

Комплекснае паміж-групавое usability-тэставанне для вольнага ПЗ. **Касцюк Д.А., Маркіна А.А.**

Брэсцкі дзяржаўны тэхнічны ўніверсітэт, dmitriykostiuk@gmail.com

A combined approach to assessing the effectiveness of user interaction with software is presented, with a combination of the psychological analysis methods and the biometric measurements. Approbation results for the comparison of free and commercial table processors are presented.

Для вымярэння эфектыўнасці работы карыстальніка ёсць два прынцыпова розных падыходу. У першым падыходзе выкарыстоўваюцца экспертныя ацэнкі і самапаведамленыя параметры; ён займае шмат часу з-за апытанняў, хронаметражу і відэапракатавання, і больш залежыць ад апрацоўцы вынікаў. Таксама шэраг параметраў, такіх як фізічная нагрузка, можа быць ацэнены толькі ўскосна. Другі падыход прадугледжвае экспрэс-ацэнку стану карыстальніка з дапамогай прыбораў. Біямэтрычныя вымярэння паддаюцца аўтаматычнай апрацоўцы, аднак дазваляюць ацэньваць эмацыйны стан і кагнітыўныя працэсы толькі па ўскосных прыкметах.

Ніжэй зроблена спроба аб'яднання падыходаў у рамках комплекснага usability-тэставання. На думку аўтараў, падобная метадыка асабліва спатрэбілася б для GNU/Linux, для якой не часта робяць поўнамаштабныя usability-даследавання.

Матэрыялы для тэставання карыстальнікаў (у асаблівасці ўніверсальныя) і праграмае забеспячэнне, якое было распрацавана і выкарыстоўвалася аўтарамі ў ходзе работы, распаўсюджваюцца пад вольнай ліцэнзіяй і даступныя ў рэпазітары праекта UXDump (<https://bitbucket.org/AsyaAliset/uxdump>).

Метрыкі

Спіс метрык уключае індыкатары выніковасці (час выканання задання, памылкі, працэнт выкананых заданняў), нагрузку (пульс, канцэнтрацыю ўвагі, фіксацыі погляду) і ўражанні рэспандэнтаў (узровень чаканняў і ўзровень задавальнення карыстальніка). Час і паспяховасць выканання задання выкарыстоўваюцца толькі ў параўнанні. Паспяховасць выканання кадуецца ў бінарным кодзе (выканаў / не выканаў).

Праблемы, з якімі сутыкнуліся карыстальнікі, рэгіструюцца як мэнэджэрам эксперыменту, так і з дапамогай відэа-захопу экрана. Пасля тэставання праводзіцца рэтраспектыва: з карыстальнікам абмяркоўваюцца заданні, якія выклікалі праблемы, прайграваецца запіс, аналізуецца рэакцыя і паводзіны ў ходзе тэсту (каб класіфікаваць праблемы, выявіць больш значныя для карыстальніка, разлічыць, колькі карыстальнікаў з імі сутыкнуліся).

Узровень чакання адлюстроўвае меркаванне карыстальніка аб прадукце і камфортнасць працы, якую ён чакае, а ўзровень задавальнення – ацэнку зручнасці выкарыстання сістэмы пасля праходжання тэсту. Гэтыя метрыкі

можна атрымаць з дапамогай стандартных міжнародных апытальнікаў System Usability Scale (SUS) і Post-Study System Usability Questionnaire (PSSUQ) [1, 2]. Акрамя гэтага, карыстальнікам прапануецца са спісу выбраць эпітэты, якія могуць апісаць іх ўражанні ад праграмнага прадукту.

Нагрузка можа ацэньвацца ў адпаведнасці з апрабаванай раней экспрэс-методыкай [3], па якой рэгіструюцца шпаркасць выканання заданняў, фізічная і разумовая нагрузка, кірунак погляду карыстальніка [4]. Для ацэнкі канцэнтрацыі ўвагі намі выкарыстоўвалася метрыка «Attention» бытавога энцефалаграфа Neurosky Mindwave, фізічная нагрузка ацэньвалася па пульсе, з дапамогай фітнес-трэкера, а для рэгістрацыі кірунку погляду ўжываліся айтрекеры фірмы Tobii. Збор і першасная апрацоўка дадзеных выконваліся ў паралельным рэжыме праграмнай сістэмай, распрацаванай у рамках праекта UXDump.

Методыка правядзення

Тэставанне было падзелена на наступныя часткі:

- Складанне сцэнарыя выкарыстання, тыпавых заданняў для выкарыстання праграмнага прадукта з нарастаючай складанасцю, каб адсачыць тэндэнцыю навучальнасці. Заданні складаюцца зыходзячы з вопыту рэспандэнтаў.
- Набор удзельнікаў адной узроставай групы, якія маюць падобны досвед працы з прадуктамі-аналагамі. Агаворваецца час правядзення кожнай стадыі эксперыменту.
- Інструктаж (прывітанне, апісанне мерапрыемстваў, мэтай даследавання, метрык і тэстаў, падпісанне пагаднення на прадастаўленне і апрацоўку персанальных дадзеных).
- Уступнае інтэрв'ю (запаўненне анкеты ўдзельніка, праверка ўзроўню валодання прадуктамі-аналагамі, правядзенне псіхалагічных тэстаў – у нашым выпадку тэсту Айзека на вызначэнне тэмпераменту і тэсту Равена на ўзровень інтэлекту).
- Дэманстрацыйны паказ работы з прадуктам.
- Чаканні ад працы з прадуктам (запаўняюцца апытальнікі SUS, абмяркоўваюцца пытанні выкарыстання і функцыянал).
- Налада сістэмы (у залежнасці ад наяўнага абсталявання, падключэнне і каліброўка біяметрычных датчыкаў, айтрекера, а таксама налада відэа-пратокалявання і захопу відэа з экрана).
- Праца з прадуктам (выкананне тэставых заданняў з вядзеннем пратаколу мэнэджэрам, дзе фіксуюцца рэакцыі карыстальніка).
- Збор выніковых уражанняў (запаўняюцца PSSUQ, праводзіцца рэтрэспектыва).

У ходзе эксперыменту менеджэры, рэгіструючы час і рэакцыю карыстальнікаў, не ўмешваюцца ў ход выканання заданняў. У нашым выпадку менеджэры падбіраліся з кола рэспандэнтаў, былі з імі знаёмыя, што магло знізіць стрэс. Адказы на пытанні «Як гэта зрабіць?» Павінны быць

расплывистыя (гэта матывуе рэспандэнта разбірацца і ня ссоўвае фокус). Таксама рэспандэнты ведаюць пра відэа-назіранне, але ўвага ў ходзе тэсту на гэтым не акцэнтуюцца.

Апрабавыя на прыкладзе таблічных працэсараў

Апрабавыя методыкі выконвалася на задачы параўнання эрганомікі таблічных працэсараў, якія ўваходзяць у склад сучасных офісных пакетаў. У якасці паддоследных выступалі студэнты 18-19 гадоў, якія атрымліваюць тэхнічную адукацыю і маюць уяўленне аб таблічным працэсары, але не выкарыстоўваюць яго штодня.

Для тэставання мы абралі Microsoft Excel 2016 і LibreOffice Calc 5.4.4. У тэставанні ўдзельнічалі 14 рэспандэнтаў.

Выбар быў абгрунтаваны наступнымі меркаваннямі:

- сучасны таблічны працэсар з'яўляецца магутным дадаткам з развітым функцыяналам і складанымі сродкамі кіравання;
- абмежаванае валоданьне ім паддоследных дазваляе ацаніць, як асаблівасці інтэрфейсу ўплываюць на засваенне функцыяналу.

Заўважым, што ribbon-інтэрфейсы (Microsoft Fluent Interface, MFI) маюць пэўныя перавагі для складаных праграм [3]. Але параўнанне класічных інтэрфейсаў і MFI часцей праводзіцца на прыкладзе тэкставых працэсараў, а таму цікава іх параўнанне на складаных задачах іншага тыпу. Акрамя таго, цікава ацаніць і рэзультатыўнасць перапрацоўкі, якую зведаў Libreoffice Calc 5.

У ходзе ўступнага інтэрв'ю тры рэспандэнта адсяліліся па тэсту Айзенка з-за нізкай эмацыянальнай ўстойлівасці або неперадолення парога праўдзівасці тэсту, а адзін – па тэсту Равена, таму што яго ўзровень інтэлекту моцна адрозніваўся ад сярэдніх паказчыкаў групы.

Праца ў Calc праходзіла больш паспяхова і хутчэй, пры меншым узроўні стрэсу. Аналіз паказаў адсутнасць змен ўжывання негатыўных прыметнікаў пры апісанні інтэрфейсу Excel да і пасля эксперыменту (8 выпадкаў), у той час як для Calc такіх выпадкаў было 8 да эксперыменту і 7 пасля. Для абодвух пакетаў была тэндэнцыя, калі рэспандэнты часцей лічылі пакет больш карысным і эфектыўным пасля тэставання, чым раней. Меркаванне аб прастаце выкарыстання Excel ў працэсе працы знізілася на 67%, а меркаванне аб прастаце выкарыстання Calc - на 25%. Аналагічна можна сказаць аб ацэнках эканоміі часу пры выкарыстанні пакета (Excel - зніжэнне на 60%, Calc - зніжэнне на 50%). Пры гэтым ёсць катэгорыя карыстальнікаў, якія цалкам справіліся з тэстам толькі ў Calc.

Асноўнымі праблемамі пры выкананні заданняў у Excel апынуліся частая адсутнасць карэктных падказак, складаны і забытаны пошук патрэбнага функцыяналу, анты-інтуітыўнасць налад карыстальніка, нізкая прыкметнасць элементаў. Пры працы з Calc назіраліся праблемы прыкметнасці некаторых элементаў кіравання і пошук імёнаў функцый па агульным даведніку.

100

10

1

0.1

Excel

Calc

Tmin, хв Tmax, хв <T>, хв tmin, хв tmax, хв <T>, хв

Малюнок 1 — Максимальнае, мінімальнае і сярэдняе значэнні часу выканання ўсяго тэста Т і часу выканання аднаго задання t, хв

200

100

0

Excel

Calc

Pmax

Pmin

M, %

Nmin, %

<N>, %

Малюнок 2 — Максимальны, мінімальны пульс Р, працэнт рэспандэнтаў, якія выканалі ўсе задання М, найменшы і сярэдні працэнт выкананых заданняў у тэсце N



Малюнок 3 — Цеплавая карты фіксацый погляду карыстальніка ў Calc (злева) і ў Excel (справа)

Заўважым таксама, што некаторыя карыстальнікі пасля працы з Excel знайшлі, што ён «асабісты», «ўладны», «некантралюемы», што ён «апякуецца» і «засмучае», а Calc – «высакаякасны», «нетрадыцыйны», «насычаны», «захапляльны», «каштоўны», і яшчэ што ён «выклікае збянтэжанасць» трохі «расчароўвае».

Спіс літаратуры

1. Sauro J. Measuring usability with the system usability scale (SUS). <https://measuringu.com/sus/> Posted on February 2, 2011.
2. Post-Study System Usability Questionnaire (PSSUQ). // UX Glossary

<http://www.conetrees.com/2010/12/ux-glossary/post-study-system-usability-questionnaire-pssuq> Posted on December 10, 2010.

3. Костюк Д.А., Латий О.О., Маркина А.А. Подход к биометрической оценке эргономики графического интерфейса пользователя // Вестник Брестского государственного технического университета. Физика, математика, информатика. -- 2016. -- № 5(101). -- С. 46--49.
4. Дубицкий А., Костюк Д., Маркина А., Фомин С. Применение айтрекеров для юзабилити-исследований в GNU/Linux // Четырнадцатая конференция разработчиков свободных программ: Тезисы докладов. – Калуга, 22--24 сентября 2017 г. М.: Базальт СПО, 2017. – С. 36–41.