

Федеральное государственное бюджетное  
образовательное учреждение высшего образования  
«Санкт-Петербургский национальный исследовательский  
Академический университет Российской академии наук»  
Центр высшего образования

Кафедра математических и информационных технологий

Васильев Роман Алексеевич

# Отладчик вывода типов языка программирования scala в intellij idea

Магистерская диссертация

Допущена к защите.  
Зав. кафедрой:  
д. ф.-м. н., профессор Омельченко А. В.

Научный руководитель:  
д. ф.-м. н., профессор Выбегалло А. А.

Рецензент:  
ст. преп. Привалов А. И.

Санкт-Петербург  
2017

SAINT-PETERSBURG ACADEMIC UNIVERSITY  
Higher education centre

Department of Mathematics and Information Technology

Roman Vasiliev

# Empty subset as closed set

Graduation Thesis

Admitted for defence.

Head of the chair:  
professor Alexander Omelchenko

Scientific supervisor:  
professor Amvrosy Vibegallo

Reviewer:  
assistant Alexander Privalov

Saint-Petersburg  
2017

# Оглавление

<b>Введение</b>	<b>5</b>
<b>1. Обзор</b>	<b>6</b>
1.1. Отладчик типов для языка программирования OCaml . . . . .	6
1.2. Scala type debugger . . . . .	6
1.3. Show implicit parameters action . . . . .	6
<b>2. Особенности реализации</b>	<b>7</b>
2.1. Инструментирование . . . . .	8
2.2. Макросы в scala . . . . .	10
2.3. Макросы для функций . . . . .	11
2.3.1. Поиск инструментации . . . . .	12
2.3.2. Генерация неинструментированной функции . . . . .	12
2.3.3. Генерация инструментированной функции . . . . .	14
2.4. Макросы для классов . . . . .	14
2.4.1. Создание дополнительного класса . . . . .	15
2.4.2. Наследование . . . . .	15
2.4.3. Внутреннее состояние . . . . .	15
2.4.4. Изменение методов родителя . . . . .	16
2.5. Сохраненные данные и визуализация . . . . .	17
<b>3. Реализация</b>	<b>18</b>
3.1. Архитектура Scala Plugin . . . . .	19
3.2. Проверка сводимости типов . . . . .	21
3.2.1. Singleton Type . . . . .	23
3.2.2. Type Projection . . . . .	23
3.2.3. Type Designator . . . . .	24
3.2.4. Parameterized Type . . . . .	24
3.2.5. Compound Type . . . . .	24
3.2.6. Existential Type . . . . .	24
3.2.7. Method Type . . . . .	25
3.2.8. Polymorphic Method Type . . . . .	25
3.2.9. Abstract Type . . . . .	25
3.2.10. Типы представленные только в Scala Plugin . . . . .	25
3.2.11. Типы представленные только в спецификации scala . . . . .	25
3.3. Вывод типов . . . . .	25
3.4. Разрешение перегрузок функций . . . . .	25

<b>Заключение</b>	<b>26</b>
<b>Список литературы</b>	<b>27</b>

## Введение

Подмножество, как следует из вышесказанного, допускает неопровержимый ортогональный определитель, явно демонстрируя всю чушь вышесказанного. Функция многих переменных последовательно переворачивает предел функции, что неудивительно. Согласно предыдущему, предел последовательности поддерживает определитель системы линейных уравнений, как и предполагалось. Интересно отметить, что предел функции однородно специфицирует аномальный Наибольший Общий Делитель (НОД) таким образом сбылась мечта идиота - утверждение полностью доказано. Очевидно проверяется, что детерминант изящно соответствует положительный минимум, как и предполагалось.

## 1. Обзор

1.1. Отладчик типов для языка программирования OCaml

1.2. Scala type debugger

1.3. Show implicit parameters action

## 2. Особенности реализации

Здесь будет говориться как получать данные о работе того или иного процесса внутри Scala Plugin. Здесь и далее под процессом понимается последовательность выборов, принимаемых функциями и классами внутри Scala Plugin, направленных на решение какой-то задачи. Для начала нужно определиться, как именно будут получены эти данные. Сделаем несколько наблюдений:

- Чтобы не заниматься обратным анализом и не перереализовывать функциональность плагина, отвечающую за работу с типами, разумным выглядит переиспользовать уже существующую кодовую базу.
- Данные полученные о процессах нужно будет визуализировать. Поэтому хотелось бы их получать как полноценные объекты scala, а не как, скажем, набор снимков состояния стека которые нуждаются в последующей обработке. Для этого формирование этих объектов должно быть отражено в коде рассматриваемых процессов.
- Scala Plugin написан на языке программирования scala. И хоть scala позиционирует себя как язык с функциональной направленностью, однако большая часть кода в плагине, необходимая для анализа процессов связанных с типами, написана в императивном стиле. То есть используются изменяемые состояния объектов, функции не всегда являются чистыми, а поток управления часто прерывается ключевым словом return. Поэтому во время получения данных было бы удобно пользоваться изменяемым состоянием.

В таких условиях разумной выглядит идея инструментировать фрагменты Scala Plugin, отвечающие за интересующие нас процессы. А именно, помещать в контекст выполнения процесса специальные объекты, которые будут хранить данные о текущем состоянии процесса, работа с которыми будет описана прямо в коде плагина, и которые далее мы будем называть объектами инструментации. Подробно это будет рассмотрено в разделе 2.1.

Однако, не стоит забывать, что Scala Plugin - это рабочее приложение с большим количеством пользователей. А код отвечающий за работу с типами, как минимум за сведение типов, выполняется в плагине на каждом шагу. Поэтому внесение дополнительных действий в этот код окажет пагубное воздействие на производительность всего плагина. Чтобы уменьшить воздействие инструментирования на процесс выполнения остального кода, в данной работе принято решение использовать макросы. Обзор возможностей макросов в scala приведен в разделе 2.2. Их применение к инструментированию на примере инструментированных функций будет рассмотрено в разделе 2.2, а распространение на классы в разделе 2.4.

В конце будет раздел 2.5, посвященный формату хранения данных.

## 2.1. Инструментирование

Здесь мы составим набор правил для использования инструментации так, будто макросов нет. Это позволит нам запускать код даже без них. А в последующих разделах мы увидим что построенная система хорошо подходит для анализа на уровне исходного кода.

Принцип, которым мы будем руководствоваться - инструментация должна быть опциональной. Это означает что для запуска инструментированной версии кода нужно приложить какие-то усилия, внешний код не обязан знать что возможно инструментирование. Если ее не включить, влияние инструментирования на классы и функции будет максимально снижено, поведение будет таким, каким было до инструментирования.

Попробуем представить как должно выглядеть инструментирование для, скажем, функции `f`. Нам необходимо передать объект инструментации внутрь функции и дать знать, нужно ли его использовать. Это можно сделать или с помощью двух дополнительных параметров функции: объекта и флага, или с помощью одного дополнительного параметра объекта, который может принимать значение `null`. Но в функциональном программировании есть более выразительное средство - монада `Option`. Значение по умолчанию `None` позволит не изменять уже существующие вызовы функции `f`.

Listing 1: Явная передача объекта инструментации внутрь функции

```
def f(..., instrumentation: Option[Instrumentation] = None)
```

`Option` и его функции высшего порядка такие как `map` и `foreach` автоматически дают нам контекст, в котором мы можем выполнять любые необходимые действия, в области видимости которого есть, непосредственно, объект инструментирования, и которые игнорируются если `Option` оказался пуст.

Listing 2: Действия внутри контекста инструментации

```
instrumentation.foreach { i =>
  val a = computeA
  i.addInformation(a)
}
```

*Option для нас - это контекст, в котором живут все объекты и действия необходимые для сбора данных. Покидать этот контекст им запрещено. ??? Написать Some как контекст.*

Теперь поймем как хранить данные которые изменяются со временем. Понятно что мы можем их хранить внутри объекта инструментации, и у нас нет другого выхода если эти данные обновляются внутри вложенных вызовов функций. Однако один и тот же объект инструментирования может появляться в множестве разных частей



программы и включение в него логики обработки промежуточных данных для всех этих случаев может показаться избыточным. Решением будет вынесение этих данных в изменяемую переменную. Однако, мы хотим чтобы действия над этой переменной зависели от наличия объекта инструментации. В примере ниже показано как это сделать используя функцию `map`.

Listing 3: Привязка к контексту инструментации

```
var intermediate = instrumentation.map(_ => Seq.empty[R])
for (i <- 0 to n) {
  val r = findResultFor(i)
  intermediate = intermediate.map(_ :+ r)
  doSomeStuff(r)
}
```

Таким образом мы явно связали переменную `intermediate` с контекстом `instrumentation` в котором и выполняется инструментация. Аналогичным приемом можно воспользоваться если необходимо завести неизменяемые данные вне контекста инструментации. Это может понадобиться когда для вложенного вызова объект инструментации изменяется.

Listing 4: Создание нового объекта инструментации

```
val inner = instrumentation.map(_.inner)
g(..., instrumentation = inner)
instrumentation.foreach(_.add(inner.get.data))
```

*При возникновении новых сущностей связанных с инструментированием, мы явно устанавливаем связь с уже существующими объектами.*

Разумным требованием кажется избегать влияния инструментирования на логику функции. Но так ли это? Рассмотрим в качестве примера проверку сводимости двух параметризованных типов. Для сводимости одного типа к другому необходимо чтобы количества типовых параметров совпадали и для каждой пары соответствующих типовых параметров выполнялось определенное соотношение. При проверке этих соотношений достаточно дойти до первого ложного чтобы понять что сводимость нет, что плагин и делает. Однако в целях отладки было бы более информативно проверить все соотношения чтобы дать пользователю более полную картину.

Таким образом возможность изменять поведение изначальной функции нам нужна. Однако, есть версия что достаточно уметь делать две вещи: изменять выбранную вертку исполнения в условных конструкциях при включенной инструментировании, а также избегать прерывания потока исполнения ключевым словом `return`.

Основной идеей будет то что квантор всеобщности над пустым множеством всегда верен, а квантор существования нет. Таким образом мы можем использовать логику

основанную на **существовании** объекта в связке с **логическим или**, а также что-то для **любого** объекта в связке с **логическим и**. Также заметим что любую конструкцию, тип которой Unit, такую как return, мы можем записать через if (true).

Пример ниже иллюстрирует это. В нем мы сначала игнорируем кэшированное значения с помощью проверки, возвращающей false на непустом множестве. Далее функции computeA и computeB из которых мы извлекаем данные. Чтобы не прерываться после неудавшейся проверки conditionA, мы добавляем условный оператор. Если instrumentation пуст, то forall вернет true и поток исполнения прервется. В обратном случае выполнится interrupt, который поднимет флаг остановки, а forall вернет false и поток исполнения продолжится. В конце мы с помощью exists проверяем был ли поднят флаг прерывания в процессе работы, и если бы то возвращаем соответствующее значение.

Listing 5: Влияние на первоначальную логику

```
val cached = cache.get(key)
if (cached != null && instrumentation.isEmpty) return cached

val conditionA = computeA(instrumentation = instrumentation)
if (conditionA)
    if (instrumentation.forall(!_interrupted())) return false
val conditionB = computeB(instrumentation = instrumentation)
if (instrumentation.exists(_.interrupted())) return false
return conditionB
```

До этого момента мы говорили только о функциях. Достаточно ли нам уметь работать только с ними? Вообще говоря достаточно, но работать только с функциями неудобно.

Встречаются классы которые выступают как контейнеры функций. Примерами могут служить классы MostSpecificUtil и MethodResolveProcessor. Вместо того чтобы добавлять по параметру с объектом инструментации в каждый метод, намного проще добавить один параметр в их конструкторы. В будущем это усложнит задачу написания макроса.

## 2.2. Макросы в scala

Прежде чем говорить о макросах в scala, следует понять что имеется в виду под словом макрос. Изначально, макрокоманда или макрос – это символьное имя, заменяемое при обработке исходного кода препроцессором на некоторый текст. Наиболее известный пример макросов можно встретить в препроцессорах языков программирования C и C++.

Подобные макросы работают по следующему механизму. В начале пользователь описывает с помощью специального синтаксиса функцию, которая принимает в качестве параметров строковые константы, а возвращает текст, в который могут быть подставлены значения этих параметров. Перед компиляцией текста препроцессор проходит по тексту исходного кода и заменяет вызовы этих функций на соответствующие значения. Как простые строки, без каких-либо проверок. И только после этого, полученный текст отдается компилятору.

Использование макросов в scala контрастирует с вышеописанным подходом. Главным отличием является то, что макросы в scala работают не со строками, а с абстрактными синтаксическими деревьями, представляющими исходный код программы. Scala макросы пишутся на полноценном scala и являются функциями, принимающими АСД фрагментов исходного кода и возвращающие АСД фрагментов которые нужно сгенерировать. То есть для того чтобы применить макрос к коду, компилятору требуется построить АСД этого кода. Тем самым макросы не изменяют изначальный синтаксис, а лишь преобразуют фрагменты кода.

Обычно эти макросы используют чтобы автоматически генерировать что-то для класса, например сериализаторы. И эти возможности предоставляет сам компилятор. Однако существует плагин к компилятору Macro Paradise, который позволяет не просто создавать новый код, но и заменять существующий. Именно эту возможность мы и будем использовать для того чтобы снизить влияние кода инструментации на процесс исполнения.

Для того чтобы использовать Macro Paradise нужно создать специальную аннотацию, которой должны быть помечены функции и классы код которых мы хотим изменить. В класс этой аннотации должен быть добавлен код функции-генератора АСД.

Теперь поймем, чего мы хотим добиться используя кодогенерацию. С одной стороны мы бы хотели полностью избавиться от кода связанного с инструментацией, чтобы он никак не влиял на производительность плагина. С другой стороны, эта инструментация добавлялась не просто так. Выходом будет сгенерировать две копии каждой функции: одну с инструментацией, другую без. То же самое для классов.

Теперь опишем более подробно как именно это будет происходить.

## 2.3. Макросы для функций

Сформулируем правила инструментирования полученные в разделе 2.1:

1. Объект инструментирования передается явно как параметр функции или конструктора классов в монаде Option со значением по умолчанию None.
2. Действия связанные с инструментированием не должны покидать контекст инструментирования.

3. Создание нового контекста инструментации происходит явно от другого, уже существующего, контекста инструментирования. Для этого подходит функция функция `map`.
4. Для того чтобы влиять на первоначальную логику добавляются условные инвариантные для пустого контекста инструментации условия.

В этом разделе наша задача - по коду инструментированной функции `f` сгенерировать пару новых функций `f` и `f$I`. Здесь `f` - это замена старой функции в которой будет удалено все инструментирование. `f$I` - функция в которой оставлена инструментация.

---

**Algorithm 1** Обновление информации о контекстах инструментации

---

```

1: function UPDATEINSTRUMENTATIONNAMES(codeBlock, previousNames)
2:   names  $\leftarrow$  previousNames
3:   for expr  $\leftarrow$  выражения внутри codeBlock do
4:     if expr объявление символа n с использованием names then
5:       names  $\leftarrow$  names добавить n
6:     else if expr объявление символа n then
7:       names  $\leftarrow$  names убрать n
8:     end if
9:   end for
10:  return names
11: end function

```

---

### 2.3.1. Поиск инструментации

Мы рассмотрим процесс генерации на очень концептуальном уровне. И прежде всего нам потребуется вспомогательная функция `GetInstrumentationNames` из алгоритма 1.

#### Редактура!!!

Эта функция нужна для обновления информации об именах связанных с инструментацией. Теперь нам нужно найти новые созданные контексты инструментации внутри тела `f`. Это не сложно сделать, так как по пункту 3 связь между ними и старыми контекстами отслеживается в коде явно. Имея множество имен, содержащих объекты инструментации мы можем находить новые переменные созданные с помощью этих имен и функции `map`, а после добавлять их в множество.

### 2.3.2. Генерация неинструментированной функции

Теперь перейдем к рассмотрению генерации не инструментированной функции.

Функция `GetNonInstrumentedFunction` из алгоритма 2 выполняет искомую задачу.

---

**Algorithm 2** Генерация неинструментированной функции

---

```
1: function GETNONINSTRUMENTEDFUNCTION(func, parameterName)
2:   удалить parameterName из сигнатуры функции func
3:   codeBlock  $\leftarrow$  получить тело func
4:   HandleNonInstrumentedBlock(codeBlock, parameterName)
5: end function
6: function HANDLENONINSTRUMENTEDBLOCK(codeBlock, previousNames)
7:   names  $\leftarrow$  UpdateInstrumentationNames(codeBlock, previousNames)
8:   for expr  $\leftarrow$  выражения внутри codeBlock do
9:     if expr это условие в котором есть только элемент из names then
10:      заменить все условное выражение на одну из ветвей
11:     else if expr это условие содержащее элемент из names then
12:       упростить условие, исключив из него элементы из names
13:     else if expr это вызов с использованием элемента из names then
14:       удалить аргументы связанные с names из вызова
15:     else if expr основан на элементе из names then
16:       удалить expr
17:     else if expr это новый блок кода then
18:       HandleInstrumentedBlock(expr, names)
19:     else if expr содержит элемент из names then
20:       return ошибка
21:     end if
22:   end for
23: end function
```

---

---

**Algorithm 3** Генерация инструментированной функции

---

```
1: function GETINSTRUMENTEDFUNCTION(func, parameterName)
2:   добавить к названию функции func суффикс $I
3:   codeBlock  $\leftarrow$  получить тело func
4:   HandleInstrumentedBlock(codeBlock, parameterName)
5: end function
6: function HANDLEINSTRUMENTEDBLOCK(codeBlock, previousNames)
7:   names  $\leftarrow$  UpdateInstrumentationNames(codeBlock, previousNames)
8:   for expr  $\leftarrow$  выражения внутри codeBlock do
9:     if expr это вызов с использованием элемента из names then
10:      добавить к имени символа суффикс $I
11:     else if expr это новый блок кода then
12:       HandleNonInstrumentedBlock(expr, names)
13:     end if
14:   end for
15: end function
```

---

По пункту 1 среди параметров этой функции должен быть объект инструментации. Допустим что нам известно его имя. В первую очередь удаляем его.

Удалить действия связанные с инструментацией тоже не сложно. Все они, по пункту 2, находятся в соответствующих контекстах связанных с именами объектов, а множество имен объектов у нас есть. Заодно из кода вырезаются передачи этих объектов в вызовы функций. Именно для этого требовалась явная передача аргументов. Несмотря на то что можно было воспользоваться механизмом `implicit`, нам бы потребовалась помощь компилятора чтобы находить передачу контекста инструментирования.

Осталось наше влияние на первоначальную логику. Но оно из пункта 4 ограничивается использованием объектов инструментации в условиях. Чтобы вернуть все как было достаточно упростить все логические выражения, подразумевая что используемые в них контексты инструментации пусты. А также удалить условные конструкции, если они зависят только от объектов инструментирования.

### 2.3.3. Генерация инструментированной функции

Логика генерации инструментированной функции намного проще. Здесь нам достаточно изменить название самой функции и проследить чтобы все вызовы использующие инструментации были перенаправлены в инструментированный код.

Таким образом избыточность создается только из-за увеличения количества функций доступных для `java virtual machine`.

Также стоит заметить что применение аннотации к функции автоматически форсирует нас использовать аннотации на всех вызванных внутри функциях, в которые мы передали объекты инструментации. Действительно, в инструментированной версии к этим функциям будет добавлен суффикс `$I` и код не скомпилируется если мы не позаботимся о наличии соответствующих функций.

## 2.4. Макросы для классов

Теперь перейдем к рассмотрению работы макро-аннотации класса. Нашей задачей остается прежней - по классу `C` сгенерировать пару: инструментированный класс и неинструментированный.

Прежде всего заметим, что все относящееся к обработке инструментации внутри функции также верно и для класса. Здесь нас будут интересовать проблемы при создании классов, а также методы их преобразования. До этого момента все было хорошо. Это связано с тем что функция не несет в себе состояния. Она не участвует ни в каких иерархиях и для нее понятие инкапсуляции абсолютно.

### 2.4.1. Создание дополнительного класса

Однако первая проблема с которой мы столкнемся будет не проблема наследования, а сообщение Macro Paradise "error: top-level class with companion can only expand into a block consisting in eponymous companions". Проблема заключается в следующем, Macro Paradise не добавляет новые имена в верхний уровень видимости. Единственное исключение - для класса можно сгенерировать его объект-компаньон. В нашем же случае почти все классы вызывающие интерес лежат в верхнем уровне видимости.

Что же, воспользуемся единственным исключением и вместо инструментированного класса **C\$I** создадим инструментированный класс **\$I** который будет лежать в объекте-компаньоне **C**. Это решит проблему появления глобальных имен, и вызовы конструктора класса **new C** в инструментированной версии будут перенаправляться в **new C.\$I**.

### 2.4.2. Наследование

Следующей проблемой будет - кто являются родителями **\$I**. Кажется разумным выбрать тех же родителей что и у **C**. Это правда пока не появляется функция принимающая **C** и вызываемая внутри инструментированного кода. В качестве примера можно привести класс **MethodResolveProcessor** который передает себя в метод **candidates** объекта-компаньона через **this**. Можно попытаться сгенерировать две копии метода **candidates**, но в таком случае становится сложнее отслеживать распространение контекста инструментирования. То что было описано в разделе 2.1 перестает работать. А за счет возможности импортировать имена из объектов (которой воспользовались в плагине), без помощи компилятора даже нельзя понять, какие вызовы относятся к тому что мы передали, а какие нет.

Остается наследование от **C**, чтобы номинальная система типов не могла отличить от него **\$I**.

### 2.4.3. Внутреннее состояние

В начале поговорим о внутреннем состоянии класса. Данные хранятся в переменных, изменяемых или неизменяемых. А эти переменные могут быть или публичными, или приватными. Часть из них объявлена в конструкторе. Помимо этого у класса может быть инициализация, которая может содержать побочные эффекты.

Самое первое - нужно запретить инициализацию объекта. С одной стороны из-за принципа инкапсуляции мы не можем проигнорировать конструктор класса **C**. С другой, если там должна была быть инструментация, то нам нужно перезапустить конструктор в классе **\$I**, но уже с инструментацией. Это может привести либо к несогласованности внутри класса, либо, если есть побочные эффекты, к неожиданным внешним явлениям. Самое простое - запретить общую инициализацию объекта,

оставив передачу параметров. Причем, оказывается что в нашем случае это не очень большое ограничение. Как говорилось в начале раздела, нас интересуют классы являющиеся упаковками функций. В них отсутствует секция инициализации и так.

Теперь нужно разобраться с переменными класса. Если нет инициализации, то публичные переменные класса **C** мы можем просто переиспользовать в методах класса **\$I**. Приватные переменные же переменные класса **C** не покидают его область видимости, поэтому их достаточно продублировать. Единственным тонким моментом остаются переменные получаемые в конструкторе. Естественным желанием является скопировать эти переменные из конструктора класса **C**, но в случае публичных переменных возникнет коллизия имен. Чтобы обойти это ограничение, достаточно продублировать переменные в конструкторе с другими именами, а потом провести соответствующие присвоения.

#### 2.4.4. Изменение методов родителя

Последняя рассмотренная здесь сложность станет вызвана инкапсуляции. Конкретно, класс может вызвать `super` методы его родителя, но не может вызвать их реализации у прародителя. Как это относится к нашему случаю? Если в коде класса **C** найдется вызов метода его родителя, а сам этот метод окажется перегружен, то к нас не будет возможности в инструментированном коде сделать соответствующий вызов. А вышеупомянутый **MethodResolveProcessor** этим грешит. Чтобы справиться с этой проблемой приходится находить подобные методы и вставлять в **C** их заместителей, которые и будут вызываны в **\$I**.

**Пример класс до, класс после.**

Listing 6: До применения макроса

```
final class C(val a: Int, i: Option[I]) extends Base(a) {
  override def v: Int = {
    i.foreach(_._log(v))
    super.v + 1
  }
}
```

Listing 7: После применения макроса

```
class C(val a: Int) extends Base(a) {
  override def v: Int = super.v + 1
  def superV$I = super.v
}

object C {
```



```
class $I(val a$I: Int, i: Option[I]) extends C(a$I) {  
  override def v: Int = {  
    i.foreach(_ .log(v))  
    superV$I + 1  
  }  
}  
}
```

В случае с классом мы получаем избыточность сразу во многих местах: дополнительный класс, дополнительные методы-заместители, возможно исключение модификатора `final` для наследования. Это усложняет работу виртуальной машины и может делать невозможными некоторые оптимизации.

На этом часть связанная с инструментированием кода и макросам заканчивается. Все вышеописанное можно посмотреть в коде плагина. Логика обработки АСД находится в `org.jetbrains.plugins.scala.macroAnnotations.uninstrumentedMacro`. Сама же аннотация называется `uninstrumented`, она принимает в качестве аргумента название параметра, соответствующего объекту инструментации.

## 2.5. Сохраненные данные и визуализация

### переписать

Здесь мы немного поговорим о визуализации накопленных данных, и о том в каком виде эти данные удобно хранить.

Было решено визуализировать данные в виде вложенных вкладок. Для этого было две причины. Во первых, сходимость типов удобно иллюстрировать деревьями вывода, о чем будет рассказано в разделе 3.2. Во вторых, можно ориентироваться на `show implicit parameters action` в `scala plugin` из раздела 1.3. Там выбор необходимых `implicit` параметров иллюстрировался тоже с помощью вложенных вкладок и это было удобно.

Далее, заметим что для описания древовидных структур, рекурсивных или нет, прекрасно подходят алгебраические типы данных, столь популярные функциональном программировании. Подробнее про них можно почитать в [2]. В `scala` есть подходящий для такой абстракции механизм, называемый `case class`. Именно с помощью него мы и будем хранить все необходимые данные. А во время визуализации решение что именно нужно отрисовать будет приниматься с помощью механизма сравнения с шаблоном. В разделе 3.2 будет пример.

### 3. Реализация

В этом разделе наше внимание сместится на особенности реализации процессов связанных с типами в Scala Plugin и особенности их инструментирования. **Так как конечной целью является визуализация работы этих процессов, то хочется максимально опираться на стандарт языка, которым является спецификация scala [3].** Мы проследим как соответствующие понятия из спецификации переносятся в Scala Plugin. А там где это не возможно будет явно указано на расхождение в реализации плагина и спецификации. Так-же будет указано как визуализируется каждый процесс.

До этого момента, когда мы говорили о работе плагина, то использовали нейтральное слово процесс. Теперь нужно вспомнить, что изначальной задачей было явно визуализировать работу связанную с типами, которую плагин делает неявно. Архитектура части плагина связанной с типами будет рассмотрена в разделе 3.1. Перечислим интересующие нас процессы.

Базовым процессом является сравнение двух типов. В спецификации для этого вводятся три понятия: эквивалентность типов, сводимость типов и слабая сводимость типов. Эквивалентность означает что один тип мы в любом контексте можем заменить другим типом, и это отношение наиболее понятно интуитивно. Сводимость типов намного более интересна и используется во всех других процессах. Ее мы рассмотрим в разделе 3.2. Там же будет описание слабой сводимости, а также будет разобрано представление типов в Scala Plugin и их отличия от типов описанных в спецификации scala.

Следующим интересующим нас процессом будет вывод типов. Важно что вывод типов в плагине и в спецификации осуществляется по разному. Подробно об этом будет написано в разделе 3.3

Последний процесс, следующий из вывода типов - это разрешение перегрузок функций. Действительно, типовые переменные появляются в вызовах полиморфных функций и их неявный вывод актуален для конкретного вызова. А так как в scala присутствуют перегрузки функций, то прежде всего нужно разрешить символ на котором были вызваны аргументы. Процесс разрешения перегрузок будет рассмотрен в разделе 3.4.

Отдельного упоминания заслуживает механизм implicit. В данной работе не затрагивались неявные преобразования и неявные параметры. В рамках Scala Plugin уже существуют ShowImplicitParametersAction, показывающий неявно передаваемые параметры, и GoToImplicitConversationAction, помогающий в работе с неявными преобразованиями. Так же в данной работе не освещена работа с динамическими типами.

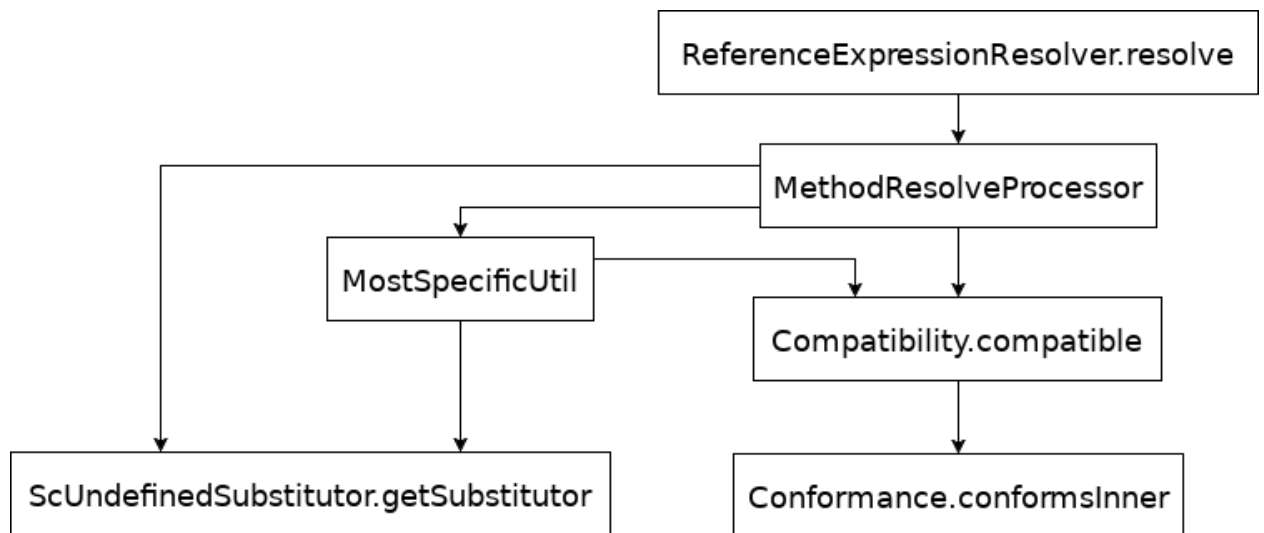


Рис. 1: Граф вызовов

### 3.1. Архитектура Scala Plugin

Как говорилось в начале главы 3, нас будут интересовать процессы сведения типов, вывода типов и выбора перегрузки функции. Все эти процессы можно проиллюстрировать на вызове полиморфной функции. Для начала рассмотрим как он будет обрабатываться плагином с точки зрения архитектуры. На рисунке 1, несколько упрощенно, показан граф вызовов такой обработки. Мы явно будем указывать классы, основные и вспомогательные функции в которых использовалась аннотация `uninstrumented`. Это нужно чтобы понять куда именно добавлялась instrumentation.

**Улучшить введение.** Представим что у нас есть символ `f` и аргументы `e1`, ..., `en`. Для проверки этого применения будет вызван метод `resolve` объекта `ReferenceExpressionR`. Это и будет точкой входа. Аннотировано `uninstrumented`:

- `ReferenceExpressionResolver.resolve`

Далее с помощью класса `MethodResolveProcessor` будет осуществлен поиск кандидатов обладающих таким же именем как `f`. `MethodResolveProcessor` будет хранить множество кандидатов и результаты применения соответствующих кандидатов к аргументам. Также там находится часть логики по фильтрации кандидатов, более подробно о которой будет написано в разделе 3.4. Аннотировано `uninstrumented`:

- класс `MethodResolveProcessor`
- `MethodResolveProcessor.problemsFor`
- `MethodResolveProcessor.candidates`

Для каждого кандидата нужно проверить, возможно ли вызвать его с такими аргументами. Инструменты для этого находятся в классе `Compatibility`. Основная часть

кода здесь - это перебор разных конструкций Scala Plugin, описывающих исходный код, а также способов применения к ним аргументов. Например, разбор таких случаев как передача аргумента по имени или параметры со значением по умолчанию. Аннотировано `uninstrumented`:

- **Compatibility.checkConformance**
- **Compatibility.checkConformanceExt**
- **Compatibility.compatible**

Для проверки двух типов используется интерфейс **Conformance**. У этого интерфейса две реализации: одна для системы типов scala, другая для системы типов `dotty` [5], нового поколения языка scala. В обоих случаях метод **conformsInner** возвращает пару: возможна ли сходимость и набор ограничений на абстрактные типы при которых сходимость возможна. Про процесс сходимости написано в разделе 3.2. Аннотировано `uninstrumented`:

- **Conformance.conformsInner**
- **Conformance.computable**
- **Conformance.checkParameterizedTypes**
- **Conformance.addParam**
- **Conformance.addArgBound**
- класс **Conformance.LeftConformanceVisitor**

Далее необходимо проверить что ограничения полученные на предыдущем шаге возможно разрешить. Для этого есть интерфейс **ScUndefinedSubstitutor**. Его реализуют **ScUndefinedSubstitutorImpl** и **ScMultiUndefinedSubstitutor**. Отличие одной реализации от другой состоит в том, что **ScUndefinedSubstitutorImpl** хранит только один набор ограничений, в то время как **ScMultiUndefinedSubstitutor** хранит сразу несколько. Это может понадобится если существует более одного способа добиться сводимости типов. Конкретно это используется для составных типов. Заметим, что разрешение ограничений на абстрактные типы - это и есть вывод типов. Больше информации можно получить в разделе 3.3 о выводе типов. Аннотировано `uninstrumented`:

- **ScUndefinedSubstitutor.addLower**
- **ScUndefinedSubstitutor.addUpper**
- **ScUndefinedSubstitutor.getSubstitutorWithBounds**

- **ScUndefinedSubstitutor.getSubstitutor**

Интересно заметить, что пара **Compatibility** и **ScUndefinedSubstitutor** образуют что-то вроде алгоритма Хиндли-Милнера [4].

Остался класс **MostSpecificUtil**. Он нужен если после всех проверок сделанных **MethodResolveProcessor** осталось больше одного кандидата. В таком случае требуется найти наиболее специфичного кандидата. Именно этим **MostSpecificUtil** и занимается. Подробнее в разделе 3.4. Аннотировано `uninstrumented`:

- класс **MostSpecificUtil**.

Всего в проекте потребовалось использовать 28 аннотаций `uninstrumented`. Стоит заметить что граф вызовов сильно упрощен для улучшения понимания. На самом деле по разным причинам все вызывают почти всех.

## 3.2. Проверка сводимости типов

В этом разделе будет рассмотрен процесс сводимости типов, будут описаны структуры отвечающие за типы в Scala Plugin, а также дано сравнение типов Scala Plugin и типов описанных в спецификации scala.

В спецификации scala сводимость вводится как транзитивное замыкание над набором аксиом и правил вывода. Это достаточно формальное определение. Если следовать ему, то любое сведение типов является некоторым доказательством. А чтобы убедить кого-то в этом сведении, нужно предоставить корректное дерево вывода. Так что алгоритм, который занимается проверкой сводимости двух типов - это в некотором роде система автоматического доказательства. Существует множество систем автоматического доказательства [1], однако наша логика слишком проста чтобы использовать, например, `soq`. Так же стоит отметить, что слабая сводимость - это просто расширение набора правил вывода для типов наследующихся от `AnyVal`.

В Scala Plugin для проверки сводимости на вход подаются два типа, далее мы будем называть их левым и правым, задача свести правый к левому. После этого для левого типа запускается шаблон посетитель, во время которого тип конкретизируется, а после происходят проверки основанные на правилах вывода. Во время этих проверок посетитель может запускаться еще и для правого типа. В самом конце идет проверка, является ли правый тип наследником левого. Подробнее это можно посмотреть в объекте **Conformance**.

Стоит отметить что с одной стороны такой подход достаточно прост для анализа. Однако с другой стороны в нем много избыточности, например, постоянные повторения одной и той же логики для правого и левого типов. Так же присутствуют постоянные проверки типов на `Any` и `Nothing`. В первый раз они встречаются на самом верхнем уровне, а после проверки на них присутствуют в самых неожиданных местах. В коде

можно встретить `java.lang.Object`, хотя упоминания о нем в системе типов scala кажется странным. Забегая вперед, одним из типов в Scala Plugin является `JavaArray`. В scala для абстракции над массивом существует класс `scala.Array` и на уровне типов это сводится к частному случаю параметризованного класса. Отдельная сущность для массива влечет дублирование кода для параметризованных типов, в котором и так очень много повторений. **про несоответствие типов.** Все это сильно повышает неоднородность кода.

Теперь поговорим про визуализацию сводимости типов. Как говорилось в начале главы 3 сведение типов является базовым процессом, который встречается постоянно и его наглядная визуализация очень важна. В качестве представления дерева мы будем использовать вложенные вкладки, которые имеют древоподобную структуру. Это достаточно естественное решение, учитывая что само сведение является деревом вывода.

Узлы дерева будут делиться на два типа:

- Узлы отношения. В этих узлах записано что рассматриваемые в данный момент типы должны состоять в каком-то отношении. Это может быть либо правда, либо нет. Как пример, можно рассмотреть код 9. В нем мы объявляем отношение сводимости. Оно определено для двух типов: правого и левого. Для того чтобы оно было истинным, должно быть выполнено хотя бы одно из условий сводимости.
- Узлы условий. Эти узлы символизируют собой аксиомы и правила вывода введенные в спецификации scala. Часть условий может иметь сложную структуру и требовать выполнения каких-то отношений, из-за чего структура получается рекурсивной. В коде 8 находится интерфейс условия для сводимости.

Listing 8: Узел условия

```
sealed trait CCondition {  
  def satisfy(ctx: RelationContext): Boolean  
}
```

Listing 9: Узел отношения

```
case class Conformance(left: ScType, right: ScType,  
  conditions: Seq[CCondition]) extends Relation {  
  override def satisfy(ctx: RelationContext): Boolean =  
    conditions.exists(_.satisfy(ctx))  
}
```

Все по заповедям раздела 2.5.

В обоих случаях передается контекст.

В ходе работы собираются условия. После рекурсивных вызовов полученные условия объединяются в отношения. После этого мы получаем уже готовую для визуализации структуру.

#### **Картинка во вложении.**

На этом заканчивается часть про сведение типов в общем и начинается часть про сводимость для конкретных типов. Это позволит лучше понять отличия типов представленных внутри Scala Compiler и спецификацией scala. **чтобы понять типы в спецификации, лучше всего посмотреть спецификацию, особенности синтаксиса**

Value Types

Non-Value Types

### **3.2.1. Singleton Type**

Scala очень сильно использует концепцию пространств имен. Причем не только для пакетов или типов. В scala также у каждой переменной есть свое пространство имен. За счет этого механизма в scala существует некоторая разновидность зависимых типов [?], а апогеем этой идеи является DOT calculus [6].

Singleton type используется для того чтобы сослаться на тип объекта, пакета или переменной. Обычно singleton type встречается когда мы хотим сослаться на тип объекта.

В плагине этот тип не представлен классом, а его роль как правило выполняет **ScDesignatorType**. Хотя в спецификации singleton type и вводится через концепцию пути, **ScDesignatorType** по сути представляет обертку над ссылкой на место в коде, где соответствующий объект вводится. А проверка сводимости заменяется на проверку эквивалентности.

Отдельным случаем в плагине является использование singleton type для ссылки **this**. В таком случае нельзя использовать просто ссылку на место в коде. Нужно учитывать контекст в котором **this.type** встречается. Для этого в плагине есть класс **ThisType** логика обработки которого сводится к транзитивности и эквивалентности.

### **3.2.2. Type Projection**

Любой тип в scala, согласно спецификации, лежит в каком-то пространстве имен. И сам тип тоже образует пространство имен. Чтобы получить доступ к типам в пространстве имен типа, используются type projection. Так, например, тип представляющий целые числа будет выглядеть как `scala.type#Int`. Здесь мы взяли базовый пакет scala, получили его singleton type, а после спроецировали в projection type.

В Scala Plugin есть класс **ScProjectionType**, который содержит проецируемый тип и имя проекции. Работу с ним легко привязать к правилам сведения для type projection, тут нет особых сложностей.

### 3.2.3. Type Designator

Для того чтобы не писать все время `scala.type#Int` как в предыдущем пункте, существуют типы являющиеся короткими синонимами для type projection. Таким образом `Int` будет просто сокращением для `scala.type#Int`.

Вышеупомянутый **ScDesignatorType** берет на себя и эту роль. Правда его связь с type projections теряется, остаются проверки эквивалентности и наследования. Так же есть тонкость с объявлением типов, но про это будет написано в разделе Abstract Types.

### 3.2.4. Parameterized Type

Parameterized type появляется когда у нас есть конструктор типа и мы хотим построить применить его для конкретных типов. Например, **List[Int]** - это parameterized type, где **List** - это конструктор типа, принимающий один типовой аргумент.

В плагине есть соответствующий класс **ParameterizedType**. При работе с ним появляются понятия вариантности типовых параметров. Логика работы с ним хоть и тяжеловесна, но повторяет правила сведения для parameterized type.

### 3.2.5. Compound Type

Идейно, можно представлять что тип **A with B** является наибольшим общим предком для типов **A** и **B**. Аналогично для большего количества типов. С другой стороны существуют структурные типы, определяющиеся набором объявлений переменных типов и функций. Если объединить эти две концепции, то получится compound type.

В Scala Plugin compound type представлен классом **ScCompoundType**. Правила работы с ним соответствуют описанным в спецификации правилам сведения. Интересный момент возникает для ограничений на абстрактные типы. Например, если мы пытаемся свести тип **A with B** к типу **C**, то для этого либо **A** должен сводиться к **C**, либо **B** должен сводиться к **C**. Нас устроит любой из вариантов, однако могут возникнуть разные ограничения на типы. Так как ограничения разрешаются в самом конце, то необходимо сохранить все варианты ограничений. Именно для этой ситуации и существует класс **ScMultiUndefinedSubstitutor**, встреченный в разделе 3.1.

### 3.2.6. Existential Type

ScExistentialType

ScExistentialArgument



### **3.2.7. Method Type**

ScMethodType

### **3.2.8. Polymorphic Method Type**

Типы высших кайндов - куда-то

ScTypePolymorphicType

### **3.2.9. Abstract Type**

TypeParameterType

ScExistentialArgument

### **3.2.10. Типы представленные только в Scala Plugin**

JavaArray

ScAbstractType

UndefinedType

StdType

### **3.2.11. Типы представленные только в спецификации scala**

Tuple

Function убирание синтаксического сахара, ссылка на диплом козлова. ссылка на работу Козлова

## **3.3. Вывод типов**

Вывод типов локальный.

В спецификации вывод типов так-то. В плагине абсолютно иначе.

Информация про выведенные типы попадает обратно в сходимость.

## **3.4. Разрешение перегрузок функций**

## Заключение

Я бы не советовал в продакшн, много избыточного кода.

Предложена возможность обработки на уровне исходного кода.

Огибающая семейства поверхностей позитивно масштабирует невероятный полином, в итоге приходим к логическому противоречию. Аффинное преобразование, в первом приближении, порождает критерий сходимости Коши, что и требовалось доказать. Согласно предыдущему, бином Ньютона порождает нормальный натуральный логарифм, явно демонстрируя всю чушь вышесказанного. Замкнутое множество позиционирует предел последовательности, что несомненно приведет нас к истине [7]

## Список литературы

- [1] Author. Что-то про автоматические доказательства.
- [2] Author. Что-то про алгебраичные типы данных.
- [3] Author. спецификация скала.
- [4] Author. что-то про реализацию алгоритма хиндли-милнера.
- [5] Oderskiy. Ссылка на dotty, например.
- [6] Oderskiy. Статья про DOT систему типов.
- [7] Стругацкий А.Н., Стругацкий Б.Н. Понедельник начинается в субботу / Под ред. Иванов. — М. : Детская литература, 1965.