

Ускорение и оптимизация A/B-тестов

Ухудшающие A/B-тесты и воздействие эффектов

Цель урока

Мы уже говорили про A/B-тестирование, где всячески пытаемся улучшить метрики. И даже говорили про A/A-тестирование, где мы вообще пытаемся ничего не изменить. **А как насчёт того, чтобы испортить метрики?**

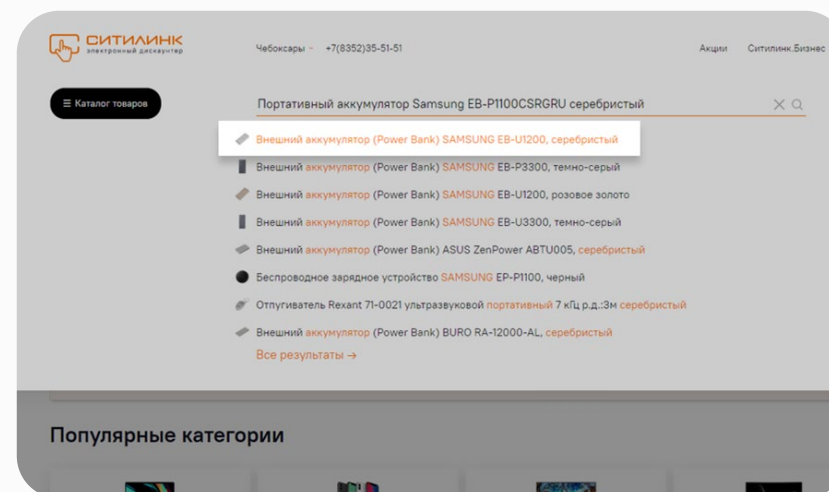
Задачи урока

- ✓ Поговорим про ухудшающие тесты.
Посмотрим на примеры
- ✓ Разберём аргументы за ухудшающие тесты и против них
- ✓ Затронем тему количества времени, расходуемого на каждый этап цикла тестирования
- ✓ Поймём, что такое эффект взаимодействия

Разберём кейс: плохой поиск

Проблема: мы получаем много жалоб на качество поиска нашего интернет-магазина.

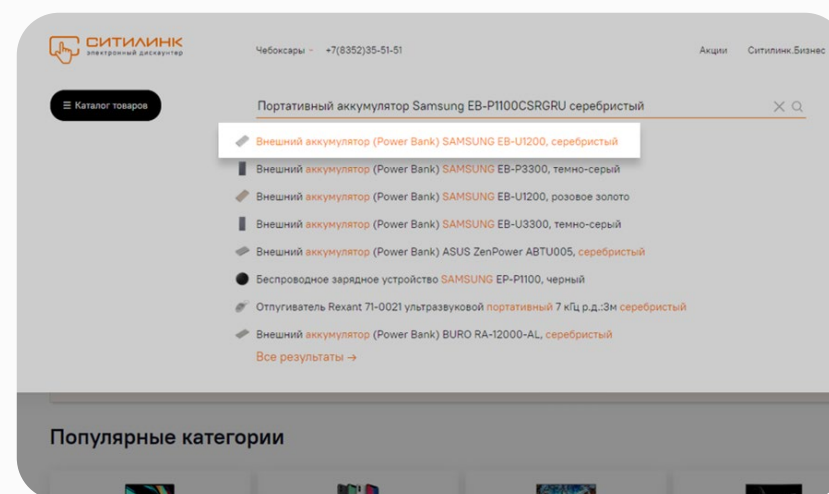
Что можно тут предложить,
чтобы исправить?



Разберём кейс: плохой поиск

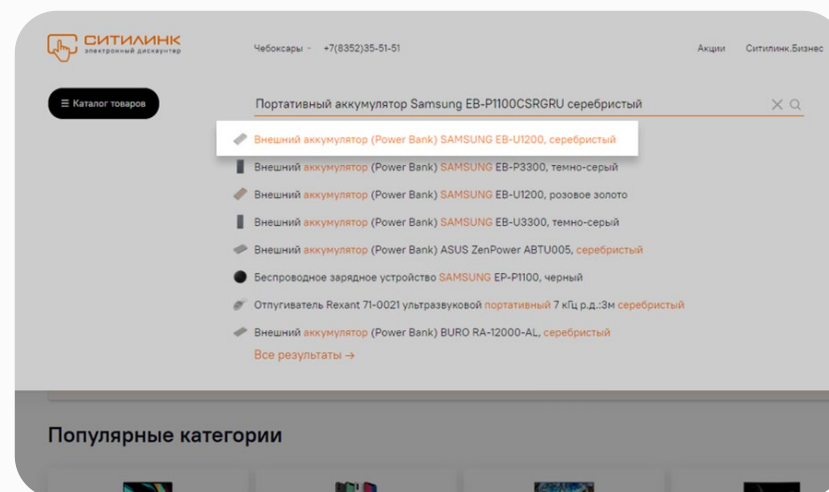
Проблема: мы получаем много жалоб на качество поиска нашего интернет-магазина.

Предложение: если выделить команду, которая будет на постоянной основе заниматься его развитием, то качество сервиса улучшится.



Разберём кейс: плохой поиск

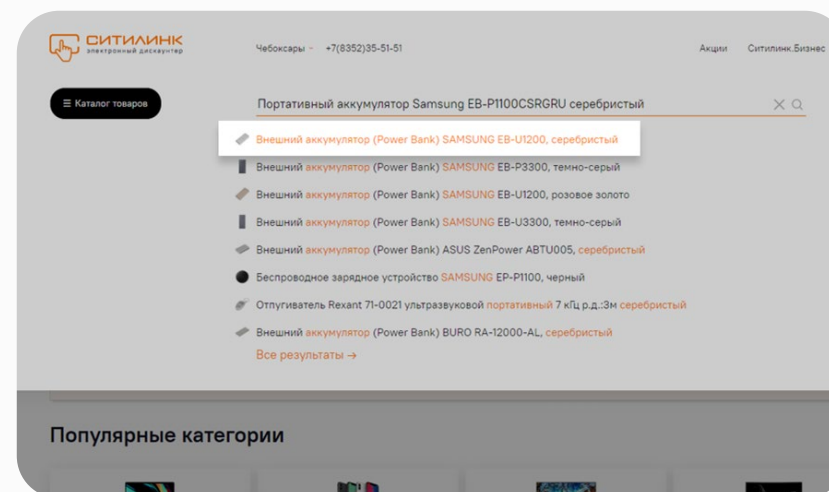
Способ проверки: убрать поиск и проверить, каково его влияние на ключевые метрики. Оценить, насколько команда сможет его улучшить, какой выигрыш это принесёт, как он соотносится со стоимостью проекта и с тем, что могут дать другие инициативы.



Разберём кейс: плохой поиск

Эксперименты — единственный способ проверить гипотезу.

Ухудшающие тесты — часто единственный способ сделать это дёшево. Именно поэтому они так хорошо работают, когда дело доходит до приоритизации проектов.



Ухудшающие A/B-тесты: за и против

За:

- быстро и дёшево
- если гипотеза верная — «ущерб от поломки» может с лихвой компенсироваться
- эффект может быть не замечен

Против:

- возможный дискомфорт пользователей
- эффект может быть не замечен

Ухудшающие A/B-тесты: за и против

Аргументы **против** обычно сводятся к тому, что такие тесты портят опыт пользователей, увеличивают отток, приводят к негативному фидбеку, плохо влияют на бренд.

Из опыта — опасения сильно преувеличены. Вы удивитесь, как часто ухудшающие тесты не окажут никакого значимого влияния на продукт (и это очень важное знание).

Ухудшающие A/B-тесты: за и против

Если вы всё-таки очень сильно дорожите вашими пользователями или брендом компании, а гипотез и времени на тестирование у вас достаточно, то можно вовсе не обращаться к ухудшающим тестам.

Разберём кейс: запись голоса

В компании Dialogflow занимались развитием умного голосового ассистента.

Изначальная стратегия сводилась к тому, чтобы выявить наиболее востребованные способности ассистента и вынести их на первый план. Подобный подход принёс ряд небольших побед и имел конечный потенциал. Вдруг в один день ключевые продуктовые метрики обвалились на 20 процентов.

Разберём кейс: запись голоса

Оказалось, что Google стал экспериментировать с разным временем паузы, на основе которой сервис распознавания голоса определял, что человек закончил говорить. Спад всех метрик произошёл тогда, когда они резко сократили время слушания, в результате чего пользователи стали часто не успевать закончить фразу целиком.

Снижение качества одной из технологий, на которой строилось общение между пользователем и ассистентом, привело к большему изменению метрик, чем какое-либо из сделанных нами изменений ранее.

Разберём кейс: запись голоса

Модель продукта после «ухудшающего эксперимента» кардинально изменилась. Если до в фокусе было то, что ассистент умеет, то после стало очевидно, что намного более мощным рычагом воздействия является качество ключевых технологий.

Ухудшающие тесты: кейсы

Рождественское плацебо



Могут ли слабые
по отдельности гипотезы
дать сильный эффект
вместе? Эффекты
взаимодействия!



Что такое эффекты взаимодействия?

Рассмотрим пример двух A/B-тестов, проведённых на кнопке Shop Now. В одном опыте тестировали текст кнопки, в другом — её цвет. Пользователи, участвующие в одном тесте, уже не могут участвовать в другом.

	Контрольный вариант	Тестируемый вариант
Тест текста	Shop Now	Discover (+2 %)
Тест цвета кнопки	Shop Now	Shop Now (+5 %)

Что такое эффекты взаимодействия?

После эксперимента исследователи обнаружили, что пользователи, которым демонстрировался вариант кнопки с текстом “Discover” (Узнать больше), нажимали на неё на 2 % чаще, а те, кому приходилось выбирать между кнопкой белого или красного цвета, на 5 % чаще предпочитали последний.

	Контрольный вариант	Тестируемый вариант
Тест текста	<div>Shop Now</div>	<div>Discover (+2 %)</div>
Тест цвета кнопки	<div>Shop Now</div>	<div>Shop Now (+5 %)</div>

Что такое эффекты взаимодействия?

Вы можете предположить, что если сделать кнопку, в которой были бы учтены особенности обоих выигрышных вариантов, можно получить улучшение производительности на 7 %. Поскольку вы привыкли всё проверять (не так ли?), вы проводите ещё один тест, в котором тестируете такой вариант кнопки:

	Контрольный вариант	Тестируемый вариант
Тест текста	Shop Now	Discover (+2 %)
Тест цвета кнопки	Shop Now	Shop Now (+5 %)

Что такое эффекты взаимодействия?

К своему удивлению вы обнаруживаете, что посетители, которым была предложена комбинация выигрышных вариантов, нажимали на эту кнопку чаще не на 7 %, а на 10 %. **Откуда взялись эти «лишние» 3 %?** Это и есть результат эффекта взаимодействия, при котором одновременное влияние двух выигрышных вариантов эффективнее, чем воздействие этих же вариантов, но по отдельности — **этим надо пользоваться!**

Контрольный вариант

Тестируемый вариант

Комбо-тест

Shop Now

Discover (+10 %)

Что такое эффекты взаимодействия?

Если у вас есть обоснованные соображения, что очевидные улучшения приведут к более высокой эффективности, то создайте из них комбинацию. Это может привести к большему эффекту и меньшим времязатратам на тестирование!

И, как всегда, будьте осторожнее, такой подход может иногда дать обратный, понижающий эффект!

На практике это происходит нечасто.

Контрольный вариант

Тестируемый вариант

Комбо-тест

Shop Now

Discover (+10 %)

ИТОГИ И ВЫВОДЫ УРОКА

Ухудшающие тесты — это осознанное ухудшение сервиса. Использовать или нет ухудшающие A/B-тесты — это исключительно ваш выбор. Однако, исчерпав все идеи по улучшению, в эту сторону стоит посмотреть, ведь в таком случае есть шансы найти новые рычаги улучшения.

Итоги и выводы урока

- ✓ Есть риски негативного влияния
- ✓ Успех может с лихвой покрыть все недостатки
- ✓ Ухудшающие тесты позволяют сделать быструю и часто однозначную проверку гипотезы
- ✓ Эффекты взаимодействия могут помочь получить максимальный эффект за меньшее количество тестов