www.czasopisma.pan.pl PAN www.journals.pan.pl

HISTORYKA. Studia Metodologiczne
T. 51, 2021 PL ISSN 0073-277X
DOI 10.24425/hsm.2021.138366
s. 83–110

Marcin Wilkowski
Instytut Badań Literackich PAN
ORCID: 0000-0003-2924-268X

STRONA INTERNETOWA JAKO ŹRÓDŁO HISTORYCZNE

THE WEB PAGE AS A HISTORICAL SOURCE

Abstract

A web page can be a useful object of analysis in historical research on the World Wide Web, but as a historical source, it does not have to be interpreted solely by its textual and visual strata. The article proposes an inclusive definition of the web page which ignores its visual content and relies on its non-obvious elements (HTTP headers and response type, URI identifier), which can be successfully used in historical study. The modular nature of the web page is the cause of many problems in building its chronology or accessing archival versions; on the other hand, it makes it possible to gain new information about the past reality. The development of Web historiography is a condition for building the historicity of the medium referred to as the"eternal now" Such historiography relies strongly on software as a tool for producing historical sources and sometimes needs to explore new time dimensions like the *stream* or *liveness*. The article explores these issues by referencing media theory and web archiving works as well as several research projects from the field of the digital humanities. Can a historiographical approach to sources inspire criticism on digital artefacts and data outside the field?

Keywords: World Wide Web, historiography, sources, web archives, data

Słowa kluczowe: World Wide Web, historiografia, źródła, archiwa Webu, dane

Niniejszy artykuł podejmuje problem wykorzystania stron internetowych jako źródła w badaniach historycznych. Po krótkim przedstawieniu konstruktywistycznego charakteru źródła historycznego przywołuję praktyki korzystania z za-

www.czasopisma.pan.pl PAN www.journals.pan.pl

sobów World Wide Web (WWW[1]) w badaniach historycznych i wskazuję, że Web może być nie tylko narzędziem, ale też przedmiotem badania historycznego, wyróżniając za Nielsem Brüggerem kierunek „history with the Web" i „history on the Web". Następnie za Lvem Manovichem przywołuję modularny charakter strony internetowej i proponuję jej definicję ignorującą tekstowe czy wizualne właściwości. Wskazuję, że uniwersalne identyfikatory, nagłówki i kod źródłowy mogą stać się wartościowym nośnikiem informacji o przeszłości i podstawą analiz historycznych. Wykorzystanie ich w badaniu wymaga jednak pośrednictwa oprogramowania, dlatego opisuję krótko jego znaczenie w wytwarzaniu źródeł historycznych. Następnie zastanawiam się, gdzie – wciąż poza tekstem – można szukać znaczeń w obrębie strony internetowej. Znajduję je między innymi w nowych warstwach wizualnych, rozpoznawanych maszynowo w oderwaniu od bezpośredniej ludzkiej percepcji, oraz w warstwie materialnej. Podkreślam, że WWW, konceptualizowana już nie jako sieć statycznych publikacji, ale strumień (danych i treści), wymaga stosowania nowych chronologii, a archiwizacje Webu i rejestrowanie zmian są niezbędne do podejmowania badań historycznych w tym medium. Na koniec zastanawiam się, czy nauka historyczna nie może odkryć swojej „aktualności" w tworzeniu warstwy historycznej w Webie oraz udostępnianiu metod poprawnej interpretacji historycznych wersji jego zasobów. Bez wątpienia badania historyczne „with the Web" / „on the Web" wymagają fundamentalnej pracy związanej z odkrywaniem/konstruowaniem nowych źródeł.

Źródło: metafora i praktyka

„Dowolny stan rzeczy staje się źródłem historycznym wtedy, gdy sytuuje się w orbicie poznania historycznego, zajmuje odpowiednie miejsce w procedurze badawczej i to wtedy tylko, gdy owo poznanie zakłada odwołanie się do swoistej empirii"[2] – pisze Wojciech Wrzosek. Chociaż źródło historyczne pełni fundamentalną rolę w praktyce historycznej, dyskusja o jego istocie, podejmowana w ramach studiów historiozoficznych i teorii historii, jest interpretowaniem pewnej metafory[3]. Wydaje się, że w wewnątrz tych dyskusji utrzymuje się pewien konsensus: źródło nigdy nie ujawnia się gotowe i skończone, jest tworzone w trakcie badania, a nie dostępne *a priori* oraz nie daje bezpośredniego i pełnego dostępu do przeszłej rzeczywistości. To ostatnie ograniczenie nie

1 „World Wide Web – powszechnie nazywane WWW, W3 lub Sieć Web – jest połączonym systemem publicznych stron internetowych dostępnych przez Internet. Sieć Web nie jest taka sama jak Internet: Sieć jest tylko jedną z wielu aplikacji zbudowanych na szczycie Internetu", zob. „World Wide Web", WebMDN Web Docs, https://developer.mozilla.org/pl/docs/Glossary/World_Wide_Web (dostęp: 20.01.2021).

2 Wojciech Wrzosek, „Losy źródła historycznego (refleksje na marginesie idei R. G. Collingwooda)", *Rocznik Antropologii Historii* 1–2, 1 (2011): 412.

3 O metaforach źródła w humanistyce zob. Ewa Kraskowska, „Źródło [Słownik poetologiczny]", *Forum Poetyki* (2015).

www.czasopisma.pan.pl PAN www.journals.pan.pl

wynika jedynie z jego potencjalnych braków – źródło może być przecież dostępne jedynie we fragmentach, bez niezbędnego kontekstu, być wynikiem intencjonalnej manipulacji czy ograniczonego spojrzenia twórcy czy twórców. Źródło daje ograniczony dostęp do przeszłej rzeczywistości także dlatego, że rzeczywistość ta była nieskończenie złożona i źródło może być jedynie przejawem czy reprezentacją jakiejś jej części. Wadę tę dostrzec można nie tylko w źródłach narracyjnych, zanurzonych w niejednoznaczności języka naturalnego, ale też w danych liczbowych. Wojciech Wrzosek przytacza tu ciekawy przykład, jakim jest problem statusu poznawczego średniej ceny transakcyjnej na dany towar wyliczonej z dokumentów historycznych. W analizach historycznych rynku i konsumpcji średnia będzie źródłem ujawniającym pewną cechę obiektywnej rzeczywistości, ale dla innych badaczy i badaczek, zainteresowanych doświadczeniem jednostkowym, pewną interpretacją, a nawet fikcją badawczą[4]. Takie dyskusje o teorii źródła zachowują niespodziewaną aktualność wobec bieżącej dyskusji o neutralności danych wyliczanych i gromadzonych z rozmaitych systemów cyfrowych[5].

Wydaje się też, że konsensus wobec źródła bazuje także na zgodzie co do tego, że ma ono niematerialny charakter, jest „bytem pomyślanym"[6], chociaż interpretowanym niekiedy z materialnych nośników. Skoro tak, dodanie do zbioru źródeł strony internetowej nie powinno być problemem – nieistotny jest fakt, że istnieje ona jedynie jako efekt pewnej komunikacji zapośredniczonej przez oprogramowanie, a jej granice są bardzo płynne. Rozpoznania teoretyczne tego problemu potwierdza praktyka: strony WWW są archiwizowane przynajmniej od połowy lat dziewięćdziesiątych (projekt Wayback Machine fundacji Internet Archive) i wykorzystywane w historycznych badaniach naukowych. Wartość poznawczą stron WWW podkreślają też ogólne polityki wobec dziedzictwa cyfrowego, wyrażane czy to w przepisach o egzemplarzu obowiązkowym, uchwalanych w niektórych krajach, czy na poziomie ustaleń międzynarodowych – myślę tu choćby o stanowisku UNESCO[7].

Źródło jako pojęcie kluczowe dla poznania historycznego dyskutowane jest na poziomie historiozofii i teorii historii, ale w konkretnych pracach historycznych przyjmuje praktyczną postać i w tej postaci poddawane jest dokładnej krytyce, dostosowanej do jego nośnika, zgodnie ze specjalizacją badawczą i w odniesieniu do zadanych pytań badawczych. W takim właśnie praktycznym kontekście chciałbym w tym artykule umieścić stronę internetową.

---

4 Wojciech Wrzosek, „Źródło historyczne jako alibi realistyczne historyka", w *Historyk wobec źródeł. Historiografia klasyczna i nowe propozycje metodologiczne*, red. Jolanta Kolbuszewska, Rafał Stobiecki (Łódź: Wydawnictwo Ibidem, 2010), 34.

5 Zob. np. *«Raw data» is an oxymoron*, red. Lisa Gitelman (Cambridge: MIT Press, 2013); Ninareh Mehrabi et al., „A survey on bias and fairness in machine learning", arXiv:1908.09635 (2019).

6 Wrzosek, *Źródło historyczne*, 32.

7 UNESCO do podlegającego ochronie dziedzictwa cyfrowego włącza „oprogramowanie i witryny internetowe występujące w szerokim i wzrastającym zakresie formatów", zob. „Karta w sprawie zachowania dziedzictwa cyfrowego" UNESCO (2003), 2.

www.czasopisma.pan.pl PAN www.journals.pan.pl

Historiografia i World Wide Web

Strony internetowe, czy też WWW, jako zasób podlegający analizie wykorzystywany jest w wielu dyscyplinach badań społecznych i humanistycznych, nie mówiąc już o naukach informatycznych. Ponieważ Web jest medium, które zmienia się nieustannie, badania te są *de facto* badaniami Webu historycznego (przeszłego). Już w momencie pozyskania zasoby internetowe mogą różnić się od tych aktualnie dostępnych online (w tzw. *live Web*). Mówienie o stronie internetowej jako źródle historycznym jest możliwe jednak dopiero, kiedy zostaje ona wykorzystana w ramach badania historycznego. Mimo że badania historyczne Webu uwzględniają wiele metod i podejść innych dziedzin (medioznawstwa, kulturoznawstwa, socjologii, informatyki czy statystyki), perspektywa interdyscyplinarna nie może zaciemniać tego, co w pisarstwie historycznym najważniejsze, czyli – przywołuję tu program Erica Hobsbawma wyrażony w jednym z jego dzieł – zrozumienia i objaśnienia wybranej przeszłości oraz jej miejsca w dziejach, zrozumienia świata w procesie określonych przemian, ukazania korzeni naszej współczesności w pokładach przeszłości i przedstawienie przeszłości jako wspólnej całości[8].

Megan Sapnar Ankerson w artykule z 2012 roku opisuje historię technologii Flash[9], ilustrując proponowany przez siebie program uprawiania historiografii Webu. W krótkiej analizie stara się pokazać Flash jako proces społeczny oparty nie tylko na rozwiązaniach informatycznych. Szuka źródeł tej technologii w krytykach kapitalistycznej transformacji Webu z połowy lat dziewięćdziesiątych, by następnie stwierdzić, że w początkach pierwszej dekady XXI wieku to Flash stał się synonimem i estetyką komercyjnego WWW. Dalej autorka rozpoznaje w postępującym kryzysie popularności Flasha jako narzędzia tworzenia interfejsów webowych zjawiska zaburzające linearności historii – kryzys mainstreamu otwiera przestrzeń dla kobiet: dizajnerek i programistek dotąd słabo widocznych w społeczności twórców Flasha. Ankerson dostrzega też zmieniającą się rolę tej technologii wobec przyjęcia przez deweloperów nowych założeń dla użyteczności i dostępności stron internetowych oraz polityki wydawcy tego oprogramowania: z narzędzia do tworzenia i publikowania animacji Flash stał się przede wszystkim narzędziem prezentacji grafik i wideo[10]. Praca Ankerson,

---

8 Eric Hobsbawm, *Wiek imperium 1875–1914*, przeł. Marcin Starnawski (Warszawa: Wydawnictwo Krytyki Politycznej, 2015), 5.

9 Flash to oprogramowanie pozwalające na tworzenie interaktywnych, multimedialnych animacji z wykorzystaniem grafiki wektorowej. Upubliczniona w 1996 r. technologia ta po 2000 r. stopniowo traciła na znaczeniu ze względu na ograniczoną użyteczność i dostępność (*accessibility*). Możliwości Flasha przejmowała nowa wersja języka HTML oraz JavaScript – w 2015 r. odtwarzacz YouTube, działający dotąd na Flashu, został przebudowany z wykorzystaniem możliwości HTML 5. 31 grudnia 2020 r. Flash przestał być oficjalnie wspierany przez producenta. W momencie publikacji artykułu Ankerson był jednak wciąż pełnoprawną, chociaż już mocno krytykowaną i przestarzałą, technologią webową.

10 Megan Sapnar Ankerson, „Writing web histories with an eye on the analog past", *New Media and Society* 14, 3 (2012): 392–396.

www.czasopisma.pan.pl PAN www.journals.pan.pl

chociaż programowa i odsłaniająca niewielki wycinek historii Webu, wydaje się wypełniać przytoczone przeze mnie wcześniej założenie Hobsbawna. To przykład historii Webu dalekiej od technologicznego determinizmu, dostrzegającej nieoczywiste relacje i odkrywającej nowych aktorów (aktorki). Wzrost i upadek popularności Flasha nie jest tu najważniejszy – liczą się zmiany w wizualności Webu i tego, jak rozumiano atrakcyjność i użyteczność stron internetowych. Ważny jest także kontekst komercjalizacji tego medium. Autorka nawiązuje do tradycji badań nad historią radia i telewizji, wskazując, że problem efemeryczności źródeł był już w historiografii podejmowany i korzystanie z nich w badaniach Webu nie jest w żaden sposób nowatorskie.

Wobec Webu i współczesnego dyskursu na temat tego medium podejście historyczne ma w pewnym sensie charakter rewolucyjny i kontrkulturowy. Protokół HTTP[11], za pomocą którego udostępniane są strony WWW, sam w sobie nie posiada żadnych mechanizmów archiwizacyjnych ułatwiających sięganie do historycznych zasobów. Dodatkowo Internet społecznie funkcjonuje w paradygmacie „wiecznego teraz"[12], kolejnych „rewolucji" technicznych i globalnych procesów rynkowych, opisywanych metaforami wersjonowania takimi jak Web 2.0[13]. Budowanie historii Webu – oczywiście „historii" w liczbie mnogiej – mogłoby być przejawem poszukiwanej przez Ewę Domańską nowej *aktualności* dyscypliny historycznej wobec wyzwań współczesnego świata[14]. Przykładowo, historiografia polskiego Webu mogłaby stać się narzędziem wspierającym nowe spojrzenie na przełom lat dziewięćdziesiątych, wykazując rozmaite postawy wobec jednokierunkowego transferu technologii (Zachód–Polska) i odkrywając podmiotowość technologiczną Polski tego okresu[15]. Oczywiście historiografia taka wymaga przesunięcia granicy, „za którą «koń-

---

[11] HTTP (*Hypertext Transfer Protocol*) to protokół odpowiedzialny za transmisję dokumentów hipermedialnych (modularnych) takich jak np. HTML. Został stworzony do komunikacji między przeglądarkami a serwerami webowymi, ale może być używany również w innych celach. HTTP opiera się na klasycznym modelu klient-serwer, gdzie klient inicjuje połączenie przez wysłanie żądania, następnie czeka na odpowiedź. Zob. MDN Web Docs, „HTTP", https://developer.mozilla.org/pl/docs/Web/HTTP (dostęp: 10.01.2021).

[12] Według Joanne Garde-Hansen Facebook jako medium funkcjonuje w dominującym w kulturze cyfrowej micie młodości i wolności od historii, zob. Joanne Garde-Hansen, „MyMemories? Personal Digital Archive Fever and Facebook", w *Save As... Digital Memories*, red. Joanne Garde-Hansen, Andrew Hoskins, Anna Reading (New York: Springer, 2009), 141.

[13] Matthew Allen, „What was Web 2.0? Versions as the dominant mode of internet history", *New Media and Society* 15, 2 (2012): 260–275.

[14] Ewa Domańska, „Perspektywy badań historycznych w Polsce wobec zmian we współczesnej humanistyce", w *Historyk wobec źródeł. Historiografia klasyczna i nowe propozycje metodologiczne*, red. Jolanta Kolbuszewska, Rafał Stobiecki (Łódź: Wydawnictwo Ibidem, 2010), 115–127.

[15] Myślę tu o perspektywie proponowanej w pracach takich jak Piotr Marecki, Tadeusz Cieślewicz, *Oni migają tymi kolorami w sposób profesjonalny. Narodziny gamedevu z ducha demosceny w Polsce* (Kraków: Ha!art, 2020) czy Mirosław Filiciak, „Gra w kapitalizm. PRL, konstruowanie pamięci i gry cyfrowe", *Widok. Teorie i praktyki kultury wizualnej* 11 (2015): 1–30. Problem cyfrowego czy też technologicznego kolonializmu rozpoznać można zarówno w transferze technologii i praw autorskich, jak i w kierunkach

www.czasopisma.pan.pl PAN www.journals.pan.pl

czy» się historia"[16] i zmusza do poruszania się w okresie ostatnich trzydziestu lat, gdzie cezurę początkową wyznacza okres budowania podstaw World Wide Web (1989–1991). Wymaga także fundamentalnej pracy związanej z odkrywaniem/konstruowaniem nowych źródeł.

Historycy, podobnie jak przedstawiciele innych dyscyplin, od lat z powodzeniem wykorzystują Internet jako wsparcie badań i narzędzie komunikacji naukowej. Digitalizacja zbiorów archiwalnych czy bibliotecznych, artykuły i monografie dostępne online (nawet jeśli nie zawsze w sposób oficjalny), specjalistyczne bazy danych, mapy cyfrowe itp. są dziś częścią infrastruktury nauk historycznych oraz podlegają dziedzinowej krytyce[17]. Literaturę na temat użycia sieci WWW w badaniach historycznych śledzić można od połowy lat dziewięćdziesiątych także w Polsce – to teksty opisujące Internet wykorzystywany jako repozytorium zdigitalizowanych dokumentów czy medium komunikacji naukowej, dydaktyki i upowszechniania literatury naukowej oraz zdigitalizowanych źródeł[18]. To praktyka „History Web", którą odróżnić należy od praktyk „Web history" będących efektem przyjęcia, że – jak pisze Niels Brügger – przeszłość Webu sama w sobie jest warta studiowania. WWW jako przedmiot badań może być przy tym wykorzystany do wyjaśniania czy ilustrowania określonych wydarzeń czy zjawisk spoza tego medium (u Brüggera – „history with the Web") lub po prostu stać się przedmiotem badań („history of the Web"), które niekoniecznie posiłkować się muszą wyłącznie źródłami cyfrowymi, ale także na przykład prasą czy nawet wywiadami historii mówionej[19]. Uprawianie historii Webu wymaga jednak nowej, odpowiedniej krytyki źródeł wytworzonych cyfrowo (*born-digital*), a więc takich, które nie posiadają analogowych, materialnych wersji pierwotnych, oraz korzystania z archiwów Webu, w których źródła takie podlegają dalszemu przekształceniu, co zresztą Brügger podkreśla, wprowadzając pojęcie źródeł odrodzonych cyfrowo (*re-born digital*)[20]. Dla Brüggera zarchiwizowana strona internetowa jest zupełnie no-

---

dystrybucji treści, zob. np. David M. Haug, „International Transfer of Technology: Lessons that East Europe can learn from the failed Third World experience", *Harvard Journal of Law & Technology* 5, 209 (1992): 209–240; Monika Stobiecka, „Archaeological heritage in the age of digital colonialism", *Archaeological Dialogues* 27, 2 (2020): 113–125.

16 Andrzej Paczkowski, „Historyk dziejów najnowszych wśród źródeł", w *Historia-dziś. Teoretyczne problemy wiedzy o przeszłości*, red. Ewa Domańska, Rafał Stobiecki, Tomasz Wiślicz (Kraków: Universitas, 2014), 89.

17 W kontekście digitalizacji zbiorów archiwalnych i bibliotecznych krytyka ta dotyczy m.in. relacji między obiektem materialnym a zdigitalizowanym, kosztów poznawczych takiego transferu, niewidzialnej pracy (*invisible labor*) związanej z procesem digitalizacji czy dominacji zbiorów dziedzictwa z instytucji anglosaskich.

18 Zob. np. serię „Historia i komputery" redagowaną przez Bohdana Ryszewskiego (wydania z 1995 i 1997 roku). W tym okresie Internet nie był jeszcze oczywistością jako medium publikowania naukowego – edycje cyfrowe źródeł planuje się publikować przede wszystkim na płytach CD lub udostępniać na terminalach bibliotecznych.

19 Niels Brügger, *The Archived Web. Doing History in the Digital Age* (Cambridge: MIT Press, 2018), 14–16.

20 Brügger, *Archived Web*, 32.

www.czasopisma.pan.pl PAN www.journals.pan.pl

wym typem dokumentu historycznego, ponieważ proces archiwizacji odbija się tu w samym artefakcie – nie mamy do czynienia z kopią, ale z zupełnie nowym obiektem analitycznym[21]. Na problem ten zwracają uwagę także Emily Maemura i współautorzy, proponując nawet listę pytań kontrolnych ułatwiających krytykę proweniencji kolekcji udostępnianych przez archiwa Webu[22].

Strona internetowa jako obiekt modularny

Aby rozważyć temat strony internetowej jako potencjalnego źródła historycznego, należy zaproponować jakąś jej definicję oraz wykazać, że zasadnie byłoby ją – w kontekście badań historycznych – wyróżnić spoza innych obiektów webowych jako obiekt podlegający badaniu. W słowniku informatycznym publikowanym przez Fundację Mozilli i przeznaczonym dla początkujących użytkowników Webu strona internetowa definiowana jest po prostu jako napisany w języku HTML[23] dokument, który może być wyświetlony w przeglądarce internetowej. Dokument taki może zawierać (*embed*) w sobie multimedia (grafikę, filmy itp.) oraz pisane w innych językach niż HTML kody określające estetykę strony i mechanizmy definiujące jej interakcję z użytkownikiem[24].

Planowanie badania historycznego z wykorzystaniem takiego obiektu napotyka od razu na podstawową trudność, ponieważ strona internetowa jest pojedynczym dokumentem HTML oraz jednocześnie siecią powiązanych ze sobą obiektów (grafik, skryptów). Cechę tę za Lvem Manovichem określić można jako modularność[25]. Modularność strony internetowej utrudnia także jej jednoznaczną definicję – trudno arbitralnie wyznaczyć granice takiego obiektu, chociaż z pewnością nie ma to większego wpływu na praktykę badań: granice wyznaczają wówczas pytania badawcze i zastosowane metody. Niekiedy sprawiają one, że strona jako wyróżniony element znika. Jak pisze Brügger, strona internetowa jest często nierozpoznawalną jednostką analityczną, ponieważ stanowi „niewidzialne" tło budujących ją elementów[26]. Historyczne źródło webowe nie istnieje *a priori*, lecz jest konstruowane.

---

21 Niels Brügger, „The Archived Website and Website Philology A New Type of Historical Document?", *Nordicom Review* 29, 2 (2008): 156.

22 Emily Maemura et al., „If these crawls could talk: Studying and documenting web archives provenance", *Journal of the Association for Information Science and Technology* 69, 10 (2018): 1230.

23 HTML (ang. *HyperText Markup Language*, hipertekstowy język znaczników) to język pozwalający na strukturyzowanie dokumentów tekstowych za pomocą określonych znaczników semantycznych. Struktura ta jest następnie interpretowana przez przeglądarkę podczas wyświetlania strony WWW.

24 MDN Web Docs, „What is the difference between webpage, website, web server, and search engine?", https://developer.mozilla.org/en-US/docs/Learn/Common_questions/Pages_sites_-servers_and_search_engines (dostęp: 5.01.2021).

25 Lev Manovich, *Język nowych mediów*, przeł. Piotr Cypryański (Warszawa: Wydawnictwa Akademickie i Profesjonalne, 2006), 96.

26 Brügger, *Archived Web*, 47.

www.czasopisma.pan.pl PAN www.journals.pan.pl

Oczywiście teoria nieograniczana szczegółowymi celami i metodami konkretnego badania musi jednak sobie poradzić z problemem odpowiedniego ujęcia tego, czym – w znaczeniu historycznym – jest strona WWW. Celem takiej pracy jest przede wszystkim rozpoznanie tego, co w stronie internetowej może być nośnikiem znaczeń i podlegać analizie. Brügger rozpoznaje w niej jednostki znaczeniowe (*textual units*) bazujące na czterech formach ekspresji: tekst, statyczne i dynamiczne obrazy (wideo) oraz dźwięki. Dodaje przy tym, że każda taka jednostka może być zbudowana z mniejszych jednostek – na przykład opublikowany na stronie tekst może składać się z paragrafów[27]. W późniejszym opracowaniu Brüggera te jednostki znaczeniowe stają się jedną z wizualnych warstw Webu (*visible Web strata*) jako elementy Webu (*web elements)*. Konstytuują one warstwę strony internetowej (*web page*), na której rozwija się warstwa witryny (a więc sieci połączonych ze sobą stron). Witryny łączą się w sfery webowe (*web sphere*s), sieci witryn wyodrębniane tematycznie, chronologicznie lub pod kątem domeny najwyższego poziomu. Najwyższą warstwą jest *Web as a whole*[28].

Brügger kopiuje ten schemat, wyodrębniając strukturę ukrytych warstw analitycznych Webu (*hidden Web strata*). Mamy tutaj do czynienia również z pięcioma poziomami analitycznymi, przy czym nośnikiem znaczeń są w nich kod źródłowy, techniczne informacje o plikach, oprogramowaniu użytym do zbudowania i publikowania strony czy o domenie[29]. W takim ujęciu badanie historyczne strony internetowej polegać może na analizie widocznej w przeglądarce treści (np. tekstu) i/lub jej kodu źródłowego (np. linków do innych stron) – wydaje się, że jest to strategia pozwalająca dobrze uchwycić modularny charakter strony WWW oraz dostrzec znaczenia budowane w inny sposób niż przez tekst.

Hayden White dostrzegł, że „zbyt często historycy traktują dane fotograficzne, kinematograficzne i video tak, jak gdyby można je było czytać tak, jak się czyta dokumenty pisane”[30], przyznając wcześniej, że „świadectwa (*evidence*) historyczne wytwarzane w naszej epoce są równie często z natury wizualne, jak oralne i pisane”[31]. Przywoływanie „natury” źródeł może być ryzykowne, skoro dokumenty tekstowe mogą być analizowane wizualnie (choćby typograficznie) lub stylometrycznie, a autentyczność historycznych fotografii cyfrowych, także tych powstałych w wyniku skanowania, można badać bez odwołania do wizualności, choćby poprzez analizy wykorzystujące rozkład Benforda. „Natura” źródeł jest u White’a wyznaczana przez ludzką percepcję i jako taka może być łatwo unieważniona choćby przez wykorzystanie oprogramowania w ich badaniu (należy przy tym zaznaczyć, że tekst White’a został opublikowany w 1988 r.). Pomimo tego jego uwaga nie traci na aktualności. Bez wątpienia

---

27 Niels Brügger, „Website history and the website as an object of study”, *New Media and Society* 11, 1–2 (2009): 121.

28 Brügger, *Archived Web*, 33–35.

29 Brügger, *Archived Web*, 35–37.

30 Hayden White, „Historiografia i historiofotia”, przeł. Łukasz Zaremba, w *Film i historia*, red. Iwona Kurz (Warszawa: Wydawnictwa Uniwersytetu Warszawskiego, 2008), 118.

31 White, *Historiografia i historiofotia*, 117.

www.czasopisma.pan.pl PAN www.journals.pan.pl

istnieje niebezpieczeństwo rozumienia i wyjaśniania nowomedialnych źródeł przez odwoływanie się do tradycyjnych ujęć i metod, przede wszystkim „przez tekst". Przyczyny tego stanu szukać można nie tylko w tym, że historia jako dyscyplina wiedzy od wieków zorientowana jest na tekst, ale też w silnym wpływie kultury tekstu i druku na kulturę komputerową[32].

Interfejsy komputerowe oraz język z nimi związany korzystają obszernie z imaginacji, pojęć, form i gatunków pochodzących z mediów przedcyfrowych, zgodnie z zasadą remediacji, o której piszą Jay David Bolter i Richard Grusin[33]. Web nie jest autonomiczny wobec kultury druku, narzuciła ona wiele metafor na to, jak wyróżniamy i nazywamy jego natywne obiekty. Na co dzień korzystamy z cyfrowych biurek i pulpitów (*desktop*) w systemach operacyjnych, porządkujemy zasoby cyfrowe wizualizowane jako pliki umieszczane w katalogach reprezentowanych ikonami nawiązującymi do teczek dokumentowych. Korzystając z Webu, korzystamy ze *stron*, które dodatkowo mogą mieć postać przewijalnego z góry na dół zwoju, a pierwsze formy publikowania online wprost nawiązywały do typów biurowych dokumentów[34]. Oczywiście historycznie doświadczenie Webu jako medium tekstowego zmienia się wobec rozwoju nowych technologii webowych. „Dokumentowe" strony z pierwszych lat WWW w połowie lat dziewięćdziesiątych ustępują miejsca kolorowym, multimedialnym prezentacjom dzięki nowym znacznikom HTML i przeglądarkom pozwalającym na ich obsługiwanie, aby dziś przyjmować postać aplikacji jako mechanizmów przetwarzających dane i z powodzeniem funkcjonować na urządzeniach mobilnych, a więc poza interfejsem komputera i tworzącymi go metaforami druku. Jednak nawet wówczas, kiedy tekst wydaje się znikać z ekranu przeglądarki czy telefonu komórkowego, jest tam wciąż obecny w postaci kodu źródłowego strony. Mimo tego definicja strony internetowej jako źródła w badaniu historycznym powinna unikać przypisywania jej „natury tekstowej", ponieważ mogłoby wpływać na wybór zestawu metod jej badania. Kiedy mamy do czynienia z tekstem, myślimy tekstem i szukamy metod, które pozwalają badać tekst. Strona internetowa nie jest jednak dokumentem tekstowym, a przynajmniej nie jest jedynie dokumentem tekstowym.

Propozycja definicji

Wydaje się zatem, że fundamentalnymi elementami definicji strony WWW byłyby 1) stwierdzenie, że jest to poprawna odpowiedź serwera na żądanie klienta (np. przeglądarki WWW) wysyłana za pomocą protokołu HTTP wraz

---

32 Warto zwrócić uwagę, że w przytoczonym wyżej poradniku Mozilli umieszczono fragment porównujący serwer z tradycyjną biblioteką. Strona internetowa ma być tutaj odpowiednikiem… książki, a witryna internetowa – działu gromadzącego książki na określony temat.

33 Jay David Bolter, Richard Grusin, *Remediation: Understanding New Media* (Cambridge: MIT Press, 2000).

34 Kevin Crowston, Marie Williams, „Reproduced and Emergent Genres of Communication on the World Wide Web", *The Information Society* 16, 3 (2000): 210–215.

z nagłówkami (*headers*), 2) odpowiedź wysyłana jest w formacie text/html czyli w postaci ustrukturyzowanego tekstu (HTML) wraz ze standardowymi nagłówkami[35] oraz 3) posiada adres URI (*Uniform Resource Identifier*, ujednolicony identyfikator zasobów[36]), unikalny identyfikator będący warunkiem jej „wyodrębnienia" jako bytu webowego. Jak pisze Michalis Vafopoulos, byty webowe to byty, które mogą być komunikowane przez WWW. Adres (identyfikator) URI jest minimalnym opisem niezmiennych elementów takiej komunikacji. URI buduje także nawigowalną przestrzeń Webu jako medium – dzięki temu identyfikatorowi można nie tylko wczytywać do przeglądarki określone zasoby, ale też je łączyć ze sobą (sieć linków) oraz komunikować na zewnątrz, na przykład w przypisach pracy naukowej[37]. Wspólna dla zestawu stron część URL może być także cechą, na której budowany będzie zbiór poddawany badaniu – będzie wyznaczała granice witryny lub strefy webowej. Element komunikacji pomiędzy serwerem a klientem, wspomniany w zaproponowanej definicji, wykorzystują także Kristen Foot i Steven Schneider. Według nich strona internetowa to „grupa elementów zestawionych przez twórcę i wyświetlonych na żądanie przez serwer"[38]. Potencjał URI jako wyznacznika unikalnego bytu webowego jest podważany przez proces personalizacji Webu, przez który treść strony WWW dostępnej pod konkretnym identyfikatorem może być uzależniona od tego, jaki użytkownik z niej korzysta[39].

Zaproponowana definicja unika odwołań do cech gatunkowych czy estetycznych oraz odrzuca metaforę tekstową narzucaną przez rzeczownik „strona". Akcentuje ona też mechanizm komunikacji między serwerem a klientem, niezbędny dla istnienia strony WWW jako dynamicznie konstruowanego obiektu. Nawet jeśli użytkownik nie ma żadnego wpływu na jej treść, w pewnym sensie stwarza ją czy kreuje przez wybranie odpowiedniego adresu URI – bez tego strona WWW pozostaje zestawem elementów przechowywanych na serwerze i informacji zapisanych w bazie danych. Z drugiej strony pewną jej wadą jest ignorowanie popularnej w latach dziewięćdziesiątych praktyki dystrybucji witryn internetowych na nośnikach takich jak dyskietki czy płyty

---

35 Nagłówki to ustandaryzowane zestawy metadanych wysyłane do serwera przez przeglądarkę przy żądaniu strony internetowej oraz w odpowiedzi do przeglądarki przez serwer. Nagłówki informują m.in. o typie dokumentu, statusie odpowiedzi serwera, standardzie formatowania treści czy wielkości odpowiedzi, zob. MDN Web Docs, „HTTP headers", https://developer.mozilla.org/en-US/docs/Web/HTTP/Headers (dostęp: 5.01.2021).

36 W powszechnym użyciu jest także akronim URL (*Uniform Resource Locator*), różniący się od URI elementem wskazującym protokół dostępu do zasobu internetowego. W praktyce publikowania webowego nie ma różnicy między URI a URL.

37 Michalis Vafopoulos, „Being, Space, and Time on the Web", *Metaphilosophy* 43, 4 (2012): 407.

38 Kirsten Foot, Steven Schneider, „Object-oriented Web Historiography", w *Web History*, red. Niels Brügger (New York: Peter Lang Inc, 2010), 69.

39 Dobrymi przykładami są tutaj URI strony głównej serwisu Facebook (https://facebook.com) czy Twitter (https://twitter.com/home) albo wyniki wyszukiwania w wyszukiwarce Google dla wybranej frazy. Dla każdego użytkownika treść tej strony będzie inna.

www.czasopisma.pan.pl PAN www.journals.pan.pl

CD, a więc *offline*[40]. Równocześnie strona internetowa jako zdefiniowany obiekt badań może być w konkretnej analizie grupowana w witryny czy sfery webowe lub stanowić nośnik mniejszych elementów znaczeniowych (np. grafiki, tekstu czy nagłówków HTTP), które mogą być następnie analizowane historycznie.

Gdzie szukać znaczeń: URI i nagłówki

Twórca WWW, Tim Berners-Lee, pisał, że „Sieć jest przestrzenią informacyjną. URI są punktami w tej przestrzeni"[41]. URI jest cechą strony internetowej pozwalającą odróżnić ją od innych i badać takie jej właściwości, jak warstwa semantyczna, nagłówki czy też powiązania strony w ramach sieci innych stron i witryn. URI warunkuje także wiarygodność (autentyczność) elementów strony i umożliwia określanie ich proweniencji.

Przykładowo, w pracy „Exploring the Memory of the First World War Using Web Archives: Web Graphs Seen from Different Angles" Valérie Beaudouin i współautorzy wykorzystują URI analizowanych przez siebie witryn (czyli grup stron internetowych) do wygenerowania sieci relacji i zanalizowania źródeł pozwalających opisać konstruowanie pamięci I wojny światowej online[42]. Ponad 9,6 mln URI jest tu nie tylko kluczem do wyróżnienia poddawanych badaniu zasobów, ale także cechą budującą sieć połączeń między witrynami. Wspomniane badanie ignoruje treści pojedynczych stron oraz wykorzystuje metody wywiadu socjologicznego do znalezienia kontekstu powstania i funkcjonowania amatorskich witryn poświęconych I wojnie światowej. W pracy Federico Nanniego poświęconej historii obecności Uniwersytetu Bolońskiego w Internecie adresy URI służą nie tylko do wyznaczenia zakresu materiału poddawanego analizie, ale określają również struktury informacyjne historycznej witryny unibo.it i umożliwiają znalezienie informacji o tej witrynie w innych miejscach Webu[43]. W pracy Brüggera i współautorów źródłami pozwalającymi opisać historię duńskiego Webu są adresy URI oraz dane wyliczane z analizowanych zarchiwizowanych witryn takie jak liczba stron i plików w witrynie czy liczba linków wychodzących poza domenę .dk i do mediów społecznościowych

---

[40] Dystrybucja taka odbywa się poza protokołem HTTP i bez URI. Wydaje się jednak, że w takich przypadkach definicję można elastycznie dopasować do koncepcji badania.

[41] Tim Berners-Lee, Dan Connolly, „Naming and Addressing", W3C Architecture Domain (1993–2006), https://www.w3.org/Addressing/ (dostęp: 5.01.2021).

[42] Valérie Beaudouin, Zeynep Pehlivan, Peter Stirling, „Exploring the Memory of the First World War Using Web Archives: Web Graphs Seen from Different Angles", w *The SAGE Handbook of Web History*, red. Niels Brügger, Ian Milligan (Los Angeles: SAGE Publications, 2018), 441–463.

[43] Federico Nanni, „Collecting Primary Sources from the Web Archives: A Tale of Scarcity and Abundance", w *The SAGE Handbook of Web History*, red. Niels Brügger, Ian Milligan (Los Angeles: SAGE Publications, 2018), 116–118.

www.czasopisma.pan.pl PAN www.journals.pan.pl

– treść poszczególnych stron nie ma tutaj żadnego znaczenia[44]. W tym przypadku odpowiednia kontekstualizacja zebranych danych pozwala przygotować interesujące opracowanie historyczne i jest ono tylko pozornie odmienne od klasycznej historiografii interpretującej źródła statystyczne[45]. Wspomniany w definicji strony internetowej nagłówek określający typ dokumentu zwracanego przez przeglądarkę[46] również może zostać użyty w charakterze źródłowym. João Miranda i Daniel Gomez w analizie archiwów portugalskiego Webu biorą pod uwagę rodzaje publikowanych w nim obiektów, a porównując dwie kwerendy wykonane w odstępie pięciu lat, są w stanie wykazać interesujące zmiany sygnalizujące nowe praktyki publikowania i korzystania z Webu (np. zwiększenie obecności online materiałów audiowizualnych). Analizując nagłówki, autorzy tworzą także historyczne statystyki odpowiedzi serwera oraz badają zmienność w średnich długościach adresów URL, która ich zdaniem jest świadectwem zmian w projektowaniu interakcji w Webie[47].

Wspomniane wyżej badania pokazują potencjał URI jako właściwości strony WWW pozwalającej badać jej relacje z innymi stronami (Beaudouin), analizować strukturę udostępnianych przez sieć stron informacji (Nanni) czy wykazywać historyczne zmiany w zasobach domeny krajowej (Brügger, Miranda), abstrahując od treści zawartych na poszczególnych stronach. Ignorowanie treści źródeł nie jest przy tym niczym nowym w praktyce historiograficznej. Niekiedy źródłem historycznym staje się przecież nie tyle warstwa semantyczna papierowych dokumentów, ale przyrastająca lub zmieniająca się ich liczba w czasie lub statystyka ich proweniencji. Kryzys gospodarczy wykazywać można studiując nagłówki gazetowe, ale też analizując przyrost liczby osób zarejestrowanych przez urzędy pracy czy nawet dane policyjne wskazujące na wzrost przestępczości. Przykładowo, Marcin Zaremba, przedstawiając problem powojennego bezrobocia, cytuje skargi wysyłane do organów Ministerstwa Bezpieczeństwa Państwowego, komentarze prasowe *Dziennika Powszechnego*, sprawozdania urzędów, dzienniki oraz zestawienia liczby bezrobotnych rejestrowanych w Urzędzie Zatrudnienia w Kielcach[48].

---

44 Niels Brügger, Jane Nielsen, Ditte Laursen, „Big data experiments with the archived Web: Methodological reflections on studying the development of a nation's Web", *First Monday* 25, 3 (2020). https://firstmonday.org/ojs/index.php/fm/article/view/10384.

45 Zob. np. Maria Bogucka, „Zboże rosyjskie na rynku amsterdamskim w pierwszej połowie XVII wieku", *Przegląd Historyczny* 53, 4 (1962).

46 MDN Web Docs, „Content-Type", https://developer.mozilla.org/en-US/docs/Web/HTTP/Headers/Content-Type (dostęp 5.01.2021).

47 João Miranda, Daniel Gomes, „Trends in Web characteristics", w *Proceedings of the Latin American Web Congress (LA-Web)*, red. Edgar Chávez, Elizabeth S. Furtado, Alberto L. Morán (Los Alamitos: IEEE Computer Society, 2009), 148–150.

48 Marcin Zaremba, *Wielka Trwoga. Polska 1944–1947* (Kraków-Warszawa: Wydawnictwo Znak, Instytut Studiów Politycznych PAN, 2012), 233–246.

www.czasopisma.pan.pl PAN www.journals.pan.pl

Gdzie szukać znaczeń: kod źródłowy

Historyk chcący sięgnąć bezpośrednio do kodu źródłowego jako źródła historycznego i szukający w nim pewnych chronologii i znaczeń z powodzeniem wykorzystać może podejścia i metody proponowane przez takie kierunki jak studia nad oprogramowaniem (*software studies*) czy krytyczne studia nad kodem źródłowym (*critical code studies*). Przedmiotem analizy historycznej mogą być nie tylko instrukcje wysyłane do komputera (algorytm), ale też język, za pomocą którego są one wyrażone. Analizą objąć może nie tylko procesy definiowane przez konkretne funkcje kodu, ale też jego specyficzną literackość, choćby w ujęciu proponowanym przez Annette Vee[49]. W humanistycznym podejściu do oprogramowania kod źródłowy niesie ze sobą nie tylko określone instrukcje informatyczne, ale też odbija kontekst kulturowy. Nickowi Montfortowi i współautorom udało się przygotować monografię jednej linii kodu źródłowego w języku BASIC, interpretując generowanie znaków w kolejnych liniach ekranu, znaczenie spacji czy nazw poleceń (PRINT, GOTO), budując odwołania historyczne (do kultury druku, mitologicznego motywu labiryntu czy eksperymentów artystycznych bazujących na losowości). Jak czytamy we wstępie do tej monografii, „kod źródłowy jest osadzony w historii tworzenia programu, jego celach, założeniach i wielu innych. Każdy symbol w kodzie źródłowym może pomóc w wyróżnieniu tych historii i uruchomieniu nowych historycznych i krytycznych badań”[50].

Historyczno-kulturowe tło powstania i ewolucji tworzącego kod źródłowy strony WWW języka HTML również może być wykorzystane w badaniach historycznych Webu. Przykładowo, definiowane w kodzie źródłowym linki można umieścić w długiej tradycji hipertekstu oraz – co proponuje Anne Helmond – dostrzegać historyczną zmienność tego, czym dla Webu było linkowanie. Link prowadzący ze strony do strony jest więc najpierw traktowany jako podstawowy element struktury sieci WWW (wzajemne wiązanie ze sobą informacji), następnie jako narzędzie wymiany i komercjalizacji (link jako polecenie konkretnej treści, płatne linki itp.), ostatecznie jako element ukrytego mechanizmu aplikacji webowych. Kategoryzacja funkcji linków pozwala Helmond opisać historię przemian Webu[51].

Kod źródłowy, a zwłaszcza język programowania jako pewna formalna i logiczna struktura, wydaje się skrajnie racjonalny i techniczny. Nie musi taki jednak

---

[49] Annette Vee, *Coding Literacy: How Computer Programming Is Changing Writing* (Cambridge: MIT Press, 2017).

[50] Nick Montfort et al., *10 PRINT CHR$(205.5+RND(1)); : GOTO 10* (Cambridge: MIT Press, 2012), 3.

[51] Anne Helmond, „A historiography of the hyperlink: Periodizing the Web through the changing role of the hyperlink”, w *The SAGE Handbook of Web History*, red. Niels Brügger, Ian Milligan (Los Angeles: SAGE Publications, 2018), 227–241. Hipertekst to pojęcie starsze od WWW, zaproponowane przez Theodora Nelsona w 1965 roku. Oznacza ono organizację informacji bazującą na sieci niezależnych, ale wzajemnie odwołujących się do siebie za pomocą hiperłączy treści (tzw. leksji).

być, co otwiera pole do interesujących interpretacji historycznych. Historia powstania tagu <BLINK> i jego recepcja w środowisku twórców stron WWW[52] czy koloru CSS rebeccapurple[53] będącego nietypową formą upamiętnienia zmarłego dziecka pozwalają traktować kod źródłowy strony także jako tekst kultury.

Zmiany w obecności tagów HTML czy zmiany treści ich atrybutów mogą być podstawą tworzenia historii dużych grup witryn internetowych. Dennis Fetterly i współautorzy wykazują na przykład duże zmienności tagów <A> (odpowiadających za linki) czy <IMG> (odpowiadający za wyświetlanie grafiki) w przypadku korpusu 150 milionów stron WWW w okresie 11 tygodni w 2002 roku[54]. Analiza zmian w obecności pewnych tagów w kodzie strony WWW przygotowana na podstawie badań dłuższego okresu może wskazywać zwiększającą się obecność materiałów wizualnych czy interaktywnych na stronie, a przez to implikować pewne stwierdzenia na temat zmian w ich projektowaniu i recepcji. Anne Helmond w badaniu historycznym ewolucji witryny internetowej New York Timesa sięgnęła po kody skryptów śledzących/marketingowych obecne w kodzie źródłowym i wykazała, jak zmieniały się komercyjne uwarunkowania funkcjonowania tego serwisu oraz dane gromadzone na temat użytkowników[55]. Oba przytoczone wyżej badania ignorowały nie tylko semantyczną treść analizowanych witryn, ale też konkretne efekty działania kodu źródłowego. W badaniu Fetterly'ego nie było istotne, jakie grafiki i jakie linki umieszczane są w interpretowanych statystycznie znacznikach HTML. Helmond nie interesowało to, w jaki sposób określone trackery wykorzystują dane użytkowników, tylko jakiego rodzaju informacje zbierają i w jakim celu.

Przywołane wyżej badania wykorzystywały modularny charakter strony internetowej do wykazania historycznych zmienności Webu, zmian w organizacji informacji publikowanych online czy nawet konstrukcji przekazu o wydarzeniach historycznych (Beaudouin). Ignorowały warstwę treści poszczególnych stron, korzystając za to z adresów URI, nagłówków i kodu źródłowego jako źródeł historycznych. Chociaż modularny charakter strony internetowej utrudnia zachowanie autentyczności i integralności badanego zasobu, szczególnie przy pracy ze zbiorami archiwów Webu, pozwala jednak na projektowanie analiz historycznych Webu odrzucających paradygmat tekstu jako jedynej warstwy niosącej znaczenie. Do tego badanie odpowiedzi serwera (statusów HTTP) czy wartości nagłówków *Content-Type* jest właściwie niepomijalną metodą historycznego ujęcia ewolucji domen krajowych, jako że zawartość strony nie

---

52 Lou Montulli, „The Humble Origins of the HTML Blink Tag", *GIZMODO* (2020). https://gizmodo.com/the-humble-origins-of-the-html-blink-tag-5903827 (dostęp: 5.01.2021).

53 rebeccapurple to kolor definiowany w specyfikacji Kaskadowych Arkuszy Stylów (ang. *Cascading Style Sheets*, CSS), języku określającym estetykę strony WWW. Jeffrey Zeldman, „The Color Purple", *Zeldman on Web and Interaction Design* (2014). http://www.zeldman.com/2014/06/10/the-color-purple/ (dostęp: 5.01.2021).

54 Dennis Fetterly et al., „A Large-Scale Study of the Evolution of Web Pages", *Software: Practice and Experience* 34, 2 (2004): 228.

55 Anne Helmond, „Historical Website Ecology. Analyzing Past States of the Web Using Archived Source Code", w *Web 25: Histories from the First 25 Years of the World Wide Web*, red. Niels Brügger (New York: Peter Lang, 2017), 139–156.

www.czasopisma.pan.pl PAN www.journals.pan.pl

jest w stanie powiedzieć wiele o jej historycznej dostępności, a nagłówki są w stanie udostępnić informację o obiektach funkcjonujących w sieci, której osią jest dokument HTML. Również kod źródłowy może być nośnikiem informacji o zmianach w charakterze witryny, relacjach z innymi serwisami czy nawet jej modelu biznesowym.

Pośrednictwo oprogramowania w konstruowaniu źródła historycznego

Konstruktywistyczna natura źródła historycznego, opisywana w pracach z teorii historii, ujawnia się w tym, że nie jest ono nigdy gotowym nośnikiem informacji o przeszłości i powstaje dynamicznie w procesie zadawania pytań i konceptualizacji. Jak pisał Jerzy Topolski, „źródło przez kontakt z historykiem staje się określoną strukturą informacyjną pozwalającą rozwijać w sposób niemal nieskończony swą moc poznawczą”[56]. Proces konstruowania zachodzi jednak już „wcześniej”, choćby na etapie selekcji materiałów, które zostaną poddane badaniu, albo jako decyzja o tym, jaka część czy warstwa wybranego materiału zostanie uznana za posiadającą wartość informacyjną dla konkretnego badania. Przykładowo, Agata Błoch i współautorzy przeanalizowali 170 tysięcy dokumentów z archiwów imperium portugalskiego przechowywanych w Arquivo Histórico Ultramarino w Lizbonie. Metody komputerowej analizy tekstu umożliwiły rezygnację z ograniczonej selekcji, jaką cechowały wcześniejsze badania. Decyzje o wyróżnieniu semantycznych elementów niezbędnych do zbudowania sieci relacji między urzędnikami aparatu zarządzania imperium zrzucono na uczenie maszynowe. Mimo że nośnikiem źródeł był tutaj klasyczny materiał piśmienniczy (dokument) oraz tekst, źródłem stały się zidentyfikowane automatycznie metodą NER (*Named-entity recognition*) wierzchołki sieci, a właściwie relacje między nimi[57]. Źródło historyczne jest w tym badaniu współkonstruowane przez oprogramowanie.

Znaczenie pośrednictwa oprogramowania w badaniach społecznych i humanistycznych bywa niekiedy ignorowane, niekiedy zaś przydaje mu się efekty dalekie od rzeczywistych. W 2008 roku jeden z redaktorów popularnego i wpływowego magazynu *Wired* prognozował, że wobec dostępności pełnych danych o rozmaitych zjawiskach społecznych oraz oprogramowania pozwalającego je wykorzystywać w badaniach zaniknie użyteczność takich narzędzi jak teoria, hipoteza, model czy testy statystyczne[58]. W takim naiwnym podejściu wielkość

---

[56] Jerzy Topolski, „Problemy metodologiczne korzystania ze źródeł literackich w badaniu historycznym”, w *Dzieło literackie jako źródło historyczne*, red. Zofia Stefanowska, Janusz Sławiński (Warszawa: Spółdzielnia Wydawnicza „Czytelnik”, 1978), 21.

[57] Agata Błoch, Demival Vasques Filho, Michał Bojanowski, „Networks from archives: Reconstructing networks of official correspondence in the early modern Portuguese empire”, *Social Networks* (2020).

[58] Chris Andreson, „The End of Theory: The Data Deluge Makes the Scientific Method Obsolete”, *Wired* (2008). https://www.wired.com/2008/06/pb-theory/ (dostęp: 5.01.2021).

www.czasopisma.pan.pl PAN www.journals.pan.pl

i kompletność danych oraz potencjał obliczeniowy oprogramowania przekraczający zdolności ludzkie umożliwiać miałby niemal doskonałe oddanie rzeczywistości. Wydaje się przy tym, że w żaden sposób nie podważałoby to konstruktywistycznej natury źródła historycznego. Źródło jest zawsze czymś więcej niż nawet najpełniejsze zestawienie najbardziej dokładnych informacji. Fundamentalną rolę pełni w jego konstrukcji intencja badacza i jego zdolność kontekstualizacji zgromadzonych informacji. Nawet najdokładniej przetranskrybowane dokumenty z Arquivo Histórico Ultramarino nie budowałyby żadnej wiedzy bez koncepcji wypracowania na ich podstawie sieci relacji między urzędnikami królestwa portugalskiego, która następnie może być podstawą opisu i interpretacji mechanizmów zarządzania imperium.

Ryszard Nycz, podsumowując i odrzucając program „wiedzy pełnej, obiektywnej, neutralnej" uzyskiwanej dzięki pośrednictwu oprogramowania i metodom ilościowym, zwraca przy tym uwagę na będące efektem takiego podejścia zmiany w metodach badań humanistycznych. Badania tekstu zastępowane są opracowaniem danych, teoria – „aplikacjami technologicznymi programów", interpretacje – wizualizacjami, modelowaniem i symulacją, ostatecznie też ma miejsce odrzucenie zainteresowania badaniem tego, co jednostkowe, oryginalne, wartościonośne i wartościotwórcze na rzecz analiz makrotendencji[59]. Metody zastosowane w badaniach obejmujących miliony adresów URI z pewnością nie pozwalają na bezstratne oddanie przeszłej rzeczywistości, jako że opierają się na mimo wszystko ograniczonych danych i dużych przybliżeniach.

Rafał Ilnicki przekonuje, że „współcześnie historia [w znaczeniu przeszłości – przyp. MW] jest wytwarzana przez media techniczne, co wynika z faktu, że coraz większa część wydarzeń jest przez nie wytwarzana i dystrybuowana, co ma charakter zarówno rejestracji, jak i interpretacji [...]"[60]. Badanie i pisanie „historii Webu" czy „historii z Webem" oznacza więc konieczność przyjęcia pośrednictwa oprogramowania. Nie jest to sytuacja nowa dla każdego, kto kiedykolwiek pisał pracę naukową w edytorze tekstu albo korzystał z wyszukiwarek artykułów naukowych. Problem leży w skali i głębokości tego pośrednictwa. W odniesieniu do klasycznej historiografii posługującej się metodami cyfrowymi podstawową krytykę pośrednictwa oprogramowania proponuje Tim Hitchcock, wskazując choćby na niedoskonałość automatycznej transkrypcji (OCR) tekstów źródłowych czy wprowadzające w błąd metody i algorytmy wyszukiwania w bazach tekstów naukowych[61]. Podobne problemy wykazać można w przypadku badań historycznych Webu, przy czym już sam obiekt tych badań, czyli strona WWW, jest efektem pośrednictwa oprogramowania. Czytelnym przykładem może być badanie wybranych stron za pomocą różnych przegląda-

---

59 Ryszard Nycz, „Nowa humanistyka w Polsce: kilka bardzo subiektywnych obserwacji, koniektur, refutacji", *Teksty Drugie* 1 (2017): 25–26.

60 Rafał Ilnicki, „Historiografia cyberkultury – od badania przeszłości do algorytmicznej analityki czasu rzeczywistego", *Kultura i Historia* 1, 25 (2014): 1–10.

61 Tim Hitchcock, „Confronting the Digital: Or How Academic History Writing Lost the Plot", *Cultural and Social History: The Journal of the Social History Society* 10, 1 (2013): 9–23.

www.czasopisma.pan.pl PAN www.journals.pan.pl

rek – niektóre elementy i mechanizmy strony mogą być różnie renderowane (interpretowane i wyświetlane), co nasuwa pytania o to, czy archiwalny Web powinien być oglądany z wykorzystaniem historycznych wersji przeglądarek[62].

Ignorowanie pośrednictwa oprogramowania w pracy badawczej to tak zwany efekt czarnej skrzynki (*black box*). Polega on na tym, że badacz udostępnia programowi określone zasoby do przetworzenia, przy czym robi to bez wiedzy na temat tego, w jaki sposób dokładnie będą one przepracowywane. Proces taki zachodzi choćby przy korzystaniu z wyszukiwarek, narzędzi do przetwarzania języka naturalnego czy narzędzi statystycznych. W kontekście pracy z zasobami Webu efekt ten polegać może na korzystaniu z zasobów bez jasno określonej proweniencji i zasad wytworzenia[63]. Wartościowymi metodami ograniczania efektu *black box* nie tylko w badaniach historycznych Webu jest dokumentowanie metod pozyskiwania i opracowywania źródeł, na przykład opisywanie zasad budowania kolekcji stron WWW podlegających analizie, oraz cytowanie wykorzystywanego oprogramowania.

Warto dodać, że problem pośrednictwa oprogramowania w konstrukcji źródeł historycznych podejmowany jest także wobec materiałów archiwalnych zdigitalizowanych i udostępnianych online. Digitalizacja nie jest procesem bezstratnym, a utrata kontaktu z materialną postacią dokumentu czy obiektu historycznego ma swoje konsekwencje. Nie zawsze jednak są one na tyle istotne, aby uwzględniać je w badaniu.

Wizualności strony internetowej

Jak widać na przytoczonych wyżej przykładach badań, strona internetowa może być zmuszona przez badaczy do wytworzenia źródła historycznego wcale nie wyłącznie na podstawie swojej treści tekstowej (czy też tekstowo-narracyjnej). Potencjalnym nośnikiem informacji historycznej może być kod źródłowy, nagłówki, odnośniki wychodzące ze strony i odnośniki do niej kierujące, ale też szeroko pojęta wizualność. Wizualność może być przy tym rozumiana dosłownie, tak jak w przypadku badania analizującego kolory jugosłowiańskiego Webu[64], jak i mniej oczywista, tak jak w badaniu Anthony'ego Cocciola, gdzie źródłem pozwalającym na spojrzenie w przeszłość WWW jest wyliczana odpowiednim algorytmem relacja między tekstem a grafiką na stronach inter-

---

62 Zob. *OldWeb.Today: JS Browser Emulation*, https://oldweb.today/#19960101/http://geocities.com/ (dostęp: 5.01.2020).

63 Na problem ten zwracają uwagę Kevin Driscoll i Shawn Walker, analizując niespójności w danych na temat wydarzeń związanych z ruchem Occupy Wall Street pozyskanych z API Twittera. Kevin Driscoll, Shawn Walker, „Working Within a Black Box: Transparency in the Collection and Production of Big Twitter Data", *International Journal of Communication* 8, 1 (2014).

64 Anat Ben-David, Adam Amram, Ron Bekkerman, „The colors of the national Web: visual data analysis of the historical Yugoslav Web domain", *International Journal on Digital Libraries* 19, 1 (2018): 95–106.

www.czasopisma.pan.pl PAN www.journals.pan.pl

netowych, wskazująca na jakościowe zmiany w percepcji Webu[65]. W badaniach kultury takie wynajdywanie wizualności proponuje od lat Lev Manovich[66].

Badanie historyczne oparte na wizualnościach Webu może więc – ale nie musi – oznaczać wpisywanie się w kierunek historii wizualnej, jeśli ta, jak pisze Dorota Skotarczak, miałaby zajmować się analizą przedstawień wizualnych[67]. Badanie wizualności nie musi przecież polegać na odczytywaniu znaczeń dostępnych ludzkiej percepcji – zachowania postaci nagranej na materiale video czy szczegółu sceny widocznej na zdjęciu. Bazować może na oddaniu komputerowi roli wyróżniania nowych elementów znaczeniowych. Przykładowo, Alise Tifentale i Lev Manovich, analizując 2,3 miliona fotografii udostępnianych w serwisie Instagram zignorowali treść i estetykę zdjęć i interpretowali je wyłącznie jako montaże pikseli[68], co pozwoliło na przykład rozpoznać zmienność wzorców w czasie (określanych tu jako rytmy wizualne, *visual rhythms*)[69]. Piksel sam w sobie nie niesie żadnego znaczenia, jest po prostu informacją o kolorze, funkcjonuje także w logice bazy danych, a nie narracji. Strona internetowa przez swoją modularność bardzo dobrze nadaje się do wynajdywania tego typu nowej wizualności – bardzo głęboko szukać możemy budujących ją elementów.

Możliwości te nie przekreślają jednak historiofotycznego podejścia, w którym strona internetowa jest niczym innym jak kolejnym wcieleniem dokumentu tekstowego zawierającego elementy wizualne czy też fotografią. Wizualność i jej znaczenie jest w takim przypadku wyodrębniona na podstawie ludzkiej percepcji, bez głębokiego pośrednictwa oprogramowania, na co zdecydował się Manovich. Podejście historiofotyczne, ignorujące właściwie modularny charakter strony internetowej, umożliwia stosowanie klasycznych narzędzi interpretacyjnych takich jak na przykład gatunek. Rozpoznanie gatunku badanego obiektu jest jedną z podstaw analizy źródłowej, także przy źródłach wizualnych.

W odniesieniu do warstwy wizualnej strony WWW rozpoznawanie gatunku wymagałoby uwzględnienia nie tylko jej cech formalnych i wizualnych, ale też otoczenia sieciowego i tego, jaką pełni ona funkcję w komunikacji w Internecie[70]. Przykładowo, analiza historyczna tweetów[71] poświęconych protestom przeciwko

---

65 Anthony Cocciolo, „Quantitative Web History Methods", w *The SAGE Handbook of Web History*, red. Niels Brügger, Ian Milligan (Los Angeles: SAGE Publications, 2018), 138–152.

66 Lev Manovich, „Selected Projects & Exhibitions", http://manovich.net/index.php/exhibitions (dostęp: 5.01.2021).

67 Dorota Skotarczak, „Kilka uwag o historii wizualnej", *Klio Polska* 8 (2016): 118.

68 Lev Manovich, Alise Tifentale, „Selfiecity: Exploring Photography and Self-Fashioning in Social Media", w *Postdigital Aesthetics: Art, Computation and Design*, red. David M. Berry, Michael Dieter (London: Palgrave Macmillan, 2014), 109–122.

69 „Instagram Cities. Planetary-Scale Visual Rhythms", https://phototrails.net/instagram-cities/ (dostęp: 5.01.2021).

70 Maciej Maryl, *Życie literackie w sieci. Pisarze, instytucje i odbiorcy wobec przemian technologicznych* (Warszawa: Instytut Badań Literackich, Fundacja Akademia Humanistyczna, 2015), 219.

www.czasopisma.pan.pl PAN www.journals.pan.pl

zaostrzaniu regulacji aborcyjnych nie powinna zakładać, że mamy oto do czynienia z jednorodnym gatunkiem strony WWW (tweetem), ale uwzględniać fakt, że w ramach określonych wspólnot czy nisz komunikacyjnych funkcjonujących w polskojęzycznej części Twittera tweet jako gatunek wypowiedzi może mieć diametralnie różne cechy. Inaczej analizie poddadzą się urzędowe wpisy publikowane przez instytucje rządowe czy policję, inaczej opinie publicystów, a inaczej tweety wysyłane „na żywo" przez osoby bezpośrednio uczestniczące w protestach. Niektóre tweety będą miały cechy asynchronicznej publikacji prasowej czy dokumentu, inne będą strumieniem (*stream*), w ramach którego interesujące historyka treści tworzone będą synchronicznie, w trakcie trwania wydarzeń. David M. Berry opisuje taką komunikację pojęciem „Web czasu rzeczywistego" (real-time Web)[72]. Chociaż na głębokim poziomie technicznym wszystkie te materiały są identyczne jako treści (*de facto* strony internetowe) publikowane w ramach tego samego serwisu, skuteczne wyróżnienie ich gatunków może pozwolić na znalezienie nowych skutecznych metod badania. W przypadku streamingu wartościowe może być uwzględnienie zmieniającej się lokalizacji tweetów czy zbadanie sieci ich redystrybucji, w przypadku wpisów politycznych ważne będą komentarze i liczba polubień, która może być interpretowana w kontekście poparcia odbiorców dla przedstawianych poglądów i opinii.

Poprawna interpretacja wizualności strony WWW wymagać może również sprawdzenia, w jaki sposób nawiązuje ona do jednego z podstawowych paradygmatów projektowania webowego. Określają one relacje dizajnu ze światem spoza przeglądarki. Podobne związki rozpoznać można zresztą w projektowaniu graficznym publikacji drukowanych – od inkunabułów imitujących pismo ręczne do szkoły szwajcarskiej w typografii. Przykładowo, skeumorfizm to podejście w projektowaniu stron WWW zakładające, że wygląd jej elementów nawiązywać powinien do wyglądu rzeczywiście istniejących obiektów[73] i przez to ułatwiać korzystanie ze strony. Wizualność strony nie jest jedynie efektem możliwości technicznych. Współczesna humanistyczna refleksja nad oprogramowaniem i Internetem podkreśla relacje między warstwą logiczną i wizualną a materialnością. Na przykład Søren Pold analizuje działanie przycisku (*button*)

---

71 Tweety (*tweets*) to krótkie publikacje użytkowników platformy społecznościowej Twitter. Ich treścią może być tekst, materiały wizualne i multimedia, odnośniki do innych stron, tweetów czy użytkowników. Twitter powszechnie wykorzystywany jest przez osoby publiczne i instytucje państwowe oraz stanowi ważne źródło zasobów dokumentujących wydarzenia polityczne.

72 David M. Berry, *The Philosophy of Software. Code and Mediation in the Digital Age* (Basingstoke - New York: Palgrave Macmillan UK, 2015), 142.

73 Skeumorfizm to paradygmat projektowania graficznego polegający na użyciu elementu dizajnu lub struktury charakterystycznego i niezbędnego dla wybranego obiektu w dizajnie lub strukturze nowego obiektu, mimo że nie służy tam już żadnemu celowi. Przykładem skeumorfizmu jest dźwięk migawki w smartfonach albo wygląd przycisku na stronie internetowej imitujący guzik mechanicznej maszyny.

www.czasopisma.pan.pl PAN www.journals.pan.pl

na stronie internetowej jako metaforę materialności przeniesioną z ludzkiego doświadczenia pracy mechanicznej maszyny, gdzie kliknięcie, czyli gest fizyczny, inicjuje fizyczny proces[74].

Materialność i przestrzeń

Czy zasadne jest zatem traktowanie stron internetowych wyłącznie jako źródeł niematerialnych, cyfrowych? Niekiedy plan badania pozwala na taką redukcję, warto zadać jednak pytanie, kiedy staje się ona wadą. Czy analizując rozwój domeny krajowej, możemy ignorować poziom dostępu do komputerów w danym społeczeństwie, koszty dostępu do Internetu i hostingu umożliwiającego publikowanie stron? Uwzględniając te czynniki w interpretacji przyrostu liczby domen w czasie, z pewnością wzbogacimy nasz opis, nawet jeśli nie da się wykazać prostej relacji na przykład między poziomem dostępu do Internetu a liczbą rejestrowanych domen.

Na stronę internetową wyświetlaną w przeglądarce i jako taką podlegającą badaniu patrzeć można jako na efekt pewnych procesów w warstwie informacyjnej i logicznej (komunikacyjnych, informatycznych) oraz materialnej (sprzętowych, społecznych). Nie jest to oczywiście nowatorskie podejście. W badaniach historycznych ujawnia się ono choćby w krytyce zewnętrznej dokumentów źródłowych, a w literaturoznawstwie w pracy nad rękopisami. Tworzenie naukowych edycji cyfrowych w standardzie TEI (Text Encoding Initiative) jest próbą wyrażenia materialności źródła za pomocą odpowiednich znaczników pozwalających wskazać na przykład skreślenia, poprawki i fizyczne wady transkrybowanego dokumentu.

R. Stuart Geiger, analizując system Wikipedii, wskazuje na rolę botów i innego wyspecjalizowanego oprogramowania spoza głównego trzonu (*core*) platformy MediaWiki i zwraca uwagę, że ten „szyty na zamówienie" kod (*bespoke code*) działać musi na serwerach pracujących równolegle do głównego serwera MediaWiki i musi być zarządzany oraz utrzymywany przez niezależnych twórców. Materialna warstwa oprogramowania, którą wyznacza Geiger, to choćby serwery konieczne do utrzymania nieustannej aktywności botów, ale także nośniki danych, system prawa autorskiego i jego praktyka czy w końcu twórcy i administratorzy tych botów oraz społeczność *open source*, w ramach której funkcjonują i tworzą[75]. Materialność oprogramowania składać się może z ogromnej liczby elementów, których jednocześnie nigdy nie da się uwzględnić w badaniu – musi ono przecież mieć wyznaczone konkretne granice, a jego efektem nie może być „opisanie wszystkiego". Sięgnięcie do niej może jednak wzbogacić wartość poznawczą źródeł internetowych. Przykładowo, badając

---

[74] Søren Pold, „Button", w *Software Studies: A Lexicon*, red. Matthew Fuller (Cambridge: MIT Press, 2006), 31–36.

[75] R. Stuart Geiger, „Bots, bespoke, code and the materiality of software platforms", *Information, Communication & Society* 17, 3 (2014): 347.

www.czasopisma.pan.pl PAN www.journals.pan.pl

z wykorzystaniem Twittera protest zorganizowany w przestrzeni miasta, historyk ma do dyspozycji nie tylko analizę tekstów poszczególnych wpisów czy relacji między użytkownikami dokumentującymi to wydarzenie, ale też dostęp do warstwy geograficznej, bo przynajmniej część tweetów publikowanych za pomocą urządzeń mobilnych dodaje do wpisu metadane geolokalizacyjne. Marco Bastos i współautorzy wykorzystali tę możliwość do przygotowania analizy protestów w ramach Brazylijskiej Wiosny (2013), co pozwoliło im wykazać, że osoby publikujące na temat tych wydarzeń nie brały w nich udziału bezpośrednio na ulicach[76]. Elementem ich analizy są mapy oraz wizualizacje geograficzne – klasyczna tekstowa narracja niekoniecznie wystarcza do prezentacji wyników badań Webu. Dane lokalizacyjne i ich przetwarzanie stały się jedną z warstw mediów społecznościowych i jako takie mogą być badane historycznie, a sensory i czujniki w urządzeniach noszonych bezpośrednio na ciele (*wearables*) i udostępniające dane online wzbogacać mogą analizowany zasób nawet o warstwę biologiczną, przy czym dostęp do niej warunkują polityki prywatności oraz zasady etyczne.

Nowe chronologie

Wspomniany wcześniej David M. Berry, pisząc o „Webie czasu rzeczywistego" (*real time Web*), podkreśla zmianę, jaką obserwować można od czasu upowszechnienia się mediów społecznościowych i urządzeń mobilnych. Web, kiedyś kojarzony głównie ze stronami WWW, jest dziś konceptualizowany jako strumień (*stream*) czy przepływ (*flow*), nieustanny, niemal synchroniczny przekaz danych i informacji. Recepcja tak scharakteryzowanego medium musi się zmieniać – *stream* nie może być czytany czy nawet oglądany, odbiorca-użytkownik raczej stara się utrzymać jakiś rodzaj łączności z różnymi źródłami nieustannie transmitowanych informacji oraz uczy się jak się pod nie podłączać i z nich rezygnować. Asynchroniczna recepcja strony WWW, którą ktoś publikuje w jednym momencie, a ktoś inny czyta czy ogląda w innym, ustępuje miejsca przekazowi w czasie rzeczywistym – webowy strumień funkcjonuje w logice *liveness* czy *nowness*[77]. Naturalnie opozycje te są raczej metaforami opisującymi Web jako medium, a nie opisują radykalnych zmian w jego konstrukcji technicznej, bo te nie nastąpiły[78]. Do tego strumień może być konsumowany także w trybie asynchronicznym, skoro media społecznościowe udostępniają wgląd w przynajmniej niektóre archiwalne publikacje swoich użytkowników. Można zadać pytanie, jak zmienia się perspektywa osoby, która – do celów

---

[76] Marco Toledo Bastos, Raquel da Cunha Recuero, Gabriela da Silva Zago, „Taking tweets to the streets: A spatial analysis of the Vinegar Protests in Brazil", *First Monday* 19, 3 (2014).

[77] David M. Berry, *Philosophy of software*, 143.

[78] *Real-time Web* mógłby podkreślać np. zwiększoną przepustowość sieci Internet oraz szerszą dostępność połączeń internetowych.

www.czasopisma.pan.pl PAN www.journals.pan.pl

analizy historycznej – korzysta najpierw z przekazu webowego w trybie *liveness*, a po pewnym czasie używa już wyłącznie jego zarchiwizowanej wersji, przetworzonej przez algorytmy mediów społecznościowych działające w sposób dla nas niewiadomy. Bez wątpienia znów mamy w tym przypadku do czynienia z fundamentalnym pośrednictwem oprogramowania w kreowaniu źródła historycznego.

Strumień cechuje klasyczna liniowa chronologia – poszczególne publikacje pojawiają się czy też przepływają jedna po drugiej. Jeśli posiadają osobne adresy URI, mogą być wyodrębnione i funkcjonować niezależnie od innych, na przykład pojawiać się w wynikach wyszukiwania. Taki status mają między innymi wpisy na Twitterze, które zarówno w postaci stron internetowych, jak i w postaci danych wyeksportowanych z interfejsu programistycznego tego serwisu mają przypisane unikalne identyfikatory i oznaczenia czasowe. Proste następstwo chronologiczne nie wyklucza jednak tego, że treści pojawiające się jedna po drugiej zachowują porządek liniowy – nawet klasyczny film, medium produkujące złudzenia chronologicznej ciągłości za sprawą przesuwania się serii klatek na materialnym nośniku, może działać w logice montażu. Poszczególne elementy strumienia mogą budować wokół siebie własne chronologie: pod każdym z tweetów pojawiać się mogą komentarze, odpowiedzi na komentarze, powiązane z nimi tweety itp. Liniowa chronologia rozdziela się tutaj na bogatą sieć połączeń i nawiązań, z których każde funkcjonuje osobno.

Vivian Sobchack, opisując kształtowanie historyczności w noworocznych podsumowaniach telewizyjnych, gdzie montuje się migawki z najważniejszych wydarzeń odchodzącego roku, zwraca uwagę na to, że każdy element takiego montażu jest kompletny, hermetyczny, czasowo nieciągły (*temporally discontinuous*) i specyficznie kodowany. Niekiedy element montażu jest ikoniczny, wykracza poza jego kontekst, może być zrozumiany bez niego i może posiadać własną temporalność. W innym przypadku jego znaczenie rozpoznawać można wyłącznie w kontekście montażu i nie posiada on żadnej historyczności[79].

Ponieważ źródło historyczne nie istnieje *a priori*, decyzja o nadawaniu znaczeń poszczególnym elementom webowego strumienia należy do badacza, zmuszonego przy tym zaakceptować pośrednictwo oprogramowania decydującego o tym, które z klatek montażu mogą być w ogóle wyróżnione jako osobne obiekty analizy. W przypadku Twittera czy historii edycji haseł w Wikipedii użytkownik ma możliwość rozpoznawania poszczególnych elementów strumienia – każdy z nich posiada unikalny adres URL oraz stosowne metadane

---

79 Rozróżnienie to proponuje Sarah Paige Baty, wprowadzając pojęcia ikon i mediafemów, zob. Vivian Sobchack, „The Insistent Fringe: Moving Images and Historical Consciousness", *History and Theory* 36, 4 (1997): 11–12. W elementach wizualnych montażu/strumienia webowego rozpoznawać można także warstwę temporalną bazującą na podstawowych właściwościach formatu cyfrowego i wytwarzaną przez oprogramowanie, nie ma ona jednak praktycznego znaczenia w budowaniu chronologii na potrzeby badania. Zob. Wolfgang Ernst, „Archiwum, przechowywanie, entropia. Tempor(e)alności fotografii", przeł. Marta Skotnicka, w *Archiwum jako projekt*, red. Krzysztof Pijarski (Warszawa: Fundacja Archeologia Fotografii, 2011), 74–75.

www.czasopisma.pan.pl PAN www.journals.pan.pl

określające czas. Taka wyraźna i łatwo dostępna warstwa czasowa w publikowaniu webowym jest jednak raczej wyjątkiem niż regułą.

M. Jones Shawn i współautorzy wykazali, że trzy czwarte zasobów webowych linkowanych w przypisach prac naukowych w badanym korpusie wraz z upływem czasu zmieniło swoją treść[80]. Zmiany te nie były rejestrowane. Efekt *content drift* wynikający z faktu, że pod stabilny adres URL przypisać można zmienne treści, jest poważnym wyzwaniem dla historycznych badań Webu, utrudniającym choćby budowanie chronologii czy nadawanie znaczeń zmianom. Niezależne archiwizowanie zmian na stronach internetowych z wykorzystaniem narzędzi takich jak NewsDiffs[81] czy Diffengine[82] może w takim przypadku mieć fundamentalne znaczenie dla badania – bez niego historyk będzie miał dostęp wyłącznie do współczesnych, bieżących wersji tych zasobów, co może uniemożliwiać podejmowanie jakichkolwiek analiz historycznych.

Dostępne w archiwach Webu kopie stron zawierać mogą elementy archiwizowane w różnym czasie ze względu na ograniczenia w metodzie crawlingu[83]. N. Brugger zwraca przy tym uwagę, że to nie moment publikacji strony i wprowadzania zmian do jej treści kształtuje dostępną do analizy chronologię, ale zazwyczaj nieregularne i przypadkowe akty archiwizacji. Dopiero na podstawie analizy rytmu archiwizacji można próbować oszacować rytm publikowania i zmian treści[84]. S.G. Ainsworth i M. Nelson opisali szereg problemów z określaniem chronologii i interpretowaniem dat archiwizacji stron dostępnych w archiwach Webu: błędne interpretacje wymuszać mogą nawet mechanizmy przeglądania zasobów. W celu usprawnienia nawigacji po zarchiwizowanej witrynie, tak aby przypominała ona korzystanie ze zwykłej witryny w *live Web*, użytkownik otrzymuje kolejne podstrony zarchiwizowane w czasie niekiedy dość odległym od czasu archiwizacji strony startowej[85].

Osobną warstwę czasową posiadają także niektóre elementy składające się na stronę internetową. Kod JavaScript, odpowiadający za reakcje strony na działania użytkownika takie jak kliknięcie, może nakładać na stronę WWW niezależną logikę czasową. Wydaje się, że uchwycenie, emulacja i analiza tego typu czasowości może mieć znaczenie na przykład w badaniach historycznych

---

80 Shawn M. Jones et al., „Scholarly Context Adrift: Three out of Four URI References Lead to Changed Content", PLoS ONE 11, 12 (2016).

81 *NewsDiffs*, https://github.com/ecprice/newsdiffs (dostęp: 5.01.2021).

82 *diffengine*, https://github.com/DocNow/diffengine (dostęp: 5.01.2021).

83 Marc Spaniol et al., „Data quality in web archiving", w *WICOW '09: Proceedings of the 3rd workshop on Information credibility on the web* (New York: Association for Computing Machinery, 2009), 19–26.

84 Metoda ta polega na indeksowaniu i pobieraniu zasobów webowych za pomocą działających automatycznie programów (botów). Bot wysyłany jest do pierwszej strony z indeksu, sczytuje dostępne na niej odnośniki do innych stron i za nimi podąża. Każda wizyta pozwala zapisać odwiedzaną stronę. *Crawling* jest metodą powszechnie stosowaną w gromadzeniu zbiorów archiwów Webu. Zob. Niels Brügger, *Archived Web*, 143–144.

85 Scott G. Ainsworth, Michael L. Nelson, „Evaluating sliding and sticky target policies by measuring temporal drift in acyclic walks through a web archive", *International Journal on Digital Libraries* 16 (2015): 129–144.

www.czasopisma.pan.pl PAN www.journals.pan.pl

sztuki nowych mediów wykorzystującej strony WWW jako nośnik projektów artystycznych. O ile archiwizacja stron internetowych metodą crawlingu często uniemożliwia odtworzenie ich dynamiki i blokuje interakcje budowane przez kod JavaScript, to alternatywna metoda proponowana przez nowojorską fundację Rhizome nastawiona jest przede wszystkim na zachowanie interakcji[86].

Potencjał historiografii Webu

Celem niniejszego artykułu było pokazanie, że WWW może być nie tylko narzędziem wspierającym badania historyczne, ale też obiektem tych badań. Strona internetowa może być w nich podstawowym nośnikiem informacji o przeszłości. Zaproponowana przeze mnie definicja strony internetowej podkreśla jej modularny charakter, który może być szansą dla historyka: uniwersalne identyfikatory, nagłówki, kod źródłowy pozwalają na podejmowanie badań wychodzących ponad interpretacje treści stron. Ponieważ oprogramowanie może pośredniczyć w konstruowaniu źródeł historycznych, w badaniach historycznych wykorzystujących zasoby WWW można polegać na percepcji maszynowej, czy to w rozpoznawaniu warstwy wizualnej stron internetowych, czy też pozyskiwaniu informacji wynikających z materialnego charakteru oprogramowania, takich jak na przykład dane geolokalizacyjne. Historiografia Webu korzystać może z zasobów, metod i narzędzi rozwijanych od połowy lat dziewięćdziesiątych przez archiwistykę WWW oraz inspirować się badaniami medioznawczymi czy socjologią i etnografią Sieci.

Być może praktykowanie historiografii Webu ma potencjał wywrotowy: nakładanie na WWW alternatywnych warstw czasowych i rejestrowanie zmian to skrajnie praktyczne przykłady wytwarzania historyczności w medium funkcjonującym w trybie „wiecznego teraz". Historiografia taka mogłaby być alternatywą dla dominującej w przekazach medialnych i silnie obecnej nawet w humanistycznej i społecznej naukowej refleksji nad WWW chronologii bazującej na wersjonowaniu (*Web 1.0*, *Web 2.0* itd.) lub kolejnych „przełomach" technologicznych. Czy, jak chcą tego autorzy *The History Manifesto*, historia mogłaby stać się dyscypliną narzucającą teorie, techniki i standardy pracy z zasobami i danymi cyfrowymi gromadzonymi i opracowywanymi w perspektywie temporalnej[87]?

---

86 Morgan McKeehan, „Symmetrical Web Archiving with Webrecorder, a Browser-based Tool for Digital Social Memory. An Interview with Ilya Kreymer", National Digital Stewardship Residency (NDSR), https://ndsr.nycdigital.org/symmetrical-web-archiving-with-webrecorder-a-browser-based-tool-for-digital-social-memory-an-interview-with-ilya-kreymer/ (dostęp: 5.01.2021).

87 David Armitage, Jo Guldi, *The History Manifesto* (Cambridge: Cambridge University Press, 2014), 104 i nast.

www.czasopisma.pan.pl PAN www.journals.pan.pl

Bibliografia

Ainsworth, Scott G., Michael L. Nelson. „Evaluating sliding and sticky target policies by measuring temporal drift in acyclic walks through a web archive". *International Journal on Digital Libraries* 16 (2015): 129–144.

Allen, Matthew. „What was Web 2.0? Versions as the dominant mode of internet history". *New Media and Society* 15, 2 (2012): 260–275.

Andreson, Chris. „The End of Theory: The Data Deluge Makes the Scientific Method Obsolete". *Wired* (2008). https://www.wired.com/2008/06/pb-theory/ (dostęp: 5.01.2021).

Ankerson, Megan Sapnar. „Writing web histories with an eye on the analog past". *New Media and Society* 14, 3 (2012): 384–400.

Armitage, David, Jo Guldi. *The History Manifesto*. Cambridge: Cambridge University Press, 2014.

Beaudouin, Valérie, Zeynep Pehlivan, Peter Stirling. „Exploring the Memory of the First World War Using Web Archives: Web Graphs Seen from Different Angles". W *The SAGE Handbook of Web History*, red. Niels Brügger, Ian Milligan, 441–463. Los Angeles: SAGE Publications, 2018.

Bastos, Marco Toledo, Raquel da Cunha Recuero, Gabriela da Silva Zago. „Taking tweets to the streets: A spatial analysis of the Vinegar Protests in Brazil". *First Monday* 19, 3 (2014).

Ben-David, Anat, Adam Amram, Ron Bekkerman. „The colors of the national Web: visual data analysis of the historical Yugoslav Web domain". *International Journal on Digital Libraries* 19, 1 (2018): 95–106.

Berry, David M. *The philosophy of software: Code and mediation in the Digital Age.* Basingstoke - New York: Palgrave Macmillan UK, 2015.

Błoch, Agata, Demival Vasques Filho, Michał Bojanowski. „Networks from archives: Reconstructing networks of official correspondence in the early modern Portuguese empire". *Social Networks* (2020) (dostęp: 10.05.2021).

Bogucka, Maria. „Zboże rosyjskie na rynku amsterdamskim w pierwszej połowie XVII wieku". *Przegląd Historyczny* 53, 4 (1962): 612–628.

Bolter, Jay David, Richard Grusin. *Remediation: Understanding New Media.* Cambridge: MIT Press, 2000.

Brügger, Niels. „The archived website and website philology: A new type of historical document?". *Nordicom Review* 29, 2 (2008): 155–175.

Brügger, Niels. *The Archived Web. Doing History in the Digital Age*. Cambridge: MIT Press, 2018.

Brügger, Niels. „Website history and the website as an object of study". *New Media and Society* 11, 1–2 (2009): 115–132.

Brügger, Niels, Jane Nielsen, Ditte Laursen. „Big data experiments with the archived Web: Methodological reflections on studying the development of a nation's Web". *First Monday* 25, 3 (2020). https://firstmonday.org/ojs/index.php/fm/article/view/10384 (dostęp: 30.06.2021).

Cocciolo, Anthony. „Quantitative Web History Methods". W *The SAGE Handbook of Web History,* red. Niels Brügger, Ian Milligan, 138–152. Los Angeles: SAGE Publications, 2018.

Crowston, Kevin, Marie Williams. „Reproduced and emergent genres of communication on the World-Wide Web". *The Information Society* 16, 3 (2000): 201–215.

Domańska, Ewa. „Perspektywy badań historycznych w Polsce wobec zmian we współczesnej humanistyce". W *Historyk wobec źródeł,* red. Jolanta Kolbuszewska, Rafał Stobiecki, 115–127. Łódź: Wydawnictwo Ibidem, 2010.

www.czasopisma.pan.pl PAN www.journals.pan.pl

Driscoll, Kevin, Shawn Walker. „Working Within a Black Box: Transparency in the Collection and Production of Big Twitter Data". *International Journal of Communication* 8, 1 (2014): 1745–1764.

Ernst, Wolfgang. „Archiwum, przechowywanie, entropia. Tempor(e)alności fotografii", przeł. Marta Skotnicka. W *Archiwum jako projekt*, red. Krzysztof Pijarski, 67–78. Warszawa: Fundacja Archeologia Fotografii, 2011.

Fetterly, Dennis et al. „A Large-Scale Study of the Evolution of Web Pages". *Software: Practice and Experience* 34, 2 (2004): 213–237.

Filiciak, Mirosław. „Gra w kapitalizm. PRL, konstruowanie pamięci i gry cyfrowe". *Widok. Teorie i praktyki kultury wizualnej* 11 (2015): 1–30.

Foot, Kirsten, Steven Schneider. „Object-oriented Web Historiography". W *Web History*, red. Niels Brügger, 62–79. New York: Peter Lang Inc, 2010.

Garde-Hansen, Joanne. „MyMemories? Personal Digital Archive Fever and Facebook". W *Save As... Digital Memories*, red. Joanne Garde-Hansen, Andrew Hoskins, Anna Reading, 135-150. New York: Springer, 2009.

Geiger, r. Stuart. „Bots, bespoke, code and the materiality of software platforms". *Information, Communication & Society* 17, 3 (2014): 342–356.

Haug, David M. „International Transfer of Technology: Lessons that East Europe can learn from the failed Third World experience". *Harvard Journal of Law & Technology* 5, 209 (1992): 209–240.

Helmond, Anne. „A historiography of the hyperlink: Periodizing the Web through the changing role of the hyperlink". W *The SAGE Handbook of Web History,* red. Niels Brügger, Ian Milligan, 227–241. Los Angeles: SAGE Publications, 2018.

Helmond, Anne. „Historical Website Ecology. Analyzing Past States of the Web Using Archived Source Code". W *Web 25: Histories from the First 25 Years of the World Wide Web*, red. Niels Brügger, 139–156. New York: Peter Lang, 2017.

Hitchcock, Tim. „Confronting the Digital: Or How Academic History Writing Lost the Plot". *Cultural and Social History: The Journal of the Social History Society* 10, 1 (2013): 9–23.

Hobsbawm, Eric. *Wiek imperium 1875–1914*, przeł. Marcin Starnawski. Warszawa: Wydawnictwo Krytyki Politycznej, 2015.

Ilnicki, Rafał, „Historiografia cyberkultury – od badania przeszłości do algorytmicznej analityki czasu rzeczywistego". *Kultura i Historia* 1, 25 (2014): 1–10.

Jones, Shawn M. et al. „Scholarly Context Adrift: Three out of Four URI References Lead to Changed Content". *PLoS ONE* 11, 12 (2016).

Kraskowska, Ewa, „Źródło [Słownik poetologiczny]". *Forum Poetyki* (2015): 98–101.

Maemura, Emily et al. „If these crawls could talk: Studying and documenting web archives provenance". *Journal of the Association for Information Science and Technology* 69, 10 (2018): 1223–1233.

Manovich, Lev. *Język nowych mediów*, przeł. Piotr Cypryański. Warszawa: Wydawnictwa Akademickie i Profesjonalne, 2006.

Manovich, Lev, Alise Tifentale. „Selfiecity: Exploring Photography and Self-Fashioning in Social Media". W *Postdigital Aesthetics: Art, Computation and Design*, red. David M. Berry, Michael Dieter, 109–122. London: Palgrave Macmillan, 2014.

Marecki, Piotr, Tadeusz Cieślewicz. *Oni migają tymi kolorami w sposób profesjonalny. Narodziny gamedevu z ducha demosceny w Polsce*. Kraków: Ha!art, 2020.

Maryl, Maciej. *Życie literackie w sieci. Pisarze, instytucje i odbiorcy wobec przemian technologicznych.* Warszawa: Instytut Badań Literackich, Fundacja Akademia Humanistyczna, 2015.

www.czasopisma.pan.pl PAN www.journals.pan.pl

McKeehan, Morgan. „Symmetrical Web Archiving with Webrecorder, a Browser-based Tool for Digital Social Memory. An Interview with Ilya Kreymer", National Digital Stewardship Residency (NDSR). https://ndsr.nycdigital.org/symmetrical-web-archiving-with-webrecorder-a-browser-based-tool-for-digital-social-memory-an-interview-with-ilya-kreymer/ (dostęp: 5.01.2021).

Mehrabi, Ninareh et al. „A survey on bias and fairness in machine learning". arXiv preprint arXiv:1908.09635 (2019).

Miranda, João, Daniel Gomes. „Trends in Web characteristics". W *Proceedings of the Latin American Web Congress (LA-Web)*, red. Edgar Chávez, Elizabeth S. Furtado, Alberto L. Morán, 146–153. Los Alamitos: IEEE Computer Society, 2009.

Montfort, Nick et al. *10 PRINT CHR$(205.5+RND(1)); : GOTO 10*. Cambridge: MIT Press, 2012.

Montulli, Lou. „The Humble Origins of the HTML Blink Tag". *GIZMODO* (2020). https://gizmodo.com/the-humble-origins-of-the-html-blink-tag-5903827 (dostęp: 5.01.2021).

Nanni, Federico. „Collecting Primary Sources from the Web Archives: A Tale of Scarcity and Abundance". W *The SAGE Handbook of Web History,* red. Niels Brügger, Ian Milligan, 112–124. Los Angeles: SAGE Publications, 2018.

Nycz, Ryszard. „Nowa humanistyka w Polsce: kilka bardzo subiektywnych obserwacji, koniektur, refutacji". *Teksty Drugie* 1 (2017): 18–40.

Paczkowski, Andrzej. „Historyk dziejów najnowszych wśród źródeł". W *Historia-dziś. Teoretyczne problemy wiedzy o przeszłości*, red. Ewa Domańska, Rafał Stobiecki, Tomasz Wiślicz, 89–98. Kraków: Universitas, 2014.

Pold, Søren. „Button". W *Software Studies: A Lexicon*, red. Matthew Fuller, 31–36. Cambridge: MIT Press, 2006.

«Raw data» is an oxymoron, red. Lisa Gitelman, Cambridge: MIT Press, 2013.

Skotarczak, Dorota. „Kilka uwag o historii wizualnej". *KlioPolska* 8 (2016): 117–130.

Sobchack, Vivian. „The Insistent Fringe: Moving Images and Historical Consciousness". *History and Theory* 36, 4 (1997): 4–20.

Spaniol, Marc et al. „Data quality in web archiving". W *WICOW '09: Proceedings of the 3rd workshop on Information credibility on the web,* 19–26. New York: Association for Computing Machinery, 2009.

Stobiecka, Monika, „Archaeological heritage in the age of digital colonialism". *Archaeological Dialogues* 27, 2 (2020): 113–125.

Topolski, Jerzy. „Problemy metodologiczne korzystania ze źródeł literackich w badaniu historycznym". W *Dzieło literackie jako źródło historyczne*, red. Zofia Stefanowska, Janusz Sławiński, 7–30. Warszawa: Spółdzielnia Wydawnicza „Czytelnik", 1978.

UNESCO. „Karta w sprawie zachowania dziedzictwa cyfrowego". 2003.

White, Hayden. „Historiografia i historiofotia", przeł. Łukasz Zaremba. W *Film i historia*, red. Iwona Kurz, 117–127. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego, 2008.

Wrzosek, Wojciech. „Losy źródła historycznego (refleksje na marginesie idei r. G. Collingwooda)". *Rocznik Antropologii Historii* 1–2, 1 (2011): 411–417.

Wrzosek, Wojciech. „Źródło historyczne jako alibi realistyczne historyka". W *Historyk wobec źródeł. Historiografia klasyczna i nowe propozycje metodologiczne*, red. Jolanta Kolbuszewska, Rafał Stobiecki, 23–38. Łódź: Wydawnictwo Ibidem, 2010.

Vafopoulos, Michalis, „Being, Space, and Time on the Web". *Metaphilosophy* 43, 4 (2012): 405–425.

Vee, Annette. *Coding Literacy: How Computer Programming Is Changing Writing*. Cambridge: MIT Press, 2017.

Zaremba, Marcin. *Wielka Trwoga. Polska 1944–1947*. Kraków-Warszawa: Wydawnictwo Znak, Instytut Studiów Politycznych PAN, 2012.

Zeldman, Jeffrey. „The Color Purple". *Zeldman on Web and Interaction Design* (2014). http://www.zeldman.com/2014/06/10/the-color-purple/ (dostęp: 5.01.2021).

Zasoby internetowe:

„Instagram Cities. Planetary-Scale Visual Rhythms". https://phototrails.net/instagram-cities/ (dostęp: 5.01.2021).

Manovich, Lev. „Selected Projects & Exhibitions". http://manovich.net/index.php/exhibitions (dostęp: 5.01.2021).

Definicje techniczne za: WebMDN Web Docs (Mozilla Developer Network): https://developer.mozilla.org/pl/docs/Web.