

DAGMARA SAWICKA

Akademia Rolnicza w Lublinie

Biblioteka Główna

## **E-pytanie, e-odpowiedź... czyli jakich badań potrzebują biblioteki przyszłości?**

*Truizmem byłoby stwierdzenie, że od informacji elektronicznej nie ma już ucieczki. Z tym faktem nikt już nawet nie dyskutuje. Hybrydyzacja bibliotek niewątpliwie będzie postępować, a być może e-biblioteki w dużym stopniu wyprą biblioteki tradycyjne. W świetle takich założeń niezbędne stają się badania zachowań użytkowników e-informacji oraz badania reakcji użytkowników na poszczególne rodzaje informacji. Co badać i jakimi metodami? Referat podejmuje próbę odpowiedzi na te pytania oraz próbę implementacji istniejących już badań użytkowników Internetu do potrzeb biblioteki akademickiej.*

Powoli urzeczywistnia się wizja elektronicznego świata. Świata wirtualnych informacji, zakupów przez e-mail, przyjaciół znanych jedynie z komputerowego monitora. Przeprowadzony w tym roku w Wielkiej Brytanii sondaż wykazał, że tylko 5 % ludzi w wieku 15-24 lat do porozumiewania używa kartki i długopisu. Pozostali korzystają wyłącznie z poczty elektronicznej i sms-ów<sup>1</sup>. Jednocześnie eksperci szacują, że 90 % dokumentów na świecie powstaje jedynie w wersji cyfrowej, przy czym aż 70 % z nich nigdy nie zmieni postaci. Nie pozostaje zatem nic innego jak pogodzić się z coraz większą dominacją informacji elektronicznej. Nie czas już debatować, jak temu zapobiec, a całą energię lepiej skupić na rozwiązywaniu problemów związanych z cyfrowymi publikacjami. Kwestie przechowywania, katalogowania czy udostępniania to cały czas jeszcze problemy otwarte, w których pozostaje wiele do zrobienia.

### ***Mierzalny Internet***

Nierozwiązany pozostaje również problem badania użytkowników rozległej sieci. Teoretycznie wydawałoby się, że nic łatwiejszego niż „zmierzyć” Internet, wszak jest on medium bardzo łatwo mierzalnym, choćby dlatego, że istnieje możliwość śledzenia wyników w czasie rzeczywistym. Toteż żadne inne medium nie posiada tak szybko opracowanych metod badawczych oraz takiego stopnia mierzalności jak Internet. Po kilku latach od powstania dostępne były badania pozwalające określić jego potencjał reklamowy i komunikacyjny<sup>2</sup>.

<sup>1</sup> *Masz coś do ukrycia? Komputer cię zdradzi*, Gazeta Wyborcza z dnia 24 kwietnia 2006, s. 33

<sup>2</sup> D. Kaznowski: *Raport strategiczny IAB Polska Internet 2004 Polska, Europa i świat*. Warszawa [2005], Toż: *op. cit.* [online] [dostęp 08.06.2006]. Dostępny w Internecie: [http://www.e-biznes.pl/raporty/pliki/raport\\_roczny\\_IAB\\_2004.pdf](http://www.e-biznes.pl/raporty/pliki/raport_roczny_IAB_2004.pdf), s. 40

Już w momencie publikacji pierwszej strony internetowej dostępne były twarde dane o liczbie wizyt użytkowników oraz o tym, kiedy i co odwiedzający oglądają na danej witrynie<sup>3</sup>. Ze względu na możliwość szybkiego pozyskania statystyk, wkrótce potem (bardzo szybko, jeśli porównać do czasu, jakiego potrzebowały na to inne media) pojawiły się pierwsze wyniki badań profili społeczno-demograficznych internautów. Koszt uzyskania wyników z badań telemetrycznych telewizji czy badań dzienniczkowych słuchalności radia jest nieporównywalnie większy niż otrzymywanie danych liczbowych dotyczących Internetu.

Jednakże Internet badany jest niemal wyłącznie pod kątem efektywności kampanii reklamowych. Całkowicie pomijane są natomiast badania naukowej informacji elektronicznej. Nadal zatem nie wiadomo, jak podawać informację, aby była łatwo odszukana, czytelna, przyjazna i zrozumiała dla każdego użytkownika lub dla określonej grupy użytkowników. Proste przemieszczenie badań dotyczących rynku reklamy na sferę informacji nie jest wykonalne. Poza słabością merytoryczną takiego rozwiązania, mogło by to mieć negatywne skutki dla instytucji zajmujących się prezentacją lub udostępnianiem informacji, a także dla użytkownika informacji. Trudno bowiem bezkrytycznie przejąć metody z innej dziedziny wiedzy, bo jeśli nawet potraktujemy informację jako towar, to jest to towar specjalnego rodzaju i stosowanie do niego metod czysto marketingowych może mieć fatalne skutki.

Całkowite odrzucenie dorobku badań Internetu spowodowałoby natomiast duże przesunięcia czasowe ze względu na konieczność stworzenia nowej metodologii. Jej skuteczność można by ocenić dopiero po kilku latach stosowania, metodą prób i błędów. Konieczne jest zatem rozwiązanie kompromisowe, będące kompilacją badań Internetu – jeśli chodzi o techniki, i dotychczasowych badań użytkowników bibliotek – w pozostałych treściach.

### ***Badania zachowań użytkowników informacji***

Technologie sprzyjają badaniu informacji, bowiem każde wejście na stronę czasopisma (bądź inną stronę informacyjną) jest liczone i mierzone pod różnymi kątami. Są to działania w pełni zautomatyzowane i dostępne najczęściej wraz z pakietem usług informacyjnych. Każdy wydawca bazy danych czy zbioru elektronicznych czasopism lub książek, zaopatruje swój produkt w mechanizmy pozwalające zmierzyć liczbę użytkowników serwisu, logowań, wejść do konkretnych czasopism, artykułów itd.

Taka analiza wykorzystania e-czasopism jest niezwykle ważna, choćby z tego powodu, że od lat większość budżetu bibliotek akademickich przeznaczana jest na wydawnictwa ciągle, które coraz częściej mają postać „e-only”. Na rynku wydawnictw naukowych obserwuje się również rosnącą tendencję do wydawania książek w wersji elektronicznej. Jednak oprócz prostego zliczenia czytelników potrzebne są pewne mechanizmy, które pozwolą zinterpretować otrzymane liczby.

H. R. Jamala<sup>4</sup> wraz ze swoimi współpracownikami, zebrał na podstawie ocen różnych autorów zajmujących się tematyką informacyjną, wady i ograniczenia badań użytkowników informacji opierających się jedynie na podstawie logowań.

<sup>3</sup> D. Kaznowski: *op.cit.*

<sup>4</sup> H. R. Jamali, D. Nicholas, P. Huntington: *The use and users of scholarly e-journals: a review of log analysis studies*, Aslib Proceedings 2005 Vol. 57 nr 6, s. 554-571

## ***Zalety analizy zachowań użytkowników na podstawie liczby logowań:***

- Ilościowe dane są zbierane automatyczne i nie podlegają żadnemu filtrowaniu. Brak ingerencji człowieka w proces ich gromadzenia gwarantuje ich obiektywność;
- Suche liczby dostarczają badaczowi bezpośrednich informacji o tym, ile milionów użytkowników weszło do systemu, natomiast nie sugerują, w jakim celu tam weszli, co mówili i myśleli logując się i co w między czasie robili. W odróżnieniu od innych badań, które można dowolnie interpretować, bądź zmieniać dane, w logowaniu nie ma możliwości wpływania na wyniki. Nie ma problemu ze stroniczymi odpowiedziami lub ich brakiem, tak jak to bywa w przypadku ankiet;
- Jeśli połączy się badania „logów” z bezpośrednim wywiadem (np. w postaci ankiety) to otrzyma się efektywne narzędzie do wykrywania rozbieżności pomiędzy tym, co użytkownicy mówią o systemie, a tym, co robią korzystając z systemu. Daje to możliwość weryfikacji odpowiedzi i ustalenia, jaki procent odpowiedzi jest „kliknięciem” w przypadkowe miejsce na formularzu. Analiza logowań stanowi również doskonałą bazę do budowania ankiety;
- Analiza logowań może być wykorzystywana do testowania hipotez. Sprawdzając na przykład częstość logowań do różnych baz z poszczególnych numerów IP, łatwo można sprawdzić, czy umiejscowienie lub konfiguracja komputerów w bibliotece jest dla użytkowników optymalna;
- Analiza logowań jest dobrym narzędziem do obiektywnej oceny efektywności takich systemów jak np. bazy czasopism. Bardzo łatwo bowiem można wyliczyć z kosztu poniesionego na zakup bazy, koszt jednego czasopisma, artykułu itp.;
- Analiza logowań jest również odpowiednią metodą do gromadzenia i porównywania informacji na temat zachowań użytkowników z poszczególnych grup. Daje badaczowi obiektywne informacje na temat czasu pobytu użytkownika na stronie, rodzaju materiału z jakiego korzystał, rodzaju wykorzystywanej nawigacji itp.

## ***Ograniczenia przydatności analizy logowań do badania użytkowników informacji:***

- Najważniejszym ograniczeniem, jeśli chodzi o wykorzystanie danych liczbowych do ustalenia profili użytkowników, jest trudność odróżnienia, które wyniki są wynikami systemu, a które wynikami internauty. Np. mierzony jest całkowity czas pobytu w systemie, nie wiadomo natomiast, ile z tego czasu było poświęcone na szukanie, ile na odpowiedź systemu, a ile na „czas do namysłu” użytkownika. Równie dobrze użytkownik mógł zalogować się w systemie i pójść zaparzyć kawę;
- Trudności z identyfikacją użytkownika. Użytkownik może korzystać z kilku różnych terminali, lub z jednego terminalu może korzystać dwóch użytkowników podczas jednego logowania. Jest to szczególnie problem terminali udostępnianych publicznie, zlokalizowanych np. w bibliotekach lub salach komputerowych na uczelniach.

Problemy z identyfikacją użytkownika mogą wpływać również z wielu innych przyczyn:

- Jedną z podstawowych będzie masowość korzystania z systemu. Ilość osób korzystających z otwartych systemów jest bowiem z reguły ogromna, zaś wielkość populacji użytkowników prowadzi do trudności z identyfikacją indywidualnych użytkowników lub wyodrębniania grup użytkowników;
  - Istnieje trudność z identyfikacją użytkowników na podstawie IP komputera. Wielu dostawców Internetu przydziela dynamiczne adresy IP. Powoduje to, że jeden użytkownik może korzystać z dwóch różnych numerów IP podczas dwóch kolejnych sesji (nawet jeśli sesje nastąpiły jedna po drugiej). Może też się zdarzyć tak, że w tym samym czasie korzysta z systemu dwóch użytkowników o jednakowych numerach IP. Trudno nawet umiejscowić geograficznie użytkowników, ponieważ adres IP zależy nie od miejsca, w którym znajduje się użytkownik, a od miejsca gdzie dane IP jest zarejestrowane;
  - Niemożliwe jest określenie faktycznej liczby sesji. Sesje bardzo często zamykają się po określonym czasie. Nie wiadomo, czy użytkownik zalogował się i cały czas z systemu korzystał, czy też zalogował się, po czym zapomniał o rozpoczętej sesji. Zaburza to znacznie wyniki badań. W celu jak największego ograniczenia ryzyka takiego zniekształcania danych, Nicholas proponuje, aby zmniejszać czas wylogowania z systemu w przypadku braku aktywności poniżej 30 minut<sup>5</sup>;
  - W błąd mogą wprowadzać badacza również serwery proxy. Np. większość firm tak konfiguruje swoje strony, żeby po ich obejrzeniu umieszczane były one w plikach tymczasowych serwera. Jeśli więc jeden użytkownik odwiedził daną stronę, a za jakiś czas wejdzie na nią po raz kolejny, bądź też inny użytkownik z niej skorzysta (nawet z innego komputera), wówczas strona jest wyświetlana z plików tymczasowych, bez logowania po raz kolejny. Zatem ruch na takiej stronie może być systematycznie zaniżany;
  - Kolejną barierą w identyfikacji użytkowników są firewalle. Jest to kombinacja rozwiązań hardwarowych i softwarowych, które blokują niektóre pliki uznawane za potencjalnie szkodliwe. Mogą one ograniczać wyświetlanie niektórych stron lub podstron. Mogą również blokować zawartość niektórych serwisów;
  - Tzw. pająki, roboty czy crowlery będące automatycznymi przeszukiwarkami, które szperają w Internecie i bazach danych w różnych celach, np. aby indeksować dane strony w wyszukiwarkach. Logują się one tak, jak każdy użytkownik. Zatem system zlicza odwiedziny takiego robota, jako odwiedziny użytkownika. Faktycznie jednak informacja z serwisu nie została przeczytana.
- Badacza, który próbuje stworzyć obraz potencjalnego użytkownika systemu ogranicza również niewiedza o tym, dlaczego użytkownik robi to, co robi. Nie da się w związku z tym określić interakcji pomiędzy systemem a użytkownikiem. „Log” użytkownika nic badaczowi nie mówi o samym użytkowniku. W jakim jest wieku, kim

<sup>5</sup> D. Nicholas [et. al.]: *Cracking the code: web log analysis*, Online & CD-ROM Review 1999 Vol. 23 nr 5, s. 263-265

jest, czego oczekuje od serwisu i czy został usatysfakcjonowany otrzymanym wynikiem. Haigh i Megarity zauważają, że same „logi” nie dają wiedzy na temat powodu zadania takiego, a nie innego pytania przez użytkownika, czy też powodu dla którego zjawił się on w danym serwisie<sup>6</sup>. Nie znamy jego reakcji na zawartość. Jedynie zaawansowane systemy analizy logowań tzw. DLA<sup>7</sup> (deep log analysis) są w stanie wyodrębnić powtórne wizyty. Powtórna wizyta zaś może być pośrednio miernikiem satysfakcji użytkownika z serwisu, jako że nieusatysfakcjonowany użytkownik rzadko powraca w to samo miejsce.

Jak wynika z powyższych stwierdzeń, na podstawie danych, które otrzymujemy ze statystyk użytkowania serwisów, może być trudne zbudowanie jednego wzorca zachowań użytkowników informacji. Zachowania te mogą się bowiem różnić (a nawet na pewno różnić się będą) w zależności od:

- użytkownika (do jakiej grupy użytkownik należy – np. grupy w zależności od wykształcenia czy znajomości języków obcych);
- rodzaju informacji (naukowa/popularno-naukowa, tekstowa/graficzna, itp.);
- rodzaju platformy, na której informacja jest udostępniana (np. CD-ROM, sieć rozległa, sieć lokalna);
- w zależności od sposobu udostępniania informacji (przez bazę danych, jako oddzielne czasopismo, poprzez system typu Onelog itp.).

Badania typu site-centric, do których należą analizy logowań, nigdy więc nie dadzą pełnego obrazu użytkownika danego serwisu. Można postawić tezę, że dopiero analiza zachowań poszczególnych grup użytkowników konkretnego rodzaju informacji na poszczególnych platformach mogłaby pozwolić na stworzenie ogólniejszego schematu zachowań internautów. Mimo więc wielu zalet tego typu badań, aby stworzyć rzetelny wzór zachowań użytkowników informacji, potrzeba czegoś jeszcze. Do suchych liczb należy dołożyć informacje pozyskane np. z ankiet wypełnianych przez użytkowników lub częściowo implementować metody badań Internetu.

Uzupełnienie badań ilościowych badaniem ankietowym umożliwia uzyskania informacji, w jaki sposób poprawić funkcjonalność witryny oraz zadowolenie internautów korzystających z serwisu. Badanie można zrealizować na dowolnej grupie użytkowników, zadając im pytania m.in. o to, jakich treści brakuje w serwisie, czy jest coś, czym badana witryna wyróżnia się na tle innych, z jakich powodów korzystają z zasobów witryny i jakich informacji przede wszystkim tam poszukują. Daje również możliwość uzupełnienia informacji społeczno-demograficznych.

<sup>6</sup> S. Haigh, J. Megarity: *Measuring web site usage: log file analysis*, Network Notes [online] 1998 nr 57 [dostęp 08.06.2006]. Dostępny w Internecie: <http://www.collectioncanada.ca/9/1/p1-256-e.html>

<sup>7</sup> D. Nicholas, P. Huntington, A. Watkinson: *Scholarly journal usage: the results of deep log analysis*, Journal of Documentation 2005 Vol. 61 nr 2, s. 248-280

## **Badanie funkcjonalności serwisu**

Od pewnego już czasu prowadzi się badania funkcjonalności witryn internetowych. Jest to jedno z trudniejszych badań, jakie realizowane jest w Internecie, ale daje wymierne efekty i dlatego warto skorzystać ze sprawdzonych wzorców w przypadku badania informacji naukowej.

Przez wiele lat sieci hipermarketów pracowały nad właściwym układem towaru na półkach, półek w regałach i regałów w sklepie, tak by zoptymalizować czas, jaki klient musi poświęcić na spacer po sklepie. Badanie funkcjonalności stron WWW przypomina nieco poszukiwanie odpowiedniego ustawienia w sklepie. Celem twórcy witryny jest taka prezentacja treści, aby użytkownik miał szansę zapoznać się z jak największą częścią oferty (np. informacji) i złożyć zamówienie lub zapytanie. Dlatego poszukuje się stale sposobów na optymalne eksponowanie zawartości serwisów. Istnieje wiele narzędzi, które wykorzystywane są do tego celu: od opinii eksperta, poprzez badania fokusowe, aż do badań typu eye-tracking. Wybór metod badawczych uzależniony jest przede wszystkim od celów, jakie chcemy osiągnąć<sup>8</sup>.

Bardzo użyteczne do badania użytkowników informacji naukowej okazałyby się badania typu eye-tracking. Jest to doświadczenie, w którym wykorzystywana jest specjalistyczna aparatura analizująca zachowania i ruchy gałki ocznej oraz sposób poruszania się kursora myszki. Zestawiając te dwie wartości w korelacji z czasem i strukturą badanej witryny, uzyskuje się informację o elementach, które skupiają uwagę lub dekoncentrują internautów. Powielając tego rodzaju badanie, zdobywane są dane, które mówią o największych błędach popełnionych na etapie projektowaniu serwisu bądź o największych zaletach witryny. Jak nietrudno się domyślić, ograniczeniem w realizacji tego typu badań, jest wysoka cena ich przeprowadzenia.

Najbardziej dostępne okazałyby się więc dla bibliotek badania fokusowe. Polegają one na moderowanej dyskusji, w której uczestniczy starannie dobrana grupa użytkowników serwisu. Moderator tak kieruje dyskusją, aby uzyskać od respondentów jak najwięcej informacji. W przypadku prowadzenia takich badań online, jest szansa na uzyskanie prawdziwych opinii użytkowników (ze względu na anonimowość i większą z reguły otwartość osób komunikujących się w sieci).

## **Czas nagli**

W ostatnich latach coraz wyraźniej widać konieczność przygotowania jednolitej i wiarygodnej metodologii do analizy i oceny zachowań w Internecie. Dyrektorzy bibliotek skrzętnie śledzą, przysyłane przez wydawców baz i czasopism online, statystyki użytkowania serwisów, ale tak naprawdę nie do końca wiedzą do czego je odnosić. Czy przeliczać na cenę jednego artykułu, czy porównywać z ceną prenumeraty drukowanej, czy może zadowolić się po prostu odpowiednią liczbą zer w polu „liczba logowań do serwisu”?

Na te pytania nie ma w tej chwili jednoznacznej odpowiedzi. Aby otrzymać pełny profil internauty, wspomniane tutaj badania statystyk użytkowania, trzeba koniecznie

<sup>8</sup> D. Kazanowski: *op. cit.*

uzupełnić o takie informacje jak: płeć, wiek, dochód czy wykształcenie czytelnika. Dopiero takie uzupełnione dane będą przydatne do stworzenia wiarygodnego wzoru zachowań użytkownika informacji. A stworzenie takiego wzoru jest z pewnością jednym z wyzwań, jakie stoją przed bibliotekami XXI wieku, niezależnie od tego, w jakim kształcie one przetrwają.

## **Bibliografia**

1. Haigh S., Megarity J.: *Measuring web site usage: log file analysis*, Network Notes 1998 nr 57 [dostęp 08.06.2006]. Dostępny w Internecie: <http://www.collectioncanada.ca/9/1/p1-256-e.html>
2. Huntington P., [et. al]: *Obtaining subject data from logfiles using deep log analysis: case study OhioLINK*, Journal of Information Science [online] 2005 [dostęp 08.06.2006]. Dostępny w Internecie: [http://www.publishing.ucl.ac.uk/2005Huntington\\_etal.pdf](http://www.publishing.ucl.ac.uk/2005Huntington_etal.pdf)
3. Jamali H. R., Nicholas D., Huntington P.: *The use and users of scholarly e-journals: a review of log analysis studies*, Aslib Proceedings 2005 Vol. 57 nr 6, s. 554-571
4. Kaznowski D.: *Raport strategiczny IAB Polska Internet 2004 Polska, Europa i świat*. Warszawa: AdPress Wydaw. Rekl. Sp. z o.o., [2005]. Toż: Kaznowski D.: op. cit., [online] [dostęp 08.06.2006]. Dostępny w Internecie: [http://www.e-biznes.pl/raporty/pliki/raport\\_roczny\\_IAB\\_2004.pdf](http://www.e-biznes.pl/raporty/pliki/raport_roczny_IAB_2004.pdf)
5. Kwarciak B.: *Cyberpsychologia sprzedaży, czyli klient bez zasłony*, Modern Marketing 2000 nr 5
6. *Masz coś do ukrycia? Komputer cię zdradzi*, Gazeta Wyborcza z dnia 24 kwietnia 2006, s. 33
7. Meller A.: *Jak Brytyjczycy badają Internet?* [online] [dostęp 08.06.2006]. Dostępny w Internecie: <http://www.internetstandard.pl/news/66366.html>
8. Meller A.: *Powrót do e-marzeń*, CEO: Magazyn kadry zarządzającej 2005 nr 11
9. Nicholas D., [et. al]: *Cracking the code: web log analysis*, Online & CD-ROM Review 1999 Vol. 23 nr 5, s. 263-265
10. Nicholas D., Huntington P., Watkinson A.: *Scholarly journal usage: the results of deep log analysis*, Journal of Documentation 2005 Vol. 61 nr 2, s. 248-280
11. *Raport IAB Polska. Polski rynek internetowy po dwóch kwartałach 2004 r.* [online] [dostęp 08.06.2006]. Dostępny w Internecie: [http://www.iab.com.pl/\\_files/\\_prezentacje/raport\\_polroczny\\_iab\\_2004.pdf](http://www.iab.com.pl/_files/_prezentacje/raport_polroczny_iab_2004.pdf)

## **E-question, e-answer... what research libraries of the future need?**

*It would be truism to claim that it is impossible to do without electronic information nowadays. This fact is beyond question. The hybridization of libraries will undoubtedly progress and it is likely that e-libraries will squeeze the traditional ones out of the market.*

*In the light of what has been assumed the research on the behaviour and the reactions of e-information users to particular types of information appears highly indispensable. What shall we do research on and what methods shall we implement?*

*This paper attempts to answer these questions as well as to apply already existing research on the Internet users for the use of the university library.*