

Testy użyteczności w praktyce

PAWEŁ GUZ

IMPAQ Sp. z o.o.

Wiśniowy Business Park, ul. 1-go Sierpnia 6A, 02-134 Warszawa

Wstęp

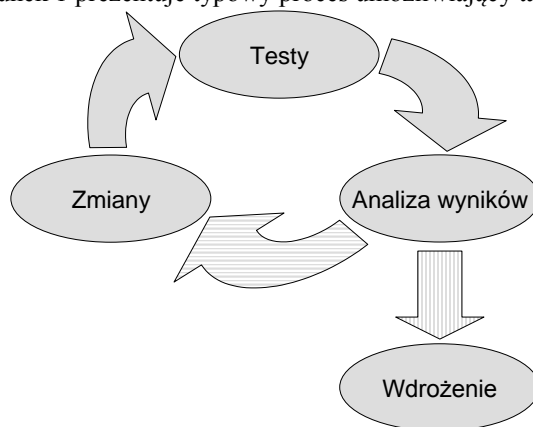
Grupa Kompetencyjna Software Usability Group (SUG) powstała w 2004 roku w celu wdrożenia w IMPAQ procesów zapewniających wysoką użyteczność tworzonego oprogramowania. Średnioterminowym celem było także oferowanie usług, obejmujących doradztwo w zakresie użyteczności oraz wykonywanie testów użyteczności.

Pierwszym zadaniem było stworzenie portalu w firmowym Intranecie, służącego rozpowszechnianiu wiedzy o użyteczności oprogramowania oraz zawierającego opis obowiązujących w firmie procesów, procedur, standardów i wskazówek dotyczących procesu User-Centered Development.

Po stworzeniu pierwszej wersji portalu, przeprowadzono testy użyteczności tego portalu. Wyniki testów prezentuje ten dokument.

Testy użyteczności – metodyka

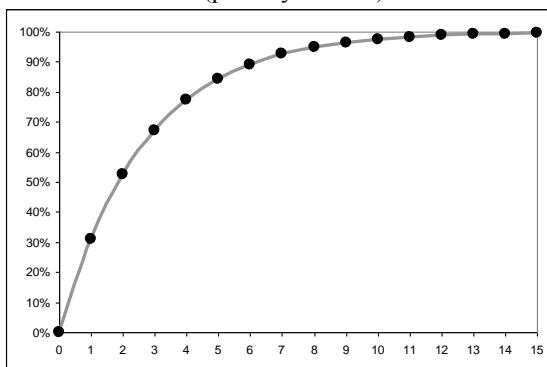
Istotnym elementem testów użyteczności jest zapewnienie możliwości formalnej oceny ich wyników. Rysunek 1 prezentuje typowy proces umożliwiający taką ocenę.



Rys. 1. Cykl zadań w procesie prowadzenia testów użyteczności

Testy użyteczności polegają na wykonaniu przez uczestników testu wcześniej zaplanowanych zadań. W trakcie wykonywania zadań mierzone są np. czas wykonania zadania, liczba wykonanych akcji, czy inne aspekty związane z produktywnością użytkowników systemu. Istotne jest także mierzenie subiektywnej oceny systemu przez uczestników testów, gdyż akceptacja oprogramowania przez końcowych użytkowników jest równie istotna jak miary produktywności.

Zgodnie z badaniami opisanymi w [1] optymalne efekty dają testy przeprowadzone przy pomocy około pięciorgo uczestników (patrz rysunek 2).



Rys. 2. Procent błędów znajdujących się przez kolejnych uczestników testów

W celu możliwości oceny użyteczności systemu, dla każdej miary zdefiniować należy minimalne wymagane wartości. Progi te umożliwiają podjęcie odpowiedniej decyzji na etapie analizy wyników testów. Jeżeli osiągnięte wartości miar przekraczają wymagane wartości, można zakończyć testy i przystąpić do wdrożenia systemu. W przeciwnym przypadku, wskazane jest wprowadzenie odpowiednich zmian i ponowne wykonanie testów użyteczności.

Planowanie testów portalu SUG

Głównym celem testów było znalezienie odpowiedzi na następujące pytania:

- Czy informacje zamieszczone w portalu są przejrzyste dla użytkowników?
- Czy znalezienie konkretnych informacji jest proste? Czy schemat nawigacyjny jest jasny?
- Jakie obszary tematyczne powinny być rozbudowane?

W tym celu zdefiniowano trzy zadania polegające na znalezieniu w portalu informacji na konkretne tematy. Każdy użytkownik zapytany został także o informacje, które oczekiwалby znaleźć na portalu. Czwartym zadaniem było znalezienie tych informacji.

Dla każdego zadania mierzono:

- ❑ Efektywność nawigacji (N) jako liczbę wykonanych akcji (W) w porównaniu do minimalnej liczby akcji koniecznej do wykonania zadania (M):

$$N = 100\% * M / W$$

- ❑ Czas wykonania zadania (T) szacowany przez prowadzącego testy w skali:
 - 0 – zadanie nie wykonane
 - 1 – zadanie wykonane wolno
 - 2 – zadanie wykonane w akceptowalnym czasie
 - 3 – zadanie wykonane szybko

Oprócz tego po zakończeniu wykonywania wszystkich zadań proszono uczestników testów o:

- ❑ ocenę nawigacji (w skali 0-3),
- ❑ ogólną ocenę portalu (w skali 0-3).

Zdefiniowano następujące oczekiwane wartości miar:

- ❑ Średnia efektywność nawigacji > 60%
- ❑ Średni czas wykonania zadania > 2
- ❑ Średnia ocena nawigacji > 2
- ❑ Średnia ogólna ocena portalu > 2

W testach wzięły udział trzy osoby.

Wyniki pierwszych testów

Tablica 1. Efektywność nawigacji (próg 60%)

Zadanie 1	15%	-
Zadanie 2	69%	+
Zadanie 3	33%	-
Zadanie 4	72%	+
Średnia	47%	-

Tablica 2. Czas wykonania zadania (próg 2)

Zadanie 1	0,7	-
Zadanie 2	2,3	+
Zadanie 3	1,0	-
Zadanie 4	2,3	+
Średnia	1,6	-

Tablica 3. Oceny subiektywne (próg 2)

Nawigacja	1,5	-
Ocena ogólna	2,0	+

Przeanalizowano także liczbę problemów zidentyfikowanych w trakcie testów:

Tablica 4. Liczba problemów

Zidentyfikowane przez wszystkich uczestników	5
Zidentyfikowane przez dwoje uczestników	5
Zidentyfikowane przez jednego uczestnika	4

Ponieważ wyniki nie osiągnęły wymaganych minimalnych wartości, przeprowadzono zmiany w systemie uwzględniające zidentyfikowane problemy, a następnie powtórzono testy.

Wyniki drugich testów

Tablica 1. Efektywność nawigacji (próg 60%)

	Wynik		Poprawa
Zadanie 1	100%	+	+567%
Zadanie 2	58%	-	-16%
Zadanie 3	54%	-	+64%
Zadanie 4	100%	+	+39%
Średnia	78%	+	+66%

Tablica 2. Czas wykonania zadania (próg 2)

	Wynik		Poprawa
Zadanie 1	2,8	+	+300%
Zadanie 2	1,7	-	-26%
Zadanie 3	1,8	-	+80%
Zadanie 4	3,0	+	+30%
Średnia	2,3	+	+44%

Tablica 3. Oceny subiektywne (próg 2)

	Wynik		Poprawa
Nawigacja	2,2	+	+47%
Ocena ogólna	2,3	+	+15%

Tablica 4. Liczba nowych znalezionych problemów

Zidentyfikowane przez wszystkich uczestników	0
Zidentyfikowane przez dwoje uczestników	2
Zidentyfikowane przez jednego uczestnika	5

Ponieważ wyniki przekroczyły wymagane minimalne wartości, cykl testów został zakończony a portal udostępniony w Intranecie.

Wnioski

Testy potwierdziły, że już przy niewielkiej liczbie uczestników (w tym przypadku trzy osoby) uzyskać można istotną poprawę użyteczności oprogramowania.

Wiarygodność testów potwierdza fakt, że 36% znalezionych problemów w trakcie pierwszych testów zidentyfikowana została przez wszystkich uczestników. W przypadku przeprowadzenia testów przez tylko dwóch użytkowników, liczba znalezionych problemów zmalałaby jedynie o 1 lub 2, jednakże dużo trudniej byłoby ocenić ich wiarygodność.

Bardzo ciekawe okazało się porównanie oceny subiektywnej z obiektywnymi wynikami testów osoby #3:

- ❑ subiektywna ocena nawigacji spadła o 50% (z 2 na 1),
- a jednocześnie:
- ❑ średni czas wykonania zadań wzrósł o 63% (z 1 na 1,63)
 - ❑ średnia efektywność nawigacji wzrosła o 145% (z 23% na 56%)

W tym przypadku wyniki obiektywne nie potwierdzają subiektywnego, negatywnego odczucia tej osoby, dotyczącego zmian w modelu nawigacji. Ponieważ końcowi użytkownicy widzieć będą tylko jedną wersję, wyniki miar obiektywnych, w połączeniu z wyższą oceną nowego modelu nawigacji przez pozostałe osoby, należy przyjąć za wiarygodne.

Literatura

1. Jacob Nielsen; „Why You Only Need to Test With 5 Users”; Jacob Nielsen’s Alertbox, March 19, 2000; <http://www.useit.com/alertbox/20000319.html>