

Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií

FIIT-5208-51166

Bc. Peter Dubec

Podpora uskutočňovania kvantitatívnych
používateľských štúdií mobilných
aplikácií s využitím sledovania pohľadu

Diplomová práca

Študijný program: Informačné Systémy

Študijný odbor: 9.2.6 Informačné Systémy

Miesto vypracovania: Ústav informatiky, informačných systémov a softvérového
inžinierstva, FIIT STU, Bratislava

Vedúci práce: prof. Ing. Mária Bieliková, PhD.

Máj 2017

ČESTNÉ PREHLÁSENIE

Čestne prehlasujem, že záverečnú prácu som vypracoval samostatne, na základe konzultácií, s použitím uvedenej literatúry a na základe svojich vedomostí a znalostí.

V Bratislave, 10.05.2017

.....

Bc. Peter Dubec

POĎAKOVANIE

Chcem sa poďakovať vedúcej mojej práce, pani profesorke Márii Bielikovej za vedenie a cenné rady pri písaní diplomovej práce.

Ďalej sa chcem poďakovať celej skupine PeWe za cenné rady a konštruktívnu kritiku.

V neposlednom rade sa chcem poďakovať mojej rodine a priateľom, ktorí ma podporovali počas môjho doterajšieho štúdia na fakulte.

Anotácia

Slovenská technická univerzita v Bratislave

FAKULTA INFORMATIKY A INFORMAČNÝCH TECHNOLOGIÍ

Študijný program: Informačné Systémy

Autor: Bc. Peter Dubec

Diplomová práca: Podpora uskutočňovania kvantitatívnych používateľských štúdií mobilných aplikácií s využitím sledovania pohľadu

Vedúci diplomovej práce: prof. Ing. Mária Bieliková, PhD.

Máj, 2017

Pre dobrú interaktívnu aplikáciu je dôležité vedieť, do akej miery spätná väzba poskytnutá používateľom zodpovedá jeho súčasnej situácii. Analýza interakcie ľudí s aplikáciami sa uskutočňuje najčastejšie formou používateľských štúdií. Používateľské štúdie v oblasti testovania použiteľnosti mobilných aplikácií sa v súčasnosti vykonávajú spravidla na individuálnej báze a organizácia kvantitatívnych používateľských štúdií mobilných aplikácií je veľmi náročná.

V rámci našej práce sme navrhli metódu používateľského testu, ktorá pomocou emulácie na počítači umožňuje efektívnejšie hromadné kvantitatívne testovanie mobilných aplikácií s využitím sledovania pohľadu. V rámci overenia našej metódy sa nám potvrdila základná hypotéza, že existujú problémy s použiteľnosťou mobilných aplikácií, ktoré je možné odhaliť pomocou emulácie na počítači. Na základe toho sme vytvorili zoznam potenciálnych generických problémov použiteľnosti, ktoré je možné odhaliť pomocou emulácie na počítači. V rámci realizovaných štúdií sa nám taktiež podarilo zistiť, že úspešnosti jednotlivých úloh, účinnosť účastníkov testovania a dĺžky trvania jednotlivých úloh nie sú nijako výrazne ovplyvnené typom zariadenia, na ktorom sa úloha vykonáva.

V práci sa zaoberáme aj podporou automatizácie vyhodnocovania používateľských štúdií pomocou zostrojenia zaznamenávačov, ktoré zaznamenávajú interakciu používateľov počas používateľských štúdií mobilných aplikácií.

Annotation

Slovak University of Technology Bratislava

FACULTY OF INFORMATICS AND INFORMATION TECHNOLOGIES

Degree Course: Information Systems

Author: Bc. Peter Dubec

Master's Thesis: Towards quantitative eye tracking user studies of mobile applications

Supervisor: prof. Mária Bieliková

2017, May

For good interactive application, it is important to know to which extent user's feedback reflects his actual state. Nowadays analysis of user's interactions with a system is realized in the form of user studies. User studies aimed at usability evaluation of mobile applications are nowadays realized on individual basis and organization of quantitative usability studies of mobile applications is very difficult.

In our work, we created user study method, which thanks to emulation on computer enables more effective quantitative usability testing of mobile applications with use of eye-tracking. Thanks to verification of our method, our main hypothesis that some usability problems can be revealed thanks to use of emulation on computer was confirmed. Based on that we created list of potential generic usability problems that can be identified with use of emulation on computer. Thanks to realized usability studies we have also determined, that task success, participant effectiveness or task duration time are not significantly affected by type of device on which task is performed.

We also focus our work on support of automatic evaluation of usability studies thanks to implementation of loggers, that log user's interactions during usability studies of mobile applications.

Obsah

1	Úvod	1
2	Interakcia ľudí s aplikáciami	5
2.1	Interakcia pomocou myši a klávesnice	5
2.2	Interakcia dotykom	7
2.3	Pohľad	8
2.4	Sledovanie pohľadu pri interakcii na mobilnom zariadení	9
2.4.1	Náhlavné sledovače pohľadu	10
2.4.2	Stojan pre mobilné zariadenia	11
2.4.3	Využitie emulátoru	12
2.5	Diskusia a zhodnotenie	14
3	Testovanie použiteľnosti a používateľské štúdie	17
3.1	Používateľské štúdie	17
3.1.1	Vyhodnocovanie použiteľnosti pomocou dotazníkov	20
3.1.2	Automatizácia vyhodnocovania používateľských štúdií	22
3.2	Použitelnosť na mobilných zariadeniach	26
3.3	Metriky využívané pri vyhodnocovaní interakcií	30
3.4	Diskusia a zhodnotenie	34
4	Metóda pre kvantitatívne používateľské štúdie mobilných aplikácií so sledovaním pohľadu	37
4.1	Kroky používateľského testu	39
4.2	Diskusia a zhodnotenie	44
5	Realizácia a overenie metódy	47
5.1	Podpora automatizácie vyhodnotenia	47
5.2	Zoznam identifikovaných potenciálnych problémov s použiteľnosťou mobilných aplikácií	48
5.3	Overenie metódy	52
5.3.1	Štúdia 1: Overenie základnej hypotézy	53
5.3.2	Štúdia 2: Overenie identifikovaného zoznamu potenciálnych problémov	56
5.4	Diskusia a zhodnotenie	63

Literatúra

A Protokol štúdie 1: Overenie základnej hypotézy

B Protokol štúdie 2: Overenie identifikovaného zoznamu potenciálnych problémov

C Technická dokumentácia

C.1 Zaznamenávač pre natívne mobilné aplikácie

C.2 Zaznamenávač pre webové mobilné aplikácie

C.3 Používateľská príručka

D Zhodnotenie plnenia plánu

D.1 Zhodnotenia plánu v rámci DP2

D.2 Zhodnotenia plánu v rámci DP3

E Obsah elektronického média

F Článok publikovaný na konferencii IIT.SRC 2017

1 Úvod

Dôležitosť mobilných aplikácií stále viac a viac narastá, a tak ako narastá počet dostupných aplikácií, tak narastá aj ich zložitosť. Stále viac a viac ľudí každodenne využíva mobilné aplikácie, a preto je potrebné vytvoriť interakciu, ktorá je jednoduchá, ako sa len dá. Efektívny dizajn umožňuje používateľovi sústrediť sa na informácie, ktoré mu umožnia naplniť jeho informačnú potrebu, alebo sú pre neho inak zaujímavé. Na druhú stranu, dobrá analýza interakcie ľudí s aplikáciami nám umožňuje vyhodnotiť použiteľnosť daných aplikácií a navrhovať aplikácie, ktoré môžu používatelia využívať efektívne.

V súčasnej dobe existuje veľké množstvo rozličných výskumných metód pre štúdium a vyhodnocovanie použiteľnosti. Medzi základné metódy patria takzvané inšpekčné metódy, ako napríklad heuristické vyhodnocovanie, kognitívna prechádzka, vyhodnocovanie vlastností alebo vyhodnocovanie konzistencie [25]. Metódou na ktorú sa sústreďme v rámci našej práce je testovanie použiteľnosti pomocou realizácie používateľských štúdií. Používateľské štúdie je možné rozdeliť do dvoch základných kategórií a to kvalitatívne a kvantitatívne.

Kvalitatívne štúdie sú založené na podrobnej analýze správania používateľov počas interakcie s aplikáciou. Slúžia na odhalenie najzávažnejších chýb a následnú interpretáciu identifikovaných chýb, identifikovanie príčin vzniku týchto chýb a prípadne aj návrh riešení pre odstránenie týchto chýb. Výsledkom nie sú spravidla žiadne štatistické údaje, pretože pri kvalitatívnych štúdiách je často počet účastníkov testovania pomerne nízky (3-5 účastníkov).

Pri kvantitatívnych štúdiách sa pokúšame dokázať alebo odvodiť špecifické fakty a to meraním rozličných atribútov ako napríklad čas, kliknutia myši alebo fixácie a následným štatistickým vyhodnotením rozličných metrík z týchto nameraných atribútov. Organizácia kvantitatívnej štúdie vyžaduje väčšie množstvo účastníkov a preto môže byť finančne značne náročnejšia v porovnaní s kvalitatívnou štúdiou. Stanovenie počtu účastníkov kvantitatívnej štúdie nie je triviálnou záležitosťou. Keďže cieľom je kvantifikovanie použiteľnosti pomocou vyhodnotenia rozličných metrík, počet účastníkov testovania určuje, aké spoľahlivé výsledky môžeme dosiahnuť.

V našej práci sa špecificky zameriavame na kvantitatívne testovanie použiteľnosti mobilných aplikácií. Tak ako aj pri klasických desktopových aplikáciách musia

vývojári klásť veľký dôraz na použiteľnosť týchto aplikácií. Keďže však ide o zariadenia s pomerne malými zobrazovacími veľkosťami, značne odlišným spôsobom interakcie a iným kontextom používania, testovanie použiteľnosti aplikácií na týchto zariadeniach prináša rôzne výzvy. Medzi tieto výzvy patrí napríklad organizovanie kvantitatívnych používateľských štúdií, alebo sledovanie pohľadu pri testovaní na mobilných zariadeniach.

Vidíme veľký potenciál v uskutočňovaní hromadných kvantitatívnych štúdií, avšak tie sa na mobilných zariadeniach nedajú realizovať vo veľkom rozsahu. Hlavným cieľom tejto práce je návrh metódy používateľského testu mobilnej aplikácie s využitím sledovania pohľadu pomocou emulácie na počítači, pričom táto metóda umožní efektívnejšie hromadné kvantitatívne testovanie mobilných aplikácií.

Keďže je naša metóda zameraná na testovanie mobilných aplikácií pomocou emulácie na počítači, je zjavné, že takouto metódou nie je možné odhaliť všetky problémy spojené s použiteľnosťou mobilných aplikácií, keďže je odstránený kontext mobilného zariadenia. Na základe toho sme v práci vytvorili zoznam potenciálnych generických problémov použiteľnosti, ktoré je možné odhaliť pomocou emulácie na počítači.

Výzvou je taktiež automatická analýza interakcie používateľa, ktorá by nám umožnila identifikovať, ktoré časti aplikácií spôsobujú používateľom problémy s ich používaním, prípadne inak ovplyvňujú používateľa, teda napríklad zabráňujú používateľom v splnení ich úloh. Za účelom podpory automatizácie vyhodnocovania sme zostrojili dva zaznamenávače, ktoré zaznamenávajú interakciu používateľov počas používateľskej štúdie.

Prácu členíme nasledovne. Prehľad možných interakcií používateľov so systémom alebo mobilnou aplikáciou uvádzame v kapitole 2. V kapitole 3 uvádzame prehľad typov používateľských štúdií, aktuálny stav automatizácie používateľských štúdií, aspekty špecifické pre použiteľnosť na mobilných zariadeniach a zoznam metrík využívaných pri vyhodnocovaní interakcií. Kapitola 4 sa venuje návrhu metódy pre uskutočňovanie hromadných používateľských štúdií mobilných aplikácií s využitím sledovania pohľadu pomocou emulácie na počítači a v Kapitole 5 uvádzame opis realizácie a overenia nami navrhutej metódy. Záver a zhodnotenie práce sa nachádza v kapitole 6.

V rámci príloh prikladáme v prílohách A a B vypracované protokoly aj s kompletným vyhodnotením pre dve realizované používateľské štúdie. Technická dokumentácia, ktorá obsahuje opis implementovaných zaznamenávačov a ich používateľskú príručku sa nachádza v prílohe C. Príloha D obsahuje zhodnotenie plnenia plánu za uplynulé semestre, príloha E opisuje obsah priloženého elektronického média a príloha F obsahuje článok publikovaný na vedeckej konferencii.

2 Interakcia ľudí s aplikáciami

Existuje veľké množstvo rozličných výskumných metód pre štúdium a vyhodnocovanie použiteľnosti. Najzaujímavejšou metódou pre našu prácu je klasické testovanie použiteľnosti, kedy sa použiteľnosť testuje tak, že aplikáciu používajú jej reprezentujúci používatelia (používateľské štúdie). Pri používateľských štúdiách sa sleduje celková interakcia človeka s aplikáciou. Od účastníka používateľskej štúdie sa získava explicitná spätná väzba, napríklad pomocou zaznamenávania vstupu z klávesnice, rozličných dotazníkov alebo prieskumov. Taktiež sa získava aj implicitná spätná väzba [32]. Implicitná interakcia predstavuje interakciu so systémom, ktorú používateľ vykonáva viacmenej podvedome. Takouto interakciou je napríklad interakcia pomocou myši, alebo taktiež interakcia pomocou pohľadu. Vďaka sledovaniu implicitnej interakcie dokážeme získať o správaní používateľov aplikácie užitočné informácie, ktoré by nám sami neprezradili [3]. Sledovanie implicitnej a explicitnej spätnej väzby slúži na zber potrebných dát, ktoré sú následne vyhodnocované.

Zber dát implicitnej spätnej väzby prebieha pomocou rozličných metód a techník. Medzi používané metódy patria sledovanie pohybu myši [23], sledovanie pohľadu [28], alebo viac špecifické metódy, ako napríklad sledovanie priemeru zrenice [27], sledovanie výrazov tváre [6], meranie vodivosti kože, sledovanie činnosti srdca, či iných fyziologických parametrov.

Skúmanie rozličných typov interakcií je dôležité, aby sme získali celkový pohľad o aktuálnom stave a správaní používateľa. Jednotlivé typy interakcií nám môžu o používateľovi prezradiť užitočné informácie a kombinovanie takto získaných informácií nám môže pomôcť objasniť správanie používateľa, čo by bolo bez sledovania týchto interakcií veľmi obtiažne.

2.1 Interakcia pomocou myši a klávesnice

Interakcia pomocou myši a klávesnice nám poskytuje detailné informácie o správaní používateľa a o tom ako interaguje s daným systémom, čo nám umožňuje jednoduchšie odhaliť problémy spojené s použiteľnosťou danej aplikácie. Príkladom môže byť klasické vyplňanie formuláru. Pokiaľ sa v aplikácii nachádza formulár, vďaka záznamom zo servera vieme iba to, či používateľ daný formulár vyplnil a čo do polí formulára vyplnil. Nemáme však žiadnu informáciu o správaní používateľa počas vyplňania tohto formulára.

Vďaka sledovaniu pohybu myši a monitorovaniu stlačených kláves na klávesnici používateľ a, by sme v tomto prípade mohli vedieť, aký čas strávil používateľ pri vyplňaní jednotlivých políček formulára a taktiež aj v akom poradí políčka formulára vyplňal. Vďaka analýze týchto dát môžeme následne identifikovať, či používateľ počas vyplňania preskočil nejaké políčka a vrátil sa k nim až dodatočne, prípadne, či používateľ potreboval na vyplnenie nejakého políčka výrazne dlhší čas v porovnaní s ďalšími políčkami.

Z dát získaných vďaka zaznamenávaniu interakcie s klávesnicou môžeme vidieť, ako používateľ vyplňal jednotlivé políčka formulára. Teda napríklad, či pri vyplňaní daného políčka formulára opätovne nezačínal zlý formát vstupných dát. Takéto správanie by indikovalo, že používateľ mal problém s vyplňaním formulára a teda vieme, že daný formulár môže spôsobovať problémy s použiteľnosťou danej aplikácie.

Vďaka monitorovaniu pohybu myši dokážeme získať dodatočné informácie o správaní používateľov a teda môžeme lepšie vytvárať modely reprezentujúce týchto používateľov, vďaka čomu môžeme následne umožniť napríklad personalizáciu obsahu [22]. Vďaka analýze dát o pohybe myši dokážeme identifikovať rozličné vzory správania používateľov. V práci [23] uskutočnili používateľskú štúdiu počas ktorej bol zaznamenávaný pohyb myši účastníkov a následne bola uskutočnená analýza zaznamenaných dát za účelom identifikácie vzorov správania. Príkladmi identifikovaných vzorov sú:

- *scrollovanie cez zoznamy* - v prípade obrazoviek, ktoré obsahujú zoznamy využívajú používatelia myš ako značku toho, na ktorej položke sa aktuálne nachádzajú.
- *váhanie medzi zaujímavými odkazmi pred kliknutím* - vďaka pohybu myšky dokážeme predpovedať aký obsah používateľ a zaujíma na danej stránke. V rámci realizovanej štúdie bolo úlohou participantov kúpiť si v elektronickom obchode ich obľúbené hudobné CD. Následne po analýze dát bolo predpovedané aké CD by si používatelia kúpili ako druhé v poradí, vzhľadom k pohybu ich myši, kde bolo vidieť, medzi ktorými položkami sa rozhodovali. Takýmto spôsobom sa až 70% položiek podarilo predpovedať správne.
- *prázdne miesta v rozhraní* zahŕňajúce prázdnych miest pri návrhu rozhrania je bežnou praktikou, pričom tieto miesta sú využívané, aby si mohli oddýchnuť

oči používateľov. Ukázalo sa, že používatelia využívajú tieto miesta aj v prípade myši podobným spôsobom. Prípadne používatelia umiestňujú myš na prázdne miesta, aby náhodou neklikli na nejaký odkaz.

- *počiatočná pozícia myši* - zaujímavou informáciou je vedieť, kde sa myš nachádza pri zobrazení danej obrazovky, pretože je vysoká pravdepodobnosť, že používateľ bude sústrediť svoju pozornosť práve v okolí tejto oblasti. V rámci štúdie sa ukázalo, že vo väčšine prípadov používatelia nechávajú myš umiestnenú tam, kde na predošlej obrazovke klikli. Na základe tejto vedomosti je možné napríklad umiestňovať obsah na nasledujúcej stránke tak, že umiestnime obsah, ktorý chceme aby zaujal používateľovu pozornosť v okolí miesta, kde bol posledný klik na predchádzajúcej obrazovke.

Monitorovanie interakcie pomocou myši nám môže taktiež poskytnúť užitočné informácie o pohľade používateľa v prípade, že nemáme k dispozícii zariadenie na sledovanie pohľadu. Viacero prác sa zaoberá vzťahom medzi pohybom myši a pohľadom používateľa [8] [31]. Ukázalo sa, že pohyb myši a pohľad spolu úzko súvisia. Štúdia [8] ukázala, že v prípade, že nastane pohyb myši, je až 75% šanca, že pohľad používateľa sa bude nachádzať v blízkosti kurzora myši.

2.2 Interakcia dotykom

Interakcia pomocou dotyku je doslova najpriamejšou formou interakcie, kde zobrazovacie zariadenie a vstupné zariadenie sú tým istým zariadením. Vďaka tomu, že vstupné a výstupné zariadenie sú tým istým, neexistuje žiadne posunutie medzi vstupom a výstupom, alebo vstupom a pohľadom používateľa, vďaka čomu je tento typ interakcie veľmi jednoduchý a obľúbený [2].

Interakcia dotykom je typ interakcie, ktorý slúži ako náhrada interakcie pomocou myši a klávesnice v prípade mobilných zariadení, alebo ako dodatočný nový typ interakcie v prípade iných zariadení. V prípade starších mobilných zariadení bola interakcia umožnená pomocou vstavanej hardvérovej klávesnice. V dnešnej dobe, kde sa do popredia dostávajú smartfóny a tablety, je interakcia s mobilným zariadením takmer výhradne umožnená iba pomocou dotyku na vstavanú dotykovú obrazovku zariadenia.

Môžeme rozlišovať medzi viacerými typmi interakcie pomocou dotyku. Základným typom je jednoduchá interakcia, kde dotyk reprezentuje klik a súvislý pohyb pred-

stavuje scrollovanie. Druhým typom je takzvaná multidotyková (angl. multi-touch) interakcia [15]. Tento typ interakcie reprezentuje interakciu pomocou dotyku, kde zariadenie je schopné rozoznávať dotyk viacerými prstami súčasne a tým pádom umožňuje pokročilejšie spôsoby interakcie napríklad pomocou rozličných gést. Treťou typom interakcie pomocou dotyku je dotyková interakcia rozoznávajúca prsty [9]. Pri takomto spôsobe interakcie je zariadenie schopné rozoznať, ktorým prstom bol dotyk realizovaný a na základe toho uskutočniť akciu.

Problémom pri interakcii dotykom na malých zariadeniach je, že niekedy si prekryjeme značnú časť zobrazovacieho zariadenia, a preto môže byť interakcia s malými elementami rozhrania značne náročná.

2.3 Pohľad

Sledovanie pohľadu pri používateľských štúdiách je technika, vďaka ktorej je výskumník schopný vedieť, kam presne sa účastník testovania pozerá v danom časovom okamihu a taktiež vidieť ako postupne sa presúva pohľad používateľa z jednej lokality na inú.

V dnešnej dobe je väčšina zariadení určených na sledovanie pohľadu založená na metóde *odraz rohovky/stred zreničky* (angl. corneal-reflection/pupil-centre) [28]. V rámci tejto metódy zariadenie na sledovanie pohľadu vysiela lúč infračerveného svetla do oka používateľa, pričom vďaka tomuto lúču sa zvýrazní zrenička používateľa a taktiež tento lúč spôsobí malý odraz v rohovke používateľa. Následne softvér pre spracovanie obrazu identifikuje stred zreničky používateľa a miesto odrazu lúča na rohovke a vypočíta vektor medzi týmito bodmi a následne pomocou aplikovania trigonometrických výpočtov dokáže určiť bod, kam sa používateľ pozerá. Obrázok 1 znázorňuje zvýraznenú zreničku a odraz, ktorý vznikol na rohovke používateľa.

Aby bolo zariadenie na sledovanie pohľadu schopné pre každého používateľa dostatočne presne určovať kam sa pozerá, musí sa prispôbiť každému používateľovi v rámci procesu kalibrácie. Dôvodom potreby procesu kalibrácie je primárne to, že každý človek má odlišné oko a zreničku a taktiež aj odlišné správanie a pohyby oka. Tento proces pri sledovaní pohľadu na počítači predstavuje zobrazovanie malej bodky na obrazovke počítača, pričom používateľ sa na daný čas zafixuje pohľadom na daný bod a následne systém vypočíta pozíciu oka pomocou predtým uvedenej metódy a uloží si túto pozíciu vzhľadom na x a y súradnice bodky nachádzajúcej sa



Obr. 1: Zvýraznená zrenička používateľ'a (angl. bright pupil) a odraz na rohovke (angl. corneal reflection) [28]

na obrazovke počítača.

V rámci sledovania pohľadu sledujeme u používateľ'ov nasledovné základné charakteristiky:

- *Fixácie*- momenty, kedy je oko relatívne v pokoji, pričom prijíma alebo kóduje informácie. Fixácie trvajú v priemere 200-250 milisekúnd pričom rozsah trvania môže byť od 100 až po 500 milisekúnd [30]
- *Sakády* - Rýchle pohyby oka medzi jednotlivými fixáciami. Dĺžka trvania sakád je v priemere 20 až 40 milisekúnd [30]

2.4 Sledovanie pohľadu pri interakcii na mobilnom zariadení

Najväčším problémom pri používateľ'ských štúdiách mobilných aplikácií s využitím sledovania pohľadu je to, že obrazovky mobilných zariadení sú veľmi malé. Štúdia [12] hovorí, že vo vzdialenosti 50cm, čo je typická vzdialenosť v akej je mobilné zariadenie pri používaní držané od očí, je potrebná iba jedna fixácia na to, aby mozog získal dostatočne presný obraz o približne jednej štvrtine obrazovky mobilného zariadenia. Na základe toho je veľmi náročné využiť zariadenie na sledovanie pohľadu pre zber detailných informácií o používateľ'ovom správaní počas testovania použiteľnosti mobilnej aplikácie, a teda pohľad nám len veľmi ťažko dokáže pomôcť pri odhaľovaní problémov s mobilnými aplikáciami.

Aj keď sledovanie pohľadu počas testovania použiteľnosti mobilných aplikácií je komplikované kvôli viacerým faktorom, existuje viacero spôsobov ako takéto zaznamenávanie pohľadu realizovať. Hlavnými technikami aplikovanými pre zaznamenávanie pohľadu pri testovaní mobilných aplikácií sú:

- Náhlavné sledovače pohľadu (angl. head-mounted eyetrackers).

- Stojany pre mobilné zariadenia.
- Konfigurácia pod stolom (angl. below-table setup).
- Využitie emulátora.

Konfigurácia pod stolom predstavuje techniku, kedy je mobilné zariadenie umiestnené pod stolom a účastník testovania interaguje priamo s mobilným zariadením, pričom obraz zo zariadenia je prenášaný na počítačovú obrazovku, ktorú sleduje, pričom táto obrazovka obsahuje aj zariadenie na sledovanie pohľadu. Avšak táto konfigurácia už nie je príliš využívaná, keďže interakcia so zariadením pod stolom nie je pre účastníka testovania príliš intuitívna, a preto sa viac využívajú vyššie uvedené spôsoby pre sledovanie pohľadu pri interakcii na mobilnom zariadení.

2.4.1 Náhlavné sledovače pohľadu

Náhlavné sledovače pohľadu (angl. head-mounted eyetrackers) umožňujú sledovanie pohľadu účastníkov testovania aj vo viac interaktívnom prostredí, kde účastníci testovania nemusia byť počas testovania na presne určenom mieste, ale môžu sa počas plnenia úloh pohybovať. Náhlavné sledovače pohľadu alebo takzvané okuliare pre sledovanie pohľadu (angl. eyetracking glasses), sú okuliare, do ktorých sú integrované kamery a infračervené svetlá. Ukážka náhlavného sledovača pohľadu je znázornená na obrázku 2.



Obr. 2: Náhlavný sledovač pohľadu¹

Náhlavné sledovače pohľadu poskytujú veľmi efektívny spôsob ako sledovať pohľad účastníkov počas vykonávania úloh, ktoré si vyžadujú ich pohyb. Okuliare pre sledovanie pohľadu zaznamenávajú dáta o pohľade účastníkov a následne sú

¹<http://www.tobii.com/product-listing/tobii-pro-glasses-2/>

tieto dáta v reálnom čase prenášané pomocou Bluetooth alebo Wi-Fi technológie do zariadenia pripojeného k týmto okuliarom, pričom pripojeným zariadením môže byť počítač, tablet alebo mobil.

Primárnou výhodou náhlavných sledovačov pohľadu je, že nám umožňujú sledovať pohľad participantov počas vykonávania úloh v prirodzenejšom kontexte používania. V prípade využitia takéhoto zariadenia pre sledovanie pohľadu počas testovania mobilných aplikácií, hlavným problémom je, že, keďže vďaka tomuto zariadeniu poskytneme participantom viac voľnosti počas testovania, získame obrovské množstvo nadbytočných dát o pohľade používateľa, ktoré následne budeme musieť filtrovať, pretože ako uvádza práca [13] participant nebudú primárne zameraní iba na sledovanie mobilného zariadenia, keďže majú možnosť sledovať širšie okolie.

Ďalším problémom je, že participant môže počas vykonávania úloh zakryť prstami, alebo rukou časť obrazovky mobilného zariadenia a aj keď náhlavné sledovače pohľadu obsahujú integrovanú kameru pre monitorovanie interakcie, nebudeme schopní vidieť interakciu používateľa so zariadením, keďže časť obrazovky môže byť pre nás skrytá. Kvôli týmto obmedzeniam sú v súčasnosti náhlavné sledovače pohľadu vhodnejšie pre kvalitatívne testovanie mobilných aplikácií avšak nie pre kvantitatívne testovanie.

2.4.2 Stojan pre mobilné zariadenia

Niektoré spoločnosti vyvíjajú sledovače pohľadu, alebo doplnky ku sledovačom pohľadu tak, aby bolo zjednodušené sledovanie pohľadu počas práce na mobilnom zariadení. Príkladom takéhoto doplnku je stojan pre mobilné zariadenia, do ktorého je umiestnený klasický samostatný sledovač pohľadu a používateľ interaguje s mobilným zariadením, ktoré je pripevnené na tento stojan. Zariadenie pre sledovanie pohľadu je umiestnené pod mobilným zariadením, aby sledovač videl participantovi priamo do očí. Stojan taktiež obsahuje scénickú kameru umiestnenú nad mobilným zariadením, ktorá slúži na monitorovanie interakcie. Takýto stojan máme dostupný aj v rámci našej školskej UXI infraštruktúry². Ukážka stojanu pre mobilné zariadenie, na ktorom je pripevnený sledovač pohľadu a mobilné zariadenie je znázornená na obrázku 3.

²<https://www.pewe.sk/uxi/>

³<http://www.tobiipro.com/product-listing/mobile-device-stand/>



Obr. 3: *Stojan pre mobilné zariadenia so sledovačom pohľadu*³

Výhodou stojanu pre mobilné zariadenia oproti náhlavným sledovačom pohľadu je, že tento stojan je vybavený klasickým sledovačom pohľadu, ktorý je schopný zaznamenávať dáta vo vyššej kvalite ako sledovač pohľadu integrovaný v okuliaroch pre sledovanie pohľadu. Aj keď pri využití stojanu pre mobilné zariadenia účastník testovania interaguje priamo s mobilným zariadením, táto interakcia je mierne menej reálna ako v prípade využitia náhlavného sledovača pohľadu, keď že v tomto prípade je mobilné zariadenie pripevnené ku stojanu, pričom používateľ môže so zariadením mierne manipulovať rotáciou zariadenia do strán, ale nemôže ho zo stojanu odstrániť. Nevýhodou taktiež je, že účastník testovania musí byť počas vykonávania úloh usadený v takmer nemennej polohe a sústrediť svoj pohľad na mobilné zariadenie, aby bol sledovač pohľadu schopný dostatočne dobre zaznamenávať dáta.

Ďalšou nevýhodou, tak ako tomu bolo aj v prípade náhlavných sledovačov pohľadu je, že keďže scénická kamera, ktorá monitoruje interakciu účastníka testovania so zariadením je umiestnená nad zariadením, môže účastník testovania počas vykonávania úloh svojou rukou prekryť časť obrazovky mobilného zariadenia a tým nám znemožniť monitorovanie interakcie so zariadením.

2.4.3 Využitie emulátoru

Ďalším spôsobom realizácie používateľskej štúdie mobilnej aplikácie s využitím sledovania pohľadu je využitie emulátoru⁴ na počítači ako náhrady mobilného zariadenia. Emulátor je zariadenie, alebo program, ktorý nám umožňuje spúšťať softvér, ktorý je určený pre špecifické zariadenia v prostredí iného zariadenia. To znamená, že vďaka využitiu emulátoru môžeme na počítači pracovať s aplikáciami určenými pre mobilné zariadenia.

⁴<http://www.merriam-webster.com/dictionary/emulator>

V súčasnej dobe existuje široká škála emulátorov určených pre počítače, ktoré emulujú prostredie operačného systému Android⁵. Príkladmi takýchto emulátorov sú: Droid4x⁶, BlueStacks⁷, Andy⁸ alebo Genymotion⁹.

Keďže môžeme na počítači pracovať s mobilnou aplikáciou, môžeme realizovať používateľskú štúdiu, pri ktorej používateľ testuje mobilnú aplikáciu v rámci emulátora na počítači, pričom na interakciu je využívaná myš a klávesnica a pohľad používateľa je zaznamenávaný pomocou zariadenia na sledovanie pohľadu, ktoré je pripojené k počítaču. Na obrázku 4 je znázornená ukážka počítača s integrovaným zariadením na sledovanie pohľadu a taktiež samostatné zariadenie X2-60 určené na sledovanie pohľadu, ktoré máme dostupné v rámci našej školskej UXI infraštruktúry¹⁰.



Obr. 4: Počítač s integrovaným zariadením na sledovanie pohľadu (vľavo)¹¹, zariadenie na sledovanie pohľadu X2-60 (vpravo)

Hlavnou výhodou konfigurácie používateľskej štúdie mobilnej aplikácie so sledovaním pohľadu s využitím emulátora na počítači je možnosť využitia zariadenia pre sledovanie pohľadu, ktoré je umiestnené pri obrazovke počítača. Vďaka takejto konfigurácii sme v porovnaní s predchádzajúcimi uvedenými riešeniami schopní zaznamenávať kvalitnejšie dáta o pohľade používateľa. Keďže používateľ vykonáva testovanie mobilnej aplikácie priamo na počítači, je po celý čas testovania usadený pred obrazovkou počítača a svoju polohu viacmenej nemení. Navyše, keďže zariadenie na sledovanie pohľadu je umiestnené priamo pri obrazovke počítača, nenastávajú problémy ako v prípade stojanu pre mobilné zariadenia, kde používateľ

⁵<https://www.android.com/>

⁶<http://www.droid4x.com/>

⁷<http://www.bluestacks.com/>

⁸<http://www.andymotion.net/>

⁹<https://www.genymotion.com/>

¹⁰<https://www.pewe.sk/uxi/>

¹¹<http://www.tobii.com/product-listing/tobii-pro-tx300/>

mohol svojimi pohybmi brániť zariadeniu na sledovanie pohľadu, aby monitorovalo jeho pohľad.

Hlavnou nevýhodou takejto konfigurácie je úplné odstránenie kontextu mobilného zariadenia. Keďže používateľ interaguje s počítačom namiesto reálneho mobilného zariadenia, je spôsob interakcie používateľa úplne iný a môže byť mierne neprirodzený. Namiesto interakcie pomocou dotyku používateľ využíva myš a klávesnicu. Navyše sú odstránené špecifické aspekty mobilného zariadenia akými sú napríklad malé zobrazovacie zariadenie, alebo náročnejšie zadávanie dlhších textových vstupov, čo znamená, že niektoré problémy spojené s použiteľnosťou sa pri takejto konfigurácií nemusia prejavovať.

2.5 Diskusia a zhodnotenie

Analyzovali sme dostupné možnosti sledovania pohľadu počas vykonávania používateľskej štúdie zameranej na testovanie mobilnej aplikácie. Medzi najvyužívanejšie spôsoby sledovania pohľadu pri takýchto štúdiách patrí využitie náhlavných sledovačov pohľadu, využitie stojanu pre mobilné zariadenie a využitie emulátora.

Po zvážení pozitív a negatív uvedených techník pre sledovanie pohľadu na mobilných zariadeniach, sme sa rozhodli využiť sledovanie pohľadu pomocou emulácie, keďže tento spôsob nám umožňuje najjednoduchšie uskutočňovať hromadné používateľské štúdie a taktiež nám umožňuje zbierať dostatočne kvalitné dáta o pohľade používateľov.

Je zjavné, že v prípade využitia emulácie ide o značne iný typ interakcie keďže používateľ neinteraguje s reálnym mobilným zariadením, ale interaguje pomocou myši a klávesnice, pričom ako zobrazovacie zariadenie je využitá obrazovka počítača. Na základe tohto faktu vzniká množstvo otázok, či je možné odhaliť problémy s použiteľnosťou mobilných aplikácií aj pomocou emulácie na počítači. Keďže v prípade emulácie je úplne odstránený kontext mobilného zariadenia, existuje množina problémov spojených s použiteľnosťou mobilných aplikácií, ktoré priamo vyplývajú z kontextu mobilného zariadenia a takéto problémy nebude možné odhaliť pomocou emulácie, a preto je určite vhodné v rámci fázy testovania mobilnej aplikácie uskutočniť testovanie aj na reálnom mobilnom zariadení.

Veľkým problémom pri testovaní mobilných aplikácií je veľká variabilita mobilných zariadení. Existuje obrovské množstvo mobilných zariadení so širokou škálou hardvérových špecifikácií a zobrazovacích veľkostí. Z finančného a ani z praktic-

kého hľadiska nie je možné testovať mobilné aplikácie na tak obrovskom množstve reálnych mobilných zariadení. Tento problém rieši práve emulácia mobilnej aplikácie na počítači, keďže v rámci emulátora si dokážeme zvoliť zo širokej škály zariadení ktoré chceme emulovať a teda dokážeme mobilnú aplikáciu otestovať na rozličných hardvérových konfiguráciach a taktiež na rozličných zobrazovacích veľkostiach. Aj na základe tohto faktu sme si zvolili spôsob testovania mobilných aplikácii pomocou emulácie na počítači ako primárny spôsob testovania využitý v rámci našej práce.

Ako bolo uvedené v tejto kapitole, existuje veľké množstvo interakcií používateľ a s rozhraním systému. V našej práci sa zameriavame na testovanie použiteľnosti na mobilných zariadeniach, avšak našim hlavným cieľom je umožnenie hromadného testovania mobilných aplikácií pomocou emulácie, a preto je pre nás veľmi zaujímavou interakcia pomocou myši a klávesnice. Je samozrejmé, že tieto typy interakcie sú značne odlišné a každá z týchto interakcií má svoje špecifiká, avšak v dnešnej dobe sú už tieto interakcie pre používateľov bežnou súčasťou ich života, a preto veríme, že sa budú vedieť vysporiadať s nahradením interakcie pomocou dotyku interakciou pomocou myši a klávesnice. Taktiež sa zameriavame aj na sledovanie pohľadu, keďže aj pohľad predstavuje spôsob interakcie používateľ a s danou aplikáciou a dokáže nám poskytnúť užitočné podrobné informácie o správaní používateľ a.

3 Testovanie použiteľnosti a používateľské štúdie

Použiteľnosť je atribút kvality, ktorý hovorí o tom, ako jednoduché sú používateľské rozhrania na používanie¹². Použiteľnosť je veľmi dôležitou súčasťou aplikácií, pretože pokiaľ je aplikácia náročná na používanie, nikto alebo len veľmi málo ľudí ju používa. Použiteľnosť je bežne určovaná na základe piatich základných faktorov:

1. *Naučiteľnosť* (angl. *learnability*) - Aké náročné je pre nových používateľov splniť základné úlohy v rámci aplikácie?
2. *Efektívnosť* (angl. *efficiency*) - Pokiaľ sa už používatelia naučili pracovať s aplikáciou, ako rýchlo a efektívne dokážu plniť úlohy?
3. *Zapamätateľnosť* (angl. *memorability*) - Pokiaľ sa používatelia po dlhšom čase vrátia späť do aplikácie, ako ľahko môžu znovu nadobudnúť predošlú efektívnosť?
4. *Chybovosť* (angl. *errors*) - Koľko chýb používatelia robia, aké závažné sú tieto chyby a ako ľahko sa používatelia dokážu z týchto chýb spamätať?
5. *Spokojnosť* (angl. *satisfaction*) - Ako uspokojujúca/príjemná je aplikácia na používanie?

3.1 Používateľské štúdie

Medzi základné metódy pre vyhodnocovanie použiteľnosti patria takzvané inšpekčné metódy, ako napríklad heuristické vyhodnocovanie, kognitívna prechádzka, vyhodnocovanie vlastností alebo vyhodnocovanie konzistencie [25]. Pre nás najzaujímavejšou metódou však je klasické testovanie použiteľnosti, kedy sa použiteľnosť testuje tak, že aplikáciu používajú jej reprezentujúci používatelia (používateľské štúdie). Aj pri takomto klasickom testovaní existuje veľké množstvo rozličných metód a odporúčaní ako vykonávať takéto testovanie. Používateľské štúdie je možné rozdeliť do dvoch základných skupín a to kvalitatívne a kvantitatívne používateľské štúdie.

Kvalitatívne používateľské štúdie

Kvalitatívne štúdie sú najčastejším typom používateľských štúdií, pretože sa ukázalo, že v porovnaní s kvantitatívnymi štúdiami sú až o štyrikrát menej finančne

¹²<https://www.nngroup.com/articles/usability-101-introduction-to-usability/>

náročné¹³. Kvalitatívne štúdie sú založené na podrobnej analýze správania používateľov počas interakcie s aplikáciou. Slúžia na odhalenie najzávažnejších chýb a následnú interpretáciu identifikovaných chýb, identifikovanie príčin vzniku týchto chýb a prípadne aj návrh riešení pre odstránenie týchto chýb.

Ide v podstate o získanie všeobecného prehľadu o použiteľnosti danej aplikácie, avšak môže to závisieť od cieľa a konkrétnej štúdie. Pri kvalitatívnych používateľských štúdiách je miera interakcie pozorovateľa s účastníkom testovania vyššia ako pri kvantitatívnych štúdiách. Pozorovatelia sledujú ako účastník používa aplikáciu, často priamo počas vykonávania úlohy interagujú s účastníkom (napríklad mu kladú otázky počas priebehu testovania) a následne po tom ako sa všetci účastníci zúčastnia štúdie, výskumníci na základe pozorovaní a analýzy získaných dát interpretujú správanie používateľov a vytvoria prioritizovaný zoznam identifikovaných problémov s použiteľnosťou danej aplikácie. Výsledkom takejto používateľskej štúdie sú identifikované základné problémy s použiteľnosťou, avšak výsledkom nie sú spravidla žiadne štatistické údaje, pretože pri kvalitatívnych štúdiách je často počet účastníkov testovania pomerne nízky (3-5 účastníkov)¹⁴.

Príkladom metodiky ako uskutočňovať kvalitatívnu používateľskú štúdiu je metóda takzvaného "zl'aveného testovania použiteľnosti (angl. discount usability testing)" [20]. Táto metóda definuje postup, ako uskutočňovať kvalitatívne používateľské štúdie svojpomocne a bez potreby špecializovaného laboratória. Odlišným prvkom voči bežným testovaniam použiteľnosti je to, že interakcia účastníkov testovania nie je okrem pozorovania moderátorom žiadnym iným spôsobom monitorovaná. Základný postup realizácie takejto používateľskej štúdie obsahuje tieto kroky:

1. *Nájdienie participantov* - zodpovedanie otázok aký typ ľudí potrebujeme, koľko ľudí potrebujeme a akú kompenzáciu im poskytneme. Pričom odporúčaný počet participantov je minimálne tri z čoho dvaja participant by mali byť reálny používatelia aplikácie (alebo reprezentatívny budúci používatelia) a jeden participant by mal byť mimo domény testovanej aplikácie.
2. *Zadefinovanie testovacieho scenára* - vytvorenie sady úloh, ktoré budú participant počas používateľskej štúdie vykonávať.

¹³<https://www.nngroup.com/articles/usability-metrics/>

¹⁴<https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/>

3. *Vytvorenie zoznamu úloh moderátora (checklist)* - zdefinovanie krokov alebo úloh, ktoré sú spojené s organizovaním a realizáciou používateľskej štúdie
4. *Uskutočnenie používateľskej štúdie* - obsahuje prípravu pred testovaním, uvítanie participantov, vyplnenie dotazníkov, realizáciu testovaného scenára a po skončení scenára polozenie dodatočných otázok pre participanta.

Kvantitatívne používateľské štúdie

Pri kvantitatívnych štúdiách sa pokúšame dokázať alebo odvodiť špecifické fakty a to meraním rozličných atribútov ako napríklad čas, kliknutia myši alebo fixácie a následným štatistickým vyhodnotením rozličných metrík z týchto nameraných atribútov. Kvantitatívne štúdie musia byť dobre organizované, aby boli výsledky dôveryhodné (zadefinovaný protokol a dostatočne veľká a reprezentujúca testovacia vzorka, aby sme mohli výsledky extrapolovať na väčšiu vzorku používateľov). Pri kvantitatívnom testovaní minimalizujeme interakciu s účastníkom testovania, keďže testovanie spravidla vykonáva väčšie množstvo účastníkov súčasne a neboli by sme schopní s každým účastníkom dostatočne interagovať. Výsledkom týchto štúdií sú kvantifikované štatistické dáta o použiteľnosti testovaných aplikácií.

Kvantitatívne používateľské štúdie nám poskytujú možnosť merať/kvantifikovať použiteľnosť rozhraní. Kvantitatívne štúdie však nie sú realizované príliš často, keďže uskutočnenie kvantitatívnej štúdie vyžaduje väčšie množstvo účastníkov a preto je finančne značne náročnejšie v porovnaní s kvalitatívnou štúdiou. Stanovenie počtu účastníkov kvantitatívnej štúdie nie je triviálnou záležitosťou. Keďže naším cieľom je kvantifikovanie použiteľnosti pomocou vyhodnotenia rozličných metrík, počet účastníkov testovania nám určuje, aké spoľahlivé výsledky môžeme dosiahnuť. Je potrebné si definovať, aká chyba (angl. margin of error) v rámci vyhodnotenia metrík je pre nás prijateľná a na základe toho si stanovíme potrebný počet účastníkov testovania¹⁵. Pri realizácii kvantitatívnej štúdie sa odporúča zahrnúť aspoň 20 účastníkov testovania, avšak závisí to od povahy realizovanej štúdie a od toho, aké sú ciele v rámci realizácie danej štúdie.

Na druhú stranu, kvantitatívna používateľská štúdia v zjednodušenej forme môže byť realizovaná aj napríklad v podobne online dotazníkov a v takom prípade je takáto štúdia finančne menej náročná v porovnaní s kvalitatívnou štúdiou. Aj pri takejto štúdii získame od používateľov kvantifikované údaje, ktoré je možné

¹⁵<https://www.nngroup.com/articles/quantitative-studies-how-many-users/>

štatisticky vyhodnotiť, avšak nezískame také informácie o použiteľnosti ako v prípade uskutočnenia kvantitatívnej používateľskej štúdie kde účastníci pracujú s testovaným rozhraním.

Použiteľnosť môžeme kvantifikovať pomocou merania rozličných metrík. Možným metrikám a ich vyhodnocovaniu sa bližšie venujeme v Kapitole 3.3 tejto práce. Medzi najčastejšie sledované metriky patria:

1. *úspešnosť splnenia úlohy (angl. task success)*
2. *dĺžka trvania úlohy (angl. task time)*
3. *miera spokojnosti účastníka testovania (angl. satisfaction)*

Po nameraní stanovených metrík zväčša prichádza na rad vyvodenie záverov na základe nameraných hodnôt. Pokiaľ by sme napríklad v rámci kvantitatívnej štúdie porovnávali dva rozličné dizajny webovej stránky, po nameraní metrík a ich porovnaní by sme zistili, ktorý dizajn je pre používateľov vhodnejší.

3.1.1 Vyhodnocovanie použiteľnosti pomocou dotazníkov

Pri vyhodnocovaní použiteľnosti je dobrou praktikou, merať nie len ako používatelia vykonávajú jednotlivé úlohy, ale taktiež sledovať čo si používatelia myslia o použiteľnosti daného rozhrania a práve na to je vhodné využiť dotazníky. V rámci realizovania používateľskej štúdie si môžeme zostrojiť vlastný dotazník, ktorý bude obsahovať otázky špecificky navrhnuté pre náš experiment, ale taktiež môžeme využiť štandardizované dotazníky navrhnuté pre vyhodnocovanie použiteľnosti.

Štandardizované dotazníky sú navrhnuté tak, aby sme pomocou nich mohli určiť použiteľnosť a spokojnosť vnímanú účastníkmi testovania pri používaní daného produktu, alebo systému. Hlavnými výhodami pri používaní štandardizovaných dotazníkov sú¹⁶:

- *Spol'ahlivosť* - miera, do akej je daný dotazník znovupoužiteľný, teda ako veľmi sú odpovede na dané otázky konzistentné. Pokiaľ podobní, alebo rovnakí používatelia vyhodnocujú použiteľnosť rovnakého rozhrania, chceme, aby ich odpovede boli veľmi podobné, alebo rovnaké. Spol'ahlivosť dotazníkov je vo všeobecnosti meraná pomocou Cronbachovej alfy, ktorá predstavuje

¹⁶<https://www.measuringu.com/blog/standardized-usability.php>

mieru vnútornej spoľahlivosti a nadobúda hodnoty z intervalu 0 až 1, pričom hodnota 0 reprezentuje veľmi nízku spoľahlivosť a hodnota 1 reprezentuje perfektnú spoľahlivosť [11]. Pri meraní použiteľnosti sa vo všeobecnosti čokoľvek čo dosahuje hodnoty Cronbachovej alfy nad 0.7 považuje za dostatočne spoľahlivé.

- *Objektívnosť a komunikatívnosť* - štandardizované dotazníky umožňujú výskumníkom nezávisle verifikovať zverejnené merania iných výskumníkov a taktiež umožňujú výskumníkom jednoduchšie prezentovať dosiahnuté výsledky, keďže tieto dotazníky sú štandardizované a výsledky sú preto jednoznačne interpretované.

Medzi najpoužívanejšie štandardizované dotazníky pre vyhodnocovanie použiteľnosti patria napríklad: *System Usability Scale (SUS)*, *Software Usability Measurement Inventory (SUMI)*, *Questionnaire User Interaction Satisfaction (QUIS)*, *Post Study System Usability Questionnaire (PSSUQ)* a *Single Ease Question (SEQ)*.

SUS

System Usability Scale predstavuje jeden z najpoužívanejších štandardizovaných dotazníkov pre vyhodnocovanie použiteľnosti. SUS pozostáva z 10 otázok, pričom možné odpovede na otázky sú reprezentované pomocou 5 bodovej Likertovej škály¹⁷, pričom najnižšia hodnota značí úplný súhlas a najvyššia hodnota značí úplný nesúhlas [7]. Príkladmi otázok z SUS dotazníka sú:

1. Myslím si, že by som systém rád používal opakovane.
2. Systém považujem za zbytočne zložitý.
3. Systém sa ľahko používa.

Používanie a aj vyhodnotenie SUS dotazníka je pomerne jednoduché, avšak interpretácia získaných výsledkov môže byť zložitá. Aj keď získané hodnoty sú z intervalu 0 až 1, nepredstavujú percentuálne hodnoty, ale malo by sa na ne nahliadať v kontexte percentilového hodnotenia. Na základe štúdie¹⁸ bolo ukázané, že SUS skóre nad hodnotou 68 môže byť považované za skóre nad priemerom, zatiaľ čo skóre pod 68 môžeme považovať za hodnotu pod priemerným skóre.

¹⁷<http://www.simplypsychology.org/likert-scale.html>

¹⁸<http://www.measuringu.com/sus.php>

Bolo dokázané, že dotazník SUS dosahuje hodnoty spoľahlivosti nad 0.8, takže môže byť považovaný za veľmi spoľahlivý. Navyše neslúži len na meranie použiteľnosti, ale taktiež aj na identifikovanie miery naučiteľnosti [34].

SUMI

Software Usability Measurement Inventory predstavuje komerčný nástroj pre meranie kvality softvéru, alebo jeho prototypu¹⁹. Dotazník SUMI pozostáva z 50 otázok, pričom pre odpovede na otázky využíva trojbodovú škálu (nesúhlasím, neviem sa rozhodnúť, súhlasím). Po vyplnení dotazníka sú výsledky vyhodnotené pomocou nástroja SUMISCO, ktorý poskytne výsledky v rámci celkovej škály ale taktiež aj výsledky pre päť dodatočných podškál, ktoré reprezentujú efektívnosť (angl. efficiency), vplyvnosť (angl. affect), schopnosť pomôcť (angl. helpfulness), ovládateľnosť (angl. controlability) a naučiteľnosť (angl. learnability) daného systému.

PSSUQ

Post Study System Usability Questionnaire slúži na meranie vnímanej spokojnosti pri používaní daného systému. Dotazník PSSUQ pozostáva zo 16 otázok, pričom pre odpovede je využitá sedem bodová škála, kde najnižšia hodnota vyjadruje silný súhlas a najvyššia hodnota vyjadruje silný nesúhlas. Tento dotazník dosahuje hodnoty spoľahlivosti 0.8 takže môže taktiež byť považovaný za veľmi spoľahlivý [34]. Výsledkom dotazníka sú 3 skóre, ktoré reprezentujú kvalitu systému, kvalitu informácií a kvalitu rozhrania. Štvrtým je globálne skóre, ktoré sa vypočíta priemerom predošlých skóre.

3.1.2 Automatizácia vyhodnocovania používateľských štúdií

Hlavným problémom pri kvalitatívnych štúdiách je, že ideme do hĺbky jednotlivých stimulov, hľadáme dôvody a príčiny vzniknutých problémov, čo môže byť v rámci jednotlivých účastníkov testovania značne odlišné a preto sú účastníci testovaní spravidla individuálne a následne po ukončení štúdie je potrebné zozbierané dáta detailne analyzovať. To spôsobuje, že kvalitatívne štúdie sú často finančne a aj časovo veľmi náročné.

Kvantitatívne štúdie sú vo väčšine prípadov finančne náročnejšie v porovnaní s kvalitatívnymi štúdiami, keďže obvykle vyžadujú výrazne väčší počet účastníkov

¹⁹<http://sumi.ucc.ie/>

testovania. Výzvou je uskutočňovanie hromadných kvantitatívnych používateľských štúdií, pričom sú používatelia podrobne sledovaní. Takéto testovanie je umožnené v špeciálnych laboratóriách, avšak vyhodnocovanie zozbieraných dát je stále veľmi časovo náročné. Výzvou je taktiež umožnenie automatického vyhodnocovania takýchto používateľských štúdií, pričom automatické vyhodnotenie by predstavovalo analýzu zozbieraných dát a štatistické vyhodnotenie vybraných metrík.

V práci [18] preskúmali celkovo 132 metód vyhodnocovania použiteľnosti, pričom sa zameriavali na možnosti ich automatizácie a zadefinovali novú taxonómiu pre klasifikáciu metód vyhodnocovania použiteľnosti vzhľadom na možnosti ich automatizácie. V rámci tejto práce rozdelili rozhrania nad ktorými sa uskutočňuje vyhodnocovanie použiteľnosti do dvoch hlavných skupín a to *WIMP* (*windows, icons, mouse, pointer*) rozhrania, čiže klasické počítačové rozhrania a *Web* rozhrania. Definovaná taxonómia rozčleňuje skúmané metódy vyhodnocovania použiteľnosti na základe štyroch faktorov:

1. *Trieda metódy* - na vysokej úrovni určuje o aký typ vyhodnocovania ide (používateľské testovanie, inšpekčná metóda, simulácia a pod.)
2. *Typ metódy* - detailnejšie určuje ako prebieha vyhodnotenie v rámci danej triedy (napríklad v prípade triedy inšpekčných metód môže ísť o heuristické vyhodnocovanie, kognitívnu prechádzku a pod.)
3. *Typ automatizácie* - opisuje aspekt vyhodnotenia, ktorý je automatizovaný. V úvahu prichádzajú štyri možné aspekty a to:
 - (a) *Žiadna časť* - žiadna časť vyhodnotenia nie je automatizovaná
 - (b) *Zaznamenávanie* (*angl. capture*) - softvér automaticky zaznamenáva dáta o interakcii
 - (c) *Analýza* (*angl. analysis*) - softvér automaticky identifikuje potenciálne problémy s použiteľnosťou
 - (d) *Kritika* (*angl. critique*) - softvér automaticky analyzuje identifikované problémy a navrhuje riešenia
4. *Úroveň úsilia* - určuje aký typ úsilia sa vyžaduje pre realizáciu danej metódy vyhodnotenia (napr. vyžaduje sa tvorba modelu, alebo vyžaduje sa používanie rozhrania používateľom a pod.)

Ivory a Hearst [18] preskúmali celkovo 132 metód vyhodnocovania použiteľnosti pričom išlo o 75 WIMP metód a 57 metód určených pre webové rozhrania. Z celkových 132 skúmaných metód iba 29 metód bolo aplikovateľných pre oba typy rozhrania.

Ukázalo sa, že z celkového počtu 132 metód vyhodnocovania, iba 33% metód v sebe zahŕňa automatické vyhodnotenie. Z toho 13% predstavuje automatické zaznamenávanie, 18% automatickú analýzu a 2% automatickú kritiku. Pre našu prácu sú zaujímavé práve prvé dva typy automatizácie a to automatizácia zaznamenávania a automatizácia analýzy. V rámci automatizácie zaznamenávania ide o automatické zaznamenávanie aktivity účastníka testovania, výstupom čoho sú rozličné logy, ktoré môžu predstavovať logy zo strany servera, alebo klientské logy, ktoré reprezentujú interakciu pomocou myši a klávesnice.

V práci uvedené metódy zahŕňajúce automatickú analýzu predstavujú automatickú analýzu získaných dát realizovanú pomocou jedného zo štyroch spôsobov:

1. *analýza založená na metrikách (angl. metric-based)* - medzi skúmané metódy patria:
 - (a) *DRUM* [21] - vyžaduje aby výskumník manuálne zaznačil začiatok a koniec jednotlivej úlohy a následne vypočíta niekoľko základných metrik akými sú *čas splnenia úlohy*, *efektívnosť používateľov* alebo čas, kedy boli používatelia produktívny.
 - (b) *MIKE UIMS* [26] - táto metóda zaznamenáva dáta o používaní a následne z nich generuje množstvo všeobecných, fyzických, logických a vizuálnych metrik, napríklad čas produktivity alebo počet fyzických operácií potrebných na splnenie úlohy.
 - (c) *AMME* [29] - vyžaduje špecificky formátované logy a manuálne vytvorený opis systému na základe čoho následne za pomoci Petriho sietí rekonštruje a analyzuje správanie používateľa. Táto metóda vypočítava napríklad počet krokov potrebných pre splnenie úlohy alebo čas, kedy boli používatelia produktívny.
2. *analýza založená na hľadaní vzorov (angl. pattern-matching)* - hľadanie opakujúcich sa vzorov správania, ktoré môžu indikovať problémy spojené s použiteľnosťou

3. *analýza založená na úlohách (angl. task-based)* - identifikovanie rozdielov medzi výskumníkom špecifikovaným modelom splnenia úlohy a modelom plnenia úlohy účastníkom testovania. Tj. sledujú aký je očakávaný postup pre splnenie úlohy a aký bol postup účastníka testovania. Príkladmi metód sú IBOT [36], QUIP [17] alebo KALDI [1].
4. *analýza založená na odvodeniach (angl. inferential)* - tieto metódy vo väčšine prípadov vyhodnocujú metriky založené na návštevnosti. Príkladmi metrick sú počet návštev na danú stránku alebo dĺžka trvania návštevy stránky.

V práci [18] išlo primárne o snahu minimalizovať potrebu testovania reálnymi používateľmi a to pomocou modelovania a simulovania používateľov pomocou rozličných prístupov.

V súčasnosti využívanou technikou pre automatické vyhodnocovanie použiteľnosti sú takzvané mapy vizuálnej nápaditosti (angl. saliency maps) [33]. Pri tejto technike aplikácia nie je testovaná reálnymi používateľmi, ale vizuálnu scénu aplikácie zanalyzuje algoritmus, ktorý vytvorí mapu vizuálnej nápaditosti. Táto mapa reprezentuje odhad toho, kam by sa používateľ pozeral. Následne je mapa vizuálnej nápaditosti analyzovaná odborníkmi, aby sa určili prípadné problémy s použiteľnosťou aplikácie. Na obrázku 5 je znázornená ukážka mapy vizuálnej nápaditosti.



Obr. 5: Mapa vizuálnej nápaditosti²⁰

Existuje viacero prístupov pre automatizáciu vyhodnocovania používateľských štúdií no vo väčšine prípadov však nejde o automatizáciu samotného vyhodnocovania, ale o automatizáciu zberu dát a prípadne o snahu o nahradenie ľudského

²⁰<http://salicon.net/demo/>

faktoru, tj. sú využité rozličné metódy pre modelovanie a simulovanie, aby boli nahradení reálny používatelia. Výzvou je podpora automatizácie vyhodnocovania používateľských štúdií, ktorá by pomohla jednoduchšie identifikovať problémy s použiteľnosťou testovaných aplikácií.

3.2 Použitelnosť na mobilných zariadeniach

Mobilné zariadenia sú spravidla zariadenia s malými zobrazovacími jednotkami poskytujúce rozličnú funkcionality. V súčasnosti väčšinu mobilných zariadení tvoria smartfóny. Smartfóny sú zariadenia, ktoré poskytujú pokročilejšie možnosti interakcie, pripojiteľnosti a funkcionality. Vývoj v oblasti aplikácií určených pre tieto zariadenia neustále napreduje, a preto je dôležité skúmať a obohacovať metodológie špecifické pre túto oblasť. Tak ako aj pri klasických desktopových alebo webových aplikáciách, vývojári musia kladť veľký dôraz na použiteľnosť týchto aplikácií.

Práca [14] uvádza zoznam pokynov, ktoré potreba dodržiavať počas vývoja mobilných aplikácií. Základné body, sú prevzaté z pokynov pre vytváranie desktopových aplikácií alebo predstavujú ich miernu modifikáciu. Príkladmi týchto bodov sú:

- Umožnite skúseným používateľom používať skratky
- Ponúknite používateľom informatívnu spätnú väzbu
- Umožnite používateľom mať kontrolu nad rozhraním
- Redukujte požiadavky na pamäť používateľov na minimum

Ďalšie body týchto pokynov sú už špecificky zamerané na vývoj aplikácií pre mobilné zariadenia. Medzi hlavné body patria:

- *Navrhujte s ohľadom na viaceré kontexty používania* - mobilné zariadenia môžu byť používané pri rozličných podmienkach a počas vykonávania rozličných aktivít. Preto je potrebné navrhovať rozhrania mobilných aplikácií tak, aby ponúkali viacero možných spôsobov používania.
- *Navrhujte pre malé zobrazovacie zariadenia* - pri návrhu rozhrania je potrebné brať do úvahy malé veľkosti zobrazovacích zariadení
- *Navrhujte s ohľadom na zníženú mieru pozornosti* - pri používaní mobilného zariadenia je používateľ často rozptýľovaný rozličnými externými vplyvmi,

a preto nemôžeme predpokladať, že bude používaniu aplikácie venovať plnú nepretržitú pozornosť.

- *Navrhujte pre rýchlosť a zotavenie* - používateľ často potrebuje prácu v aplikácii odložiť a vrátiť sa k nej až po určitej dobe, alebo niekedy používateľ má veľa málo času na to, aby použil aplikáciu, a preto nesmie dlho čakať na zapnutie aplikácie a ani používanie aplikácie nesmie byť veľa časovo náročné.
- *Umožnite personalizáciu* - mobilné zariadenia sú obvykle využívané iba jedným používateľom, a preto je dobré poskytnúť im možnosť určitej miery personalizácie.

Princípy a metódy využívané na vyhodnocovanie použiteľnosti mobilných aplikácií sú primárne založené na technikách a testoch určených pre desktopové prostredia. Techniky pre vyhodnocovanie použiteľnosti mobilných zariadení čelia trom hlavným problémom:

1. *Malé veľkosti obrazoviek mobilných zariadení* - kvôli technickým obmedzeniam je veľa náročné podrobne sledovať aktivitu používateľov, a preto zbierané dáta sú vo väčšine prípadov kvalitatívneho charakteru.
2. *Takmer úplná absencia softvérových nástrojov špecificky orientovaných na mobilné zariadenia* - nedokážeme jednoducho zaznamenávať kliknutia na konkrétne elementy alebo vstup z klávesnice, tak ako pri desktopových prostrediach.
3. *Ďalšie problémy vyplývajúce z kontextu mobilných zariadení* - používateľ a aj samotné zariadenie sú takmer neustále v pohybe, čo môže mať vplyv na ovládanie zariadenia a navyše úlohy používateľ a sú s veľkou pravdepodobnosťou prerušované vplyvmi z externého prostredia.

Na základe tretieho uvedeného problému sa môže zdať otázne, či dáta získané v laboratórnych podmienkach sú relevantné. Boli uskutočnené viaceré štúdie pre porovnanie laboratórneho testovania a testovania pri reálnom používaní a ukázalo sa, že vo väčšine prípadov najviac relevantné dáta pochádzajú práve z laboratórneho testovania [19]. Navyše pri reálnom testovaní v teréne sa vo veľa prípadoch zbierajú iba kvalitatívne a nie kvantitatívne dáta, keďže zbieranie kvantitatívnych dát z mobilného zariadenia v teréne je technicky pomerne náročné. Na druhú stranu však

štúdiá [13] hovorí, že počas testovania používania mobilného zariadenia v teréne sa podarilo odhaliť výrazne väčšie množstvo problémov s použiteľnosťou. V tomto prípade išlo primárne o problémy vyplývajúce z používania mobilného zariadenia v teréne, kde na používateľa vplývajú rozličné externé faktory.

V rámci testovania použiteľnosti na mobilných zariadeniach v laboratórnych podmienkach existujú dva základné spôsoby realizácie takejto používateľskej štúdie. Prvým spôsobom je realizácia klasickej používateľskej štúdie, kde účastníci testovania vykonávajú úlohy na reálnom mobilnom zariadení. Druhým spôsobom je využitie emulátora na počítači, kde používatelia pracujú s mobilnou aplikáciou priamo na počítači, pričom na interakciu využívajú myš a klávesnicu. Keďže v prípade používateľskej štúdie, kde je pre testovanie mobilnej aplikácie využitý emulátor je značne iný typ interakcie so zariadením a navyše je úplne odstránený kontext mobilného zariadenia, vzniká veľa otázok, či je takýmto testovaním vôbec možné odhaliť problémy spojené s použiteľnosťou mobilnej aplikácie. Týmito otázkami sa zaoberalo viacero štúdií a ukázalo sa, že aj s využitím emulátora je možné odhaliť problémy spojené s použiteľnosťou mobilných zariadení.

V práci [12] skúmali možnosť využitia emulátora pri uskutočňovaní kvantitatívnych používateľských štúdií so sledovaním pohľadu, zameraných na testovanie použiteľnosti mobilných aplikácií. V rámci práce autori skúmali, či kvalita dát o pohľade je ovplyvnená veľkosťou projekcie.

Bol realizovaný experiment, v rámci ktorého boli sledované dve metriky a to *priemerná dĺžka fixácií* a *celkový počet fixácií*. Účastníci testovania vykonávali úlohy na troch rôznych typoch zariadenia: (1) klasický desktopový počítač, (2) desktopový počítač s dotykovou obrazovkou a (3) mobilné zariadenie. V prípade desktopových konfigurácií boli navyše zvolené tri možné projekčné veľkosti testovanej aplikácie, ktoré boli vzhľadom na veľkosť reálneho mobilného zariadenia v pomere 1:1, 2:1 a 3:1.

Pre účely testovania bola vytvorená mobilná aplikácia určená na overenie vizuálnych schopností účastníkov testovania. Táto aplikácia zobrazovala účastníkom testovania mriežku o rozmere 6x6, ktorá obsahovala samé čísla a jedno skryté písmeno. Úlohou účastníkov bolo niekoľkokrát nájsť skryté písmeno v danej mriežke.

V rámci vyhodnotenia experimentu sa ukázalo, že kvalita dát o pohľade je

ovplyvnená veľkosťou projekcie. Ukázalo sa, že čím je veľkosť projekcie väčšia, tým je priemerná dĺžka fixácií kratšia a na druhú stranu, čím je veľkosť projekcia menšia, tým je počet fixácií menší. To znamená, že pri väčšej veľkosti projekcie dokážeme zbierať detailnejšie dáta o pohľade. Ukázalo sa však, že identifikované trendy nie sú lineárne a teda nebolo definitívne určené, ktorá projekčná veľkosť je najvhodnejšia.

Autori v práci naznačili aj vzťah veľkosti projekcie k výkonnosti používateľa pri úlohách zameraných na vizuálne schopnosti (angl. visual skill tasks) a taktiež k odhaleniu problémov s použiteľnosťou. Tento vzťah však nie je vyhodnotený a autori to uvádzajú ako prácu na ďalšie obdobie.

V práci [4] sa autori zamerali na porovnanie troch rozličných prístupov pre identifikovanie problémov s použiteľnosťou na mobilných zariadeniach. Skúmanými kontextami v rámci tejto štúdie boli:

1. Využitie emulátora v špecializovanom laboratóriu.
2. Testovanie na mobilnom zariadení v špecializovanom laboratóriu.
3. Testovanie s mobilným zariadením v teréne.

Pre porovnanie týchto troch prístupov bola uskutočnená používateľská štúdia, ktorej sa zúčastnilo 36 účastníkov, ktorí boli rozdelení do troch skupín po 12 účastníkov. Tj. pre overenie každého z prístupov bola realizovaná používateľská štúdia s 12 účastníkmi, pričom zadané úlohy (7 úloh) pre účastníkov boli identické v prípade všetkých štúdií.

V rámci vyhodnotenia tejto štúdie sa sledovalo viacero metrík. Sledovanými metrikami boli:

1. *Účinnosť úlohy (angl. Task effectiveness)* - ukázalo sa, že v rámci troch sledovaných kontextov nebola účinnosť jednotlivých úloh štatisticky signifikantne odlišná.
2. *Efektívnosť účastníkov testovania (angl. User efficiency)* - ukázalo sa, že efektívnosť v rámci troch sledovaných kontextov nebola pre tri úlohy štatisticky signifikantne odlišná. V prípade porovnania testovania v laboratóriu pomocou emulátora a testovania v laboratóriu pomocou mobilného zariadenia pri štyroch úlohách bola efektívnosť účastníkov v prípade využitia mobilného

zariadenia štatisticky signifikantne odlišná (nižšia). Efektívnosť bola definovaná ako pomer úspešnosti a dĺžky trvania úlohy. V prípade úloh, kde došlo k signifikantnému rozdielu však bola účinnosť rovná hodnote 100%, znamená to teda, že rozdielnym faktorom bola dĺžka trvania úloh. Teda používateľom na reálnom zariadení trvalo splnenie úloh dlhší čas ako v prípade využitia emulátoru.

3. *Miera spokojnosti (angl. satisfaction)* - pre identifikáciu miery spokojnosti bol využitý štandardizovaný dotazník SUS. Ukázalo sa, že pri testovaní v laboratóriu s využitím reálneho zariadenia bola celková miera spokojnosti účastníkov signifikantne nižšia v porovnaní s dvoma ďalšími kontextami testovania.
4. *Pracovná záťaž (angl. workload)* - pre meranie pracovnej záťaže bol využitý dotazník NASA TLX [16]. Ukázalo sa, že v rámci troch sledovaných kontextov nebola vnímaná pracovná záťaž štatisticky signifikantne odlišná.
5. *Množstvo identifikovaných problémov s použiteľnosťou* - celkovo sa pomocou troch rozličných kontextov testovania podarilo odhaliť 43 unikátnych problémov s použiteľnosťou. Množstvo problémov identifikovaných v rámci jednotlivých kontextov bolo nasledovné: (1) testovanie pomocou mobilného zariadenia v laboratóriu - 41 problémov, (2) testovanie pomocou mobilného zariadenia v teréne - 38 problémov, (3) testovanie pomocou emulátoru v laboratóriu - 35 problémov.

Celkovo teda Betiol a Cybis [4] ukázali, že aj s využitím emulátoru je možné odhaliť problémy spojené s použiteľnosťou mobilných zariadení. Nie je však takto možné odhaliť všetky možné problémy s použiteľnosťou a navyše niekedy môže aj malá zmena v zobrazení viesť k značne odlišným výsledkom. V prípade jednej úlohy, kde emulátor zobrazoval o jeden riadok obsahu viac ako mobilné zariadenie a úlohou bolo nájsť informáciu v obsahu, úspešnosť splnenia tejto úlohy v rámci emulátoru bola značne vyššia v porovnaní s mobilnými zariadeniami.

3.3 Metriky využívané pri vyhodnocovaní interakcií

Kvantitatívne používateľské štúdie nám poskytujú možnosť merať/kvantifikovať použiteľnosť rozhraní pomocou vyhodnocovania rozličných metrick. Medzi základné metriky, ktoré sú vyhodnocované pri kvantitatívnych používateľských štú-

diách patria *úspešnosť* (angl. *success rate*), tj. či používateľ splnil alebo nespĺnil danú úlohu, *čas potrebný na splnenie úlohy*, *miera chybovosti*, *spokojnosť používateľa*, *kliknutia a zobrazenia* alebo *miera konverzií* [35]. Pri kvantitatívnych používateľských štúdiách je možné sledovať aj rozličné špecifické metriky, akou je napríklad kognitívna záťaž používateľa, alebo emócie používateľa [27].

Metriky založené na sledovaní pohľadu

V rámci sledovania pohľadu môžeme aj v závislosti od dostupného zariadenia na sledovanie pohľadu okrem sledovania fixácií a sakád monitorovať aj dodatočné charakteristiky, akými sú napríklad počet žmurknutí alebo veľkosť zreničky, ktoré môžeme následne využiť napríklad na identifikovanie kognitívnej záťaže používateľa. Alebo môžeme zo základných charakteristík odvodiť dodatočné charakteristiky, akou je napríklad snímacia cesta (angl. *scanpath*), ktorá predstavuje súvislú sekvenciu sakád a fixácií (sakáda-fixácia-sakáda...).

Práce [28], [5] uvádzajú prehľad metrick založených na sledovaní pohľadu a taktiež opis významu týchto jednotlivých metrick. Zoznam týchto metrick spoločne aj s ich opisom je uvedený v Tabuľke 1.

Tabuľka 1: *Metriky založené na sledovaní pohľadu*

Metrika	Opis
Amplitúda sakád	Dlhšie sakády indikujú zmyslupnejšie podnety, pretože pozornosť používateľa je pritiahnutá už z diaľky.
Celkový počet fixácií	Väčší celkový počet fixácií indikuje menej efektívne hľadanie.
Celkový počet fixácií pred prvou fixáciou na danú oblasť záujmu	Pokiaľ je počet fixácií pred prvou fixáciou na danú oblasť záujmu vysoký, znamená to, že daná oblasť je málo výrazná. Táto metrika určuje spozorovateľnosť danej oblasti záujmu.
Čas do prvej fixácie na cieľový element	Nižší čas potrebný na prvú fixáciu na cieľový element značí, že element je dostatočne výrazný. Táto metrika určuje spozorovateľnosť (angl. <i>noticeability</i>) danej oblasti záujmu.
Čas od prvej fixácie na cieľový element po jeho zvolenie	Táto metrika hovorí o rozpoznateľnosti cieľového elementu. Menšia dĺžka trvania indikuje väčšiu zmyslupnosť cieľového elementu.
Dĺžka snímacej cesty	Väčšia dĺžka snímacej cesty indikuje menej efektívne hľadanie.
Dĺžka trvania fixácie	Väčšia dĺžka trvania fixácie indikuje, že používateľ má problém so spracovaním informácie, alebo je pre neho daný objekt viac zaujímavý.

Dĺžka trvania snímacej cesty	Väčšie dĺžky trvania snímacích ciest indikujú menej efektívne snímanie.
Matica prechodov	Matica prechodov znázorňuje poradie hľadania v kontexte oblastí záujmu. V prípade že používateľ prechádza medzi dvoma oblasťami záujmu tam a späť značí to mieru neistoty u používateľa.
Opakované fixácie	Opakované spätné fixácie na cieľový element naznačujú, že cieľový element je málo viditeľný alebo málo zmysluplný.
Percento používateľov, ktoré sa zafixuje na danú oblasť záujmu	Pokiaľ sa na oblasť záujmu, ktoré je pre danú úlohu dôležitá fixuje málo používateľov, znamená to, že táto oblasť záujmu je málo výrazná. Táto metrika určuje spozorovateľnosť danej oblasti záujmu.
Počet fixácií na cieľový element	Pomer počtu fixácií na cieľový element a celkového počtu fixácií. Nižší pomer indikuje nižšiu efektívnosť hľadania.
Počet fixácií na danú oblasť záujmu	Väčší počet fixácií na danú oblasť záujmu indikuje, že je viac nápadná alebo dôležitejšia pre používateľa v porovnaní s inými oblasťami záujmu.
Počet návštev pohľadom na cieľový element (angl. number of gaze visits) pred jeho zvolením	Táto metrika hovorí o rozpoznateľnosti cieľového elementu. Dobre rozpoznateľný cieľový si vyžaduje iba jednu návštevu pohľadom na to, aby sme rozpoznali, že ide o cieľový element.
Pohľad (angl. gaze)	Pohľad je definovaný ako suma všetkých fixácií v rámci danej oblasti. Najviac sa využíva na porovnávanie záujmu medzi jednotlivými oblasťami.
Počet sakád	Väčší počet sakád indikuje hľadanie
Pomer počtu sakád a fixácií	Tento pomer nám určuje, či používateľ viac prehládaval obsah alebo spracovával informácie.
Priestorová hustota fixácií	Fixácie koncentrované v rámci malej oblasti indikujú zamerané a efektívne hľadanie. Fixácie rovnomerne rozložené v priestore indikujú neefektívne hľadanie.
Priestorová hustota snímacích ciest	Menšia hustota indikuje priamejšie hľadanie.
Smer snímacej cesty	Smer snímacej cesty nám môže hovoriť o stratégii prehládavania, ktorú využíva daný používateľ. Napríklad či zoznamy prehládava zhora nadol alebo zdola nahor.
Smerové posuny sakád	Sakáda ktorá sa od predošlej sakády líši o viac ako 90 stupňov naznačuje rapídnu zmenu smeru, čo môže znamenať, že sa zmenil cieľ používateľa, alebo rozhranie nespĺňa očakávania používateľa.
Spätné sakády (angl. Regressive saccades)	Spätné sakády indikujú menej zmysluplné podnety

Metriky sledované pri používateľských štúdiách

V práci [10] vytvorili prehľad metrík, ktoré sú bežne sledované počas používateľských štúdií. Základné metriky, ktoré sú najčastejšie sledované sú uvedené v Tabuľke 2.

Tabuľka 2: Metriky sledované pri používateľských štúdiách

Metrika	Opis
Dĺžka trvania úlohy (angl. Task time)	Čas potrebný na splnenie úlohy.
Efektívnosť (angl. efficiency)	<p>Miera do akej systém umožňuje splnenie zadaných úloh rýchlo a efektívne. Jedným zo spôsobov pre výpočet miery efektívnosti pre jednotlivé úlohy je výpočet efektívnosti ako podielu miery ukončenia úlohy a mediánu času potrebného pre správne ukončenie úlohy. Ďalším spôsobom je výpočet relatívnej celkovej efektívnosti, ktorá reprezentuje mieru efektívnosti vzhľadom na všetky úlohy. Výpočet relatívnej celkovej efektívnosti (RCE) môže byť realizovaný pomocou nasledovného vzťahu²¹:</p> $RCE = \frac{\sum_{j=1}^R \sum_{i=1}^N n_{ij} t_{ij}}{\sum_{j=1}^R \sum_{i=1}^N t_{ij}} \times 100\%$ <p>Pričom N reprezentuje počet úloh a R počet účastníkov. Kde n_{ij} reprezentuje splnenie úlohy i používateľom j a t_{ij} reprezentuje čas potrebný na splnenie úlohy i používateľom j.</p>
Frekvencia identifikovaných problémov	Histogram množstva výskytov identifikovaných problémov.
Frekvencia identifikovaných problémov vzhľadom na účastníka.	Táto metrika nám hovorí, aký je priemer počtu identifikovaných problémov vzhľadom na účastníka.
Frekvencia identifikovaných problémov vzhľadom na úlohu.	Priemer počtu identifikovaných problémov vzhľadom na úlohu.
Frekvencia identifikovaných unikátnych problémov	Po uskutočnení používateľskej štúdie identifikujeme problémy spojené s použiteľnosťou. Táto metrika nám hovorí o tom, koľko unikátnych problémov sme identifikovali.
Miera spokojnosti (angl. satisfaction)	Miera do akej sú používatelia spokojní s používaním daného systému. Miera spokojnosti sa meria po ukončení úlohy alebo po skončení celého testu pomocou štandardizovaných dotazníkov. Štandardizované dotazníky sú opísané v Kapitole 3.1.1 tejto práce.
Naučiteľnosť (angl learnability)	Hovorí o tom, ako rýchlo je nový používateľ schopný začať používať daný systém efektívne a bez chýb

²¹<http://ui-designer.net/usability/efficiency.htm>

Prístupnosť (angl. accessibility)	Miera použiteľnosti systému pre ľudí s rozličnými obmedzeniami
Účinnosť (angl. effectiveness)	Presnosť a úplnosťou s akou špecifickí používatelia splnia špecifické ciele. Jedným zo spôsobom ako určiť mieru účinnosti je vypočítanie miery ukončenia úloh, pomocou nasledovného vzťahu ²² : $\text{účinnosť} = \frac{\text{počet úspešne ukončených úloh}}{\text{celkový počet vykonaných úloh}} \times 100\%$ Na základe štúdie realizovanej nad 115 testovaniami použiteľnosti pozostávajúcich z 1189 unikátnych úloh sa ukázalo, že priemernou mierou ukončenia úloh je hodnota 78% ²³
Úspešnosť splnenia úlohy (angl. Task success)	Metrika, ktorá reprezentuje či používateľ splnil alebo nesplnil danú úlohu. Hodnota metriky môže byť binárna (splnil / nesplnil), ale aj viacúrovňová (napríklad splnil / čiastočne splnil / nesplnil)
Zapamätateľnosť (angl. memorability)	Miera, do akej je používateľ schopný si zapamätať používanie systému tak, aby sa nemusel znovu nič učiť keď chce systém použiť po určitej dobe

V rámci interakcie pomocou myši môžeme taktiež sledovať rozličné metriky, ktoré sú viacmenej iba miernou modifikáciou predošlých metrík. Príkladmi metrík ktoré môžeme sledovať v rámci interakcie pomocou myši sú:

- Počet kliknutí.
- Čas strávený myšou nad danou oblasťou záujmu.
- Čas potrebný do prvého kliknutia na cieľový element.
- Percento používateľov, ktoré prešlo myšou ponad cieľový element.

3.4 Diskusia a zhodnotenie

Pre dobrú interaktívnu aplikáciu je nevyhnutné, aby bola dobre použiteľná a teda umožňovala používateľom dosahovať svoje ciele rýchlo a efektívne. Aby sme vedeli, či daná aplikácia neobsahuje problémy s použiteľnosťou, je potrebné ju testovať s reálnymi používateľmi v rámci používateľských štúdií a to buď kvalitatívnych alebo kvantitatívnych.

V kontexte mobilných aplikácií je takisto nevyhnutné snažiť sa odhaliť a minimalizovať chyby spojené s použiteľnosťou danej aplikácie. Vzhľadom k

²²<http://usabilitygeek.com/usability-metrics-a-guide-to-quantify-system-usability/>

²³<http://www.measuringu.com/blog/task-completion.php>

obmedzeniam spojeným s mobilnými zariadeniami uvedenými v tejto kapitole práce je potrebné dbať na použiteľnosť mobilných zariadení ešte viac ako v prípade desktopových aplikácií.

V súčasnej dobe testovanie použiteľnosti mobilných aplikácií prebieha primárne dvoma hlavnými spôsobmi a to buď ako testovanie v špecializovanom laboratóriu alebo ako testovanie v teréne. Oba spôsoby majú svoje výhody aj nevýhody. Mnoho prác diskutuje, ktorý spôsob je vhodnejší pre testovanie mobilných aplikácií a teda dokáže odhaliť viac problémov s použiteľnosťou [19] [13].

Hlavnou výhodou testovania v laboratóriu je možnosť zbierať kvalitnejšie dáta, keďže pre zaznamenávanie interakcie používateľa s aplikáciou dokážeme použiť presnejšie zariadenia a zároveň je používateľ počas vykonávania úlohy v pokoji a plne koncentrovaný na danú úlohu. Nevýhodou testovania v laboratóriu však je, že používateľ využíva aplikáciu v menej reálnom kontexte. Pri reálnom používaní je používateľ v pohybe alebo je rušený rozličnými externými vplyvmi okolia.

Ukázalo sa, že vo väčšine prípadov najzávažnejšie problémy s použiteľnosťou sa odhalia práve pomocou laboratórneho testovania. Na druhú stranu testovanie v teréne pomáha odhaliť viac špecifické problémy spojené s kontextom využívania mobilného zariadenia v reálnom prostredí [19]. Testovanie v teréne je však časovo náročnejšie ako testovanie v laboratóriu a taktiež si vyžaduje veľké množstvo príprav. Navyše dáta získané vďaka takémuto testovaniu majú skôr kvalitatívny a nie kvantitatívny charakter. V našej práci sa zameriavame na kvantitatívne testovanie mobilných aplikácií pomocou testovania v špecializovaných laboratóriách, pričom našim hlavným cieľom je umožnenie hromadného testovania mobilných aplikácií s využitím sledovania pohľadu pomocou emulácie na počítači. Vidíme veľký potenciál v umožnení takéhoto spôsobu testovania mobilných aplikácií.

Betiol a Cybis [4] ukázali, že aj s využitím emulátoru je možné identifikovať problémy s použiteľnosťou mobilnej aplikácie. Na druhú stranu autori ukázali, že v niektorých prípadoch môže využitie emulátoru ovplyvniť efektívnosť účastníkov testovania alebo úspešnosť splnenia jednotlivých úloh. V uvedenej práci sa však autori nezameriavajú na to, aké typy problémov je možné odhaliť takýmto spôsobom a teda ani na to, kedy je vhodné využiť takýto spôsob testovania použiteľnosti mobilných aplikácií pomocou emulácie na počítači. Vidíme priestor na pomerne veľký možný prínos našej práce a teda aj našim cieľom je práve identifikovanie typov problémov s použiteľnosťou, ktoré je možné odhaliť pomocou emulácie na počítači a kedy je vhodné takýto spôsob testovania využiť.

Po uskutočnení používateľskej štúdie je potrebné zozbierané dáta analyzovať, aby sme identifikovali problémy spojené s použiteľnosťou. Proces analýzy dát je obvykle časovo veľmi náročný, a preto výzvou je podpora automatizácie vyhodnocovania používateľských štúdií. V súčasnej dobe prebiehajú snahy o automatizáciu samotných používateľských štúdií, pričom ide primárne o využitie rozličných simulácií so snahou o nahradenie testovania s reálnymi používateľmi. Príkladom automatizácie používateľských štúdií sú napríklad mapy vizuálnej nápaditosti, kde algoritmus určí, ktoré miesta grafického rozhrania by lákali pozornosť používateľov.

Takmer žiadne práce sa nezaoberajú automatizáciou vyhodnotenia dát zozbieraných počas používateľských štúdií [18]. V rámci našej práce sa zameriavame na kvantitatívne štúdie, ktoré v sebe zahŕňajú väčšie množstvo účastníkov testovania. Jedným z cieľov našej práce je návrh spôsobu podpory automatického vyhodnotenia kvantitatívnej používateľskej štúdie. To napomôže k jednoduchšiemu identifikovaniu problémov s použiteľnosťou a zredukuje čas potrebný na analýzu zozbieraných dát.

V práci sme taktiež analyzovali širokú škálu metrík, ktoré sa využívajú pre vyhodnocovanie interakcií, pričom sme sa špecificky zamerali na metriky využívané pri používateľských štúdiách zameraných na testovanie mobilných aplikácií a taktiež na metriky využívané pre sledovanie pohľadu používateľov. V práci uvádzame komplexný zoznam najvyužívanejších metrík pričom pri každej metrike uvádzame na čo môže byť využitá alebo ako môže byť vypočítaná.

4 Metóda pre kvantitatívne používateľské štúdie mobilných aplikácií so sledovaním pohľadu

V súčasnej dobe prebieha testovanie mobilných aplikácií primárne vo forme kvantitatívnych používateľských štúdií. Takéto testovanie prebieha v špecializovanom laboratóriu, kde účastníci testovania pod dohľadom moderátora testovania vykonávajú zadané úlohy, pričom je zaznamenávaná ich aktivita a problémy s použiteľnosťou aplikácie sú identifikované primárne na základe pozorovania akcií účastníkov testovania.

Hlavným cieľom našej práce je návrh metódy používateľského testu, ktorá umožňuje efektívne hromadné testovanie mobilných aplikácií a zahŕňa v sebe podporu automatického vyhodnotenia, pričom sa špecificky zameriavame na kvantitatívne používateľské štúdie s využitím sledovania pohľadu.

Analyzovali sme dostupné metódy sledovania pohľadu používateľov počas testovania mobilných aplikácií. V rámci dostupných možností sme pre každú možnosť zvážili jej pozitívne a negatívne stránky a na základe týchto faktorov sme sa pre našu prácu rozhodli využiť sledovanie pohľadu pomocou emulácie na počítači. Jedným z hlavných dôvodov prečo sme zvolili tento spôsob je, že vidíme veľký potenciál v uskutočňovaní hromadných kvantitatívnych štúdií, ktoré sa však na mobilných zariadeniach v takom rozsahu nedajú realizovať. Navyše nám hromadné testovanie pomocou emulácie umožňuje vykonávať UX infraštruktúra dostupná u nás na fakulte vo Výskumnom centre používateľského zážitku a interakcie [24].

Používateľské štúdie sú bežne definované scenárom definujúcim postupnosť krokov, na základe ktorých používateľská štúdia prebehne. Naša metóda definuje tieto kroky pre hromadné kvantitatívne používateľské štúdie mobilných aplikácií so sledovaním pohľadu. Naša metóda reprezentuje špecifickú oblasť testovania mobilných aplikácií, keďže testovanie neprebieha na mobilnom zariadení, ale pomocou emulácie na počítači. Je preto nevyhnutné obohatiť kroky bežnej používateľskej štúdie o nami definované kroky, ktoré umožnia uskutočnenie takejto štúdie. Tieto kroky definujú správnu prípravu testovacieho prostredia, ale taktiež aj overenie vhodnosti využitia testovania pomocou emulácie a po ukončení testovania následné overenie správnosti identifikovaných problémov.

Nami vytvorená metóda používateľského testu pozostáva z týchto krokov:

1. Zadefinovanie scenára testovania
2. **Overenie vhodnosti využitia emulátoru pre testovanie**
3. Vytvorenie protokolu testovania
4. Vytvorenie podkladov pre účastníkov testovania
5. **Príprava testovacieho prostredia**
6. Uskutočnenie pilotného testovania
7. Uskutočnenie používateľskej štúdie na základe vytvoreného protokolu
8. Analýza zozbieraných dát
9. **Vytvorenie usporiadaného zoznamu identifikovaných problémov a overenie správnosti identifikovaných problémov**

Kroky zvýraznené hrubým písmom reprezentujú kroky špecifické pre našu metódu, ktorými sa naša metóda odlišuje od bežných prístupov. V rámci nami definovaného používateľského testu účastníci vykonávajú úlohy v mobilnej aplikácii, ktorá je emulovaná na počítači.

Na základe uskutočnenej analýzy automatického vyhodnocovania používateľských štúdií, sme si stanovili požiadavku, že v rámci podpory automatického vyhodnotenia je potrebné detailne zaznamenávať informácie o interakcii používateľa. Počas vykonávania testu sa zaznamenáva nasledovná spätná väzba:

1. pohyb myši
2. kliknutia myši
3. pohľad používateľa

Navyše sme si stanovili požiadavku, že pri zaznamenávaní kliknutí myši je potrebné identifikovať aj na aký prvok/element aplikácie interakcia nastala. Takáto znalosť môže byť veľmi prínosná v rámci podpory automatického vyhodnotenia používateľského testu, najmä pri analýze sekvencií interakcií.

Počas analýzy rozličných typov mobilných aplikácií sme identifikovali štyri najčastejšie typy implementácií mobilných aplikácií a to vzhľadom na operačný

system zariadenia, pre ktoré sú určené. Týmto operačnými systémami sú OS Android²⁴, iOS²⁵ a Windows Phone²⁶. Štvrtým typom mobilných aplikácií sú hybridné aplikácie, ktoré vo väčšine prípadov predstavujú webové aplikácie prispôbené pre mobilné zariadenia. Mobilné aplikácie určené pre konkrétny mobilný operačný systém majú svoju špecifickú implementáciu, a preto musí byť pre každý z týchto operačných systémov implementovaný odlišný zaznamenávač (angl. logger), ktorý umožní zaznamenávanie horeuvedenej spätnej väzby. Nami implementované zaznamenávače bližšie opisujeme v kapitole 5.1.

4.1 Kroky používateľského testu

Zadefinovanie scenára testovania

Prvým krokom v rámci realizácie používateľského testu je zadefinovanie scenára testovania. Scenár reprezentuje presný opis jednotlivých úloh, ktoré budú účastníci testovania počas testovania vykonávať. Je potrebné mať presne špecifikované čo je cieľom realizácie danej používateľskej štúdie a na základe toho špecifikovať úlohy pre účastníkov testovania, aby sme na základe dát získaných počas vykonávania týchto úloh získali také výstupy, aké očakávame.

Overenie vhodnosti využitia emulátoru pre testovanie

Keďže nami navrhovaná metóda nie je vhodná na identifikáciu všetkých problémov s použiteľnosťou, ktoré by mohli nastať počas používania mobilnej aplikácie, nasledujúcim krokom po zadefinovaní scenára používateľskej štúdie je overenie vhodnosti využitia emulátoru pre testovanie.

V rámci kroku definície scenára sme si presne stanovili, akú funkcionálnosť mobilnej aplikácie plánujeme testovať v pripravovanej štúdií. Následne treba prejsť nami vytvorenými zoznamami potenciálnych problémov, ktoré sú uvedené v Kapitole 5.2 tejto práce a odhadnúť, aké typy problémov by sa mohli vyskytnúť počas plánovanej štúdie.

V prípade, že zistíme, že značná väčšina problémov, ktoré by mohli nastať pochádza práve z množiny problémov, ktoré nie sú vhodné pre testovanie pomocou našej metódy (Tabuľka 4), tak potom pre takúto štúdiu nie je vhodné využiť

²⁴<https://www.android.com/>

²⁵<https://www.apple.com/sk/ios/ios-10/>

²⁶<http://www.windowscentral.com/windows-phone>

testovanie pomocou emulátoru, ale je treba zväžiť testovanie pomocou reálneho mobilného zariadenia.

Vytvorenie protokolu testovania

Vytvorenie protokolu testovania je krok, ktorý môže byť vykonávaný paralelne s krokom definovania scenára testovania. Protokol testovania predstavuje ucelený obraz o plánovanej používateľskej štúdií. V rámci protokolu testovania je potrebné si zdefinovať, čo je cieľom realizovanej používateľskej štúdie a aké výstupy od danej štúdie očakávame. Následne je potrebné v rámci protokolu zdefinovať presný priebeh používateľskej štúdie.

Protokol by mal obsahovať scenár testovania, ktorý obsahuje úlohy zadané pre účastníkov testovania, ale taktiež by protokol mal obsahovať aj scenár vytvorený pre moderátora testovania. Tento scenár by mal obsahovať detailný opis všetkých úloh, ktoré má moderátor vykonať v rámci prípravy používateľskej štúdie a taktiež počas priebehu samotnej štúdie a po ukončení štúdie.

V rámci vytvárania protokolu testovania je potrebné zdefinovať:

- *Ciele realizovanej štúdie* - aké sú ciele realizácie danej štúdie a vyhodnotením akých metrick chceme tieto ciele dosiahnuť.
- *Špecifické požiadavky na účastníkov testovania* - napríklad vek, pohlavie alebo vzdelanie a zamestnanie.
- *Potrebný počet účastníkov* - závisí od požiadaviek konkrétnej štúdie. Keďže realizujeme kvantitatívnu štúdiu, našim zámerom je získanie štatisticky signifikantných výsledkov, na základe čoho sa odporúča realizácia takejto štúdie aspoň s 20 účastníkmi testovania²⁷. Avšak ako už bolo v našej práci uvedené, počet účastníkov testovania závisí od povahy konkrétnej štúdie a taktiež aj od toho, aké presné výsledky očakávame.
- *Požiadavky na mobilnú aplikáciu* - rôzne požiadavky špecifické pre testovanú mobilnú aplikáciu, vzhľadom na emuláciu na počítači. Napríklad požadovaná zobrazovacia veľkosť v rámci emulátoru alebo zdefinovanie spôsobu zobrazenia systémových tlačidiel.

²⁷<https://www.nngroup.com/articles/quantitative-studies-how-many-users/>

Ukážka protokolu, ktorý sme vypracovali pri realizácii používateľskej štúdie v rámci tejto práce sa nachádza v Prílohách A a B.

Vytvorenie podkladov pre účastníkov testovania

Keďže realizovaná štúdia predstavuje používateľské testovanie s väčším počtom účastníkov, ktorí vykonávajú úlohy paralelne, nie je možné aby moderátor testovania komunikoval s každým účastníkom samostatne. Preto je nevyhnutné pripraviť pre účastníkov testovania všetky podklady, ktoré by mohli počas testovania potrebovať. Najdôležitejším podkladom pre účastníkov testovania je zoznam úloh, ktoré majú vykonávať a prípadné dodatočné inštrukcie. Aké všetky podklady je potrebné pre účastníkov testovania pripraviť, je závislé od špecifickej štúdie.

Príprava testovacieho prostredia

Pred realizáciou používateľskej štúdie pripravíme a nastavíme prostredie, v ktorom budú účastníci testovania vykonávať úlohy. Keďže ide o hromadné testovanie mobilnej aplikácie pomocou emulácie na počítači s využitím sledovania pohľadu, testovanie bude prebiehať na viacerých počítačoch súčasne. Proces, prípravy testovacieho prostredia pozostáva z viacerých kôrkov:

1. *Inštalácia nástrojov pre získavanie spätnej väzby od používateľa* - implementovanie správneho zaznamenávača do mobilnej aplikácie, vzhľadom na typ danej aplikácie (Príloha C), alebo inštalácia softvéru pre sledovanie pohľadu.
2. *Inštalácia prostredia PC* - inštalácia emulátoru na každý počítač na ktorom bude prebiehať testovanie²⁸ a inštalácia mobilnej aplikácie do emulátoru²⁹.
3. *Nastavenie prostredia* - nastavenie softvéru pre zaznamenávanie pohľadu a nastavenie emulátoru³⁰ (napríklad požadovaná zobrazovacia veľkosť, alebo zobrazenie systémových tlačidiel)

Uskutočnenie pilotného testovania

Po tom ako je vytvorený protokol testovania a testovacie prostredie je pripravené a nastavené pre potreby štúdie, pred samotnou hromadnou používateľskou štúdiou

²⁸https://docs.genymotion.com/Content/01_Get_Started/Installation.htm

²⁹https://docs.genymotion.com/Content/01_Get_Started/Basic_steps.htm

³⁰https://docs.genymotion.com/Content/03_Virtual_Devices/Managing_virtual_devices/Configuring_a_virtual_device.htm

uskutočníme pilotné testovanie. V rámci pilotného testovania uskutočníme používateľskú štúdiu podľa zadaného protokolu s malou vzorkou používateľov (odporúčajú sa minimálne 2 používatelia, aby sme overili paralelné testovanie). Po uskutočnení pilotného testovania analyzujeme zistenia identifikované počas tohto testovania a v prípade potreby upravíme protokol testovania, aby sme odstránili identifikované nedostatky.

Uskutočnenie používateľskej štúdie na základe vytvoreného protokolu

Po uskutočnení pilotného testovania a prípadnom zapracovaní zmien do protokolu testovania, nasleduje uskutočnenie hromadnej používateľskej štúdie podľa zadaného protokolu s požadovaným počtom účastníkov testovania.

Analýza zozbieraných dát

Po ukončení používateľskej štúdie je potrebné zozbierané dáta analyzovať za účelom identifikovania problémov spojených s použiteľnosťou testovanej aplikácie. Tento krok obsahuje aj vyhodnotenie zadaných metrík.

Vytvorenie usporiadaného zoznamu identifikovaných problémov a overenie správnosti identifikovaných problémov

Ako výstup z predošlého kroku máme k dispozícii zoznam identifikovaných problémov, ktoré sa odhalili počas realizovanej štúdie. Keďže sme štúdiu realizovali pomocou emulácie na počítači, kde je interakcia so zariadením značne odlišná ako v prípade mobilného zariadenia a aj samotný mobilný kontext je silno obmedzený, je potrebné overiť správnosť identifikovaných problémov a odstrániť tie problémy, ktoré nie sú validné.

Je potrebné preskúmať celý zoznam identifikovaných problémov a pokúsiť sa každý problém namapovať na problém v rámci nami vytvorených zoznamov potenciálnych problémov, ktoré sú uvedené v Kapitole 5.2. V prípade, že problém namapujeme na problém z Tabuľky 3 (problémy identifikovateľné pomocou emulácie), tak môžeme tento problém považovať za validný. Avšak pokiaľ problém namapujeme na problém z Tabuľky 4 (problémy neidentifikovateľné pomocou emulácie), tak tento problém nie je validný a musíme ho odstrániť zo zoznamu identifikovaných problémov.

Príkladom takéhoto nevalidného problému môže byť: "stránka/aplikácia na interakciu reaguje veľ mi pomaly". Tento problém bol síce identifikovaný na základe dát z uskutočnenej používateľskej štúdie, avšak rýchlosť aplikácie a taktiež aj spätnej väzby aplikácie môže byť značne ovplyvnená prostredím emulátoru. Z toho dôvodu nemôžeme tento problém považovať za validný a je potrebné ho odstrániť zo zoznamu identifikovaných problémov.

Pokiaľ nastane taký prípad, kedy problém identifikovaný počas realizovanej štúdie nie je možné namapovať na žiaden problém uvedený v Kapitole 5.2, je potrebné svojpomocne na základe skúseností určiť, či daný problém je validný, alebo mohol byť spôsobený kontextom emulátoru.

Po preskúmaní celého zoznamu identifikovaných problémov a odstránení nevalidných problémov, máme k dispozícii kompletný zoznam problémov s použiteľnosťou, ktoré boli identifikované v rámci realizovanej štúdie. Posledným krokom je usporiadanie identifikovaných problémov na základe ich závažnosti. Pri určovaní závažnosti problému je vhodné brať do úvahy frekvenciu výskytu daného problému v kombinácii s jednou z využívaných škál pre určovanie závažnosti problému³¹.

Jakob Nielsen využíva 5 bodovú stupnicu reprezentujúcu závažnosť daného problému³². Stupnica je nasledovaná (v zátvorkách sú uvedené hodnoty reprezentujúce závažnosť daného stupňa):

1. Nesúhlasím s tým, že ide o problém s použiteľnosťou (0)
2. Kozmetický problém (potrebná oprava iba ak zvýši čas) (1)
3. Menší problém (oprava by mala mať nízku prioritu) (2)
4. Väčší problém (oprava by mala mať vysokú prioritu) (3)
5. Katastrofa z hľadiska použiteľnosti (maximálna priorita) (4)

V prípade, že by sme závažnosť chceli vyjadriť ako kombináciu frekvencie problému a vyššie uvedenej stupnice, môžeme celkovú závažnosť problému vyjadriť ako súčin frekvencie výskytu daného problému a hodnoty závažnosti pridelenej z vyššie uvedenej stupnice [4]. Napríklad ak by sme danému problému pridelili stupeň závažnosti "Menší problém", ktorý nadobúda hodnotu '2' a tento problém sa

³¹<http://www.measuringu.com/blog/rating-severity.php>

³²<https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/>

vyskytol v prípade 10 účastníkov testovania, tak celková závažnosť problému by bola rovná hodnote 20 ($2 \cdot 10$).

4.2 Diskusia a zhodnotenie

Navrhli sme metódu kvantitatívneho testovania mobilných aplikácií s využitím sledovania pohľadu, pomocou emulácie na počítači, pričom sa špecificky zameriavame na hromadné používateľské štúdie. Naša metóda v porovnaní s existujúcimi prístupmi umožňuje kvantitatívne hromadné testovanie mobilných aplikácií, vďaka čomu umožňuje realizáciu používateľských štúdií mobilných aplikácií v značne väčšom rozsahu. Zároveň vďaka využitiu zariadenia na sledovanie pohľadu určeného pre počítače, môžeme vďaka našej metóde získať detailnejšie informácie o pohľade používateľov, ako v prípade využitia iných techník akými sú napríklad stojany pre mobilné zariadenia so sledovačmi pohľadu.

Cieľom tejto práce bol návrh metódy pre kvantitatívne testovanie mobilných aplikácií, keďže s využitím súčasných prístupov je organizácia kvantitatívnych štúdií na mobilných zariadeniach časovo veľmi náročná. Táto metóda nie je striktno obmedzená iba na kvantitatívne testovanie, ale taktiež by mohla byť aplikovaná aj pre kvalitatívne testovanie.

V rámci návrhu našej metódy, sme definovali kroky používateľského testu, ktoré definujú priebeh používateľskej štúdie realizovanej pomocou našej metódy. Tieto kroky predstavujú obohatenie bežne využívaných postupov o kroky špecifické pre našu metódu a definujú správnu prípravu testovacieho prostredia, ale taktiež aj overenie vhodnosti využitia testovania pomocou emulácie a po ukončení testovania následné overenie správnosti identifikovaných problémov. Pre jednotlivé kroky uvádzame detailný opis činností, ktoré je potrebné v rámci daného kroku realizovať.

Pri návrhu metódy sme sa zamerali aj na možnosti podpory automatizácie vyhodnocovania používateľských štúdií. V rámci podpory automatizácie sme uskutočnili analýzu automatického vyhodnocovania používateľských štúdií a stanovili sme požiadavky pre zaznamenávače, ktorých cieľom je zaznamenávanie aktivity používateľov. Zaznamenávače okrem kliknutí používateľov, identifikujú aj element, na ktorý nastala interakcia. Takáto znalosť má vysokú pridanú hodnotu a môže byť veľmi prínosnou z hľadiska podpory automatizácie vyhodnocovania používateľských štúdií, pretože nám umožní automaticky detegovať, či na daný objekt/element nastala interakcia. Vďaka umožneniu takejto automatickej analýzy sa následne

zjednoduší automatizácia vyhodnocovania.

Počas analýzy rozličných typov mobilných aplikácií sme identifikovali štyri najčastejšie typy implementácií mobilných aplikácií a to vzhľadom na operačný systém zariadenia pre ktoré sú určené. Miernym obmedzením je, že mobilné aplikácie určené pre konkrétny mobilný operačný systém majú svoju špecifickú implementáciu, a preto musí byť pre každý z týchto operačných systémov implementovaný odlišný zaznamenávač.

Navrhnutá metóda má aj svoje ohraničenia. Keďže ide o testovanie mobilných aplikácií pomocou emulácie na počítači, je zjavné že je do vysokej miery odstránený kontext mobilného zariadenia. Na základe tohto faktoru je jasné, že naša metóda nedokáže odhaliť všetky problémy spojené s použiteľnosťou mobilných aplikácií, ale iba takú ich časť, ktorá nie je úzko spätá s kontextom mobilného zariadenia. V rámci našej práce sme sa zamerali aj na riešenie tejto problematiky a vytvorili sme zoznam generických problémov s použiteľnosťou, u ktorých predpokladáme, že ich je možné odhaliť s využitím emulátoru (kapitola 5.2).

V tejto práci sme navrhli metódu testovania mobilných aplikácií pomocou emulácie na počítači, ktorá umožní hromadné testovanie mobilných aplikácií a tým zníži dĺžku trvania týchto štúdií. Avšak tak ako sme uviedli, nebudeme schopní pomocou tejto metódy odhaliť všetky problémy s použiteľnosťou. Otvára sa teda možnosť kombinácie našej metódy s bežnými prístupmi. Tejto možnosti sa v našej práci nevenujeme, avšak nechávame ju otvorenú ako možnosť pre ďalší rozvoj našej práce.

V rámci kombinácie metód navrhujeme rozloženie, kedy by časť účastníkov testovania vykonávala testovanie mobilnej aplikácie pomocou emulácie na počítači a zároveň by menšia časť účastníkov vykonala testovanie na mobilnom zariadení. Takouto kombináciou našej metódy s bežnými prístupmi, by bolo možné ušetriť čas pomocou hromadnej emulácie na počítači a zároveň aj odhaliť problémy, ktoré môžu byť špecifické pre kontext mobilného zariadenia.

5 Realizácia a overenie metódy

Prvým krokom v rámci realizácie našej metódy, bol výber vhodného emulátoru pre realizáciu hromadných používateľských štúdií. Analyzovali sme dostupné emulátory, ktoré slúžia na emuláciu operačného systému Android, keďže sme sa rozhodli na tento systém zamerať v našej práci, vzhľadom na to, že ide o najpoužívanejší mobilný operačný systém³³. V rámci testovania viacerých dostupných riešení sme zvažovali viacero faktorov a to napríklad: (1) dostatočný výkon emulátoru aby nebola práca s rozhraní nijako obmedzujúca, (2) podpora rozličných zobrazovacích veľkostí alebo (3) možnosť neplatenej licencie. Po otestovaní viacerých riešení a zvážení uvedených faktorov sme zvolili emulátor Genymotion ako emulátor, ktorý budeme využívať v našej práci.

5.1 Podpora automatizácie vyhodnotenia

V rámci našej metódy účastníci používateľského testu vykonávajú úlohy v mobilnej aplikácii, ktorá je emulovaná na počítači. V rámci podpory automatického vyhodnotenia sa počas vykonávania testu sa zaznamenáva nasledovná spätná väzba:

1. pohyb myši
2. kliknutia myši
3. pohľad používateľa

Pre umožnenie zaznamenávania uvedenej spätnej väzby sme implementovali dva typy zaznamenávačov:

1. *Zaznamenávač pre natívne aplikácie* - tento zaznamenávač predstavuje modul/knižnicu pre OS Android, ktorú je možné jednoduchým spôsobom vložiť do akejkoľvek natívnej aplikácie určenej pre operačný systém Android. Zaznamenávač zaznamenáva všetky udalosti (angl. events) a zaznamenané dáta odosiela na server. Ako server pre odosielanie dát sme využili voľne dostupnú službu Keen.IO³⁴, ktorá slúži ako jednoduché API pre prijímanie eventov.

³³<https://www.statista.com/statistics/272698/global-market-share-held-by-mobile-operating-systems-since-2009/>

³⁴<https://keen.io/>

Zaznamenaný event môže byť dvoch typov a to *kliknutie* alebo *scroll*. V rámci zaznamenaného eventu zaznamenávač vždy určí súradnice v rámci aplikácie na ktorých daný event nastal. Navyše pre každý event zaznamenávač určí na akej obrazovke daný event nastal a taktiež pokiaľ je to možné, tak určí konkrétny element aplikácie, na ktorom event nastal (napríklad konkrétne tlačidlo alebo položka menu).

2. *Zaznamenávač pre webové aplikácie* - tento zaznamenávač predstavuje knižnicu implementovanú v jazyku Javascript, ktorá slúži na zaznamenávanie eventov v rámci webových aplikácií. Zaznamenávač je taktiež možné veľmi ľahko vložiť do akéhokoľvek projektu a zaznamenávač následne zaznamenáva a odosiela eventy na server identicky ako v prípade zaznamenávača pre natívne aplikácie. Tento zaznamenávač nie je schopný určiť, na ktorej obrazovke event nastal, ale na druhú stranu je vždy schopný unikátne identifikovať element aplikácie, na ktorom event nastal.

Vďaka využitiu týchto zaznamenávačov je možné zbierať spätnú väzbu počas vykonávania používateľskej štúdie, pričom zaznamenané dáta môžu neskôr byť využité pre automatické vyhodnotenie uskutočnenej používateľskej štúdie.

5.2 Zoznam identifikovaných potenciálnych problémov s použiteľnosťou mobilných aplikácií

Tak ako sme už uviedli v predošlých kapitolách interakcia s aplikáciou emulovanou na počítači je značne odlišná od interakcie s reálnym mobilným zariadením. Pomocou takejto emulácie sa značne eliminuje mobilný kontext, ktorý je úzko spätý s mobilnými aplikáciami. Je preto zjavné, že pri používateľskej štúdii mobilnej aplikácie s využitím emulátoru nie je možné odhaliť všetky problémy spojené s použiteľnosťou danej aplikácie.

Na základe tohto faktoru sme vykonali rozsiahlu analýzu, v rámci ktorej sme vytvorili dva zoznamy potenciálnych problémov s použiteľnosťou, ktoré sa môžu objaviť počas testovania mobilnej aplikácie. Pri vytváraní týchto zoznamov sme vychádzali z dostupnej literatúry a internetových zdrojov, ktoré sú zamerané na problematiku

testovania použiteľnosti^{35 36 37} [20], [35]. Taktiež sme analyzovali viaceré používateľské štúdie zamerané na mobilné aplikácie, ktoré boli realizované u nás na fakulte a mali sme prístup k výslednému zoznamu identifikovaných problémov. V rámci týchto zoznamov sme sa snažili jednotlivé špecifické problémy generalizovať, aby boli schopné reprezentovať viacero podobných problémov.

Výsledkom tejto analýzy sú dva zoznamy identifikovaných potenciálnych problémov s použiteľnosťou mobilnej aplikácie. V tabuľke 3 sú uvedené problémy, ktoré predpokladáme, že je možné odhaliť pomocou emulácie na počítači, zatiaľ čo tabuľka 4 obsahuje problémy, ktoré nie je možné odhaliť pomocou emulácie. Problémy uvedené v nasledujúcich tabuľkách nie sú na rovnakej úrovni granularity, keďže sme sa snažili o identifikáciu čo najväčšieho množstva potenciálnych problémov. Niektoré problémy uvedené v tabuľkách predstavujú zoskupenia/generalizáciu viacerých možných problémov a iné zas reprezentujú špecifické problémy.

Tabuľka 3: *Problémy identifikovateľné za pomoci emulácie*

Názov	Opis
Absencia navigačných prvkov	<ul style="list-style-type: none"> • používateľ sa nevie dostať z aplikácie bez hardvérového tlačidla • používateľ sa dostal na obrazovku, kde chýba možnosť vrátenia sa na hlavnú obrazovku • používateľ nevie, kde v aplikácii sa zrovna nachádza • používateľ nevie nájsť menu (navigáciu), keď je zoscrollovaný nižšie na obrazovke
Dlhé formuláre	formulár obsahuje príliš veľa prvkov a preto je pre používateľa a odpudzujúci
Informatívnosť obsahu	používateľ nevidí na obrazovke všetky potrebné informácie (alebo napr. obrázok sa nezmesť na obrazovku)
Nájditeľnosť (angl. findability)	jednoduché nájdenie obsahu, alebo funkcionality, o ktorej používatelia predpokladajú že je prítomná
Nejednoznačnosť identity	nepochopenie zamerania webu/aplikácie počas iniciálnej interakcie

³⁵<https://www.nngroup.com/>

³⁶<https://www.usability.gov/>

³⁷<http://www.leafdigital.com/class/lessons/usability/3.html>

Nejednoznačnosť navigačných prvkov	<ul style="list-style-type: none"> • navigačné prvky nie sú pre nových používateľov úplne jasné pretože obsahujú nejednoznačné pomenovania (používateľ nevie, kam ho navigačný prvok zavedie) • navigačné prvky nie sú pre nových používateľov úplne jasné, pretože sú reprezentované pomocou neznámych ikon bez dodatočného textového vysvetlenia
Nepochopenie funkcionality	<ul style="list-style-type: none"> • používateľ má nesprávne predpoklady o správaní aplikácie • používateľ v aplikácii očakáva funkcionality, ktorá tam nie je
Nepochopenie obsahu	<ul style="list-style-type: none"> • používateľ nesprávne chápe obsah, pretože v aplikácii sú používané doménovo špecifické termíny • používateľ nesprávne chápe obsah, pretože texty sú nevhodne zvolené • používateľ zle interpretuje grafické prvky aplikácie (napr. graf bez legendy a pod.)
Nesprávne využitie komponentov rozhrania	využitie komponentov rozhrania pre nesprávnu/neočakávanú funkcionality
Nevhodne označené vstupné elementy	vstup nie je dobre označený a používateľ nevie, čo má byť obsahom daného pol'a
Nevhodne umiestnená funkcionality	používateľ niekde očakával určitú funkcionality, ale reálne sa nachádzala inde
Nevhodne umiestnený obsah	používateľ očakáva iný obsah (očakával že na stránke X nájde obsah Y ale nebolo tomu tak)
Nevhodne vizualizované interaktívne elementy	<ul style="list-style-type: none"> • používateľ nevie o tom, že klikateľný element je klikateľný • používateľ nevie určiť, prečo v danom okamihu link/tlačidlo nie je klikateľné (je deaktivovaný)
Nevhodne vizualizované vstupné elementy	vstup nezobrazuje celý zadaný obsah, a preto má používateľ problém skontrolovať zadanú hodnotu

Nevhodne vizualizované navigačné prvky	používateľ nevidí rozdiel medzi navštíveným a nenavštíveným odkazom
Nevhodná spätná väzba komponentu rozhrania	komponent rozhrania nedáva dostatočnú spätnú väzbu (napr. používateľ nevie či potvrdil vyhľadávanie a pod.)
Objaviteľnosť (angl. discoverability)	objavenie novej funkcionality, o ktorej používatelia doteraz nevedeli
Preddefinovaný formát vstupu	vstup vyžaduje špecifické formátovanie vstupu, ktoré je pre používateľa ťažko pochopiteľné
Preplnenie navigačnými prvkami	stránka/aplikácia obsahuje nadmerné množstvo navigačných prvkov, a preto je problematické identifikovať správny prvok pre požadovanú akciu
Problémy spojené so zadávaním vstupu	<ul style="list-style-type: none"> pri zadávaní vstupu nie je klávesnica prispôbená očakávanej hodnote, a preto má používateľ problém zadať správny vstup, alebo ho zadáva neefektívne, klávesnica sa automaticky neschová
Problémy spojené s personalizáciou	aplikácia si nepamätá personalizovaný stav predchádzajúcej návštevy
Slabá výraznosť dôležitého obsahu	obsah, ktorý je pre používateľa veľmi podstatný je málo výrazný, a preto si ho používateľ nevšimol (notifikácia, chybová hláška a pod.)

Tabuľka 4: Problémy neidentifikovateľné za pomoci emulácie

Názov	Opis
Malé klikateľné plochy	prvky, alebo plochy ktoré sú klikateľné, sú príliš malé a používateľ má problém s nimi interagovať
Nedostupnosť (angl. inaccessibility)	slabá podpora používania pre ľudí so zdravotnými problémami
Nesprávna voľba farieb	farby pozadia a textu sú nevhodne zvolené, a preto je text ťažko čitateľný
Nesprávna voľba veľkosti písma	veľkosť písma je príliš malá/veľká, a preto je text ťažko čitateľný
Ovládateľnosť vnorených navigácií	používateľ má problém ovládať vnorené navigácie (submenu / dropdown)
Preplnenie textom	obrazovka obsahuje prílišné množstvo textu, a preto je veľmi ťažko čitateľný

Problémy spojené s rýchlosťou	<ul style="list-style-type: none"> • stránka/aplikácia sa dlho načítava, pretože obsahuje príliš veľké obrázky, príliš veľký obsah a pod. • stránka/aplikácia na interakciu reaguje veľmi pomaly
Problémy spojené s ergonomiou rozhrania	držanie zariadenia a jeho ovládanie (palec vs. iné prsty) - napr. používateľ nedočiahne na prvok rozhrania, alebo si pri používaní aplikácie prekryva časť aplikácie
Problémy vyplývajúce z funkcionality vyžadujúcej kameru	problémy spojené s používaním kamery
Problémy vyplývajúce z funkcionality vyžadujúcej aby bolo zariadenie v pohybe	problémy vyplývajúce z kontextu, že zariadenie je v pohybe
Problémy vyplývajúce z funkcionality využívajúcej telefonovanie	problémy spojené s telefonovaním
Problémy vyplývajúce z funkcionality vyžadujúcej multitouchové gestá	problémy vyplývajúce z využitia multitouchových gest
Problémy vyplývajúce z funkcionality vyžadujúcej špecifické senzory	problémy spojené s využívaním špecifických senzorov
Zlé umiestnenie klikateľných plôch	- prvky alebo plochy ktoré sú klikateľné sú príliš blízko pri sebe

Samozrejme uvedené tabuľky potenciálnych problémov neobsahujú všetky možné typy problémov, ktoré by počas testovania mohli nastať a sú otvorené pre doplnenie ďalších problémov. Tabuľky obsahujú iba tie typy problémov, ktoré sa nám podarilo počas našej analýzy identifikovať.

5.3 Overenie metódy

Existujú špecifické problémy spojené s použiteľnosťou mobilných aplikácií, ktoré môžu vyplývať priamo z kontextu mobilného zariadenia a tieto problémy neodhalíme pomocou emulácie na počítači. Naším cieľom nie je pomocou emulácie identifikovať všetky problémy, ale umožniť efektívnejšie hromadné testovanie mobilných aplikácií, počas ktorého odhalíme mnohé problémy s použiteľnosťou danej aplikácie.

Pre overenie nami navrhnutej metódy používateľského testu sme uskutočnili dve používateľské štúdie, ktorých cieľom bolo dokázať možnosť identifikovania problémov spojených s použiteľnosťou mobilnej aplikácie pomocou emulácie mobilnej aplikácie na počítači s využitím sledovania pohľadu a overenie nami definovaného zoznamu potenciálnych problémov u ktorých predpokladáme, že ich je možné odhaliť pomocou emulácie.

V prvej štúdii, ktorú sme realizovali účastníci pracovali s identickou mobilnou aplikáciou v rámci mobilného zariadenia a emulátoru. Hlavným cieľom tejto používateľskej štúdie bolo overenie základnej hypotézy, že existujú problémy s použiteľnosťou mobilnej aplikácie, ktoré je možné odhaliť pomocou emulácie na počítači.

V druhej realizovanej štúdii účastníci pracovali s dvoma rozličnými mobilnými aplikáciami, pričom s jednou pracovali priamo na mobilnom zariadení a s druhou pomocou emulácie na počítači. Scenár tejto štúdie bol navrhnutý tak, aby sa vyskytli konkrétne problémy s použiteľnosťou daných aplikácií. Hlavným cieľom tejto štúdie bolo overenie nami definovaného zoznamu potenciálnych problémov u ktorých predpokladáme, že ich je možné odhaliť pomocou emulácie.

5.3.1 Štúdia 1: Overenie základnej hypotézy

Protokol testovania, ktorý obsahuje podrobný opis a vyhodnotenie realizovanej štúdie sa nachádza v Prílohe A.

Ciele štúdie

Hlavným cieľom tejto používateľskej štúdie bolo overenie základnej hypotézy, že existujú problémy s použiteľnosťou mobilnej aplikácie, ktoré je možné odhaliť pomocou emulácie na počítači.

Prostredie pre realizáciu štúdie

V rámci návrhu tejto používateľskej štúdie sme realizovali analýzu ktorej cieľom bolo nájsť bežne používanú mobilnú aplikáciu, ktorú by sme využili ako testovaciu aplikáciu. Hlavnými požiadavkami pre hľadajúcu mobilnú aplikáciu bolo aby daná aplikácia: (1) poskytovala dostatok obsahu a funkcionality, aby bolo možné

vytvoriť primerane komplikovaný scenár, (2) patrila medzi bežne využívané aplikácie, (3) obsahovala problémy s použiteľnosťou. Na základe definovaných kritérií sme ako testovaciu aplikáciu zvolili mobilnú aplikáciu Česko-Slovenskej filmovej databázy (CSFD)³⁸.

Úlohou každého účastníka testovania bolo vykonať 7 krátkych úloh v mobilnej aplikácii CSFD. Zadefinované úlohy boli rozdelené na dve sady úloh, pričom jedna sada vždy obsahovala úlohu prihlásenia do aplikácie a ďalšie tri úlohy. Účastníci testovania boli rozdelení do dvoch skupín, pričom jedna skupina vykonala prvú sadu úloh na reálnom mobilnom zariadení a druhú sadu na počítači, zatiaľ čo druhá skupina účastníkov vykonala prvú sadu úloh na počítači a druhú sadu na reálnom mobilnom zariadení. Takéto rozdelenie úloh a účastníkov testovania do skupín nám umožnilo porovnanie problémov identifikovaných na mobilnom zariadení s problémami identifikovanými pomocou emulácie na počítači.

Účastníci

Tento štúdie sa zúčastnilo 18 účastníkov, pričom išlo o študentov našej fakulty z rozličných ročníkov, ktorí boli testovaní individuálne. Hlavným dôvodom prečo sme sa rozhodli pre individuálne testovanie bol fakt, že sme potrebovali vyššiu mieru interakcie s účastníkmi testovania, aby sme ich mohli detailne sledovať a tak lepšie pochopiť ako ľudia pracujú s mobilným zariadením a zároveň ako sú schopní pracovať s mobilnou aplikáciou emulovanou na počítači.

Výsledky a zhodnotenie

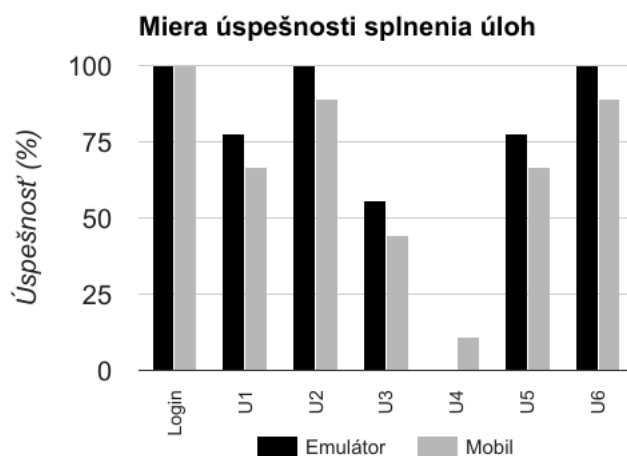
V rámci vyhodnotenia realizovanej používateľskej štúdie sme sledovali viacero metrík. Jednou z metrík ktoré sme sledovali je úspešnosť splnenia úloh. Táto metrika nám hovorí o tom, či daný účastník testovania splnil alebo nespľnil danú úlohu. My sme vyhodnotili celkovú úspešnosť všetkých účastníkov testovania vzhľadom na danú úlohu (teda koľko percent účastníkov testovania úspešne splnilo danú úlohu).

Na Obrázku 6 môžeme vidieť výsledky vyhodnotenia úspešnosti splnenia jednotlivých úloh, pričom rozlišujeme, či bola úloha vykonaná na mobilnom zariadení alebo pomocou emulácie na počítači, aby sme boli schopní porovnať, či typ zariadenia na ktorom bola úloha vykonávaná ovplyvňuje úspešnosť úloh. Experimentu

³⁸<https://play.google.com/store/apps/details?id=cz.csfd.csfdroid>

sa zúčastnilo 18 účastníkov testovania, čiže každú úlohu vykonalo 9 účastníkov na mobilnom zariadení a 9 účastníkov pomocou emulácie na počítači.

Vidíme, že medzi úspešnosťami úloh vzhľadom na typ zariadenia nie je žiaden výrazný rozdiel. Vo väčšine prípadov je úspešnosť vykonania úloh na počítači vyššia o jedného účastníka, čo si vysvetlíjeme tým, že práca s emulátorom bola pre účastníkov testovania menej prirodzená a aj odozvy zariadenia boli mierne pomalšie ako v prípade mobilného zariadenia. Kvôli tejto skutočnosti účastníci testovania pracovali na počítači trochu pomalšie a teda venovali úlohám viac pozornosti.



Obr. 6: Miera úspešnosti splnenia úloh

V rámci používateľskej štúdie sa podarilo identifikovať 19 unikátnych problémov z čoho 17 problémov bolo identifikovaných pomocou emulácie na počítači a 16 problémov pri interakcii s mobilným zariadením. Z celkovej množiny 19 identifikovaných problémov bolo 14 problémov identifikovaných na oboch typoch zariadenia. Pomocou emulácie na počítači sa podarilo odhaliť 87.5% problémov, ktoré boli odhalené pomocou mobilného zariadenia.

Vďaka tomuto experimentu sa nám podarilo potvrdiť našu základnú hypotézu ktorou bolo, že pomocou emulácie mobilných aplikácií na počítači dokážeme odhaliť určité problémy spojené s ich použiteľnosťou. V rámci vyhodnotenia uskutočnenej používateľskej štúdie sme identifikovali niekoľko problémov spojených s použiteľnosťou danej aplikácie a väčšinu problémov, ktoré sa podarilo identifikovať pomocou realizácie na mobilnom zariadení sa podarilo identifikovať aj pomocou

emulácie na počítači.

Navyše sa nám pomocou tohto experimentu podarilo ukázať, že úspešnosti jednotlivých úloh, účinnosť účastníkov testovania a dĺžky trvania jednotlivých úloh nie sú nijako výrazne závislé od typu zariadenia na ktorom je úloha vykonávaná.

5.3.2 Štúdia 2: Overenie identifikovaného zoznamu potenciálnych problémov

Protokol testovania, ktorý obsahuje podrobný opis realizovanej štúdie a kompletné vyhodnotenie sa nachádza v Prílohe B.

Ciele štúdie

V rámci našej práce sme vytvorili zoznam potenciálnych problémov s použiteľnosťou mobilných aplikácií. Tento zoznam sa nachádza v Kapitole 5.2. Zoznam je tvorený z dvoch častí a to problémov, ktoré predpokladáme, že je možné odhaliť pomocou emulácie a problémov, ktoré nie je možné odhaliť pomocou emulácie.

Problémy, ktoré sú uvedené ako tie, ktoré nie je možné odhaliť pomocou emulácie sú jednoznačné a nie je potrebné ich experimentálne overovať. Ide totiž o problémy priamo vyplývajúce z kontextu mobilného zariadenia, alebo naopak kontextu emulátoru. Príkladom sú problémy vyplývajúce z funkcionality vyžadujúcej kameru, alebo špeciálne senzory. Keďže takúto funkcionality v rámci emulátoru nemôžeme využiť, tak je zjavné, že takéto problémy nemôžu byť identifikované pomocou emulácie. Druhým príkladom, môžu byť problémy spojené s rýchlosťou. Je zjavné, že emulátor si vyžaduje značný výkon počítača na ktorom beží a preto aj plynulosť behu aplikácie je silno ovplyvnená týmto faktorom. Na základe toho, pokiaľ by sme počas testovania aplikácie pomocou emulácie zistili, že aplikácia pracuje pomaly, nemohli by sme sa na toto zistenie plne spoľahnúť, pretože aplikácia môže v rámci emulátora bežať pomaly, avšak v rámci mobilného zariadenia môže fungovať bezproblémovo. Pre overenie takéhoto problému by bolo potrebné overenie na mobilnom zariadení.

Hlavným cieľom realizácie tejto používateľskej štúdie bolo overenie vhodnosti zaradenia čo najväčšieho množstva potenciálnych problémov do zoznamu problémov, ktoré je možné identifikovať pomocou emulácie. Navyše cieľom našej práce je

vytvorenie metódy, ktorá bude v sebe zahŕňať podporu automatického vyhodnotenia. V rámci podpory automatického vyhodnotenia sme vytvorili dva zaznamenávače slúžiace na zaznamenávanie aktivity používateľov pri práci s mobilnou aplikáciou. Druhým cieľom realizácie tejto štúdie bolo zároveň aj overenie správnosti fungovania implementovaných zaznamenávačov.

Prostredie pre realizáciu štúdie

Pri návrhu tejto používateľskej štúdie sme brali do úvahy fakt, že potrebujeme, aby účastníci testovania pracovali s mobilnou aplikáciou na mobilnom zariadení a taktiež aj s mobilnou aplikáciou na emulátore, aby sme následne mohli porovnať identifikované problémy v oboch prostrediach. Na základe tohto faktoru sme sa rozhodli nechať účastníkov testovania pracovať s dvoma mobilnými aplikáciami, pričom s jednou pracovali na svojom mobilnom zariadení a s druhou pomocou emulátoru. Vďaka tomu sme eliminovali faktor učenia sa, kedy by sa používatelia v rámci jedného prostredia oboznámili s danou aplikáciou a naučili by sa ako ju používať a následne pri práci v druhom prostredí s identickou aplikáciou by výsledky mohli byť značne ovplyvnené ich nadobudnutou znalosťou danej aplikácie.

Keďže sme chceli, aby účastníci pracovali v dvoch prostrediach s dvoma odlišnými aplikáciami a potrebovali sme problémy identifikované v rámci danej aplikácie pri práci na mobilnom zariadení porovnávať s problémami identifikovanými pomocou emulácie, rozhodli sme sa pre organizáciu danej používateľskej štúdie v dvoch behoch.

V rámci prvého behu pracovali účastníci s mobilnou aplikáciou číslo 1 v rámci emulátoru na počítači a následne s mobilnou aplikáciou číslo 2 na svojom mobilnom zariadení. V rámci druhého behu boli prostredia pre dané aplikácie vymenené. Čiže účastníci druhého behu pracovali s aplikáciou číslo 2 v rámci emulácie na počítači a s aplikáciou číslo 1 na svojom mobilnom zariadení. Vďaka takémuto spôsobu realizácie používateľskej štúdie sme získali dáta vhodné pre vyhodnotenie, či daný problém s použiteľnosťou je, alebo nie je vhodný pre identifikáciu pomocou emulácie.

Na základe uvedených potrieb štúdie sme si stanovili nasledovné požiadavky pre nájdenie vhodných dvoch aplikácií na testovanie:

1. Obe aplikácie musia poskytovať dostatok obsahu a funkcionality, aby bolo možné vytvoriť primerane komplikovaný scenár testovania.
2. Obe aplikácie musia byť vhodné pre identifikovanie čo najväčšieho množstva potenciálnych problémov s použiteľnosťou.
3. Obe aplikácie musia mať voľne dostupný zdrojový kód, aby sme mohli dané aplikácie mierne modifikovať. Do oboch aplikácií musíme byť schopní vložiť nami vytvorené zaznamenávače a prípade do aplikácii manuálne zaniest problémy s použiteľnosťou ak to bude potrebné.
4. Jedna aplikácia by mala byť natívna a druhá webová, aby sme overili oba nami vytvorené zaznamenávače.

Na základe stanovených požiadaviek pre aplikácie testovania, sme pre túto používateľskú štúdiu zvolili aplikácie VirtualFIIT³⁹ (webová mobilná aplikácia) a RedReader⁴⁰ (natívna mobilná aplikácia).

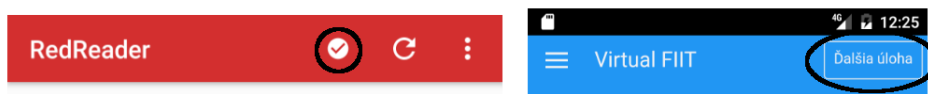
Keďže sme v oboch aplikáciách potrebovali robiť modifikácie (vlozenie zaznamenávačov a zavedenie niektorých problémov s použiteľnosťou) bolo potrebné obe aplikácie sfunkčnit' na lokálnom vývojom prostredí. Pri tomto kroku sme v rámci oboch aplikácií narazili na závažné technické problémy, ktoré sa nám však nakoniec podarilo odstrániť.

Pri návrhu scenára testovania sme pri oboch aplikáciách narazili na potrebu získavať od používateľov spätnú väzbu počas vykonávania jednotlivých úloh (odpovede na otázky). Pre uľahčenie práce účastníkov testovania sme sa rozhodli túto funkcionality implementovať priamo do testovaných aplikácií.

Do oboch aplikácií sme doplnili tlačidlo, ktorým účastníci značia, že ukončili práve vykonávanú úlohu a začínajú pracovať na nasledovnej úlohe (obr. 7). Vďaka tejto funkcionalite máme zabezpečené zaznamenávanie času začiatku a konca vykonávania jednotlivých úloh. Navyše pokiaľ daná úloha vyžaduje od účastníka testovania odpoveď, po stlačení tlačidla pre potvrdenie ukončenia úlohy je účastník vyzvaný, aby priamo v aplikácii zadal odpoveď na práve vykonanú úlohu a táto odpoveď sa automaticky odošle na server.

³⁹<https://play.google.com/store/apps/details?id=sk.stuba.fiit.virtfiit>

⁴⁰<https://play.google.com/store/apps/details?id=org.quantumbadger.redreader>



Obr. 7: Tlačidlá pre potvrdenie ukončenia úlohy

Pri návrhu scenára testovania (úloh pre účastníkov testovania) sme sa snažili v rámci oboch aplikácií identifikovať čo najväčšie množstvo problémov s použiteľnosťou a následne sme úlohy pre účastníkov testovania formulovali tak, aby bola vysoká pravdepodobnosť, že počas vykonávania testovania narazia na dané problémy.

V rámci vybraných aplikácií sme vykonali niekoľko modifikácií, ktoré predstavovali podporu testovania, alebo umožňovali identifikáciu problémov s použiteľnosťou (tj. manuálne sme zanesli do aplikácií problémy s použiteľnosťou).

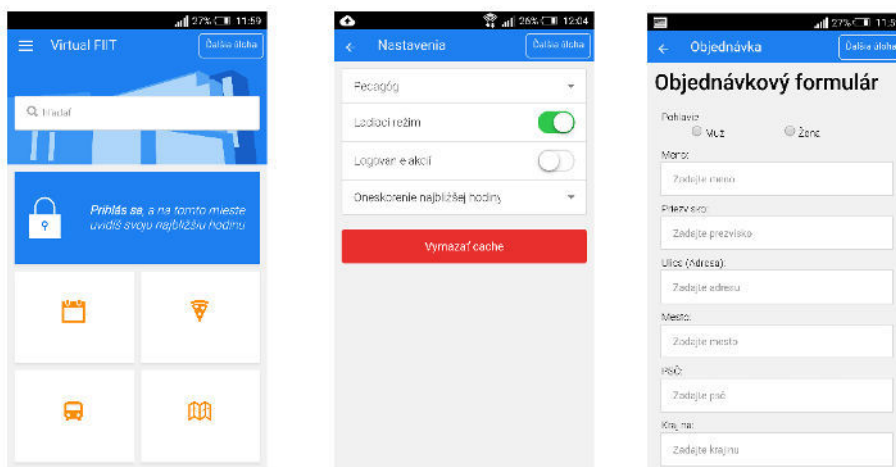
V rámci aplikácie RedReader jedinou modifikáciou bolo doplnenie tlačidla pre značenie práve vykonanej úlohy a začiatku práce na ďalšej úlohe. V rámci aplikácie VirtualFIIT sme vykonali nasledovné zmeny:

1. Doplnenie tlačidla pre značenie práve vykonanej úlohy a začiatku práce na ďalšej úlohe.
2. Odstránenie popisov ikoniek z hlavnej obrazovky (Obrázok 8a).
3. V navigácii zmena názvu položky 'Harmonogram' na 'Rozpis'.
4. Preusporiadanie položiek v rámci obrazovky 'Odkazy'.
5. Doplnenie tlačidla 'Logovanie akcií' v nastaveniach aplikácie (Obrázok 8b).
6. Vytvorenie komplexného formuláru. Obrazovka 'Objednávka' znázornená na Obrázku 8c.

Úlohy pre účastníkov testovania sme navrhovali s ohľadom na to, aby sa nám podarilo overiť čo najväčší počet správnosti zaradenia problémov s použiteľnosťou ako problémov, ktoré je možné odhaliť pomocou emulácie na počítači.

V rámci navrhnutého scenára testovania sme testovali nasledovné problémy:

- formulár obsahuje príliš veľa prvkov a preto je pre používateľa odpudzujúci
- jednoduché nájdenie obsahu alebo funkcionality o ktorej používateľa predpokladajú že je prítomná



Obr. 8: Zľava: a) Modifikovaná hlavná obrazovka aplikácie VirtualFIIT, b) Doplnené tlačidlo 'Logovanie akcií' v nastaveniach aplikácie VirtualFIIT, c) Obrazovka 'Objednávkový formulár' v aplikácii VirtualFIIT

- nepochopenie zamerania webu/aplikácie počas iniciálnej interakcie
- navigačné prvky nie sú pre nových používateľov úplne jasné pretože obsahujú nejednoznačné pomenovania (používateľ nevie, kam ho navigačný prvok zavedie)
- navigačné prvky nie sú pre nových používateľov úplne jasné pretože sú reprezentované pomocou neznámych ikon bez dodatočného textového vysvetlenia
- používateľ nesprávne chápe obsah (v aplikácii sú používané doménovo špecifické termíny)
- vstup nie je dobre označený a používateľ nevie, čo má byť obsahom daného pol'a
- používateľ niekde očakával určitú funkcionálnosť, ale reálne sa nachádzala inde
- používateľ očakáva iný obsah (očakával že na stránke X nájde obsah Y ale nebolo tomu tak)
- používateľ nevie určiť, prečo v danom okamihu link/tlačidlo nie je klikateľné (je deaktivovaný)

- vstup nezobrazuje celý zadaný obsah a preto má používateľ problém skontrolovať zadanú hodnotu (musí napr. scrollovať)
- objavenie novej funkcionality o ktorej používatelia doteraz nevedeli
- vstup vyžaduje špecifické formátovanie vstupu, ktoré je pre používateľa ťažko pochopiteľné

Účastníci

Štúdia prebehla v dvoch sedeniach, v ktorých sa spoločne zúčastnilo 24 účastníkov testovania, pričom išlo o študentov našej fakulty.

Výsledky a zhodnotenie

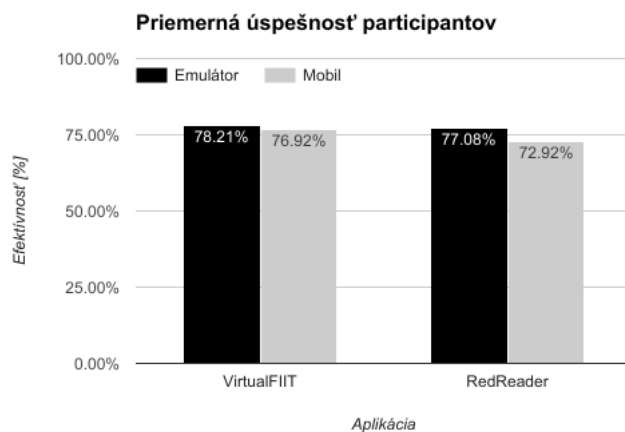
V rámci vyhodnotenia tejto štúdie sme primárne sledovali dve metriky a to: *Priemerná úspešnosť účastníkov testovania* a *Počet identifikovaných problémov s použiteľnosťou*.

Úspešnosť jedného účastníka testovania bola určená ako jeho úspešnosť splnenia úloh (teda koľko percent úloh daný účastník vykonal správne). Určili sme priemernú úspešnosť účastníkov testovania ako priemer hodnôt jednotlivých účastníkov pre jednotlivé aplikácie a taktiež vzhľadom na typ zariadenia na ktorom bola úloha vykonaná. Vypočítané priemerné úspešnosti účastníkov testovania môžeme vidieť na obrázku 9.

V rámci aplikácie VirtualFIIT bola priemerná úspešnosť participantov v rámci mobilného zariadenia 76.92%, zatiaľ čo v rámci emulátoru 78.21%. Vidíme, že ide o rozdiel 1.29%. Aplikovali sme Welch's t-test⁴¹ aby sme identifikovali, či je daný rozdiel štatisticky signifikantný. Na základe zistených hodnôt *t-value: 0.11041* a *p-value: 0.913002* sa ukázalo, že daný rozdiel nie je štatisticky signifikantný.

To že nám štatistický test hovorí, že daný rozdiel nie je štatisticky signifikantný, nemusí nutne znamenať, že tu rozdiel skutočne nie je. Indikuje nám to iba to, že naše dáta nie sú dostatočné na to, aby potvrdili, že rozdiel existuje. Na základe toho sme uskutočnili aj na testovanie ekvivalencie. Pre overenie sme vypočítali 95% confidence interval, ktorý nadobudol hodnoty: *-10.34 : 12.9*. Vidíme, že tento interval prekročil hranicu 0 a teda rozdiel nie je štatisticky signifikantný. Avšak

⁴¹<http://www.statisticshowto.com/welchs-test-for-unequal-variances/>



Obr. 9: Priemerná úspešnosť účastníkov testovania

pokiaľ sa zameriavame na testovanie ekvivalencie, je potrebné si stanoviť hranicu ekvivalencie, teda taký rozdiel v nameraných hodnotách, kedy výsledok ešte budeme považovať za ekvivalentný. To závisí od povahy konkrétnej štúdie a v našom prípade je ťažké určiť kedy by sme mohli výsledok považovať za ekvivalentný. To by bolo potrebné určiť vzhľadom na povahu testovanej aplikácie. V rámci našich výsledkov (na základe vypočítaného confidence intervalu) vidíme, že keby stanovíme hranicu ekvivalencie rovnú 13%, tak náš confidence interval spadne do tejto hranice z oboch strán a teda môžeme namerané výsledky považovať za ekvivalentné.

V rámci aplikácie RedReader bola priemerná úspešnosť účastníkov v rámci mobilného zariadenia 72.92%, zatiaľ čo v rámci emulátoru 77.08%. Vidíme, že ide o rozdiel 4.16%. Opäť sme aplikovali Welch's t-test aby sme identifikovali, či je daný rozdiel štatisticky signifikantný. Na základe zistených hodnôt $t\text{-value}$: 0.36521 a $p\text{-value}$: 0.720414 sa ukázalo, že daný rozdiel nie je štatisticky signifikantný. Pre zistenie ekvivalencie sme taktiež vypočítali 95% confidence interval, ktorý nadobudol hodnoty: -7.24 : 15.58. Vidíme, že tento interval prekročil hranicu 0 a teda rozdiel nie je štatisticky signifikantný. Opäť však vidíme, že tento interval je pomerne široký a hranicu ekvivalencie by sme v tomto prípade museli stanoviť až na 15.6%.

Vďaka detailnej analýze nahrávok sedenia účastníkov testovania sme vytvorili

zoznamy problémov spojených s použiteľnosťou daných aplikácií, pričom sme rozlišovali, či bol daný problém identifikovaný na mobilnom zariadení alebo pomocou emulácie na počítači. Podarilo sa nám spoločne pre obe aplikácie identifikovať 21 problémov s použiteľnosťou. Z toho na mobilnom zariadení bolo identifikovaných všetkých 21 problémov a pomocou emulátoru 16 problémov. Z celkových 21 problémov sa teda na oboch zariadeniach podarilo identifikovať 16 problémov čo predstavuje 76%.

Hlavným cieľom tejto štúdie bolo overenie nami vytvoreného zoznamu potenciálnych problémov s použiteľnosťou mobilných aplikácií. Vybrali sme 13 problémov zo zoznamu problémov u ktorých predpokladáme, že ich je možné odhaliť pomocou emulácie a 1 problém zo zoznamu problémov u ktorých predpokladáme, že ich nebude možné odhaliť pomocou emulácie. Tieto problémy sme manuálne zanesli do testovaných aplikácií a sledovali sme, či budú odhalené.

V prípade jedného problému, ktorý bol zvolený zo zoznamu problémov u ktorých predpokladáme, že ich nebude možné odhaliť pomocou emulácie sa nám potvrdila správnosť jeho zaradenia. Ide o problém: “Farby pozadia a textu sú nevhodne zvolené a preto je text ťažko čitateľný.” Môžeme vidieť, že pomocou mobilného zariadenia sa podarilo identifikovať 4 výskyty tohto problému, zatiaľ čo pomocou emulátoru ani jeden.

Zo zoznamu 13 generických problémov u ktorých predpokladáme, že ich bude možné odhaliť pomocou emulácie sa nám správnosť ich zaradenia potvrdila u 12 problémov. To znamená, že sa tento problém podarilo identifikovať aj pomocou mobilného zariadenia a taktiež aj pomocou emulácie na počítači.

5.4 Diskusia a zhodnotenie

V rámci našej metódy sme pre podporu automatizácie vyhodnocovania používateľských štúdií implementovali dva zaznamenávače, slúžiace na zaznamenávanie interakcie používateľov. Prvý zaznamenávač je určený pre natívne mobilné aplikácie s operačným systémom Android a druhý je určený pre webové mobilné aplikácie. Tieto zaznamenávače zaznamenávajú kliknutia používateľov a pokiaľ je to možné, tak identifikujú na aký element interakcia nastala.

Keďže je naša metóda zameraná na testovanie mobilných aplikácií pomocou emulácie na počítači, tak ako sme už viac krát uviedli, je zjavné, že takouto metódou nie je možné odhaliť všetky problémy spojené s použiteľnosťou mobilných apli-

kácií. Vykonali sme rozsiahlu analýzu, v rámci ktorej sme vytvorili dva zoznamy potenciálnych problémov s použiteľnosťou, ktoré sa môžu objaviť počas testovania mobilnej aplikácie. Prvý zoznam obsahuje problémy, ktoré predpokladáme, že je možné odhaliť pomocou emulácie na počítači, zatiaľ čo druhý obsahuje problémy, ktoré nie je možné odhaliť pomocou emulácie. Vďaka využitiu týchto zoznamov je následne možné určiť, či je vhodné využiť našu metódu pre plánované testovanie mobilnej aplikácie a následne po ukončení štúdie je možné vďaka týmto zoznamom identifikovať validitu zistených problémov s použiteľnosťou.

Pre overenie nami navrhutej metódy používateľského testu sme uskutočnili dve používateľské štúdie.

Prvej štúdie sa zúčastnilo 18 účastníkov a jej hlavným cieľom bolo overenie základnej hypotézy, že existujú problémy s použiteľnosťou mobilnej aplikácie, ktoré je možné odhaliť pomocou emulácie na počítači. Vďaka tomuto experimentu sa nám podarilo potvrdiť našu základnú hypotézu keďže zo 16 problémov identifikovaných pomocou mobilného zariadenia bolo 14 problémov identifikovaných aj pomocou emulátoru, čo predstavuje 87.5%. Navyše sa nám pomocou tohto experimentu podarilo ukázať, že úspešnosti jednotlivých úloh, účinnosť účastníkov testovania a dĺžky trvania jednotlivých úloh nie sú nijako výrazne ovplyvnené typom zariadenia na ktorom je úloha vykonávaná.

Druhej štúdie sa zúčastnilo 24 účastníkov a jej hlavným cieľom bolo overenie nami definovaného zoznamu potenciálnych problémov u ktorých predpokladáme, že ich je možné odhaliť pomocou emulácie. Zo zoznamu 13 generických problémov u ktorých predpokladáme, že ich bude možné odhaliť pomocou emulácie sa nám správnosť ich zaradenia potvrdila u 12 problémov. V prípade jedného problému zo zoznamu problémov u ktorých predpokladáme, že ich nebude možné odhaliť pomocou emulácie sa nám taktiež potvrdila správnosť jeho zaradenia.

6 Záver a zhodnotenie

V rámci našej práce sme navrhli metódu používateľského testu, ktorá vďaka využitiu emulácie na počítači umožňuje efektívnejšiu realizáciu kvantitatívnych používateľských štúdií s využitím sledovania pohľadu, zameraných na testovanie mobilných aplikácií. Zároveň vďaka využitiu zariadenia na sledovanie pohľadu určeného pre počítače, môžeme vďaka našej metóde získať detailnejšie informácie o pohľade používateľov, ako v prípade využitia iných techník akými sú napríklad stojany pre mobilné zariadenia so sledovačmi pohľadu.

Keďže je naša metóda zameraná na testovanie mobilných aplikácií pomocou emulácie na počítači, je zjavné, že takouto metódu nie je možné odhaliť všetky problémy spojené s použiteľnosťou mobilných aplikácií. Vykonali sme rozsiahlu analýzu, v rámci ktorej sme vytvorili dva zoznamy potenciálnych problémov s použiteľnosťou, ktoré sa môžu objaviť počas testovania mobilnej aplikácie. Prvý zoznam obsahuje problémy, ktoré predpokladáme, že je možné odhaliť pomocou emulácie na počítači, zatiaľ čo druhý obsahuje problémy, ktoré nie je možné odhaliť pomocou emulácie. Vďaka využitiu týchto zoznamov je následne možné určiť, či je vhodné využiť našu metódu pre plánované testovanie mobilnej aplikácie a následne po ukončení štúdie je možné vďaka týmto zoznamom identifikovať validitu zistených problémov s použiteľnosťou.

Za účelom podpory automatického vyhodnocovania používateľských štúdií sme implementovali dva zaznamenávače, ktoré slúžia na zaznamenávanie interakcie v mobilných aplikáciách počas testovania použiteľnosti. Jeden zaznamenávač je určený pre natívne mobilné aplikácie, zatiaľ čo druhý je určený pre zaznamenávanie aktivity vo webových mobilných aplikáciách. Okrem zaznamenávania kliknutí používateľov, pokiaľ je to možné, tak zaznamenávače identifikujú konkrétny element, na ktorom nastala interakcia. Takáto znalosť má vysokú pridanú hodnotu a môže byť veľmi prínosnou z hľadiska podpory automatizácie vyhodnocovania používateľských štúdií.

V rámci overenia nami vytvorenej metódy sme realizovali dve používateľské štúdie. Hlavným cieľom prvej štúdie bolo overiť našu základnú hypotézu, že pomocou emulácie mobilných aplikácií na počítači dokážeme odhaliť určité problémy spojené

s ich použiteľnosťou. V rámci vyhodnotenia uskutočnenej používateľskej štúdie sme identifikovali 19 unikátnych problémov spojených s použiteľnosťou danej aplikácie a 87.5% problémov, ktoré sa podarilo identifikovať pomocou realizácie na mobilnom zariadení sa podarilo identifikovať aj pomocou emulácie na počítači. Tieto výsledky dostatočne podporili našu prvotnú hypotézu a uistili nás v tom, že naša metóda je schopná odhaliť vybrané problémy s použiteľnosťou mobilnej aplikácie.

Navyše sa nám pomocou tohto experimentu podarilo ukázať, že úspešnosti jednotlivých úloh, účinnosť účastníkov testovania a dĺžky trvania jednotlivých úloh nie sú nijako výrazne ovplyvnené typom zariadenia na ktorom je úloha vykonávaná.

Hlavným cieľom druhej realizovanej používateľskej štúdie bolo overenie vhodnosti zaradenia čo najväčšieho množstva potenciálnych problémov do zoznamu problémov, ktoré je možné identifikovať pomocou emulácie. Cieľom realizácie tejto štúdie bolo zároveň aj overenie správnosti fungovania implementovaných zaznamenávačov. Zo zoznamu 13 generických problémov u ktorých predpokladáme, že ich bude možné odhaliť pomocou emulácie sa nám správnosť ich zaradenia potvrdila u 12 problémov. V prípade jedného problému zo zoznamu problémov u ktorých predpokladáme, že ich nebude možné odhaliť pomocou emulácie sa nám taktiež potvrdila správnosť jeho zaradenia.

V rámci našej metódy nebudeme schopní odhaliť všetky problémy s použiteľnosťou, keďže je odstránený kontext mobilného zariadenia. Otvára sa teda možnosť kombinácie našej metódy s bežnými prístupmi. Tejto možnosti sa v našej práci nevenujeme, avšak nechávame ju otvorenú ako možnosť pre ďalší rozvoj našej práce. V rámci kombinácie metód by časť účastníkov testovania vykonávala testovanie mobilnej aplikácie pomocou emulácie a zároveň by menšia časť účastníkov vykonala testovanie na mobilnom zariadení. Takouto kombináciou by bolo možné ušetriť čas pomocou hromadnej emulácie na počítači a zároveň aj odhaliť problémy, ktoré môžu byť špecifické pre kontext mobilného zariadenia.

Ďalším otvoreným výskumným problémom, v rámci ktorého vidíme možné vylepšenie našej práce je skúmanie vplyvu zobrazovacích veľkostí. Je pravdepodobné, že v rámci rozličných zobrazovacích veľkostí v rámci emulácie na počítači bude správanie používateľov mierne odlišné a teda budú ovplyvnené aj výsledky štúdie. Otvoreným výskumným problémom je nájdenie optimálnej zobrazovacej veľkosti, v rámci ktorej by bolo možné získať dostatočne detailné dáta o pohľade

používateľov a zároveň, by výsledky štúdie najviac korelovali s identickou štúdiou na mobilnom zariadení.

Literatúra

- [1] AL-QAIMARI, G., AND MCROSTIE, D. *KALDI: A Computer-Aided Usability Engineering Tool for Supporting Testing and Analysis of Human-Computer Interaction*. Springer Netherlands, Dordrecht, 1999, pp. 337–355.
- [2] ALBINSSON, P.-A., AND ZHAI, S. High precision touch screen interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2003), CHI '03, ACM, pp. 105–112.
- [3] ATTERER, R., WNUK, M., AND SCHMIDT, A. Knowing the user's every move: User activity tracking for website usability evaluation and implicit interaction. In *Proceedings of the 15th International Conference on World Wide Web* (New York, NY, USA, 2006), WWW '06, ACM, pp. 203–212.
- [4] BETIOL, A. H., AND DE ABREU CYBIS, W. *Human-Computer Interaction - INTERACT 2005: IFIP TC13 International Conference, Rome, Italy, September 12-16, 2005. Proceedings*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005, ch. Usability Testing of Mobile Devices: A Comparison of Three Approaches, pp. 470–481.
- [5] BOJKO, A. *Eye tracking the user experience*. Rosenfeld Media, 2013.
- [6] BRANCO, P., FIRTH, P., ENCARNAÇÃO, L. M., AND BONATO, P. Faces of emotion in human-computer interaction. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA, 2005), CHI EA '05, ACM, pp. 1236–1239.
- [7] BROOKE, J. Sus-a quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.
- [8] CHEN, M. C., ANDERSON, J. R., AND SOHN, M. H. What can a mouse cursor tell us more?: Correlation of eye/mouse movements on web browsing. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA, 2001), CHI EA '01, ACM, pp. 281–282.
- [9] COLLEY, A., AND HÄKKILÄ, J. Exploring finger specific touch screen interaction for mobile phone user interfaces. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design* (New York, NY, USA, 2014), OzCHI '14, ACM, pp. 539–548.
- [10] COURSARIS, C. K., AND KIM, D. J. A meta-analytical review of empirical mobile usability studies. *J. Usability Studies* 6, 3 (May 2011), 11:117–11:171.

- [11] CRONBACH, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 3 (1951), 297–334.
- [12] CUADRAT SEIX, C., VELOSO, M. S., AND SOLER, J. J. R. Towards the validation of a method for quantitative mobile usability testing based on desktop eyetracking. In *Proceedings of the 13th International Conference on Interacción Persona-Ordenador* (New York, NY, USA, 2012), INTERACCION '12, ACM, pp. 49:1–49:8.
- [13] DUH, H. B.-L., TAN, G. C. B., AND CHEN, V. H.-H. Usability evaluation for mobile device: A comparison of laboratory and field tests. In *Proceedings of the 8th Conference on Human-computer Interaction with Mobile Devices and Services* (New York, NY, USA, 2006), MobileHCI '06, ACM, pp. 181–186.
- [14] GONG, J., AND TARASEWICH, P. Guidelines for handheld mobile device interface design. In *In Proceedings of the 2004 DSI Annual Meeting* (2004).
- [15] HAN, J. Y. Multi-touch interaction wall. In *ACM SIGGRAPH 2006 Emerging Technologies* (New York, NY, USA, 2006), SIGGRAPH '06, ACM.
- [16] HART, S. G., AND STAVELAND, L. E. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Human Mental Workload*, P. A. Hancock and N. Meshkati, Eds., vol. 52 of *Advances in Psychology*. North-Holland, 1988, pp. 139 – 183.
- [17] HELFRICH, B., AND LANDAY, J. A. Quip: Quantitative user interface profiling, 1999.
- [18] IVORY, M. Y., AND HEARST, M. A. The state of the art in automating usability evaluation of user interfaces. *ACM Comput. Surv.* 33, 4 (Dec. 2001), 470–516.
- [19] KAIKKONEN, A., KEKÄLÄINEN, A., CANKAR, M., KALLIO, T., AND KANKAINEN, A. Usability testing of mobile applications: A comparison between laboratory and field testing. *J. Usability Studies* 1, 1 (Nov. 2005), 4–16.
- [20] KRUG, S. *Rocket Surgery Made Easy: The Do-It-Yourself Guide to Finding and Fixing Usability Problems*, 1st ed. New Riders Publishing, Thousand Oaks, CA, USA, 2009.
- [21] MACLEOD, M., AND RENGGER, R. The development of drum: A software tool for video-assisted usability evaluation. In *In Proceedings of HCI'93* (1993), Cambridge University Press, pp. 293–309.
- [22] MOBASHER, B., COOLEY, R., AND SRIVASTAVA, J. Automatic personalization based on web usage mining. *Commun. ACM* 43, 8 (Aug. 2000), 142–151.

- [23] MUELLER, F., AND LOCKERD, A. Cheese: Tracking mouse movement activity on websites, a tool for user modeling. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA, 2001), CHI EA '01, ACM, pp. 279–280.
- [24] MÓRO, R., DARÁŽ, J., AND BIELIKOVÁ, M. Visualization of gaze tracking data for ux testing on the web. In *Proceedings of DataWiz 2014: Data Vizualisation Workshop* (2014).
- [25] NIELSEN, J. Usability inspection methods. In *Conference Companion on Human Factors in Computing Systems* (New York, NY, USA, 1994), CHI '94, ACM, pp. 413–414.
- [26] OLSEN, D. R., AND HALVERSEN, B. W. Interface usage measurements in a user interface management system. In *Proceedings of the 1st Annual ACM SIGGRAPH Symposium on User Interface Software* (New York, NY, USA, 1988), UIST '88, ACM, pp. 102–108.
- [27] PALINKO, O., KUN, A. L., SHYROKOV, A., AND HEEMAN, P. Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (New York, NY, USA, 2010), ETRA '10, ACM, pp. 141–144.
- [28] POOLE, A., AND BALL, L. J. Eye tracking in human-computer interaction and usability research: Current status and future. In *Prospects*, Chapter in C. Ghaoui (Ed.): *Encyclopedia of Human-Computer Interaction*. Pennsylvania: Idea Group, Inc (2005).
- [29] RAUTERBERG, M. A petri net based analyzing and modelling tool kit for logfiles in human-computer interaction. *Proceedings 'Cognitive Systems Engineering in Process Control–CSEPC'96* (1996), 268–275.
- [30] RAYNER, K., FOORMAN, B. R., PERFETTI, C. A., PESETSKY, D., AND SEIDENBERG, M. S. How psychological science informs the teaching of reading. *Psychological Science in the Public Interest* 2, 2 (2001), 31–74.
- [31] RODDEN, K., AND FU, X. Exploring how mouse movements relate to eye movements on web search results pages. *Web Information Seeking and Interaction* (2007), 29–32.
- [32] SCHMIDT, A. Implicit human computer interaction through context. *Personal Technologies* 4, 2 (2000), 191–199.
- [33] SHEN, C., AND ZHAO, Q. *Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VII*. Springer International Publishing, Cham, 2014, ch. Webpage Saliency, pp. 33–46.

- [34] SURE, M. Questionnaires for usability: A systematic literature review. Master's thesis, Linköping University, Sweden, 2014.
- [35] TULLIS, T., AND ALBERT, W. *Measuring the User Experience, Second Edition: Collecting, Analyzing, and Presenting Usability Metrics*, 2nd ed. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2013.
- [36] ZETTLEMOYER, L. S., ST. AMANT, R., AND DULBERG, M. S. Ibots: Agent control through the user interface. In *Proceedings of the 4th International Conference on Intelligent User Interfaces* (New York, NY, USA, 1999), IUI '99, ACM, pp. 31–37.

A Protokol štúdie 1: Overenie základnej hypotézy

Protokol experimentu

Kontext experimentu

Meno a priezvisko Peter Dubec

Supervisor prof. Mária Bieliková

Názov projektu Podpora uskutočňovania kvantitatívnych používateľských štúdií mobilných aplikácií s využitím sledovania pohľadu

Kľúčové slová sledovanie pohľadu, testovanie mobilných aplikácií, kvantitatívne testovanie, emulácia

Stručný opis projektu V dnešnej dobe je testovanie použiteľnosti mobilných aplikácií s využitím sledovania pohľadu uskutočňované ako klasické testovanie použiteľnosti. Taktiež aktuálne je veľmi náročné uskutočňovať testovanie použiteľnosti mobilných aplikácií pomocou viacerých účastníkov testovania súčasne. Na základe toho sme schopní zbierať počas testovania iba malé množstvo dát, a preto ide skôr o kvalitatívne a nie kvantitatívne testovanie. Naším hlavným cieľom je umožniť hromadné testovanie mobilných aplikácií pomocou emulácie na počítači a zbieraním dát počas interakcie na počítači. Aj keď interakcia s počítačom je odlišná ako interakcia s mobilným zariadením, predpokladáme, že existujú určité problémy s použiteľnosťou mobilných aplikácií, ktoré budeme schopní identifikovať aj pomocou takejto emulácie. Taktiež predpokladáme, že vďaka emulácii budeme schopní zbierať kvalitnejšie dáta zo sledovača pohľadu.

Stručný opis projektu v AJ Nowadays usability testing of mobile applications with use of eye-tracking is realized as traditional usability testing. It is very difficult to perform mobile usability testing on multiple participants simultaneously, therefore in most cases these tests are qualitative and not quantitative because we are able to collect only small amount of data. Our main goal is to make bulk testing of mobile applications possible by means of simulating mobile applications on PC and collecting data from PC. Interaction with PC is quite different than interaction with smartphone, but we believe that there are certain usability problems which we will be able to identify also during emulation on PC and also we will be able to collect more precise data from eye-tracking thanks to bigger displaying screen.

Príprava experimentu

Cieľ experimentu Cieľom experimentu je uskutočniť testovanie mobilnej aplikácie na mobilnom zariadení a pomocou emulácie na počítači, aby sme získali potrebné dáta pre overenie našich hypotéz.

Hypotézy

1. Pomocou emulácie mobilných aplikácií na počítači dokážeme odhaliť určité problémy spojené s ich použiteľnosťou.

Parametre experimentu

1. Testovanie mobilnej aplikácie na mobilnom zariadení
2. Testovanie mobilnej aplikácie pomocou emulácie na počítači

Účastníci 18 participantov vo veku 19-23 rokov. (študenti FIIT STU)

Sledované metriky

- Úspešnosť splnenia úloh
- Čas trvania jednotlivých úloh

Scenár experimentu

Účastníci experimentu majú za úlohu vykonať 7 definovaných úloh v mobilnej aplikácii CSFD. Predpokladaná dĺžka trvania 15min jeden účastník.

- **Pred vstupom účastníka (pripraví moderátor)**
 - Nachystať úvodný dotazník
 - Náhodne zvolíť 3 úlohy, ktoré vykoná účastník na mobilnom zariadení a 3 ktoré vykoná na počítači (+ úloha prihlásenia v oboch prípadoch - budú definované dve trojice úloh a náhodný výber bude predstavovať voľbu, ktorú trojicu úloh participant vykoná na mobilnom zariadení a ktorú na počítači)
 - Nastaviť Tobii studio na test **Prihlásenie**
 - Nastaviť Tobii Studio na predkalibračnú obrazovku, kde sa zadáva meno účastníka
 - **Príprava aplikácie:**
 - Otvoriť v smartfóne alebo na počítači CSFD aplikáciu, nastaviť ju na úvodnú obrazovku
 - Odstrániť TV stanicu Dajto zo zoznamu TV staníc
 - Odstrániť z hlavnej obrazovky položku "Práve vychází na DVD"
 - Odstrániť film Avatar zo zoznamu "Chci vidět"
 - **Ak testovanie na mobile:**
 - Položiť na smartfón kalibračnú doštičku
 - Skontrolovať správne uchytanie okulografu
 - Otvoriť webcam control, resp. overiť pripojenie a správne nakalibrovanie kamery cez test preview
- **Uvedenie účastníka:**
 - Účastník sa usadí na stoličku
 - Účastníkovi sa vysvetlí experiment
 - Cieľ: overujeme aplikáciu CSFD
 - Postup: jednoduché úlohy na porozumenie a prácu s rozhraním, časť úloh bude vykonaná na mobile a časť úloh na počítači, predpokladaná dĺžka celého sedenia cca 15min
 - Upozorniť: priebeh experimentu zaznamenávame pre účely neskoršej analýzy, ale na nič iné
 - Upozorniť: "nie je cieľom experimentu vás skúšať, nič, čo dnes spravíte, nemôže byť zle a naopak nám to pomôže pri vylepšení testovanej aplikácie",
 - Upozorniť: "aplikácia, ktorá bude testovaná, môže obsahovať chyby, preto nie je problém, ak sa na nejakej úlohe 'zaseknete', vtedy stačí povedať, že neviete, ako v úlohe ďalej pokračovať a ideme na ďalšiu úlohu "
 - Upozorniť: "počas experimentu využijememe kameru na zaznamenávanie pohľadu - ukázať účastníkovi, kde sa kamera nachádza"

- Moderátor vysvetlí účastníkovi ako správne sedieť počas experimentu a ako ovládať zariadenie tak, aby bolo zaznamenávanie bezproblémové
- Vyplní sa úvodný dotazník (účastník postupne odpovie na moderátorove otázky)
- Moderátor dá účastníkovi priestor na prípadné otázky pred experimentom

● **ÚLOHY:**

1. Úloha 1 (Prihlásenie)

- Prihláste sa do aplikácie CSFD a vráťte sa na hlavnú obrazovku aplikácie.
 - Meno: uxfiit
 - Heslo: uxfiit1

2. Pre režiséra filmu Schindlerov zoznam, zistite aké bolo jeho najstaršie dielo v rámci jeho scénaristickej filmografie.

- **Správna odpoveď:** Fighter Squad (1961)
- **Možné problémy:**
 - Používateľ nebude vedieť identifikovať meno režiséra
 - V rámci filmografie režiséra používateľ nenájde správny zoznam

3. Pridajte film Avatar do zoznamu filmov, ktoré si chcete pozrieť a následne uveďte počet položiek v tomto zozname.

- **Možné problémy:**
 - Používateľ nebude vedieť pridať film do zoznamu filmov, ktoré si chce pozrieť
 - Používateľ nebude vedieť nájsť zoznam filmov, ktoré si chce pozrieť

4. Zistite názov 5. Najobľúbenejšieho seriálu podľa CSFD

- **Správna odpoveď:** Game of Thrones
- **Možné problémy:**
 - Používateľ bude mať problém nájsť správny rebríček

5. Zistite, čo sa bude vysielat' na televíznej stanici "Dajto" dňa 30. Apríla o 20:15

- **Správna odpoveď:** Blbý a blbší
- **Možné problémy:**
 - Používateľ nebude vedieť pridať novú stanicu do zoznamu TV kanálov
 - Používateľ bude mať problém nájsť požadovaný dátum a čas

6. Na úvodnú obrazovku pridajte položku "Právě vychází na DVD"

- **Možné problémy:**
 - Používateľ nebude vedieť pridať novú položku na hlavnú obrazovku

7. Nájdite zoznam všetkých kín, ktoré sa nachádzajú v meste Praha a uveďte názov prvého kina v tomto zozname

- **Správna odpoveď:** Cinestar Praha - Anděl
- **Možné problémy:**

- Používateľ nebude vedieť nájsť zoznam kín, kde je možné vyhľadávať na základe názvu mesta.

Priebeh experimentu

Pilot 26.4.2016

Hlavnou úlohou v rámci pilotného experimentu bolo nastavenie potrebnej konfigurácie v oboch laboratóriách, aby boli používatelia schopní počas experimentu vykonávať úlohy a aby bola ich aktivita korektne zaznamenávaná. V rámci realizácie experimentu na PC bolo potrebné na dané zariadenie nainštalovať emulátor, aby sme boli schopní na počítači spustiť mobilnú aplikáciu a zadať požadované úlohy v rámci nástroja Tobii. Pre realizáciu experimentu na mobilnom zariadení bolo potrebné nakonfigurovať mobilné zariadenie, nainštalovať požadovanú aplikáciu a taktiež nakonfigurovať nástroj Tobii aby bol schopný korektne zaznamenávať interakciu s mobilným zariadením.

Po úspešnej konfigurácii bol pilotný experiment realizovaný pomocou jedného účastníka, ktorý vykonal zadanú sadu úloh na mobilnom zariadení a následne druhú sadu úloh na počítači. V rámci tohto testovania sme sledovali, či sú úlohy pre účastníkov splniteľné, sledovali sme, či sú zariadenia správne nakonfigurované aby zaznamenávali interakciu a taktiež sme odhadli čas, aký si bude potrebné vyhraďiť na jedného účastníka počas realizácie ostrej štúdie.

Samotné vykonávanie úloh pilotným účastníkom sme nezahrnuli do finálneho vyhodnotenia uskutočnenej používateľskej štúdie.

Sedenie 1 27.4.2016

Prvého sedenia používateľskej štúdie sa zúčastnilo celkovo 18 účastníkov a sedenie prebiehalo celý deň. Účastníci vykonávali úlohy jednotlivo a jednému účastníkovi trvalo splnenie všetkých úloh podľa zadaného scénára približne 20-30min pričom účastník vykonával sadu úloh na mobilnom zariadení a sadu úloh priamo na počítači pomocou emulácie mobilnej aplikácie.

Zhodnotenie experimentu

Výsledky experimentu

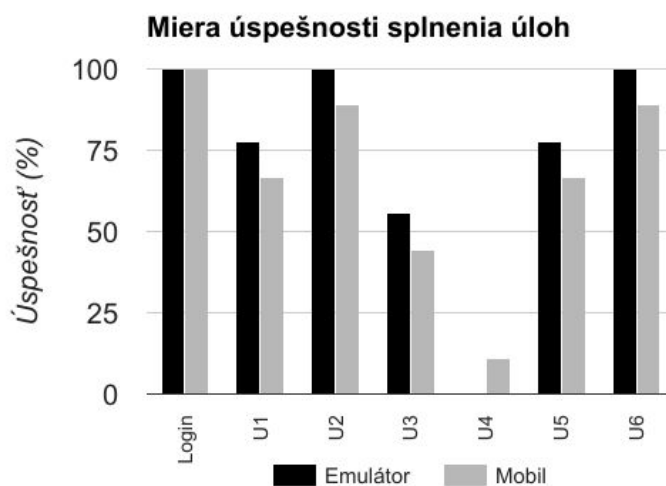
V rámci vyhodnotenia experimentu sme sledovali tri základné metriky a to:

- **Úspešnosť splnenia úloh**
- **Dĺžka trvania úloh**
- **Počet identifikovaných problémov s použiteľnosťou**

Úspešnosť splnenia úloh

Jednou z metrík ktoré sme sledovali je úspešnosť splnenia úloh. Táto metrika nám hovorí o tom, či daný používateľ splnil alebo nespľnil danú úlohu. My sme vyhodnotili celkovú úspešnosť všetkých participantov vzhľadom na danú úlohu (teda koľko percent participantov úspešne splnilo danú úlohu).

Na obrázku 1 môžeme vidieť výsledky vyhodnotenia úspešnosti splnenia jednotlivých úloh, pričom rozlišujeme, či bola úloha vykonaná na mobilnom zariadení alebo pomocou emulácie na počítači, aby sme boli schopí porovnať, či typ zariadenia ovplyvňuje úspešnosť úloh. Experimentu sa zúčastnilo 18 participantov, čiže každú úlohu vykonalo 9 participantov na mobilnom zariadení a 9 participantov pomocou emulácie na počítači. Vidíme, že medzi úspešnosťami úloh vzhľadom na typ zariadenia nie je žiaden výrazný rozdiel. Vo väčšine prípadov je úspešnosť vykonania úloh na počítači vyššia o jedného participanta, čo si vysvetľujeme tým, že práca s emulátorom bola pre participantov menej prirodzená a aj odozvy zariadenia boli mierne pomalšie ako v prípade mobilného zariadenia. Kvôli tejto skutočnosti participantí pracovali na počítači trochu pomalšie a teda venovali úlohám viac pozornosti.

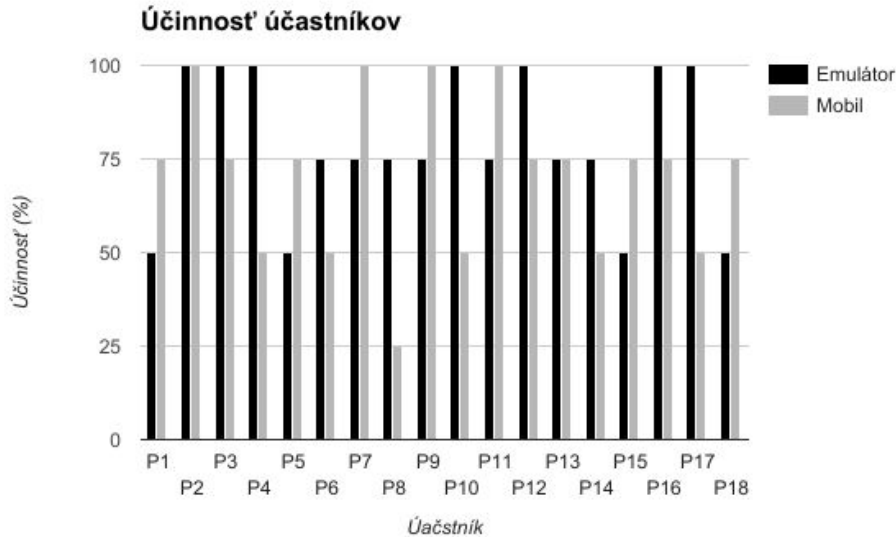


Obrázok 1: Miera úspešnosti splnenia úloh

Taktiež sme sledovali účinnosť (angl. effectiveness) jednotlivých participantov (presnosť a úplnosť s akou používateľ splnil zadané úlohy, tj. percento úloh, ktoré participant splnil spomedzi všetkých vykonaných úloh). Účinnosť participantov sme taktiež sledovali oddelene vzhľadom na typ zariadenia, aby sme boli schopní určiť, či typ zariadenia pri vykonávaní úloh výrazne neovplyvnil účinnosť jednotlivých participantov. Ako môžeme vidieť na Obrázku 2 účinnosť participantov je mierne ovplyvnená typom zariadenia, avšak je to veľmi špecifické pre daného participanta. U 7 participantov bola účinnosť vyššia pri práci na mobilnom zariadení, u 9 participantov bola účinnosť vyššia pri práci na počítači a u zvyšných 2 participantov bola miera účinnosti rovnaká na oboch zariadeniach.

Celkovo participantí vykonávali 7 unikátnych úloh, pričom jedna úloha bolo prihlásenie do aplikácie, čo vykonávali participantí na oboch zariadeniach. Na jednom zariadení teda používateľ vykonal celkovo 4 úlohy a teda splnenie/nespĺnenie jednej úlohy sa prejavilo v zmene účinnosti na danom zariadení o hodnotu 25%. Celková priemerná účinnosť dosiahnutá všetkými participantmi na počítači je rovná hodnote 76%, zatiaľ čo hodnota priemernej účinnosti na mobilnom zariadení je rovná 67%. Môžeme teda vidieť že ide o rozdiel 9%. Pre zistenie štatistickej signifikancie sme aplikovali Welch's t-test na základe ktorého sa nám ukázalo, že identifikovaný rozdiel nie je štatisticky signifikantný. Získané

hodnoty sú: $t\text{-value} = 0.59606$, $p\text{-value} = 0.562213$. Navyšme sme vypočítali 95% confidence interval, ktorý nadobudol hodnoty: $-6.5 : 25.5$. Keďže tento interval prekročil hodnotu 0, taktiež vidíme, že zistený rozdiel nie je štatisticky významný.

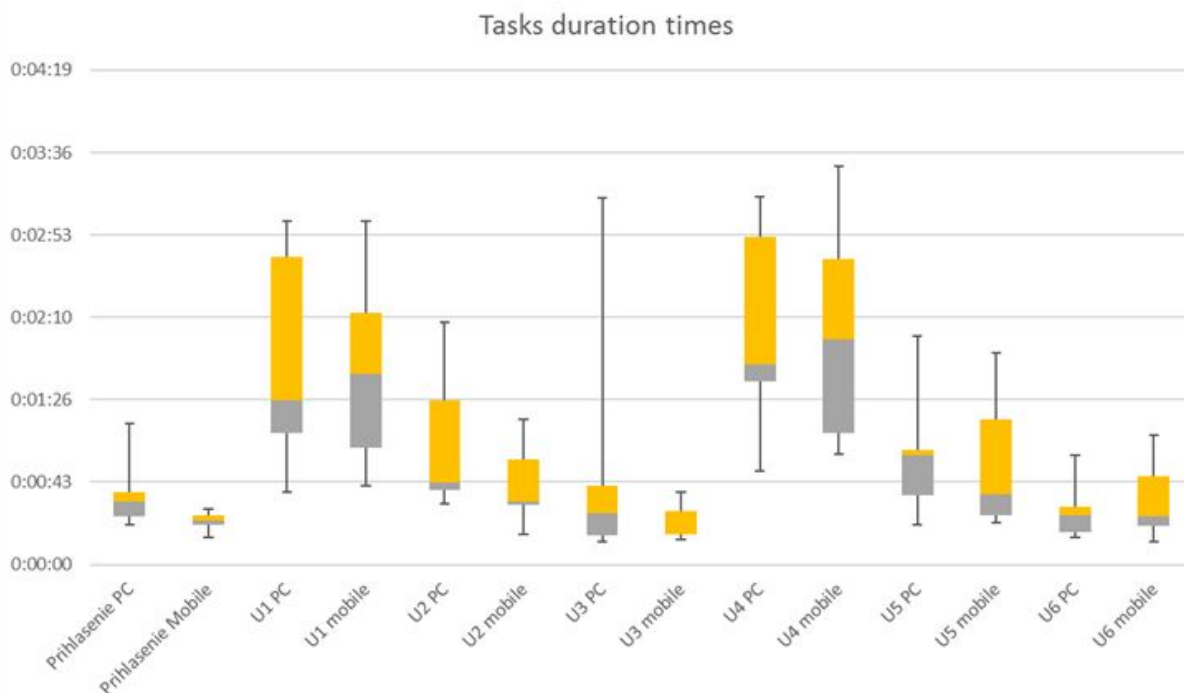


Obrázok 2: Účinnosť participantov

Dĺžka trvania úloh

V rámci vyhodnotenia experimentu sme ako jednu zo základných metrick sledovali dĺžku trvania jednotlivých úloh, pričom sme taktiež rozlišovali či používateľ vykonal úlohu na mobilnom zariadení alebo na počítači. Sledovali sme, či je dĺžka trvania jednotlivých úloh výrazne ovplyvnená typom zariadenia na ktorom bola úloha vykonávaná. Tak ako môžeme vidieť na obrázku 3 medián dĺžky trvania jednotlivých úloh je veľmi podobný v prípadoch oboch zariadení, na základe čoho môžeme určiť, že typ zariadenia výrazne neovplyvňuje dĺžku trvania úloh. Vo väčšine úloh však môžeme vidieť, že medián dĺžky trvania úloh vykonávaných na počítači je mierne vyšší ako pri úlohách vykonávaných na mobilnom zariadení. Vysvetlenie tejto situácie sme už poskytli pri opise úspešnosti úloh. Keďže interakcia s emulovanou aplikáciou na počítači je pre používateľov mierne menej intuitívna ako práca s mobilným zariadením a taktiež odozva na interakcie je mierne pomalšia, je zjavné, že dĺžky trvania úloh vykonávaných na počítači sú mierne dlhšie ako v prípade mobilného zariadenia.

Priemerná dĺžka trvania splnenia úlohy na emulátore je 70s, zatiaľ čo na mobilnom zariadení 61s, takže ide o rozdiel 9s. Pre zistenie štatistickej významnosti sme taktiež aplikovali Welch's t-test na základe ktorého sa nám ukázalo, že identifikovaný rozdiel nie je štatisticky významný. Získané hodnoty sú: $t\text{-value} = 0.30232$, $p\text{-value} = 0.767582$. Navyšme sme vypočítali 95% confidence interval, ktorý nadobudol hodnoty: $-20 : 37.5$. Keďže tento interval prekročil hodnotu 0, taktiež vidíme, že zistený rozdiel nie je štatisticky významný.



Obrázok 3: Dĺžka trvania úloh

Identifikované problémy spojené s použiteľnosťou

Vďaka detailnej analýze nahrávok sedenia participantov sme vytvorili zoznam problémov spojených s použiteľnosťou danej aplikácie, pričom sme rozlišovali, či bol daný problém identifikovaný na mobilnom zariadení alebo pomocou emulácie na počítači. Tabuľka 1 obsahuje kompletný zoznam identifikovaných problémov s použiteľnosťou, alebo problémových situácií, ktoré nastali počas realizácie experimentu.

Tabuľka 1 - identifikované problémy s použiteľnosťou

Typ chyby	Úloha	Opis	Počet výskytov na PC	Počet výskytov na mobile	Celkový počet výskytov
Väčší problém	4	Participant nevie nájsť možnosť prídania TV stanice do TV programu (očakáva ju priamo pri tv programe)	9	8	17
Väčší problém	5	Participant nebol schopný pridať na hlavnú stránku novú sekciu	2	3	5
Menší problém	4	Participant hľadal TV stanicu v rámci vyhľadávania	7	7	14
Menší problém	1	Zoznam filmografií osoby je príliš dlhý, keďže jednotlivé podkategórie sú uvedené za sebou ako súčasť zoznamu a participant si delenie na kategórie všimol iba náhodou po dlhšom scrollovaní v rámci tohto zoznamu, alebo si delenie na podkategórie ani nevšimol.	7	6	13
Menší problém	2	Participant nerozumie rozdielu medzi položkami menu "Oblúbené" a "Chci vidieť"	4	5	9
Menší problém	1	Participant v menu hľadal položku "hľadať"	4	2	6

		(najsťôr si nevšimol vrchnú lištu určenú na vyhľadávanie)			
Menší problém		Keď sa používateľ nachádza na konkrétnej položke menu (napr. "Hlavná stránka"), vysunie menu a chce znovu kliknúť na túto položku, aplikácia nijako nereaguje a ani menu sa neschová, takže pokiaľ participant nevie že sa aktuálne na tejto položke už nachádza má pocit, že aplikácia nereaguje	4	2	6
Menší problém	Prihlásenie	Participant na prihlásenie využil položku menu "Muj účet", ktorá ho po vyplnení údajov prihlásila ale na obrazovke sa zobrazila chybová hláška "Pri načítaní dat došlo k chybe".	2	3	5
Menší problém	2	Participant očakáva v položke menu "Chci videt" možnosť pridať položku priamo cez tento zoznam	3	2	5
Menší problém	1	Participant si najskôr nevšimol pri detaile postavy hornú lištu kde sa mohol prepnúť na položku "filmografie"	1	2	3
Menší problém	6	Participanta miatlo, že mal pri vyhľadávaní kina k dispozícii 2 vyhľadávania (aj to v hornej lište a na vyhľadávanie používal to)	0	1	1
Kozmetický problém	5	Participant očakáva tlačidlo na pridanie novej sekcie priamo na hlavnej stránke	2	4	6
Kozmetický problém	1	Participant v hornej vyhľadávacej lište hľadal tlačidlo potvrdenia zadaného dopytu.	3	0	3
Kozmetický problém	5	Pri pridaní položky cez nastavenia si nebol participant istý, či sa jeho voľba uložila alebo nie	1	2	3
Kozmetický problém	2	Participant očakával v aplikácii zoznam "Chci videt" na úvodnej obrazovke aplikácie	2	0	2
Kozmetický problém		Participant si nevšimol, že menu je scrollovatelne	2	0	2
Kozmetický problém	3	Participant očakáva rebríčky na hlavnej obrazovke aplikácie	1		1
Kozmetický problém	Prihlásenie	Participant očakával, že po kliknutí na logo aplikácie v hornej lište bude presmerovaný na hlavnú stránku	0	1	1
Nesúhlasím s tým, že ide o problém s použiteľnosťou	3	Participant uviedol položku z nesprávneho rebríčka	2	5	7

V rámci používateľskej štúdie sa podarilo identifikovať 19 unikátnych problémov z čoho 17 problémov bolo identifikovaných pomocou emulácie na počítači a 16 problémov pri interakcii s mobilným zariadením. Pomocou emulátoru bolo identifikovaných 89% problémov, zatiaľ čo pomocou mobilného zariadenia 84% problémov. Ide teda o rozdiel 5%. Pre zistenie štatistickej signifikancie sme aplikovali N-1 Two proportion test a ukázalo sa, že zistený rozdiel nie je štatisticky signifikantný. Nadobudnutá hodnota *p-value*: 0.635826. Z celkovej množiny 19 identifikovaných problémov bolo 14 problémov identifikovaných na oboch typoch zariadenia z čoho môžeme usúdiť, že väčšinu problémov ktoré sa podarilo identifikovať pri

interakcii na mobilnom zariadení sa podarilo identifikovať aj pomocou emulácie na počítači, čo potvrdzuje našu základnú hypotézu. Dôvodom problémov, ktoré boli identifikované iba na jednom type zariadenia je mierne odlišný spôsob interakcie zo zariadením, avšak tieto problémy nastali iba u malého počtu používateľov a nepredstavovali závažné problémy. Najzávažnejšie problémy ktoré boli identifikované a nastávali u najväčšieho počtu participantov sa podarilo identifikovať pomocou oboch typov zariadení.

Neočakávané udalosti

V rámci realizácie používateľskej štúdie na mobilnom zariadení bola iníciaľna konfigurácia veľmi náročná, keďže na nahrávanie obrazovky mobilného zariadenia (a teda celkovej interakcie) je využívaná externá kamera. Nastavenie tejto externej kamery bolo pomerne časovo zdĺhavé, keďže bolo potrebné nájsť takú konfiguráciu pri ktorej by bola kamera schopná dostatočne kvalitne zaznamenávať obrazovku mobilného zariadenia pri daných svetelných podmienkach a kontraste obrazovky mobilného zariadenia,

Navyše počas pilotného experimentu na mobilnom zariadení sme zistili, že nástroj Tobii nebol schopný zaznamenávať zvuk. Preto sme počas realizácie ostrého sedenia nahrávali zvuk externe pomocou dodatočného mobilného zariadenia. Nakoniec sa však pri vyhodnocovaní experimentu ukázalo, že nástroj Tobii zvuk nahral, iba ho nebol schopný na danom počítači prehrať.

Ďalším problémom, ktorý nastával počas realizácie štúdie na mobilnom zariadení bolo, že po ukončení nahrávania sedenia v rámci nástroja Tobii takmer vždy nástroj Tobii spadol a následne teda výsledná nahrávka sedenia bola značne dlhšia ako bolo reálne sedenie, keďže na konci vždy obsahovala prázdnu časť nahrávky.

Čo sa podarilo

Vďaka tomuto experimentu sa nám podarilo potvrdiť našu základnú hypotézu ktorou bolo, že pomocou emulácie mobilných aplikácií na počítači dokážeme odhaliť určité problémy spojené s ich použiteľnosťou. V rámci vyhodnotenia uskutočnenej používateľskej štúdie sme identifikovali niekoľko problémov spojených s použiteľnosťou danej aplikácie a takmer všetky problémy, ktoré sa podarilo identifikovať pomocou realizácie na mobilnom zariadení sa podarilo identifikovať aj pomocou emulácie na počítači.

Samozrejme existujú špecifické problémy spojené s použiteľnosťou, ktoré môžu vyplývať priamo z kontextu mobilného zariadenia a teda tieto problémy nemôžu nastať počas emulácie na mobilnom zariadení. Avšak naším cieľom nie je pomocou emulácie identifikovať všetky potencionálne problémy, ale umožniť efektívnejšie hromadné testovanie mobilných aplikácií počas ktorého chceme odhaliť základné problémy s použiteľnosťou danej aplikácie. Čo sa nám vďaka realizácii tohto experimentu potvrdilo, že možné je.

Navyše sa nám pomocou tohto experimentu podarilo ukázať, že úspešnosti jednotlivých úloh, účinnosť participantov a dĺžky trvania jednotlivých úloh nie sú nijako výrazne závislé od typu zariadenia na ktorom je úloha vykonávaná.

**B Protokol štúdie 2: Overenie identifikovaného zoznamu
potenciálnych problémov**

Protokol experimentu

Kontext experimentu

Meno a priezvisko Peter Dubec

Supervisor prof. Mária Bieliková

Názov projektu Podpora uskutočňovania kvantitatívnych používateľských štúdií mobilných aplikácií s využitím sledovania pohľadu

Kľúčové slová sledovanie pohľadu, testovanie mobilných aplikácií, kvantitatívne testovanie, emulácia

Stručný opis projektu V dnešnej dobe je testovanie použiteľnosti mobilných aplikácií s využitím sledovania pohľadu uskutočňované ako klasické testovanie použiteľnosti. Taktiež aktuálne je veľmi náročné uskutočňovať testovanie použiteľnosti mobilných aplikácií pomocou viacerých účastníkov testovania súčasne. Na základe toho sme schopní zbierať počas testovania iba malé množstvo dát, a preto ide skôr o kvalitatívne a nie kvantitatívne testovanie. Naším hlavným cieľom je umožniť hromadné testovanie mobilných aplikácií pomocou emulácie na počítači a zbieraním dát počas interakcie na počítači. Aj keď interakcia s počítačom je odlišná ako interakcia s mobilným zariadením, predpokladáme, že existujú určité problémy s použiteľnosťou mobilných aplikácií, ktoré budeme schopní identifikovať aj pomocou takejto emulácie. Taktiež predpokladáme, že vďaka emulácii budeme schopní zbierať kvalitnejšie dáta zo sledovača pohľadu.

Stručný opis projektu v AJ Nowadays usability testing of mobile applications with use of eye-tracking is realized as traditional usability testing. It is very difficult to perform mobile usability testing on multiple participants simultaneously, therefore in most cases these tests are qualitative and not quantitative because we are able to collect only small amount of data. Our main goal is to make bulk testing of mobile applications possible by means of simulating mobile applications on PC and collecting data from PC. Interaction with PC is quite different than interaction with smartphone, but we believe that there are certain usability problems which we will be able to identify also during emulation on PC and also we will be able to collect more precise data from eye-tracking thanks to bigger displaying screen.

Príprava experimentu

Cieľ experimentu Cieľom experimentu je uskutočniť kvantitatívne testovanie mobilnej aplikácie na mobilnom zariadení a pomocou emulácie na počítači, aby sme získali potrebné dáta pre overenie našich hypotéz.

Hypotézy

1. Pomocou emulácie mobilných aplikácií na počítači dokážeme odhaliť väčšinu problémov spojených s použiteľnosťou danej aplikácie, ktoré sa podarí odhaliť pomocou mobilného zariadenia, v prípade, že nejde o problémy úzko späté s kontextom mobilného zariadenia.

Parametre experimentu

1. Testovanie mobilnej aplikácie na mobilnom zariadení
2. Testovanie mobilnej aplikácie pomocou emulácie na počítači

Účastníci 24 participantov (študenti a zamestnanci FIIT STU)

Sledované metriky

- Úspešnosť splnenia úloh
- Priemerná úspešnosť participantov
- Počet identifikovaných problémov s použiteľnosťou

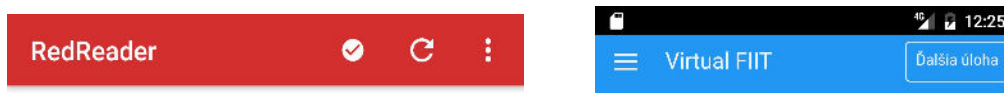
Scenár experimentu

Experiment bude prebiehať v dvoch behoch, pričom participantí budú rozdelení tak, aby jednotlivé behy obsahovali rovnomerný počet účastníkov testovania. Ide o hromadné testovanie, takže všetci účastníci jedného behu budú pracovať paralelne. Účastníci budú počas testovania pracovať s dvoma mobilnými aplikáciami, pričom s jednou budú pracovať na svojom vlastnom mobilnom zariadení a s druhou pomocou emulácie na počítači. Dôvodom využitia dvoch aplikácií je eliminácia faktoru učenia sa, ktorý by následne mohol ovplyvniť výsledky experimentu.

Experiment prebieha v dvoch behoch, aby bolo možné obe aplikácie otestovať aj na mobilných zariadeniach a aj pomocou emulácie, aby sme následne mohli porovnať identifikované problémy s použiteľnosťou. Účastníci prvého behu experimentu budú testovať aplikáciu VirtualFIIT na svojich mobilných zariadeniach a aplikáciu RedReader pomocou emulácie na počítači. Účastníci druhého behu experimentu budú mať testované aplikácie vymenené t.j. budú testovať aplikáciu RedReader na svojich mobilných zariadeniach a následne aplikáciu VirtualFIIT pomocou emulácie na počítači.

Aktivita účastníkov vykonávaná na PC je nahrávaná pomocou infraštruktúry Tobii, zatiaľ čo aktivita účastníkov pracujúcich na svojich mobilných zariadeniach je zaznamenávaná pomocou softvéru určeného na nahrávanie obrazovky mobilného zariadenia. Navyše aktivita účastníkov priamo v testovaných aplikáciách je zaznamenávaná pomocou nami implementovaných zaznamenávačov, ktoré boli vložené priamo do testovaných aplikácií. Tieto zaznamenávače zaznamenávajú kliknutia a scroll používateľov a zaznamenané dáta automaticky zasielajú na server.

Tak ako je znázornené na Obrázku 1, do oboch aplikácií bolo doplnené tlačidlo, ktorým účastníci značia, že ukončili práve vykonávanú úlohu a začínajú pracovať na nasledovnej úlohe. Vďaka tejto funkcionalite máme zabezpečené zaznamenávanie času začiatku a konca jednotlivých úloh. Navyše pokiaľ daná úloha vyžaduje od participanta odpoveď, po stlačení tlačidla pre potvrdenie ukončenia úlohy je participant vyzvaný aby priamo v aplikácii zadal odpoveď na práve vykonanú úlohu a táto odpoveď sa automaticky odošle na server.



Obrázok 1: Tlačidlá pre potvrdenie ukončenia úlohy (RedReader - check, VirtualFIIT - Ďalšia úloha)

Účastníci experimentu majú za úlohu vykonať 13 definovaných úloh v mobilnej aplikácii VirtualFIIT a následne 8 Úloh v mobilnej aplikácii RedReader. Pred samotným vykonávaním úloh účastníci testovania najskôr vyplnia úvodný dotazník.

Predpokladaná dĺžka trvania 40min.

- **Pred vstupom účastníkov:**

- Nastaviť projekciu úvodnej prezentácie
- Zapnúť potrebný počet počítačov
- Na každom PC:
 - Zapnúť Android emulátor
 - Do emulátora nainštalovať potrebnú aplikáciu
 - Nastaviť Tobii studio na test prislúchajúcej aplikácie
 - Nastaviť Tobii Studio na predkalibračnú obrazovku, kde sa zadáva meno účastník

- **Pred začatím štúdie:**

- Previesť účastníkov testovania procesom nainštalovania potrebnej aplikácie do svojho mobilného zariadenia
- Overenie, či každý účastník má vo svojom zariadení nainštalovaný softvér pre nahrávanie obrazovky

- **Úvodný dotazník:**

1. Zadajte Váš AIS login
2. Aký OS využívate na svojom mobilnom zariadení?
 - Android
 - iOS
 - Iné
3. Poznáte a používate aplikáciu VirtualFIIT?
 - Poznám a používam
 - Poznám ale nepoužívam
 - Nepoznám a nepoužívam
4. Poznáte a používate službu Reddit?
 - Poznám a používam
 - Poznám ale nepoužívam
 - Nepoznám a nepoužívam
5. Používate/použili ste už mobilnú aplikáciu RedReader?
 - Áno
 - Nie
6. Na škále od 1 do 5, tak ako v škole, ohodnoťte svoju znalosť angličtiny. (1 najlepšie, 5 najhoršie)

• **ÚLOHY:**

Aplikácia VirtualFIIT

1. Spustíte aplikáciu VirtualFiit a prihlásite sa do aplikácie s využitím vašich AIS údajov.
 - a. **Testuje sa:** nájdiťnosť/objaviteľnosť funkcionality
2. Na základe úvodnej stránky stručne opíšte na čo slúži daná aplikácia.
 - a. **Testuje sa:** Nejednoznačnosť identity
3. Zistíte, aké sú doobedňajšie otváracie hodiny knižnice počas letných prázdnin.
 - a. **Testuje sa:** Nájdiťnosť (obsahu)
4. Zvoľte si v aplikácii, že ste pedagóg (Bieliková Mária, prof. Ing., PhD.)
 - a. **Testuje sa:** Objaviteľnosť (objavenie funkcionality v nastaveniach aplikácie)
5. Bez toho aby ste na ňu klikli, uveďte ako si interpretujete prvú zo štyroch veľkých ikon nachádzajúcich sa v dolnej časti hlavnej obrazovky. Čo podľa Vás znamená a kam očakávate, že by Vás zaviedla, keby ste na ňu klikli?
 - a. **Testuje sa:** Nepochopenie navigačných prvkov (z aplikácie budú odstránené popisy pri týchto ikonách)
6. Bez toho aby ste na ňu klikli uveďte ako si interpretujete odkaz "Rozpis" nachádzajúci sa v navigácii. Tj. čo očakávate, že sa skrýva pod týmto odkazom?
 - a. **Testuje sa:** Nejednoznačnosť navigačných prvkov (položka Harmonogram bude premenovaná na rozpis)
7. Zistíte, aká je aktuálna verzia aplikácie.
 - a. **Testuje sa:** nesprávna voľba farieb (farba pozadia a písma bude zvolená tak, aby bol text ťažko čitateľný)
8. Zapnite v aplikácii možnosť logovania akcií.
 - a. **Testuje sa:** používateľ nevie určiť, prečo v danom okamihu link/tlačidlo nie je klikateľné (disabled) - tlačidlo nie je klikateľné pokiaľ sa najskôr nezapne možnosť "Ladiaci režim"
9. Nájdite v aplikácii odkaz na iMHD (stránka mestskej hromadnej dopravy) a uveďte, kde sa tento odkaz nachádza.
 - a. **Testuje sa:** používateľ niekde očakával určitú funkcionality, ale reálne sa nachádzala inde (pravdepodobne to bude hľadať na hlavnej stránke pod ikonkou MHD = reálne sa odkaz nachádza v sekcii "odkazy")
10. Navigujte sa do časti aplikácie "Objednávka" avšak nevyplňajte daný formulár. Uveďte, ako na Vás pôsobí daný formulár a či si viete predstaviť, že by ste ho bežne používali.
 - a. **Testuje sa:** dlhé formuláre (do aplikácie bude doplnená položka menu "objednávka", ktorá bude obsahovať stránku s veľmi dlhým a komplexným formulárom.
11. Uveďte aká hodnota je uvedená v rámci formulára Objednávka v poli: "Škola"
 - a. Vstup (angl. input) nezobrazuje celý zadaný obsah a preto má používateľ problém skontrolovať zadanú hodnotu (musí napr. scrollovať)
12. Vyplňte formulár objednávka.
 - a. **Testuje sa:**
 - i. input vyžaduje špecifické formátovanie vstupu, ktoré je pre používateľa ťažko pochopiteľné
 - ii. input nie je dobre označený a používateľ nevie, čo má byť obsahom daného poľa
13. Odhlásite sa z aplikácie.
 - a. **Testuje sa:** nájdiťnosť / objaviteľnosť funkcionality

Aplikácia RedReader

1. Spustíte aplikáciu a zvolíte, že chce aplikáciu využívať ako anonymný používateľ. Na základe úvodnej stránky stručne opíšete na čo slúži daná aplikácia.
 - a. **Testuje sa:** Nejednoznačnosť identity
2. Prihlásite sa do aplikácie
 - a. **Testuje sa:** Nájdiťnosť/Objaviteľnosť funkcionality
3. Chcete vyhľadať špecifický článok podľa jeho názvu. Uveďte kde v aplikácii by ste očakávali funkciu na vyhľadávanie a kde sa naozaj táto funkcia v aplikácii vyskytuje.
 - a. **Testuje sa:** používateľ niekde očakával určitú funkcionality, ale reálne sa nachádzala inde
4. Bez toho aby ste naň klikli uveďte ako si interpretujete odkaz "Custom Location" nachádzajúci sa na hlavnej stránke aplikácie. Tj. čo očakávate, že je ukryté pod týmto odkazom?
 - a. **Testuje sa:** Nepochopenie navigačných prvkov
5. Zistíte, meno používateľa, ktorý je autorom ľubovoľného príspevku a taktiež uveďte aká je hodnota jeho link karma (angl. Link Karma)
 - a. **Testuje sa:**
 - i. Nájdiťnosť funkcionality
 - ii. Nájdiťnosť obsahu
6. Povoľte v aplikácii možnosť zobrazovania NSFW náhľadov a uveďte, čo rozumiete pod pojmom NSFW.
 - a. **Testuje sa:**
 - i. objaviteľnosť funkcionality
 - ii. Misinterpretácia obsahu (používateľ nesprávne chápe obsah - v aplikácii sú používané doménovo špecifické termíny)
7. Vyberte si ľubovoľný príspevok a označte, že sa Vám tento príspevok nepáči a teda ho chcete skryť (angl. hide)
 - a. **Testuje sa:** Nájdiťnosť funkcionality
8. Uveďte, kde by ste hľadali informáciu o aktuálnej verzii aplikácie (nie kde ste ju našli, ale kde by ste ju očakávali) a uveďte aká je aktuálna verzia.
 - a. **Testuje sa:**
 - i. používateľ očakáva iný obsah (očakával že na stránke X nájde obsah Y ale nebolo tomu tak)
 - ii. Nájdiťnosť obsahu

Priebeh experimentu

Pilot 7.12.2016

Hlavnou úlohou v rámci pilotného testovania bola príprava testovacieho prostredia na všetkých potrebných počítačoch. Na každý počítač bolo potrebné nainštalovať Android emulátor a následne do každého emulátoru nainštalovať obe aplikácie, ktoré budú predmetom testovania. Po tom ako každý emulátor mal nainštalované potrebné aplikácie bolo potrebné všetky aplikácie na všetkých počítačoch otestovať, či nami implementované zaznamenávače vložené do testovaných aplikácii korektne zaznamenávajú dáta a zasielajú ich na server. V rámci tohto kroku nenastali žiadne problémy a všetky aplikácie odosieli dáta na server korektne.

Pre zaznamenávanie obrazovky počítača a taktiež pohľadu účastníkov sme sa rozhodli využiť infraštruktúru Tobii. Na každom počítači bolo potrebné vytvoriť projekt a korektne nasetupovať Tobii. (To sme urobili na jednom PC a využili možnosť automatického exportu vytvoreného projektu na ďalšie počítače).

V rámci overenia scenára testovania sa pilotného testovania zúčastnili dvaja účastníci (študenti FIIT). Účastníci pilotného testovania prešli scenárom presne tak, ako by boli súčasťou ostrej štúdie.

V rámci pozorovania účastníkov pilotného testovania a ich spätnej väzby vplynuli nasledovné zmeny:

- Úvodný dotazník:
 - Jedna otázka bola odstránená pretože pôsobila duplicitne.
 - K otázke “Ohodnoťte svoju úroveň angličtina na škále 1-5 (1 najlepšie, 5 najhoršie)” bol pridaný dodatok “tak ako v škole”, pretože škála bola chybné interpretovaná.
 - V otázke “Používate/použili ste už mobilnú aplikáciu Reddit?” bol Reddit zmenený na konkrétnu aplikáciu ktorá sa ide testovať, teda “RedReader”.
- Aplikácie:
 - Zistilo sa, že aplikácia RedReader nesprávne zaznamenávala poradie práve vykonávanej úlohy.
 - Zistilo sa, že aplikácia VirtualFIIT nefungovala na starších verziách OS Android.
- Scénár testovania:
 - VirtualFIIT:
 - Do otázok 5 a 6 bolo na úvod doplnené “Bez toho aby ste naň klikli”, pretože účastníci pilotného testovania ani nedočítali zadanie otázky a už klikali v aplikácii.
 - Do otázky 10 bolo explicitne doplnené, aby dotazník v tejto úlohe nevypíňali, pretože účastník pilotného testovania ho automaticky začal vypíňať.
 - RedReader
 - Do otázky 4 bolo na úvod taktiež doplnené “Bez toho aby ste naň klikli”.

Ostrá štúdia 8.12.2016 - prvý beh

Prvého behu ostrej štúdie sa zúčastnilo 6 účastníkov testovania. V rámci tohto behu účastníci vykonávali najskôr úlohy v emulovanej aplikácii VirtualFIIT na počítači a následne vykonávali úlohy v aplikácii RedReader na svojom mobilnom zariadení. Toto sedenie prebehlo bez väčších komplikácií a všetky dáta sa zaznamenávali na server.

V rámci priebehu sedenia sa stalo niekoľko drobných neočakávaných okolností, ktoré však neovplyvnili celkovú štúdiu. Príkladom je napríklad skutočnosť, že viacerým účastníkom testovania sa omylom podarilo vynechať niektorú úlohu, takže nevykonali úlohy v správnom poradí. Toto nespôsobuje žiaden problém z hľadiska validity štúdie, avšak na základe tohto faktu si bude štúdia vyžadovať mierne dlhší čas pre vyhodnotenie, keďže pri niektorých účastníkoch neboli odpovede na server odoslané v správnom poradí.

V rámci priebehu testovania sa dvom účastníkom v rámci práce s emulátorom podarilo stlačiť na klávesnici klávesu “END”, čo malo za dôsledok ukončenie emulátora. O takomto správaní emulátora sme nemali žiadne poznatky a účastníkov druhého behu sme na túto skutočnosť vopred upozornili. Táto

skutočnosť však taktiež nijako neovplyvnila validitu experimentu, keďže emulátor sme potom znovu zapli a účastníci pokračovali v riešení nasledujúcich úloh. Avšak na základe tejto skutočnosti bude vyhodnotenie experimentu znovu o niečo časovo náročnejšie.

Ostrá štúdia 8.12.2016 - druhý beh

Druhého behu ostrej štúdie sa zúčastnilo 7 účastníkov testovania. V rámci tohto behu účastníci vykonávali najskôr úlohy v emulovanej aplikácii RedReader na počítači a následne vykonávali úlohy v aplikácii VirtualFIIT na svojom mobilnom zariadení. Toto sedenie prebehlo bez väčších komplikácií a všetky dáta sa zaznamenávali na server

V tomto sedení sa aj napriek upozorneniam znovu zopakovali niektoré problémy, ktoré sa vyskytli počas prvého behu, avšak validita experimentu nebola ohrozená.

Druhá ostrá štúdia 19.04.2017 - prvý beh

Prvého behu druhej ostrej štúdie sa zúčastnilo 6 účastníkov testovania. V rámci tohto behu účastníci vykonávali najskôr úlohy v emulovanej aplikácii VirtualFIIT na počítači a následne vykonávali úlohy v aplikácii RedReader na svojom mobilnom zariadení. Toto sedenie prebehlo bez väčších komplikácií a všetky dáta sa zaznamenávali na server.

Na začiatku sedenia nastalo menšie zdržanie, pretože 3 účastníkom testovania sa nedarilo do mobilu nainštalovať testovanú aplikáciu. To sa nakoniec všetkým účastníkom podarilo až po reštarte mobilného zariadenia. Počas priebehu testovania sa opäť zopakovalo pár prípadov, kedy participanti nesprávne zaznačili poradie odpovedaných otázok.

Druhá ostrá štúdia 19.04.2017 - druhý beh

Druhého behu ostrej štúdie sa zúčastnilo 6 účastníkov testovania. V rámci tohto behu účastníci vykonávali najskôr úlohy v emulovanej aplikácii RedReader na počítači a následne vykonávali úlohy v aplikácii VirtualFIIT na svojom mobilnom zariadení. Toto sedenie prebehlo bez väčších komplikácií a všetky dáta sa zaznamenávali na server

V tomto sedení sa aj napriek upozorneniam znovu zopakovali niektoré problémy, ktoré sa vyskytli počas prvého behu, avšak validita experimentu nebola ohrozená.

Zhodnotenie experimentu

Pôvodným plánom bolo mať v rámci tejto štúdie minimálne 20 účastníkov testovania, avšak vzhľadom k termínu prvého testovania sa podarilo získať iba 13 účastníkov testovania. Na základe toho sme sa rozhodli experiment replikovať v rámci letného semestra a zorganizovali sme ďalšie sedenie, ktorého sa zúčastnilo 12 účastníkov. Celkovo sa nám podarilo získať 25 účastníkov testovania, avšak keďže našim zámerom je porovnávanie výsledkov vzhľadom na typ zariadenia, potrebovali sme aby bolo zastúpenie počtu účastníkov v rámci jednotlivých skupín rovnomerné. V rámci účastníkov testovania sa z hľadiska výsledkov nevyskytol žiaden outlier a preto aby sme získali rovnomerné zastúpenie účastníkov bol z vyhodnotenia odstránený jeden účastník s najpriemernejšími výsledkami.

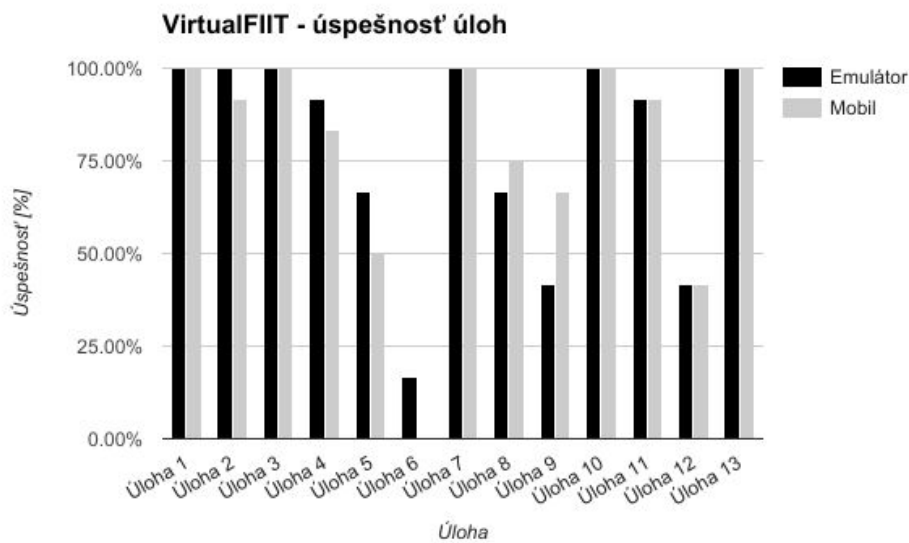
V rámci vyhodnotenia experimentu sme sledovali tri základné metriky a to:

- **Úspešnosť splnenia úloh**
- **Priemerná úspešnosť participantov**
- **Počet identifikovaných problémov s použiteľnosťou**

Úspešnosť splnenia úloh

Jednou z metrických ktoré sme sledovali je úspešnosť splnenia úloh. Táto metrika nám hovorí o tom, či daný používateľ splnil alebo nespĺnil danú úlohu. My sme vyhodnotili celkovú úspešnosť všetkých participantov vzhľadom na danú úlohu (teda koľko percent participantov úspešne splnilo danú úlohu).

Výsledky vyhodnotenia úspešnosti splnenia jednotlivých úloh v rámci aplikácie VirtualFIIT môžeme vidieť na obr.1, a v rámci aplikácie RedReader na obr.2, pričom rozlišujeme, či bola úloha vykonaná na mobilnom zariadení alebo pomocou emulácie na počítači, aby sme boli schopí porovnať, či typ zariadenia ovplyvňuje úspešnosť úloh. Experimentu sa zúčastnilo 24 participantov, čiže každú úlohu vykonalo 12 participantov na mobilnom zariadení a 12 participantov pomocou emulácie na počítači. Vidíme, že medzi úspešnosťami úloh vzhľadom na typ zariadenia nie je žiaden výrazný rozdiel.



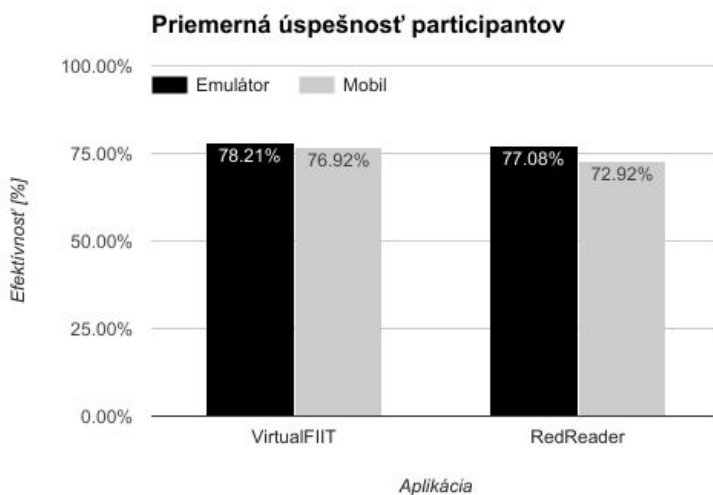
Obrázok 1: Miera úspešnosti splnenia úloh v rámci aplikácie VirtualFIIT



Obrázok 2: Miera úspešnosti splnenia úloh v rámci aplikácie RedReader

Priemerná úspešnosť účastníkov

Ďalšou sledovanou metrikou bola priemerná úspešnosť účastníkov. Úspešnosť jedného účastníka bola určená ako jeho úspešnosť splnenia úloh (teda koľko percent úloh daný účastník vykonal správne). Určili sme priemernú úspešnosť účastníkov ako priemer hodnôt jednotlivých účastníkov pre jednotlivé aplikácie a taktiež vzhľadom na typ zariadenia na ktorom bola úloha vykonaná. Vypočítané priemerné úspešnosti účastníkov môžeme vidieť na obrázku 3.



Obrázok 3: priemerná úspešnosť účastníkov

V rámci aplikácie VirtualFIIT bola priemerná úspešnosť participantov v rámci mobilného zariadenia 76.92%, zatiaľ čo v rámci emulátoru 78.21%. Vidíme, že ide o rozdiel 1.29%. Aplikovali sme Welch's t-test aby sme identifikovali, či je daný rozdiel štatisticky významný. Na základe zistených hodnôt *t-value*: 0.11041 a *p-value*: 0.913002 sa ukázalo, že daný rozdiel nie je štatisticky významný. Samozrejme, to že nám štatistický test hovorí, že daný rozdiel nie je štatisticky významný, nemusí nutne znamenať, že tu rozdiel skutočne nie je. Indikuje nám to iba to, že naše dáta nie sú dostatočné na to, aby potvrdili, že rozdiel existuje. Na základe toho sme sa pozreli aj na testovanie ekvivalencie. Pre overenie sme vypočítali 95% confidence interval, ktorý nadobudol hodnoty: -10.34 : 12.9. Vidíme, že tento interval prekročil hranicu 0 a teda rozdiel nie je štatisticky významný. Avšak pokiaľ sa zameriava na testovanie ekvivalencie, je potrebné si stanoviť hranicu ekvivalencie, teda taký rozdiel v nameraných hodnotách, kedy výsledok ešte budeme považovať za ekvivalentný. To závisí od povahy konkrétnej štúdie a v našom prípade je ťažké určiť kedy by sme mohli výsledok považovať za ekvivalentný. To by bolo potrebné určiť vzhľadom na povahu testovanej aplikácie. V rámci našich výsledkov (na základe vypočítaného confidence intervalu) vidíme, že keby stanovíme hranicu ekvivalencie rovnú 13%, tak náš confidence interval spadne do tejto hranice z oboch strán a teda môžeme namerané výsledky považovať za ekvivalentné.

Hranica ekvivalencie rovná 13% je samozrejme pomerne veľká, čo je spôsobené pomerne veľkou šírkou nášho confidence intervalu. To je však zapríčinené malým počtom účastníkov testovania a predpokladáme, že s väčším počtom účastníkov testovania by sa tento interval zúžil a teda by sme boli schopní znížiť aj hranicu ekvivalencie. Toto tvrdenie opierame o poznatky, ktoré sme nadobudli v rámci vyhodnotenia tejto štúdie. Ako bolo uvedené táto štúdia bola najskôr organizovaná s 12 účastníkmi a až neskôr sme pre získanie väčšieho počtu účastníkov zorganizovali druhé sedenie. Po ukončení prvého sedenia sme taktiež uskutočnili vyhodnotenie a v tom prípade 95% confidence interval pre priemernú úspešnosť 12 participantov v rámci aplikácie VirtualFIIT nadobudol hodnoty: -11.78 : 14.35. Vidíme teda, že zorganizovaním ďalšieho sedenia a získaním väčšieho počtu účastníkov testovania sa nám podarilo tento interval zúžiť.

V rámci aplikácie RedReader bola priemerná úspešnosť participantov v rámci mobilného zariadenia 72.92%, zatiaľ čo v rámci emulátoru 77.08%. Vidíme, že ide o rozdiel 4.16%. Aplikovali sme Welch's t-test aby sme identifikovali, či je daný rozdiel štatisticky významný. Na základe zistených hodnôt *t-value*: 0.36521 a *p-value*: 0.720414 sa ukázalo, že daný rozdiel nie je štatisticky významný. Pre zistenie ekvivalencie sme taktiež vypočítali 95% confidence interval, ktorý nadobudol hodnoty: -7.24 : 15.58. Vidíme, že tento interval prekročil hranicu 0 a teda rozdiel nie je štatisticky významný. Opäť však vidíme, že tento interval je pomerne široký a hranicu ekvivalencie by sme v tomto prípade museli stanoviť až na 15.6%.

Identifikované problémy spojené s použiteľnosťou

Vďaka detailnej analýze nahrávok sedenia participantov sme vytvorili zoznamy problémov spojených s použiteľnosťou daných aplikácií, pričom sme rozlišovali, či bol daný problém identifikovaný na mobilnom zariadení alebo pomocou emulácie na počítači. Tabuľka 1 obsahuje kompletný zoznam identifikovaných problémov s použiteľnosťou, alebo problémových situácií, ktoré nastali počas realizácie experimentu v rámci aplikácie VirtualFIIT, zatiaľ čo tabuľka 2 obsahuje problémy identifikované v rámci aplikácie RedReader.

Tabuľka 1 - identifikované problémy s použiteľnosťou v rámci aplikácie VirtualFIIT

ID	Typ chyby	Úloha	Opis	Počet výskytov na PC	Počet výskytov na mobile	Celkový počet výskytov
1	Väčší problém	6	Používateľ zle interpretuje navigačný prvok, lebo je reprezentovaný pomocou nejednoznačného označenia ("Rozpis")	10	12	22
2	Väčší problém	12	Používateľ nebol schopný správne vyplniť formulár, pretože jeden input vyžadoval špecifické formátovanie vstupu, ktoré nebolo nikde uvedené	7	7	14
3	Väčší problém	5	Používateľ zle interpretuje navigačný prvok, lebo je reprezentovaný pomocou ikony bez označenia ("ikonka kalendára")	4	6	10
4	Väčší problém	8	Používateľ nevie určiť, prečo v danom okamihu link/tlačidlo "logovanie akcií" nie je klikateľné	4	3	7
5	Väčší problém	4	Používateľ nebol schopný nájsť požadovanú funkcionality voľby pedagóga	1	2	3
6	Väčší problém	2	Používateľovi na základe iniciálnej interakcie nie je jasné, aký je účel aplikácie	0	1	1
7	Menší problém	10	Formulár je príliš dlhý a preto na používateľa pôsobí odpudzujúco a nechcel by ho pravidelne vyplňať	7	5	12
8	Menší problém	9	Používateľ očakával odkaz na MHD pod ikonku autobusu, avšak reálne sa nachádzala v menu v sekcii odkazy	7	4	11
9	Menší problém	12	Input "Kontakt" v rámci formuláru je nejednoznačne označený a pričom sa ako vstup očakával emailový kontakt používateľ zadal svoj telefonický kontakt	4	4	8

10	Menší problém	11	Input formuláru obsahoval príliš dlhú hodnotu a používateľ mal problém správne uviesť hodnotu tohto poľa	1	1	2
11	Menší problém	2	Používateľ po prihlásení nebol presmerovaný na hlavnú stránku ale na inú obrazovku a preto túto obrazovku považoval za úvodnú obrazovku	0	1	1
12	Kozmetický problém	7	Text verzie aplikácie bol napísaný svetlosivou farbou na bielom pozadí a preto bol pre používateľa ťažko čitateľný	0	5	5

Tabuľka 2 - identifikované problémy s použiteľnosťou v rámci aplikácie RedReader

ID	Typ chyby	Úloha	Opis	Počet výskytov na PC	Počet výskytov na mobile	Celkový počet výskytov v
13	Väčší problém	4	Používateľ zle interpretuje navigačný prvok, lebo je reprezentovaný pomocou nejednoznačného označenia ("Custom location")	7	9	16
14	Väčší problém	5	Používateľ nebol schopný v rámci aplikácie nájsť požadovanú funkčnosť pre zobrazenie okna s potrebnými metaúdajmi (bolo potrebné dlhé stlačenie a dlhé podržanie na základe čoho sa otvorilo kontextové okno)	9	5	14
15	Väčší problém	3	Používateľ nebol schopný nájsť funkciu vyhľadávania	5	3	8
16	Väčší problém	7	Používateľ nebol schopný otvoriť kontextové okno pre skrytie zvoleného príspevku (Bolo potrebné stlačenie a dlhé podržanie)	3	4	7
17	Väčší problém	1	Používateľovi na základe iniciálnej interakcie nie je jasné, aký je účel aplikácie	2	3	5

18	Väčší problém	5	Používateľ nebol schopný správne identifikovať požadované metaúdaje	0	1	1
19	Väčší problém	8	Používateľ nebol schopný nájsť informáciu o aktuálnej verzii aplikácie	0	1	1
20	Menší problém	3	Používateľ očakával funkciu vyhľadávania priamo v hornej lište aplikácie na úvodnej obrazovke, no reálne sa nachádzala inde	10	10	20
21	Menší problém	6	Používateľ nerozumie doménovo špecifickému termínu "NSFW"	3	3	6

V rámci tejto používateľskej štúdie sa nám v rámci oboch aplikácií podarilo spoločne identifikovať 21 problémov s použiteľnosťou. Z toho na mobilnom zariadení bolo identifikovaných všetkých 21 problémov a pomocou emulátoru 16 problémov. Z celkových 21 problémov sa teda na oboch zariadeniach podarilo identifikovať 16 problémov čo predstavuje 76%.

Hlavným cieľom tejto štúdie bolo overenie nami vytvoreného zoznamu potenciálnych problémov s použiteľnosťou mobilných aplikácií. Vybrali sme 13 problémov zo zoznamu problémov u ktorých predpokladáme, že ich je možné odhaliť pomocou emulácie a 1 problém zo zoznamu problémov u ktorých predpokladáme, že ich nebude možné odhaliť pomocou emulácie. Tieto problémy sme manuálne zanesli do testovaných aplikácií a sledovali sme, či budú odhalené. Tabuľka 3 zobrazuje namapovanie nami identifikovaných generických problémov na problémy, ktoré boli identifikované počas používateľskej štúdie.

V prípade jedného problému, ktorý bol zvolený zo zoznamu problémov u ktorých predpokladáme, že ich nebude možné odhaliť pomocou emulácie sa nám potvrdila správnosť jeho zaradenia. Ide o problém: "Farby pozadia a textu sú nevhodne zvolené a preto je text ťažko čitateľný." Naozaj môžeme vidieť, že pomocou mobilného zariadenia sa podarilo identifikovať 4 výskyty tohto problému, zatiaľ čo pomocou emulátora ani jeden.

Zo zoznamu 13 generických problémov u ktorých predpokladáme, že ich bude možné odhaliť pomocou emulácie sa nám správnosť ich zaradenia potvrdila u 12 problémov. To znamená, že sa tento problém podarilo identifikovať aj pomocou mobilného zariadenia a taktiež aj pomocou emulácie na počítači. V prípade jedného problému sa nám jeho zaradenie nepodarilo potvrdiť, pretože sa vyskytol iba na mobilnom zariadení. Ide o problém: "Nepochopenie zamerania webu/aplikácie počas iniciálnej interakcie". Avšak tento problém sa vyskytol iba jedenkrát v rámci každej z testovaných aplikácií a keďže ide o malý počet výskytov, nevyradíme ho zo zoznamu problémov u ktorých predpokladáme, že by ich bolo možné identifikovať pomocou emulácie, ale overenie zaradenia tohto problému by si vyžadovalo dodatočné experimentálne overenie.

Tabuľka 3 - sledované generické problémy a ich mapovanie na identifikované problémy

Kategória problému	Opis problému	ID problemov	Počet výskytov na PC	Počet výskytov na mobile
Nejednoznačnosť identity	nepochopenie zamerania webu/aplikácie počas iniciálnej interakcie	6, 17	0	2
Objaviteľnosť (Discoverability)	objavenie novej funkcionality o ktorej používatelia doteraz nevedeli	5, 14, 16	7	8
Nejednoznačnosť navigačných prvkov	navigačné prvky nie sú pre nových používateľov úplne jasné pretože obsahujú nejednoznačné pomenovania (používateľ nevie, kam ho navigačný prvok zavedie)	1, 13	7	10
Nejednoznačnosť navigačných prvkov	navigačné prvky nie sú pre nových používateľov úplne jasné pretože sú reprezentované pomocou neznámych ikon bez dodatočného textového vysvetlenia	3	3	4
Nesprávna voľba farieb	farby pozadia a textu sú nevhodne zvolené a preto je text ťažko čitateľný	12	0	4
Nevhodne vizualizované interaktívne elementy	používateľ nevie určiť, prečo v danom okamihu link/tlačidlo nie je klikateľné (disabled)	4	3	3
Nevhodne umiestnený obsah	používateľ očakáva iný obsah (očakával že na stránke X nájde obsah Y ale nebolo tomu tak)	8	3	2
Dlhé formuláre	formulár obsahuje príliš veľa prvkov a preto je pre používateľa odpudzujúci	7	5	2
Nevhodne vizualizované vstupné elementy	input nezobrazuje celý zadaný obsah a preto má používateľ problém skontrolovať zadanú hodnotu (musí napr. scrollovať)	10	1	1
Preddefinovaný formát vstupu	input vyžaduje špecifické formátovanie vstupu, ktoré je pre používateľa ťažko pochopiteľné	2	6	4
Nevhodne označené vstupné elementy	input nie je dobre označený a používateľ nevie, čo má byť obsahom daného poľa	9	1	3

Nájditeľnosť (Findability)	jednoduché nájdenie obsahu alebo funkcionality o ktorej používateľa predpokladajú že je prítomná	15, 18, 19	3	4
Nevhodne umiestnená funkcionality	používateľ niekde očakával určitú funkcionality, ale reálne sa nachádzala inde	20	4	6
Nepochopenie obsahu	používateľ nesprávne chápe obsah (v aplikácii sú používané doménovo špecifické termíny)	21	2	1

Čo sa podarilo

Vďaka tomuto experimentu sa nám podarilo ukázať, že typ použitého zariadenia (emulátor alebo mobilné zariadenie) nemá výrazný vplyv na metriky priemernej úspešnosti úloh a taktiež priemernej úspešnosti účastníkov.

V rámci tejto používateľskej štúdie sa nám v rámci oboch aplikácií podarilo spoločne identifikovať 21 problémov s použiteľnosťou. Z toho na mobilnom zariadení bolo identifikovaných všetkých 21 problémov a pomocou emulátoru 16 problémov. Z celkových 21 problémov sa teda na oboch zariadeniach podarilo identifikovať 16 problémov čo predstavuje 76%.

Hlavným cieľom tejto štúdie bolo overenie nami vytvoreného zoznamu potenciálnych problémov s použiteľnosťou mobilných aplikácií. Vybrali sme 13 problémov zo zoznamu problémov u ktorých predpokladáme, že ich je možné odhaliť pomocou emulácie a 1 problém zo zoznamu problémov u ktorých predpokladáme, že ich nebude možné odhaliť pomocou emulácie. Správne zaradenie problému u ktorého predpokladáme, že ho nebude možné odhaliť pomocou emulácie sa nám podarilo experimentálne overiť. Z 13 vybraných problémov zo zoznamu problémov u ktorých predpokladáme, že ich je možné odhaliť pomocou emulácie, sa nám podarilo experimentálne overiť správnosť zaradenia 12 z týchto problémov.

C Technická dokumentácia

C.1 Zaznamenávač pre natívne mobilné aplikácie

Zaznamenávač pre natívne aplikácie predstavuje modul/knižnicu pre OS Android, ktorú je možné jednoduchým spôsobom vložiť do akejkoľvek natívnej aplikácie. Zaznamenávač zaznamenáva všetky akcie používateľ a aplikácie (eventy) a zaznamenané dáta odosiela na server. Ako server pre odosielanie dát sme využili voľne dostupnú službu Keen.IO⁴², ktorá slúži ako jednoduché API pre prijímanie eventov. Zaznamenaný event môže byť dvoch typov a to *kliknutie* alebo *scroll*.

Event zaznamenaný pomocou tohto zaznamenávača obsahuje nasledovné atribúty:

- *label* - text elementu na ktorom daný event nastal, pokiaľ sa podarí extrahovať (napr. text tlačidla a pod.)
- *screen* - názov obrazovky na ktorej event nastal (aktuálne názov danej Android Activity)
- *timestamp* - časová pečiatka kedy event nastal (formát ISO 86012)⁴³
- *type* - typ eventu: click/scroll
- *toXAbsolute* - v prípade že išlo o scroll, absolútna X súradnica, kde event skončil
- *toX* - v prípade že išlo o scroll, X súradnica, kde event skončil
- *toYAbsolute* - v prípade že išlo o scroll, absolútna Y súradnica, kde event skončil
- *toY* - v prípade že išlo o scroll, Y súradnica, kde event skončil
- *uuid* - unikátny identifikátor zariadenia na ktorom event nastal
- *xAbsolute* - absolútna X súradnica miesta kde event nastal
- *x* - X súradnica miesta kde event nastal
- *yAbsolute* - absolútna Y súradnica miesta kde event nastal

⁴²<https://keen.io/>

⁴³<http://www.iso.org/iso/home/standards/iso8601.htm>

- y - Y súradnica miesta kde event nastal

Ukážka zdrojového kódu

Ukážka 1: Ukážka implementácie zaznamenávača pre natívne mobilné aplikácie

```
1 public class MyWindowCallback implements Window.Callback{
2     //global definitions
3     Window.Callback localCallback;
4     Activity context;
5     int x, y, absX, absY, offsetX, offsetY, lastEventType;
6     String ts, screen;
7
8     //constructor
9     public MyWindowCallback(Window.Callback localCallback, Activity
10         context) {
11         this.localCallback = localCallback;
12         this.context = context;
13         this.offsetX = 0;
14         this.offsetY = 0;
15     }
16
17     @Override
18     public boolean dispatchTouchEvent(MotionEvent event) {
19         if (event.getAction() == MotionEvent.ACTION_DOWN ||
20             event.getAction() == MotionEvent.ACTION_MOVE) {
21
22             this.lastEventType = event.getAction();
23
24             //log scroll event
25             if (this.lastEventType == MotionEvent.ACTION_MOVE) {
26                 if (Logger.scrollEvent == null) {
27                     Logger.scrollEvent = new MyClickEvent(
28                         "Scroll",
29                         "Unknown",
30                         Logger.lastClick.getX(),
31                         Logger.lastClick.getY(),
32                         Logger.lastClick.getAbsX(),
33                         Logger.lastClick.getAbsY(),
34                         Logger.lastClick.getEventScreen()
35                         Logger.lastClick.getTimestamp()
36                     );
37                 }
38             }
39
40             //detect click event attributes
41             View view = context.findViewById(android.R.id.content);
42             int[] screenLocation = new int[2];
```



```

43     view.getLocationOnScreen(screenLocation);
44     Float x = event.getRawX() - screenLocation[0];
45     Float y = event.getRawY() - screenLocation[1];
46     TimeZone tz = TimeZone.getTimeZone("UTC");
47     final DateFormat df = new SimpleDateFormat("yyyy-MM-dd'T'
48         HH:mm:ss'Z'");
49     df.setTimeZone(tz);
50     this.ts = df.format(new Date());
51
52     this.screen = context.getClass().getSimpleName();
53     this.x = Math.round(x);
54     this.y = Math.round(y);
55     this.absX = Math.round(event.getRawX());
56     this.absY = Math.round(event.getRawY());
57
58     View target = getClickedView(view, Math.round(x),
59         Math.round(y));
60
61     //no specific target detected
62     if (target == null) {
63         Logger.lastClick = new MyClickEvent(
64             "Raw",
65             "Unknown",
66             this.x + this.offsetX,
67             this.y + this.offsetY,
68             this.absX,
69             this.absY,
70             this.screen,
71             this.ts
72         );
73     } else {
74         }
75     } else if (event.getAction() == MotionEvent.ACTION_UP) {
76         if (this.lastEventType == MotionEvent.ACTION_DOWN) {
77             //Logger.stack.put(Logger.lastClick.stringify());
78         } else if (this.lastEventType == MotionEvent.ACTION_MOVE &&
79             Logger.scrollEvent != null) {
80             Logger.scrollEvent.setScrollCoordinates(
81                 Logger.lastClick.getX(),
82                 Logger.lastClick.getY(),
83                 Logger.lastClick.getAbsX(),
84                 Logger.lastClick.getAbsY()
85             );
86         }
87     }
88     Logger.lastClick = null;
89     return localCallback.dispatchTouchEvent(event);
90 }

```

C.2 Zaznamenávač pre webové mobilné aplikácie

Zaznamenávač pre webové mobilné aplikácie predstavuje knižnicu implementovanú v jazyku Javascript, ktorá slúži na zaznamenávanie eventov v rámci webových aplikácií. Zaznamenávač je možné veľmi ľahko vložiť do akéhokoľvek projektu a zaznamenávač následne zaznamenáva a odosiela eventy na server identicky ako v prípade zaznamenávača pre natívne aplikácie. Zaznamenávač nie je schopný určiť, na ktorej obrazovke event nastal, ale na druhú stranu je vždy schopný unikátne identifikovať element aplikácie na ktorom event nastal.

Event zaznamenaný pomocou tohto zaznamenávača obsahuje nasledovné atribúty:

- *button* - hodnota reprezentujúca, ktoré tlačidlo myši bolo stlačené ⁴⁴
- *buttons* - hodnota reprezentujúca, ktoré tlačidlá myši boli stlačené. Ak ich bolo stlačených viac, hodnota predstavuje súčet reprezentatívnych hodnôt⁴⁵
- *clientX* - X súradnica v rámci klienta
- *clientY* - Y súradnica v rámci klienta
- *pageX* - X súradnica v rámci stránky
- *pageY* - Y súradnica v rámci stránky
- *screenHeight* - výška obrazovky
- *screenWidth* - šírka obrazovky
- *screenX* - X súradnica v rámci obrazovky
- *screenY* - Y súradnica v rámci obrazovky
- *target* - cieľový element na ktorý nastal event
 - *className* - triedy cieľového elementu
 - *id* - id cieľového elementu
 - *name* - atribút name cieľového elementu
 - *path* - XPATH k cieľovému elementu

⁴⁴<https://developer.mozilla.org/en-US/docs/Web/API/MouseEvent/button>

⁴⁵<https://developer.mozilla.org/en-US/docs/Web/API/MouseEvent/buttons>

- *tagName* - názov HTML tagu cieľového elementu (div, span, button atd.)
 - *targetHeight* - výška cieľového elementu
 - *targetWidth* - šírka cieľového elementu
 - *title* - atribút title cieľového elementu
 - *type* - atribút type cieľového elementu
 - *value* - text elementu na ktorom daný event nastal, pokiaľ sa podarí extrahovať (napr. text tlačidla a pod.)
- *timestamp* - časová pečiatka kedy event nastal (formát ISO 86012)
 - *type* - typ eventu click/scroll
 - *windowHeight* - výška okna
 - *windowWidth* - šírka okna

Ukážka zdrojového kódu

Ukážka 2: Ukážka implementácie zaznamenávača pre webové mobilné aplikácie

```

1  var Logger = (function () {
2      var topics = {};
3      var hasProp = topics.hasOwnProperty;
4
5      return {
6          init: function() {
7              var rootElement = document.documentElement;
8              rootElement.addEventListener('click',
9                  this.clickListener.bind(this));
10             window.addEventListener('scroll', this.scrollListener.bind
11                 (this));
12         },
13         getPathTo: function(element) {
14             if (element.id !== '')
15                 return "//*[@id='"+element.id+"']";
16
17             if (element===document.body)
18                 return element.tagName.toLowerCase();
19
20             var count= 0;
21             var elementSiblings= element.parentNode.childNodes;
22             for (var i= 0; i<elementSiblings.length; i++) {
23                 var elementSibling= elementSiblings[i];

```

```

22
23     if (elementSibling===element)
24         return this.getPathTo(element.parentNode) +
25             '/' +
26             element.tagName.toLowerCase() +
27             '[' + (count + 1) + ']';
28
29     if (elementSibling.nodeType===1 &&
30         elementSibling.tagName === element.tagName) {
31         count++;
32     }
33 }
34 },
35 clickListener: function (e) {
36     var event = this.processEvent(e);
37     this.publish('click', event);
38 },
39 scrollListener: function (e) {
40     var event = this.processEvent(e);
41     this.publish('scroll', event);
42 },
43 subscribe: function(topic, topicCallback) {
44     if(!hasProp.call(topics, topic)) topics[topic] = [];
45
46     var i = topics[topic].push(topicCallback) -1;
47
48     return {
49         remove: function() {
50             delete topics[topic][i];
51         }
52     };
53 },
54 publish: function(topic, content) {
55     if(!hasProp.call(topics, topic)) return;
56
57     topics[topic].forEach(function(topicCallback) {
58         topicCallback(content != undefined ? content : {});
59     });
60 }
61 }
62 }) ();

```

C.3 Používateľská príručka

Používateľská príručka slúži ako návod pre využitie nami vytvorených zaznamenaných akcií z mobilných aplikácií.

Zaznamenávač pre natívne mobilné aplikácie

Pre zaznamenávanie eventov z natívnej mobilnej aplikácie pomocou nami implementovaného zaznamenávača je potrebné vykonať nasledovné kroky:

1. Pridať súbor *logger.aar* do projektu ako nový modul. Príklad cez prostredie Android Studio: (1) File -> New -> New Module, (2) Import .JAR/.AAR package, (3) vybrať súbor *logger.aar* a zadať názov *logger*
2. V rámci súboru *settings.gradle* pridať riadok *include ':logger'*
3. V rámci súboru *build.gradle* do časti *dependencies* pridať riadok *compile project(':logger')*
4. Vo všetkých Aktivitách, ktoré môžu byť vstupným bodom do aplikácie inicializovať zaznamenávač tak, ako na Ukážke 3

Ukážka 3: Inicializácia zaznamenávača pre natívne mobilné aplikácie

```
1 //import Logger
2 import com.example.logger.Logger;
3
4 /*
5  * initialize logger
6  * @params:
7  * - uuid - unique device identifier
8  * - true/false - send events to Keen API
9  */
10 Logger.start(uuid, true);
```

Zaznamenávač pre webové mobilné aplikácie

Pre zaznamenávanie eventov z webovej mobilnej aplikácie pomocou nami implementovaného zaznamenávača je potrebné vykonať nasledovné kroky:

1. Vložiť súbor *logger.js* do koreňového adresára projektu
2. Importovať zaznamenávač do projektu. Napr. v *index.html* v sekcii *head* pomocou riadku `<script src='logger.js'></script>`
3. Spustiť zaznamenávač (napr. v *app.js*) a vytvoriť callback pre spracovanie eventov, tak ako je znázornené na Ukážke 4

Ukážka 4: Inicializácia zaznamenávača pre webové mobilné aplikácie

```
1 //start Logger
2     Logger.init();
3     //listen for click events
4     Logger.subscribe('click', function(e) {
5         console.log(e);
6     });
7     //listen for scroll events
8     Logger.subscribe('scroll', function(e) {
9         console.log(e);
10    });
```

D Zhodnotenie plnenia plánu

D.1 Zhodnotenia plánu v rámci DP2

V rámci tohto semestra som pracoval na vyhodnocovaní uskutočnenej prvej používateľskej štúdie, vytvorení zoznamu potenciálnych problémov s použiteľnosťou mobilných aplikácií, implementácii zaznamenávačov pre zaznamenávanie aktivity v mobilných aplikáciách a taktiež na návrhu a realizácii druhej používateľskej štúdie za účelom overenia správnosti identifikovaného zoznamu potenciálnych problémov s použiteľnosťou mobilných aplikácií.

V rámci plánu ktorý som si definoval pre tento semester bol prvým kontrolným bodom dátum 29.6.2016, do kedy som si stanovil za cieľ vyhodnotenie uskutočnenej používateľskej štúdie. Tento bod sa mi podarilo splniť včas, keďže sa počas vyhodnocovania nevyskytli žiadne neočakávané udalosti.

Druhým kontrolným bodom bol dátum 30.9.2016 do kedy som mal v pláne vytvorenie zaznamenávača pre OS Android. Tento plán sa mi nepodarilo dodržať, keďže nemám príliš veľké skúsenosti s vývojom mobilných aplikácií. Narazil som na implementačné problémy, ktorých vyriešenie mi trvalo pomerne veľmi dlhú dobu. Ukázalo sa, že implementácia zaznamenávača v rámci natívnej mobilnej aplikácie, ktorý by bol schopný odchytať všetky interakcie v rámci danej aplikácie nie je triviálnou záležitosťou.

Navyše počas analýzy mobilných aplikácií som narazil na fakt, že pomerne veľkú časť aplikácií tvoria nie len natívne aplikácie ale aj webové mobilné aplikácie. Na základe tohto faktu som sa rozhodol pre vytvorenie druhého zaznamenávača, ktorý by umožňoval zaznamenávanie aktivity účastníkov testovania aj v takýchto aplikáciách. Keďže v tomto prípade išlo o webové technológie, ktorým sa venujem, implementácia takéhoto zaznamenávača nebola časovo príliš náročná.

Keďže implementácia zaznamenávača pre natívne aplikácie mi trvala pomerne dlhú dobu, začal som paralelne s implementáciou zaznamenávača pracovať aj na ďalšej úlohe ktorou bolo vytvorenie zoznamu potenciálnych problémov s použiteľnosťou. Tento bod v pôvodnom pláne nebol, avšak ukázal sa ako nevyhnutný pre tvorbu našej metódy používateľského testu, keďže na základe tohto zoznamu sa určuje či použitie našej metódy je alebo nie je vhodné pre danú používateľskú štúdiu.

Posledným kontrolným bodom, resp. dvoma bodmi v rámci plánu bol návrh, realizácia a vyhodnotenie druhej používateľskej štúdie.

Počas vytvárania zoznamu potenciálnych problémov sme sa zhodli, že realizácia tejto štúdie bude vhodná na overenie tohto zoznamu a zároveň aj na overenie vytvorených zaznamenávačov. Preto som najskôr musel dokončiť stanovené úlohy a až následne som mohol začať pracovať na návrhu používateľskej štúdie.

Návrh a realizácia tejto štúdie mi taktiež zabrala značne väčší čas ako som pôvodne plánoval. Hlavným dôvodom bolo, že v rámci štúdie som mal v pláne testovať dve aplikácie, ktoré som musel najskôr sfunkčnit' na lokálnom vývojom prostredí, aby som mohol v týchto aplikáciách spraviť potrebné modifikácie. Pri tomto kroku som však narazil na niekoľko technických problémov, ktorých odstránenie sa z niekoľkých hodín predĺžilo až na niekoľko dní. Hlavným problémom na ktorý som narazil pri oboch zvolených aplikáciách, bola nekompatibilita závislostí (angl. dependencies), ktoré tieto aplikácie využívali. V čase vývoja týchto aplikácií boli zrejme využívané iné verzie závislostí, ako tie, ktoré sa mne nainštalovali v rámci iniciálnej inštalácie. Následne tieto aplikácie ani nebolo možné spustiť a bolo časovo pomerne náročné identifikovať, čo tento problém spôsobovalo.

Druhým hlavným problémom bolo to, že tieto aplikácie neboli takmer vôbec dokumentované. Na základe toho bola práca s nimi pomerne náročná.

Nakoniec sa mi štúdiu podarilo zrealizovať až 8.12.2016 a preto vyhodnotenie štúdie presúvam do plánu pre ďalší semester, ktorý sa nachádza v Tabuľke 5.

Tabuľka 5: Plán práce na ďalšie obdobie

Dátum	Opis
16.1.2017	Vyhodnotenie uskutočnenej používateľskej štúdie
31.1.2017	Úprava zaznamenávača pre natívne aplikácie tak, aby bol schopný zaznamenávať všetku aktivitu používateľov
15.2.2017	Výber vhodných metrík pre automatickú analýzu
12.3.2017	Návrh a implementácia prototypu skriptov pre automatickú analýzu

25.3.2017	Návrh používateľskej štúdie za účelom overenia automatického vyhodnotenia
16.4.2017	Realizácia a vyhodnotenie používateľskej štúdie
-	Na základe výsledkov uskutočnenej štúdie zlepšovanie výsledkov podpory automatického vyhodnocovania a následná realizáciu štúdie pre ďalšie overenie. Alebo návrh a realizácia štúdie pre identifikáciu optimálnej zobrazovacej veľkosti v rámci emulátora. Následne finalizácia diplomovej práce podľa potrieb.

D.2 Zhodnotenia plánu v rámci DP3

Pôvodná téma našej práce znela: *"Podpora automatického vyhodnocovania používateľských štúdií založených na sledovaní pohľadu"*. V rámci analýzy sme sa rozhodli špecificky zamerať na oblasť testovania mobilných aplikácií, keďže sa ukázalo, že v rámci tejto oblasti existuje pomerne málo prístupov. Navyše sme sa špecificky zamerali na hromadné kvantitatívne používateľské štúdie mobilných aplikácií, ktorých realizácia je v rámci súčasných prístupov veľmi náročná. Na základe toho prešla po obhajobe DP2 naša práca revíziou a téma práce sa špecifikovala na: *"Podpora uskutočňovania kvantitatívnych používateľských štúdií mobilných aplikácií s využitím sledovania pohľadu"*.

Na základe týchto skutočností sme sa značne odklonili od plánu definovaného v rámci predošlého semestra.

Prvým kontrolným bodom bol dátum 16.1.2017 do kedy sme si stanovili ako cieľ vyhodnotenie uskutočnenej druhej používateľskej štúdie. Tento cieľ sa nám podarilo splniť a zistené výsledky podporili našu testovanú hypotézu.

Druhým kontrolným bodom bol dátum 31.1.2017, ktorého cieľom bola úprava zaznamenávača pre natívne mobilné aplikácie. V rámci realizovanej používateľskej štúdie sa ukázalo, že tento zaznamenávač mal problém so zaznamenávaním interakcie v prípade špecifickej situácie, kedy sa v aplikácii vyskytovali prekrývajúce sa okná (ang. pop-up / modal). Keďže sme sa v rámci našej práce odklonili od pôvod-

ného zámeru, ktorým bola automatizácia vyhodnocovania, rozhodli sme sa tomuto problému nevenovať príliš veľké množstvo času, keďže išlo o veľmi špecifickú situáciu. Nefunkčnosť zaznamenávača v prípade takejto špecifickej situácie sme preskúmali, avšak nepodarilo sa nám nájsť riešenie pre tento problém.

Ďalšie kontrolné body sa primárne zameriavali na automatizáciu vyhodnocovania a keďže sme sa od toho zámeru odklonili, týmto bodom sme sa už nevenovali.

Pôvodné vyhodnotenia realizovaných používateľských štúdií obsahovali vyhodnotenie zvolených metrick a vytvorenie zoznamov identifikovaných problémov. Rozhodli sme sa zistené výsledky podporiť štatistickým vyhodnotením a preto sme spätne pre obe realizované štúdie uskutočnili štatistické vyhodnotenie.

Keďže druhá používateľská štúdia bola realizovaná koncom semestra, zúčastnilo sa jej iba 12 účastníkov testovania. Na základe toho sme sa rozhodli replikovať túto štúdiu a zorganizovali sme druhý beh tejto štúdie, ktorý bol identický s predošlým behom. Pri prípravách tejto štúdie sme narazili na problém, kde oficiálny server jednej z testovaných aplikácií, stratil svoju HTTPS certifikáciu a na základe toho aplikácia nebola schopná aktualizovať svoje dáta a preto ju nebolo možné ani spustiť. Museli sme sa teda venovať tomuto problému, keďže išlo o replikáciu pôvodnej štúdie, nemohli sme nájsť inú testovaciu aplikáciu, ale museli sme zabezpečiť, aby boli využité identické aplikácie ako v prípade prvého behu. Tento problém sa nám nakoniec podarilo ošetriť a prebehol druhý beh štúdie, ktorého sa taktiež zúčastnilo 12 účastníkov. Následne sme dáta od nových účastníkov vyhodnotili a opäť sme realizovali štatistické vyhodnotenie, teraz už s 24 účastníkmi testovania.

V neposlednom rade sme sa venovali písaniu článku na študentskú vedeckú konferenciu IIT.SRC 2017, príprave článku na vedeckú konferenciu MobileHCI 2017 a finalizácií dokumentu diplomovej práce.

E Obsah elektronického média

K dokumentu priložené elektronické médium má nasledovnú štruktúru:

/doc

- dokumenty diplomovej práce

/doc/latex

- súbory dokumentácie vo formáte Latex

/doc/iit-src.pdf

- článok publikovaný na študentskej vedeckej konferencii IIT.SRC 2017

/doc/iit-src-poster.pdf

- poster prezentovaný na študentskej vedeckej konferencii IIT.SRC 2017

/src

- zdrojové kódy implementovaných zaznamenávačov

/src/native-logger

- zdrojové kódy zaznamenávača pre natívne mobilné aplikácie

/src/web-logger.js

- zdrojové kódy zaznamenávača pre webové mobilné aplikácie

readme.txt

- popis obsahu média

F Článek publikovaný na konferenci IIT.SRC 2017

Towards Quantitative Eye-tracking User Studies of Mobile Applications

Peter DUBEC*

*Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
peter.dubec93@gmail.com*

Abstract. Nowadays eye-tracking user studies aimed at usability evaluation of mobile applications are realized on individual basis and mostly produce just qualitative data. Main goal of our work is to improve effectiveness of quantitative usability testing of mobile applications with the use of eye-tracking. We propose a user study method, which enables quantitative eye-tracking usability testing of mobile applications thanks to emulation on computer.

Introduction

Importance and complexity of mobile applications is nowadays still rising. Effective application design enables a user to fully concentrate solely on information which helps him to fulfill information need. On the other hand, good analysis of user's interaction with application helps us to better evaluate application's usability and design applications, which can be used more effectively.

Multiple research methods are being used for studying and evaluation of application's usability. In this paper we are focusing solely on usability testing based on user studies. Two main types of user studies are known - qualitative and quantitative. Qualitative user studies are based on detailed monitoring of user behavior during interaction with application and their output is in most cases list of identified usability problems. Since number of test participants is usually very small (3-5 participants) we are not able to obtain any statistically significant data. Quantitative usability studies are realized with bigger sample of participants and their main goal is to prove or derive specific facts by measuring and evaluating different metrics. Output of quantitative usability studies are quantified statistical data about application's usability.

Main problem of quantitative usability studies is that detailed monitoring of users interaction is quite challenging. Thanks to use of eye-tracking we are able to obtain detailed information about users behavior even during quantitative usability studies.

Usability testing on mobile devices

Principles and methods used for mobile usability evaluation are primarily based on techniques used for desktop environments. Techniques used for mobile devices are facing three main problems [4]:

1. Small display sizes - detailed user monitoring is very challenging.
2. Lack of software tools specifically oriented on mobile devices - we are not able to easily monitor user's interaction.
3. Other problems resulting from context of mobile devices - user and device are almost always moving and that can have big impact on interaction.

Regarding second mentioned point, we can find multiple online tools such as UXCam¹, allowing us to monitor user's interaction on mobile device. However these tools are not able to provide detailed monitoring of user's interaction which can be achieved thanks to eye-tracking.

Based on third mentioned point it can be questionable whether laboratory testing of mobile applications usability can be considered relevant. Multiple studies were focused on comparison of field and laboratory testing of mobile applications and they showed that in many cases most relevant data were obtained during laboratory testing [5]. On the other hand, study [3] shows that during field testing significantly greater

* Master study programme in field: Information Systems

Supervisor: Professor Mária Bielíková, Institute of Informatics, Information Systems and Software Engineering, Faculty of Informatics and Information Technologies STU in Bratislava

¹ <https://uxcam.com/>

amount of usability problems were discovered. However, in this case most of problems discovered were usability problems which arise thanks to field usage of mobile device, where participant was influenced by multiple external factors. Accordingly to studies [1, 3, 5] we can not clearly define whether more usability problems can be discovered in laboratory or field testing, because it clearly depends on type of defined tasks, but it is clear that laboratory testing can be considered relevant.

Eye-tracking on mobile devices

Usability testing and eye-tracking on mobile devices introduces multiple challenges. As study [2] shows on such small display sizes from distance of 50 cm only one fixation is necessary for the brain to get accurate image about approximately quarter of the display.

Most commonly used techniques for mobile eye-tracking are: 1) Head-mounted eyetrackers, 2) The below-table setup, 3) Mobile stand (stand alone eyetracker) and 4) Emulator setup.

We have decided to use emulator setup. Main reason for choosing emulator setup was, that we see great potential in quantitative usability studies of mobile applications with use of eye-tracking, but such studies can not be realized on real mobile devices on such scale. However infrastructure available at our faculty allows us to organize quantitative user studies on such scale thanks to emulation on PC [6].

Since type of interaction with emulator is quite different during usability testing of mobile applications, question whether usability problems can be revealed arise. Comparison of three different approaches for mobile usability testing was realized in study [1]. Studied approaches were: 1) emulator setup in specialized laboratory, 2) testing on mobile device in specialized laboratory and 3) field testing with mobile device. To compare this three approaches, usability study with 36 participants was conducted and following numbers of usability problems were identified: 1) laboratory testing with mobile device - 41 problems, 2) field mobile device testing - 38 problems and 3) laboratory emulator setup - 35 problems. This study shows that usability problems with mobile application can be discovered even with use of emulator setup.

Method for quantitative eyetracking usability testing of mobile applications

We have analyzed multiple approaches of eye-tracking on mobile devices and after considering pros and cons

we have decided to use eye-tracking of mobile applications thanks to use of PC emulation. We focus our work on mobile operation system Android² and we have chosen Genymotion³ as emulator used for our method, however any emulator can be used.

In this work we propose method for quantitative usability testing of mobile applications with use of eye-tracking which consists of the following steps:

1. Test scenario definition
2. Emulator use suitability verification
3. Test protocol definition
4. Participant material creation
5. Test environment setup + eyetracking setup
6. Pilot testing
7. Usability testing based on specified protocol
8. Data analysis
9. Creation of prioritized usability problems list and verification of identified usability problems

Test scenario definition. Test scenario represent exact description of tasks, which will be performed by test participants.

Emulator use suitability verification. We have defined two lists of possible usability problems of mobile applications. First list contains usability problems that can be identified with the use of emulation such as: *absence of navigation elements, long forms, content informativeness, discoverability or findability.*

Second list consists of problems, that can not be identified with use of emulation. These problems are mostly strictly related to context of mobile devices. Examples of such problems are: *problems associated with use of camera, problems associated with specific mobile device sensors, wrong use of colors or small clickable elements.*

After scenario definition it is necessary to verify emulator suitability for specified study by going through our identified lists of potential usability problems and identifying which problems are possible to occur. When majority of potential problems is from list of problems that can be identified with use of emulation, than use of our method using emulation is appropriate, otherwise it is not.

Test protocol definition and participant material creation. This step can be realized in parallel with

² <https://www.android.com/>

³ <https://www.genymotion.com/>

test scenario creation. Test protocol represents complete picture about planned usability study. In test protocol main goal of conducting usability study, expected outputs and exact course of planned usability study should be defined. During test protocol definition it is necessary to also define specific requirements for participants and required number of participants.

Test environment setup. Process of environment setup consists of multiple steps: 1) emulator installation, 2) mobile application installation, 3) emulator setup (required projection size or other settings) and 4) eye-tracking software configuration.

Pilot testing and usability testing based on specified protocol. Before conducting usability study with specified number of participants pilot testing with minimal number of two participants should be conducted. During pilot testing participants should perform specified tasks as they would during normal study and main goal of this testing is to verify whether test protocol is correctly specified and whether everything is working as expected. After pilot testing and incorporation of necessary changes if some were identified during pilot testing, normal study with specified number of participants can be conducted.

Data analysis. After conducting of usability study collected data should be analyzed and defined metrics should be evaluated. In this step also list of identified usability problems should be created.

Creation of prioritized usability problems list and verification of identified usability problems. One of outputs from previous step is list of identified usability problems. Since study was conducted using emulation on PC, correctness of identified problems should be validated. It is necessary to go through list of identified problems and map each problem on list of potential usability problems that we have defined. If problem is mapped on item from list of problems that can not be identified with use of emulation, that problem should be removed from list of identified problems, because it can not be considered valid.

Method validation

To validate our proposed method we have conducted two usability studies. Main goal of the first study was verification of our main hypothesis that some usability problems can be identified even with use of emulation. Main goal of the second study was to verify our defined list of potential usability problems.

Study 1. This study was conducted with 18 participants from which each performed seven defined tasks (login + six tasks) in specified mobile application. Half of participants performed first four tasks on mobile device and second four tasks on desktop PC thanks to use of emulation. Second half of participants performed first four tasks on PC and second four tasks on mobile device. After conducting of this study we have evaluated multiple metrics such as: *task success*, *task duration* or *participant effectiveness*.

Figure 1 shows evaluation of task success rate for each performed task accordingly to device on which task was performed. As we can see task success rates are very similar for both devices. However, in many tasks success rate on emulator is higher (difference is one participant). Our explanation for this phenomena is, that interaction on emulator is quite different and therefore participants were more focused. This is also confirmed thanks to task duration times analysis as we can see on Figure 2.

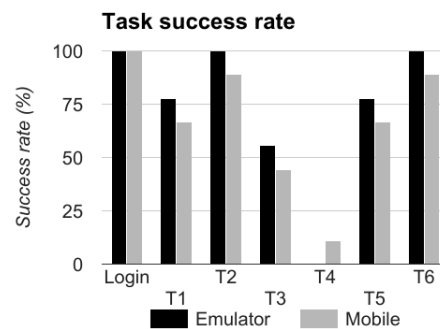


Figure 1. Study 1 - task success rates

In this study we have identified 19 unique usability problems from which 17 was identified on mobile device and 16 thanks to emulation. From 19 identified problems 14 problems were identified on both devices (74%). This confirmed our initial hypothesis. Moreover thanks to this study we showed that type of used device does not have any significant influence on task success, task duration or participant effectiveness.

Study 2. Main goal of this study was to verify correctness of list of potential usability problems that can be identified thanks to emulation. Correctness of problems that can not be identified is not necessary to validate because in most cases it is evident. This study was conducted with 13 participants. Participants performed two sets of specified tasks in two mobile applications. One on mobile device and one desktop PC with emulator. During this study 20 unique usability problems were identified. Twenty problems were

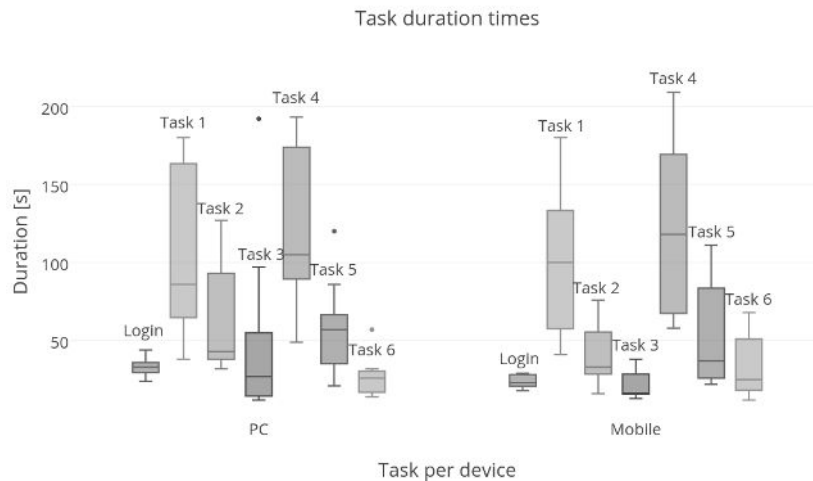


Figure 2. Study 1 - task execution times

identified on mobile device and fifteen on emulator.

In this study we have focused mainly on identification of fourteen usability problems from our lists of potential problems that we have manually injected into tested applications. One problem was from list of problems that can not be identified with use of emulator, and this was confirmed because four participants identified this problem on mobile device, however zero with use of emulator. Remaining thirteen problems were from list of problems that can be identified thanks to emulation. Twelve of this were confirmed because they occurred on mobile device and also with use of emulator. One problem did not occur at any device.

Conclusions

Our aim was to improve effectiveness of quantitative usability testing of mobile applications with the use of eye-tracking. In this paper we proposed a method for more effective quantitative usability testing of mobile applications with use of eye-tracking. We have also defined two lists of potential usability problems with mobile applications, where one list contains problems that can be identified with use of emulation and second contains problems that can not be identified with use of emulation.

We validated our method thanks to conducting two usability studies. First study proved our main hypothesis that some usability problems can be identified even with use of emulation. Second study verified our defined list of potential usability problems.

As future work we plan to find best projection size, which will help us to obtain eye-tracking data of best quality and also will not directly affect participant's task execution.

Acknowledgement: This contribution was created with kind support of ČSOB Foundation and is partial result of the project University Science Park of STU Bratislava, ITMS 26240220084, co-funded by the ERDF.

References

- [1] Betiol, A.H., de Abreu Cybis, W.: Chap. Usability Testing of Mobile Devices: A Comparison of Three Approaches. In: *Human-Computer Interaction - INTERACT 2005: IFIP TC13 Int. Conf., Rome, Italy, Proc.*. Springer Berlin, 2005, pp. 470–481.
- [2] Cuadrat Seix, C., Veloso, M.S., Soler, J.J.R.: Towards the Validation of a Method for Quantitative Mobile Usability Testing Based on Desktop Eyetracking. In: *Proc. of the 13th Int. Conf. on Interacción Persona-Ordenador. INTERACCION '12*, New York, ACM, 2012, pp. 49:1–49:8.
- [3] Duh, H.B.L., Tan, G.C.B., Chen, V.H.h.: Usability Evaluation for Mobile Device: A Comparison of Laboratory and Field Tests. In: *Proc. of the 8th Conference on Human-computer Interaction with Mobile Devices and Services. MobileHCI '06*, New York, NY, USA, ACM, 2006, pp. 181–186.
- [4] Gong, J., Tarasewich, P.: Guidelines for handheld mobile device interface design. In: *In Proc. of the 2004 DSI Annual Meeting*, 2004.
- [5] Kaikkonen, A., Kekäläinen, A., Cankar, M., Kallio, T., Kankainen, A.: Usability Testing of Mobile Applications: A Comparison Between Laboratory and Field Testing. *J. Usability Studies*, 2005, vol. 1, no. 1, pp. 4–16.
- [6] Móro, R., Daráz, J., Bielíková, M.: Visualization of Gaze Tracking Data for UX Testing on the Web. In: *Proceedings of DataWiz 2014: Data Visualisation Workshop*, 2014.