

dr inż. Marcin Krzesaj¹

Zakład Ekonomiki Informacji, Wydział Ekonomiczny
Uniwersytet Opolski

Ocena jakości informacji w wybranych polskich katalogach stron WWW

WSTĘP

Informacja we współczesnym świecie stała się kolejnym, istotnym zasobem gospodarczym i czynnikiem wytwórczym. Wyodrębnienie informacji wiarygodnych i rzetelnych stanowi istotny problem, szczególnie dla przeciętnego użytkownika Internetu, który nie posiada wystarczającej wiedzy i nie jest w stanie samodzielnie ocenić pozyskane informacje. W sieci WWW panuje chaos informacyjny, a poziom dostępnej informacji o zły lub bardzo zły jakości gwałtownie rośnie.

J. Unold zauważa, że pozyskiwanie informacji w sieci angażuje dwie zasadnicze formy: przeglądanie i wyszukiwanie [Unold, 2011, s. 165]. Z przeprowadzonych badań wykorzystania technologii informacyjno-telekomunikacyjnych w gospodarstwach domowych w 2015 roku w celach komunikacyjnych wynika [Spoleczeństwo informacyjne..., (http)], że trzecią najczęściej wykonywaną za pośrednictwem Internetu czynnością było wyszukiwanie informacji o towarach i usługach – korzystało z niej 42,2% respondentów w wieku 16–74 lata.

Jedno z najbardziej rozpowszechnionych podejść do klasyfikacji narzędzi zdobywania informacji w sieci wymienia ich dwie kategorie: katalogi i wyszukiwarki [Maciejewski, 2004, s. 124, 129]. W innym ujęciu większość narzędzi pozyskiwania informacji w Internecie można zakwalifikować do jednej z trzech kategorii: katalogi ogólne, katalogi specjalistyczne oraz wyszukiwarki [Unold, 2011, s. 165]. Oprócz nich wyróżnia się narzędzia specjalistyczne oraz hybrydowe, np. metawyszukiwarki, roboty internetowe, portale wertykalne.

Z problemami natury jakościowej borykają się zarówno wyszukiwarki, jak i katalogi stron WWW. Autor w niniejszej publikacji skoncentrował się na katalogach stron WWW. Jakość udostępnianych przez nie informacji jest kontro-

¹ Adres korespondencyjny: Uniwersytet Opolski, ul Ozimska 46a, 45-058 Opole; tel. (77) 40 16 869; e-mail: marcin.krzesaj@uni.opole.pl.

wersyjna. Podobnie z deklarowanym celem powstania, który często jest niejasny a postawione przez ich właścicieli zadania nie zawsze są realizowane, np. wyszukiwanie wartościowych stron o zweryfikowanej jakości. Na jakość informacji w tworzonych katalogach może wpływać przyjęty model projektowania katalogów. Spotykaną praktyką jest tworzenie wpisów w modelu otwartym przez grupę wolontariuszy (ochotników), którzy za darmo recenzują proponowane witryny. Przykładowo w darmowym katalogu *Dmoz* 91 385 edytorów wpisało 3 945 185 stron w 1 030 534 kategoriach (stan na dzień 21.06.2016 r.), a wpisy dokonywane są w 90 językach. Często w katalogach stron brak jest jakiegokolwiek kontroli jakości zamieszczanych i publikowanych informacji, np. oceny jakości stron wystawiane są przez użytkowników Internetu.

Celem artykułu jest ocena jakości informacji w wybranych polskich katalogach stron WWW. Na wstępie artykułu zaprezentowano charakterystykę, klasyfikację i ranking popularności katalogów stron. Następnie przedstawiono metodykę badania oceny jakości informacji w katalogach stron WWW. Na podstawie zidentyfikowanych kryteriów opracowano narzędzie w postaci kwestionariusza ankietowego zawierające zestaw pytań podlegających ocenie. W drugiej części artykułu zaprezentowano wyniki pilotażowego pomiaru jakości informacji w wybranych polskich katalogach stron WWW. Artykuł kończy podsumowanie zawierające wnioski oraz wskazówki dotyczące dalszych kierunków badań.

CHARAKTERYSTYKA KATALOGÓW STRON WWW

Katalogi stron WWW zajmują czołowe miejsce w procesie pozyskiwania informacji w Internecie. Umożliwiają one kontekstowe i strukturalizowane przeglądanie zawartości Internetu. Wyszukiwarki natomiast pozwalają na szukanie według zadanych słów i kluczowych fraz, aczkolwiek bez kontekstu. Według J. Unolda przeglądanie jest bardziej efektywne w przypadku katalogów stron WWW natomiast wyszukiwanie kiedy wprowadzamy słowa lub frazy do wyszukiwarek internetowych.

Najczęściej spotykany katalog stron WWW zawiera zbiór odnośników do stron internetowych, posegregowanych tematycznie na kategorie przez zespół redakcyjny danego serwisu. Katalogi z reguły mają strukturę drzewiastą zawierającą listę głównych kategorii. W kategoriach opisane są strony WWW, które podlegają subiektywnej ocenie przeprowadzanej przez zespół redakcyjny danego katalogu. Poszczególne kategorie tematyczne posiadają niższe stopnie hierarchii zawężające zakres poszukiwania. Dla przykładu w katalogu stron WWW portalu Onet.pl, znajduje się 17 kategorii tematycznych, np. „Biznes i ekonomia”, „Rozrywka”, „Motoryzacja”, „Społeczeństwo”, „Firmy według Branż”. Dodatkową funkcjonalnością jest możliwość wyszukiwania stron z danego regionu poprzez wybór kategorii „Według regionów”.

Katalogi stron WWW pozwalają zarówno na przeglądanie, jak i wyszukiwanie informacji. Wykorzystując mechanizm linków oraz hierarchiczną strukturę katalogu użytkownik może przechodzić od kategorii ogólnych do podkategorii, zawężając zakres poszukiwań. Większość katalogów WWW ma wbudowane własne menu wyszukiwania, co pozwala na realizację drugiej formy pozyskiwania informacji w Internecie. Z reguły oprócz wyszukiwania informacji w bazie katalogu WWW można pozyskiwać wyniki wyszukiwania z najpopularniejszych światowych silników wyszukiwawczych.

W literaturze można odnaleźć różne klasyfikacje katalogów stron WWW ze względu na ich tematykę (np. ogólna, specjalistyczna), modele projektowania katalogów (zamknięty, otwarty), sposób prowadzenia (np. moderowany, niemoderowany, płatny, bezpłatny), czy też konfigurację (np. odmienne metody rejestracji stron WWW). Analizując dostępne katalogi stron WWW można zaproponować ich poniższą typologię:

- katalogi ogólne (np. *katalog.interia.pl*, *katalog.wp.pl*),
- katalogi tematyczne – branżowe, specjalistyczne (np. *motolog.pl*, *dirbud.pl*),
- katalogi stron SEO (np. *katalogseo.pl*, *seodirect.pl*).

Tabela 1 zawiera ranking katalogów utworzony na podstawie odwiedzin internautów z obszaru Polski przypadający na każdy pierwszy kwartał 2013–2016.

Tabela 1. Ranking katalogów stron WWW

Lp.	Katalog	1 kwartał 2013 r.	1 kwartał 2014 r.	1 kwartał 2015 r.	1 kwartał 2016 r.
1	start24.pl	23,21%	29,97%	35,47%	39,19%
2	katalog.onet.pl	33,17%	28,33%	27,83%	29,26%
3	katalog.wp.pl	36,58%	35,97%	26,70%	23,68%
4	dmoz.org	1,65%	2,36%	4,7%	2,41%
5	katalog.hoga.pl	0,66%	0,89%	1,39%	2,05%
6	katalog.interia.pl	1,64%	0,76%	2,03%	0,97%
7	gery.pl	0,69%	0,54%	0,53%	1,04%
8	stronastartowa.pl	0,22%	0,69%	0,24%	0,79%
9	katalog.prv.pl	0,00%	0,00%	0,30%	0,23%
10	dodaj.pl	0,06%	0,01%	0,09%	0,14%

Źródło: opracowanie własne na podstawie badań firmy Gemius SA, gemiusTraffic.

Według danych PBI/Gemius przypadających na pierwszy kwartał 2016 roku pierwsze miejsce w rankingu popularności katalogów stron WWW wśród polskich użytkowników zajmuje *start24.pl* z 39,19% udziałem, na drugim miejscu znajduje się *katalog.onet.pl* z 29,26% udziałem, trzecie miejsce zajmuje *katalog.wp.pl* z 23,68% udziałem.

METODYKA BADAŃ

Zaproponowana metodyka oceny jakości informacji w katalogach stron WWW składa się z czterech etapów: konceptualizacja jakości informacji, określenie kryteriów jakości informacji, określenie miar jakości dla zadanych obiektów informacyjnych oraz pomiar jakości informacji. Ze względu na ograniczone ramy niniejszej publikacji w pracy przedstawione jedynie streszczenie zastosowanej metodyki. Pełny opis metodyki oceny jakości informacji w katalogach stron WWW przedstawiono w publikacji *Information quality evaluation methodology for web page directories* [Krzesaj, 2016a, s. 185–194].

Określając badany problem w pracy przyjęto następującą za J. Ruževičius i A. Gedminaitė definicję jakości informacji [Ruževičius, Gedminaitė, 2007, s. 19]: „Jakość informacji – to suma własności i wymiarów informacji pozwalająca zapewnić lub przekroczyć oczekiwania i wymagania oraz wyrażone i nieujawnione potrzeby użytkownika – pracownika wiedzy”. Powyższa definicja ma interdyscyplinarny charakter i umożliwia jej ujęcie w jak najszerszym kontekście. W literaturze przedmiotu różni autorzy proponują odmienne definicje jakości informacji [Unold, 2011; English, 1999; Kahn i in., 2002].

Na podstawie charakterystyki katalogów stron WWW określono specyfikę możliwych problemów z jakością informacji. Przedstawione wady i zalety katalogów umożliwiły zidentyfikowanie zestawu dziesięciu cech jakości informacji: kompletność/wystarczalność, aktualność, spójność, wiarygodność, rzetelność, jednoznaczność, bezpieczeństwo informacji, zgodność strukturalna, dokładność i użyteczność [Krzesaj, 2016a, s. 189].

W kolejnym etapie wymienione cechy posłużyły do rozpoznania kryteriów oceny jakości informacji badanych obiektów. Do określenia kryteriów wykorzystano zestaw ogólnych metodyk badania jakości informacji medycznej dostępnej w Internecie, na podstawie których zidentyfikowano dziewięć kryteriów: referencje/kompetencje, ujawnienie informacji, aktualność informacji, atrybucja, dostępność, transparentność treści, polityka prywatności oraz ochrona danych, interaktywność i kryteria techniczne (nawigacja, linki, poprawność działania). Analizę doboru kryteriów przedstawiono w artykule pt. *Wybrane narzędzia i metody ewaluacji jakości informacji na stronach WWW* [Krzesaj, 2016b, s. 115–127].

Dalszy etap prac polegał na próbie przypisania rozpoznanym cechom jakości informacji wymienionych kryteriów. W tabeli 2 przedstawiono zestawienie rozpoznanых cech informacji oraz przypisanych im kryteriów oceny jakości informacji.

Zidentyfikowane cechy jakości informacji dokładność i spójność nie znalazły odzwierciedlenia w kryteriach, co wynika ze specyfiki badanych obiektów. Kolejnym etapem oceny jest określenie miar jakości informacji dla zaprezentowanych kryteriów.

Tabela 2. Zestawienie rozpoznanych cech informacji oraz przypisanych im kryteriów oceny jakości informacji

Lp.	Rozpoznana cecha jakości informacji	Kryteria oceny	Grupa kryteriów
1	Rzetelność	Referencje/ Kompetencje	Jakościowe
2	Wiarygodność	Ujawnienie informacji	
3	Aktualność	Aktualność informacji	
4	Wiarygodność	Atrybucja	
5	Kompletność/Wystarczalność	Dostępność	
6	Jednoznaczność	Transparentność treści	
7	Bezpieczeństwo	Polityka prywatności oraz ochrona danych	
8	Zgodność strukturalna	Kryteria techniczne	Poza jakościowe
9	Użyteczność	Interaktywność	

Źródło: opracowanie własne.

Do oceny jakości informacji wykorzystano badanie ankietowe. Zidentyfikowanych kryteriom przypisano jedenaście pytań badawczych z obszaru jakości informacji w katalogach stron WWW (tabela 3).

Tabela 3. Zestawienie kryteriów oceny jakości informacji oraz przypisanych im pytań badawczych

Lp.	Kryteria	Pytania badawcze
1	Referencje/ Kompetencje	1. Czy zawarto referencje/kwalifikacje recenzenta/właściciela danego katalogu stron WWW?
2	Ujawnienie informacji	2. Czy umieszczono informację o właścicielu katalogu stron WWW? 3. Czy katalog stron WWW ma jednoznacznie sformułowane cele? 4. Czy katalog stron WWW zawiera informacje o źródle jego finansowania?
3	Aktualność informacji	5. Czy znajdują się dane o aktualizacji informacji w katalogu?
4	Atrybucja	6. Czy katalog stron zawiera system ocen wpisywanych stron internetowych?
5	Dostępność	7. Czy katalog stron WWW zawiera metainformacje w opisie zbioru stron internetowych?
6	Transparentność treści	8. Czy umieszczono informacje o polityce publikacji materiałów?
7	Polityka prywatności oraz ochrona danych	9. Czy umieszczono informacje o polityce prywatności, reklamowej oraz ochronie danych osobowych?
8	Interaktywność	10. Czy katalog stron ma wbudowane mechanizmy pozwalające na interakcje z użytkownikiem?
9	Kryteria techniczne (nawigacja, linki, poprawność działania)	11. Czy zbiory linków do stron internetowych umieszczone są w sposób ustrukturalizowany?

Źródło: opracowanie własne.

Narzędzie badawcze w postaci kwestionariusza ankietowego zawiera dwie sekcje. Pierwsza sekcja składa się z jedenastu pytań szczegółowych (do oceny użyto pięciostopniowej skali Likerta). Każde z pytań dodatkowo zawiera objaśnienia w postaci pytań pomocniczych (wytycznych). Poniżej przedstawiono pytanie 3 wraz ze wskazówkami.

Czy katalog stron WWW ma jednoznacznie sformułowane cele?

Wskazówka – należy poszukać następujących informacji:

- Czy określono jakiego rodzaju informacje zawiera katalog stron?
- Dla kogo informacje umieszczone w katalogu są przeznaczone?

Wytyczne do oceny punktowej pytania:

- 5: Tak – katalog stron ma jasno postawione cele.
- 2–4: Częściowo – katalog stron ma postawione cele, ale są one niejasne lub niepełne.
- 1: Nie – katalog stron nie zawiera żadnych informacji dotyczących celów.

Druga sekcja zawiera jedno pytanie odnoszące się do ogólnej oceny danego katalogu. Ogólna ocena dokonywana jest na podstawie ocen uzyskanych z pytań szczegółowych w pierwszej sekcji.

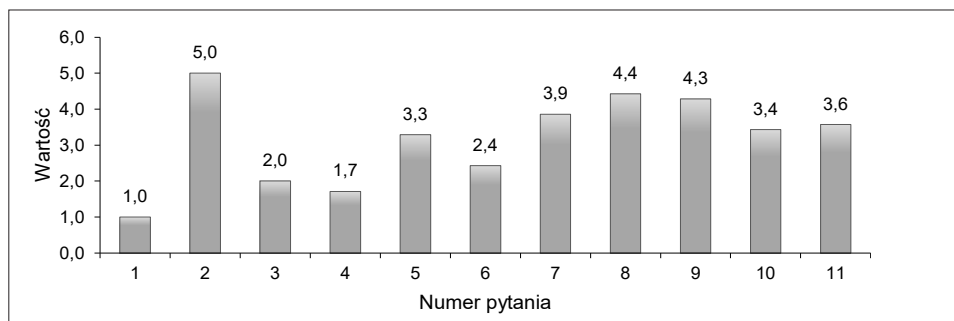
WYNIKI BADAŃ

Badanie pilotażowe przeprowadzono w czerwcu 2016 roku na siedmiu polskich katalogach stron WWW. Dobór próby dokonano na podstawie rankingu popularności katalogów z tabeli 1. Z dziesięciu najpopularniejszych katalogów wykluczono trzy: *Dmoz*, który nie jest polskim katalogiem oraz katalog *start24.pl* i *stronastartowa.pl* – nie stanowią zbioru stron lecz jedynie formę strony startowej. Na podstawie odpowiedzi na jedenaście pytań dokonano ogólnej oceny jakości informacji w badanych katalogach stron WWW, która kształtuje się w jednym z trzech poziomów:

- wysoka (5),
- umiarkowana (3),
- niska (1).

Trzy katalogi otrzymały ocenę wysoką (5): katalog.wp.pl, katalog.interia.pl i gery.pl. Ocena wynika z większości uzyskanych odpowiedzi na poziomie 4 lub 5. Ogólna wysoka ocena oznacza, że katalog stron jest dobrej jakości – stanowi użyteczne źródło informacji. Katalogi: katalog.onet.pl, katalog.hoga.pl i dodaj.pl otrzymały ogólną ocenę umiarkowaną (3). Uzyskana ogólna umiarkowana ocena oznacza, że katalog stron jest właściwej jakości – stanowi przydatne źródło informacji, ale posiada pewne ograniczenia. Uzyska ocena wynika z większości odpowiedzi na poziomie 3 lub w zbliżonej liczbie pytań na poziomie wysokim 5 i niskim 1. Na podstawie większości uzyskanych odpowiedzi na poziomie 2 lub 1 jedynie katalog prv.pl otrzymał ogólną niską ocenę (1). Ogól-

na niską ocenę oznacza, że katalog stron jest słabej jakości – posiada poważne braki i nie jest użyteczny. Rys. 1 zawiera średnie wartości odpowiedzi dla poszczególnych pytań szczegółowych.



Rys. 1. Katalogi stron WWW – średnia wartość odpowiedzi dla jedenastu pytań szczegółowych (kolejność pytań zgodnie z tabelą 3)

Źródło: opracowanie własne. Próba badawcza n=7.

Wszystkie przebadane katalogi stron WWW zawierały szczegółowe informacje o właścicielu strony, które były umieszczone w sposób przejrzysty i łatwy do odnalezienia (5,0). Na przeciwnym biegunie znalazły się odpowiedzi dotyczące kwalifikacji osób uprawnionych do publikacji informacji, które kształtowały się na bardzo niskim poziomie (1,0). Żaden z badanych katalogów stron nie zawiera takich informacji. Na niskim poziomie kształtują się informacje dotyczące celu katalogu stron i jego przeznaczenia (2,0) oraz umieszczonej w serwisie informacji o źródle jego finansowania (1,7). Na dość niskim poziomie kształtował się system ocen stron internetowych (2,4). Trzy z siedmiu katalogów nie posiadają takiej funkcjonalności, natomiast pozostałe zawierają taki system, lecz jego zasady funkcjonowania są niejasne, np. brak jest kryteriów, na podstawie których dokonano oceny lub opinie wystawiane są przez użytkowników katalogu, a nie osoby z odpowiednimi kwalifikacjami. W badanej próbie na umiarkowanym poziomie znajduje się kryterium odnoszące się do aktualizacji informacji w katalogu (3,3) oraz opisie informacyjnym zbioru stron internetowych (3,9). Dwa kryteria: wbudowane mechanizmy pozwalające na interakcje z użytkownikiem (3,4) oraz strukturalizowany sposób prezentacji zbiorów linków w katalogu (3,6) kształtują się na bardzo zbliżonym poziomie. Dość wysoką ocenę uzyskały informacje dotyczące zasad publikacji materiałów w badanych katalogach (4,4). Większość z ich zawiera regulaminy napisane w sposób jednoznaczny i precyzyjny. Zbliżoną ocenę uzyskało kryterium odnoszące się do polityki prywatności oraz ochrony danych osobowych (4,3). W przypadku katalogów stron zawierających reklamy ocenie podlegała informacja o polityce dotyczącej reklamy.

WNIOSKI

Podjęta w artykule próba oceny jakości informacji w katalogach stron WWW wydaje się ważna zarówno z praktycznego, jak i naukowego punktu widzenia. Utworzone narzędzie oceny jakości katalogów stron WWW rekomendowane jest głównie dla użytkowników katalogów stron WWW choć może być używane przez właścicieli katalogów (np. w celu badania jakości informacji w konkurencyjnych katalogach). W wymiarze naukowym stworzona metodologia wpisuje się w nurt badań nad rozwojem pomiaru jakości informacji w sieci Internet.

Przeprowadzone badanie pilotażowe dało odpowiedź na zasadnicze pytanie: czy udostępniane przez katalogi informacje posiadają cechy dobrej jakości takie jak, np.: aktualność, spójność, kompletność, czy wiarygodność? Trzy katalogi otrzymały ogólną ocenę wysoką (5), kolejne trzy ocenę umiarkowaną (3), a jeden otrzymał ocenę niską (1). Ogólna ocena jakości nie jest adekwatna co do ich popularności odwiedzin. Według danych zaprezentowanych w tabeli 1 *katalog.onet.pl* zajmuje drugie miejsce w rankingu z 29,26% udziałem, a w przeprowadzonym badaniu otrzymał ogólną ocenę umiarkowaną (3). Katalog *gery.pl* z 1,04% udziałem zajął w rankingu siódme miejsce, a uzyskał ocenę wysoką (5). Przedstawione rozbieżności między ogólną oceną jakości a pozycją w rankingu mogą wynikać z różnych przesłanek, lecz ich identyfikacja wymaga dalszych pogłębionych badań. W celu potwierdzenia uzyskanych wyników w kolejnym etapie badań należy zwiększyć próbę badawczą katalogów stron WWW.

BIBLIOGRAFIA

- English L., 1999, *Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits*, John Wiley & Sons Inc., New York.
- Kahn B.K., Strong D.M., Wang R.Y., 2002, *Information Quality Benchmarks: Product and Service Performance*, Communications of the ACM, Vol. 45, No 4, <http://dx.doi.org/10.1145/505248.506007>.
- Krzesaj M., 2016a, *Information quality evaluation methodology for web page directories*, Information Systems In Management, Vol. 5(2), SGGW, Warszawa.
- Krzesaj M., 2016b, *Wybrane narzędzia i metody ewaluacji jakości informacji na stronach WWW* [w:] *Informacja – dobro publiczne czy prywatne?*, red. A. Czerwiński, A. Jańdziak, M. Krzesaj, Wydawnictwo Uniwersytetu Opolskiego, Opole.
- Maciejowski T., 2004, *Firma w Internecie*, Oficyna Ekonomiczna, Kraków.
- Raport GUS, 2015, *Spółeczeństwo informacyjne w Polsce. Wyniki badań statystycznych z lat 2011-2015*, <http://stat.gov.pl/obszary-tematyczne/nauka-i-technika-spoleczenstwo-informacyjne/spoleczenstwo-informacyjne/> (dostęp: 22.06.2016 r.).
- Ruževičius J., Gedminaitė A., 2007, *Business Information Quality and its Assessment*, „Engineering Economics”, nr 2 (52).
- Unold J., 2011, *Teoretyczno-metodologiczne podstawy przetwarzania informacji w cyberprzestrzeni*, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław.

Streszczenie

Artykuł prezentuje ocenę jakości informacji w wybranych polskich katalogach stron WWW. W pierwszej części artykułu zaprezentowano charakterystykę, klasyfikację i ranking popularności katalogów stron. Następnie przedstawiono metodykę badania oceny jakości informacji w katalogach stron WWW. Do pomiaru jakości informacji zaproponowano kwestionariusz ankietowy. W drugiej części artykułu przedstawiono wyniki ewaluacji jakości informacji w wybranych katalogach stron WWW. Należy podkreślić, że wyniki badań są ograniczone ze względu na niewielką reprezentatywność próby. Artykuł kończy podsumowanie zawierające wnioski oraz wskazówki dotyczące dalszych kierunków badań.

Słowa kluczowe: jakość informacji, katalogi stron internetowych, ocena jakości informacji

Information quality evaluation in the selected Polish web page directories*Summary*

The article presents information quality evaluation in the selected polish web page directories. In the first part of the paper present: characteristics, classification and popularity ranking web page directories. Subsequently was shown the methodology of information quality assessment in the web page directories. To measure the information quality was proposed survey questionnaire. In the second part of the article presents the results of evaluating the quality of information in the selected directories. It should be emphasized that the results are limited due to the small representativeness of the sample. The article ends with a summary containing conclusions and guidelines for further research directions.

Keywords: information quality, quality in the web page directories, information quality evaluation

JEL: D82, D83, L86