

Władysław Marek Kolasa

Polskie czasopisma humanistyczne w perspektywie SEO

Wstęp

W potocznej świadomości pozycjonowanie serwisów internetowych (ang. *Search Engine Optimization*, SEO) kojarzona jest z tajemniczymi zabiegami, wskutek których witryna zawsze pojawia się na czele rankingów wyszukiwarki. Jest to świadomość fałszywa, gdyż wysoką pozycję na stronie wyników (SERP) można uzyskać tylko dla określonej podstrony, w odpowiedzi na określone słowa kluczowe i jest ona zależna od szeregu innych czynników ujętych w algorytmie wyszukiwarki. Obecnie w algorytmie Google'a uwzględnianych jest ok. 200 czynników o różnej wadze, które można zgrupować w cztery kategorie obejmujące: słowa kluczowe, czynniki on-page (niezależne od webmastera); czynniki off-page (zależne od webmastera) oraz metryki reputacji. Warto też podkreślić, że rola poszczególnych składników nie jest do końca znana oraz stale ewoluuje (m.in. ze względu na cyklicznie pojawiające się poprawki algorytmu, np. Panda Update, Penguin Update, Hummingbird Update i in.). Google z oczywistych powodów trzyma w tajemnicy znaczenie poszczególnych czynników, ujawniając jedynie zbiory dobrych praktyk (Webmaster Guidelines), zaś jedyną systematyczną wiedzę na ten temat pochodzi z przeprowadzanych cyklicznie badań zespołów eksperckich (MOZ, Searchmetrics i in.). Mimo wspomnianych ograniczeń SEO jest dziedziną (bardziej praktyczną niż naukową), która dopracowała się wielu teorii oraz grupuje pokaźne zespoły, które świadczą usługi w tym zakresie. Mimo rozmaitych zastrzeżeń praktykę SEO można zoperacjonalizować i przedstawić wiele empirycznych narzędzi do jej weryfikacji.

Warto też na wstępie postawić pytanie: czy stosowanie zasad SEO do właściwego funkcjonowania w sieci jest niezbędne? Czy praktyka SEO dotyczy także serwisów naukowych (w tym czasopism) w podobnej mierze co serwisów handlowych lub newswych? Odpowiedź na to pytanie nie jest jednoznaczna. Nie można bowiem zaprzeczyć tezie, że również bez stosowania zasad SEO każda poprawna (nawet akceptowalna) w sensie technicznym strona będzie wyświetlana w przeglądarce użytkownika, który zna jej URL. Jednakże badania przekonują, że jedynie niewielka część witryn jest w taki sposób wywoływana. Zdecydowanie częściej trafiamy na witryny internetowe z SERP-ów (ang. *Search Engine Results Page*) wyszukiwarki, czyli stron wyświetlających listę adresów znalezionych za pomocą wyszukiwarki, lub ze stron referencyjnych, gdzie została podlinkowana. W efekcie, aby skutecznie uczestniczyć w dyskursie, należy zadbać o widoczność serwisu w wyszukiwarce.

O roli wyszukiwarek w badaniach naukowych przekonuje wiele badań. Zgodnie z ustaleniami uczonych z Interdyscyplinarnego Centrum Modelowania Matematycznego i Komputerowego Uniwersytetu Warszawskiego w roku 2014 aż 80,9% polskich naukowców wskazało Google jako główne narzędzie pozyskiwania informacji o publikacjach (Szprot, 2014). Z drugiej strony, liczne badania zachowań użytkowników przekonują, że jedynie obecność na pierwszej stronie wyników daje realną szansę na odwiedzinę. Na pierwszej stronie wyników kończy bowiem wyszukiwanie aż 71,3% użytkowników (Tabela 1). 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

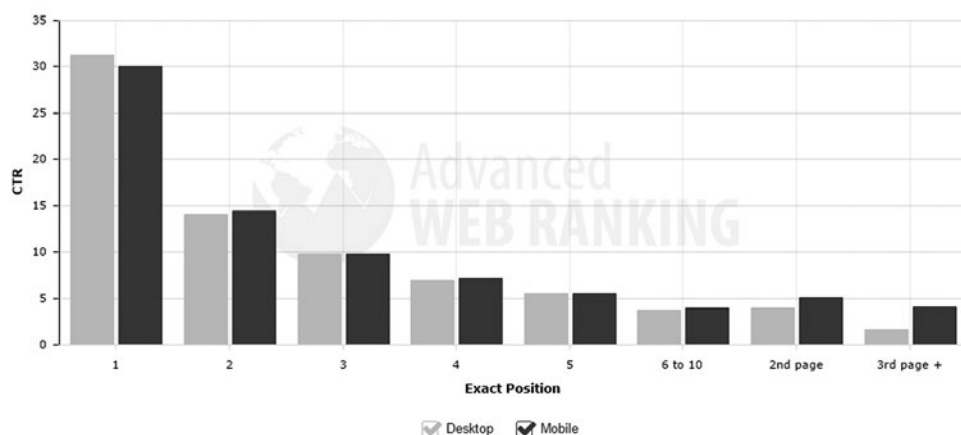
Tabela 1. Współczynnik klikalności (click-through rate) w wyszukiwaniu organicznym Google (VII 2014)

Pozycja	SERP – pierwsza strona						Druga strona	Trzecia strona
	1	2	3	4	5	6 do 10		
CTR	31,24	14,04	9,85	6,97	5,5	3,73	3,99	1,6
Σ%	31,24	45,28	55,13	62,1	67,6	71,33	75,32	76,92

Przeliczenia własne na podstawie: <http://moz.com/blog/google-organic-click-through-rates-in-2014>.

zagląda tylko 3,99% użytkowników, a na kolejne jedynie 1,6%. Warto równocześnie zauważyć, że użytkownicy urządzeń mobilnych zachowują się w podobny sposób do użytkowników komputerów stacjonarnych (wykres 1); oba typy dostępu różni jedynie nieco większy odsetek kliknięć na drugiej i kolejnych stronach SERP-ów płynący ze smartfonów i tabletów.

Wykres 1. Urządzenia a współczynnik klikalności w wyszukiwaniu organicznym Google (VII 2014)



- This chart shows the organic click-through rates for searches coming from 88,535 keywords for 3,185 websites.
- You can compare the CTR for searches coming from desktop devices (web) versus mobile devices.

Źródło: <https://www.advancedwebranking.com/ctrstudy/> (2016.04.03).

Przedmiot badań. Metody. Próba badawcza

Zasadniczym celem niniejszego artykułu jest próba odpowiedzi na pytanie: czy i w jakim stopniu polskie czasopisma humanistyczne stosują zasady SEO (*Search Engine Optimization*), a jeśli tak, to czy wpływają one na wyniki wyszukiwania? W tym celu polskie czasopisma humanistyczne przebadano w czterech przekrojach, obejmujących kolejno: 1. badanie słów kluczowych (*keyword research, keyword usage & keyword density*), 2. analizę czynników on-page (*content & technology*);

3. analizę czynników off-page (*backlinks & social media*) oraz 4. analizę metryk reputacji (*brand & 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 & keyword density*). W każdym przypadku analizie poddano po dwie podstrony: czołówkę serwisu oraz jeden artykuł w widoku abstrakt. Analiza czołówki zmierzała do ustalania poziomu widoczności marki (tj. tytułu czasopisma), z kolei analiza strony artykułu – poziomu widoczności jednostkowych metadanych (autora i tytułu pracy oraz słów kluczowych z abstraktu). Materiał empiryczny do analizy aspektów on-page wygenerowano za pomocą narzędzi serwisu SEORH oraz SEOquake; z kolei czynniki off-page analizowano głównie narzędziami serwisu Ahrefs oraz MOZ Open Site Explorer. Punktem odniesienia w diagnozie czynników były aktualne raporty korelacyjne firmy Searchmetrics *SEO Ranking Factors and Rank Correlations* (Tober, Hennig & Furch, 2014; Tober, Furch, Londenberg, Massaron & Grundmann, 2015) oraz *MOZ Search Engine Ranking Factors* (MOZ, 2015). Badania empiryczne przeprowadzono w okresie od stycznia do marca 2016 roku.

Wyczerpujące przebadanie całego korpusu polskich czasopism humanistycznych byłoby bardzo pracochłonne, przeto badania przeprowadzono na dwu próbach: część aspektów przebadano z użyciem tzw. dużej próby 60 tytułów, wybranych nielosowo spośród 807 pism z obszaru nauk humanistycznych zarejestrowanych na częściach C i B Wykazu czasopism punktowanych z 2015 roku; część na tzw. małej próbie ($n = 16$ i $n = 8$), którą tworzyły wysoko punktowane kwartalniki wybrane z pierwszej próby.

Duża próba była konieczna, aby zachować swoistą łączność tematyczną z inną pokrewną pracą (Siuda, 2016a), w której opisano strategie promocyjne polskich czasopism humanistycznych, w szczególności zdiagnozowano instytucjonalne uwarunkowania ich promocji (Siuda, Stachura, 2016) oraz poddano analizie ich obecność w Internecie (Kozłowski, Bomba, 2016). Autorzy wskazanej pracy skonstruowali na potrzeby badań próbę (nielosową i niereprezentatywną) opartą na dwóch czynnikach: (1) liczbie punktów przypisanych pismu w Wykazie czasopism punktowanych MNiSW z grudnia 2013 roku oraz (2) posiadaniu czynnej strony internetowej.

Dobór był zatem oparty na założeniu, że w próbie muszą się znaleźć pisma o każdej liczbie punktów, choć nie stosowano zasady proporcjonalności (Siuda, 2016b). W niniejszych badaniach także przyjęto takie założenia, zaś jedyną różnicą była synchronizacja bieżącej listy z nowym Wykazem czasopism punktowanych MNiSW (z grudnia 2015 roku). Na finalną listę złożyły się zatem: 23 roczniki, 14 półroczników, 22 kwartalniki i jeden dwumiesięcznik; których rozkład tematyczny był następujący: literaturoznawstwo, językoznawstwo, kulturoznawstwo (20 tytułów); filozofia, religioznawstwo, teologia (13), historia, historia sztuki, archeologia (9), nauki humanistyczne i społeczne – zakres ogólny (15), inne pisma humanistyczne. Dużą próbę ($n = 60$) ze względu na pracochłonność wykorzystano tylko w kilku badaniach cząstkowych (np. architektura witryn i czynniki on-page).

Przeważającą część testów wykonano na małej próbie ($n = 16$). Próba taka była niezbędna do wykonania badań porównawczych, te zaś wymagały skonstruowania jednorodnego korpusu pism. Uznano, że ten warunek spełniają kwartalniki, które wybrano z dużej próby (21 tytułów). Ponieważ zbiór okazał się zbyt duży, aby uniknąć pracochłonnych badań, ograniczono ich listę wyłącznie do tytułów o ugruntowanej pozycji, cieszących się w środowisku prestiżem, co zoperacjonalizowano jako warunek posiadania w rankingu MNiSW za rok 2015 co najmniej 10 punktów (16 tytułów)¹. Również i ta lista uległa później redukcji, gdyż podczas badania ruchu w wyszukiwarkach (Google Adwords) tylko połowa z nich (8 witryn)² generowała zauważalny ruch i istniały dla nich dające się przetwarzać dane empiryczne. Z tej listy wyeliminowano jeszcze jeden tytuł („Biuletyn Historii Sztuki”), gdyż w finalnej grupie było inne, podobne pismo tego samego wydawcy – „Pamiętnik Teatralny”.

.....
¹ „Avant”, „Biuletyn Historii Sztuki”, „Diametros – An Online Journal of Philosophy”, „Ethos. Kwartalnik Instytutu Jana Pawła II KUL”, „Filozofia Nauki” (Philosophy of Science), „Kultura Popularna”, „Miscellanea Anthropologica et Sociologica”, „Pamiętnik Teatralny”, „Poznań Studies in Contemporary Linguistics”, „Praktyka Teoretyczna”, „Przegląd Religioznawczy”, „Studia Humanistyczne AGH”, „Studia Semiotyczne”, „Studia Socjologiczne”, „Studies in Polish Linguistics”, „Zeszyty Naukowe Uniwersytetu Jagiellońskiego Prace Etnograficzne”.

² „Biuletyn Historii Sztuki”, „Ethos”, „Filozofia Nauki”, „Pamiętnik Teatralny”, „Poznań Studies in Contemporary Linguistics”, „Przegląd Religioznawczy”, „Studia Humanistyczne AGH”, „Studies in Polish Linguistics”.

W badaniach zastosowano jeszcze kilka ograniczeń. Wszelkie dane na temat pozycjonowania odnoszono tylko do jednej wyszukiwarki – Google. Powód był prosty. Na polskim rynku wyszukiwarek jest ona od wielu lat absolutnym liderem. Według rankingu GemiusTraffic już od początku 2007 roku co najmniej 90% wszystkich wyszukiwań z Polski pochodziło z Google, zaś w ciągu ostatnich pięciu lat udział Google stabilnie oscylował w przedziale 95–98% (Gemius 2016). Drugie ograniczenie zastosowano przy badaniu dostępności pełnych tekstów – w analizach i testach uwzględniano bowiem wyłącznie teksty w formacie PDF (choć nie był to format jedyny). Również w tym przypadku argumentem była powszechność. Na przykład według badaczy z ICM aż 96,9% polskich czasopism udostępnianych w sposób otwarty serwuje zasoby w formacie PDF; uznano więc, że ich pominięcie nie zdeformuje obrazu badań.

Architektura witryn

Analiza zmierzająca do oceny witryn w kontekście *Search Engine Optimization* wymaga na wstępie udzielenia odpowiedzi na pytanie: czy badane serwisy mają prawidłową architekturę i czy ich zasób jest poprawnie udostępniany; w szczególności czy jest on prawidłowo rozdzielony na podstrony (poszczególne artykuły) i zaopatrzony w odpowiednie metadane czytelne dla maszyn (Nahotko, 2010; Sapa, 2009). Szczególne znaczenie ma przy tym przedostatnia właściwość, gdyż jedynie poprawnie skategoryzowany zasób daje gwarancję, że zostanie właściwie zidentyfikowany i zindeksowany przez roboty wyszukiwarek (*crawlers*). Praktyczną realizację tego modelu, który funkcjonuje jako niepisany standard, można obserwować u czołowych światowych edytorów, m.in. Springera czy Elsewiera, a także u większości wydawców stosujących profesjonalne oprogramowanie (np. Open Journal System) (Kulczycki, 2013c). Nieco upraszczając, można założyć, że warunki te spełnią te czasopisma, które potrafią generować zasób zgrupowany w co najmniej trzech następujących kategoriach (reprezentowanych jako podstrony):

- A. Pojedynczy zeszyt (tom, wolumin);
- B. Spis treści pojedynczego zeszytu;
- C. Tytuł/autor/abstrakt artykułu i link do jego treści.

Ocena tak zdefiniowanej architektury polskich czasopism humanistycznych (Tabela 2) nie prowadzi do zbyt optymistycznych wniosków. Wprawdzie informację o pojedynczym zeszycie (A) zaopatrzoną zwykle w spis treści (B) serwują prawie wszystkie redakcje (odpowiednio: 96% i 96%), wszelako szczegółową informację o artykułach (C) udostępnia ich znacznie mniejszy odsetek (47–80%). Ostatni rodzaj informacji redakcje często realizują tylko częściowo (C1, C2, C3). Najczęściej serwowana jest podstrona zawierająca tytuł i nazwisko autora publikacji (80%), w wielu przypadkach wzbogacona w link prowadzący do PDF-a (73%), lecz jedynie w 47% zawiera abstrakt i słowa kluczowe (C2). Warto jednocześnie dodać, że stosunkowo wysoki wskaźnik linków do pełnej treści (C3) nie oznacza, że pełne teksty są prawidłowo udostępniane. Wydawcy bowiem stosunkowo często tworzą linki od różnych metadanych (numer, spis treści, tytuł artykułu) do nierozdzielonego na artykuły PDF-a, co jest wprawdzie akceptowalne przy użytkowaniu manualnym, lecz zupełnie niedopuszczalne przy komunikacji maszynowej.

Tabela 2. Kategorie podstron polskich czasopism humanistycznych

Próba/Kategoria	A. Zeszyt / tom/numer		B. Spis treści C1. Tytuł/aut.		C. Tytuł/autor/abstrakt/PDF					
					C2. Abstrakt		C3. PDF			
	N	%	N	%	N	%	N	%	N	%
Duża próba $n=60$	58	96	58	96	48	80	28	47	44	73
Mała próba $n=16$	16	100	16	100	14	87	11	69	12	75

Obliczenia własne.

Trudno jednoznacznie wskazać, jaki odsetek polskich czasopism humanistycznych ma w pełni poprawną architekturę, gdyż serwowanie abstraktu na osobnej podstronie (47%) nie jest równoznaczne z wymogiem jego prawidłowego osadzenia w strukturze HTML/XML. Gwarancję tę daje natomiast zastosowanie profesjonalnego systemu informatycznego. Wykonana *ad hoc* analiza przekonuje, że ok. 33% redakcji stosuje do obsługi archiwum profesjonalny system, z czego 15% instalacji obsługuje Open Journal

System (OJS)³, zaś ok. 18% wykorzystuje inne oprogramowanie. W analizowanej próbie (60 tytułów) zidentyfikowano 5 takich systemów (11 instalacji): DSpace (2 tytuły) oraz 4 dedykowane, które stosowali profesjonalni wydawcy lub przez repozytoria: Wydawnictwo Uniwersytetu Jagiellońskiego (4), Index Copernicus (2), CEEOL (2), DeGruyter (1).

Osobną kwestią blisko związaną z architekturą jest formuła dostępu do pełnych tekstów artykułów. Analiza dużej próby ($n = 60$) wykazała, że wśród 73% redakcji udostępniających teksty 13% oferuje się w modelu płatnym, zaś 60% – bez ograniczeń (Open Access). Dane te okazały się zbieżne z wynikami badaczy z ICM, którzy badając duży korpus (1304 tytuły), obliczyli, że odsetek pism dających dostęp do pełnych tekstów, wliczając w to tytuły płatne i stosujące czasowe embargo, wynosi 67,8% (Szprot, 2014). Zbieżne rezultaty przynoszą także analizy małej próby (Tabela 2). Podobnie wysoki jest w tym przypadku odsetek pism udostępniających osobną podstronę zawierającą informację o pojedynczym zeszycie (A), spisie treści (B) oraz pełny tekst (C3), odpowiednio: 100%, 100% i 75%. Znacząco wyższy jest jednakże odsetek pism serwujących abstrakty (69%), co jest w dużej mierze związane z faktem, że pisma tej grupy częściej stosują specjalistyczne oprogramowanie do obsługi archiwum (56%), z czego 18% przypada na OJS i aż 38% na systemy dedykowane.

Nie roszcząc sobie prawa do pełnej precyzji, można stwierdzić, że jedynie 33–47% polskich czasopism humanistycznych prawidłowo udostępnia swój zasób i ma realną szansę, aby być prawidłowo zindeksowanych w wyszukiwarkach.

Słowa kluczowe (*keyword research & keyword usage*)

Jednym z głównych obszarów analiz SEO jest badanie słów kluczowych. Wynikiem tych analiz jest odpowiedź na dwa pytania: (1) w jaki sposób użytkownicy wyszukiwarek trafiają do naszych zasobów (jakich **słów kluczowych używają i ile zapytań generują**)?

.....

³ Warto dodać, że obliczenia niniejsze znacznie korygują dotychczasowe ustalenia. Według badaczy z ICM (Szprot, 2014) w 2013 roku Open Journal System wykorzystywało 41 redakcji (czyli 2,7% korpusu czasopism naukowych); nowsze źródła wskazują natomiast: 6,8% (Rozkosz, 2014a) lub 9% (Drabek, 2015).

oraz (2) w jaki sposób nasze podstrony są oznaczone tymi słowami kluczowymi? Strategia poszukiwania odpowiedzi na każde z tych pytań jest inna; wymaga też użycia innych metod i narzędzi.

Badanie słów kluczowych (*keyword research*)

W pierwszym przypadku wszelkich danych 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, ponieważ dzięki Google Adwords ma dostęp do kompletnego materiału badawczego (za okres ostatnich dwóch lat), a zatem nie musi ograniczać się do próby badawczej (Geddes, 2014; Kaushik, 2010).

Badanie potencjalnego ruchu generowanego przez użytkowników czasopism naukowych już na wstępie napotyka poważne trudności. Ponieważ zasoby udostępniane przez czasopisma naukowe (artykuły) mają jednostkowy, unikalny charakter, rozproszenie kwerend musi być duże. Ponadto zapytania wysyłane przez użytkowników do przeglądarki (zwykle frazy z tytułu artykułu i nazwisko autora) są długie (wielowyrzowe). Analiza takich obiektów, jakkolwiek technicznie niełatwa, znana jest w praktyce badawczej i dość dobrze opisana. Podwaliny teoretyczne dla tego problemu (znanego w literaturze jako „long tail”) stworzył Chris Anderson podczas prac nad analizą zasobów serwisów Amazon.com i Netflix. Wynikiem odkrycia był najpierw sygnałny tekst opublikowany w „Wired Magazine” (Anderson, 2004), a następnie szereg *case study* i pogłębionych monografii (np. Anderson, 2006, tłum. 2008). W myśl koncepcji Andersona posiadanie unikalnego, ale szerokiego asortymentu może w odpowiednio długim czasie prowadzić do uzyskania sumarycznie większych obrotów niż te osiągnięte na najpopularniejszych, masowo sprzedawanych towarach. Zasada „long tail” jest przeto sprzeczna z jedną z głównych zasad klasycznej ekonomii (zasada Pareto), w myśl której firmy uzyskują ok. 80% przychodów dzięki sprzedaży jedynie 20% swojego asortymentu.

Powyższa hipoteza nie oznacza bynajmniej, że obiekty podlegające prawidłom „long tail” nie mają cech, które można badać empirycznie. Jakkolwiek zapytania dotyczące głównego asortymentu (artykułów) są bardzo rozproszone, wszelako skupienie wyników następuje

na marce obiektu – czyli w naszym przypadku: tytule czasopisma. Analiza danych historycznych Google (wszystkie zapytania za okres ostatnich dwóch lat [marzec 2014 – luty 2016]) jednoznacznie przekonuje, że zainteresowanie użytkowników jest ledwie symboliczne; przy czym spośród 16 badanych tytułów jedynie połowa (8) generuje zauważalny ruch – średnio 830 zapytań miesięcznie (Tabela 3).

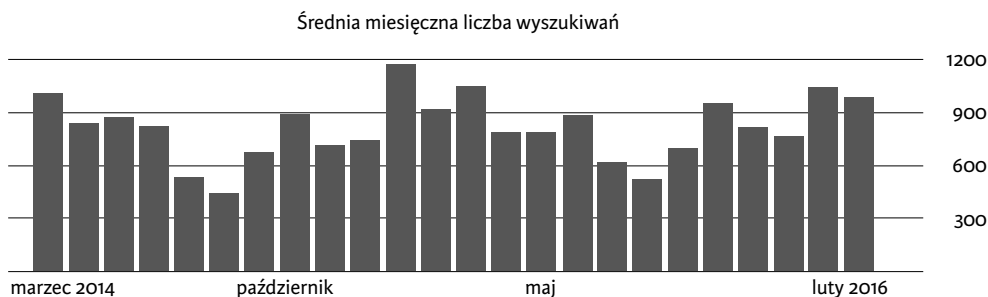
Tabela 3. Liczba wyszukiwań wybranych polskich czasopism humanistycznych w Google

Keyword	Średnia liczba wyszukiwań miesięcznie
Filozofia Nauki Kwartalnik	260
Biuletyn Historii Sztuki	170
Przegląd Religioznawczy	110
Studia Humanistyczne AGH	90
Poznań Studies in Contemporary Linguistics	90
Pamiętnik Teatralny	70
Studies in Polish Linguistics	30
Ethos (Kwartalnik)	10
Razem	830

Obliczenia własne. Źródło danych: Google Adwords (pomiar 6.04.2016).

Ruch generowany przez użytkowników i skierowany na poszczególne „marki” jest wprawdzie niewielki, lecz długookresowa analiza (dwa lata) przekonuje, że jest on względnie stały i podlega niewielkim wahaniom, które następują zgodnie z rokiem akademickim (nasilenie na początku semestrów i wyraźny spadek w okresie wakacyjnym) (Wykres 2). Badania pozwoliły dodatkowo uzyskać kilka charakterystyk przekrojowych użytkowników polskich czasopism humanistycznych. Statystyczny czytelnik korzysta najczęściej z komputera stacjonarnego (90,8%), rzadziej wykorzystuje do tego celu smartfon (6,7%), a najrzadziej tablet (2,4%). Rozkład geograficzny zapytań świadczy także o lokalności polskich periodyków humanistycznych, gdyż dominujący ruch pochodzi z Polski (79,8%), nieco zapytań płynie z państw ościennych: Niemcy 2,2%, Ukraina 2,0%, zaś 16,0% przypada na pozostałe kraje.

Wykres 2. Dynamika wyszukiwań wybranych polskich czasopism humanistycznych w Google (2014–2016)



Źródło: Google Adwords. Uwzględnione tytuły – zob. Tabela 3. Okres analizy (marzec 2014 – luty 2016).

Ruch kierowany do czasopism naukowych (czołówki witryn głównych), jakkolwiek mierzalny, jest bardzo mały, co potwierdza hipotezę, że większość zapytań jest nakierowanych na poszczególne obiekty (artykuły). Wiele poszlak potwierdza tę supozycję, sytuując czasopisma naukowe jako klasyczny przykład modeli biznesowych opartych na „long tail”.

Stosowanie słów kluczowych (*on-page keyword usage & keyword density*)

Kolejnym krokiem w analizie SEO jest zbadanie, na jakie słowa kluczowe pozycjonują się same redakcje. 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 do algorytmu Google poprawki Panda Update, strony przeładowane nieadekwatnymi słowami kluczowymi są dość skutecznie identyfikowane i zaniżane w rankingu.

Strony zgodne z SEO tworzy się zwykle w dwu krokach. Najpierw następuje redakcja merytoryczna (redaktor przygotowuje materiał zgodny z tradycyjnymi regułami pisarskimi), następnie zaś do pracy wkracza webwriter, który dodaje w odpowiednich miejscach i w odpowiednim kontekście słowa kluczowe. Tradycyjnym miernikiem poprawnego pozycjonowania jest tzw. gęstość słów kluczowych

(*keyword density* [dalej KD]), którego optymalna wartość winna oscylować w przedziale 1–3% [według niektórych 1–4,5%] (Patil Swati, Pawar & Patil Ajay, 2013; Wang & Zhang, 2011). Duże znaczenie ma także umiejscowienie słów i fraz kluczowych w kodzie HTML. Największe znaczenie przypisuje się wystąpieniom słów w tytule strony (*tag <title>*) oraz nagłówkach strukturalnych (tagach H1, H2, H3), mniejsze natomiast, jeśli wystąpią w sekcji *<body>*. Słowa i frazy użyte w części *<body>* wzmacnia się ponadto tagami strukturalnymi: pogrubienie **, kursywa **, lista nieuporządkowana ** i lista numerowana ** oraz słowa użyte w kotwicach linków (*anchors*).

Ten aspekt analizy z uwagi na pracochłonność przeprowadzono na małej próbie ($n = 7$)⁴. Z uwagi na rozproszenie słów kluczowych w badanym korpusie zbadano, jak na stronach głównych czasopism pozycjonowana jest marka (frazą tytułowa [tytuł czasopisma] lub jej znacząca część) oraz jakie słowo kluczowe jest najwyżej w rankingu (TOP1).

Wyniki analizy zestawiono w tabeli 4, wskazując kolejno dla każdej badanej frazy lub słowa kluczowego: miejsca, gdzie wystąpiło (Found), liczbę powtórzeń na stronie (Repeat) oraz jej gęstość (KD). Zestawienie zdaje się przekonywać, że większość redakcji prawidłowo pozycjonuje „czołówkę” własnego serwisu, choć rzadko robi to w sposób optymalny. Jakkolwiek w dwu przypadkach („Filozofia Nauki” – 4,73% i „Studies in Polish Linguistics” – 4,60%) wartość KD przekroczyła zalecane optimum o 50% oraz w jednym („Poznań Studies in Contemporary Linguistics” – 0,84%) było ono nieznacznie zaniżone, wszelako w pozostałych przypadkach gęstość przyjmowała poprawne wartości i była powtarzana odpowiednią liczbą razy (2–10). Nie oznacza to bynajmniej, że badane strony były w pełni zoptymalizowane. Jedyne redaktorzy czterech witryn umieścili frazę „marki” w tytule stron (T), a nieliczni w innych zalecanych miejscach: dwie witryny w metatagach „description” (D) oraz dwie witryny w nagłówkach najwyższego poziomu (H1). Do kilku wniosków prowadzi też analiza głównego słowa kluczowego na stronie (TOP1). Nieco ponad połowa witryn dobrała je optymalnie, było bowiem częścią składową

.....
⁴ Ponieważ „Biuletyn Historii Sztuki” i „Pamiętnik Teatralny” mają tego samego wydawcę i prezentują podobny sposób realizacji, w dalszych analizach uwzględniono tylko ten ostatni („Pamiętnik Teatralny”).

Tabela 4. Użycie słów kluczowych na „czołówkach” wybranych polskich czasopism humanistycznych

Tytuł	Word		„Marka”		Keyword TOP1			
	N	Found	Repeat	KD %	Keyword	Found	Repeat	KD %
Ethos* Kwartalnik	130	K	2	1,54	pawła	TDH1B	4	3,08
Filozofia* Nauki*	169	TDKH1B	4	4,73	nauki	TDKH1B	5	2,96
Pamiętnik* Teatralny*	1717	B	11	1,28	prof.	B	17	0,99
Poznań* Studies* in Contemporary* Linguistics*	132	TB	2	0,84	current	TH1B	6	2,27
Przegląd Religioznawczy*	235	B	10	2,35	religioznawczy	B	10	2,35
Studia* Humanistyczne* AGH*	389	TB	5	3,86	studia	TB	5	1,29
Studies* in Polish* Linguistics*	522	TDH1B	8	4,60	polish	TDH1B	10	1,92

Narzędzie: SEOQake i SEORCH (pomiar 9.04.2016). Uwzględniono słowa kluczowe w różnych przypadkach gramatycznych. Gwiazdką (*) oznaczono te słowa kluczowe z fraz wielowyrazowych, które były pozycjonowane na testowanych stronach. Oznaczenie w kolumnach Found: K – słowo użyte w sekcji *meta keywords*; T – słowo użyte w sekcji *meta title*; D – słowo użyte w sekcji *meta description*; H1 – słowo użyte w nagłówku H1; B – słowo użyte w sekcji *body*.

frazy „marka” (słowa: „nauki”, „religioznawczy”, „studia”, „polish”). W trzech przypadkach rangę głównego słowa kluczowego zyskały frazy niezwiązane z marką („pawła”, „prof.”, „current”). Można przypuszczać, że stwierdzone niedoskonałości (w znacznej części) są skutkiem niesamodzielności badanych witryn, wskutek czego część metadanych wymusza serwis główny lub użyty CMS. W badanej próbie tylko dwie („Ethos” i „Filozofia Nauki”) były w pełni niezależne, pozostałe zaś funkcjonowały jako podserwis w domenie organizatora lub wydawcy.

Zbyt mała próba badawcza nie uprawnia do szerszych generalizacji, lecz kilka poszlak pozwala stwierdzić, że redakcje polskich periodyków humanistycznych raczej nie przywiązują większej wagi do zasad webwritingu. Można też zaryzykować tezę, że skoro w badanym korpusie prawie połowa witryn (skądinąd cenionych

kwartalników) zawiera jakieś uchybienia i niedoskonałości, zapewne ich odsetek w ogólnym korpusie pism polskich jest większy.

Wiele interesujących wniosków płynie z analizy KD dla podstron zawierających poszczególne artykuły, czyli tzw. widoku abstrakt. Ponieważ z informacyjnego punktu widzenia są to najważniejsze komponenty witryny czasopisma naukowego, winny być podstawowym i najważniejszym elementem oceny każdego archiwum online. Zgodnie z wcześniejszym ustaleniem podstrony poszczególnych artykułów ze swej istoty mają jednostkowy, unikalny charakter i **są klasycznym** obiektem lokującym w tzw. długim ogonie fraz wyszukiwawczych (*long tail*). Właściwość ta wymusza użycie adekwatnych fraz kluczowych do ich pozycjonowania (zwykle również długich i unikalnych). Identyfikacja ich nie jest jednak trudna, gdyż frazą jednoznacznie wskazującą na artykuł jest jego tytuł właściwy (ewentualnie wraz z podtytułem) oraz nazwisko autora.

W celu sprawdzenia poprawności użycia słów kluczowych w widoku abstrakt zbadano użycie słów kluczowych w pierwszym artykule wydanym w pierwszym zeszycie 2015 roku⁵ (Tabela 5). Z analizy wyłania się dość klarowny obraz stanu pozycjonowania podstron w widoku abstrakt. Wszystkie witryny wykorzystujące profesjonalne oprogramowanie archiwalne („Filozofia Nauki” – CEEOL; „PSLC” – DeGruyter; „Studia Humanistyczne AGH” – OJS i „Studies in Polish Linguistics” – system Wydawnictwa UJ) względnie prawidłowo pozycjonowały tytuły udostępnianych artykułów. Potwierdza to w szczególności liczba powtórzeń (*repeat*), miejsce wystąpienia (*found*) oraz wartość KD. Gęstość wprowadzie w dwu przypadkach znacznie przekroczyła optimum (1–3%), lecz praktyka SEO potwierdza, że przy długich frazach jest to zabieg dopuszczalny. Nieco gorzej wypadł test sprawdzający pozycjonowanie nazwiska autora – gdyż tylko połowa archiwów obsługiwanych przez systemy eksponowała go prawidłowo (OJS i system Wydawnictwa UJ). Znacznie gorzej wypadły testy archiwów nieużytkujących dedykowanych systemów; witryny te albo nie miały podstron w widoku archiwum („Przegląd Religioznawczy”), albo były one ukryte za linkiem chronionym

.....
⁵ W dwu przypadkach („Pamiętnik Teatralny”, „Przegląd Religioznawczy”) wystąpiły problemy z dostępem do numeru 1/2015, wobec czego wykorzystano artykuł z ostatniego zeszytu wydanego w 2014 roku.

hasłem („Pamiętnik Teatralny”). Wskutek tego artykuły obu pism są zupełnie niewidzialne dla wyszukiwarek. Brak systemu nie zawsze jednak oznacza brak widoczności dla robotów wyszukiwarek, o czym świadczy przypadek kwartalnika „Ethos”, który w miarę prawidłowo udostępnia zasób w widoku abstrakt, mimo iż do obsługi strony wykorzystuje uniwersalny CMS.

Tabela 5. Użycie słów kluczowych w wybranych publikacjach w widoku abstrakt

Czasopismo – publikacja	Word N	Nazwisko autora			Tytuł		
		Found	Repeat	KD %	Found	Repeat	KD %
[A] Ethos – Jerzy Szymik: Succisa virescit. Do czego Europie potrzebna dziś teologia (1)	341	B	2	0,59	B	1	2,31
[B] Filozofia Nauki – Igor Hanzel: Idealizations, Ceteris Paribus Clauses, Idealizational Laws (2)	266	H3	1	0,37	H2	1	2,23
[C] Pamiętnik Teatralny – Dobrochna Ratajczakowa: Comédie polonaise, czyli O komediach przepolszczonych (4/2014) (1)	0	0	0	0	0	0	0
[D] PSLC – Krisda Chaemsaitong: Positioning self and others in the courtroom. Person markers in the opening statement (3)	1033	H2	3	0,29	TDH1	8	10,00
[E] Przegląd Religioznawczy – Marcin Z. Paszke – Uwagi na temat iluminacji tzw. boskich łodzi w starożytnej Mezopotamii (4/2014) (1)	0	0	0	0	0	0	0
[F] Studia Humanistyczne AGH – Dorota Rancew-Sikora: Opowiadanie w społecznym układzie stołu. Analiza konwersacyjna spotkań rodzinnych (4)	248	TDB	5	2,01	TDB	4	8,06
[G] Studies in Polish Linguistics – Paweł Rutkowski, Anna Kuder, Małgorzata Czajkowska-Kisil, Joanna Łacheta: The Structure of Nominal Constructions in Polish Sign Language (PJM): A Corpus-based Study (5)	357	H3B	3	3,35	H2	1	3,64

Systemy do obsługi archiwum: (1) – brak systemu; (2) – CEEOL; (3) – DeGruyter; (4) – OJS; (5) – Wydawnictwo UJ.

Efekty pozycjonowania widoku abstrakt przetestowano za pomocą wyszukiwarki Google i jej podserwisu Google Scholar, wyszukując tytułowe frazy (opcja: wyszukiwanie dokładne). Wyniki testów wskazują na wysoką korelację pomiędzy widzialnością w Google a stosowaniem dedykowanego systemu (Tabela 6). Każdy rekord serwowany przez profesjonalny CMS (opisy: [B, D, E, G]) pojawił się bowiem co najmniej kilkakrotnie w wynikach Google, był poprawnie identyfikowany przez Google Scholar i wskazywał na PDF-y (z wyjątkiem przypadków, gdy PDF nie był publicznie dostępny [B]). Testy rekordów serwowanych z witryn bez specjalistycznego archiwum dały rezultaty zupełnie odwrotne, gdyż pojawiały się co najwyżej na stronach Google, nie były natomiast indeksowane w Google Scholar. Warto zauważyć, że jakkolwiek do Google można powiązać dowolne archiwa i strony (Kulczycki, 2013b), dobre rezultaty przynosi stosowanie systemów dedykowanych⁶. Test wykazał ponadto niską przydatność operatora (*filetype*), gdyż w kilku przypadkach (D, G) PDF-y poprawnie powiązane z Google Scholar nie były identyfikowane z pozycji uniwersalnej wyszukiwarki Google.

W tym kontekście warto przytoczyć ciekawe spostrzeżenie naukowców, które poczyniono na marginesie głównego nurtu badań. Jakkolwiek znaczna część polskich czasopism humanistycznych w miarę poprawnie funkcjonuje w Google Scholar, jednak korpus czasopism referencyjnych (cytujących opublikowane w nich artykuły) jest bardzo mały albo słabo zindeksowany. Świadczą o tym m.in. sumaryczny pomiar ich cytowalności (Tabela 7), który wydaje się bardzo zaniżony (maksymalna wartość pięcioletniego indeksu Hirscha dla najlepiej cytowanego tytułu wynosi zaledwie 6). Trudno wszelako oczekiwać, że w żadnym z testowanych czasopism nie ukazał się w ciągu ostatnich pięciu lat żaden wybitny (tj. wysoko cytowany) artykuł. Na poparcie tej hipotezy warto przytoczyć jeszcze przykłady dwu prestiżowych polskich czasopism, których wskaźniki wynoszą odpowiednio: „Pamiętnik Literacki” (2, 2) i „Kwartalnik Historyczny” (1, 1).

.....

⁶ Scholar Inclusions – <https://partnerdash.google.com/partnerdash/d/scholarinclusions> (1.04.2016).

Tabela 6. Wynik wyszukiwań podstron w widoku abstrakt (Google/Google Scholar)

Tytuł	Google		Google Scholar			
	Czasopismo – publikacja	Liczba trafień	Filetype: pdf	SERP	Wersje	PDF
[A] Ethos – Jerzy Szymik: Succisa virescit. Do czego Europie potrzebna dziś teologia (1)		6	1	0	0	0
[B] Filozofia Nauki – Igor Hanzel: Idealizations, Ceteris Paribus Clauses, Idealizational Laws (2)		69	0	1	5	0
[C] Pamiętnik Teatralny – Dobrochna Ratajczakowa: Comédie polonaise, czyli O komediach przepolszczonych (4/2014) (1)		1	0	0	0	0
[D] PSLC – Krisda Chaemsaitong: Positioning self and others in the courtroom: Person markers in the opening statement (3)		5	0	1	6	3
[E] Przegląd Religioznawczy – Marcin Z. Paszke – Uwagi na temat iluminacji tzw. boskich łodzi w starożytnej Mezopotamii (4/2014) (1)		1	0	0	0	0
[F] Studia Humanistyczne AGH – Dorota Rancew-Sikora: Opowiadanie w społecznym układzie stołu. Analiza konwersacyjna spotkań rodzinnych (4)		6	2	1	1	1
[G] Studies in Polish Linguistics – Paweł Rutkowski, Anna Kuder, Małgorzata Czajkowska-Kisil, Joanna Łacheta: The Structure of Nominal Constructions in Polish Sign Language (PJM): A Corpus-based Study (5)		6	0	1	2	2

Systemy do obsługi archiwum: (1) – brak systemu; (2) – CEEOL; (3) – DeGruyter; (4) – OJS; (5) – Wydawnictwo UJ.

Czynniki on-page (*content & technology*)

Osobnym obszarem badań nad pozycjonowaniem stron jest analiza czynników (w większości technicznych) zależnych od webmastera. Już w najstarszych oficjalnych poradnikach Google dla webmasterów (obecnie Google Search Console)⁷, podręcznikach SEO (np. Danowski,

.....

⁷ Google Centrum pomocy Search Console – <https://support.google.com/webmasters> (4.04.2016).

Tabela 7. *Metryki Google Scholar wybranych polskich periodyków humanistycznych*

Tytuł	Google Scholar	
	Indeks h5*	Mediana h5**
[A] „Ethos”	0	0
[B] „Filozofia Nauki”	3	3
[C] „Pamiętnik Teatralny”	0	0
[D] „Poznań Studies in Contemporary Linguistics”	6	8
[E] „Przegląd Religioznawczy”	0	0
[F] „Studia Humanistyczne AGH”	2	4
[G] „Studies in Polish Linguistics”	0	0

Źródło: Google Scholar Metrics – <https://scholar.google.pl/intl/pl/scholar/metrics.html>; * Indeks h5 – pięcioletni indeks Hirscha; ** Mediana h5 – średnia liczba cytowań dla prac, które tworzą jej indeks h5.

Makarczuk, 2007, 2009, 2011), analizach naukowych (Patil Swati, Pawar & Patil Ajay, 2013) oraz w większości serwisów raportujących (np. SEORCH)⁸ podkreślano wagę czterech czynników optymalizacyjnych. Były to kolejno: tytuł w HTML (*tag <title>*), opis strony (*<meta name=description content=... >*), tekst strony (*<body>*) oraz liczba linków wychodzących oraz wewnętrznych. Ich optymalne wartości zmieniały się w czasie, obecnie zaś 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 400–500 słów oraz pewną liczbę linków wewnętrznych (do kilkudziesięciu) oraz linków wychodzących (nie więcej niż 30)⁹. 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

⁸ SEORCH – <http://seorch.eu/> (4.04.2016).

⁹ Podano za wytycznymi zaimplementowanymi w najnowszych wtyczkach Yoast SEO dla WordPressa – <https://yoast.com/> (4.04.2016).

udokumentowana (Danowski, Makarczuk, 2011), więc nie warto przytaczać uzasadnień. Warto natomiast podkreślić, że w ostatnich rankingach Searchmetrics i MOZ (Tober, Hennig & Furch, 2014; Tober, Furch, Londenberg, Massaron & Grundmann, 2015) ich rola systematycznie maleje i według najnowszego rankingu MOZ cała grupa tytułowych czynników (*Page-Level Keyword-Agnostic Features*) ułożyła się dopiero na czwartej pozycji (MOZ, 2015).

Analiza kilku polskich periodyków humanistycznych (Tabela 8) przekonuje, że powyższe cztery zasady są raczej znane i stosowane. Duża część testowanych witryn ma optymalnie dobrany tytuł (A, C, D), niektóre wprawdzie są nieco za krótkie, choć mieszczące się w dopuszczalnym zakresie (B, F, G) i jedynie w jednym przypadku („Przegląd Religioznawczy”) jest on ewidentnie za krótki. Niepokojącą praktyką jest natomiast dość częste pomijanie opisu z sekcji meta (<meta...description>), co dotyczy ponad połowy przypadków (C, D, E, F); bywa, że opis jest za długi (G) lub za krótki (B) i jedynie w jednym przypadku ma on poprawne parametry (A). Analizowany tag nie jest wprawdzie bezpośrednio uwzględniany w algorytmie wyszukiwarki (o czym wielokrotnie informowali rzecznicy Google), lecz pełni funkcję psychologiczną, gdyż tekst opisu (z pogrubieniem słów kluczowych) pojawia się na stronach SERP-ów. W testowanej próbie tylko trzy witryny miały optymalną długość tekstu w sekcji <body> (E, F, G), w jednym przypadku był on zdecydowanie za długi jak na czołówkę (C), w trzech

Tabela 8. Główne czynniki on-page wybranych polskich czasopism humanistycznych

Tytuł	Długość tytułu	Meta desc.	Text in body	Links internal	Links external	Przyjazne linki
[A] „Ethos”	53	132	119	15	1	Nie
[B] „Filozofia Nauki”	28	28	166	27	10	Nie
[C] „Pamiętnik Teatralny”	47	0	1611	90	9	Tak
[D] „PSCL”	58	0	122	13	2	Tak
[E] „Przegląd Religioznawczy”	14	0	239	17	6	Tak
[F] „Studia Humanistyczne AGH”	24	0	385	24	1	Tak
[G] „Studies in Polish Linguistics”	29	453	483	66	5	Tak

przypadkach zaś zdecydowanie za krótki (A, B, D). Świadczy to, że webmasterzy nie doceniają tego parametru, choć istnieje wiele dowodów, że witryny za krótkie mogą zostać zakwalifikowane jako *thin content* i stracić wysokie pozycje w rankingu. Wszystkie testowane serwisy umieściły na stronie głównej wiele linków wewnętrznych. Fakt ów należy ocenić pozytywnie, z dwu powodów. Po pierwsze – odpowiednie kotwice linków (*anchor text*) skutecznie pozycjonują właściwe słowa kluczowe, po wtóre – linki wewnętrzne na stronie startowej ułatwiają nawigację dla użytkowników i robotów. Nieco inny obraz przynosi analiza linków wychodzących. W większości przypadków redakcje bardzo oszczędnie linkują do partnerów (szczególnie: A, D, F), a tym samym tracą szansę na poprawne ułożenie w sieci zależności tematycznych Google (te tworzą właśnie linki zewnętrzne); tracą ponadto szansę na to, że administrator podlinkowanej witryny odwdzięczy się im tym samym (linkiem zwrotnym, ang. *backlink*) – który – podkreślmy – jest obok słów kluczowych jednym z najważniejszych czynników w algorytmie Google. Duże znaczenie dla poprawnego pozycjonowania stron mają również tzw. przyjazne linki, gdyż pozwalają na umieszczenie słów kluczowych bezpośrednio w URL. Wykonana analiza przekonuje jednak, że właściwość ta (powszechnie dostępna w każdym współczesnym CMS-ie, który potrafi wykorzystać moduł *rewrite* serwera Apache) jest często wykorzystywana (C, D, E, F, G), choć w odosobnionych przypadkach (A, B) webmasterzy nie zauważają jej zalet.

Kolejnym elementem klasycznego audytu SEO jest analiza struktury kodu HTML oraz elementów pozycjonujących. W strukturze poprawnie zaprojektowanej strony www winien obowiązkowo wystąpić jeden nagłówek H1, który może (ale nie musi) się dzielić na H2, te z kolei na H3 itd. Dłuższe strony winny mieć przeto jasno podzieloną strukturę. Nie chodzi przy tym wyłącznie o czytelność, dla którego lektura ustrukturalizowanego tekstu jest łatwiejsza i przyspiesza selekcję, lecz w pierwszym rzędzie o przetwarzanie maszynowe i indeksowanie. Wykonane *ad hoc* badania (Tabela 9) ujawniają, że webmasterzy badanych witryn nie przywiązują większej wagi do struktury witryn. Wprawdzie większość badanych serwisów ma poprawną strukturę (A, B, D, F), lecz w połowie przypadków (A, B) nie jest ona ani zbyt rozbudowana (mało tagów H2/H3), ani nasycona wyróżnieniami (**, **). Trzy spośród siedmiu

badanych witryn obarczone były ponadto poważnymi błędami strukturalnymi (brak lub powielenie H1). W efekcie jedynie strukturę dwu serwisów (F, D) można uznać za poprawną i optymalnie ukształtowaną. Dodajmy, że podobną analizę przeprowadzono także na dużej próbie ($n = 60$) i uzyskano zbieżne wyniki. W pełni poprawne tagi H1 i H2 ma 36,7% witryn, struktura 25,0% jest obciążona niewielkimi błędami, zaś 38,3% jest zdeformowanych poprzez poważne uchybienia (np. brak lub powielone tagi H1).

Tabela 9. Analiza struktury kodu wybranych polskich czasopism humanistycznych

Tytuł	Elementy HTML	H1	H2	H3	Błędy strukt.	<p>		
[A] „Ethos”	76	1	0	0	Nie	1	0	2
[B] „Filozofia Nauki”	98	1	4	1	Nie	30	1	0
[C] „Pamiętnik Teatralny”	445	0	0	0	Tak	39	8	22
[D] „PSCŁ”	148	1	5	0	Nie	9	5	1
[E] „Przegląd Religioznawczy”	168	0	4	0	Tak	3	1	3
[F] „Studia Humanistyczne AGH”	218	1	1	2	Nie	7	7	5
[G] „Studies in Polish Linguistics”	400	2	4	3	Tak	10	17	16

Obliczenia własne. Narzędzie: SEORCH – <http://seorch.eu/> (5.04.2014).

Duże znaczenie dla poprawnego funkcjonowania stron www mają dodatkowe atrybuty towarzyszące tagom odpowiadającym za wstawienie grafik oraz hiperłączy. W przypadku grafiki jest to element ALT, 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 z problemami wzroku. Dla indeksowania i przeglądarek rola atrybutu ALT 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 testowanym korpusie czasopism humanistycznych (Tabela 10) pozwala na raczej pozytywną ocenę witryn w tym względzie, gdyż tylko w sporadycznych przypadkach wystąpiły braki.

Podobną rolę w hiperłączach pełni tekst kotwicy (ang. *anchor text*), czyli ciąg znaków znajdujący się pomiędzy otwierającym i zamykającym znacznikiem linku. Teksty anchorów w sposób bezpośredni wpływają, na jakie słowa kluczowe pozycjonowany jest dany URL. Teksty te winny być przeto starannie dobrane i najlepiej, aby były zgodne z pozycjonowaną frazą. Wykonany test (Tabela 10) przekonuje, że redakcje polskich czasopism humanistycznych doceniają rolę anchorów i powszechnie go stosują. Warto jednocześnie zwrócić uwagę, że związek anchorów i URL dotyczy wszystkich kategorii linków, a zatem zarówno prowadzących do naszego serwisu, gdzie webmaster zna słowa, na jakie chce pozycjonować własną stronę, jak i zewnętrznych, gdzie webmaster arbitralnie dobiera słowo, na jakie pozycjonuje zewnętrzną stronę (inny serwis). Występuje też sytuacja odwrotna: zewnętrzne serwisy zamieszczając hiperłącza do podstron naszego serwisu (backlinki), arbitralnie dobierają anchora (frazy i słowa kluczowe), na jaki pozycjonują naszą witrynę. Ponieważ zależności pomiędzy anchorami a linkami są wielostronne, webmasterzy nie mają wpływu na wszystkie frazy (słowa kluczowe), na które pozycjonowane będą ich strony. W efekcie anchora związane z określonym serwisem mają skłonność do dużego rozrzutu.

Kolejnym 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 (HTML 5), a już w październiku 2014 r. osiągnęła ona status rekomendacji W3C. Od początku prac nad specyfikacją znalazła ona wielu zwolenników, głównie ze względu na szereg nowych możliwości, jakie HTML 5 oferuje. Nowy standard jest bowiem bardziej semantyczny od HTML 4, m.in. umożliwi obsługę struktury na poziomie kodu, a w ślad za tym zagwarantuje nawigację w dokumencie (nowe tagi: `<header>`, `<main>`, `<article>`, `<aside>`, `<footer>`, `<nav>`). Dzięki tym funkcjonalnościom powstaje m.in. możliwość łatwego tworzenia stron responsywnych, czyli dostosowanych

do różnych urządzeń. Strona taka wyświetla się poprawnie zarówno na dużych ekranach, jak również na smartfonach i tabletach. Z kolei standard OGP (Open Graph Protocol) ma związek z 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. Rozwijają przeto semantykę kodowania, dzięki czemu łatwiejsze stają się przetwarzanie i ekstrakcja zasobów dostępnych w sieci. Mikroformaty znalazły szerokie zastosowanie na stronach SERP-ów i są znane m.in. jako recenzje (popularne gwiazdki) przy przebojach kinowych. Ich zastosowanie jest jednak znacznie szersze, gdyż mogą w sposób jednoznaczny identyfikować np. twórców z Google Scholar.

Tabela 10. *Obrazy i hiperłącza w strukturze wybranych polskich czasopism humanistycznych*

Tytuł	Obrazy			Hiperłącza wewnętrzne/zewnętrzne		
	N	Brak ALT	Brak Title	N	Brak Anchor	Brak Title
„Ethos”	1	0	1	15	0	14
„Filozofia Nauki”	1	1	1	37	0	36
„Pamiętnik Teatralny”	3	0	3	99	1	98
„PSCL”	1	0	1	15	0	4
„Przegląd Religioznawczy”	1	1	1	23	0	23
„Studia Humanistyczne AGH”	4	1	4	25	0	25
„Studies in Polish Linguistics”	6	0	3	71	0	69
Razem	17	3	14	285	1	269

Obliczenia własne. Narzędzie: SEORCH – <http://seorch.eu/> (5.04.2014).

Wykonana *ad hoc* analiza wybranych polskich czasopism humanistycznych nie prowadzi do budujących konkluzji. Redakcje testowanych czasopism są konserwatywne i z dużą ostrożnością podchodzą do nowinek technicznych. W żadnym z siedmiu testowanych serwisów nie używano bowiem ani HTML5, ani OGP, ani mikroformatów. Ich brak nie jest jednak sam w sobie ułomnością,

ale wskutek ich braku (HTML5) trudno jest zrealizować postulat responsywności, w ślad za tym utrudniony jest kontakt z użytkownikami pracującymi na urządzeniach mobilnych. Wśród testowanych serwisów tylko jeden („Studies in Polish Linguistics”) ma wdrożoną responsywność, którą wszelako realizuje starą i pracochłonną metodą (mobilna wersja strony serwowana przez podstawięcie [*user agent*]). Z drugiej strony, witryny polskich czasopism humanistycznych nie wydają się radykalnie zapóźnione w rozwoju, o czym świadczy m.in. powszechne odejście od przestarzałych standardów. Z analizy dużej próby badawczej ($n = 60$) wynika, że witryny czasopism humanistycznych niemal powszechnie (95,0%) zaniechały umieszczania kłopotliwych dla współczesnych przeglądarek ramek.

Pewne znaczenie dla skuteczności pozycjonowania ma kilka innych parametrów technicznych strony związanych z wydajnością serwisu. Parametry te mają charakter ściśle informatyczny (Tabela 11). Nikogo nie trzeba przekonywać, że nawet najlepsza strona www nie zyska sympatyków, gdy będzie wolno (lub irytująco wolno) wyświetlana, ma ona też znaczenie dla robotów Google. W kontekście SEO właściwość ta nabiera jeszcze większego znaczenia, gdyż kliknięciami z pozycji strony wyników wyszukiwarki najczęściej rządzi tzw. ekonomia uwagi/percepcji (Piekarski, 2014). Niezdecydowany użytkownik SERP-ów zwykle błyskawicznie dokonuje selekcji i wolno wyświetlająca się strona traci swą pozycję już na starcie, gdyż podczas ładowania się wolnej strony zniecierpliwiony czytelnik albo znajdzie inne trafienie (inną witrynę), albo podaży dalej, zapominając wrócić do ładującego się wyniku. Wykonana *ad hoc* analiza szybkości działania stron polskich czasopism humanistycznych (Tabela 11, kolumna: Speed desktop/mobile) prowadzi do budujących konkluzji. Wszystkie testowane strony w wyszukiwaniu prowadzonym z komputerów stacjonarnych (*desktop*) osiągnęły wysokie noty: co najmniej 70 punktów na 100 możliwych (a ponad połowa zbliżyła się do 90 punktów lub je przekroczyła). Nieco gorszy wynik osiągnęły w wyszukiwaniu mobilnym (najsłabsze lokowały się na poziomie 56–59 punktów, a najszybsze 78–79). Blisko skorelowane z powyższymi danymi są dwa inne parametry (wielkość pliku [*filesize*] oraz stosunek tekstu do kodu [*code text ratio*]). Optymalna wartość pierwszego nie powinna przekraczać 32 kB, zaś

Tabela 11. Wybrane parametry techniczne polskich czasopism humanistycznych

Tytuł	Filesize kB	Code text ratio	HTML Error	Speed desktop /mobile	Robotx txt	Sitemap xml	Canonical tags
„Ethos”	4,69	18,40%	0	74/59	1	1	0
„Filozofia Nauki”	6,01	20,65%	31	72/69	1	0	0
„Pamiętnik Teatralny”	36,72	36,71%	0	92/79	1	1	0
„PSCL”	14,27	5,72%	0	90/78	1	0	0
„Przegląd Religioznawczy”	20,57	9,30%	14	92/79	1	0	0
„Studia Humanistyczne AGH”	13,86	18,98%	1	70/56	1	1	0
„Studies in Polish Linguistics”	21,54	17,86%	10	88/75	1	1	0

Obliczenia własne. Narzędzie: SEORCH – <http://seorch.eu/> (5.04.2014).

drugi powinien być wyższy niż 25%. Analiza danych przekonuje, że większość badanych serwisów (6 na 7) serwuje strony poprawnej wielkości, choć nie są one zoptymalizowane (tylko 1 na 7 legitymuje się właściwym współczynnikiem objętości tekstu do kodu). Bardziej precyzyjne dane dostarcza analiza dużej próby ($n = 60$), z których wynika, że średnia wartość wskaźnika *text code ratio* wynosi 19,05% i jedynie 40% witryn legitymuje się poprawnym wskaźnikiem (powyżej 20,00). Oznacza to, że strony polskich czasopism humanistycznych (ale postawmy tezę: nie tylko one) są przeładowane kodem rozmaitych wtyczek, w szczególności współpracujących z sieciami społecznościowymi i kodami śledzenia.

Aby usprawnić współpracę witryn z crawlerami wyszukiwarek, administratorzy serwisów winni udostępnić dwa specyficzne pliki sterujące. Pierwszy (*robots.txt*) wskazuje listę dozwolonych robotów oraz ich uprawnienia, drugi (*sitemap.xml*) jest mapą witryny przygotowaną specjalnie dla robota. Jakkolwiek pliki te nie są obowiązkowe, ich obecność wpływa na efektywność indeksowania. Z przeprowadzonych testów (Tabela 11) wynika, że administratorzy testowanych witryn powszechnie stosują pierwszy z nich (*robots.txt*), rzadko

natomiast drugi (sitemap.xml). Zbliżone rezultaty przyniosła analiza dużej próby ($n = 60$), z której wynikało, że plik robots.txt ma 53,3% witryn, zaś sitemap.xml – 26,7%.

Ważnym zagadnieniem z punktu widzenia SEO jest ponadto problem powielających się treści lub funkcjonowania tej samej witryny na kilku różnych domenach. Reakcja robota na powielającą treść jest jednoznaczna – zostawia w indeksie tylko jeden rekord (zwykle najstarszy), pozostałe zaś usuwa z listy wyników. Ponieważ administratorzy nierzadko tworzą takie strony ze względów w pełni uzasadnionych (np. wersje mobilne, wersje do wydruku itp.), zachodzi konieczność instruowania robotów o poprawnej tożsamości. Na pomoc w tym przypadku przychodzi instrukcja `rel="canonical"`, informująca robota, która wersja strony (niezależnie od wywołania) jest wariantem głównym. Instrukcja ta pełni ponadto inne dobroczynne funkcje, m.in. pozwala scalić ruch do tej samej strony wywołanej na kilka różnych sposobów (np. z `www` i bez `www`). Przeprowadzone *ad hoc* badania (Tabela 11) prowadzą do konkluzji, że rozwiązanie to jest niemal nieznanym webmasterom tworzącym serwisy polskich pism humanistycznych. Tezy tej jednak nie potwierdza analiza dużej próby ($n = 60$), z której wynika, że tagi kanoniczne ma 43,3% witryn.

Linki przychodzące (*backlinks*)

Najstarszy algorytm tworzenia rankingów w wyszukiwarkach internetowych bazował na analizie linków. Zasada jego działania była prosta, ponieważ opierała się na założeniu, że witryna, do której prowadzi więcej linków, jest ważniejsza od witryny, do której liczba linków jest mniejsza. Z czasem zasada ta została zmodyfikowana i prócz liczby linków wyszukiwarki zliczały ich referencje (skąd linki pochodzą). Wówczas wartość linku pochodzącego z witryny, która sama ma dużo linków przychodzących, była większa niż z witryny, która ma ich mniej. Na zasadzie tej opierał się też PageRank (PR) opracowany w 1998 roku przez jednego z twórców Google – Larry'ego Page'a. Idea PR była z czasem wielokrotnie modyfikowana i jego rola w algorytmie znacząco spadła (dziś jest ledwie symboliczna). Modyfikacje miały przede wszystkim związek

z nieetycznymi zachowaniami webmasterów, którzy 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 linków (Google, 2016). Obecnie backlinki wciąż mają duże znaczenie w algorytmie, ale sposób ich zliczania i mapowania pozostaje tajemnicą. Rozmaite poszlaki wskazują, że rolę PageRank przejął nowy wskaźnik (TrustRank) bazujący na identycznym rozwiązaniu stosowanym ongiś w wyszukiwarce Yahoo! (Krishnan & Raj, 2006). W efekcie w algorytmie Google backlinki wciąż mają duże znaczenie, ale zasada ich zliczania jest inna. Prawdopodobnie aktualnie link przychodzący nadaje wskazywanej witrynie część własnego autorytetu (*authority*), zaufania (*trust*) i rangi (*rank*). Zatem dzisiaj nie liczba, ale jakość linków odgrywa pierwszorzędną rolę. Taką supozycję potwierdza wiele serwisów analitycznych, m.in. Majestic SEO czy Ahrefs oraz społeczności eksperckie (MOZ, Searchmetrics), które tworzą własne odpowiedniki wskaźnika TrustRank. Warto jednocześnie dodać, że częściowo rolę backlinków przejęły inne czynniki algorytmu Google (sygnały sieci społecznych), które będą scharakteryzowane w ostatnim podrozdziale.

Powyższą hipotezę potwierdzają dane pochodzące z Ahrefs – prestiżowego serwisu do analizy backlinków (Tabela 12), gdzie zestawiono rozmaite kategorie linków oraz wskaźnik jakościowy URL Rating. Analiza danych jednoznacznie wskazuje, że pomiędzy liczbą linków (w dowolnym przekroju) a metryką reputacji (URL Rating) nie istnieje żadna współzależność. Niezależnie więc, czy uwzględnimy w obliczeniach ogólną liczbę backlinków (Backlinks Total) lub ich poszczególne kategorie (Dofollow – czyli tzw. linki liczące się, Gov – linki pochodzące z domen rządowych, Edu – linki pochodzące z witryn naukowych i edukacyjnych) czy liczbę domen, z których one pochodzą (Referring Domain), ich liczba nie jest skorelowana ze wskaźnikami jakościowymi.

Powyższa uwaga pozostaje w wyraźnej sprzeczności z literaturą, gdyż w większości podręczników SEO (Danowski, Makarczuk, 2007, 2009, 2011 i in.), a nawet w części poważnych serwisów analitycznych (MOZ Open Site Explorer; monitorbacklinks.com i in.) wciąż podkreślana jest przesadnie duża rola liczby linków i sugerowany

jest precyzyjny pomiar ich zróżnicowania. Mimo powyższych uwag webmasterzy winni monitorować własną witrynę (Google Webmaster Tools [GWT], Google Analytics [GA]), gdyż jest to jedyny sposób na precyzyjne rozpoznanie, jak użytkownicy trafiają na naszą witrynę i jak korzystają z jej treści. Analiza danych z dużej próby ($n = 60$) przekonuje jednak, że nie jest to zjawisko powszechne, gdyż z kodów śledzenia Google Analytics korzysta 36,7% polskich czasopism humanistycznych i prawdopodobnie podobna liczba ma profil na GWT.

Tabela 12. Linki przychodzące do witryn wybranych polskich czasopism humanistycznych

Tytuł	Ahrefs backlink checker					URL Rating
	Backlinks Total	Referring Domain	Backlink Types			
			Dofollow	Gov	Edu	
„Ethos”	131 903	80	131 223	0	425	26
„Filozofia Nauki”	253	32	205	0	43	22
„Pamiętnik Teatralny”	6	3	6	0	3	13
„PSCL”	4 327	496	2 565	2	336	17
„Przegląd Religioznawczy”	40	6	38	0	35	7
„Studia Humanistyczne AGH”	661	14	644	0	8	15
„Studies in Polish Linguistics”	76	12	24	0	9	14

Źródło: Obliczenia własne na bazie serwisu Ahrefs (Ahrefs backlink checker) – <https://ahrefs.com/> (4.04.2016).

Metryki autorytetu, zaufania i reputacji (*authority, trust, rank*)

W poprzednim podrozdziale wspomniano o stopniowej ewolucji PageRank Google i jego prawdopodobnym zastąpieniu przez wskaźnik wzorowany na TrustRank opracowany przez naukowców z Yahoo! (Krishnan & Raj, 2006). Informację tę już w 2007 roku potwierdził rzecznik prasowy Google (Cutts, 2007), a następnie była wielokrotnie komentowana (Patel & Patel, 2014; Bailyn, 2014 i in.). Ponieważ dokładna zasada działania i kalkulacji tego wskaźnika jest tajemnicą handlową Google, społeczności eksperckie i firmy SEO 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, przyjmując wartości 0–100. Z kolei MozRank (*link popularity*) odzwierciedla znaczenie danej strony w Internecie skalkulowane poprzez liczbę 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 logarymicznie i mogą przybrać wartości z zakresu 0–10.

Zestawienie metryk autorytetu testowanych witryn (Tabela 13) pozwala stwierdzić, że większość serwisów legitymuje się wysokimi wskaźnikami autorytetu/reputacji zliczanymi wg metody MOZ (witryny: A, B, D, F); średnie wartości DA otrzymały witryny „Studies in Polish Linguistics” (G) oraz „Pamiętnik Teatralny” (C), które jednak równolegle legitymowały się wysokimi wartościami MR i MT. Niskie wartości uzyskała jedynie witryna „Przeglądu Religioznawczego”, która w okresie przeprowadzania testów miała przejściowe kłopoty z domeną. 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. Warto jednocześnie dodać, że w środowisku badaczy SEO panuje opinia, że w algorytmie Google autorytet jest blisko skojarzony z pojęciami „władza” i „wiedza”. Potwierdza tę supozycję praktyka, gdyż najwyższe wartości DA, MR i MT mają strony naukowe i rządowe.

Pewien wpływ na zaobserwowane oceny autorytetu i zaufania ma także wiek domeny. Potwierdzają tę tezę w szczególności przykłady witryn działających długo na tej samej domenie (A i B), podczas gdy większość witryn młodszych osiąga słabsze wyniki (C, E, G). Pomiar pozostaje w zgodzie ze starą zasadą SEO, że „stabilne trwanie buduje reputację i zaufanie”, co wszelako nie oznacza, że witryny młode zawsze długo muszą pracować na zaufanie. Poświadczają tę hipotezę dwa przykłady witryn mających czteroletni staż (D, F), które jednak poprzez dobrą współpracę (backlinki z renomowanych źródeł) uzyskały wysokie wartości metryk.

Tabela 13. *Metryki autorytetu wybranych polskich czasopism humanistycznych*

Tytuł	Internet Archive		MOZ			Ahrefs	
	Domain Age	Dom. Auth.	Page Auth.	Moz Rank	Moz Trust	URL Rating	Dom. Rating
[A] „Ethos”	10	63	46	4,67	5,11	26	69
[B] „Filozofia Nauki”	11	72	42	4,31	4,72	22	66
[C] „Pamiętnik Teatralny”	6	37	23	5,44	5,52	13	51
[D] „PSCL”	4	70	35	4,00	4,23	17	65
[E] „Przegląd Religioznawczy”	4	13	17	4,28	0,00	7	39
[F] „Studia Humanistyczne AGH”	4	71	31	4,32	4,77	15	65
[G] „Studies in Polish Linguistics”	3	32	27	5,02	5,14	14	52

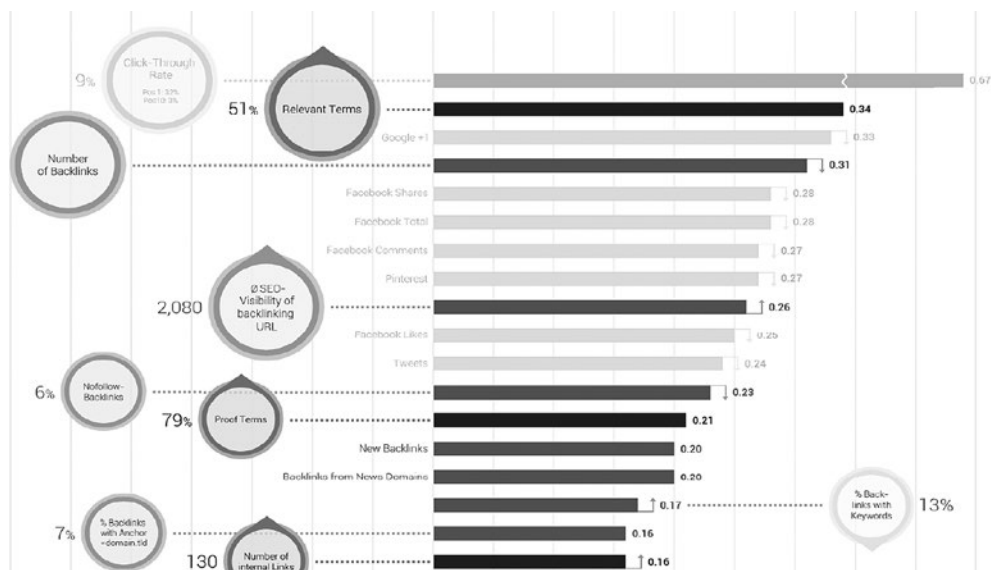
Obliczenia własne na bazie serwisów: Internet Archive; MOZ Open Site Explorer, Ahrefs (7.04.2016).

Sygnaly z sieci społecznościowych (social)

Najnowszymi czynnikami, które zaczęły odgrywać istotną rolę w algorytmach wyszukiwarek, są sygnały z sieci społecznych. Jakkolwiek pomiar tych sygnałów wprowadzono do algorytmu dopiero po 2010 roku, zrobiły one zawrotną karierę i od kilku lat lokują się na szczycie rankingów korelacji (Tober, Hennig & Furch, 2014; Tober, Furch, Londenberg, Massaron, Grundmann & 2015; MOZ, 2015). W rankingu Searchmetrics za rok 2014 w grupie 15 najwyższej cenionych czynników sygnały z social mediów stanowiły połowę (na wykresie 3. serie oznaczone jasnym odcieniem). 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).

Pomiar sygnałów wysyłanych przez sieci społeczne, wskazujących na polskie czasopisma humanistyczne (Tabela 14), prowadzi do bardzo pesymistycznych wniosków. Badane periodyki są bowiem niemal nieobecne w tej przestrzeni.

Wykres 3. Google Ranking Factors 2014 (Searchmetrics)



Źródło: Tober, Hennig & Furch, 2014.

Wśród siedmiu testowanych pism tylko jeden tytuł – „Studia Humanistyczne AGH” – posiadał profil na Facebooku. Wprawdzie konto istnieje od ponad trzech lat i zgromadziło wokół siebie społeczność (248 fanów), jednakże aktywność administratora jest bardzo niska, by nie rzec żadna (4–5 postów rocznie), więc nie generuje ona zauważalnego ruchu (24 sygnały, w tym: 8 udostępnień, 15 polubień, 1 komentarz). W przypadku pozostałych ruch jest jeszcze mniejszy, gdyż redakcje nie posiadają profili, stąd i liczba uzyskanych sygnałów jest ledwie symboliczna (łącznie 23). Warto dodać, że problem nieobecności polskich pism humanistycznych w mediach społecznościowych nie dotyczy jedynie popularnych sieci (Facebook, Twitter, Google Plus), lecz także portali naukowych (Rozkosz, 2014b; Kulczycki, 2013a). Wykonany *ad hoc* pomiar udowodnił, że tylko jedna redakcja („Filozofia Nauki”) posiada profil w Academia.edu (31 followers), zaś w ResearchGate nie funkcjonuje żaden badany tytuł. Owocuje to podobnie niskim ruchem w Academia.edu, gdzie udostępniono łącznie 45 artykułów opublikowanych w tytułach z badanego korpusu, zaś w ResearchGate tylko 18 („Filozofia Nauki” – 10; „Pamiętnik Teatralny” – 4; „Studia Humanistyczne AGH” – 2 oraz „Przegląd Religioznawczy” – 2).

Tabela 14. Sygnały z sieci społecznych do wybranych czasopism humanistycznych

Strona	Profil/liczba fanów			Sygnały do strony głównej**						
	FB	G+	Tw	Facebook			Twitter	Google+	Inne	Razem
				Shares	Likes	Comm	Tweets	Pluses		
[A] „Ethos”	0	0	0	1	0	0	0	0	0	1
[B] „Fil. Nauki”	0	0	0	10	1	3	0	0	0	14
[C] „Pam. Teatr.”	0	0	0	1	0	1	0	0	0	2
[D] „PSCL”	0	0	0	0	0	0	0	0	0	0
[E] „Prz. Relig.”	0	0	0	0	0	0	0	0	0	0
[F] „St. Hum. AGH”	248	0	0	8	15	1	0	0	0	24
[G] „St. Pol. Ling.”	0	0	0	1	4	1	0	0	0	6
Razem	248	0	0	21	20	6	0	0	0	47

** Linki do czołówki stanowią zwykle ok. 25–35% sygnałów do domeny.

Źródło danych: serwis <http://www.seocentro.com/tools/search-engines/link-popularity.html> (4.04.2016).

Jakkolwiek trudno zdiagnozować powód tego stanu rzeczy, wydaje się, że jest on skutkiem pasywności redakcji, które nie doceniają lub lekceważą komunikację w sieciach społecznościowych.

Wnioski

Zaprezentowane badania nie były wykonane na reprezentatywnej próbie, przeto nie uprawniają do formułowania kategoriycznych wniosków. Praca ma charakter przede wszystkim jakościowy, a zaobserwowane prawidłowości mają charakter hipotetyczny. Większość testów przeprowadzono na tzw. małej próbie ($n = 16$ i $n = 8$), do której zakwalifikowano wyłącznie kwartalniki mające w rankingu MNiSW za rok 2015 co najmniej 10 punktów, w ślad za tym większość wniosków odnosi się do periodyków z tej grupy, tj. czasopism o ugruntowanej pozycji, cieszących się w środowisku znacznym prestiżem.

Badania nie pozwoliły jednoznacznie odpowiedzieć na postawione we wstępie pytanie. Stosowanie zasad SEO przez polskie pisma humanistyczne jest bowiem rozmaicie realizowane w różnych obszarach. Nie ulega wątpliwości, że pozycjonowanie czasopism (nie tylko humanistycznych) nie jest czynnością prostą, ponieważ ze względu na znaczne rozproszenie zapytań (o poszczególne artykuły) kwerendy lokują się na pozycjach *long tail*. Z drugiej strony, popyt na informacje naukowe jest w porównaniu z zapytaniami o charakterze usługowo-handlowymi bardzo mały, co sprawia, że globalna liczba zapytań (nawet o markę, tj. tytuł czasopisma) jest niska; choć jak przekonały badania, względnie stała. Z uwagi na unikalny charakter obiektów (artykuły) redakcje powinny je pozycjonować na frazę tytułową. Badania przekonały, że postulat ten jest realizowany przede wszystkim w serwisach wykorzystujących profesjonalne oprogramowanie archiwalne (OJS, dSpace) lub deponowanych u dużych wydawców (Wydawnictwo UJ) lub dostawców (CEEOL). Artykuły z tych źródeł były bowiem prawidłowo zidentyfikowane i przetwarzane w Google Scholar. Analiza czynników on-page (zależnych od twórców) przekonuje, że webmasterzy i administratorzy badanych serwisów w dużym stopniu stosują klasyczne zasady SEO, witryny są posadowione na szybkich i wydajnych serwerach, wszelako rzadko lub wcale stosują nowe standardy (HTML5, OGP itp.) i nie dbają o użytkowników korzystających z urządzeń mobilnych. Badanie czynników off-page (niezależnych od webmasterów) doprowadziło do kilku dalszych konkluzji. Większość badanych czasopism 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”. Ponadto analiza backlinków wykazała, że obecnie nie istnieje żadna korelacja pomiędzy ich liczbą a metrykami reputacji i autorytetu. Zdecydowanie negatywny wynik przyniosła analiza obecności polskich czasopism humanistycznych w przestrzeni sieci społecznych, mimo iż sygnały płynące z tych źródeł są obecnie bardzo cenne dla wyszukiwarek i lokują się na czele rankingów korelacyjnych. Pomiar sygnałów wysyłanych przez sieci społeczne, wskazujących na polskie czasopisma humanistyczne, przekonuje, że są one nieobecne w tej przestrzeni.

Literatura

- Anderson, C. (2004). The Long Tail. *Wired Magazine*, 10(XI). Pozyskane z: <http://www.wired.com/2004/10/tail/>
- Anderson, C. (2006). *The long tail: why the future of business is selling less of more*. New York: Hyperion. Pol. tłum. (2008): *Długi ogon: ekonomia przyszłości – każdy konsument ma głos*. Poznań: Media Rodzina.
- Bailyn, E. (2014). *PageRank vs. TrustRank in Google's Algorithm*. Pozyskane z: <http://firstpagesage.com/seo-blog/seo-basics/pagerank-vs-trustrank-in-googles-algorithm/>
- Bednarek-Michalska, B. (2012). *Modele biznesowe otwartego publikowania naukowego: informator dla polskich wydawców uczelnianych*. Toruń: Stowarzyszenie EBIB. Pozyskane z: http://open.ebib.pl/ojs/index.php/wydawnictwa_zwarte/article/view/111/203
- Cutts, M. (2007). *Trust Rank Explained [Suresh Babu Interviews Matt Cutts in Pubcon 2007]*. Pozyskane z: <https://youtu.be/p8mUXQzwEvs>
- Danowski, B., Makarczuk, M. (2007). *Pozycjonowanie i optymalizacja stron WWW*. Gliwice: Wydawnictwo Helion.
- Drabek, A. (2015). Dostępność polskich czasopism naukowych w sieci, *Nowa Biblioteka*, 1, 47–56.
- Geddes, B. (2014). *Advanced Google AdWords*. Indianapolis: John Wiley & Sons.
- Gemius (2016). *Wyszukiwarki – silniki. Ranking gemius Traffic*. Pozyskane z: <http://www.ranking.pl/pl/rankings/search-engines.html>
- Google (2016). *Google Algorithm Change History*. Pozyskane z: <https://moz.com/google-algorithm-change>
- Kaushik, A. (2010): *Web analytics 2.0. Świadome rozwijanie witryn internetowych*. Gliwice: Helion.
- Kozłowski, T., Bomba, R. (2016). Polskie czasopisma naukowe z zakresu nauk humanistycznych i społecznych w Internecie. W: P. Siuda (red.): *(Nie)obecna nauka. Strategie promocyjne polskich czasopism naukowych z zakresu nauk humanistycznych i społecznych* (s. 46–85). Poznań: Wydawnictwo Naukowe Instytutu Filozofii UAM.
- Krishnan, V., & Raj, R. (2006). Web Spam Detection with Anti-Trust Rank. *Proceedings of the Second International Workshop on Adversarial Information Retrieval on the Web – AIRWeb 2006*, 6, 37–40. Pozyskane z: <http://airweb.cse.lehigh.edu/2006/krishnan.pdf>

- Kulczycki, E. (2013a): *Altmetrics – dlaczego warto interesować się alternatywnymi metrykami*. Pozyskane z: http://ekulczycki.pl/warsztat_badacza/altmetrics-dlaczego-warto-interesowac-sie-alternatywnymi-metrykami/
- Kulczycki, E. (2013b). *Jak dodać prace do Google Scholar i zwiększyć liczbę cytowań oraz indeks Hirscha. Poradnik dla początkujących*. Poznań: Stowarzyszenie EBIB. Pozyskane z: http://open.ebib.pl/ojs/index.php/wydawnictwa_zwarte/article/view/115/0
- Kulczycki, E. (2013c): *Otwarte czasopisma. Zakładanie czasopism naukowych oraz transformacja czasopism zamkniętych*. Toruń: Stowarzyszenie EBIB. Pozyskane z: http://open.ebib.pl/ojs/index.php/wydawnictwa_zwarte/article/view/112/205
- MOZ (2015). *Search Engine Ranking Factors 2015. Expert Survey and Correlation Data*. Pozyskane z: <https://moz.com/search-ranking-factors>
- Nahotko, M. (2010). *Komunikacja naukowa w środowisku cyfrowym. Globalna biblioteka cyfrowa w informatycznej infrastrukturze nauki*. Warszawa: Wydawnictwo SBP.
- Patel, J., & Patel, S. (2014). Efficient calculation of PageRank using TrustRank and Anti-TrustRank. *International Journal of Engineering Development and Research*, 2(2), 1895–1898.
- Patil Swati, P., Pawar, B. V., & Patil Ajay, S. (2013). Search Engine Optimization: A Study. *Research Journal of Computer and Information Technology Sciences*, 1(1), 10–13.
- Piekarski, K. (2014). *Ekonomia percepcji. Mechanizmy selekcjonowania informacji w internecie*. Praca doktorska. Katowice: Uniwersytet Śląski w Katowicach. Pozyskane z: <http://www.sbc.org.pl/Content/126980/doktorat3505.pdf>
- Rozkosz, E. (2014a). Open Journal Systems – wartość dodana czasopisma. *Biuletyn EBIB*, 4(149). Pozyskane z: <http://open.ebib.pl/ojs/index.php/ebib/article/view/246/419>
- Rozkosz, E. (2014b). Serwisy społecznościowe dla naukowców (SSN) na przykładzie ResearchGate i Academia.edu. *Forum Bibliotek Medycznych*, 2(7), 16–24.
- Sapa, R. (2009). *Metodologia badań obszaru pośredniczenia w komunikacji naukowej z perspektywy nauki o informacji*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Siuda, P. (red.). (2016a). *(Nie)obecna nauka. Strategie promocyjne polskich czasopism naukowych z zakresu nauk humanistycznych*

- i społecznych*. Poznań: Wydawnictwo Naukowe Instytutu Filozofii UAM.
- Siuda, P. (2016b). Wprowadzenie. Czy humanistyka i nauki społeczne są (nie)obecne? W: P. Siuda (red.): *(Nie)obecna nauka. Strategie promocyjne polskich czasopism naukowych z zakresu nauk humanistycznych i społecznych* (s. 4–10). Poznań: Wydawnictwo Naukowe Instytutu Filozofii UAM.
- Siuda, P., Stachura, K. (2016): Instytucjonalne uwarunkowania promocji czasopism naukowych z zakresu nauk humanistycznych i społecznych. W: P. Siuda (red.): *(Nie)obecna nauka. Strategie promocyjne polskich czasopism naukowych z zakresu nauk humanistycznych i społecznych* (s. 11–32). Poznań: Wydawnictwo Naukowe Instytutu Filozofii UAM.
- Szprot, J. (red.). (2014). *Otwarta nauka w Polsce 2014. Diagnoza*. Warszawa: Wydawnictwa ICM. Pozyskane z: <http://pon.edu.pl/index.php/nasze-publicacje?pubid=13>
- Tober M., Furch D., Londenberg K., Massaron L., & Grundmann J. (2015). *Search Ranking Factors and Rank Correlations 2015*. San Mateo: Searchmetrics. Pozyskane z: <http://www.searchmetrics.com/knowledge-base/ranking-factors/>
- Tober, M., Hennig, L., & Furch, D. (2014). *SEO Ranking Factors and Rank Correlations 2014*. San Mateo: Searchmetrics. Pozyskane z: <http://www.searchmetrics.com/knowledge-base/ranking-factors-2014/>
- Wang, F., Li, Y., & Zhang, Y. (2011). An Empirical study on the Search Engine Optimization Technique and Its Outcomes, *Proc. 2nd International Conference on AIMSEC*, 2767–2770. doi: 10.1109/AIMSEC.2011.6011361