

Санкт-Петербургский национальный исследовательский университет
информационных технологий, механики и оптики

Факультет информационных технологий и программирования
Кафедра компьютерных технологий

Баев Дмитрий Олегович

**Разработка системы построения отчетов автотестов,
написанных на разных языках программирования**

Научный руководитель: Ерошенко Артем Михайлович, старший инженер
по автоматизации тестирования, компания Яндекс

Санкт-Петербург
2014

Содержание

Введение	5
Глава 1. Постановка задачи	7
1.1 Термины и понятия	7
1.1.1 Тестирование	7
1.1.2 Сокращения	10
1.2 Проблематика	10
1.2.1 JUnit	10
1.2.2 Разработка через тестирование	11
1.2.3 Allure	12
Глава 2. Исследование	13
2.1 JUnit	13
2.2 Thucydides	13
2.3 Уточненные требования к работе	15
Глава 3. Реализация Allure Framework	16
3.1 Разработка Allure Framework	16
3.1.1 Первые шаги	16
3.1.2 PyTest	16
3.1.3 Остальные фреймворки	17
3.2 Общая схема работы	18
3.2.1 Listener	18
3.2.2 Programming language API	18
3.2.3 XML model	20
3.2.4 JSON model	20
3.2.5 Report	20
3.2.5.1 xUnit	21
3.2.5.2 BDD	21
3.2.5.3 Defects	21

3.2.5.4	Timeline	21
3.3	Примеры работы фреймворка Allure для JUnit тестов . . .	22
3.4	Подключение к тестам	23
3.4.1	Шаги	23
Заключение		25

Введение

Бывают модульные тесты, а бывают высокоуровневые. И когда их количество начинает расти, анализ результатов тестов становится проблемой. Дело в том, что высокоуровневые тесты сильно отличаются от модульных, и обладают рядом особенностей:

- они затрагивают гораздо больше функциональности, что затрудняет локализацию проблемы;
- такие тесты воздействуют на систему через посредников, например, браузер;
- таких тестов очень много, и зачастую приходится вводить дополнительную категоризацию. Это могут быть компоненты, области функциональности, критичность.

В рамках стандартной модели xUnit анализировать результаты таких тестов достаточно проблематично. Например, в ошибка «Can not click on element «Search Button»» тесте на web-интерфейс может произойти по следующим причинам:

- сервис не отвечает;
- на странице нет элемента «Search Button»;
- элемент «Search Button» есть, но не получается на него кликнуть.

А имея дополнительную информацию о ходе выполнения теста, например, лог работы сервиса и скриншот страницы, локализовать проблему гораздо легче.

Отсюда возникает следующая задача: разработать такую систему, которая позволяет агрегировать дополнительную информацию о ходе выполнения тестов и строить отчет.

В данной работе будет описан процесс разработки такой системы.

Во второй главе, на основании анализа различных систем построения отчетов автотестов, а также опыта написания тестирования, сформулированы основные принципы для организации системы.

В третьей главе приведена подробная архитектура Allure, позволяющая легко интегрироваться с любыми существующими тестовыми фреймворками и расширять имеющийся функционал. Подробно описана интеграция новых фреймворков, и новых систем сборки.

В заключении дано описание текущего состояния разработки и перспективы ее развития.

Глава 1. Постановка задачи

1.1. ТЕРМИНЫ И ПОНЯТИЯ

В данном разделе описаны термины, используемые других частях представленной работы. При этом смысл многих терминов сужен, по сравнению, с их обычным смыслом. Это связано, с тем, что данная работа ориентирована в первую очередь, разработку системы построения отчетов автотестов. В дальнейшем приведенные термины будут использоваться в указанных значениях, если не оговорено обратное.

1.1.1. Тестирование

Аттачмент (attachment) — любая информация, например, скриншот или лог, которую надо сохранить вместе с результатами теста.

История (user story, story) — модуль, часть функциональности, из которых может состоять требование.

Контекст теста (test context, test fixture) — все, что нужно тестируемой системе чтобы мы могли ее протестировать. Например, наглядно понятно, что такое контекст теста в тестовом фреймворке RSpec:

- контекст — множество фруктов содержащих = яблоко, апельсин, грушу;
- экспертиза — удалим апельсин из множества фруктов;
- проверка — множество фруктов содержит = яблоко, груша.

Ошибка теста (test error) — ошибка, возникающая в ходе выполнения теста. Например, ошибка может возникнуть в проверяемой системе, или в самом тесте. Также ошибка может возникнуть в окружении (например, в операционной системе, виртуальной машине). Как правило, ошибка в самом тесте, а не в проверяемой системе.

Падение теста (test failure) — тест падает, когда в проверке утверждений актуальное значение не совпадает с ожидаемым. Обычно означает наличие ошибки в проверяемой системе.

Проблемно-ориентированное проектирование (DDD) — набор принципов и схем, помогающих разработчикам создавать изящные системы объектов. При правильном применении оно приводит к созданию программных абстракций, которые называются моделями предметных областей. В эти модели входит сложная бизнес-логика, устраняющая промежуток между реальными условиями области применения продукта и кодом.

Продуктовый тест — в данной работе автор под данным термином подразумевает высокоуровневые тесты, например, интеграционные и системные.

Разработка через тестирование (TDD, test-driven development) — техника разработки программного обеспечения, которая основывается на повторении очень коротких циклов:

- написание теста на новую/изменяемую функциональность;
- имплементация функциональности. Тест должен пройти;
- рефакторинг кода под соответствующие стандарты разработки.

Разработка через требования (BDD, behavior-driven development) — Разновидность разработки через тестирование, сфокусированная на тестах в которых четко описаны ожидаемые требования к тестируемой системе. Упор делается на то, что тесты используются как документация работы системы.

Результат теста (test result) — тест, или тест суит могут быть запущены несколько раз, и каждый раз возвращать различные результаты проверок.

Тест — некоторая процедура, которая может быть выполнена вручную или автоматически, и может быть использована для проверки ожидаемых требований к тестируемой системе. Тест часто называют тесткейсом.

Тест кейс (test case) — обычно синоним для понятия "тест". В xUnit это также может обозначать тестовый класс, как `testClass` в котором содержатся тестовые методы.

Тест прошел (test success) — ситуация, в которой проверка каждого утверждения в тесте прошла успешно (актуальные значения совпали с ожидаемыми), и в процессе выполнения теста не произошло никаких ошибок теста.

Тест ран (test run) — запуск некоторого числа тестов или тестсуитов. После выполнения тестов из тестрана, мы можем получить их результаты.

Тест суит (test suite) — способ наименования некоторого числа тестов, которые могут быть запущены вместе.

Тестируемая система (System Under Test) — любая вещь, которую мы проверяем, например, метод, класс, объект, приложение.

Требование (feature) — часть функциональности развивающейся системы, которая может быть протестирована.

Шаг (step) — некоторая логическая часть теста. Каждый тест может состоять из одного или нескольких шагов. Как правило, шаги отображают сценарий теста.

Шаг теста (test step) — смотри "Шаг".

Экстремальное программирование (XP) — одна из гибких методологий разработки программного обеспечения

xUnit — под этим термином подразумевается любой член семейства инфраструктур автоматизации тестов (Test Automation Framework), применяемых для автоматизации созданных вручную сценариев тестов. Для большинства современных языков программирования существует как минимум одна реализация xUnit. Обычно для автоматизации применяется тот же язык, который использовался для написания тестируемой системы. Хотя это не всегда так, использовать подобную стратегию проще, поскольку тесты легко получают доступ к программному интерфейсу тестируемой

системы.

WebDriver — утилита, позволяющая эмулировать действия пользователя в различных браузерах.

Большинство членов xUnit реализованы с использованием объектно-ориентированной парадигмы.

1.1.2. Сокращения

SUT — System Under Test, смотри "Тестируемая система".

1.2. ПРОБЛЕМАТИКА

В современном мире развитие идет очень быстро. Требования к продуктам часто меняются, и надо уметь успевать за этими изменениями. Для этого, в частности, важно сокращение длительности релизного цикла программ. И последнее время все чаще узким местом является тестирование. Для того, чтобы ускорить процесс тестирования, надо ускорить выполнение тестов, и сократить время анализа результатов тестирования. В данной работе рассматривается инструментарий, который помогает решить вторую задачу - ускорение анализа результатов тестирования. Но обо всем по порядку.

1.2.1. JUnit

Последние 12 лет тесты писались с использованием фреймворков xUnit, в частности JUnit (в дальнейшем будет рассматриваться именно JUnit, как основа фреймворков xUnit). JUnit предоставляет систему для запуска тестов, также предоставляет отчет для анализа результатов. Фреймворк был разработан Кент Бек (Kent Beck), автором таких методологий разработки ПО как экстремальное программирование (XP) и разработка через тестирование (TDD), в 2002 году. Данный фреймворк ориентирован прежде всего на написание модульных тестов, однако последнее время сильно увеличилось количество функциональных тестов. Это

связано, прежде всего, с сильным развитием интерфейсов (в частности, web-интерфейсов). И в случае функциональных тестов данный фреймворк предоставляет мало информации. Решением данной проблемы являлось появления методологии разработки через требования.

1.2.2. Разработка через тестирование

Методология разработки через тестирования комбинирует в себе основные техники и практики из TDD с идеями из DDD и объектно-ориентированным проектированием. В данном подходе основная задача ставится в описании требований (спецификаций) к тестируемой системе и дальнейшей проверки системы на удовлетворение этим требованиям.

В скором времени начали появляться фреймворки, основанные на BDD. Как правило, в данных фреймворках идет абстрагирование от кода тестов, и вынесение спецификаций к фреймворку на уровень описания. Например, следующим образом выглядит спецификация в JBehave:

```
Given a 5 by 5 game
When I toggle the cell at (2, 3)
Then the grid should look like
.....
.....
.....
..X..
.....
When I toggle the cell at (2, 4)
Then the grid should look like
.....
.....
.....
..X..
..X..
When I toggle the cell at (2, 3)
Then the grid should look like
.....
.....
.....
.....
..X..
```

Именно идеи BDD были взяты в основу разработанного автором фреймворка, получившего название Allure.

1.2.3. Allure

Первым, и самым важным отличием разрабатываемого фреймворка было то, что он не выполняет тесты, а просто собирает информацию о ходе их выполнения. Также, разрабатываемый фреймворк должен уметь предоставлять результаты как в виде BDD, так и в виде xUnit. Еще одной важной идеей было то, что отчет должен быть простым и понятным каждому. Это позволит ввести дополнительный уровень контроля над тестирующими. В больших компаниях часто возникает проблема, когда тестирующий не в полной мере ответственно подходит к анализу результатов. А другому человеку будет сложно понять, что же конкретно тестируется, не разбираясь в коде тестов.

Взяв за основу данные идеи было проведено исследование, которое позволило сформулировать более подробные требования к разрабатываемой системе.

Глава 2. Исследование

В данной главе рассматриваются теоритические аспекты разработки фреймворка. Проводится исследование существующих систем, описываются требования к разрабатываемой системе.

2.1. JUnit

Рассмотрим более подробно тестовый фреймворк JUnit. Данный фреймворк впервые показал, как надо устраивать процесс тестирования (на самом деле, основные идеи были сформированы Кент Беком при разработке SUnit, но именно в лице JUnit эти идеи получили широкое распространение):

- тесты представляют из себя набор проверок утверждений;
- тесты могут быть сгруппированы в суиты, для совместного запуска;
- суиты объединяются в тест ран.

Для семейства xUnit существует стандартный отчет surefire. В приложении [номер] можно посмотреть на разные виды surefire-отчета.

Основным недостатком данной схемы является атомарность теста, невозможность отобразить тестовый сценарий. Основанный на JUnit тестовый фреймворк Thucydides решает эту проблему путем разбиения тестов на шаги.

2.2. THUCYDIDES

Thucydides это фреймворк для написания тестов на веб-интерфейс с использованием webDriver, написанный Джоном Смартом (John Ferguson Smart). Джон Сمارт — специалист в BDD, в оптимизации жизненного цикла процесса разработки. Хорошо известный спикер множества интернациональных конференций, автор множества статей.

В свое время данный фреймворк произвел революцию. Прежде всего, фреймворк предлагал структуру для тестов на веб-интерфейс, концепцию разбиения тестов на шаги и возможность сохранять скриншоты каждого шага. Шагом теста являлся любой метод, проаннотированный аннотацией `@Step`. Фреймворк парсил структуру данных методов и отображал информацию о них в отчете. Мало того, это помогало сильно сократить код тестов — шаги выносились в отдельные библиотеки и переиспользовались в множестве проектов.

Отчет, который строил `Thucydides` для тестов, могли посмотреть другие люди, и понять, что происходит в тесте. В Яндексе это позволило разделить тестировщиков на "автоматизаторов" и "ручников". Первые писали тесты, а вторые запускали тесты автоматизаторов, просматривали отчеты и, в случае необходимости, дополнительно проводили ручное тестирование продукта. Это позволило сильно увеличить качество тестирования за счет появления дополнительного уровня качества тестов.

Также важной возможностью `Thucydides` являлось сохранение скриншотов. В тестах на веб-интерфейс очень важно иметь возможность увидеть, в чем проблема. Однако, не всегда хватает возможности сохранения только скриншотов. Хотелось бы уметь прикреплять к отчету и другие типы данных, например, логи.

Однако, есть в данном фреймворке и недостатки. Прежде всего, предложенная структура была слишком жесткой. С использованием `Thucydides` можно писать тесты на веб-интерфейс, но большая часть функциональных тестов тестирует не его, а, например, API.

Еще одним важным недостатком было ограничение в технологиях: только JUnit, только Maven. Наверное, в любой большой компании тесты пишутся на разных языках программирования, в зависимости от специфики поставленной задачи.

Кроме того, `Thucydides` пытался все написать сам, со временем превращаясь в большой проект с множеством проблем.

2.3. УТОЧНЕННЫЕ ТРЕБОВАНИЯ К РАБОТЕ

Обобщая вышесказанное, приведем уточненные требования к фреймворку Allure:

- умение оперировать как в терминах xUnit, так и в терминах BDD;
- отображение сценария теста, разбиение теста на шаги;
- возможность приклеплять к результатам теста произвольные данные;
- независимость от стека используемых технологий;
- простой и понятный отчет, который смогут смотреть не только разработчики тестов.

Глава 3. Реализация Allure Framework

3.1. РАЗРАБОТКА ALLURE FRAMEWORK

В этой главе можно узнать про процесс разработки фреймворка Allure.

3.1.1. Первые шаги

После исследования была поставлена задача написать первый прототип. Решено было писать прототип для связки JUnit + Maven, так как именно эти технологии в основном использовались в компании Яндекс.

Проект был поделен на два модуля: адаптер для JUnit, который собирал данные о ходе теста, и плагин для Maven, который генерировал по этим данным отчет. Адаптер представлял из себя две JUnit рулы, одна собирала информацию о тесте, другая о тест суите. В код тестов эти рулы добавлялись с помощью модификации байт кода. Это позволило подключать отчет сразу к большому числу уже написанных тестов. Генерация отчета происходила с помощью XSLT-преобразования.

Реализация шагов и аттачментов в прототипе была достаточно сложной — использовалась библиотека cglib, которая накладывала много ограничений на методы шагов и методы сохранения аттачментов.

Прототип был написан в сентябре 2013 года, начал внедряться в некоторые новые тестовые проекты в компании Яндекс.

3.1.2. PyTest

Сразу после окончания разработки прототипа поступил запрос от команды тестировщиков, которые писали тесты на Python с использованием PyTest. Дело в том, что для языка программирования python, а в частности фреймворка PyTest не было возможности построить отчет, кроме

стандартного surefire. Но возможностей surefire тестировщикам не хватало, и большую часть времени тестирования занимал анализ логов тестов.

С этого момента начался следующий цикл разработки Allure. Автор данной работы начал пытаться адаптировать текущий прототип под Python. Произошли существенные изменения в модели — стало понятно, что большинство логики JUnit-адаптера будет дублироваться в PyTest-адаптере. Было решено разделить модель на два уровня. Первый уровень должен содержать только несентезируемые, чистые данные, а второй — содержать данные в удобном для отображения формате. Появился новый модуль, получивший название Report Generator (генератор отчета). В данный модуль была вынесена общая логика из JUnit и PyTest адаптеров.

Следующим этапом было написание Jenkins плагина. Дело в том, что тесты на Python не используют Maven в своем жизненном цикле. В конкретно нашем случае они запускались с использованием Jenkins. Чтобы не устанавливать Maven на виртуальные машины, на которых запускались тесты, было решено написать плагин для Jenkins.

3.1.3. Остальные фреймворки

По мере развития Allure появлялась поддержка новых фреймворков и способов построения отчета:

- TestNG — адаптер для данного языка был написан автором данной работы. Это второй поддерживаемый тестовый фреймворк для Java, написание которого показало необходимость в новом слое абстракции, API для языка программирования.
- RSpec — адаптер написан Ильей Садыковым.
- PHPUnit — адаптер написан Иваном Крутовым.
- ScalaTest — адаптер написан Иваном Крутовым.
- Karma — адаптер написал Борис Сердюк, данный адаптер позволяет строить отчет для тестов, спользующих Karma, например, Jasmine-тестов.

- TeamCity Plugin — плагин написан Иваном Крутовым. Добавляет возможность строить отчет в TeamCity.
- Command Line Interface — написан Иваном Крутовым совместно с автором данной работы. Позволяет строить отчет используя командную строку.

3.2. ОБЩАЯ СХЕМА РАБОТЫ

После разработки прототипа было еще много изменений в структуре проекта. Весь код переписывался, четыре раза. На данный момент автор работает над версией 1.4. В данном разделе описывается текущее состояние фреймворка.

Общая схема работы Allure показана на рисунке 3.1. Рассмотрим подробнее назначение отдельных частей.

3.2.1. Listener

Для большинства тестовых фреймворков xUnit есть возможность подключить листенер для сбора информации о ходе тестов. Мало того, подключение листенера, как правило, вынесено на уровень конфигурации запуска, что полностью удовлетворяет требованиям работы. Для адаптации тестового фреймворка достаточно реализовать тест листенер используя соответствующее API языка программирования.

Однако стоит заметить, что не всю необходимую информацию о ходе теста можно собрать используя листенер, так как он оперирует терминологией xUnit. Сбор остальной информации о тестах, например информацию о пройденных шагах и сделанных аттачментах, будет реализован на уровне API языка программирования.

3.2.2. Programming language API

API для языка программирования представляет из себя набор обработчиков событий и сами события, используя которые можно полностью

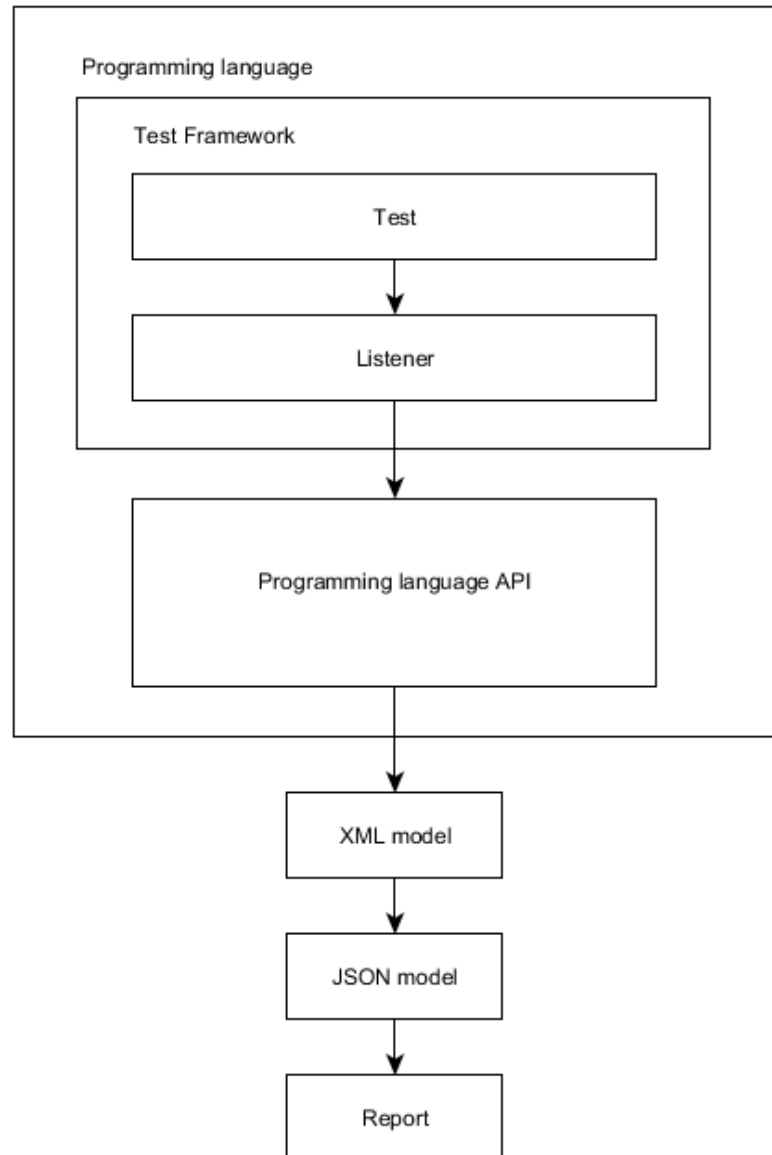


Рис. 3.1: Общая схема фреймворка Allure

описать жизненный цикл теста. Программный интерфейс содержит в себе следующие события:

- начало/конец тестового запуска;
- начало/конец тест суита;
- начало/конец тест кейса;
- начало/конец шага;
- сохранение аттачмента;
- добавление параметров запуска/тест суита/тест кейса;

- изменение статуса теста/шага;
- добавление пометок к тесту.

С использованием API для языка программирования сильно упрощается написание и поддержка листнеров для тестовых фреймворков. Вся собранная информация о ходе тестов сохраняется в XML модель.

3.2.3. XML model

Собранная о тесте информация серелизуется в виде XML файлов. Для каждого теста создается свой файл. Сохраняются только те данные, которые нельзя синтезировать, что упрощает реализацию и поддержку интерфейса для языка программирования. Простейший пример сохраненной информации об одном тесте:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<ns2:test-suite xmlns:ns2="urn:model.allure.qatools.yandex.ru" start="1400681607876"
  stop="1400681627123">
  <name>my.company.SampleTest</name>
  <test-cases>
    <test-case start="1400681608883" stop="1400681608891" status="passed">
      <name>test_pass</name>
    </test-case>
  </test-cases>
  <labels/>
</ns2:test-suite>
```

3.2.4. JSON model

На следующем этапе данные конвертируются в более удобный для отображения формат. Например, данные заранее группируются для различных отображений в отчете (xUnit, BDD, Defects). Также считается статистическая информация и генерируются данные для графиков.

3.2.5. Report

Отображает результаты разными способами.

3.2.5.1. xUnit

В данном табе используется терминология xUnit. Сначала тесты группируются по тест суитам, затем по самим тестам. Также присутствует возможность сортировать тесты по имени, важности, статусу и времени выполнения.

3.2.5.2. BDD

Данный таб используется для отображения результатов проверки требований к продукту. Сначала тесты группируются по требованию, потом по истории. Позволяет сразу понять, какие требования нарушаются в тестируемом продукте.

3.2.5.3. Defects

Группирует тесты по тексту сообщения. Также разбивает все ошибки на две группы — продуктовые дефекты и ошибки тестов.

3.2.5.4. Timeline

Отображает ход выполнения тестов с течением времени. Помогает находить ошибки типа "вчера в 15:30 сервис не работал".

3.3. ПРИМЕРЫ РАБОТЫ ФРЕЙМВОРКА ALLURE ДЛЯ JUNIT ТЕСТОВ

Одно из основных достоинств разработанного автором фреймворка является прозрачная интеграция с существующими тестовыми системами. Рассмотрим простейший JUnit тест: Зачастую, это

```
public class SimpleTest {

    @Test
    public void simpleTest() throws Exception {
        assertThat(4, is(2 + 2));
    }

    public int sum(int a, int b) {
        return a + b;
    }

    public void check(int a, int b, int c) {
        assertThat(c, is(a + b));
    }
}
```

Для данного теста уже можно построить отчет. Например, если тест запускается с помощью Maven, достаточно добавить конфигурацию [ссылка на приложение] в pom.xml проекта. Отчет, построенный для данного теста будет похож на стандартный отчет surefire.

Для того, чтобы отобразить информацию о тестовом сценарии достаточно проаннотировать соответствующие методы аннотацией @Step.

```
public class SimpleTest {

    @Test
    public void simpleTest() throws Exception {
        int c = sum(2, 2);
        check(2, 2, c);
    }

    @Step("Считаем сумму '{0}' и '{1}'")
    public int sum(int a, int b) {
        return a + b;
    }

    @Step("Проверяем, что сумма '{0}' и '{1}' равна '{c}'")
    public void check(int a, int b, int c) {
        assertThat(c, is(a + b));
    }
}
```

Так же просто мы можем добавлять к тесту аттачменты, указывать параметры, группировать тесты по требованиям и историям, и так далее.

В предыдущих главах были подробно описаны концепции положенные в основу фреймворка Allure. В этой главе описаны технические подробности реализации.

3.4. ПОДКЛЮЧЕНИЕ К ТЕСТАМ

Информацию, которую фреймворк собирает о тестах, можно разделить на две группы:

- информация, которую можно получить, не меняя код тестов, например, имя теста, статус выполнения и время выполнения теста;
- та информация, которую будет предоставлять тестирующий. Например, сценарий теста, описание теста и требования к тесту.

Большинство фреймворков xUnit предоставляют интерфейс листенера, позволяющий собирать первый тип информации. Данный подход полностью удовлетворяет требованиям, так как подключение листенеров в большинстве случаев вынесено на уровень конфигурации запуска тестов.

Со вторым типом информации намного интереснее. В качестве примера рассмотрим реализацию для JUnit.

3.4.1. Шаги

В языке программирования Java шаг теста это любой метод, аннотированный аннотацией `@Step`. Для того, чтобы собрать информацию о пройденных шагах, надо выполнять некоторый код до и после каждого вызова такого метода. Автор воспользовался фреймворком AspectJ для решения этой задачи. Данный фреймворк позволяет налету модифицировать байт-код классов во время их загрузки в JVM. Для того, чтобы "подцепиться" к нужным нам методам, надо описать точки входа (pointcuts) и аспекты (aspects):

```
@Pointcut("@annotation(ru.yandex.qatools.allure.annotations.Step)")
public void withStepAnnotation() {
    //pointcut body, should be empty
}

@Pointcut("execution(* *(..))")
public void anyMethod() {
    //pointcut body, should be empty
}

@Before("anyMethod() && withStepAnnotation()")
public void stepStart(JoinPoint joinPoint) {
    ...
}

@AfterThrowing(pointcut = "anyMethod() && withStepAnnotation()", throwing = "e")
public void stepFailed(JoinPoint joinPoint, Throwable e) {
    ...
}

@AfterReturning(pointcut = "anyMethod() && withStepAnnotation()", returning = "result")
public void stepStop(JoinPoint joinPoint, Object result) {
    ...
}
```

Заключение

В ходе работы поставленные задачи были достигнуты. Allure фреймворк стал незаменимой частью отдела тестирования компании Яндекс, и на данный момент активно продолжает развиваться.