

Wyszukiwarki i ich rola w kształtowaniu Internetu

Wstęp

Internet jest opisywany przez entuzjastów jako czynnik demokratyzujący przestrzeń medialną, zdominowaną przez wielkie, komercyjne media masowe. Wielość kanałów komunikacji, które współtworzą Internet, interaktywność, i zatarcie granic między komunikacją masową a personalną czynią z niego medium wyjątkowe. Jednak optymistyczne oczekiwania towarzyszyły także wczesnemu rozwojowi radia i telewizji (Introna i Nissenbaum 2000).

Media masowe to tak zwane "push media". Instytucje komunikowania masowego selekcionują informację, którą później "popychają" w stronę odbiorców - decydując do jakiej informacji widzowie będą mieć dostęp. Internet to - teoretycznie - "pull medium", które pozwala użytkownikowi selekcionować i "przyciągać" informacje (McQuail i Suine 1998). Według wizjonerów Internet powinien dążyć do ideału przestrzeni publicznej, w której istnieje skrajna demokratyzacja funkcji nadawcy oraz uniwersalny dostęp do różnorodnych źródeł informacji. Wraz z komercjalizacją Internetu w połowie lat 90-ych pojawiły się próby przekształcenia Internetu w "push medium", w którym ruch internetowy skupiałby się w obrębie kilku portali będących kompleksowym, ale selektywnym źródłem informacji.

W naszej pracy chcemy pokazać, że funkcję selekcionującą informację *gate-keeperów* pełnią także internetowe wyszukiwarki. Wyszukiwanie informacji w złożonej - wbrew potocznym poglądom - strukturze sieci wymaga neutralnych narzędzi zapewniających równy dostęp do wszystkich obszarów Internetu. Tymczasem różne aspekty działania wyszukiwarek prowadzą do skrzywienia (*bias*), które systematycznie faworyzuje jedno źródło informacji kosztem innych.

Przywołując teksty Marshalla McLuhana można powiedzieć, że wyszukiwarki spełniają w Internecie funkcję przedłużenia ludzkich zmysłów. Dzięki nim uzyskujemy orientację w nieznanym przestrzeni, możemy poruszać się po terenach, do których nie mieliśmy wcześniej dostępu. Jednocześnie są to przedłużenia, które "wprowadzają głębokie i trwałe zmiany w człowieku i [...] przekształcają jego środowisko" (McLuhan 1995). Wyszukiwarki nie są neutralne. Przedstawiają użytkownikowi swoją własną ofertę zawartości Internetu, którą można zweryfikować jedynie poprzez wykorzystanie innej wyszukiwarki. Środowisko sieci - w tej postaci, jaka jest dostępna Internautom - jest, zatem formowane i stale przekształcane poprzez technologię i ideologię immanentnie wpisana w strukturę systemów wyszukiwania.

Kształt Internetu

Internet to *“elektroniczna sieć sieci, która łączy ludzi i informacje poprzez komputery i inne elektroniczne urządzenia, umożliwiając komunikację międzyludzką i pobieranie informacji”* (DiMaggio et al. 2001, 307). Nie posiada jednolitej architektury zarówno na poziomie technicznej infrastruktury, jak i informatycznego kodu. Poszczególne sieci tworzące Internet to różnorakie kanały przekazu informacji, od form komunikacji personalnej do mediów masowych. Różnorodność oznacza, że *“w przypadku Internetu, w każdym punkcie tradycyjnego modelu procesu komunikacji może znajdować się jedna, kilka lub wiele komunikujących się jednostek”* (Morris i Ogan 1996). Wspólny jest jedynie podstawowy protokół przekazu danych oraz podłączony do sieci komputer, będący interfejsem.

Zmienna natura i nieustanny wzrost utrudniają zebranie rzetelnych ilościowych danych. Według wiarygodnych szacunków w roku 2001 liczba stron WWW przekroczyła miliard (Kleinberg i Lawrence 2001), choć według niektórych źródeł istnieje nawet 2,5 miliardów stron. W Internecie istnieje ponad 30 milionów zarejestrowanych nazw domen (www.domainstats.com) a liczba Internautów wynosi ponad 400 milionów (Castells 2001).

Internet jest często widziany jako bliski habermasowskiej wizji idealnej przestrzeni publicznej, w której dochodzi do racjonalnej komunikacji. Kolejną wartością, obok powszechności dostępu i demokratyzacji funkcji nadawcy, jest ilość dostępnej informacji. *“Jako powszechnie dostępne źródło niezliczonej informacji, Sieć obiecuje liczne korzyści”* (Introna i Nissenbaum 2000). Swoboda dostępu i wymiany informacji była ważną wartością dla wielu osób i środowisk odpowiedzialnych za kształt Internetu, szczególnie zanim do Sieci masowo wkroczyły podmioty komercyjne (Castells 2001).

Krytycy wskazywali równolegle na wady i patologie Internetu. Wystarczy wspomnieć o istnieniu tzw. Digital Divide, czyli znaczących nierówności w dostępie do mediów elektronicznych oraz o przeciążeniu związanym z nadmiarem dostępnej informacji. Niemniej sama architektura sieci nie jest postrzegana jako źródło tych problemów. Zdecentralizowany i samo-organizujący się Internet jest widziany jako przestrzeń pozbawiona hierarchicznej struktury i zarządzania, a więc w pewnym sensie demokratyczna (Kleinberg i Lawrence 2001). Zagrożeniem z tego płynącym może być co najwyżej brak kontroli społecznej.

Badania struktury sieci

Można, ignorując treść, opisywać cechy Internetu na podstawie analizy rozmieszczenia stron i połączeń między nimi: *“Nieuregulowany wzrost prowadzi do wielkiej złożonej sieci, [...] topologia tego grafu [sieci - AT] determinuje strukturę połączeń w Sieci, a tym samym łatwość odnajdywania danych”* (Albert et al. 1999, 130) . Wizji Internetu jako przestrzeni zdecentralizowanej, uniwersalnie dostępnej i demokratycznej odpowiada abstrakcyjny model sieci bliskich sobie elementów, które są gęsto, równomiernie i w pełni ze sobą powiązane. Taka struktura odpowiadałaby teorii tzw. “małego świata”. Opracował ją w 1967 roku Stanley Milgram, który udowodnił empirycznie, że dwa losowo wybrane elementy w ogromnych sieciach społecznych (takich jak zglobalizowane społeczeństwo czy społeczność Internetu) dzieli łańcuch jedynie kilku pośredników. W roku 1999 aparat teorii “małych światów” został użyty do analizy wycinka Internetu. Autorzy badania dowiedli regularności struktury oraz stosunkowo niewielkiego rozmiaru Internetu. Zdaniem badaczy w pozornie ogromnej i chaotycznej strukturze Internetu dowolne dwie strony są oddalone łańcuchem co najwyżej dziesiętnastu innych stron. *“Na szczęście, zaskakująco mała średnica sieci oznacza, że cała ta informacja znajduje się zaledwie kilka kliknięć stąd”* (Albert et al. 1999, 130).

W tym samym czasie przeprowadzono badania, które przeczyły tym wynikom. W 1999 r. wykonano dwie analizy wycinków Internetu liczących po kilkadziesiąt milionów stron, które wykazały, że strony

internetowe są nierównomiernie połączone. Przeważająca większość stron posiada niewiele odnośników łączących je z resztą sieci. Z kolei nieproporcjonalnie mały procent stron posiada skrajnie dużo odnośników, tworząc centralne węzły sieci. Kolejnym krokiem były badania struktury połączeń wykonane w roku 2000 przez zespół IBM Research z Almaden na zbiorze około 200 milionów stron skatalogowanych przez wyszukiwarkę Altavista, wraz z 1,5 miliardami odnośników (Broder et al. 2000). Na tej podstawie potwierdzono wcześniejsze ustalenia oraz sformułowano tzw. "teorię muszki" (*bow-tie theory*), która głosi, że Internet jest strukturą nierównomiernie połączoną, którą z grubsza można podzielić na cztery "obszary" o odmiennej charakterystyce połączeń. W środku znajduje się centralne jądro (*central core*) odpowiadające węzłowi muszki. Składa się na nie 30% wszystkich stron, które są ściśle wzajemnie powiązane, umożliwiając łatwe przemieszczanie się Internautom. Po bokach znajdują się: przestrzeń OUT powiązana z jądrem jedynie odnośnikami wychodzącymi od niego oraz przestrzeń IN, z której odnośniki biegną wyłącznie do jądra. W każdej z tych przestrzeni znajduje się po 24% wszystkich stron WWW. W ostatniej grupie, obejmującej 22% stron, znajdują się "wąsy" połączone wyłącznie z przestrzenią IN lub OUT, "tuby" łączące obydwie przestrzenie z pominięciem jądra lub "wyspy" stron nie powiązanych z żadną z powyższych grup (Broder et al. 2000). Taka struktura oznacza, że niektóre obszary sieci nie są w ogóle połączone - istnieje jedynie 24% szansy, że między dowolnymi dwoma stronami będzie istniała ścieżka połączeń. Nawet, gdy istnieje połączenie, to długość łańcucha, choć średnio wynosi szesnaście, może sięgać nawet pięciuset odnośników.

Złożona topologia sieci oraz bariery utrudniające przemieszczanie się po niej oznacza, że Internet nie jest medium w pełni demokratycznym, jeśli chodzi o dostęp do informacji. W takiej strukturze, z jednej strony rośnie rola wyszukiwarek, które są alternatywą do żmudnego wędrowania w poszukiwaniu potrzebnej informacji. Z drugiej strony możemy spodziewać się, że ich działanie będzie o wiele mniej skutecznie.

Mechanizmy działania wyszukiwarek

Powszechne zapotrzebowanie na wyspecjalizowane systemy ułatwiające poruszanie się po Internecie i przeszukiwanie jego zawartości pojawiło się wraz z rozwojem World Wide Web (1991) i przeglądarek internetowych takich jak Netscape Navigator (1994). Wcześniej istniały katalogi indeksujące informacje gromadzone na poszczególnych serwerach, np. Archie. Pierwsze wyszukiwarki i katalogi WWW powstały w 1994 roku (World Wide Web Worm, Yahoo!, Lycos). Początkowo funkcjonowały przede wszystkim jako instrumenty służące do katalogowania zawartości Sieci. Gwałtowny rozwój Internetu, wzrost liczby użytkowników, szybkie poszerzanie się zasobów WWW, a także stopniowa komercjalizacja sieci przyczyniły się również do zmiany charakteru wyszukiwarek, które zaczęły przekształcać się w portale oferujące dodatkowe usługi. Proces wyszukiwania został sprowadzony do jednej z wielu funkcji wrót Internetu, jak określają swoją rolę portale.

Ewolucja wyszukiwarek przyniosła wymierne korzyści dla użytkowników. Obudowanie programów wyszukujących treścią na bieżąco aktualizowaną przez redaktorów portalu pozwoliło na uwzględnienie wiadomości i informacji, które w przeciwnym wypadku dotarłyby do odbiorcy ze znacznym opóźnieniem. Dodatkowe funkcje umożliwiające personalizację zawartości pozwoliły użytkownikowi na indywidualną selekcję otrzymywanych treści.

Rozrastanie się Internetu wymaga stałego doskonalenia systemów wyszukiwania, które nie tylko rejestrują nowe strony, ale również aktualizują dane dotyczące dokumentów znajdujących się już w bazie danych. Chociaż stale zwiększa się liczba informacji dostępnych w sieci, możliwości poznawcze użytkownika Internetu pozostają na tym samym poziomie - zazwyczaj jesteśmy w stanie przejrzeć tylko kilka pierwszych

pozycji z listy wyników wyszukiwania. O nieadekwalności tych możliwości oraz niedoskonałości wyszukiwarek może świadczyć fakt, że prawie 80% Internautów deklaruje, że frustruje ich niezdolność znalezienia potrzebnej informacji (Bergman 2001).

Programy przeszukujące Internet można podzielić na aplikacje, które korzystają ze specjalnego programu komputerowego, tzw. crawlera (zwanego również robotem lub pajakiem), oraz takie, które przedstawiają wyniki w postaci katalogów tworzonych przez redaktorów.

Crawlers generują wyniki automatycznie, ich praca polega na przeszukiwaniu milionów stron internetowych i katalogowaniu zawartości na podstawie założonych kryteriów. Podążając śladem odnośników (linków) umieszczonych na stronach przenoszą się na kolejne witryny. Crawlers regularnie odwiedzają te same strony, co pozwala na rejestrowanie zmian wprowadzanych przez autorów serwisów. Roboty analizują przede wszystkim adresy internetowe, treść nagłówków i zawartość tzw. strony głównej. Na tej podstawie zostają następnie wygenerowane "słowa-klucze", które ułatwiają segregowanie treści. Dane o ścieżkach, którymi podążają programy przeszukujące są pilnie strzeżone przez właścicieli wyszukiwarek. Dlatego trudno określić, jakie kryteria decydują o ścieżkach pajaków a także sprecyzować częstotliwość odwiedzania poszczególnych stron.

Wyszukiwarki wykorzystujące crawlers tworzą indeksy informacji zebranych przez pajaka, w których znajdują się kopie wszystkich stron przez niego odwiedzonych. Bardzo ważną funkcję pełni oprogramowanie wyszukiwarki - program, który przesiewa miliony stron zarejestrowanych w indeksie w celu znalezienia odpowiedzi na zadane zapytanie oraz uszeregowania ich w odpowiedniej kolejności. Mimo, że wszystkie wyszukiwarki oparte na crawlerach wykorzystują powyższe narzędzia, każda z nich w inny sposób ustala parametry wyszukiwania. Różnią się one liczbą skatalogowanych stron, częstością wizyt crawlera. Dlatego używając różnych wyszukiwarek otrzymujemy różne wyniki.

Katalogi tworzone przez redaktorów zawierają wybrane przez nich, "wartościowe" adresy oraz strony zgłoszone przez webmasterów. Zgłoszenie strony do katalogu nie jest równoznaczne z umieszczeniem jej w bazie danych. Kryteria przyjęte przez redaktorów nie są zazwyczaj dokładnie określone, a strona może zostać odrzucona bez podania uzasadnienia. Większe szanse na przyjęcie do katalogu mają strony, które zawierają dużą liczbę odnośników, a także te, które należą do rzadkiej kategorii (Sullivan 2001). Proces katalogowania stale się wydłuża i dziś trzeba czekać ok. 6 miesięcy na włączenie nowej strony do indeksu a ponowna aktualizacja stron jest dokonywana bardzo rzadko (Introna i Nissenbaum 2000). Indywidualne indeksowanie niesie ze sobą niebezpieczeństwo subiektywności i tendencyjności.

Obecnie stosuje się również mieszane metody archiwizowania zawartości Internetu. Taką formułę przyjmują zazwyczaj portale, które wykorzystują zarówno wyniki zebrane przez crawlers, jak i katalogi konstruowane przez redaktorów. W ten sposób działa największy portal internetowy Yahoo!, który współpracuje z wyszukiwarką Google.

Bardzo ważnym etapem procesu wyszukiwania jest przypisywanie wynikom pozycji w rankingu. Przeszukując miliony stron internetowych program sortuje wyniki odpowiednio do ich zgodności/adekwatności (*relevancy*) z zadaniem zapytaniem. Pierwsza dziesiątka adresów powinna obejmować najbardziej trafne rezultaty. Jak wskazują badania, to właśnie te 10-20 dokumentów przyciąga największą uwagę użytkownika, który bardzo rzadko sięga do wyników na dalszych pozycjach (Introna i Nissenbaum 2001). Dlatego udoskonalanie metod pozycjonowania rezultatów wyszukiwania stało się jednym z najważniejszych celów, jakie stawiają przed sobą twórcy wyszukiwarek.

W związku z tym, że programy wyszukujące nie są w stanie naprawdę zrozumieć pytania i operują zestawieniami słów-kluczy, sortowanie wyników polega na zastosowaniu określonego zestawu reguł - algorytmu pilnie chronionego przez każdą firmę. Najczęściej stosuje się kryteria lokalizacji i częstotliwości występowania danych słów-kluczy na stronie. Stąd najwyższą pozycję uzyskują dokumenty, które zawierają szukane słowo w adresie URL, w nagłówku lub w pierwszym paragrafie. Im częściej dane słowo występuje na stronie tym wyższa ocena adekwatności strony.

Bardzo istotnym kryterium, które wpływa na ocenę danej strony jest liczba odnośników, które łączą ją z innymi dokumentami, bądź też kierują użytkownika do tej strony z innych lokalizacji. Strony zawierające więcej tzw. linków są uznawane za ważniejsze i lepsze. Jedną z najprostszych metod oceny adekwatności jest popularność danej strony, czyli liczba odwiedzin. Wyniki, do których odnośniki nie generują wielu odwiedzin spadają na niższe pozycje w rankingu (Sullivan 2001).

Wiele z powyższych sposobów pozycjonowania wyników umożliwia manipulowanie zawartością strony w celu zwiększenia jej szans pojawienia się w rankingu. Wielu webmasterów projektuje swoje strony z myślą o działaniu crawlera. W pierwszym okresie działanie robotów zdarzało się tzw. "zaśmiecanie" witryny słowami-kluczami (*spamdexing*) lub tworzenie fałszywych odnośników. Próba ominięcia podobnych pułapek jest system PageRank™ opracowany przez twórców jednej z najpopularniejszych obecnie wyszukiwarek - Google. PageRank™ klasyfikuje wyniki na podstawie liczby odnośników prowadzących na stronę: link ze strony A do strony B jest interpretowany jako głosowanie strony A na stronę B. Dodatkowo Google analizuje również stronę, która oddaje głos. Głosy oddane ze stron ważnych są bardziej istotne i przyczyniają się do wzrostu znaczenia nowych stron. Ostateczny wynik przeszukiwania to połączenie zgodności z zapytaniem i oceny wystawionej witrynie przez PageRank™.

Jednak również ten system nie jest nieomylny. W styczniu 2001 jednym z wyników wyszukiwania hasła "*dumb motherfucker*" była strona poświęcona prezydentowi Bushowi. Pomyłka ta, która na krótki okres podważyła dotychczasową reputację Google, wynikała z działania PageRank™. Mimo, że ani na stronie, ani w jej kodzie HTML nie było wyrażenia "*dumb motherfucker*", adres ten znalazł się wysoko w rankingu wyników, ponieważ wyrażenie zostało użyte na stronie zawierającej odnośnik do witryny o prezydencie (Manjoo 2001).

Niewielka wyszukiwarka Yep.com ocenia jakość i popularność stron na podstawie danych zebranych przez program HitBox, który analizuje częstotliwość umieszczania strony w zakładkach, ilość czasu spędzanego na stronie i powracalność Internautów. W przeciwieństwie do Google adekwatność nie jest dla Yep.com kluczowym kryterium. W konsekwencji całkowitego zautomatyzowania kategoryzacji wyników Yep.com w ogóle nie uwzględnia zawartości stron, a jedynie ich popularność. Np. po wpisaniu hasła "Żyd" okazuje się, że strona Muzeum Żydowskiego została oceniona niżej niż antysemitcka witryna Global Domination (Chaudhry 2000).

Bardzo wysoka atrakcyjność stron znajdujących się na pierwszych miejscach wyników wyszukiwania zachęcała twórców wyszukiwarek do wprowadzania możliwości wykupienia pozycji adresu na liście bądź konkretnych słów-kluczy. Wyszukiwarki, jako przedsięwzięcia komercyjne, muszą poszukiwać źródeł zysku. Dlatego większość z nich oferuje banery reklamowe połączone ze słowami kluczami, promuje strony własne bądź opłacone przez sponsora, a także wprowadza sponsorowane treści umieszczone w obrębie listy wyników. Coraz częściej jednak zdarza się, że małe firmy rezygnują całkowicie z własnego katalogowania stron i w zamian przedstawiają wyniki wykupione przez właścicieli stron (*paid placement links*). Na stronach takich wyszukiwarek jak GoTo.com, iWon i Canada.com pierwszeństwo w rankingu mają adresy sponsorowane. Wpisując w

wyszukiwarce Google hasło Canada otrzymamy na pierwszym miejscu odnośnik do oficjalnej strony rządu Kanady, natomiast Canada.com przedstawi nam oferty turystyczne na terenie Kanady (Sullivan 2001). Serwis GoTo zawiera katalog jedynie opłaconych odnośników, a pozycje rankingowe są wystawiane na licytację. Reklamodawcy wykupują wysokie miejsca poprzez przebijanie oferty konkurentów, a GoTo pobiera opłaty tylko wtedy, gdy użytkownik odwiedzi daną stronę. (Bayers 2001).

Takie praktyki nie tylko przyczyniają się do spadku zaufania użytkowników do danej wyszukiwarki, lecz również wprowadzają w błąd osoby, które nie są świadome, że dane wyniki nie pokazują prawdziwej zawartości Sieci. Podobną opinię podziela grupa konsumentów, którzy w lipcu 2001 złożyli przed Federalną Komisją Handlu oficjalny protest, w którym oskarżają 8 wyszukiwarek¹ o łamanie zakazu umieszczania niejawnych reklam (*deceptive advertising*). Organizacja Commercial Alert domaga się wyraźnego oddzielenia treści sponsorowanych od zawartości katalogów i karania wyszukiwarek, które przemycają wśród wyników zawołowane reklamy (Lyman 2001). Jak pokazuje *Search Engines Strategies Report 2001* większość wyszukiwarek wprowadziła do swoich usług ofertę *paid placement*. Do wyjątków, które oparły się temu trendowi należy zaliczyć Open Directory Project i Northern Light. Próbę wyjścia ze ślepej uliczki płatnych rankingów podjęła również wyszukiwarka Google, która wprowadziła system, który na pierwszym miejscu stawia efektywność wyszukiwania. Google nie sprzedaje pozycji PageRank™, lecz umieszcza sponsorowane linki w zależności od ich adekwatności. Im lepsza strona, tym wyższa cena.

Badanie opublikowane w dzienniku *Nature* wykazało, że wyszukiwarki nie docierają do wszystkich zasobów Internetu. Żadna z nich nie indeksuje więcej niż 16% zawartości Sieci a wszystkie razem rejestrują ok. 42% dokumentów. Trzy największe obecnie wyszukiwarki obejmują swym zasięgiem odpowiednio – Google 1,35 miliarda, Fast 575 milionów i Northern Light 327 milionów dokumentów (Bergman 2001). Wyszukiwarki mają tendencję do rejestrowania stron z większą liczbą odnośników i częściej odsyłają do treści dokumentów w Stanach Zjednoczonych.

Głęboka sieć

Istnieją również obszary, które ze względu na specyficzną architekturę są całkowicie niedostępne dla crawlerów. To tzw. “niewidzialna” albo “głęboka sieć” - strony z założenia “niewidoczne” dla konwencjonalnych wyszukiwarek. Termin ten nie obejmuje stron, które potencjalnie mogłyby być skatalogowane, choć nie są.

Sieć WWW składała się początkowo wyłącznie z tzw. stron statycznych (*static pages*) o stosunkowo stałej zawartości i stabilnej strukturze katalogów. Obecnie ogromny zbiór stron statycznych tworzy tzw. sieć powierzchniową (*surface web*), potencjalnie dostępną dla crawlerów (Bergman 2001). Rozwój sieci doprowadził do powstania stron generowanych dynamicznie (*dynamic pages*). Na taką stronę składa się baza danych, z której skrypt generujący stronę wyświetla jedynie wyselekcjonowane informacje. Taki model pozwala, bez strat w wygodzie użytkownika, oszczędzić przestrzeń na serwerach.

Głęboka Sieć jest niewidzialna, gdyż crawlery nie radzą sobie z dynamicznym charakterem tworzących ją stron. Są to witryny, które wyświetlają daną stronę w odpowiedzi na zapytanie zadane przez Internautę wewnętrznej wyszukiwarce, umieszczonej zazwyczaj na stronie głównej witryny. Natrafiając na taką stronę crawler kataloguje stronę główną, będącą jedynie interfejsem. Dane zapisane w bazie danych są dla niego niewidzialne. Przykładami takich stron są wirtualne sklepy, bazy wyników badań naukowych, czy też - paradoksalnie - same wyszukiwarki, które nie potrafią skatalogować zawartości własnego indeksu. Tak, więc do

¹ Są to: Alta Vista, AOL Time Warner i Netscape searcher, Microsoft i MSN search, Direct Hit Technologies, iWon Inc., LookSmart, Terra Lycos

Głębokiej Sieci możemy też zaliczyć wyspecjalizowane, prawie nieznane wyszukiwarki, które indeksują treść pominiętą przez większe wyszukiwarki. Ich liczbę szacuje się na nawet kilkadziesiąt tysięcy (Bergman 2001).

Według badań zawartości Głębokiej Sieci, są w niej równomiernie reprezentowane różnorodne tematy, informacja ma wyższą jakość niż w sieci powierzchniowej i wykazuje szybsze tempo przyrostu (Bergman 2001). Według szacunkowych obliczeń Głęboka Sieć zawiera ponad 500 razy więcej pojedynczych dokumentów i prawie 400 razy więcej megabajtów danych niż powierzchniowy Internet, umieszczonych w ponad 200 tysiącach baz danych (Bergman 2001).

Obecnie istnieją już wyspecjalizowane katalogi witryn Głębokiej Sieci, a nawet wyszukiwarki pozwalające przeszukiwać na raz zawartość wielu "ukrytych" baz danych. Jednak część z nich jest płatna, a wszystkie są mało znane wśród Internautów. Musimy wziąć pod uwagę, że - jak pokazują badania - większość Internautów używa tylko jednej wyszukiwarki, wychodząc z błędnego założenia, że wszystkie indeksują tę samą treść (Sullivan 2000). W praktyce Głęboka Sieć jest w znacząco mniejszym stopniu dostępna dla przeciętnego Internauty niż sieć powierzchniowa.

Najpopularniejsze wyszukiwarki powoli zaczynają doceniać wartość zawartych w Głębokiej Sieci informacji. W tych staraniach prowadzi Google, który skupia się na poszerzeniu swoich indeksów o formaty danych wcześniej nie uwzględniane przez crawlery indeksujące niemal wyłącznie dokumenty HTML. W zeszłym roku na stronach Google'a pojawiła się wyszukiwarka wiadomości z sieci Usenet, a podstawowe oprogramowanie zostało rozbudowane o mechanizm wyszukujący pliki PDF. W najbliższym czasie Google zindeksuje także pliki audio i video dostępne w sieci.

Wnioski

Dla wielu analityków mediów powstanie Internetu symbolizowało nadejście nowej ery w komunikacji masowej i zapowiadało rewolucyjne przemiany w dotychczasowych relacjach pomiędzy nadawcą a odbiorcą. Dzięki technicznym możliwościom Sieci znacznie obniżyły się koszty wyprodukowania i wyemitowania przekazu. Zniknęły również bariery, które utrudniały odbiorcy skomentowanie i publiczną ocenę docierających do niego treści. W Sieci każdy odbiorca jest potencjalnym nadawcą i swym własnym wydawcą.

Na przekór opinii McLuhana, że *"Wszystkie media istnieją po to, aby skłaniać nas do sztucznego postrzegania świata i narzucać nam arbitralne wartości"* (McLuhan 1964) oczekiwano, że nowe medium, które charakteryzuje się otwartością i zdecentralizowaniem, nie będzie potrzebować tzw. dozorców, czyli *gate-keeperów*, którzy tradycyjnie pełnili funkcje selekcyjnych treści. Pojęcie *gate-keeper*, wprowadzone w roku 1950 przez White'a (Goban-Klas 2000), określa aktora, który na pewnym etapie procesu komunikacji dokonuje selekcji emitowanych idei i informacji. Jak pisze Paul Levinson, media takie jak prasa, radio, telewizja pełnią rolę dozorców, przede wszystkim ze względu na ograniczenia techniczne - wyraźne limity czasu emisji przeznaczonego na dany program czy też ilości stron poświęconych danej tematyce. Autorzy, którym zależy na upublicznieniu własnych wytworów muszą, zatem zabiegać o względy redaktorów i wydawców, którzy decydują o tym, które materiały zostaną opublikowane. Także odbiorca jest skazany na arbitralne wybory nadawców - może jedynie wybierać pomiędzy ograniczoną liczbą dostępnych tytułów.

Internet zlikwidował te ograniczenia, umożliwiając całkowitą swobodę wejścia i wyjścia w przestrzeni komunikacyjnej. Oczywiście, przy założeniu, że dysponuje się dostępem do Sieci. Zniknęło, zatem zapotrzebowanie na pośredników, którzy umożliwiali wcześniej transmisję treści. Jednak łatwość dostępu do kanału komunikacji, jakim jest Internet, zaowocowała nadmiarem publikowanych treści, co, przy braku selekcji i kontroli jakości, przyczyniło się do zaśmiecenia Sieci. Poruszanie się po przestrzeni Internetu zależy nie tylko od znajomości obsługi komputera i oprogramowania, lecz również od opanowania dużego zasobu wiedzy o architekturze Internetu i sposobach nawigowania po jego zawartości. O ile na wcześniejszym etapie rozwoju

mediów to autor treści był zmuszony do poszukiwania środków przekazu, które nagłośniłyby jego wytwory, o tyle w erze Internetu to odbiorca spotyka się z trudnościami z powodu zalewu informacji. Niewiele osób posiada szczegółową wiedzę na temat adresów URL wszystkich interesujących go stron a większość początkujących Internautów nie rozumie nawet zasad, które określają tworzenie tych adresów. Sama znajomość poszczególnych wyszukiwarek, ich zalet i różnic jest kapitałem, który wpływa znacząco na skuteczność wyszukiwania i końcowe zadowoleniu z jego wyników. Może on prowadzić do znaczących nierówności w wykorzystaniu Internetu.

Paul Levinson twierdzi, że w sieci nie grozi nam przeładowanie informacją, ponieważ mamy pomocników - *“...odpowiednie struktury ułatwiające poszukiwanie, liczne programy, czasem widoczne, czasem działające pod powierzchnią, które służą wyszukiwaniu, sortowaniu, łączeniu, magazynowaniu, odtwarzaniu, cytowaniu i monitowaniu tekstu umieszczonego w sieci”* (Levinson 1999, 208). Rzeczywiście, w teorii programy wyszukiwujące powinny spełniać funkcję przejrzystego katalogu lub “przyjaznego bibliotekarza”, który pozwala nam dotrzeć do przepastnych zasobów sieci. Jednak w pracy tej stawiamy tezę, że wyszukiwarki i portale także pełnią rolę *gate-keeperów*, dozorców. Nastąpiło jedynie przeniesienie punktu ciężkości ich działania – dziś internetowi selekcyonerzy nie decydują już o tym, co w ogóle znajdzie się w sieci, ale skupiają się na skierowaniu uwagi odbiorcy na konkretne treści (Hargittai 2000).

Według badań z 1999 roku 0.5% wszystkich stron skupia 80% ruchu Internautów, 1000 stron jest odpowiedzialnych za 55% tego ruchu, a jedno z każdych pięciu odwiedzin trafia na strony siedmiu największych firm internetowych: Yahoo, Microsoft, Excite, eBay, CMGI, Disney i AOL (Waxman 2000). Zaskakująco mała jest liczba stron odwiedzanych w ciągu miesiąca przez przeciętnego użytkownika - wzrosła ona z 14 w lipcu 1997 do 25 w grudniu 1999. Autor cytowanego badania wyraża zdziwienie, że *“w obecności milionów stron oraz wielu różnych wyszukiwarek użytkownicy ograniczają swoją uwagę do tak niewielu witryn. Być może użytkownicy przejawiają przywiązanie z powodów poznawczych lub z powodu wierności danej domenię”* (Montgomery i Faloustos 2000). Koncentracja ruchu internetowego w obrębie niewielkiej liczby stron oraz dominacja kilku potężnych witryn świadczą o tym, że Internet ma już cechy “push medium” (McQuail i Siune 1998).

Najpopularniejsze strony internetowe to w dużej mierze kompleksowe portale oraz wyszukiwarki (czasem zintegrowane z portalami). Strategia tych pierwszych opiera się na utrzymaniu Internauty w przestrzeni własnej witryny i dążeniu do wyłączności w dostarczaniu Internaucie informacji. Wyszukiwarki z kolei mają na celu umożliwienie jak najszybszego dotarcia do innych stron – więc to one tak naprawdę kształtują obraz Internetu, który dociera do użytkownika.

Paradoks wynikający z ich funkcjonowania polega na tym, że bez nich wędrówka po sieci byłaby praktycznie niemożliwa, ale korzystanie z nich pociąga za sobą przyjęcie arbitralnej wizji przestrzeni Internetu. Subiektywna selekcja rozpoczyna się już na etapie pracy robotów, które poruszają się trasami wyznaczonymi przez tajne algorytmy, różne dla każdego systemu wyszukiwającego. Algorytmy wyszukiwarek mają wskazywać strony o największej adekwatności i wiarygodności, a nie prezentować szeroki wachlarz różnorodnych źródeł informacji. Żadne obiektywne kryteria nie decydują też o wyborach dokonywanych przez redaktorów konstruujących katalogi. Na ostateczny wynik wyszukiwania często wpływa także strategia marketingowa danej firmy i oczekiwania reklamodawców.

Tak, więc wyszukiwarki nie są neutralnymi narzędziami. Jak każdy zmysł, ich percepcja jest selektywna. W środowisku Internetu władza wyszukiwarek – podkreślanie znaczenia jednych stron, a osłabianie innych – jest znacząca. Można, więc mówić o polityce i ukrytej ideologii działania wyszukiwarek, które

uwzględniają przede wszystkim kryteria rentowności przedsięwzięcia i adekwatności wskazywanego źródła informacji, a w mniejszym stopniu realizują ideały doskonałego wyszukiwania i prezentacji różnorodnych informacji.

Polityka wyszukiwarek sprzyja przede wszystkim właścicielom dużych i znanych witryn internetowych. *“Wyszukiwarki systematycznie podkreślają istnienie komercyjnych stron WWW, witryn mogących przyciągnąć do gustu szerokiego gronu użytkowników, a także stron wspieranych przez silnych aktorów ekonomicznych”* (Introna i Nissenbaum 2000). Trend ten, prowadząc do ograniczenia zawartości Internetu – faktycznej lub dostępnej Internaucie – podważa wizję Sieci jako dobra publicznego.

Na polityce wyszukiwarek tracą przede wszystkim użytkownicy, którzy zostają pozbawieni wiedzy o istnieniu innych, mniej komercyjnych i często bardziej wartościowych zasobów. Tracą też właściciele stron, które z powodu skrzywienia mechanizmu wyszukiwania i wskazywania stron nie przyciągają uwagi Internautów.

Utopijnym wizjom Internetu przeczy sama jego struktura: bardzo złożona i nie sprzyjająca przepływowi informacji. W tej sytuacji wyszukiwarki są aktorami, którzy potencjalnie mogą zapewnić Internautom swobodny dostęp do różnorodnych źródeł informacji. Jednak w chwili obecnej czynniki ekonomiczne, ciągle niedoskonałe metody indeksowania oraz rozwiązania techniczne powodują, że wyszukiwarki wzmacniają trend przekształcający Internet w “push medium”, w którym *gate-keeperzy* zamiast pełnić funkcję pośrednika raczej zakłócają swobodny dostęp do informacji.

Bibliografia

- Albert, R., H. Jeong i A.-L. Barabasi, 1999, *Diameter of the World Wide Web*, Nature, Vol. 401, 9 September 1999, s. 130
- Bayers, Ch. 2001 *I'm Feeling Lucky* Wired 10, 2001 <http://www.wired.com/wired/archive/9.10/google.html>
- Bergman, M. K., 2001, *Deep Content. White paper*,
<http://www.brightplanet.com/deepcontent/BPTechnologyWhitePaper.pdf>
- Brin, S. i P. Lawrence, 1998 *The Anatomy of a Large-Scale Hypertextual Web Search Engine*,
<http://www7.scu.edu.au/programme/fullpapers/1921/com1921.htm>
- Broder, A., R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, Janet Wiener, 2000, *Graph structure in the web. Proceedings of the 9th International World Wide Web Conference (WWW9)*, <http://www.almaden.ibm.com/cs/k53/www9.final/>
- Castells, M., 2001, *The Internet Galaxy*, Oxford University Press, Oxford
- Chaudry, L. 2000 *Online Hate Gets Five Stars* Wired News 25.05.2000,
<http://www.wired.com/news/politics/0,1283,36499,00.html>
- DiMaggio, P., E. Hargittai, W. R. Neuman i J. P. Robinson, 2001, *Social Implications of the Internet*, "Annual Review of Sociology", 27: 307-36
- Elliot, K. 2001 *Search Engine Strategies 2001 Report*,
<http://websearch.about.com/library/weekly/aa082101a.htm>
- Goban-Klas, T., 2000, *Media i komunikowanie masowe. Teorie i analizy prasy, radia, telewizji i Internetu*, Wydawnictwo Naukowe PWN
- Introna, Lucas D. i H. Nissenbaum, *Shaping the Web: Why the politics of search engines matters*, "The Information Society", nr. 16 (3)
- Kleinberg, J. i S. Lawrence, 2001, *The Structure of the Web*, Science, Vol. 294, 30 November 2001, s. 1849-50
- Lawrence, S., i Giles, L. 1999, *Accessibility of information on the Web*, Nature, Vol. 400. s.107-109
- Levinson, P., 1999 *Miękkie ostrze*, MUZA S.A., Warszawa
- Lyman, J., 2001 *Group Charges Search Engines Results Deceptive* NewsFactor Network,
<http://www.newsfactor.com/perl/printer/12047>
- Manjoo, F., 2001 *Google Link is Bush League* Wired News 25.01.2001,
<http://www.wired.com/news/technology/0,1282,41401,00.html>
- McLuhan, M., 1964 *Understanding Media: The Extensions of Man*, McGraw-Hill, New York
----- 1995 *Wybór tekstów* Red. Eric McLuhan, Frank Zingrone, Zysk i S-ka, Warszawa 2001
- McQuail, D. i K. Siune, *Media Policy. Convergence, Concentration & Commerce*, Sage Publications, London
- Montgomery, A. i C. Faloustos, 2000, *Trends and Patterns of WWW Browsing Behavior*,
<http://www.andrew.cmu.edu/user/alm3/papers/web%20trends.pdf>
- Morris, M. i C. Ogan, 1996, *The Internet as Mass Medium*, Journal of Computer-Mediated Communication, Vol. 1, Nr. 4, <http://www.ascusc.org/jcmc/vol1/issue4/vol1no4.html>
- Nowak, M. *Wyszukiwanie w Internecie* Chip 2/2002
- Sullivan, D. 2001, *How Search Engines Rank Web Pages*, <http://searchenginewatch.com/webmasters/rank.html>
-----*Pay for Placement*, <http://searchenginewatch.com/resources/paid-listings.html>
-----*Can Portals Resist the Dark Side?*, <http://searchenginewatch.com/sereport/01/06-darkside.html>
-----2000, *Search Satisfaction and Behavior Results Released*,
<http://www.searchenginewatch.com/sereport/00/04-npd.html>
-----2001, *The Search Engine Report*, www.searchenginewatch.com
- Waxman, J., 2000, *Internet Trends Report 1999 Review*, Alexa Research,
<http://www.startupdna.com/reports/ALEXA-trendsreport4Q99.pdf>