

# Michał T. Szczepański

---

Cyfrowa kolekcja czasopism  
humanistycznych (Bazhum)  
<http://www.muzhp.pl/bazhum>

---

Przegląd Historyczny 101/1, 101-106

---

2010

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej [bazhum.muzhp.pl](http://bazhum.muzhp.pl), gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach  
dozwolonego użytku.

MICHAŁ T. SZCZEPAŃSKI  
Muzeum Historii Polski  
w Warszawie

## Cyfrowa kolekcja czasopism humanistycznych (Bazhum) <http://www.muzhp.pl/bazhum>

Ogromna część dorobku nauk humanistycznych i społecznych została utrwalona poprzez publikację na łamach czasopism naukowych. Tworzą one obszerny zasób informacji i wiedzy, którego wartości nie sposób przecenić. Niestety jest on w tej samej mierze rozległy, co trudno dostępny. Ograniczenia narzuca sama forma publikacji — większość tekstów istnieje wyłącznie w postaci papierowej. Korzystanie z pełnego korpusu jest ideą fantastyczną nawet w dużych ośrodkach akademickich. Brakuje efektywnych narzędzi umożliwiających szybkie przeszukiwanie i przetwarzanie zawartości. W 1928 r. Marceli Handelsman rozwijając pojęcie heurystyki jako „działu metodologii, którego najłatwiej się nauczyć” dzielił się praktyczną refleksją: „wszelkie poszukiwania winny być prowadzone systematycznie, planowo i celowo, a mimo to nigdy nie można być pewnym, czy dadzą one rezultat oczekiwany — —. Tem większe więc znaczenie posiadać musi sumienne przygotowanie techniczne, umożliwiające planowość poszukiwań”<sup>1</sup>. Pedagogiczny ton nabiera nowego znaczenia w dobie cyfrowej i sieciowej komunikacji, gdy sumienne przygotowanie techniczne nie nadąża za narastającym zasobem wiedzy.

Ograniczony lub utrudniony dostęp do treści naukowej niekorzystnie wpływa na przyswajanie i rozwój wiedzy, a właśnie korzystanie z polskich czasopism naukowych wymaga pokonywania różnorodnych trudności. Warto przy okazji zauważyć, że problemy związane z zapewnieniem dostępu do aktualnych wyników badań lub swobodnej partycypacji w dyskusji naukowej (a więc kwestie wyznaczające rytm debat o otwartym dostępie do zasobów naukowych) w polskich warunkach stanowią ledwie fragment zagadnienia otwartości zasobów naukowych. Z punktu widzenia dyscyplin humanistycznych i społecznych równie istotna wydaje się potrzeba zapewnienia w pełni funkcjonalnego dostępu do starszego dorobku naukowego. Środkiem do zapewnienia większej a nawet całkowitej otwartości tych zasobów może być tylko pełna digitalizacja czasopism i budowa zasobów

---

<sup>1</sup> M. H a n d e l s m a n, *Historyka. Zasady metodologii i teorii poznania historycznego. Podręcznik dla szkół wyższych*, wyd. 2, Warszawa 1928, s. 41.

cyfrowych<sup>2</sup>. Wymaga to stworzenia infrastruktury w postaci baz danych i repozytoriów, nastawionej na wspieranie bieżącej działalności badawczej i dydaktycznej. W kontekście tych zmian forma publikacji periodycznej powinna zostać przemyślana na nowo. Zarazem jest to okazja do skonstatowania przemian w obiegu treści naukowej na obszarze nauk humanistycznych i społecznych, niezależnie od obserwacji czynionych na gruncie nauk przyrodniczych<sup>3</sup>.

Obecny stan opracowania i udostępniania polskich czasopism naukowych (humanistycznych i społecznych) w formie cyfrowej jest niezadowolający<sup>4</sup>. Stanowi to bodziec do poszukiwania rozwiązań zaawansowanych pod względem technologicznym i informacyjnym, a zarazem łatwych w adaptacji do bieżących potrzeb i praktyki naukowej. Założenia funkcjonalne dla narzędzi wspomagających korzystanie z czasopism należałoby wyprowadzać z obserwacji rzeczywistych sposobów pracy z tekstami w periodykach<sup>5</sup>. Trzeba zatem wziąć pod uwagę, że materiały publikowane w czasopismach ukazują się w ramach uporządkowanej struktury z wyraźnie wyodrębnionymi szczeblami hierarchii (rocznik, tom, zeszyt, numer itp.). Odwzorowanie tego układu w procesie opracowania cyfrowego polega na naśladowaniu relacji określonych przez wydawcę. Tylko pozornie banalny jest fakt, że czytelnicy na ogół nie przywiązują wagi do odmian wydawniczych artykułów z czasopism. Różnice w edycjach (np. odmienne egzemplarze tego samego numeru) traktowane są jako mutacje od podstawowego wzoru, bez istotnej wartości poznawczej. Najwidoczniej stopień intuicyjnego abstrahowania od fizycznego nośnika treści jest większy niż w przypadku książki. Dlatego też identyfikacja zawartości periodyków odbywa się głównie poprzez odniesienie do właściwego miejsca w hierarchii publikacji. Wreszcie, na koniec, istotny przedmiot zainteresowania dla większości czytelników stanowi pojedynczy artykuł, rzadziej cały numer lub tom. Strukturalny układ zawartości nie jest postrzegany jako część przekazywanej wiedzy, pełni tylko funkcję porządkującą.

Wychodząc od tych spostrzeżeń można stwierdzić, że krytycznym momentem w digitalizacji czasopism będzie wytworzenie struktury opisowej dotyczącej kompletnej zawartości danego tytułu. Powinna ona zawierać wykaz wszystkich tekstów opublikowanych na łamach periodyku wraz z informacjami identyfikującymi poszczególne elementy (metadane bibliograficzne), a także opis relacji wią-

<sup>2</sup> Na ten temat m.in. M. Niezgodka, *Polska Biblioteka Wirtualna Nauki: program i jego perspektywy*, „Archeion”, t. CVII, 2004, s. 354 nn.

<sup>3</sup> A. Odlyzko, *The future of scientific communication*, [w:] *Access to Publicly Financed Research: The Global Research Village*, t. III, Amsterdam 2000, s. 273 nn.

<sup>4</sup> Raport o digitalizacji dóbr kultury opracowany na zlecenie Ministerstwa Kultury i Dziedzictwa Narodowego na Kongres Kultury Polskiej 2009, znany też jako *Program digitalizacji dóbr kultury oraz gromadzenia, przechowywania i udostępniania obiektów cyfrowych w Polsce 2009–2020* (Warszawa, wrzesień 2009), rozdz. 2, omawia kwestię nader pobieżnie i ogólnikowo.

<sup>5</sup> Wokół prostoty i wykorzystania intuicji dotyczących pracy z tradycyjną dokumentacją D. Geirnter, *Mechaniczne piękno. Kryterium estetyczne w informatyce*, Warszawa 1999, s. 57.

żących poszczególne składowe<sup>6</sup>. Czasopismo zostanie wyobrażone w postaci hierarchicznej struktury roczników, tomów, zeszytów itp. podporządkowanej tytułowi jako jednostce nadrzędnej, gdzie artykuły tworzą podstawowy, najniższy szczebel hierarchii. Dalszy proces polega na poszerzaniu warstwy opisowej (metadanych) oraz przyłączaniu pełnych tekstów w wersji cyfrowej. Wytworzenie kompletnego rejestru zawartości na wstępie opracowania cyfrowego gwarantuje funkcjonalność kolekcji niezależnie od etapu działań. Zaspokaja ona pewne potrzeby informacyjne jako bibliograficzna baza danych, zarazem oddaje do dyspozycji konstrukcję elastyczną i otwartą na dodatkowe zasoby. Jest właściwym punktem wyjścia do budowy w pełni wartościowego narzędzia naukowego, jakimi jest baza pełnotekstowa<sup>7</sup>.

Bazy zawartości czasopism ze względu na wysoki stopień ustrukturyzowania i powiązania danych tworzą zasób o dużym potencjale heurystycznym. Zakres ich użyteczności jest zdeterminowany przez dostęp do narzędzi wspomagających konstruowanie kwerend oraz przetwarzanie danych (wyszukiwarki, algorytmy, kwantyfikatory logiczne). Podstawowymi warunkami efektywnych działań są jednak zakres i koherencja materiału zgromadzonego w kolekcji. Opracowanie czasopism w pełnym zakresie — od początku ukazywania się do ostatnich numerów — wzmacnia spójność danych. Podobnie zasoby obejmujące treści naukowe w ramach obszernych dziedzin wiedzy (np. czasopisma medyczne, przyrodnicze lub humanistyczne) mają przewagę nad bazami konstruowanymi według wytycznych subiektywnych, niejako zewnętrznych wobec gromadzonych danych (np. wybranych problemów badawczych, tematów, pojęć geograficznych itp.). Ze względu na efektywność kwerend wskazane byłoby tworzenie zbiorów obejmujących dużą liczbę tytułów<sup>8</sup>.

Zasada otwartego dostępu (*open access*) do treści naukowych jest ważnym punktem odniesienia przy tworzeniu kolekcji. Swobodny i nielimitowany dostęp do wszystkich publikacji (natychmiast lub po upływie okresu karencji) przyczynia się do rozpowszechnienia i tworzenia wiedzy. Jest korzystny dla autorów, czytelników i wszystkich użytkowników kolekcji. Nie trzeba dodawać, że prezentacja tekstów powinna odbywać się z całkowitym poszanowaniem praw autorskich. Z uwagi na podstawowe znaczenie dla obiegu informacji naukowej bazy powinny

---

<sup>6</sup> Cf. *Standardy w procesie digitalizacji obiektów dziedzictwa kulturowego*, red. G. PłoS z a j s k i, Warszawa 2008, s. 66; R. T. P r i n k e, *Standardy i formaty danych w digitalizacji zasobów archiwalnych i bibliotecznych*, „Archeion”, t. CVII, 2004, s. 219 nn.

<sup>7</sup> Cf. J. S a d o w s k a, *Polskie dziedziny bibliograficzne bazy danych w perspektywie lokalnej i globalnej*, [w:] *Bibliograficzne bazy danych : kierunki rozwoju i możliwości współpracy*, Bydgoszcz, 27–29 maja 2009, EBIB Materiały konferencyjne, 19, Warszawa 2009, <<http://www.ebib.info/publikacje/matkonf/mat19/sadowska.php>>.

<sup>8</sup> Cf. M. K o w a l s k a, *Zasoby czasopiśmiennicze w polskich bibliotekach cyfrowych i bibliograficznych bazach danych — koegzystencja, konkurencja czy kooperacja?* [w:] *ibidem*, <<http://www.ebib.info/publikacje/matkonf/mat19/kowalska.php>>.

być oparte na konstrukcjach informatycznych o otwartym kodzie źródłowym, z zachowaniem szerokiej możliwości ingerowania i wprowadzania modyfikacji.

W kontekście zarysowanych obserwacji, jako odpowiedź na wyłaniające się postulaty powstał projekt skonstruowania bazy informatycznej obejmującej zawartość polskich czasopism naukowych z zakresu nauk humanistycznych i społecznych. Został on zainicjowany przed trzema laty jako przedsięwzięcie badawcze Muzeum Historii Polski i obecnie jest prowadzony przy współpracy z Interdyscyplinarnym Centrum Modelowania Matematycznego i Komputerowego na Uniwersytecie Warszawskim pod roboczą nazwą Bazhum. Zgromadzony w bazie zespół treści jest dostępny w internecie w sposób bezpłatny i bez ograniczeń pod adresem <http://www.muzhp.pl/bazhum>. Witryna umożliwia wgląd w rejestr treści ponad stu czasopism humanistycznych. Odnotowane są wszystkie materiały opublikowane na łamach periodyków, a zatem artykuły, artykuły recenzyjne, recenzje, zapiski, miscellanea, nekrologi, różnego rodzaju kroniki i wspominki oraz każda inna forma publikacji. Opisywanie zasobów w tak szerokim zakresie wynika z charakteru kolekcji jako narzędzia wspomagającego pracę naukową. Interpretacja treści oraz jej ocenę merytoryczną pozostawiono użytkownikowi (badaczowi). Rejestr, co do zasady, ma odwzorowywać układ tekstów pomieszczonych w zeszytach przez redakcję. Czasopisma są opracowane w ujęciu retrospektywnym, tak aby można było z nich korzystać w pełnym zakresie od pierwszych numerów aż do ostatnich, najnowszych. Informacje o najnowszych publikacjach dodawane są na bieżąco w miarę ukazywania się tytułów na rynku. Baza bibliograficzna jest stopniowo wzbogacana o pełne teksty artykułów z niektórych czasopism, w tym przede wszystkim o zawartość „Przeglądu Historycznego”. Wraz z narastaniem zasobu pełnotekstowego Bazhum będzie podążał w kierunku wytyczonym przez takie kolekcje zawartości czasopism jak np. JSTOR. Z jednym jednak zastrzeżeniem: ma on pozostać bazą powszechnie dostępną i bezpłatną.

Bazhum gromadzi przede wszystkim treści o charakterze naukowym, przy czym za naukowe uznawane są wszystkie czasopisma spełniające to kryterium w oczach współczesnych powstawianiu tych opracowań. Rejestracji podlega zawartość czasopism polskich, język publikacji nie ma znaczenia. W kolekcji znajdują się periodyki ukazujące się na ziemiach polskich w obecnych lub w dawnych granicach państwa polskiego, jak również wytworzonych przez polskie instytucje za granicą. Czasopisma ukazujące się dawniej, a dziś już niekontynuowane, wchodzą do zasobu na tych samych zasadach, jak tytuły o nieprzerwanej ciągłości. Wreszcie zakres dyscyplin humanistycznej i społecznej wyznaczają dziś uznane znaczenia tych pojęć. Bazhum gromadzi opisy bibliograficzne tekstów (np. artykułów lub recenzji). Zasady rejestracji metadanych opierają się na polskiej normie bibliograficznej oraz formacie MARC 21<sup>9</sup>. Wersje pełnotekstowe są ściśle powią-

<sup>9</sup> Cf. *Format MARC 21 rekordu bibliograficznego dla artykułu z czasopisma: wersja 2008*, oprac. G. Jaroszewicz, M. Krynicka, B. Wieczorkowska, Warszawa 2008; M. Janowska, *Opis bibliograficzny artykułów (interpretacja postanowień PN-N-01152)*, Warszawa 1997.

zane z odpowiednimi metadanymi. Teksty w postaci plików elektronicznych w formacie PDF można przeglądać w internecie oraz kopiować. Po uruchomieniu kwerendy baza przeszukuje warstwę opisową (metadane bibliograficzne), a następnie również wszystkie dostępne zasoby pełnotekstowe. Rozbudowane możliwości konstruowania zapytań czynią zbędnymi stosowanie słów kluczowych oraz systemu haseł przedmiotowych<sup>10</sup>.

Baza Bazhum jest umieszczona w repozytorium cyfrowym stworzonym przez ICM, które pozwala na przechowywanie, przetwarzanie i udostępnianie treści cyfrowych w internecie. Zasoby można przeglądać i przeszukiwać za pomocą przeglądarki internetowej. Podstawowym ekranem wyświetlającym się po wejściu na stronę repozytorium jest panel wyszukiwania. W górnej strefie znajduje się pasek szybkiego wyszukiwania oraz menu poziome z pozycjami „Szukaj” i „Przełóżaj”. Pod hasłem „Pomoc” znajduje się obszerny opis wszystkich funkcji dostępnych w serwisie. Pasek szybkiego wyszukiwania i menu poziome są stale widoczne i dostępne z każdej strony repozytorium. Kwerendy tworzymy przez wpisanie słów do pól w panelu wyszukiwania. Wpisanie haseł do pól o mniej lub bardziej szczegółowym opisie pozwala na konstruowania poszukiwań o różnym stopniu złożoności. Zasoby można przeglądać pod kątem udostępnionych czasopism oraz autorów i wydawców opracowanych treści. W przypadku czasopism po uzyskaniu dostępu do wykazu zarejestrowanych tytułów mamy możliwość zapoznania się z kompletną strukturą publikacji w rozbiciu na lata i tomy. Nawigowanie po wykazie umożliwia dotarcie do poszczególnych artykułów, które są podstawowym, z tego punktu widzenia najniższym elementem hierarchii.

Program załączony do bieżącego numeru „Przełóżaj Historycznego” ma być również egzemplifikacją niektórych cech i możliwości tkwiących w kolekcji Bazhum. Jest to baza zawartości czasopisma za okres 1905–2009 (sto roczników) zawierająca opisy bibliograficzne wszystkich artykułów i analogicznych jednostek tekstowych z tego okresu<sup>11</sup>. Została ona wyposażona w interfejs do przetwarzania danych. W ten sposób umożliwia przeglądanie i przeszukiwanie opisów według swobodnie określonych kryteriów. Pozwala także na eksportowanie i drukowanie wyników kwerendy zgodnie z preferowanym sposobem zapisu bibliograficznego w jednym z kilku dostępnych formatów. Program działa na płycie CD, nie wymaga instalacji na komputerze. Podstawowa witryna programu składa się z paska tytułowego, belki funkcyjnej oraz dwóch paneli do prezentowania treści. Zasadnicza

---

<sup>10</sup> Na temat metod opracowania rzeczowego na tle potrzeb użytkowników oraz w kontekście wyszukiwania pełnotekstowego instruktywnie W. Klenczon, *Indeksowanie zawartości treściowej w bibliograficznych bazach danych. Bibliograficzne bazy danych : kierunki rozwoju i możliwości współpracy*, Bydgoszcz, 27–29 maja 2009, EBIB Materiały konferencyjne, 19, Warszawa 2009, <<http://www.ebib.info/publikacje/matkonf/mat19/klenczon.php>>.

<sup>11</sup> Baza znajduje się na płycie załączonej do niniejszego numeru „Przełóżaj Historycznego” [przyp. red.].

organizacja informacji jest taka sama na każdym etapie pracy. W pasku tytułu znajdują się nazwy i logotypy instytucji twórców aplikacji. Belka funkcyjna zawiera przyciski uruchamiające główne funkcje programu. Panele dzielą pomiędzy siebie treść w taki sposób, że lewy wyświetla informacje nadrzędne i ogólne, prawy — szczegółowe i konkretne. Przycisk „pomoc” na belce funkcyjnej przenosi użytkownika do opisu wszystkich funkcji dostępnych w programie. W lewym panelu ukazują się nazwy rozdziałów, w prawym — w zależności od wybranego rozdziału — tekst objaśnienia. Opcja pomocy zawiera instrukcję, w jaki sposób należy korzystać z dostępnych funkcji oraz jakich efektów należy się po nich spodziewać.

Projekt Bazhum stanowi próbę zarejestrowania, opracowania i udostępnienia znacznej części dawnego i współczesnego dorobku polskiej nauki. Jest to próba stworzenia wygodnej, prostej w użyciu i szeroko dostępnej platformy dostępu do obszernego zasobu informacji zgromadzonej w periodykach naukowych z dziedzin humanistycznych i społecznych. Poprzez ułatwienie poszukiwań bibliograficznych Bazhum umożliwia skupienie uwagi na czynnościach ściśle badawczych, wzmacniając intelektualny aspekt tworzenia i przyswajania wiedzy. Priorytetem dla bazy jest jej użyteczność do pracy badawczej i dydaktycznej, dlatego też zaspokojenie potrzeb użytkownika (np. badacza, nauczyciela, studenta) stanowi właściwe kryterium oceny słuszności przyjętych założeń, zwłaszcza gdy chodzi o zawartość kolekcji, szczegółowość dostępnej informacji bibliograficznej, możliwości konstruowania kwerend itp. Projekt jest w fazie rozbudowy: powiększa się lista opracowanych czasopism i poszerza dostęp do artykułów. Trwają prace nad zespołem narzędzi interfejsowych wspomagających pracę na zasobie danych.