

# Annales Universitatis Paedagogicae Cracoviensis

Studia ad Bibliothecarum Scientiam Pertinentia 21 (2023)

ISSN 2081-1861

DOI 10.24917/20811861.21.27

**Krzysztof Wieniecki**

Uniwersytet Mikołaja Kopernika w Toruniu

ORCID 0000-0002-9794-2736

## Europejskie inicjatywy archiwizacji Internetu – przegląd i ocena benchmarkingowa wybranych serwisów WWW

### Wstęp

Instytucje pamięci (biblioteki, archiwa i galerie) znane są od wieków i do dziś pełnią rozmaite funkcje (m.in. informacyjną, kulturową, społeczną, artystyczną). W przypadku tradycyjnych źródeł informacji wykształciły one skuteczne metody archiwizowania, magazynowania oraz katalogowania. Postęp technologiczny oraz praktycznie nieograniczony dostęp do informacji cyfrowej sprawiają jednak, że coraz częściej odczuwalna jest potrzeba wypracowania podobnych metod dla zasobów cyfrowych, w tym stron internetowych. Powszechnie wiadomo, że strony internetowe są z różnych powodów modyfikowane lub nawet usuwane z przestrzeni sieciowej. Dzieje się tak za sprawą umyślnego działania, aktualizacji treści, zakończenia projektu, zmian technologicznych, rozwoju sieci lub zaniedbania administratora strony. Tymczasem pierwotne wersje stron WWW mogą stanowić przydatny materiał badawczy (np. w procesach badania zmieniającej się na przestrzeni lat architektury informacji). W tym kontekście, coraz częściej postulowana, archiwizacja Internetu wydaje się dobrą strategią zachowania treści internetowych oraz udostępniania ich na określone potrzeby użytkowników.

Problem archiwizacji zasobów internetowych od wielu lat jest przedmiotem rozważań środowisk naukowych, archiwistów – praktyków i informatyków. Tematykę tę na gruncie polskim podejmowali w swoich pracach m.in. Lidia Derfert-Wolf, która opisała genezę i wyzwania stojące przed internetową archiwizacją<sup>1</sup>; Grzegorz Gmiterek, który przybliżył istotę procesów archiwizacyjnych<sup>2</sup>, Bartłomiej Konopa<sup>3</sup>, który scharakteryzował najstarsze europejskie inicjatywy archiwizujące Web, Aneta Januszko-Szakiel, która przedstawiła propozycję programu

---

1 L. Derfert-Wolf, *Archiwizacja Internetu – wniosku i rekomendacje z kilku raportów*, „Biuletyn EBIB” 2017, nr 172, s. 1–11.

2 G. Gmiterek, *Długoterminowa archiwizacja zasobów cyfrowych*, „Res Historica” 2013, nr 35, s. 213–235.

3 B. Konopa, *Witryna internetowa jako materiał archiwalny*, „Archiwa – Kancelarie – Zbiory” 2016, nr 7, s. 113–120.

ochrony zasobów cyfrowych zgromadzonych w polskich instytucjach bibliotecznych<sup>4</sup>, Marcin Wilkowski, który jest inicjatorem Archiwum Społecznego Polskiego Webu<sup>5</sup>, oraz Wojciech Woźniak, który zaprezentował dotychczasowy stan archiwizacji Internetu w kontekście rozważań pojęciowych związanych z rozumieniem terminu „archiwizacja Internetu”<sup>6</sup>. Głównym wnioskiem formułowanym przez badaczy był postulat utworzenia projektu, który zabezpieczyłby wszystkie dostępne treści w Internecie, jednocześnie magazynując je w niezmiennym postaci i udostępniając zainteresowanym internautom w ramach ruchu społecznego i wolnego dostępu do informacji<sup>7</sup>.

Potrzeba stworzenia takiego rozwiązania cały czas powraca, a wskutek zdarzeń geopolitycznych związanych z konfliktami, zagrożeniem do swobodnego dostępu do informacji i zalewem dezinformacji jest o tyle pożądana, co trudna do rzeczywistej realizacji. Potwierdzeniem tezy o użyteczności archiwizacji, choćby w obszarze utrwalania tożsamości narodowej i kulturowej, jest inicjatywa podjęta przez Ukrainian Research Institute w marcu 2022 r. W związku z agresją wojsk rosyjskich na terytorium Ukrainy przeprowadzono archiwizację kolekcji zasobów ukraińskich stron internetowych, gromadząc w serwisie Archive-it ich pokaźną liczbę<sup>8</sup>.

Celem niniejszego artykułu jest prezentacja 14 europejskich archiwów Internetu oraz ich ocena pod względem cech treściowych i funkcjonalnych. Autor ma nadzieję, że omówienie wybranych inicjatyw będzie stanowić inspirację w procesie uruchamiania polskiego archiwum Internetu, a wyłonienie mocnych i słabych strony badanych serwisów pozwoli uniknąć podobnych błędów podczas jego projektowania.

## Strategie archiwizowania Webu

Analizując rolę i sens utrwalania treści stron internetowych, należy w pierwszej kolejności zdefiniować termin „archiwizacja Internetu”, wskazać rodzaje tego procesu oraz elementy związane ze sposobami zbierania danych, ich przetwarzania, przechowywania oraz udostępniania.

Według przewodnika archiwizacji treści internetowych *Web Archiving Guidance*, wydanego przez Narodowe Archiwum Zjednoczonego Królestwa Brytyjskiego, *archiwizacja Webu to*

proces pobierania ze środowiska wirtualnego stron internetowych i zawartych w nich informacji oraz magazynowania ich w odpowiednim archiwum [tłum. autora].

---

4 A. Januszko-Szakiel, *Długoterminowa archiwizacja zasobów cyfrowych – program dla polskich bibliotek*, „Przegląd Biblioteczny” 2011, nr 79, s. 211–230.

5 Archiwum Społeczne Polskiego Webu, Informacje, 2023, [on-line:] <https://aspw.pl/informacje> – 27.09.2023.

6 W. Woźniak, *Archiwizacja Internetu – próba podsumowania dotychczasowych prac i ustaleń*, „Archiwa – Kancelarie – Zbiory” 2019, nr 10, s. 75–98.

7 L. Derfert-Wolf, dz. cyt., s. 1–11.

8 Ukrainian Research Institute, *Russia’s war on Ukraine, 2022*, [on-line:] <https://archive-it.org/collections/18886> – 27.09.2023.

Z przywołanej definicji wynika, że archiwizacja treści internetowych jest podobna do tradycyjnej archiwizacji dokumentów papierowych lub drukowanych, w której informacje są wybierane, przechowywane, zabezpieczane i udostępniane zainteresowanej społeczności<sup>9</sup>. Internet zawiera ogromne zasoby informacji, które uniemożliwiają ręczne przechwytywanie i archiwizowanie danych. Archiwiści Webu do kolekcjonowania zasobów cyfrowych zazwyczaj wykorzystują zautomatyzowane procesy ich zbierania. Ta aktywność wykonywana jest przez roboty indeksujące, które przeszukują określoną wirtualną strefę, kopiują i zapisują informacje podczas pobierania wersji stron internetowych. Zarchiwizowane strony internetowe i zawarte w nich informacje są następnie udostępniane w archiwum dla użytkowników, którzy nie tylko mają możliwość podglądu strony internetowej, ale również swobodnego przemieszczania się po niej oraz manipulowania interfejsem poprzez wybieranie podstron i wyświetlanie lub pobieranie zawartych na stronie materiałów. W inicjatywę archiwizacji treści internetowych najczęściej zaangażowane są biblioteki narodowe, archiwa krajowe oraz różne grupy, które archiwizują treści w sposób szczegółowy, mając na względzie utrwalenie dziedzictwa kulturowego (takie grupy reprezentuje m.in. mniejszość baskijska oraz mniejszość katalońska)<sup>10</sup>.

Strategie archiwizowania zasobów sieciowych można traktować jako „proces poszukiwania, gromadzenia i organizacji źródeł informacji w celu zabezpieczenia ich przed zanikiem w WWW”<sup>11</sup>. Pod terminem „strategia” kryje się sposób, w jaki podmioty zajmujące się archiwizacją Webu będą posługiwać się podczas tworzenia wirtualnej kolekcji. Wybór strategii determinuje późniejszy kształt oraz znaczenie informacyjne zarchiwizowanych zasobów<sup>12</sup>. Janne Nielsen w opracowaniu *Using Web Archives in Research* przywołuje kilka rodzajów strategii wykorzystywanych w procesie archiwizacji:

- *broad crawl* (szerokie przeszukiwanie zasobów),
- *national/regional domain crawl* (gromadzenie zasobów z uwzględnieniem kryterium narodowego lub regionalnego),
- *selective harvesting* (gromadzenie wybranych zasobów),
- *event harvesting* (gromadzenie zasobów związanych z określonym wydarzeniem),
- *thematic harvesting* (gromadzenie zasobów dotyczących określonej tematyki)<sup>13</sup>.

Wiele archiwów nie tylko korzysta z wymienionych strategii, ale stosuje podobne określenia do opisu procesu archiwizacji zasobów. Ta wielość terminów i ich podobieństwo sprawiają, że wspomniane strategie można w zasadzie sprowadzić do dwóch rodzajów gromadzenia zasobów, tj. **gromadzenia masowego** (ang. *bulk harvesting*), które dotyczy strategii *broad crawl*, oraz **gromadzenia selektywnego**

9 The National Archives, *Web Archiving Guidance*, 2011, [on-line:] <https://cdn.nationalarchives.gov.uk/documents/information-management/web-archiving-guidance.pdf> – 27.09.2023.

10 B. Konopa, *Archiwizacja Webu w Europie*, „Archeion” 2020, nr 121, s. 1–21.

11 M. Jankowska, *Biblioteki akademickie – trendy dotyczące zasobów elektronicznych*, [w:] *Informacja dla nauki a świat zasobów cyfrowych*, red. H. Ganińska, Poznań 2008, s. 168.

12 B. Konopa, *Strategia selektywna jako narzędzie w archiwizacji Webu. Analiza wybranych przykładów*, „Archiwa – Kancelarie – Zbiory” 2020, nr 11, s. 97–118.

13 J. Nielsen, *Using web archives in research: an introduction*, NetLab, Aarhus 2016, s. 15–25.

(ang. *selective harvesting*), które obejmuje strategię *national/regional domain crawl*, *selective harvesting*, *event harvesting* oraz *thematic harvesting*<sup>14</sup>. Zaproponowany podział będzie stanowić punkt odniesienia do określania strategii gromadzenia danych przez badane w dalszej części artykułu europejskie inicjatywy archiwizowania Internetu.

Gromadzenie masowe opiera się na archiwizowaniu konkretnej domeny lub grupy domen, jak ma to miejsce w przypadku narodowej archiwizacji stron internetowych. Z uwagi na prostotę metody oraz możliwości wykorzystania technologii metoda ta jest bardzo zautomatyzowana w procesach i obsłudze. Wykorzystywanie robotów indeksujących sprowadza udział człowieka w kolekcjonowaniu treści do kontrolera wbudowanego oprogramowania. Zaletą stosowania strategii gromadzenia masowego jest zbudowanie dużego zbioru kolekcji stron internetowych przy stosunkowo małym nakładzie pracy. Jednakże taka praktyka rodzi również mankamenty w postaci niekompletności zbioru, co jest wynikiem niemożliwości ręcznego sprawdzenia zgromadzonych elementów nowej kolekcji zasobów internetowych<sup>15</sup>.

Strategia gromadzenia selektywnego polega na ręcznym typowaniu określonych stron internetowych oraz innych obiektów cyfrowych, które następnie poddaje się procesowi archiwizacji. Wyboru dokonuje się na podstawie wymienionych kryteriów (ze względu na rodzaje domen, tematykę, wydarzenie lub indywidualną selekcję). Zaletą takiego rozwiązania jest z pewnością dokładność kontroli nad treściami oraz ich precyzyjne wytypowanie do powstającej kolekcji. Ponadto zgromadzone elementy są łatwiejsze do opisanie metadanymi i innymi elementami etykietowania zasobów i ich opisywania. Wadą selektywnej strategii gromadzenia danych jest większe zaangażowanie człowieka jako głównego decydenta do gromadzenia, weryfikacji, opisu, archiwizacji i udostępniania każdego znalezionej i odpowiadającym kryteriom zasobu<sup>16</sup>.

### Internet Archive – pionierska inicjatywa archiwizowania zasobów WWW

Najstarszą i do dziś największą inicjatywą archiwizacji zasobów internetowych jest założone w 1996 roku Internet Archive<sup>17</sup>, a ściślej – należące do tego przedsięwzięcia – narzędzie Wayback Machine<sup>18</sup>. Założycielem serwisu jest amerykański bibliotekarz Brewster Kahle, a jego ideą gromadzenie oraz udostępnianie zasobów internetowych różnego typu. Kolekcje archiwum obejmują nie tylko strony internetowe (sekcja *Web*), ale również pliki graficzne (*Images*), pliki audio (*Audio*), filmy (*Video*), książki (*Books*) oraz różnego typu oprogramowania (*Software*). Aktualnie (stan na 2 października 2023 r.) zbiory Internet Archive liczą:

- a) 840 miliardów różnych wersji stron internetowych,
- b) 38 milionów książek lub tekstów,
- c) 15 milionów nagrań audio (w tym ponad 240 tysięcy nagrań koncertów na żywo),

---

14 B. Konopa, *Strategia selektywna...*, s. 97–113.

15 Tamże.

16 Tamże.

17 Internet Archive, *Introduction*, 2023, [on-line:] <https://archive.org> – 27.09.2023.

18 Tamże.

- d) ponad 11 milionów filmów (w tym ponad 2,4 miliona nagrań programów telewizyjnych),
- e) 4,7 miliona plików graficznych,
- f) ponad milion udostępnionych oprogramowań<sup>19</sup>.

Za symboliczny początek rozpoczęcia idei archiwizowania Internetu można uznać artykuł Brewstera Kahle pt. *Preserving the Internet*<sup>20</sup>, w którym opisał on idee oraz misję stworzenia Internet Archive w następujący sposób:

Większość społeczeństw przywiązuje dużą wagę do przechowywania śladów swojego dziedzictwa kulturowego. Bez takich śladów cywilizacja nie ma pamięci i mechanizmu, który pozwala uczyć się na dawnych błędach oraz powodzeniach. Współczesny świat produkuje dużo informacji w postaci elektronicznej. Naszym obowiązkiem jest próba zachowania tych danych i stworzenia biblioteki internetowej dla naukowców, badaczy i wszystkich zainteresowanych użytkowników Internetu<sup>21</sup>.

W następstwie powstania Internet Archive w 1996 r. powołano do życia inne inicjatywy, takie jak: Australia's Web Archive, Tasmanian Web Archive i szwedzkie archiwum Kulturarw<sup>22</sup>. Jednakże pomimo stosunkowo długiej historii ich istnienia, zasoby amerykańskiego Internet Archive są zdecydowanie większe, a jego kolekcje cieszą się większą popularnością. W 2004 roku powstała organizacja European Archive (późniejsza Internet Memory Foundation), której jednym z celów było utworzenie wielkiego europejskiego archiwum internetowego we współpracy z Internet Archive<sup>23</sup>. Niestety pomysł nie został zrealizowany. W tym samym czasie zaczęły powstawać jednak oddzielne archiwa europejskie, które przechowywały zbiory internetowe dla danego kraju lub regionu, wykorzystując przy tym narzędzia, procesy i strategie kojarzone z istniejącymi od dłuższego czasu archiwami, a zwłaszcza z rozwiązaniami Internet Archive.

Podstawowym narzędziem dostępnym w Internet Archive jest Wayback Machine. To internetowy magazyn zasobów World Wide Web – głównie stron internetowych oraz ich dodatkowych elementów, takich jak: pliki graficzne, odnośniki, materiały do pobrania na stronie oraz kody źródłowe<sup>24</sup>. Nazwa narzędzia odnosi się bezpośrednio do amerykańskiej kreskówki, w której główni bohaterowie zaprojektowali maszynę do podróżowania w czasie<sup>25</sup>. Wayback Machine jako rozwiązanie do utrwalenia zasobów internetowych rozwija się od 1996 roku. Zasoby, które gromadzi archiwum, są pozyskiwane za pomocą metody crawlingu, co oznacza, że

19 Internet Archive, *About, 2023*, [on-line:] <https://archive.org> – 27.09.2023.

20 B. Kahle, *Preserving the Internet*, "Scientific American" 1997, nr 276, s. 82–83.

21 P. Koncewicz, *Ludzie którzy przechowują Internet, czyli Wayback Machine w pigułce*, 2022, [on-line:] <https://antyweb.pl/internet-archive-wayback-machine-w-pigulce> – 27.09.2023.

22 M. Konopka-Nowina, *Problematyka zarządzania informacją w procesie archiwizacji zasobów*, „Media, Kultura, Komunikacja Społeczna” 2014, nr 10, s. 30–43.

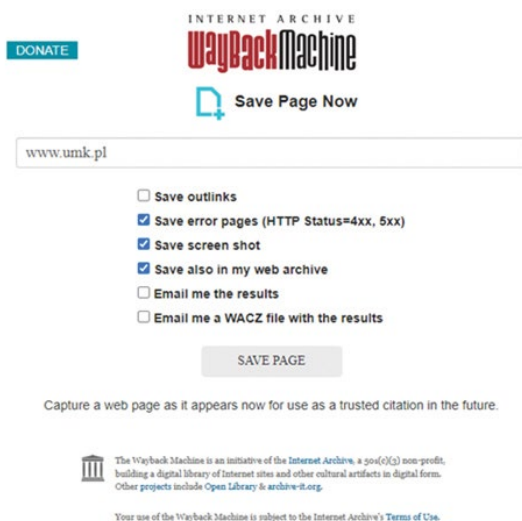
23 L. Derfert-Wolf, dz. cyt., s. 1–11.

24 M. Wilkowski, *Wayback Machine – podstawy wykorzystania*, „Biuletyn EBIB” 2017 nr 172, s. 1–3.

25 K. Gmerek, *Archiwa internetowe po obu stronach Atlantyku – Internet Archive, Wayback Machine oraz UK Web Archive*, „Biuletyn EBIB” 2012, nr 128, s. 1–12.

magazynowanie obiektów dotyczy jedynie publicznie dostępnych obiektów internetowych, które dodatkowo nie mogą być zablokowane przez indeksowanie odczytywania ich za pomocą wpisów do robots.txt<sup>26</sup>. Do 2023 roku strona zgromadziła ponad 818 miliardów archiwalnych wersji stron internetowych. Proces archiwizacji strony internetowej może zostać zapoczątkowany przez wolontariuszy pracujących dla organizacji Internet Archive oraz każdego użytkownika archiwum, który uzyska dostęp do narzędzia poprzez rejestrację konta. Proces archiwizowania stron internetowych za pomocą Wayback Machine może być również stosowany automatycznie z uwzględnieniem odpowiednich ram czasowych. Zapoczątkowanie procesu archiwizacyjnego dla zasobu WWW wiąże się z odpowiednimi technikami oraz opcjami, które użytkownik zobowiązany jest zaznaczyć ręcznie (zob. rys. 1). Wśród zaproponowanych opcji możliwe jest zachowanie wszelkich linków zewnętrznych, wszelkich błędów, graficznego interfejsu strony oraz opcjonalnie umieszczenie kopii archiwalnej wersji strony w wirtualnej kolekcji użytkownika, a następnie wysłanie potwierdzenia w postaci wiadomości e-mail.

Rysunek 1. Wayback Machine – wprowadzenie strony internetowej do zarchiwizowania



Źródło: *Wayback Machine*, [on-line:] <https://archive.org/web/> – 4.07.2023.

Proces archiwizacji strony internetowej może trwać od kilku do kilkudziesięciu minut w zależności od ilości zasobów, dostępnych w zaproponowanej do archiwizowania stronie WWW. Podstawowymi trudnościami w archiwizowaniu stron poprzez Wayback Machine są błędy odnoszące się bezpośrednio do crawlerów, niezapisujących się materiałów zewnętrznych na stronach internetowych oraz błędnie indeksowane daty archiwizacji stron internetowych.

---

26 M. Wilkowski, dz. cyt., s. 2–3.

## Rysunek 2. Wayback Machine – wynik archiwizacji strony i napotkane błędy

INTERNET ARCHIVE  
wayBackMachine

DONATE

Saving page <http://www.umk.pl> ✔ Done!

A snapshot was captured. Visit page: </web/20230704083122/https://www.umk.pl/>  
A screen shot was captured. View screen shot  
</web/20230704083122/http://web.archive.org/screenshot/http://www.umk.pl/>

<https://www.umk.pl/>  
<https://www.umk.pl/grafika/IDUB.png>  
<https://www.umk.pl/grafika/hr.png>  
<https://www.umk.pl/css/style.css?n=1688459482>  
<https://www.umk.pl/grafika/logo-umk.png>  
<https://www.umk.pl/grafika/logo-Yerun3.png>  
<https://portal.umk.pl/uploads/5d47829edfed2.jpg>  
<https://portal.umk.pl/uploads/nagrodyumkfofmikolajkuras-20.jpg>  
<https://portal.umk.pl/uploads/kolawisniewskastrategia-portal.jpg>  
<https://portal.umk.pl/uploads/badaniaKopernikanskiefilm-portal2.jpg>

If something goes wrong please click [here](#) to send us an error report. Downloaded elements: 36

**Saving outlinks and their embedded resources**

<http://bip.umk.pl/> Error! Please try again in -46 min. Crawling this host is paused because they notified us that they are overloaded right now.  
</web/20230704083139/https://glos.umk.pl/> ✔ Done!  
<https://absolwent.umk.pl/zwiazak-sw-jakuba>  
Error! Please try again in -43 min. Crawling this host is paused because they notified us that they are overloaded right now.  
</web/20230704083139/https://apcz.umk.pl/> ✔ Done!  
<https://idub.umk.pl/> Error! Please try again in -34 min. Crawling this host is paused because they notified us that they are overloaded right now.  
</web/20230704083136/https://kopernik550.umk.pl/> ✔ Done!  
<https://pl-pl.facebook.com/UniersytetMikolajKopernika>  
Error! Please try again in -41 min. Crawling this host is paused because they notified us that they are overloaded right now.

Źródło: Wayback Machine, [on-line:] <https://archive.org/web> – 4.07.2023.

Internet Archive jako organizacja non-profit często angażuje się w akcje, które mają charakter społeczny oraz charytatywny. Ponieważ wiodącą misją serwisu jest utrwalenie dziedzictwa cyfrowego i zapewnienie każdemu otwartego i wolnego dostępu do informacji, toteż aktywności o charakterze społecznym dotyczą tematyki związanej z równym dostępem do informacji, ruchem Open Source oraz zachowaniem wolności i swobody wymiany danych w Internecie.

Jedną z inicjatyw, która realizuje misję otwartości, jest WayForward Machine, który powstał z okazji 25-lecia funkcjonowania Internet Archive<sup>27</sup>. WayForward Machine to wyszukiwarka mająca zastosowanie przeciwne do Wayback Machine, bowiem ma pokazywać przyszłość zasobów Internetu (zamiast zasobów zarchiwizowanych)<sup>28</sup>. Po wpisaniu do wyszukiwarki odpowiedniej strony internetowej użytkownik otrzymuje chaotyczne treści reklamowe, informacje na temat wolności przepływu informacji w sferze internetowej oraz szereg komunikatów związanych z zablokowaniem sieci oraz oszustwami informacyjnymi. Projekt ma pełnić rolę uświadamiającą powagę misji zachowania wolności Internetu oraz być przestrożą dla użytkowników w kwestii podawania zbyt wielu informacji osobistych obcym podmiotom. Wy tłumaczenie zasadności powstania takiej inicjatywy znajduje się na specjalnej osi czasu, zaprojektowanej na 25-lecie funkcjonowania Internet Archive, która opisuje negatywną perspektywę następnych 25 lat funkcjonowania Internetu (lata 2021–2046)<sup>29</sup>. Oś czasu obejmuje hipotetyczne niepożądane zdarzenia, które

27 Internet Archive, *Anniversary*, 2021, [on-line:] <https://anniversary.archive.org> – 27.09.2023.

28 Internet Archive, *WayForward Machine*, 2021, [on-line:] <https://wayforward.archive.org> – 27.09.2023.

29 Internet Archive, *IA 2046*, 2021, [on-line:] <https://wayforward.archive.org/ia2046> – 27.09.2023.

w przyszłości mogą dotknąć społeczeństwo informacyjne, m.in. uchwalenie restrykcyjnych ustaw związanych z prawem autorskim, powstanie globalnego monopolisty informacyjnego, dostęp do informacji jedynie za opłatą w postaci pieniędzy lub w wyniku sprzedaży własnych danych osobowych, likwidacja bibliotek publicznych, wprowadzenie do szkół podręczników wydawanych przez monopolistę informacyjnego, które w swojej zawartości posiadają głównie materiały reklamowe i propagandowe, szereg zmian klimatycznych, które będą skutkowały migracją na szeroką skalę, zniszczenie ostatniej kopii publikacji *Rok 1984* George'a Orwella, likwidacja tradycyjnych materiałów czytelnicych oraz bunt społeczeństwa informacyjnego, którego marzeniem będzie powrót do wolnego Internetu i swobodnego dostępu do informacji<sup>30</sup>.

Inną inicjatywą prowadzoną przez Internet Archive jest blog, na którym ukazują się informacje o nowych projektach, zarchiwizowanych zasobach oraz wykorzystaniu zasobów archiwalnych przez różne podmioty<sup>31</sup>. Internet Archive oferuje również możliwość wolontariatu obejmującego bibliotekarzy, naukowców, informatyków oraz osoby, które chcą być bezpośrednio zaangażowane w utrwalanie cyfrowych dóbr kulturowych oraz ochronę informacji. Wolontariat wiąże się z wyszukiwaniem informacji, programowaniem rozwiązań pozyskiwania informacji, katalogowaniem zasobów oraz ich indeksowaniem, a także organizowaniem inicjatyw i spotkań<sup>32</sup>.

Internet Archive jest również inicjatorem i organizatorem szeregu bezpośrednich inicjatyw, mających związek z bibliotekami, funkcjonowaniem społeczeństwa informacyjnego oraz promowaniem wartości edukacyjnych. Przykładami takich projektów są: Political TV Ad Archive, Building Libraries Together, Open Library, Scanning Services, Offline Archive, Bookmobile, BookServer, Petabox oraz Open Community Networks<sup>33</sup>.

### **Badania wybranych europejskich inicjatyw archiwizowania Internetu: założenia, cele i metody**

Przedmiotem analiz w niniejszym artykule są wybrane europejskie archiwa Internetu. W ich wyborze posłużono się zestawieniem pochodzących z artykułu *Archiwizacja Webu w Europie – narodowe archiwa Sieci* autorstwa Bartłomieja Konopy. Pod uwagę wzięto jednak wyłącznie te inicjatywy, które funkcjonują co najmniej 10 lat oraz cieszą się rozbudowaną kolekcją zasobów. Ich wykaz zawiera tabela 1.

W badaniach serwisów wykorzystano metodę analizy zawartości stron internetowych oraz benchmarking. Analiza zawartości stron internetowych (zwana również metodą heurystyczną) polega na ocenie cech jakościowych serwisów z uwzględnieniem określonych kryteriów treściowych i funkcjonalnych. Z kolei benchmarking jest ekspercką metodą badawczą w swej naturze podobną do metody heurystycznej. Zakłada on działania, których celem jest osiągnięcie przewagi nad

---

30 Tamże.

31 Internet Archive, *Blog*, 2023, [on-line:] <https://blog.archive.org> – 27.09.2023.

32 Internet Archive, *Volunteers*, 2023, [on-line:] <https://archive.org/about/volunteer-positions.php> – 27.09.2023.

33 Internet Archive, *Projects*, 2023, [on-line:] <https://archive.org/projects> – 27.09.2023.



konkurencją<sup>34</sup>. Celowi temu służy dokonywanie porównań usług i produktów, które znajdują się w ofercie podmiotów przodujących w określonej branży<sup>35</sup>. Ocena stron WWW metodą benchmarkingu, podobnie jak metoda jakościowo-heurystyczna, obejmuje stworzenie przez osobę oceniającą odpowiedniego zbioru kryteriów, według których badany serwis będzie oceniany. Badania te służą do wskazania lidera spośród badanych serwisów internetowych, toteż do prawidłowego przebiegu badań dobiera się przynajmniej dwa serwisy, które w następnym etapie są oceniane pod względem treściowym i funkcjonalnym<sup>36</sup>.

Tabela 1. Archiwa Internetu wybrane do badań

Lp.	Nazwa archiwum	Obszar działalności	Rok założenia	Badana strona internetowa
1.	Archives de l'Internet	Francja	2002	<a href="https://www.bnf.fr/fr/archives-de-linternet">https://www.bnf.fr/fr/archives-de-linternet</a>
2.	Archivo de la Web Española	Hiszpania	2009	<a href="http://www.dl-e.es/openwayback/wayback/index.jsp">http://www.dl-e.es/openwayback/wayback/index.jsp</a>
3.	Arquivo PT	Portugalia	2008	<a href="http://www.arquivo.pt/?l=en">www.arquivo.pt/?l=en</a>
4.	Eesti veebiarhiivi	Estonia	2010	<a href="https://veebiarhiiv.digar.ee">https://veebiarhiiv.digar.ee</a>
5.	Hrvatski arhiv weba	Chorwacja	2004	<a href="https://haw.nsk.hr/en">https://haw.nsk.hr/en</a>
6.	Netarkivet	Dania	2005	<a href="http://www.kb.dk/find-materiale/samlinger/netarkivet">www.kb.dk/find-materiale/samlinger/netarkivet</a>
7.	Ondarenet	Hiszpania (Kraj Basków)	2007	<a href="http://www.ondarenet.kultura.ejgv.euskadi.eus:8085/ondarenet">www.ondarenet.kultura.ejgv.euskadi.eus:8085/ondarenet</a>
8.	PADICAT	Hiszpania (Katalonia)	2005	<a href="https://padicat.cat">https://padicat.cat</a>
9.	Suomalainen verkkoarkisto	Finlandia	2006	<a href="https://verkkoarkisto.kansalliskirjasto.fi/va">https://verkkoarkisto.kansalliskirjasto.fi/va</a>
10.	UK WebArchive	Wielka Brytania	2004	<a href="https://www.webarchive.org.uk/en/ukwa">https://www.webarchive.org.uk/en/ukwa</a>
11.	Vefsafn IS	Islandia	2004	<a href="https://vefsafn.is/is">https://vefsafn.is/is</a>
12.	Webarchief van Nederland	Holandia	2007	<a href="https://rijksoverheid.sitearchief.nl/#archive">https://rijksoverheid.sitearchief.nl/#archive</a>
13.	WebArchiv	Czechy	2000	<a href="https://webarchiv.cz/en">https://webarchiv.cz/en</a>
14.	WebArchive	Irlandia	2011	<a href="https://archive-it.org/home/nli">https://archive-it.org/home/nli</a>

Źródło: opracowanie własne.

Nie istnieje jeden ściśle określony katalog kryteriów, które należy przyjąć, oceniając strony WWW. W publikacji *Benchmarking w doskonaleniu serwisów WWW bibliotek akademickich* Remigiusz Sapa do oceny polskich i amerykańskich stron internetowych wśród kryteriów treściowych zaproponował informacje o bibliotece,

34 R. Sapa, *Benchmarking w doskonaleniu serwisów WWW bibliotek akademickich*, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków 2005, s. 37–38.

35 E. Głowacka, *Studium kompleksowego zarządzania jakością (TQM) w bibliotekoznawstwie i informacji naukowej*, Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, Toruń 2000, s. 100.

36 R. Sapa, *Benchmarking w doskonaleniu...*, s. 47–49.

zasoby elektroniczne dostępne wyłącznie dla uprawnionych użytkowników, zasoby elektroniczne dostępne publicznie, usługi informacyjne, przestrzeń komunikacji, obsługę interesantów<sup>37</sup>. Z kolei do cech funkcjonalnych zaliczył przejrzystość, komunikatywność, spójność i konsekwencję, prewencyjność, elastyczność, aktualność, narzędzia nawigacji, sprawność, sekwencyjność<sup>38</sup>.

Autor niniejszego artykułu, inspirując się badaniami R. Sapy, również zdecydował się na ocenę kryteriów formalnych i funkcjonalnych. Pod uwagę zostały wzięte następujące kryteria oceny:

1. Cechy odnoszące się do treści serwisów:
  - a) informacje o archiwum – informacje związane z identyfikacją archiwum, takie jak: informacje ogólne, kalendarium, osie czasu oraz charakterystyka zasobów,
  - b) zasoby archiwum – wielkość gromadzonych zasobów, tematyka archiwizowanych materiałów i ich ogólna jakość,
  - c) kompletność oraz dokładność informacji – poprawność odnośników, źródeł oraz metadanych zasobów,
  - d) komunikacja z użytkownikiem – możliwości i formy kontaktu na linii archiwum – użytkownik, z uwzględnieniem kontaktu wirtualnego i tradycyjnego.
2. Cechy odnoszące się do funkcjonalności serwisów:
  - a) design i struktura serwisu – subiektywne odczucia podczas przeglądania strony, jej architektury informacji oraz szaty graficznej,
  - b) dostępność serwisu – dostosowanie stron internetowych dla osób z niepełnosprawnościami,
  - c) narzędzia nawigacji – obecność narzędzi nawigacyjnych, łatwość ich obsługi oraz elementów związanych z zaawansowanym wyszukiwaniem informacji,
  - d) aktualność serwisu – częstotliwość zmian oraz wprowadzania dokumentów do archiwum.

Każda z cech punktowana była w skali 0–3. Najniższa ocena 0 oznaczała brak występowania danej cechy lub zdecydowane trudności w jej odnalezieniu. Ocena 3 była oceną najwyższą i określała obecność wskazanych elementów na stronie internetowej, zgodnych z kryteriami i ich charakterystyką. Każde archiwum cyfrowe mogło otrzymać maksymalnie 60 punktów za daną cechę, a łącznie 480 punktów za wszystkie kryteria. Zebrane dane umożliwiły określenie stopnia natężenia poszczególnych cech<sup>39</sup>. Materiał badawczy posłużył do sformułowania wniosków końcowych, wskazując mocne strony i słabości serwisów internetowych.

Badania 14 europejskich inicjatyw archiwizacji Internetu przeprowadzono na przełomie maja oraz czerwca 2023 roku. W badaniu poza autorem niniejszego artykułu uczestniczyło 20 studentów drugiego roku studiów uzupełniających kierunku architektura informacji Uniwersytetu Mikołaja Kopernika w Toruniu,

---

37 R. Sapa, dz. cyt., s. 151–158.

38 Tamże, s. 159–160.

39 Natężenie cechy – stosunek ocen uzyskanych przez wszystkie serwisy za daną cechę do sumy maksymalnych ocen przyznawanych za tę cechę, wyrażony w procentach. Tamże, s. 55.

którzy wcześniej zaliczyli 15-godzinny kurs z zakresu archiwizacji Internetu (moduł wzbogacający na tym kierunku). Grupa oceniająca była zatem zaznajomiona ze zjawiskiem archiwizowania Webu, a jej umiejętności oceniania stron internetowych – z uwagi na studiowany kierunek – cechowały się wysokim poziomem zaawansowania.

## Wyniki badań

Rezultaty badania benchmarkingowego zaprezentowano w formie tabel w kolejności alfabetycznej nazw internetowych archiwów. Punktacja dla jednej cechy mieści się w skali od 0 do 60 punktów, a ocena całego serwisu internetowego w skali od 0 do 480 punktów.

Tabela 2. Wyniki badań archiwum francuskiego

Nazwa archiwum	Archives de l' Internet		
	Punktacja	Średnia ocen użytkowników	Natężenie cechy
<b>Kryteria treściowe</b>			
informacje o archiwum	41	2,05	68,33%
zasoby archiwum	38	1,9	63,33%
kompletność i dokładność informacji	33	1,65	55%
komunikacja z użytkownikiem	36	1,8	60%
Suma:	148/240	-	-
<b>Kryteria funkcjonalne</b>			
design i struktura serwisu	33	1,65	55%
dostępność serwisu	28	1,4	46,67%
narzędzia nawigacji	31	1,55	51,67%
aktualność serwisu	44	2,2	73,33%
Suma:	136/240	-	-
Całkowita punktacja:	284/480 (59,1%)	-	-

Źródło: opracowanie własne.

Tabela 3. Wyniki badań archiwum hiszpańskiego

Nazwa archiwum	Archivo de la Web Española		
	Punktacja	Średnia ocen użytkowników	Natężenie cechy
<b>Kryteria treściowe</b>			
informacje o archiwum	21	1,05	35%
zasoby archiwum	12	0,6	20%
kompletność i dokładność informacji	13	0,65	21,67%
komunikacja z użytkownikiem	7	0,35	11,66%
Suma:	53/240	-	-

Nazwa archiwum	Archivo de la Web Española		
	Punktacja	Średnia ocen użytkowników	Natężenie cechy
<b>Kryteria funkcjonalne</b>			
design i struktura serwisu	7	0,35	11,66%
dostępność serwisu	5	0,25	8,33%
narzędzia nawigacji	13	0,65	21,67%
aktualność serwisu	14	0,7	23,33%
Suma:	39/240	-	-
Całkowita punktacja:	92/480 (19,17%)	-	-

Źródło: opracowanie własne.

Tabela 4. Wyniki badań archiwum portugalskiego

Nazwa archiwum	Arquivo PT		
	Punktacja	Średnia ocen użytkowników	Natężenie cechy
<b>Kryteria treściowe</b>			
informacje o archiwum	44	2,2	73,33%
zasoby archiwum	41	2,05	68,33%
kompletność i dokładność informacji	39	1,95	65%
komunikacja z użytkownikiem	30	1,5	50%
Suma:	154/240	-	-
<b>Kryteria funkcjonalne</b>			
design i struktura serwisu	36	1,8	60%
dostępność serwisu	22	1,1	36,67%
narzędzia nawigacji	40	2	66,67%
aktualność serwisu	41	2,05	68,33%
Suma:	139/240	-	-
Całkowita punktacja:	293/480 (61,04%)	-	-

Źródło: opracowanie własne.

Tabela 5. Wyniki badań archiwum estońskiego

Nazwa archiwum	Eesti veebiarhiivi		
	Punktacja	Średnia ocen użytkowników	Natężenie cechy
<b>Kryteria treściowe</b>			
informacje o archiwum	12	0,6	20%
zasoby archiwum	28	1,4	46,67%
kompletność i dokładność informacji	26	1,3	43,33%
komunikacja z użytkownikiem	6	0,3	10%
Suma:	72/240	-	-

Nazwa archiwum	Eesti veebiarhiivi		
	Punktacja	Średnia ocen użytkowników	Natężenie cechy
<b>Kryteria funkcjonalne</b>			
design i struktura serwisu	8	0,4	13,33%
dostępność serwisu	9	0,45	15%
narzędzia nawigacji	16	0,8	26,67%
aktualność serwisu	26	1,3	43,33%
Suma:	59/240	-	-
Całkowita punktacja:	131/480 (27,29%)	-	-

Źródło: opracowanie własne.

Tabela 6. Wyniki badań archiwum chorwackiego

Nazwa archiwum	Hrvatski arhiv weba		
	Punktacja	Średnia ocen użytkowników	Natężenie cechy
<b>Kryteria treściowe</b>			
informacje o archiwum	47	2,35	78,33%
zasoby archiwum	34	1,7	56,67%
kompletność i dokładność informacji	32	1,6	53,33%
komunikacja z użytkownikiem	38	1,9	63,33%
Suma:	151/240	-	-
<b>Kryteria funkcjonalne</b>			
design i struktura serwisu	36	1,8	60%
dostępność serwisu	38	1,9	63,33%
narzędzia nawigacji	44	2,2	73,33%
aktualność serwisu	34	1,7	56,67%
Suma:	152/240	-	-
Całkowita punktacja:	303/480 (63,13%)	-	-

Źródło: opracowanie własne.

Tabela 7. Wyniki badań archiwum duńskiego

Nazwa archiwum	Netarkivet		
	Punktacja	Średnia ocen użytkowników	Natężenie cechy
<b>Kryteria treściowe</b>			
informacje o archiwum	38	1,9	63,33%
zasoby archiwum	37	1,85	61,67%
kompletność i dokładność informacji	36	1,8	60%
komunikacja z użytkownikiem	37	1,85	61,67%
Suma:	148/240	-	-

Nazwa archiwum	Netarkivet		
	Punktacja	Średnia ocen użytkowników	Natężenie cechy
<b>Kryteria funkcjonalne</b>			
design i struktura serwisu	32	1,6	53,33%
dostępność serwisu	21	1,05	35%
narzędzia nawigacji	35	1,75	58,33%
aktualność serwisu	33	1,65	55%
Suma:	121/240	-	-
Całkowita punktacja:	269/480 (56,04%)	-	-

Źródło: opracowanie własne.

Tabela 8. Wyniki badań archiwum baskijskiego

Nazwa archiwum	Ondarenet		
	Punktacja	Średnia ocen użytkowników	Natężenie cechy
<b>Kryteria treściowe</b>			
<b>informacje o archiwum</b>	<b>25</b>	<b>1,25</b>	<b>41,67%</b>
zasoby archiwum	30	1,5	50%
kompletność i dokładność informacji	19	0,95	31,67%
komunikacja z użytkownikiem	11	0,55	18,33%
Suma:	85/240	-	-
<b>Kryteria funkcjonalne</b>			
design i struktura serwisu	4	0,2	6,67%
dostępność serwisu	9	0,45	15%
narzędzia nawigacji	25	1,25	41,67%
aktualność serwisu	21	1,05	35%
Suma:	59/240	-	-
Całkowita punktacja:	144/480 (30%)	-	-

Źródło: opracowanie własne.

Tabela 9. Wyniki badań archiwum katalońskiego

Nazwa archiwum	PADICAT		
	Punktacja	Średnia ocen użytkowników	Natężenie cechy
<b>Kryteria treściowe</b>			
informacje o archiwum	35	1,75	58,33%
zasoby archiwum	31	1,55	51,67%
kompletność i dokładność informacji	24	1,2	40%
komunikacja z użytkownikiem	28	1,4	46,67%
Suma:	118/240	-	-

Nazwa archiwum	PADICAT		
	Punktacja	Średnia ocen użytkowników	Natężenie cechy
<b>Kryteria funkcjonalne</b>			
design i struktura serwisu	10	0,5	16,67%
dostępność serwisu	21	1,05	35%
narzędzia nawigacji	26	1,3	43,33%
aktualność serwisu	16	0,8	26,67%
Suma:	73/240	-	-
Całkowita punktacja:	191/480 (39,79%)	-	-

Źródło: opracowanie własne.

Tabela 10. Wyniki badań archiwum fińskiego

Nazwa archiwum	Suomalainen verkkoarkisto		
	Punktacja	Średnia ocen użytkowników	Natężenie cechy
<b>Kryteria treściowe</b>			
informacje o archiwum	39	1,95	65%
zasoby archiwum	33	1,65	55%
kompletność i dokładność informacji	29	1,45	48,33%
komunikacja z użytkownikiem	24	1,2	40%
Suma:	125/240	-	-
<b>Kryteria funkcjonalne</b>			
design i struktura serwisu	21	1,05	35%
dostępność serwisu	19	0,95	31,67%
narzędzia nawigacji	24	1,2	40%
aktualność serwisu	33	1,65	55%
Suma:	97/240	-	-
Całkowita punktacja:	222/480 (46,25%)	-	-

Źródło: opracowanie własne.

Tabela 11. Wyniki badań archiwum brytyjskiego

Nazwa archiwum	0,5		
	Punktacja	Średnia ocen użytkowników	Natężenie cechy
<b>Kryteria treściowe</b>			
informacje o archiwum	48	1,95	80%
zasoby archiwum	35	1,65	58,33%
kompletność i dokładność informacji	32	1,45	53,33%
komunikacja z użytkownikiem	37	1,2	61,67%
Suma:	152/240	-	-

Nazwa archiwum	0,5		
	Punktacja	Średnia ocen użytkowników	Natężenie cechy
<b>Kryteria funkcjonalne</b>			
design i struktura serwisu	27	1,05	45%
dostępność serwisu	36	0,95	60%
narzędzia nawigacji	37	1,2	61,67%
aktualność serwisu	23	1,65	38,33%
Suma:	123/240	-	-
Całkowita punktacja:	275/480 (57,29%)	-	-

Źródło: opracowanie własne.

Tabela 12. Wyniki badań archiwum islandzkiego

Nazwa archiwum	Vefsafn IS		
	Punktacja	Średnia ocen użytkowników	Natężenie cechy
<b>Kryteria treściowe</b>			
informacje o archiwum	32	1,6	53,33%
zasoby archiwum	27	1,35	45%
kompletność i dokładność informacji	20	1	33,33%
komunikacja z użytkownikiem	26	1,3	43,33%
Suma:	105/240	-	-
<b>Kryteria funkcjonalne</b>			
design i struktura serwisu	31	1,55	51,67%
dostępność serwisu	23	1,15	38,33%
narzędzia nawigacji	32	1,6	53,33%
aktualność serwisu	17	0,85	28,33%
Suma:	103/240	-	-
Całkowita punktacja:	208/480 (43,33%)	-	-

Źródło: opracowanie własne.

Tabela 13. Wyniki badań archiwum holenderskiego

Nazwa archiwum	Webarchief van Nederland		
	Punktacja	Średnia ocen użytkowników	Natężenie cechy
<b>Kryteria treściowe</b>			
informacje o archiwum	19	0,95	31,67%
zasoby archiwum	38	1,9	58,33%
kompletność i dokładność informacji	23	1,15	38,33%
komunikacja z użytkownikiem	25	1,25	41,67%
Suma:	105/240	-	-



Nazwa archiwum	Webarchief van Nederland		
	Punktacja	Średnia ocen użytkowników	Natężenie cechy
<b>Kryteria funkcjonalne</b>			
design i struktura serwisu	24	1,2	40%
dostępność serwisu	18	0,9	30%
narzędzia nawigacji	26	1,3	43,33%
aktualność serwisu	44	2,2	73,33%
Suma:	112/240	-	-
Całkowita punktacja:	217/480 (45,21%)	-	-

Źródło: opracowanie własne.

Tabela 14. Wyniki badań archiwum czeskiego

Nazwa archiwum	WebArchiv		
	Punktacja	Średnia ocen użytkowników	Natężenie cechy
<b>Kryteria treściowe</b>			
informacje o archiwum	42	2,1	70%
zasoby archiwum	31	1,55	51,67%
kompletność i dokładność informacji	31	1,55	51,67%
komunikacja z użytkownikiem	33	1,65	55%
Suma:	137/240	-	-
<b>Kryteria funkcjonalne</b>			
design i struktura serwisu	5	0,25	8,33%
dostępność serwisu	20	1	33,33%
narzędzia nawigacji	30	1,5	50%
aktualność serwisu	45	2,25	75%
Suma:	100/240	-	-
Całkowita punktacja:	237/480 (49,30%)	-	-

Źródło: opracowanie własne.

Tabela 15. Wyniki badań archiwum irlandzkiego

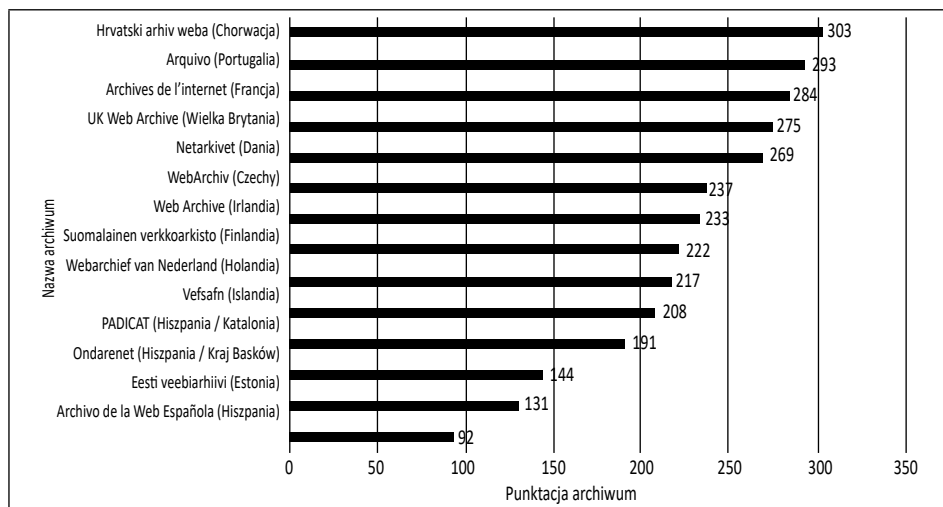
Nazwa archiwum	Web Archive		
	Punktacja	Średnia ocen użytkowników	Natężenie cechy
<b>Kryteria treściowe</b>			
informacje o archiwum	35	1,75	58,33%
zasoby archiwum	39	1,95	65%
kompletność i dokładność informacji	33	1,65	55%
komunikacja z użytkownikiem	31	1,55	51,67%
Suma:	138/240	-	-

Nazwa archiwum	Web Archive		
	Punktacja	Średnia ocen użytkowników	Natężenie cechy
<b>Kryteria funkcjonalne</b>			
design i struktura serwisu	21	1,05	35%
dostępność serwisu	11	0,55	18,33%
narzędzia nawigacji	27	1,35	45%
aktualność serwisu	36	1,8	60%
Suma:	95/240	-	-
Całkowita punktacja:	233/480 (48,54%)	-	-

Źródło: opracowanie własne.

Na podstawie danych przedstawionych w tabelach można wskazać najwyżej oraz najniżej ocenione serwisy internetowe pod względem zaproponowanych kryteriów. Serwisy, które zostały ocenione najwyżej, uzyskując najwyższą liczbę punktów, a tym samym najwyższy wskaźnik procentowy, to strony WWW: chorwackiego Hrvatski arhiv weba (63,13%), portugalskiego Arquivo PT (61,04%) oraz francuskiego Archives de l'internet (59,10%). Najniżej ocenione strony internetowe, które nie osiągnęły powyżej 30% wszystkich ocenianych cech należały do: hiszpańskiego Archivo de la Web Española (19,17%), estońskiego Eesti veebiarhiivi (27,29%) oraz baskijskiego Ondarenet (30%). Ranking serwisów internetowych z uwzględnieniem sumarycznej liczby punktów prezentuje rys. 3. Warto zwrócić uwagę, że rozbieżność pomiędzy najwyżej ocenioną stroną internetową a najniżej ocenionym serwisem wynosi aż 211 punktów (w 480 punktowej skali).

Rysunek 3. Ranking badanych archiwów internetowych z uwzględnieniem sumarycznej liczby punktów za cechy treściowe i funkcjonalne

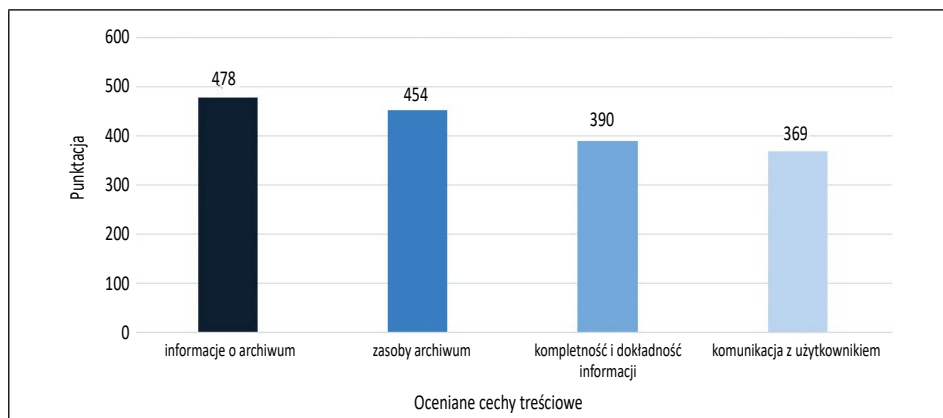


Źródło: opracowanie własne.

Zidentyfikowane w trakcie badań problemy poszczególnych stron internetowych wynikały często nie z barier technologicznych, ale z podstawowych zaniedbań na etapie projektowania i tworzenia strony internetowej. Przykładem strony internetowej, która nie uplasowała się najwyżej w rankingu mimo zdobycia ponad 40 punktów w niektórych cechach, jest czeski serwis WebArchiv. Z racji, iż serwis ten jest najstarszy, otrzymał tylko pięć punktów w kryterium „design i struktura serwisu”. Gdyby wprowadzić minimalne zmiany, które są niejako wymuszone przez rozwijający się sektor stron internetowych, strona mogłaby uzyskać uśrednioną notę na poziomie 30 punktów w tym konkretnym kryterium, co skutkowałoby tym, iż awansowałaby o jedną pozycję, przekraczając barierę zgodności serwisu na poziomie około 54%.

Na rys. 4 i 5 zaprezentowano punktację, jaką serwisy otrzymały za cechy treściowe i funkcjonalne.

Rysunek 4. Sumaryczna punktacja przyznana wszystkim ocenianym archiwom za cechy treściowe



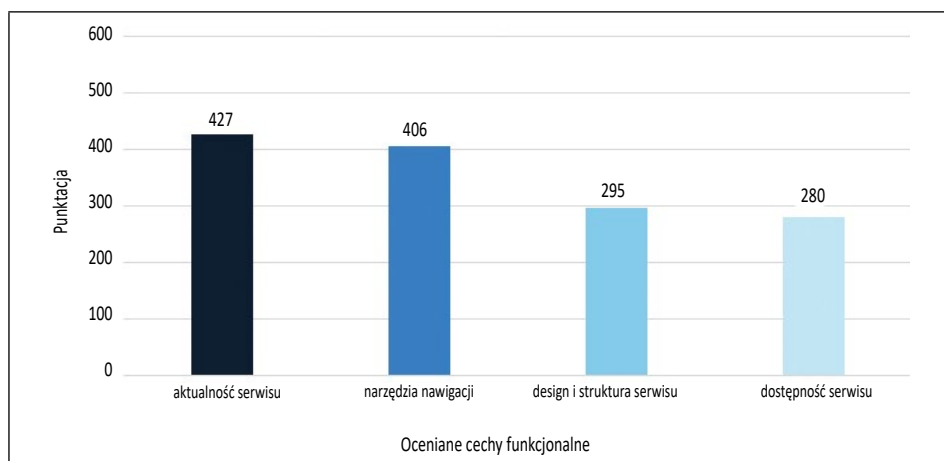
Źródło: opracowanie własne.

Jak ilustruje rys. 4, różnica między najwyżej i najniżej ocenioną cechą treściową wyniosła jedynie 109 punktów (w 840 punktowej skali), co oznacza, że użytkownicy ocenili serwisy w stopniu zbliżonym. Z uwagi na łatwość w znalezieniu informacji najwyżej ocenianą cechą okazały się „informacje o archiwum”, które były również najwyżej ocenianą cechą w zestawieniu wszystkich ośmiu kryteriów badania. Najniżej ocenianą cechą wśród kryteriów treściowych była „komunikacja z użytkownikiem”, co świadczy o problemach użytkowników z dotarciem do tego typu informacji podczas procesu wyszukiwania lub po prostu o braku jej występowania na stronach archiwów.

Zestawiając kryteria funkcjonalności z treściowymi, można zauważyć minimalną rozbieżność ocen pomiędzy najwyżej i najniżej ocenianym kryterium – wynosi ona 147 punktów (na 840). Szczególną uwagę należy poświęcić cechom „dostępność serwisów” oraz „design i struktura serwisów”. Te dwie cechy zostały ocenione najniżej i jako jedyne z ośmiu wszystkich kryteriów nie przekroczyły 300 punktów. Ich niskie oceny mogą wynikać z „wiekowości” serwisów – najstarsza strona internetowa powstała w 2000 roku, a najmłodsza w 2011. W czasach ich powstawania

trendy w tworzeniu stron internetowych były zdecydowanie inne i nie wymuszały na administratorze ciągłych zmian. Biorąc po uwagę fakt, że „aktualność serwisów” została oceniona na 454 punkty, należy zasugerować administratorom ocenianych serwisów wprowadzenie pewnych zmian i w tym obszarze. Największą słabością stron WWW europejskich inicjatyw archiwizacji Internetu okazała się „dostępność serwisu”. Problem ten może być spowodowany – podobnie jak w przypadku „designu i struktury strony” – „wiekowością” serwisów, ale również charakterem podmiotów, które tymi serwisami zarządzają. Należy jednak mieć nadzieję, że popularyzacja archiwizowania treści w Internecie niejako wymusi na administratorach stron internetowych wdrożenie odpowiednich rozwiązań, by strona internetowa ich przedsięwzięcia spełniała określone standardy dla użytkowników.

Rysunek 5. Sumaryczna punktacja przyznana wszystkim ocenianym archiwom za kryteria funkcjonalne



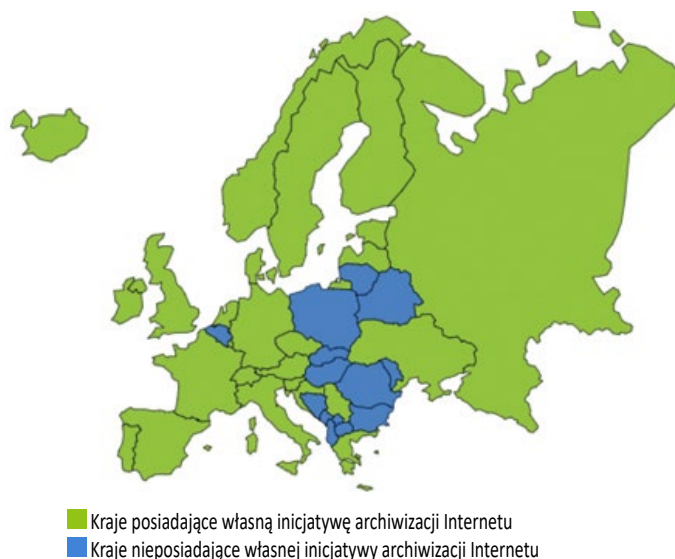
Źródło: opracowanie własne.

Przeprowadzone badania wykazały, że w przypadku tworzenia nowej strony internetowej dla archiwum Internetu w danym państwie należy zwrócić szczególną uwagę na kwestie związane z funkcjonalnością stron. Aspekty związane z ogólną strukturą strony, jej szatą graficzną oraz dostosowaniem do potrzeb grup osób, które mogą być wykluczone cyfrowo, powinny być głównym założeniem podczas procesu planowania i tworzenia strony internetowej. Ważną kwestią powinno być również odpowiednie umiejscowienie na stronie WWW danych kontaktowych, by w ten sposób ułatwić użytkownikom m.in. zgłaszanie sugestii dotyczących funkcjonowania danego archiwum. Ponieważ nisko ocenione elementy można wykluczyć poprzez odpowiednie planowanie strony internetowej oraz rozwiązania projektowe, dobrym rozwiązaniem może tu być głębsza współpraca archiwistów z administratorami oraz architektami informacji. Cieszy natomiast fakt, że jakość zasobów, informacje o podmiotach archiwizujących Web oraz narzędzia nawigacyjne zostały ocenione dość wysoko.

## Internetowa archiwizacja w Polsce

Biorąc pod uwagę zasięg funkcjonowania archiwów Internetu w Europie, można zaryzykować stwierdzenie, że Polska jest największym państwem na Starym Kontynencie, które nie posiada własnej inicjatywy archiwizacji zasobów internetowych.

Rysunek 5. Podział Europy z uwzględnieniem krajów posiadających i nieposiadających archiwów Internetu



Źródło: opracowanie własne.

Potrzeba powstania polskiej inicjatywy archiwizowania Webu jest przedmiotem licznych dyskusji podejmowanych przez środowiska naukowe, zwłaszcza związane z gospodarką oraz archiwistyką<sup>40</sup>. W dniach 16–17 listopada 2021 roku Archiwa Państwowe zorganizowały „Forum archiwizacji zasobów polskiego Internetu”, które stało się miejscem dyskusji i wymiany poglądów na temat przyszłości zachowania cyfrowego dziedzictwa narodowego w Polsce. Uczestnikami tej konferencji byli przedstawiciele europejskich archiwów WWW, Archiwów Państwowych, reprezentanci ośrodków akademickich oraz administracji publicznej<sup>41</sup>.

W literaturze przedmiotu pojawiają się także propozycje modeli archiwizowania treści internetowych. Dla przykładu, Sebastian Gaca proponuje zastosowanie modelu brytyjskiego w tworzeniu polskiej inicjatywy archiwizacji Webu, zakładając, iż taka inicjatywa powinna być inicjatywą państwową, za którą – zgodnie z modelem brytyjskim – stałoby Ministerstwo Kultury i Dziedzictwa Narodowego we współpracy z Biblioteką Narodową i Narodowym Archiwum Cyfrowym. W modelu tym

40 W gospodarce, *Archiwizacja Internetu to realny problem*, 2018, [on-line:] <https://wgospodarce.pl/informacje/44931-archiwizacja-internetu-to-realny-problem> – 27.09.2023.

41 Archiwa Państwowe, *Forum archiwizacji zasobów polskiego Internetu*, 2021, [on-line:] <https://archiwa.gov.pl/forum-archiwizacji-zasobow-polskiego-internetu> – 27.09.2023.

resort kultury byłby odpowiedzialny za ustalenie strategii powstania archiwum, zaś archiwizacja zasobów przebiegałaby z wykorzystaniem infrastruktury Narodowego Archiwum Cyfrowego, mediów społecznościowych oraz zasobów Internet Archive. Archiwizacja miałaby odbywać się na zasadzie podziału na materiały urzędowe i nieurzędowe z otwartym dostępem dla użytkowników<sup>42</sup>.

W listopadzie 2020 roku uruchomiono inicjatywę Archiwum Społeczne Polskiego Webu (ASPW)<sup>43</sup>, której celem jest indeksowanie polskiej domeny krajowej oraz wspieranie badań historycznych nad polskim Webem. Archiwum udostępnia do wolnego wykorzystania zbiory skategoryzowanych adresów URL z odpowiednimi metadanymi. Kopie archiwalnych zindeksowanych stron można znaleźć za pomocą interfejsu projektu MementoWeb<sup>44</sup>. Inicjatorem przedsięwzięcia jest wspomniany w artykule badacz internetowej archiwizacji Marcin Wilkowski.

Warty odnotowania jest fakt, iż w polskiej przestrzeni cyfrowej istnieje strona internetowa HistoriaStron.pl, która gromadzi zasoby archiwalnych wersji stron internetowych, jednakże wyświetla je na podstawie zgromadzonych danych przez Internet Archive oraz Google Cache. Strona ta wyświetla zgromadzone wcześniej archiwalne wersje stron internetowych z wyżej wymienionych źródeł oraz aktualną wersję danej strony WWW. Ponadto na stronie znajduje się informacja o charakterze edukacyjnym, która krótko przedstawia definicje archiwalnej wersji strony WWW oraz podaje zalety prowadzenia internetowego archiwum<sup>45</sup>.

Internet Archive w swoich zbiorach gromadzi również polskie strony internetowe oraz kolekcje cyfrowe z pozostałych archiwów w języku polskim. Przykładem polskich treści są materiały cyfrowe załączane do numerów czasopisma „CyberMycha”, które ukazywało się na rynku polskim w latach 2000–2008<sup>46</sup>. W zasobach Internet Archive można również odnaleźć nagrania programów informacyjnych transmitowanych przez Telewizję Polską<sup>47</sup>.

### Potencjalne obszary wykorzystania zasobów archiwów stron internetowych

Zasoby archiwalne, gromadzone w postaci tradycyjnej i cyfrowej w archiwach mogą być cennym źródłem wiedzy oraz pożądanym materiałem edukacyjnym. Działalność muzeów, archiwów i bibliotek dowodzi, iż gromadzenie materiałów i ich udostępnianie pełni bardzo ważną społeczną rolę. Wirtualne zasoby archiwalne, które dotyczą głównie sfery internetowej, mogą również odegrać dużą rolę we współczesnym świecie, stając się cennym materiałem badawczym, informacyjnym i edukacyjnym.

---

42 S. Gaca, *Propozycja modelu archiwizacji polskiego Internetu na bazie doświadczeń brytyjskich*, „Officina Historiae” 2021, t. 4, nr 1, s. 79–90.

43 Archiwum Społeczne Polskiego Webu, *Informacje*, 2023, [on-line:] <https://aspw.pl/informacje> – 27.09.2023.

44 MementoWeb, *Guide*, 2023, [on-line:] <https://timetravel.mementoweb.org/guide/api> – 27.09.2023.

45 Historia Stron, *O projekcie*, 2023, [on-line:] <http://historiastron.pl> – 27.09.2023.

46 Creator: CyberMycha, *Personal collection*, 2023, [on-line:] <https://archive.org/search?query=creator%3A%22CyberMycha%22> – 27.09.2023.

47 Creator: Telewizja Polska, *Personal collection*, 2023, [on-line:] <https://archive.org/details/telewizja-polska> – 27.09.2023.

Zdaniem autora niniejszego artykułu archiwalne zasoby internetowe mogą znaleźć swoje zastosowanie w trzech zasadniczych obszarach, takich jak:

- a) działalność informacyjna – wyszukiwanie informacji, infobrokerstwo oraz ekologia informacji,
- b) utrwalanie dziedzictwa narodowego i kulturowego,
- c) działalność naukowa.

Istnienie archiwalnych wersji stron internetowych w momencie, kiedy domena wygasła lub strona została zamknięta i nie posiada aktualnej wersji, pozostaje jedynym śladem oraz cennym materiałem informacyjnym do ponownego wykorzystania. Zasoby archiwalnych wersji stron internetowych mogą zatem być nowym i unikalnym źródłem pozyskiwania informacji, klasyfikowanym tuż obok rejestrów publicznych, repozytoriów cyfrowych, katalogów firm oraz baz danych. Ze względu na specyficzny charakter treści archiwalnych wersji stron internetowych mogą one służyć do weryfikacji informacji oraz wskazywać zmiany zachodzące w ich obrębie w określonych ramach czasowych. Takie działanie może być wykorzystywane do odpowiedniej selekcji informacji pod kątem rzetelności i aktualności.

Archiwa Internetu mogą pełnić w przyszłości funkcje nowoczesnych muzeów oraz wypierać rolę tradycyjnych archiwów. Zasoby wymienionych w artykule archiwów są często obiektem aktywności społecznej w wielu aspektach (np. obywatelskim, rozrywkowym, politycznym lub artystycznym). Aktywność użytkowników i praca z zasobami archiwalnymi determinuje je jako element dziedzictwa narodowego i kulturowego. Z uwagi na dynamikę sytuacji geopolitycznej archiwizacja Internetu może być uniwersalnym sposobem na ratowanie dziedzictwa kulturowego w postaci cyfrowej. Przykładem takiego działania jest rozwiązanie, zaproponowane przez Środowisko naukowe w zakresie ratowania danych cyfrowych projektu Saving Ukrainian Cultural Heritage Online (SUCHO)<sup>48</sup>, w wyniku czego powstało narzędzie, które umożliwia kopiowanie stron internetowych oraz całej ich zawartości, zapisywanej na rzeczowych urządzeniach pamięci masowej. Projekt ten może objąć wszelkie zasoby instytucji dziedzictwa kulturowego (archiwa, biblioteki, galerie sztuki oraz muzea)<sup>49</sup>. Archiwizacja Internetu może pełnić ważną funkcję obywatelską z powodu społecznej roli oprogramowania, które jest podstawą podmiotowości w sieci. Otwartość i kontrola nad oprogramowaniem umożliwia komunikację, publikowanie lub cenzurowanie informacji wraz z jej filtrowaniem i selekcjonowaniem<sup>50</sup>. Efekt społeczny archiwizacji Webu może być podyktowany umożliwieniem wolnego dostępu do cyfrowego dziedzictwa oraz zapewnieniem bezpieczeństwa zasobów cyfrowych, które z biegiem czasu mogą stać się cennym materiałem historycznym<sup>51</sup>.

Archiwizacja Internetu może również odgrywać znaczącą rolę w aktywności artystycznej. Archiwum stanowi rodzaj formy kulturowej, która od dawna jest

---

48 J. Adamus-Kowalska, *Zastosowanie technologii archiwizacji Internetu w identyfikacji i ratowaniu cyfrowego dziedzictwa kulturowego Ukrainy*, „Annales Universitatis Paedagogicae Cracoviensis Studia ad Bibliothecarum Scientiam Pertinentia” 2022, nr 20, s. 596–611.

49 Q. Dombrowski, *SUCHO Project*, 2023, [on-line:] <https://www.sucho.org> – 27.09.2023.

50 M. Wilkowski, *Oddolne archiwizacje Internetu jako działania społeczne*, „Archiwa – Kancelarie – Zbiory” 2015, nr 6, s. 207–220.

51 Tamże.

przedmiotem zainteresowania sztuki<sup>52</sup>. Wszelkie kolekcje w cyfrowych archiwach mogą pełnić rolę wirtualnych eksponatów, a internetowa forma udostępniania dzieł sztuki może promować artystów i poszerzać grono ich odbiorców. Potwierdzają to zasoby Internet Archive, które obejmują autorskie prace graficzne, fotografie, filmy, teksty oraz pliki audio, wpisujące się bez wątpienia w działalność artystyczną.

## Podsumowanie

Archiwa internetowe zabezpieczają strony internetowe i umożliwiają dostęp do nich, bez względu na utratę głównego źródła. Dla wielu naukowców pozostają one nieznanne. Istnieje jednak kilka charakterystycznych cech archiwów internetowych, które mogą stać się przesłanką do wykorzystywania ich jako narzędzi badawczych, a mianowicie:

1. Archiwa sieciowe zawierają treści, które mogą zniknąć z bieżącej sieci.
2. W aktywności naukowej powszechne jest cytowanie stron internetowych. Naukowcy często odnoszą się do pierwotnych źródeł, które znikają lub ich adresy URL są zmieniane, i w konsekwencji dostęp do nich jest utrudniony, a w niektórych przypadkach niemożliwy.
3. Historia sieci ilustruje również znaczącą część kultury. Okresowe przechwytywanie stron internetowych pokazuje ich zmiany, dostosowanie pod nowe trendy oraz umożliwia ich badanie pod kątem np. architektury informacji.
4. Może nastąpić sytuacja, że niektórych dokumentów nie będzie można oddzielić od narzędzi i platform, przy pomocy których je wygenerowano. Ujednolicenie wyświetlania treści przez archiwa ułatwi ich ponowne odtworzenie.
5. Internetowe archiwa oferują głównie tematyczne kolekcje danych, które mogą być analizowane za pomocą nowych metod masowej eksploracji danych. Daje to zupełnie nowe możliwości badań dla naukowców i stanowi cenny materiał informacyjny<sup>53</sup>.

Zdaniem autora artykułu zasoby Internet Archive i innych wymienionych w artykule archiwów można z powodzeniem wykorzystać w celu promowania działalności naukowej (np. konferencji naukowych, publikacji naukowych), co może przyczynić się do szybkiej wymiany wyników badań, a w konsekwencji rozwoju nauki<sup>54</sup>.

W Polsce narasta potrzeba utworzenia oddzielnego archiwum Webu, które może powstać przy współpracy prywatnych podmiotów, instytucji państwowych oraz ośrodków naukowych i ich przedstawicieli. W procesie jego powstawania należy zwrócić uwagę na stworzenie łatwej w obsłudze strony internetowej, która będzie dostępna dla wszystkich zainteresowanych użytkowników sieci. Uwzględniając wyniki przeprowadzonych badań, w odniesieniu do takiej strony internetowej można sformułować pewne grupy rekomendacji, które muszą, powinny i mogą być uwzględnione podczas jej projektowania. Są to:

---

52 Wilkowski M., *Oddolne archiwizacje Internetu...*, s. 207–220.

53 A. Kugler, T. Beinert, A. Schoger, *Archiwizacja Internetu jako usługa naukowa*, „Biuletyn EBIB” 2017, nr 172, s. 1–8.

54 Wszelkie źródła cyfrowe, wykorzystane przy powstawaniu niniejszego artykułu, zostały zarchiwizowane w serwisie Internet Archive w osobistej kolekcji tematycznej. Dla większości z nich była to pierwsza archiwizacja w historii.



- a) rekomendacje najwyższej rangi: dostępność stron internetowych oraz ogólna struktura serwisu, jej szata graficzna oraz design,
- b) rekomendacje warte uwagi: komunikacja wirtualna i tradycyjna z użytkownikiem oraz kompletność i dokładność informacji zawartych w odnośnikach, źródłach i metadanych,
- c) kryteria neutralne: aktualność serwisu oraz narzędzia nawigacji,
- d) kryteria najmniejszej troski: informacje ogólne o archiwum oraz zasoby archiwum.

## Bibliografia

- Adamus-Kowalska J., *Zastosowanie technologii archiwizacji Internetu w identyfikacji i ratowaniu cyfrowego dziedzictwa kulturowego Ukrainy*, „Annales Universitatis Paedagogicae Cracoviensis Studia ad Bibliothecarum Scientiam Pertinentia” 2022, nr 20, s. 596–611.
- Archiwa Państwowe, *Forum archiwizacji zasobów polskiego Internetu*, 2021, [on-line:] <https://archiwa.gov.pl/forum-archiwizacji-zasobow-polskiego-internetu> – 27.09.2023.
- Archiwum Społeczne Polskiego Webu, *Informacje*, 2023, [on-line:] <https://aspw.pl/informacje> – 27.09.2023.
- Derfert-Wolf L., *Archiwizacja Internetu – wnioski i rekomendacje z kilku raportów*, „Biuletyn EBIB” 2017, nr 172, s. 1–11.
- Dombrowski Q., *SUCHO Project, 2023*, [on-line:] <https://www.sucho.org> – 27.09.2023.
- Gaca S., *Propozycja modelu archiwizacji polskiego Internetu na bazie doświadczeń brytyjskich*, „Officina Historiae” 2021, t. 4, nr 1, s. 79–90.
- Głowacka E., *Studium kompleksowego zarządzania jakością (TQM) w bibliotekoznawstwie i informacji naukowej*, Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, Toruń 2000.
- Gmerek K., *Archiwa internetowe po obu stronach Atlantyku – Internet Archive, Wayback Machine oraz UK Web Archive*, „Biuletyn EBIB” 2012, nr 128, s. 1–12.
- Gmiterek G., *Długoterminowa archiwizacja zasobów cyfrowych*, „Res Historica” 2013, nr 35, s. 213–235.
- Historia stron, [on-line:] <http://historiastron.pl> – 27.09.2023.
- Internet Archive, [on-line:] <https://archive.org> – 27.09.2023.
- Jankowska M., *Biblioteki akademickie – trendy dotyczące zasobów elektronicznych*, [w:] Informacja dla nauki a świat zasobów cyfrowych, red. H. Ganińska, Poznań 2008, s. 168.
- Januszko-Szakiel A., *Długoterminowa archiwizacja zasobów cyfrowych – program dla polskich bibliotek*, „Przegląd Biblioteczny” 2011, nr 79, s. 211–230.
- Kahle B., *Preserving the Internet*, “Scientific American” 1997, nr 276, s. 82–83.
- Koncewicz P., *Ludzie którzy przechowują Internet, czyli Wayback Machine w pigułce*, 2022, [on-line:] <https://antyweb.pl/internet-archive-wayback-machine-w-pigulce> – 27.09.2023.
- Konopa B., *Archiwizacja Webu w Europie*, „Archeion” 2020, nr 121, s. 1–21.

- Konopa B., *Strategia selektywna jako narzędzie w archiwizacji Webu. Analiza wybranych przykładów*, „Archiwa – Kancelarie – Zbiory” 2020, nr 11, s. 97–118.
- Konopa B., *Witryna internetowa jako materiał archiwalny*, „Archiwa – Kancelarie – Zbiory” 2016, nr 7, s. 113–120.
- Konopka-Nowina M., *Problematyka zarządzania informacją w procesie archiwizacji zasobów*, „Media, Kultura, Komunikacja Społeczna” 2014, nr 10, s. 30–43.
- Kugler A., Beinert T., Schoger A., *Archiwizacja Internetu jako usługa naukowa*, „Biuletyn EBIB” 2017, nr 172, s. 1–8.
- MementoWeb, *Guide*, 2023, [on-line:] <https://timetravel.mementoweb.org/guide/api> – 27.09.2023.
- Nielsen J., *Using web archives in research: an introduction*, NetLab, Aarhus 2016, s. 15–25.
- Sapa R., *Benchmarking w doskonaleniu serwisów WWW bibliotek akademickich*, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków 2005.
- The National Archives, [on-line:] <https://cdn.nationalarchives.gov.uk/documents/information-management/web-archiving-guidance.pdf> – 27.09.2023.
- Ukrainian Research Institute, *Russia’s war on Ukraine*, [on-line:] <https://archive-it.org/collections/18886> – 27.09.2023.
- W gospodarce, *Archiwizacja Internetu to realny problem*, 2018, [on-line:] <https://wgospodarce.pl/informacje/44931-archiwizacja-internetu-to-realny-problem> – 27.09.2023.
- Wilkowski M., *Oddolne archiwizacje Internetu jako działania społeczne*, „Archiwa – Kancelarie – Zbiory” 2015, nr 6, s. 207–220.
- Wilkowski M., *Wayback Machine – podstawy wykorzystania*, „Biuletyn EBIB” 2017, nr 172, s. 1–3.
- Woźniak W., *Archiwizacja Internetu – próba podsumowania dotychczasowych prac i ustaleń*, „Archiwa – Kancelarie – Zbiory” 2019, nr 10, s. 75–98.

## Streszczenie

Archiwizacja zasobów globalnej sieci jest procesem, który w różnych częściach świata przybiera odmienne formy i natężenie. Artykuł przedstawia strategię gromadzenia archiwalnych danych w Internecie, historię powstania internetowych archiwów oraz problemy archiwizacji Webu w Polsce. Jego zasadniczym celem jest ocena 14 wybranych europejskich archiwów Internetu, ukazanie ich mocnych i słabych stron zarówno w warstwie treściowej, jak i funkcjonalnej. Zastosowano metodę heurystyczną i benchmarkingową. Ocenie poddano ogółem osiem kryteriów – cztery treściowe i cztery funkcjonalne (informacje o archiwum; zasoby archiwum; kompletność oraz dokładność informacji; komunikacja z użytkownikiem; design i struktura serwisu; dostępność serwisu; narzędzia nawigacji; aktualność serwisu – częstotliwość zmian oraz wprowadzania dokumentów do archiwum). Badania przeprowadzono na przełomie maja i czerwca 2023 r. W badaniu poza autorem niniejszego artykułu uczestniczyło 20 studentów drugiego roku studiów uzupełniających kierunku architektura informacji Uniwersytetu Mikołaja Kopernika w Toruniu. Z przeprowadzonych badań wynika, że pod względem treściowym oceniane serwisy plasują się na podobnym poziomie. Różnice ujawniają się w ocenie kryteriów funkcjonalnych. Najwyżej ocenianą cechą treściową były „informacje o archiwum”, najniżej – „komunikacja z użytkownikiem”. Z kolei w przypadku cech funkcjonalnych najlepiej wypadła „aktualność serwisu”, najgorzej – „dostępność serwisów” oraz „design i struktura serwisów”. W kontekście braku polskiego archiwum Internetu autor postuluje, by

w przypadku jego projektowania zwrócić szczególną uwagę na kwestie związane z ogólną strukturą strony, jej szatą graficzną, dostosowaniem do potrzeb grup osób, które mogą być wykluczone cyfrowo, a także na odpowiednie umiejscowienie formularzy kontaktowych.

Słowa kluczowe: architektura informacji, archiwizacja Internetu, benchmarking, dziedzictwo kulturowe, europejskie archiwa

## **European Internet archiving initiatives – review and benchmarking assessment of selected websites**

### **Abstract**

Archiving the resources of the global Web is a process that takes different forms and intensities in different parts of the world. The article presents strategies for collecting archival data on the Internet, the history of the creation of Internet archives and the problems of Web archiving in Poland. Its main goal is to evaluate 14 selected European Internet archives, showing their strengths and weaknesses, both in terms of content and functionality. The method in this article was based on heuristic and benchmarking methods. A total of eight criteria were assessed – four content-related and four functional (information about the archive; archive resources; completeness and accuracy of information; communication with the user; design and structure of the website; accessibility of the website; navigation tools; timeliness of the website – frequency of changes and entries of documents into the archive). The research was conducted at the turn of May and June 2023. Apart from the author of this article, 20 students in their second year of supplementary studies in information architecture at the Nicolaus Copernicus University in Toruń participated in the study. The conducted research shows that, in terms of content, the assessed websites are at a similar level. Differences are revealed in the assessment of functional criteria. The highest rated content feature was “information about the archive”, the lowest – “communication with the user”. In turn, in the case of functional features, the best scores were for “website timeliness”, and the worst for “service availability” and “service design and structure”. Due to the lack of a Polish Internet archive, the author postulates that future designers should pay special attention to issues related to the general structure of the website, its graphic design, adaptation to the needs of groups of people who may be digitally excluded, as well as the appropriate location of contact forms.

Keywords: information architecture, Internet archiving, benchmarking, cultural heritage, European archives