

Лекция № 12. Юзабилити-тестирование.

Под юзабилити-тестированием понимают любой эксперимент, направленный на измерение качества пользовательского интерфейса или поиск конкретных проблем в нем.

Качество интерфейса, в соответствии с определением Международной организации стандартизации (ISO) определяется, в первую очередь, удобством применения программного продукта. Таким образом, можно определить юзабилити-тестирование как тестирование интерфейса на удобство применения.

Удобство применения определяется тремя факторами:

- эффективность работы пользователя;
- продуктивность;
- удовлетворенность.

Т.е. тестирование интерфейса на удобство применения проводится для того, чтобы оценить насколько он эффективен, рентабелен и довольны ли им пользователи.

Юзабилити-тестирование позволяет:

- понять, насколько хорошо или плохо работает интерфейс;
- сравнить качество старого и нового интерфейса;
- найти проблемные фрагменты интерфейса.

Для организации и проведения юзабилити-тестирования необходимо выполнить следующие действия.

1. Определить цели и задачи тестирования.
2. Выделить значимые метрики интерфейса.
3. Выбрать метод (группу методов) для проведения тестирования, разработать методику проведения тестирования.
4. Разработать тестовые сценарии.
5. Определить значимые характеристики целевой аудитории и подобрать респондентов.
6. Провести необходимые подготовительные мероприятия.
7. Провести тестирование.
8. Проанализировать результаты.
9. Составить отчет.

Рассмотрим подробнее эти этапы.

1. Цели и задачи тестирования.

Прежде чем планировать и проводить тестирование на удобство применения продукта, следует четко определить цели и задачи, стоящие перед ним. Эта задача лежит на владельцах продукта, специалистах, составляющих планы, проектировщиках и разработчиках.

Бут выявил четыре фактора, составляющих удобство применения: полезность (продуктивность), эффективность, простота изучения и отношение пользователей. Все операционные определения удобства должны затрагивать один или более этих факторов.

Факторы, определяющие удобство применения.

Полезность – степень, до которой продукт позволяет пользователю достичь стоящей перед ним цели (Могут ли пользователи использовать продукт?). Оценка мотивации пользователя в использовании данного продукта. Оценка полезности, как правило, осуществляется с помощью оценки качества выполненной работы.

Эффективность – насколько успешно продукту удается содействовать пользователю в выполнении стоящей перед ним задачи (насколько хорошо пользователь может выполнить стоящую перед ним задачу с помощью данного продукта?). Оценка эффективности, как правило, осуществляется с помощью оценки качества выполненной работы.

Простота изучения. Пользователи могут начать использовать продукт, приобретая определенный уровень знаний после прохождения тренинга (Насколько хорошо обучены пользователи?). Оценка простоты изучения, как правило, осуществляется с помощью оценки качества выполненной работы.

Отношение пользователей – восприятие, ощущения и мнения пользователей по поводу изучения и использования данного продукта (Что пользователи думают об удобстве применения продукта?). Оценка, как правило, осуществляется по отзывам пользователей (устной или письменной обратной связи).

Цели и задачи, стоящие перед разработкой удобства применения, должны определяться для всех программных продуктов. *Цели* – это обеспечение преимуществ продукта перед конкурирующими в области простоты изучения, эффективности, гибкости и т.д., т.е. напрямую связаны с перечисленными факторами.

Цели сами по себе не подлежат непосредственной оценке – они должны делиться на задачи.

Задачи – уточнение целей, они более конкретны и детальны, их можно оценить и измерить. Достижение одной цели может потребовать решения множества задач. Задачи

должны быть выстроены таким образом, чтобы содержать информацию по конкретным действиям или операциям.

Пример характеристики задач.

Задача удобства применения	Критерий	Качество работы	Условия
После 4-часового тренинга 90% пользователей в состоянии выполнить заказ клиента в течение 5 мин.	90% пользователей; в течение 5 мин.	Выполнить заказ клиента	После 4-часового тренинга
<i>Полезность:</i> После выполнения пяти сценариев задач 90% пользователей будут в состоянии успешно выполнить задачу	90% пользователей; выполнить задачу	Успешно выполнить задачу	После выполнения пяти сценариев задач
<i>Эффективность:</i> После выполнения пяти сценариев задач 75% пользователей будут в состоянии успешно выполнить задачу в течение 10 мин.	75% пользователей; в течение 10 мин.	Успешно выполнить задачу	После выполнения пяти сценариев задач
<i>Простота изучения:</i> После 4-часового тренинга все пользователи достигнут определенного уровня владения продуктом	Все пользователи; определенный уровень владения продуктом	Успешно овладеть продуктом	После 4-часового тренинга
<i>Отношение пользователей:</i> После выполнения пяти сценариев задач 85% пользователей оценят степень своей удовлетворенности продуктом на 5,5 и выше баллов (по 7-балльной системе)	85% пользователей; степень удовлетворенности 5,5 баллов по 7-балльной системе	Степень удовлетворенности	После выполнения пяти сценариев задач

2. Значимые метрики.

Как правило, тестирование на удобство применения включает два типа методов оценки: количественные и качественные.

Методы оценки работы, подразумевающие подсчет действий, определение полноты выполнения задачи, подсчет времени, затраченного на это выполнение, ошибок и обращений за помощью. Такие методы называют *количественными*.

Субъективные методы, включающие сбор устных и письменных сообщений пользователей об их восприятии, мнениях, суждениях, предпочтениях, а также степени удовлетворенности от системы и их собственной выполненной работы. Эти методы носят название *качественных*.

В зависимости от того, какие задачи стоят в ходе тестирования, выбираются метрики, т.е. способы оценки и соответствующие им критерии. Их можно получить уже из формулировок задач (см. таблицу).

Эффективность работы пользователя определяется количественными показателями:

- скорость работы пользователя (метрики: длительность выполнения операций, время на обнаружение ошибок, время на исправление ошибок, количество команд на используемую операцию, длительность поиска сведений в документации, снижение производительности при длительной работе и т.д.);
- скорость обучения (метрики: количество и частота обращений к справочной системе, длительность периода между началом работы в системе и точкой, когда скорость работы перестает расти, разница в количестве ошибок/скорости работы у опытных и неопытных пользователей и т.д.);
- количество человеческих ошибок (метрики: процент операций, вызвавших ошибку, среднее число ошибок на операцию у опытных пользователей, количество не обнаруженных и неисправленных пользовательских ошибок и пр.).

Продуктивность характеризуется такими параметрами как:

- успешность (метрики: отношение выполненных тестовых заданий к невыполненным или выполненным полностью неправильно);
- мощность интерфейса (метрики: соотношение задач пользователя и задач, решаемых программным продуктом);
- нагрузка на пользователя (ментальная нагрузка – метрики: объем клавиатурного ввода, GOMS).

Удовлетворенность:

- оценивается в ходе бесед, опросов, анкетирования.

Чем больше пользователей зависят от неудобного в применении проекта, тем сложнее его изменить. Помните о феномене WYKIWYL – что вам знакомо, то вам и нравится. Пользователи могут адаптироваться к плохо «скроенным» проектам, хотя они не должны этого делать.

Помимо указанных основных анализируемых показателей, бывает полезно оценить сохраняемость навыков работы с системой:

- сохранение навыков работы с системой (метрики: разница в скорости работы/количестве ошибок у пользователя после часа работы с системой и у того же пользователя в начале использования после длительного перерыва).

3. Методы юзабилити-тестирования.

Существует три основных метода юзабилити-тестирования: пассивное наблюдение, поток сознания и активное вмешательство.

Метод пассивного наблюдения

Респондент выполняет тестовые задания без вмешательства со стороны тестировщика, его действия анализируются (во время теста или после, по протоколам), что позволяет как найти проблематичные фрагменты, так и замерить эргономические характеристики интерфейса.

Метод «поток сознания»

Организационно аналогичен методу пассивного наблюдения, но респондента при этом просят также устно комментировать свои действия. Затем комментарии анализируются. Главный недостаток метода – измерение количественных характеристик интерфейса не всегда возможно.

Метод активного вмешательства

В этом методе юзабилити-специалист после каждого действия респондента спрашивает его, почему респондент действует именно так; на каждом экране экспериментатор спрашивает, как именно респондент понимает назначение и функции этого экрана. Этот метод ближе к фокусированному интервью, чем к собственно тестированию. Понятно, что при активном вмешательстве никакие измерения попросту невозможны, но зато объем получаемых качественных данных наиболее велик.

Помимо описанных собственных методов в юзабилити-тестировании применяются также методы общенаучные, такие как:

- наблюдение;
- проведение опросов и тестов;
- контекстуальные опросы;
- эвристические оценки;
- работа с выделенными группами;
- лабораторное тестирование.

4. Тестовые сценарии

Тестовые сценарии формируются на основе реальных задач пользователя. Поскольку таких задач в повседневной работе пользователя может быть очень большое число, для проведения тестирования пользовательские задачи выбираются с учетом следующих правил:

- все задачи должны быть реальными, обычными;
- выбираются наиболее важные задачи (частотные – используемые часто или всеми пользователями);
- выбираются задачи, подозрительные с точки зрения юзабилити-специалиста;
- обязательно тестируются задачи критические (приводящие к финансовым потерям, представляющие потенциальную опасность и пр.).

Тестовый сценарий состоит из следующих компонентов:

- 1) тестовое задание пользователю;
- 2) признак успешности выполнения задания;
- 3) значимые эргономические метрики с указанием критериев успешности;
- 4) точка входа (с какой точки интерфейса пользователь должен начать выполнение задания);
- 5) сценарии выполнения задания (возможны несколько альтернативных способов).

Как отмечалось выше, тестовые задания разрабатываются в соответствии с пользовательскими задачами (задачами предметной области). К формулировкам тестовых заданий предъявляются следующие требования:

- однозначность (респондент должен точно понимать, что от него требуется);
- полнота (в задании должны присутствовать все данные, необходимые для его выполнения);

- краткость (формулировка задания должна быть минимальной по объему, чтобы время чтения не влияло существенно на оценку времени выполнения задания у разных пользователей);
- отсутствие подсказок (тестовое задание формулируется в терминах предметной области, без указания используемых интерфейсных элементов; не должно содержать процедурных подсказок о последовательности действий).

Для каждого респондента формируется набор тестовых заданий таким образом, чтобы время сеанса тестирования не превысило допустимые нормативы: рекомендуемое – 40 минут, предельно допустимое – 1,5 часа.

5. Подбор респондентов

Важной частью любого тестирования является правильный подбор пользователей (и достаточного их числа). Участники теста должны быть типичными представителями пользователей данного продукта. Их количество зависит от многих факторов: времени, ресурсов, проекта теста, типа тестируемых задач и вида статистического анализа результатов.

При выборе респондентов для тестирования сначала удобно определить общие требования к респондентам, а уж потом подбирать респондентов из целевой аудитории, пользуясь сформированными требованиями.

Общие правила подбора респондентов.

1) Опыт работы - если оптимизируется интерфейс существующей системы, половина респондентов должна иметь опыт работы (на них можно определить проблемы переучивания при внедрении), а половина нет (на них определяется скорость обучения). Если существуют конкурирующие системы, лучше другая пропорция: треть с опытом работы с предыдущей версией, другая треть – с опытом использования конкурирующих систем, оставшаяся треть – без опыта работы с системой.

2) Уровень компьютерной грамотности - предпочтительным выбором является реальный, т.е. совпадающий с опытом целевой аудитории, уровень у трех четвертей респондентов и низкий уровень – у оставшейся четверти (на ней можно определить большее количество проблем).

3) Возраст - оптимальная пропорция: три четверти респондентов имеет возраст целевой аудитории системы, оставшаяся четверть старше (на ней можно определить большее количество проблем).

4) Пол - стоит увеличить количество женщин среди респондентов по сравнению с пропорцией в целевой аудитории, поскольку на женщинах легче определять проблемы при внедрении (женщины, в целом, медленнее обучаются, но, обучившись, лучше работают).

5) Уровень эмоциональной открытости респондентов - чем более скован респондент, тем меньше он способен сказать вам ценного. Даже определив наличие проблемы, вы не сможете добиться от него никакой информации о том, что эту проблему вызвало.

Количество респондентов определяется сложностью системы и количеством пользовательских задач, наличием нескольких категорий пользователей и планируемой длительностью тестирования.

Для небольших систем с одной категорией пользователей может быть достаточно 5-8 респондентов.

Если у системы несколько категорий пользователей, то 5-8 респондентов на каждую категорию.

Если тестируется сложная система, то формируются несколько групп респондентов для различных групп пользовательских задач.

6. Организационные мероприятия

На этом этапе решаются вопросы: где будет проводиться юзабилити-тестирование, каким образом будут привлекаться респонденты и как будет оплачиваться их участие, какое аппаратное и программное обеспечение будет использовано, сколько времени будет проводиться тестирование, какое количество тестируемых должно быть привлечено. После решения перечисленных вопросов составляется график проведения тестирования.

7. Проведение тестирования

Существует два подхода к организации рабочего места для юзабилити-тестирования: стационарное рабочее место и мобильное.

Минимальный набор средств тестирования:

- ноутбук
- веб-камера
- микрофон
- программа записи содержимого экрана
- секундомер
- распечатанные тестовые задания респондентам

- распечатанные анкеты
- бумага, ручка
- другое необходимое оборудование и расходные материалы.

В ходе тестирования выполняется следующая процедура.

- 1) включите запись и усадите респондента за компьютер
- 2) введите респондента в задачу
- 3) выясните у него его ожидания от системы
- 4) протестируйте интерфейс
- 5) выясните, насколько оправдались ожидания респондента
- 6) завершите тест.

Не забудьте по завершении тестирования поблагодарить респондента за участие в тестировании.

8. Анализ результатов теста

Анализировать результаты можно как во время, так и после теста. Анализ во время тестирования имеет достоинства и недостатки. К достоинствам относится то, что он:

- позволяет сэкономить время на этапе анализа, т.к. часть анализа производится во время более раннего этапа;
- дает наиболее непосредственное впечатление от теста, что позволяет увидеть проблемы, не замечаемые никаким другим способом.

Есть и недостатки:

- не позволяет фиксировать более одного эргономического показателя за раз, да и то на практике можно замерять только скорость работы пользователя и число человеческих ошибок (правда, это самые востребованные показатели);
- возможен только при значительном опыте юзабилити-специалиста;
- невозможен, если тест проводит один специалист, а анализ делает другой (с другой стороны, этому другому человеку обязательно пригодятся наблюдения того, кто проводил тест).

Анализ после тестирования лишен этих недостатков и достоинств. Он позволяет внимательно и вдумчиво проанализировать материал, независимо от количества и сущности измеряемых показателей.

9. Составление отчета по результатам юзабилити-тестирования

После проведения тестирования необходимо передать обнаруженные сведения заказчику. Как правило, лучшим средством для этого является более или менее формальный отчет.

Оптимальная структура отчета

1. Резюме (общая характеристика интерфейса, достоинства, недостатки, версия продукта, кто и когда проводил тест)
2. Основные проблемы (интерфейсные проблемы, проявляющиеся по всему интерфейсу – по убыванию степени важности)
3. Частные проблемы (проблемы, проявляющиеся на отдельных экранах – по убыванию степени важности)
4. Количественные данные (если они собирались – отдельная таблица для каждого показателя, возможны графики, иллюстрации, скриншоты экранов)
5. Приложение 1. Методика эксперимента и условия теста
6. Приложение 2. Описание тестовых сценариев
7. Приложение 3. Описание респондентов.

Надежность и достоверность результатов тестирования.

Необходимо отметить, что юзабилити-тестирование в корне отличается от тестирования в гуманитарных науках (социология, психология).

Можно организовать проведение юзабилити-тестирования таким образом, чтобы получить статистически значимые результаты, но ценность полученных результатов не окупит затрат времени и финансов. Т. о. оценивать значимость результатов статистическими методами не имеет смысла. Тем не менее при планировании и проведении тестирования необходимо учитывать требования надежности и достоверности результатов тестирования

Надежность теста состоит в том, что один и тот же тест может проводиться снова и снова и будет демонстрировать одинаковые результаты.

Тест считается достоверным, если точно измерено то, что должно быть измерено. Сравнительное тестирование на удобство применения продуктов может быть мало надежным.

Можно сформулировать следующие рекомендации по тестированию:

- сравнительные тесты могут представить полезную информацию при исследовании удобства применения новых версий программных продуктов;

- используйте задачи общего характера и области известных проблем в качестве базовых задач, чтобы произвести точные сравнительные измерения между версиями продуктов;
- применяйте стандартные измерения удобства применения, например, качество выполнения работ пользователем (выполняемость задач и показатель успешности их выполнения), степень удовлетворенности пользователей (данные по степени удовлетворенности и предпочтений);
- используйте внешних, независимых производителей для планирования и проведения тестирования на удобство применения, чтобы уменьшить субъективность восприятия.

Тесты на удобство применения должны разрабатываться, а их результаты анализироваться для того, чтобы совершенствовать интерфейсы продуктов и повышать производительность пользователей, а не для сравнения и сопоставления с другими продуктами и интерфейсами.

Следует помнить, что тестирование на удобство применения является неотъемлемой частью итерационного процесса разработки и проектирования программного продукта.

В бюджете тестирование на удобство применения и его оценка должны рассматриваться как часть рабочих расходов. На проведение такого тестирования должно выделяться от 5 до 10% от общего бюджета. Как и другие рабочие расходы, они впоследствии окупятся повышением доходов, связанных с улучшением качества продукта.