



Министерство науки и высшего образования Российской Федерации
федеральное государственное бюджетное образовательное
учреждение высшего профессионального образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)» (МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ «Робототехника и комплексная автоматизация»
КАФЕДРА «Системы автоматизированного проектирования (РК-6)»

РАСЧЁТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

к выпускной квалификационной работе

на тему

«Разработка механизма вывода типов с использованием системы
типов Хиндли-Милнера»

Студент РК6-75Б
группа

подпись, дата

Никитин В.Л.
ФИО

Руководитель ВКР

подпись, дата

Соколов А.П.
ФИО

Консультант

подпись, дата

Соколов А.П.
ФИО

Нормоконтролёр

подпись, дата

Грошев С.В.
ФИО

Москва, 2024

Министерство науки и высшего образования Российской Федерации
федеральное государственное бюджетное образовательное
учреждение высшего профессионального образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)» (МГТУ им. Н.Э. Баумана)

УТВЕРЖДАЮ

Заведующий кафедрой РК-6
индекс

_____ А.П. Карпенко

« _____ » _____ 2024 г.

ЗАДАНИЕ

на выполнение выпускной квалификационной работы

Студент группы: РК6-75Б

Никитин Владимир Леонидович

(фамилия, имя, отчество)

Тема выпускной квалификационной работы: Разработка механизма вывода типов с использованием системы типов Хиндли-Милнера

Источник тематики (кафедра, предприятие, НИР): кафедра

Тема выпускной квалификационной работы утверждена распоряжением по факультету РК № _____ от « _____ » _____ 2024 г.

Техническое задание

Часть 1. Анализ актуальности.

Должен быть выполнен анализ существующих система систем типов в современных языках программирования. Также должна быть обоснована актуальность разработки в целом.

Часть 2. Математическая постановка задачи, разработка архитектуры программной реализации, программная реализация.

Должна быть описана система типов и реализована в языке программирования Kodept в качестве механизма вывода типов

Часть 3. Проведение тестирования.

Должно быть проведено тестирование разработанной в ходе работы над курсовым проектом программы.

Оформление выпускной квалификационной работы:

Расчетно-пояснительная записка на 29 листах формата А4.

Перечень графического (иллюстративного) материала (чертежи, плакаты, слайды и т.п.):

<i>количество: 5 рис., 1 табл., 7 источн.</i>

Дата выдачи задания «01» октября 2024 г.

В соответствии с учебным планом выпускную квалификационную работу выполнить в полном объеме в срок до « » 2024 г.

Студент

подпись, дата

Никитин В.Л.
ФИО

Руководитель выпускной квалификационной работы

подпись, дата

Соколов А.П.
ФИО

Примечание. Задание оформляется в двух экземплярах: один выдается студенту, второй хранится на кафедре.

Министерство науки и высшего образования Российской Федерации
федеральное государственное бюджетное образовательное
учреждение высшего профессионального образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)» (МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ РК
КАФЕДРА РК-6
ГРУППА РК6-75Б

УТВЕРЖДАЮ
Заведующий кафедрой РК-6
индекс
_____ А.П. Карпенко
« _____ » _____ 2024 г.

КАЛЕНДАРНЫЙ ПЛАН выполнения выпускной квалификационной работы

Студент группы: РК6-75Б
Никитин Владимир Леонидович

(фамилия, имя, отчество)

Тема выпускной квалификационной работы: Разработка механизма вывода типов
с использованием системы типов Хиндли-Милнера

№ п/п	Наименование этапов вы- пускной квалификационной работы	Сроки выполнения этапов		Отметка о выполнении	
		план	факт	Должность	ФИО, подпись
1.	Задание на выполнение работы. Формулировка проблемы, цели и задач работы	18.02.2024	18.02.2024	Руководитель ВКР	Соколов А.П.
2.	1 часть: аналитический обзор ли- тературы	18.02.2024	31.03.2024	Руководитель ВКР	Соколов А.П.
3.	Утверждение окончательных формулировок решаемой про- блемы, цели работы и перечня задач	28.02.2024	28.02.2024	Заведующий кафедрой	А.П. Карпенко

№ п/п	Наименование этапов выпускной квалификационной работы	Сроки выполнения этапов		Отметка о выполнении	
		план	<i>факт</i>	Должность	ФИО, подпись
4.	2 часть: математическая постановка задачи, разработка архитектуры программной реализации, программная реализация	31.03.2024	<i>31.03.2024</i>	Руководитель ВКР	Соколов А.П.
5.	3 часть: проведение вычислительных экспериментов, отладка и тестирование	30.04.2024	<i>30.04.2024</i>	Руководитель ВКР	Соколов А.П.
6.	1-я редакция работы	31.05.2024	<i>31.05.2024</i>	Руководитель ВКР	Соколов А.П.
7.	Подготовка доклада и презентации	17.06.2024	<i>17.06.2024</i>		
8.	Заключение руководителя	15.06.2024	<i>15.06.2024</i>	Руководитель ВКР	Соколов А.П.
9.	Допуск работы к защите на ГЭК	15.06.2024	<i>15.06.2024</i>	Нормоконтролер	С.В. Грошев
10.	Внешняя рецензия	12.06.2024	<i>12.06.2024</i>		
11.	Защита работы на ГЭК	19.06.2024	<i>19.06.2024</i>		

Студент _____ Никитин В.Л. Руководитель ВКР _____ Соколов А.П.
подпись, дата ФИО подпись, дата ФИО

НАПРАВЛЕНИЕ НА ЗАЩИТУ ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ

Председателю
Государственной Экзаменационной Комиссии № _____

факультета «Робототехника и комплексная автоматизация» МГТУ им. Н.Э. Баумана

Направляется студент Фамилия Имя Отчество группы РК6-81Б

на защиту выпускной квалификационной работы Тема

Декан факультета подпись декана «11» мая 2020 г.

Справка об успеваемости

Студент Фамилия Имя Отчество за время пребывания в МГТУ имени Н.Э. Баумана с 2017 г. по 2020 г. полностью выполнил учебный план со следующими оценками: отлично – [процент] %, хорошо – [процент] %, удовлетворительно – [процент] %.

Инспектор деканата от руки

Отзыв руководителя выпускной квалификационной работы

запросить у руководителя
Студент Фамилия И.О. в процессе выполнения ВКР проявил себя как ... Результаты, полученные в процессе реализации задания, позволили сделать вывод о ... целесообразности/нецелесообразности выбранных путей решения поставленной задачи, ... невозможности применения ... Работа выполнена автором самостоятельно, в полном объеме, в полном соответствии с заданием и календарным планом. Несмотря на сделанные замечания студент достоин «отличной» оценки... и присвоения звания бакалавр техники и технологий по направлению «Информатика и вычислительная техника».

запросить у руководителя
Руководитель ВКР _____ «____» _____ 2020 г.
(подпись) А.П. Соколов (ФИО) (дата)
Студент _____ «____» _____ 2020 г.
(подпись) И.О. Фамилия (ФИО) (дата)
подписать

РЕФЕРАТ

выпускная квалификационная работа: 29 с., 5 рис., 1 табл., 7 источн.

ТЕОРИЯ ТИПОВ, ЯЗЫКИ ПРОГРАММИРОВАНИЯ, КОМПИЛЯТОРЫ, ФУНКЦИОНАЛЬНОЕ ПРОГРАММИРОВАНИЕ, СИСТЕМА ТИПОВ ХИНДЛИ-МИЛНЕРА.

Работа посвящена реализации механизма вывода типов для языка программирования Kodept. Программирование выстроено вокруг глубокой математической теории. Благодаря этому появляются возможности для оптимизации, развития и улучшения языков посредством применения математики. Одним из важных применений является теория типов, которая помогает программисту в написании кода. В последнее время все больше и больше языков почерпывают что-то из этой области. Применение мощной системы типов позволяет зачастую снизить количество ошибок, возникающих при разработке.

Тип работы: выпускная квалификационная работа.

Тема работы: *«Разработка механизма вывода типов с использованием системы типов Хиндли-Милнера».*

Объект исследования: система типов.

Основная задача, на решение которой направлена работа: реализация алгоритма вывода типов на основе выбранной системы типов.

Цели работы: реализация системы вывода и проверки типов

В результате выполнения работы: 1) спроектировано представление абстрактного синтаксического дерева в компиляторе; 2) реализован семантический анализатор; 3) показано, что компилятор успешно может вывести тип функции

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	9
1 Постановка задачи	12
1.1 Концептуальная постановка задачи	12
1.2 Математическая постановка задачи	12
2 Программная реализация	16
2.1 Архитектура	16
2.2 Проблема хранения абстрактного синтаксического дерева	18
2.3 Организация доступа к элементам абстрактного синтаксического дерева	20
2.4 Реализация алгоритма W	21
3 Тестирование и отладка	23
ЗАКЛЮЧЕНИЕ	24
Литература	25
ПРИЛОЖЕНИЯ	27
А	27
Б	28

ВВЕДЕНИЕ

В современном программировании становится все более важным сколько будет потрачено времени на создание того или иного продукта. В это число входит как время, потраченное непосредственно на создание приложения, так и время, потраченное на его поддержку. Поэтому инструменты, применяемые программистом в повседневной работе, должны всячески помочь ему в этом.

Пожалуй, самым главным таким инструментом является компилятор. Разработчики компиляторов прикладывают большие усилия, чтобы язык программирования отвечал требованиям надежности и скорости. При создании инструмента такого рода важно правильно выбирать и проектировать каждую часть. Одной из основных таких частей является то, как в языке программирования взаимодействуют друг с другом типы.

В высокоуровневых языках программирования типы окружают разработчика повсюду. Чем более развитая система типов, там больше можно выразить, используя ее, а значит, если она надежна и подкреплена математической основой, то в программе станет меньше ошибок. Кроме того, в таком случае программы можно будет применять в качестве доказательств для различных теорий [1]. Сейчас такое уже применяет компания Intel при проектировании новых алгоритмов умножения или деления.

Актуальна проблема высокого порога входа в некоторые функциональные языки программирования, например Haskell и др. Он завышен, так как в них применяются сложные математические теории, повлиявшие и на синтаксис языка, и на всю его идеологию в целом. Поэтому появилась идея создать язык программирования, вобравший в себя идеи функциональных языков, но при этом сохранивший C-подобный синтаксис. Кроме того, он задумывался еще и как способ изучить всю цепочку создания компилятора. На сегодняшний день уже разработан синтаксис языка (листинг 1), стабилизировано его внутреннее представление (*абстрактное синтаксическое дерево (AST)*), а также ведется работа над компилятором.

Абстрактным синтаксическим деревом называют структуру данных, получаемую после синтаксического анализа. В ней связываются элементы синтаксиса языка: параметры функции, объявление переменной и т.д.

```

1  module Main =>
2  fun greeting(name: String) => "Hello " + name + "!"
3  fun main => print(greeting("world"))

```

Компилятор в языке программирования обычно разделен на несколько частей (рис. В.1). Одной из них является семантический анализ. В него входят анализ областей видимости объявлений, *проверка (и вывод) типов* и др.

Проверкой типов называют процесс, когда тем или иным образом проверяется правильность типа выражения согласно системе типов языка. В ходе научно-исследовательской работы была выбрана система типов Хиндли-Милнера. Алгоритм W [2], реализующий ее, позволяет также проводить вывод типов.

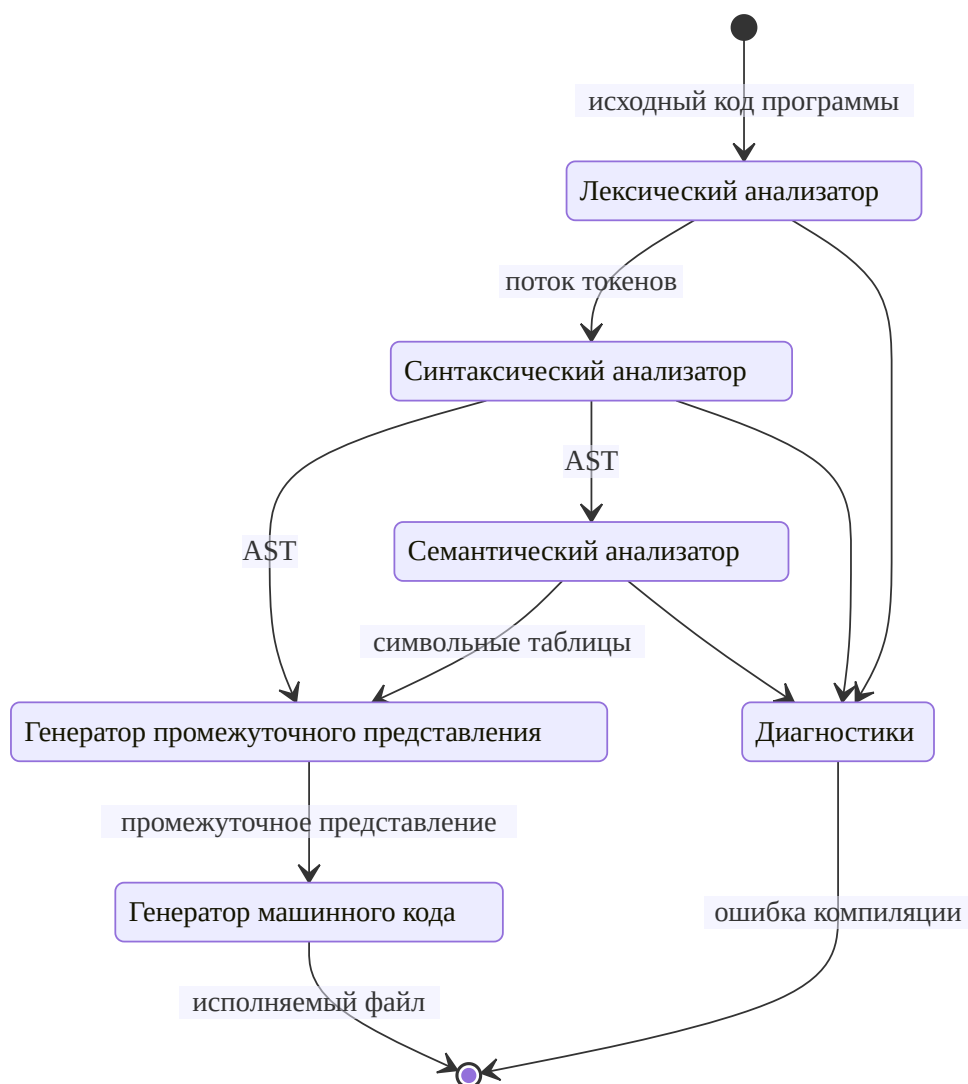


Рисунок В.1. Схема состояний работы компилятора

Целью курсового проекта является реализация механизма вывода и проверки типов для языка программирования Kodept в качестве его дальнейшего развития.

1 Постановка задачи

1.1 Концептуальная постановка задачи

Объект разработки: система типов

Цель: реализовать систему вывода и проверки типов

Задачи:

- 1) спроектировать представления AST в компиляторе,
- 2) реализовать анализатор областей видимости,
- 3) написать алгоритм для вывода типов.

1.2 Математическая постановка задачи

1.2.1 Теория типов

В разделе представлена информация о специальном разделе математики - теории типов [3]. Освящены важные понятия - *терм*, *тип*, *суждение* и *система типов*.

Терм x - чаще всего элемент языка программирования, будь то переменная, константа, вызов функции и др. Например, в Haskell, термами будут: лямбда-функция $\lambda x \rightarrow x + 1$, определение переменной `let x = "Hello" in ()` и т.д. Как можно заметить, термы могут включать в себя другие термы.

Типом A обозначается метка, приписываемая объектам, например объекты на натюрмортах принадлежат к типу (классу) «фрукты». Обычно каждому терму соответствует определенный тип - $x : A$. Типы позволяют строго говорить о возможных действиях над объектом, а также формализовать взаимоотношения между ними.

Система типов же, определяет правила взаимодействия между типами и термами. В программировании это понятие равноценно понятию типизация.

С помощью суждений можно создавать логические конструкции и *правила вывода*. Именно благодаря этому теория типов активно применяется в компиляторах в фазе статического анализа программы, как для вывода, так и для проверки соответствия типов. Более того, согласно изоморфизму Карри-Ховарда [4] (таблица 2), программы могут быть использованы для доказательства логических

высказываний. Такие доказательства называют автоматическими, и они широко применяются среди таких языков, как Agda, Coq, Idris.

Таблица 2. Изоморфизм Карри-Ховарда

Логическое высказывание	Язык программирования
Высказывание, F, Q	Тип, A, B
Доказательство высказывания F	$x : A$
Высказывание доказуемо	Тип A обитаем
$F \implies Q$	Функция, $A \rightarrow B$
$F \wedge Q$	Тип-произведение, $A \times B$
$F \vee Q$	Тип-сумма, $A + B$
Истина	Единичный тип, \top
Ложь	Пустой тип, \perp
$\neg F$	$A \rightarrow \perp$

Тип T обитаем (англ. inhabitat), если выполняется следующее: $\exists t : \Gamma \vdash t : T$

Наборы суждений образуют предположения (англ. assumptions), которые образуют контекст Γ . Правила вывода записываются следующим образом, например правило подстановки:

$$\frac{\Gamma \vdash t : T_1, \Delta \vdash T_1 = T_2}{\Gamma, \Delta \vdash t : T_2} \quad (1.1)$$

Выражение 1.1 можно трактовать следующим образом: если в контексте Γ терм t имеет тип T_1 , а в контексте Δ тип T_1 равен типу T_2 , то можно судить, что при наличии обоих контекстов, терм t имеет тип T_2 .

1.2.2 Система типов Хиндли-Милнера

В результате работы над научно-исследовательской работой, было принято решение использовать систему типов Хиндли-Милнера. Среди прочих ее особенностей, важно отметить то, что она способна вывести наиболее общий тип выражения, основываясь на аннотациях типов программиста и окружающем контексте. Предложено использовать её небольшую модификацию с добавлением типов-объединений и некоторых примитивных типов.

Наиболее классическим алгоритмом в этой области является так называемый алгоритм W [2].

Для определения системы типов необходимо 3 составляющие: набор термов, набор типов и набор суждений.

Термы:

$a, b, c ::=$

x	(переменная)
$\lambda x.a$	(лямбда-функция)
$a(b)$	(применение аргумента к функции)
$let\ a = b\ in\ c$	(объявление переменной)
$1, 2, 3, \dots$	(целочисленный литерал)
$1.1, 1.2, 10.0, \dots$	(вещественный литерал)
(a, b)	(объединение)

Типы:

$\iota ::=$	(примитивный тип)
$Integral$	(целочисленный)
$Floating$	(вещественный)
$\tau, \sigma ::=$	(мономорфный тип)
ι	
T	(переменная типа)
$\tau \rightarrow \sigma$	(функциональный тип)
(τ, σ)	(тип-объединение)
Λ	(пользовательский тип)
$\alpha ::=$	(полиморфный тип)
τ	
$\forall a.\alpha$	(параметрический тип)

Благодаря полиморфным типам имеется возможность определять обобщенные функции. Самый простой пример - функция id . Она имеет следующий тип:

$id : \forall a. a \rightarrow a$. Таким образом ее можно вызвать и с аргументом-числом, и с аргументом-функцией.

Работа алгоритма W строится на основе следующих суждений:

$$\overline{\Gamma \vdash x : \sigma} \quad (\text{TAUT})$$

$$\frac{\Gamma \vdash x : \sigma, \sigma' < \sigma}{\Gamma \vdash x : \sigma'} \quad (\text{INST})$$

Запись $\sigma' < \sigma$ означает, что тип σ' более конкретный, чем σ .

$$\frac{\Gamma \vdash x : \sigma, a \notin \text{free}(\Gamma)}{\Gamma \vdash x : \forall a. \sigma} \quad (\text{GEN})$$

$$\frac{\Gamma \vdash f : \tau \rightarrow \tau', x : \tau}{\Gamma \vdash f(x) : \tau'} \quad (\text{COMB})$$

$$\frac{\Gamma \cup x : \tau \vdash y : \tau'}{\Gamma \vdash \lambda x. y : \tau \rightarrow \tau'} \quad (\text{ABS})$$

$$\frac{\Gamma \vdash x : \sigma, \Gamma \cup y : \sigma \vdash z : \tau}{\Gamma \vdash (\text{let } y = x \text{ in } z) : \tau} \quad (\text{LET})$$

$$\frac{\Gamma \vdash x_1 : \tau_1, x_2 : \tau_2, x_3 : \tau_3, \dots}{\Gamma \vdash (x_1, x_2, x_3, \dots) : (\tau_1, \tau_2, \tau_3, \dots)} \quad (\text{TUPLE})$$

Исходя из этих суждений, алгоритм W составляет так называемое дерево вывода. Если дерево построить удалось, то написанная программа считается верно типизированной.

2 Программная реализация

2.1 Архитектура

В структуре программы почти любого компилятора можно выделить следующие части:

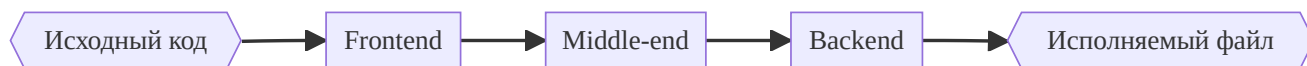


Рисунок 2. Архитектура большинства современных компиляторов

Сопоставляя эти части с рис. В.1 можно прийти к такому выводу: 1) к frontend части относятся лексический и синтаксический анализ, 2) к middle-end части - семантический анализ и генерация промежуточного представления, 3) к backend части - генерация машинного кода. Разработка frontend части была завершена ещё до этой работы и не будет подробно раскрываться.

Проект разрабатывается с использованием языка программирования Rust. Этот язык предлагает надежный концепт управления памятью, не имея при этом сборщика мусора [5]. Кроме того, он соперничает по скорости с C и C++ и применяется в довольно широком спектре приложений. Основные преимущества выбора этого языка:

- Rust работает быстрее за счёт использования мощных оптимизаторов, а так же применяет более строгие требования к разработке в целом,
- он предоставляет больше гарантий разработчику, как посредством его системы типов, так и другими средствами, например, borrow checker,
- система сборки создает нативный файл программы - его можно запустить, не имея на машине специальных сред выполнения.

Работать с большими проектами в разы удобнее и эффективнее при грамотном разбиении на модули. В экосистеме Rust такие модули именуются крейтами (англ. crates). На диаграмме ниже представлено разбиение на модули проекта Kodept. При этом сплошной линией выделено отношение модулей (от независимых к зависимым), пунктирной линией отмечена передача данных во время работы программы (от начала к концу). Под диагностиками необходимо понимать сообщения компилятора об исходной программе.

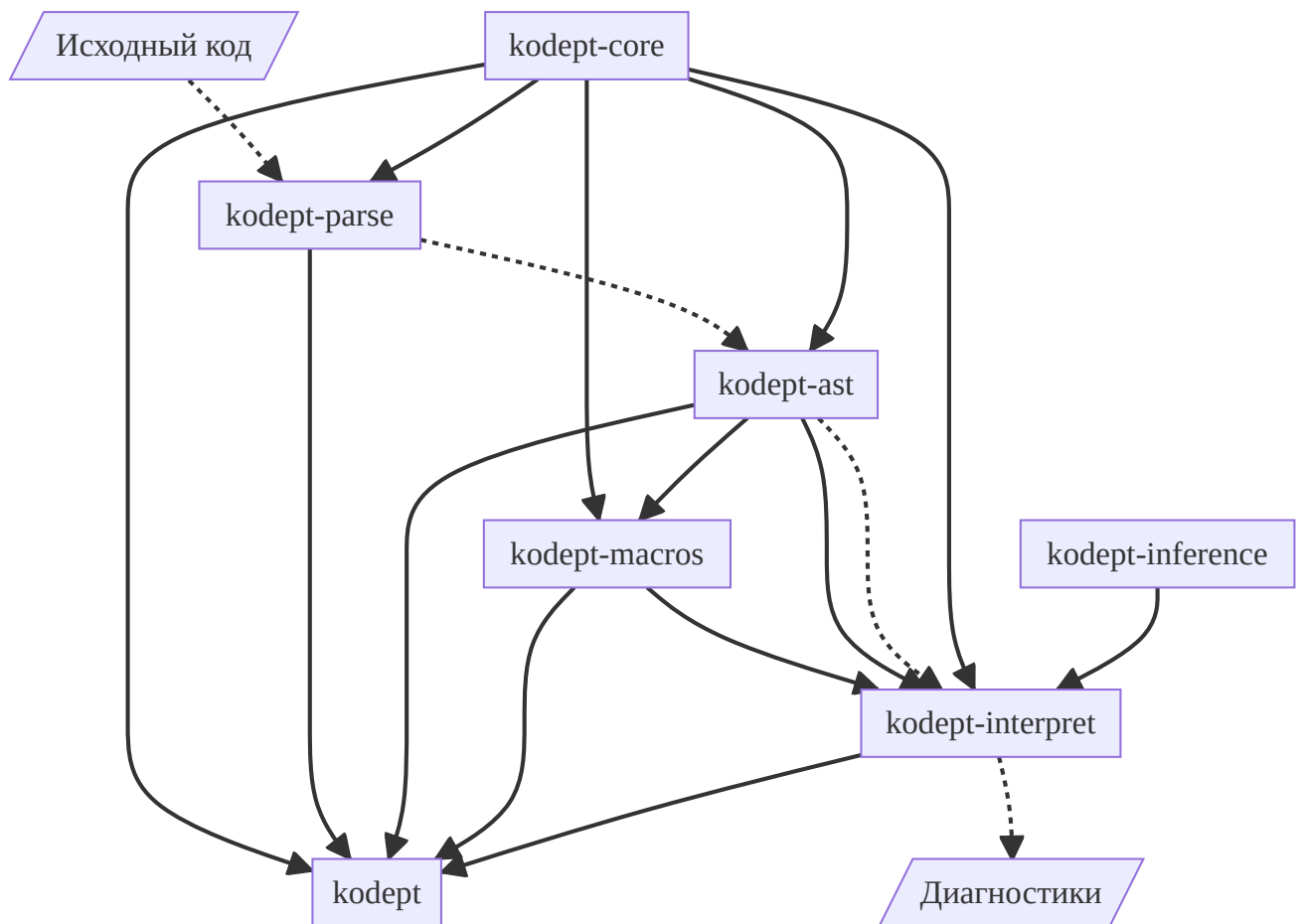


Рисунок 3. Иерархия модулей в проекте

В рамках этой работы внимание будет сконцентрировано вокруг модулей `kodept-ast`, `kodept-interpret` и `kodept-inference`. Однако, дадим кратное описание остальных модулей:

- `kodept-core` отвечает за определение основных структур данных, необходимых в остальных частях приложения, в частности, в нем определена структура *дерева разбора*,
- `kodept-parse` отвечает за лексический и синтаксический анализ, определяя набор синтаксических анализаторов (парсеров), которые генерируют дерево разбора,
- `kodept-macros` нужен для работы с AST и создания диагностик,
- `kodept` является корневым модулем и обеспечивает запуск стадий компилятора для входных файлов.

Дерево разбора - подробная версия AST, куда включается вся информация, полученная от парсеров. Нужно для восстановления конкретной точки в исходном коде при создании диагностики.

В модуле `kodept-ast` определена структура AST и его хранения. Рассмотрим некоторые проблемы, возникшие при проектировании этого модуля.

2.2 Проблема хранения абстрактного синтаксического дерева

Обычно структура AST реализована в программе в виде вложенных друг в друга структур с данными, описывающими тот или иной синтаксис языка (листинг 2.1).

Листинг 2.1. Псевдокод структур, представляющей собой описание синтаксиса функции в языке программирования.

```
1  struct Function {
2      vector<Parameter> parameters;
3      Type return_type;
4      Statement body;
5  }
6  struct Parameter {
7      string name;
8      Type type;
9  }
10 struct Type {
11     string identifier;
12 }
13 struct Statement {
14     // ...
15 }
```

Однако при таком подходе возникают определенные трудности. Чтобы написать обход AST, необходимо использовать исключительно рекурсивную реализацию. Это может стать проблемой при формировании большого AST, что его обход приведет к переполнению стека. Кроме того, гораздо большее неудобство возникает и при непосредственной реализации: нужно добавить по отдельной функции обхода для каждого узла дерева. При добавлении нового синтаксиса править код придется сразу в нескольких местах, а это усложняет поддерживаемость.

Поэтому необходимо было придумать более оптимизированный вариант для хранения AST. В итоге было решено переписать имеющиеся структуры так, чтобы они включали только те данные, которые непосредственно относятся к ним, а также включали цифровой идентификатор.

Листинг 2.2. Псевдокод структур после преобразования.

```
1  struct Function {  
2      int id;  
3  }  
4  struct Parameter {  
5      string name;  
6      int id;  
7  }  
8  struct Type {  
9      string identifier;  
10     int id;  
11 }  
12 struct Statement {  
13     // ...  
14     int id;  
15 }
```

Такая композиция позволяет хранить все объекты этих структур в одном месте, линейно. А с помощью индексов моделировать между ними взаимосвязь. Проще говоря, все свелось к хранению обычного графа, где вершины - идентификаторы. С таким видом гораздо удобнее работать, так как алгоритм перебора оперирует числами. Также хранение небольших структур в одном месте повышает вероятность попадания в кеш-память.

При применении «графового» хранения нарушается непосредственная взаимосвязь между структурами. Таким образом, становится непонятно, какие структуры могут быть в каких изначально были «вложены». Но это легко решается с помощью системы типов Rust. А именно: вводятся так называемые свойства принадлежности. Например, функция в качестве вложенных данных могла иметь вектор параметров, возвращаемый тип и тело. Значит, структура Parameter может выступать в качестве «ребенка» структуры Function. Также в структуру Function

добавляются методы для доступа к объявленным «детям». Таким образом сохраняется безопасность при работе с AST.

Листинг 2.3. Упрощенное представление структур после всех доработок на языке Rust

```
1 struct Function {
2     id: usize
3 }
4
5 struct Parameter {
6     id: usize,
7     name: string
8 }
9
10 impl HasChildrenMarker<Vec<Parameter>> for Function {
11     // ...
12 }
13
14 impl Function {
15     fn parameters(&self, /* ... */) -> Vec<&Parameter> {
16         // ...
17     }
18 }
```

Для наглядного изучения AST, было добавлено сохранение дерева в файл формата DOT [6]. В вершинах графа расположены имена синтаксических конструкций и числовой идентификатор. В ребрах - специальный тег, призванный отличать одни вершины от других. На рисунке А.1 изображена визуализация AST программы с листинга А.1.

2.3 Организация доступа к элементам абстрактного синтаксического дерева

С помощью модуля `kodept-macros` организовывается анализ AST. Во время этого элементы, составляющие дерево, могут быть изменены. К сожалению, из-за специфики Rust, в момент времени может существовать только 1 ссылка на объект, допускающая изменение (*exclusive mutable access*). Иногда разработчику

не обойтись без разделяемого изменяемого состояния. Для этого прибегают к использованию внутренней изменяемости (англ. interior mutability).

Interior mutability позволяет изменять внутренние данные неизменяемой структуры. Самым популярным способом является использование указателя с подсчетом ссылок (reference-counted pointer, RC) вкупе с проверкой заимствований во время выполнения (RefCell). Аналогом из C++ можно считать использование `shared_ptr`.

Листинг 2.4. Упрощенная реализация структур Rc и RefCell на псевдокоде

```
1 struct Rc<T> {  
2     int strong_count;  
3     int weak_count;  
4     T data;  
5 }  
6  
7 struct RefCell<T> {  
8     T data;  
9     int borrows_count;  
10 }
```

При анализе AST элементы могут поменяться, а структура - нет. Получается, что граф - неизменяемая структура, но его внутренние данные изменяемы. Это как раз подходит под понятие interior mutability. Но использование проверок во время выполнения, а также хранение дополнительных счетчиков для RC плохо влияет на производительность и использование памяти.

Предлагается использовать вместо привычной комбинации из RC + RefCell *zero-cost* абстракцию GhostCell [7]. Термином zero-cost или «нулевых затрат» называют полное отсутствие затрат во время выполнения - все необходимые преобразования выполняет компилятор во время компиляции. Основной идеей GhostCell является разделение ответственности за хранение и модификацию данных. Вводится специальный токен, который дает право изменять данные.

2.4 Реализация алгоритма W

За реализацию алгоритма отвечает модуль `kodect-inference`. В нем также находится определение необходимых термов и типов. Согласно рисунку 3, этот мо-

дуль не зависит ни от каких других. Система типов и алгоритм вывода могут быть определены, используя собственную модель. Таким образом, достаточно реализовать функцию по конвертации элементов AST в элементы модели.

2.4.1 Анализ областей видимости

Задачей этого анализа является разбиение AST на области видимости (англ. *scopes*), при этом строится так называемое дерево областей. Оно состоит из идентификаторов узлов AST, где каждый такой узел начинает новую область. Внутри области можно найти все объявленные переменные и функции, и удостовериться, что не используются необъявленные.

Например, для программы с листинга А.1, дерево областей будет таким:

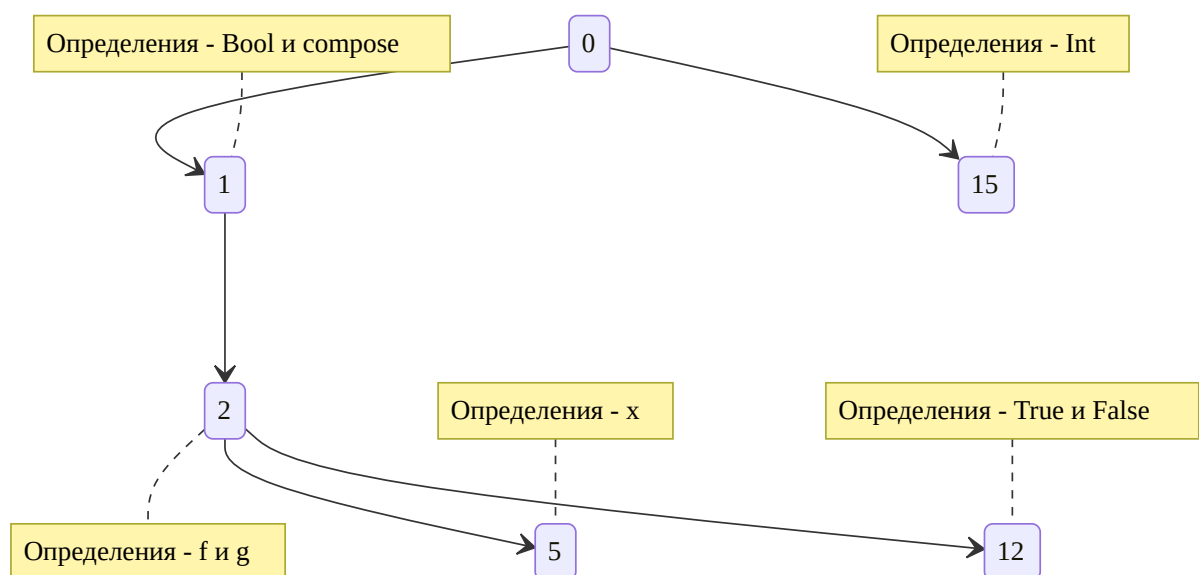


Рисунок 4. Дерево областей видимости

Для каждого определения из любой области можно применить алгоритм W.

В процессе преобразования модели AST (листинг Б.1) в термы (1.2.2), также образуются предположения. Они играют роль ограничений. Например, запись `val a: Int = expr` добавит предположение $a : Int$. Таким образом, алгоритм вывода типов проверит, что тип переменной `a` совпадает с типом выражения `expr`.

3 Тестирование и отладка

При разработке приложения не обойтись без тестирования. Оно помогает выявить различные ошибки и исправить их. Популярным вариантом тестирования является модульное тестирование (unit тестирование). Unit test - функция или набор функций, который проверяет корректность работы отдельного нетривиального куска программы.

В ходе разработки компилятора были написаны модульные тесты, они покрывают большое количество кода и успешно выполняются. Для запуска можно использовать систему сборки Cargo. Из папки с проектом следует запустить следующую команду: `cargo test --all-features --all --lib --no-fail-fast`. Cargo соберет проект и последовательно запустит все модульные тесты.

Компилятор Kodept является консольным приложением, поэтому для него был разработан интерфейс командной строки (CLI). С помощью него можно настроить вид выходных данных, поведение работы и др.

На текущий момент поддерживается 3 команды: справка по использованию, генерация графа AST в формате DOT и анализ файла.

Ранее уже были приведены примеры работы команды по генерации графа (A.1) для листинга A.1. Продемонстрируем работу механизма вывода типов на примере этого же кода. Для этого тоже можно воспользоваться Cargo: `cargo run -- -d examples/test.kd`. В результате в консоль будет выведены тип функции `compose`:

```
cargo run - -d examples/test.kd
```

```
kodept_interpret::type_checker: [compose:  $\forall a, b, c \Rightarrow (b \rightarrow a) \rightarrow (c \rightarrow b) \rightarrow c \rightarrow a$ ]
```

Действительно, если преобразовать эту функцию в термы, то получится $\lambda f, g. \lambda x. f(g(x))$. Тип этого выражения действительно совпадает с выведенным типом.

ЗАКЛЮЧЕНИЕ

В результате данной работы был реализован механизм вывода типов для языка программирования Kodept. Показано, что программа действительно может правильно определить тип выражения. Однако развитие проекта на этом не останавливается. Планируется реализация следующего шага в компиляции программы - генератора промежуточного представления.

В ходе работы рассмотрены некоторые проблемы, возникшие при работе с абстрактным синтаксическим деревом. Приведены способы их решения. Успешно решены все поставленные задачи, а именно:

- 1) сформирована модель абстрактного синтаксического дерева,
- 2) реализован семантический анализатор, включающий в себя анализатор областей видимости, преобразователь в термы системы типов Хиндли-Милнера и непосредственно механизм вывода типов на основе этой системы.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Голованов Вячеслав. Насколько близко компьютеры подошли к автоматическому построению математических рассуждений? [Электронный ресурс]. 2020. (Дата обращения 22.04.2024)). URL: <https://habr.com/ru/articles/519368/>.
- 2 Urban Christian, Nipkow Tobias. Nominal verification of algorithm W // From Semantics to Computer Science. Essays in Honour of Gilles Kahn / под ред. G. Huet, J.-J. Lévy, G. Plotkin. Cambridge University Press, 2009. С. 363–382.
- 3 Milner Robin. A theory of type polymorphism in programming // Journal of computer and system sciences 17. 1978. С. 348–375.
- 4 Свиридов Сергей. Теория типов [Электронный ресурс]. 2023. (Дата обращения 19.04.2024). URL: <https://habr.com/ru/articles/758542/>.
- 5 badcasedaily1. Как работает управление памятью в Rust без сборщика мусора [Электронный ресурс]. (Дата обращения 15.04.2024). URL: <https://habr.com/ru/companies/otus/articles/787362/>.
- 6 GraphViz Documentation [Электронный ресурс]. 2022. (Дата обращения 24.04.2024). URL: <https://www.graphviz.org/documentation/>.
- 7 GhostCell: separating permissions from data in Rust / Joshua Yanovski, Hoang-Hai Dang, Ralf Jung [и др.] // Proc. ACM Program. Lang. New York, NY, USA, 2021. aug. Т. 5, № ICFP. 30 с. URL: <https://doi.org/10.1145/3473597>.

Выходные данные

Никитин В.Л.. Разработка механизма вывода типов с использованием системы типов Хиндли-Милнера по дисциплине «Модели и методы анализа проектных решений». [Электронный ресурс] — Москва: 2024. — 29 с. URL: <https://sa2systems.ru:88> (система контроля версий кафедры РК6)

Постановка:



доктор физико-математических наук, Соколов А.П.

Решение и вёрстка:




студент группы РК6-75Б, Никитин В.Л.

2024, осенний семестр

АКТ
проверки выпускной квалификационной работы

Фамилия Имя Отчество
(Фамилия, имя, отчество)

Объем заимствования составляет 2 % текста, что с учетом корректного заимствования соответствует / не соответствует требованиям к ВКР _____
ненужное зачеркнуть *бакалавра, специалиста, магистра*



 (подпись)

С.В. Грошев

 (ФИО)

И.О. Фамилия
(ФИО)

(ПОДПИСЬ)

подписать

руки

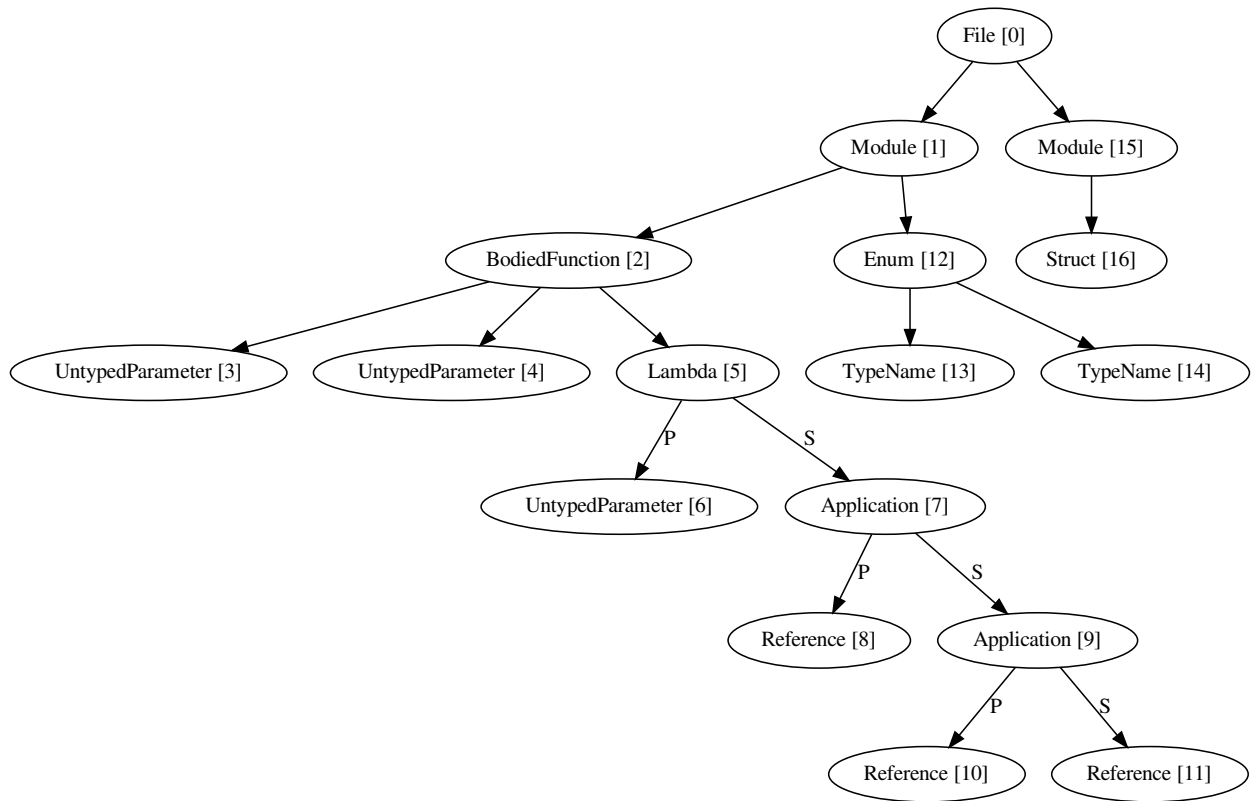


Рисунок А.1. Изображение структуры абстрактного синтаксического дерева

Листинг А.1. Исходная программа на языке Kodept

```

1 module Testing {
2   fun compose(f, g) => \x => f(g(x))
3
4   enum struct Bool { True, False }
5 }
6
7 module Testing2 {
8   struct Int
9 }
  
```

Б

Листинг Б.1. Имена всех элементов, составляющих абстрактное синтаксическое дерево

```

1  File, // root element
2  Module, // module name rest
3  Struct, // struct name(params) rest
4  Enum, // enum name rest
5  TypedParameter, // name: type
6  UntypedParameter, // name
7  Variable, // val name: type
8  InitializedVar, // val name: type = expr
9  BodiedFunction, // fun name(params) => expr
10 ExpressionBlock, // expr1; expr2; ...
11 Application, // expr(expr)
12 Lambda, // \binds => expr
13 Reference, // name
14 Access, // expr.expr
15 Number, // number literal
16 Char, // char literal
17 String, // string literal
18 Tuple, // (expr1, expr2, ...)
19 If, // if expr => expr другие ветки
20 Elif, // elif expr => expr
21 Else, // else expr
22 Binary, // binary operator: +, -, *, /, %, ^
23 Unary, // unary operator: -, +, !, ~
24 AbstractFunction, // abstract fun name(params): type
25 ProdType, // (type1, type2, ...)

```
