

GROMADZENIE DANYCH O UŻYTKOWNIKACH NA POTRZEBY PERSONALIZACJI PORTALI INTERNETOWYCH

Celina M. Olszak, Kornelia Olszówka

Wprowadzenie

Liczba użytkowników portali internetowych stale rośnie, a ich wymagania w stosunku do tych rozwiązań są coraz większe. Oczekują oni indywidualnego traktowania oraz szybkiej i przyjaznej obsługi. W tej sytuacji ważnym wyzwaniem dla projektantów stała się personalizacja portali internetowych. Personalizacja jest mechanizmem, który pozwala na dostosowanie komputerowej interakcji do potrzeb i możliwości użytkownika. W stworzeniu spersonalizowanego portalu pomocne okazują się profile użytkownika.

Profil użytkownika reprezentuje potrzeby informacyjne użytkownika będącego w interakcji z systemem [Dani92]. Zawiera on wiedzę o preferencjach użytkownika, jego umiejętnościach, zdolnościach, celach, potrzebach oraz zainteresowaniach. Stworzenie takiego repozytorium wiedzy wymaga korzystania z różnych metod i technik gromadzenia informacji. Teoria i praktyka wypracowała wiele interesujących propozycji w tym zakresie. W niniejszym artykule dokonano ich syntetycznego przeglądu.

1. Metody zbierania informacji o użytkownikach

Personalizacja portali internetowych wymaga zbierania szczegółowych informacji o użytkowniku, dotyczących jego

zainteresowań, roli jaką pełni w organizacji, uprawnień decyzyjnych oraz dokonywanych zakupów. Przyjmuje się, że profil użytkownika powinien być tworzony przy minimalnym udziale człowieka. Najczęściej stosowanymi metodami przy budowie profili jest profilowanie jawne (explicit profile) oraz profilowanie niejawne (implicit profile).

Profilowanie jawne (profilowanie wprost) polega na uzyskiwaniu informacji bezpośrednio od użytkownika poprzez wypełnienie przez niego kwestionariusza (np. przy zakładaniu konta pocztowego, przy wszelkiego rodzaju ankietach, sondach). Zaletą tej techniki jest umożliwienie użytkownikowi samodzielnego określenia zawartości portalu. Natomiast wadą jest jej duża subiektywność. Można spotkać się także z niechęcią użytkowników do wypełniania dodatkowych formularzy.

Profilowanie niejawne (domniemane, nie wprost) polega na obserwowaniu zachowań i reakcji użytkownika. Na ich podstawie tworzy się schematy przeglądania stron internetowych oraz dokonywanych przez nich zakupów [Wojc99]. Dane uzyskane w ten sposób są zapisywane przede wszystkim w plikach logach, plikach cookies lub modułach ścieżek sesyjnych (*session tracking modules*). W profilowaniu niejawnym dostosowanie prezentowanej zawartości portalu jest przenoszone z użytkownika na serwer WWW (agenta informacyjnego). Głównym celem agenta informacyjnego jest ustalenie struktury profilu oraz procesu jego tworzenia i aktualizacji tak, aby jak najlepiej odwzorowywał on zainteresowania użytkownika [Staw07]. W tworzeniu profili zazwyczaj rekomenduje się łączenie obu metod.

2. Rodzaje informacji o użytkownikach wykorzystywane w personalizacji portali internetowych

Aby stworzyć zadawalający profil użytkownika należy zgromadzić cały wachlarz różnorodnych informacji o nim. Generalnie, można podzielić je na dwie grupy: typowe dane o użytkowniku oraz dane mówiące o sposobie interakcji użytkownika z portalem.

2.1 Dane o użytkowniku portali internetowych

Na dane o użytkowniku składają się następujące kategorie: dane demograficzne, wiedza i umiejętności użytkownika, zainteresowania i preferencje użytkownika oraz zamierzenia i cele użytkownika.

Dane demograficzne. W tej kategorii można wyróżnić dane: osobowe (imię, nazwisko, adres zamieszkania, telefon), geograficzne (kod pocztowy, kraj, miejscowość), klasyfikujące klienta (częstotliwość korzystania z usług i produktów), psychograficzne i socjologiczne (dane określające tryb życia) oraz dane na temat wieku, płci, wykształcenia, dochodu oraz przynależności kulturowej. W portalach internetowych wykorzystuje się zazwyczaj podstawowe dane demograficzne oraz dane dotyczące zakupów realizowanych przez użytkownika.

Wiedza oraz umiejętności użytkownika portalu. Do ważnych informacji na potrzeby personalizacji zalicza się wiedzę użytkowników na temat pojęć, związków między pojęciami, faktów oraz reguł dotyczących dziedziny, która jest przedmiotem prezentacji [KoKo01]. W spersonalizowanym portalu powinno się brać pod uwagę indywidualną wiedzę i umiejętności użytkownika, szczególnie w przypadku zapoznawania go z nowymi informacjami, produktami oraz usługami. Tak jak w przypadku innych systemów, w których stawia się na przyjazną

obsługę, tak i tutaj, sposób prezentacji informacji należy dopasowywać do możliwości poznawczych i percepcyjnych człowieka [Olsz00].

Preferencje oraz zainteresowania użytkowników. Z powodu zróżnicowanych zainteresowań użytkowników ta kategoria danych jest bardzo istotna w procesie dostosowywania portalu do użytkownika. Często zdarza się, że produkty oferowane w jednej grupie użytkowników mogą nie wywoływać zainteresowania u innych osób. Aby zapobiec takim sytuacjom należy wykorzystywać różne techniki segmentacji klientów, które pozwalają wyodrębnić podobne grupy z uwagi na określone atrybuty.

Zamierzenia i cele użytkowników. Spersonalizowany portal na podstawie analizy zachowań klientów np. ich przemieszczania się po stronach www, dokonanych zakupach powinien w miarę szybko ustalić cel i dalsze zamierzenia użytkownika. Może odesłać go do innych analogicznych miejsc w sieci lub też dalej go obserwować, celem wychwycenia pewnych prawidłowości w jego zachowaniu. Czasami może okazać się to bardzo trudne lub wręcz niemożliwe z uwagi na fakt, że nie wszyscy użytkownicy sieci www mają jasno sprecyzowane cele, kiedy przeglądają określone strony (często robią to zupełnie przypadkowo).

2.2 Dane o sposobie interakcji użytkownika z portalem

Interakcja użytkownika z portalem stanowi niezwykle cenne źródło informacji o nim.

Zachowanie użytkownika podczas przeglądania portalu. Jednym ze wskaźników, służącym do badania zachowania użytkownika jest określenie efektywnego czasu przeglądania przez niego danej strony

www. Wbrew pozorom nie jest to proste zadanie. Wynika to z faktu, że niezwykle trudno jest określić, czy w danym czasie użytkownik rzeczywiście cały czas był obecny przy komputerze. Zazwyczaj przyjmuje się, że informacja jest mało interesująca dla użytkownika, gdy czas przeglądania strony www jest niższy od przyjętego i wcześniej ustalonego progu. Łatwiejsze jest określenie czasu przeglądania multimediiów strumieniowych. Obejrzenie całej prezentacji jest zazwyczaj równoznaczne z jego dużym zainteresowaniem danym tematem. Fakt ciągłego klikania i przesuwania myszką na wybranej stronie www może również wskazywać na zainteresowanie danej osoby określonym obiektem [KoKo01].

Akcje wyborów użytkownika. Jest to sposób, w jaki użytkownicy mogą wchodzić w interakcję z portalem. Najczęściej akcja wyborów polega na wybieraniu odpowiedniego odnośnika lub elementu menu.

Zakupy i akcje związane z zakupami. Zakupy dokonane przez użytkowników w sklepach internetowych w sposób istotny wskazują na ich zainteresowania i pewne preferencje. Niektóre portale internetowe na podstawie dokonanych zakupów proponują użytkownikom podobne lub powiązane produkty. O pewnych preferencjach użytkowników świadczy już sam fakt wkładania produktów do wirtualnego koszyka (które niekoniecznie użytkownik kupi), rozwiązywanie quizów, zapytania o bardziej szczegółowe informacje na temat oferowanych produktów lub nowości [KoKo01].

Systemy oceniania. W niektórych portalach internetowych istnieje możliwość oceniania określonych artykułów, produktów, dokumentów itp. Należy zauważyć, że już samo sporządzenie przez użytkownika

takiej oceny może wskazywać na jego określone zainteresowania. Typowym sposobem oceniania jest skala binarna, która pozwala sklasyfikować obiekt jako „interesujący” lub „nieinteresujący”. Skala ocen może być także numeryczna (np. skala od 1 do 5) lub symboliczna (np. „bardzo dobry” aż do „słaby”). Mankamentem systemów oceniania jest ich spory subiektywizm, ponadto wielu użytkowników ignoruje takie metody [KoKo01].

Pozostałe akcje. Wiele cennych informacji o użytkowniku można uzyskać analizując np. jak często łączył się z daną stroną www, jak długo zatrzymywał się w danym miejscu oraz jakie inne sekwencje działań wykonywał.

Sprzęt i oprogramowanie. Użytkownicy portali internetowych są w posiadaniu różnorodnego sprzętu i oprogramowania, który ma istotny wpływ na sposób ładowania się i prezentację danego serwisu. Aby zapobiec ewentualnym zakłóceniom w działaniu portalu należy rejestrować m.in. takie dane jak: wersja przeglądarki (przeglądarki internetowe różnią się znacznie zakresem możliwości), dostępność uaktualnień, skrypty Javy (zastosowanie apletów Javy pozwala na interakcję ze stroną internetową na wyższym poziomie), technologia Flash (oglądanie pewnych prezentacji na stronie jest możliwe tylko w odtwarzaczu Macromedia Flash Player). Częstość pobierania przez użytkownika najnowszych wersji oprogramowania może świadczyć o jego sporych doświadczeniach i umiejętnościach np. w zakresie obsługi komputera.

2.3 Inne źródła danych

Wiele cennych danych o użytkownikach można wydobyć analizując pliki cookies, logi serwera, a także dane, które wprowadzili oni do systemu podczas rejestracji.

Pliki cookies jest to porcja informacji (w formacie pliku tekstowego) wysyłana przez serwer do przeglądarki użytkownika, która je przechowuje i udostępnia na polecenie serwera. Pliki te pozwalają na zidentyfikowanie użytkownika i śledzenie jego preferencji, co umożliwia optymalne projektowanie stron internetowych. Ułatwiają one nawigację i w znacznym stopniu zwiększają komfort użytkownika przy przeglądaniu danej strony www. Cookies pomagają również w identyfikacji szczególnie popularnych elementów oferowanych przez portal. Przykładami zastosowania cookies mogą być: powitanie użytkownika na stronie jego imieniem lub pseudonimem, zapamiętania stanu koszyka w sklepie internetowym (po to, aby klient mógł przerwać zakupy i wrócić do nich później), wyświetlanie reklam produktów, którymi użytkownik może być zainteresowany.

Logi odwiedzin (*Access logs*) są to pliki tworzone przez oprogramowanie zainstalowane na serwerze sieciowym obsługującym portal. W rejestrze odwiedzin zapisywane są szczegóły zachowań użytkowników. Każda linia dostępu zawiera informacje o czasie, poleceniu i nazwie żadanego przez użytkownika pliku. W rejestrze zapisywane są również informacje o użytkowniku, zebrane na podstawie zapytań skierowanych do wyszukiwarki, z której korzysta. Najistotniejszymi dla analizy użytkowników informacjami zawartymi w logu serwera jest nazwa lub adres IP komputera, z którego nastąpiło

odwołanie, data, czas oraz nazwa pliku zamawianego przez użytkownika [WoZa07], [Wojc99]. W większości rejestrów odwiedzin są zapisywane także informacje na temat typu przeglądarki użytkownika, nazwy domeny, z której się łączy oraz nazwy ostatniej strony, jaką odwiedził [Ziem05].

Duże znaczenie ma także rejestracja użytkowników. W tym przypadku dane zbierane są z wykorzystaniem interaktywnych formularzy. Najczęściej użytkownik musi podać swoje dane personalne, miejsce pracy, zainteresowania, preferencje odnośnie zakupów itp. Kluczowe znaczenie ma takie zaprojektowanie procesu rejestracji, aby jak najwięcej użytkowników podało swoje dane i aby były one prawdziwe.

Podsumowanie

Budowanie spersonalizowanych portali internetowych, czyli aplikacji odpowiadających indywidualnym potrzebom użytkowników jest dzisiaj koniecznością. Użytkownicy potrzebują nie tylko szybkiej obsługi, ale także indywidualnego traktowania, włączania do projektowania nowych produktów i usług. Budowa spersonalizowanych portali wymaga sięgnięcia do różnorodnych technologii: od prostych baz danych poprzez oprogramowania agentów do kolaboratywnego filtrowania. Aby owe technologie mogły być efektywnie wykorzystane, należy przede wszystkim zadbać o staranne, rzetelne i etyczne zbieranie danych o użytkowniku. Jest to zadanie niezwykle trudne, wymagające wielu badań i weryfikacji.

Literatura

- [Wojc99] Wojciechowski M., Odkrywanie wzorców zachowań użytkowników WWW, Materiały konferencyjne POLMAN'99, OWN, Poznań, 1999.
- [Staw07] Stawarz T. Teoretyczne podstawy adaptacyjnych stron, techniki odkrywania wiedzy stosowane do ich personalizacji oraz modułowa implementacja takich rozwiązań, http://datamining.home.pl/_pdfy/sta04teo.pdf.
- [KoKo01] Kobsa A, Koenemann J, Pohl W (2001), Personalized Hypermedia Presentation Techniques for Improving Online Customer Relationships, The Knowledge Engineering Review 16(2), 111-155.
- [Adam07] Adamczyk J., Zbawienna personalizacja, <http://www.emarketing.pl/artyk/artyk53.php>, 23.04.2007.
- [Olsz00] Olszak C. M., Zarys metodologii multimedialnych systemów wspomaganie decyzji w zarządzaniu, AE, Katowice 2000.
- [Ziem05] Ziemia E., Metodologia budowy serwisów internetowych dla zastosowań gospodarczych, AE, Katowice 2005.
- [WoZa07] Wojciechowski, M., Zakrzewicz, M., *Automatyczna personalizacja serwerów WWW z wykorzystaniem metod eksploracji danych*, <http://www.cs.put.poznan.pl/mwojciechowski/papers/szczyrk02b.pdf>, 23.04.2007
- [BMNS07] Berendt B., Mobasher B., Nakagawa, M., Spiliopoulou, M., The impact of site structure and user environment on session reconstruction in Web analysis,

<http://maya.cs.depaul.edu/~mobasher/papers/webkdd02.pdf>, 23.04.2007

- [Dani92] Daniłowicz C., Modele systemów wyszukiwania informacji uwzględniającego preferencje użytkowników końcowych, Prace Naukowe BGiONT nr. 6, monograf (3), Wrocław 1992.

Informacje o autorach

Prof. dr hab. inż. Celina M. Olszak
Mgr Kornelia Olszówka
Katedra Informatyki Ekonomicznej
Akademia Ekonomiczna
ul. Bogucicka 3
40-226 Katowice – Polska
Numer telefonu (fax) +48/32/2577610
e-mail: olszak@ae.katowice.pl
olszowka@ae.katowice.pl