

Studiu de caz în evaluarea formativă a utilizabilității unui sit web din administrația publică locală

Dragoș Daniel Iordache, Ruxandra Dora Marinescu, Maria Gheorghe-Moisii, Costin Pribeanu

Institutul Național de Cercetare-Dezvoltare în Informatică – ICI București

Bd. Mareșal Averescu nr.8-10, București

{iordache, doruma, maria, pribeanu}@ici.ro

REZUMAT

Utilizabilitatea aplicațiilor informatice din administrația publică este o cerință prioritară pentru implementarea cu succes a societății informaționale. Din păcate, pe plan național nu există încă o practică curentă de evaluare a produselor informatice înainte de a fi instalate la utilizator și nici evaluatori cu experiență în utilizabilitate. Scopul acestui articol este de a prezenta o metodologie și un studiu de caz în evaluarea formativă a utilizabilității siturilor web din administrația publică. Procedura de evaluare se bazează pe două metode: evaluare euristică și testare cu utilizatori.

Cuvinte cheie

Utilizabilitate, evaluare formativă, evaluare euristică.

Clasificare ACM

D.2.2: Design tools and techniques. H5.2 User interfaces.

INTRODUCERE

În momentul de față, calitatea produselor și serviciilor informatice este o cerință prioritară pentru implementarea cu succes a societății informaționale. Acest deziderat necesită o abordare centrată pe utilizator a procesului de proiectare, care să conducă la o îmbunătățire substanțială a utilizabilității aplicațiilor informatice.

Pe plan național, utilizabilitatea este un subiect de interes în cadrul unor proiecte de cercetare [1, 8, 22]. Nu există încă o practică curentă de evaluare a utilizabilității produselor informatice înainte de a fi instalate la utilizator și nici evaluatori cu experiență în utilizabilitate.

În funcție de momentul și scopul evaluării, evaluarea utilizabilității poate fi formativă sau sumativă. Evaluarea formativă a utilizabilității este efectuată iterativ, pe parcursul ciclului de dezvoltare a unei aplicații informatice, cu scopul identificării și remedierii problemelor de utilizabilitate cât mai devreme [24].

Scopul acestui articol este de a prezenta o metodologie și un studiu de caz în evaluarea formativă a utilizabilității siturilor web din administrația publică. Procedura de evaluare se bazează pe două metode: evaluare euristică și testare cu utilizatori.

Metodologia a fost elaborată în cadrul unui proiect din Planul Sectorial de cercetare a MCSI și a fost apoi utilizată, cu rol experimental, în patru studii de caz de evaluare euristică și două studii de caz de evaluare prin testare cu utilizatori. În cadrul acestei lucrări se prezintă și

se analizează comparativ rezultatele obținute în evaluarea aceluiași sit web din administrația publică locală, cu fiecare dintre cele două metode.

Restul acestui articol este organizat după cum urmează. În secțiunea următoare se prezintă câteva preocupări recente și tendințe actuale în evaluarea utilizabilității. Accentul este pus pe problematica asociată cu evaluarea formativă a utilizabilității: scopul evaluării, metode utilizate, măsuri colectate și mod de descriere, consolidarea problemelor de utilizabilitate, gradul de încredere și validitate. În cea de a treia secțiune se prezintă succint metodologia (elemente caracteristice), studiul de caz și se analizează rezultatele și limitele studiului. Lucrarea se încheie cu concluzii și direcții de continuare a cercetărilor.

PREOCUPĂRI ȘI TENDINȚE ACTUALE

Evaluarea formativă a utilizabilității

Sistemele interactive sunt proiectate în mod iterativ. Este dificil, practic imposibil, parcurgerea etapelor din ciclul de dezvoltare o singură dată. În acest sens, metodele de evaluare formativă a utilizabilității sunt utilizate în cadrul unei bucle de proiectare - evaluare.

O problemă de utilizabilitate a fost definită ca orice aspect al interfeței cu utilizatorul despre care se presupune (se observă) că ar crea utilizatorului dificultăți sau nemulțumiri, raportat la un indicator important al utilizabilității: ușurință în învățare, ușurință în operare, rată de erori, satisfacție subiectivă [18].

După gradul de severitate, adică impactul pe care îl au dificultățile sau nemulțumirile utilizatorului, problemele de utilizabilitate se pot clasifica în trei categorii:

- Severe: probleme care nu permit ca utilizatorul să îndeplinească sarcina sau care rezultă în pierderi importante de date sau timp.
- Moderate: probleme care au un impact semnificativ asupra îndeplinirii sarcinii dar pentru care utilizatorul poate găsi o soluție.
- Minore: probleme care sunt iritante pentru utilizator dar care nu au un impact semnificativ asupra îndeplinirii sarcinii.

De regulă, în evaluarea formativă a utilizabilității se colectează măsuri cantitative (numărul de probleme de utilizabilitate pe niveluri de severitate: major, moderat și minor) și măsuri calitative (descrierea detaliată a fiecărei probleme și a contextului în care apare).

Într-o lucrare recentă, Hvannberg et al. [12], au utilizat un format structurat de descriere, care a fost adaptat după [4]

și care conține : identificatorul problemei, descriere, dificultățile întâmpinate de utilizator, contextul specific (localizarea problemei în interfață), cauzele posibile (ce anume este greșit în proiect), euristicile utilizate și gradul de severitate (impactul, pe trei niveluri : major, moderat și minor).

Un aspect important al oricărei abordări în evaluarea formativă este utilitatea pe care o au rezultatele evaluării asupra procesului de dezvoltare. În esență, întrebarea este în ce măsură rezultatele sunt valorificate ulterior în procesul de proiectare și care este gradul de recuperare al costurilor activității de evaluare. Așa cum arată Molich et al. [16], raportul de evaluare trebuie să fie atât util cât și utilizabil. Din acest motiv este necesară includerea, în descrierea problemelor de utilizabilitate, a recomandărilor de remediere.

Metode de evaluare formativă

Problemele de utilizabilitate pot fi identificate atât prin inspecție de utilizabilitate (evaluare de tip expert) cât și prin testare cu utilizatori (evaluare centrată pe utilizator). Între metodele de inspecție a utilizabilității, cele mai cunoscute sunt: evaluarea euristică, inspecția cognitivă și evaluarea bazată pe recomandări.

Evaluarea euristică este o metodă efectuată de un număr redus de evaluatori, care examinează o interfață cu utilizatorul, judecă respectarea unui set de principii de utilizabilitate (euristici) și elaborează o listă de probleme de utilizabilitate clasificate pe categorii de severitate, corespunzător impactului estimat asupra performanțelor utilizatorului sau acceptanței [15,18].

Setul de euristici utilizat este la alegerea analistului iar rezultatele evaluării sunt determinate în primul rând de abilitățile acestuia de a le utiliza. În [12] au fost utilizate alternativ două seturi de principii: euristicile lui Jacob Nielsen și cele derivate din ingineria cognitivă de către Gerhard-Powels [7] și au concluzionat că al doilea set necesită un plus de experiență. Bastien și Scapin [3] au elaborat un set de criterii ergonomice, care poate fi folosit atât pentru proiectare cât și pentru evaluare. Ulterior, setul de criterii ergonomice a fost adaptat pentru situri web de comerț electronic și pentru aplicații bazate pe realitate îmbogățită / realitate virtuală [14, 2].

În studiul făcut de Nielsen [19] au fost comparate 7 seturi de euristici din perspectiva explicării problemelor de utilizabilitate. O analiză factorială arată că 25 de factori (cu o contribuție de peste 1% fiecare) explică 62% din varianță. Primii primii 7 factori (care grupează peste 65 de euristici) explică cca. 30% din varianță. De asemenea, rezultatele studiului arată că în primele 10 euristici care explică toate problemele de utilizabilitate, trei aparțin autorului, celelalte fiind din alte seturi. Procentajul de probleme explicate variază între 23% și 2%.

Testarea cu utilizatori este definită ca o metodă empirică de evaluare a utilizabilității implicând participanți având caracteristici apropiate de cele ale utilizatorilor reali ai produsului care va fi evaluat. Prin testarea cu utilizatori se înregistrează comportamentul utilizatorului cu ajutorul unor tehnici specifice cum sunt observarea sau protocoalele verbale. O tehnică frecvent utilizată în

evaluarea formativă a utilizabilității este protocolul de « gândire cu voce tare » (TAP – *Think Aloud Protocol*).

Avantajele principale ale protocolului gândirii cu voce tare constau într-o mai bună înțelegere a modelului mental al utilizatorului și a interacțiunii cu produsul: ce fac utilizatorii, de ce fac ceea ce fac și cum fac [20]. TAP permite utilizatorilor să exprime, într-o secvență de pași cum utilizează un produs pentru a îndeplini sarcinile solicitate.

Studiile efectuate de Usability Professional Association – UPA [25,26] arată că în anul 2009, cele mai frecvent utilizate metode în practica proiectării centrate pe utilizator au fost studiile utilizator (anchete, interviuri) 75%, evaluarea euristică (inspecții de utilizabilitate) 74% și testarea cu utilizatori 68%. O comparație față de situația existentă în 2007, arată o scădere a ponderii pe care o au utilizarea evaluării euristice (-3%) și testării cu utilizatori (9%) și o creștere a interesului față de studiile utilizator (+1%).

Consolidarea problemelor de utilizabilitate

Consolidarea problemelor de utilizabilitate este o activitate importantă a evaluării formative în care sunt implicați mai mulți evaluatori și are ca scop agreeerea unei liste unice de probleme de utilizabilitate. Consolidarea presupune eliminarea duplicatelor, agreeerea asupra gradului de severitate și a descrierii fiecărei probleme în parte. În acest scop se pot folosi diferite metode [11].

Procesul de consolidare a problemelor de utilizabilitate presupune două tipuri de operații [13]:

- *filtrarea*: eliminarea duplicatelor din listele de probleme identificate de către un singur evaluator
- *integrarea*: combinarea problemelor de utilizabilitate dintre diferite liste.

În cazul testării cu utilizatori, există două etape de consolidare: consolidare individuală și consolidare colaborativă. În fiecare etapă se efectuează, după caz atât operații de filtrare cât și de integrare.

În consolidarea individuală se analizează, de către fiecare evaluator, problemele de utilizabilitate identificate de mai mulți utilizatori. În a doua etapă, doi sau mai mulți evaluatori, care au utilizat aceleași rapoarte de testare, cad de acord asupra unei liste unice de probleme de utilizabilitate pentru fiecare sarcină.

Aspecte privind gradul de încredere

În ultimi ani au fost efectuate numeroase cercetări privind validitatea și eficacitatea metodelor de evaluare a utilizabilității. Încrederea se referă la măsura în care evaluările independente produc același rezultat. Validitatea se referă la măsura în care problemele detectate în cadrul evaluării sunt aceleași cu cele care apar în practica utilizării sistemului.

Pentru măsurarea gradului de încredere se recomandă doi indicatori: rata de detecție și coeficientul de agreeere între oricare doi evaluatori (any-two-agreement) [12].

În studiul lui Connell și Hammond [5], coeficientul mediu de agreeere a variat între 5% și 9%.

Studiul făcut de Jakob Nielsen [17] arată că gradul de încredere variază mult în funcție de expertiză și de

numărul de evaluatori. În medie, un evaluator începător a identificat 22% din problemele de utilizabilitate existente într-o aplicație, un evaluator experimentat 41% (aproape dublu) iar un evaluator cu dublă expertiză (utilizabilitate și domeniul aplicației) 60% din problemele de utilizabilitate.

În același studiu se arată că în medie, un singur evaluator începător identifică 20% din problemele de utilizabilitate, trei evaluatori cca. 40% iar 5 evaluatori începători 50% din problemele de utilizabilitate. Pentru a depăși pragul de 75% este nevoie de o echipă de 15 evaluatori (foarte mult și practic imposibil de asigurat). În mod similar, se estimează că este suficient un număr de 3 evaluatori experimentați pentru a identifica 75% și 5 evaluatori experimentați pentru a identifica 85% din problemele de utilizabilitate dintr-o interfață.

Aspecte privind validitatea

Evaluarea euristică produce un set de probleme de utilizabilitate care sunt considerate *anticipate* în timp ce testarea cu utilizatori identifică probleme de utilizabilitate *reale*. Diferența între cele două seturi o reprezintă problemele *fals pozitive* (alarme false, identificate de evaluarea euristică dar neconfirmate de testarea cu utilizatori) și *fals negative* (problemele de utilizabilitate reale care nu au fost anticipate de evaluarea euristică).

Pentru măsurarea validității, eficacității și eficienței evaluării euristice au fost propuși mai mulți indicatori [9]:

- validitate: probleme anticipate-confirmate / probleme anticipate.
- completitudine (thoroughness): probleme anticipate-confirmate / probleme reale
- eficacitate generală (overall effectiveness): validitate * completitudine.

Calcularea acestor indicatori presupune evaluarea cu ambele metode, pe baza aceluiași scenariu de utilizare, fapt care mărește costurile evaluării.

METODOLOGIE ȘI STUDIU DE CAZ

Componentele metodologiei și activități de evaluare

Metodologia elaborată în acest proiect respectă prevederile standardului ISO/IEC 14598-5:1998 privind structura unui proces de evaluare :

- Stabilirea cerințelor de evaluare: stabilirea scopului evaluării, a produselor țintă și specificarea modelului utilizabilității;
- Specificarea evaluării: specificarea produsului țintă, a contextului și scenariilor de utilizare, alegerea metodelor și tehnicilor, selectarea metricilor, stabilirea nivelurilor de apreciere pentru acestea și stabilirea cerințelor de măsurare.
- Proiectarea evaluării: elaborarea planului de evaluare.
- Executarea evaluării: măsurarea caracteristicilor, colectarea datelor și analiza preliminară a rezultatelor.
- Concluzia evaluării: finalizarea analizei și raportarea rezultatelor.

În general, o metodă de evaluare este o procedură care descrie acțiunile executate de către evaluator pentru a obține rezultatul măsurării sau verificării specificate a unui

produs specificat sau unor componente specificate ale unui produs.

Metodologia elaborată în acest proiect integrează două metode de evaluare formativă: evaluarea euristică și testarea cu utilizatori.

Aplicarea a două tipuri de metode de evaluare formativă este benefică din mai multe motive:

- Complementaritatea metodelor.
- Creșterea gradului de încredere și a validității rezultatelor.
- Creșterea eficienței activității de remediere a problemelor de utilizabilitate.
- Îmbunătățirea expertizei de utilizare a fiecărei metode în parte.
- Pregătirea evaluării sumative.

Metodele sunt complementare întrucât se bazează pe testarea de către două categorii de utilizatori:

- Experți în utilizabilitate
- Utilizatori reprezentativi.

Pe de altă parte, în cazul testării cu utilizatori pot fi colectate și alte tipuri de măsuri: măsuri ale eficacității (rata de îndeplinire a sarcinii, numărul de erori), ale eficienței (timpul de execuție) și satisfacției (prin aplicarea de chestionare, cu grad variabil de complexitate). În acest fel, se obțin date cantitative / calitative suplimentare față de cele obținute de evaluarea euristică.

Repetarea evaluării cu o altă metodă, dar utilizând același set de sarcini relevante, astfel încât rezultatele să fie comparabile, mărește încrederea în rezultatele obținute.

Fiecare dintre cele două metode necesită competențe specifice. Compararea rezultatelor obținute permite analiza procedurii, identificarea greșelilor de evaluare și îmbunătățirea practicii de evaluare. Validarea rezultatelor evaluării euristice prin testare cu utilizatori permite și o comparație corectă a rezultatelor obținute de către fiecare expert în parte, fapt care îmbunătățește expertiza individuală.

Metodologia integrată de evaluare a utilizabilității cuprinde următoarele activități:

- Specificarea evaluării: analiza cerințelor de evaluare, specificarea sarcinilor de testare, planificarea și pregătirea evaluării, specificarea modelului utilizabilității.
- Evaluarea euristică. Se utilizează setul extins de criterii ergonomice care este aplicat prin verificarea îndeplinirii recomandărilor de utilizabilitate. În acest sens, evaluarea are un caracter sistematic și structurat.
- Evaluarea prin testare cu utilizatori. Comportamentul utilizatorilor este înregistrat audio (limitare impusă de resursele tehnice achiziționate prin proiect) și prin observare. Se utilizează protocolul „gândire cu voce tare”. Înregistrările sunt apoi analizate de evaluatori care extrag și documentează problemele de utilizabilitate.
- Analiza comparativă a rezultatelor obținute prin cele două metode, calcularea indicatorilor de validitate, completitudine și eficiență a evaluării euristice.

Având în vedere practica curentă în evaluarea utilizabilității, analiza cerințelor de evaluare și specificarea evaluării au fost integrate într-o singură activitate, care precede aplicarea celor două metode. În această etapă, planificarea evaluării se referă la ordinea temporală a aplicării celor două metode.

Rezultatele obținute se analizează din două perspective:

- A dezvoltatorului, care trebuie să efectueze remediarea deficiențelor constatate la evaluare.
- A evaluatorului, care trebuie să analizeze eficacitatea și eficiența activității de evaluare.

Analiza comparativă se completează cu o discuție asupra performanțelor de evaluare cu fiecare metodă, atât la nivel individual (rată individuală de detecție) cât și la nivel de echipă (rată medie de detecție și coeficient de agreare).

Evaluarea euristică a fost efectuată cu un set de 24 de euristici, structurat în 6 categorii (criterii ergonomice generale) și care este prezentat în Tabelul 1.

Tabelul 1. Euristici de evaluare

Ghidarea utilizatorului	
1	Vizibilitatea stării sistemului
2	Incitare (prompting)
3	Feedback imediat
4	Grupare / distincție prin format
5	Grupare / distincție prin amplasare
6	Lizibilitate
Efort de operare	
7	Concizie
8	Recunoaștere în loc de reamintire
9	Acțiuni minimale
10	Densitatea informației
Adaptabilitate și control	
11	Flexibilitatea și eficiența utilizării
12	Gradul de experiență
13	Acțiuni explicite
14	Controlul utilizatorului
Management erori	
15	Prevenirea erorilor
16	Calitatea mesajelor de eroare
17	Corectarea erorilor
Consistență și standarde	
18	Consistență
19	Conformitate cu standarde și reglementări
20	Semnificație coduri
Compatibilitate	
21	Compatibilitate cu utilizatorul
22	Compatibilitate cu sarcina de lucru
23	Asistență și documentație
24	Proiectare estetică / experiență plăcută

Setul de euristici a fost elaborat prin integrarea setului de criterii ergonomice al lui Bastien și Scapin [3] și euristicii lui Jakob Nielsen [17], între care există o suprapunere destul de mare. Având în vedere că cele două seturi originale de euristici au fost validate [19, 23], putem considera că gradul de acoperire al setului integrat de euristici este suficient de mare, atât în ceea ce privește puterea explicativă cât și în identificarea problemelor individuale de utilizabilitate.

Descrierea detaliată a problemelor de utilizabilitate urmează formatul propus în [12], cu mențiunea că după descrierea cauzelor sunt incluse și sugestiile de remediere.

Studiu de caz în evaluarea euristică

Obiectul evaluării a fost situl web al Primăriei orașului Pitești aflat la adresa web: <http://www.primariapitesti.ro>.

Sarcinile stabilite pentru evaluare sunt cele din Tabelul 2.

Tabelul 2. Sarcini de testare

1.	(T1) Disponibilitatea documentelor
2.	Consultare formulare/acte
3.	Accesare documente
4.	Completare formulare
5.	(T2) Înregistrare client nou pentru plata online
6.	Înregistrare client
7.	Accesare cont
8.	Solicitare ajutor

Evaluarea a fost realizată în perioada mai-noiembrie 2009 de către o echipă formată din 4 evaluatori.

Rezultatele obținute după identificarea individuală a problemelor de utilizabilitate sunt prezentate în Tabelul 3.

Tabelul 3 Probleme de utilizabilitate identificate pe evaluator

Sarcină	Total				Din care majore			
	Ev1	Ev2	Ev3	Ev4	Ev1	Ev2	Ev3	Ev4
T1	5	5	3	5	3	1	2	1
T2	2	4	1	1	2	4	1	1
Total	7	9	4	6	5	5	3	2

Așa cum se observă, primul evaluator a identificat în total 12 probleme de utilizabilitate (5 majore), al doilea evaluator a identificat în total 14 probleme (5 majore), al treilea evaluator a identificat în total 7 probleme (4 majore) iar al patrulea evaluator a identificat în total 8 probleme (2 majore).

Listele individuale de probleme de utilizabilitate au fost consolidate pe sarcina, pe baza criteriului „modificări similare”. În total, lista consolidată de probleme de utilizabilitate cuprinde 20 de probleme de utilizabilitate dintre care 5 probleme majore. Lista este prezentată în Tabelul 3.

Tabelul 3 Lista consolidată de probleme unice de utilizabilitate

Sarcina	Total	majore	moderate	minore
T1	15	3	7	5
T2	5	2	2	1
TOTAL	20	5	9	6

Din punctul de vedere al criteriilor ergonomice care nu au fost respectate, majoritatea problemelor de utilizabilitate anticipate de evaluarea euristică, respectiv 13 (65%), se datorează ghidării necorespunzătoare a utilizatorului.

Rata de detecție individuală, calculată după consolidarea colaborativă, a variat între 20% și 40% cu o medie de 31.25%. Este o rată medie destul de mică, întrucât pentru patru evaluatori, rata medie de detecție este de cel puțin 25%. Coeficientul de agreare mediu a fost 11.47%.

Studiu de caz în testarea cu utilizatori

Având în vedere caracterul experimental al studiului, a fost evaluat același sit web, prin testarea cu utilizatori a acelorași sarcini.

La testare au participat 5 utilizatori, 3 de sex masculin și 2 de sex feminin, cu vârste cuprinse între 29-58 de ani (M=37.4). Toți utilizatorii au experiență anterioară în

lucrul cu calculatorul. Totdată, nici un utilizator nu a fost familiar cu situările evaluate.

Aplicația a fost accesată pe un sistem desktop. Înregistrarea protocoalelor verbale s-a realizat cu reportofon digital Philips LFH 7780.

Evaluarea a fost realizată în ianuarie 2010 de către o echipă formată din 3 evaluatori. Sarcinile au fost prezentate în scris și oral utilizatorilor de către facilitatori pe parcursul sesiunilor de testare. Cei trei evaluatori au fost alternativ facilitatori pe parcursul sesiunilor de testare.

În prima etapă, primul evaluator a identificat în total 20 probleme de utilizabilitate dintre care 5 majore, al doilea al 22 probleme dintre care 5 majore iar al treilea 20 probleme dintre care 6 majore.

După consolidarea pe sarcină, rezultatele obținute de fiecare evaluator sunt prezentate în Tabelul 5.

Tabelul 5. Probleme după consolidarea individuală

Sarcină	Total			Din care majore		
	Ev1	Ev 2	Ev 3	Ev 1	Ev 2	Ev 3
T1	6	4	5	1	2	1
T2	4	5	3	1	2	2

După consolidarea colaborativă au rămas în total 11 probleme de utilizabilitate, dintre care 2 majore, 8 moderate și una minoră. Lista este prezentată în Tabelul 6.

Tabelul 6 Lista consolidată de probleme unice de utilizabilitate

Sarcina	Total	majore	moderate	minore
T1	6	1	4	1
T2	5	1	4	0
TOTAL	11	2	8	1

Două dintre problemele de utilizabilitate au fost identificate de 4 utilizatori, alte două de către 3 utilizatori iar restul de 7 probleme de utilizabilitate de către un singur utilizator. Un utilizator a identificat 6 probleme (54.54%), alt utilizator a identificat 5 probleme (45.45%), alți doi utilizatori au identificat câte 4 probleme (36.36%) și ultimul utilizator a identificat numai două probleme de utilizabilitate (18.18%).

O analiză a criteriilor ergonomice nerespectate arată că cele mai multe probleme de utilizabilitate se datorează efortului de operare (45.45%) și lipsei de ghidare (27.27%).

Rata de detecție individuală a fost ridicată, variind între 63.64% și 90.91%. Rata de detecție medie a fost de 72.73%. Coeficientul de agreere mediu a fost de 67.07%.

Atât rata medie de detecție cât și coeficientul de agreere au valori foarte ridicate, mai ales dacă le comparăm cu rezultatele evaluării euristice. O explicație este faptul că s-au testat numai două sarcini (destul de simple) iar numărul de utilizatori a fost mic (5 este numărul minim de utilizatori), fapt care a condus la identificarea unui număr redus de probleme de utilizabilitate.

Analiza comparativă a rezultatelor

La sarcina T1, din cele 15 probleme de utilizabilitate anticipate, au fost confirmate 2 de către utilizatori. Un număr de 4 probleme de utilizabilitate reale nu au fost identificate în evaluarea euristică.

La sarcina T2, din cele 5 probleme de utilizabilitate anticipate, au fost confirmate 2 de către utilizatori. Un număr de 3 probleme de utilizabilitate reale nu au fost identificate în evaluarea euristică. Indicatorii de validitate, completitudine și eficiență au următoarele valori: $V=20\%$, $C=36.36\%$ și $EG=7.27\%$.

Diferența între problemele de utilizabilitate anticipate și cele reale este destul de mare. Pe de o parte, numărul problemelor de utilizabilitate omise arată limitele evaluării euristice și utilitatea testării cu utilizatori. Pe de altă parte, evaluatorii au tendința de a anticipa multe probleme minore / locale în timp ce în testarea cu utilizatori se identifică probleme majore / globale [4]. Există o complementaritate între cele două metode de care se recomandă să se țină seama, mai ales atunci când numărul de utilizatori este mic.

Atât gradul de încredere cât și validitatea măsurate după formulele menționate anterior, trebuie privite cu prudență. Numărul problemelor anticipate, a celor confirmate și a celor reale depind atât de numărul și expertiza evaluatorilor cât și de numărul de utilizatori.

Limite ale studiului

Acest studiu reprezintă prima experimentare sistematică în România a unei metodologii de utilizabilitate. În mod inerent, există o serie de limitări legate de resursele disponibile, caracterul experimental (de laborator), noutatea activității și expertiza în evaluare.

O limită a studiului este numărul redus de utilizatori care au testat numai două sarcini și numărul mic de evaluatori, care au condus la identificarea unui număr mic de probleme de utilizabilitate.

O altă limită este lipsa de experiență a evaluatorilor (dintre care numai unul a mai efectuat evaluare euristică într-un proiect anterior). Pe de altă parte, toți trei au evaluat pentru prima dată pe baza analizei protocoalelor verbale. Așa cum s-a menționat anterior, cu numai trei evaluatori începători se pot detecta cca. 40% din problemele de utilizabilitate existente într-o interfață.

CONCLUZII

Concepția de integrare care stă la baza metodologiei utilizate are la bază utilizarea a două tipuri de metode de evaluare formativă și are multiple avantaje, între care cele mai importante sunt:

- Complementaritatea metodelor, respectiv combinarea unei metode bazate pe inspecție (evaluare de tip expert) cu testarea cu utilizatori (evaluare centrată pe utilizator).
- Creșterea gradului de încredere și a validității rezultatelor, prin confirmarea problemelor de utilizabilitate anticipate de către testarea cu utilizatori și prin utilizarea aceluiași măsuri cantitative și calitative în ambele metode.
- Creșterea eficienței activității de remediere a problemelor de utilizabilitate, prin utilizarea unei metode mai puțin costisitoare (evaluarea euristică) ca evaluare preliminară (în scopul eliminării erorilor de programare).

Atât evaluarea euristică cât și testarea cu utilizatori produc descrieri ale problemelor de utilizabilitate pe baza unui format unic. Din acest motiv, efortul de descriere este redus, întrucât o mare parte din probleme, respectiv cele anticipate și confirmate, sunt descrise în evaluarea euristică.

Analiza rezultatelor fiecărei metode are și scopul de a furniza un feedback util asupra performanțelor fiecărui evaluator, raportat la performanțele echipei de evaluare. Analiza comparativă a rezultatelor obținute de cele două metode permite calcularea indicatorilor de eficacitate și eficiență a evaluării euristice și oferă o confirmare privind validitatea problemelor de utilizabilitate identificate în evaluarea euristică.

Se intenționează efectuarea unui nou studiu de caz în aplicarea metodologiei integrate, cu un număr mai mare de utilizatori.

Confirmare

Această lucrare a fost finanțată din proiectul de cercetare PS MCSI 49/2008.

REFERINȚE

- Andrei, D.M., Mureșan, A.D. (2008) Atribute hedonice și pragmatice în determinarea experienței utilizatorului de telefonie mobilă. Buraga, S.C. & Juvină, I. (Ed.) *Interacțiune Om-Calculator 2008*. MatrixROM, 133-138.
- Bach, C. & Scapin, D.L., 2003. Adaptation of Ergonomic Criteria to Human-Virtual Environments Interactions. *Proceedings of Interact'03*. IOS Press, 880-883.
- Bastien, C. & Scapin, D.L. (1993) *Evaluating a User Interface with Ergonomic Criteria*. INRIA Report, Roquencourt.
- Cockton, G., Woolrych, A. (2001). Understanding inspection methods: lessons from an assessment of heuristic evaluation. Blandford, A., Vanderdonckt, J., Gray, P.D. (Eds.), *Proceedings of People and Computers XV*. Springer-Verlag, 171-182
- Connell, I.W., Hammond, N.V., 1999. Comparing usability evaluation principles with heuristics. In: Sasse, A., Johnson C. (Eds.), *Proceedings of the 7th IFIP international conference on Human-computer Interaction*, INTERACT'99, IOS Press, Edinburgh.
- Dumas, J., Redish, J. *A practical guide to usability testing*. Intellect, 1999.
- Gerhardt-Powals, J., 1996. Cognitive engineering principles for enhancing human-computer performance. *International Journal of Human-Computer Interaction* 8, 189-211.
- Guran, A.M., Cojocar, G.S. (2008) Abordări în evaluarea automată a utilizabilității. Studiu comparativ. *Revista Română de Interacțiune Om-Calculator*, 1(1), 73-84.
- Hartson, H.R., Andre, T.S., Williges, R.C., (2001). Criteria for evaluating usability evaluation methods. *International Journal of Human-Computer Interaction* 13, 373-410
- Hertzum, M., Jacobsen, N.E. (2001) „The evaluator effect: A chilling fact about usability evaluation methods”. *International Journal of Human-Computer Interaction* 13, 421-443.
- Hornbaek, K & Frokjaer, E. (2008). Comparison of techniques for matching of usability problem descriptions. *Interacting with Computers* 20, 505-514.
- Hvannberg, E.T. and Law, E.L.-C., Larusdotir, M.C. (2007) „Heuristic Evaluation: Comparing ways of finding and reporting usability problems, *Interacting with Computers* 19, 255-240.
- Law, E. Hvannberg, E.T. (2008): Consolidating usability problems with novice evaluators. *Proceedings of NordiCHI 2008*. ACM Press, 495-498.
- Leulier, C., Bastien, C., Scapin, D. (1998) *Commerce & Interactions*. INRIA Report.
- Molich, R., and Nielsen, J. (1990). Improving a human-computer dialogue, *Communications of the ACM* 33(3), 338-348.
- Molich, R., Jeffries, R., Dumas, J. (2007) Making usability recommendations usable. *Journal of Usability Studies*, 2(1), 162-179.
- Nielsen, J. (1992). Finding usability problems through heuristic evaluation. P. Bauersfeld, J. Bennett & G. Lynch (Eds.), *Proceedings of CHI'92*, ACM Press..
- Nielsen, J. (1993). *Usability Engineering*. Academic Press, New York.
- Nielsen, J. (1994). “Enhancing the explanatory power of usability heuristics”. *Proceedings of CHI'94*, ACM Press, 152-158.
- Nielsen, J. (1997). Usability Testing. *Handbook of Human Factors and Ergonomics*. G. Salvendy, John Wiley and Sons, 1543-1568.
- Nielsen, J. (2000) Why You Only Need to Test with 5 Users. *Alertbox*, March 19, 2000.
- Pribeanu, C., Iordache, D.D., Balog, A. (2008) Evaluarea utilizabilității unui scenariu de învățare a biologiei implementat pe o platformă de realitate îmbogățită. *Revista Română de Interacțiune Om-Calculator*, 1(1), 39-56.
- Scapin, D. & Bastien, C. (1997) Ergonomic criteria for evaluating the ergonomic quality of interactive systems. *Behaviour & Information Technology*, 16(4/5), 220-231.
- Theofanos, M. & Quesenbery, W. (2005) “Towards the Design of Effective Formative Test Reports”. *Journal of Usability Studies*, Issue 1, Vol.1, 27-45
- *** UPA 2007 Survey. Usability Professionals Association.
- *** UPA 2009 Survey. Usability Professionals Association.