



► **Monika Jabłońska** – absolwentka Wydziału Nauk Politycznych i Dziennikarstwa Uniwersytetu im. Adama Mickiewicza w Poznaniu na kierunku dziennikarstwo i komunikacja społeczna w specjalności reklama i promocja oraz nowe media. Zainteresowania badawcze skupione wokół procesów związanych z konwergencją mediów, komunikacją medialną i wizualną, kompetencjami medialnymi użytkowników sieci oraz wokół szeroko pojmowanych nowych mediów.

## Personalizacja internetu – zagrożenie czy naturalny proces rozwoju sieci?

**Monika Jabłońska**

[jablonska.monika@amu.edu.pl](mailto:jablonska.monika@amu.edu.pl)

UNIwersytet im. Adama Mickiewicza w Poznaniu

### ABSTRAKT

Przedmiotem artykułu jest ukazanie problematyki personalizacji internetu oraz zwrócenie uwagi na korzyści i zagrożenia, które sieć ze sobą niesie. Główny akcent został położony na wyjaśnienie procesów związanych z dostosowywaniem zawartości internetu do indywidualnych potrzeb jego użytkowników. W tym celu podjęłam próbę odpowiedzi na pytania: czy termin „globalna wioska” stanowi aktualną metaforę komunikacji w internecie? Co jest źródłem personalizacji internetu, jaką pełni rolę oraz jakie zagrożenia ze sobą niesie? Czy możemy uniknąć zjawiska personalizacji?

W artykule poddano analizie dane zawarte w raportach StatCounter Global Stats, a także przeprowadzono własne badania empiryczne w zakresie funkcjonowania użytkownika w przestrzeni wirtualnej. Znacząca część artykułu oparta jest na wiedzy oraz kilkuletnim doświadczeniu własnym w zakresie optymalizacji stron internetowych pod kątem SEO (*search engine optimization*) oraz marketingu internetowego – SEM (*search engine marketing*).

**SŁOWA KLUCZOWE: PERSONALIZACJA INTERNETU, DOSTOSOWYWANIE WYNIKÓW WYSZUKIWANIA, BEZPIECZEŃSTWO W SIECI, ALGORYTMY GOOGLE, FILTROWANIE ZAWARTOŚCI SIECI**

### ABSTRACT

The subject of the article shows the problem of personalization of the Internet and it concerns the benefits and threats that it poses. The main emphasis has been put on explaining the process related to adaptation of the Internet content to individual needs of network users. The author has attempted to answer the questions: Is the term of the “global village” a current metaphor of the Internet communication? What is the source of personalization of the Internet, what role does it play and what threats does it pose? Moreover, can we avoid the phenomenon of personalization?

In the article, data contained in StatCounter Global Stats reports have been analyzed, as well as own empirical research in the field of user functioning in the virtual space has been carried out. A significant part of the article is based on author’s knowledge and several years of experience in the field of SEO (*search engine optimization*) and the Internet marketing – SEM (*search engine marketing*).

**KEY WORDS: THE INTERNET PERSONALIZATION, CUSTOMIZING SEARCH RESULTS, NETWORK SECURITY, GOOGLE ALGORITHMS, FILTERING THE CONTENT OF THE NETWORK**

W pierwszych latach działania internetu w Polsce bariera dostępu była na tyle duża, że mogli z niego korzystać wyłącznie nieliczni. Wraz z rozwojem technologii grono użytkowników poszerzało się, a nowe medium stawało się z roku na rok coraz bardziej popularne. Szacuje się, że pierwszy milion użytkowników osiągnięty został w 1997 roku (Baran, 2013, s. 77). Wówczas Radosław Katarzyniak, Przemysław Kazienko oraz Janusz Sobeczki w publikacji zatytułowanej *Wybrane problemy wyszukiwania informacji w sieci Internet*, podchodząc z pewną dozą ostrożności do ocen poszczególnych procesów zachodzących w internecie, jak również samego medium, określili go „abstrakcyjnym obszarem pojęciowym charakteryzującym się niezwykle dużą złożonością oraz dynamiką” (Katarzyniak, Kazienko, Sobeczki, 1997).

Dziś, patrząc przez pryzmat zmian, które zaszły w internecie, można stwierdzić, że był to rewolucyjny wynalazek, który wpłynął przede wszystkim na zmianę funkcjonowania społeczeństwa, nie tylko w środowisku wirtualnym. Walery Pisarek postawił tezę, że internet jest medium, które „jak żadne inne w historii ma znaczący wpływ na większość sfer życia człowieka” (Pisarek, 2009, s. 7). Trudno się z nim nie zgodzić, zwłaszcza w kontekście komunikowania. Możliwości, które dał internet, pobudziły w jego użytkownikach poczucie otwartości i bliskości całego świata. Stworzona w latach 60. XX wieku przez Marshalla McLuhana teoria „globalnej wioski”, dotycząca telegrafu, na nowo pojawiła się w dyskursie naukowym, stanowiąc metaforę dla braku granic terytorialnych w kontekście używania mediów elektronicznych (McLuhan, 2001, s. 179). Warto jednak zastanowić się, czy ewolucja internetu i rozmaite procesy, które wpłynęły na jego funkcjonowanie, nie spowodują, że w XXI wieku wspomniana metafora będzie się dezaktualizować. Kluczowe zdaje się tu zestawienie kategorii technicznych możliwości z realnymi ograniczeniami, mające bezpośredni wpływ na komunikację masową w internecie. Z technicznego punktu widzenia internet umożliwia niczym nieskrępowany dostęp do informacji i masową komunikację na skalę globalną, jednak zwrócenie uwagi w kierunku ograniczeń, które są narzucane użytkownikom sieci, unaocznia, że w wielu sytuacjach status „globalnej wioski” traci rangę, przekształcając się co najwyżej w „wioskę lokalną”.

Barieri, które powstają na drodze swobodnego dostępu do informacji w przestrzeni wirtualnej, mają wielowymiarowy charakter, a u ich podstaw leżą zarówno czynniki techniczne warunkowane przez rozmaite przedsiębiorstwa z branży informatycznej, jak i behawioralne inicjowane przez samych użytkowników sieci. Skutkiem tego ci drudzy wybierają na co dzień

takie narzędzia, które w najbardziej optymalny sposób dostosowane są do ich potrzeb. Statystyki prowadzone przez StatCounter Global Stats wskazują, że dla ponad 98% polskich internautów słowo „Google” stanowi synonim wyszukiwania (StatCounter, 2018). Stale udoskonalane algorytmy pozwalają na szybkie i trafne zwracanie użytkownikom pożądaných wyników, co z kolei jest warunkiem koniecznym do utrzymania wiodącej pozycji. Przodownicy innowacji na przestrzeni lat wypracowali sprawną strukturę i stworzyli setki algorytmów, które pozwalają na natychmiastową personalizację wyników wyszukiwania jednostkowego użytkownika, co jeszcze bardziej umocniło ich pozycję na rynku.

## PERSONALIZACJA INTERNETU

„Personalizacja” to słowo, które w XXI wieku nabrało szczególnego znaczenia. Potrzeba nadawania produktom i usługom osobistego charakteru stała się tak silna i tak powszechna, że z trudem dziś można wyodrębnić sektory, których to zjawisko nie dotyczy. Jednostki, które zwykle są postrzegane jako element ogółu społeczeństwa, z chęcią doświadczają zindywidualizowanych form zaspokajania ich potrzeb. W przeszłości personalizacja zwrócona była wyłącznie w stronę elit, utwierdzając ich tym samym w poczuciu wyjątkowości. Na przełomie XX i XXI wieku dzięki rozwijającemu się postępowi technologicznemu proces personalizacji ewoluował, nabywając przy tym zdecydowanie bardziej inkluzywnego charakteru. Towarzyszył temu swoisty paradoks, który miał wpływ na wypracowanie nowej jakości – rozpoczęła się era wdrażania modelu biznesowego opartego na zasadzie indywidualizmu, skierowanego do klienta masowego. Zastosowanie tegoż modelu umożliwiło oferowanie szerokiej grupie odbiorców rozmaitych produktów i usług bezpośrednio dostosowanych do indywidualnych potrzeb każdej jednostki. W krótkim czasie praktyka ta zaczęła być stosowana w wielu branżach na całym świecie. Obecnie przedsiębiorstwa umożliwiają na przykład dowolne zestawienie kanałów tematycznych w telewizji cyfrowej, zakup biżuterii z osobistym akcentem w postaci graweru bądź własnej kompozycji elementów czy nawet, wydawać by się mogło, tak prozaiczną usługę, jaką jest podanie ulubionej kawy na wynos w kubku opatrzonym imieniem zamawiającego. Obserwując dynamikę postępu działań związanych z personalizacją, należy także zwrócić uwagę na rolę, jaką odgrywa ona w sektorach IT oraz E-commerce, gdzie procedura dostosowywania do jednostkowego konsumenta nie dotyczy

wyłącznie usług czy produktów, ale całej struktury funkcjonowania rozmaitych narzędzi, dzięki którym użytkownik odnosi wrażenie, że zostały one stworzone wyłącznie z myślą o nim. Początki personalizacji internetu sięgają lat 90. XX wieku, kiedy to twórcy nieistniejącej już wyszukiwarki AltaVista<sup>1</sup> stworzyli algorytm, który pozwalał na wprowadzanie słów kluczowych w wielu językach obcych (włącznie z tymi, które nie były zapisywane w alfabecie łacińskim), a także w inteligentny sposób potrafił nie uwzględniać we frazach wielowyrazowych spójników bądź wyrażeń nieistotnych dla danego wyszukiwania (Battelle, 2005, s. 45). Na przykład: wpisując pytanie „Where is London?” eliminował wyrazy „where” oraz „is”, dzięki czemu w naturalny sposób zawężył obszar poszukiwań co najmniej o kilka milionów witryn, w których z pewnością by się pojawiły (Seymour, 2011, s. 51).

Dziś personalizacja internetu to coś znacznie więcej. Należy o niej mówić w kontekście podejmowania wszelkich zabiegów związanych z dostosowywaniem zwracanej przez wyszukiwarki zawartości do indywidualnych potrzeb każdego użytkownika na podstawie podjętych przez niego w przeszłości działań w przestrzeni wirtualnej. Personalizacja oznacza odejście w niepamięć czasów, kiedy internet był jednakowy dla wszystkich, gdy dwoje użytkowników wpisujących w wyszukiwarkach te same frazy otrzymywało identyczne wyniki. Okazuje się jednak, że coś, co miało ulepszyć sposób pozyskiwania informacji, jednocześnie ograniczyło do nich dostęp. Każdy z nas, przemierzając wirtualne szlaki, buduje własną, jedyłą i niepowtarzalną strukturę informacji, którą wyszukiwarki zapisują, a pochodzące z nich dane przetwarzają na miliony sposobów. Eli Pariser w swojej książce *The Filter Bubble: What the Internet Is Hiding from You* nazywa tę strukturę niewidzialną „filtrującą bańką”. Jest to swego rodzaju zamknięta przestrzeń, do której docierają już przefiltrowane informacje, dedykowane indywidualnym użytkownikom. Co więcej, działania, które internauci podejmują w sieci, znacząco zawężają przestrzeń informacji, które do nich docierają, przy jednoczesnym nieświadomym udostępnianiu własnych danych. Z perspektywy konsumenta sieci możemy mówić o dychotomicznym ujęciu doświadczania personalizacji internetu: na płaszczyźnie intencjonalnej i nieintencjonalnej (Skwarski, 2004, s. 4). Obie te płaszczyzny obejmują swoim zasięgiem wszystkich użytkowników sieci i nieustannie oddziałują na nich w jednakowym czasie, przynosząc jednak

<sup>1</sup> AltaVista była jedną z najpopularniejszych wyszukiwarek na świecie, zanim pojawiła się wyszukiwarka Google. Pobrane z: ICSTI Insights, [http://www.icsti.org/IMG/pdf/INSIGHT\\_2010\\_JULY.pdf](http://www.icsti.org/IMG/pdf/INSIGHT_2010_JULY.pdf) (07.01.2019).

różne skutki. Różnice podejmowanych w ramach ich funkcjonowania działań związane są z kompetencjami konsumentów, potrzebami oraz intencjami. Świadome dopasowywanie parametrów witryn będzie odbywało się na przykład za pośrednictwem bezpośredniego wskazywania wyszukiwarce ulubionych stron, akceptowania zasad polityki cookies<sup>2</sup>, zapisów do newsletterów, wyrażania zgód na przesyłanie i wyświetlanie rozmaitych komunikatów czy aprobowania wykorzystywania danych osobowych przez administratorów stron podanych w trakcie wypełniania formularzy i po zalogowaniu do wybranych serwisów. W ramach podejmowania kolejnych kroków w sieci, na płaszczyźnie intencjonalnej, użytkownik jest informowany o działaniach związanych z przetwarzaniem konkretnych danych i w każdym momencie może podjąć decyzję o rezygnacji i wycofaniu się.

Zgoła odmienna sytuacja występuje w przypadku personalizacji na płaszczyźnie nieintencjonalnej. W trakcie codziennej eksploracji przestrzeni wirtualnej konsumenci internetu wielokrotnie poddawani są działaniom algorytmów, które skrupulatnie gromadzą o nich dane bez ich wiedzy. Robią to poprzez śledzenie historii wyszukiwań, zakupów, analizowanie korespondencji mailowej, a nawet nagrywanie fragmentów rozmów prowadzonych przez smartfony. Przyczyny takiego postępowania związane są bezpośrednio z ekwiwalentem finansowym otrzymywanym przez przedsiębiorstwa medialne w zamian za emisję reklam bądź handel danymi umożliwiającymi precyzyjne konstruowanie przekazów reklamowych bądź politycznych. Każdy użytkownik decydujący się na korzystanie z darmowych narzędzi, które udostępnia monopolista w swojej dziedzinie (jak Google czy Facebook), zaprasza ich tym samym do swojej przestrzeni prywatnej, wyrażając zgodę na przetwarzanie danych, zarówno tych, które przekazali świadomie, jak i tych, które zostały pozyskane bez ich wiedzy. Przetwarzanie zgromadzonych za pośrednictwem algorytmów informacji,

<sup>2</sup> Polityka cookies dotyczy przechowywania i wykorzystywania przez wyszukiwarkę internetową niewielkich fragmentów tekstu wprowadzonych przez internautę. Za sprawą takich działań użytkownik przełączający się pomiędzy stronami czy wracający do przeglądanych wcześniej stron nie musi ponownie wprowadzać danych, gdyż zostają one dodane automatycznie przez wyszukiwarkę. W 2012 roku Unia Europejska nałożyła na wszystkie państwa członkowskie obowiązek wprowadzenia zmian dotyczących informowania na stronach internetowych o prowadzeniu takich praktyk. Rok później w Polsce została znowelizowana ustawa – Prawo telekomunikacyjne, w której w myśl art. 173 właściciele witryn www są zobligowani do zamieszczenia komunikatu o przechowywaniu i wykorzystywaniu danych użytkowników końcowych. Pobrane z: <https://ico.org.uk/for-organisations/guide-to-pecr/cookies-and-similar-technologies> (07.01.2019).

takich jak: dane demograficzne, zachowania konsumenckie czy zainteresowania użytkowników, wiąże się z ich przechowywaniem, interpretowaniem oraz wykorzystaniem na potrzeby własne oraz zewnętrzne firm, które je zbierają (Ratalewska, Wierzbicka, 2017, s. 322). Co istotne, kwestie braku prywatności i udostępniania danych dotyczą nie tylko użytkowników wybranych narzędzi, ale wszystkich osób, które nawiązują z nimi jakąkolwiek komunikację w formie elektronicznej. W 2013 roku organizacja Consumer Watchdog wniosła pozew do sądu przeciwko firmie Google Inc., zarzucając jej naruszanie prywatności osób korespondujących z użytkownikami poczty Gmail. W odpowiedzi, jaką wystosowali w piśmie o odrzucenie pozwu, adwokaci światowego potentata zaznaczyli, że nie istnieją rozsądne powody, dla których osoba, która wysła wiadomość na skrzynkę pocztową oferowaną przez darmowy serwis, mogłaby oczekiwać jakiegokolwiek prywatności (Orliński, 2013, s. 96).

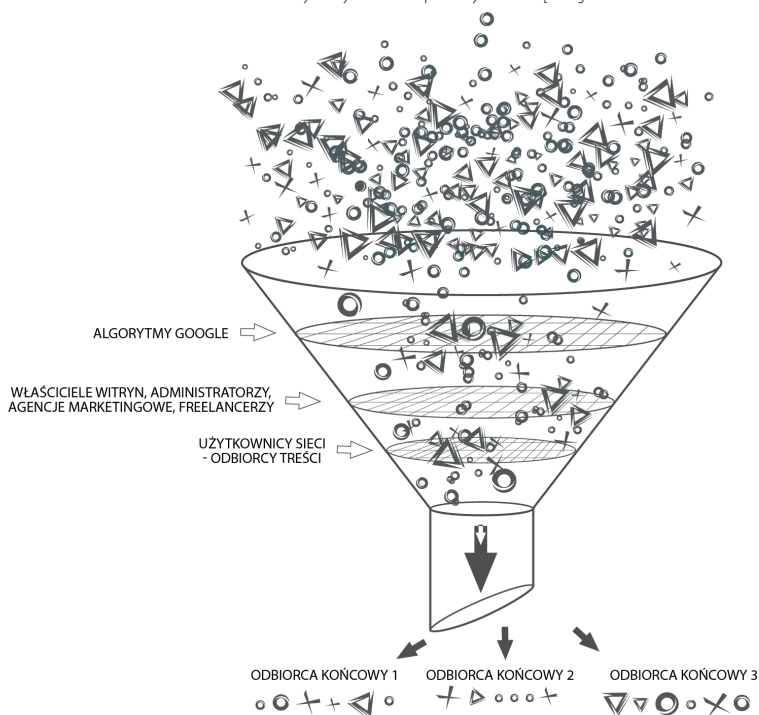
Aby lepiej zrozumieć skalę tego zjawiska i wynikające z niego konsekwencje, warto zapoznać się ze strukturą działania procesu wyszukiwania i podmiotami mającymi na niego wpływ. Ze względu na bardzo wysoki odsetek użytkowników wyszukiwarki Google w Polsce przedstawiona niżej analiza mechanizmów funkcjonowania bazy zindeksowanych stron internetowych oparta została na najważniejszym produkcie przedsiębiorstwa Google.

## ALGORYTMY GOOGLE A FILTROWANIE WYNIKÓW

Obserwując postępujące zjawisko personalizacji internetu, należy w pierwszej kolejności zwrócić uwagę na wielowymiarowość działań warunkujących ostateczny kształt otrzymywanych wyników wyszukiwania. Wyznaczają je bowiem kroki podejmowane w tym samym czasie przez trzy grupy podmiotów (rysunek 1). Pierwszą grupę stanowią odbiorcy treści, którzy poszukują informacji w sieci. Jak wcześniej wspomniano, na umiejscowienie poszczególnych wyników, które oddaje im wyszukiwarka, wpływa między innymi historia zapytań wprowadzonych przez nich w przeszłości czy chociażby ich lokalizacja. Drugą grupą są właściciele stron internetowych, którzy samodzielnie bądź za pośrednictwem osób trzecich (firm zewnętrznych czy freelancerów) podejmują rozmaite działania optymalizacyjne (SEO – *search engine optimization*) i marketingowe (SEM – *search engine marketing*), mające na celu podnoszenie pozycji strony www w wynikach wyszukiwania. Wszelkie przedsięwzięte kroki,

zarówno przez pierwszą, jak i drugą grupę, są wynikiem konieczności dostosowania się do zasad stanowiących przez grupę trzecią, najważniejszą pod względem warunkowania wahań pozycji poszczególnych stron. Tworzy ją całe zaplecze strukturalno-personalne firmy Google. I to właśnie oni za pośrednictwem aplikowanych algorytmów filtrują odpowiednio witryny z bogatego katalogu całej przestrzeni wirtualnej, bezpośrednio wpływając na ostateczny kształt rankingu stron będącego odpowiedzią na wprowadzane słowa bądź frazy kluczowe. Filtrowanie wyników przypomina lejek, do którego trafiają poszczególne fragmenty sieci determinowane przez działania wskazanych wcześniej grup podmiotów, by w ostatecznym rozrachunku przesłać wyniki spersonalizowane pod konkretnego użytkownika.

**Rysunek 1.** Podmioty determinujące hierarchię wyników wyszukiwania stron www zwracanych użytkownikom przez wyszukiwarkę Google.



Źródło: opracowanie własne

Algorytmy Google to programy komputerowe działające zgodnie z zaprogramowanym wcześniej zbiorem setek określonych zasad i procedur, dzięki którym wyszukiwarka wie, w jaki sposób ma przeanalizować każdą stronę internetową, a następnie ocenić ją, aby w końcowym etapie procesu

zaindeksować ją i ustanowić jej odpowiednie miejsce w hierarchii dla złożonych zapytań. Częstotliwość ich aktualizacji oraz konieczność wdrażania nowych produktów wynika przede wszystkim z rosnącej potrzeby segregacji informacji, które pojawiają się w sieci. Dobrze skrojona wyszukiwarka przetwarza w każdej sekundzie niewyobrażalną ilość danych, porządkując powszechnie panujący chaos w internecie i starając się oddać wyniki jak najlepiej pasujące do złożonych przez użytkowników zapytań. Niewątpliwym sukcesem było stworzenie mechanizmu, który będzie z powodzeniem oddawał doskonale spasowane wyniki bez względu na to, czy będą one pochodnymi pojedynczych słów kluczowych, czy tzw. długiego ogona wyszukiwania.

Aby zobrazować funkcjonowanie tej struktury, można ją przyrównać do rynku walutowego – strony internetowe, podobnie jak waluty na giełdzie, zajmują określone miejsce w hierarchii i poddawane są różnym czynnikom zewnętrznym, które decydują o codziennych wahaniach i wyznaczaniu nowej pozycji strony w wynikach wyszukiwania. Przyczyn niespodziewanych, a znaczących fluktuacji na rynku witryn należy doszukiwać się przede wszystkim w stale ulepszanych i aktualizowanych algorytmach. Pierwszym algorytmem zaaplikowanym wyszukiwarce Google był PageRank<sup>3</sup>. Ze skutkiem natychmiastowym wpłynął on na globalny ranking wszystkich stron internetowych i zapoczątkował światową gonitwę o miejsce na podium w wynikach wyszukiwania (Page, Brin, 1998, s. 15). Skupiał on swoje działania przede wszystkim na analizie treści witryny, przyznając jej odpowiednią wartość w 10-stopniowej skali<sup>4</sup>. O wartości strony decydowała z kolei ilość cytowań, czyli linków wychodzących oraz przychodzących, i jakość stron, z których pochodzą. Im większy węzeł powiązań, tym większa wartość witryn (*ibidem*). Z biegiem lat, wraz ze stale wzrastającą liczbą nowych stron internetowych, zwiększającą się świadomością technologiczną użytkowników internetu i rosnącym zainteresowaniem podejmowania działań w kierunku podnoszenia widoczności witryny w sieci, twórcy algorytmu zdali sobie sprawę, że PageRank nie był doskonałym narzędziem i łatwo było go zmanipulować i oszukać.

<sup>3</sup> Nazwa algorytmu wbrew pozorom nie pochodzi od angielskiego słowa „strona” (page), a od nazwiska jednego z twórców algorytmu i założyciela Google Larry’ego Page’a. Sam algorytm został opatentowany w 1998 roku, jednak nie przez jego twórców, lecz przez Uniwersytet Stanforda ze względu na to, iż narzędzie to powstało w ramach projektu zaliczeniowego jednego z przedmiotów. Pobrane z: <https://patents.google.com/patent/US6285999> (07.01.2019).

<sup>4</sup> Liczba 10 określała najwyższą wartość strony, zaś wyrażenie N/A wskazywało na całkiem nową witrynę.

Zanim do tego doszło, okazało się, że pomysłowość internautów wybiegała znacznie dalej niż sięgały granice wyobraźni twórców algorytmu. Przestrzeń wirtualna w mgnieniu oka wypełniła się sztucznymi tworam stworzonymi wyłącznie na potrzeby generowania linków. Tworzono tzw. farmy linków/treści, czyli strony internetowe poświęcone danej tematyce, gromadzące tysiące artykułów branżowych zawierających określone słowa kluczowe. Z zamieszczonych w artykułach wybranych fraz tworzone były hiperłącza odnoszące się do strony, której pozycja miała się docelowo podnieść. Im bardziej „zadbana farma”, tym wyższą generowała wartość wskaźników w rankingu dla linków zewnętrznych. Ale na tym nie koniec, internauci poszli znacznie dalej. Konieczność tworzenia setek artykułów wymagała czasu, pomysłu i bogatego zasobu słownictwa, ponieważ każdy powstały artykuł musiał być unikatowy i niepowtarzalny. Stworzenie mieszerek tekstów, które pozwalały na zachowanie cech unikatowości danemu artykulowi wyłącznie poprzez wymianę kilku bądź kilkunastu słów na ich odpowiednie synonimy, miało ułatwić ten proces.

Na przestrzeni lat Google wprowadził setki algorytmów oraz ich aktualizacji, które z różnym nasileniem wpływały na warunkowanie pozycji poszczególnych wyników wyszukiwania. Część z nich została oficjalnie uwzględniona w rocznych raportach, natomiast pozostałe pominięto bez żadnego uzasadnienia. Obecność kolejnych aktualizacji można było potwierdzić jedynie poprzez systematyczną obserwację i analizę wahań na rynku witryn. Największy wpływ wywarł tzw. zwierzyniec Google’a, czyli wprowadzone kolejno w latach 2011–2014 roku algorytmy Panda, Pinwin, Koliber oraz Gołąb (Mierzyński, 2018).

Algorytm Panda wdrożony został w lutym 2011 roku jako narzędzie pozostające w opozycji do witryn o treści niskiej jakości. Dziesiątki tysięcy stron www spadło w rankingu w wyniku wprowadzenia nowego algorytmu. Założeniem Pandy jest wyróżnianie w rankingach stron, które odznaczają się wysoką jakością, zaś karanie tych, których wartości informacyjne i strukturalne są wątpliwe (farmy linków, strony z zaburzonym stosunkiem reklam względem treści, strony z ubogą warstwą tekstową) (Slegg, 2015). Za treści wysokiej jakości uznawane są przede wszystkim artykuły unikatowe i oryginalne. Panda zwraca szczególną uwagę na teksty będące plagiatem bądź duplikatem, a także na warstwę językową, w której pojawiają się liczne powtórzenia (Slegg, 2016). Innowacyjność ta ukazała swoje możliwości w zakresie sprawdzania zawartości treści pod kątem merytorycznym. Sprowadzało się to bowiem do analizy tekstu i wykrywania niepowiązanych z tematyką artykułu zamieszczonych słów kluczowych linkujących

do zewnętrznych stron. Na przestrzeni lat algorytm był wielokrotnie ulepszany, a ostatnia jego aktualizacja, już blisko trzydziesta z kolei, została zarejestrowana przez użytkowników w lipcu 2015 roku. Każda z nich miała na celu przede wszystkim ustawienie wysoko w hierarchii stron, które będą wartościowym i korzystnym materiałem dla użytkowników, nie zaś dla właścicieli.

Algorytm Pingwin, wprowadzony w kwietniu 2012 roku, miał za zadanie ograniczanie zaufania do stron, za pośrednictwem których dochodzi do manipulowania naturalnym obiegiem wymiany linków (Mierzyński, 2018). Pingwin został doskonale skonstruowany pod kątem weryfikacji jakości zewnętrznych hiperłączy. Analizuje nie tylko liczbę nowych odnośników, które pojawiły się w krótkim czasie, ale potrafi również skontrolować ich jakość i z powodzeniem wykryć próby oszustwa. Ocenie podlegają również źródła czy sposoby zdobywania linków. Za niepożądane działanie uznawane jest prowadzenie niezróżnicowanej strategii optymalizacji witryn.

Algorytm Koliber pojawił się w sierpniu 2013 roku i zrewolucjonizował metodologię pozyskiwania pożądanych wyników na trzech płaszczyznach, dając jednocześnie początek innowacyjnemu procesowi wyszukiwania semantycznego (*ibidem*). Koliber to algorytm, na który czekała znaczna większość użytkowników sieci. W pierwszej kolejności usprawnił proces wyszukiwania informacji poprzez wyświetlanie poszukiwanych treści już na pierwszej stronie, pomijając konieczność wchodzenia na strony z zaprezentowanych kolejno linków. Po drugie, okazał się nieocenionym wsparciem w odnajdywaniu trafnych wyników względem lokalizacji użytkownika. Jednak jego najistotniejszym założeniem miało być lepsze zrozumienie zapytań na długi ogon wyszukiwania. Na przykład, wpisanie frazy: „w jakim miejscu w Warszawie najlepiej zjeść pizzę?” oznacza dla wyszukiwarki poszukiwanie restauracji w najbliższej okolicy użytkownika składającego zapytanie oraz najwyżej ocenianej pizzerii w Warszawie.

Algorytm Gołąb, wdrożony w 2014 roku, można uznać poniekąd za rozwinięcie Kolibra. Mechanizm jego działania związany jest bezpośrednio z lokalizacją internautów, co znacząco usprawniło proces wyszukiwania lokalnego (*ibidem*). Na przykład po wprowadzeniu frazy: „gabinet kosmetyczny” wyszukiwarka oddaje indeks stron nie tylko dopasowanych tematycznie, ale także zgodnych z ustaloną przez nią lokalizacją użytkownika. Równie istotne, co użyteczne, okazało się zintegrowanie wyników wyszukiwania z lokalizacją pożądanego miejsca/produktu na mapie. Dzięki takiemu rozwiązaniu po wpisaniu słów kluczowych bądź pełnej frazy



wyszukiwarka oddaje wyniki z linkami do stron www wraz z zaznaczoną dokładną ich pozycją na mapie z możliwością wyboru trasy dojazdu.

Nie tylko przedsiębiorstwo Google stara się o zindywidualizowane traktowanie internautów. Niedługo po wprowadzeniu algorytmu PageRank w imperium Marka Zuckerberga został uruchomiony algorytm EdgeRank, który po dziś dzień decyduje za nas, które informacje na Facebooku są najważniejsze i najbardziej interesujące, kogo lubimy bardziej, a kogo mniej, wyświetlając na prywatnej tablicy tylko przefiltrowany fragment treści z całego ogromu, który mógłby do nas dotrzeć. Wielokrotnie modyfikowany EdgeRank, podobnie jak algorytmy Google'a, rozwinął się w kierunku ulepszania parametrów oddawania w pierwszej kolejności wyników powiązanych z użytkownikami lokalizacją. Zmiany te dostrzec można chociażby poprzez wizytę na podstronie Facebook Marketplace, czyli usłudze umożliwiającej dokonywanie zakupów oraz handel wymienny między społecznością Facebooka. Zanim użytkownik zdąży sprecyzować zapotrzebowanie, narzędzie filtrujące automatycznie wyświetla produkty bądź usługi oferowane w jego najbliższym otoczeniu.

Warto w tym miejscu zwrócić uwagę na przekaz budowany przez wspomniane wyżej podmioty. Dają one jasno do zrozumienia, że wszelkie podejmowane przez nie działania i aktualizacje robione są z myślą o użytkownikach – „im więcej danych o sobie nam przekażesz, tym precyzyjniej dopasujemy nasze usługi do twoich potrzeb”. Przedsiębiorstwa, takie jak: Facebook, Instagram, Snapchat, Google, Microsoft czy Apple, zaproponowały konsumentom usługi najwyższej jakości, które zrewolucjonizowały metody komunikacji, ale – co najważniejsze – uzależniły ich od swoich narzędzi. Kompatybilne urządzenia, łatwe w użyciu i współpracujące ze sobą aplikacje, pozwalające na zarządzanie plikami z wielu miejsc i urządzeń jednocześnie, usługi poczty elektronicznej, społeczna przestrzeń swobodnej wymiany informacji w postaci tekstów, zdjęć, plików audio czy wideo, to zaledwie kropla w morzu ofert aplikacji i usług udostępnianych internautom za darmo. Ale czy na pewno całkowicie za darmo? Cena, którą muszą zapłacić, z perspektywy pojedynczego użytkownika może wydawać się pozornie niewielka, stanowią ją bowiem dane, które użytkownicy wielokrotnie wprowadzają w rozmaitych formularzach rejestracyjnych. Na ogół większość internautów jest już przyzwyczajona do „ekshibicjonizmu” w sieci w lekkiej formie poprzez dzielenie się fragmentami swojej sfery prywatnej. Odbywa się to za pośrednictwem publikowania zdjęć, preferencji, upodobań, sympatii, antypatii, czyli wszelkich czynności związanych z codziennym korzystaniem z rozmaitych portali społecznościowych.

W rzeczywistości są to informacje, które w dzisiejszym świecie są droższe od surowców. O tym, jak wielkie zagrożenie niosą ze sobą powyższe działania, świadczą słowa jednego z inżynierów Google w odniesieniu do ich hasła reklamowego „We’re not evil”: „We’re not evil. We try really hard not to be evil. But if we wanted to, man, could we ever!” (Pariser, 2011).

Nawet przyjmując najbardziej optymistyczne założenie, że intencje związane z przetwarzaniem danych oraz filtrowaniem wyników wyszukiwania nie są złe, to i tak nie zmienia faktu wynikających z nich negatywnych konsekwencji ponoszonych przez użytkowników.

## NEGATYWNE ASPEKTY PERSONALIZACJI

Eli Pariser wskazuje trzy podstawowe negatywne aspekty wynikające ze stosowania filtrów informacyjnych w sieci (Pariser, 2011, s. 9).

1. „W bańce jesteśmy sami” – na świecie istnieje wiele osób, których zainteresowania są zbliżone do naszych. Różnice, które determinują wygląd naszej bańki, to droga, którą pokonaliśmy w internecie, odkrywając czy rozwijając swoje zainteresowania: witryny, które odwiedzaliśmy; słowa kluczowe i frazy, które wpisywaliśmy w wyszukiwarce, by zdobyć informacje; kontakty, które nawiązaliśmy w mediach społecznościowych; zakupy, jakich dokonaliśmy, i wiele innych. Każdy internauta obiera inną drogę, wobec czego informacje, które zostaną przekazane, będą dedykowane mu indywidualnie.
2. „Wkroczenie do bańki nie jest świadome” – sięgając po konkretną prasę, możemy spodziewać się, że wszelkie treści będą sympatyzowały z określonymi poglądami, natomiast wpisując w wyszukiwarce daną frazę, nie wiemy, jakich wyników oczekiwać.
3. „Bańka jest transparentna” – w większości przypadków internauci nie zdają sobie sprawy z istnienia filtrów przepuszczających przez sito informacje, które do nich docierają. Zwykle satysfakcja wynikająca z przeprowadzonych poszukiwań pozostaje na wysokim poziomie, ponieważ odnajdują to, co mieści się w kręgu ich zainteresowań.

Sporządzona przez Parisera lista nie wyczerpuje jednak katalogu negatywnych aspektów związanych z filtrowaniem i personalizowaniem internetu, który można uzupełnić jeszcze o kilka aspektów. Zagrożenie, które – jak się wydaje – powinno pojawić się w pierwszej kolejności, to brak bezpieczeństwa informacji – jeden podmiot przechowuje bogaty zbiór danych



na temat internauty (lokalizacja, zainteresowania, preferencje, ale też loginy, hasła i kody). Inny negatywny aspekt to powtarzalność – wielokrotne ukazywanie treści, z którą użytkownik zdążył już się zapoznać. Kolejny to informacyjna segregacja. Joseph Turow w książce *The Daily You* zwraca uwagę na dyskryminację części internetowej społeczności, nazywając ich „wielkimi przegranymi”. Wskazuje, że dyskryminacja dotyczy nie konkretnych grup, a jednostek, które są w zindywidualizowany sposób traktowane gorzej (Turow, 2013). Następnym jest brak efektu zaskoczenia – sugerowanie użytkownikowi treści wyłącznie z kręgu jego zainteresowań zabija ciekawość. I w końcu efekt potwierdzenia – „pojawia się, kiedy mamy dostęp wyłącznie do informacji, które mogą nas zainteresować. Ludzie nieświadomie wybierają te źródła informacji, które potwierdzają ich wizję świata” (Górecki, 2011). Nieustanne korzystanie z tych samych źródeł może powodować radykalizowanie poglądów i blokadę na świeże spojrzenie na pewne kwestie.

## PODSUMOWANIE

Personalizacja w internecie niejako definiuje sposób funkcjonowania jednostki w społeczeństwie i stanowi wyznacznik kierunku rozwoju komunikacji nie tylko w przestrzeni wirtualnej, lecz także w świecie rzeczywistym. Pomimo wielu negatywnych aspektów, które niesie ze sobą stosowanie mechanizmów filtrujących informacje, jest to proces, bez którego swobodne i zręczne odnalezienie pożądaných wyników wyszukiwania jest praktycznie niemożliwe. Ilość informacji pojawiających się w sieci wzrasta w niewyobrażalnym tempie. W 2011 r. internauci generowali codziennie blisko milion wpisów na blogach, ponad 50 mln tweetów i ponad 210 mld wiadomości poczty elektronicznej (Górecki, 2011). W 2018 roku szacowano, że liczba osób korzystających z internetu przekroczyła 4 mld (Kemp, 2018), a tylko w wyszukiwarce Google rejestrowano niemal 6,5 mld zapytań w ciągu doby (Website hosting rating, 2018). Personalizacja jest naturalnym procesem rozwoju sieci, który przede wszystkim ułatwia i usprawnia komunikację i pozyskiwanie pożądaných treści, produktów czy usług w internecie, a nieustannie poprawiane i ulepszane algorytmy znacząco obniżają barierę wejścia do świata skrojonego na miarę potrzeb każdego z nas. Negatywne aspekty, które wynikają z personalizacji, są jednak znaczące i mogą w konsekwencji doprowadzić do wielu zagrożeń, ale należy przy tym zaznaczyć, że skala owego zagrożenia jest każdorazowo

determinowana czynnościami podejmowanymi przez samych użytkowników sieci. Mogą oni wpływać na obniżanie poziomu niebezpieczeństwa bądź jego całkowite niwelowanie za pośrednictwem konkretnych działań związanych z ochroną, takich jak: szyfrowanie korespondencji, stosowanie bezpiecznych haseł, przetrzymywanie dokumentów lokalnie czy – wydawać by się mogło – tak prozaiczne działanie, jak rotacyjne używanie dostępnych na rynku wyszukiwarek. Naturalnie każdy użytkownik może całkowicie zrezygnować z darmowych narzędzi, z których korzystanie wiąże się z udostępnianiem swoich danych, jednak niewielu zdecyduje się na tak radykalne rozwiązanie w obliczu wymiernych korzyści, które w zamian otrzymują.

## BIBLIOGRAFIA

- Baran, D. (2013). Internet w Polsce. W: K. Pokorna-Ignatowicz (red.), *Polski system medialny 1989–2011* (s. 77). Kraków: Oficyna Wydawnicza AFM.
- Battelle, J. (2005). *The Search How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*. New York: Penguin Books.
- Górecki, P. (2011). *Złapani w sieci*. Pobrane z: <https://www.newsweek.pl/wiedza/zlapani-w-siec/y1s0wp6> (1.01.2019).
- Katarzyniak, R.P., Kazienko, P.R., Sobecki, J.F. (1997). Wybrane problemy wyszukiwania informacji w sieci Internet. W: *IV Krajowe Forum Informacji Naukowej i Technicznej. Zakopane 2–5 września 1997 r.* (s. 152). Warszawa: Polskie Towarzystwo Informacji Naukowej.
- Kemp, S. (2018). *Digital in 2018: world's internet users pass the 4 billion mark*. Pobrane z: <https://wearesocial.com/uk/blog/2018/01/global-digital-report-2018> (12.01.2019).
- ICSTI – The International Council for Scientific and Technical Information (2010). Next generation search. W: *ICSTI Insights*. Pobrane z: [http://www.icsti.org/IMG/pdf/INSIGHT\\_2010\\_JULY.pdf](http://www.icsti.org/IMG/pdf/INSIGHT_2010_JULY.pdf) (07.01.2019).
- McLuhan, M. (2001). Galaktyka Gutenberga. W: E. McLuhan, F. Zingrone (red.), *Wybór tekstów*. (s. 179). Poznań: Zys i S-ka.
- Method for node ranking in a linked database*. Pobrane z: <https://patents.google.com/patent/US6285999> (10.11.2018).

- Mierzyński, K. (2018). *Historia aktualizacji algorytmów wyszukiwarki Google*. Pobrane z: <https://www.sunrisesystem.pl/blog/2155-historia-aktualizacji-algorytmow-wyszukiwarki-google.html#2000> (10.11.2018).
- Orliński, W. (2013). *Internet. Czas się bać*, Warszawa: Agora.
- Pisarek, W. (2009). *Wstęp do nauki o komunikowaniu*, Warszawa: WAiP.
- Page, L., Brin, S. (1998). *The pagerank citation ranking: Bringing order to the web*, Stanford Digital Library Technologies Project.
- Pariser, E. (2011). *The Filter Bubble: What The Internet is Hiding From You*, New York: Penguin Books.
- Ratalewska, M., Wierzbicka, A. (2017). Personalizacja w Internecie – innowacyjna strategia na rynku MŚP?. *Przedsiębiorczość i zarządzanie*, XVIII(12), cz. I, s. 322.
- Roszkowski, M. (2007). Folksonomia jako narzędzie społecznego tagowania. *Warsztaty Bibliotekarskie*, 4. Pobrane z: <http://www.pedagogiczna.edu.pl/warsztat/2007/4/070404.htm> (10.11.2018).
- Seymour, T.J. (2011). History Of Search Engines. *International Journal of Management & Information Systems*, 15(4), s. 51
- Skwarski, P. (2004). *Personalizacja internetu. Nowe oblicze sieci*. Pobrane z: <http://www.ipipan.waw.pl/~subieta/prezentacje%20studenckie/Personalizacja%20Internetu%20-%20Skwarski.ppt> (10.11.2018).
- Slegg, J. (2015). *Google Panda Update: Everything We Know About Panda 4.2*. Pobrane z: <http://www.thesempost.com/google-panda-update-everything-we-know-about-panda-4-2/> (15.11.2018).
- Slegg, J. (2016). *Understanding Google Panda: Definitive Algo Guide for SEOs*. Pobrane z: <http://www.thesempost.com/understanding-google-panda-definitive-algo-guide-for-seos/> (15.11.2018).
- StatCounter Global Stats, *Search Engine Market Share Poland, Jan-Dec 2018*. Pobrane z: <http://gs.statcounter.com/search-engine-market-share/all/poland/#monthly-201801-201812-bar> (10.01.2019).
- Turow, J. (2013). *The Daily You. How the New Advertising Industry is Defining Your Identity and Your Worth*. New Haven and London: Yale University Press.
- Website hosting rating (2018), *100+ internet stats and facts for 2018*. Pobrane z: <https://www.websitehostingrating.com/internet-statistics-facts-2018> (22.11.2018).

