



ISSN 2080-1807

TORUŃSKIE STUDIA BIBLIOLOGICZNE
2013, nr 1 (10)

Natalia Pamuła-Cieślak

Instytut Informatyki i Bibliologii
Uniwersytet Mikołaja Kopernika w Toruniu
e-mail: napa@umk.pl

Analiza czynników wpływających na jakościowe cechy wyników w wyszukiwarce Google

DOI: 10.12775/TSB.2013.006

STRESZCZENIE: W artykule podjęto próbę identyfikacji czynników mających istotny wpływ na jakościowe cechy rezultatów uzyskanych w wyszukiwarce Google. Wzięto pod uwagę algorytmy, zaproponowaną architekturę listy odpowiedzi, politykę koncernu oraz faktyczne zachowania użytkowników wykorzystujących to narzędzie. Wskazano istotne czynniki dla jakościowego postrzegania wyszukanej informacji, a także niedostatki – zarówno po stronie narzędzia, korporacji, jak i użytkowników wyszukiwarki.

SŁOWA KLUCZOWE: Google, jakościowe cechy wyników wyszukiwania, wyszukiwanie informacji.

Wprowadzenie

Narzędziem, które zmonopolizowało Internet, w aspekcie udostępniania informacji uporządkowanej w pewien określony sposób, jest bez wątpienia wyszukiwarka internetowa. Odsetek osób korzystających z tego narzędzia wzrósł w ostatnim dziesięcioleciu o 21% (podczas gdy w styczniu 2002 r. 52% Amerykanów określało siebie jako użytkowników wyszukiwarek, w lutym 2012 r. z wyszukiwarek korzystało już 73%

obywateli Stanów Zjednoczonych¹). W Polsce 92% internautów używa wyszukiwarek². Wyszukiwarką dominującą na internetowym rynku informacji jest Google³. Znajomość tej marki wyszukiwawczej, a nawet familiarny stosunek do niej, podkreśla używanie jej nazwy w formie czasownika oznaczającego czynność wyszukiwania – w języku angielskim to *googling* i *google*, w języku polskim czasownik ten przybiera formy: *wygooglować*, *googlować* itp.

Od użytkownika informacji pozyskanej on-line (poziomu jego wykształcenia w zakresie informacji literacy, ale także indywidualnych zdolności w tym zakresie) zależy, jak postrzega on proces oceny informacji oraz kwestie jej cech jakościowych. Badania wskazują na to, że sporą część odpowiedzialności za jakość i wiarygodność wyszukanej informacji użytkownicy przerzucają na narzędzie, którym się posługują⁴. Jak wykazano, w większości przypadków tym narzędziem jest Google. W niniejszym opracowaniu autorka stara się zidentyfikować i opisać cechy wyszukiwarki Google, które mają znaczący wpływ na jakościowe cechy prezentowanych wyników, oraz odnieść je do zachowań wyszukiwawczych internautów korzystających z tego narzędzia.

¹ K. Purcell, J. Brenner, L. Rainie, *Search Engine Use 2012* [on-line]. Pew Internet & American Life Project [dostęp 31 marca 2012]. Dostępny w World Wide Web: <http://pewinternet.org/Reports/2012/Search-Engine-Use-2012.aspx>.

² Internauci w Polsce to 60% obywateli powyżej 16 roku życia, w tym 97% uczniów i studentów oraz 89% osób z wyższym wykształceniem. Por. A. Jeran, *Internauta i jego aktywność – o paradoksie (nie)kompetencji polskich internautów na podstawie „Diagnozy społecznej 2011”*, „E-mentor” 2012, nr 4 (46), s. 34–39.

³ 83% Amerykanów wskazywało, że używa właśnie tej wyszukiwarki. Za: K. Purcell, J. Brenner, L. Rainie, dz. cyt. W Polsce ten odsetek jest jeszcze wyższy – 88,4% zapytań polskich internautów kierowano do Google. Por. *Wyszukiwarki udział zapytań* [on-line]. Megapanel PBI/Gemius [dostęp 31 marca 2012]. Dostępny w World Wide Web: <http://www.audience.gemius.pl/pages/display/wyszukiwarki-udzial>.

⁴ B. Pan i in., *In Google we trust: user's decisions on rank, position and relevance*, „Journal of Computer-Mediated Communication” 2007, no. 12, s. 803; E. Hargittai i in., *Trust online: young adults' evaluation of web content*, „International Journal of Communication” 2010, no. 4, s. 479.

PageRank i zespół algorytmów indeksujących

PageRank jest najbardziej znanym i najstarszym algorytmem indeksującym wyszukiwarki Google. Jest to parametr, który według wielu opinii przyczynił się do jej ogromnego sukcesu rynkowego, szybko dając jej pozycję wiodącego narzędzia wyszukiwawczego on-line. PageRank opiera się na analizie i punktowaniu powiązań hipertekstowych dokumentów w sieci WWW. W 1998 r., gdy koncern Google startował ze swoją wyszukiwarką, istniejące wówczas silniki wyszukiwawcze (Inktomi, AltaVista, Yahoo!) posiłkowały się przede wszystkim statystyczną analizą słów w tekście strony i metadanymi z jej nagłówka, na ich podstawie ustalając kolejność wyświetlanych wyników. Dodatkowo wspierano się algorytmami czuwającymi nad próbami sztucznego podnoszenia pozycji witryn w rankingach. PageRank wniósł do wyszukiwania nową jakość, punktuąc wysoko serwisy internetowe, do których z sieci prowadzi wiele odsyłaczy (w późniejszym okresie w samym PageRanku wprowadzono wiele zmian, bez naruszania naczelnej zasady działania), wspierając go innymi algorytmami jakościowymi⁵ – badającymi np. jakość odsyłaczy. PageRank sprawił, że miernikiem prezentowanych na stronie informacji stał się jej odbiorca – użytkownik Internetu, który „linkując” stronę internetową, oddaje na nią głos, podwyższając jej wskaźnik PageRank. W ten sposób to społeczność internetowa, budując sieć odsyłaczy, ma decydujący wpływ na kształt list rankingowych w wyszukiwarce. Jest to pewnego rodzaju sprzężenie zwrotne między użytkownikami wyszukiwarki (społeczność internautów) a jej twórcami i projektantami. Google dostarcza treści najbardziej oczekiwanej przez użytkowników (to, co sami w systemie odsyłaczy uważają za godne polecenia), a użytkownicy uznają, że treści te są odpowiednie i wiarygodne⁶. Twórcy raportu Centrum Badań nad Kulturą Popularną Szkoły Wyższej Psychologii Społecznej nazywają to zjawisko *autorytetem kolektywnym*⁷, który w Internecie dotyczy nie

⁵ *Algorytmy jakościowe* to termin stosowany przez koncern Google na określenie algorytmów poprawiających postrzeganie wyników wyszukiwania przez użytkowników, eliminujących z wysokich pozycji rankingowych strony niespełniające ustalonych przez koncern kryteriów. Autorka przyjmuje to określenie w niniejszym opracowaniu.

⁶ E. Hargittai i in., dz. cyt., s. 470.

⁷ *Młodzi i media. Nowe media a uczestnictwo w kulturze. Raport Centrum Badań nad Kulturą Popularną Szkoły Wyższej Psychologii Społecznej* [on-line]. Szkoła Wyższa

tylko wyszukiwarki, ale też wielu źródeł i zasobów informacyjnych tworzonych często przez anonimowy zbiór użytkowników. Clay Shirky ten sam mechanizm określa z kolei mianem *autorytetu algorytmu*⁸.

Należy podkreślić, że zasady działania PageRanku dają efekt w postaci pętli: „lepszy staje się coraz lepszy”. Oznacza to, że dla zapytań popularnych, często zadawanych wyszukiwarce, na wysokich pozycjach list rankingowych dominuje kilka witryn internetowych⁹. Szczególnie widoczny jest tu efekt dominacji Wikipedii, która pojawia się na wysokich miejscach list odpowiedzi zawsze, gdy przedmiot zapytania został w niej zdefiniowany. Dzieje się tak nawet wtedy, gdy mechanizmy weryfikacyjne samej encyklopedii wskazują, że hasło jest niekompletne lub niewłaściwie udokumentowane przez odsyłacze do innych źródeł.

Polityka jakości* Google, czyli algorytmy jakościowe

Jakość indeksowanych i wyświetlanych na wysokich pozycjach rankingowych zasobów sieciowych jest deklarowaną cechą w oficjalnych stanowiskach twórców i administratorów wyszukiwarki Google¹⁰. Zdając sobie sprawę ze zjawisk takich, jak pozycjonowanie i optymalizacja (ang. *Search Engines Optimization – SEO*), oraz działań marketingowych (ang. *Search Engines Marketing – SEM*), które są prowadzone w stosunku do wyszukiwarek (by podnieść pozycję określonych witryn w rankingach, a więc

Psychologii Społecznej [dostęp 31 marca 2013]. Dostępny w World Wide Web: <http://www.mim.swps.pl/>.

* Podobnie jak termin *algorytmy jakościowe*, koncern Google często stosuje określenie *polityka jakości* w stosunku do działań, które poprawiają postrzeganie wyników wyszukiwania przez użytkowników wyszukiwarki Google. Autorka przyjmuje w pracy to pojęcie, próbując ukazać jego przejawy w działaniu opisywanego narzędzia wyszukiwawczego.

⁸ C. Shirky, *A Speculative Post on the Idea of Algorithmic Authority*. W: Clay Shirky [Weblog] [on-line] 15.11.2009 [dostęp 31 marca 2013]. Dostępny w World Wide Web: <http://www.shirky.com/weblog/2009/11/a-speculative-post-on-the-idea-of-algorithmic-authority/>.

⁹ B. Pan i in., dz. cyt., s. 802.

¹⁰ *Jakość* pojmowana jest tu jako merytoryczne zasoby poświęcone dowolnemu zjawisku, dobrze opisane i umiejscowione na witrynie. Za zasoby o niskiej jakości przyjmuje się webspam, a zatem zasoby niemające spójnej i ważnej treści merytorycznej, których funkcją jest sprzedaż reklam sieciowych w dowolnej formie.



także podnieść współczynnik dotarcia do użytkownika z informacją lub ofertą), firma Google drobiazgowo sprecyzowała własną politykę w tym zakresie. Polityka jakości znajduje odbicie w kształcie list rankingowych i istotnie wpływa na wybory użytkowników (walka o wysokie miejsca na liście odpowiedzi dla konkretnych zapytań to walka o prawdziwe zainteresowanie odbiorców). „Skuteczna optymalizacja usprawnia indeksowanie witryn, a przy tym poprawia dostępność poszczególnych stron i ułatwia ich znalezienie. Optymalizacja może np. polegać na właściwym dobraniu słów kluczowych, aby upewnić się, że na stronie znajdują się najbardziej odpowiednie wyrazy, a nie tylko fachowe terminy, których zwykły użytkownik¹¹ nigdy nie wpisze do wyszukiwarki”¹². Powyższy cytat ukazuje, w jaki sposób koncern Google pojmuje politykę jakości – wpływanie na jakościowe cechy wyników wyszukiwania to dostosowanie wyświetlanych na wysokich pozycjach zasobów do umiejętności, gustów i potrzeb użytkownika, który z wyszukiwarki Google korzysta. Podkreśla to popularny charakter tego narzędzia, podczas gdy użytkownicy kierują do niego różne rodzaje pytań, także tych o charakterze specjalistycznym.

Pozycjonowanie i optymalizację zasobów Google postrzega dwutorowo, dzieląc ją na dobrą i potrzebną (ang. *white hat*) oraz szkodliwą i niepotrzebną (ang. *black hat webspam*). Pierwsza z nich „ułatwia użytkownikom korzystanie z witryny, wspomaga tworzenie wysokiej jakości treści i przyspiesza ładowanie stron, co jest korzystne zarówno dla użytkowników, jak i dla wyszukiwarek”¹³. Szkodliwe pozycjonowanie jest definiowane jako działanie mające zapewnić witrynie, „by [...] znalazła się wyżej, niż na to zasługuje”¹⁴. Przykładami niedozwolonego i zwalczanego przez Google pozycjonowania są maskowanie i przekiero-

¹¹ Sformułowanie *zwykły użytkownik* ukazuje, że na potrzeby rozwijania narzędzia wyszukiwawczego Google firma ta stworzyła narzędzia analizy zachowań pojedynczych użytkowników wyszukiwarki, a także grupowych preferencji różnych typów jej użytkowników. Por. zamieszczony w niniejszym artykule fragment pt. *Personalizacja i komercyjny charakter wyszukiwarki*.

¹² M. Cutts, *Kolejny krok w kierunku promowania witryn wysokiej jakości*. W: *Blog Google Polska* [on-line]. 25.04.2012 [dostęp 31 marca 2013]. Dostępny w World Wide Web: <http://googlepolska.blogspot.com/2012/04/kolejny-krok-w-kierunku-promowania.html>.

¹³ Tamże.

¹⁴ Tamże.

wywanie¹⁵, sztuczne zawyżanie znaczących słów kluczowych na stronie, umieszczanie słów kluczowych niezwiązanych z treścią, ukryty tekst lub odsyłacze, kopiowanie i wielokrotne publikowanie tej samej treści, tworzenie szkodliwych witryn (wyłudających dane użytkowników, zawierających wirusy lub inne szkodliwe oprogramowanie), tworzenie systemów wymiany linków (tzw. *farmy linków*), które nie mają na celu nic poza zwiększeniem wskaźnika PageRank¹⁶. Należy zauważyć, że ta polityka jest konsekwentnie utrzymywana w wyszukiwarce przez trzy główne nurty działań zaprezentowane w tabeli 1.

Tabela 1. Charakterystyka polityki jakości w wyszukiwarce Google

Przejawy polityki jakości Google	Opis podejmowanych działań
Wsparcie dla webmasterów	Wskazówki (techniczne, jakościowe i inne) dla twórców i administratorów witryn internetowych. Poza listą praktyk zakazanych webmasterzy są instruowani co do polityki jakościowej: „Twórz strony z myślą o użytkownikach, a nie wyszukiwarkach. Nie wprowadzaj swoich użytkowników w błąd. Unikaj sztuczek mających na celu poprawienie rankingu w wyszukiwarkach. Zastosuj praktyczną zasadę: zastanów się, czy możesz bez oporów opisać swoje działania wydawcom konkurencyjnej witryny lub pracownikowi Google. Inny przydatny test to zadanie sobie pytania: «Czy to przynosi korzyść moim użytkownikom? Czy warto byłoby to robić, gdyby nie istniały wyszukiwarki?». Zastanów się, co czyni Twoją witrynę wyjątkową, wartościową lub angażuje użytkowników. Spraw, by wyróżniała się wśród innych podobnych witryn” ¹⁷ .
Algorytmy jakościowe	Kluczowym algorytmem Google, który przyczynia się do eliminowania webspamingu w wynikach wyszukiwania, jest algorytm Panda – wprowadzony na początku 2011 r. Google deklaruje, że poprawił on wyniki wyszukiwania w przypadku niemal 12% wszystkich zapytań zadawanych wyszukiwarce,

¹⁵ Maskowanie to tworzenie dwóch odrębnych stron – jednej przeznaczonej dla wyszukiwarek, a drugiej dla użytkowników – i wzajemne przekierowywanie. W rezultacie indeksowanie zasobu dla wyszukiwarki nie daje realnego obrazu strony, z której może w rzeczywistości korzystać użytkownik.

¹⁶ *Wskazówki dla webmasterów* [on-line] [dostęp 31 marca 2013]. Dostępny w World Wide Web: <http://support.google.com/webmasters/bin/answer.py?hl=pl&answer=35769#3>.

¹⁷ Tamże.

Tabela 1. Charakterystyka polityki jakości (cd.)

Przejawy polityki jakości Google	Opis podejmowanych działań
	choć nie są znane szczegóły jego działania ¹⁸ . Kolejny ważny dla jakości rankingu algorytm wdrożono na początku 2012 r. Jest to algorytm wykrywający układ stron internetowych, obniżający pozycję witryn, zawierających zbyt mało treści w części witryny wyświetlającej się od razu na ekranie użytkownika (górną część strony domowej). Ma to wyeliminować z wysokich miejsc rankingowych witryny skupiające się na wyświetlaniu reklam. Szacunkowe dane wskazują na to, że aktualizacja ta wpłynie na około 1% wszystkich zapytań do wyszukiwarki ¹⁹ . W 2012 r. do wyszukiwarki wdrożono jeszcze jeden algorytm, którego szczegółów Google nie chce zdradzać, by nie ułatwiać działań spammerom sieciowym. Szacuje się, że globalnie wpłynie on na poprawę jakości wyników w przypadku około 3% zapytań, choć w przypadku języków, w których natężenie spamu on-line występuje częściej (w tym dla języka polskiego), zmiana może być zauważalna w przypadku 5% wyników ²⁰ .
Polityka karania niewłaściwych praktyk	Poza działającymi algorytmami walczącymi z webspamem Google reaguje na zgłoszenia o nieuczciwym pozycjonowaniu witryn i ręcznie eliminuje je z wyników lub drastycznie obniża ich pozycję w rankingu ²¹ .

Źródło: opracowanie własne.

Personalizacja i komercyjny charakter wyszukiwarki

PageRank i polityka jakości to niejedyne mechanizmy istotnie wpływające na wyniki wyszukiwania. Duże znaczenie ma też szereg działań i roz-

¹⁸ A. Singhal, M. Cutts, *Finding more high-quality sites in search*. W: *Google Official Blog* [on-line]. 25.02.2011 [dostęp 31 marca 2013]. Dostępny w World Wide Web: <http://googleblog.blogspot.com/2011/02/finding-more-high-quality-sites-in.html>.

¹⁹ M. Cutts, *Udoskonalony algorytm uwzględniający wygląd witryn*. W: *Blog Google Polska* [on-line]. 24.01.2012 [dostęp 31 marca 2013]. Dostępny w World Wide Web: <http://googlepolska.blogspot.com/2012/01/udoskonalony-algorytm-uwzględniajacy.html>.

²⁰ M. Cutts, *Kolejny krok...*

²¹ K. Szymański, [*#Google #spam #SWL*]. W: *Google+* [on-line]. 14.05.2012 [dostęp 31 marca 2013]. Dostępny w World Wide Web: <https://plus.google.com/+KasparSzymanski/posts/8Le2Reyp4Uc>.

wiązań mających na celu poznanie potrzeb i zachowań użytkowników wyszukiwarki Google, nie tylko dotyczących wyszukiwania i pozyskiwania informacji, koncern interesuje się bowiem także aktywnością i działaniami podejmowanymi w każdej oferowanej przez siebie usłudze. Wyszukiwarki oraz użytkowników dotyczy mechanizm zwany personalizacją wyszukiwania. Działa on na kilku poziomach. Personalizacja anonimowa odnosi się do osób nielogujących się w żadnej usłudze Google, a korzystających z wyszukiwarki. Może ona przejawiać się na trzy sposoby:

- Analiza demograficzna – prowadzona przez firmę Google w oparciu o założenie, że przedstawiciele różnych krajów mają różne potrzeby informacyjne oraz charakterystyczne zachowania wyszukiwawcze – np. konstruowanie zapytań, swoiste potrzeby informacyjne, zarówno w codziennej aktywności prywatnej, jak i zawodowej, naukowej, edukacyjnej, której dotyczą wyszukiwania specjalne. Dane demograficzne o zachowaniu użytkowników są zbierane i analizowane, co ma później odbicie w prezentowanych na terenie poszczególnych krajów wyników wyszukiwania dla poszczególnych zapytań, a także na sugestie w wyszukiwaniu dynamicznym (podpowiedzi najczęściej wyszukiwanych fraz w usłudze Google Instant Search).
- Geolokalizacja – użytkownicy są dokładnie lokalizowani w przestrzeni geograficznej, by wyniki wyszukiwania były lepiej dostosowane do miejsca ich przebywania (inne pozycjonowanie usług i instytucji lokalnych, reklama regionalna).
- Personalizacja wyników wyszukiwania i reklam ze względu na analizę zachowań wyszukiwawczych konkretnego użytkownika, anonimowego ze względu na brak profilu w usługach Google wymagających logowania. Dane o historii wyszukiwania dla użytkowników anonimowych Google pozyskuje, analizując pliki *cookies* i własne pliki wyszukiwarki²² zostawiane w pamięci urządzenia, na którym z sieci korzysta użytkownik.

²² P. Stanisławski, A. Singhal, *To, czym zajmujemy się w Google, to budowanie użytecznej inteligencji*. Gazeta.pl [on-line] z dn. 19.10.2012 [dostęp 31 marca 2013]. Dostępny w World Wide Web: http://technologie.gazeta.pl/internet/1,104665,12701537,_To_czym_zajmujemy_sie_w_Google_u_to_budowanie_uzytecznej.html.



Personalizacja szczegółowa użytkowników opiera się na analizie historii wyszukiwania oraz wszystkich zachowaniach i informacjach generowanych przez zalogowanych użytkowników w sieci usług i produktów Google. Szczególnie dużo cennej informacji o użytkownikach Google pozyskuje z analizy wysyłanych i przychodzących wiadomości elektronicznych wraz z załącznikami (dotyczy to posiadaczy poczty elektronicznej Gmail), dokumentów tworzonych i edytowanych w pakiecie biurowym Google Docs, plików przechowywanych „w chmurze” w usłudze Google Drive, terminów wpisywanych do Kalendarza Google. Nie bez znaczenia jest też analiza informacji nietekstowej, tj. multimediów i obrazów w usługach, takich jak: Picasa, YouTube, Drive²³.

Personalizacja ma dwa zasadnicze cele. Pierwszym jest zwiększenie i utrzymanie zainteresowania użytkowników wyszukiwarką (dobór wyników wyszukiwania i reklam do ich potrzeb). Drugi ma ścisły związek z komercyjnym charakterem korporacji, której 96% dochodu według danych z 2011 r.²⁴ pochodzi z reklam, a głównym nośnikiem i generatorem usług reklamowych jest wyszukiwarka – monopolista na internetowym rynku wyszukiwawczym. Jeśli reklama jest ściśle powiązana z potrzebami werbalizowanymi przez użytkowników w różny sposób (często nie zdają oni sobie sprawy z analizy wiadomości e-mail czy prywatnych dokumentów przez koncern), istnieje duże prawdopodobieństwo kliknięcia na spersonalizowane treści komercyjne, co przekłada się bezpośrednio na notowane zyski spółki. Drobiazgowo dobierane treści sponsorowane znacząco wpływają na percepcję wyników wyszukiwania przez użytkowników – 62% z nich nie odróżnia treści komercyjnej od faktycznie generowanych wyników wyszukiwania²⁵. Wskazuje to, jak bardzo zjawisko personalizacji – skorelowane z wyświetlaniem treści reklamowej (w stosunku do której nie obowiązują takie mechanizmy wyszukiwarki, jak PageRank i polityka jakości) – jest powiązane z jakością, obiektywizmem i wiarygodnością zasobów na liście odpowiedzi Google.

²³ Tamże.

²⁴ *United States Securities and Exchange Commission, Annual Report Pursuant to Section 13 or 15(d) of the Securities Exchange Act of 1934 for the fiscal year ended December 31, Washington DC 2011, s. 26.*

²⁵ D. Fallows, *Search Engine Users*, Washington DC 2005, s. II.

Formy prezentacji informacji

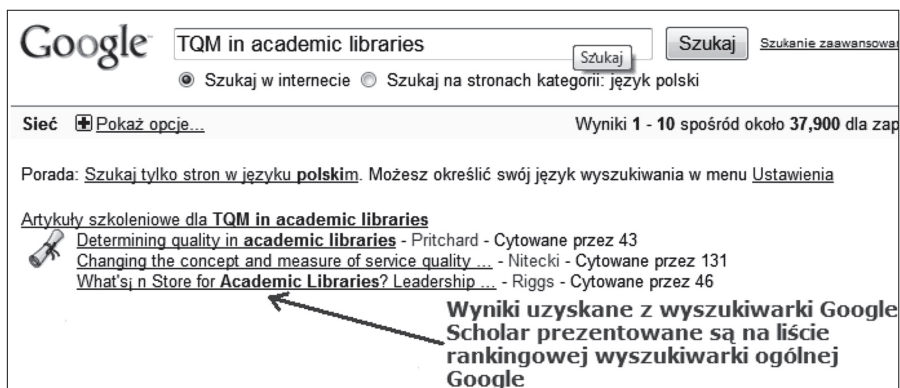
Na przestrzeni 14 lat funkcjonowania w sieci firma Google zmieniała wygląd wyszukiwarki oraz listy rankingowej wielokrotnie, dotyczy to także różnych dostępnych usług i opcji. Nigdy nie były to jednak radykalne zmiany, interfejs ma od zawsze prosty i przejrzysty wygląd, a ranking odpowiedzi to zestawienie wyszukanych witryn w formie stronicowanej listy. Zmianie podlegały tylko opcje, z których użytkownik mógł skorzystać, oraz ich położenie względem listy uzyskanych rezultatów. Dwa elementy formy prezentacji wyników mogą być uznane za przełomowe rozwiązania w dziejach koncernu i mieć wpływ na ocenę jakościowych cech prezentowanych rezultatów wyszukiwania. Te elementy to:

- uniwersalne wyszukiwanie – integracja informacji ze wszystkich dostępnych usług informacyjnych oraz wyszukiwarek specjalnych Google,
- próby przełamania linearnego charakteru rankingu odpowiedzi przez proste wizualizacje mające na celu ułatwienie doprecyzowania zapytania na podstawie dostępnych odpowiedzi wynikających z analizy kontekstowej wyników wyszukiwania (*Wonder Wheel* oraz *Timeline* – obie już nie funkcjonują – oraz najnowsze rozwiązanie kontekstowe – *Knowledge Graph* dostępne w chwili powstawania niniejszego artykułu jedynie w anglojęzycznej wyszukiwarce Google).

Uniwersalne wyszukiwanie jest rozwiązaniem przełomowym, gdyż pozwoliło w pewnym stopniu wyeliminować rozproszenie informacji (o różnym charakterze i sposobie indeksowania) w różnych interfejsach wyszukiwawczych. Mimo że dotyczy ono jedynie integracji informacji indeksowanej w usługach informacyjno-wyszukiwawczych Google, samo w sobie jest odpowiedzią na problem Ukrytego Internetu (zasobów nieindeksowanych w wyszukiwarkach o charakterze specjalnym, ze szczególnym uwzględnieniem zawartości baz danych) i przywiązania użytkowników do korzystania z jednego narzędzia wyszukiwawczego (wyszukiwarka ogólna Google) niezależnie od faktycznych potrzeb informacyjnych (np. informacji o charakterze akademickim, specjalistycznym). Dzięki wyszukiwaniu uniwersalnemu wyniki są kompilowane w zależności od zadanego zapytania – na pierwszej stronie rankingu odpowiedzi



pojawiają się odsyłacze do innych usług, np. Google Grafika, Mapy, Scholar czy materiały multimedialne zamieszczone w YouTube.



Rysunek 1. Prezentacja materiałów o charakterze naukowym (Google Scholar) w wyszukiwarce Google

Źródło: Hasło „TQM in academic libraries” [on-line]. Google.pl [dostęp 31 marca 2013]. Dostępny w World Wide Web: <http://daa.pl/4ks8>.

Znaczenie takiej formy prezentacji informacji jest z jakościowego punktu widzenia ogromne – do tej pory użytkownicy często ograniczali się do korzystania wyłącznie z wyszukiwarki sieciowej, dziś za jej pomocą mogą trafić na zasoby spoza tradycyjnie indeksowanego WWW, bez konieczności pamiętania i obsługiwania narzędzi wyspecjalizowanych. Na rysunku 1 przedstawiono fragment uniwersalnego rankingu ze wskazaniem na wyszukiwarkę naukową Google Scholar, która dla wskazanego zapytania prezentuje wyniki najtrafniejsze, mające charakter specjalistyczny, pochodzące także z tradycyjnie nieindeksowanych baz danych (wspomniany Ukryty Internet).

Przełamanie linearności rankingu jest bardzo istotne wobec faktu, że użytkownicy korzystają jedynie ze znikomej jego części w postaci kilku pierwszych odsyłaczy. Oznacza to, że pozostała część rezultatów traci na znaczeniu, a kolejne ekrany wynikowe mogłyby przestać funkcjonować. Rozwiązania wizualizujące otrzymane rezultaty pozwalają na proste zmiany w zapytaniu, a więc zawężanie i doprecyzowywanie otrzymanych rezultatów. W przeszłości były nimi wspomniani już

Timeline i Koło informacji (ang. *Wonder Wheel*). Najnowsze rozwiązanie graficzne – *Knowledge Graph* daje (na razie na anglojęzycznym rynku wyszukiwawczym) ogromne możliwości. Działa ono w oparciu o dwa ogromne źródła informacji: Freebase, która jest olbrzymią bazą wiedzy i powiązań semantycznych między obiektami (firma Google jest jej właścicielem), oraz powszechną Wikipedię (z której Google korzysta i którą wspiera finansowo). Korzystając z rozwiązań semantycznych, Google wyświetla nie tylko odsyłacze do stron i dokumentów w postaci rankingu, ale także informacje o szukanej osobie, obiekcie czy zjawisku wraz z powiązaniem. W przypadku pojęć o charakterze homonimów pozwala na wybór kontekstu wyszukiwania w formie zbliżonej do odsyłacza „Zobacz też” i dopiero wówczas prezentuje właściwe wyniki (zob. rysunek 2).

The screenshot displays a search result for "Washington" with a Knowledge Graph interface. On the left, there is a Wikipedia summary of the state of Washington, including its location and history. Below this, there are several news articles with thumbnails and brief descriptions, such as "Wizards vs. Hawks: Washington's 10th straight loss in Atlanta" and "Former USC tight end Gerald Washington dreams of championship". On the right side, there is a map of Washington state with a circular inset showing the state seal. Below the map, the word "Washington" is prominently displayed, followed by the word "State". A detailed description of the state's location and capital is provided. Below this, there is a list of "Points of interest" including Olympic National Park, Space Needle, Mount Rainier National Park, and Woodland Park Zoo. At the bottom, there is a section titled "See results about" which lists the University of Washington and the District of Columbia (DC).

Rysunek 2. *Knowledge Graph* dla zapytania „Washington”

Źródło: Hasło „Washington” [on-line]. Google.com [dostęp 31 marca 2013]. Dostępny w World Wide Web: <http://daa.pl/4ks9>.

Czynnik ludzki

Na cechy jakościowe rezultatów uzyskanych w wyszukiwarce ogromny wpływ mają zachowania jej użytkowników, którzy przez interakcję z systemem, jak również stały zespół zachowań nie tylko generują listy odpowiedzi, ale też wpływają na postawy twórców tych narzędzi.

Wiele badań wskazuje na to, że użytkowników, w stosunku do procesów wyszukiwawczych, a także zachowań informacyjnych w sieci WWW w ogóle, cechuje zasada najmniejszego wysiłku²⁶. Prostota i szybkość to te cechy wyszukiwania, które są dla użytkowników najważniejsze – wśród zapytań kierowanych do wyszukiwarek dominują te proste, zadane w podstawowym interfejsie, bez prób filtrowania i użycia dodatkowych opcji wyszukiwawczych²⁷. Jednym z kryteriów wiarygodności strony i oceny jakości publikowanej na niej treści jest wysoka pozycja w rankingu odpowiedzi wyszukiwarki Google, nawet wtedy, kiedy zupełnie nie zaspokaja ona potrzeb informacyjnych użytkowników oraz zawiera informacje, której nie potwierdzają inne źródła²⁸. Badania pokazują, że część internautów uważa, że to wyszukiwarka, a nie inne zasoby informacyjne i treść wyszukanej strony, jest właściwym narzędziem oceny informacji i wiarygodności witryn²⁹. Bywa, że cechuje ich brak refleksji, że nie odwołują się do własnej wiedzy i doświadczeń, a także innych narzędzi i zasobów informacyjnych – ocena zostaje powierzona współużytkownikom i współtwórcom zasobów treściowych sieci WWW, co w rzeczywistości oznacza scedowanie tej funkcji na algorytmy wyszukiwarek³⁰. W parze z zaufaniem do PageRanku i innych algorytmów jakościowych idzie nieznamość zasad ich działania, polityki Google oraz czynników wpływających na wysoką pozycję rankingową witryn³¹, jak również zasad rządzących treściami reklamowymi publikowanymi na listach odpowie-

²⁶ S. Cisek, R. Sapa, *Komunikacja naukowa w Internecie – mity i rzeczywistość* [on-line]. E-prints in Library and Information Science [dostęp 31 marca 2013]. Dostępny w World Wide Web: http://eprints.rclis.org/bitstream/10760/9148/1/cisek_sapa_2006.pdf.

²⁷ Tamże; J. R. Griffins, P. Brophy, *Student searching behavior and the Web: use of academic resources and Google*, „Library Trends” 2005, vol. 53, iss. 4, s. 540–541; D. Fal-lows, dz. cyt., s. IV.

²⁸ E. Hargittai i in., dz. cyt., s. 479; B. Pan, dz. cyt., s. 801–802.

²⁹ E. Hargittai i in., dz. cyt., s. 479.

³⁰ *Młodzi i media...*

³¹ E. Hargittai i in., dz. cyt., s. 484; B. Pan, dz. cyt., s. 801–802.

dzi³². Do podobnych konkluzji prowadzi analiza raportu OCLC opublikowanego w marcu 2010 r. Wynika z niego, że poddani badaniom uczniowie i studenci byli zadowoleni z wyników uzyskanych w wyszukiwarkach, oczekiwali prostych i intuicyjnych interfejsów wyszukiwawczych, a ocena jakości uzyskanych wyników była jednym z trzech zasadniczych problemów raportowanych przy korzystaniu z informacji. Nie rozumieli oni także prawdziwej natury Internetu³³. Z dozą umiarkowanego optymizmu warto jednak odnotować, że im ważniejsze (w sensie osobistym) jest dla użytkownika zadanie wyszukiwawcze, tym bardziej skłonny jest on weryfikować jakość wyszukanej informacji. Jeszcze większą wagę użytkownicy przykładają do jakościowej strony wyszukanej informacji, jeśli od poszukiwań i ich wyników mogą zależeć istotne kwestie życiowe (np. skutki zdrowotne³⁴). Taką refleksją wykazują się także internauci, którzy w ważnych sprawach zaufali wysoko umiejscowionym w rankingu wynikom wyszukiwawczym i ponieśli tego konsekwencje³⁵.

Podsumowanie

Powszechność marki Google, jej wszechobecność i monopolistyczny charakter ma wpływ nie tylko na to, jaka treść jest najbardziej dostępna dla użytkowników, ale także na to, co uznają za obowiązujące, wiarygodne i wysokojakościowe. Użytkownikom brakuje nie tylko wiedzy i refleksji dotyczącej podstawowych mechanizmów jakościowych rządzących wyszukiwarką, brakuje im też wiedzy o prawach, które rządzą Internetem, będącym zarówno ich środowiskiem informacyjnym, jak i obszarem funkcjonowania narzędzi wyszukiwawczych, z których korzystają i którym zaufali. Jednocześnie wdrażane w wyszukiwarce Google rozwiązania technologiczne i deklarowana polityka jakościowa wydają się niewystar-

³² D. Fallows, dz. cyt., s. II.

³³ L. S. Connaway, T. J. Dickey, *The Digital Information Seeker: Report of the Findings from Selected OCLC, RIN, and JISC User Behaviour Projects* [on-line]. Joint Information Systems Committee [dostęp 31 marca 2013]. Dostępny w World Wide Web: <http://www.jisc.ac.uk/media/documents/publications/reports/2010/digitalinformationseekerreport.pdf>.

³⁴ E. Hargittai i in., dz. cyt., s. 485–486.

³⁵ *Młodzi i media...*

czające, opierają się głównie na wizerunku narzędzia wyszukiwawczego i jego mechanizmów rankingowych. Informacje błędne, niekompletne, zniekształcone są powielane przez internautów (szczególną rolę odgrywają tu serwisy społecznościowe) oraz przez automaty zajmujące się ciągłym generowaniem i powielaniem treści elektronicznej do celów komercyjnych (wspomniany już webspaming). Oprócz prób walki z nieuczciwą optymalizacją i marketingiem kierowanym do wyszukiwarek, Google nie stosuje żadnych mechanizmów analizujących i punktujących faktyczną wartość informacyjną witryn, gdyż wymaga to zaangażowania zarówno nowoczesnych i kosztownych technologicznie rozwiązań, jak i intelektu ludzkiego, który jest równie kosztowny. Twórcy wyszukiwarki Google, monopolisty i świetnie prosperującego przedsięwzięcia komercyjnego, przyjęli model biznesowy opierający się na sprawdzonych, powtarzalnych zachowaniach internautów i dostosowaniu się do ich wyborów. Dopóki użytkownicy powracają do wyszukiwarki, mają z nią więź emocjonalną³⁶ oraz generują ruch w płatnych opcjach reklamowych oferowanych przez korporację, polityka jakościowa będzie koncentrowała się na eliminowaniu niechcianych przez społeczność wyników, a nawet funkcji czy architektury listy wynikowej, czego dowodem są liczne zmiany w tym obszarze odnotowywane na przestrzeni ostatnich lat.

Proponowanym we wszystkich niemal opracowaniach na ten temat wyjściem z impasu jest powszechna i ciągła edukacja informacyjna użytkowników sieci (co dziś równa się edukacji całych społeczeństw). Można zauważyć, że działania edukacyjne w obszarze środowiska internetowego, wsparte negatywnymi doświadczeniami użytkowników (niepotwierdzającymi skuteczności algorytmów jakościowych w stosunku do wiarygodności wyszukanej informacji), czynią wyłom w zadowoleniu użytkowników, którzy w przytaczanych tu badaniach deklarują pewną niepewność co do autorytetu zbiorowości oraz wskazują na problemy z oceną informacji pozyskaną on-line.

Wydaje się, że dobrych rozwiązań w obszarze poprawy cech jakościowych rezultatów wyszukiwawczych możliwych do uzyskania w wyszukiwarce Google nie można uzyskać, nie ingerując w samo narzędzie, co zależy wyłącznie od jej twórców i ich zapatrywania na to zagadnienie. Obszarem, w którym możliwe są działania, jest wspomniana wcześniej

³⁶ Tamże, s. 480.



edukacja informacyjna oraz nowe technologie, które pojawiają się lub będą się pojawiać zarówno na komercyjnym rynku wyszukiwawczym, jak i na rynku specjalistycznym – wszędzie tam, gdzie parametry jakościowe wyników mają szczególne znaczenie.



Analysis of factors affecting the quality of results in Google search engine

ABSTRACT: The paper aims to identify factors that have a significant impact on the quality characteristics of the results obtained in the Google global search engine. Several factors were considered, e.g. the impact of algorithms, the results page architecture, the Google company policies and the actual behaviour of the Google search engine users. Significant factors and some shortcomings in terms of information quality were identified, both on the tool's and corporation's side and the users' side.

KEYWORDS: Google, information retrieval, qualitative characteristics of search results.

