



Łukasz Lipiński

Cloud Technologies

lukasz.lipinski@cortb.pl

Michał Bernardelli

Szkoła Główna Handlowa w Warszawie

mbernard@sgh.waw.pl

Anonimowość w Internecie – identyfikacja płci użytkowników na podstawie historii odwiedzanych stron internetowych

Wielkość ruchu internetowego, zaawansowane technologicznie narzędzia wykorzystywane na co dzień przez współczesnych ludzi oraz zwykła niewiedza przeciętnych użytkowników Internetu, często sprawiają złudne wrażenie anonimowości. Tymczasem możliwości śledzenia użytkowników, ich zachowań, lokalizacji, czy preferencji są – nawet w granicach obowiązującego prawa – bardzo rozwinięte. Mając odpowiednie narzędzia lub uprawnienia, możliwa jest nawet dokładna identyfikacja użytkownika. Z punktu widzenia biznesowego istotne jest uzyskanie jak najbardziej szczegółowych informacji o użytkowniku. Pozwala to na lepiej spersonalizowaną komunikację, co z kolei zwiększa szansę np. na skuteczniejsze wykorzystanie skłonności zakupowych użytkownika. Do podstawowych charakterystyk opisujących klienta należą z pewnością takie cechy jak płeć oraz wiek.

Celem badania było przedstawienie algorytmu identyfikacji płci polskich użytkowników Internetu na podstawie historii odwiedzanych przez nich witryn internetowych. Proponowana metoda wykorzystuje dane z profili użytkowników zawierające adresy stron internetowych oraz częstości ich odwiedzin przez danego użytkownika. Do opracowania algorytmu wykorzystane zostały dane dziesiątków tysięcy



Polaków oraz milionów odwiedzin stron internetowych z roku 2017. Skuteczność poprawnej identyfikacji płci z użyciem proponowanej metody, oscyluje w granicach 80%, bez uwzględniania dodatkowych danych o użytkowniku, np. używanego sprzętu, systemu operacyjnego, przeglądarki, godzin korzystania z Internetu, itp.

Skuteczność działania, prostej w swojej koncepcji metody, po pierwsze skłania do refleksji na temat anonimowości w Internecie, a po drugie otwiera duże możliwości personalizacji, która może być wykorzystana np. w marketingu internetowym czy rynku e-commerce. Dodatkową wartością dodaną artykułu jest przedstawienie niektórych informacji na temat preferencji najchętniej odwiedzanych stron przez Polaków.