

Annales Universitatis Paedagogicae Cracoviensis

Studia ad Bibliothecarum Scientiam Pertinentia XIV (2016)

ISSN 2081-1861

DOI 10.24917/20811861.14.2

Władysław Marek Kolasa

Serwisy biblioteczne w perspektywie *search engine optimization*

Wstęp

Pozycjonowanie serwisów internetowych (ang. *search engine optimization*, SEO) bywa wśród użytkowników rozmaicie rozumiane. Lektura popularnych opracowań i blogów przekonuje, że funkcjonuje w tym zakresie wiele mitów¹. Istnieją poglądy jakoby SEO była zlepkiem technik informatycznych, które sprawiają, że pozycjonowana witryna zawsze pojawia się na czele rankingów wyszukiwarki. Jest to przekonanie fałszywe, z co najmniej dwóch powodów: po pierwsze, wysoką pozycję na stronie wyników (ang. *search engine results page*, SERP) można uzyskać tylko dla określonej podstrony serwisu, po wtóre – pozycjonowanie odbywa się zawsze na ściśle określone słowa kluczowe (wliczywszy także ich warianty gramatyczne). Pozycja serwisu na liście adresów znalezionych za pomocą wyszukiwarki zależy ponadto od szeregu innych czynników ujętych w jej algorytmie. Aktualnie algorytm Google uwzględnia ok. 200 czynników o rozmaitej wadze, które można zgrupować w cztery kategorie. Pierwsza obejmuje słowa kluczowe, druga – tzw. czynniki *on-page* (czyli elementy zależne od webmastera), trzecia – tzw. czynniki *off-page* (niezależne od webmastera), zaś czwarta – metryki reputacji. Warto też podkreślić, że rola poszczególnych składników nie jest do końca znana oraz stale ewoluuje; głównie ze względu na cyklicznie pojawiające się poprawki algorytmu, np. Panda Update, Penguin Update, Hummingbird Update².

Twórcy wyszukiwarek z oczywistych powodów trzymają w tajemnicy znacznie poszczególnych czynników algorytmu, ujawniając jedynie publicznie zbiory dobrych praktyk (np. *Google Webmaster Guidelines*)³, zaś empiryczna wiedza na ten temat pochodzi jedynie z przeprowadzanych cyklicznie badań zespołów eksperckich.

¹ C. Shepard, *10 SEO Myths that Friggin' Tick Me Off*, MOZ Blog (March 9th, 2015), [online] <https://moz.com/blog/seo-myths> [dostęp 21.05.2016]; M. Kliszczak, *Pozycjonowanie (SEO) w internecie – fakty i mity*, Eactive Blog (11/02/2014), [online] <https://www.eactive.pl/pozycjonowanie-stron/pozycjonowanie-w-internecie-fakty-i-mity/> [dostęp 10.05.2016]; Ł. Iwanek, *10 mitów na temat SEO*, „e-Gospodarka” 2014, [online] <http://www.egospodarka.pl/116534,10-mitow-na-temat-SEO.html> [dostęp 10.05.2016].

² E. Enge, S. Spencer, J. Stricchiola, *SEO, czyli sztuka optymalizacji witryn dla wyszukiwarek*, Wyd. 3. Gliwice 2016, s. 535–550.

³ *Google Webmaster Guidelines*, [online] <https://support.google.com/webmasters/answer/35769> [dostęp 20.05.2016].

Aktualnie najwyższym prestiżem wśród praktyków SEO cieszą się dwie społeczności: grupa skupiona wokół amerykańskiego serwisu MOZ (dawniej SEOmoz)⁴, który został założony w 2004 roku przez Randa Fishkina i Gillian Muessig, oraz jego niemieckiego odpowiednika – firmy Searchmetrics⁵, powołanej w 2005 roku przez Marcusa Tobera. Pomimo pewnych ograniczeń (SEO wciąż pozostaje dziedziną bardziej praktyczną niż naukową), na polu tym wypracowano już wiele teorii, które są wykorzystywane do budowy narzędzi analitycznych i z powodzeniem stosowane w branży świadczącej usługi w zakresie pozycjonowania serwisów.

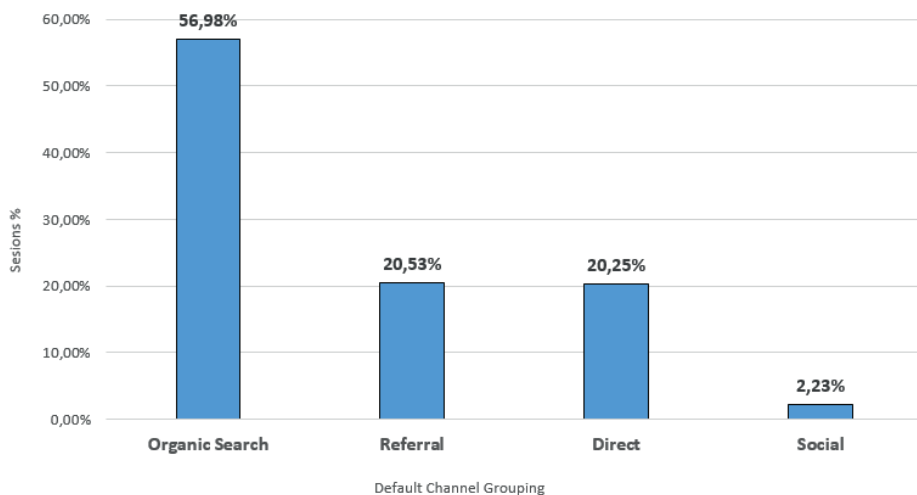
Zanim przejdziemy do sedna artykułu warto jeszcze postawić dwa pytania: po pierwsze, czy stosowanie zasad SEO do właściwego funkcjonowania w sieci jest niezbędne, po drugie, czy praktyka SEO dotyczy także serwisów instytucji *non-profit* (w tym bibliotek) w takim samym stopniu, co serwisów handlowych? Odpowiedź na pierwsze pytanie nie jest jednoznaczna. Nie można bowiem zaprzeczyć tezie, że również bez stosowania zasad SEO każda poprawna strona będzie wyświetlana w przeglądarce użytkownika, jeśli ten zna jej adres URL. Badania przekonują jednak, że jedynie niewielka część witryn jest w taki sposób wywoływana. Wykonany doraźnie test pozyskiwania ruchu przez typowy polski serwis biblioteczny (wykres 1), wykonany na próbie danych z trzech miesięcy ($n = 38087$ sesji), przekonuje, że co piąty użytkownik (20,25%) trafia tam poprzez bezpośrednie wywołanie (*direct*). Pewien odsetek (20,53%) pochodzi ze stron referencyjnych, gdzie witryna została podlinkowana (*referral*) oraz z odwołań pochodzących z sieci społecznościowych (2,25%, *social*). Zdecydowanie największy ruch generuje jednak wyszukiwarka (*organic search*), skąd pochodzi większość wszystkich wywołań (56,98%).

Należy jednak podkreślić, że przedstawionego rozkładu nie należy traktować uniwersalnie. Źródła pochodzenia ruchu są związane z typem serwisu (charakterem udostępnianej informacji i świadczonych usług) oraz rodzajem użytkownika (kontakt doraźny lub stały). Z pewnością więc inne (wyższe) będą proporcje odwołań do podstron serwisów handlowych, usługowych, gdzie z reguły użytkownik trafia doraźnie, a poszczególne produkty/usługi ma indywidualną podstronę. Podstrony tego typu serwisów są zwykle bardzo unikalne i nietrwałe, toteż szansa na wyłanianie bezpośrednie i pośrednie jest znikoma. Na przeciwnym biegunie lokują się serwisy o charakterze naukowym, kulturalnym czy urzędowym, które cechuje większa trwałość, renoma (związana z marką) oraz wyrobiony czytelnik (często fachowiec). Ostatni typ portalu ma, ze względu na wyliczone uprzednio walory, zdecydowanie większe szanse na podlinkowanie przez bazy danych i inne serwisy (*referral*) oraz zapis w zakładkach przeglądarki (*direct*). Bez wątplenia serwisy biblioteczne należą do ostatniej z wymienionych grup. Warto jednocześnie dodać, że te naukowe, kulturalne i urzędowe zwykle nie korzystają z płatnego marketingu w wyszukiwarkach, stąd na ogół nie generują ruchu reklamowego (*advertising search*), typowego dla serwisów handlowo-usługowych.

Kolejną ważną zasadą SEO, z którą muszą liczyć się twórcy serwisów, jest konieczność takiego pozycjonowania, aby podstrona znalazła się w pierwszej dziesiątce wyników wyszukiwarki. Badania zachowań użytkowników, szczególnie współczynnika klikalności (*click-through rate*, CRT), przekonują, że jedynie obecność na

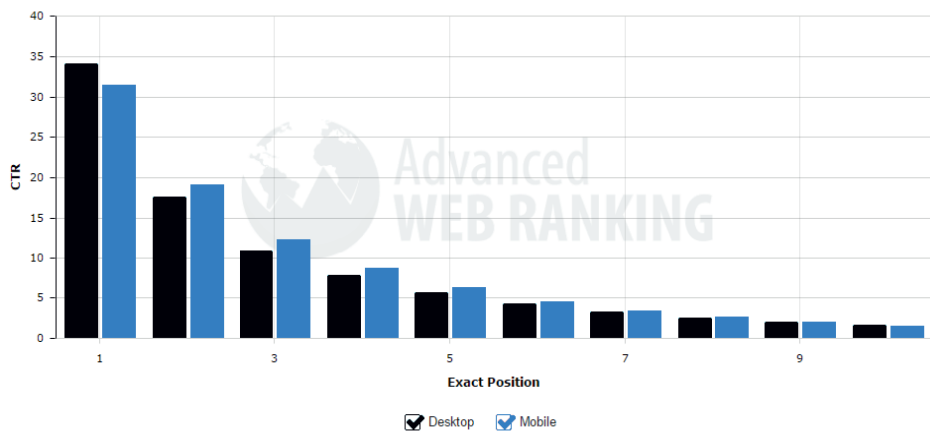
⁴ MOZ, [online] <https://moz.com> [dostęp 20.05.2016].

⁵ Searchmetrics, [online] <http://www.searchmetrics.com> [dostęp 20.05.2016].



Wyk. 1. Źródła pozyskiwania ruchu w serwisie bibliotecznym (na przykładzie MBC*)

Źródło: obliczenia własne na bazie danych z Google Analytics. *Przedmiot pomiaru: Małopolska Biblioteka Cyfrowa [<http://mbc.malopolska.pl/>], dane za okres marzec-maj 2016. n = 38087 sesji.



- This chart shows the organic click-through rates for searches coming from 6,170,005 keywords for 56,894 websites.
- You can compare the CTR for searches coming from desktop devices (web) versus mobile devices.

Wyk. 2. Współczynnik klikalności (*click-through rate*) w wyszukiwaniu organicznym Google (maj 2016)

Źródło: <https://www.advancedwebranking.com/ctrstudy/> [dostęp 30.05.2016].

pierwszej stronie wyników daje realną szansę na odwiedzinę. Z cyklicznych raportów MOZ wynika, że na pierwszej stronie kończy aż 71,3% użytkowników. Największą szansę na kliknięcie ma wynik na czele listy – 31,24%, drugi – 14,04%, trzeci – 9,85%, czwarty – 6,97%, piąty – 5,5%. Na pozycje 9–10 przypada łącznie 3,73%, na drugą stronę SERP-ów zagląda tylko 3,99% użytkowników, a na kolejne jedynie

1,6%⁶. Warto równocześnie zauważyć, że w podobny sposób zachowują się użytkownicy komputerów stacjonarnych (wykres 2). Oba typy dostępu różni jedynie nieco mniejszy odsetek kliknięć pozycji pierwszej oraz nieznacznie większy na pozycjach: drugiej, trzeciej, czwartej i piątej.

Przedstawione argumenty przemawiają za tezą, że aby skutecznie uczestniczyć w dyskursie internetowym należy zadbać o dobrą widoczność serwisu w wyszukiwarce. Wprawdzie pozycjonowanie serwisów instytucji *non-profit* (w tym bibliotek) nie wydaje się bezwzględnie konieczne (43,02% ruchu pochodzi spoza wyszukiwarki), aczkolwiek trudno bagatelizować ponad połowę wywołań (56,98%), które prowadzą z wyników organicznych wyszukiwarek.

Przedmiot badań. Metody. Próba badawcza

Podstawowym celem niniejszego artykułu jest próba diagnozy stopnia znajomości i stosowania zasad SEO w polskich serwisach bibliotecznych. Analizy zmierzają do uzyskania odpowiedzi na pytanie: czy twórcy serwisów internetowych największych polskich bibliotek implementują zasady SEO, a jeśli tak, to czy wpływają one na ich pozycje w popularnych wyszukiwarkach? Ponieważ wyczerpująca analiza całego korpusu polskich serwisów bibliotecznych byłaby bardzo pracochłonna, do badań wybrano próbę, którą tworzyło 13 witryn dużych bibliotek ulokowanych w czterech głównych miastach akademickich (Warszawa, Kraków, Poznań, Wrocław – tab. 1). Dobór witryn nie był przypadkowy. Z każdego miasta wybrano po trzy najczęściej odwiedzane serwisy, które wyłoniono za pomocą narzędzia Google AdWords. W trzech przypadkach (Kraków, Wrocław, Poznań) były to triady: biblioteka uniwersytecka, biblioteka politechniki oraz biblioteka publiczna; zaś w Warszawie, z uwagi na wyjątkową pozycję Biblioteki Narodowej, cztery biblioteki (biblioteka uniwersytecka, biblioteka politechniki, biblioteka publiczna oraz narodowa). Wybór miast także nie był przypadkowy. Dalsze badania przekonują, że z czterech wyliczonych ośrodków płynie prawie połowa całego ruchu internetowego z wyszukiwarki Google do polskich portali bibliotecznych.

Wybrane serwisy przebadano w czterech przekrojach i opisano w sześciu osobnych podrozdziałach obejmujących kolejno: badanie słów kluczowych (*keyword research, keyword usage and keyword density*), analizę czynników *on-page* (*content and technology*), analizę czynników *off-page* (*backlinks and social media*) oraz analizę metryk reputacji (*brand and confidence metrics*). W poszczególnych częściach analizy wykorzystano wiele narzędzi i metod. Do badań pierwszego obszaru (*keyword research*) wykorzystano dane z serwisu Google AdWords oraz dane wygenerowane z kodu HTML (*keyword usage and keyword density*). Materiał empiryczny do analizy aspektów *on-page* zyskano przy pomocy narzędzi serwisu SEORH oraz SEOquake; z kolei czynniki *off-page* analizowano głównie narzędziami serwisu Ahrefs oraz MOZ Open Site Explorer.

Zależnie od kontekstu w artykule zastosowano kilka jednostek analizy i pomiaru: przy badaniu słów kluczowych były to dane zawarte we wszystkich podstronach

⁶ P. Petrescu, *Google Organic Click-Through Rates in 2014*, MOZ Blog (October 1st 2014), [online] <http://moz.com/blog/google-organic-click-through-rates-in-2014> [dostęp 30.05.2016].

Tab. 1. Wykaz testowanych serwisów bibliotecznych

Miasto	Biblioteka	Skrót	URL
Warszawa	Biblioteka Narodowa	BN	www.bn.org.pl
	Biblioteka Główna Politechniki Warszawskiej	BG PW	www.bg.pw.edu.pl
	Biblioteka Uniwersytecka w Warszawie	BUW	www.buw.uw.edu.pl
	Biblioteka Publiczna m.st. Warszawy	BPmsW	www.koszykowa.pl
Kraków	Biblioteka Jagiellońska	BJ	www.bj.uj.edu.pl
	Biblioteka Główna Akademii Górniczo-Hutniczej	BG AGH	www.bg.agh.edu.pl
	Wojewódzka Biblioteka Publiczna w Krakowie	WBP Krak.	www.rajska.info
Poznań	Biblioteka Raczyńskich w Poznaniu	B. Racz.	www.bracz.edu.pl
	Biblioteka Uniwersytecka w Poznaniu	BUAM	lib.amu.edu.pl
	Biblioteka Politechniki Poznańskiej	B. PPozn.	library.put.poznan.pl
Wrocław	Centrum Wiedzy i Informacji Naukowo-Technicznej Politechniki Wrocławskiej	CWINT	biblioteka.pwr.wroc.pl
	Biblioteka Uniwersytecka we Wrocławiu	BUWr.	www.bu.uni.wroc.pl
	Miejska Biblioteka Publiczna we Wrocławiu	MBP Wr.	www.biblioteka.wroc.pl

* W kolejnych tabelach użyto skrótów

serwisów, gdyż takie dane dostarcza Google AdWords. W rozdziale dotyczącym stosowania słów kluczowych analizie poddano po trzy podstrony: czołówkę serwisu oraz dwie najczęściej odwiedzane podstrony (TOP1 i TOP2), zaś w dalszej części artykułu – tylko dane pochodzące lub prowadzące do „czołówki” (strony startowej). W trakcie projektowania badań uznano, że punktem odniesienia będzie tylko jedna wyszukiwarka – Google. Powód był prosty, od wielu lat jest absolutnym liderem. Według rankingu GemiusTraffic już od początku 2007 roku, co najmniej 90% wszystkich wyszukiwań z Polski pochodziło z Google, zaś w ciągu ostatnich 5 lat udział Google stabilnie oscylował w przedziale 95–98%⁷.

Dużym wyzwaniem przy kategoryzacji danych i ich interpretacji był dobór odpowiednich teorii. Te bowiem w dziedzinie SEO – o czym pisano wyżej – zmieniają się bardzo często i ewoluują wraz z algorytmem wyszukiwarek. Ponieważ większość regularnych związków wykrywa się drogą indukcji, w diagnozie czynników posłużono się najnowszymi raportami korelacyjnymi firmy Searchmetrics: *SEO Ranking Factors and Rank Correlations* i społeczności MOZ: *Search Engine Ranking Factors*⁸ oraz aktualnymi podręcznikami⁹.

Przyjęta strategia wynikała głównie z nikłego stanu badań w rozważanym obszarze. W polskim piśmiennictwie naukowym istnieje wprawdzie kilka opracowań, lecz są one przestarzałe lub ogólnikowe. Taki charakter ma np. artykuł autorstwa

⁷ Wyszukiwarki–silniki. *Ranking gemius Traffic*, [online] <http://www.ranking.pl/pl/rankings/search-engines.html> [dostęp 4.04.2016].

⁸ M. Tober, L. Hennig, D. Furch, *SEO Ranking Factors and Rank Correlations 2014*, Searchmetrics, San Mateo 2014, [online] <http://www.searchmetrics.com/knowledge-base/ranking-factors-2014/> [dostęp 4.04.2016]; M. Tober, D. Furch, K. Londenberg, L. Massaron, J. Grundmann, *Search Ranking Factors and Rank Correlations 2015*, Searchmetrics, San Mateo 2015, [online] <http://www.searchmetrics.com/knowledge-base/ranking-factors/> [dostęp 4.04.2016]; *Search Engine Ranking Factors 2015. Expert Survey and Correlation Data*, MOZ Blog, [online] <https://moz.com/search-ranking-factors> [dostęp 4.04.2016].

⁹ E. Enge, S. Spencer, J. Stricchiola, *SEO, czyli sztuka optymalizacji...*

Bożeny Jaskowskiej¹⁰, gdzie autorka wykonała prostą analizę czynników SEO (aktualnych w roku publikacji) oraz przedstawiła poziom ich realizacji w 27 polskich serwisach bibliotecznych, na podstawie danych uzyskanych drogą ankietową. Wśród prac nowszych warto wspomnieć o studium Adama Jachimczyka i Magdaleny Chrapek¹¹, w którym badacze poddali analizie ponad 300 katalogów *Presell Pages*, szukając odpowiedzi na to, czy czynnik ten wpływa na jakość stron mierzoną wskaźnikiem PageRank. Tekst jest wprawdzie poprawny metodologicznie, lecz ma nikłą wartość praktyczną. Po pierwsze dlatego, że w środowisku praktyków SEO od dawna wiadomo, że rola *Presell Pages*, podobnie jak katalogów SEO, jest ledwie symboliczna, zaś wprowadzona w 2012 roku poprawka Penguin Update traktuje pozyskane stamtąd linki jako spam¹². Po wtóre, od wielu lat rola PageRank w zasadzie nie liczyła się już w algorytmie, co *de facto* w 2016 roku potwierdził Google, usuwając w marcu 2016 roku dostęp danych dla Toolbar PageRank¹³.

Tematyka SEO jest wprawdzie dosyć często podejmowana w piśmiennictwie zagranicznym, lecz bardzo rzadko są to prace należące do obszaru badawczego informatologii. Baza *Web of Science* rejestruje 1216 dokumentów z zakresu SEO, w tym tylko 17 z dziedziny *information sciences and library science*, podobnie SCOPUS (odpowiednio: 2185 i 6) oraz LISA (108 i 33). Ich przydatność w niniejszym artykule okazała się ograniczona, gdyż zdecydowana większość to wycinkowe *case study*, dotyczące regionów¹⁴ lub usług¹⁵, zaś opracowania ogólne lub syntetyczne¹⁶ operują na wysokim poziomie ogólności.

¹⁰ B. Jaskowska, *Efektywny marketing SEO witryny bibliotecznej w wyszukiwarkach*, „Przegląd Biblioteczny” 2009, z. 2, s. 229–238.

¹¹ A. Jachimczyk, M. Chrapek, *Katalogi artykułów (Presell Pages) – analiza stron internetowych wykorzystywanych w SEO*, „Przegląd Biblioteczny” 2015, nr 2, s. 209–225.

¹² G. Gabe, *Penguin 3.0 Analysis – Penguin Tremors, Recoveries, Fresh Hits, and Crossing Algorithms*, [blog] „G-Squared Interactive Marketing Blog” (October 27 2014), [online] <http://www.gsqi.com/marketing-blog/penguin-3-analysis-findings> [dostęp 30.05.2016]; T. Sadowski, *Penguin 2.1 – zwany również Pingwinem 5 – co to jest, skąd się wzięło i jak sobie z nim radzić*, GoogleWatchers SEO BLOG (6 października 2013), [online] <http://googlewatchers.info/google/penguin-2-1-co-to-jest> [dostęp 30.05.2016].

¹³ D. Sullivan, *RIP Google PageRank score: A retrospective on how it ruined the web*, Search Engine Land [blog] (March 9, 2016), [online] <http://searchengineland.com/rip-google-pagerank-retrospective-244286> [dostęp 30.05.2016].

¹⁴ S. Lee, W. Jang, E. Lee, S. G. Oh, *Search engine optimization: A case study using the bibliographies of LG Science Land in Korea*, „Library Hi Tech” 2016, vol. 34, iss. 2, s. 197–206.

¹⁵ M. Gasparotto, *Search Engine Optimization for the Research Librarian: A Case Study Using the Bibliography of U.S. Latina Lesbian History and Culture*, „Practical Academic Librarianship: The International Journal of the SLA Academic Division” 2014, vol. 4, no. 1, s. 15–34, [online] <https://journals.tdl.org/pal/index.php/pal/article/view/6971> [dostęp 25.05.2016].

¹⁶ D. Scott, *White Hat Search Engine Optimization (SEO): Structured Web Data for Libraries*, „Partnership: The Canadian Journal of Library and Information Practice and Research” 2015, 10(1); D. Onaifo, D. Rasmussen, *Increasing libraries' content findability on the web with search engine optimization*, „Library Hi Tech” 2013, vol. 31, iss. 1, s. 87–108 [dostęp 25.05.2016]; A. Kenning, P. O'Brien, B. Rossmann, *Managing search engine optimization: An introduction for library administrators*, „Journal of Library Administration” 2013, vol. 53, no. 2–3, s. 177–188.

Słowa kluczowe

Jednym z głównych obszarów analiz SEO jest badanie słów kluczowych. Zasadniczym celem tych analiz jest udzielenie odpowiedzi na dwa pytania: (A) w jaki sposób użytkownicy wyszukiwarek trafiają do naszych zasobów (jakich słów kluczowych używają, jak formułują zapytania) oraz (B) w jaki sposób nasze podstrony są oznaczone tymi słowami kluczowymi (jak gęsto i w jakich sekcjach HTML)? Strategia poszukiwania odpowiedzi na każde z tych pytań jest inna, wymaga też użycia innych metod i narzędzi oraz materiału empirycznego¹⁷.

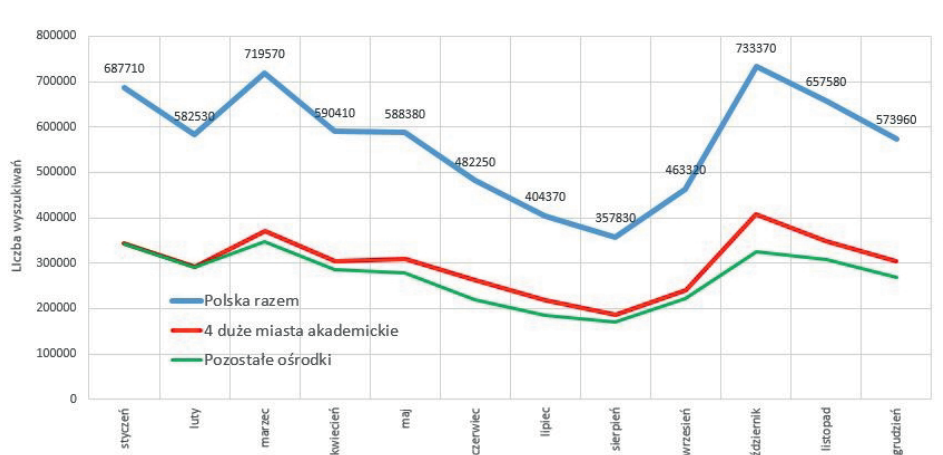
Badanie słów kluczowych

W pierwszym przypadku danych empirycznych do tego typu analiz (w literaturze określanych jako *keyword research*) dostarcza narzędzie Google AdWords, zaś postępowanie badawcze sprowadza się do wykorzystania rozmaitych metod statystycznych i technik używanych przez specjalistów z zakresu badania rynku. Warto równocześnie dodać, że badacz słów kluczowych znajduje się w sytuacji komfortowej, bowiem dzięki Google AdWords ma dostęp do kompletnego materiału badawczego (za ostatnie 2 lata), a zatem nie musi ograniczać się do próby badawczej. Aby wyselekcjonować ruch skierowany do portali bibliotecznych, w pierwszym kroku należy wyłonić korpus słów kluczowych, które pozostają z nimi w związku kognitywnym. W naszym przypadku był to zabieg nieskomplikowany, gdyż uznano, że wystarczy robocze założenie, że każdy biblioteczny serwis zawiera w swojej nazwie słowo *biblioteka* (nazwa rodzajowa) oraz kilka innych słów określających nazwę własną (*miejska, wojewódzka* itp.), i/lub pełniącą funkcję marki (*Jagiellońska, Narodowa, Norwida, AGH* itd.), ponadto lokalizację (*Kraków, Warszawa* itd.).

Dzięki danym Google AdWords udało się ustalić, że w 2014 roku polscy internauci wprowadzili do wyszukiwarki 6 841 280 zapytań zawierających słowo *biblioteka* (wykres 3, seria niebieska). Dodajmy, że badano wyłącznie ruch płynący z obszaru Polski, w języku polskim. W skali miesiąca takich zapytań kierowano średnio ok. 570 000, choć w poszczególnych miesiącach ruch podlegał pewnym wahaniom: styczeń – 687 710; luty – 582 530, marzec – 719 570, kwiecień – 590 410, maj – 588 380, czerwiec – 482 250, lipiec – 404 370, sierpień – 357 830, wrzesień – 463 320, październik – 733 370, listopad – 657 580, grudzień – 573 960. Warto zauważyć, że był on wyraźnie skorelowany z rokiem akademickim/szkolnym (nasilenie na początku semestrów i wyraźny spadek w okresie wakacyjnym), co jest wyraźną przesłanką, że duża liczba zapytań miała związek z kształceniem i była kierowana przez studentów i młodzież szkolną.

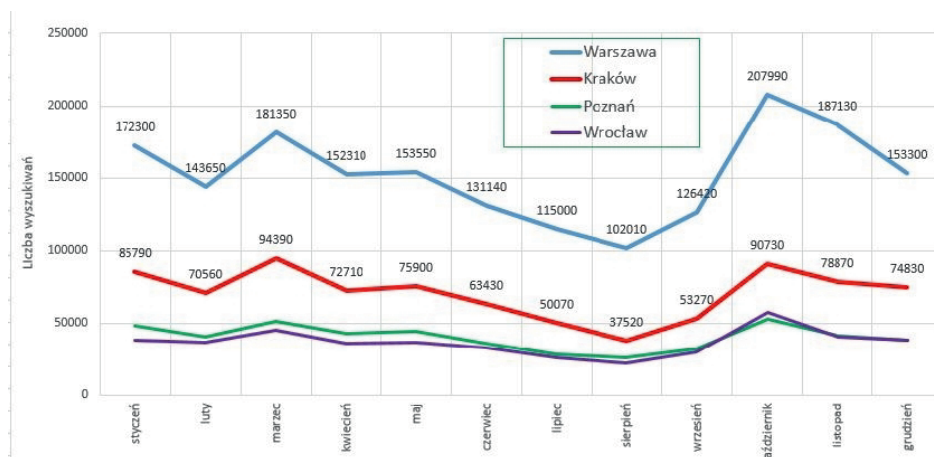
Wiele interesujących danych przynosi analiza geografii zapytań. Bliższa analiza zapytań przekonuje, że ruch skierowany do serwisów bibliotecznych rozpada się na dwie duże, porównywalne grupy: 4 duże miasta akademickie (Warszawa, Kraków, Wrocław, Poznań – 3 593 010 zapytań rocznie) oraz pozostałe ośrodki (3 248 270) – por. wykres 3, serie: czerwona i zielona). Dodajmy, że charakterystyka obu grup jest zbliżona i naśladuje trend ogólny, co oznacza, że obie części są odseparowane

¹⁷ E. Enge, S. Spencer, J. Stricchiola, *SEO, czyli sztuka optymalizacji...*, s. 169–228.



Wyk. 3. Wyszukiwanie korpusu frazy biblioteka... w roku 2014 roku (Polska)

Źródło: obliczenia własne na bazie danych z Google AdWords. Dostęp 25 lutego 2015 r. n = 7 411 387 zapytań, 801 fraz



Wyk. 4. Wyszukiwanie korpusu frazy *biblioteka...* w roku 2014 roku (główne miasta)

Źródło: obliczenia własne na bazie danych z Google AdWords. Dostęp 25 lutego 2015 r. n = 3 593 010 zapytań, 801 fraz

i obrazują te same zjawiska. Konstatacja ta potwierdza ponadto, że dokonany na wstępie wybór ośrodków był słuszny.

Analiza ruchu internetowego, nakierowanego na serwisy biblioteczne w czterech dużych polskich miastach akademickich (wykres 4), prowadzi do dalszych konstatacji. Najwięcej badanych zapytań wychodziło z Warszawy (średnio 152 179 miesięcznie), nieco mniej z Krakowa (70 673) oraz po około 25% z Poznania i Wrocławia (odpowiednio: 40 051 i 36 515). Dane nie zaskakują. Zaobserwowany ruch zdaje się być mocno skorelowany z potencjałem naukowym, edukacyjnym i demograficznym oraz (prawdopodobnie) z liczbą bibliotek w poszczególnych miastach.



Wyk. 5. Główne słowa kluczowe współwystępujące z frazą *biblioteka* (Polska)

Źródło: obliczenia własne na bazie danych z Google AdWords. n = 7 411 387 zapytań, 801 fraz. Uwzględniono frazy występujące co najmniej 100 razy. Generator wizualizacji: <http://www.wordle.net/create>. [dostęp 25.02.2015].

Zdecydowanie więcej informacji istotnych z punktu widzenia SEO przynosi analiza zgromadzonych fraz (tzw. korpusu słów kluczowych). W grupie 7 411 387 analizowanych zapytań Google AdWords zidentyfikował aż 801 różnych fraz zawierających słowo *biblioteka* (w różnych przypadkach gramatycznych i na różnych pozycjach w zapytaniu).

Lista wprowadzie nie jest kompletna, gdyż Google AdWords zlicza tylko frazy, które wystąpiły co najmniej 10 razy, niemniej ich zaniedbanie nie powinno zdeformować obrazu badań. W grupie 801 fraz użyto ponad 3000 innych słów (w tym spójników i skrótowców), które zgodnie z założeniem najczęściej określały nazwę własną (*publiczna*, *pedagogiczna*, *uniwersytecka* itp.), pełniły funkcję marki (*Raczyńskich*, *AGH*, *SGGW* itd.) oraz informowały o lokalizacji (*Gdańsk*, *Bydgoszcz* itd.). Wykaz 150 najczęściej używanych słów współwystępujących ze słowem *biblioteka* zobrazowano na wykresie 5 jako chmurę tagów.

Dalsza analiza pełnego korpusu z uwagi na jego rozmiar jest trudna do wykonania. Zatem do ustalenia dalszych zależności użyto lokalnego korpusu słów – wpisywanych przez użytkowników z Krakowa (wykres 6).

Z chmury tagów nietrudno odczytać, w jaki sposób użytkownicy kojarzą nazwy bibliotek i serwisów bibliotecznych. Główną rolę prócz słowa *biblioteka* pełnią słowa kojarzone z marką (*Jagiellońska*, *Podgórska*, *Krowoderska*), które w wielu przypadkach przejawiają skłonność do skracania (obok *Jagiellońska* w użyciu jest *BJ* i *UJ*). Dodajmy, że w przypadku bibliotek uczelnianych tendencja ta jest bardzo silna, stąd na czele rankingów widnieją niemal wyłącznie akronimy: AGH, UJ, PK, UEK,



Wyk. 6. Główne słowa kluczowe współwystępujące ze słowem *biblioteka* (Kraków)

Źródło: obliczenia własne na bazie danych z Google Adwords. $n=71\ 160$ zapytań, 800 fraz. Uwzględniono 150 najczęściej występujących słów. Generator wizualizacji: <http://www.wordle.net/create>. [dostęp 25.02.2015].

UP, UR, które odsyłają odpowiednio do: Akademii Górniczo-Hutniczej, Uniwersytetu Jagiellońskiego, Politechniki Krakowskiej, Uniwersytetu Pedagogicznego, Uniwersytetu Ekonomicznego i Uniwersytetu Rolniczego. W przypadku rodzajowych nazw własnych użytkownicy na ogół podają pełną frazę lub słowo (*miejaska, pedagogiczna*). Czasami w świadomości internautów funkcjonują też inne skojarzenia, np. *rajska*, które dotyczą lokalizacji (ul. Rajska, gdzie znajduje się siedziba Wojewódzkiej Biblioteki Publicznej w Krakowie) oraz jej domeny: **.rajska.info*.

Dokładniejszy obraz rozproszenia fraz przedstawiono w tabeli 2, gdzie uszeregowano je malejąco według frekwencji (słowo *biblioteka* w drodze wyjątku postawione na czele listy). Z zestawienia nietrudno odczytać, że ranking przypomina krzywą logarytmiczną o dużej progresji (zob. wykres 7). Trzy pozycje z czoła listy generują 20,47% ogólnej liczby zapytań, jeśli uwzględnimy 7 będzie ich 35,71%, zaś do osiągnięcia połowy ogólnej liczby zapytań wystarczy 14 fraz.

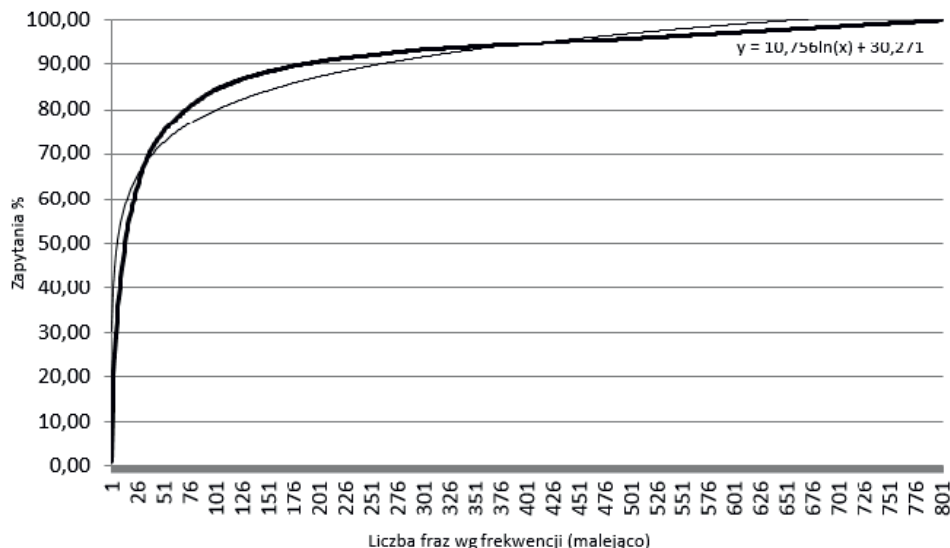
Tab. 2. Główne frazy ze słowem kluczowym *biblioteka* dla wyszukiwań z Krakowa

	Keyword	Średnio wyszukiwań miesięcznie		Razem
		N	%	Σ%
1	biblioteka	880	1,19	1,19
2	biblioteka jagiellońska	9900	13,35	14,54
3	biblioteka rajska	4400	5,93	20,47
4	biblioteka agħ	3600	4,85	25,32
5	biblioteka uj	2900	3,91	29,23
6	biblioteka pk	2400	3,24	32,47

7	biblioteka uek	2400	3,24	35,71
8	biblioteka podgórska	1600	2,16	37,86
9	krowoderska biblioteka publiczna	1600	2,16	40,02
10	rajska biblioteka	1600	2,16	42,18
11	podgórska biblioteka publiczna	1600	2,16	44,34
12	biblioteka up	1600	2,16	46,49
13	nowohucka biblioteka publiczna	1300	1,75	48,25
14	uj biblioteka	1300	1,75	50,00
15	biblioteka krowoderska	1300	1,75	51,75
	Pozostałe 785 frazy łącznie	35780	48,25	100,00
	Razem	74160	100,00	

Źródło: obliczenia własne na bazie danych z Google AdWords. Dostęp 25 lutego 2015 r. n = 71 160 zapytań, 800 fraz.

Konstatacja ta ma duże znaczenie dla twórców serwisów, w szczególności web-writerów, gdyż wskazuje, że aby skutecznie pozyskać ruch wystarczy wykonać kilka podstron pozycjonowanych na najczęściej używane frazy i słowa kluczowe. Aby poprawnie wypozycjonować stronę startową najczęściej wystarczą 2 słowa (*biblioteka* i akronim), np. witryna BG Uniwersytetu Ekonomicznego 97% ruchu pozyskuje na słowach *biblioteka* i *uek*; Biblioteka Politechniki Krakowskiej – 89% (*biblioteka*, *pk*); BG Akademii Górniczo-Hutniczej – 87% (*biblioteka*, *agh*), BG Uniwersytetu Pedagogicznego – 82% (*biblioteka*, *up*); BG Uniwersytetu Rolniczego – 74% (*biblioteka*, *ur*). W niektórych przypadkach występują inne kombinacje, np. Wojewódzka Biblioteka



Wyk. 7. Frekwencja fraz przy wyszukiwaniu serwisów bibliotecznych z Krakowa

Źródło: obliczenia własne na bazie danych z Google AdWords. Dostęp 25 lutego 2015 r. n = 71 160 zapytań, 800 fraz.

Publiczna w Krakowie – 68% ruchu pozyskuje na słowach: *biblioteka, rajska*. Czasami zaś istnieje konieczność użycia większej liczby słów, np. serwis Biblioteki Jagiellońskiej 98% ruchu skupia się na 4 słowach (*biblioteka, jagiellońska, uj, bj*).

Badania pozwoliły dodatkowo uzyskać kilka charakterystyk przekrojowych użytkowników serwisów bibliotecznych. Statystyczny czytelnik korzysta najczęściej z komputera stacjonarnego lub notebooka (80,6%), rzadziej wykorzystuje do tego celu smartfon (15,4%), a najrzadziej tablet (4%). Rozkład geograficzny zapytań świadczy natomiast o lokalności polskich serwisów bibliotecznych, gdyż dominujący ruch pochodzi z Polski (87,7%) i tylko nieco ponad 12% zapytań płynie z państw ościennych.

Stosowanie słów kluczowych

Kolejnym krokiem w analizie SEO jest ustalenie, na jakie słowa kluczowe twórcy serwisów pozycjonują własne podstrony. Analizy tego typu określa się w anglojęzycznej literaturze, jako *on-page keyword usage and keyword density*. W poprawnie zaprojektowanym serwisie każda z podstron jest pozycjonowana na kilka określonych słów kluczowych, przy czym jej zawartość (treść) powinna być z nimi zgodna, tzn. powinien zachodzić pomiędzy nimi związek kognitywny. Warto jednocześnie dodać, że od 2011 roku tj. od wprowadzania poprawki Google Panda Update strony przeładowane nieadekwatnymi słowami kluczowymi są dość skutecznie identyfikowane i zaniżane w rankingu. Strony przyjazne SEO tworzy się zwykle w dwu krokach. Najpierw redaktor przygotowuje materiał zgodny z tradycyjnymi regułami pisarskimi (redakcja merytoryczna), następnie do pracy wkracza webwriter, który w odpowiednich miejscach i w odpowiednim kontekście dodaje słowa kluczowe. Tradycyjnym miernikiem poprawnego pozycjonowania jest tzw. gęstość słów kluczowych (*keyword density* [KD]), którego optymalna wartość winna oscylować w przedziale 1–3% [według niektórych 1–4,5%]¹⁸. Duże znaczenie ma także umiejscowienie słów kluczowych w kodzie HTML. Największe znaczenie przypisuje się wystąpieniom słów w tytule strony (<title>) oraz nagłówkach strukturalnych (<H1>, <H2>, <H3>), mniejsze zaś, jeśli wystąpią w sekcji <body> i w kotwicach linków (anchory). Słowa i frazy użyte w <body> wzmacnia się ponadto tagami strukturalnymi, m.in. poprzez pogrubienie lub kursywę .

Ten aspekt analizy z uwagi na pracochłonność ograniczono do zbadania trzech podstron każdego serwisu: strony startowej (tzw. „czołówki”, która była zwykle stroną najczęściej wywoływana z SERP-ów [TOP1]) oraz dwu kolejnych podstron znajdujących się najwyżej w rankingu (TOP2 i TOP3). Wyniki analizy zestawiono w tabeli 3, wskazując kolejno dla każdej badanej frazy lub słowa kluczowego:

¹⁸ P. Patil Swati, B. V. Pawar, S. Patil Ajay, *Search Engine Optimization: A Study*, „Research Journal of Computer and Information Technology Sciences” 2013, 1(1), s. 10–13, [online] http://www.retawprojects.com/uploads/search_engine.pdf [dostęp 16.05.2016]; F. Wang, Y. Li, Y. Zhang, *An Empirical study on the Search Engine Optimization Technique and Its Outcomes*, [w:] *Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), Proceedings of 2nd International Conference on AIMSEC*. Piscataway 2011, s. 2767–2770 [dostęp 4.04.2016].

miejsca, gdzie wystąpiło (*found*), liczbę powtórzeń na stronie (*repeat*) oraz jej gęstość (*keyword density*).

Tab. 3. Użycie słów kluczowych na stronach startowych wybranych polskich serwisów bibliotecznych

Miasto	Biblioteka	Total word	Found in	Keyword density			
				A. „biblioteka”		B. „marka”	
				N	%	N	%
Warszawa	BN	413	TDK	11	2,66	5	1,21
	BG PW	305	TDK	15	4,92	5	1,64
	BUW	1033	TDK	34	3,29	20	1,94
	BPmsW	1629	TDK	39	2,39	24	1,47
Kraków	BJ	795	T	35	4,40	18	2,26
	BG AGH	94	T	3	3,19	5	5,32
	WBP Krak.	1825	TDK	43	2,36	24	1,32
Poznań	B. Rac.	460	TDK	12	2,61	12	2,61
	BUAM	499	TDK	23	4,61	9	1,80
	B. PPozn.	1033	T	34	3,29	6	0,58
Wrocław	CWINT	603	TDK	10	1,66	7	1,16
	BUWr	667	T	18	2,70	11	1,65
	MBP Wr.	1210	DK	16	1,32	30	2,48

Narzędzie: SEOqake i SEORCH [pomiar 2015.02.30]. Uwzględniono słowa kluczowe w różnych przypadkach gramatycznych. Oznaczenia stosowane w kolumnie Found in: K – słowo użyte w sekcji *meta keywords*; T – w sekcji *meta title*; D – w sekcji *meta description*.

Zestawienie zdaje się przekonywać, że większość redakcji prawidłowo pozycjonuje „czołówkę” własnego serwisu, choć nie zawsze robi to w sposób optymalny. Wprawdzie KD dla frazy *biblioteka...* w dwu przypadkach (BG Politechniki Warszawskiej – 4,92% i Biblioteka Uniwersytecka w Poznaniu – 4,61%) przekroczyło o ponad 50% zalecane maksimum, wszelako w pozostałych przypadkach gęstość przyjmowała poprawne wartości. Była ona też (z wyjątkiem AGH – 3) powtarzana większą od optymalnej (10–43) liczbę razy, lecz usprawiedliwiało to ich rozmiar (*total word*). Zbliżone uwagi można sformułować w odniesieniu do pozycjonowania marki, która (nie licząc Bibliotek: AGH i Politechniki Poznańskiej) oscylowała w granicach 1,21–2,61. Nie oznacza to, że badane strony były w pełni zoptymalizowane. Wprawdzie prawie wszyscy redaktorzy umieścili główne słowa kluczowe w tytule stron (T), lecz jedynie 9 z 13 nie umieściło ich w metatagach <description> (D). Z obrazu tego niekorzystnie wyłamuje się tylko Miejska Biblioteka Publiczna we Wrocławiu, która nie umieściła słów kluczowych w metatagu <title>.

Do kilku wniosków prowadzi też analiza doboru dwu głównych słów kluczowych na stronach (TOP1, TOP2, TOP3). Tylko nieco ponad połowa witryn dobrała je optymalnie. Na początku przyjrzyjmy się bibliotekom akademickim. Przypomnijmy, że zgodnie z ustaleniami z poprzedniej części artykułu największy ruch do ich witryn płynie w wykorzystaniem pary słów: *biblioteka* + skrót marki. Zasadę tę zastosowały jedynie dwa serwisy (BUW i BUWr.), w kolejnych trzech (BG PW, BJ, BUAM) zamiast skrótowca podano w pełnej formie słowo znaczące wyjęte z nazwy. W pozostałych przypadkach dobór słów był nieadekwatny: brak słowa *biblioteka* (AGH,

Tab. 4. Widoczność słów kluczowych na witrynach wybranych polskich serwisów bibliotecznych

	Biblioteka	Słowo kluczowe					Różnica TOP2- TOP3
		TOP1		TOP2		TOP3	
		Word	Visibility	Word	Visibility	Visibility	
Warszawa	BN	biblioteka	58	narodowa	53	52	1
	BG PW	biblioteka	66	politechniki	54	54	0
	BUW	buw	76	biblioteka	62	51	11
	BPmsW	powiat	87	biblioteka	82	78	4
Kraków	BJ	biblioteka	78	Jagiellońska	61	51	10
	BG AGH	główna	61	agh	50	48	2
	WBP Krak.	krakowie	79	biblioteka	75	63	12
Poznań	B. Racz.	biblioteka	53	raczyńskich	53	38	15
	BUAM	biblioteka	96	uniwersytecka	79	77	2
	B. PPozn.	put	42	poznan	42	41	1
Wrocław	CWINT	pwr	49	wroc	38	37	1
	BUWr.	biblioteka	55	buwr	51	50	1
	MBP Wr.	filia	118	biblioteka	89	52	37

Źródło: obliczenia własne. Narzędzie: SEORCH [pomiar 2015.02.30]. Uwzględniono słowa kluczowe w różnych przypadkach gramatycznych.

B. PPozn.). Wśród bibliotek nieakademickich sytuacja kształtowała się podobnie: w pełni poprawnie dobrały słowa trzy księżnice (BN, WBP Krak., B.Racz.), zaś w dwu przypadkach w TOP1 zabrakło słowa znaczącego z nazwy (BPmsW pozycjonowała się na słowo *powiat*, zaś MBP Wr. – na słowo *filia*). Można przypuszczać, że niektóre niedoskonałości są skutkiem niesamodzielnosci badanych witryn (biblioteki akademickie), wskutek czego część metadanych wymusza główny serwis uczelniany lub użyty CMS (B. PPozn., CWINT).

Bardziej obiektywnych danych na temat poprawności pozycjonowania dostarcza metodologia stosowana w serwisie SEORCH, który określa widoczność stron poprzez specjalny wskaźnik *visibility* (tab. 4). Zgodnie z zastosowanym algorytmem optymalne pozycjonowanie wymaga, aby jego wartość spełniała łącznie dwa warunki: $TOP1 \geq 70$ oraz różnica $(TOP2 - TOP3) \geq 40$. W świetle powyższych ustaleń ocena analizowanych witryn wypada słabo. Wprawdzie wiele serwisów stosunkowo dobrze sytuuje się pod względem widoczności TOP1: blisko połowa (6) legitymuje się wskaźnikiem większym od zalecanego (≥ 70), w kolejnych pięciu przybrał on wartość nieco zaniżoną ($70 > 50$), ale akceptowalną (BG PW, BG AGH, BUWr, BN, B. Racz.), i tylko w dwóch przypadkach był zdecydowanie za niski (B. PPozn., CWINT). Dużo gorzej wypada analiza, w której bierzemy pod uwagę drugi warunek (różnica $TOP2 - TOP3 \geq 40$), nie spełnia jej bowiem żaden analizowany serwis.

Zbyt mała próba badawcza nie uprawnia do szerszych generalizacji, lecz rozmaite poszlaki pozwalają stwierdzić, że redakcje polskich serwisów bibliotecznych nie przywiązują większej wagi do zasad webwritingu. Można też zaryzykować tezę, że skoro w badanym korpusie prawie połowa witryn (skądinąd największych polskich księżnic) zawiera jakieś uchybienia i niedoskonałości zapewne ich odsetek w ogólnym korpusie bibliotek polskich jest większy.

Czynniki *on-page* (zawartość i technologia)

Osobnym obszarem badań pozycjonowania stron jest analiza czynników zależnych od webmastera (w literaturze określanych zwykle *on-page content and technology*). Już w najstarszych oficjalnych poradnikach Google dla webmasterów (obecnie Google Search Console)¹⁹, podręcznikach SEO²⁰, analizach naukowych²¹ oraz w większości serwisów raportujących (np. SEORCH)²² podkreślano wagę czterech czynników optymalizacyjnych. Były to kolejno: tytuł w sekcji <head> kodu HTML (<title>... </title>), opis strony (<meta name=description content=... />), tekst strony (<body>... </body>) oraz liczba linków wychodzących oraz wewnętrznych. Ich optymalne wartości zmieniały się w czasie, zaś aktualnie dominuje przekonanie, że optymalny tytuł winien mieć maksymalnie 64 znaki (i zawierać słowa kluczowe), opis strony – 160 znaków (i zawierać słowa kluczowe), tekst strony – co najmniej 400 słów oraz pewną liczbę linków wewnętrznych (do kilkudziesięciu) oraz linków wychodzących²³. W ostatnich latach w praktyce audytu SEO pojawił się także kolejny czynnik – obecność tzw. przyjaznych linków, czyli tworzonych przez CMS-y końcówek adresów URL, które zawierają słowa kluczowe. Rola poszczególnych czynników jest dość dobrze udokumentowana w podręcznikach, więc nie ma konieczności przytaczania uzasadnień. Warto natomiast podkreślić, że w ostatnich rankingach Searchmetrics i MOZ²⁴ ich rola systematycznie maleje i według najnowszego rankingu MOZ grupa tytułowych czynników (*Page-Level Keyword-Agnostic Features*) znalazła się dopiero na czwartej pozycji²⁵.

Tab. 5. Główne czynniki *on-page* wybranych polskich serwisów bibliotecznych

Miasto	Biblioteka **	Długość tytułu [znaki]	Meta desc. [znaki]	Tekst w body [znaki]	Linki wewnętrzne [N]	Linki zewnętrzne [N]	Przyjazne linki
Warszawa	BN	19	140	413	89	25	Tak
	BG PW	51	43	305	47	12	Tak
	BUW	36	36	1033	61	24	Nie
	BPmsW	82	108	1629	209	20	Tak

¹⁹ Google Centrum Pomocy Search Console, <https://support.google.com/webmasters> [dostęp 30.05.2016].

²⁰ B. Danowski, M. Makarczuk, *Pozycjonowanie i optymalizacja stron WWW*, Wydawnictwo Helion, Gliwice 2007 (wyd. 2 – 2009, wyd. 3 – 2011).

²¹ P. Patil Swati, B. V. Pawar, S. Patil Ajay, *Search Engine Optimization: A Study*, „Research Journal of Computer and Information Technology Sciences” 2013, 1(1), s. 10–13, [online] http://www.retawprojects.com/uploads/search_engine.pdf [dostęp 16.05.2016].

²² SEORCH, [online] <http://seorch.eu/> [dostęp 4.04.2016].

²³ E. Enge, S. Spencer, J. Stricchiola, *SEO, czyli sztuka optymalizacji...*

²⁴ M. Tober, L. Hennig, D. Furch, *SEO Ranking Factors and Rank Correlations 2014*, Searchmetrics, San Mateo 2014, [online] <http://www.searchmetrics.com/knowledge-base/ranking-factors-2014/> [dostęp 4.04.2016]; M. Tober, D. Furch, K. Londenberg, L. Massaron, J. Grundmann, *Search Ranking Factors and Rank Correlations 2015*, Searchmetrics, San Mateo 2015, [online] <http://www.searchmetrics.com/knowledge-base/ranking-factors/> [dostęp 4.04.2016].

²⁵ *Search Engine Ranking Factors 2015. Expert Survey and Correlation Data*, MOZ Blog, [online] <https://moz.com/search-ranking-factors> [dostęp 4.04.2016].

Kraków	BJ	68	0	795	86	29	Tak
	BG AGH	17	0	94	26	7	Nie
	WBP Krak.	87	685	1825	256	42	Tak
Poznań	B. Racz.	34	34	460	87	6	Nie
	BUAM	43	35	499	63	16	Nie
	B. PPozn.	73	0	1033	278	12	Nie
Wrocław	CWINT	146	23	603	15	12	Tak
	BUWr.	37	0	667	117	30	Tak
	MBP Wr.	11	41	1210	161	21	Tak

Źródło: obliczenia własne. Narzędzie: SEORCH [pomiar 2015.02.30]. ** Rozwiązania skrótów – zob. tab. 1.

Analiza wybranych polskich serwisów bibliotecznych (tab. 5) przekonuje, że powyższe cztery zasady są raczej znane i stosowane. Prawie połowa testowanych witryn ma optymalnie dobrany tytuł [BG PW, BUW, BJ, B. Racz., BUAM, BUWr], niektóre tytuły są wprawdzie nieco za długie, choć mieszczą się w dopuszczalnym zakresie [BPmsW, WBP Krak., B. PPozn.], jedynie w czterech przypadkach jest on ewidentnie za krótki (BN, BG AGH, MBP Wr.) lub za długi (CWINT.). Niepokojącą praktyką są natomiast liczne błędy w sekcji <meta name=description.... Co czwarty testowany serwis pomija ten element [BJ, BG AGH, P PPozn., BUWr.], często bywa on za krótki [BG PW, BUW, B. Racz., BUAM, CWINT, MBP Wr.] lub za długi [WBP Krak.]. W efekcie tylko w dwu przypadkach ma on poprawne parametry [BN, BPmsW]. Dodajmy, że analizowany tag nie jest wprawdzie bezpośrednio uwzględniany w algorytmie wyszukiwarki, lecz pełni funkcję psychologiczną, gdyż tekst opisu (z pogrubieniem słów kluczowych) pojawia się na stronach SERP-ów.

Duże znaczenie dla poprawnego funkcjonowania serwisu ma optymalnie dobrana objętość (tekst zamknięty w sekcji <body>). Dominują opinie, że najlepiej, aby oscylował w przedziale: 400–800 słów. Testy udowodniły, że niespełna połowa stron startowych badanych witryn ma optymalnie dobraną długość [BN, BJ, B.Racz., BUAM, CWINT, BUWr.], w pięciu przypadkach był on zdecydowanie za długi jak na „czołówkę” [BUW, BPmsW, WBP Krak., B. PPozn., MBP Wr.], zaś w dwu – zdecydowanie za krótki [BG PW, BG AGH]. Świadczy to o tym, że webmasterzy nie doceniają tego parametru, choć istnieje wiele dowodów, że witryny zbyt krótkie mogą zostać zakwalifikowane przez Google jako tzw. cienka treść (*thin content*) i stracić w rankingu²⁶.

Wszystkie testowane serwisy umieściły na stronie głównej wiele linków wewnętrznych (*links internal*). Fakt ów należy ocenić pozytywnie z dwu powodów. Po pierwsze – kotwice linków (*anchor text*) skutecznie pozycjonują właściwe słowa kluczowe, po wtóre – linki wewnętrzne na stronie startowej ułatwiają nawigację dla użytkowników i robotów. Podobny obraz przynosi analiza linków wychodzących (*links external*). W wielu przypadkach redakcje chętnie linkują do partnerów bibliotecznych, a tym samym pomagają wyszukiwarce w poprawnym ulokowaniu serwisu w sieci zależności tematycznych (tworzą je głównie linki zewnętrzne). Duże znaczenie dla poprawnego pozycjonowania stron mają również tzw. przyjazne linki, gdyż pozwalają na umieszczenie słów kluczowych bezpośrednio w URL. Wykonana

²⁶ E. Enge, S. Spencer, J. Stricchiola, *SEO, czyli sztuka optymalizacji...*, s. 286.

analiza przekonuje jednak, że właściwość ta (powszechnie dostępna w każdym współczesnym CMS-e) jest często wykorzystywana (8 z 13), wszelako prawie 40% serwisów nie zauważa jej zalet [BUW, BG AGH, B.Racz., BUAM, B.PPozn.].

Tab. 6. Analiza struktury kodu wybranych polskich serwisach bibliotecznych

Miasto	Biblioteka**	Elementów HTML	H1	H2	H3	Błędy struktury	<p>		
Warszawa	BN	373	1	6	0	Nie	9	3	33
	BG PW	387	0	4	4	Tak	4	15	12
	BUW	633	0	0	0	Tak	26	2	4
	BPMSW	1030	0	0	8	Tak	67	2	33
Kraków	BJ	535	5	1	0	Tak	11	9	15
	BG AGH	290	1	4	0	Tak	14	0	6
	WBP Krak.	1439	0	6	8	Tak	102	33	27
Poznań	B. Racz.	699	0	0	1	Tak	9	18	0
	BUAM	404	1	3	0	Nie	1	2	11
	B. PPozn.	1192	0	3	0	Tak	34	16	30
Wrocław	CWINT	437	0	0	0	Tak	12	9	1
	BUWr.	758	0	9	0	Tak	2	1	30
	MBP Wr.	986	0	7	13	Tak	31	3	21

Źródło: obliczenia własne. Narzędzie: SEORCH [pomiar 2015.02.30]. ** Rozwiązania skrótów – zob. tab. 1.

Ważnym elementem audytu SEO jest analiza struktury kodu HTML oraz elementów pozycjonujących. W strukturze poprawnie zaprojektowanej strony winien wystąpić obowiązkowo jeden nagłówek H1, który może (ale nie musi) się dzielić na H2, te z kolei na H3 itd. Dłuższe strony powinny mieć jasno określoną strukturę. Dodajmy, że nie chodzi wyłącznie o wygodę czytelnika, dla którego lektura ustrukturalizowanego tekstu jest łatwiejsza i przyspiesza selekcję, lecz w pierwszym rzędzie o przetwarzanie maszynowe i indeksowanie. Wykonane badania (tab. 6) ujawniają, że webmasterzy serwisów bibliotecznych nie przywiązują większej wagi do struktury witryn. Zdecydowana większość stron startowych ma poważne błędy w strukturze: w 9 na 13 przypadków brakuje H1 lub występuje on wielokrotnie, brakuje tagów niższego poziomu (H2/H3) oraz wyróżnień (,). W efekcie tylko strukturę dwu serwisów [BN, BUAM] można uznać za poprawną i optymalnie ukształtowaną.

Tab. 7. Obrazy i hiperłącza w strukturze wybranych polskich czasopism humanistycznych

Miasto	Biblioteka**	Obrazy			Hiperłącza wewnętrzne / zewnętrzne		
		N	Brak ALT	Brak Title	N	Brak Anchor	Brak Title
Warszawa	BN	24	1	17	114	0	104
	BG PW	10	5	7	59	0	55
	BUW	15	9	15	85	2	81
	BPmsW	34	1	34	229	0	192
Kraków	BJ	12	9	11	115	0	103
	BG AGH	3	0	3	33	7	31
	WBP Krak.	38	3	16	298	0	202

Poznań	B. Rac.	76	22	56	93	14	68
	BUAM	5	2	5	79	1	61
	B. PPozn.	76	52	54	290	0	270
Wrocław	CWINT	11	11	11	27	0	27
	BUWr.	28	21	18	147	0	87
	MBP Wr.	42	30	34	182	1	176

Źródło: obliczenia własne. Narzędzie: SEORCH [pomiar 2015.02.30]. ** Rozwiązania skrótów – zob. tab. 1.

Duże znaczenie dla poprawnego funkcjonowania stron www mają dodatkowe atrybuty towarzyszące tagom odpowiadającym za wstawienie grafik oraz hiperłączy. W przypadku grafiki jest to element ATL, czyli opis alternatywny wyświetlany w przypadku, gdy obrazek nie może zostać wczytany. Dzięki temu, można korzystać z graficznej nawigacji, nawet gdy obrazek nie został załadowany. ALT jest ponadto wykorzystywany przez specjalne programy dla osób niewidomych i niedowidzących. Rola tego atrybutu dla indeksowania i przeglądarek jest jeszcze większa. Ponieważ roboty wyszukiwarek nie analizują obrazów pod względem ich zawartości, atrybut ALT jest jedynym substytutem obrazu, który indeksują. W ślad za tym atrybut ten przejmuje rolę słowa kluczowego w wyszukiwaniu obrazów. Analiza poprawności użycia tego atrybutu w badanym korpusie serwisów (tab. 7) pozwala na pozytywną ocenę witryn, gdyż tylko w sporadycznych przypadkach wystąpiły braki.

Podobną rolę w hiperłączach pełni tekst kotwicy (*anchor text*), czyli ciąg znaków znajdujący się pomiędzy otwierającym i zamykającym znacznikiem linku. Anchory w sposób bezpośredni wpływają na słowa kluczowe, na które pozycjonowany jest dany URL. Teksty te winny być starannie dobrane i zgodne z pozycjonowaną frazą. Wykonany test (tab. 7) przekonuje, że redakcje polskich serwisów bibliotecznych doceniają rolę anchorów i powszechnie go stosują. Warto jednocześnie zwrócić uwagę, że związek anchorów i adresów URL dotyczy wszystkich kategorii linków, zarówno prowadzących do naszego serwisu, gdzie webmaster zna słowa, na jakie chce pozycjonować własną stronę, jak i zewnętrznych, gdzie webmaster arbitralnie dopiera słowo, na jakie pozycjonuje zewnętrzną stronę (inny serwis).

Następnym elementem audytu SEO są nowe standardy. Na obecnym etapie rozwoju sieci www warto sprawdzić tylko niektóre (m.in. HTML5, Open Graph Protocol i mikroformaty). Od 2008 roku rozpoczęto testowanie nowej wersji języka do tworzenia i prezentowania stron internetowych (HTML5) i już w październiku 2014 roku osiągnęła ona status rekomendacji W3C. Od początku prac specyfikacja ta znalazła wielu zwolenników, głównie ze względu na szereg nowych możliwości (m.in. umożliwiała obsługę struktury na poziomie kodu). Z kolei standard OGP (Open Graph Protocol) ma związek w sieciami społecznymi i jego użycie sprawia, że witryny są wyświetlane i przetwarzane w portalach społecznościowych w sposób kontrolowany. Mikroformaty natomiast mają za zadanie wzbogacić istniejące dane w sposób zrozumiały zarówno dla maszyn, jak i ludzi (np. gwiazdki z oceną filmów); rozwijają semantykę kodowania, dzięki czemu łatwiejsze stają się przetwarzanie i ekstrakcja zasobów dostępnych w sieci. Wykonana analiza stosowania nowych standardów nie prowadzi do budujących wniosków. Wprowadzie na czołówkach 5 z 13 serwisów w deklaracjach kodu widnieje element `<!DOCTYPE html>` (czyli użycie HTML5), niemniej tylko w dwu przypadkach są wykorzystywane jego możliwości (BPmsW, MBP Wr.). Jeszcze gorzej prezentuje się stan stosowania innych specyfikacji: tylko

jeden serwis używa microformatów (BG PW), zaś standardu OGP – żaden. Warto w tym miejscu zauważyć, że brak nowych standardów sam w sobie nie jest ułomnością, jednak niesie za sobą poważne konsekwencje. Przykładowo bez zastosowania HTML5 trudno jest zrealizować postulat responsywności, a w ślad za tym utrzymać kontakt z użytkownikami pracującymi na smartfonach i tabletach. Wśród testowanych serwisów tylko dwa (BPmsW, MBP Wr.) oferowały pełną responsywność i poprawnie funkcjonowały na urządzeniach mobilnych.

Tab. 8. Wybrane parametry techniczne polskich serwisów bibliotecznych

Miasto	Biblioteka**	Filesize [kB]	Code text ratio	Speed desktop	Robotx txt	Sitemap xml	Canonical tags
Warszawa	BN	21,31	15,26	61	1	1	0
	BG PW	27,87	7,22	46	0	0	1
	BUW	38,41	19,16	74	0	0	0
	BPMSW	14,93	14,93	55	1	0	0
Kraków	BJ	57,51	10,74	77	1	0	1
	BG AGH	27,54	2,54	84	1	0	0
	WBP Krak.	95,74	14,29	38	1	0	0
Poznań	B. Racz.	67,03	4,93	31	1	0	0
	BUAM	38,88	9,08	74	1	0	0
	B. PPozn.	94,22	8,00	43	1	0	0
Wrocław	CWINT	25,39	7,46	74	1	1	0
	BUWr.	59,86	12,16	69	1	1	0
	MBP Wr.	57,99	13,95	34	1	0	1

Źródło: obliczenia własne. Narzędzie: SEORCH [pomiar 2015.02.30]. ** Rozwiązania skrótów – zob. tab. 1.

Duże znaczenie dla skuteczności pozycjonowania serwisu ma kilka parametrów technicznych, które rzutują na wydajność (tab. 8). Nie ulega wątpliwości, że nawet najlepsza witryna nie zyska sympatyków, gdy będzie wolno wyświetlana. Podobne znaczenie ma wydajność także dla robotów wyszukiwarki. W kontekście SEO właściwość ta nabiera jeszcze większego znaczenia, gdyż badania przekonują, że wobec nadmiaru informacji zachowania użytkowników podlegają tzw. ekonomii uwagi²⁷. Niezdecydowany użytkownik wyszukiwarki znajdujący się na stronie wyników (SERP-ów) zwykle błyskawicznie dokonuje selekcji, więc szansa na wnikliwą lekturę z wolno wyświetlającej się strony jest niewielka. Podczas ładowania takiej strony zniecierpliwiony czytelnik albo znajdzie inne trafienie (inną witrynę), albo podąży dalej, zapominając wrócić do ładującego się wyniku. Analiza szybkości działania stron polskich portali bibliotecznych (tab. 8, kol. *speed desktop*) nie prowadzi do budzących konkluzji. Prawie połowa (6) serwisów nie osiągnęła nawet 60 punktów (w skali 100–stopniowej); dalszych 6 legitymowało się akceptowalną szybkością (60–79 punktów) i tylko jeden (BG AGH) osiągnął bardzo dobry rezultat (84). Słaba lub przeciętna wydajność badanych serwisów jest – jak się wydaje – w dużej mierze zawiniona przez webmasterów. Udowadnia to analiza dwu kolejnych parametrów:

²⁷ K. Piekarski, *Ekonomia percepcji. Mechanizmy selekcjonowania informacji w internecie* [praca doktorska], Uniwersytet Śląski w Katowicach, Katowice 2014, [online] <http://www.sbc.org.pl/Content/126980/doktorat3505.pdf> [dostęp 4.04.2016].

wielkość pliku (*filesize*) oraz stosunek tekstu do kodu (*code text ratio*). Według serwisu SEORCH optymalna wartość pierwszego nie powinna przekraczać 32 kB, zaś drugi powinien być wyższy niż 25%. W testowanym korpusie strony startowe, ponad połowy serwisów, co najmniej dwukrotnie przekroczyły pierwszy parametr (niektóre znacznie: WBP Krak., B. PPozn.), zaś właściwej proporcji kodu do tekstu nie miała żadna. Wynik ten oznacza, że witryny serwisów bibliotecznych są przeładowane wtyczkami, kodami śledzenia oraz modułami do współpracy z sieciami społecznymi.

Pełny audyt SEO wymaga, aby sprawdzić poprawność działania specyficznych plików sterujących. Pliki powinni tworzyć administratorzy, aby usprawnić współpracę witryn z *crawlerami* wyszukiwarek. Pierwszy (*robots.txt*) – wskazuje listę dozwolonych robotów oraz ich uprawnienia, drugi (*sitemap.xml*) – jest mapą witryny przygotowaną specjalnie dla robota. Testy udowodniły, że *robots.txt* jest niemal powszechnie stosowany (11 z 13 witryn), podczas gdy bagatelizowane są korzyści płynące z posiadania i aktualizacji pliku *sitemap.xml* (tylko 3 z 13).

Pokrewnym zagadaniem audytu jest zbadanie metod, jakimi serwis rozwiązuje problem powielających się treści (*duplicate content*) lub jej funkcjonowania na kilku różnych domenach. Zagadnienie jest bardzo istotne, gdyż reakcja robota wyszukiwarki na powielającą treść jest jednoznaczna – zostawia w indeksie tylko jeden rekord (zwykle najstarszy), zaś pozostałe usuwa z listy wyników²⁸. Ponieważ czasem pojawia się potrzeba stworzenia takich stron z powodów w pełni uzasadnionych (np. wersje do wydruku, mobilne, inne wersje językowe), zachodzi wówczas konieczność instruowania robotów o poprawnej tożsamości. Realizują to zadanie umieszczone w kodzie strony tagi kanoniczne (instrukcja *rel="canonical"*), informujące robota, która wersja strony niezależnie od wywołania jest wariantem głównym. Przeprowadzone badania (tab. 8, kol. *canonical tags*) przekonują, że rozwiązania to jest niemal nieznanym webmasterom badanych serwisów, gdyż używają go tylko 3 witryny (BG PW, BJ i MBP Wr.). W efekcie nie dysponują zabezpieczeniem wobec *duplicate content* i nie korzystają z innych dobroczynnych funkcji tagów kanonicznych, w szczególności scalania ruchu do tej samej strony wywołanej na kilka różnych sposobów (np. z *www* i bez *www*, z *http* i bez *http* itd.).

Linki przychodzące

Linki przychodzące (*backlinks*) wykorzystywano do tworzenia rankingów już w algorytmach najstarszych wyszukiwarek. Zasada ich wykorzystania była prosta. Opierała się na założeniu, że witryna do której prowadzi więcej linków jest wartościowsza od witryny, do której liczba linków jest mniejsza. Z czasem regułę tę zmodyfikowano i prócz liczby linków wyszukiwarki zliczały ich referencje (skąd pochodzą). Wówczas wartość linku pochodzącego z witryny, która posiada dużo linków przychodzących była większa niż z witryny, która ma ich mniej²⁹. Na zasadzie tej opierał się też PageRank (PR) opracowany w 1998 roku przez twórców Google. Idea PR była później wielokrotnie modyfikowana i z czasem jego rola w algorytmie znacząco spadła (dziś jest ledwie symboliczna). Modyfikacje miały związek przede wszystkim z nieetyczny-

²⁸ E. Enge, S. Spencer, J. Stricchiola, *SEO, czyli sztuka optymalizacji...*, s. 287–290.

²⁹ T. Seymour, D. Frantsvog, S. Kumar, *History of search engines*, „International Journal of Management and Information Systems” 2011, vol. 15, no. 4, s. 47–59, [online] <http://www.minot.com/tom/JIS.pdf> [dostęp 19.05.2016].

mi zachowaniami webmasterów, który chcąc wypozycjonować swój serwis tworzyli tzw. farmy linków, katalogi SEO i strony zaplecza. W 2012 roku Google wprowadził do algorytmu poprawkę Penguin Update (systematycznie udoskonalaną), która definitywnie rozwiązała problem nielegalnych linków³⁰. Aktualnie backlinki wciąż mają duże znaczenie w algorytmie, ale sposób ich zliczania i mapowania pozostaje tajemnicą. Różne poszlaki wskazują, że rolę PageRank przejął nowy wskaźnik (TrustRank), bazujący na rozwiązaniu stosowanym ongiś w wyszukiwarce Yahoo!³¹. Aktualnie w algorytmie Google backlinki wciąż mają duże znaczenie, ale zasada ich zliczania jest inna. Prawdopodobnie każdy link przychodzący nadaje wskazywanej witrynie część własnego autorytetu (*authority*), zaufania (*trust*) i rangi (*rank*). Na obecnym etapie rozwoju algorytmu to nie ilość, ale jakość linków odgrywa pierwszorzędną rolę. Taką supozycję potwierdza wiele serwisów analitycznych, m.in. Majestic SEO czy Ahrefs oraz społeczności eksperckie (MOZ, Serchmetrics), które tworzą własne odpowiedniki wskaźnika TrustRank. Warto jednocześnie dodać, że częściowo rolę backlinków przejęły inne czynniki algorytmu Google (sygnały sieci społecznych).

Tab. 9. Linki przychodzące do witryn wybranych serwisów bibliotecznych

Miasto	Biblioteka**	Ahrefs Backlink Checker					
		Backlinks Total	Referring Domain	Backlink Types			URL Rating
				Dofollow	Gov	Edu	
Warszawa	BN	1056200	2725	839801	242	42701	80
	BG PW	12398	289	12321	6	545	39
	BUW	191257	905	142423	153	3822	50
	BPMSW	47218	243	46555	9	3023	34
Kraków	Bib. Jag.	88890	1143	40429	140	18195	50
	BG AGH	3390	208	3286	1	554	33
	WBP Kr.	194231	299	193697	20	290	40
Poznań	B. Racz.	25356	305	25063	60	119	40
	BUAM	106149	452	105815	43	552	42
	B. PPozn	141563	124	140918	0	532	20
Wrocław	CWINT	5997	44	5855	1	252	18
	BUWr.	347269	500	288502	22	1463	43
	MBP Wr.	8170	285	7209	2	35	38

Źródło. Obliczenia własne na podstawie danych z serwisu Ahrefs [Ahrefs backlink checker], <https://ahrefs.com> Dostęp 30 lutego 2015 r. ** Rozwiązania skrótów – zob. tab. 1.

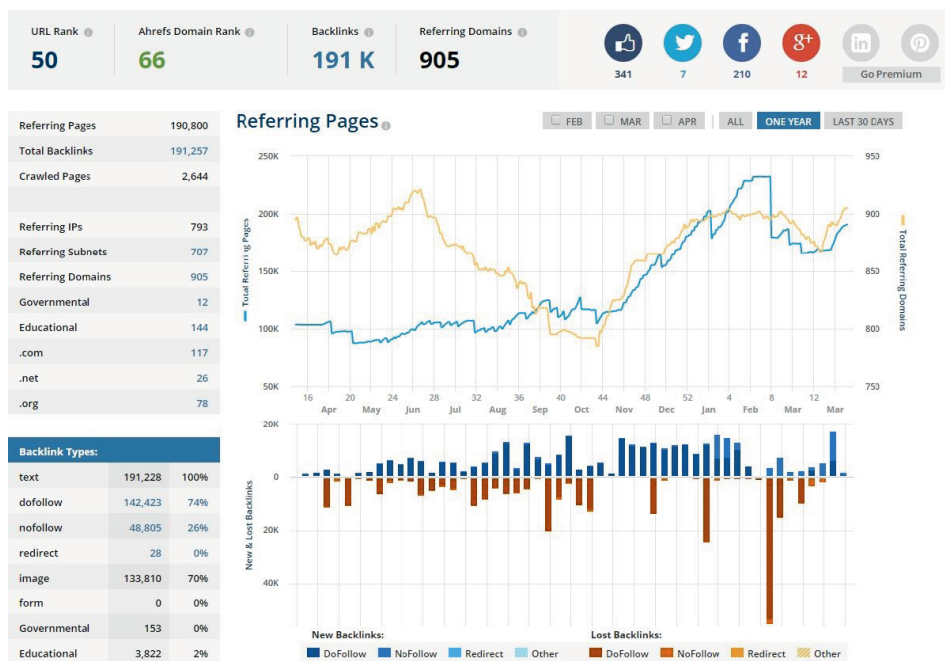
Ukazana wyżej hipoteza potwierdza dane z Ahrefs (tab. 9). Dla zobrazowania problemu zestawiono w niej główne kategorie i parametry linków oraz wskaźnik jakościowy *URL Rating*. Analiza danych nie pozostawia wątpliwości: pomiędzy liczbą linków (w dowolnym przekroju) a metryką reputacji (*URL Rating*) nie istnieje

³⁰ *Google Algorithm Change History*, [online] <https://moz.com/google-algorithm-change> [dostęp 4.04.2016].

³¹ V. Krishnan, R. Raj, *Web Spam Detection with Anti-Trust Rank*, „Proceedings of the Second International Workshop on Adversarial Information Retrieval on the Web – AIRWeb” 2006, vol. 6, s. 37–40, [online] <http://airweb.cse.lehigh.edu/2006/krishnan.pdf> [dostęp 4.04.2016].

żadna współzależność. Niezależnie czy uwzględnimy w obliczeniach ogólną liczbę backlinków (Backlinks Total), czy ich poszczególne kategorie (Dofollow – o niezerowej wartości, Gov – linki z domen rządowych, Edu – linki z witryn naukowych i edukacyjnych) czy liczbę domen, z których one pochodzą (Referring Domain), ich suma nie jest skorelowana ze wskaźnikami jakościowymi.

Warto jednocześnie podkreślić, że uwaga powyższa pozostaje w sprzeczności z klasycznym audytem SEO, gdyż w wielu opracowaniach³² wciąż podkreślana jest przesadnie duża rola liczby linków i sugerowany jest precyzyjny pomiar ich zróżnicowania (por. wykres 8).



Wyk. 8. Linki przychodzące do serwisu Biblioteki Uniwersyteckiej w Warszawie w 2014 roku

Źródło: Ahrefs, <https://ahrefs.com/site-explorer>. Dostęp 20 kwietnia 2015 r.

Mimo powyższych uwag webmasterzy winni monitorować własną witrynę (Google Search Console [GSC], Google Analytics [GA]), gdyż jest to jedyny sposób precyzyjnego rozpoznania, w jaki sposób użytkownicy trafiają na nią i jak korzystają z jej treści. Wykonany doraźnie test przekonuje, że administratorzy serwisów bibliotecznych powszechnie korzystają z GA i posiadają profile na GCS. Wśród 13 testowanych serwisów jedynie dwa (BG AGH i MBP Wr.) nie używały kodów śledzenia GA w HTML i prawdopodobnie również nie używały GSC.

³² B. Danowski, M. Makarczuk, *Pozycjonowanie i optymalizacja stron WWW*, Wydawnictwo Helion, Gliwice 2007 (wyd. 2 – 2009, wyd. 3 – 2011).

Metryki autorytetu, zaufania i reputacji

Wraz z ewolucją poglądów na temat stosowania backlinków przy obliczaniu rangi witryn rola PageRank w algorytmie Google systematycznie malała. Różne analizy udowodniły, że jest on podatny na manipulację³³. Wiele poszlak wskazuje, że już w połowie pierwszej dekady XXI wieku Google zastąpił go innym wskaźnikiem, który wzorował na TrustRank Yahoo!³⁴. Enuncjacje takie pojawiły się w 2007 roku na oficjalnym blogu Google³⁵, a następnie były wielokrotnie komentowane³⁶. Ponieważ zasada jego działania i kalkulacji jest tajemnicą handlową Google, społeczności eksperckie wprowadziły własne metody jego kalkulacji. Powszechnie uznawane w środowisku SEO są wskaźniki udostępniane przez MOZ oraz Ahrefs. Społeczność MOZ operuje czterema wskaźnikami: Domain Authority (DA), Page Authority (PA), MozRank (MR) i MozTrust (MT). Autorytet domeny (DA) i autorytet strony (PA) rosną liniowo i przyjmują wartości 0–100. Z kolei MozRank (*link popularity*) odzwierciedla znaczenie danej strony skalkulowane poprzez ilość i jakość innych stron, które się do nich odwołują. Metryka MozTrust (*trustscore*) działa na podobnej zasadzie, ale mierzy zaufanie linków przychodzących. Dwa ostatnie wskaźniki rosną logarytmicznie i mogą przybrać wartości z zakresu 0–10.

Zestawienie metryk autorytetu polskich serwisów bibliotecznych (tab. 10) pozwala stwierdzić, że zdecydowana większość z nich legitymuje się wysokimi wskaźnikami zliczanymi wg metody MOZ. Jednocześnie można zaobserwować ostry podział na biblioteki akademickie (wraz z BN) i pozostałe. Metryki autorytetu tych pierwszych osiągnęły bowiem bardzo wysokie wartości (DA oscyluje w przedziale 63–73), podczas gdy serwisy pozostałych bibliotek (publicznych) MOZ ocenił niemal dwukrotnie niżej (34–41). Obserwacja ta potwierdza obiegową opinię, że w algorytmie Google autorytet jest blisko skojarzony z pojęciami „władza” i „wiedza”. Do zbieżnych rezultatów prowadzi analiza wskaźników jakościowych serwisu Ahrefs (Domain Rating, URL Rating), które są odpowiednikami DA i PA i przyjmują wartości 0–100. Wartości obu indeksów są silnie skorelowane, co daje asumpt do postawienia hipotezy, że nowe wskaźniki jakościowe Google są mierzalne i dają się zobiektywizować. Z drugiej strony warto zwrócić uwagę, że trudno doszukać się takiej korelacji pomiędzy obiema miarami autorytetu domeny (DA, Domain Rating) a PageRank,

³³ Z. Gyöngyi, P. Berkhin, H. Garcia-Molina, J. Pedersen, *Link spam detection based on mass estimation*, [w:] *Proceedings of the 32nd international conference on Very large data bases*. Seoul 2006, s. 439–450, [online] <http://ilpubs.stanford.edu:8090/697/1/2005-33.pdf> [dostęp 10.04.2016].

³⁴ V. Krishnan, R. Raj, *Web Spam Detection with Anti-Trust Rank*, „Proceedings of the Second International Workshop on Adversarial Information Retrieval on the Web – AIRWeb” 2006, vol. 6, s. 37–40, [online] <http://airweb.cse.lehigh.edu/2006/krishnan.pdf> [dostęp 4.04.2016].

³⁵ M. Cutts, *Trust Rank Explained [Suresh Babu Interviews Matt Cutts in Pubcon 2007]*, YouTube (11 Dec. 2007), [online] <https://youtu.be/p8mUXQzwEvs> [dostęp 03.04.2016].

³⁶ J. Patel, S. Patel, *Efficient calculation of PageRank using TrustRank and Anti-TrustRank*, „International Journal of Engineering Development and Research” 2014, vol. 2, iss. 2, s. 1895–1898, [online] <http://www.ijedr.org/papers/IJEDR1402098.pdf> [dostęp 16.05.2016]; E. Bailyn, *PageRank vs. TrustRank in Google's Algorithm*, [Blog] „FirstPageSage” (29 July 2014), [online] <http://firstpagesage.com/seo-blog/seo-basics/pagerank-vs-trustrank-in-googles-algorithm/> [dostęp 4.04.2016].

Tab. 10. Metryki autorytetu wybranych polskich serwisów bibliotecznych

Miasto	Biblioteka**	Google		Internet Archive		MOZ					Ahrefs	
		Liczba podstron	Page Rank	Domain Age	Dom. Auth.	Page Auth.	Moz Rank	Moz Trust	URL Rating	Dom. Rating		
Warszawa	BN	3590	8	18	66	72	6,56	6,33	80	62		
	BG PW	1750	6	18	68	58	5,09	6,16	39	66		
	BUW	1990	7	18	71	63	5,71	6,23	50	66		
	BPMSW	2800	6	10	39	48	5,39	5,89	34	51		
Kraków	BJ	8320	7	20	73	65	5,49	6,22	50	66		
	BG AGH	1900	6	14	73	54	4,98	5,96	33	65		
	WBP Kr.	7520	6	7	41	49	5,83	6,11	40	52		
	B. Rac.	3970	4	15	34	44	5,30	5,62	40	51		
Poznań	BUAM	2080	7	20	64	60	5,13	6,22	42	66		
	B. Ppoz	6030	5	5	63	46	4,23	5,76	20	63		
	CWINT	868	6	6	67	57	5,14	6,14	18	66		
	BUWr.	13100	7	20	66	63	5,51	6,22	43	65		
Wrocław	MBP Wr.	2940	4	14	39	45	5,41	5,66	38	52		

Obliczenia własne na bazie serwisów: Internet Archive, MOZ Open Site Explorer, Ahrefs, Google [2015.02.30] ** Rozwiązania skrótów – zob. tab. 1.

a tym bardziej pomiędzy jakimkolwiek wskaźnikiem jakościowym a liczbą podstron zaindeksowanych przez Google.

Pośrednio duży wpływ na metryki autorytetu i zaufania ma także wiek domeny. W badanym korpusie witryn czynnik ten nie odegrał jednak większej roli, gdyż wszystkie badane serwisy działały stosunkowo długo na tych samych domenach (średnio 14 lat). Najmłodszy badany serwis (B. P Pozn.) funkcjonował nieprzerwanie od 5 lat, zaś dwa najstarsze (BUWr i BUAM) od 20 lat.

Sygnaly z sieci społecznych

Historia sieci społecznościowych jest stosunkowo krótka, lecz ich wpływ na rozwój internetu okazał się ogromny. Jeszcze pod koniec pierwszej dekady XXI wieku linki udostępniane przez użytkowników Facebooka czy Twittera były w branży SEO traktowane jak odnośniki ze zwykłych stron. Jednak wraz z rozwojem sieci (w 2011 roku liczba aktywnych użytkowników portali społecznościowych na świecie przekroczyła 1 miliard)³⁷ dostrzeżono, że niosą one ze sobą nieco inne znaczenie. Social media w odróżnieniu od mediów oficjalnych są spontanicznym nośnikiem opinii publicznej. Są postrzegane przez użytkowników jako niezależne i w ślad za tym bardziej wiarygodne³⁸. Odmienność tę dostrzegli także specjaliści z Google, uwzględniając od 2010 roku ich specjalny wpływ w algorytmie wyszukiwarki. Od tego czasu zrobiły one w branży SEO zawrotną karierę i od kilku lat lokują się na szczycie rankingów korelacji³⁹. W rankingu Searchmetrics za rok 2014 w grupie 15 najwyższ cenionych czynników sygnały z social mediów stanowiły połowę. Najwyższą lokatę miały: polubienia +1 (Google Plus) [0,33], udostępnienia FB (Facebook Shares) [0,28], łączna liczba sygnałów z FB (Facebook Total) [0,28], komentarze FB (FaceBook Comments) [0,27], piny (Pinterest) [0,27], polubienia FB (FaceBook Likes) [0,25] oraz tweety (Tweets) [0,24]⁴⁰.

Specjalne badania udowodniły, że twórcy polskich serwisów bibliotecznych dostrzegli doniosłą rolę mediów społecznych i wykorzystują je w działalności marketingowej (tab. 11). Największym zainteresowaniem cieszy się Facebook (FB). Wszystkie testowane serwisy prowadzą na fanpage FB, na którym na bieżąco publikują newsy. Jakkolwiek aktywność na poszczególnych profilach jest różna, długofalowa działalność liderów (BUW – 724 sygnałów, WBP w Krakowie – 271, B. Rac. – 162) udowadnia, że poświęcona praca przekłada się na adekwatną liczbę sygnałów (udostępnienia, polubienia, komentarze).

³⁷ *Number of social network users worldwide from 2010 to 2020*, [online] <http://www.statista.com/statistics/278414> [dostęp 16.05.2016].

³⁸ J. van Dijck, *The culture of connectivity: A critical history of social media*, University Press, Oxford 2013; E. Enge, S. Spencer, J. Stricchiola, *SEO, czyli sztuka optymalizacji witryn dla wyszukiwarek*, Wyd. 3. Gliwice 2016.

³⁹ M. Tober, D. Furch, K. Londenberg, L. Massaron, J. Grundmann, *Search Ranking Factors and Rank Correlations 2015*, Searchmetrics, San Mateo 2015, [online] <http://www.searchmetrics.com/knowledge-base/ranking-factors/> [dostęp 4.04.2016]; *Search Engine Ranking Factors 2015. Expert Survey and Correlation Data*, MOZ Blog, [online] <https://moz.com/search-ranking-factors> [dostęp 4.04.2016].

⁴⁰ M. Tober, L. Hennig, D. Furch, *SEO Ranking Factors and Rank Correlations 2014*, Searchmetrics, San Mateo 2014, [online] <http://www.searchmetrics.com/knowledge-base/ranking-factors-2014/> [dostęp 4.04.2016].

Tab. 11. Sygnaty z sieci społecznych do wybranych serwisów bibliotecznych

Biblioteka	Profil/liczba fanów				Sygnaty do strony głównej**							Razem
	FB	G+	Tw.	Facebook			Titter Tweets	Google+ Pluses	Inne sygnaty			
				Shares	Likes	Comm.						
BN	16223	0	0	66	27	25	5	10	1	134		
BG PW	883	0	0	50	24	20	0	5	0	99		
BUW	13063	2	0	210	341	152	7	12	2	724		
BPmsW	1389	0	0	26	18	3	1	4	0	52		
BJ	2987	0	0	82	42	12	4	14	0	154		
BG AGH	3031	12	0	9	0	0	0	1	0	10		
WBP Kr.	8440	138	0	80	18	13	3	157	0	271		
B. Rac.	2689	1	0	78	46	34	2	2	0	162		
BUAM	4067	0	0	58	24	9	2	5	0	99		
B. PPozn	1634	0	182	11	0	0	2	3	0	16		
CWINT	327	26	32	3	4	3	1	21	0	32		
BUWr.	1037	4	4	59	0	3	1	6	0	69		
MBP Wr.	1712	0	0	59	25	10	2	3	0	99		

Źródło: Social Media Shares Checker – <http://www.seocentro.com/tools/search-engines/link-popularity.html> [2015.02.30]. ** Rozwiązania skrótów – zob. tab. 1.

Dodajmy jednocześnie, że aby uzyskać dobre wyniki w rankingach wyszukiwarek, przy pomocy portali społecznościowych, nie wystarczy poprzestać na zgromadzeniu licznej rzeszy sympatyków profilu, potrzebna jest ciągła systematyczna praca. Przekonuje o tym przykład BN, która mimo dużej liczby fanów (16223) wypracowała nieadekwatnie mało sygnałów (134), czyli mniej niż BJ (154), która zrzesza ich przeszło pięciokrotnie mniej (2987). Badania przekonują ponadto, że serwisy inne niż FB nie cieszą się w środowisku bibliotecznym większym zainteresowaniem. Profile na Google + ma wprawdzie 6 jednostek, lecz w wielu przypadkach pełnią one tylko rolę wizytówek, gdyż nie zaobserwowano na ich osiach aktywności. Chlubnym wyjątkiem na tym tle jest tylko WBP w Krakowie, której profil zgromadził 138 fanów i wygenerował adekwatny ruch (157). Jeszcze mniejszym zainteresowaniem cieszy się Twitter, z którego korzystają 3 jednostki, zaś tylko jedna (B. PPozn.) zgromadziła zauważalną społeczność (182 fanów).

Wnioski

Badania nie były wykonane na próbie reprezentatywnej, przeto nie uprawniają do formułowania kategoriycznych wniosków. Praca ma przede wszystkim charakter jakościowy, zaś zaobserwowane prawidłowości są hipotetyczne. Większość testów (poza słowami kluczowymi) przeprowadzono jedynie na małej próbie (13 serwisów), które dobrano w sposób nielosowy: po trzy serwisy z największą liczbą odwiedzin ulokowane w czterech największych ośrodkach akademickich (w Warszawie dodatkowo BN). Jakkolwiek nie było to zamierzone, w próbie znalazły się trzy jednorodne grupy: 4 duże biblioteki uniwersyteckie (BUW, BJ, BUAM, BUWr); 4 politechniczne (BG PW, BG AGH, B. PPozn., CWINT) oraz 4 publiczne (BPmsW, WBP Krak., B. Rac., MBP Wr.). W efekcie większość sformułowanych wniosków odnosi się głównie do serwisów tworzonych przez biblioteki tych właśnie typów.

Na wstępie zbadano ruch generowany przez użytkowników oraz używane frazy i słowa kluczowe. Badania nie zaskoczyły. Prawie połowę ruchu internetowego płynącego do portali bibliotecznych generują użytkownicy 4 największych ośrodków akademickich. Najwięcej badanych zapytań wychodziło z Warszawy (średnio 152 179 miesięcznie), nieco ponad połowę mniej – z Krakowa (70 673) i po około 25% z Poznania i Wrocławia (odpowiednio: 40 051 i 36 515). Analiza chronologiczna ujawniła ponadto, że ruch ten jest wyraźnie skorelowany z rokiem akademickim/szkolnym (nasilenie na początku semestrów i wyraźny spadek w okresie wakacyjnym), co skłania ku tezie, że duża liczba zapytań do portali bibliotecznych ma związek z kształceniem i jest kierowana przez studentów i młodzież szkolną. Przy badaniu słów kluczowych stwierdzono, że frazy najczęściej używane są mocno spolaryzowane; jeśli ułożymy je w porządku malejącym, to krzywa sumy zapytań przyjmuje charakterystykę logarytmiczną. Wprawdzie pełny korpus fraz współwystępujących ze słowem *biblioteka* był dość długi – 801 pozycji, wszelako 14 fraz najczęściej używanych generowało ponad 50% wszystkich zapytań. Szczegółowe badania pozwoliły ponadto na ustalenie, że w przypadku bibliotek uczelnianych najwięcej ruchu ściąga kombinacja dwu słów kluczowych (*biblioteka* i akronim uczelni), choć występują w tym zakresie pewne wyjątki.

Badania poszczególnych aspektów pozycjonowania nie pozwoliły jednoznacznie odpowiedzieć na postawione we wstępie pytanie. Stosowanie zasad SEO przez badane serwisy jest bowiem rozmaicie realizowane w różnych obszarach. Analiza stosowania słów kluczowych w badanej próbie ujawniła, że większość webmasterów prawidłowo (lub w sposób akceptowalny) pozycjonuje „czołówkę” własnego serwisu, choć nie zawsze robi to w sposób optymalny. Stwierdzone uchybienia, poszczególne błędy w strukturze HTML przekonują jednak, że redakcje polskich serwisów bibliotecznych nie przywiązują większej wagi do zasad webwritingu.

Analiza czynników *on-page* (zależnych od twórców) daje asumpt do stwierdzenia, że administratorzy serwisów bibliotecznych znają i w dużym stopniu stosują klasyczne zasady SEO (tytuł, opis strony, długość <body> oraz poprawność linków i obrazów). Dbają tym także o techniczną stronę swoich badanych serwisów, które są na ogół osadzone na serwerach o słabej lub średniej wydajności. Niemniej rzadko stosują nowe standardy (HTML5, OGP itp.) i nie dbają o użytkowników korzystających z urządzeń mobilnych (przeszło 80% serwisów nie jest responsywnych).

Badanie czynników *off-page* (niezależnych od webmasterów) doprowadziło do kilku dalszych konkluzji. Większość serwisów bibliotecznych cieszy się w sieci wysokim autorytetem. Udowodniły to badania metryk autorytetu, zaufania i reputacji (*authority, trust, rank*), które jednocześnie potwierdziły obiegową opinię, że dla Google autorytet jest blisko kojarzony z pojęciami „władza” i „wiedza”. Analizy metryk ujawniły jednocześnie ostry podział na biblioteki akademickie i publiczne. Metryki autorytetu tych pierwszych osiągnęły bowiem bardzo wysokie wartości np. Domain Authority oscyloowało w granicach 63–73, podczas gdy serwisy bibliotek publicznych oceniono niemal dwukrotnie niżej (34–41). Analiza backlinków potwierdziła przypuszczenie, że aktualnie nie istnieje żadna korelacja pomiędzy ich liczbą a metrykami reputacji i autorytetu. Natomiast analiza obecności serwisów bibliotecznych w przestrzeni sieci społecznych wykazała, że webmasterzy znają i doceniają ich marketingową rolę. Absolutnym liderem w tym zakresie jest Facebook, na którym wszystkie witryny prowadzą fanpage i publikują newsy. Wprawdzie zaangażowanie twórców poszczególnych profili jest różne, niemniej długofalowa działalność liderów (BUW, WBP w Krakowie) przekonuje, że poświęcona praca przekłada się na adekwatną liczbę sygnałów.

Bibliografia

- Bailyn E., *PageRank vs. TrustRank in Google's Algorithm*, [Blog] „FirstPageSage” (29 July 2014), [online] <http://firstpagesage.com/seo-blog/seo-basics/pagerank-vs-trustrank-in-googles-algorithm/> [dostęp 4.04.2016].
- Cutts M., *Trust Rank Explained [Suresh Babu Interviews Matt Cutts in Pubcon 2007]*, YouTube (11 Dec. 2007), [online] <https://youtu.be/p8mUXQzwEvs> [dostęp 3.04.2016].
- Danowski B., Makarczuk M., *Pozycjonowanie i optymalizacja stron WWW*, Wydawnictwo Helion, Gliwice 2007 (wyd. 2 – 2009, wyd. 3 – 2011).
- Enge E., Spencer S., Stricchiola J., *SEO, czyli sztuka optymalizacji witryn dla wyszukiwarek*, Wydawnictwo Helion, Gliwice 2016.

- Gabe G., *Penguin 3.0 Analysis – Penguin Tremors, Recoveries, Fresh Hits, and Crossing Algorithms*, [blog] “G-Squared Interactive Marketing Blog” (October 27 2014), [online] <http://www.gsqi.com/marketing-blog/penguin-3-analysis-findings> [dostęp 30.05.2016].
- Gasparotto M., *Search Engine Optimization for the Research Librarian: A Case Study Using the Bibliography of U.S. Latina Lesbian History and Culture*, „Practical Academic Librarianship: The International Journal of the SLA Academic Division” 2014, vol. 4, no. 1, s. 15–34, [online] <https://journals.tdl.org/pal/index.php/pal/article/view/6971> [dostęp 25.05.2016].
- Google Algorithm Change History*, [online] <https://moz.com/google-algorithm-change> [dostęp 4.04.2016].
- Gyöngyi Z., Berkhin P., Garcia-Molina H., Pedersen J., *Link spam detection based on mass estimation*, [w:] *Proceedings of the 32nd international conference on Very large data bases*. Seoul 2006, s. 439–450, [online] <http://ilpubs.stanford.edu:8090/697/1/2005-33.pdf> [dostęp 10.04.2016].
- Iwanek Ł., *10 mitów na temat SEO*, „e-Gospodarka” 7 października 2014, [online] <http://www.egospodarka.pl/116534,10-mitow-na-temat-SEO.html> [dostęp 10.05.2016].
- Jachimczyk A., Chrapek M., *Katalogi artykułów (Presell Pages) – analiza stron internetowych wykorzystywanych w SEO*, „Przegląd Biblioteczny” 2015, nr 2, s. 209–225.
- Jaskowska B., *Efektywny marketing SEO w witrynie bibliotecznej w wyszukiwarkach*, „Przegląd Biblioteczny” 2009, z. 2, s. 229–238.
- Kenning A., O’Brien P., Rossmann B., *Managing search engine optimization: An introduction for library administrators*, „Journal of Library Administration” 2013, vol. 53, no. 2–3, s. 177–188.
- Kluszczak M., *Pozycjonowanie (SEO) w internecie – fakty i mity*, Eactive Blog, [online] <https://www.eactive.pl/pozycjonowanie-stron/pozycjonowanie-w-internecie-fakty-i-mity> [dostęp 10.05.2016].
- Krishnan V., Raj R., *Web Spam Detection with Anti-Trust Rank*, „Proceedings of the Second International Workshop on Adversarial Information Retrieval on the Web – AIRWeb” 2006, vol. 6, s. 37–40, [online] <http://airweb.cse.lehigh.edu/2006/krishnan.pdf> [dostęp 4.04.2016].
- Lee S., Jang W., Lee E., Oh S.G., *Search engine optimization: A case study using the bibliographies of LG Science Land in Korea*, „Library Hi Tech” 2016, vol. 34 iss. 2, s. 197–206, [dostęp 30.05.2016].
- Onaifo D., Rasmussen D., *Increasing libraries’ content findability on the web with search engine optimization*, „Library Hi Tech” 2013, vol. 31 iss. 1, s. 87–108, [dostęp 25.05.2016].
- Patel J., Patel S., *Efficient calculation of PageRank using TrustRank and Anti-TrustRank*, „International Journal of Engineering Development and Research” 2014, vol. 2 iss. 2, s. 1895–1898, [online] <http://www.ijedr.org/papers/IJEDR1402098.pdf> [dostęp 16.05.2016].
- Patil Swati P., Pawar B. V., Patil Ajay S., *Search Engine Optimization: A Study*, „Research Journal of Computer and Information Technology Sciences” 2013, 1(1), s. 10–13, [online] http://www.retawprojects.com/uploads/search_engine.pdf [dostęp 16.05.2016].
- Petrescu P., *Google Organic Click-Through Rates in 2014*, MOZ Blog (October 1st 2014), [online] <http://moz.com/blog/google-organic-click-through-rates-in-2014> [dostęp 30.05.2016].
- Piekarski K., *Ekonomia percepcji. Mechanizmy selekcjonowania informacji w internecie* [praca doktorska], Uniwersytet Śląski w Katowicach, Katowice 2014, [online] <http://www.sbc.org.pl/Content/126980/doktorat3505.pdf> [dostęp 4.04.2016].

- Sadowski T., *Penguin 2.1 – zwany również Pingwinem 5 – co to jest, skąd się wzięło i jak sobie z nim radzić*, GoogleWatchers SEO BLOG (6 października 2013), [online] <http://google-watchers.info/google/penguin-2-1-co-to-jest> [dostęp 30.05.2016].
- Scott D., *White Hat Search Engine Optimization (SEO): Structured Web Data for Libraries*, „Partnership: The Canadian Journal of Library and Information Practice and Research” 2015, vol. 10, no. 1, [online] https://zone.biblio.laurentian.ca/bitstream/10219/2513/1/seo_for_libraries.pdf [dostęp 25.05.2016].
- Search Engine Ranking Factors 2015. Expert Survey and Correlation Data*, MOZ Blog, [online] <https://moz.com/search-ranking-factors> [dostęp 4.04.2016].
- Seymour T., Frantsvog D., Kumar S., *History of search engines*, „International Journal of Management and Information Systems” 2011, vol. 15, no. 4, s. 47–59, [online] <http://www.minot.com/tom/JIS.pdf> [dostęp 19.05.2016].
- Shepard C., *10 SEO Myths that Friggin' Tick Me Off*, MOZ Blog (March 9th, 2015), [online] <https://moz.com/blog/seo-myths> [dostęp 21.05.2016].
- Sullivan D., *RIP Google PageRank score: A retrospective on how it ruined the web*, Search Engine Land [blog] (March 9, 2016), [online] <http://searchengineland.com/rip-google-pagerank-retrospective-244286> [dostęp 30.05.2016].
- Tober M, Hennig L., Furch D., *SEO Ranking Factors and Rank Correlations 2014*, Searchmetrics, San Mateo 2014, [online] <http://www.searchmetrics.com/knowledge-base/ranking-factors-2014/> [dostęp 4.04.2016].
- Tober M., Furch D., Londenberg K., Massaron L., Grundmann J., *Search Ranking Factors and Rank Correlations 2015*, Searchmetrics, San Mateo 2015, [online] <http://www.searchmetrics.com/knowledge-base/ranking-factors/> [dostęp 4.04.2016].
- van Dijck J., *The culture of connectivity: A critical history of social media*, University Press, Oxford 2013.
- Wang F., Li Y., Zhang Y., *An Emphirical study on the Search Engine Optimization Technique and Its Outcomes*, [w] *Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), Proceedings of 2nd International Conference on AIMSEC*. Piscataway 2011, s. 2767–2770, [dostęp 4.04.2016].
- Wyszukiwarki–silniki. Ranking gemius Traffic*, [online] <http://www.ranking.pl/pl/rankings/search-engines.html> [dostęp 4.04.2016].

Library services in the perspective of search engine optimization

Abstract

This article attempts to answer the question whether the websites of large Polish libraries apply principles of SEO (Search Engine Optimization). The study was conducted in four areas, including as follows: 1. study of key words (key word research, key word usage & key word density); 2. analysis of the on-page factors (content & technology); 3. analysis of off-page factors (backlinks & social media); 4. analysis of metrics reputation (brand & confidence metrics). The first area of research is based on the data generated from Google Adwords for the year 2014 (7.411.387 queries, 801 phrases); for subsequent areas material was provided by technical analysis of 13 most visited websites of libraries located in four major academic cities (Warsaw, Krakow, Poznan, Wroclaw). In the article the author examined in detail key words/phrases, by which users search for local academic libraries, and also examined how the selected libraries' websites implement SEO guidelines. The point of reference in the diagnosis of

current factors were correlation reports of companies: Searchmetrics (*SEO Ranking Factors and Rank Correlations 2014, 2015*) and MOZ (*Search Engine Ranking Factor 2015*). Empirical research was based on data as of February 2015.

Key words: libraries, Poland, search engine optimization, SEO

Władysław Marek Kolasa
Pedagogical University of Cracow
Institute of Information Science