



ISSN 2080-1807

TORUŃSKIE STUDIA BIBLIOLOGICZNE
2019, nr 1 (22)

Bartłomiej Konopa*

Uniwersytet Mikołaja Kopernika w Toruniu

Archiwum Państwowe w Bydgoszczy

e-mail: bartlomiejkonopa@gmail.com

Zachowanie różnorodności Sieci w archiwach Webu

DOI: <http://dx.doi.org/10.12775/TSB.2019.003>

STRESZCZENIE: Web charakteryzuje się niezwykle różnorodnością, zarówno jeśli chodzi o występujące w nim formy i treści, jak i samych jego użytkowników. W swojej trzydziestoletniej historii stał się bardzo popularnym i ważnym elementem współczesnego świata, przez co wzrasta zainteresowanie nim wśród badaczy różnych dziedzin. Pomimo tego, że zamieszczane jest w nim coraz więcej informacji, to w dużym tempie część z nich jest zamieniana na inne lub nieodwracalnie tracona. Remedium na to może stanowić rozwijana od połowy lat 90. XX w. archiwizacja Webu, która pozwala na gromadzenie i zachowanie zasobów Sieci. W artykule podjęto rozważania, w jakim zakresie stosowane metody, rozwiązania i strategie umożliwią zabezpieczenie występującej w WWW różnorodności.¹

SŁOWA KLUCZOWE: archiwizacja Webu, archiwa Webu, zachowanie dziedzictwa kulturowego.

* Uczestnik studiów doktoranckich z zakresu historii, prowadzonych na Wydziale Nauk Historycznych Uniwersytetu Mikołaja Kopernika w Toruniu.

Wstęp

Istotna rola Internetu w dzisiejszych czasach coraz częściej jest dostrzegana i nie trzeba nikogo o niej przekonywać. Internet stał się ważnym elementem w życiu człowieka, umożliwiając mu szybką komunikację i dostęp do informacji, przez co znacząco wpłynął na rozwój i funkcjonowanie dzisiejszego społeczeństwa, zwłaszcza w świecie „Zachodu”. Jeden z jego kluczowych elementów stanowi oparty o hipertekst i hiperłącza system World Wide Web, skrótowo nazywany WWW lub Web. W ciągu trzech dekad zyskał on niezwykłą popularność i wraz z innymi usługami globalnej Sieci znajduje wiele zastosowań w codziennym życiu, zarówno jako narzędzie pracy oraz źródło rozrywki, jak i obiekt badań naukowych. Liczba użytkowników Webu wciąż wzrasta, i, choć dominują wśród nich ludzie młodzi, to korzystają z niego również politycy, dziennikarze, artyści, naukowcy, działacze społeczni oraz przedstawiciele wielu innych środowisk. Jest on coraz powszechniejszym miejscem prowadzenia debaty publicznej, publikowania prac naukowych i danych badawczych, rozpowszechniania dzieł literackich czy artystycznych. Wiele z tych aktywności przybiera nowe, nieznane wcześniej formy. Ze względu na swoje obecne znaczenie zawartość Webu traktowana bywa jako swoiste dziedzictwo kulturowe, które na bieżąco dokumentuje współczesne społeczeństwo i świat, w którym ono żyje.

Duża popularność, powszechny i łatwy dostęp oraz szerokie spektrum możliwości dostarczanych przez WWW skutkuje jego znaczącą, wewnętrzną różnorodnością. Można mówić przynajmniej o jej trzech wymiarach. Pierwszym z nich może być, wynikająca z rozpowszechnienia wykorzystywania Internetu, rozwoju Webu 2.0 oraz mediów społecznościowych różnorodność użytkowników, zarówno twórców, jak i odbiorców. Drugim – różnorodność form, począwszy od statycznych witryn urzędów, poprzez prywatne blogi, kończąc na portalach angażujących duże grupy ludzi, np. artystów. Ostatnim jej rodzajem może być różnorodność treści, która za pomocą Webu jest publikowana. W tym miejscu można postawić pierwsze pytanie: skoro Sieć jest ważnym elementem współczesnego świata to czy możliwe jest jej zachowanie wraz z jej różnorodnością?



Niekiedy Web jest rozpatrywany jako archiwum lub biblioteka sama w sobie, gromadząca i przechowująca olbrzymie ilości danych, w których jej użytkownicy mogą zdobyć wiedzę na niemal każdy temat¹. Taka opinia może być efektem przeświadczenia, iż informacja raz zamieszczona gdzieś w Sieci zostaje w niej na zawsze. Dodatkowo może być ona umacniana przez rozważanie w jej kontekście prawa do zapomnienia oraz próby wartościowania publikowanych w niej treści. Jednak jak pokazują badania, a także codzienne obserwacje i doświadczenia korzystania z zasobów WWW, nie ma gwarancji, że zachowają one swoją pierwotną formę oraz lokalizację. Internet charakteryzują się niezwykle dynamiką, wiele treści zmienia się oraz jest zastępowane innymi lub znika pozostawiając po sobie informację o błędzie 404². Ich trwałość jest zatem nieduża i wraz z szybkim rozwojem technologii oraz zastępowaniem ich przez nowsze, a także ewolucją samego wykorzystywania Internetu (choćby wzrastająca popularność aplikacji mobilnych i dominująca dziś pozycja mediów społecznościowych) ta niewielka trwałość nie pozwala uznać Webu za miejsce umożliwiające zachowanie informacji o otaczającej rzeczywistości w długiej perspektywie. Sieć daje odpowiedzi na wiele pytań i zaspokaja potrzeby jej użytkowników, jednak mają one charakter bieżący, zgodny z tym, co aktualnie jest opublikowane³.

Archiwizacja Webu

Skoro Web stanowi tak ważny element współczesnego świata, jest wykorzystywany przez badaczy z wielu różnych dziedzin i może budzić

¹ Takie rozważanie przedstawiono m.in. w artykule A. Sobczak, *Internet jako globalne archiwum społeczne – rozważania na temat roli Internetu w dokumentowaniu dziejów ludzkości*, [w:] *Toruńskie Konfrontacje Archiwalne. Tom IV. Nowa archiwistyka – archiwa i archiwistyka w ponowoczesnym kontekście kulturowym*, red. W. Chorążyczewski, W. Piasek, A. Rosa, Toruń 2014, s. 237–247.

² Kod/błąd 404 Not Found – jeden z kodów odpowiedzi protokołu http informujący o tym, że klient połączył się z serwerem do którego skierował zapytanie, jednak pod wskazanym adresem nie znajduje się żaden plik. Wykaz kodów odpowiedzi: Kod odpowiedzi HTTP [online] [dostęp 29 maja 2019]. Dostępny w World Wide Web: https://pl.wikipedia.org/wiki/Kod_odpowiedzi_HTTP.

³ J. Nielsen, *Using Web Archives in Research – an Introduction* [online], NetLab 2016, s. 6–7 [dostęp 18 marca 2019]. Dostępny w World Wide Web: <http://www.netlab.dk/>

ich zainteresowanie w przyszłości, próby zachowania go wydają się być koniecznością. Działalnością, dzięki której można ten cel osiągnąć, podejmowaną w różnych częściach świata od połowy lat 90., jest archiwizacja Webu. Pozwala ona na tworzenie i gromadzenie źródeł, które w przyszłości mogą być wykorzystywane do poznania dawnej, niefunkcjonującej już Sieci. Czy umożliwi ona zachowanie i dostrzeżenie charakteryzującej jej wewnętrznej różnorodności, o której wspomniano wcześniej? Aby móc odpowiedzieć na to pytanie, należy przyjrzeć się temu, w jaki sposób zasoby archiwalnego Webu są gromadzone, a następnie w jaki sposób można do nich dotrzeć. Z tego punktu widzenia istotne są stosowane metody, strategie, a także zakres podejmowanych działań.

Archiwizację Webu można rozumieć dwojako: szeroko, jako każde rozmyślnie zachowanie zasobów sieciowych⁴, lub wąsko, jako działalność mającą na celu gromadzenie danych, przechowywanie oraz udostępnianie ich osobom zainteresowanym⁵. W dalszej części rozważań nad tym zagadnieniem zostanie zastosowana jego druga, węższa definicja. Obejmuje ona przede wszystkim funkcjonujące już na świecie różnorodne archiwa Webu oraz podobne projekty, które można rozumieć jako usługi, mające na celu przygotowanie źródeł dotyczących dawnego WWW, z których w przyszłości korzystać będą badacze, dziennikarze lub inni zainteresowani. Ich działalność przeprowadzana jest przeważnie na większą skalę, często z misją zachowania jak największego wycinka Webu rozumianego jako dziedzictwo kulturowe, co powinno uwzględnić jego różnorodność i złożoność. Tak rozumiana archiwizacja pozwala wykluczyć gromadzenie zasobów na potrzeby indywidualnych badań, które może nie służyć później szerszemu udostępnianiu i mieć zdecydowanie węższy zakres.

wp-content/uploads/2016/10/Nielsen_Using_Web_Archives_in_Research.pdf; teŝe, *Recording the Web*, [w:] *Towards a Philosophy of Digital Media*, red. A. Romele, E. Terrone, Palgrave Macmillan, Cham 2018, s. 53–54.

⁴ N. Brügger, *Web Archiving – Between Past, Present, and Future*, [w:] *The Handbook of Internet Studies*, red. M. Consalvo, C. Ess, Blackwell Publishing Ltd 2011, s. 25.

⁵ *ISO/DTR 14873 Information and documentation – Statistics and Quality Indicators for Web Archiving* [online], 2012, s. VI [dostęp 18 marca 2019]. Dostępny w World Wide Web: http://netpreserve.org/resources/IIPC_project-SO_TR_14873_E_2012-10-02_DRAFT.pdf.

Gromadzenie zasobów Sieci

Omawiana działalność może być realizowana za pomocą dwóch podstawowych metod – zasoby sieciowe mogą być gromadzone od strony serwera oraz od strony klienta. W przypadku archiwów Sieci stosowana jest druga metoda, która przeprowadzana jest z wykorzystaniem specjalnych robotów internetowych nazywanych crawlerami lub harvesterami. Zadaniem tego oprogramowania jest poruszanie się po hiperłączach i pobieranie danych z serwera poprzez skierowanie do niego odpowiedniego zapytania (czyli tak, jak robi to człowiek korzystający z przeglądarek internetowych)⁶. Sposób, w jaki wykonują one zrzuty stron WWW, wpływa na charakter powstałych w trakcie archiwizacji źródeł (Niels Brügger określa je mianem *reborn digital*⁷), a także na to, co może być gromadzone. Rozwiązanie to ma swoje ograniczenia, ponieważ crawlery nie zawsze radzą sobie z każdym rodzajem zawartości Webu, m.in. z treściami generowanymi dynamicznie, skryptami JavaScript lub technologią Flash, a także Siecią głęboką. Dodatkowo wgląd do części zasobów może być zablokowany, chociażby poprzez konieczność zalogowania się lub odpowiednim ustawieniem pliku robots.txt, który odpowiedzialny jest za dostęp crawlerów do witryny lub jej części⁸.

Jak można zauważyć, przeprowadzana w ten sposób archiwizacja Webu posiada swoje ograniczenia, które nie pozwalają zgromadzić wszelkich zamieszczanych w nim treści. Jest to również niemożliwe ze względu na czasochłonność tego procesu oraz konieczność przygotowania olbrzymiej przestrzeni dyskowej, na której można będzie przechowywać powstałe niebotyczne ilości danych. Zachowanie pełnego obrazu rzeczywistości WWW jest zadaniem niewykonalnym, co wymusza poszukiwanie innych rozwiązań, które pozwolą na wgląd w funkcjonowanie dawnej Sieci. Przede wszystkim można podejmować próby zarchiwizowania

⁶ F. Thouvenin, P. Hettich, H. Burkert, U. Gasser, *Remembering and Forgetting in the Digital Age*, Cham 2018, s. 91.

⁷ N. Brügger, *When the present web is later the past: web historiography, digital history and internet studies* [online], *Historical Social Research*, 37, 4, 2012, s. 104, 107–109 [dostęp 18 marca 2019]. Dostępny w World Wide Web: www.jstor.org/stable/41756477.

⁸ F. Thouvenin, P. Hettich, H. Burkert, U. Gasser, dz. cyt., s. 92; *ISO/DTR 14873*, s. 8–9.

jak największego wycinka Webu, który ukazałby jego złożoność i różnorodność. Jest to jedna ze strategii wykorzystywanych w archiwistyce webowej nazywana *bulk harvesting*, czyli gromadzenie masowe. W trakcie jej stosowania crawler porusza się po hiperłączach, począwszy od początkowej listy adresów URL (tzw. *seed list*), i wykonuje zrzuty witryn, które odpowiadają ustalonym kryterium (znajdują się na odpowiedniej domenie najwyższego poziomu, mają określony rozmiar etc.). Powstają w ten sposób olbrzymie ilości danych, które nie posiadają jednak konkretnego, ustrukturyzowanego układu⁹.

Najbardziej rozpoznawalnym przykładem realizacji tego założenia jest amerykańska fundacja Internet Archive, która jest jedną z pierwszych inicjatyw tego rodzaju na świecie i obecnie posiada największe zasoby zarchiwizowanego Webu. Jej założyciel, Brewster Kahle, postawił przed nią niemożliwe zadanie – zachowania całej publicznie dostępnej Sieci, z którego w miarę możliwości stara się ona wywiązywać. W ramach licznych projektów Internet Archive (np. ogólne archiwizacje Webu, zabezpieczanie odnośników z Wikipedii) gromadzi różnorodne wycinki Sieci, które pozwalają na przeglądanie milionów archiwalnych wersji witryn WWW pochodzących aż drugiej połowy lat 90. XX w¹⁰. Niekiedy mogą być to treści, których nie można już znaleźć w żywej Sieci, jak na przykład zasoby pochodzące z zamkniętej w 2010 r. domeny jugosłowiańskiej¹¹, jednakże sposób działania robota archiwizującego, a także ogrom tego przedsięwzięcia, sprawia, że często nie wiadomo, co dokładnie trafia do zasobu tego archiwum. Świadectwem tego jest znalezienie w zasobach Internet Archive witryn pochodzących z domeny Korei Północnej, które znalazły się tam w bliżej nieokreślony sposób, a poznanie go wymagało pogłębionych poszukiwań¹².

⁹ J. Nielsen, *Using web archives*, s. 21; _ISO/DTR 14873, s. 8–9.

¹⁰ J. Nielsen, *Using web archives*, s. 35–36.

¹¹ Badania nad zachowaniem zasobów z jugosłowiańskiej domeny zaprezentowano w artykule: A. Ben-David, *What does the Web remember of its deleted past? An archival reconstruction of the former Yugoslav top-level domain* [online], „New Media & Society” 2016, R. 18(7), s. 1103–1119 [dostęp 18 marca 2019]. Dostępny w World Wide Web: <https://journals.sagepub.com/doi/abs/10.1177/1461444816643790>.

¹² M. Wilkowski, *Archiwum Webu jako black box* [online] [dostęp 17 marca 2019]. Dostępny w World Wide Web: <https://webarch.uw.edu.pl/blog/archiwum-webu-jako-black-box/>.

Strategię masowego gromadzenia zasobów Webu wykorzystują również inne archiwa działające w na świecie, w tym także w Europie. Jednak ze względu na fakt, iż prowadzą je instytucje dziedzictwa, takie jak biblioteki narodowe, zakres archiwizacji jest mniejszy, przede wszystkim ze względu na ich misję oraz przepisy prawa. Tego rodzaju projekty zainteresowane są zachowaniem wycinka Sieci powiązanego z daną narodowością i w związku z tym przy wykorzystaniu omawianej strategii ograniczają się do zachowania materiałów zarejestrowanych na krajowych domenach najwyższego poziomu. Ilość pozyskiwanych przez nie danych jest zdecydowanie mniejsza, więc oczekiwać można od nich większej przejrzystości zbiorów¹³. *Seed list* pozyskiwane są od instytucji zarządzających domenami krajowymi, które mogą być rozszerzane o dodatkowe adresy URL znajdujące się na innych domenach, ponieważ użytkownicy z danego kraju mogą z nich równie dobrze korzystać. Ograniczenie się tylko do archiwizacji wyłącznie jednej TLD mogłoby nie oddać pełnego obrazu wycinka Sieci wyodrębnionego na podstawie kryterium narodowości. Taką praktykę stosuje się m.in. w Danii, gdzie archiwizowane są witryny na podstawie lokalizacji geograficznej lub wskazań użytkowników. Podejście to nosi znamiona innej strategii wykorzystywanej w archiwizacji Webu¹⁴.

Strategia selektywna, bo to o niej mowa, znacząco różni się od gromadzenia masowego, przede wszystkim innym i większym nakładem pracy człowieka (w przypadku wcześniej omawianej praktyki prowadzi się ona głównie do nadzorowania pracy crawlera), co skutkuje także zachowaniem mniejszej ilości zasobów. Jej istotą jest subiektywny wybór zasobów do archiwizacji na podstawie wcześniej ustalonych kryteriów, co pozwala na większą kontrolę nad tym, co zostanie objęte tym procesem oraz jego dokładnością. Dzięki temu możliwe jest również nadanie zasobowi struktury poprzez tworzenie katalogów tematycznych lub specjalnych kolekcji. Archiwizacja selektywna może być stosowana jako uzupełnienie gromadzenia masowego, co pozwoli uzupełnić luki

¹³ F. Thouvenin, P. Hettich, H. Burkert, U. Gasser, dz. cyt., s. 94–95.

¹⁴ S. Schostag, E. Fønss-Jørgensen, *Webarchiving: Legal Deposit of Internet in Denmark. A Curatorial Perspective* [online], „Microform & Digitization Review” 2012, R. 41, nr 3–4, s. 111 [dostęp 17 marca 2019]. Dostępny w World Wide Web http://netarkivet.dk/wp-content/uploads/Artikel_webarkivering1.pdf.

powstałe w jego czasie, chociażby poprzez częstsze gromadzenia witryn charakteryzujących się dużą zmiennością, jak np. strony z informacjami, a także ważniejszych lub unikatowych, trudniejszych do uchwycenia zasobów¹⁵. Strategia ta może również funkcjonować samodzielnie, co ma miejsce w przypadku projektów, którym przepisy prawa nie pozwalają na szersze gromadzenie (jak np. w Holandii¹⁶, czy do 2013 r. w Wielkiej Brytanii¹⁷), o mniejszej skali lub gdy domena krajowa nie istnieje, co ma miejsce w przypadku Ondarenet działającym w Kraju Basków¹⁸.

Podjęcie selektywne można podzielić na dwa rodzaje. Pierwszym z nich jest archiwizacja tematyczna, w trakcie której zasoby dobierane są na podstawie wcześniejszych kryteriów i następnie przydzielane do dużych katalogów tematycznych (takich jak np. sztuka, rząd, edukacja etc.), które rozbudowywane są w sposób ciągły¹⁹. Drugim jest tworzenie mniejszych kolekcji specjalnych, na przykład bardziej skonkretyzowanych zbiorów tematycznych, poświęconych wąskiemu zagadnieniu. Przykłady takich kolekcji można znaleźć w zbiorach cyfrowych Biblioteki Kongresu USA²⁰ lub United Kingdom Web Archive²¹. Innym ich rodzajem jest *event harvesting*, czyli gromadzenie zasobów webowych i budowanie kolekcji dotyczących konkretnych, zarówno zaplanowanych (np. wybory lub Igrzyska Olimpijskie), jak i przypadkowych (np. katastrofy naturalne, zamachy terrorystyczne etc.). Zarówno kolekcje tematyczne, jak i *event harvesting* tworzone są przez określony czas²².

¹⁵ F. Thouvenin, P. Hettich, H. Burkert, U. Gasser, dz. cyt., s. 95–96.

¹⁶ *Selection* [online] [dostęp 17 marca 2019]. Dostępny w World Wide Web: <https://www.kb.nl/en/organisation/research-expertise/long-term-usability-of-digital-resources/web-archiving/selection>.

¹⁷ *Collection guides. UK Web Archive* [online] [dostęp 17 marca 2019]. Dostępny w World Wide Web: <https://www.bl.uk/collection-guides/uk-web-archive>.

¹⁸ F. P. Vernalte, S. M. Maciá, *Capturing the Basque Web* [online], metariały z konferencji LIDA 2009, Dubrownik i Zadar (Chorwacja), 25–29 maja 2009, s. 8–9 [dostęp 17 marca 2019]. Dostępny w World Wide Web: http://eprints.rclis.org/13164/1/EN_Lida_paper_Ondarenet_APA.pdf.

¹⁹ J. Nielsen, *Using web archives*, s. 22.

²⁰ Kolekcje archiwalny witryn WWW w zbiorach Biblioteki Kongresu USA: *Digital Collections* [online] [dostęp 17 marca 2019]. Dostępny w World Wide Web: <https://www.loc.gov/collections/?fa=original-format:archived+web+site>.

²¹ *Topics and Themes* [online] [dostęp 17 marca 2019]. Dostępny w World Wide Web: <https://www.webarchive.org.uk/en/ukwa/collection>.

²² J. Nielsen, *Using web archives*, s. 21–22.

Ciekawym rozwiązaniem, wykorzystywanym w ramach archiwizacji selektywnej, jest możliwość rozdzielenia doboru materiałów pomiędzy podmioty uczestniczące w projekcie, a także zaangażowanie w ten proces użytkowników. Przykładem rozdzielenia tej kompetencji pomiędzy instytucje partycypujące jest australijskie archiwum Webu PANDORA, w przypadku którego każdy z 10. uczestników przedsięwzięcia samodzielnie wybiera zasoby do archiwizacji na podstawie własnych wytycznych. Każdy z nich ma określony zakres działalności: Biblioteka Narodowa Australii zajmuje się materiałami o charakterze ogólnonarodowym, biblioteki stanowe związanymi z ich regionem, National Gallery of Australia ze sztukami wizualnymi itd.²³ Podobne rozwiązanie stosuje Biblioteka Narodowa Hiszpanii, która współpracuje z instytucjami we Wspólnotach Autonomicznych w celu budowania kolekcji dawnego Webu poświęconym poszczególnym rejonom²⁴.

Częstą praktyką jest również umożliwienie użytkownikom zgłaszania własnych propozycji zasobów do archiwizacji. Z ich pomocy, poprzez zamieszczony na portalu WWW odpowiedni formularz, korzystają m.in. archiwa Webu w Wielkiej Brytanii²⁵, Danii²⁶ oraz Chorwacji²⁷. Zgłoszenie wymaga, oprócz podania URL, także dodatkowych informacji, które następnie są weryfikowane przez odpowiedniego pracownika, który następnie zadecyduje, czy dana witryna zostanie zachowana. Podobną funkcję, nazwaną *Save Page Now*, udostępnia Internet Archive. Pozwala ona na wskazanie strony internetowej do archiwizacji poprzez podanie jej adresu URL, jednak propozycje te nie są kontrolowane. Prowadzi to do wykorzystywania tej możliwości niezgodnie z jej przeznaczeniem, czego

²³ *PANDORA partners* [online] [dostęp 25 lutego 2019]. Dostępny w World Wide Web: <http://pandora.nla.gov.au/partners.html>.

²⁴ *Autonomous Communities. Collections* [online] [dostęp 25 lutego 2019]. Dostępny w World Wide Web: <http://www.bne.es/en/Colecciones/ArchivoWeb/Subcolecciones/ComunidadesAutonomas.html>.

²⁵ *Save a UK website* [online] [dostęp 25 lutego 2019]. Dostępny w World Wide Web: <https://www.webarchive.org.uk/en/ukwa/info/nominate>.

²⁶ *Du kan hjælpe Netarkivet* [online] [dostęp 25 lutego 2019]. Dostępny w World Wide Web: <http://netarkivet.dk/du-kan-hjaelpe-netarkivet/>.

²⁷ *Obrazac za prijavu online publikacija* [online] [dostęp 25 lutego 2019]. Dostępny w World Wide Web: http://haw.nsk.hr/en/registration_form.

przykładem jest używanie jej przez Państwo Islamskie w celu ukrywania w zasobach fundacji swoich materiałów propagandowych²⁸.

Korzystnie na zachowanie różnorodności Webu wpłynąć może również mnogość projektów podejmujących się jego gromadzenia. Oprócz głównego potentata jakim jest wspomniana kilkakrotnie amerykańska fundacja Internet Archive oraz narodowych archiwów WWW, które działają w wielu krajach europejskich (ale także w Nowej Zelandii²⁹ czy Singapurze³⁰), powstaje wiele mniejszych projektów. Przyczyny ich powstawania są różne, może to być chęć zachowania ważnych dla ich twórców zasobów Webu lub udokumentowanie jakiegoś zjawiska, ale także gromadzenie materiałów dla potrzeb badawczych. Za przykład takiego archiwum posłużyć może prowadzone od sierpnia 2001 r. DACHS, czyli Digital Archive for Chinese Studies prowadzone przez ośrodki w Heidelbergu oraz Leiden³¹, oraz działające od 2005 r. LAWAP, czyli Latin America Web Archive Project działający na Uniwersytecie Teksasu w Austin³². Prowadzenie takiej działalności ułatwiają serwisy oferujące usługę archiwizacji Webu (m.in. Archive-It należące do Internet Archive), co umożliwia budowanie własnych kolekcji witryn WWW instytucjom, które nie mogą robić tego samodzielnie³³. Przybywa także inicjatyw o charakterze oddolnym, wśród których wskazać można między innymi na działalność kolektywu Archive Team czy próby stworzenia anarchistycznego archiwum ruchu Occupy Wall Street³⁴.

²⁸ B. Konopa, *Propaganda ISIS ukrywana w Wayback Machine* [online] [dostęp 27 lutego 2019]. Dostępny w World Wide Web: <https://webarch.uw.edu.pl/blog/propaganda-isis/>.

²⁹ *New Zealand Web Archive* [online] [dostęp 27 lutego 2019]. Dostępny w World Wide Web: <https://natlib.govt.nz/collections/a-z/new-zealand-web-archive>.

³⁰ *Web Archive Singapore* [online] [dostęp 27 lutego 2019]. Dostępny w World Wide Web: <http://eresources.nlb.gov.sg/webarchives/landing-page>.

³¹ *DACHS – Digital Archive for Chinese Studies* [online] [dostęp 2 marca 2019]. Dostępny w World Wide Web: https://www.zo.uni-heidelberg.de/boa/digital_resources/dachs/index_en.html; o powstaniu i założeniach DACHS: H. E. Lecher, *Small Scale Academic Web Archiving: DACHS*, [w:] *Web Archiving*, red. J. Masanès, Berlin-Heidelberg 2006, s. 213–225.

³² *Latin American Web Archiving Project* [online] [dostęp 2 marca 2019]. Dostępny w World Wide Web: <http://lanic.utexas.edu/project/archives/>.

³³ *About Archive-It* [online] [dostęp 3 marca 2019]. Dostępny w World Wide Web: <https://archive-it.org/blog/learn-more/>.

³⁴ Więcej o oddolnych inicjatywach: M. Wilkowski, *Oddolne archiwizacje Internetu jako działania społeczne*, „Archiwa-Kancelarie-Zbiory” 2015, nr 6(8), s. 207–220.

Udostępnianie archiwalnego WWW

Zgromadzenie i przechowywanie olbrzymich zbiorów archiwalnego Webu, nie jest równoznaczne z zachowaniem jego różnorodności, ponieważ bez podjęcia odpowiednich działań może ona zniknąć w masie informacji. Niezbędne jest przygotowanie narzędzi, które pozwolą na przeszukiwanie i wykorzystanie zebranych materiałów, a także nadanie im struktury oraz opatrzenie ich odpowiednimi metadanymi pozwalającymi na pełne zrozumienie danego spektrum materiałów. Już sama liczba projektów, które podejmują się tego zadania, może utrudniać orientację w tym, co zostało zachowane oraz gdzie należy poszukiwać interesujących danego użytkownika zasobów. Dodatkową przeszkodę może stanowić fakt, iż wiele archiwów nie publikuje logów z pracy crawlerów ani swoich polityk selekcji, w wyniku czego stwierdzenie, co dokładnie się w nich znajduje, jest skomplikowane lub wręcz niemożliwe. Jednym z możliwych rozwiązań tego problemu jest przygotowanie agregatorów, które umożliwiłyby przeszukiwanie rozproszonych zasobów³⁵. Przykładem takiej inicjatywy jest projekt Memento, w ramach którego przygotowano wyszukiwarkę Time Traveler pozwalającą poszukiwać witryny w zbiorach kilkunastu archiwów Webu, a także wykonywać na ich podstawie jej rekonstrukcje³⁶.

Poszczególne archiwa oraz pozostałe projekty umożliwiają mniej lub bardziej otwarty dostęp do zasobów poprzez swoje portale internetowe (kwestia ta wynika przede wszystkim z sytuacji legislacyjnej, co wychodzi poza ramy niniejszego artykułu) i oferują różne narzędzia pozwalające na dotarcie do poszukiwanych obiektów. Podstawowym rozwiązaniem, zwłaszcza w przypadku projektów wykorzystujących strategię gromadzenia masowego, jest wyszukiwanie za pomocą adresów URL, które znaleźć można m.in. na witrynie Internet Archive³⁷, ale także archiwum

³⁵ A. AlSum, M. C. Weigle, M.L. Nelson, H. Van de Sompel, *Profiling web archive coverage for top-level domain and content language* [online], „International Journal on Digital Libraries” 2018, R. 14, nr 3–4, s. 149–152 [dostęp 3 marca 2019]. Dostępny w World Wide Web: <https://link.springer.com/article/10.1007/s00799-014-0118-y>.

³⁶ *About the Time Travel Service* [online] [dostęp 3 marca 2019]. Dostępny w World Wide Web: <http://timetravel.mementoweb.org/about/>.

³⁷ *Internet Archive* [online] [dostęp 5 marca 2019]. Dostępny w World Wide Web: <http://web.archive.org/>.

chorwackiego³⁸, hiszpańskiego³⁹, estońskiego⁴⁰ oraz wielu innych. Nie jest to rozwiązanie idealne, ponieważ użytkownik musi wcześniej znać adresy interesujących go zasobów⁴¹. Jest to dobrze widoczne na badaniach przeprowadzonych przez Anat Ben-David, która rekonstruując domenę jugosłowiańską bazowała na trzech listach adresów URL pozyskanych z zewnętrznych źródeł⁴².

Metoda ta nie umożliwia w pełni eksplorowania różnorodności danych Sieci zgromadzonej w zasobach archiwalnych Webu, dlatego też poszukiwane są dodatkowe rozwiązania, które pozwolą na sprawniejsze ich przeszukiwanie. Coraz więcej projektów umożliwia wykorzystanie wyszukiwania pełnotekstowego. Taką funkcję oferuje od 2016 r. Internet Archive (jest ona jednak ograniczona wyłącznie do stron głównych)⁴³, a także inne projekty, w tym m.in. kataloński Padicat⁴⁴ czy też brytyjskie archiwum rządowego Webu⁴⁵. Jest to możliwość oczekiwana i chętnie wykorzystywana przez użytkowników, zwłaszcza ze względu na jej obecność w popularnych wyszukiwarkach internetowych, takich jak Google. Wiele z nich pozwala na zastosowanie dodatkowych filtrów, np. pozwalających na wskazanie zakresu chronologicznego, formatu danych (tekst, obraz, wideo etc.) oraz konkretnej domeny, która ma być przeszukiwana. Ze względu na wymaganą moc obliczeniową oraz swój poziom skomplikowania efektywne funkcjonowanie wyszukiwarek pełnotekstowych jest

³⁸ *Harvesting of the national web domain (.hr)* [online] [dostęp 5 marca 2019]. Dostępny w World Wide Web: <http://haw.nsk.hr/en/harvesting-of-the-national-web-domain>.

³⁹ *Bienvenido al Archivo de la Web Española* [online] [dostęp 5 marca 2019]. Dostępny w World Wide Web: <http://www.dl-e.es/openwayback/wayback/>.

⁴⁰ *Eesti Veebiarhiiv* [online] [dostęp 5 marca 2019]. Dostępny w World Wide Web: <http://veebiarhiiv.digar.ee/>.

⁴¹ M. Costa, M. J. Silva, *Evaluating Web Archive Search Systems*, [w:] *Web Information Systems Engineering – WISE 2012. 13th International Conference, Paphos, Cyprus, November 28–30, 2012. Proceedings*, red. X. S. Wang, I. Cruz, A. Delis, G. Huang, Berlin-Heidelberg 2012, s. 442.

⁴² A. Ben-David, dz. cyt., s. 1103–1119.

⁴³ V. Goel, *Beta Wayback Machine – Now with Site Search!* [online] [dostęp 5 marca 2019]. Dostępny w World Wide Web: <https://blog.archive.org/2016/10/24/beta-wayback-machine-now-with-site-search/>.

⁴⁴ *Padicat The Web Archive of Catalonia* [online] [dostęp 5 marca 2019]. Dostępny w World Wide Web: <https://www.padicat.cat/en>.

⁴⁵ *Search* [online] [dostęp 5 marca 2019]. Dostępny w World Wide Web: <https://webarchive.nationalarchives.gov.uk/search/>.

jednak utrudnione i nie zawsze spełnia ono oczekiwania. Inną wprowadzaną metodą jest wyszukiwanie przy użyciu metadanych, które są generowane w trakcie archiwizacji, zwłaszcza tej przeprowadzanej w sposób selektywny⁴⁶. Przykład takiej wyszukiwarki znaleźć można na witrynie heidelberskiej części projektu DACHS⁴⁷.

Oprócz wyszukiwarek możliwe jest przeglądanie zasób archiwów Webu za pomocą zamieszczonych na ich witrynach katalogów oraz kolekcji specjalnych, które powstają przy zastosowaniu strategii selektywnej. Katalogi umożliwiają wyszukiwanie materiałów alfabetycznie lub na podstawie kilku lub kilkunastu grup rzeczowych (np. sztuka, historia, polityka, sport) wytypowanych przez twórców archiwum, które następnie mogą dzielić się na dodatkowe podkategorie. Ich przykłady znaleźć można w m.in. w australijskim projekcie PANDORA⁴⁸, archiwum chorwackim⁴⁹ oraz baskijskim⁵⁰. Podobny charakter mają kolekcje specjalne, które, jak już wcześniej wspomniano, dostępne są chociażby w brytyjskim archiwum Webu oraz zasobach biblioteki kongresu USA, a także w serwisie Archive-It⁵¹. Obie formy udostępniania archiwalnego Webu mają postać wykazu obiektów, opatrzonym mniej lub bardziej rozbudowanym opisem. Najczęściej składa się on z tytułu, daty archiwizacji oraz linku prowadzącego do zarchiwizowanej witryny, rzadziej zawiera on dodatkowe słowa kluczowe, kategorie lub szczegółową charakterystykę.

Poszczególne archiwa i projekty poszukują innych rozwiązań, które umożliwiłyby na szerszy dostęp do zarchiwizowanych zasobów. Dobrym pomysłem jest zamieszczenie informacji o zasobach w agregatorach (takich jak np. wcześniej wspomniany Time Travel będący częścią projektu Memento). Dzięki odpowiedniemu przystosowaniu meta danych, część

⁴⁶ M. Costa, M. J. Silva, dz. cyt., s. 442.

⁴⁷ *Search DACHS* [online] [dostęp 9 marca 2019]. Dostępny w World Wide Web: https://www.zo.uni-heidelberg.de/boa/digital_resources/dachs/search_en.html.

⁴⁸ *PANDORA – Preserving and Accessing Networked Documentary Resources Of Australia* [online] [dostęp 9 marca 2019]. Dostępny w World Wide Web: <https://pandora.nla.gov.au/>.

⁴⁹ *Croatian Web Archive* [online] [dostęp 9 marca 2019]. Dostępny w World Wide Web: <http://haw.nsk.hr/en>.

⁵⁰ *Ondarenet* [online] [dostęp 9 marca 2019]. Dostępny w World Wide Web: <http://www.ondarenet.kultura.ejgv.euskadi.eus:8085/ondarenet/>.

⁵¹ *Explore All Archives* [online] [dostęp 11 marca 2019]. Dostępny w World Wide Web: <https://archive-it.org/explore?show=Collections>.

zasobów z Chorwackiego Archiwum Webu dostępna jest w Europie, wśród innych zbiorów z całej Europy. Jest to obecnie jedyny projekt współpracujący z tą biblioteką cyfrową⁵². Inne archiwa, prowadzone przez biblioteki narodowe, mogą zamieszczać w ich katalogach informacje o swoich zbiorach. W katalogu Biblioteki Narodowej Irlandii dostępne są archiwalne witryny WWW gromadzone w ramach tematycznych kolekcji, a także samodzielnie⁵³. Również australijski serwis Trove, służący do przeszukiwania różnorodnych zasobów z tego kraju, pozwala jednocześnie przeszukiwać zasoby z projektu PANDORA, Archiwum Webu Rządu Australijskiego oraz z masowej archiwizacji domeny .au⁵⁴. Rozwijane są również bardziej zaawansowane rozwiązania, której mają na celu ułatwienie dostępu i szersze wykorzystania zasobów dawnego Webu. Na potrzeby Big UK Data Arts and Humanities Project opracowano prototypowy silnik wyszukiwawczy SHINE, pozwalający na przeszukiwanie zbiorów z zastosowaniem zaawansowanych filtrów oraz prostą wizualizację danych w postaci analizy trendów⁵⁵. Także singapurskie archiwum Webu zamieściło na swojej witrynie podobne narzędzie, pozwalające generować wykresy ukazujące częstotliwość pojawiania się danych słów lub fraz⁵⁶.

Podsumowanie

Tak jak już to zostało podkreślone na początku niniejszego artykułu, dążenie do zachowania Webu wydaje się być koniecznością, natomiast jego archiwizacja pozwala na realizację tego zamiaru. Sama Sieć oraz jej zawar-

⁵² K. Holub, I. Rudomino, *A decade of web archiving in the National and University Library in Zagreb* [online], materiały z konferencji IFLA WLIC 2015, Kapsztad (RPA), 11–20 sierpnia 2015, s. 7 [dostęp 11 marca 2019]. Dostępny w World Wide Web: <http://library.ifla.org/1092/1/090-holub-en.pdf>.

⁵³ *Web archive of the National Library of Ireland, [electronic resource] / 2011 – compiled by the National Library of Ireland* [online] [dostęp 11 marca 2019]. Dostępny w World Wide Web: <http://catalogue.nli.ie/Collection/vtls000510774>.

⁵⁴ *Archived websites (1996 – now)* [online] [dostęp 11 marca 2019]. Dostępny w World Wide Web: <https://trove.nla.gov.au/website>.

⁵⁵ *About SHINE* [online] [dostęp 11 marca 2019]. Dostępny w World Wide Web: <https://www.webarchive.org.uk/shine>.

⁵⁶ *Data Visualizer* [online] [dostęp 11 marca 2019]. Dostępny w World Wide Web: <http://eresources.nlb.gov.sg/webarchives/data-visualizer>.

tość pełnią niezwykle rolę w dzisiejszym świecie i wywierają znaczący wpływ na jego wygląd, a w związku z tym mogą budzić zainteresowanie obecnie, jak i w przyszłości. Rozwiązania stosowane w jego gromadzeniu i możliwości technologiczne nie pozwalają jednak zachować całego WWW, choć nie musi być to uznane za konieczne. Przede wszystkim można podważyć sensowność tak szeroko zakrojonej działalności, chociażby ze względu na przepastną ilość danych, a także wartość informacyjną zasobów Webu, ponieważ duża ich część może zostać uznana wręcz za pozbawioną tej wartości. Ponadto badacze przyzwyczaili się do tego, że nie mają dostępu do kompletnego zestawu źródeł do swoich prac i muszą korzystać z reprezentacyjnej, mniejszej lub większej, próbki.

Obecnie archiwa Webu wypracowały metody, umożliwiające działania w skali makro (gromadzenie masowe) i mikro (archiwizacja selektywna), które pozwalają zachować zarówno duże, rozbudowane wycinki zasobów Sieciowych, jak i małe kolekcje tematyczne oraz pojedyncze witryny o dużej wartości. Oba podejścia dają możliwość zachowania występujących w Webie różnorodności, zarówno w zakresie treści i form, a także ich użytkowników. Z rozwojem archiwizacji mediów społecznościowych, takich jak Facebook lub Twitter, można wiązać nadzieje na pozyskanie więcej zasobów z trzeciego z wymienionych obszarów. Wiele projektów łączy omawiane rozwiązania i obok dużych zbiorów danych pobieranych z domeny krajowej, budują małe kolekcje tematyczne lub przeprowadzają *event harvesting*. Poszukiwane są również sposoby na zwiększenie zasięgu gromadzenia zasobów Sieci, m.in. poprzez zaangażowanie dodatkowych, wyspecjalizowanych instytucji, a także samych użytkowników archiwów. Tak zakrojone działania powinny pozwolić na zachowanie pełnego obrazu Webu wraz z jego wewnętrznym zróżnicowaniem. Będzie to oczywiście wymagało dalszego rozwijania narzędzi i metod, które umożliwią nadążanie za postępującym rozwojem technologii, w tym wspomnianych już mediów społecznościowych czy aplikacji mobilnych.

Innym problemem w rozpatrywanym powyżej zagadnieniu jest dostęp do zasobów i możliwość jak najszerszego ich wykorzystania. Zgromadzenie olbrzymiej ilości witryn WWW i innych zasobów pochodzących z Sieci, bez zapewnienia odpowiednich narzędzi, sprawi, że zachowana w nich różnorodność zostanie niezauważona. Obecnie przeszukiwanie możliwe jest za pomocą wyszukiwarek oraz poprzez przeglądanie kata-

logów tematycznych. Pierwsze z tych rozwiązań wymaga dopracowania i dalszego rozwijania, ponieważ w tym momencie ich skuteczność nie jest zadowalająca. Drugie natomiast jest łatwiejsze i bardziej przejrzyste, jednak obejmuje przeważnie mały wycinek zbiorów danego archiwum. Przydatne będzie też opracowanie i rozbudowanie agregatorów, takich jak wymieniany Time Travel, które pozwolą na jednoczesny dostęp do zasobów rozproszonych pomiędzy liczne projekty. Pomocne będą także dodatkowe narzędzia, które umożliwią szersze wykorzystanie archiwalnego Webu. Przykładem tego może być chociażby przytaczana wcześniej wizualizacja danych, ale przydatne byłoby również udostępnienie API, dzięki któremu zainteresowani mogliby swobodnie korzystać ze zgromadzonych w archiwum informacji.

W powyższym artykule zwrócono uwagę przede wszystkim na metody i strategie gromadzenia wykorzystywane w działalności większych archiwów Webu, zwłaszcza tych o charakterze narodowym, natomiast obok nich powstaje i funkcjonuje już wiele mniejszych projektów. To również powinno korzystnie wpłynąć na zachowanie różnorodności, zwłaszcza, że część z nich ma charakter oddolny i społeczny, dzięki czemu może zwrócić uwagę na zasoby pomijane przez projekty o charakterze instytucjonalnym⁵⁷. Skala zjawiska, jakim jest archiwizacja Webu, wciąż się wzrasta, przyjmuje ono różne formy i obejmuje coraz więcej rozmaitych obszarów. Stawarza to szanse na zachowanie Sieci wraz z szerokim spektrum zachodzących w niej zjawisk oraz jej wewnętrzną różnorodnością, na co pozwalają obecnie już istniejące i wciąż rozwijane narzędzia. Wyzwaniem, które stoi przed archiwami Webu jest spowodowanie, aby zgromadzone w nich bogactwo informacji było widoczne i dostępne dla zainteresowanych. Konieczne będzie zwiększenie świadomości na temat posiadanych przez nie zbiorów oraz przygotowania lepszych sposobów ich przeszukiwania, a także opracowania metod pozwalających na efektywne ich wykorzystywanie⁵⁸.

⁵⁷ Przykład takiej inicjatywy może stanowić kolektyw Archiveteam: *Archiveteam* [online] [dostęp 16 marca 2019]. Dostępny w World Wide Web: https://www.archiveteam.org/index.php?title=Main_Page.

⁵⁸ Szereg wyzwań stojących przed archiwizacją Webu oraz ich możliwe rozwiązania przedstawiono w raporcie: E.T Mayer, A. Thomas, R. Schroeder, *Web Archives: The Future(s)*

Bibliografia**Opracowania**

- AlSum Ahmed, Weigle Michele C., Nelson Michael L., Sompel Herbert Van de, *Profiling web archive coverage for top-level domain and content language* [online], „International Journal on Digital Libraries” 2018, R. 14, nr 3–4, s. 149–152 [dostęp 3 marca 2019]. Dostępny w World Wide Web: <https://link.springer.com/article/10.1007/s00799-014-0118-y>.
- Brügger Niels, *Web Archiving – Between Past, Present, and Future*, [w:] *The Handbook of Internet Studies*, red. Mia Consalvo, Charles Ess, Malden 2011, s. 24–42.
- Brügger Niels, *When the present web is later the past: web historiography, digital history and internet studies* [online], „Historical Social Research” 2012, R. 37, nr 4, s. 104, 107–109 [dostęp 18 marca 2019]. Dostępny w World Wide Web: www.jstor.org/stable/41756477.
- Ben-David Anat, *What does the Web remember of its deleted past? An archival reconstruction of the former Yugoslav top-level domain* [online], „New Media & Society” 2016, R. 18(7), s. 1103–1119 [dostęp 18 marca 2019]. Dostępny w World Wide Web: <https://journals.sagepub.com/doi/abs/10.1177/1461444816643790>.
- Costa Miguel, Silva Mário J., *Evaluating Web Archive Search Systems*, [w:] *Web Information Systems Engineering – WISE 2012. 13th International Conference, Paphos, Cyprus, November 28–30, 2012. Proceedings*, red. X. Sean Wang [i in.], Berlin-Heidelberg 2012, s. 440–454.
- Holub Karolina, Rudomino Ingeborg, *A decade of web archiving in the National and University Library in Zagreb* [online], materiały z konferencji IFLA WLIC 2015, Kapsztad (RPA), 11–20 sierpnia 2015 [dostęp 11 marca 2019]. Dostępny w World Wide Web: <http://library.ifla.org/1092/1/090-holub-en.pdf>.
- Lecher Hanno E., *Small Scale Academic Web Archiving: DACHS*, [w:] *Web Archiving*, red. Julien Masanès, Berlin-Heidelberg 2006, s. 213–225.

[online], 2011, s. 8–24 [dostęp 18 marca 2019]. Dostępny w World Wide Web: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1830025.

- Mayer Eric T., Thomas Arthur, Schroeder Ralph, *Web Archives: The Future(s)* [online], 2011 [dostęp 16 marca 2019]. Dostępny w World Wide Web: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1830025.
- Nielsen Janne, *Recording the Web*, [w:] *Towards a Philosophy of Digital Media*, red. Alberto Romele, Enrico Terrone, Cham 2018, s. 51–76.
- Nielsen Janne, *Using Web Archives in Research – an Introduction* [online], Net-Lab 2016 [dostęp 18 marca 2019]. Dostępny w World Wide Web: http://www.netlab.dk/wp-content/uploads/2016/10/Nielsen_Using_Web_Archives_in_Research.pdf.
- Schostag Sabine, Fønss-Jørgensen Eva, *Webarchiving: Legal Deposit of Internet in Denmark. A Curatorial Perspective* [online], „Microform & Digitization Review” 2012, R. 41, nr 3–4, s. 111 [dostęp 17 marca 2019]. Dostępny w World Wide Web: http://netarkivet.dk/wp-content/uploads/Artikel_webarkivering1.pdf.
- ISO/DTR 14873 Information and documentation – Statistics and Quality Indicators for Web Archiving* [online], 2012, s. VI [dostęp 18 marca 2019]. Dostępny w World Wide Web: http://netpreserve.org/resources/IIPC_project-SO_TR_14873_E_2012-10-02_DRAFT.pdf.
- Sobczak Anna, *Internet jako globalne archiwum społeczne – rozważania na temat roli Internetu w dokumentowaniu dziejów ludzkości*, [w:] *Toruńskie Konfrontacje Archiwalne. Tom IV. Nowa archiwistyka – archiwa i archiwistyka w ponowoczesnym kontekście kulturowym*, red. Waldemar Chorążyczewski, Wojciech Piasek, Agnieszka Rosa, Toruń 2014, s. 237–247.
- Thouvenin Florent, Hettich Peter, Burkert Herbert, Gasser Urs, *Remembering and Forgetting in the Digital Age*, Cham 2018.
- Vernalte Francisca Pulgar, Maciá Sonia Marcos, *Capturing the Basque Web* [online], metariały z konferencji LIDA 2009, Dubrownik i Zadar (Chorwacja), 25–29 maja 2009 [dostęp 17 marca 2019]. Dostępny w World Wide Web: http://eprints.rclis.org/13164/1/EN_Lida_paper_Ondarenet_APA.pdf.
- Wilkowski Marcin, *Oddolne archiwizacje Internetu jako działania społeczne*, „Archiwa-Kancelarie-Zbiory” 2015, nr 6(8), s. 207–220.

Witryny internetowe

- About Archive-It* [online] [dostęp 3 marca 2019]. Dostępny w World Wide Web: <https://archive-it.org/blog/learn-more/>.



- About SHINE* [online] [dostęp 11 marca 2019]. Dostępny w World Wide Web: <https://www.webarchive.org.uk/shine>.
- About the Time Travel Service* [online] [dostęp 3 marca 2019]. Dostępny w World Wide Web: <http://timetravel.mementoweb.org/about/>.
- Archived websites (1996 – now)* [online] [dostęp 11 marca 2019]. Dostępny w World Wide Web: <https://trove.nla.gov.au/website>.
- Archiveteam* [online] [dostęp 16 marca 2019]. Dostępny w World Wide Web: https://www.archiveteam.org/index.php?title=Main_Page.
- Autonomous Communities. Collections* [online] [dostęp 25 lutego 2019]. Dostępny w World Wide Web: <http://www.bne.es/en/Colecciones/ArchivoWeb/Subcolecciones/ComunidadesAutonomas.html>.
- Bienvenido al Archivo de la Web Española* [online] [dostęp 5 marca 2019]. Dostępny w World Wide Web: <http://www.dl-e.es/openwayback/wayback/>.
- Croatian Web Archive* [online] [dostęp 25.02.2019]. Dostępny w World Wide Web: <http://haw.nsk.hr/en>.
- DACHS – Digital Archive for Chinese Studies* [online] [dostęp 2 marca 2019]. Dostępny w World Wide Web: https://www.zo.uni-heidelberg.de/boa/digital_resources/dachs/index_en.html.
- Du kan hjælpe Netarkivet* [online] [dostęp 25 lutego 2019]. Dostępny w World Wide Web: <http://netarkivet.dk/du-kan-hjaelpe-netarkivet/>.
- Eesti Veebiarhiiv* [online] [dostęp 5 marca 2019]. Dostępny w World Wide Web: <http://veebiarhiiv.digar.ee/>.
- Goel Vinay, *Beta Wayback Machine – Now with Site Search!* [online] [dostęp 5 marca 2019]. Dostępny w World Wide Web: <https://blog.archive.org/2016/10/24/beta-wayback-machine-now-with-site-search/>.
- Internet Archive* [online] [dostęp 5 marca 2019]. Dostępny w World Wide Web: <http://web.archive.org/>.
- Digital Collections* [online] [dostęp 17 marca 2019]. Dostępny w World Wide Web: <https://www.loc.gov/collections/?fa=original-format:archived+web+site>.
- Kod odpowiedzi HTTP* [online] [dostęp 29 maja 2019]. Dostępny w World Wide Web: https://pl.wikipedia.org/wiki/HTTP_404.
- Konopa Bartłomiej, *Propaganda ISIS ukrywana w Wayback Machine* [online] [dostęp 27 lutego 2019]. Dostępny w World Wide Web: <https://webarch.uw.edu.pl/blog/propaganda-isis/>.
- Latin American Web Archiving Project* [online] [dostęp 2 marca 2019]. Dostępny w World Wide Web: <http://lanic.utexas.edu/project/archives/>.

- New Zealand Web Archive* [online] [dostęp 27 lutego 2019]. Dostępny w World Wide Web: <https://natlib.govt.nz/collections/a-z/new-zealand-web-archive>.
- Ondarenet* [online] [dostęp 9 marca 2019]. Dostępny w World Wide Web: <http://www.ondarenet.kultura.ejgv.euskadi.eus:8085/ondarenet/>.
- Padicat The Web Archive of Catalonia* [online] [dostęp 5 marca 2019]. Dostępny w World Wide Web: <https://www.padicat.cat/en>.
- PANDORA – Preserving and Accessing Networked DOcumentary Resources Of Australia* [online] [dostęp 25 lutego 2019]. Dostępny w World Wide Web: <https://pandora.nla.gov.au/>.
- Save a UK website* [online] [dostęp 25 lutego 2019]. Dostępny w World Wide Web: <https://www.webarchive.org.uk/en/ukwa/info/nominate>.
- Selection* [online] [dostęp 17 marca 2019]. Dostępny w World Wide Web: <https://www.kb.nl/en/organisation/research-expertise/long-term-usability-of-digital-resources/web-archiving/selection>.
- Topics and Themes* [online] [dostęp 17 marca 2019]. Dostępny w World Wide Web: <https://www.webarchive.org.uk/en/ukwa/collection>.
- Web archive of the National Library of Ireland, [electronic resource] / 2011 – compiled by the National Library of Ireland* [online] [dostęp 11 marca 2019]. Dostępny w World Wide Web: <http://catalogue.nli.ie/Collection/vtls000510774>.
- Web Archive Singapore* [online] [dostęp 27 lutego 2019]. Dostępny w World Wide Web: <http://eresources.nlb.gov.sg/webarchives/landing-page>.
- Wilkowski Marcin, *Archiwum Webu jako black box* [online] [dostęp 17 marca 2019]. Dostępny w World Wide Web: <https://webarch.uw.edu.pl/blog/archiwum-webu-jako-black-box/>.

Preserving the diversity of World Wide Web in Web archives

ABSTRACT: The Web is characterized by an extraordinary diversity, both among its users, as well as its forms and content. In its thirty-year history it has become a very popular and important element of the modern world, which increases interest among researchers in various fields. Despite the fact that more and more information is being published on the Web, a large part of it is being turned into another or irretrievably lost. The remedy for this may be the Web Archive, deve-

loped since the mid-1990s, which allows the collection and maintenance of the Network's resources. The article discusses the extent to which the used methods, solutions and strategies will enable securing diversity occurring in the Web.

KEYWORDS: Web archiving, Web archives, preserving cultural heritage.

