PRZEGLĄD KULTUROZNAWCZY
NR 1 (19) 2014, s. 44–55
doi: 10.4467/20843860PK.13.005.2854
www.ejournals.eu/Przeglad-Kulturoznawczy/

Magdalena Szpunar
Akademia Górniczo-Hutnicza w Krakowie

Sieć ukryta a sieć widzialna. O zasobach WWW nieindeksowanych przez wyszukiwarki

Invisible *vs.* Visible Web. On Non-indexed WWW Resources

Abstract: With the use the Internet for information search and especially the widespread use for this purpose search tools are related two stereotypes. The first concerns the belief that search for content online is an act extremely simple, trivial, requiring no special skills. The second myth refers to the same search engines – treated as instruments, allowing you to reach to any information that is online. These simplified and stereotyped beliefs are not limited to the realm of declarative, but manifest themselves in concrete expression of the information through the Internet behavior. The purpose of the present article is to overcome common stereotypes associated with the use of search tools, but also to show the hidden part of the Internet.

Key words: hidden web, deep web, surface web, visible web, the principle of least effort, Google hegemony, non-indexed resources

Z użytkowaniem internetu[1] w celach informacyjnych, a szczególnie powszechnym wykorzystywaniem w tym celu narzędzi wyszukiwawczych, związane są dwa stereotypy. Pierwszy z nich dotyczy przekonania, iż wyszukiwanie treści online jest czynnością niezwykle prostą, trywialną, niewymagającą specjalnych umiejętności. Drugi mit odnosi się do samych wyszukiwarek – traktowanych jako omnipotencyjne instrumentarium, pozwalające na dotarcie do każdej informacji, która znajduje się online. Te uproszczone i szablonowe przekonania nie ograniczają się jedynie do sfery deklaratywnej, ale objawiają się w konkretnych, przejawianych przez internautów zachowaniach informacyjnych. O kluczowej roli wyszukiwarek w procesie wyszukiwania świadczy zakorzenienie w języku angielskim takich terminów jak *googling* i *to google*, a w języku polskim guglować[2] czy wyguglować,

[1] Według rekomendacji Rady Języka Polskiego termin „internet" można pisać zarówno małą, jak i wielką literą. Wielką literę stosujemy wtedy, gdy mamy na myśli konkretną sieć globalną, pojęcie to funkcjonuje tutaj jako nazwa własna – takie rozróżnienie pojawia się na przykład w *Wielkim słowniku ortograficznym języka polskiego* pod redakcją Andrzeja Markowskiego (1999). Internet pisany małą literą sugeruje rozumienie sieci jako medium, a nie jako systemu (łączności komputerowej). W niniejszej pracy zastosowano pisownię zgodną z drugim modelem rozumienia tego pojęcia.

[2] Jak podaje Randall Stross, już po pięciu latach funkcjonowania firmy Google termin *google* został oficjalnie uznany przez *American Dialect Society* – zob. R. Stross, *Planeta Google*, przeł. A. Wojtaszczyk, O. Wojtaszczyk, Studio EMKA, Warszawa 2009, s. 9. W internetowej poradni językowej afiliowanej przez PWN postuluje

traktowanych jako synonimiczne wobec procesu przeczesywania sieci. Wiele badań wskazuje, że odpowiedzialność za jakość i wiarygodność znalezionej informacji online użytkownicy transmitują na narzędzia wyszukiwawcze, którymi się posiłkują[3]. Mechanizm ten Evgeny Morozov określa mianem solucjonizmu[4] – jesteśmy usatysfakcjonowani tym, że technologia informacyjna umożliwia nam szybki dostęp do informacji, co implikuje przekonanie, że narzędzia te (w tym wyszukiwarki) są dobre same w sobie. Co ciekawe, wiele badań pokazuje, że użytkownicy, mimo świadomości innych źródeł informacji online, z reguły ograniczają się do wyszukiwarek, traktując je jako satysfakcjonujące i adekwatne do ich potrzeb[5]. Zwraca się uwagę, iż zachowaniami informacyjnymi w sieci kieruje zasada najmniejszego wysiłku (*the principle of least effort*)[6], która powoduje, że użytkownicy internetu z łatwością obniżają swoje wymagania, by móc szybko i w sposób nieskomplikowany dotrzeć do interesującej treści[7]. Zgodnie z zasadą minimaksu internauta chce jak najszybciej i jak najniższym kosztem osiągnąć zamierzony przez siebie rezultat, co powoduje, że mimowolnie obniża on standardy, zadowalając się ofertą wyszukiwarki, którą obdarza zaufaniem, implikując sytuację, w której odpowiedzialność za to, jakie treści i w jakiej formie są jej dostarczane, jest przenoszona z jednostki na narzędzie, którym się posługuje. Niestety to, co indeksują wyszukiwarki, a przez to to, co *de facto* jest dostępne online, stanowi jedynie niewielki odsetek treści, które znajdują się w przestrzeni WWW. Pole uwagi[8] jednostki jest zatem mocno ograniczone do tego, co zaproponuje jej narzędzie stanowiące dla wielu internautów jedyną możliwą bramę wejścia do internetowego świata.

Kluczowa rola wyszukiwarek w naszej codzienności sprawia, że uprawnione jest określanie ich za Jayem Bolterem jako technologii definiujących. Przypomnijmy, iż: „Technologia definiująca definiuje lub redefiniuje rolę człowieka w odniesieniu do przyrody. Obiecując zastąpienie człowieka (lub grożąc nim), komputer podsuwa nam nową definicję człowieka jako procesora informacji”[9]. Owa technologia definiująca, której mianem proponuję określać wyszukiwarki, kształtuje nasz sposób myślenia i postrzegania świata. Wyszukiwarki

się, aby termin „guglować” pisać fonetycznie, czyli tak, jak go słyszymy; zob. http://poradnia.pwn.pl/lista.php?id=8070 (dostęp: 9.11.2013).

3 B. Pan, H. Hembrooke, T. Joachmis *et al.*, *In Google We Trust: Users' Decisions on Rank, Position and Relevance*, „Journal of Computer-Mediated Communication” 2007, no. 12.

4 E. Morozov, *To Save Everything, Click Here: The Folly of Technological Solutionism*, Penguin Group, London 2013.

5 J. Brophy, D. Bawden, *Is Google Enough? Comparison of an Internet Search Engine with Academic Library Resources*, „Aslib Proceedings: New Information Perspectives” 2005, vol. 57, s. 498–512.

6 J. Griffiths, P. Brophy, *Student Searching Behaviour and the Web: Use of Academic Resources and Google*, „Library Trends” 2005, vol. 53, s. 539–554.

7 D. Nicholas, T. Dobrowolski, R. Withey *et al.*, *Digital Information Consumers, Players and Purchasers: Information Seeking Behavior in the New Digital Interactive Environment*, „Aslib Proceedings: New Information Perspectives” 2003, vol. 55, s. 23–31.

8 Problematykę związaną z ograniczeniami pola uwagowego szeroko omawiam w artykule *Bogactwo informacji a ubóstwo uwagi w cyfrowej kulturze nadmiaru*, [w:] M. Kaczmarczyk, D. Rott (red.), *Problemy konwergencji mediów*, Verbum, Praga 2013, s. 293–302.

9 J.D. Bolter, *Człowiek Turinga. Kultura Zachodu w wieku komputera*, przeł. T. Goban-Klas, PIW, Warszawa 1990, s. 43.

stają się technikami autorytarnymi, o których jeszcze w latach 60. XX wieku pisał Lewis Mumford[10], centralizując władzę i kontrolę, stają się formą władzy samej w sobie. Przypominają Innisowski bias[11], o którym kanadyjski badacz pisał: „używanie przez dłuższy czas jakiegoś środka komunikacji określa w pewnej mierze kształt przekazywanej wiedzy, a gdy jego oddziaływanie staje się dominujące, prowadzi w końcu do stworzenia cywilizacji, która z upływem czasu z coraz większym trudem zachowuje żywotność i elastyczność, aż pojawi się nowe medium, o nowych możliwościach, które da początek nowej cywilizacji"[12]. Harold Innis wskazywał, iż dominująca technologia pełni rolę kluczową, gdyż stanowi podstawę wszystkich procesów społeczno-politycznych, a każdy nowy środek komunikowania klasa rządząca wykorzystuje w procesach dystrybucji wiedzy. Zatem dominujące medium nie tylko umożliwia transmisję i utrwalanie informacji, ale znacznie więcej – modyfikuje istniejące w danym społeczeństwie systemy wiedzy.

Jak trafnie skonstatował teoretyk mediów Marshall McLuhan, każdy wynalazek ma „dwojaką naturę – jest zarówno dobrodziejstwem, jak i przekleństwem"[13], zatem technologia (w tym i internetowa), dając nam coś, niewątpliwie równie wiele odbiera, co znajduje swoje odzwierciedlenie chociażby w tetradzie praw mediów (*laws of media*)[14], sformułowanej przez kanadyjskiego badacza. McLuhanowskie prawa mediów uświadamiają nam, że wynalazki techniczne nie są wobec nas neutralne, przekształcają bowiem swoich użytkowników[15]. Podobne konstatacje odnajdziemy u Neila Postmana, który jeszcze silniej akcen-

---

10 L. Mumford, *Authoritarian and Democratic Technics*, „Technology and Culture" 1964, vol. 5, s. 1–8.

11 Bias należy rozumieć jako nastawienie, nachylenie określonych wartości, przekonań, dokonujące się pod wpływem kluczowego medium. Twórca tego terminu Harold Innis uważał, iż bias tworzy dominującą metafizykę danej epoki. W polskich tłumaczeniach termin ten funkcjonuje jako nastawienie, nachylenie czy skłonność.

12 H. Innis, *Nachylenie komunikacyjne*, „Communicare. Almanach Antropologiczny. Oralność/Piśmienność" 2007, s. 10.

13 K. Loska, *Dziedzictwo McLuhana – między nowoczesnością a ponowoczesnością*, Rabid, Kraków 2001, s. 103.

14 McLuhanowska tetrada praw mediów sprowadza się do czterech pytań, które można zadać wobec każdego artefaktu, w tym i technologii internetowej: 1. Co dany artefakt wzmacnia, nasila, umożliwia bądź przyspiesza? 2. Jeśli pewien aspekt sytuacji powiększa się bądź wzmaga, to tym samym zanikają dawne warunki lub sytuacja niewzmożona. Cóż zatem zostaje odrzucone albo zanika za sprawą nowego „organu"? 3. Jakie wcześniejsze działania i pomoce powracają bądź pojawiają się ponownie za sprawą nowej formy? Jaka dawna podstawa, która wcześniej zanikła, zostaje przywrócona i zawiera się w nowej formie? 4. Kiedy się ją rozciągnie do granic możliwości (kolejne działanie komplementarne), nowa forma będzie odwracała swoje pierwotne cechy. W jakim stopniu zatem nowa forma może się odwrócić? Wyjaśnienia mechanizmu działania owych praw eksplikuje McLuhan, odwołując się do telefonu. Narzędzie to jego zdaniem wzmacnia: komunikację interpersonalną, dostępność, czas reakcji, sprzyja zanikaniu prywatności, anonimowości i budek telefonicznych, pozwoliło odzyskać: kulturę plemienną, przestrzeń akustyczną; odwrócenie wprowadziło poprzez przywiązanie (telefon jako smycz). Samochód z kolei zdaniem Kanadyjczyka wzmocnił naszą szybkość, przyczynił się do zaniku powozów, pozwolił „odzyskać" epokę rycerską, a odwrócenie oznacza korki uliczne. Zob. M. McLuhan, E. McLuhan, *Laws of Media*, The New Science, Toronto 1988, s. 168–171. Tetradę praw mediów w odniesieniu do internetu można znaleźć w publikacji autorki: *Nowe-stare medium. Internet między tworzeniem nowych modeli komunikacyjnych a reprodukowaniem schematów komunikowania masowego*, IFiS PAN, Warszawa 2012, s. 41–44.

15 Por. M. McLuhan, *Wybór tekstów*, przeł. E. Różalska, J. Stokłosa, Zysk i S-ka, Poznań 2001, s. 547.

tuje deterministyczną rolę narzędzi, którymi przychodzi nam się posługiwać: „w każdym narzędziu tkwią pewne założenia ideologiczne, pewna predyspozycja do konstruowania świata takiego raczej niż innego, cenienia jednej rzeczy bardziej niż innej, wzmacniania jednego znaczenia, jednej zdolności, jednej postawy bardziej niż innej”[16]. Postmanowski technopol ukazuje triumf techniki nad kulturą i choć rozwiązanie techniczne – w tym przypadku wyszukiwarki internetowe – są narzędziem wytworzonym przez ludzi, koncepcja deterministyczna nakazuje je postrzegać jako instrumentarium autonomiczne, niezależne od woli jednostki. Jak zauważa bowiem Halavais: „Niektóre witryny przyciągają więcej uwagi niż inne i z pewnością nie dzieje się tak przez przypadek. Wyszukiwarki nie tylko przyczyniają się do selekcji bardziej znaczących witryn, lecz także znajdują się pod ich wpływem”[17]. Ów wpływ to poddanie się dyktatowi rynku i komercjalizacji[18], o które jeszcze do niedawna nikt internetowego giganta nie posądzał. Wydaje się jednak, że w tym przypadku realizuje się model biznesowy: darmowa usługa – komercyjna (korporacyjna) kontrola. Wyszukiwarki stają się współczesnymi gatekeeperami[19], wytwarzając iluzję niczym nieograniczonego wyboru. Wolność w internecie ma charakter pozorny i jak trafnie konstatuje Wojciech Orliński: „Dopiero internet spełnił odwieczne marzenie cenzorów o cenzurze tak doskonałej, że odbiorcy nie są świadomi jej istnienia”[20].

Głęboka sieć (*Deep Web*), zwana inaczej siecią ukrytą (*Hidden Web*) lub siecią niewidzialną (*Invisible Web*), stanowi tę część World Wide Web, która nie jest indeksowana przez standardowe wyszukiwarki. Po raz pierwszy termin *sieć niewidzialna* został użyty już w 1994 roku – a więc znacznie wcześniej, nim internet wszedł w etap upowszechnienia – przez Jilla Ellswortha dla opisu informacji niewidocznych dla konwencjonalnych wyszukiwarek. Przez niewidoczność należy rozumieć sytuację, w której wyszukiwarka ma dostęp do kodu danej strony, ale nie potrafi go zinterpretować. W 1996 roku pojęcie to pojawiło się także w notce prasowej Bruce’a Mounta i Matthew B. Kolla[21]. Jednakże do powszech-

---

16 N. Postman, *Technopol. Triumf techniki nad kulturą*, przeł. A. Tanalska-Dulęba, Muza, Warszawa 2004, s. 26.

17 A. Halavais, *Wyszukiwarki internetowe a społeczeństwo*, przeł. T. Płudowski, Wydawnictwo Naukowe PWN, Warszawa 2012, s. 81.

18 Warto wskazać, że w 2006 roku dochód z reklam Google’a wyniósł 10,5 miliarda dolarów, podaję za: L. Gorman, D. McLean, *Media i społeczeństwo. Wprowadzenie historyczne*, przeł. A. Sadza, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków 2010, s. 291. Halavais pisze wprost: „producenci komercyjni sprawują niepodważalną kontrolę nad uwagą odbiorców. Patrząc na listę najpopularniejszych witryn, łatwo dojść do wniosku, że Sieć należy do sfery przedsięwzięć komercyjnych”, za: *idem*, *Wyszukiwarki internetowe a społeczeństwo*, *op. cit.*, s. 103.

19 Mechanizm ten określam jako gatekeeping technologiczny – szeroko na ten temat piszę w artykule: *Wokół koncepcji gatekeepingu. Od gatekeepingu tradycyjnego do technologicznego*, [w:] I.S. Fiut (red.), *Człowiek w komunikacji i kulturze*, Wydawnictwo Aureus, Kraków 2013, s. 52–61. Klasyczna w medioznawstwie koncepcja gatekeepingu została opracowana w 1943 roku przez Kurta Lewina. Gatekeeper pełni rolę selekcjonera, decydując o tym, jakie informacje i w jakiej formie (filtrowanie) będą rozpowszechniane. Poprzez decyzje o tym, jakie informacje znajdą się w obiegu, a jakie nie, gatekeeperzy mogą realnie kontrolować poziom wiedzy w społeczeństwie; por. K. Lewin, *Frontiers in Group Dynamics*, „Human Relations” 1947, vol. 1, no. 2.

20 W. Orliński, *Internet. Czas się bać*, Agora SA, Warszawa 2013, s. 80.

21 B. Mount, M. Koll, *PLS introduces AT1, the first ‘second generation’ Internet search service*, 1996, http://web.archive.org/web/19971021232057/www.pls.com/news/pr961212_at1.html (dostęp: 7.11.2013).

nego obiegu określenie to wprowadził i spopularyzował w 2001 roku Mike Bergman, autor przełomowej pracy opublikowanej w „Journal of Electronic Publishing”[22]. Wyszukiwanie informacji w sieci WWW[23] porównał on do przeciągania siatki na powierzchni oceanu, zwracając uwagę, iż wiele treści może być w sieci złowionych, ale równie wiele z nich nie udaje się wydobyć. Przekonanie o nieograniczonym dostępie do informacji online jest złudne. Wyszukiwarki, którym przypisywano swobodę wyboru, realnie ograniczają naszą samodzielność, wyobraźnię i możliwość decydowania. Wyszukiwarki *de facto* stają się nieformalnymi autorytetami, tworem, który „sprawnie i spolegliwie informując, decyduje o zachowaniu wielu osób”[24]. Skoro zatem narzędzia te działają sprawnie, a jak pokazują badania, są przez korzystających przez nich użytkowników dodatkowo oceniane jako wiarygodne i rzetelne[25], zwalniają nas niejako z procesu myślenia i decydowania, sprawiając, iż słowa McLuhana „kształtujemy narzędzia, a potem one kształtują nas”[26] nie straciły na swej aktualności.

Głęboka sieć stanowi przeciwieństwo sieci powierzchniowej (*Surface Web*), zwanej inaczej siecią widoczną (*Visible Web*) czy siecią indeksowalną (*Indexable Web*), do której dostęp jest osiągalny z poziomu wyszukiwarek. Według statystyk początkiem listopada 2013 roku zindeksowano co najmniej 13,5 miliarda stron[27]. Bergman na podstawie analiz prowadzonych w marcu 2000 roku wykazał, że w sieci głębokiej znajduje się od 400 do 550 razy więcej informacji niż w powszechnie dostępnej sieci WWW. Ponadto dowiódł, iż zawiera ona 7500 terabajtów informacji w stosunku do 19 terabajtów informacji sieci powierzchniowej. Bergman zauważył, że głęboka sieć jest najdynamiczniej rozwijającą się kategorią sieci WWW, zawierającą informacje specjalistyczne, rzetelne i weryfikowane częściej niż te, które wyłuskiwane są przez tradycyjne szperacze. Ponad połowa tego typu informacji znajduje się w specjalistycznych bazach danych, a 95% z nich jest dostępnych za darmo, bez konieczności uiszczania opłaty czy subskrypcji. Z punktu widzenia przedstawicieli nauk humanistycznych sieć niewidzialna wydaje się szczególnie ważna. Największy odsetek jej zasobów stanowią bowiem treści z obszaru humanistyki (13,5%), mediów i wiadomości (12,2%) oraz informa-

22 M. Bergman, *The Deep Web: Surfacing Hidden Value*, „Journal of Electronic Publishing” 2001.

23 Warto zwrócić uwagę na pewne nieścisłości w wielu publikacjach naukowych, które podejmują problematykę głębokiej sieci. Niektórzy autorzy nieprawidłowo posiłkują się terminem „ukryty internet” zamiast „ukryta sieć”; zob. N. Pamuła-Cieślak, *Ukryty Internet – jeśli nie wyszukiwarka, to co?*, „EBIB” 2004, nr 7; *eadem*, *Typologia zasobów ukrytego Internetu*, „Przegląd Biblioteczny” 2006, z. 2. Fenomen ukrytej sieci dotyczy zasobów World Wide Web, a nie znacznie pojemniejszego i niesłusznie synonimicznie z nim używanego terminu „internet”, który obejmuje nie tylko strony internetowe, ale także programy pocztowe, programy do komunikowania się typu Jabber, Skype – ich efekt głębokiej sieci nie obejmuje.

24 W. Gogołek, *Komunikacja sieciowa. Uwarunkowania, kategorie i paradoksy*, ASPRA, Warszawa 2010, s. 270.

25 Według badań prowadzonych w USA 73% ankietowanych uważa, że informacje, które wyłuskują dla nich szperacze, są dokładne i wiarygodne, a 66% jest zdania, że wyszukiwarki są rzetelnym i obiektywnym źródłem informacji. Zob. K. Purcell, J. Brenner, L. Rainie, *Search Engine Use 2012*, „Pew Internet & American Life” 2012.

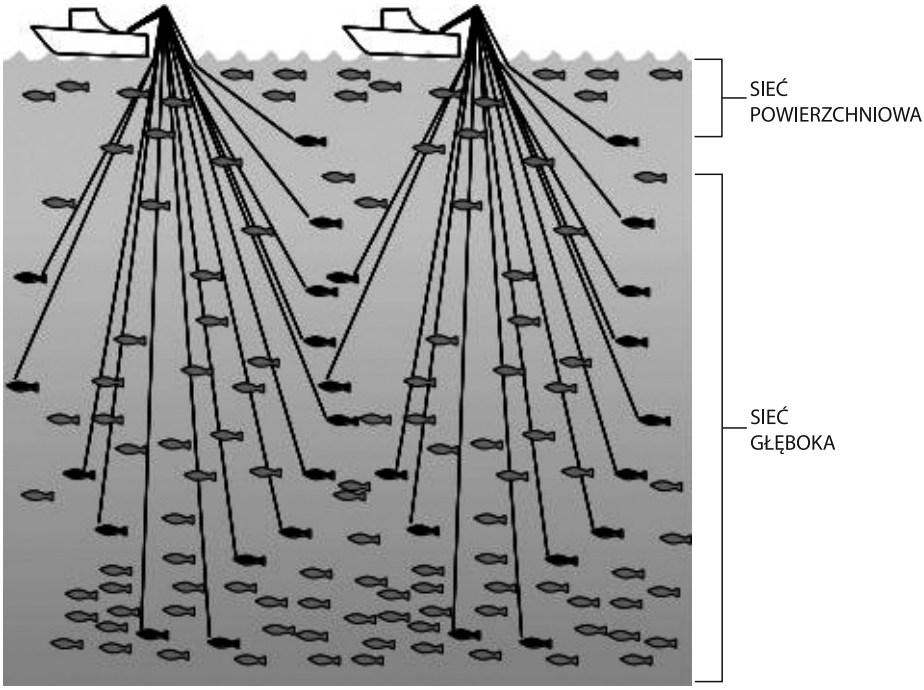
26 M. McLuhan, *Zrozumieć media. Przedłużenia człowieka*, WNT, przeł. N. Szczucka, Warszawa 2004, s. 17.

27 Zob. http://www.worldwidewebsize.com (dostęp: 8.11.2013).

tyki (6,9%)[28]. Można też zauważyć, że treści, które funkcjonują w obrębie sieci niewidzialnej, cechuje wysoka jakość informacji, rzetelność i profesjonalizm[29], często bowiem powstają one z inicjatywy ekspertów i specjalistów z danej dziedziny[30].

Konstatacje Bergmana były poddawane krytyce. Niektórzy badacze zarzucali mu przeszacowanie wielkości sieci ukrytej[31], jednakże aktualne szacunki potwierdzają, iż to, do czego docieramy dzięki szperaczom, stanowi niewielki odsetek tego, co znajduje się w obrębie sieci WWW. Paul Gil dowodzi, że jedynie 10% sieci WWW zostało zindeksowane, a więc jest dostępne z poziomu wyszukiwarek, zaś 90% stanowi sieć ukryta[32].

Rysunek 1. Sieć widzialna i sieć ukryta



Źródło: opracowanie własne na podstawie: M. Bergman, *The Deep Web: Surfacing Hidden Value*, „Journal of Electronic Publishing" 2001.

28 N. Pamuła-Cieślak, *Ukryty Internet*…, *op. cit.*

29 J. Devine, F. Egger-Sider, *Beyond Google: The Invisible Web in the Academic Library*, „The Journal of Academic Librarianship" 2004, vol. 30, s. 265–269.

30 D. Szumilas, *Kop głębiej! Google to nie wszystko*, „Magazyn Internet" 2005, nr 8, s. 60–63.

31 D. Lewandowski, P. Mayr, *Exploring the Academic Invisible Web*, 2006, http://arxiv.org/pdf/cs/0702103.pdf (dostęp: 8.11.2013).

32 P. Gil, *What Is the 'Invisible Web'?*, 2013, http://netforbeginners.about.com/cs/secondaryweb1/a/secondaryweb.htm (dostęp: 8.11.2013).

Warto w tym miejscu zwrócić uwagę, że sieć głęboka nie jest ciemnym internetem (*Dark Internet*) ani też Darknetem, który *de facto* stanowi jego część. Ciemny internet, zwany także ciemną przestrzenią adresową, oznacza wszystkie niedostępne hosty sieciowe w Internecie. Jednym z przykładów ciemnego internetu są archaiczne strony funkcjonujące jeszcze w MILNECIE[33], czasami tak stare jak sam ARPANET, niewłączone do rozwijającej się architektury współczesnego internetu[34]. Darknet z kolei stanowi sieć anonimową, w której połączenia odbywają się wyłącznie pomiędzy zaufanymi osobami, czasami określanymi jako *Przyjaciele* (F2F)[35]. Darknet różni się od innych rozproszonych sieci *peer-to-peer*[36] tym, że dzielenie się zasobami w jego obrębie jest całkowicie anonimowe (adresy IP nie są publicznie udostępniane)[37]. Jak zauważa Jessica Wood, Darknet jest często związany z komunikacją politycznych dysydentów, ale jest także wykorzystywany do działań niezgodnych z prawem.

Głęboki internet nie jest zatem zakulisową, nieoficjalną, ukrytą z premedytacją po to, by umożliwić działania nielegalne czy nieuczciwe, częścią sieci. Stanowi tę część, która z różnych względów nie jest indeksowana przez najpopularniejsze wyszukiwarki, co w znaczący sposób utrudnia dotarcie do treści znajdujących się w jej obrębie. Główna przyczyna istnienia ukrytego internetu tkwi w samych mechanizmach wyszukująco-indeksujących szperaczy. Strony połączone z sobą hiperlinkami można zobrazować za pomocą struktury grafu[38]. Jak wskazuje Natalia Pamuła-Cieślak, struktura grafu nie jest dokładnie znana, więc opracowanie algorytmu, który wyszuka i zindeksuje wszystkie strony internetowe, jest trudne. Po pierwsze, przeczesywanie internetu przez boty[39] metodą grafu sprawia, że roboty nie docierają do stron, do których nie prowadzą żadne linki, a po drugie, każdy z nich wybiera różne drogi, co sprawia, że automaty różnych wyszukiwarek rejestrują odmienne zbiory danych[40]. Autorka zauważa, że istnienia tak rozległego internetu ukrytego należy upatrywać w tym, iż standardowe wyszukiwarki zostały zaprojektowane do indeksowania stron

---

33 MILNET (*Military Network*) – część ARPANET-u wyznaczona do przesyłania jawnych treści Amerykańskiego Departamentu Obrony. Wyłączony z sieci ARPANET w 1983 roku ze względów bezpieczeństwa, po udostępnieniu ARPANET-u społeczności akademickiej.

34 A. Hissey, *The Dark Internet*, http://www.crt.net.au/About/ETopics/Archives/darkint.html (dostęp: 7.11.2013).

35 *Friend-to-Friend* – typ sieci *peer-to-peer*, w której użytkownicy dokonują bezpośrednich połączeń wyłącznie z osobami, które znają. Dla uwierzytelnienia danego użytkownika mogą być stosowane hasła lub podpisy cyfrowe. Uwierzytelnianie służy potwierdzeniu zadeklarowanej tożsamości osoby uczestniczącej w procesie komunikacji. Banki uwierzytelniają tożsamość danej osoby np. poprzez podanie danych, które są przypisane do konkretnej jednostki – jej daty urodzenia, nazwiska panieńskiego etc.

36 *Peer-to-peer* – to rodzaj zdecentralizowanej i rozproszonej architektury sieci, w której poszczególne węzły w sieci stanowią sami użytkownicy, bez potrzeby scentralizowanej koordynacji przez serwery.

37 J. Wood, *The Darknet: A Digital Copyright Revolution*, „Richmond Journal of Law and Technology" 2010, nr 16.

38 J. Rafa, *Co każdy internauta wiedzieć powinien*," „Internet". Dodatek do: „Poradnik Praktyczny" 2002, cz. 6, s. 1–12.

39 Internet bot, zwany również robotem WWW, pająkiem internetowym lub crawlerem, to oprogramowanie, które wykonuje zautomatyzowane zadania przez Internet.

40 N. Pamuła-Cieślak, *Typologia zasobów ukrytego Internetu*…, *op. cit.*

opartych na języku HTML (HyperText Markup Language), a więc takich, które mają postać tekstu statycznego.

Warto dodać, że zindeksowanie danej strony zależne jest także od samego użytkownika. Witryna, by została uwzględniona w indeksie wyszukiwarki, musi być zgłoszona do jej katalogu. Brak takiego zgłoszenia skutkuje tym, że nie jest ona przez szperacze uwzględniana, chyba że w sieci są jakieś odnośniki do niej. Przyczyną pomijania witryny przez boty jest także jej nieprawidłowa semantyka – złe nagłówki, brak lub niewłaściwe znaczniki w kodzie strony (dokumencie HTML).

Strony, które znajdują się w strefie głębokiej sieci, są także ignorowane przez crawlery wtedy, gdy nie prowadzą do nich żadne hiperłącza, gdy strony mają charakter prywatny – dostęp do nich wymagania rejestracji i logowania, co powoduje, że zasoby chronione są hasłem. Witryna staje się dla pająka internetowego „niewidoczna" też wtedy, gdy jej zawartość jest ograniczona mechanizmami, których celem jest redukowanie działania robotów na przykład poprzez CAPTCHA[41]. Kluczowe znaczenie ma także to, czy strona jest zbudowana wyłącznie w HTML-u, czy bazuje także na treściach dynamicznie pobieranych z serwerów WWW w technologii Flash bądź Ajax.

Kilka lat temu niektóre witryny, ze względu na częstość działania internetowych botów, przynależały czasowo do niewidzialnej sieci. Jeszcze w 2001 aktualizowanie przez pająki internetowe treści zajmowało trzy, cztery miesiące[42], w 2003 roku około miesiąca[43], obecnie aktualizacja treści odbywa się niemal na bieżąco. Wykorzystywane przez Google Googleboty działają w dwojaki sposób – dokonując głębokiego pełzania (*deep crawl*) i świeżego pełzania (*fresh crawl*)[44]. Głębokie pełzanie polega na przechodzeniu od linku do linku i dodawaniu możliwie wielu informacji do indeksu. Proces ten odbywa się średnio raz w miesiącu. Świeże pełzanie odbywa się nawet kilka razy dziennie, a jego celem jest odwiedzenie stron zmieniających się często i zaktualizowanie informacji, które się na niej znajdują. Warto dodać, iż Google umożliwia dostosowanie częstości odwiedzin bota na witrynie w zależności od własnych potrzeb.

W obrębie struktury WWW można wyróżnić cztery poziomy, z których dwa mają charakter dostępny, dwa zaś przynależą do sfery niewidzialnego internetu. Na pierwszym poziomie widzialnego internetu znajdują się witryny o relatywnie stałej tematyce, rzadko zmieniające swą zawartość. Poziom drugi to zazwyczaj strony tematyczne, poświęcone zwykle jednemu zagadnieniu. Trzeci poziom głębokiej sieci zawiera dynamiczne bazy da-

41 CAPTCHA – stanowi mechanizm zabezpieczający strony WWW, którego celem jest weryfikacja, czy działania na danej stronie podejmował człowiek, czy automat. Polega ona najczęściej na odczytaniu treści (krótkich wyrazów) z obrazka, które są dość łatwo interpretowane przez człowieka, zaś automatom sprawiają kłopot. Czasem CAPTCHA polega na udzieleniu odpowiedzi na proste zadanie: „Ile jest cztery razy cztery?" czy „Podaj datę wybuchu II wojny światowej". Zastosowanie go zapobiega działaniom spamerskim, zakładaniu kont na portalach przez automaty oraz chroni przed reklamami na blogach.

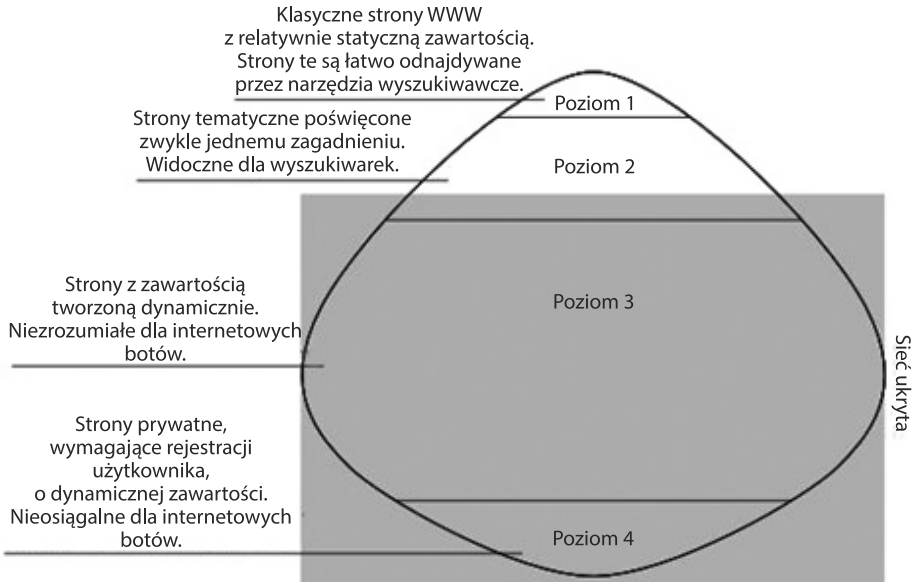
42 M. Bergman, *The Deep Web: Surfacing Hidden Value*, *op. cit.*

43 Ch. Sherman, G. Price, *The Invisible Web. Uncovering Information Sources Search Engines Can't See*, Medford, NJ 2003.

44 *Google Webmaster Central. Get data about crawling, indexing and search traffic. Increase traffic to your site* (dostęp: 8.11.2013).

nych, które nie są „zrozumiałe" dla internetowych pełzaczy, zaś czwarty przynależy do stron o dynamicznej zawartości, wymagających rejestracji użytkownika, często prywatnych. Do sfery czwartej zaliczyć można witryny przynależące do sieci prywatnej (*The Private Web*) oraz sieci zastrzeżonej (*The Proprietary Web*)[45]. Dostęp do tych stron jest ograniczony przez hasło, co uniemożliwia crawlerowi jej zindeksowanie. Możliwa jest również sytuacja, w której webmaster strony uniemożliwia botowi dostęp do niej, np. poprzez użycie metatagu NOARCHIVE lub NON[46] w kodzie HTML. Sieć zastrzeżona limituje dostęp do strony użytkownikom, którzy dokonali subskrypcji, co blokuje do niej dostęp bota.

Rysunek 2. Poziomy World Wide Web



Źródło: opracowanie własne na podstawie: http://netforbeginners.about.com/library/diagrams/n4layers.htm (dostęp: 8.11.2013).

Czy zatem jesteśmy skazani na hegemonię Google i wyłączanie z naszego pola uwagi ogromnych zasobów treści online? Niekoniecznie. Nieindeksowanie tak znacznych informacji funkcjonujących w przestrzeni WWW nie wynika ze złej woli internetowego giganta, ale raczej z ograniczonych możliwości serwerów tej firmy. Być może za kilka lat firma ta zainwestuje w wydajniejsze serwery, co pozwoli na indeksowanie nie tylko statycznego tekstu, ale także zasobów dynamicznych. Warto dodać, że twórcy wyszukiwarek nieustan-

[45] Ch. Sherman, G. Price, *The Invisible Web…*, *op. cit.*
[46] Nieskuteczny w przypadku Googlebota.

nie udoskonalają możliwości narzędzi wyszukiwawczych, włączając w proces indeksowania coraz to nowsze typy źródeł, co pozwala na redukcję obszaru ukrytego internetu.

Należy również uświadomić sobie fakt, iż to, że dane treści nie są osiągane z poziomu wyszukiwarki, nie oznacza, że dotarcie do nich jest niemożliwe. Czasami wystarczy skorzystanie z multiwyszukiwarki[47], ze specjalistycznych katalogów tematycznych czy dziedzinowych, a także bibliotek wirtualnych[48], co pozwala dotrzeć do nieindeksowanej przez tradycyjne wyszukiwarki treści. Innym rozwiązaniem umożliwiającym znalezienie zasobów znajdujących się w głębokim internecie są tak zwane *subject gateways*, które stanowią posegregowane według dziedzin przewodniki po treściach znajdujących się online. Selekcji, oceny i katalogowania tychże treści dokonują profesjonalni bibliotekarze lub eksperci dziedzinowi[49]. Jednakże nawet najdoskonalsze i najefektywniejsze rozwiązania technologiczne nie powinny zwalniać jednostki z krytycyzmu wobec źródła, z którego czerpie informacje.

Reasumując, pozwolę sobie sięgnąć do wymownego cytatu z książki Randalla Strossa *Planeta Google*: „W każdym wieku – węgla, stali, ropy – jest taki surowiec, który definiuje jego historyczne znaczenie. W naszym są nim informacje, a Google stał się ich górującym nad innymi zarządcą"[50]. Niezwykle aktualne wydają się też konstatacje Michela Foucaulta, który zauważył, że ten, kto ma wiedzę, ten ma władzę. Władanie i reglamentowanie dostępu do wiedzy funkcjonującej w sferze online przypada współcześnie internetowemu gigantowi – firmie Google, która w wielu wymiarach zarządza naszymi umysłami. Trudno obecnie wyobrazić sobie, by Google utraciła swoją monopolistyczną pozycję na rynku wyszukiwawczym. Statystyki użytkowania i zaufania wobec dominującej wyszukiwarki pozwalają na konstatację, że stała się ona współczesnym informacyjnym hegemonem. Choć internet zgodnie z jego libertariańskimi korzeniami postrzegano jako przeciwwagę dla ograniczających nas mediów masowych, w których kluczową rolę odgrywają informacyjni gatekeeperzy, szybko okazało się, że te same mechanizmy, choć w nieco innej formie, obserwować można w pozornie wolnościowym, niezależnym i egalitarnym internecie[51], czyli medium, które zostało zdominowane przez mechanizmy wyszukiwawcze agregujące i ogniskujące naszą uwagę na pewnych kwestiach, z pominięciem innych. Implikuje to sytuację, w której jesteśmy jeszcze bardziej sterowani i manipulowani niż w klasycznych mediach masowych, co do których takie formy sprawowania kontroli nad naszymi umysłami są łatwiej wychwytywane i obserwowalne niż w przypadku medium, którego immamentną cechą miał być brak jakiejkolwiek kontroli. Wolnością w przypadku użytkowania nowych mediów może być jedynie

[47] Multiwyszukiwarki przeszukują jednocześnie zasoby OA (*Open Archives*) wszystkich lub wybranych repozytoriów, czasopism i bibliotek cyfrowych, np. ScientificCommons – indeksuje metadane i treści publikacji (ponad 38 mln z 1269 repozytoriów), Harvester2 – multiwyszukiwarka *Public Knowledge Project*, która przeszukuje zasoby około 2600 archiwów, repozytoriów i czasopism OA (ponad 2 mln publikacji), OAIster – wyszukiwarka dokumentów elektronicznych, m.in. książek, artykułów, plików audio i wideo, grafiki itp.

[48] Por. N. Pamuła-Cieślak, *Ukryty Internet…*, *op. cit.*, oraz *eadem*, *Typologia zasobów ukrytego Internetu*, *op. cit.*

[49] L. Derfert-Wolf, *Serwisy tematyczne o kontrolowanej jakości w Internecie* – subject gateways, „Biuletyn EBIB" 2004, nr 6.

[50] R. Stross, *op. cit.*, s. 11.

[51] Zob. M. Szpunar, *Nowe-stare medium…*, *op. cit*.

krytycyzm i odrzucenie bezgranicznego zaufania wobec technologii, która zawiaduje naszą uwagą. Im częściej będziemy przenosić odpowiedzialność za własne decyzje na narzędzia, którymi się posługujemy, tym częściej to technologia będzie stawać się narzędziem sprawowania kontroli i władzy, której milczącemu przyzwoleniu będziemy się poddawać. Po raz kolejny w historii rozwoju mediów okazuje się, że praktyki kulturowe nie nadążają za zmianami w obszarze technologii, co implikuje sytuację, w której hipoteza opóźnienia kulturowego w dobie mediów cyfrowych nie straciła swej mocy eksplanacyjnej. Wydaje się, iż po etapie zachwytu nad nieograniczonymi możliwościami wyszukiwania informacji online zaawansowani użytkownicy sieci, po zaznajomieniu się z mechanizmami działania szperaczy, przekonują się, jak wiele wartościowych informacji znajduje się poza ich polem uwagowym online. Niestety, tych refleksyjnych i krytycznych internautów jest ciągle zbyt niewielu, by móc przełamać hegemonię internetowego giganta w zakresie wyszukiwania informacji.

Bibliografia

Bergman M., *The Deep Web: Surfacing Hidden Value*, „Journal of Electronic Publishing" 2001.
Bolter J.D., *Człowiek Turinga. Kultura Zachodu w wieku komputera*, przeł. T. Goban-Klas, PIW, Warszawa 1990.
Brophy J., Bawden D., *Is Google Enough? Comparison of an Internet Search Engine with Academic Library Resources*, „Aslib Proceedings: New Information Perspectives" 2005, vol. 57, s. 498–512.
Derfert-Wolf L., *Serwisy tematyczne o kontrolowanej jakości w Internecie – subject gateways*, „Biuletyn EBIB" 2004, nr 6.
Devine J., Egger-Sider F., *Beyond Google: The Invisible Web in the Academic Library*, „The Journal of Academic Librarianship" 2004, vol. 30, s. 265–269.
Griffiths J., Brophy P., *Student Searching Behaviour and the Web: Use of Academic Resources and Google*, „Library Trends" 2005, vol. 53, s. 539–554.
Gogołek W., *Komunikacja sieciowa. Uwarunkowania, kategorie i paradoksy*, ASPRA, Warszawa 2010.
Gorman L., McLean D., *Media i społeczeństwo. Wprowadzenie historyczne*, przeł. A. Sadza, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków 2010.
Halavais A., *Wyszukiwarki internetowe a społeczeństwo*, przeł. T. Płudowski, Wydawnictwo Naukowe PWN, Warszawa 2012.
Hargittai E., Fullerton L., Menchen-Trevino E., Thomas K.Y., *Trust Online: Young Adults' Evaluation of Web Content*, „International Journal of Communication" 2010, no. 4.
Innis H., *Nachylenie komunikacyjne*, „Communicare. Almanach Antropologiczny. Oralność/Piśmienność" 2007.
Lewin K., *Frontiers in Group Dynamics*, „Human Relations" 1947, vol. 1, no. 2.
Loska K., *Dziedzictwo McLuhana – między nowoczesnością a ponowoczesnością*, Rabid, Kraków 2001.
Markowski A. (red.), *Wielki słownik ortograficzny języka polskiego*, Wilga, Warszawa 1999.
McLuhan M., McLuhan E., *Laws of Media, The New Science*, Toronto 1988.
McLuhan M., *Wybór tekstów*, przeł. E. Różalska, J. Stokłosa, Zyski S-ka, Poznań 2001.
McLuhan M., *Zrozumieć media. Przedłużenia człowieka*, przeł. N. Szczucka, Wydawnictwo Naukowo-Techniczne, Warszawa 2004.

Morozov E., *To Save Everything, Click Here: The Folly of Technological Solutionism*, Penguin, 2013.
Mumford L., *Authoritarian and Democratic Technics*, „Technology and Culture" 1964, vol. 5.
Nicholas D., Dobrowolski T., Withey R. *et al.*, *Digital Information Consumers, Players and Purchasers: Information Seeking Behaviour in the New Digital Interactive Environment*, „Aslib Proceedings: New Information Perspectives" 2003, vol. 55, s. 23–31.
Orliński W., Internet. Czas się bać, Agora SA, Warszawa 2013.
Pamuła-Cieślak N., *Typologia zasobów ukrytego Internetu*, „Przegląd Biblioteczny" 2006, z. 2.
Pamuła-Cieślak N., *Ukryty Internet – jeśli nie wyszukiwarka, to co?*, „EBIB 2004", nr 7.
Pan B., Hembrooke H., Joachmis T. *et al.*, *In Google We Trust: Users' Decisions on Rank, Position, and Relevance*, „Journal of Computer-Mediated Communication" 2007, no. 12.
Postman N., *Technopol. Triumf techniki nad kulturą*, przeł. A. Tanalska-Dulęba, Muza, Warszawa 2004.
Purcell K., Brenner J., Rainie L., *Search Engine Use 2012*, „Pew Internet & American Life" 2012.
Rafa J., *Co każdy internauta wiedzieć powinien*, „Internet". Dodatek: „Poradnik Praktyczny" 2002, cz. 6, s. 1–12.
Sherman Ch., Price G., *The Invisible Web. Uncovering Information Sources Search Engines Can't See*, Medford, NJ 2003.
Stross R., *Planeta Google*, przeł. A. Wojtaszczyk, O. Wojtaszczyk, Studio EMKA, Warszawa 2009.
Szpunar M., *Bogactwo informacji a ubóstwo uwagi w cyfrowej kulturze nadmiaru*, [w:] M. Kaczmarczyk, D. Rott (red.), *Problemy konwergencji mediów*, Verbum, Praga 2013.
Szpunar M., *Nowe-stare medium. Internet między tworzeniem nowych modeli komunikacyjnych a reprodukowaniem schematów komunikowania masowego*, IFiS PAN, Warszawa 2012.
Szpunar M., *Wokół koncepcji gatekeepingu. Od gatekeepingu tradycyjnego do technologicznego*, [w:] I.S. Fiut (red.), *Człowiek w komunikacji i kulturze*, Wydawnictwo Aureus, Kraków 2013.
Szumilas D., *Kop głębiej! Google to nie wszystko*, „Magazyn Internet" 2005, nr 8.
Wood J., *The Darknet: A Digital Copyright Revolution*, „Richmond Journal of Law and Technology" 2010, nr 16.

Netografia

Gil P., *What Is the 'Invisible Web'?*, 2013, http://netforbeginners.about.com/cs/secondaryweb1/a/secondaryweb.htm (dostęp: 8.11.2013).
Google Webmaster Central. Get data about crawling, indexing and search traffic. Increase traffic to your site (dostęp: 8.11.2013).
Hissey A., *The Dark Internet*, http://www.crt.net.au/About/ETopics/Archives/darkint.html (dostęp: 7.11.2013).
Lewandowski D., Mayr P., *Exploring the Academic Invisible Web*, http://arxiv.org/pdf/cs/0702103.pdf (dostęp: 8.11.2013).
Mount B., Koll M., *PLS introduces AT1, the first 'second generation' Internet search service*, 1996, http://web.archive.org/web/19971021232057/www.pls.com/news/pr961212_at1.html (dostęp: 7.11.2013).
http://poradnia.pwn.pl/lista.php?id=8070 (dostęp: 9.11.2013).
http://www.worldwidewebsize.com (dostęp: 8.11.2013).