Test IT × Девелоника FAI

Сервис аналитики прогонов

Пайплайн для анализа причин падений

Спикер:

Дмитрий Ермакович

Старший разработчик группы разработки интеграций Test IT («Девелоника» fabricaONE.AI)

План доклада

1 Проблематика и ключевая идея сервиса

2 Трудности и их решения

3 Что получилось и как сделать у себя

4 Ограничения реализации и выводы

Знакомая ситуация

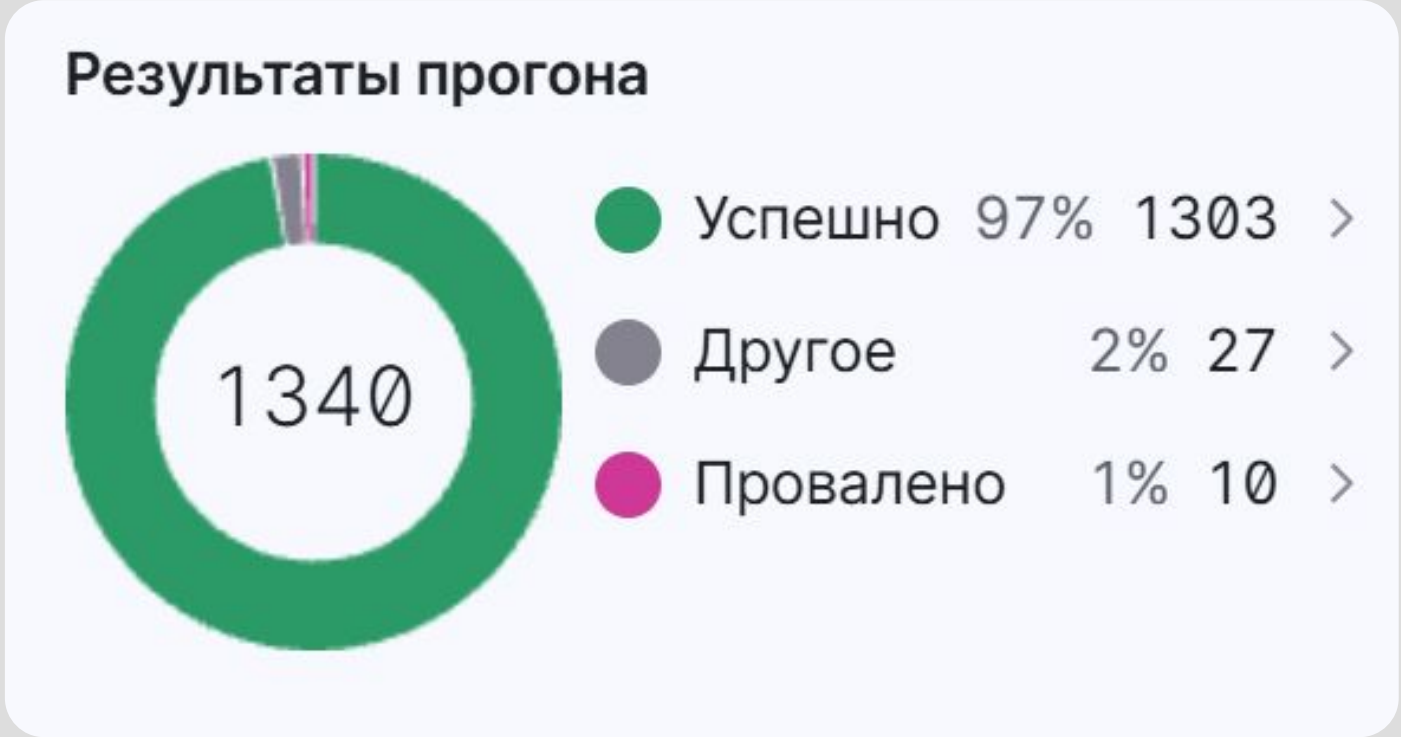
1 Активная разработка

2 Команды меняют смежные куски

3 В E2E /инфре что-то забыли обновить

4 Как итог — новые падения автотестов

Знакомая ситуация

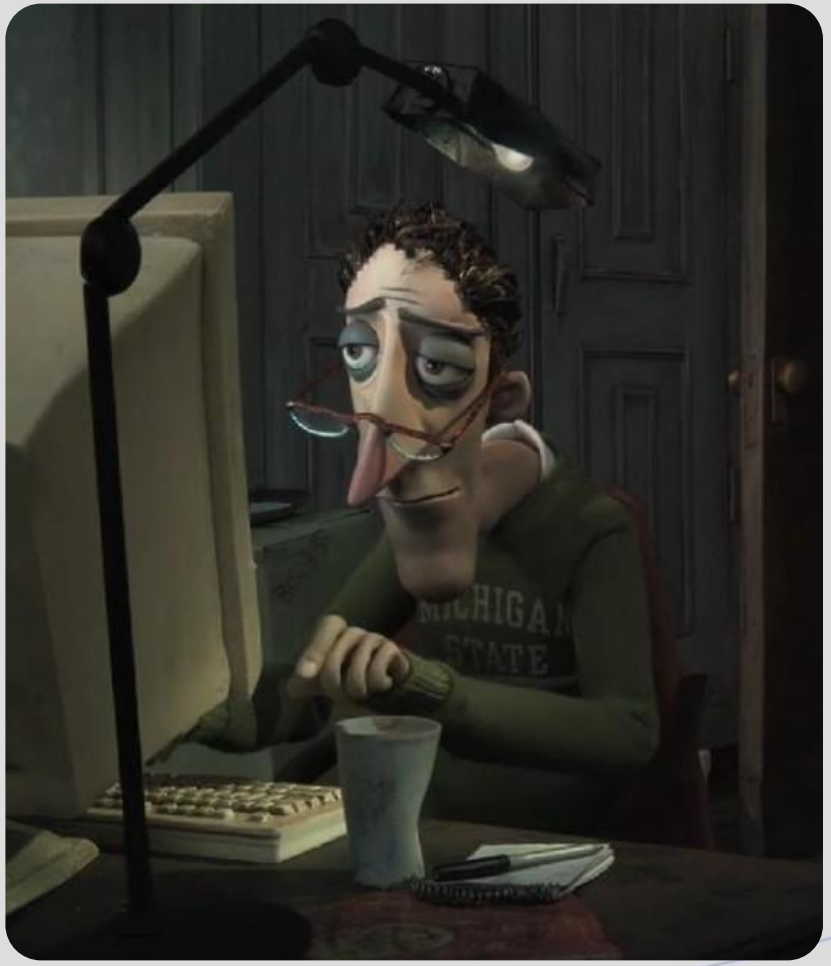


Сводка по прогону



Дежурный собирает картину руками

- Что в стеке, шагах, вложениях?
- Что менялось в коде?
- Похоже ли на вчерашние падения?
- Это инфра, продукт или тесты?



Откуда взялась идея?

Инженеры качества нашей команды





С чем мы работаем

- Адаптеры проливают результаты в прогон
- Артефакты тест-результата внутри TMS*
- Дальше все читаем по API

**Платформа управления тестированием*

Если у вас нет TMS*

Можно
начать
с JUnit XML



This XML file does not appear to have any style information associated with it. The document tree is shown below.


```
<testsuites analysisId="11" analysisDate="09-02-2024 17:20:28">
  <testsuite name="Sample TSQL" description="Path: C:\ProgramData\Novalys\VisualExpert\Sample\Sample TSQL, Mask: .sql,.veperf, Scan Subfolders: False" size="1801"
   reliability="28" security="2" maintainability="1225" comments="6.33 %" codeReviewProfile="Light" tests="79" failures="78">
    <testcase name="@@IDENTITY should not be used" coderuleId="VETSQLRULE26" activeIssues="1" codeReviewQualityLevel="D" severity="Major" type="Maintainability">
      <failure message="Active issues are more then desired limit."/>
    </testcase>
    <testcase name="A primary key should be specified during table creation" coderuleId="VETSQLRULE57" activeIssues="5" codeReviewQualityLevel="D" severity="Critical"
     type="Maintainability">
      <failure message="Active issues are more then desired limit."/>
    </testcase>
    <testcase name="All branches in a conditional structure should not have exactly the same implementation" coderuleId="VETSQLRULE14" activeIssues="1"
     codeReviewQualityLevel="D" severity="Major" type="Bug">
      <failure message="Active issues are more then desired limit."/>
    </testcase>
    <testcase name="ANSI_NULLS, ANSI_PADDING and CONCAT_NULL_YIELDS_NULL should not be configured" coderuleId="VETSQLRULE25" activeIssues="3" codeReviewQualityLevel="D"
     severity="Major" type="Maintainability">
      <failure message="Active issues are more then desired limit."/>
    </testcase>
    <testcase name="Boolean checks should not be inverted" coderuleId="VETSQLRULE51" activeIssues="2" codeReviewQualityLevel="D" severity="Minor" type="Maintainability">
      <failure message="Active issues are more then desired limit."/>
    </testcase>
    <testcase name="CASE expressions should end with ELSE clauses" coderuleId="VETSQLRULE62" activeIssues="9" codeReviewQualityLevel="D" severity="Critical"
     type="Maintainability">
      <failure message="Active issues are more then desired limit."/>
    </testcase>
    <testcase name="CASE expressions should not have too many WHEN clauses" coderuleId="VETSQLRULE37" activeIssues="1" codeReviewQualityLevel="D" severity="Major"
     type="Maintainability">
      <failure message="Active issues are more then desired limit."/>
    </testcase>
    <testcase name="CASE input expressions should be invariant" coderuleId="VETSQLRULE4" activeIssues="3" codeReviewQualityLevel="D" severity="Critical" type="Bug">
      <failure message="Active issues are more then desired limit."/>
    </testcase>
    <testcase name="CATCH clauses should do more than rethrow" coderuleId="VETSQLRULE50" activeIssues="1" codeReviewQualityLevel="D" severity="Minor"
     type="Maintainability">
      <failure message="Active issues are more then desired limit."/>
    </testcase>
    <testcase name="COALESCE, IIF, and CASE input expressions should not contain subqueries" coderuleId="VETSQLRULE27" activeIssues="3" codeReviewQualityLevel="D"
     severity="Major" type="Maintainability">
```

С чего мы начали

- Собрали все по падению в прогоне
- Шаги, вложения, связанные мануальные тесты
- Отправили в модель и ждали разбор

Проблема 1

Слишком много всего

- Объем большой и шумный
- Результат ненадежный — полная зависимость на модель

Что решили делать?

01 → Выделить главное из контекста

02 → Классифицировать падение

03 → Точечно применить модель

Проблема 2

Шумный стектрейс



```
2  Assert.Fail failed. Condition failed: Attribute 'Global attribute_1ams7esqe0' value not equals to expected → Expected: '"'.
3  Actual: '"1a"'
4  at Microsoft.VisualStudio.TestTools.UnitTesting.Assert.Fail(String message) in /_/src/TestFramework/TestFramework/Assertions/Assert.Fail.cs:line 25
5    at E2E.Common.Utility.AssertTimeout.AssertCondition(Func`1 check, Func`1 msg, Nullable`1 pollingTimeoutSeconds) in /builds/testit/e2e_tests/testit/E2E.Common/Utility/AssertTimeout.cs:line 304
6    at E2E.Common.Utility.AssertTimeout.AreEqual[T](Func`1 check, Func`1 expected, String msg, Nullable`1 pollingTimeout)
7    at E2E.Common.Utility.AssertTimeout.AreEqual[T](Func`1 check, T expected, String msg, Nullable`1 pollingTimeout)
8    at E2E.Common.TestIT.Asserts.WorkItemCardAsserts.AssertAttributeValue(CreateWorkItemModalPage workItemModalPage, String attributeName, String expectedValue)
9    at TestIT.Tests.Tests.Attributes.ProjectCustomAttributeTests.ClearingValuesOfProjectAttributes() in /builds/testit/e2e_tests/testit/TestIT.Tests/Tests/Attributes/ProjectCustomAttributeTests.cs:line 1382
10   at Microsoft.VisualStudio.TestPlatform.MSTest.TestAdapter.Execution.TestMethodInfo.ExecuteInternalAsync(Object[] arguments, CancellationTokenSource timeoutTokenSource) in /_/src/Adapter/MSTestAdapter.PlatformServices/Execution/TestMethodInfo.cs:line 408
```

Что оставляем из стека

- Тип ошибки и сообщение
- Первый «свой» кадр
- ! Кадры фреймворков и библиотек — пропускаем

Как выкидываем чужой шум

```
static readonly string[] FrameDenyPrefixes =
[
    "at System.",
    "at Xunit.",
    "at NUnit.Framework.",
    "at FluentAssertions.",
    "at org.junit.assert.",
    "at printjob.browser.ChromiumBrowser",
    "at Microsoft.VisualStudio.TestPlatform.MSTest"
];
```

Типы, которые ищем правилами

Инфра — Timeout, сеть, 5xx коды ошибок

Продукт — Валидация, прикладной код

Тесты — Waits, flaky, внутрянка фреймворка

Unknown — В иных случаях

Правило — это вес и категория

```
new(Infrastructure, "infra_timeout",
    KeywordRegex("timeout", "timed out", "deadline exceeded"),
    weight: 0.34),

new(Unknown, "assert_mismatch_unresolved",
    KeywordRegex("assertion failed", "expected:", "but was:"),
    weight: 0.46),
```

Проблема 3

Недостаток контекста — модели нужен код

Надо понять, какой автотест упал
и какие части системы он затрагивает

Путь из стека → GitLab

- Достаем пути из стека
- Ищем близкие коммиты по зоне
- Выделяем нужные данные, а не полные файлы

Как бы хотели улучшить?

Сейчас

Скрипты и GitLab API



В планах

Что-то вроде GitNexus: поиск по графу

Как это выглядит в разборе

Evidence
• ClassificationSignal: 59% — Deterministic classification: Product (rules: assert_fail).
• ModelCited: 32% — Ошибка `Assert.Fail` в `WorkItemCardAsserts.AssertAttributeValue`: ожидаемое '', фактическое '1a'.
• ModelCited: 32% — Скриншот (`ClearingValuesOfProjectAttributes-20-05-2026-03-23-16-883.png`) подтверждает, что 'Global attribute_zf7kx7u4...' имеет значение '1a'.
• ModelCited: 32% — Коммит `gitc-7e47d62d2fb8` вносит изменения в сущности проектов и кастомные атрибуты (TMS-38733).

Как это выглядит в разборе

Evidence
• ClassificationSignal: 59% — Deterministic classification: Product (rules: assert_fail).
• ModelCited: 32% — Ошибка `Assert.Fail` в `WorkItemCardAsserts.AssertAttributeValue`: ожидаемое '', фактическое '1a'.
• ModelCited: 32% — Скриншот (`ClearingValuesOfProjectAttributes-20-05-2026-03-23-16-883.png`) подтверждает, что 'Global attribute_zf7kx7u4...' имеет значение '1a'.
• ModelCited: 32% — Коммит `gitc-7e47d62d2fb8` вносит изменения в сущности проектов и кастомные атрибуты (TMS-38733).

Что хотим дальше

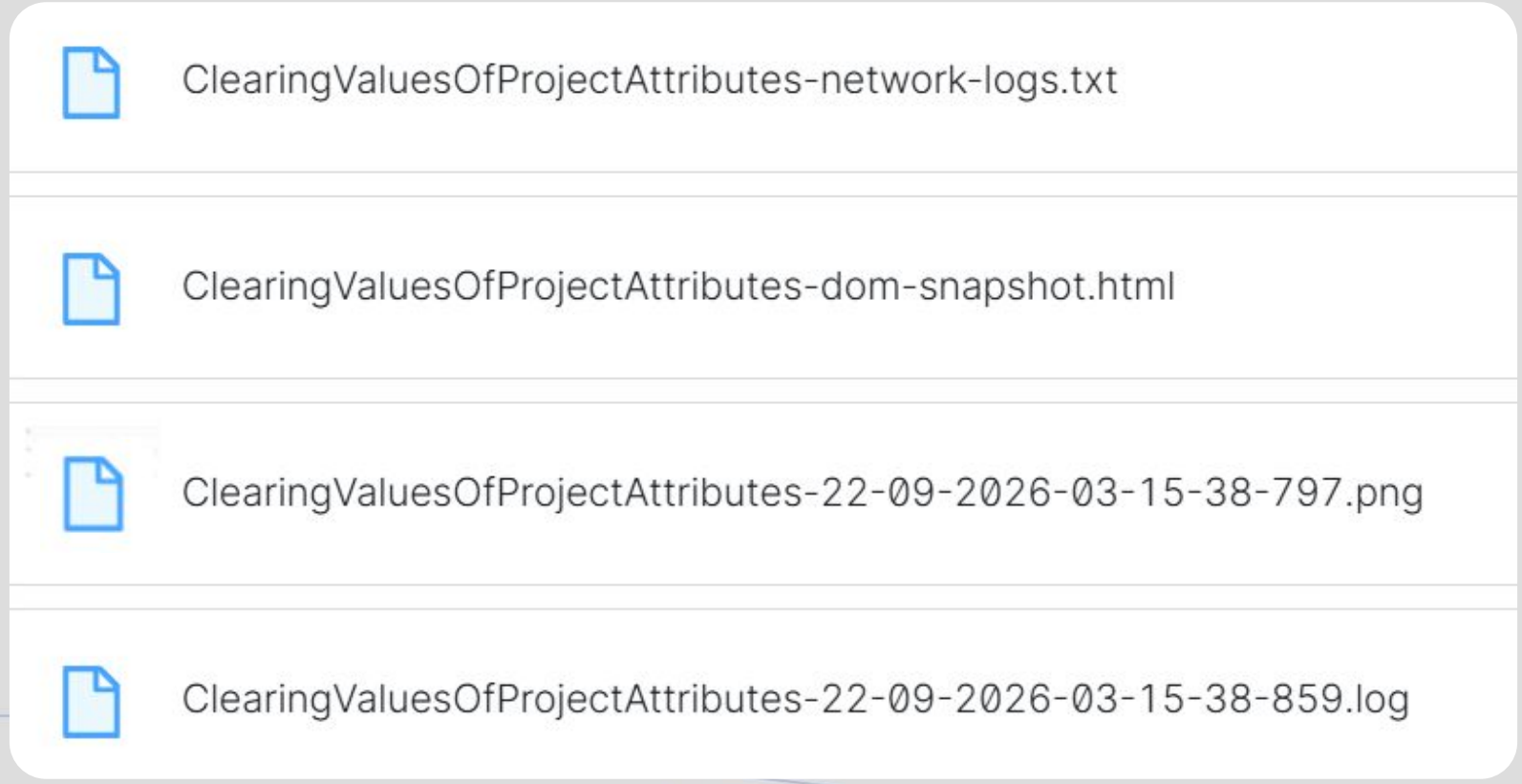


Хранить разобранные падения и классификации в БД.

Пока в проработке

Проблема 4

Большие вложения занимают контекст



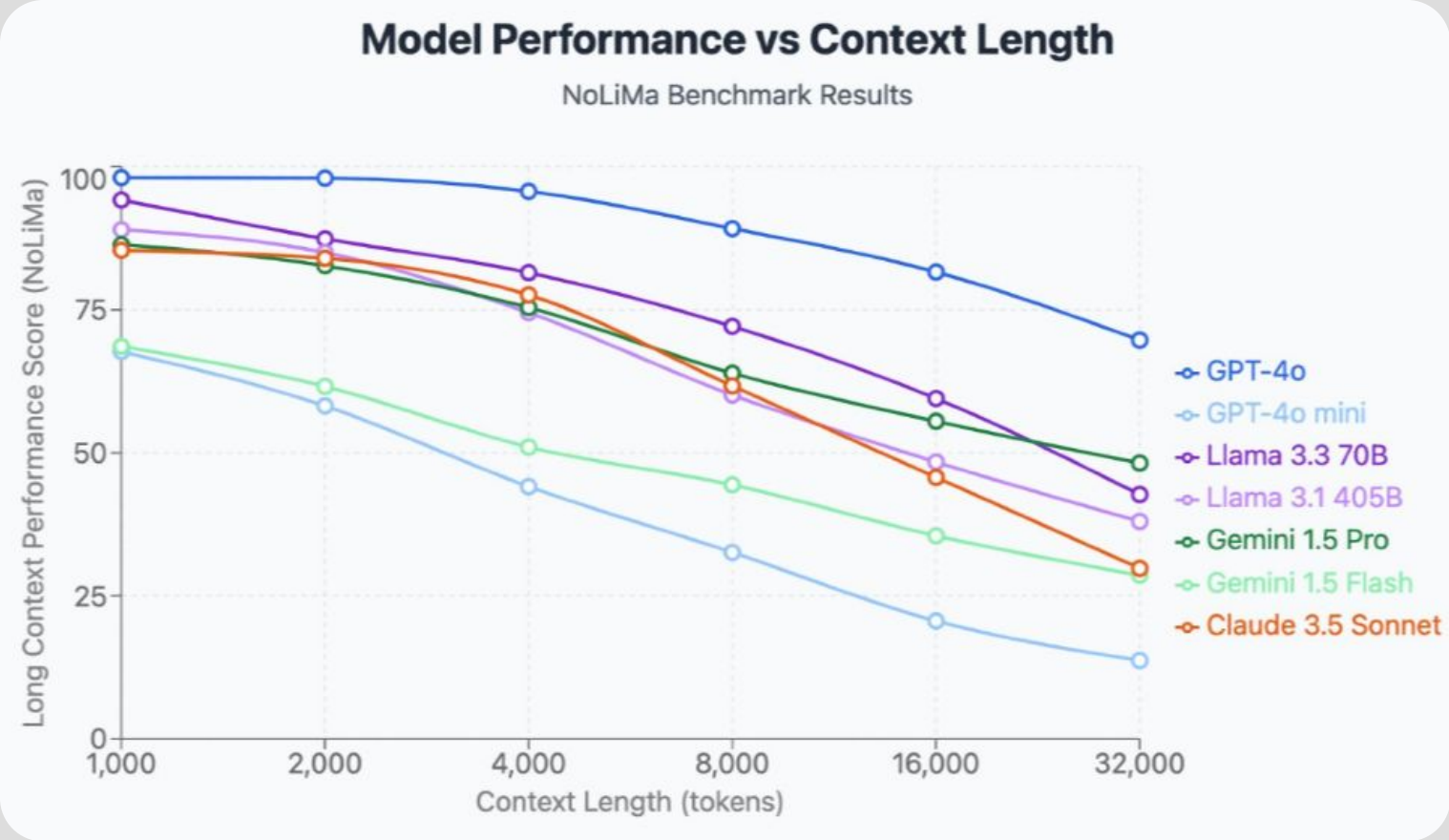
Решение: суммаризация вложения

```
payloads = SplitAttachments(extracted, maxChars)

foreach (payload in payloads)
{
    // отдельный структурированный вызов:
    // вытащи суть, которая относится к падению
    // результат : summary по каждому вложению
}
```

Проблема 5

Иногда контекст даже после саммари файлов великоват



Источник: NoLiMa: Long-Context Evaluation Beyond Literal Matching

Решение

Разбиваем большие контексты на чанки

```
// если контекст больше 256к символов
const maximumChunkCharacters = 80_000;
payloadChunks = SplitWithoutLoss(...)
// разбиваем, группируя по размерам куски контекста
// после обработки каждого контекста - склеиваем факты о них
```

Итог: Два этапа работы с моделью

1. Гипотеза

Классификация, стек, git, сжатые вложения

Формирование гипотезы

2. Рекомендации

Гипотеза + результаты классификатора

Формирование рекомендаций

Промежуточные итоги

Выдохнем и выделим главное

- Сначала - сбор инфы и работа с моделью
- Далее - итеративные улучшения по контексту



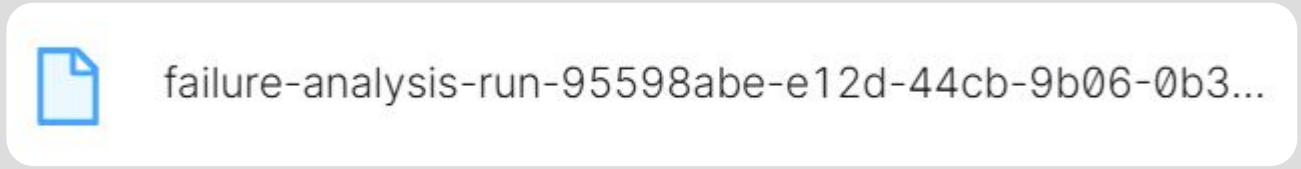
Рекомендации в отчете

Group suggested actions
• Проанализировать изменения в коммите `gitc-7e47d62d2fb8` (feat(custom-attributes): TMS-38733), так как он затрагивает функционал кастомных атрибутов.
• Проверить шаги теста, отвечающие за очистку значений атрибутов, на корректность выполнения и наличие ожиданий (waits).
• Перезапустить тест для исключения влияния состояния окружения (dirty state).

И обоснование причины

Probable cause (engineering signal)
Регрессия в бэкенде при очистке значений атрибутов. Недавний коммит `gitc-7e47d62d2fb8` (feat(custom-attributes): TMS-38733) мог изменить логику обработки сущностей или значений по умолчанию, из-за чего атрибут не очищается или сохраняет старое значение ('1a'). Тест ожидает пустую строку, но получает заполненное значение, что видно на скриншоте и в логе ошибки.

Куда кладем результат



Общий



Индивидуальный

Чего стоит опасаться

– галлюцинации и недостаток внимания инженера

Обоснование уверенности	Доказательства
Изменение поведения API эндпоинта настроек автотестов. Сервис начал возвращать статус 400 (BadR... | retrieval ~0% → aggregate 5%.	ClassificationSignal: Deterministic classification: Infrastructure (rules: infra_timeout). --- FlakySignal: Flaky heuristics score 0.28; signals: timing_or_race_language --- ModelCited: Assert.AreEqual failed. Expected:<PaymentRequired>. Actual:<BadRequest>. at E2E.Common.TestIT.Asserts.ApiAsserts.AssertStatusCodeOfAutotestsSettings ---

Сводка по прогону

	T	U	V	
	Обоснование уверенности	**Доказательства**	**Рекомендуемые действия**	**Под**
7	Ошибка вызвана таймаутом или недоступностью зависимостей. | retrieval ~0% → aggregate 7%.	ClassificationSignal: Deterministic classification: Infrastructure (rules: infra_timeout, infra_auth). --- FlakySignal: Flaky heuristics score 0.00; signals:	Проверьте логи и состояние зависимостей, указанных в стэктрейсе. Убедитесь, что тестовый стенд и окружение доступны и работают корректно. Проверьте настройки таймаутов и задержек в тесте и инфраструктуре.	
6	Проблема с таймаутом или недоступностью зависимостей. | retrieval ~0% → aggregate 6%.	ClassificationSignal: Deterministic classification: Infrastructure (rules: infra_timeout, infra_auth). --- FlakySignal: Flaky heuristics score 0.00; signals:	Проверьте журнал и конфигурацию зависимостей, чтобы убедиться в их доступности и корректности. Проверьте настройки таймаутов и убедитесь, что они соответствуют ожидаемому поведению приложения. Выполните ручное тестирование, чтобы подтвердить, что проблема не связана с тестовым окружением.	
0	Ошибка вызвана проблемами с инфраструктурой или зависимостями. | retrieval ~0% → aggregate 10%.	ClassificationSignal: Deterministic classification: Infrastructure (rules: infra_timeout, infra_auth). --- FlakySignal: Flaky heuristics score 0.00; signals:	Проверьте логи и журналы для получения дополнительной информации о причине ошибки. Убедитесь, что тестовое окружение настроено правильно и соответствует требованиям. Проверьте зависимости и их версии, чтобы исключить конфликты или несовместимости. Проверьте наличие известных дефектов или проблем в продукте, связанных с данной ошибкой.	

Что если ИИ рекомендует опасное действие?

Пример: почистить сущности в БД, что-то подправить наживую

Фильтр «опасных» рекомендаций

+ логирование на будущее

```
if (action.Safety ==
RcaRecommendedActionSafety.RejectedDestructive)
{
    warnings.Add(...)
    rejectedUnsafeAction = true;
    continue;
}
```

Маленькая проблемка

При повторных запусках берем в расчет свои же артефакты



Решение

Игнорируем свои же артефакты

```
IgnoredAttachmentNameMasks =
[
    "failure-analysis*.md",
    "failure-analysis*.xlsx",
];

if (IsMatch(attachment.Name, masks))
    skip;
```

Проблема 7

Как тестировать агентский сервис, промпты?

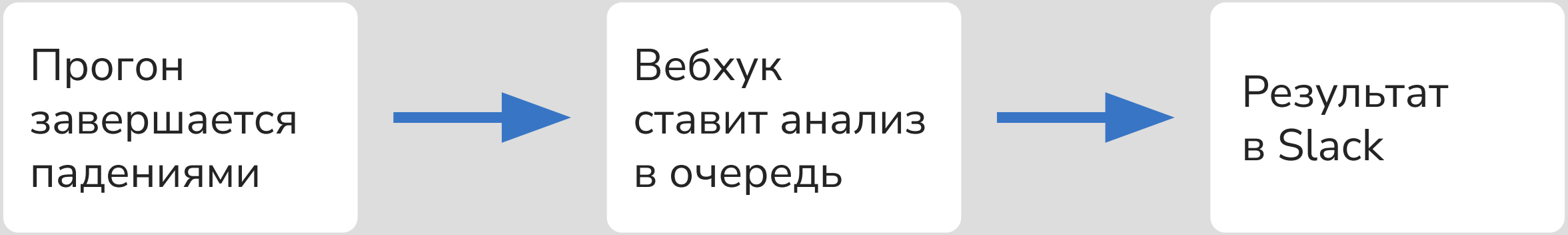


Решение

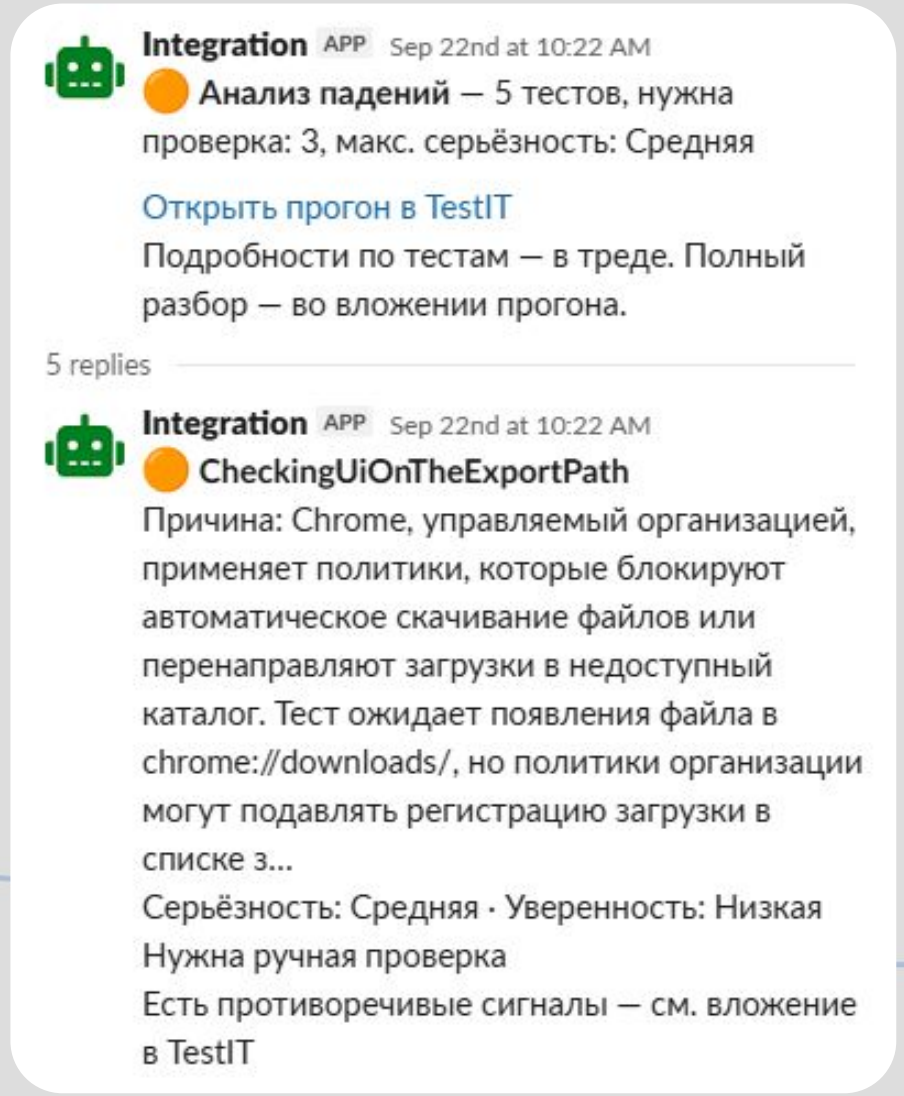
Готовим тест-кейсы под анализ и матчим результаты

```
// evaluation/rca/v1/cases/regression_assert_001.json
{
  "fixture": "fixtures/regression_assert_001.signal.json",
  "expectations": {
    "expectedPrimaryClassification": "ProductRegression",
    "expectedRootCauseKeywords": ["assert"]
  }
}
```

Немного автоматизации



Интеграция в рабочие процессы



На чем экономит дежурный?

- Сбор информации
- Чтение логов
- Поиск места в коде автотестов и продукта

По словам коллег

Было

до 15 минут
на одно падение

Стало

~1-3 минуты
на решение по карточке

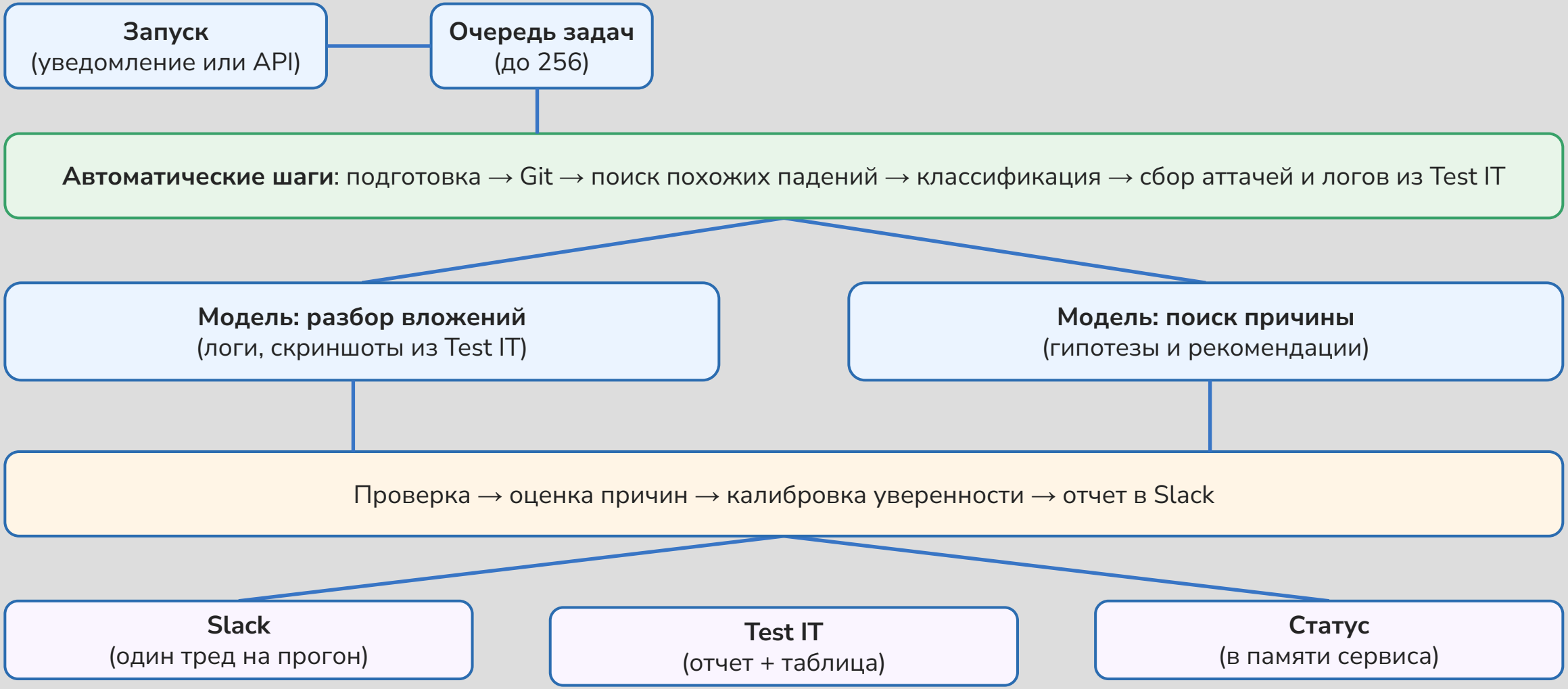
Дежурный решает, как поступить

- Завести задачу на доработку?
- Перезапустить тест-ран?
- Эскалировать проблему разработчикам?

С чего начать у себя

- MVP: собрать контекст падения, вызов модели
- Нормализация стека и классификатор
- Интеграции, вложения, тестирование сервиса, уведомления
- Продолжить работу над итеративными улучшениями

Итоговая архитектура решения



Доп фича:
сравнение двух прогонов

- Запускается вручную
- Сравниваем строки данных тестов
- Распределяем по корзинам: новые / упавшие / исправленные

Запуск


Первый тест-ран ID
84589d3f-ad88-4824-9bfd-fcc41b64ff20
Второй тест-ран ID
f83f01cc-9d31-4971-8681-78c7cb033ca6
Сравнить

Результаты сравнения

Fixed Tests (1)

AutoFillOfCompleteDateColumnNotErasesValueWhenStoppingTestRun (30786)

- **TestResultId:** b35887a4-179b-4fcf-b9dd-2f7f4a0c9eeb
- **TestName:** AutoFillOfCompleteDateColumnNotErasesValueWhenStoppingTestRun
- **ExternalId:** 30786
- Links:

Failed Tests (9)

EditCheckList (1559)

- **TestResultId:** 7ac40a98-97a5-469c-9db9-507fe8bbfc2a
- **TestName:** EditCheckList
- **ExternalId:** 1559
- Links:
- **Message:** Tested on 'https://e2e.auto03.cloud.testit.ru/' Assert.Fail failed. Condition failed: Steps not equals to expected -> Expected: '[{ "Id": "1", "Action": "Changed", "Expected": "", "TestData": "", "Comments": "" }, { "Id": "2", "Action": " Auto Action ", "Expected": "", "TestData": "", "Comments": "" }, { "Id": "3", "Action":

Проблема 8

Параллельные запуски могут приводить к ошибкам обработки

Лимиты модели, API, ресурсы

Решение

Постановка в очередь обработки + рестарты для надежности

```
POST /api/internal/analyze-run-comparison
{ newerRunId, olderRunId, maxEnqueueCount }

outcome = comparer.GetComparison(newer, older)
plan = enqueuePlanner.BuildPlan(...)

foreach (candidate in plan.Candidates)
    queue.WriteAsync(workItem)
```

Ограничения сейчас

- Каждое падение разбираем отдельно, без группировки
- Не анализируем скриншоты
- Нет обратной связи «верно / неверно» от инженера
- Не читаем требования и контекст задач из трекера
- Коммиты — только кратко, без полного кода

Test IT FAI
× Девелоника

@TAIPAN_W

Спасибо за внимание!

testit.software

t.me/testit_tms

vk.com/testittms