

Zeit ist Geld – App Evaluation in nur 5 Minuten

Erfahrungen einer Felduntersuchung und Richtlinien für App Evaluationen

Claudia Nass

Fraunhofer IESE
Fraunhofer-Platz 1
67663 Kaiserslautern
claudia.nass@iese.fraunhofer.de

Dominik Pascal Magin

Fraunhofer IESE
Fraunhofer-Platz 1
67663 Kaiserslautern
dominik.magin@iese.fraunhofer.de

Abstract

Eine der größten Herausforderungen der nutzerzentrierter Entwicklung in interaktiven Systemen ist die Einbeziehung einer großen Anzahl an Nutzern in die Evaluationsphase des Designprozesses. In diesem Artikel werden Erfahrungen einer 5-minütigen Evaluation pro Nutzer vorgestellt, mit der eine mobile crowdsourcing-App im Rahmen der FIFA Weltmeisterschaft 2014 mit mehr als 110 Nutzern in drei verschiedenen Städten weltweit evaluiert wurde. Darauf basierend geben Experten Richtlinien für die Durchführung einer leichten und schnellen Evaluation, mit deren Ergebnissen die Qualität einer App mit geringem Aufwand stark erhöht werden kann.

Keywords

App Evaluation, Feldstudie, Nutzertest, Evaluationsrichtlinien, mobile Software-Systeme, mobile App

Einleitung

Die Nutzung von mobilen Software Systemen ist sowohl im privaten Leben als auch in der Businesswelt längst angekommen (BITKOM (1) 2014, BITKOM (2) 2014, *Haufe Online Redaktion 2012*, IDC 2013). Neue Technologien, wie *Wearables* und *Beacons*, verbessern die Nutzererfahrung, besonders durch die Verwendung von personalisierten, ortsbezogenen Dienstleistungen (Buenen und Teje 2015). Viele Unternehmen arbeiten an der Entwicklung von mobilen Software-Systemen bzw. investieren Zeit und Geld in den Test von mobilen Software-Systemen. Angeles (2012) berichtet, dass 40 % des Budgets während der Entwicklung eines mobilen Software-Systems für Tests ausgegeben werden. Nach Buenen

und Teje (2015) fokussieren Tests von mobilen Software-Systemen auf Funktionalität, Sicherheit und Performanz. In ihren Studien mit Softwareentwicklungsunternehmen ist aber der Anteil von Usability-Tests mobiler Anwendungen von 36 % in 2013 auf 45 % in 2014 gestiegen (Buenen und Teje 2015). Bei einer Evaluation von mobilen Anwendungen müssen die Ziele genau analysiert werden, um die richtige Mischung aus Techniken und Methoden zu finden. Es liegt jedoch auf der Hand, dass man mit einer Testart nicht alle Ziele der Teststrategie bzw. des Qualitätsplans erreichen kann (Kuniavsky 2003).

Für Nutzertests zur Messung von Usability und User Experience stehen heutzutage quantitative und qualitative Methoden zur Verfügung (Kuniavsky 2003). Diese können sowohl in kontrollierten Umgebungen als auch in realen Nutzungskontexten durchgeführt werden. Die Meinung hinsichtlich der geeignetsten Methode geht stark auseinander. Zum Beispiel sagt Lang (2013), dass der Einsatz von Nutzerstudien im Bereich mobiler Software-Systeme in realen Nutzungskontexten überbewertet sei und dass man nicht immer Tests in dem Nutzungskontext durchführen muss. Dies basiert auf statistischen Hinweisen die besagen, dass 73% der Nutzer ihr Smartphone im Bett bzw. auf der Couch benutzen (IGD 2015), was dafür sprechen würde Labortests vorzuziehen, da die Nutzung des mobilen Software-Systems in einem Raum mit reproduzierbaren Einflussfaktoren stattfindet (im Bett / auf der Couch). Dieselbe IGD-Studie (2015) sagt aber auch, dass 69 % der Nutzer während der Arbeitszeit, 51 % auch am Wochenende, und 42 % während des Pendelns ihr Smartphone verwenden. Dies bedeutet, würde man immer Labortests vorziehen, würden viele weitere Nutzungskontexte nicht berücksichtigt werden.

Die Betrachtung der Hauptziele eines Nutzertests von mobilen Software-Systemen ist von enormer Wichtigkeit, d.h. das Verstehen wie Nutzer mit einem System im realen Kontext und mit anderen alltäglichen Gegenständen interagieren, ob das System die Nutzerbedürfnisse in den entsprechenden Nutzungssituationen erfüllt und ob das System der Denkweise und Einstellungen der Nutzer entspricht (Kuniavsky 2003). Diese Ziele können ausschließlich durch Feldstudien erreicht werden. Im Gegensatz zur Argumentation von Lang (2013) finden wir es extrem relevant die Evaluation des mobilen Software-Systems in realen Kontext als Feldstudien durchzuführen.

Feldstudien von mobilen Software-Systemen haben interessante Spitznamen bekommen: *Guerrilla and Lightweight Testing* (Mankelov, 2012), *In-The-Wild Testing* (User Zoom (1), 2015), or *Fly on the Wall-Study* (Optimal Experience, 2015). Alle diese Arten von Studien besagen, dass Evaluationen vor Ort, nämlich genau da wo die Nutzer das mobilen Software-System verwenden, tiefliegende Erkenntnisse über die Nutzer und wie sie mit dem System umgehen bringt. Hierfür sprechen auch die drei oben genannte Referenzen: Nutzertests im realen Kontext sind einfach und schnell durchzuführen, sie sind flexibel und können kleine oder große Stichproben einbeziehen, sie erlauben die Kombination von quantitativen und qualitativen Messverfahren und erlauben die Beobachtungen der Interaktion zwischen mobilen Geräten und realen Welt (User Zoom (2), 2015).

Diese Faktoren waren ausschlaggebend Feldstudien durchzuführen und eigene Erfahrungen zu sammeln. In diesem Artikel stellen wir Erkenntnisse und Richtlinien einer unserer Studie vor, die Usability- und UX-Researcher dabei unterstützen sollen, eine schnelle Feldstudie

durchzuführen. Als Beispiel für die dargestellten Richtlinien verweisen wir auf eine Nutzerstudie, die während der FIFA Weltmeisterschaft 2014 durchgeführt wurde.

Kontext der Feldstudie

Gerade vor einem Fußballspiel ist die Zeit, die Besucher bereit sind für eine Umfrage zu investieren, sehr gering. Die Suche nach dem besten Platz, das Kaufen von Verpflegung, das Spiel von Beginn an mitzubekommen sind die entscheidenden Faktoren. Es ist daher nötig eine Feldstudie so kurz wie möglich zu gestalten, aber dennoch alle wichtigen Aspekte abzudecken. Die vorgestellte Feldstudie wurde im Rahmen des Projekts RESCUER (*Reliable and Smart Crowdsourcing Solution for Emergency and Crisis Management*) durchgeführt. Das Ziel von RESCUER ist die Entwicklung einer intelligenten, computer-basierten Lösung für die Unterstützung der Notfall- und Krisenverwaltung bei Großveranstaltungen und in industriellen Anlagen. Die in RESCUER entwickelte Lösung besteht aus vier verschiedenen Komponenten (*Mobile Crowdsourcing Solution*, Datenanalyse Komponente, *Emergency-Reaction-Toolkit* und Kommunikationsinfrastruktur).

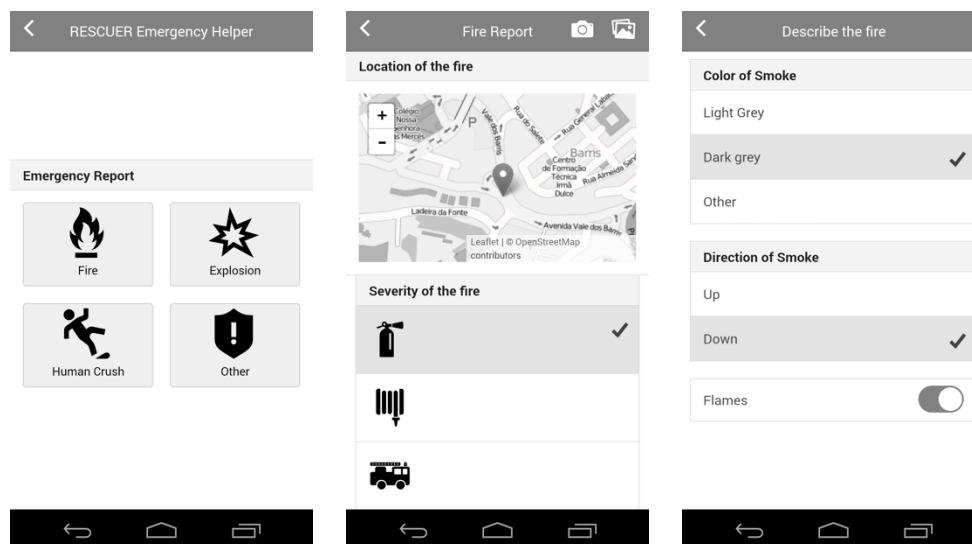


Abbildung 1: Screens von der RESCUER App

Die Mobile Crowdsourcing Komponente, auch RESCUER App in Abbildung 1, hat das Ziel, betroffene Personen in einer Gefahrensituation zu unterstützen und (detaillierte) Informationen über den Notfall an die Kommandozentrale weiterzugeben. Betroffene können eine Notfallmeldung mithilfe der RESCUER App absetzen und, sobald sie in der Lage sind, detaillierte Information über den Vorfall versenden. Zum Beispiel, im Fall eines Feuers während einer großen Veranstaltung können die Benutzer mittels der RESCUER App die genaue Position des Feuers zeigen, den Schweregrad und weitere Eigenschaften des Feuers

beschreiben. Durch die RESCUER App können die Betroffenen auch Anweisungen der Kommandozentrale empfangen, um z.B. sich in Sicherheit zu bringen.

Die Felduntersuchung war so aufgebaut, dass die Teilnehmer einen Notruf mit Hilfe der RESCUER App absetzen und dazu zusätzliche Informationen über den Vorfall mitschicken mussten. Als Unterstützung zur Durchführung der Aufgaben haben die Teilnehmer ein Bild bekommen, auf dem Feuer in einem Stadion zu sehen war. Die Teilnehmer wurden durch zwei Personen während der Nutzung begleitet und nach einer bestimmten Nutzungszeit der RESCUER App haben sie einen Fragebogen ausgefüllt. Insgesamt haben über drei Standorte hinweg 112 potentielle Nutzer (57 % weiblich, Altersdurchschnitt = 26 Jahre, Min = 13, Max = 68, SD = 11.67) an der Felduntersuchung teilgenommen. Eine komplette Beschreibung der Studie und Ergebnisse kann in Nakagawa et al. (2015) gelesen werden. Im folgenden Abschnitt konzentrieren wir uns auf konkrete Aspekte der Studie und diskutieren anhand diesen Richtlinien und Empfehlung für ähnliche, kurze Feldstudien.

Erfahrungen und Richtlinien für kurze Feldstudien mobiler Software-Systeme

In diesem Abschnitt werden unsere Erfahrungen und daraus abgeleitete Richtlinien für kurze Feldstudien mobiler Software-Systeme vorgestellt. Sie sind in drei Gruppen unterteilt: (1) Ziele und Messverfahren, (2) Vorbereitung und Ablauf und (3) Grundlagen und Offensichtliches.

Ziele und Messverfahren

Ziel der Feldstudie war die Evaluation der Usability, Ästhetik und Akzeptanz der App im Kontext von Großveranstaltungen. Dafür wurden sowohl quantitative als auch qualitative Messverfahren verwendet, um Verbesserungspotential für die RESCER App abzuleiten. Tabelle 1 zeigt welche Messverfahren für jedes Evaluationsziel angewendet wurden.

Evaluationsziele	Messverfahren
Effektivität	Anzahl von erfolgreich durchgeführten Aufgaben
Pragmatische und hedonische Qualitäten	Mini AttrakDiff (Bertram et al., 2010)
Demographische Daten	Personalisierter Fragebogen
Verbesserungspotential	Notizen des Beobachters

Tabelle 1: Evaluationsziele und Messverfahren

Ziel der Studie stark eingrenzen

In dem gesamte Evaluationsplan der RESCUER App (Nakagawa et al. 2014) wurde anhand der ISO NORM *Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE)* (ISO/IEC 25010:2011) ein breites Spektrum an Qualitätsmerkmalen festgelegt, an denen sich die RESCUER App orientieren muss. Diese Merkmale betrachten sowohl Produktqualitätsaspekte (*Product Quality*), wie funktionale Vollständigkeit, Verbrauchsverhalten, Benutzbarkeit (Bedienbarkeit, Attraktivität und Erlernbarkeit), Übertragbarkeit (Anpassbarkeit und Installierbarkeit), als auch Nutzungsqualitätsaspekte (*Quality in Use*), wie Effektivität, Effizienz, Risikofreiheit, Zufriedenheit (Nützlichkeit, Vertrauen, Kontextanpassbarkeit). Diese Qualitätsaspekte können nur anhand von Tests und Evaluationen überprüft werden. Wie schon vorher erwähnt, es ist nicht möglich in einer Studie alle relevanten Qualitätsmerkmalen zu evaluieren.

Für die Feldstudie der RESCUER App wurde nur ein Ausschnitt aller Qualitätsmerkmale ausgesucht, vgl. Tabelle 1. Besonders wenn die Qualitätsmerkmale anhand eines Fragebogens gemessen werden, müssen die Ziele stark reduziert werden. Aufgrund der kurzen Durchführungszeit der Untersuchung ist es von enormer Wichtigkeit, sich auf Ziele die durch die Nutzung des Systems beobachtet werden können zu konzentrieren. **Konzentrieren Sie sich auf die wesentlichen Qualitätsattribute, die man durch die Interaktion mit dem System überprüfen kann.**

Verhaltensbasierte Messverfahren sollen bevorzugt werden

Rohrer (2014) präsentiert ein 3-dimensionales Framework zur Auswahl von UX-Research Methoden. In diesem Framework klassifizierte er verschiedene UX-Research Methoden in drei Dimensionen:

1. Verhaltensbasiert (was machen die Leute?) vs. haltungsbasiert (was sagen die Leute?)
2. Qualitative (warum und wie kann man korrigieren?) vs. Quantitative (wie viele Leute und wie viel unterscheiden sie sich?)
3. Nutzungskontext (reale Kontext oder kontextnahe Nutzung des Produkts, Drehbuch-geschriebene Nutzung, ohne Nutzung des Produkts während der Studie, oder Mischform)

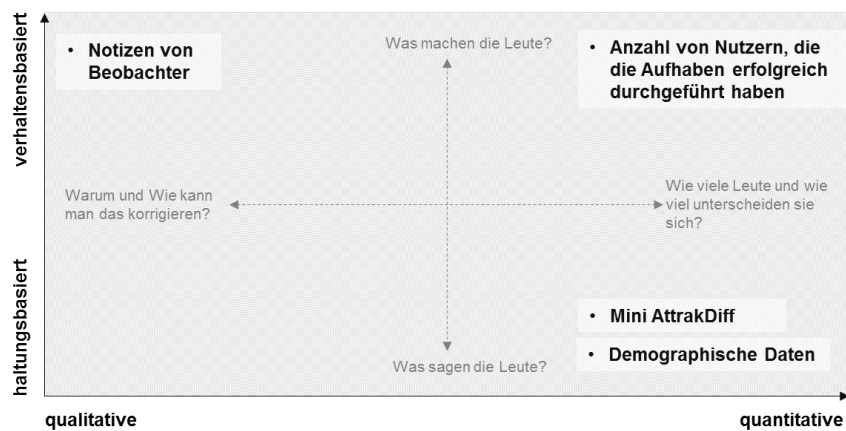


Abbildung 2: Klassifizierung der Messverfahren der RESCUER App Feldstudie (nach Rohrer 2014)

Für die Überprüfung der Qualitätsmerkmale der RESCUER App wurde eine Mischung von verschiedenen Messverfahren erstellt, vgl. Tabelle 1. Die genannten Messverfahren beziehen sich auf unterschiedliche Arten von Daten und Erhebungsmethoden. Abbildung 2 zeigt eine Klassifizierung der angewendeten Methoden nach Rohrer (2014).

Der Vergleich von verhaltensbasierten und haltungsbasierten Messverfahren zeigt auch kontroverse Ergebnisse. Verhaltensbasierte Messverfahren (Anzahl von Nutzern, die die Aufgaben erfolgreich durchgeführt haben) zeigen schwere Nutzungsprobleme, während die haltungsbasierten Messverfahren (Mini AttrakDiff) auf eine sehr positive Erfahrung mit der RESCUER App hinweisen. Es lies sich kein Zusammenhang zwischen verhaltensbasierten und haltungsbasierten Ergebnissen nachweisen. Abbildung 3 zeigt die Anzahl an Nutzern, die die Aufgaben erfolgreich durchgeführt haben. Wie bei Aufgabe 2 und Aufgabe 4 beobachtet wurde, konnte knapp die Hälfte der Teilnehmer die Aufgabe nicht erfolgreich durchführen. Abbildung 4 zeigt aber eine extrem positive Bewertung der RESCUER App anhand von Mini AttrakDiff.

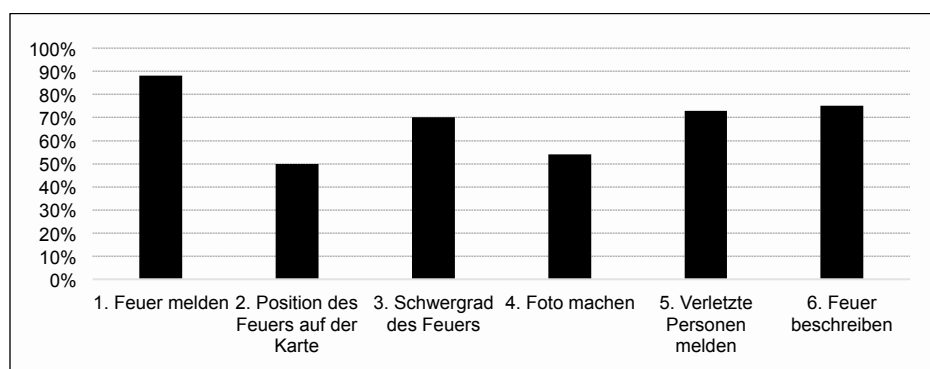


Abbildung 3: Anzahl von Nutzern, die die Aufgaben erfolgreich durchgeführt haben (N = 112)

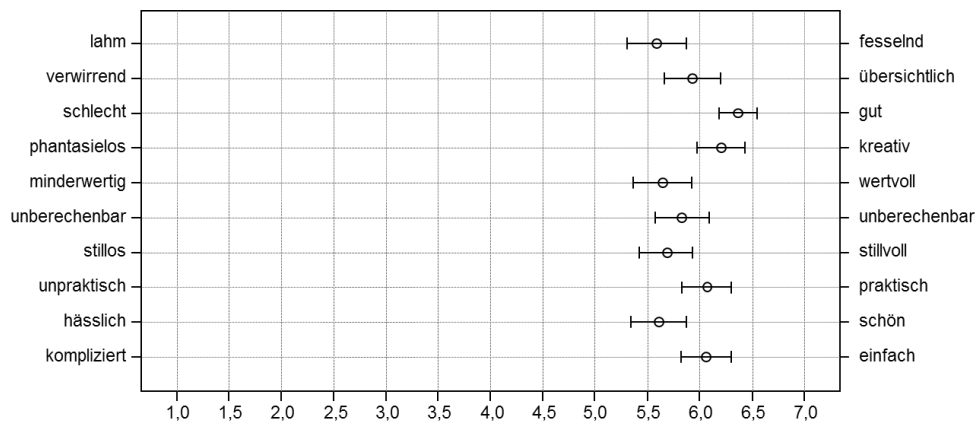


Abbildung 4: Berechneter Mittelwert von Mini AttrakDiff (N = 112)

Die Nutzung von haltungsbasierten Messfahren hat in dieser Feldstudie kein weiteres Ergebnis für die Verbesserung der RESCUER App geliefert. Folgende Faktoren haben aus unserer Sicht dazu beigetragen:

- Extrem kurze Nutzungszeit des Systems: Die Teilnehmer in der Feldstudie haben die RESCUER App nur für eine kurze Zeit verwendet. Der gesamte Testverlauf beträgt nur 5 Minuten, die Nutzung der RESCUER App selbst lag bei ca. 3 Minuten pro Benutzer. Haltungsbasierte Messverfahren basieren auf der Meinung der Person bezüglich des Systems. Aber offensichtlich bietet die kurze Nutzungszeit nicht die Gelegenheit eine realistische Meinung gegenüber dem System zu bilden.
- Einfluss der Untersuchungssituation: Die Teilnehmer füllten die Fragebogen mit Hilfe einer Evaluationskarte (Abbildung 5) aus. Obwohl Moderator und Beobachter nicht über die Schulter der Teilnehmer geschaut haben, wurden sie beobachtet: soziale Erwünschtheit, Reaktivität können das Ausfüllen der Fragebogen beeinflusst haben.
- Die Teilnehmer waren in Eile: wie schon erwähnt, wurde die Studie unmittelbar vor einem Spiel der FIFA Weltmeisterschaft 2014 durchgeführt. Die Teilnehmer waren in Eile sich gute Plätze zu suchen und Verpflegung zu organisieren. Die Eile beim Ausfüllen des Fragebogens kann auch die Ergebnisse beeinflussen haben.

Konkrete Hinweise für die Verbesserung der RESCUER App kamen aus den Kommentaren der Beobachter, die das Verhalten und Schwierigkeiten der Benutzer andeuteten. All dies führte dazu, dass in unserer Feldstudie sich besonders die verhaltensbasierten Messverfahren bewährt haben. **Wir empfehlen die Nutzung von verhaltensbasierten Messverfahren für kurze Feldstudien, da sie konkrete Hinweise zur Verbesserung des Systems liefern.**

Durchzuführende Aufgaben angemessen zuschneiden, sodass diese leicht zu verstehen sind und zu dem aktuellen Kontext in Beziehung stehen

Analysieren Sie alle Aufgaben, die man mit der App durchführen kann (z.B. von App Herunterladen, First-Login, Login, Hauptfunktionalitäten, Konfiguration) und beachten Sie dabei, wo diese Aufgaben durch die Hauptzielgruppe durchgeführt werden. Priorisieren Sie, welche Aufgaben Sie testen möchten und ob es möglich ist, alles in kurzer Zeit zu testen. Ist dies nicht möglich, bereiten sie mehrere Aufgabensätze vor und lassen Sie diese durch verschiedene Probanden durchführen, so dass jeder Proband nicht zu lange in den Test involviert ist.

In unserem Beispiel wurden sechs verschiedene Aufgaben getestet (vgl. Abbildung 3). Diese sechs Aufgaben wurden in drei Aufgabensätze geteilt. Die Intention dahinter war, die Evaluationszeit gering zu halten und die Studienteilnehmer nicht zu überfordern. Tabelle 2 zeigt die drei Aufgabensätze, die verwendet wurden. Einerseits waren wir erfolgreich, da die Evaluationszeit kurz gehalten wurde. Andererseits haben viele Teilnehmer die anderen Funktionalitäten einfach so verwendet, obwohl sie nicht dazu aufgefordert wurden, die Aufgabe durchzuführen.

Aufgabensatz 1	Aufgabensatz 2	Aufgabensatz 3
Aufgaben: 1. Informieren Sie, dass Sie Feuer sehen. 2. Informieren Sie, dass die Feuerstelle sich auf der anderen Seite des Stadions befindet. 3. Informieren Sie über den Schweregrad des Feuers 4. Machen Sie ein Bild des Feuers.	Aufgaben: 1. Informieren Sie, dass Sie Feuer sehen. 2. Informieren Sie, dass Sie verletzte Personen sehen. 3. Beschreiben Sie die Eigenschaften des Feuers. 4. Machen Sie ein Bild des Feuers.	Aufgaben: 1. Informieren Sie, dass Sie Feuer sehen. 2. Informieren Sie, dass sich die Feuerstelle auf der anderen Seite des Stadions befindet. 3. Informieren Sie, dass Sie verletzte Personen sehen. 4. Machen Sie ein Bild des Feuers.

Tabelle 2: Aufgabensätze

Wir empfehlen weiterhin sich einen angemessenen Aufgabensatz zu überlegen und dies entsprechend zu testen, so dass die Feldstudie in kurzer Zeit durchgeführt werden kann. Benötigt ein Aufgabensatz zu lange, so erwägen Sie diesen in mehrere kleinere sinnvolle Sätze aufzuteilen.

Vorbereitung und Ablauf

Maximale Dauer von 5 Minuten pro Teilnehmer

Mit einer kurzen Durchführungszeit erreichen Sie eine höhere Aufmerksamkeit der Probanden. Erfahrungsgemäß sinkt mit fortschreitender Umfragedauer die Aufmerksamkeit und Motivation von Probanden. Durch das Unterteilen zu großer Aufgaben und Ziele in

mehrere kleine Aufgaben können Sie die Umfragedauer entsprechend reduzieren. Achten Sie darauf, dass die Ermittlung Studienrelevanter Daten (demographischer Daten, ...) nicht zu lange dauert.

Evaluation als Team durchführen: ein Moderator und ein Beobachter

Gerade im Hinblick auf eine schnelle Studiendurchführung ist es notwendig, die Evaluation als Team durchzuführen – ein Moderator und ein Beobachter sind notwendig. Der Moderator ist für die Moderation durch die Studie zuständig. Er stellt Fragen zu den zu erhebenden Daten, beschreibt die Ziele, stellt die Aufgaben vor und leitet den/die Teilnehmer durch die Studie. Der Beobachter unterstützt den Moderator, indem er ihm benötigte Hilfsmittel reicht, das Interviewprotokoll führt, den Fragebogen ausfüllt und Auffälligkeiten bei der Durchführung der Aufgaben notiert. Vor der Studiendurchführung sollten mehrere Tests bzw. Durchläufe des Settings erprobt werden, um einen optimalen Ablauf sicherzustellen. **Wir empfehlen solch eine Studie immer paarweise durchzuführen. Dies hilft bei der Unterstützung der Teilnehmer und bei der Durchführung der Studie. Dies führt zu einer höheren Qualität der Daten.**

Unterstützendes Material entsprechend vorbereiten

Feldstudien mobiler Software-Systeme bringen eine ganz spezielle Situation für die Untersuchungssituation mit sich. Wenn man mobil agiert muss man unterstützendes Material sehr bedacht zusammenstellen. Für unsere Feldstudie haben wir eine physische Evaluationskarte im A5-Format entwickelt (Abbildung 5). Der Beobachter konnte auf einer Seite der Evaluationskarte seine Kommentare notieren und auch welche Aufgaben durch den Teilnehmer erfolgreich durchgeführt wurden oder nicht. Auf der Rückseite derselben Karte waren die Fragebogen für die Teilnehmer. Alle Daten der Teilnehmer waren zusammen auf einer Karte. Diese wurden nummeriert und danach für die Auswertung digitalisiert.

OBSERVER: Task Set 1

Device: Evaluation Nr.: TS1/

App Presentation Without demonstration With demonstration

Questions to be taken into account here are questions related to the understanding or use of the app. Questions or requests regarding the instruction do not count!

Task 1: Report that you see a fire.

accomplished without further questions
 not accomplished after 1-2 questions
 after 3-4 questions
 after 5 or more questions

Task 2: Inform that the fire spot is on the other side of the stadium.

accomplished without further questions
 not accomplished after 1-2 questions
 after 3-4 questions
 after 5 or more questions

Task 3: Inform the severity of the fire.

accomplished without further questions
 not accomplished after 1-2 questions
 after 3-4 questions
 after 5 or more questions

Task 4: Take a picture of the fire.

accomplished without further questions
 not accomplished after 1-2 questions
 after 3-4 questions
 after 5 or more questions

Comments

PARTICIPANT: Task Set 1

Ihr Urteil!
 Bitte geben Sie mit Hilfe der folgenden Wortpaare Ihren Eindruck des RESCUER app wieder.
 Bitte kreuzen Sie nur jeweils ein Kästchen an!

Bitte ausfüllen... © Adina Gf2 0 min (deutsch) Marc Hassenzahl www.gripoff.de

	1	2	3	4	5	6	7	
einfach	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	kompliziert
hässlich	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	schön
praktisch	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unpraktisch
stilvoll	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	stillos
voraussagbar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unberechenbar
minderwertig	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	wertvoll
phantasielos	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	kreativ
gut	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	schlecht
verwirrend	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	übersichtlich
lahm	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	fesselnd

Würden Sie diese app benutzen, um Rettungskräfte während ein Notfall in einem großen Veranstaltung zu unterstützen? Ja Nein

Würden Sie diese App benutzen, um sich in einem Notfall in einem großen Veranstaltung in Sicherheit zu bringen? Ja Nein

Haben Sie ein Smartphone? Ja Nein

Haben Sie Erfahrung mit Notfallsituationen? Ja Nein

Alter: Weiblich Männlich

www.project-rescuer.org

Abbildung 5: In der Feldstudie angewendete Evaluationskarte: Beobachter-Seite (links) und Nutzer-Seite (rechts)

Die Evaluationskarte hat sich als ein praktisches Instrument für die Feldstudie erwiesen. Das Evaluationsteam wurde durch die Karte einerseits unterstützt und andererseits wurden alle Daten auf dieser Karte festgehalten. **Wir empfehlen die Entwicklung ähnlicher Materialien, die sowohl den Ablauf als auch die spätere Auswertung der Daten unterstützen und organisieren.**

Die gesamte Dauer der Studie soll vor Ort nicht mehr als 2 Stunden betragen

Planen Sie die Studie mit einer begrenzten Dauer und setzen Sie sich auch ein Ziel, wie viele Leute Sie in dieser Zeit ansprechen möchten. Das erhöht die Motivation und die Überwindung die Passanten anzusprechen. In unserer Studie haben wir als Ziel 1,5 Stunden und 20 Probanden pro Team festgelegt. Pro Durchlauf hat jedes Team durchschnittlich 17 Probanden die Anwendung evaluieren lassen. Am Ende des Tages, nach 1,5 Stunden, haben 50 Personen an der Evaluation teilgenommen. Beachten Sie hierbei den nötigen Aufwand: Teilnehmer aufsuchen, Leute ansprechen, sich vorstellen, App erklären, Anweisung geben, Fragebogen erklären, immer höflich sein, schnell sein, Geschenke verteilen, noch ein bisschen Smalltalk mit allen und weiter geht es – es ist gar nicht so einfach, dies über mehrere Stunden beizubehalten.

Grundlagen und Offensichtliches

Identifizieren des realen Nutzungskontexts und durchführen der Evaluation direkt vor Ort und explizite Selektion der Zielgruppe

Lassen Sie eine Anwendung bei einer Evaluation mit echten Endbenutzern immer in ihrem realen Nutzungskontext testen. Das bedeutet, wenn die App für eine große Veranstaltung entwickelt wird, lohnt es sich wenig, sie während einer Veranstaltung zu testen, wenn nur 100 Leute eingeladen sind. Wenn eine App im Kontext von Bushaltstellen geplant wird, ist es nicht ratsam in der Mitte des Uni-Campus zu testen, nur weil Studenten den öffentlichen Verkehr nehmen. Es spielen hierbei nicht nur der Ort oder die Zielgruppe, sondern auch andere Einflussfaktoren wie beispielsweise die kognitive Belastung des Benutzers oder andere Umwelteinflüsse eine Rolle.

Direkte und freundliche Ansprache verwenden

Der erste Satz ist alles! Überlegen Sie sich im Voraus ganz genau, wie Sie die Leute ansprechen möchten. Es muss etwas Interessantes und Provokatives sein, so dass die Leute anhalten um mit Ihnen zu reden. In unserem Beispiel mit der RESCUER App haben wir die Benutzer so angesprochen: „Haben Sie sich schon einmal vorgestellt wie Sie sich verhalten würden wenn es hier im Stadion auf einmal brennt?“ Die Leute waren meisten überrascht und interessiert und wir konnten direkt unsere RESCUER App vorstellen. Das Ausprobieren der App seitens der Benutzer ging dann praktisch von allein los. Als Belohnung und kleine Motivation an der Studie teilzunehmen, verteilen Sie am Ende kleine Geschenke. Dies können Süßigkeiten, Eintrittskarten oder Werbegeschenke wie beispielsweise Kugelschreiber sein.

Sich für die Zielgruppe kenntlich machen (z.B. durch Uniform oder Namensschild)

Leute sind eher ängstlich, wenn Unbekannte auf sie zukommen, daher ist es sehr wichtig, zu zeigen, dass man Teil einer Organisation ist. Die Nutzung von gleichen T-Shirts, Namensschildern oder sogar ein kleiner Stand mit einem Plakat, wo evtl. mehr Information über die Arbeit präsentiert wird, sorgen für Vertrauen. Dies allein hilft jedoch nicht, dass die Leute anhalten und mit dir sprechen.

Schlussfolgerung

In diesem Artikel werden Erfahrungen und Richtlinie zur Erstellung schneller Feldstudien mobiler Software-Systeme vorgestellt. Als Beispiel wurde eine Feldstudie ausgewählt, die im Kontext der FIFA Weltmeisterschaft 2014 mit mehr als 110 Probanden durchgeführt wurde. Die kurze Dauer der Studie ermöglichte es dennoch den Usability und UX-Experten, viele potentielle Nutzer des Systems zu beobachten und entsprechend Verbesserungspotential des mobilen Software-Systems zu erkennen.

Kurzevaluationen sind eine gute Möglichkeit, innerhalb weniger Stunden eine Anwendung von vielen potentiellen Endnutzern testen zu lassen. Durch ein intelligentes Studiendesign ist es hierdurch möglich, einen großen Funktionsumfang der Anwendung von Benutzern evaluieren zu lassen. Die hier vorgestellten Richtlinien geben einen Einblick, welche Erfahrungen wir während der Vorbereitung und Durchführung verschiedener Studien gesammelt haben.

Schnelles und qualitatives Nutzer-Feedback führt zu einer kontinuierlichen Verbesserung von mobilen Software-System und passt gut auf die kurzzyklische Entwicklung typischer mobiler Software-Systeme. Wichtige Punkte die es zu beachten gilt, sind eine starke Einschränkung der Ziele und die Auswahl von angemessenen Messverfahren. Vor allem bei kritischen Qualitätsmerkmalen wie Sicherheit, Safety oder Risikoreduktion stellt sich die Frage, ob schnelle Evaluationen empfehlenswert sind. Eine schnelle Feldstudie kann auch wertvolle Hinweise für solche Qualitätsmerkmale bringen. Wichtig ist dabei den Fokus der Studie von Beginn an zu betrachten, so wie die entsprechenden Aufgaben so vorzubereiten, dass solche Attribute überprüft werden können. Nichtsdestotrotz ist die Durchführung von mehreren Studien wichtig, um alle Aspekte eines mobilen Software-Systems zu evaluieren und zu überprüfen.

Wir werden die Erfahrungen aus weiteren Studien zur Verbesserung dieser Richtlinien verwenden. Die RESCUER App wird ebenfalls kontinuierlich weiterentwickelt und es ist eine zweite große Evaluation mit Endnutzern geplant. Diesmal wird die Nutzung nicht mehr moderiert, da Performanz und Portabilität des gesamten Systems in Fokus sind.

Danksagung

Diese Arbeit wurde im Rahmen des Projekts RESCUER durchgeführt. RESCUER wird durch die Europäische Kommission (Förderkennzeichen: 614154) und den Brasilianischen Nationalen Rat für Wissenschaftliche und Technologische Entwicklung CNPq/MCTI (Förderkennzeichen: 490084/2013-3) gefördert. Wir bedanken uns bei allen Studienteilnehmern, welche die Durchführung dieser Studie ermöglicht haben.

Literatur

- Angeles, S. (2012). *How Much Does it Cost to Make an App?: An Infographic*. Gelesen am 27.05.2015 von <http://www.ideatoappster.com/how-much-does-it-cost-to-make-an-app-an-infographic/>.
- Bertram, D., Blum, R., Bürger, S., Ecker, M., Henkel, T., Katsaros, V. (2010). *Usability-Testbericht*. Hochschule der Medien, Stuttgart,
- BITKOM (1). (2014). *Jeder zweite Smartphone-Nutzer installiert zusätzliche Apps*. Gelesen am 27.05.2015 von http://www.bitkom.org/de/themen/54894_78415.aspx.
- BITKOM (2). (2014). *App-Markt wächst rasant*. Gelesen am 27.05.2015 von http://www.bitkom.org/de/themen/54894_79327.aspx.
- Buenen, M. und Teje, M. (2015). *World Quality Report 2014-15*. Capgemini, HP, Sogeti. United Kingdom.

- Haufe Online Redaktion. (2012). *Acquisa*. Mobile ist auch im Mittelstand etabliert. Gelesen am 27.05.2014 von http://www.haufe.de/marketing-vertrieb/dialogmarketing/technik-mobile-ist-auch-im-mittelstand-etabliert_126_116250.html.
- IDC. (2013). *IDC-Studie: Deutsche Unternehmen setzen auf mobile Apps zur Verbesserung ihrer Geschäftsprozesse*. Gelesen am 27.05.2015 von <http://idc.de/de/ueber-idc/press-center/56517-idc-studie-deutsche-unternehmen-setzen-auf-mobile-apps-zur-verbesserung-ihrer-geschäftsprozesse>.
- IGD. (2015). *IDG Global Mobile 2014 Survey*. Gelesen am 25.05.2015 von <http://idgknowledgehub.com/mobileidg/idg-mobile-survey/>.
- ISO. ISO/IEC 25010:2011 *Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE)* - System and software quality models, 2011.
- Kuniavsky, M. (2003). *Observing the User Experience: A Practitioner's Guide to User Research*. San Francisco: Morgan Kaufmann Publishers, Inc.
- Lang, T. (2013). *Eight Lessons in Mobile Usability Testing*. Gelesen am 25.05.2015 von <http://uxmag.com/articles/eight-lessons-in-mobile-usability-testing>.
- Mankelov, T. (2012). *10 Tips for Mobile Usability Testing*. Gelesen am 25.05.2015 von <http://optimalexperience.com/2012/01/10-tips-for-mobile-usability-testing/>
- Nakagawa, E. Y., Santos, D. S., de Oliveira, L. B. R., Duran, A. (2015). *Evaluation report of the Mobile Crowdsourcing Solution including basic information gathering*. Public Project Deliverable. Available on http://www.rescuer-project.org/?page_id=11037
- Nakagawa, E. Y., Santos, D. S., de Oliveira, L. B. R., Ocaña, A. B. M., Torres, J. (2014). *Evaluation Model and General Evaluation Plan*. Public Project Deliverable. Available on http://www.rescuer-project.org/?page_id=11037
- Optimal Experience. (2015). *Field studies*. Gelesen am 25.05.2014 von <http://optimalexperience.com/what-we-do/field-studies/>.
- User Zoom (1). (2015). *5 Reasons for Going Remote with Mobile Website Usability Testing*. Gelesen am 28.05.2015 von http://www.userzoom.com/mobile-usability-testing-blog/5-reasons-for-going-remote-with-mobile-website-usability-testing/#_ftn3.
- User Zoom (2). (2015). *Mobile In-The-Wild Testing*. Gelesen am 27.05.2015 von <http://www.userzoom.com/mobile-in-the-wild-testing/>.