

Katarzyna Gmerek  
Biblioteka Uniwersytecka w Poznaniu

## **Archiwa internetowe po obu stronach Atlantyku — Internet Archive, Wayback Machine oraz UK Web Archive**

**Streszczenie:** Artykuł przedstawia dwa archiwa stron internetowych, ze Stanów Zjednoczonych i Wielkiej Brytanii, różniące się zasadami gromadzenia i sposobem korzystania. Wayback Machine jest przedsięwzięciem prywatnym, opartym na własnej fundacji, korzystającym ze współpracy wolontariuszy z wielu krajów świata. Gromadzi światowe strony WWW, nie poddając ich selekcji ani cenzurze. UK Web Archive jest dziełem bibliotek brytyjskich, pragnących realizować ideę egzemplarza obowiązkowego brytyjskich stron internetowych (niedziałającego dotychczas oficjalnie z powodu niedopracowanego prawa). Strony są starannie dobierane i poddawane ocenie merytorycznej przez bibliotekarzy według kryterium ważności i przydatności.

**Słowa kluczowe:** archiwizacja Internetu, archiwizacja stron WWW, archiwa internetowe, Wayback Machine, Internet Archive, UK Web Archive

### **Internet Archive Wayback Machine**

Amerykański projekt Internet Archive (IA, <http://www.archive.org/><sup>1</sup>) jest zarejestrowany w Kalifornii jako biblioteka i sam określa się skromnie jako „biblioteka internetowa” (*internet library*). Jak deklarują twórcy, *do jego celów należy oferowanie stałego dostępu do kolekcji historycznych, istniejących w formacie cyfrowym, badaczom, historykom, uczonym, osobom niepełnosprawnym i szeroko pojętej publiczności*<sup>2</sup>. Liczba samych zdigitalizowanych tekstów sięgnęła we wrześniu 2011 r. trzech mln, z czego dwa wprowadzili pracownicy IA, zaś pozostały mln — stowarzyszeni ochotnicy z całego świata, którym udostępniono odpowiednie oprogramowanie i miejsce na serwerach w USA. Archiwum zawiera ponadto bogate zbiory audiowizualne, np. stare filmy czy archiwalne nagrania muzyczne. Słynie z kolekcji zespołu Grateful Dead, ale można znaleźć tam prawie każdy typ materiałów, od nagrań spektakli teatralnych i audiobooków, po wideo muzyczne, filmy i nagrania dźwiękowe, często na licencji Creative Commons. Są także grafiki i stare programy komputerowe (jak zawartość archiwum Tucows z 2004 r.).

Internet Archive udostępnia również serwis Wayback Machine (WM, <http://www.archive.org/web/web.php>), gromadzący dawne wersje stron internetowych z całego świata, od 1996 r.<sup>3</sup> Ilość miejsca zajętego przez materiały Wayback Machine doszła już prawie do dwóch petabajtów i rozrasta się w tempie 20

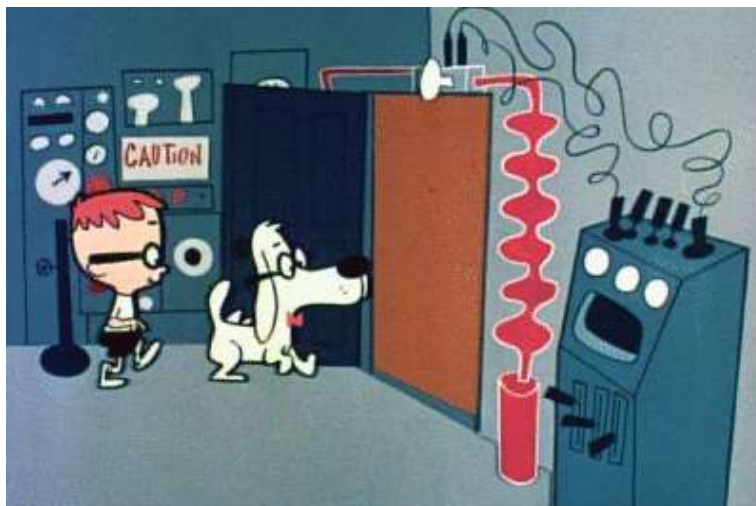
---

<sup>1</sup> Wszystkie odesłania do stron internetowych przedstawiają wersję aktualną w dn. 08.02.2012 r.

<sup>2</sup> *About the Internet Archive* [on-line]. [Dostęp 07.01.2012]. Dostępny w World Wide Web: <http://www.archive.org/about/about.php>.

<sup>3</sup> Nazwa Wayback Machine (co można przełożyć jako Machina Czasu) pochodzi ze starej amerykańskiej kreskówki dla dzieci *The Rocky & Bullwinkle Show*, której bohater stworzył wehikul czasu WABAC (wym. jak *way back* czyli droga powrotna). Zob. *The Rocky and Bullwinkle Show*. W: *Wikipedia, The Free Encyclopedia* [on-line]. [Dostęp 07.01.2012]. Dostępny w World Wide Web: [http://en.wikipedia.org/wiki/The\\_Rocky\\_and\\_Bullwinkle\\_Show](http://en.wikipedia.org/wiki/The_Rocky_and_Bullwinkle_Show).

terabajtów na miesiąc. W chwili pisania tego artykułu, twórcy bazy mogli pochwalić się zarchiwizowaniem ponad 150 mld stron (szczegółowe statystyki na stronie <http://www.archive.org/stats/> dotyczą tylko korzystania ze stron Archiwum i Wayback Machine). Baza jest rozproszona i mieści się na serwerach w trzech miastach kalifornijskich, a jej lustro (mirror) w Egipcie, w Bibliotheca Alexandrina ([http://www.bibalex.org/InternetArchive/IA\\_en.aspx](http://www.bibalex.org/InternetArchive/IA_en.aspx)).



Rys. 1. Wayback Machine

Źródło: *Waybackmachine3.png*. *Wikipedia: the free encyclopedia [on-line]*. [Dostęp 07.01.2012].  
Dostępny w World Wide Web: <http://en.wikipedia.org/wiki/File:Waybackmachine3.png>.

Twórca Internet Archive, Brewster Kahle (ur. 1960, absolwent MIT), był jednym z twórców Open Content Alliance. Jest też jednym z najbardziej wymownych krytyków polityki Google, zmierzającej jego zdaniem do monopolizacji bibliotek cyfrowych i ich komercjalizacji. B. Kahle wynalazł wiele intratnych narzędzi sieciowych, które sprzedał takim gigantom jak AOL czy Amazon, uzyskując w ten sposób czas i środki na niekomercyjną działalność związaną z tworzeniem cyfrowego archiwum publicznie dostępnych materiałów, w tym witryn internetowych<sup>4</sup>. Nie bez znaczenia jest fakt, że jego najbardziej znanym (i kontrowersyjnym) dziełem jest Alexa — wyszukiwarka dostarczająca informacji na temat generowanego ruchu do innych stron internetowych, zaopatrzona w robota internetowego indeksującego mld stron WWW. Alexa instaluje się jako pasek narzędziowy przeglądarki i zbiera informacje o stronach, po których porusza się internauta (nie jest to mechanizm ukryty — każdy, kto go instaluje, powinien przeczytać informacje i wiedzieć, że jego aktywność będzie analizowana). Powszechnie używany Google Toolbar również zbiera informacje o naszych wyszukiwaniach. Alexa jest dziś własnością Amazona, jednakże Wayback Machine wciąż korzysta z jej zasobów. Mała popularność Alexy w Polsce i całej Europie Środkowej może w pewnym stopniu przyczyniać się do mniejszej liczby stron z naszej części świata w Wayback Machine. Podobnie jak Google, Internet Archive nie udostępnia publicznie szczegółowych danych na temat działania jego oprogramowania. Wiadomo tylko ogólnie, że przez całą dobę przeczesuje sieć i

---

<sup>4</sup> The Internet's librarian. *The Economist* [on-line]. 05.03.2009 [Dostęp 07.01.2012]. Dostępny w World Wide Web: <http://www.economist.com/node/13174399>.

wykonuje kopie stron. Musi zbierać też dane z Alexy i od użytkowników odwiedzających archiwum, o czym będzie mowa dalej.

Alexa nie jest podstawowym programem archiwizującym strony dla Wayback Machine. Programiści IA stworzyli do tego celu własne oprogramowanie typu *open source*, na razie w wersji beta. Na stronie WM zaleca się, by uwzględnić swoją witrynę w katalogach stron WWW i tym samym zwiększyć jej szansę na zarchiwizowanie. Wynika z tego, że program przeszukuje przede wszystkim internetowe spisy stron na dany temat i z nich czerpie dane o tym, co ma archiwizować (zatem trochę inaczej niż roboty Google, które teoretycznie archiwizują wszystko). Internauci mogą zwrócić się do serwisu z prośbą o zarchiwizowanie ich strony. Internet Archive, w tym i Wayback Machine, jest wierne idei *open access*. Istnieje zatem możliwość skorzystania z usługi archiwizacji stron na życzenie. Oferta ta jest skierowana głównie do instytucji edukacyjnych, muzeów oraz niezależnych badaczy. Służy do tego dedykowana aplikacja internetowa Archive-It, którą można dokładniej poznać na sieciowych seminariach organizowanych dla chętnych przez zespół IA. Zarchiwizowane strony są przechowywane na serwerach IA. Usługa jest dostępna na zasadzie subskrypcji, zatem należy przypuszczać, że nie jest darmowa<sup>5</sup>.

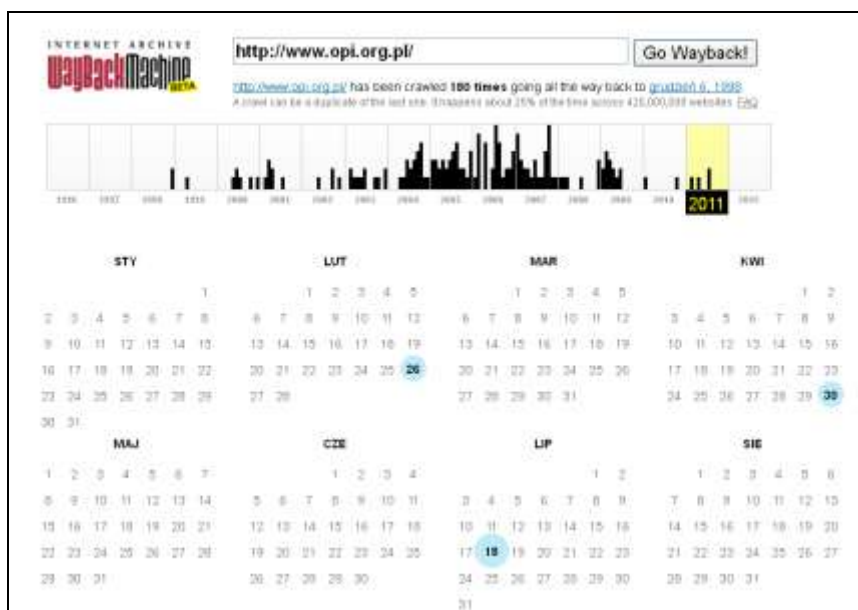


Rys. 2. Wayback Machine

Źródło: Wayback Machine [on-line]. [Dostęp 07.01.2012]. Dostępny w World Wide Web:  
<http://www.archive.org/web/web.php>.

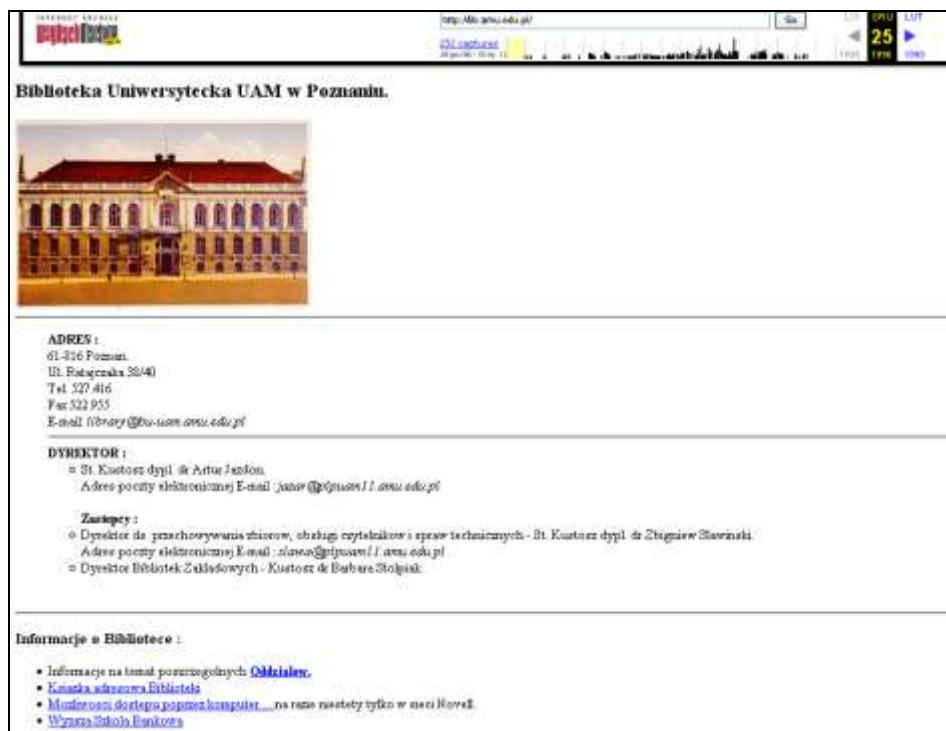
Szukając w Wayback Machine kopii archiwalnej jakiegokolwiek strony, musimy znać jej dokładny adres URL. Nie jest dostępne wyszukiwanie według słów kluczowych ani przeglądanie w indeksach, dlatego nie można analizować liczby znalezionych stron pod kątem ich zróżnicowania tematycznego. Kiedy wprowadzimy adres strony, to — o ile strona została zarchiwizowana — wyświetla się kalendarz pokazujący, kiedy roboty pobierały ze strony dane (rys. 3.). Użytkownik może przemieszczać się po kalendarzu wzdłuż linii lat albo według miesięcy danego roku i wybierać widoki stron według potrzeb.

<sup>5</sup> E-mail autorki artykułu do Internet Archive w sprawie ewentualnej opłaty pozostał bez odpowiedzi.



Rys. 3. Wayback Machine — rezultat wyszukiwania  
 Źródło: Wayback Machine [on-line]. [Dostęp 07.01.2012]. Dostępny w World Wide Web: <http://www.archive.org/web/web.php>.

Po wybraniu daty otrzymujemy kopię archiwalną witryny, jaka została zachowana przez roboty dla Wayback Machine. Na górze ekranu zawsze wyświetlana jest informacja o adresie URL oraz dacie archiwizacji (rys. 4.).



Rys. 4. Wayback Machine — strona Biblioteki Uniwersyteckiej w Poznaniu z 25.12.1996 r.  
 Źródło: Wayback Machine [on-line]. [Dostęp 07.01.2012]. Dostępny w World Wide Web: <http://www.archive.org/web/web.php>.

Z doświadczeń autorki w przeglądaniu zasobów Wayback Machine wynika, że głębokość archiwizowania stron i podstron bywa różna. W trakcie wertowania zarchiwizowanych stron „Rzeczypospolitej” okazało się, że kliknięcie w tytuł artykułu na pierwszej stronie skierowało na inną, a u góry ekranu ukazała się uwaga: *Wayback Machine nie zarchiwizowało tej strony. To jest zrzut zrobiony 0 minut temu na żywo, który w ciągu kilku miesięcy stanie się częścią permanentnego archiwum.* Wynika z tego, że przeglądanie zarchiwizowanych stron i dalej podstron powoduje, że WM dodaje przeglądane strony do zarchiwizowanych wcześniej. Oczywiście tylko wtedy, kiedy są one jeszcze dostępne za darmo. W ten sposób strony cieszące się większym zainteresowaniem są archiwizowane dokładniej.

W wielu krajach świata, choć chyba niedostatecznie w Polsce, Wayback Machine cieszy się dużą popularnością, która przyczyniła się także do powstania rozmaitych problemów. Jednym z nich były procesy sądowe, wytoczone przez osoby czy instytucje, które nie chciały, aby zawartość ich stron była archiwizowana. Dziś WM respektuje takie żądania — wystarczy plik *no robots* na serwerze, a oprogramowanie WM nie będzie próbowało obejść tych zabezpieczeń. Innym problemem są żądania, aby usunęto zarchiwizowane strony, których treść nie podobała się powodowi. IA usunęło na żądanie scjentologów stronę poświęconą polemice z nimi, choć nikt nie usuwał stron propagujących tę religię. Ostatnio wiele kontrowersji wzbudza wzmożona aktywność radykalnych użytkowników islamskich, umieszczających w Internet Archive tak dużo materiałów, jak to tylko możliwe. Dotyczy to wszelkich dokumentów — audiowizualnych i tekstów oraz stron WWW.

Coraz częściej wykorzystuje się Internet, w tym WM, do celów sądowych, jednakże zwykle zachodzi potrzeba udowodnienia, że archiwum jest wiarygodne, a strona nie uległa modyfikacji podczas przechowywania na innym serwerze. Sędziowie mogą mieć w tej sprawie różne zdania, choć wydaje się, że strony internetowe coraz śmielej są przyjmowane jako materiał dowodowy w procesach sądowych. Jako ciekawostkę można dodać, że jednym z takich procesów był spór Telewizji Polskiej z pewnym operatorem telewizji satelitarnej w USA<sup>6</sup>.

Można tu wspomnieć także o problemie natury informatycznej — im nowsze strony, tym trudniej je przechowywać i prezentować w oryginalnej formie. Spowodowane jest to specyfiką nowych metod tworzenia stron WWW, zwykle interaktywnych, a także powiązanych skryptami i aplikacjami, których maszyna już nie archiwizuje, bo musiałaby skopiować zbyt dużo z serwera (de facto roboty archiwizują głównie tekst i obrazki bezpośrednio osadzone na stronie tradycyjnymi metodami).

Wayback Machine jest niezwykle ciekawym zjawiskiem. Siłą i zarazem słabością tego archiwum są jego ogromne rozmiary oraz pewien „wszystkoizm” — ambitna próba zebrania w jednej bazie stron internetowych z całego świata. Archiwum rozrasta się i potrzebuje coraz więcej miejsca na przechowywanie olbrzymiej ilości danych oraz coraz więcej funduszy na utrzymanie. Szlachetna intencja równego

---

<sup>6</sup> LEVITT C., ROSCH M. Making Internet searches part of due diligence. *Los Angeles Lawyer* [on-line]. Feb 2007, Vol. 29, No. 12 [Dostęp 07.01.2012]. Dostępny w World Wide Web: <http://www.lacba.org/Files/LAL/Vol29No12/2349.pdf>.

traktowania wszystkich internautów przez udostępnianie im serwerów jest ostatnio wystawiana na ciężką próbę konfrontacji z radykałami, nie zawsze o przekonaniach demokratycznych. Rozwój techniki informatycznej również komplikuje zadanie archiwizowania stron, ponieważ ich architektura uniemożliwia proste kopiowanie. Czas pokaże, jak B. Kahle i jego współpracownicy poradzą sobie z tymi wszystkimi problemami. Pewne jest jednak, że Wayback Machine, jak i cały serwis Internet Archive, to przedsięwzięcie wyjątkowe.

## UK Web Archive

Istniejące od 2004 r. UK Web Archive (UKWA<sup>7</sup>, <http://www.webarchive.org.uk/ukwa/>) współtworzą obecnie cztery organizacje: British Library, Biblioteka Narodowa Walii, Joint Information Systems Committee (JISC) oraz Wellcome Library (londyńska biblioteka medyczna). UKWA współpracuje także z innymi instytucjami przy tworzeniu kolekcji specjalnych. Archiwum posiada także interfejs w języku walijskim, którym w 2004 r. mówiło na co dzień 340 tys. Brytyjczyków. W UK jest więcej mniejszości językowych, ale tylko Narodowa Biblioteka Walii wspiera serwis. Biblioteka Narodowa Szkocji wycofała się z niego. Archiwum jest bardzo odmienne od Wayback Machine jeśli chodzi o zasięg zbieranego materiału, ponieważ jak wskazuje nazwa, jest to baza dotycząca wyłącznie stron internetowych Zjednoczonego Królestwa Wielkiej Brytanii i Irlandii Północnej.



Rys. 5. UK Web Archive

Źródło: UK Web Archive: preserving UK websites [on-line]. [Dostęp 07.01.2012]. Dostępny w World Wide Web: <http://www.webarchive.org.uk/ukwa/>.

<sup>7</sup> UK Web Archive statistics [on-line]. [Dostęp 07.01.2012]. Dostępny w World Wide Web: <http://www.webarchive.org.uk/ukwa/statistics>.

Kryteria doboru stron do archiwizacji w UKWA zostały określone następująco:

- materiał powinien być przydatny do badań naukowych przez dłuższy okres i odzwierciedlać różnorodność zainteresowań, działań oraz ogólnie życia w Zjednoczonym Królestwie,
- materiał powinien ukazywać innowacyjność stron internetowych.

Doborem zajmują się pracownicy instytucji tworzących UKWA, kierujący się polityką zgodną z polityką gromadzenia we własnych bibliotekach i własnymi kompetencjami (np. National Library of Wales zbiera strony walijskie). Wybrane strony mają ukazywać różne przejawy życia kulturalnego, politycznego, naukowego, religijnego i ekonomicznego. Twórcy bazy przyjmują rekomendacje internautów, ale sami podejmują decyzję w sprawie archiwizacji. Ta arbitralność jest kolejną różnicą pomiędzy „demokratycznym” archiwum amerykańskim, a bardziej zdyscyplinowanym, ale zdecydowanie autorytarnym archiwum brytyjskim.

Inną istotną różnicą formalną między WM a UKWA, która wprawdzie na razie nie ma zastosowania w praktyce, jest zasada gromadzenia zasobów, według której archiwum UKWA deklaruje działanie. Zgodnie z brytyjską Ustawą o egzemplarzu obowiązkowym (e.o.) z 2003 r. zasadniczo (*in principle*) biblioteki, które mają zagwarantowane prawo do e.o., powinny być uprawnione do gromadzenia stron internetowych bez uzyskiwania zgody właścicieli<sup>8</sup>. Jednakże w ustawie brakuje precyzyjnych zapisów, zatem w UKWA nie archiwizuje się niczego bez pozwolenia. Twórcy za każdym razem wysyłają (poczta elektroniczną lub zwykłą) prośbę o zgodę właścicieli stron, które chcą zarchiwizować. Automatycznie archiwizuje się strony, których właściciel zamieści informację o licencji Creative Commons lub w inny sposób zadeklarował, iż wyraża zgodę na rozpowszechnianie treści strony pod pewnymi warunkami. W przyszłości, jeśli prawo na to zezwoli, strony będą archiwizowane bez zgody właścicieli. W British Library pracuje się obecnie nad techniczną stroną takiego rozwiązania.

Powiązanie archiwum z egzemplarzem obowiązkowym upraszcza kryterium doboru, a zarazem — ze względu na nowość problemu — może prowokować liczne pytania. Tradycyjny e.o. to również gazety i czasopisma, których wydania internetowe są płatne w części lub całości. Archiwa i biblioteki mogłyby je udostępniać tylko w obrębie intranetu bądź dla zalogowanych czytelników, ale takie rozwiązanie musieliby zaakceptować właściciele czasopism. Należałoby się też zastanowić, które biblioteki udostępniłyby e.o. wydawnictw ciągłych przechowywany na serwerach UKWA — tylko British Library czy wszystkie brytyjskie biblioteki akademickie? Pojawia się też problem natury polityczno-prawnej. Ze względów historycznych i kulturowych wybrane biblioteki Zjednoczonego Królestwa i Republiki Irlandzkiej przesyłają sobie nawzajem e.o. druków zwartych. Zasada ta nie obowiązuje w UKWA, gdzie zarchiwizowano bardzo niewiele stron irlandzkich, choć w rzeczywistości jest dużo stron WWW dotyczących zwłaszcza Irlandii Północnej, posadowionych na serwerach nie tylko w Republice, lecz także w USA. Gromadzenie materiałów dotyczących danego kraju jest tradycją wszystkich bibliotek narodowych na świecie. Twórcy UKWA, którego filarem jest British Library, nie deklarują jednak takich zamiarów.

---

<sup>8</sup> *UK Web Archive. Legislative Status* [on-line]. [Dostęp 07.01.2012]. Dostępny w World Wide Web: [http://www.webarchive.org.uk/ukwa/info/about#legislative\\_status](http://www.webarchive.org.uk/ukwa/info/about#legislative_status).

UK Web Archive używa programów typu robot internetowy (*web crawler*), które podobnie jak roboty Google czy Alexy nieustannie przeczesują Internet. Narzędzie typu open source o nazwie Web Curator Tool (WCT) powstało przy współpracy National Library of New Zealand oraz British Library, na bazie również wolnego oprogramowania Heritrix, stworzonego w Internet Archive. Podobnie jak Wayback Machine, UKWA też ma kłopoty z archiwizacją stron bardziej skomplikowanych technicznie, tzn. zawierających skrypty Javy, wideo strumieniowe albo opierających swą funkcjonalność na wyszukiwarce wbudowanej w stronę zamiast mapy serwisu. Twórcy archiwum deklarują, że pracują nad rozwiązaniami komputerowymi, które pozwoliłyby archiwizować również i te witryny. Poza tym na razie roboty ściągają tylko określoną liczbę danych. Jeśli napotkają ich więcej, opuszczają stronę, co ma zapobiegać zatrzymaniu robota przez złośliwe oprogramowanie (niektóre strony WWW zawierają takie pułapki). Witryny archiwizowane są zasadniczo w odstępach półrocznych. Jeśli strony przypisze się do którejś z kolekcji specjalnych, wtedy treść jest archiwizowana częściej, jeśli uznano to za celowe<sup>9</sup>.

Statystyki na stronie UKWA mówią, że dotychczas (do 21.01.2012 r.) zarchiwizowano 9821 tytułów oraz 42 009 kopii wszystkich stron WWW, co daje przeciętnie cztery wersje jednej strony w różnych okresach czasu<sup>10</sup>. Cały materiał zajmuje 10,92 TB na dyskach komputerów (UKWA wynajmuje serwery komercyjne). Jest to wiele, ale wciąż mało w porównaniu z rozmiarami sieci. Zoë Corbyn podaje, że całą sieć brytyjską szacuje się na około 4 do 5 mln stron, więc około 10 tys. stron w UKWA to niewielki odsetek. Przytacza również wypowiedź kierowniczki archiwizacji sieci w British Library, która skarży się, że tylko 30% respondentów odpowiada pozytywnie na prośbę o zgodę na archiwizację strony<sup>11</sup>.

Niewątpliwą innowacją w funkcjonalności UK Web Archive w stosunku do Wayback Machine jest to, że brytyjskie archiwum można przeszukiwać, i to na kilka sposobów. Generalnie dostęp do zasobów UKWA odbywa się poprzez:

- wyszukiwanie według:
  - wyrażen z możliwością ograniczenia wyników do konkretnej daty, tematu (*subject*) czy kolekcji (*special collection*),
  - znanego adresu URL lub tytułu strony, również z możliwością ograniczenia wyników do daty, tematu czy kolekcji,
- przeglądanie:
  - indeksów,
  - dziedzin (*browse by subject*),
  - kolekcji (*browse by special collection*),
  - tytułów stron (*browse by website title*).

---

<sup>9</sup> Więcej szczegółów technicznych w: VILE D. Lessons from the library: behind the UK's web archive. *Computer Weekly* [on-line]. 21.03.2011 [Dostęp 21.01.2012]. Dostępny w World Wide Web: <http://www.computerweekly.com/news/1280097326/Lessons-from-the-library-Behind-the-UKs-web-archive>.

<sup>10</sup> *UK Web Archive statistics* [on-line]. 22.01.2012 [Dostęp 21.01.2011]. Dostępny w World Wide Web: <http://www.webarchive.org.uk/ukwa/statistics>.

<sup>11</sup> CORBYN Z. Memory failure detected. *The Times Higher Education* [on-line]. 01.09.2011 [Dostęp 21.01.2011]. Dostępny w World Wide Web: <http://www.timeshighereducation.co.uk/story.asp?storycode=417285>.



Wyniki wyszukiwania nie są niestety kumulowane, co powoduje w rezultacie setki wystąpień tej samej strony i dotarcie do innej wymaga wyświetlania list czasem kilkuset pozycji. Przeglądanie indeksów pozwala na wertowanie siedmiu szerokich kategorii: sztuka i nauki humanistyczne; biznes, ekonomia i przemysł; edukacja i badania naukowe; rząd, prawo i polityka; medycyna i zdrowie; nauki ścisłe i technologia oraz społeczeństwo i kultura. Jest też możliwość przeglądania 37 kolekcji specjalnych zarówno bardzo ogólnych, jak i szczegółowych, np. wybory powszechne w 2010 r. podzielone na 11 (!) podsekcji. W każdej kolekcji znajduje się 50-400 witryn. Przykładowe kolekcje to: blogi, kryzys gospodarczy, problemy kobiet i mężczyzn, zdrowie psychiczne, igrzyska olimpijskie i paraolimpijskie. Niektóre strony w kolekcjach były archiwizowane jednorazowo, bo przestały być aktywne, jednakże uznano je za wartościowe. Inne są archiwizowane z częstotliwością zgodną z decyzją opiekunów określonych kolekcji.

Ostatecznym rezultatem wyszukiwania czy przeglądania są miniaturki zarchiwizowanych stron, które można dalej przeglądać w takiej wersji, w jakiej zostały skopiowane. Jeśli uda nam się trafić na stronę archiwizowaną wielokrotnie, rezultat będzie podobny jak na rys. 6. Strona festiwalu filmowego w Edynburgu archiwizowana była w latach 2005-2006. Później tego zaniechano, choć festiwal wciąż działa. Być może powodem jest wycofanie się Scottish National Library z projektu UK Web Archive?

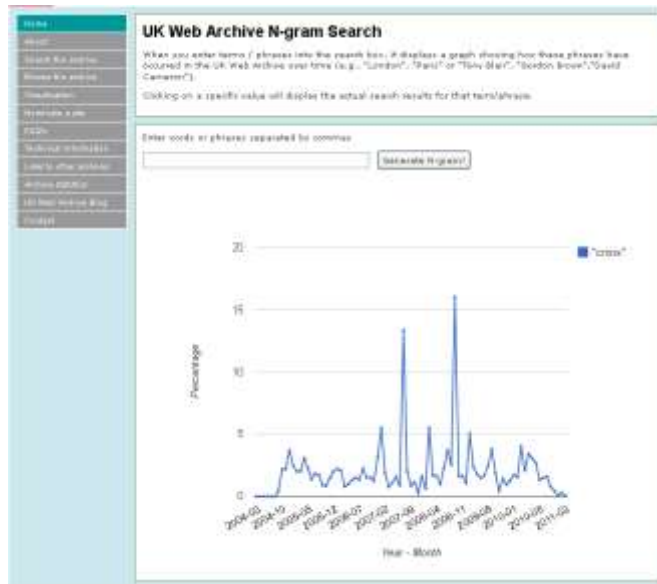


Rys. 6. Rezultaty szukania Edinburgh International Film Festival w UK Web Archive  
Źródło: UK Web Archive: preserving UK websites [on-line]. [Dostęp 07.01.2012]. Dostępny w World Wide Web: <http://www.webarchive.org.uk/ukwa/>.

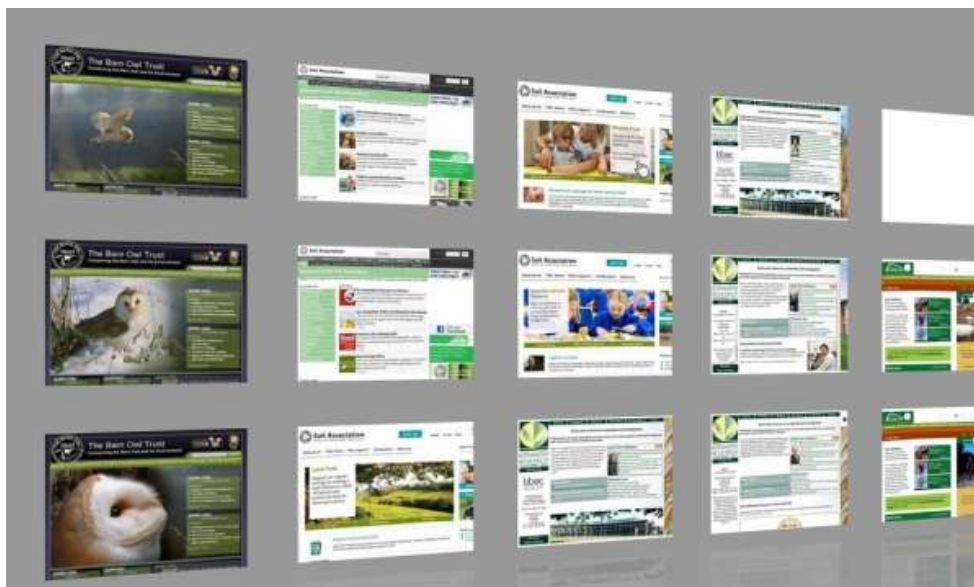
UK Web Archive dostarcza dodatkowo atrakcyjne formy wizualizacji rezultatów wyszukiwania:

- generowanie wykresów częstotliwości występowania terminów/wyrażeń na witrynach zarchiwizowanych w UKWA (rys. 7.),

- chmury tagów występujących na razie w jednej kolekcji “General Election 2005 Collection”,
- efektowna prezentacja kolekcji specjalnych — miniaturki witryn w 3D (rys. 8.).



Rys. 7. Częstotliwość występowania słowa „crisis” w Internecie brytyjskim w latach 2004-2011  
Źródło: UK Web Archive: preserving UK websites [on-line]. [Dostęp 07.01.2012]. Dostępny w World Wide Web: <http://www.webarchive.org.uk/ukwa/>.



Rys. 8. Kolekcja British Countryside w UK Web Archive, widok 3D  
Źródło: UK Web Archive: preserving UK websites [on-line]. [Dostęp 07.01.2012]. Dostępny w World Wide Web: <http://www.webarchive.org.uk/ukwa/>.

## Zakończenie

Porównanie Internet Archive i UKWA pod względem podobieństwa zarchiwizowanych zasobów nie przyniosło sukcesów. Jedyne w UKWA

zarchiwizowane strony na temat republikanów i lojalistów irlandzkich pochodziły z naukowego archiwum konfliktu północnoirlandzkiego CAIN lub ze stron rządowych oraz partii politycznych. UKWA archiwizuje BBC News i serwis CAIN, tak więc prezentowany jest głównie oficjalny pogląd na konflikt, z niewielką liczbą skromnie zarchiwizowanych blogów. Strony rządowe są w archiwum brytyjskim zdecydowanie najliczniej archiwizowane — szukając czegoś na temat polityki lub zdrowia, trzeba przejrzeć listę pierwszych kilkuset stron państwowych, aby dotrzeć do prywatnych.

Do dyskusji i przemyśleń czytelników pozostaje problem, która metoda jest lepsza — ścisła polityka doboru zasobów w imię poszanowania nie tylko prawa, lecz także pewnych zasad, czy kompletny jej brak w imię pełnej wolności. A może powinna istnieć jakaś trzecia droga? Łatwo krytykować bibliotekarzy brytyjskich za nadmierną ostrożność. Nie ulega wątpliwości, że jeśli archiwum jest prowadzone przez organizację finansowaną przez państwo, to państwo może stawiać warunki. Pytanie, jak daleko może posunąć się taka ingerencja. Respektowanie prawa jest oczywiste, lecz istnieją także zasady (tzw. dobre praktyki chyba dopiero są w trakcie tworzenia, bo idea takich archiwów jest dość nowa) lub — co już mniej oczywiste — aktualne wytyczne zgodne z bieżącą polityką. Ostatnie rewelacje na temat materiałów Al Kaidy w zbiorach Internet Archive też dają do myślenia na temat granic wolności słowa w Internecie. Czas pokaże, czy powstaną kiedykolwiek jakieś globalne wspólne wytyczne dotyczące polityki gromadzenia archiwów internetowych. Nie wydaje się to bardzo prawdopodobne i być może tak jest lepiej... Jedno jest pewne, świat cyfrowy pozbawiony archiwów stron byłby skazany na żywot motyla lub, jak to określono w dokumentach europejskich, cyfrowe średniowiecze<sup>12</sup>.

#### **Bibliografia:**

- [1] *About the Internet Archive* [on-line]. [Dostęp 07.01.2012]. Dostępny w World Wide Web: <http://www.archive.org/about/about.php>.
- [2] CORBYN Z. Memory failure detected. *The Times Higher Education* [on-line]. 01.09.2011 [Dostęp 21.01.2011]. Dostępny w World Wide Web: <http://www.timeshighereducation.co.uk/story.asp?storycode=417285>.
- [3] LEVITT C., ROSCH M. Making Internet searches part of due diligence. *Los Angeles Lawyer* [on-line]. Feb 2007, Vol. 29, No 12 [Dostęp 07.01.2012]. Dostępny w World Wide Web: <http://www.lacba.org/Files/LAL/Vol29No12/2349.pdf>.
- [4] Naszym celem jest, aby Europa miała zapewniony cyfrowy renesans, zamiast pogrążyć się w cyfrowym średniowieczu. W: *The New Renaissance* [on-line]. Brussels, 10 January 2011 [Dostęp 17.01.2012]. Dostępny w World Wide Web: [http://ec.europa.eu/culture/documents/report\\_comite\\_des\\_sages.pdf](http://ec.europa.eu/culture/documents/report_comite_des_sages.pdf).
- [5] The Internet's librarian. *The Economist* [on-line]. 05.03.2009 [Dostęp 07.01.2012]. Dostępny w World Wide Web: <http://www.economist.com/node/13174399>.
- [6] *UK Web Archive. Legislative Status* [on-line]. [Dostęp 07.01.2012]. Dostępny w World Wide Web: [http://www.webarchive.org.uk/ukwa/info/about#legislative\\_status](http://www.webarchive.org.uk/ukwa/info/about#legislative_status).
- [7] *UK Web Archive statistics* [on-line]. 22.01.2012 [Dostęp 21.01.2011]. Dostępny w World Wide Web: <http://www.webarchive.org.uk/ukwa/statistics>.

---

<sup>12</sup> Naszym celem jest, aby Europa miała zapewniony cyfrowy renesans, zamiast pogrążyć się w cyfrowym średniowieczu. W: *The New Renaissance* [on-line]. Brussels, 10 January 2011 [Dostęp 17.01.2012]. Dostępny w World Wide Web: [http://ec.europa.eu/culture/documents/report\\_comite\\_des\\_sages.pdf](http://ec.europa.eu/culture/documents/report_comite_des_sages.pdf).

- [8] VILE D. Lessons from the library: behind the UK's web archive. *Computer Weekly* [on-line]. 21.03.2011 [Dostęp 21.01.2012]. Dostępny w World Wide Web: <http://www.computerweekly.com/news/1280097326/Lessons-from-the-library-Behind-the-UKs-web-archive>.

---

Gmerek, K. Archiwa internetowe po obu stronach Atlantyku — Internet Archive, Wayback Machine oraz UK Web Archive. W: *Biuletyn EBIB* [online] 2012, nr 1 (128), *Archiwizacja Internetu* [Dostęp: 24.02.2012] Dostępny w World Wide Web: [http://www.nowyebib.info/images/stories/numery/128/128\\_gmerek.pdf](http://www.nowyebib.info/images/stories/numery/128/128_gmerek.pdf). ISSN 1507-7187.