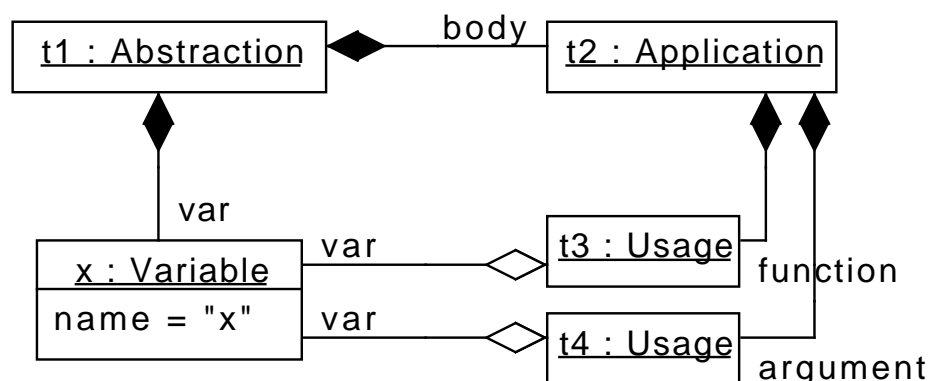


Рис. 1.2: Графическое представление для терма  $\lambda x . x x$

Объекты обозначаются прямоугольниками. В верхней части прямоугольника располагается подчеркнутый текст: это идентификатор объекта и *имя класса*. О классах подробно говорится ниже, а пока лишь заметим, что объекты, относящиеся к одному классу, имеют одинаковую структуру. Значения свойств изображаются либо под горизонтальной чертой в самом прямоугольнике (как для объекта  $x$ ), либо в виде ребер графа. Ближе к концу ребра написано имя свойства, а в начале ребра изображается ромб: незакрашенный, если ребро соответствует ссылке<sup>2</sup>, и закрашенный, если значением свойства является объект, то есть имеет место *встраивание*<sup>3</sup> — один объект является частью другого. Таким образом, если рассматривать только ребра с закрашенными ромбами, из графа объектов выделяется дерево, соответствующее текстовой нотации, введенной выше (см. Рис. 1.1).

<sup>2</sup>Что соответствует *агрегации* в терминах UML.

<sup>3</sup>*Композиция* — в терминах UML

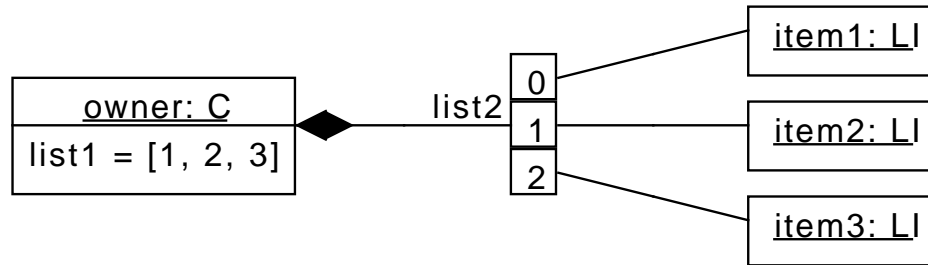


Рис. 1.3: Графическое представление упорядоченных списков

Если значением свойства является множество, оно изображается несколькими ребрами с одинаковой пометкой. В случае списка важен порядок следования элементов, поэтому мы отклоняемся от нотации UML и вводим промежуточный блок с индексами, от ячеек которого отходят ребра к элементам списка. Пример использования данной нотации приведен на Рис. 1.3: список чисел `list1` записан непосредственно в теле объекта, а список (встраиваемых) объектов `list2` изображен графически.

### 1.1.2. Конгруэнтность и правильно построенные термы

**Определение 1.2** (Конгруэнтность). Отношение *конгруэнтности* на модельных термах,  $\cong \subset \text{MT} \times \text{MT}$ , есть минимальное отношение эквивалентности, обладающее следующими свойствами:

1. Константы конгруэнтны сами себе:

$$x \cong x, \text{ для любого } x \in \{\text{null}\} \cup \mathbb{Z} \cup \mathbb{B} \cup \Sigma^*;$$

2. Ссылки на конгруэнтные идентификаторы конгруэнтны:

$$@x \cong @y, \text{ если } x \cong y;$$

3. Списки сравниваются поэлементно:

$$[x_1, \dots, x_n] \cong [y_1, \dots, y_n], \text{ если } x_i \cong y_i \text{ для любого } i \in [1..n];$$



4. Множества сравниваются поэлементно, без учета порядка:

$\{x_1, \dots, x_n\} \cong \{x'_{\pi(1)}, \dots, x'_{\pi(n)}\}$ , если существует перестановка  $\pi$  размера  $n$  такая, что  $x_i \cong x'_{\pi(i)}$  при  $i \in [1; n]$ .

5. Объекты сравниваются без учета порядка свойств:

$$\mathbf{object} \ id_1 : c_2 \left\{ \begin{array}{l} p_1 = v_1, \\ \dots, \\ p_n = v_n \end{array} \right\} \cong \mathbf{object} \ id_2 : c_2 \left\{ \begin{array}{l} p'_{\pi(1)} = v'_{\pi(1)}, \\ \dots, \\ p'_{\pi(n)} = v'_{\pi(n)} \end{array} \right\},$$

если  $id_1 \cong id_2$ , а также  $c_1 \cong c_2$ , и существует перестановка  $\pi$  размера  $n$  такая, что  $p_i \cong p'_{\pi(i)}$  и  $v_i \cong v'_{\pi(i)}$ .

Приведем несколько примеров:

- $@abc \cong @abc$ , поскольку идентификаторы конгруэнтны;
- $[1, 2] \not\cong [2, 1]$ , при сравнении списков порядок элементов важен;
- а при сравнении множеств — не важен:  $\{1, 2\} \cong \{2, 1\}$ ;
- также не важен порядок свойств при сравнении объектов:

$$\mathbf{object} \ abc : @c \left\{ \begin{array}{l} a = b \\ c = d \end{array} \right\} \cong \mathbf{object} \ abc : @c \left\{ \begin{array}{l} c = d \\ a = b \end{array} \right\},$$

- но значения свойств важны:

$$\mathbf{object} \ abc : @c \left\{ \begin{array}{l} a = b \\ c = d \end{array} \right\} \not\cong \mathbf{object} \ abc : @c \left\{ \begin{array}{l} c = \textcolor{red}{b} \\ a = \textcolor{red}{d} \end{array} \right\}.$$

Введем функцию  $\mathcal{S}(t)$ , вычисляющую множество всех поддер-  
мов терма  $t$ :

$$\mathcal{S}(t) \stackrel{\text{def}}{=} \{t\} \cup \left\{ \begin{array}{ll} \mathcal{S}(id) \cup \mathcal{S}(c) \cup \bigcup_i \mathcal{S}(p_i) \cup \bigcup_i \mathcal{S}(v_i), & t = \mathbf{object} \ id : c \{p_1 = v_1, \dots, p_n = v_n\} \\ \mathcal{S}(r), & t = @r \\ \bigcup_i \mathcal{S}(x_i), & t = [x_1, \dots, x_n] \text{ или } t = \{x_1, \dots, x_n\} \\ \emptyset, & \text{в остальных случаях} \end{array} \right. \quad (1.1)$$

Пример:  $\mathcal{S}(\text{object } a : @b \{1 = [\text{null}, \text{false}]\}) = \{$   
 $a, \quad @b, \quad b, \quad 1,$   
 $[\text{null}, \text{false}], \quad \text{null}, \quad \text{false}$   
 $\text{object } a : @b \{1 = [\text{null}, \text{false}]\}\}.$

Как правило, нам будут нужны множества подтермов какого-то определенного вида, например, множество подтермов термина  $t$ , являющихся ссылками, мы будем обозначать  $\mathcal{S}_{ref}(t)$ . Пример:

$$\mathcal{S}_{ref}(\text{object } a : @b \{1 = [\text{null}, \text{false}]\}) = \{@b\}.$$

Аналогично  $\mathcal{S}_{set}(t)$  — для подтермов-множеств и  $\mathcal{S}_{obj}(t)$  — для подтермов-объектов.

**Определение 1.3.** Модельный терм  $t$  называется *правильно построенным*, если выполняются следующие условия:

1. Среди подтермов  $t$  не существует двух объектов с конгруэнтными идентификаторами;
2. В  $\mathcal{S}_{obj}(t)$  не существует объекта, содержащего два свойства с конгруэнтными именами;
3. В  $\mathcal{S}_{set}(t)$  не существует множества, содержащего два конгруэнтных элемента.

Так, например, терм  $[\text{object } a : @c \{\}, \text{object } a : @c \{x = y\}]$  не является правильно построенным (пункт 1). Терм  $\text{object } a : @c \{a = b, a = c\}$  тоже не является правильно построенным, но по другой причине см. пункт 2. Согласно пункту 3, терм  $\{1, 2, 1\}$  также не является правильно построенным.

### 1.1.3. Ссылочные контексты и пред-модели

**Определение 1.4** (Ссылочный контекст). Будем называть *ссылочным контекстом* частичную функцию  $\Psi : \text{MT} \rightarrow \text{MT}$ , обладающую следующими свойствами:

1. Если  $\Psi(id)$  определено, то  $\Psi(id) = \{\mathbf{object} \ id' : \dots \{\dots\}\}$ , где  $id \cong id'$ ;
2.  $\Psi(id) \cong \Psi(id')$  тогда и только тогда, когда  $id \cong id'$ .

Другими словами, ссылочный контекст “находит” объект по его идентификатору. Для удобства мы будем писать  $\Psi(id) = \perp$  в случае, если  $\Psi(id)$  не определено.

Обозначим множество всех подтермов, *упомянутых* в контексте  $\Psi$ , то есть являющихся подтермами элементов области определения  $\Psi$  и/или области значений  $\Psi$ , через  $\mathcal{S}(\Psi)$ . Аналогично вводятся  $\mathcal{S}_{ref}(\Psi)$  и  $\mathcal{S}_{obj}(\Psi)$ .

**Определение 1.5** (Замкнутый контекст). Будем называть ссылочный контекст  $\Psi$  *замкнутым*, если

$$o = \mathbf{object} \ id : \dots \{\dots\} \in \mathcal{S}_{obj}(\Psi) \Rightarrow \Psi(id) = o$$

и

$$r = @id \in \mathcal{S}_{ref}(\Psi) \Rightarrow \Psi(id) \neq \perp$$

Все объекты, упомянутые в замкнутом контексте, могут быть получены по своим идентификаторам, и каждая ссылка ссылается на существующий объект. Из этого, в частности, следует, что в таком контексте каждому объекту присвоен уникальный идентификатор. Так, например, следующий контекст является замкнутым (здесь функция рассматривается как множество пар вида “аргумент  $\mapsto$  результат”):

$$\Psi(id) = \{a \mapsto \mathbf{object} \ a : @a \{\}\}$$

Приведем также пример незамкнутого контекста:

$$\Psi(id) = \{a \mapsto \mathbf{object} \ a : @b \{c = \mathbf{object} \ b : @a \{\}\}\}$$

в последнем примере нарушаются оба требования определения 1.5:

- объект, упомянутый в контексте, не принадлежит области его значений;
- ссылка, упомянутая в контексте, не принадлежит его области определения.

**Определение 1.6.** Пусть  $\Psi_1$  и  $\Psi_2$  – согласованные ссылочные контексты:

$$\forall x \in \mathbf{dom}(\Psi_1), y \in \mathbf{dom}(\Psi_2) : x \cong y \Rightarrow \Psi_1(x) \cong \Psi_2(y).$$

Тогда их *объединение* определяется следующей формулой:

$$(\Psi_1 \cup \Psi_2)(id) \stackrel{\text{def}}{=} \begin{cases} \Psi_1(id), & \Psi_1(id) \neq \perp \\ \Psi_2(id), & \Psi_2(id) \neq \perp \\ \perp, & \text{в противном случае.} \end{cases}$$

По модельному терму  $t$  можно построить ссылочный контекст, упоминающий все объекты, входящие в  $t$ :

$$\Psi_t \stackrel{\text{def}}{=} \{id \mapsto o \mid o = \mathbf{object} \ id : \dots \{ \dots \} \in \mathcal{S}_{obj}(t)\}$$

**Определение 1.7** (Пред-модель). Будем называть правильно построенный модельный терм  $t$  *пред-моделью в контексте*  $\Psi_0$ , если контекст  $\Psi_0 \cup \Psi_t$  является замкнутым.

Пред-модель обладает важным свойством: любой объект в ней однозначно определяется своим идентификатором, и любая ссылка указывает на известный объект.

## 1.2. Классы и типы

На практике для удобства навигации по моделям и контроля над их корректностью вводятся дополнительные ограничения (в виде системы типов), регулирующие структуру моделей. В этом разделе мы вводим понятие *класса* для объектов, которое является основой для таких ограничений.

**Определение 1.8** (Класс). *Классом* называется кортеж  $\langle abs, id, S, P \rangle$ , где

1.  $abs$  — признак абстрактного класса, принимает значения `true` или `false`; Если  $abs = \text{false}$ , класс называется *конкретным*, иначе — *абстрактным*,
2.  $id$  — *идентификатор* класса (имя),
3.  $S$  — множество классов, называемых *предками* данного класса (*суперклассов*),
4.  $P$  — множество пар вида  $p : \tau$ , где  $p$  — *идентификатор* свойства, а  $\tau$  — *тип* свойства (множество типов определяется ниже).

Множество  $P$  не содержит различных пар, в которых совпадают идентификаторы свойств.

Для удобства, мы будем обозначать конкретный класс  $\langle \text{false}, id, S, P \rangle$  через `class id : S {P}`, а абстрактный класс  $\langle \text{true}, id, S, P \rangle$  — через `class id : S {P}`. Когда значения признака абстрактного класса не важно или однозначно определяется из контекста, мы будем использовать обозначение `class id : S {P}`.

**Определение 1.9.** *Перечислением* называется кортеж  $\langle id, \{L_1, \dots, L_n\} \rangle$ , где  $id$  — *идентификатор* перечисления, а  $L_i$  — *литералы* перечисления.

Мы будем обозначать перечисления следующим образом: `enum id {L1, ..., Ln}`.

**Определение 1.10.** Множество типов  $\mathcal{T}$  описывается следующим об-

разом:

$$\begin{aligned} \mathcal{T} ::= & \text{Char} \mid \text{String} \mid \text{Int} \mid \text{Bool} \\ & \mid \mathcal{T}^? \mid \{\mathcal{T}^+\} \mid \{\mathcal{T}^*\} \mid [\mathcal{T}^+] \mid [\mathcal{T}^*] \\ & \mid \text{val}(c) \mid \text{ref}(c), \text{ где } c \text{ — идентификатор некоторого класса} \\ & \mid \text{enum}(e), \text{ где } e \text{ — идентификатор некоторого перечисления} \\ & \mid \top \end{aligned}$$

Неформально говоря, приведенное определение описывает следующие разновидности типов:

- Char, String, Int, Bool обозначают примитивные типы: символы, строки, целые числа и булевские значения, соответственно.
- Тип  $\tau^?$  допускает значение null или значение типа  $\tau$ , например,  $\text{Char}^?$  соответствует множеству значений  $\Sigma \cup \{\text{null}\}$ .
- $\{\tau^*\}$  обозначает тип множеств, содержащих ноль или более элементов типа  $\tau$ . Например,  $\{\text{String}^*\}$  — это множества строк, а  $\{\{\text{Int}^*\}^*\}$  — множества множеств целых чисел.
- Аналогично,  $[\tau^*]$  обозначает списки.
- Если вместо “\*” используется “+”, то множество или список должны содержать не менее одного элемента. Например,  $\{\tau^+\}$  — непустые множества, составленные из элементов  $\tau$ .
- $\text{val}(c)$ , где  $c$  — некоторый класс (точнее, его идентификатор), соответствует *встраиваемому* объекту класса  $c$ . Напомним, *встраиванием* называется ситуация, при которой один объект входит в другой как подтерм, например:  $\text{object } a : @c \{a = \text{object } b : @c \{\}\}$ .
- $\text{ref}(c)$  обозначает тип ссылок на объекты класса  $c$ , то есть значений вида  $@a$ , где  $a$  — идентификатор объекта, принадлежащего классу  $c$  (см. ниже).

- $\text{enum}(e)$ , где перечисление  $e$  определяется как  $\text{enum } e \{L_1, \dots, L_n\}$  обозначает тип, значениями которого являются литералы  $L_1, \dots, L_n$  и только они.
- $\top$  (“Top”) — супертип всех остальных типов (см. ниже).

### 1.2.1. Графическая нотация для классов

Наравне с текстовой, мы будем использовать для классов графическую нотацию, основанную на диаграммах классов языка UML (см. Рис. 1.4). Класс обозначается прямоугольником, разделенным на две секции: верхняя секция содержит имя класса (*курсивом* в случае абстрактного класса), а нижняя — его свойства.

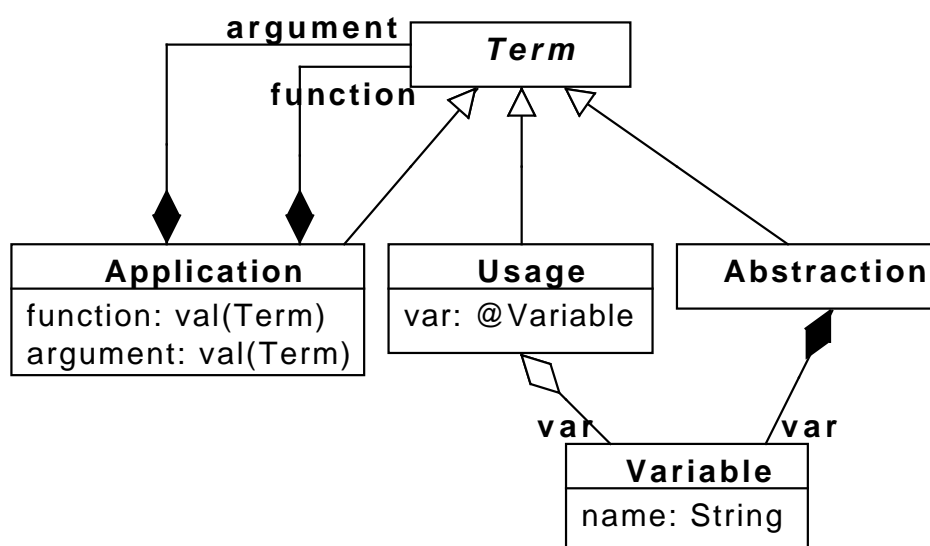


Рис. 1.4: Графическая нотация для классов  $\lambda$ -исчисления

Для обозначения отношения “подкласс-суперкласс” или *наследования* на диаграммах используются ребра с треугольником на конце, обозначающим суперкласс. Заметим, что циклов по наследованию и встраиванию на диаграмме классов быть не должно.

Как и в случае с объектами, некоторые свойства могут обозначаться ребрами (вместо или одновременно с записью в теле класса). Начало ребра обозначается закрашенным или незакрашенным ромбом — для встраивания ( $\text{val}(\tau)$ ) и ссылок ( $\text{ref}(\tau)$ ) соответственно, а имя свойства пишется ближе к концу ребра, причем для обозначения множеств и списков используется нотация, проиллюстрированная на Рис. 1.5. Для обозначения свойства  $\text{bset} : \{\text{val}(B)^*\}$  на диаграм-



Рис. 1.5: Нотация для множеств и списков на диаграмме классов

ме изображается ребро от класса  $A$ , в котором объявлено свойство, к классу  $B$ , ближе к концу ребра пишется имя свойства —  $\text{bset}$  заключенное в фигурные скобки вместе со знаком “\*”, обозначающие что тип свойства — множество объектов класса  $B$ . На Рис. 1.5 начало ребра, изображающего свойство  $\text{bset}$  отмечено закрашенным ромбом, следовательно элементы множества встраиваются в объекты класса  $A$ , что завершает описание типа свойства —  $\{\text{val}(B)^*\}$ . Аналогично, для обозначения списка  $\text{clist} : [\text{ref}(C)^+]$  начало ребра отмечено незакрашенным ромбом, а в конце ребра мы пишем  $[\text{clist}+]$ .



Рис. 1.6: Пример графической нотации для перечислений



Перечисления также отображаются в виде прямоугольников (см. Рис. 1.6), но над идентификатором пишется *стереотип* «enumeration», а под чертой перечисляются литералы.

### 1.2.2. Типы как множества значений

Для типа  $\tau$ , множество упомянутых в данном типе классов определяется следующим образом:

$$\overline{C}(\tau) \stackrel{\text{def}}{=} \begin{cases} \{c\}, & \tau \in \{\mathbf{val}(c), \mathbf{ref}(c)\} \\ \overline{C}(\sigma), & \tau \in \{[\sigma^+], [\sigma^*], \{\sigma^+\}, \{\sigma^*\}, \sigma^?\} \\ \emptyset, & \text{в остальных случаях} \end{cases}$$

Пусть дано множество классов

$$C = \{\mathbf{class} \ id_i : S_i \{p_j^i : \tau_j^i \mid j \in [1; m_i]\} \mid i \in [1; n]\},$$

обозначим  $\overline{S}(C) \stackrel{\text{def}}{=} C \cup \overline{S}(\bigcup_1^n S_i)$  множество всех классов из  $C$ , их предков, предков их предков и т.д.

Аналогично определяется множество всех свойств класса  $c = \mathbf{class} \ id : S \{p_i : \tau_i\}$ . Кроме свойств  $p_i$ , объявленных в самом  $c$ , рассматриваются также свойства, *унаследованные* из суперклассов:  $\overline{P}(c) \stackrel{\text{def}}{=} \bigcup_{i=1}^n \{p_j^i : \tau_j^i \mid j \in [1; m_i]\} \cup \overline{P}(\overline{S}(c))$ . Определим также множество всех классов, встраиваемых в  $c$ :

$$\overline{C}_{comp}(c) \stackrel{\text{def}}{=} \{c' \mid \exists (p : \tau) \in \overline{P}(c), \text{ где } \tau \in \{\mathbf{val}(c'), \{\mathbf{val}(c')^+\}, [\mathbf{val}(c')^+]\}\}.$$

**Определение 1.11.** Конечный набор классов  $C$  называется *корректным*, если выполняются следующие свойства:

1. Все предки классов из  $C$  содержатся в  $C$ :  $\overline{S}(C) = C$ .
2. Все типы свойств классов из  $C$  упоминают только классы из  $C$ : пусть  $c \in C$  и  $\overline{P}(c) = \{p_i : \tau_i\}$ , тогда  $\bigcup_i \overline{C}(\tau_i) \subseteq C$ .

3. Никакой класс из  $C$  не является прямо или косвенно предком самого себя: если  $c = \text{class } id : S \{ \dots \} \in C$ , то  $c \notin \bar{S}(S)$ .
4. Никакой класс из  $C$  прямо или косвенно не встраивается сам в себя:  $c \notin \bar{C}_{comp}(c)$ .

На типах, порожденных корректным набором классов  $C$ , вводится отношение “подтип” ( $\preceq$ ) — это наименьшее транзитивное рефлексивное отношение, обладающее следующими свойствами<sup>4</sup>:

$$\frac{\text{class } c : S \{ \dots \} \in C \quad s \in S}{\text{val}(c) \preceq \text{val}(s) \wedge \text{ref}(c) \preceq \text{ref}(s)} \text{ subclass}$$

Другими словами, подклассы порождают подтипы. Кроме того, тип  $\tau^?$  шире типа  $\tau$ , поскольку свойства этого типа могут принимать на одно значение — `null` — больше<sup>5</sup>:

$$\frac{}{\tau \preceq \tau^?} \text{ nullable}$$

Аналогично для списков и множеств:

$$\frac{\tau \preceq \sigma \quad \iota, \kappa \in \{+, *\} \quad \iota \leq \kappa}{\{\tau^\iota\} \preceq \{\sigma^\kappa\}} \text{ set} \quad \frac{\tau \preceq \sigma \quad \iota, \kappa \in \{+, *\} \quad \iota \leq \kappa}{[\tau^\iota] \preceq [\sigma^\kappa]} \text{ list}$$

Здесь отношение порядка на знаках  $+$  и  $*$  определяется правилом

$$+ < *.$$

Классы и типы используются для описания множеств допустимых модельных термов. Пусть зафиксирован замкнутый ссылочный контекст  $\Psi$ . Ниже мы определим функцию  $\llbracket \bullet \rrbracket_\Psi^C : \mathfrak{T} \rightarrow 2^{\text{MT}}$ , которая для корректного набора классов  $C$  сопоставляет каждому типу множество его значений. Для этого мы сначала введем вспомогательную функцию  $\lceil \bullet \rceil_\Psi : \mathfrak{T} \rightarrow 2^{\text{MT}}$ , которая возвращает все модельные

<sup>4</sup>Здесь и далее, следуя нотации изображения правил вывода в математической логике, мы записываем *посылки* над горизонтальной чертой, а *заключение* — под чертой; справа курсивом указывается имя или номер, на который мы будем ссылаться в тексте.

<sup>5</sup>Правила, не имеющие посылок над горизонтальной чертой, являются *аксиомами*, то есть выполняются для любых значений метапеременных.

термы, упомянутые в контексте  $\Psi$ , принадлежащие к данному типу и ни к какому другому, то есть не согласует отношение “подтип” с отношением вложенности множеств:  $\tau \preceq \rho \not\Rightarrow \lceil \tau \rceil_\Psi \subseteq \lceil \rho \rceil_\Psi$ . Эта функция задается следующим образом. Типу  $\text{val}(c)$  ставится в соответствие множество  $\lceil \text{val}(c) \rceil_\Psi$  всех объектов из  $\mathcal{S}_{obj}(\Psi)$ , в которых ссылкой на класс является терм  $@_c$ , и имеющих все свойства из множества  $\overline{P}(c)$  и только их, причем значение каждого свойства имеет тип, указанный для данного свойства в классе:

$$\frac{\text{class } c : S \{ \dots \} \quad \overline{P}(c) = \{p_1 : \tau_1, \dots, p_n : \tau_n\}}{\lceil \text{val}(c) \rceil_\Psi = \{ \text{object } id : @_c \{p_i = v_i\} \in \mathcal{S}_{obj}(\Psi) \mid v_i \in \llbracket \tau_i \rrbracket_\Psi^C \}} \text{ objects}$$

Другие типы характеризуются следующим образом:

$$\frac{\text{class } c : S \{ \dots \}}{\lceil \text{ref}(c) \rceil_\Psi = \{ @id \in \mathcal{S}_{ref}(\Psi) \mid \Psi(id) \in \lceil \text{val}(c) \rceil_\Psi \}} \text{ refs}$$

$$\overline{\lceil \{\tau^+\} \rceil_\Psi} = \overline{\{ \{x_1, \dots, x_n\} \mid x_i \in \lceil \tau \rceil_\Psi \}} \text{ set}$$

$$\overline{\lceil [\tau^+] \rceil_\Psi} = \overline{\{ [x_1, \dots, x_n] \mid x_i \in \lceil \tau \rceil_\Psi \}} \text{ list}$$

$$\overline{\lceil \{\tau^*\} \rceil_\Psi} = \overline{\lceil \{\tau^+\} \rceil_\Psi \cup \{\{\}\}} \text{ eset} \quad \overline{\lceil [\tau^*] \rceil_\Psi} = \overline{\lceil [\tau^+] \rceil_\Psi \cup \{[]\}} \text{ elist}$$

$$\overline{\lceil \tau^? \rceil_\Psi} = \overline{\lceil \tau \rceil_\Psi \cup \{\text{null}\}} \text{ null}$$

$$\overline{\lceil \text{Int} \rceil} = \overline{\mathbb{Z}} \text{ int} \quad \overline{\lceil \text{Bool} \rceil} = \overline{\mathbb{B}} \text{ bool}$$

$$\overline{\lceil \text{Char} \rceil} = \overline{\Sigma} \text{ char} \quad \overline{\lceil \text{String} \rceil} = \overline{\Sigma^*} \text{ string}$$

$$\frac{\text{enum } e \{L_1, \dots, L_n\}}{\lceil \text{enum}(e) \rceil = \{L_1, \dots, L_n\}} \text{ enum}$$



регламентируют структуру своих экземпляров, гарантируя наличие тех или иных свойств у объектов.

### 1.3. Модели и мета-модели

Выше мы потребовали, чтобы экземпляры класса  $C = \text{class } c : S\{P\}$ , имели вид `object id : @c{...}`, то есть содержали ссылку на идентификатор класса  $c$ . Если класс  $C$  сам не является объектом, то контекст, в котором упоминаются его экземпляры, не может быть замкнутым. Для того, чтобы устранить это неудобство, мы представим классы в виде объектов и сделаем, таким образом, синтаксис модельных термов замкнутым.

На Рис. 1.8 приведена диаграмма, охватывающая основные аспекты представления классов в виде объектов: приведены представления класса `Class`, экземпляры которого обозначают классы, класса `Type` — для типов и `Property` — для свойств. Пунктирные ребра на этой диаграмме проведены не между классами и объектами, а между объектами, представляющими классы, и объектами, являющимися экземплярами этих классов. Также проведено пунктирное ребро между ребром, изображающим одно из значений свойства `@properties` объекта `Class`, и объектом описывающим это свойство. Отметим, что идентификаторы свойств на этой диаграмме имеют вид `@x`, то есть являются ссылками на объекты, описывающие сами свойства. Такое кодирование наиболее удобно, поскольку позволяет, например, узнать тип свойства, просто получив описывающий его объект по идентификатору.

У некоторых объектов на диаграмме Рис. 1.8 не указаны идентификаторы. Это сделано для краткости: эти идентификаторы нигде не используются, соответственно, единственное, что требуется — это чтобы они не совпадали ни с какими другими идентификаторами объектов. Поскольку такие несовпадающие идентификаторы всегда можно выбрать, мы можем позволить себе не изображать их на рисунке.

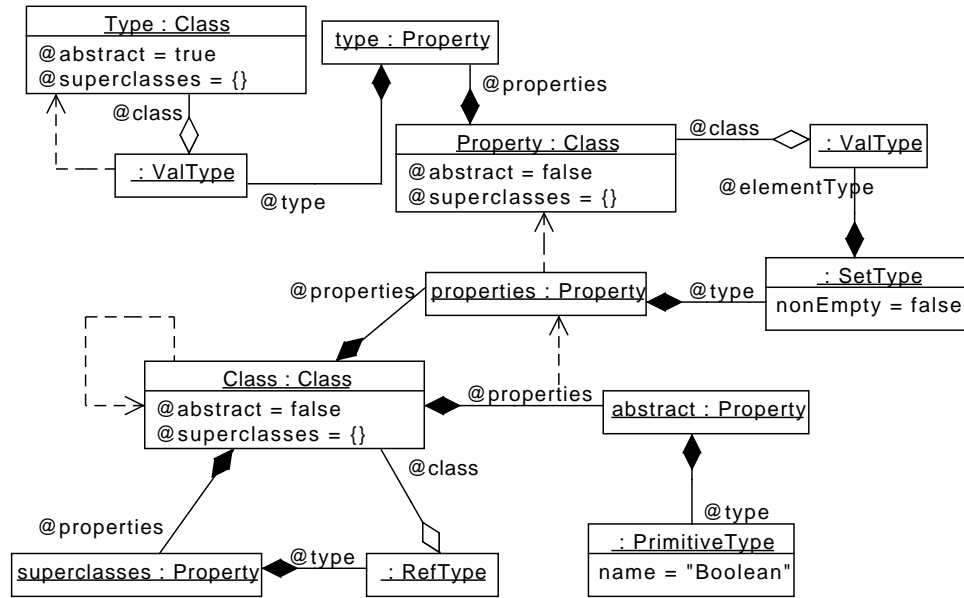


Рис. 1.8: Классы как объекты

Формальное правило для преобразования объектного представления в классы (см. Приложение ??), описанные выше, записывается довольно длинно и довольно тривиально, поэтому мы не приводим его здесь целиком. Основная идея состоит в том, что объекты вида  $\text{object } c : @\text{Class } \{p\}$  преобразуются в классы вида  $\text{class } c : S \{P\}$ , где  $S$ ,  $P$  и признак абстрактности вычисляются из значений соответствующих свойств из  $p$ .

Представление типов приведено на Рис. 1.9 и Рис. 1.10. Оно также непосредственно трансформируется в типы, определенные выше (см. Приложение ??).

**Определение 1.13** (Типовой контекст). Пусть  $m$  является предмоделью в некотором ссылочном контексте  $\Psi_0$ , и  $\Psi = \Psi_0 \cup \Psi_m$ . Если по множеству всех экземпляров класса  $\text{Class}$ , упомянутых в  $\Psi$ , строится корректный набор классов  $C$ , то функция

$$\mathfrak{T}_m(t) = \begin{cases} \tau, & \text{где } t \in \llbracket \tau \rrbracket_{\Psi}^C, \\ \perp, & \text{в противном случае,} \end{cases}$$

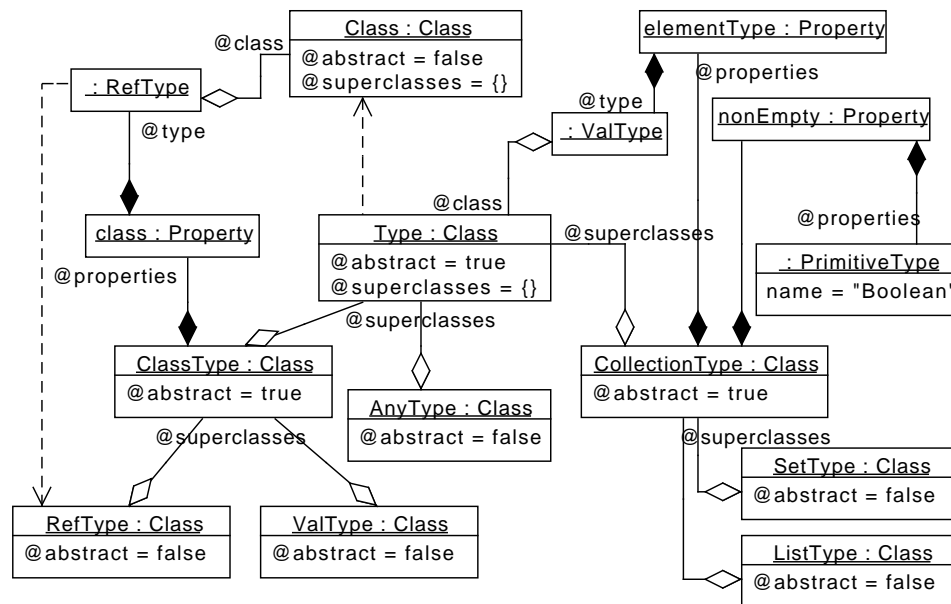


Рис. 1.9: Классы для типов (I)

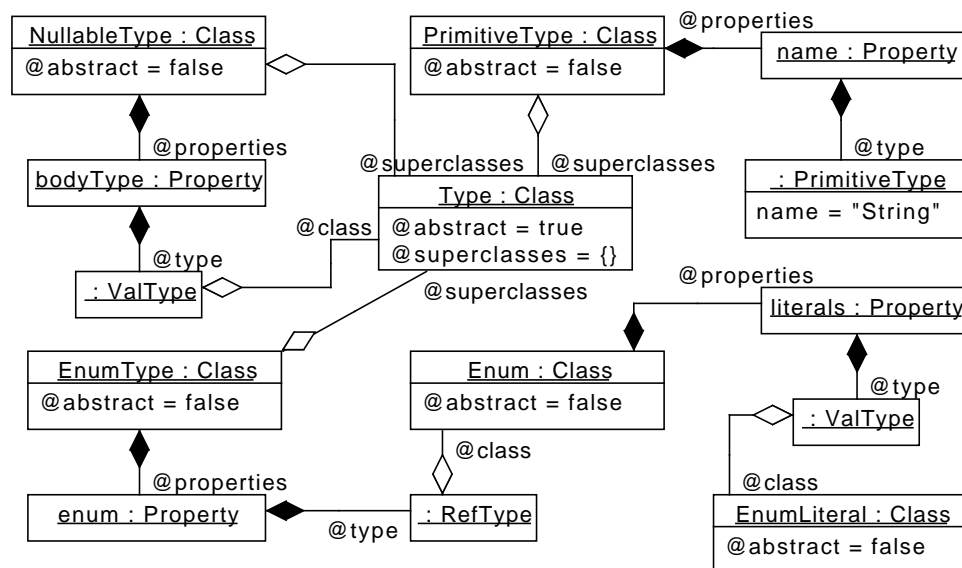


Рис. 1.10: Классы для перечислений типов (II)

называется *типовым контекстом*, порожденным  $m$ . Термы  $t$  для которых  $\mathfrak{T}_m(t) \neq \perp$ , называются *типизируемыми* в контексте  $\mathfrak{T}_m$ .

Заметим, что из типизируемости терма  $m$  следует типизируемость всех его подтермов. Заметим также, что  $m$  может не содержать описания классов, составляющих  $C$ , а лишь ссылаться на них, поскольку сами объекты, описывающие эти классы, могут быть упомянуты в контексте  $\Psi_0$ .

Основная идея определения 1.13 в том, что “правильные” термы являются типизируемыми. В дальнейшем мы будем работать только с такими термами:

**Определение 1.14 (Модель).** Пусть типовой контекст  $\mathfrak{T}_m$  построен по пред-модели  $m$  (в контексте  $\Psi_0$ ) согласно определению 1.13. Если  $\mathfrak{T}_m(m) \neq \perp$ , то  $m$  называется *моделью* в контексте  $\Psi_0$ .

Отметим следующее важное наблюдение:

**Предложение 1.1.** Модельный терм  $\mathfrak{M}$ , изображенный на Рис. 1.8, Рис. 1.9 и Рис. 1.10 (в совокупности) является моделью в пустом контексте  $\Psi_\emptyset$ :  $\Psi_\emptyset(id) \equiv \perp$ .

*Доказательство.* Для доказательства этого утверждения необходимо убедиться в том, что

- а)  $\mathfrak{M}$  является пред-моделью в контексте  $\Psi_\emptyset$ ,
- б) из  $\mathfrak{M}$  извлекается корректный набор классов,
- в)  $\mathfrak{T}_{\mathfrak{M}} \neq \perp$ .

Эти утверждения проверяются непосредственно. □

Фактически, ссылочный контекст  $\Psi_{\mathfrak{M}}$  является минимальным контекстом, необходимым для извлечения классов из любого модельного терма. Это связано с тем, что любое описание классов будет содержать ссылки на подтермы  $\mathfrak{M}$ , такие как `Class`, `Property`, `ValType`



и т.д. Чтобы проиллюстрировать это соображение, рассмотрим представление классов для  $\lambda$ -исчисления, приведенных на Рис. 1.4, в виде объектов: см. Рис. 1.11. Заметим, что все ссылки на подтермы  $\mathcal{M}$  на

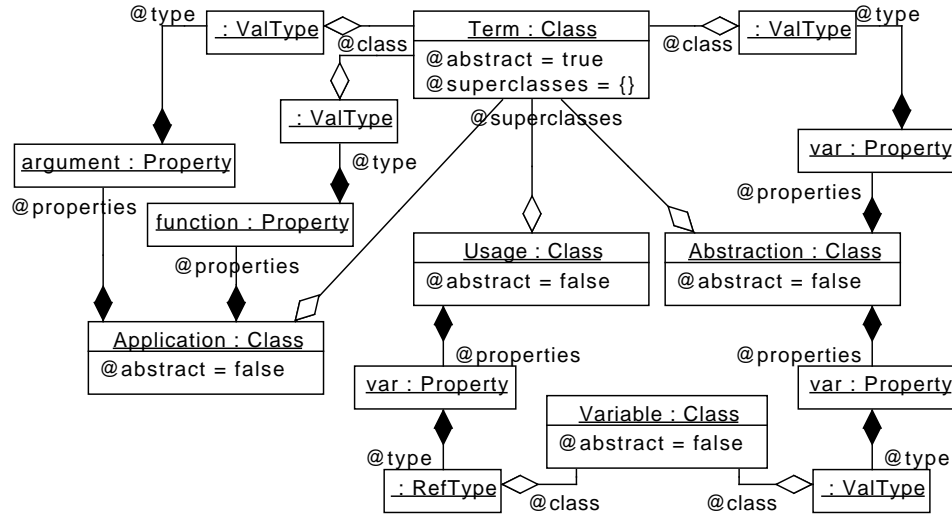


Рис. 1.11: Классы для  $\lambda$ -исчисления (см. Рис. 1.4) в виде объектов

Рис. 1.11 являются не значениями свойств объектов, а лишь ссылками на классы и маркерами свойств, то есть они лишь играют роль “разметки” для данных, задавая их структуру. Эта “разметка” называется *метаинформацией*. Такая ситуация является типичной: хотя объекты могут ссылаться на классы как на равноправные себе сущности, первоочередная роль классов в том, чтобы контролировать структуру термов. Удобно отделять описание классов от данных, структуру которых они регламентируют, поэтому мы вводим понятие *метамодели* как множества всех классов, составляющих метаинформацию для данной модели. Для этого нам понадобится определить несколько вспомогательных понятий.

Пусть  $m$  является моделью в контексте  $\Psi_0$ , и пусть  $\mathcal{M}_0(m)$  — множество всех описаний классов, на которые указывают ссылки на классы из объектов в  $\mathcal{S}_{obj}(m)$ :

$$\mathcal{M}_0(m) \stackrel{\text{def}}{=} \{ \Psi_m(cid) \mid \text{object id} : @cid \{ \dots \} \in \mathcal{S}_{obj}(m) \}.$$

Фактически,  $\mathcal{M}_0(m)$  — это множество тех классов, которые регламентируют структуру модели  $m$ .

Пусть  $\mathcal{S}_{obj}^{\overline{\mathcal{M}}}(t)$  — множество всех объектов, упомянутых в терме  $t$  и термах, на которые  $t$  ссылается, но не только в качестве метаинформации (мы считаем, что сам терм  $t$  упомянут в замкнутом контексте  $\Psi_m$ ):

$$\mathcal{S}_{obj}^{\overline{\mathcal{M}}}(t) \stackrel{\text{def}}{=} \begin{cases} \{t\} \cup \mathcal{S}_{obj}^{\overline{\mathcal{M}}}(id) \cup \bigcup_{i=1}^n \mathcal{S}_{obj}^{\overline{\mathcal{M}}}(v_i), & \text{при } t = \mathbf{object} \ id : c \{p_1 = v_1, \dots, p_n = v_n\}, \\ \bigcup_{i=1}^n \mathcal{S}_{obj}^{\overline{\mathcal{M}}}(x_i), & \text{при } t = \{x_1, \dots, x_n\} \text{ или } t = [x_1, \dots, x_n], \\ \mathcal{S}_{obj}^{\overline{\mathcal{M}}}(\Psi_m(a)), & \text{при } t = @a, \\ \emptyset, & \text{в остальных случаях} \end{cases}$$

Функция  $\mathcal{S}_{obj}^{\overline{\mathcal{M}}}(\bullet)$  естественным образом распространяется на множества термов:  $\mathcal{S}_{obj}^{\overline{\mathcal{M}}}(T) = \bigcup_{t \in T} \mathcal{S}_{obj}^{\overline{\mathcal{M}}}(t)$ .

**Определение 1.15** (Метамодель). Если  $m$  — модель, то множество  $\mathcal{M}(m) = \mathcal{S}_{obj}^{\overline{\mathcal{M}}}(\mathcal{M}_0(m))$  называется *метамоделью* для  $m$ .

$\mathcal{M}(m)$  — это множество всех объектов, непосредственно вовлеченных в описание классов из  $\mathcal{M}_0(m)$ , эти объекты составляют полную метаинформацию для  $m$ . Некоторая громоздкость функции  $\mathcal{S}_{obj}^{\overline{\mathcal{M}}}(T)$  объясняется тем, что сами объекты из  $\mathcal{M}(m)$  тоже имеют метаинформацию (из  $\mathfrak{M}$ ), которая может быть не задействована непосредственно при описании структуры  $m$ . Функция  $\mathcal{S}_{obj}^{\overline{\mathcal{M}}}(T)$  призвана исключить эту “метаинформацию для метаинформации”.

Ниже мы называем метамоделью не только множество  $\mathcal{M}(m)$ , но и терм-множество

$$\{t_1, \dots, t_n\},$$

где  $t_i$  — элементы множества  $\mathcal{M}(m)$ .

**Предложение 1.2.** *Метамодель  $\mathcal{M}(m)$  (как терм) является моделью в контексте  $\Psi_{\mathfrak{M}}$ .*

*Доказательство.* В самом деле,  $\mathcal{M}_0(m)$  состоит из *описаний классов*, которые удовлетворяют структурным требованиям классов из  $\mathfrak{M}$ , то есть  $\mathcal{S}_{obj}^{\overline{\mathcal{M}}}(\mathcal{M}_0(m)) = \mathcal{M}(m)$  состоит только из объектов, типизируемых в контексте  $\mathfrak{T}_{\mathfrak{M}}$ . С другой стороны, функция  $\mathcal{S}_{obj}^{\overline{\mathcal{M}}}(\bullet)$  гарантирует разрешение всех ссылок, упомянутых прямо или косвенно в  $\mathcal{M}_0(m)$ , то есть обеспечивает замкнутость контекста  $\Psi_{\mathcal{M}(m)} \cup \Psi_{\mathfrak{M}}$ .  $\square$

**Предложение 1.3.** *Метамодель  $\mathcal{M}(\mathfrak{M})$  совпадает с  $\mathfrak{M}$ .*

*Доказательство.* Это утверждение проверяется непосредственно.  $\square$

Предложение 1.3 показывает, что терм  $\mathfrak{M}$  имеет важную особенность: *классы в нем контролируют структуру своего собственного описания*<sup>7</sup>. Такая роль  $\mathfrak{M}$  настолько важна, что заслуживает отдельного термина:

**Определение 1.16** (Метаметамодель). Модельный терм  $\mathfrak{M}$  называется *метаметамоделью*.

## § 2. Языки как множества моделей

Традиционно *языком* называется множество строк, составленных символами некоторого конечного алфавита [2, ?]. Такой подход вполне оправдан при рассмотрении вопросов теории сложности [?], алгебраических свойств операций над языками [?, ?] и задач синтаксического анализа [2]. Однако язык, предназначенный для использования человеком, в первую очередь представляет собой *систему понятий и отношений между ними*, и при рассмотрении таких языков

<sup>7</sup>Но не только: например, они же контролируют структуру описания классов на Рис. 1.11

целесообразно отталкиваться не от текстового *конкретного* синтаксиса, роль которого обычно весьма важна, но не первостепенна, а от какого-то более структурированного описания, не зависящего от визуального представления предложений языка. Такое структурированное представление, как правило, называют *абстрактным синтаксисом*. Сразу заметим, что этот термин сильно перегружен значениями. Так наиболее распространенное понятие *абстрактного синтаксического дерева* (Abstract Syntax Tree, AST [?, 2]) не соответствует представлениям об абстрактном синтаксисе языка: абстрактное представление даже весьма простых программ, не будет являться деревом (см. Рис. 1.2).

Мы используем следующее определения языка:

**Определение 1.17 (Язык).** Пусть  $M$  — модель и  $\mathcal{M}(M) = \mathfrak{M}$  (элементами  $M$  являются описания классов, следовательно  $M$  может выступать в качестве метамодели). Множество  $\mathcal{L}(M)$ , состоящее из всех моделей  $m$ , таких что  $\mathcal{M}(m) = M$ , называется *языком*, порожденным метамоделью  $M$  или просто *языком*  $M$ .

Возникает вопрос о том, как связано традиционное определение понятия *язык* с тем, которое используем мы. В данном разделе мы покажем, во-первых, что для любого контекстно-свободного языка существует естественное представление в виде языка, порожденного некоторой метамоделью [?], а во-вторых, что контекстно-свободная грамматика  $G$ , снабженная аннотациями определенного вида [83], может быть механически преобразована в метамодель  $M_G$ , такую, что  $\mathcal{L}(M_G)$  является в точности множеством абстрактных синтаксических деревьев для языка, порожденного грамматикой  $G$ .

## 2.1. Нотация для контекстно-свободных грамматик

Для описания контекстно-свободных грамматик чаще всего применяется нотация Бэкуса-Наура (Backus-Naur Form, BNF, [?]), поз-

воляющая задавать продукции, в правой части которых стоят последовательности терминальных и нетерминальных символов. Поскольку множество контекстно-свободных языков замкнуто относительно регулярных операций объединения и итерации, мы можем рассматривать более удобную для приложений расширенную нотацию Бэкуса-Наура (Extended BNF, EBNF [36]). Существует несколько версий конкретного синтаксиса EBNF, и мы используем вариант, основанный на нотации ANTLR [63]<sup>1</sup>. В этой нотации правая и левая части продукции разделяются двоеточием, и в правой части могут использоваться следующие операции (A и B обозначают грамматические выражения):

- конкатенация: A B;
- объединение: A | B;
- итерация Клинни: A\* (повторение ноль или более раз);
- повторение один или более раз: A+;
- необязательное вхождение: A?;
- группировка: (A).

Следуя нотации ANTLR, мы пишем названия терминальных символов заглавными буквами (например NAME), а нетерминальных — начиная со строчной буквы, согласно конвенции “CamelCase” (например, longName). Ключевые слова и знаки операций, также соответствующие терминальным символам, пишутся в двойных кавычках.

В качестве примера использования данной нотации рассмотрим контекстно-свободную грамматику для упрощенного синтаксиса заголовков методов в языке JAVA [38] (см. Лист. 2.1). Метод может быть либо *абстрактным*, тогда он должен быть помечен ключевым словом `abstract` и после списка параметров следует точка с запятой, либо

<sup>1</sup>Полное описание используемой нами нотации приведено в Приложении 1.

```

methodHeader
  : "abstract" signature ";"
  | signature body
  ;

signature : type NAME "(" parameterList? ")";

parameterList : parameter ("," parameter)*;

type : "void" | NAME;

parameter : type NAME;

body : "{" "}" ;

```

Листинг 2.1: Упрощенный синтаксис описания методов в языке JAVA.

конкретным, тогда после списка параметров следует *тело* метода в фигурных скобках (для простоты в нашем примере тело всегда пустое).

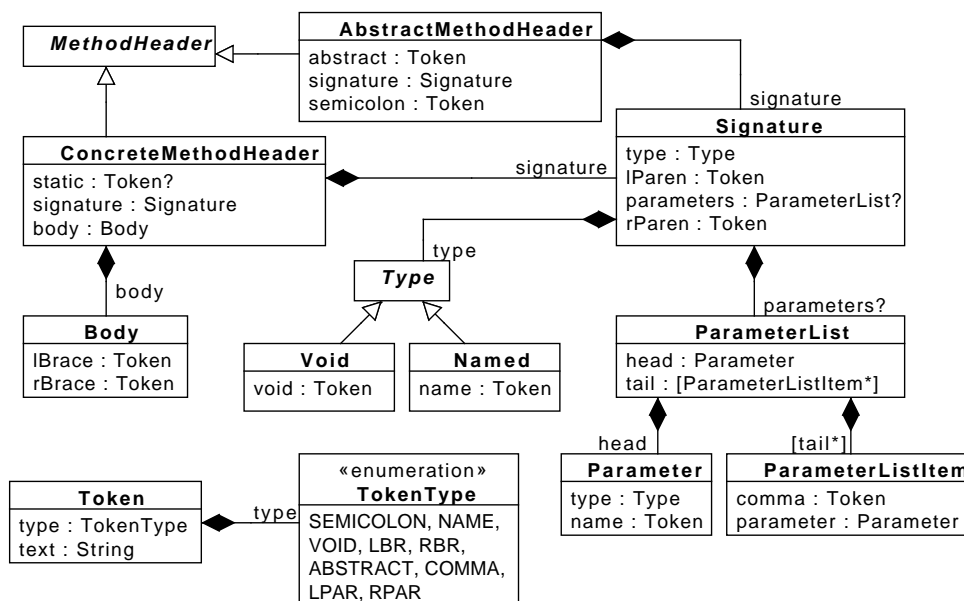


Рис. 2.1: Метамодел, соответствующая Рис. 2.1.

## 2.2. Метамодел, эквивалентные грамматикам

На Рис. 2.1 приведена диаграмма метамодели, построенной по грамматике из Лист. 2.1. Эта метамодель получена механически, и ни-

же мы приведем процедуру построения таких метамodelей, но сначала рассмотрим этот пример. Большинство классов на Рис. 2.1 соответствуют нетерминалам грамматики и имеют свойства, соответствующие составляющим правых частей соответствующих продукций, так что экземплярами данной метамodelи являются в точности деревья разбора, соответствующие предложениям языка, порождаемого грамматикой из Лист. 2.1. Рассмотрим пример такого предложения:

```
void example(String s) {}
```

Соответствующая модель приведена на Рис. 2.2. Из рисунка видно,

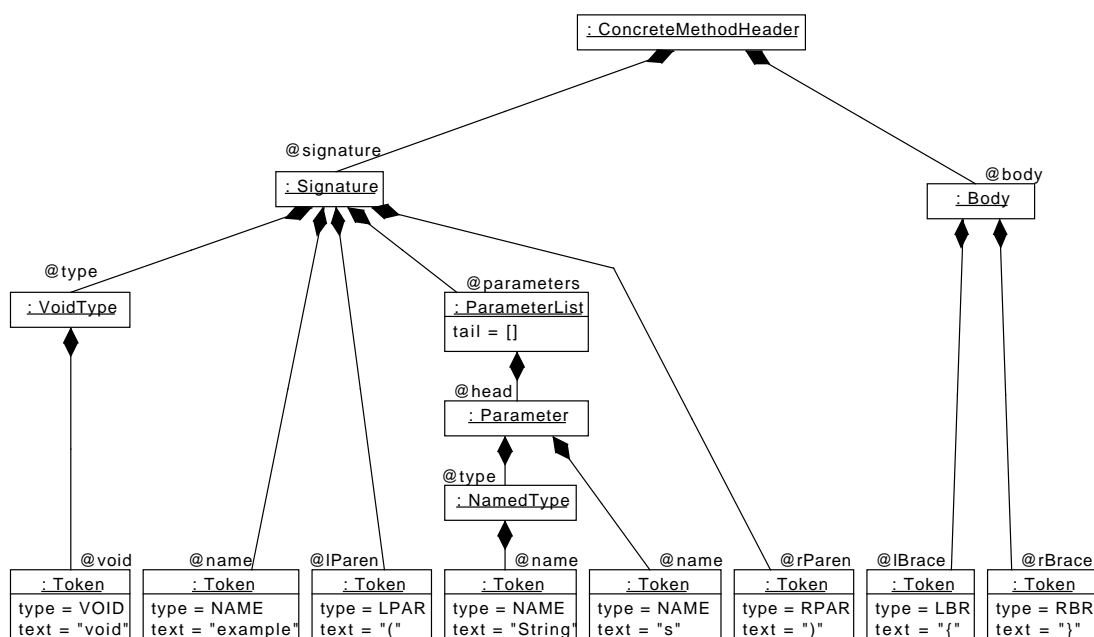


Рис. 2.2: Диаграмма объектов дерева разбора.

что дерево встраивания объектов прямо соответствует дереву разбора. Объекты класса `Token`, занимающие нижний уровень дерева, соответствуют лексемам (каждая лексема хранит текст, из которого она была получена в поле `text` и соответствующий терминальный символ в поле `type`), а на объекты на остальных уровнях дерева соответствуют применениям продукций.

Заметим, что по дереву разбора легко можно восстановить

текст, из которого оно было получено, следовательно можно утверждать, что при преобразовании в модель *сохраняется вся полезная информация*, имевшаяся в исходном тексте<sup>2</sup>. Отсюда следует Предложение 1.4.

**Предложение 1.4.** *Метамодель на Рис. 2.1 в качестве описания языка эквивалентна грамматике из Лист. 2.1.*

Осталось показать, что по любой контекстно-свободной грамматике можно построить эквивалентную метамодель. Для этого достаточно применить следующую процедуру.

!!!!!!

---

<sup>2</sup>В нашем примере мы не учитываем пробельные символы, поскольку грамматика их игнорирует, но это легко исправить.



## Глава 2.

# Расширение предметно-ориентированных языков механизмами композиции

## § 1. Автоматическое построение языков, поддерживающих типизированные макроопределения

В настоящем разделе описан метод, позволяющий по описанию языка построить описание более богатого языка, поддерживающего композицию с помощью типизированных макроопределений. Мы будем называть такой пополненный язык *макро-языком*, построенным на базе исходного языка.

**Замечание 2.1** (О терминологии). В англоязычной литературе используется термин *macro* [24, ?, 74]. В качестве русского перевода этого термина употребляется либо слово “макрос”, являющееся в сущности сленговым и полученное транслитерацией множественного числа “*macros*”, либо слово “макроопределение”, соответствующее более узкому по смыслу термину “*macro definition*”. Близкой по смыслу альтернативой является термин “шаблон” (англ. “*template*”), используемый в языке C++ [77] и ряде других [?, 26, 13].

Мы считаем термин “*macro*” более подходящим для наших целей, и будем, следуя традиции, использовать более формальный вариант его перевода — макроопределение.

### 1.1. Неформальное описание механизма композиции, основанного на макроопределениях

Наиболее известными системами, использующими макроопределения, являются язык программирования LISP [75] и препроцессор языка C [44, ?]. В обоих случаях макроопределения служат для обогащения языка новыми конструкциями, которые преобразуются в базовые конструкции языка во время компиляции<sup>1</sup>, этот процесс называется *разворачиванием* макроопределений.

Приведем пример использования макроопределений в языке C: пусть нам требуется реализовать односвязные списки. Для этого в первую очередь необходимо описать *структуру* элемента списка, например, для списка целых чисел эта структура будет выглядеть следующим образом:

```
struct IntList {
    struct IntList* next;
    int data;
};
```

Для списка элементов другого типа, например, строк, структура будет очень похожей, но тип поля `data` будет отличаться:

```
struct StrList {
    struct StrList* next;
    char* data;
};
```

Дублировать описания для каждого нового типа элементов не очень удобно, поэтому мы напишем макроопределение, которое по данному типу элемента генерирует описание соответствующей структуры<sup>2</sup>:

```
#define DEFLIST(name, type) struct name { \
    struct name *next; \
    type data; \
};
```

После директивы `#define` следует *имя* макроопределения, а за ним в скобках — *параметры*. Весь остальной текст — это *тело* макроопределения (знак “\” используется для подавления перевода строки).

<sup>1</sup>Понятие “время компиляции” в данном контексте является собирательным и противопоставляется понятию “время выполнения программы”.

<sup>2</sup>Задача, которую мы решаем в этом примере с помощью макроопределений языка C, более эффективно решается в языке C++ с помощью *шаблонных структур*, которые можно рассматривать как узкоспециализированную разновидность макроопределений.

Для того, чтобы использовать данное макроопределение, достаточно написать его имя и передать в скобках значения параметров (то есть *аргументы*), например:

```
DEFLIST(StrList, char*)
```

В результате *разворачивания* данного определения аргументы будут подставлены в тело вместо соответствующих параметров и мы получим определение структуры `StrList`, приводившееся выше. Аналогично можно получить определения структуры `IntList`, а также структуры элемента списка для любого типа. Заметим, что разворачивание происходит во время компиляции и результатом является исходный текст на языке C, который, в свою очередь, транслируется в машинный код, причем транслятор ничего не знает о том, были использованы макроопределения или нет.

Обобщая сведения, приведенные в данном примере, можно выделить следующие свойства, присущие механизму макроопределений<sup>3</sup>:

- Макроопределение состоит из *списка параметров* и *тела* и имеет уникальное *имя*.
- Тело макроопределения содержит конструкции на целевом языке (в нашем примере — на языке C) и ссылки на параметры. Также тело может содержать обращения к другим макроопределениям.
- При разворачивании ссылки на параметры в теле макроопределения заменяются значениями соответствующих аргументов.
- Разворачивание происходит во время компиляции программы.

Макроопределения представляют собой достаточно гибкий механизм повторного использования. В принципе, этот механизм не зависит от целевого языка, в который разворачиваются макроопределения.

---

<sup>3</sup>Такое обобщение имеет смысл, поскольку в различных языках и системах макроопределения функционируют схожим образом.

В частности, в языке C поддержка макроопределений обеспечивается препроцессором CPP — независимым программным средством, обрабатывающим исходный код *до* начала работы собственно компилятора языка C. Препроцессор CPP может работать с любым текстом, не только с исходным кодом на языке C, следовательно он (или аналогичный механизм) может применяться для повторного использования и в других языках, в частности в предметно-ориентированных, делая их более пригодными для использования в промышленных проектах.

Однако чисто текстовый препроцессор обладает одним важным недостатком: корректность результата разворачивания макроопределений никак не гарантируется, поскольку препроцессор манипулирует простым текстом и “не знает” о синтаксисе целевого языка.

Вернемся к примеру макроопределения, приведенному выше. Если программист допустит ошибку при использовании макроопределения `DEFLIST`, а именно перепутает порядок аргументов, что случается не так уж редко, препроцессор послушно выполнит свою работу:

```
DEFLIST(char*, StrList)
```

превратится в

```
struct char* { // error: expected '{' before 'char'
    struct char* *next;
    StrList data;
};
```

Получившийся код синтаксически некорректен, и компилятор, получив этот текст на вход, выдаст сообщение об ошибке:

```
DEFLIST(char*, StrList) // error: expected '{' before 'char'
```

В итоге ошибка программиста обнаружена, но сообщение, выданное компилятором, записано в терминах программы, полученной после разворачивания макроопределений, и совсем не помогает программисту исправить ситуацию. Чтобы разобраться, в чем проблема, придется вручную рассмотреть код, полученный на выходе препроцессора, что является существенным затруднением при разработке больших проектов. Описанная здесь проблема является основной причиной, по

которой профессиональные программисты зачастую избегают широкого использования возможностей макроопределений в программах на С [?].

Итак, чисто текстовый препроцессор позволяет легко обеспечить поддержку макроопределений в любом языке, но не обеспечивает своевременного обнаружения ошибок, что затрудняет разработку. В данной главе мы рассмотрим метод реализации макроопределений, который также пригоден для любого языка, но обеспечивает контроль корректности аргументов макроопределений с помощью специальной системы типов, что позволяет избежать проблем, присущих чисто текстовым препроцессорам. Такие макроопределения мы будем называть *шаблонами* (templates).

### 1.1.1. Шаблоны в языках, порожденных метамоделями

Вернемся к примеру описания структуры элементов списка в языке С. Для начала рассмотрим абстрактный синтаксис такого описания; соответствующая метамодель приведена на Рис. 1.1. Мы не ста-

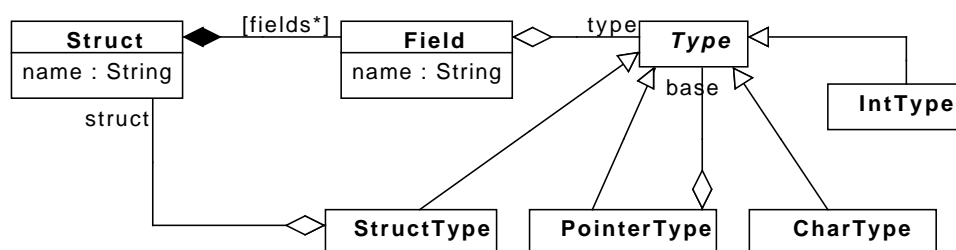


Рис. 1.1: Метамодель абстрактного синтаксиса описаний структур в языке С

вим целью рассмотрение возможностей языка С во всем их многообразии, поэтому наша метамодель позволяет оперировать лишь весьма ограниченным набором типов: структурами (**Struct** и **StructType**), указателями (**PointerType**) и примитивными типами `int` (**IntType**) и `char` (**CharType**).

На Рис. 1.2 приведена диаграмма, соответствующая описанию структуры `IntList`, уже приводившемуся выше:

```
struct IntList {
    struct IntList* next;
    int data;
};
```

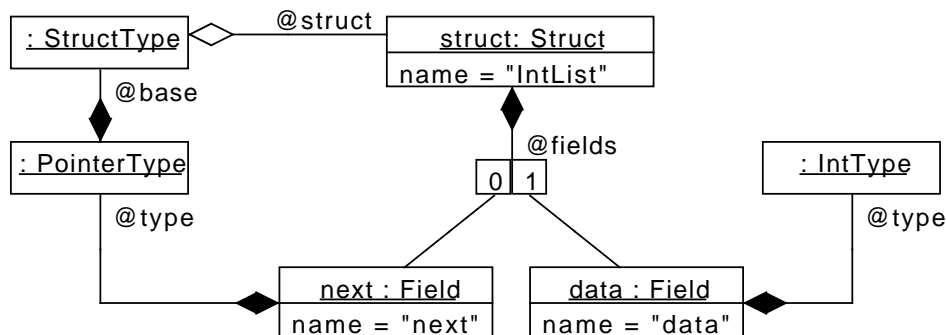


Рис. 1.2: Модель, описывающая структуру `IntList`

Объект класса `Struct` хранит список, состоящий из объектов класса `Field`, каждый из которых имеет тип и имя.

Теперь преобразуем модель на Рис. 1.2 в шаблон, аналогичный макроопределению `DEFLIST` из предыдущего раздела:

```
#define DEFLIST(name, type) struct name { \
    struct name *next; \
    type data; \
};
```

Что для этого нужно сделать? Нужно добавить специальные объекты, представляющие структуру шаблона. На Рис. 1.3 эти объекты выделены зеленым цветом фона. Рассмотрим новую диаграмму подробнее. Корневым элементом дерева встраивания является объект `DEFLIST` класса `Abstraction` — это и есть шаблон, он содержит список *параметров*, состоящий из двух объектов класса `Variable`, и *тело* — объект `struct`. Значение свойства `@name` объекта `struct` изменилось по сравнению с Рис. 1.2: если раньше это была строка `IntList`, то теперь это объект класса `VariableUsage`, который ссылается на объект `name`

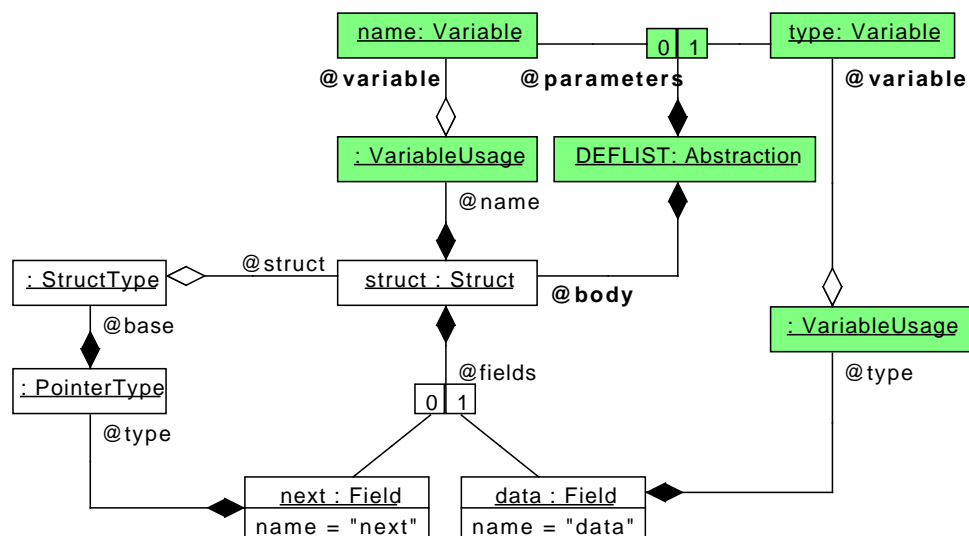


Рис. 1.3: Шаблон описания структуры элемента списка

класса `Variable`. Это соответствует использованию параметра теле шаблона. Аналогично изменилось значение свойства `@type` объекта `data`: теперь это тоже объект класса `VariableUsage`, ссылающийся на параметр `type`.

Процедура разворачивания просто заменяет объекты `VariableUsage` в теле шаблона значениями соответствующих аргументов и получается модель, не содержащая “шаблонных” объектов (зеленого цвета). Чтобы придать параметрам значения, требуется применить шаблон к соответствующим аргументам. На Рис. 1.4 при-

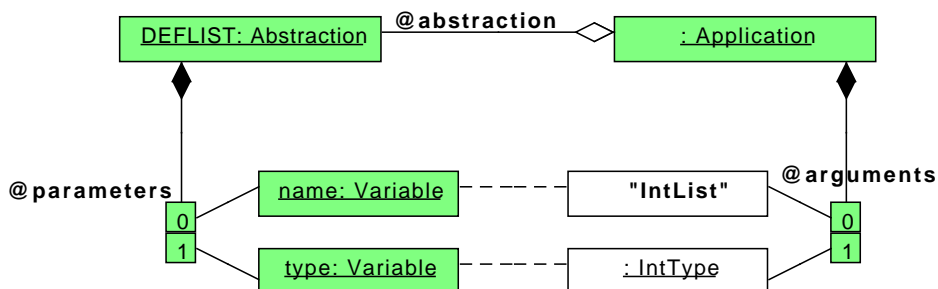


Рис. 1.4: Применение шаблона

ведена диаграмма, соответствующая применению шаблона `DEFLIST`, определенного выше, к аргументам `IntList` и `object : @IntType{}`. Объект класса `Application` (*применение*) содержит ссылку на применяемый шаблон и список аргументов. Соответствие между параметрами шаблона и аргументами устанавливается с помощью индексов в списках: нулевой аргумент соответствует нулевому параметру, первый — первому и т.д. На рисунке параметры шаблона соединены с соответствующими аргументами пунктирными линиями.

Заметим, что результатом разворачивания применения шаблона на Рис. 1.4 будет в точности модель на Рис. 1.2, аналогично тому как разворачивание применения макроопределения

```
DEFLIST(IntList, int)
```

дает описание структуры `IntList`.

Заметим также, что аргументами шаблона в принципе могут быть и “шаблонные” объекты. Так на Рис. 1.5 показано применения

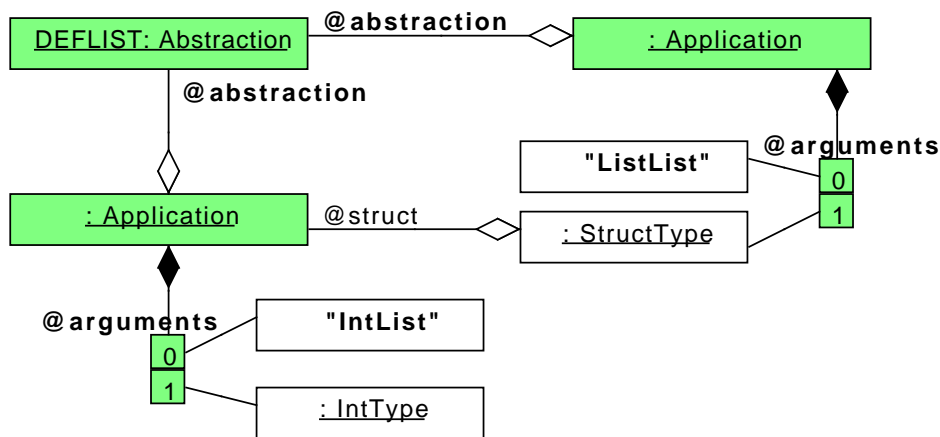


Рис. 1.5: Использование применения шаблона в аргументах

шаблона `DEFLIST` к списку аргументов, один из которых, в свою очередь, тоже является применением шаблона `DEFLIST`. В результате разворачивания шаблонов в этом примере получится описание структуры `ListList` элементов списка, состоящего из списков целых чисел.



Чтобы облегчить понимание, мы позволили себе некоторую вольность, приводя модельные термины с шаблонами на диаграммах данном разделе. Рассмотрим, например, Рис. 1.3: объекты `struct` и `data`, изображенные на этом рисунке, не удовлетворяют требованиям метамодели, приведенной на Рис. 1.1, поскольку эта метамодель определяет свойство `name` класса `Struct` как строковое, а на нашей диаграмме оно хранит объект класса `VariableUsage`; аналогично для свойства `type` объекта `data`.

Такое положение вещей неудивительно: язык, порожденный метамоделью на Рис. 1.1, не поддерживает шаблоны, а для того, чтобы добавить в язык поддержку нового механизма, нужно как минимум пополнить его новыми конструкциями. В последующих разделах мы опишем метод, позволяющую пополнять языки автоматически. Процесс разработки языка с поддержкой шаблонов на основе уже существующего языка схематически представлен на Рис. 1.6. Метамодель и

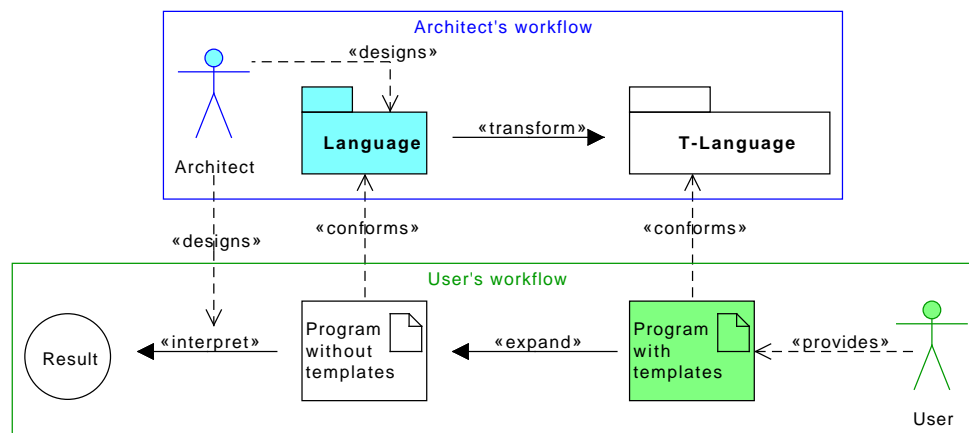


Рис. 1.6: Разработка и использование языков с поддержкой шаблонов

интерпретирующая семантика языка *Language* разрабатывается архитектором, после чего к метамодели применяется автоматизированная процедура трансформации, которую мы опишем ниже, в результате чего получается новая метамодель, поддерживающая шаблоны. Поль-

зователь создает программу, удовлетворяющую требованиям пополненной метамодели, в своей программе он может использовать шаблоны. К этой программе применяется процедура разворачивания шаблонов, включающая в себя проверку типов аргументов (эту процедуру мы также опишем ниже). В результате получается программа, удовлетворяющая исходной метамодели языка *Language*, к которой применима интерпретирующая семантика, разработанная архитектором. Таким образом, интерпретирующая семантика пополненного языка *T-Language* получается автоматически как композиция процедуры разворачивания и интерпретирующей семантики исходного языка.

### 1.1.2. Базовый язык шаблонов

Примеры, приведенные в предыдущем разделе, дают общее представление о том, как можно пополнить существующий язык (в данном случае — подмножество языка C, позволяющее описывать структуры) механизмом шаблонов. В данном разделе мы приступаем к формальному описанию этого метода.

Структура “шаблонных” объектов в примерах из предыдущего раздела соответствует упрощенной метамодели, приведенной на Рис. 1.7. Определению шаблона соответствует класс *Abstraction*,

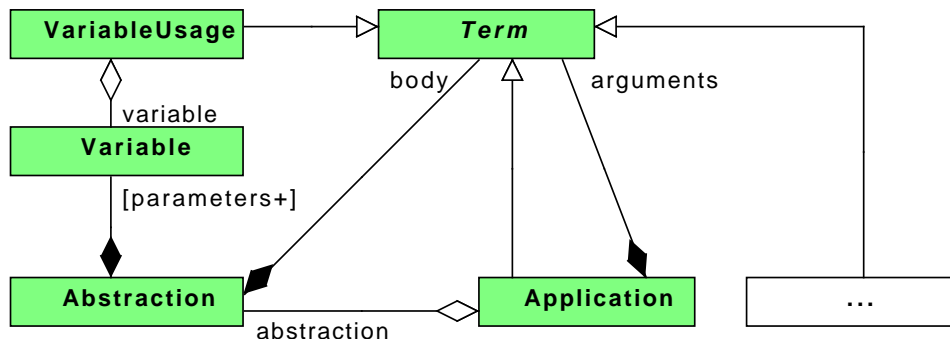


Рис. 1.7: Упрощенная метамодель базового языка шаблонов

объекты которого содержат *тело* (body) и список *параметров*

(parameters). Телом шаблона может являться объект класса `Term`, мы будем называть такие объекты *шаблонными термами*. На Рис. 1.7 изображено два конкретных подкласса абстрактного класса `Term`: `Application` (применение шаблона) и `VariableUsage` (использование переменной), структуру которых мы обсуждали в предыдущем разделе.

Заметим, что Рис. 1.3 не удовлетворяет требованиям данной метамодели: в качестве тела шаблона указан объект класса `Struct`, не являющегося подклассом класса `Term`. Эта проблема сродни той, которую мы обсуждали в конце предыдущего раздела: базовый язык шаблонов, порождаемый метамodelью, изображенной на Рис. 1.7, не поддерживает конструкций, специфичных для описания структур. Поэтому при построении шаблонного языка необходимо не просто добавить классы для шаблонов в исходную метамodelь, но и изменить структуру метамодели так, чтобы конструкции исходного языка стали шаблонными термами.

На Рис. 1.7 изображен подкласс класса `Term`, вместо идентификатора имеющий многоточие. Он символизирует шаблонные конструкции, полученные из структур исходного языка. На Рис. 1.8 изображены классы метамодели, полученной в результате пополнения языка описания структур шаблонными конструкциями. Из рисунка видно, что каждый класс исходной метамодели (Рис. 1.1) теперь является подклассом класса `Term`, и все свойства в также имеют значения типа `ref (Term)` или `val (Term)`. Это позволяет в качестве значения любого свойства, включая свойства классов на Рис. 1.7, использовать как конструкции исходного языка (теперь являющиеся шаблонными термами), так и “чисто-шаблонные” термы, то есть применения шаблонов и ссылки на параметры.

## 1.2. Генерация языка шаблонов

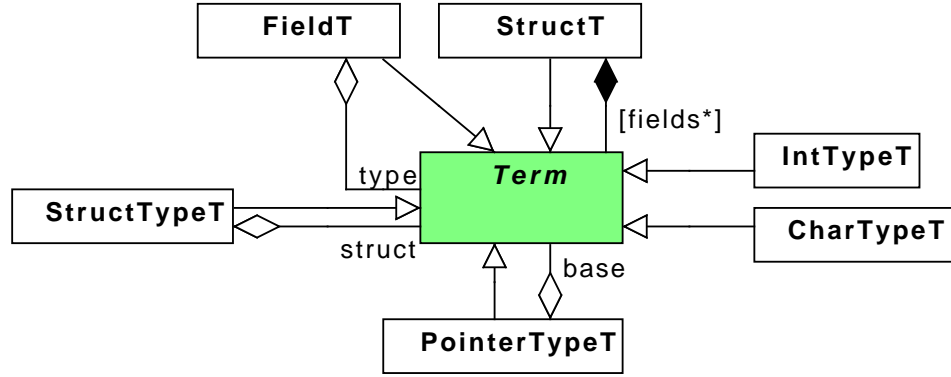


Рис. 1.8: Мета-модель пополненного языка описания структур

Для того, чтобы построить язык шаблонов на основе заданного языка  $L$ , необходимо расширить базовую мета-модель, алгоритм разворачивания и систему типов конструкциями и правилами, специфичными для этого языка.

Пусть целевая мета-модель языка  $L$  обозначается  $\mathcal{M}$ , тогда мета-модель соответствующего языка шаблонов,  $\mathcal{TM}$ , состоит из классов, полученных применением преобразования  $\mathcal{TC}(\bullet)$ , описанного на Рис. 1.9, к каждому классу мета-модели  $\mathcal{M}$ . Эта мета-модель использует как классы базовой мета-модели шаблонов ( $\text{Term}$ ), так и примитивные типы и перечисления мета-модели  $\mathcal{M}$ .

$$\begin{aligned}
 \mathcal{TC}(\text{class } \langle S, R, A \rangle) &= \text{class } \langle \text{Term}, \mathcal{TR}(R), \mathcal{TA}(A) \rangle \\
 \mathcal{TC}(C^*) &= \mathcal{TC}(C)^* \\
 \mathcal{TC}(C^+) &= \mathcal{TC}(C)^+ \\
 \mathcal{TC}(C^?) &= \mathcal{TC}(C)^? \\
 \mathcal{TR}(\text{ref}\langle r : T \rangle) &= \text{ref}\langle r : \text{Term} \rangle \\
 \mathcal{TR}(\text{var}\langle r : T \rangle) &= \text{val}\langle r : \text{Term} \rangle \\
 \mathcal{TA}(\text{attr}\langle a : T \rangle) &= \text{attr}\langle a : T \rangle
 \end{aligned}$$

Рис. 1.9: Преобразование конструкций языка в шаблонные выражения

Преобразование  $\mathcal{TC}(\bullet)$  сопоставляет каждому классу соответствующий тип шаблонного выражения. При этом ссылки перенаправляются на класс  $\text{Term}$ , поскольку вместо конкретного объекта может

выступать шаблонное выражение.

### 1.3. Семантика языка шаблонов

Семантика языка шаблонов задается операцией *разворачивания*, описанной в композиционном стиле правилами на Рис. 1.10. Мы обозначаем результат разворачивания шаблонов в выражении  $e$  как  $\mathcal{I}_\gamma[e]$ , где  $\gamma$  (“*среда*”) является множеством значений параметров шаблонов вида  $p = e$ , где  $p$  — параметр, а  $e$  — шаблонное выражение. Как видно из правила *app-inst*, когда разворачивается применение шаблона, среда пополняется информацией о текущих значениях параметров, при этом вызов происходит по значению, то есть аргументы разворачиваются до обработки тела вызываемого шаблона.

$$\frac{\frac{\{p = e\} \subseteq \gamma}{\mathcal{I}_\gamma[<?p>] = e} \text{ var-inst}}{\frac{\text{template}(T <p_1, \dots, p_n> \{b\})}{\gamma' = \bigcup_{i=1}^n \{p_i = \mathcal{I}_\gamma[a_i]\} \quad \mathcal{I}_\gamma[<T a_1, \dots, a_n>] = \mathcal{I}_{\gamma \cup \gamma'}[b]} \text{ app-inst}$$

Рис. 1.10: Базовая семантика языка шаблонов

Для обеспечения функционирования специфичных конструкций процедура разворачивания шаблонов пополняется правилами следующего вида:

$$\frac{t = \mathcal{TC}(C) @id \{r_i = rv_i, a_j = av_j\}}{\mathcal{I}_\gamma[t] = \mathcal{TC}(C) @id' \{r_i = \mathcal{I}_\gamma[rv_i], a_j = av_j\}} \text{ ds-inst}(C)$$

Задача таких правил — развернуть шаблонные выражения, находящиеся внутри специфичных конструкций, поэтому все эти правила однотипны и просто осуществляют рекурсивные вызовы на значениях всех ссылок, выходящих из данного объекта. Поскольку в графе объектов могут быть циклы, в процессе разворачивания поддерживается служебное множество уже обработанных объектов, что соответствует стандартному алгоритму обхода графа в глубину [?].

Согласно приведенному выше правилу, результатом разворачивания шаблонного выражения является другое шаблонное выражение, полученное разворачиванием всех ссылок на параметры и применений шаблонов. Будем называть такие выражения имеющими *нормальную форму*. Для того, чтобы получить из выражения  $e$  в нормальной форме экземпляр мета-модели  $\mathcal{M}$ , необходимо его преобразовать. Такое преобразование (обозначаемое  $\mathcal{M}(e)$ ) весьма просто: для каждого объекта класса  $\mathcal{TC}(C)$  создается объект класса  $C$  с той же структурой, то есть ссылки трансформируются рекурсивно, а атрибуты копируются. Ниже мы будем подразумевать выполнение данного преобразования после разворачивания шаблонов.

#### 1.4. Система типов

Для того, чтобы гарантировать, что результат применения шаблона будет корректным с точки зрения целевой мета-модели, мы определяем систему типов для языка шаблонов. Базовые правила этой системы типов приведены на Рис. 1.11, они должны быть дополнены специфичными правилами для поддержки конкретного языка. Отношение  $\preceq$  является специфичным для расширяемого языка и не определяется в базовой системе типов, а лишь используется.

В правилах *list* и *set* использовано отношение  $Item(x, \tau)$ , означающее, что  $x$  может быть элементом коллекции  $\tau$ . Определение этого отношения дано на том же рисунке и сводится к тому, что внутри коллекции типа  $\tau$  можно использовать не только одиночные значения этого типа, но и переменные, сами являющиеся коллекциями, что соответствует возможности вставки сразу нескольких элементов.

Система типов также пополняется правилами для специфичных конструкций. Эти правила имеют следующий вид:

$$\frac{\begin{array}{l} x = \mathcal{TC}(C) @id \{r_i = v_i; a_j = v'_j\} \quad r_i = \mathcal{TR}(\rho_i : \tau_i) \\ \mathcal{TM} \Vdash x : \mathcal{TC}(C) \quad \Gamma \vdash v_i : \tau_i \end{array}}{\Gamma \vdash x : C} \text{ ds-type}(C)$$

$$\begin{array}{c}
\frac{}{\Gamma \cup \{v : \tau\} \vdash \langle ?v \rangle : \tau} \text{ var} \\
\\
\frac{\Gamma \cup \bigcup_{i=1}^n \{p_i : \tau_i\} \vdash b : \sigma}{\Gamma \vdash \mathbf{template}(T \langle p_1 : \tau_1, \dots, p_n : \tau_n \rangle : \sigma \{b\})} \text{ abstr} \\
\\
\frac{\Gamma \vdash \mathbf{template}(T \langle p_1 : \tau_1, \dots, p_n : \tau_n \rangle : \sigma \{b\}) \quad \forall i : [1 : n]. \Gamma \vdash a_i : \tau_i}{\Gamma \vdash \langle T a_1, \dots, a_n \rangle : \sigma} \text{ appl} \\
\\
\frac{}{\Gamma \vdash \mathbf{NULL} : \tau^?} \text{ null} \quad \frac{\Gamma \vdash x : \tau}{\Gamma \vdash x : \tau^?} \text{ relax} \quad \frac{\Gamma \vdash x : \tau^+}{\Gamma \vdash x : \tau^*} \text{ relax}^+ \\
\\
\frac{\Gamma \vdash x : \tau \quad \tau \preceq \sigma}{\Gamma \vdash x : \sigma} \text{ subtype} \\
\\
\frac{}{\Gamma \vdash [] : \tau^*} \text{ elist} \quad \frac{\forall i \in [1 : n]. \Gamma \vdash \mathbf{Item}(t_i, \tau)}{\Gamma \vdash [t_1, \dots, t_n] : \tau^+} \text{ list} \\
\\
\frac{}{\Gamma \vdash \{\} : \tau^*} \text{ eset} \quad \frac{\forall i \in [1 : n]. \Gamma \vdash \mathbf{Item}(t_i, \tau)}{\Gamma \vdash \{t_1, \dots, t_n\} : \tau^+} \text{ set} \\
\\
\frac{\Gamma \vdash t : \tau}{\Gamma \vdash \mathbf{Item}(t, \tau)} \text{ item} \quad \frac{\Gamma \vdash v : \tau^*}{\Gamma \vdash \mathbf{Item}(\langle ?v \rangle, \tau)} \text{ item}^* \quad \frac{\Gamma \vdash v : \tau^+}{\Gamma \vdash \mathbf{Item}(\langle ?v \rangle, \tau)} \text{ item}^+
\end{array}$$

Рис. 1.11: Базовая система типов языка шаблонов

Однотипность этих правил также объясняется их рекурсивной природой: они нужны только для того, чтобы проверить типы в шаблонных выражениях внутри данного объекта, если он сам удовлетворяет требованиям мета-модели  $\mathcal{TM}$ .

Заметим, что шаблонное выражение, являющееся объектом класса  $\mathcal{TC}(C)$  типизируется самим классом  $C$ . Это необходимо для соблюдения требований к наследованию. Поскольку в мета-модели  $\mathcal{TM}$  все классы наследуются от класса `Term`, отношения наследования в ней не соответствуют таким же отношениям в  $\mathfrak{M}$ . Поэтому в правиле *subtype* на Рис. 1.11 отношение  $\preceq$  задается мета-моделью  $\mathfrak{M}$ , а не  $\mathcal{TM}$ , и типы тоже берутся из  $\mathfrak{M}$ .

Чтобы показать, что ограничения, накладываемые системой типов адекватны требованиям мета-модели, докажем следующую лемму.

**Лемма 2.1** (О нормальных формах). *Если шаблонное выражение  $e$*

имеет нормальную форму и  $\vdash e : \tau$ , то  $\Psi, \mathfrak{M} \Vdash \mathfrak{M}(e) : \tau$ .

*Доказательство.* Достаточно заметить, что в дереве вывода для  $\vdash e : \tau$  правила  $var$ ,  $abstr$ ,  $appl$ ,  $item^*$  и  $item^+$  не встречаются, а остальные правила в системе типов для шаблонов имеют прямые аналоги в системе требований мета-модели на Рис. ??.

Дополнительно к правилам типизации, на язык шаблонов накладывается следующее требование: *Никакой шаблон не должен быть достижим из своего тела по ссылкам.* Это требование гарантирует отсутствие рекурсии в определениях шаблонов.

## 1.5. Структурная корректность

Система типов накладывает ограничения на шаблоны. Шаблоны и шаблонные выражения, удовлетворяющие правилам типизации, при разворачивании порождают конечные объекты, удовлетворяющие требованиям целевой мета-модели. Для того, чтобы убедиться в этом, покажем, что для системы типов и семантики языка шаблонов выполняются свойства *сохранения типов* и *нормализации* [65].

Как отмечалось выше, необходимым условием корректного поведения является тот факт, что функция разворачивания получает на вход выражение, в котором соблюдаются правила системы типов. Это условие формализовано в следующем определении.

**Определение 2.1.** Среда  $\gamma$  называется *согласованной с контекстом*  $\Gamma$ , если все ее элементы имеют допустимые типы:

$$\forall p : \{p = e\} \subseteq \gamma \Rightarrow \begin{cases} \{p : \tau\} \subseteq \Gamma \\ \Gamma \vdash e : \tau \end{cases}$$

Теперь докажем первое из упомянутых выше свойств.

**Теорема 2.2** (О сохранении типов). *Если среда  $\gamma$  согласована с контекстом  $\Gamma$  и  $\Gamma \vdash e : \tau$ , то  $\Gamma \vdash \mathcal{I}_\gamma[e] : \tau$ . Другими словами, преобразование  $\mathcal{I}[\bullet]$  сохраняет типы.*



*Доказательство.* Будем вести индукцию по определению  $\mathcal{I}_\gamma [e]$ .

**База.** Правило  $ds-inst(C)$  сохраняет типы, поскольку из объекта класса  $\mathcal{TC}(C)$  получается объект того же класса, а правило  $ds-type(C)$  выводит тип из класса объекта.

**Переход.** В правиле  $app-inst$  среда расширяется, и нам необходимо показать, что результат согласован с контекстом. Это обеспечивается требованиями к типам параметров, накладываемыми правилом  $appl$  и способом формирования контекста в правиле  $abstr$ . Из этого по предположению индукции следует, что правило  $app-inst$  сохраняет типы.

Так же по предположению правило  $var-inst$  сохраняет типы.  $\square$

Свойство нормализации формулируется следующим образом:

**Теорема 2.3** (О нормализации). *Если среда  $\gamma = \cup\{p_i = e_i\}$  согласована с контекстом  $\Gamma$ , все  $e_i$  имеют нормальную форму и  $\Gamma \vdash e : \tau$ , то результат вычисления  $\mathcal{I}_\gamma [e]$  имеет нормальную форму.*

*Доказательство.* Будем вести индукцию по определению  $\mathcal{I}_\gamma [\square \bullet]$ .

**База.** Результат применения правила  $var-inst$  имеет нормальную форму, поскольку все элементы среды имеют нормальную форму. **Пе-**

**реход.** В правиле  $app-inst$  среда пополняется значениями, имеющими нормальную форму, поэтому по предположению индукции вызовы  $\mathcal{I}_\gamma [a_i]$  и  $\mathcal{I}_{\gamma \cup \gamma'} [b]$  заканчиваются за конечное число шагов и результат имеет нормальную форму.

Правило  $ds-inst(C)$  заменяет значения ссылок результатами разворачивания, имеющими нормальную форму, следовательно и результат применения правила имеет нормальную форму.  $\square$

Требование о том, чтобы все элементы среды имели нормальную форму выполняется для пустой среды, следовательно теорема применима для разворачивания шаблонных выражений, применяемых на практике.

Мы показали, что функция  $\mathcal{I}[\bullet]$  корректно разворачивает все шаблонные параметры и применения шаблонов, а также что она не нарушает структурной корректности с точки зрения мета-модели. Результатом применения этой функции всегда является константное шаблонное выражение, которое, как отмечалось выше, тривиальным образом преобразуется в экземпляр мета-модели  $\mathcal{M}$ . Таким образом, описанный здесь механизм шаблонов работает корректно.

## 1.6. Вывод типов

Как отмечалось выше, в большинстве случаев типы в объявлениях шаблонов можно опускать, поскольку они могут быть реконструированы автоматически. Для этого применяется алгоритм, аналогичный использованному в GRAMMATIC<sup>SDT</sup> (см. Главу ?? и [65]). Тип переменной выводится исходя из двух соображений: (а) какие аргументы ей присваиваются в применениях шаблона и (б) в каком контексте она используется в теле шаблона, то есть, если на использование переменной указывает ссылка  $\mathcal{TR}(R)$ , то учитывается тип ссылки  $R$ . Если система ограничений, построенная таким образом, не имеет решения, генерируется сообщение об ошибке типизации. Такой подход позволяет для переменной найти наиболее широкий тип объектов, которые могут быть подставлены на ее место.

Возникающие в процессе реконструкции неоднозначности разрешаются следующим образом: если параметр используется непосредственно внутри коллекции типа  $T$ , он получает тип  $T^*$  и позволяет добавить в эту коллекцию ноль или более элементов. Исключение составляет случай, когда параметр является единственным элементом коллекции, в которой мета-модель требует наличия хотя бы одного элемента: в этом случае параметр получает тип  $T^+$ .

## Приложение 1. Нотация для контекстно-свободных грамматик

```

grammar{[Symbol*]}
  : symbol{>>}*;

symbol{Symbol}
  : NAME{>name} (":" expression{>>expressions})+
  ;

expression{abstract Expression}
  : singleCharacter
  : positiveRange
  : negativeRange
  : stringLiteral
  : option
  : iteration
  : alternative
  : sequence
  : empty
  : symbolReference
  : "(" expression{>} ")"
  ;

positiveRange{Alternative}
  : "[" rangeEntry{>>expressions} "]"
  : "[" rangeEntry{>>expressions} rangeEntry{>>expressions}+ "]"
  ;

negativeRange{NegativeCharacterRange}
  : "[" "^" rangeEntry{>>ranges}+ "]"
  ;

singleCharacter{CharacterRange}
  : CHARACTER {>from >to}
  ;

rangeEntry{CharacterRange}
  : singleCharacter
  : CHARACTER{>from} "-" CHARACTER{>to}
  ;

stringLiteral{StringLiteral}
  : STRING{~>value}
  ;

option{Option}
  : expression{>body} "?"
  ;

iteration{Iteration}
  : expression{>body} iterationKind{>kind}
  ;

iterationKind{enum IterationKind}

```

```

: "*" {ZERO_OR_MORE}
: "+" {ONE_OR_MORE}
;

alternative {Alternative}
: expression{>>expressions} ("|" expression{>>expressions}) *
;

sequence {Sequence}
: expression{>>expressions} expression{>>expressions} +
;

empty {Empty}
: "#empty"
;

symbolReference {SymbolReference}
: NAME{~~>symbol} // ~~> for late resolve
;

```

## Список литературы

- [1] *Agesen, O.* Adding type parameterization to the java language / O. Agesen, S. N. Freund, J. C. Mitchell // *SIGPLAN Not.* — 1997. — Vol. 32, no. 10. — Pp. 49–65.
- [2] *Aho, A. V.* Compilers: principles, techniques, and tools / A. V. Aho, R. Sethi, J. D. Ullman. — Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1986.
- [3] *Aksit, M.* Compiler generation based on grammar inheritance. — 1990. <http://doc.utwente.nl/19862/>.
- [4] Apache Derby. — <http://db.apache.org/derby/>.
- [5] The asf+sdf meta-environment: A component-based language development environment / M. G. J. van den Brand, A. v. Deursen, J. Heering et al. // CC '01: Proceedings of the 10th International Conference on Compiler Construction. — London, UK: Springer-Verlag, 2001. — Pp. 365–370.
- [6] Aspect-oriented programming / G. Kiczales, J. Lamping, A. Mendhekar et al. // In ECOOP. — SpringerVerlag, 1997.
- [7] *Ammann, U.* Invasive Software Composition / U. Ammann. — Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2003.
- [8] *Atkinson, C.* Model-driven development: A metamodeling foundation / C. Atkinson, T. Kühne // *IEEE Softw.* — 2003. — Vol. 20, no. 5. — Pp. 36–41.
- [9] *Baader, F.* Term rewriting and all that / F. Baader, T. Nipkow. — New York, NY, USA: Cambridge University Press, 1998.

- [10] *Bagge, A.* DSAL = library+notation: Program transformation for domain-specific aspect languages / A. Bagge, K. K. T. // First Domain Specific Aspect Language Workshop (DSAL'06). — 2006.
- [11] *Berners-Lee, T.* Uniform resource identifiers (uri) generic syntax: Tech. Rep. RFC 2396 / T. Berners-Lee, R. Fielding, L. Masinter: 1998. — <http://www.ietf.org/rfc/rfc2396.txt>.
- [12] *Booch, G.* Object-Oriented Analysis and Design with Applications (3rd Edition) / G. Booch. — Redwood City, CA, USA: Addison Wesley Longman Publishing Co., Inc., 2004.
- [13] *Booch, G.* Unified Modeling Language User Guide, The (2nd Edition) (Addison-Wesley Object Technology Series) / G. Booch, J. Rumbaugh, I. Jacobson. — Addison-Wesley Professional, 2005.
- [14] *Budinsky, F.* Eclipse Modeling Framework / F. Budinsky, S. A. Brodsky, E. Merks. — Pearson Education, 2003.
- [15] *Ciric, M. D.* Parsing in different languages / M. D. Ciric, S. R. Rancic // *FACTA UNIVERSITATIS*. — 2005. — Vol. 18, no. 2. — Pp. 299–307.
- [16] The compost, compass, inject/j and recoder tool suite for invasive software composition: Invasive composition with compass aspect-oriented connectors. / D. Heuzeroth, U. Ammann, M. Trifu, V. Kuttruff // GTTSE / Ed. by R. Lämmel, J. Saraiva, J. Visser. — Vol. 4143 of *Lecture Notes in Computer Science*. — Springer, 2006. — Pp. 357–377.
- [17] *Dekel, U.* Towards a standard family of languages for matching patterns in source code / U. Dekel, T. Cohen, S. Porat // SWSTE '03: Proceedings of the IEEE International Conference on Software-Science, Technology & Engineering. — Washington, DC, USA: IEEE Computer Society, 2003. — P. 10.

- [18] *Denny, J. E.* Ielr(1): practical lr(1) parser tables for non-lr(1) grammars with conflict resolution / J. E. Denny, B. A. Malloy // SAC '08: Proceedings of the 2008 ACM symposium on Applied computing. — New York, NY, USA: ACM, 2008. — Pp. 240–245.
- [19] *Design Patterns: Elements of Reusable Object-Oriented Software* / E. Gamma, R. Helm, R. Johnson, J. Vlissides. — Addison-Wesley, 1995.
- [20] *Dinkelaker, T.* Untangling crosscutting concerns in domain-specific languages with domain-specific join points / T. Dinkelaker, M. Monperrus, M. Mezini // DSAL '09: Proceedings of the 4th workshop on Domain-specific aspect languages. — New York, NY, USA: ACM, 2009. — Pp. 1–6.
- [21] *Eli: a complete, flexible compiler construction system* / R. W. Gray, S. P. Levi, V. P. Heuring et al. // *Communications of the ACM*. — 1992. — Vol. 35, no. 2. — Pp. 121–130.
- [22] *Fowler, M.* Domain specific languages. — <http://martinfowler.com/dslwip/>. — 2010.
- [23] *Gagnon, E. M.* SableCC, an object-oriented compiler framework / E. M. Gagnon, L. J. Hendren // TOOLS '98: Proceedings of the Technology of Object-Oriented Languages and Systems. — Washington, DC, USA: IEEE Computer Society, 1998. — Pp. 140–154.
- [24] *Ganz, S. E.* Macros as multi-stage computations: type-safe, generative, binding macros in macroml / S. E. Ganz, A. Sabry, W. Taha // ICFP '01: Proceedings of the sixth ACM SIGPLAN international conference on Functional programming. — New York, NY, USA: ACM, 2001. — Pp. 74–85.

- [25] Gärtner, F. The PretzelBook. —  
<http://www.informatik.tu-darmstadt.de/BS/Gaertner/pretzel/dist>  
 1998.
- [26] Gradecki, J. D. Mastering Apache Velocity / J. D. Gradecki,  
 J. Cole. — Wiley, 2003.
- [27] Grimm, R. Better extensibility through modular syntax / R. Grimm //  
*SIGPLAN Not.* — 2006. — Vol. 41, no. 6. — Pp. 38–51.
- [28] Haskell 98 Language and Libraries: The Revised Report / Ed. by  
 S. P. Jones. — <http://haskell.org/>, 2002. — September. — P. 277.  
<http://haskell.org/definition/haskell98-report.pdf>.
- [29] Havinga, W. Prototyping and composing aspect languages /  
 W. Havinga, L. Bergmans, M. Aksit // ECOOP '08: Proceedings of  
 the 22nd European conference on Object-Oriented Programming. —  
 Berlin, Heidelberg: Springer-Verlag, 2008. — Pp. 180–206.
- [30] Hedin, G. JastAdd: an aspect-oriented compiler construction system /  
 G. Hedin, E. Magnusson // *Science of Computer Programming*. —  
 2003. — Vol. 47, no. 1. — Pp. 37–58.
- [31] Henry, K. A crash overview of groovy / K. Henry // *Crossroads*. —  
 2006. — Vol. 12, no. 3. — Pp. 5–5.
- [32] Human-usable textual notation. —  
<http://www.omg.org/spec/hutn/>.
- [33] Implementation of multiple attribute grammar inheritance in the tool  
 lisa / M. Mernik, V. Žumer, M. Lenič, E. Avdičaušević // *SIGPLAN*  
*Not.* — 1999. — Vol. 34, no. 6. — Pp. 68–75.
- [34] Ingalls, D. H. H. The smalltalk-76 programming system design and  
 implementation / D. H. H. Ingalls // POPL '78: Proceedings of the 5th



- ACM SIGACT-SIGPLAN symposium on Principles of programming languages. — New York, NY, USA: ACM, 1978. — Pp. 9–16.
- [35] Introduction to Algorithms / T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein. — 2nd edition. — The MIT Press, 2001.
- [36] ISO. ISO/IEC 14977:1996: Information technology — Syntactic metalanguage — Extended BNF / ISO. — 1996. — P. 12.  
<http://www.iso.ch/cate/d26153.html>.
- [37] ISO/IEC 9075:1992, Database Language SQL. —  
<http://www.contrib.andrew.cmu.edu/~shadow/sql/sql1992.txt>. — 1992.
- [38] The Java Language Specification, Third Edition / J. Gosling, B. Joy, G. Steele, G. Bracha. — 3 edition. — Amsterdam: Addison-Wesley Longman, 2005. — June. — P. 688.
- [39] JetBrains. Meta Programming System (MPS). —  
<http://www.jetbrains.com/mps>. — 2009.
- [40] Johnson, S. C. Yacc: Yet another compiler-compiler: Tech. rep. / S. C. Johnson: Bell Laboratories, 1979.
- [41] Jouault, F. Km3: A dsl for metamodel specification / F. Jouault, J. Bézivin // FMOODS / Ed. by R. Gorrieri, H. Wehrheim. — Vol. 4037 of *Lecture Notes in Computer Science*. — Springer, 2006. — Pp. 171–185.
- [42] Jouault, F. Tcs:: a dsl for the specification of textual concrete syntaxes in model engineering / F. Jouault, J. Bézivin, I. Kurtev // GPCE '06: Proceedings of the 5th international conference on Generative programming and component engineering. — New York, NY, USA: ACM, 2006. — Pp. 249–254.

- [43] *Kastens, U.* Ordered attributed grammars / U. Kastens // *Acta Informatica*. — 1980. — Vol. 13, no. 3. — Pp. 229–256.
- [44] *Kernighan, B. W.* The C programming language / B. W. Kernighan, D. M. Ritchie. — Prentice-Hall, Englewood Cliffs, N.J. :, 1978. — Pp. x, 228 p. ;.
- [45] *Khedker, U.* Data Flow Analysis: Theory and Practice / U. Khedker, A. Sanyal, B. Karkare. — Boca Raton, FL, USA: CRC Press, Inc., 2009.
- [46] *Klint, P.* Toward an engineering discipline for grammarware / P. Klint, R. Lämmel, C. Verhoef // *ACM Trans. Softw. Eng. Methodol.* — 2005. — Vol. 14, no. 3. — Pp. 331–380.
- [47] *Klint, P.* Term rewriting meets aspect-oriented programming. / P. Klint, T. van der Storm, J. J. Vinju // Processes, Terms and Cycles / Ed. by A. Middeldorp, V. van Oostrom, F. van Raamsdonk, R. C. de Vrijer. — Vol. 3838 of *Lecture Notes in Computer Science*. — Springer, 2005. — Pp. 88–105.
- [48] *Knuth, D. E.* Semantics of context-free languages / D. E. Knuth // *Theory of Computing Systems*. — 1968. — June. — Vol. 2, no. 2. — Pp. 127–145.
- [49] *Kodaganallur, V.* Incorporating language processing into java applications: A JavaCC tutorial / V. Kodaganallur // *IEEE Software*. — 2004. — Vol. 21, no. 4. — Pp. 70–77.
- [50] *Lämmel, R.* Recovering Grammar Relationships for the Java Language Specification / R. Lämmel, V. Zaytsev // Ninth IEEE International Working Conference on Source Code Analysis and Manipulation. — IEEE, 2009. — Сентябрь. — Pp. 178–186. — Full version submitted for journal publication.

- [51] *Latendresse, M.* Regreg: a lightweight generator of robust parsers for irregular languages / M. Latendresse // *Reverse Engineering, Working Conference on.* — 2003. — Vol. 0. — P. 206.
- [52] *Levine, J.* Flex & Bison / J. Levine. — O'Reilly, 2009.
- [53] *McPeak, S. G.* Elkhound: A fast, practical glr parser generator: Tech. rep. / S. G. McPeak. — Berkeley, CA, USA: 2003.
- [54] Meta-object facility. — <http://www.omg.org/mof/>.
- [55] Model driven architecture. — <http://www.omg.org/mda/>.
- [56] *Moon, D. A.* Programming language for old timers. — <http://users.rcn.com/david-moon/PL0T>.
- [57] *Mossenbock, H.* Coco/r - a generator for fast compiler front ends: Tech. rep. / H. Mossenbock: 1990.
- [58] *Nickel, U.* The fujaba environment / U. Nickel, J. Niere, A. Zündorf // ICSE '00: Proceedings of the 22nd international conference on Software engineering. — New York, NY, USA: ACM, 2000. — Pp. 742–745.
- [59] *Odersky, M.* Programming in Scala: [a comprehensive step-by-step guide] / M. Odersky, L. Spoon, B. Venners. — artima, 2008. — Vol. 1. ed., version 5. — P. 736.
- [60] On language-independent model modularisation / F. Heidenreich, J. Henriksson, J. Johannes, S. Zschaler. — 2009. — Pp. 39–82.
- [61] Open fortran parser. — <http://fortran-parser.sourceforge.net>.
- [62] An overview of AspectJ / G. Kiczales, E. Hilsdale, J. Hugunin et al. // ECOOP '01: Proceedings of the 15th European Conference on Object-Oriented Programming. — London, UK: Springer-Verlag, 2001. — Pp. 327–353.

- [63] *Parr, T.* The Definitive ANTLR Reference: Building Domain-Specific Languages / T. Parr. Pragmatic Programmers. — First edition. — Pragmatic Bookshelf, 2007. — Май.
- [64] *Parr, T. J.* Antlr: A predicated- :::: Ll(k) :::: Parser generator / T. J. Parr, R. W. Quong // *SPE*. — 1995. — Vol. 25, no. 7. — Pp. 789–810.
- [65] *Pierce, B. C.* Types and programming languages / B. C. Pierce. — Cambridge, MA, USA: MIT Press, 2002.
- [66] PostgreSQL. — <http://www.postgresql.org/>.
- [67] *Rebernak, D.* A tool for compiler construction based on aspect-oriented specifications / D. Rebernak, M. Mernik // COMPSAC '07: Proceedings of the 31st Annual International Computer Software and Applications Conference. — Washington, DC, USA: IEEE Computer Society, 2007. — Pp. 11–16.
- [68] Revised report on the algorithmic language algol 68 / A. van Wijngaarden, B. J. Mailloux, J. E. L. Peck et al. // *Acta Inf.* — 1975. — Vol. 5. — Pp. 1–236.
- [69] *Safonov, V. O.* Using Aspect-Oriented Programming for Trustworthy Software Development / V. O. Safonov. — New York, NY, USA: Wiley-Interscience, 2008.
- [70] Separation of concerns in compiler development using aspect-orientation / X. Wu, B. R. Bryant, J. Gray et al. // SAC '06: Proceedings of the 2006 ACM symposium on Applied computing. — New York, NY, USA: ACM, 2006. — Pp. 1585–1590.
- [71] *Sheard, T.* Template meta-programming for haskell / T. Sheard, S. P. Jones // Haskell '02: Proceedings of the 2002 ACM SIGPLAN

- workshop on Haskell. — New York, NY, USA: ACM, 2002. — Pp. 1–16.
- [72] *Shonle, M.* Xaspects: an extensible system for domain-specific aspect languages / M. Shonle, K. Lieberherr, A. Shah // OOPSLA '03: Companion of the 18th annual ACM SIGPLAN conference on Object-oriented programming, systems, languages, and applications. — New York, NY, USA: ACM, 2003. — Pp. 28–37.
- [73] Silver: an extensible attribute grammar system / E. Van Wyk, D. Bodin, J. Gao, L. Krishnan // *ENTCS*. — 2008. — Vol. 203, no. 2. — Pp. 103–116.
- [74] *Skalski, K.* Meta-programming in Nemerle. — <http://nemerle.org/metaprogramming.pdf>. — 2004.
- [75] *Steele Jr., G. L.* Common LISP: the language / G. L. Steele, Jr. — Newton, MA, USA: Digital Press, 1984.
- [76] Stratego/xt 0.17. a language and toolset for program transformation / M. Bravenboer, K. T. Kalleberg, R. Vermaas, E. Visser // *Sci. Comput. Program.* — 2008. — Vol. 72, no. 1-2. — Pp. 52–70.
- [77] *Stroustrup, B.* The C++ Programming Language / B. Stroustrup. — Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2000.
- [78] *Van Wyk, E.* Implementing aspect-oriented programming constructs as modular language extensions / E. Van Wyk // *Sci. Comput. Program.* — 2007. — Vol. 68, no. 1. — Pp. 38–61.
- [79] *Varró, D.* Metamodeling mathematics: A precise and visual framework for describing semantics domains of uml models / D. Varró, A. Pataricza // UML '02: Proceedings of the 5th

International Conference on The Unified Modeling Language. — London, UK: Springer-Verlag, 2002. — Pp. 18–33.

- [80] Weaving a debugging aspect into domain-specific language grammars / H. Wu, J. Gray, S. Roychoudhury, M. Mernik // SAC '05: Proceedings of the 2005 ACM symposium on Applied computing. — New York, NY, USA: ACM, 2005. — Pp. 1370–1374.
- [81] Wöß, A. LL(1) conflict resolution in a recursive descent compiler generator / A. Wöß, M. Löberbauer, H. Mössenböck // JMLC / Ed. by L. Böszörményi, P. Schojer. — Vol. 2789 of *Lecture Notes in Computer Science*. — Springer, 2003. — Pp. 192–201.
- [82] Xml metadata interchange. — <http://www.omg.org/xmi/>.
- [83] Behrens, H. Xtext User Guide. — [http://www.eclipse.org/Xtext/documentation/0\\_7\\_2/xtext.pdf](http://www.eclipse.org/Xtext/documentation/0_7_2/xtext.pdf). — 2009.
- [84] Zdun, U. A dsl toolkit for deferring architectural decisions in dsl-based software design / U. Zdun // *Information and Software Technology*. — 2010. — Vol. 52. — Pp. 733–748.