

METODY I TECHNIKI EWALUACJI INTERAKCJI CZŁOWIEK-KOMPUTER, A TRADYCYJNY PODZIAŁ BADAŃ: JAKOŚCIOWE VS ILOŚCIOWE

Joanna KWIATKOWSKA

Streszczenie: Niniejszy artykuł stanowi próbę stworzenia klasyfikacji metod badań użyteczności zgodnie z tradycyjną typologią badań: jakościowe vs. ilościowe. Autor podjął się niniejszego zadania ze względu na rosnące znaczenia testowania jakości użytkowej produktów informatycznych oraz rozwój dziedziny *interakcja człowiek-komputer* (ang. *Human-Computer Interaction*, HCI), której integralną składową stanowi użyteczność oraz brak w literaturze przedmiotu podobnych opracowań.

Słowa kluczowe: badania jakościowe, badania ilościowe, użyteczność, interakcja człowiek-komputer, serwis internetowy.

1. Wprowadzenie

Przydatność serwisu internetowego dla danej instytucji określa jakość, która w standardzie ISO 9126:1991 została zdefiniowana jako [...] *całość cech i właściwości produktu programowego wpływająca na jego zdolność do zaspokojenia określonych lub implikowanych potrzeb* [3]. Oceniając jakość produktu informatycznego często zwraca się uwagę na następujące jego cechy:

- funkcjonalność (ang. *functionality*),
- niezawodność (ang. *reliability*),
- użyteczność (ang. *usability*),
- efektywność (ang. *efficiency*),
- łatwość utrzymania (ang. *maintainability*),
- przenośność (ang. *portability*).

Jedną z charakterystyk wymienionych powyżej – użyteczność, znalazła się w ścisłym kręgu zainteresowań badaczy zajmujących się prężnie rozwijającą się dziedziną nauki: interakcja człowiek-komputer (ang. *human-computer interaction*, HCI). Istnieje wiele definicji użyteczności (ang. *usability*), często określanej również jako jakością użytkową.

Według normy ISO 9241-11 jest to *miara wydajności, efektywności i satysfakcji z jaką dany produkt może być używany przez określonych użytkowników dla osiągnięcia określonych celów w określonym kontekście użycia*.

Ekspert w dziedzinie użyteczności - Jakob Nielsen, uważa z kolei, że jest to atrybut jakości, który mówi o łatwości użytkowania produktu interaktywnego i może oznaczać także zbiór metod używanych do doskonalenia użyteczności produktu w procesie projektowania [3]. Nielsen rozszerza ponadto swoją definicję wprowadzając pięć elementów, które charakteryzują użyteczność:

- łatwość nauki, przyswajalność (ang. *learnability*),
- efektywność (ang. *efficiency*),

- zapamiętywalność (ang. *memorability*),
- błędy (ang. *errors*),
- satysfakcja (ang. *satisfaction*).

Marcin Sikorski, autor pierwszego polskiego podręcznika akademickiego z dziedziny HCI wyróżnia dwie składowe jakości użytkowej, do których należą:

- satysfakcja użytkownika, która wynika z *rzeczywistego* korzystania z produktu informatycznego,
- oraz kompatybilność – jednoczesne zapewnienie jakości technicznej i ergonomicznej [3].

W literaturze można spotkać się również z definicjami użyteczności zaproponowanymi przez takich specjalistów jak Krug [5], Szulc [6], czy Karwatka [7], których interpretacja jest tożsama.

Użyteczność jest zatem stopniem satysfakcji użytkownika, który został osiągnięty podczas pracy z danym systemem informatycznym (serwisem internetowym). Wysoki poziom użyteczności to łatwość pracy z systemem, komfort jego użycia, osiągnięcie za jego pośrednictwem postawionych celów.

Ze względu na rosnącą konkurencję, a także sytuację jaka utrzymuje się w gospodarce po okresie kryzysu w latach 2009-2010, przedsiębiorstwa które poprzez wdrożenie produktu informatycznego (w tym serwisu internetowego) oczekują odniesienia takich korzyści jak podniesienie skuteczności sprzedaży, zwiększenie świadomości marki, czy też zdobycie lojalności klienta, stanęły przed koniecznością optymalizacji użyteczności i określenia mierzalnych celów projektów informatycznych. Prowadzenie badań użyteczności na każdym etapie procesu projektowania systemu informatycznego zapewnia możliwość dokonania w nim szybkich zmian, co zmniejsza ryzyko niepowodzenia całego przedsięwzięcia.

Użyteczność – potrzeba klasyfikacji technik badawczych

Użyteczność jest dziedziną interdyscyplinarną, która odwołuje się do takich nauk jak ekonomia, psychologia, informatyka, kognitywistyka i wiele innych. W niektórych technikach *pomiaru* użyteczności wykorzystywane są nawet osiągnięcia medycyny. Istnieje zatem wiele metod ewaluacji interakcji człowiek-komputer. Najczęściej tego typu badania klasyfikowane są po prostu według podziału na wszelkiego rodzaju testy z udziałem użytkowników oraz audyty eksperckie (analizy heurystyczne). Liczne agencje badawcze oferują swoim klientom przeprowadzanie badań bez wskazania, jakie cele dzięki ich realizacji można osiągnąć. Ponadto autorka niniejszego artykułu nie spotkała się w literaturze przedmiotu z odniesieniem do tradycyjnej typologii badań, gdzie wyróżnia się podział na metody ilościowe i jakościowe. W związku z zaprezentowanymi powyżej argumentami, a także obserwowanym wzrostem popularności realizacji badań użyteczności w Polsce, autorka dostrzega konieczność stworzenia klasyfikacji metod i technik badawczych użyteczności zgodnej z podziałem: badania jakościowe vs. Ilościowe. Próbę realizacji tego zadania stanowi niniejszy artykuł.

Badania jakościowe vs. Ilościowe

Wybór metody badawczej, która zostanie wykorzystana diagnozy danego produktu informatycznego powinien zależeć od tego, czego badacz poprzez jej zastosowanie chce się dowiedzieć. W literaturze można spotkać wiele dysput głoszących wyższość badań ilościowych względem jakościowych [11] i vice versa.

Badania ilościowe reprezentują tzw. *twarde* podejście metodologiczne. Są obiektywne, opierają się na statystyce i pewnej strukturze. Celem tego typu badań jest udowodnienie

postawionej wcześniej hipotezy, ale niestety wielu procesów nie da się w ten sposób zweryfikować.

Metody ilościowe zakładają istnienie reguł, dokonują ich pomiaru, opisują relacje między poszczególnymi zmiennymi, wykrywają zależności.

Minusy metod ilościowych:

- często nie uwzględniają *czynnika ludzkiego*,
- podejście statystyczne może być zawodne w kontekście testowania hipotez w oparciu o dane [11],
- często pewne badane czynniki są określane *z góry*, co przesądza o wynikach takich badań.

Do metod badań ilościowych należą wszelkiego rodzaju ankiety, wywiady telefoniczne (ang. *Computer Assisted Telephone Interview - CATI*), wywiady bezpośrednie (ang. *Computer Assisted Personal Interviewing - CAPI*), wywiady realizowane za pośrednictwem Internetu (ang. *Computer Assisted Web Interviewing - CAWI*).

Wyniki badań ilościowych są obiektywne i udzielają odpowiedzi na pytanie *ile*. Z tego względu znajdują najczęściej zastosowanie w przypadku badania serwisów o charakterze sprzedażowym [16]. Obecnie badania ilościowe użyteczności stanowią tylko sposób identyfikacji problemów, których dalsza analiza realizowana jest z wykorzystaniem metod jakościowych.

Badania jakościowe często uważa się za przeciwieństwo badań ilościowych. Ich celem jest przede wszystkim odzwierciedlenie punktu widzenia uczestnika badania. Zastosowanie tego typu badań w przypadku testowania użyteczności ma przede wszystkim na celu określenie przyczyn zachowania użytkownika podczas pracy z określonym produktem informatycznym, ma również pomóc zidentyfikować problemy, a także nawet twórczo rozwinąć dany produkt (udoskonalić go). Denzin i Lincoln zaproponowali następującą definicję badań jakościowych:

Badanie jakościowe jest usytuowaną aktywnością, która umieszcza obserwatora w świecie. Składa się z zespołu interpretacyjnych, materialnych praktyk, które czynią świat widzialnym. Praktyki te przekształcają świat. Przeobrażają go w serie reprezentacji, takich jak notatki terenowe, wywiady, rozmowy, fotografie, nagrania i własne uwagi. Na tym poziomie badania jakościowe to interpretatywne, naturalistyczne podejście do świata. Oznacza to, że badacze jakościowi badają rzeczy w ich naturalnym środowisku, próbując nadać sens lub interpretować zjawiska przy użyciu terminów, którymi posługują się badani ludzie [13].

Badanie jakościowe często określane jest jako *dociekanie* (ang. *inquiry*), czy też podejście interpretatywne.

Do najczęściej stosowanych metod badań jakościowych należą wywiady swobodne, indywidualne wywiady pogłębione (IDI), zogniskowane wywiady grupowe (FGI), obserwacje uczestniczące, metody fotograficzne (wizualne) oraz etnograficzne.

Wyników badań jakościowych nie można poddać obróbce statystycznej, mają one najczęściej charakter subiektywny i udzielają odpowiedzi na pytanie *dlaczego*.

Użyteczność jest dziedziną naukową, w przypadku której coraz większa liczba młodych badaczy wykorzystuje w swoich dociekaniach metody jakościowe lub ich połączenie z metodami ilościowymi. Celem badań użyteczności jest odnalezienie sensu zachowania się użytkownika, a badania jakościowe mają na celu właśnie opracowanie wzorców postępowania badanego obiektu.

Istnieje wiele sposobów łączenia tych dwóch typów badań, np. mające na celu wzajemną kontrolę wyników, dostarczanie inspiracji, czy też stanowiące tylko uzupełnienie.

Alan Bryman wskazał aż jedenaście sposobów łączenia badań ilościowych i jakościowych:

- triangulacja, która ma na celu uwiarygodnienie wyników,
- wspieranie jednego podejścia przez drugie,
- łączenie w celu uzyskania szerszego horyzontu badawczego,
- badanie cech strukturalnych danego zjawiska stosując metody ilościowe, a aspekty procesualne – metody jakościowe,
- przedstawienie punktu widzenia badanych – metody jakościowe, a badaczy – metody ilościowe,
- wyniki badań jakościowych można uzupełnić o wyniki badań ilościowych, co niweluje problem reprezentatywności,
- można wykorzystać wyniki jakościowe do interpretacji relacji między zmiennymi w zbiorach danych ilościowych,
- łączenie podejścia jakościowego z ilościowym może ukazać relację między poziomem mikro i makro badanego zagadnienia,
- różne badania można stosować na różnych etapach procesu badawczego,
- tworzenie form hybrydowych [2].

Tashakkori i Teddlie nazwali dążenia do utworzenia połączenia między podejściem jakościowym i ilościowym *metodologią trzeciej drogi* [2].

2. Metody i techniki ewaluacji interakcji człowiek-komputer

W niniejszej części artykułu autor dokona próby klasyfikacji i krótkiej prezentacji znanych mu technik i metod ewaluacji interakcji człowiek-komputer zgodnie z tradycyjnym podziałem badań na metody jakościowe i ilościowe.

Celem wprowadzenia miar użyteczności jest wskazanie problemów oraz zidentyfikowanie ich przyczyn, porównanie różnych wersji rozwiązań danego produktu informatycznego, czy też zbadanie satysfakcji użytkownika podczas pracy z danym systemem (lub na przykład serwisem informatycznym). Miary użyteczności umożliwiają również pomiar wskaźnika ROI – zwrot z inwestycji (ang. *Return on Investment*) oraz są przydatne podczas kontroli postępów w realizacji projektu informatycznego.

Tullis i Albert stworzyli klasyfikację miar użyteczności ze względu na mierzoną jednostkę:

- miary wydajności (ang. *performance metrics*),
- miary problemów (ang. *issue-based metrics*),
- miary fizjologiczne i behawioralne (ang. *behavioral and physiological metrics*),
- miary łączone i porównawcze (ang. *combined and comparative metrics*).

Ponieważ jest to jedyny podział metod ewaluacji użyteczności z jakim autorka spotkała się w literaturze przedmiotu, zostanie on wykorzystany w celu przedstawienia i skategoryzowania badań ilościowych.

Metody ilościowe

Miary wydajności

Pierwszą grupą metryk, które można zaliczyć do metod ilościowych s miary wydajności, które stanowią bardzo cenne źródło informacji dla badacza użyteczności.

Wskaźniki wydajności powstają w oparciu o zachowania użytkowników, którzy realizują przygotowane przez eksperta użyteczności zadania na podstawie określonego scenariusza. Miary wydajności identyfikują skuteczność pewnego rozwiązania (lub brak tej skuteczności). W celu poszerzenia wiedzy o przyczyny takiego stanu badacz powinien rozszerzyć swoje dochodzenia o dodatkowe obserwacje (obserwacje użytkowników zalicza się do **metod jakościowych**).

Można wyróżnić następujące miary wydajności:

- wykonanie (ukończenie) zadania (ang. *task success*) – miara podstawowa. Sprawdza efektywność ukończenia przez użytkownika przygotowanych dla niego zadań. Istnieją dwa typy tej miary: sukces binarny (ang. *binary success*) oraz poziom wykonania zadania (ang. *levels of success*),
- czas wykonania zadania (ang. *time-on-task*) – pomiar czasu niezbędnego do realizacji danego zadania,
- błędy (ang. *errors*) – pomiar ilości błędów popełnionych podczas wykonywania zadania,
- efektywność (ang. *efficiency*) – pomiar wysiłku użytkownika, jaki musi on włożyć w celu realizacji przygotowanego dla niego zadania,
- zdolność nauki (przyswajalność) (ang. *learnability*) – np. różnice w czasie wykonania zadania wynikające z powtórnego korzystania z systemu.

Pomiar wykonania zadania jest możliwy tylko wówczas, kiedy badacz odpowiednio zdefiniuje kryteria zadania przygotowanego dla użytkownika. Tego typu pomiar dokonywany jest najczęściej w laboratorium, przy zastosowaniu papierowego formularza lub

narzędzi on-line. Przykładem dobrze sformułowanego zadania jest następujące polecenie: *wskaż kiedy kończy się rekrutacja na studia na Uniwersytecie Warszawskim (Wydział Zarządzania, kierunek – Zarządzanie, studia dzienne I stopnia)*. Przykładem niepoprawnie zdefiniowanego zadania może być natomiast następujące sformułowanie: *jakie w tym roku kierunki studiów cieszyły się największą popularnością wśród studentów*.

Bardziej restrykcyjną formą miary wykonania zadania jest sukces binarny. Jeżeli użytkownik poprawnie zrealizuje zadanie, w formularzu badania dla danej zmiennej wpisywana jest wartość 1, natomiast nieukończenie zadania oznacza przypisanie wartości 0. Mając tak przygotowany formularz można bardzo łatwo obliczyć średnią wykonania zadania, a także przedział ufności. Czytelną formą przedstawienia wyników jest też przygotowanie zestawienia procentowej realizacji poszczególnych zadań przez użytkownika. Obliczenie statystycznie znaczących różnic między poszczególnymi zadaniami można sprawdzić wykonując np. analizę wariancji ANOVA (ang. *analysis of variance*).

Wykonanie zadania można przedstawić również wprowadzając jego odpowiednie poziomy (miara: poziomy wykonania zadania). W ten sposób możemy zbadać częściowe wykonanie zadania przez użytkownika. Aby zastosować tę miarę, badacz musi zdefiniować odpowiednie poziomy wykonania zadania wg których dokona oceny. Można wyróżnić następujące poziomy:

- przedziały (zakresy), które osiągnął użytkownik realizując dane zadanie,
- zakresy, które wynikają z wcześniejszych doświadczeń w pracy z badanym produktem informatycznym,
- zakresy, które pomagają ocenić, czy użytkownik realizując dane zadanie robi to w optymalny sposób.

Osiągnięcie poszczególnego poziomu można zaznaczyć stosując sześciostopniową skalę ukończenia zadania:

- prawidłowe wykonanie całego zadania (bez pomocy moderatora),
- prawidłowe wykonanie całego zadania (z pomocą moderatora),
- częściowe wykonanie zadania (bez pomocy moderatora),
- częściowe wykonanie zadania (z pomocą moderatora),
- nieukończenie zadania (użytkownik uważa, że zrealizował zadania, ale tak nie jest – mylne przekonanie),
- nieukończenie zadania (użytkownik zrezygnował z realizacji zadania, gdyż przerosło ono jego możliwości).

W celu obliczenia wartości średniej poziomu wykonania zadania, stopniom zaprezentowanym powyżej można przypisać odpowiednie wagi.

Innym podejściem umożliwiającym pomiar poziomu wykonania zadania jest uwzględnienie doświadczenia użytkownika. Można to zrobić stosując skalę dokonującą pomiaru, czy zadanie przygotowane dla użytkownika sprawiało mu jakieś problemy, np.:

- brak problemów,
- niewielkie problemy (użytkownik wykonał zadanie, ale popełnił drobne błędy lub doszedł do rozwiązania nie wprost),
- znaczące problemy (użytkownik wykonał zadanie, ale realizując je popełnił rażące błędy),
- fiasko realizacji zadania.

Prowadząc pomiar wykonania zadania badacz może mieć trudności z oceną sytuacji, kiedy użytkownik dobrze rozwiązuje badanie, ale finalnie podaje błędną odpowiedź. Również problematyczne jest określenie momentu, w którym badacz powinien przerwać użytkownikowi wykonanie zadania (użytkownik podejmuje kolejne próby i mimo to nie może osiągnąć celu wykonania zadania).

Istotnych informacji na temat użyteczności produktu dostarcza kolejna miara: czas wykonania zadania. Szybsza realizacja zadania jest dla użytkownika satysfakcjonująca, a to z kolei przekłada się na zwiększenie skuteczności, wydajności, czyli przynosi zwrot z inwestycji (ROI, ang. *return on investment*) i prowadzi do oszczędności kosztów. Czas wykonania zadania to po prostu przedział czasowy mierzony od momentu rozpoczęcia realizacji zadania do jego zakończenia. Tullis i Albert wymieniają wiele narzędzi do automatycznego pomiaru czasu, np. Ergo Browser, Data Logger, Bailey's, Usability Testing Environment (UTE). Pomiar czasu wykonania zadania można dodatkowo wzbogacić obliczając takie miary jak średnia geometryczna, kres górny (ang. *upper bound*), kres dolny (ang. *lower bound*), czy przedział ufności czasu wykonania poszczególnych zadań.

Pomiar czasu wykonania zadania można dodatkowo wzbogacić wprowadzając proggi czasowe realizacji poszczególnych zadań. Mogą one być ustalone przez badacza.

W przypadku pomiaru czasu wykonania zadania problem stanowią wyniki osób, które nie ukończyły zadania – uwzględniać je, czy nie. Często na przebieg realizacji zadania znacząco wpływa decyzja, czy badacz ma poinformować uczestnika testu o tym, że czas w jakim wykona on zadanie jest mierzony, czy jest to wykonywane niejawnie.

Jeżeli chcemy dowiedzieć się co przyczynia się do znacznego spadku wydajności badanego produktu informatycznego i w związku z tym generuje duże koszty, warto dokonać pomiaru błędów. Pomiar błędów najczęściej odbywa się poprzez obserwację użytkownika, analizę nagrań video, czy wypełnienie formularza zawierającego listę

możliwych błędów jakie może popełnić badana osoba, która została opracowana przez prowadzącego badanie. Jeżeli seria zadań jest tak skonstruowana, że można w każdym z nich popełnić jeden błąd, wówczas po obliczeniu częstości występowania błędu w każdym zadaniu można wskazać te z nich, które są najbardziej kłopotliwe dla użytkowników. Istnieją również inne podejścia prezentacji danych:

- odsetek uczestników, którzy popełnili dany błąd,
- obliczenie całkowitego błędu w badaniu,
- ustalenie prognozy ilości błędów dla danego zadania i pomiar tylko tych przypadków, gdzie próg został przekroczony.

Jeżeli w danym zadaniu można popełnić kilka błędów, wówczas warto obliczyć tzw. wskaźnik błędów. Konstruuje się go sumując liczbę błędów dla danego zadania i dzieląc tę sumę przez całkowitą liczbę możliwych błędów. Do innych sposobów prezentacji pomiarów błędów wielokrotnych w zadaniu należą:

- obliczenie liczby błędów, jaką popełnił pojedynczy użytkownik w poszczególnych zadaniach,
- próg ilości błędów dla każdego zadania lub pojedynczego użytkownika,
- wprowadzenie wag dla poszczególnych błędów dla pojedynczego zadania, co umożliwi obliczenie średniej miary błędów.

Ilość kroków, jakie użytkownik musi wykonać w celu realizacji danego zadania, stanowi kolejną miarę wydajności jaką jest efektywność. Projektanci interakcji dążą do opracowywania takich rozwiązań funkcjonalnych, gdzie liczba akcji, jakie należy podjąć w celu realizacji określonego zadania, jest możliwie zminimalizowana. Pomiar efektywności stanowi zatem wskaźnik, który określa czy dane rozwiązanie interakcji należy jeszcze uprościć, czy jest ono satysfakcjonujące. Pomiar efektywności może odbywać się np. poprzez zliczenie ruchów myszką jakie wykonał użytkownik w celu zrealizowania zadania. Badacz może stworzyć listę innych akcji podejmowanych przez użytkownika, które będą składać się na miarę efektywności (np. wprowadzenie tekstu w odpowiednie miejsce, przewinięcie strony, etc).

Innym sposobem pomiaru efektywności jest obliczenie współczynnika zagubienia na stronie (ang. *lostness*), który został zaproponowany w 1996 przez Pauline Smith [14]. Wygląda on następująco:

$$L = \sqrt{\left(\frac{N}{S} - 1\right)^2 + \left(\frac{R}{N} - 1\right)^2} \quad (1)$$

gdzie:

- **L** – współczynnik zagubienia,
- **N** – liczba różnych stron odwiedzonych podczas wykonania zadania,
- **S** – liczba wszystkich stron odwiedzonych podczas wykonania zadania (całkowita liczba stron), wliczając ponowne wizyty na tych samych stronach,
- **R** - minimalna liczba stron, które należy odwiedzić, aby wykonać zadanie.

Według badań Smith, użytkownicy zaczynają odczuwać zagubienie na stronie dla wartości wskaźnika powyżej 0,42 [14]. Idealna wartość zagubienia wynosi 0 [1].

Efektywność można potraktować również jako proporcję dwóch miar: wykonania zadania do średniego czasu wykonania zadania.

Aby obliczyć liczbę zadań ukończonych pozytywnie w określonej jednostce czasu (np. minuta) można podzielić sumę wszystkich zakończonych pozytywnie zadań przez

całkowity czas spędzony na realizacji wszystkich zadań (zakończonych zarówno sukcesem jak i niepowodzeniem).

Kolejna miara – zdolność nauki (przyswajalność) pozwala ocenić, czy użytkownik nabiera biegłości w obsłudze danego produktu informatycznego w czasie. Aby dokonać pomiaru zdolności nauki, badacz w pierwszej kolejności musi ustalić interwał czasowy pomiędzy kolejnymi próbami, czyli odwiedzinami badanej strony. Następnie w ramach pojedynczej próby dokonywane są pomiary czasu wykonania zadania, efektywności, błędów, etc., czyli pomiary metryk wydajności. Miara zdolności nauki jest zatem zagregowaną wartością danej metryki wydajności ze wszystkich prób, w których prowadzono badanie.

Wiele zachowań jakie wykazują użytkownicy podczas badania (np. śmiech, komentowanie niektórych rozwiązań, mimika twarzy) może zostać zmierzonych i zaprezentowanych

w postaci miary psychologicznej i behawioralnej. Albert i Tullis zaproponowali podział tego typu miar ze względu na ich charakter werbalny lub niewerbalny.

Pomiar zachowań werbalnych polega na zliczeniu wypowiedzianych przez użytkownika komentarzy pozytywnych i negatywnych dotyczących badanego serwisu, których proporcja pozwala na utworzenie odpowiedniego współczynnika. Badacz przystępując do pomiaru musi skatalogować zachowania werbalne i przypisać je do jednej z trzech grup: pozytywne, negatywne, neutralne. Dzięki takiemu podziałowi można obliczyć częstość występowania w danym badaniu komentarzy danej kategorii.

Do mierzalnych zachowań niewerbalnych można zaliczyć mimikę twarzy użytkownika, śledzenie ruchu gałek ocznych, zmianę średnicy źrenicy, przewodność skóry, czy częstość akcji serca.

Mimikę twarzy użytkownika można zbadać mierząc aktywność określonych mięśni twarzy (wahania potencjałów włókien mięśniowych – polaryzacja i depolaryzacja), do czego stosuje się urządzenie zwane elektromiogramem. Istnieją również próby wprowadzenia oprogramowania komputerowego, które automatycznie rozpoznaje wyrazy twarzy i przypisuje je odpowiednim emocjom. Jednym z przykładów takich systemów jest program FaceReader opracowany przez firmę Noldus [17]. Potrafi on rozpoznać sześć różnych emocji: zadowolenie, smutek, strach, obrzydzenie, zdziwienie, złość. Działanie programu polega na analizie obrazu video (automatycznie wykrywa na obrazie twarz i dokonuje jej klasyfikacji, realizuje modelowanie 3D twarzy uwzględniając 491 punktów kluczowych i dokonuje analizy wyrazów twarzy nawet wtedy, gdy użytkownik nie patrzy na wprost kamery rejestrującej obraz).

Do coraz częściej wykorzystywanych w badaniach użyteczności urządzeń należą eyetrackery. Ich funkcjonalność polega na monitorowaniu miejsca, na które w danym momencie patrzy użytkownik. Zasada działania tych urządzeń opiera się na odbiciu przez powierzchnię oka podczerwieni, która emitowana jest przez eyetracker. Dzięki danym zebranych przez eyetracker możliwe jest obliczenie ilu badanych użytkowników skupiło swoją uwagę na kluczowym elemencie oraz czas trwania fiksacji wzroku.

Eyetrackery umożliwiają dodatkowo pomiar średnicy źrenicy użytkownika. Rozszerzanie się jej i zwężanie nie oznacza jedynie fizjologicznej reakcji na światło, lecz dodatkowo dostarcza informacji sygnalizujących o zachodzących procesach kognitywnych. Pomiar reakcji źrenicy nazywany jest pupilometrią.

Pomiar przewodności skóry (ang. *Galvanic Skin Response, GSR*) wykonywany za pomocą urządzenia zwanego psychogalwanometrem, a także pomiar zmienności rytmu serca (ang. *heart rate variability, HRV*) także wykorzystywane są w badaniach

użyteczności stron internetowych, gdyż obrazują reakcje na stres i odczuwanie silnych emocji.

Mając do dyspozycji wiele różnych miar można pokusić się o stworzenie nowych metryk przez kombinację tych istniejących. Zazwyczaj można spotkać się z dwoma podejściami prowadzącymi do powstania nowych miar: połączenie kilku metryk w jedną lub porównanie niektórych miar z np. wynikami eksperckimi (wzorcowymi).

Często utworzenie nowej metryki sprowadza się do zmiany formy prezentacji danych. Na przykład w celu uproszczenia czytelności wyników badania użyteczności, np. przeznaczonych dla managerów, ustala się próg dla metryki wykonanie zadania i czas wykonania zadania, powyżej którego przyjmuje się, że użytkownik zrealizował zadanie pomyślnie. W ten sposób można wprowadzić nową metrykę, np. *pomyślnie ukończenie zadania*, która przyjmuje dwie wartości: tak/nie. Odczyt tak zaprezentowanych danych, np. przez kadry zarządzające jest prostszy i umożliwia szybsze podjęcie decyzji.

J. Sauro i E. Kindlund (2005) [18] zaproponowali standaryzowaną, podsumowującą i pojedynczą miarę użyteczności: SUM (ang. *single usability metric*), która stanowi odpowiednie połączenie kilku miar. Celem SUM jest jednoznaczne określenie użyteczności badanego produktu informatycznego.

J. Sauro i E. Kindlund zaproponowali również adaptacje metodologii Sześć Sigma (ang. *Six Sigma*) do oceny użyteczności. Sześć Sigma jest metodą zarządzania jakością stosowaną w największych światowych koncernach.

Badacze użyteczności często dokonują także pomiaru problemów których doświadczają użytkownicy podczas pracy z danym produktem informatycznym, np. niedostrzeżenie tego, co użytkownik powinien dostrzec, dezorientacja, niezrozumienie jakiegoś elementu. W ten sposób wyodrębniono grupę miar problemów.

Aby pomiar problemów był efektywny, badacz użyteczności musi przede wszystkim dobrze zdefiniować samo pojęcie problemu oraz świetnie znać badany produkt. Problem mogą stanowić wszelkie przeszkody uniemożliwiające pozytywne ukończenie zadania, mogą to być zachowania wynikające z dezorientacji czy podjęcia błędnego działania, etc.

Najprostszym sposobem identyfikacji błędów jest wnikliwa analiza komentarzy jakich udzielili użytkownicy podczas badania. Często wprowadza się tzw. wskaźniki ciężkości błędów (ang. *severity ratings*), dzięki którym można ustalić ważność zidentyfikowanych problemów.

Wskaźnik ciężkości błędów może reprezentować poziomy problemów, których podział wynika z doświadczenia użytkownika, np. małe błędy, błędy umiarkowane, katastrofalne. Może również uwzględniać wpływ, jakie dane problemy mają na realizację celów biznesowych, czy na doświadczenia użytkownika.

Problemy jakich doświadcza użytkownik można również skategoryzować, np. problemy związane z nawigacją, architekturą informacji, funkcjonalnością, etc.

Miarę problemów może stanowić obliczenie średniej liczby błędów na każdego uczestnika badania, czy też obliczenie odsetka użytkowników, którzy mieli problem z tym samym elementem. Jeżeli dane badanie jest powtarzane w kilku sesjach, warto zliczyć unikatowe problemy, aby móc następnie porównać wyniki.

Miarami, które najlepiej obrazują postrzeganie przez użytkownika danego produktu informatycznego są miary dostarczane przez samego użytkownika, które często określa się danymi subiektywnymi (ang. *subjective data*) lub danymi preferencji (ang. *preference data*). Do dwóch najczęściej stosowanych metod zbierania tego typu danych należą skala Likert'a oraz skala zróżnicowania semantycznego (ang. *semantic differential scale*).

Skala Likert'a oparta jest zazwyczaj na 5-stopniowej skali według której użytkownicy wyrażają swoją opinię na temat przygotowanego przez badacza twierdzenia. Poszczególne stopnie skali przybierają najczęściej następującą formę:

- zdecydowanie się nie zgadzam,
- nie zgadzam się,
- nie mam zdania,
- zgadzam się,
- zdecydowanie zgadzam się.

Skala zróżnicowania semantycznego umożliwia zaprezentowanie par bipolarnych/przeciwstawnych na dwóch różnych końcach, np.:

Słaby ○ ○ ○ ○ ○ ○ ○ Silny

Piękny ○ ○ ○ ○ ○ ○ ○ Brzydki

Między tymi przeciwstawnymi przymiotnikami zazwyczaj stosuje się 5- lub 7-stopniową skalę.

Do innych miar dostarczanych przez użytkownika należą:

- Miara oczekiwań (ang. *expectation measure*) – użytkownicy przed podejściem do badania wypełniają formularz przedstawiający ich oczekiwania względem systemu (np. poziom trudności). Po przeprowadzeniu badania ocena jest wykonywana ponownie, dzięki czemu można otrzymać dwa wskaźniki: wskaźnik oczekiwania oraz wskaźnik doświadczenia (ang. *expectation rating i experience rating*),
- SUS (ang. *system usability scale*) – formularz oceny opracowany przez John'a Brook'a. Składa się z 10 twierdzeń i 5-stopniowej skali oceny dla każdego z nich. 5 twierdzeń posiada zabarwienie pozytywne, a 5 - negatywne.

Metody jakościowe

Najcenniejszą metodą zdobycia wiedzy na temat badanego rozwiązania informatycznego są obserwacje użytkowników, które zalicza się też do grupy testów z udziałem użytkowników. Najczęściej obserwacje użytkowników służą wyjaśnianiu dylematów przed jakimi stoją projektanci, które dotyczą np. wybranych interakcji. Badacz obserwujący zachowanie użytkownika powinien mieć przygotowany scenariusz badania, który zawiera pytania stanowiące zadania dla użytkownika jakie musi on wykonać w badanym systemie w celu udzielenia na nie odpowiedzi.

Kolejna metoda jakościowa, analiza heurystyczna, zaliczana jest do metod analizy eksperckiej. Zazwyczaj stosowana jest w celu dokonania oceny istniejącego już serwisu lub dostarcza informacji niezbędnych do przygotowania jego nowszej wersji. Często uważa się, że jest to jedna z tańszych i należąca do najszybszych metod oceny użyteczności. Analiza heurystyczna pozwala stwierdzić w jakim stopniu badana strona jest zgodna z przyjętymi standardami użyteczności. Jest to rodzaj ewaluacji, w trakcie której użytkownik nie bierze udziału, a ekspert prowadzący audyt nie zalicza się do grona potencjalnych odbiorców (klientów) badanego serwisu.

Analiza heurystyczna została wprowadzona przez pioniera użyteczności – Jakoba Nielsena przy współudziale Ralfa Molicha (1990) [20]. Zakłada ocenę np. serwisu internetowego przez kilku niezależnych ekspertów, z których każdy mając do dyspozycji ujednoczoną listę kontrolną sygnalizuje o występowaniu odbiegających od normy rozwiązaniach (zdarzeniach, etc.) i wnosi sugestie środków (kroków) naprawczych. Jakob Nielsen sugeruje, że efektywne przeprowadzenie badania powinno odbywać się z udziałem trzech doświadczonych ekspertów. Wynikiem realizacji analizy heurystycznej jest opracowanie zbiorczego raportu, który zawiera listę błędów często uszeregowanych pod względem ich znaczenia (np. błędy krytyczne, umiarkowane, etc.).

Przeгляд kognitywny (ang. *cognitive walkthrough*) jest metodą, która polega na zrealizowaniu krok po kroku scenariusza wykonania zadania w na przykład badanym serwisie internetowym. W ten sposób grupa badaczy prowadzących przeгляд kognitywny (np. projektanci serwisu, osoby które zleciły jego wykonanie, etc.) wcielają się w rolę użytkowników tego serwisu i starają się naśladować ich zachowanie.

Sortowanie kart (ang. *card sorting*) to kolejna metoda jakościowa. Stosuje się ją zazwyczaj przed przystąpieniem do pracy nad tzw. *klikanym* prototypem. Jest to metoda wpisująca się metodologię projektowania zorientowanego na użytkownika (ang. *user centred design*, UCD). Jej celem jest sprawdzenie w jaki sposób potencjalni użytkownicy danego serwisu internetowego (systemu informatycznego) postrzegają zaproponowany przez projektantów układ treści. Użytkownicy podczas badania dokonują kategoryzacji lub grupowania informacji, opracowują słowa kluczowe i przedstawiają swój sposób myślenia.

3. Podsumowanie

Jak zaprezentowano w niniejszym artykule, badania użyteczności stanowią gamę różnych metod badawczych. Oczywiście autorka nie przedstawiła wszystkich istniejących technik ewaluacji, a jedynie ich podstawowe rodzaje. Celem dokonanego przeglądu było opracowanie podziału miar użyteczności ze względu na ich charakter jakościowy lub ilościowy.

Literatura

1. Tullis T., Albert B.: *Measuring the User Experience*, Morgan Kaufmann, 2008.
2. Flick U.: *Projektowanie badania jakościowego*, Wydawnictwo Naukowe PWN, Warszawa, 2010.
3. Sikorski M.: *Interakcja człowiek-komputer*, Wydawnictwo PJWSTK, Warszawa, 2010.
4. Kobyliński A.: *ISO/IEC 9126 – Analiza modelu jakości produktów programowych*, artykuł dostępny on-line pod adresem http://www.swo.ae.katowice.pl/_pdf/92.pdf
5. Krug S.: *Nie każ mi myśleć!*, Wyd. Helion, Gliwice, 2010.
6. Szulc M., Jastrzębowicz A.: *Optymalizacja witryny internetowej Banku Zachodniego WBK S.A. na podstawie testów użyteczności*, Inżynieria Oprogramowania w procesach Integracji Systemów Informatycznych, Górski J., Orłowski C. [red.], Wyd. PWNT, 2010.
7. Karwatka T.: *Usability w biznesie. Co kieruje Twoim klientem?*, Wyd. Helion, Gliwice, 2009.
8. Frontczak T.: *Marketing internetowy w wyszukiwarkach*, Wyd. Helion, Gliwice, 2006. (Użyteczność – tłumaczenie angielskiego słowa usability. Inne tłumaczenie, stosowane zamiennie, to funkcjonalność)
9. Polska edycja książki Jakoba Nielsena *Projektowanie funkcjonalnych serwisów internetowych* (tytuł oryginalny: *Designig Web Usability*), Wyd. Helion, Gliwice, 2003.
10. Chmielarz W.: *Analiza porównawcza usług bankowości elektronicznej w wybranych bankach w Polsce po okresie kryzysu 2009-2010*.
11. Silverman D.: *Porządzenie badań jakościowych*, Wydawnictwo Naukowe PWN, Warszawa, 2010.
12. Maison D.: *Jakościowe metody badań marketingowych. Jak zrozumieć konsumenta*, Wydawnictwo Naukowe PWN, Warszawa.

13. Denzin N., Lincoln Y.S.: Metody badań jakościowych, tom I, Wydawnictwo Naukowe PWN, Warszawa 2009, strona 13.
14. Smith P.A.: Towards a practical measure of hypertext usability, *Interacting with Computers*, Volume 8, Issue 4, December, 1996, strony 365-381.
15. Metryka na piątek, 25 luty 2011r., artykuł dostępny pod adresem <http://uxlabs.pl/category/metryka-na-piatek/>
16. Mozol A.: Gdy pojawi się problem – kurs usability cz. IX, 22 lipca 2011r. Artykuł dostępny on-line pod adresem: <http://www.symetryczna.pl/usability/gdy-pojawi-sie-problem-%E2%80%93kurs-usability-cz-ix/>
17. Strona internetowa firmy Noldus, producenta oprogramowania FaceReader: <http://www.noldus.com/human-behavior-research/products/facereader>.
18. Sauro J., Kindlund E.: A method to standardize usability metrics into a single score, artykuł dostępny on-line pod adresem: <http://www.measuringusability.com/papers/p482-sauro.pdf>
19. Sauro J., Kindlund E.: Making sense of usability metrics: usability and Six Sigma, artykuł dostępny on-line pod adresem: http://www.measuringusability.com/papers/sauro-kindlund_paper.pdf
20. Jakob Nielsen – hasło w Wikipedii. Artykuł dostępny on-line pod adresem: http://pl.wikipedia.org/wiki/Jakob_Nielsen#Heurystyki_Nielsena

Mgr inż. Joanna KWIATKOWSKA
Instytut Ekonometrii i Informatyki, Wydział Zarządzania
Politechnika Częstochowska
42-202 Częstochowa, ul. Dąbrowskiego 69
tel.: 668-806-068
e-mail: joan.kwiatkowska@gmail.com