

## Todo list

Санкт-Петербургский государственный политехнический  
университет  
Институт информационных технологий и управления  
Кафедра компьютерных систем и программных технологий

Диссертация допущена к защите  
зав. кафедрой

\_\_\_\_\_ В.Ф. Мелехин

«\_\_\_\_» \_\_\_\_\_ 2014 г.

## **ДИССЕРТАЦИЯ на соискание ученой степени МАГИСТРА**

**Тема: Инструментальная среда для анализа  
программных систем**

Направление: 230100 – Информатика и вычислительная техника  
Магистерская программа: 230100.68.15 – Технологии проектирования  
системного и прикладного программного обеспечения

Выполнил студент гр. 63501/13

\_\_\_\_\_ А.М. Половцев

Научный руководитель,  
к. т. н., доц.

\_\_\_\_\_ В.М. Ицыксон

Консультант по нормоконтролю,  
ст. преп.

\_\_\_\_\_ С.А. Нестеров

Эта страница специально оставлена пустой.

Санкт-Петербургский государственный политехнический  
университет  
Институт информационных технологий и управления  
Кафедра компьютерных систем и программных технологий

УТВЕРЖДАЮ  
Зав. кафедрой  
\_\_\_\_\_, В.Ф. Мелехин  
«\_\_\_» \_\_\_\_\_ 2014 г.

**ЗАДАНИЕ**  
**на магистерскую диссертацию**  
**студента Половцева Александра Михайловича**

1. **Тема работы:** Инструментальная среда для анализа программных систем.
2. **Срок сдачи студентом законченной работы:** хх.хх.2014
3. **Содержание расчетно-пояснительной записки:**
  - 1) Анализ подходов и средств построения инструментальных средств
  - 2) Постановка задач и требований к среде
  - 3) Проектирование архитектуры инструментальной среды
  - 4) Реализация инструментальной среды
  - 5) Тестирование разработанной программной системы
4. **Дата выдачи задания:** хх.хх.2013

Руководитель: \_\_\_\_\_ (В.М. Ицыксон)  
Задание принял к исполнению «хх» сентября 2013:  
\_\_\_\_\_ (А.М. Половцев)

Эта страница специально оставлена пустой.

# РЕФЕРАТ

Отчет, 71 стр., 17 рис., 2 табл., 19 ист., 2 прил.

# ABSTRACT

Report, 71 pages, 17 figures, 2 tables, 19 references, 2 appendices





# СОДЕРЖАНИЕ

<b>ВВЕДЕНИЕ</b> . . . . .	11
<b>1. АНАЛИЗ ПОДХОДОВ К ПОСТРОЕНИЮ ИНСТРУМЕНТАЛЬНЫХ СРЕДСТВ</b> . . . . .	13
1.1. Методы повышения качества . . . . .	13
1.2. Классификация методов обеспечения качества . . . . .	14
1.3. Модели программных систем . . . . .	16
1.4. Метамодел . . . . .	18
1.5. Анализ существующих решений . . . . .	19
1.5.1. Критерии сравнения . . . . .	19
1.5.2. Обзор средств . . . . .	20
1.5.3. Выводы . . . . .	26
<b>2. ПОСТАНОВКА ЗАДАЧИ</b> . . . . .	29
2.1. Формулирование требований к инструментальной среде . . . . .	29
2.2. Выбор пути решения . . . . .	30
2.3. Вывод . . . . .	31
<b>3. ПРОЕКТИРОВАНИЕ АРХИТЕКТУРЫ ИНСТРУМЕНТАЛЬНОЙ СРЕДЫ</b> . . . . .	33
3.1. Структура программной системы . . . . .	33
3.2. Проектирование метамодел . . . . .	34
3.2.1. Стандарт MOF . . . . .	35
3.2.2. Иерархия моделей MOF . . . . .	36
3.2.3. Структура MOF . . . . .	37
3.2.4. Сериализация метамодел . . . . .	39
3.3. Проектирование архитектуры преобразователей . . . . .	41
3.4. Проектирование графического интерфейса . . . . .	43
3.5. Вывод . . . . .	44
<b>4. РАЗРАБОТКА ИНСТРУМЕНТАЛЬНОЙ СРЕДЫ</b> . . . . .	45
4.1. Выбор средств разработки . . . . .	45
4.2. Структура проекта . . . . .	45
4.3. Разработка метамодел . . . . .	47
4.3.1. Проведение операций над метамоделью . . . . .	51
4.3.2. Выбор фреймворка для сериализации . . . . .	54

4.3.3.	Вывод . . . . .	57
4.4.	Разработка преобразователей . . . . .	58
4.4.1.	Выбор генератора лексического и синтаксического анализаторов . . . . .	58
4.4.2.	Правила использования генератора парсеров ANTLR . . . . .	61
4.4.3.	Реализация преобразователя для языка Java . . . . .	62
<b>ЗАКЛЮЧЕНИЕ . . . . .</b>		<b>65</b>
<b>СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ . . . . .</b>		<b>67</b>
<b>ПРИЛОЖЕНИЕ А. ПРИМЕР МЕТАМОДЕЛИ, СЕРИАЛИЗОВАННОЙ В XML . . . . .</b>		<b>69</b>
<b>ПРИЛОЖЕНИЕ Б. ГРАММАТИКА ЯЗЫКА JAVA ДЛЯ ПОСТРОЕНИЯ МОДЕЛИ АНАЛИЗИРУЕМОЙ ПРОГРАММНОЙ СИСТЕМЫ . . . . .</b>		<b>71</b>

# ВВЕДЕНИЕ

В данной работе рассматривается подход к автоматизации процесса проведения анализа и верификации программных систем с целью повышения характеристик качества.

С развитием вычислительных систем и ростом в них доли программной составляющей, сложность разрабатываемых программ постоянно возрастает. Также, вследствие большой конкуренции на рынке программного обеспечения, постоянно снижаются сроки разработки новых версий ПО. Эти факторы неизбежно ведут к снижению качества выпускаемых продуктов.

Падение уровня качества является проблемой, особенно если программное обеспечение задействовано в критически важных сферах человеческой деятельности, например, медицине и космонавтике, так как наличие в них ошибок ведет к большому материальному ущербу и даже человеческим жертвам. Поэтому задача повышения качества является одной из самых актуальных в сфере информационных технологий.

Одними из способов повышения качества программ являются аудит, статический анализ и формальные методы, которые часто реализуются в виде инструментальных средств. При разработке данных средств часто решаются похожие задачи, такие как:

- Построение моделей программы, например, абстрактного синтаксического дерева, графа потока управления, графа программных зависимостей и т.д. (модели программ рассмотрены в подразделе 1.3)
- Построение различных метрик программного кода
- Реинжиниринг программного обеспечения (оптимизация, рефакторинг и т.п.)
- Визуализация свойств программной системы
- и т.п.

Обычно эти задачи решаются вручную каждый раз при создании анализаторов или проведения верификации программы. В данной работе предлагается способ автоматизации решения этих задач на осно-

ве использования представления программы, не зависящего от языка написания ее исходного кода.

В разделе 1 приведены способы повышения качества, используемые в них модели ПО, предложен подход к автоматизации решения задач повышения качества, рассмотрены существующие решения. Раздел 2 посвящен формулированию требований к инструментальной среде и постановке задач. В разделе 3 рассматривается архитектура инструментальной среды и всех ее составляющих. Разработка системы приведена в разделе 4. Раздел 5 посвящен тестированию разработанной системы.

# 1. АНАЛИЗ ПОДХОДОВ К ПОСТРОЕНИЮ ИНСТРУМЕНТАЛЬНЫХ СРЕДСТВ

## 1.1. Методы повышения качества

Существует две группы подходов по обеспечению качества программного обеспечения [1]:

1. Подходы, основанные на синтезе ПО
2. Подходы, основанные на анализе уже созданного ПО

Подходы, основанные на синтезе ПО, используют различные формализации во время проектирования системы, таким образом позволяя избежать ошибок на более поздних этапах разработки.

Данные формализации включают в себя:

- формальные спецификации
- формальные и неформальные описания различных аспектов программной системы
- архитектурные шаблоны и стили
- паттерны проектирования
- генераторы шаблонов программ
- генераторы программ
- контрактное программирование
- аннотирование программ
- верификация моделей программ с использованием частичных спецификаций
- использование моделей предметной области для автоматизации тестирования программ

Подходы, основанные на анализе уже созданного ПО, используются для повышения качества уже созданного ПО, что позволяет улучшить огромное количество уже разработанных программных систем, имеющих проблемы с уровнем качества.

Данная группа подходов оперирует функциональными и нефункциональными требованиями к разработанной системе для доказательства соответствия или приведения контрпримеров, показывающих несоответствие поставленным требованиям.

## 1.2. Классификация методов обеспечения качества

Обычно выделяют следующие базовые классификации методов обеспечения качества [1]:

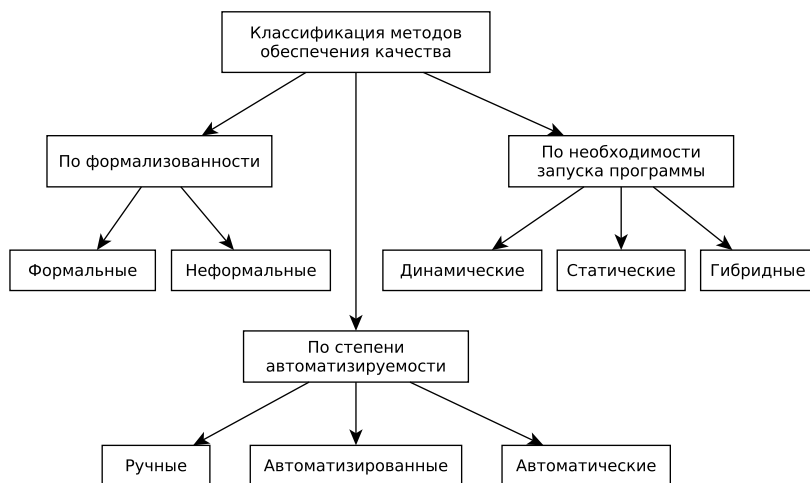


Рисунок 1.1. Схема используемой классификации методов верификации

По формализованности процедур обеспечения качества методы подразделяются на следующие категории:

**Формальные методы** позволяют создавать формальные функциональные спецификации и модели архитектуры систем, а также

осуществлять их преобразование в программы с последующей верификацией [2]. Корректность полученных результатов гарантируется математическим аппаратом. К таким методам относятся, например, дедуктивная верификация, проверка моделей и абстрактная интерпретация.

Эти методы можно применить только к тем свойствам, которые можно выразить в рамках некоторой математической модели. Построение этой модели не автоматизируется, а провести анализ таких моделей может лишь специалист. Однако сама проверка свойств может быть автоматизирована и позволяет находить даже самые сложные ошибки.

**Неформальные методы** позволяют находить ошибки, используя различные артефакты жизненного цикла системы, в отличие от формальных методов, и позволяет находить большое множество разновидностей ошибок. К ее недостаткам можно отнести невозможность автоматизации [3]. Примером неформального метода является аудит.

По степени автоматизируемости процедур методы подразделяются на:

**Ручные методы** - группа методов, не подлежащая автоматизации. К этому классу методов обычно относятся все неформальные методы (аудит, парное программирование).

**Автоматизированные методы** предполагают частичную автоматизацию процесса повышения качества. Данные методы предполагают проведение ручных действий (например, по подготовке модели в методе проверки моделей) и запуск автоматизированных процедур (над подготовленными моделями).

**Автоматические методы** не требуют вмешательства специалистов во время своего выполнения, что позволяет легко интегрировать их в жизненный цикл разработки ПО. Примером таких методов является модульное регрессионное тестирование.

По необходимости запуска анализируемой программы выделяют следующие классы методов:

**Статические методы** - это процесс выявления ошибок и недочетов в исходном коде программ или каких-либо иных артефактов (например, формальные спецификации и аннотации). От остальных методов их отделяет то, что статические методы использует только исходные тексты программы, что позволяет обнаруживать ошибки на стадии написания кода. Таким образом, при анализе отсутствует спецификация программы - описание того, что она делает. Это умень-

пает множество обнаруживаемых ошибок, но позволяет полностью автоматизировать процесс анализа. Примером статических методов является статический анализ.

**Динамические методы** используются для анализа и оценки свойств программной системы по результатам ее реальной работы. Одними из таких методов являются тестирование и анализ трасс исполнения.

Для применения данных методов необходимо иметь работающую систему (или ее прототип), поэтому их нельзя использовать на ранних стадиях разработки. Также данные методы позволяют найти только те ошибки в ПО, которые проявляются в его работе.

**Гибридные методы** объединяют в себе элементы некоторых способов повышения качества, описанных выше. Примерами таких методов являются тестирование на основе моделей (model driven testing) [4] и мониторинг формальных свойств (runtime verification) [5]. Цель таких методов - объединить преимущества уже используемых подходов.

### 1.3. Модели программных систем

Одной из важнейших составляющих анализа программных систем является построение модели. Без нее анализатор будет вынужден непосредственно оперировать с исходным кодом, что влечет за собой усложнение процедур анализа и самого анализатора в целом.

В зависимости от способа построения и назначения модели, они могут различаться по структуре и сложности и обладать различными свойствами. Существуют следующие виды моделей [1]:

- Структурные модели
- Поведенческие модели
- Гибридные модели

Структурные модели во основном используют информацию о синтаксической структуре анализируемой программы, в то время как поведенческие - информацию о динамической семантике. Гибридные модели используют оба этих подхода.



## Структурные модели

### 1. Синтаксическое дерево

Синтаксическое дерево является результатом разбора программы в соответствии с формальной грамматикой языка программирования. Вершины этого дерева соответствуют нетерминальным символам грамматики, а листья - терминальным.

### 2. Абстрактное синтаксическое дерево

Данная модель получается из обычного синтаксического дерева путем удаления нетерминальных вершин с одним потомком и замены части терминальных вершин их семантическими атрибутами.

## Поведенческие модели

### 1. Граф потока управления

Граф потока управления представляет потоки управления программы в виде ориентированного графа. Вершинами графа являются операторы программы, а дуги отображают возможный ход исполнения программы и связывают между собой операторы, выполняемые друг за другом.

### 2. Граф зависимостей по данным

Граф зависимостей по данным отображает связь между конструкциями программы, зависимыми по используемому данным. Дуги графа соединяют узлы, формирующие данные, и узлы, использующие эти данные.

### 3. Граф программных зависимостей

Данная модель объединяет в себе особенности графа потока управления и графа зависимости по данным. В графе программных зависимостей присутствуют дуги двух типов: информационные дуги отображают зависимости по данным, а дуги управления соединяют последовательно выполняемые конструкции.

### 4. Представление в виде SSA

Однократное статическое присваивание (static single assignment) - промежуточное представление программы, которое обладает следующими свойствами:

- Всем переменным значение может присваиваться только один раз.
- Вводится специальный оператор  $\phi$ -функция, который объединяет разные версии локальных переменных.
- Все операторы программы представляются в трехоперандной форме.

## Гибридные модели

### 1. Абстрактный семантический граф

Данная модель является расширением абстрактного синтаксического дерева путем добавления дуг, отражающих некоторые семантические свойства программы, например, такие дуги могут связывать определение и использование переменной или определение функции и ее вызов.

## 1.4. Метамоделей

Обобщенное решение задач анализа и верификации основывается на использовании *метамоделей* [6]. В подразделе 1.3 были приведены возможные модели, которые могут быть извлечены из описания системы на каком-либо языке программирования и отображающие различные ее свойства. Аналогичным образом, как модели определяют язык описания системы, метамоделей определяют язык описания моделей.

Для решения задач проведения различного вида анализа необходимо разработать такую мета модель, которая бы описывала необходимый набор моделей. Наиболее подходящей мета моделью является описание на уровне сущностей анализируемой системы (объектов или функций) и взаимодействия между ними (вызов функции, обращение к полю объекта). Детализация взаимодействия между сущностями при этом должна соответствовать поставленным задачам, т.е. существует проблема компромисса между степенью детализации мета модели (количеством информации, которое можно из нее извлечь) и уровнем абстракции (независимостью от способа описания анализируемой системы). С увеличением степени детализации повышается мощность модели, но тем самым увеличивается ее сложность, так как необходимо учитывать специфические для каждого языка возможности описания взаимодействия объектов программной системы. С ро-

стом же уровня абстракции уменьшается класс задач, для которых применима данная метамодель.

Таким образом, приведенные ранее проблемы можно решить, разработав метамодель, которая бы описывала набор моделей, хранящий необходимое количество информации для проведения анализа, но обладающей достаточной простотой для описания широкого круга программных систем.

## 1.5. Анализ существующих решений

На данный момент существует лишь небольшое количество инструментов, использующих метамоделирование. Необходимо сравнить их и выявить их пригодность к решению поставленных задач.

### 1.5.1. Критерии сравнения

Для сравнения существующих средств введем следующие критерии оценки:

№	Критерий	Описание
1	Извлечение моделей программ	Возможность получения различных моделей (AST, CFG) из анализируемой программы
2	Подсчет метрик	Возможность расчета метрик
3	Визуализации	Наличие различных визуализаций свойств системы
4	Свойства метамодели	Свойства используемой метамодели - независимость от языка, простота использования
5	API	Язык программирования, на котором написан API системы
6	Открытость	Возможность бесплатного использования и наличие открытых исходных кодов

Таблица 1.1. Критерии сравнения средств метамоделирования

### 1.5.2. Обзор средств

#### Фреймворк Moose

Moose является платформой для анализа программ и поддерживает большое количество различных видов анализа [7]:

1. Построение и визуализация метрик.
2. Обнаружение клонов.
3. Построение графа зависимостей между пакетами.
4. Вывод словаря, используемого в проекте.
5. Поддержка браузеров исходного кода.

Moose использует целое семейство метамodelей под названием FAMIX. Данное семейство обладает довольно сложной структурой, упрощенный вид которой приведен на рис 1.2:

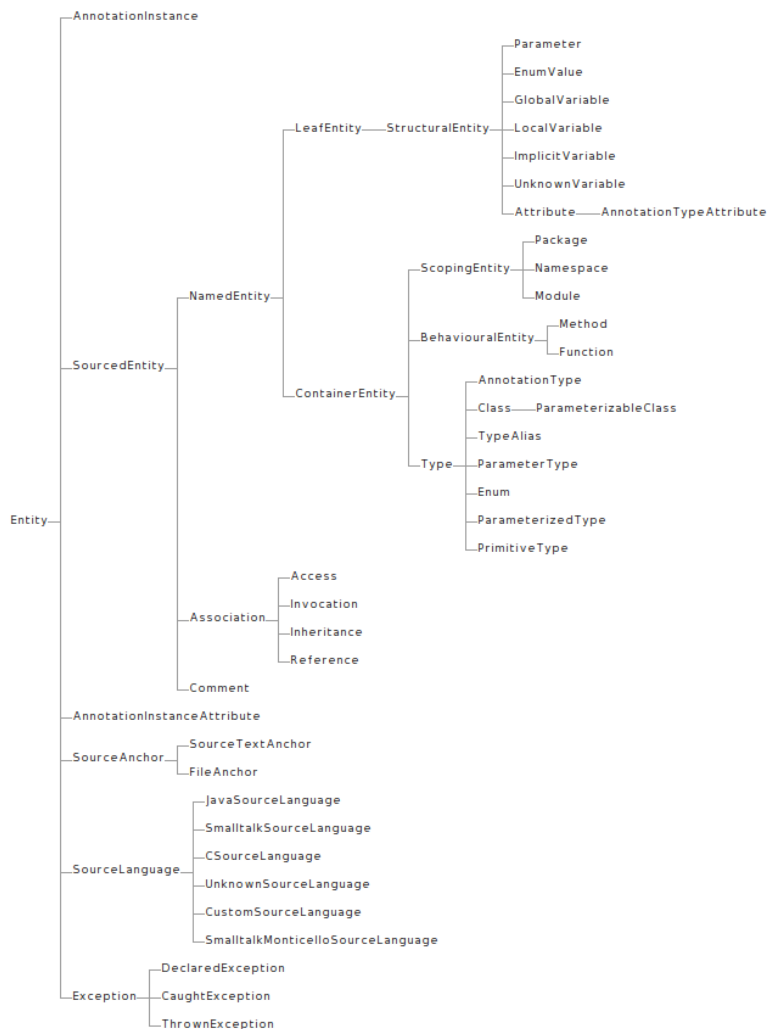


Рисунок 1.2. Структура метамodelей семейства FAMIX

Анализ программы происходит следующим образом:

1. Импортрование входных данных. Импортрование может про-

исходить как при помощи встроенных средств (Moose поддерживает Smalltalk, XML и MSE), так и при помощи сторонних средств.

2. После импортирования данные приводятся к одной из метамodelей семейства FAMIX.
3. Применение заданных алгоритмов анализа.

Архитектура средства приведена на рис 1.3

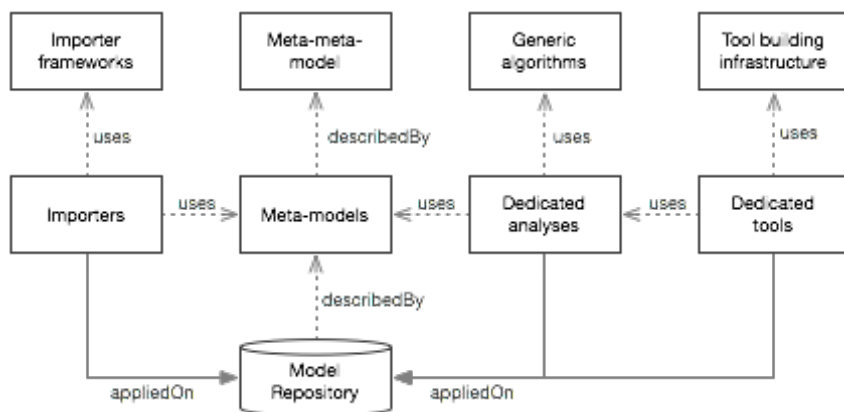


Рисунок 1.3. Архитектура фреймворка Moose

К недостаткам данной среды, можно отнести то, что Moose нацелен в первую очередь на задачи реинжиниринга и обладает слишком избыточной и громоздкой метамodelью для разработки на ее основе алгоритмов статического анализа и верификации.

## Средство SMIILE

SMIILE - среда, предназначенная для вычисления метрик анализируемой системы [8].

В качестве метамodelи SMIILE использует представление в виде eCST (enriched Concrete Syntax Tree), которое представляет собой дерево разбора программы с добавлением универсальных узлов, что

позволяет сделать дерево разбора независимым от входного языка программирования.

На рис 1.4 изображена архитектура данного средства:

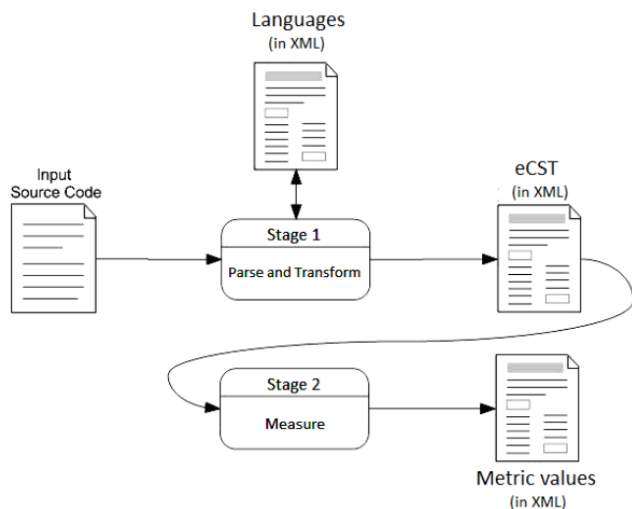


Рисунок 1.4. Архитектура SMILE

Анализ программы происходит в две фазы:

#### 1. Фаза 1

- Определение языка программирования, на котором написана анализируемая программа.
- Вызов парсера этого языка для построения CST и преобразование в eCST.
- Вывод результата в формате XML.

#### 2. Фаза 2

- Считывание eCST из файла.
- Подсчет метрик.
- Сохранение результата в формате XML.

Средство SMILE на данный момент не разрабатывается и, помимо подсчета метрик, не предоставляет никаких других возможностей.

## Фреймворк LLVM

LLVM - фреймворк для анализа и трансформации программ, путем предоставления информации для трансформаций компилятору во время компиляции, линковки и исполнения [9].

LLVM использует промежуточное представление, в основе которого лежит представление в виде SSA. Промежуточное представление является набором RISC-подобных команд и содержит дополнительную информацию более высокого уровня, например информацию о типах и графе потока управления.

Данный фреймворк обладает следующими особенностями:

1. Сохранение информации о программе даже во время исполнения и между запусками.
2. Предоставление информации пользователю для профилирования и оптимизации.
3. Промежуточное представление не зависит от языка программирования.
4. Возможность оптимизации всей системы в целом (после этапа линковки).

Архитектура LLVM приведена на рис 1.5:

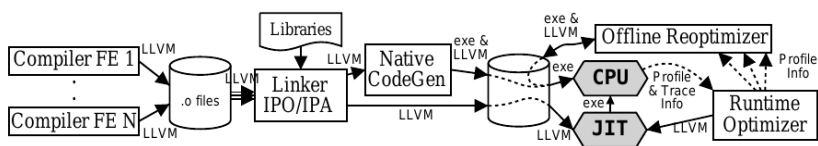


Рисунок 1.5. Архитектура LLVM

Front-end компиляторы транслируют исходную программу в промежуточное представление LLVM, которое затем компонуется LLVM-линкером. На этой стадии может проводиться межпроцедурный анализ. Получившийся код затем транслируется в машинный код для целевой платформы.



Однако, промежуточное представление, используемое в LLVM, обладает недостаточной полнотой описания, необходимой для подсчета метрик и визуализации свойств программной системы.

## Фреймворк ULF-Ware

ULF-Ware является средством для генерации кода из модели, созданной при помощи языка SDL (Specification and Description Language) [6]. Данное средство использует метамодель под названием CeeJay, которая предназначена для описания систем на языках Java и C++.

Архитектура ULF-Ware приведена на рис 1.6.

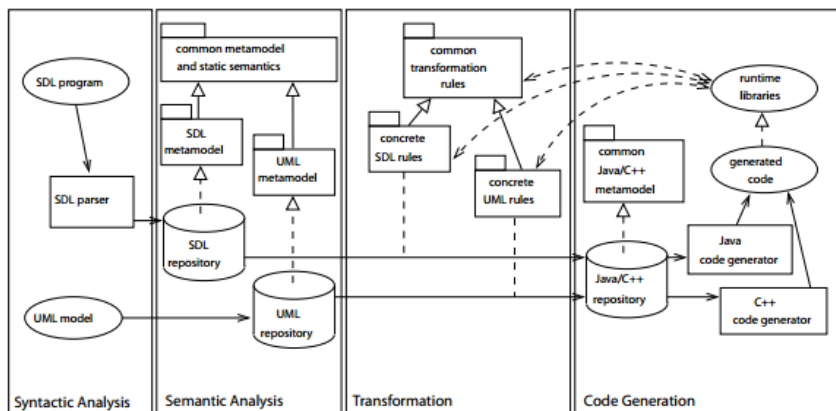


Рисунок 1.6. Архитектура ULF-Ware

Генерация кода происходит следующим образом: компилятор SDL создает модель программы, полученную из SDL-описания, соответствующую метамодели SDL. После этого происходит трансформация исходной модели в экземпляр метамодели CeeJay, который обладает всей необходимой информацией для генерации кода на языке Java или C++.

Ограничением данного средства и метамодели CeeJay является тот факт, что, в силу специфики задачи, данная метамодель может применяться только для языков Java и C++.

### 1.5.3. Выводы

Результаты сравнения существующих средств на основе критериев из подразд. 1.5.1, приведены в таблице 1.2.

На основе проведенного анализа было выявлено, что наиболее подходящим средством (из рассмотренных) для решения поставленных проблем является среда Moose. Однако, метамодель, используемая Moose отличается высокой сложностью взаимодействия и работы с ней, так как она предназначена для решения слишком широкого круга задач. К тому же данное средство ориентировано в основном на визуализацию свойств, мало связанных с задачами статического анализа и верификации.

Поэтому, для решения задач построения моделей программ и подсчета метрик было принято решение о разработке языконезависимой метамодели и инструментальной среды, позволяющей извлекать эти модели и визуализировать свойства анализируемого ПО.

Название	Модели	Метрики	Визуализации	Свойства	API	Лицензия
Moose	-	Метрики размера (количество методов, количество атрибутов, количество свойств)	Визуализация метрик, наличие клонов, зависимости между пакетами и т.д.	Не зависит от языка, высокий уровень сложности использования	Smalltalk	BSD и MIT
SMILE	-	Метрики Холстеда, OO-метрики, цикломатическая сложность и т.д.	-	Не зависит от языка	- (средство не доступно)	-
LLVM	SSA, CFG	-	-	Не зависит от языка, имеет модель в виде SSA	C++	NCSA Open Source License
ULF-Ware	-	-	-	Только языки Java и C++	- (средство не доступно)	-

Таблица 1.2. Результаты сравнения существующих решений



## 2. ПОСТАНОВКА ЗАДАЧИ

Как было сказано в разд. 1.3, инструменты для проведения статического анализа и верификации используют различные модели программ для облегчения процедур анализа. Однако, данные модели зависят от языка, на котором написана система, а, следовательно, при таком подходе невозможно обобщить разработанные алгоритмы.

К тому же проблема зависимости процедур анализа программно-го обеспечения от исходного текста остро стоит не только при проведении верификации, но и при решении задач реинжиниринга и оптимизации, а именно:

- Построение метрик
- Визуализация свойств системы
- Поиск клонов
- Анализ истории проекта

### 2.1. Формулирование требований к инструментальной среде

Таким образом, исходя из описанных проблем, ставится задача разработки среды, которая бы позволила абстрагировать алгоритмы анализа и реинжиниринга от языка описания системы. Это позволит применять разработанные средства для гораздо более широкого круга систем, автоматизировав процесс извлечения модели.

Ядром всей системы и средством абстракции является метамодель, которая должна обладать следующими свойствами:

1. Независимость от языка описания системы - метамодель должна поддерживать несколько парадигм программирования (как минимум структурную и объектно-ориентированную парадигмы) и обладать достаточной полнотой для описания специфических конструкций конкретного языка.
2. Расширяемость - возможность добавлять новые сущности по мере необходимости.

3. Простота в использовании - метамодель должна предоставлять удобный API для облегчения обхода структуры модели и реализации различных процедур анализа ПО. Исходя из возможного изменения структуры метамодели API должен предусматривать возможность работы с добавленными узлами.
4. Полнота - метамодель должна содержать достаточное количество информации для извлечения различных моделей, описанных в подразд. 1.3, а также других свойств анализируемой системы (например, метрик).

Для демонстрации возможностей метамодели необходимо разработать инструментальную среду, позволяющую выполнять следующие действия:

1. Визуализация извлеченных моделей
2. Подсчет метрик
3. Визуализация дополнительных свойств системы (например, диаграммы классов для объектно-ориентированной системы)
4. Импорт систем на языках Java и C (являющимися наиболее популярными языками программирования и реализующие две разные парадигмы)

## 2.2. Выбор пути решения

Как видно из обзора в разд. 1.5, ни одно из рассмотренных средств не отвечает поставленным требованиям, поэтому было принято решение о написании собственной инструментальной среды. При этом предполагается решить следующие задачи:

1. Проектирование структуры метамодели, отвечающей всем поставленным требованиям
2. Проектирование графической инструментальной среды, поддерживающей все необходимые виды визуализации
3. Реализация и тестирование программной системы
4. Анализ полученных результатов

## 2.3. Вывод

В результате успешного выполнения задач, приведенных в подразд. 2.2, необходимо получить метамодель, позволяющую описывать программные системы, абстрагируясь от использованного языка программирования, а также инструментальную среду для визуализации ее различных свойств.





### 3. ПРОЕКТИРОВАНИЕ АРХИТЕКТУРЫ ИНСТРУМЕНТАЛЬНОЙ СРЕДЫ

При проектировании архитектуры инструментальной среды необходимо учитывать следующие особенности:

- Часть программы, зависящую от языка программирования, на котором написано анализируемое ПО, необходимо вынести в отдельную составляющую для упрощения добавления новых парсеров языков, тем самым увеличивая множество доступных систем для анализа.
- Необходимо отделить метамодель от процедур анализа. Таким образом можно расширять набор процедур анализа и визуализации, не изменяя структуру метамодели.
- Крайне сложно найти компромисс между уровнем абстракции и степенью детализации метамодели, поэтому необходимо предусмотреть возможность расширения, путем добавления новых узлов в существующую метамодель.

#### 3.1. Структура программной системы

На основе анализа поставленных требований, существующих решений и перечисленных замечаний была предложена следующая архитектура программной системы:

Как видно из рис. 3.1 система состоит из трех частей:

1. Центральная часть системы - *метамодель*. Над ней производятся все операции по проведению анализа и графического отображения свойств анализируемой системы.
2. *Преобразователи* отвечают за импортирование исходного кода анализируемой системы и его преобразование в промежуточное представление (сериализованную метамодель). Преобразователи являются единственной частью системы, зависящей от языка программирования, на котором написана анализируемая система. Однако, так как инструментальная среда оперирует только с промежуточным представлением, то преобразователи могут

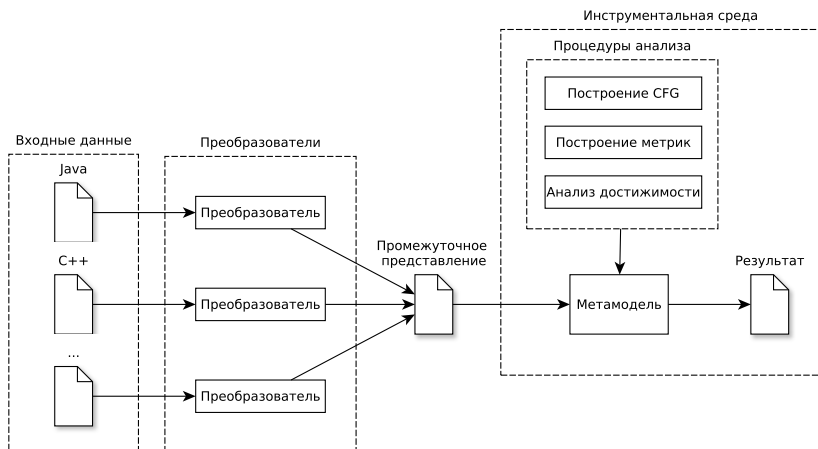


Рисунок 3.1. Архитектура программной системы

поставляться сторонними разработчиками, тем самым избавляя программиста от необходимости поддерживать все возможные языки.

3. *Процедуры анализа* вместе с *метамоделью* являются ключевыми составляющими инструментальной среды. Инструментальная среда содержит графический интерфейс пользователя, на котором в том или ином виде отображается результат проведенного анализа.

## 3.2. Проектирование метамоделей

Существует два подхода к разработке метамоделей:

1. Разработка архитектуры “с нуля”
2. Использование стандартных средств описание метамоделей, т.н. мета-метамоделей.

Каждый из двух подходов обладает своими достоинствами и недостатками, а именно:

1. Собственная метамодель позволит сильно упростить ее структуру, специализировав ее под нужды инструментальной среды, что увеличит производительность разрабатываемых алгоритмов над метамodelью и облегчит ее использование. Однако данный подход обладает очень высокими рисками, так как данная составляющая является ключевой в разрабатываемой среде, и ошибки в ее проектировании могут привести к провалу проекта.
2. Использование стандартных средств менее подвержено рискам, но, в силу универсальности данного подхода, получившаяся метамодель может быть слишком громоздкой в использовании.

В результате анализа было принято решение об использовании стандартной архитектуры, но с незначительными изменениями, чтобы скомбинировать достоинства обоих подходов к проектированию.

### **3.2.1. Стандарт MOF**

Рассмотрим подробнее существующий стандарт разработки метамodelей:

В 1997 году группой OMG был создан унифицированный язык моделирования (UML). UML позволяет описывать различные компоненты и артефакты системы, тем самым упрощая процесс разработки [10].

Для описания конструкций языка UML был разработан фреймворк MOF (Meta Object Facility) [11]. В дальнейшем обе эти концепции вошли в подход, называемый Model Driven Architecture (MDA) [12]. Данный подход к разработке ПО вводит дополнительный уровень абстракции, позволяющий описывать структуру и поведение разрабатываемой системы, не завися от нижележащей используемой технологии.

Таким образом, MOF является мета-метамodelью для описания метамodelей (например, метамodelи UML). Аналогично расширенной форме Бэкуса-Наура (РБНФ), которая задает грамматику языка программирования, MOF позволяет задавать структуру и абстрактный синтаксис метамodelи.

На данный момент последней версией стандарта является версия 2.4.2, выпущенная в 2014 году [11].

### 3.2.2. Иерархия моделей MOF

Архитектура MOF определена в контексте иерархии моделей, на верхнем уровне которой и находится MOF. Данная иерархия выглядит следующим образом:

**M3** - слой мета-метамоделей

**M2** - слой метамоделей

**M1** - слой моделей

**M0** - слой времени исполнения

Слои M3 и M2 так же называются *слоями спецификации языков* [13]. Каждый слой является уровнем абстракции - чем ниже слой, тем конкретнее описывается система.

На слое M3 находится только одна метамодель - MOF, задача которой - описание метамоделей. На уровне M2 располагается множество метамоделей, которые являются экземплярами MOF. Слой моделей содержит пользовательское описание системы. Слой M0 содержит объекты системы во время ее исполнения.

Ниже приведен пример иерархии для конкретной системы:

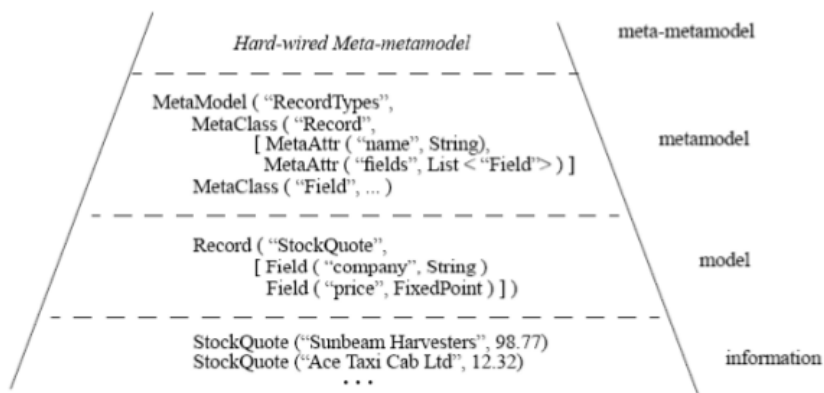


Рисунок 3.2. Пример иерархии моделей

Теоретически, иерархию можно дополнить дополнительными слоями над слоем МЗ, однако MOF позволяет рекурсивно описывать вышележащие слои, поэтому слои выше МЗ не имеют смысла, так как содержат в себе ту же самую мета-метамодель.

### 3.2.3. Структура MOF

MOF обладает модульной структурой, каждый модуль называется *пакетом*. Пакеты можно так же объединять в пакеты, образуя иерархию пакетов. Существует два пакета верхнего уровня - MOF и UML infrastructure library.

Стоит отметить, что MOF использует те же понятия для описания сущностей, что и UML, таким образом, между пакетами образуется следующий вид отношения:

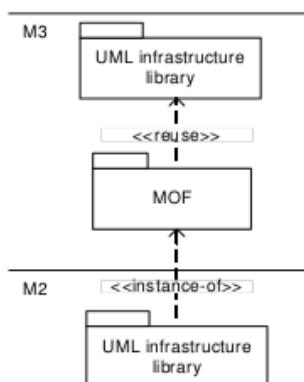


Рисунок 3.3. Взаимосвязь между пакетами MOF верхнего уровня

Данные пакеты включают в себя множество подпакетов, дерево которых приведено на рис.3.4.

Опишем некоторые из приведенных пакетов:

### Пакет Abstractions

Данный пакет содержит элементы, которые затем используются во всех других пакетах. В нем содержатся такие элементы как выражения, литералы, пространства имен, отношения и т.д.

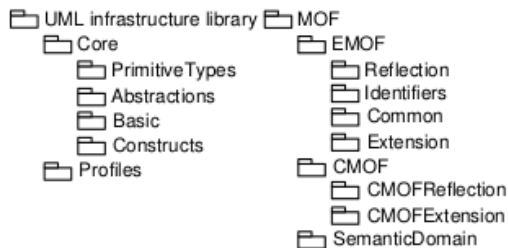


Рисунок 3.4. Дерево пакетов MOF

## Пакет Basic

Пакет Basic предназначен исключительно для облегчения использования и объединяет в себе сущности из нескольких других пакетов.

## Пакет Constructs

В данном пакете содержатся конкретные экземпляры классов из пакета Abstractions, например, классы различных отношений, конструкций языка и т.д.

## Пакет EMOF

EMOF является объединяет в себе базовые концепции MOF, такие как Reflection (возможность доступа к свойствам описываемого объекта), Element (базовая единица метамодели, являющаяся экземпляром какого-либо класса), Common (пакет, поддерживающий коллекции экземпляров Element).

Структура данного пакета приведена на рис. 3.5:

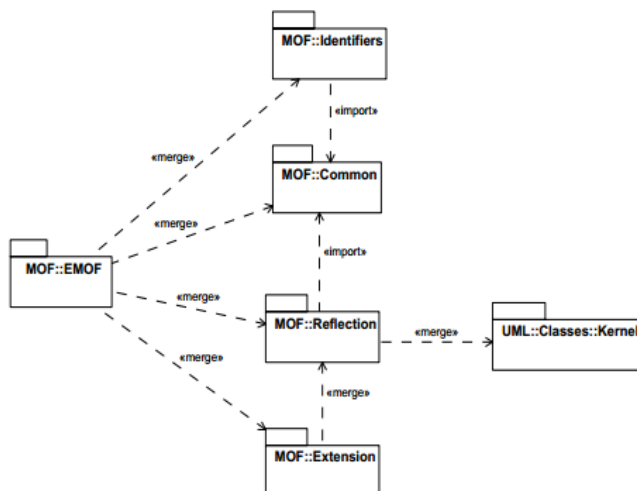


Рисунок 3.5. Структура EMOF

## Пакет CMOF

Данный пакет включает в себя пакет EMOF с некоторыми дополнениями (например, он расширяет пакет Reflection, добавляя в него новые операции над сущностями).

На рис. 3.6 приведен краткий пример MOF-совместимой метамодели для описания электрических схем [13].

Класс **Component** является базовым классом для всех сущностей, включая **ElectricalCircuit**, которая отображает саму электрическую схему. Класс **Wire** показывает вид отношений между объектами электрической схемы.

### 3.2.4. Сериализация метамодели

Для сериализации UML моделей и их передачи между различными средствами OGM был разработан формат XMI. Но так как UML является совместимой со стандартом MOF, XMI может быть использован для сериализации любой MOF-совместимой метамодели [14].

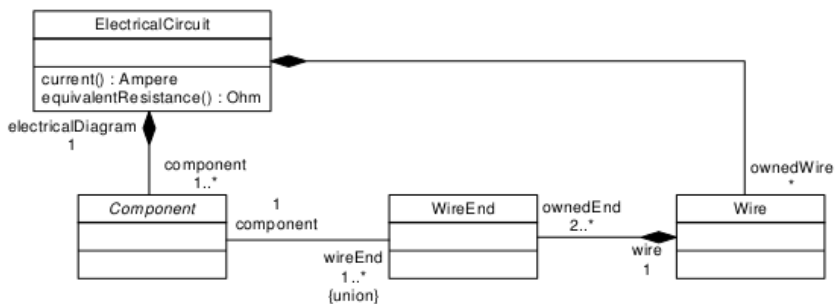


Рисунок 3.6. Метамодель для описания электрических схем

В стандарт XMI входят следующие составляющие:

1. Набор правил DTD для отображения метамodelей в XML документ
2. Правила описания метаданных
3. Схему XML для XMI документов

Пример отношения между системой на языке программирования, ее модели и сериализованной модели в XMI приведен на рис. 3.7:

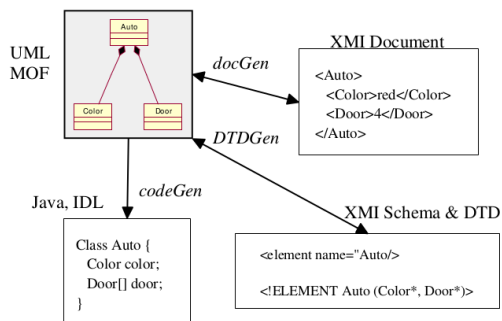


Рисунок 3.7. Отношение между MOF и XMI



Приведем пример сериализованной UML модели в формате XML:

header		<pre>&lt;XMI xmi.version="1.1" xmlns:uml="org.omg/uml1.3"&gt;   &lt;XMI.header&gt;     &lt;XMI.documentation&gt;       An example of an auto.     &lt;/XMI.documentation&gt;     &lt;XMI.metamodel name="UML" version="1.3"       href="uml1.3.xmi" /&gt;     &lt;XMI.model name="Cars" version="1.0" /&gt;   &lt;/XMI.header&gt;   &lt;XMI.content&gt;     &lt;Class name="Car"&gt;       &lt;Class.ownedElements&gt;         &lt;Attribute name="make"/&gt;         &lt;Attribute name="model"/&gt;         &lt;Operation name="drive"/&gt;       &lt;/Class.ownedElements&gt;     &lt;/Class&gt;   &lt;/XMI.content&gt; &lt;/XMI&gt;</pre>		metadata and version
content				

Рисунок 3.8. Пример сериализованной UML-модели

Таким образом, выбранный формат данных естественным образом ложится на разработанную архитектуру, а его стандартизованность облегчает взаимодействие между инструментальной средой и сторонними преобразователями. Это является неоспоримым преимуществом по сравнению с другими форматами (JSON, формат сериализации объектов JAVA и т.д.).

### 3.3. Проектирование архитектуры преобразователей

Задача преобразователей - отображение исходного кода программы в метамодель. Таким образом каждый преобразователь - это транслятор языка программирования, на котором написана анализируемая программа.

Как и любой транслятор, преобразователь может иметь несколько фаз выполнения [15]:

1. Лексический анализ - выделение лексем во входном тексте программы
2. Синтаксический анализ - построение синтаксического дерева по входному набору лексем в соответствии с грамматикой языка

3. Семантический анализ - проверка семантических условий языка (например, соответствие типов, наличие объявления используемой переменной и т.д.)
4. Оптимизация - преобразование промежуточного представления программы с целью повышения производительности, уменьшения размера генерируемого кода, объема используемой памяти и т.п.
5. Генерация кода - получение результата выполнения трансляции

Так как задачей преобразователя является генерация метамодели, некоторые фазы не являются необходимыми (например, фаза оптимизации). Также, можно опустить фазу семантического анализа, что сильно упростит код преобразователя, но скажется на его простоте использования, так как пользователю будет необходимо проверять корректность анализируемой системы сторонними средствами (например, при помощи компилятора языка программирования, на котором написана анализируемая система).

Получившаяся структура преобразователя изображена на рис. 3.9.

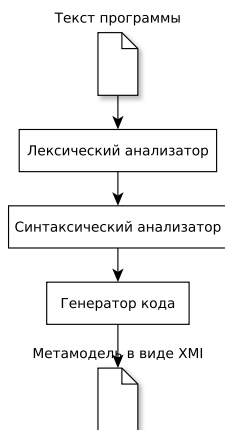


Рисунок 3.9. Архитектура преобразователя

Существует два подхода к построению лексических (лексеров) и синтаксических (парсеров) анализаторов:

1. Написание кода лексера и парсера вручную (например, парсеров, использующие метод рекурсивного спуска или метод функциональных комбинаторов).
2. Использование генераторов парсеров (например, табличных парсеров)

К достоинствам первого метода можно отнести повышенное быстродействие, пониженную ресурсоемкость и меньший объем кода лексера и парсера. К недостаткам данного метода можно отнести высокую трудоемкость и сложность написания.

Использование генераторов парсеров значительно упрощает написание, тем самым уменьшая количество возможных ошибок и дефектов. Однако сгенерированные парсеры обладают худшей производительностью и предъявляют повышенные требования к объему используемой памяти.

На основе анализа достоинств и недостатков обоих методов, было принято решение об использовании автоматической генерации кода парсеров и лексеров разрабатываемых преобразователей. В случае нехватки производительности или объема предоставляемой памяти по результатам профилирования возможна разработка новых преобразователей с использованием вручную написанных парсеров и лексеров.

Результатом работы преобразователя является сериализованная метамодель в формате XML. Так как структура XML-файла полностью отражает структуру метамодели, для ее сериализации рационально использовать различные фреймворки для преобразования объектов программы в формат XML.

### **3.4. Проектирование графического интерфейса**

При проектировании графической составляющей инструментальной среды необходимо решить следующие проблемы:

1. Унификация процедур анализа над метамodelью для обеспечения возможности добавления новых процедур по мере необходимости

## 2. Удобство отображения моделей сложных программных систем

Так как большинство процедур анализа строится на обходе структуры метамодели, то для решения первой проблемы целесообразно применить паттерн “Посетитель” (Visitor) [16]. Применение данного паттерна позволит абстрагировать метамодель от алгоритмов над ней, выделив обход ее структуры в отдельный класс.

Необходимые виды визуализации, описанные в п. 2, имеют вид дерева или графа, поэтому для решения второй проблемы можно применить следующие способы уменьшения размера отображаемой модели системы:

1. Сворачивание отдельных узлов и примыкающих к ним дуг
2. Масштабирование всего графа относительно размера панели отображения визуализаций

## 3.5. Вывод

В результате была разработана архитектура инструментального средства. Вся программная система разбита на три составляющие:

1. Преобразователи, в чью задачу входит импортирование исходного кода анализируемой программы и создание сериализованной метамодели. Данный подход позволяет отделить языкозависимую часть от всей системы в целом.
2. Метамодель, которая является абстрактным представлением системы, над которой затем проводятся все процедуры анализа и визуализации.
3. Все процедуры анализа были вынесены в отдельную составляющую и отделены от метамодели, что позволит легко расширять комплект имеющихся процедур.

При проектировании графического интерфейса была учтена возможность извлечения крупных моделей из и были приняты меры по улучшению их визуализации.

## 4. РАЗРАБОТКА ИНСТРУМЕНТАЛЬНОЙ СРЕДЫ

В данном разделе рассматривается реализация программной системы в соответствии с поставленными требованиями и спроектированной архитектурой.

### 4.1. Выбор средств разработки

Разработка среды велась на языке Java 7 версии (OpenJDK Runtime Environment (IcedTea 2.4.7) [17]).

Сборка проекта производится при помощи фреймворка для автоматизации сборки Apache Maven. Данный фреймворк позволяет декларативным образом описывать процесс сборки проекта, фокусируясь на его структуре, используя различные плагины для организации фаз построения [18]. Вся информация содержится в XML-файле под названием `pom.xml` (Project Object Model). Одной из отличительной особенностей Maven является наличие центрального репозитория, что сильно упрощает управление зависимостями разрабатываемого проекта. Добавляя необходимые библиотеки в `pom.xml`, они будут автоматически скачены из центрального репозитория и размещены в локальном репозитории для дальнейшего использования в цикле сборки проекта.

### 4.2. Структура проекта

В соответствии с архитектурой среды, приведенной в п.п. 3.1, проект разбит на 3 составляющих:

1. Мета модель
2. Преобразователи
3. Инструментальная среда (графический интерфейс, процедуры визуализации и анализа)

Проект разделен на три модуля, связь между которыми изображена на рис. 4.1.

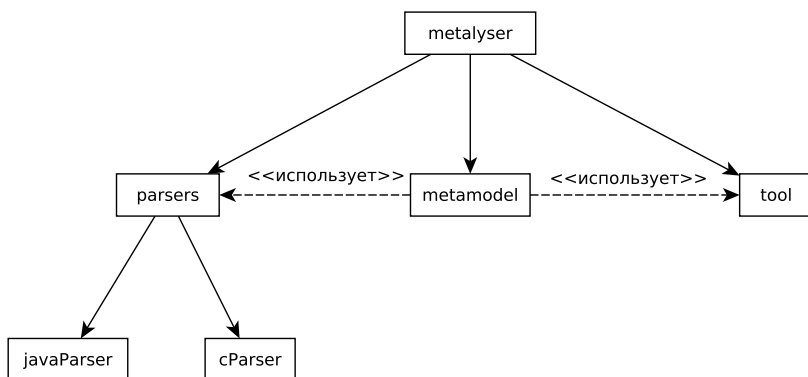


Рисунок 4.1. Структура maven-проекта

Рассмотрим подробнее приведенную структуру:

1. Модуль верхнего уровня называется **metalyser** (Meta-Analyser) и является корнем всего проекта. В нем размещаются зависимости и плагины maven, необходимые для сборки всей программной системы.
2. Модуль **parsers** является родительским модулем для преобразователей языка Java и C, поставляемых вместе с разработанной средой.
3. Модуль **javaParser** является преобразователем для языка Java. Аналогично, модуль **cParser** используется для импортирования систем, написанных на языке C.
4. Модуль **metamodel** является реализацией метамодели, описывающей анализируемую систему. Данный модуль реализован в виде библиотеки классов и не предназначен для непосредственного использования (он не содержит Main-класса для запуска), а используется в виде зависимости в остальных модулях системы.
5. Модуль **tool** является реализацией инструментальной среды.

### 4.3. Разработка метамодели

На основе анализа возможной архитектуры метамодели (см. п.п. 3.2), было принято взять за основу архитектуру, предложенную стандартом MOF. Недостатком данной архитектуры является ее громоздкость: полученные метамодели охватывают широкий спектр возможных задач, что делает их достаточно объемными и сложными в использовании. Исходя из требований к системе, было решено модифицировать исходную архитектуру, а именно были убраны отдельные классы для отображения отношений между объектами. Вместо этого ссылки на зависимые объекты хранятся непосредственно в самих классах. Из-за этого теряется семантическая информация о видах отношений, но намного сокращается количество используемых классов, что сильно упрощает использование библиотеки.

Таким образом, полученная метамодель не является полностью MOF-совместимой, но, если в дальнейшем развитии проекта будет необходима такая совместимость для взаимодействия со сторонними приложениями, не составит труда привести ее к полной совместимости.

Принимая во внимание все перечисленные замечания, была разработана соответствующая структура метамодели (краткая UML-диаграмма классов приведена на рис. 4.2). Стоит отметить, что на данной диаграмме, для уменьшения ее размеров, были опущены все методы, отношения между классами и некоторые наследники основных суперклассов.

Рассмотрим подробнее некоторые из классов:

#### Entity

Листинг 4.1. Интерфейс Entity

```
1 public interface Entity {  
2     public void accept(final Visitor visitor);  
3 }
```

От интерфейса **Entity** наследуются все классы метамодели. Класс содержит лишь один метод - **accept**, который необходим для обхода дерева, которое представляет из себя построенная модель. Таким образом для проведения анализа метамодели был реализован паттерн “Посетитель” (подробнее про обход дерева объектов описано в п.п. 4.3.1).

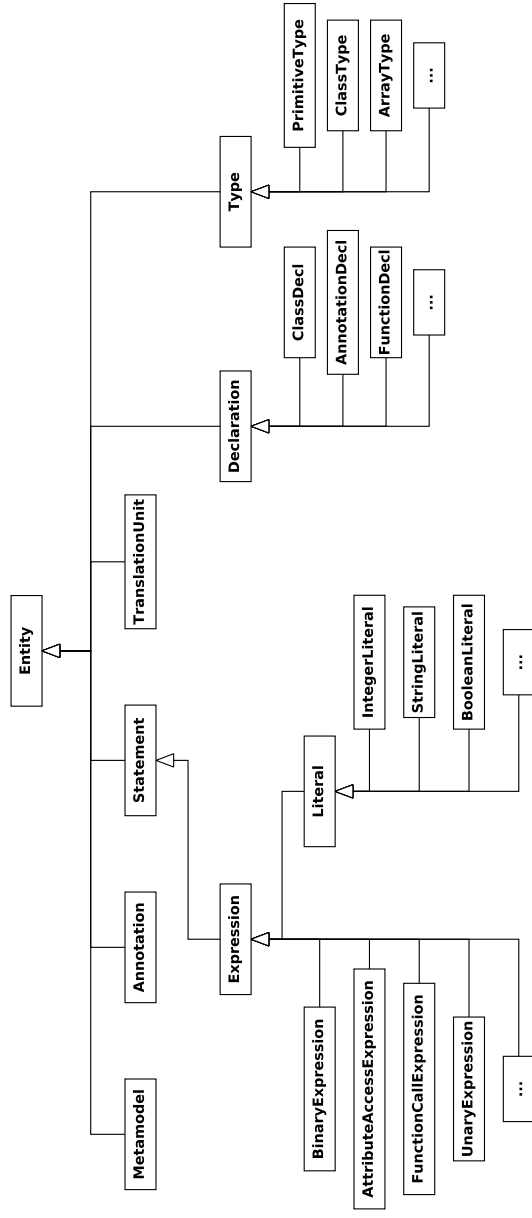


Рисунок 4.2. Упрощенная диаграмма классов метамодели



## Metamodel

Листинг 4.2. Класс Metamodel

```
1 public class Metamodel implements Entity {
2     private final List<TranslationUnit> units;
3
4     ...
5 }
```

Данный класс является контейнером верхнего уровня для объектов типа `TranslationUnit`. Он предназначен для сборки результатов работы преобразователей над каждым из файлов с исходным кодом анализируемой программной системы.

## TranslationUnit

Листинг 4.3. Класс TranslationUnit

```
1 public class TranslationUnit implements Entity {
2     private final List<Import> imports;
3     private final List<Declaration> types;
4
5     ...
6 }
```

Класс `TranslationUnit` отображает одну единицу трансляции при работе преобразователя. Обычно ей соответствует один файл с исходным кодом (за исключением, например, языка C, где в каждый файл на этапе работы препроцессора вставляется текст используемых заголовочных или любых других файлов, подключенных при помощи директивы `#include`). Таким образом, обычно при трансляции всей анализируемой системы в результате работы преобразователя будет получено несколько объектов типа `TranslationUnit`, которые затем должны быть объединены в один объект типа `Metamodel`.

## Annotation

Листинг 4.4. Класс Annotation

```
1 public class Annotation implements Entity {
2     private final String name;
3     private final Map<String, Entity> values;
4     private final Entity value;
5 }
```

```
6      ...
7  }
```

Данный класс предназначен для описания аннотаций (например, в языке Java) или им подобных элементов (например, декораторов в языке Python). Если аннотация имеет параметры, то существует два варианта их хранения:

1. Если параметров несколько, то они помещаются в контейнер `values`. Необходимо учитывать, что в данном случае параметры являются именованными.
2. Если аннотация принимает один параметр, то он может быть именованным и помещаться в поле `value`.

## Statement

Класс `Statement` является базовым для всех выражений в исходном коде программы. Примером наследников являются классы `IfStatement` (ветвление вида “if-else”) и `WhileStatement` (цикл “while”).

## Expression

Листинг 4.5. Класс `Expression`

```
1 public abstract class Expression implements Statement {
2     protected final Type type;
3
4     ...
5 }
```

Данный класс по своему назначению аналогичен классу `Statement`, но, в отличие от него, предназначен для описания выражений, имеющих тип. Примером таких выражений может быть обращение к переменной (класс `VariableReferenceExpression`), арифметические выражения (классы `BinaryExpression` и `UnaryExpression`), вызов функции (класс `FunctionCallExpression`) и т.д. Следует отметить, что типизированные выражения также являются обычными выражениями и могут использоваться аналогично другим наследникам класса `Statement`.

## Type

Данный класс предназначен для описания различного вида типов, как примитивных, так и пользовательских. Наследниками этого класса являются, например, классы `PrimitiveType` (примитивные типы) и `ArrayType` (массивы).

## Literal

Класс `Literal` отображает различного вида литералы в исходном коде программы. Примерами литералов могут являться целочисленные литералы (1, 42, 88), строковые литералы (“foo”, “barbaz”) и т.д.

## Declaration

Листинг 4.6. Класс `Declaration`

```
1 public abstract class Declaration implements Entity {
2     protected final String name;
3     protected Visibility visibility;
4     protected final List<String> modifiers;
5     protected final List<Annotation> annotations;
6
7     ...
8 }
```

Класс `Declaration` является суперклассом для объявлений типов и глобальных переменных. Каждое объявление может быть помечено аннотациями (`annotations`) или различными модификаторами (например, `final` в языке Java или `const` в C). Один вид модификаторов, а именно модификаторы доступа, вынесены в тип `Visibility`. Это сделано для облегчения подсчета некоторых видов метрик, а также потому что в большинстве языков программирования назначение этих модификаторов совпадает (поэтому данное допущение не ударит по универсальности метамодели). Примерами наследников данного класса являются классы `ClassDecl` (объявление класса) и `FunctionDecl` (объявление функции).

### 4.3.1. Проведение операций над метамоделью

Существует два подхода к реализации паттерна “Посетитель” в языке Java:

1. “Классический” подход
2. Подход с использованием рефлексии в языке Java (Java Reflection API [17])

Рассмотрим оба варианта подробнее:

### “Классический” подход

В данном случае в классе-“посетителе” необходимо реализовывать методы `visit` для каждого типа иерархии объектов. К недостаткам данного подхода можно отнести очень малую гибкость - при изменении иерархии классов, над которой совершается обход, аналогичные изменения необходимо вносить в каждый из классов-“посетителей”.

### Подход с использованием рефлексии

Подход с использованием рефлексии использует метод языка Java `Java.lang.Class.getMethod()` для поиска метода `visit` для типа объекта-параметра. Достоинствами данного метода является отсутствие необходимости реализации метода `visit` для каждого возможного класса из иерархии: пользователь должен реализовать метод только для тех классов, которые необходимы для той или иной процедуры, производимой над этой иерархией. Недостатком данного подхода является пониженное быстродействие по сравнению в “классическим” подходом (из-за низкой производительности Reflection API).

Так как метамодель, в соответствии с требованиями, должна быть расширяемой и позволять легко добавлять новые классы к уже существующим, то было принято решение использовать подход с использованием рефлексии.

Исходя из вышесказанного был разработан интерфейс `Visitor`:

Листинг 4.7. Базовый класс для обхода метамодели

```
1 public interface Visitor {  
2     public void dispatch(final Entity entity);  
3     public void navigate(final Entity entity);  
4 }
```

Метод `dispatch` вызывается на корневом объекте метамодели (чаще всего это экземпляр класса `Metamodel`, но это условие не является обязательным) и инициирует процедуру обхода. Метод `navigate`

предназначен для определения порядка обхода вложенных элементов метамодели: для каждого класса, унаследованного от **Entity**, необходимо переопределить метод **accept**, в котором пользователь может задавать, в каком порядке будут посещены поля этого класса. Также поддерживается навигация непосредственно из класса, реализующего интерфейс **Visitor**: для необходимо переопределить метод **navigate**, чтобы он оставался пустым.

Примером реализации интерфейса **Visitor** является класс **VisitorAdapter**:

Листинг 4.8. Реализация интерфейса **Visitor**

```
1 public class VisitorAdapter implements Visitor {
2     @Override
3     public void dispatch(final Entity entity) {
4         if (entity == null) return;
5         try {
6             final Method m = getClass().getMethod("visit",
7                 new Class<?>[] { entity.getClass() });
8             m.invoke(this, new Object[] { entity });
9         } catch (NoSuchMethodException ex) {
10             // протоколирование ошибки
11         } catch (IllegalAccessException |
12             IllegalArgumentException |
13             InvocationTargetException ex) {
14             // протоколирование ошибки
15         }
16         navigate(entity);
17     }
18     @Override
19     public void navigate(Entity entity) {
20         entity.accept(this);
21     }
22 }
```

Переопределенный метод **dispatch** работает следующим образом:

1. Получение метода **visit** для класса объекта, переданного в качестве параметра.
2. В случае отсутствия такого метода, происходит запись сообщения с соответствующей информацией в лог, но ход программы не прерывается.
3. Вызов метода на объекте-парамetre.

4. Вызов метода `navigate` для дальнейшего обхода объектов метамодели. В данном случае используется навигация по-умолчанию, описанная в классах метамодели.

### 4.3.2. Выбор фреймворка для сериализации

Для языка Java существует множество фреймворков для генерации XML-файлов, самыми популярными из них являются:

1. Java Architecture for XML Binding (JAXB)
2. Simple
3. XStream
4. XMLBeans

Во всех фреймворках, кроме XMLBeans, описание правил сериализации производится при помощи аннотирования необходимых элементов классов.

### Фреймворк JAXB

JAXB представляет собой стандарт, описывающий преобразование Java объектов в XML (маршалинг) и обратно (демаршалинг). Так как JAXB является стандартом, то существует несколько его реализаций, например, Metro, EclipseLink MOXy, JaxMe. Отличительной особенностью данного стандарта является наличие его реализации в составе Java SE начиная с версии 6, тем самым исключается необходимость внедрения дополнительных зависимостей в проект.

Пример использования аннотаций JAXB:

Листинг 4.9. Пример использования фреймворка JAXB

```
1 @XmlRootElement
2 public class Customer {
3     String name;
4     int age;
5     int id;
6
7     @XmlElement
8     public void setName(String name) {
9         this.name = name;
10    }
```

```

11
12     @XmlElement
13     public void setAge(int age) {
14         this.age = age;
15     }
16
17     @XmlAttribute
18     public void setId(int id) {
19         this.id = id;
20     }
21
22     ...
23 }

```

XML-файл после сериализации:

Листинг 4.10. Полученный XML-файл

```

1 <?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
2 <customer id="100">
3     <age>23</age>
4     <name>Alexander</name>
5 </customer>

```

## Фреймворк Simple

Фреймворк Simple организован таким образом, чтобы как можно больше снизить необходимость конфигурирования, что уменьшает количество используемых аннотаций в коде. Еще одной отличительной особенностью данного фреймворка является наличие возможности десериализации неизменяемых объектов, что актуально для большинства классов разработанной метамодели.

Пример использования:

Листинг 4.11. Пример использования фреймворка Simple

```

1 @Root
2 public class Example {
3     @Element
4     private String text;
5
6     @Attribute
7     private int index;
8
9     ...
10 }

```

Полученный XML-файл:

Листинг 4.12. Полученный XML-файл

```
1 <example index="123">
2   <text>Example message</text>
3 </example>
```

## Фреймворк XStream

XStream по своим возможностям практически ничем не отличается от уже описанной ранее реализации JAXB. К недостаткам данного фреймворка можно отнести отсутствие подробной документации.

Пример использования:

Листинг 4.13. Пример использования фреймворка XStream

```
1 @XStreamAlias("message")
2 class RendezvousMessage {
3
4     @XStreamAlias("type")
5     @XStreamAsAttribute
6     private int messageType;
7
8     @XStreamImplicit(itemFieldName="part")
9     private List<String> content;
10
11     ...
12 }
```

Полученный XML-файл:

Листинг 4.14. Полученный XML-файл

```
1 <message type="15">
2   <part>firstPart</part>
3   <part>secondPart</part>
4 </message>
```

## Фреймворк XMLBeans

Фреймворк XMLBeans является частью проекта Apache XML. В отличие от всех остальных фреймворков, XMLBeans создает Java-классы на основе схем, описанных на языке XSD (XML Schema Definition), что делает его непригодным к использованию в разрабатываемом проекте.



### 4.3.3. Вывод

Таким образом, после проведения анализа, для сериализации метамодели был выбран фреймворк Simple. Ключевым фактором является возможность данного фреймворка сериализовывать неизменяемые объекты, а наличие сторонней зависимости не является проблемой, благодаря системе сборки Maven.

Пример использования фреймворка Simple в разработанной метамодели:

Листинг 4.15. Пример использования фреймворка Simple в метамодели

```
1 public abstract class Declaration implements Entity {
2     @Element(required = false)
3     protected final String name;
4     @Element
5     protected Visibility visibility;
6     @ElementList(required = false)
7     protected final List<String> modifiers;
8     @ElementList(required = false)
9     protected final List<Annotation> annotations;
10
11     public Declaration(
12         @Element(name = "name")
13         final String name,
14         @Element(name = "visibility")
15         final Visibility visibility,
16         @ElementList(name = "modifiers")
17         final List<String> modifiers,
18         @ElementList(name = "annotations")
19         final List<Annotation> annotations) {
20
21         ...
22     }
23
24     ...
25 }
```

Аннотация `@Element` обозначает один тэг в итоговом XML-документе, где аргумент `required` указывает на то, что данный тэг не является обязательным. Аннотация `@ElementList` описывает коллекцию тэгов. Так как класс `Declaration` является неизменяемым, то для десериализации необходимо предоставить конструктор, позволяющий инициализировать все поля класса, при этом каждый параметр необходимо пометить соответствующей аннотацией с указанием имени поля.

В данном проекте используется версия Simple 2.7.1, выпущенная

в 2013 году.

Пример сериализованной метамодели приведен в приложении А.

## 4.4. Разработка преобразователей

При разработке архитектуры преобразователей (п.п.3.3) было принято решение об использовании генераторов парсеров для проведения лексического и синтаксического анализа при построении модели анализируемой программы.

### 4.4.1. Выбор генератора лексического и синтаксического анализаторов

Наиболее популярными генераторами парсеров (которые активно разрабатываются и поддерживаются) на языке Java на данный момент являются следующие программные средства:

- ANTLR (ANother Tool for Language Recognition)
- JavaCC (Java Compiler Compiler)
- Jparsec

## ANTLR

ANother Tool for Language Recognition - генератор синтаксических и лексических анализаторов. Данное средство генерирует нисходящий анализатор на основе LL(\*) грамматики. Достоинствами данного генератора является комбинированная грамматика (единая для парсера и лексера), наличие среды разработки (ANTLRWorks), подробная диагностика ошибок, а так же способность разбирать леворекурсивные грамматики (которые обычно недопустимы при построении LL-анализаторов) [19].

Для описания грамматики используется форма, близкая к РВНФ:

Листинг 4.16. Пример грамматики, используемой в средстве ANTLR

```
1  grammar T; //имя грамматики
2  //нетерминальные символы:
3  msg : 'name' ID ';'
4      {
5          System.out.println("Hello ,_" + $ID.text + " !");
```

```

6         } ;
7 //терминальные символы
8 ID: 'a'..'z' + ; //произвольное (но >=1) количество букв
9 WS: (' ' | '\n' | '\r' )+; // пробел, перенос строки, табуляция

```

## JavaCC

Как и средство ANTLR, JavaCC генерирует парсер на основе рекурсивного спуска, но с использованием LL(1) грамматики, что увеличивает производительность сгенерированного кода, но также увеличивает сложность написания грамматики.

Пример грамматики:

Листинг 4.17. Пример грамматики, используемой в средстве JavaCC

```

1 SKIP: { " " | "\t" | "\n" | "\r" }
2 TOKEN: { "(" | ")" | "+" | "*" | <NUM: ([ "0"-"9" ])+> }
3
4 void S(): {} { E() <EOF> }
5 void E(): {} { T() ("+" T())* }
6 void T(): {} { F() ("*" F())* }
7 void F(): {} { <NUM> | "(" E() ")" }

```

## Jparsec

Jparsec так же используется для генерации парсера на основе рекурсивного спуска. Главной отличительной особенностью данного генератора является тот факт, что он использует комбинаторы синтаксического анализа. Это значит, что он не использует отдельный файл с грамматикой в форме РБНФ, на основе которого генерируется код парсера на языке Java. Вместо этого каждому нетерминальному символу грамматики ставится в соответствие объект языка Java, а операции по составлению данного символа - альтернатива (символ “[” в РБНФ), зависимость (последовательность лексем) и т.д., описываются при помощи методов этого объекта.

Приведем пример парсера, использующего данный фреймворк:

Листинг 4.18. Пример парсера, использующего фреймворк Jparsec

```

1 private static final Terminals OPERATORS =
2     Terminals.operators("+", "-", "*", "/", "(" , ")");
3
4 static final Parser<> TOKENIZER =

```

```

5      Parsers.or(Terminals.DecimalLiteral.TOKENIZER,
6                  OPERATORS.tokenizer());
7
8      static Parser<?> term(String... names) {
9          return OPERATORS.token(names);
10     }
11
12     static final Parser<BinaryOperator> WHITESPACE_MUL =
13         term("+", "-", "*", "/").not().retn(BinaryOperator.MUL);
14
15     static <T> Parser<T> op(String name, T value) {
16         return term(name).retn(value);
17     }
18
19     static Parser<Double> calculator(Parser<Double> atom) {
20         Parser.Reference<Double> ref = Parser.newReference();
21         Parser<Double> unit =
22             ref.lazy().between(term("("), term(")")).or(atom);
23         Parser<Double> parser = new OperatorTable<Double>()
24             .infixl(op("+", BinaryOperator.PLUS), 10)
25             .infixl(op("-", BinaryOperator.MINUS), 10)
26             .infixl(op("*", BinaryOperator.MUL).or(WHITESPACE_MUL),
27                 20)
28             .infixl(op("/", BinaryOperator.DIV), 20)
29             .prefix(op("-", UnaryOperator.NEG), 30)
30             .build(unit);
31         ref.set(parser);
32         return parser;
33     }

```

Использование данного подхода исключает необходимость фазы генерации кода при сборке проекта, а разработанный код в несколько раз меньше по объему, чем сгенерированный. К недостаткам данного фреймворка можно отнести сложную реализацию необходимой грамматики и неудовлетворительное информирование об ошибках при проведении разбора.

Таким образом, самым совершенным средством для разработки синтаксических и лексических анализаторов является фреймворк ANTLR, в силу простоты использования и неплохой эффективности полученных анализаторов.

В данном проекте используется версия ANTLR 4.2.2, выпущенная в 2014 году.

## 4.4.2. Правила использования генератора парсеров ANTLR

Общая структура файла для описания грамматики ANTLR выглядит следующим образом:

Листинг 4.19. Структура грамматики ANTLR

```
1 grammar Name; // имя грамматики
2 options {...}
3 import ... ;
4 tokens {...}
5 @actionName {...}
6
7 rule1 // правила грамматики
8 ...
9 ruleN
```

Секция **import** предназначена для импортирования других файлов с правилами грамматики, тем самым позволяя создавать иерархичную структуру ANTLR-проекта.

Секция **tokens** позволяет явно задавать типы лексем лексического анализатора. На практике это означает, что будет сгенерировано перечисление (`enum`), которое затем может быть использовано в коде парсера для отсылке к той или иной лексеме.

Правила вида **@действие ...** позволяют использовать набор определенных действий, поддерживаемых ANTLR. Примерами являются действие **@header** для описания заголовочной части сгенерированного кода парсера и действие **@after** для генерации кода, который будет выполнен непосредственно перед вызовом определенного правила грамматики. Пример использования приведен в листинге 4.20:

Листинг 4.20. Использование действий ANTLR

```
1 grammar Count;
2 // объявление пакета, в котором располагается код парсера
3 @header {
4 package foo;
5 }
6
7 @members {
8 int count = 0; // объявление члена класса парсера
9 }
10
11 list
12 // вывод количества чисел во входной последовательности лексем
13 @after {System.out.println(count + "_ints");}
```

```

14 : INT {count++;} ( ' , ' INT {count++;} ) *
15 ;
16
17 INT : [0-9]+ ;

```

Секция **options** задает параметры генерации кода, например, язык на котором будет сгенерирован полученный парсер.

ANTLR позволяет создавать смешанные грамматики, т.е. в одном и том же файле задаются как правила синтаксического анализа, так и лексического. Правила разбора лексем имеют вид “ИМЯ : правило”, где правая часть задает название правила, а левая - правило его разбора. Аналогичный формат имеют правила для нетерминальных символов с той лишь разницей, что имена лексем задаются заглавными буквами, а нетерминалов - прописными.

#### 4.4.3. Реализация преобразователя для языка Java

В основу реализации парсера легла грамматика языка Java, представляемая вместе с исходным кодом фреймворка ANTLR. Грамматика была модифицирована для упрощения построения модели анализируемой программы (текст грамматики приведен в приложении Б).

Код построения модели программы интегрирован в текст грамматики. ANTLR делается следующим образом (в листинге 4.21 приведен пример построения объекта, отображающего объявление класса в исходном коде анализируемой программы):

Листинг 4.21. Пример построения модели

```

1  classDeclaration
2  returns [ClassDecl result]
3  locals [
4      List<TemplateDecl> templates = Collections.emptyList(),
5      List<Type> inherits = new LinkedList<>()
6  ]
7  :   'class' Identifier
8      (   typeParameters
9          { $templates = $typeParameters.result; }
10         )?
11      (   'extends' type
12          { $inherits.add($type.result); }
13         )?
14      (   'implements' typeList
15          { $inherits.addAll($typeList.result); }
16         )?
17      classBody

```

[illegible]





## ЗАКЛЮЧЕНИЕ



# СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. М.И. Глухих, В.М. Ицыксон. Программная инженерия. Обеспечение качества программных средств методами статического анализа. — Санкт Петербург : Издательство Политехнического университета, 2011.
2. Ковалёв С. П. Применение формальных методов для обеспечения качества вычислительных систем // Вестник Новосибирского государственного университета. — 2004. — Т. IV, № 2. — С. 49–74.
3. В.В. Кулямин. Методы верификации программного обеспечения. — Институт системного программирования РАН, 2008.
4. Automation of GUI testing using a model-driven approach / Marlon Vieira, Johanne Leduc, Bill Hasling et al. // AST '06: Proceedings of the 2006 international workshop on Automation of software test. — New York, NY, USA : ACM, 2006. — P. 9–14.
5. Barnett Mike, Schulte Wolfram. Spying on Components: A Runtime Verification Technique // Proc. of the Workshop on Specification and Verification of Component- Based Systems OOPSLA 2001. — 2001.
6. Piefel M. A Common Metamodel for Code Generation // Proceedings of the 3rd International Conference on Cybernetics and Information Technologies, Systems and Applications. — 2006.
7. Nierstrasz Oscar, Ducasse Stéphane, Gîrba Tudor. The story of moose: an agile reengineering environment. — ACM, 2005. — P. 1–10.
8. A Programming Language Independent Framework for Metrics-based Software Evolution and Analysis / Črt Gerlec, Gordana Rakić, Zoran Budimac, Marjan Heričko // Computer Science and Information Systems. — 2012. — Sep. — Vol. 9, no. 3. — P. 1155–1186.
9. Lattner Chris, Adve Vikram. LLVM: A Compilation Framework for Lifelong Program Analysis & Transformation // Proceedings of the 2004 International Symposium on Code Generation and Optimization (CGO'04). — Palo Alto, California, 2004. — Mar.
10. Fowler Martin. UML Distilled: A Brief Guide to the Standard Object Modeling Language. — 3 edition. — Boston, MA : Addison-Wesley, 2003. — ISBN: 978-0-321-19368-1.
11. Meta Object Facility (MOF) 2.0 Core Specification. — 2003. — Ver-

sion 2.

12. MDA Guide Version 1.0.1 : Rep. / Object Management Group (OMG) ; Executor: J. Miller, J. Mukerji : 2003.
13. Overbeek J.F. Meta Object Facility (MOF): investigation of the state of the art. — 2006. — June.
14. Object Management Group (OMG). XML Meta-Data Interchange XMI 2.1.1. — formal/2007-12-01. — 2007.
15. Aho A., Sethi R., Ullman J. Compilers: Principles, Techniques and Tools. — Addison-Wesley, 1986.
16. Design patterns: elements of reusable object-oriented software / Erich Gamma, Richard Helm, Ralph Johnson, John Vlissides. — Pearson Education, 1994.
17. The Java Language Specification, Java SE 7 Edition / James Gosling, Bill Joy, Guy Steele et al. — Addison Wesley, 2013.
18. Maven Documentation : Rep. / Apache Software Foundation ; Executor: Apache Maven Project : Apache Maven Project, 2001 - 2005. — URL: <http://maven.apache.org/guides/index.html>.
19. Parr Terence. The Definitive ANTLR Reference: Building Domain-Specific Languages. — Pragmatic Bookshelf, 2007. — P. 376. — ISBN: 0978739256.

# ПРИЛОЖЕНИЕ А

## ПРИМЕР МЕТАМОДЕЛИ, СЕРИАЛИЗОВАННОЙ В XML

Листинг А.1. Пример сериализованной метамодели



# ПРИЛОЖЕНИЕ Б ГРАММАТИКА ЯЗЫКА JAVA ДЛЯ ПОСТРОЕНИЯ МОДЕЛИ АНАЛИЗИРУЕМОЙ ПРОГРАММНОЙ СИСТЕМЫ

Листинг Б.1. Грамматика языка Java