

Cyfrowa Biblioteka Druków Ulotnych Polskich i Polski Dotyczących z XVI, XVII i XVIII wieku – prezentacja projektu i jego zastosowania w pracy naukowej oraz dydaktyce

WŁODZIMIERZ GRUSZCZYŃSKI¹, MACIEJ OGRODNICZUK²

¹ Szkoła Wyższa Psychologii Społecznej

² Instytut Podstaw Informatyki Polskiej Akademii Nauk

Streszczenie

Projekt „Cyfrowa Biblioteka Druków Ulotnych Polskich i Polski Dotyczących z XVI, XVII i XVIII w.” miał na celu stworzenie dostępnej nieodpłatnie w Internecie biblioteki zawierającej możliwie wszystkie zachowane i opisane w literaturze starodruki mające charakter gazet ulotnych. Zostały one zebrane głównie na podstawie zawierającej opisy prawie 2 tys. jednostek bibliograficznych pracy Konrada Zawadzkiego „Gazety ulotne polskie i Polski dotyczące od XVI-XVIII wieku”, rozszerzonej podczas kwerendy archiwalnej o pozycje odkryte lub opisane już po jej wydaniu. Dla dużej liczby druków (niespełna 70%) udostępniono jednocześnie w formacie DjVu ich skany wykonane z mikrofilmów znajdujących się w zbiorach Biblioteki Narodowej w Warszawie. Wybrane druki (polsko- i niemieckojęzyczne) zostały skomentowane pod względem historycznym, medioznawczym i/lub lingwistycznym w celu objaśnienia mniej znanych realiów, których dotyczy tekst zamieszczony w danym druku, oraz nieczytelnych dla współczesnego odbiorcy skrótów, metafor i symboli. W wypadku tekstów tłumaczonych lub wzorowanych na tekstach obcojęzycznych, wskazane zostały teksty źródłowe (lub za takie uchodzące) oraz teksty pochodne, by ułatwić porównywanie różnojęzycznych wersji informacji dotyczących tego samego wydarzenia, a wywodzących się zapewne z jednego źródła. Tego rodzaju komentarze nie są obecne w dotychczasowych bibliotekach cyfrowych i stanowią novum projektu. Osobnym wynikiem projektu było dostosowanie platformy EPrints do wyszukiwania druków według różnych kryteriów, zarówno takich, które standardowo wykorzystywane są w opisach bibliograficznych, jak i mniej typowych: tematycznych, gatunkowych i innych. W oparciu o tekst oryginalny uzupełniony o źródła wykryte w czasie trwania projektu wykonano ponadto elektroniczny skład typograficzny rozszerzonej wersji bibliografii Zawadzkiego. Projekt został zrealizowany w Instytucie Dziennikarstwa Uniwersytetu Warszawskiego i był dofinansowany ze środków Ministerstwa Kultury i Dziedzictwa Narodowego oraz Fundacji na rzecz Rozwoju Szkolnictwa Dziennikarskiego. Zbiory Biblioteki dostępne są pod adresem www.cbdu.pl oraz za pośrednictwem Federacji Bibliotek Cyfrowych. Biblioteka została zaprojektowana z myślą o konkretnych zastosowaniach naukowych i dydaktycznych. Potrzeba takiej biblioteki została zauważona w trakcie prac nad „Słownikiem języka polskiego XVII i 1. poł. XVIII w.”. Przez redakcję tego słownika systematycznie wykorzystywane są polskojęzyczne druki zawarte w bibliotece. CBDU znalazła też zastosowanie w badaniach nad dawnymi przekładami medialnymi prowadzonymi w Instytucie Lingwistyki Stosowanej i w Instytucie Germanistyki UW. W dydaktyce CBDU przydaje się przede wszystkim na studiach dziennikarskich w ramach przedmiotu „Historia prasy polskiej”.

Słowa kluczowe: druki ulotne, gazety ulotne, starodruki, biblioteka cyfrowa, EPrints

1. Prezentacja projektu

„Cyfrowa Biblioteka Druków Ulotnych Polskich i Polski Dotyczących z XVI, XVII i XVIII w.” powstała dzięki grantowi Ministerstwa Kultury i Dziedzictwa Narodowego (dofinansowanemu przez Fundację na rzecz Rozwoju Szkolnictwa Dziennikarskiego) w ramach programu Dziedzictwo Kulturowe, a w nim tzw. priorytetu 4. Tworzenie zasobów cyfrowych dziedzictwa kulturowego, ze środków pochodzących z Funduszu Promocji Kultury. Zadanie realizowane było w okresie od 1 stycznia 2009 r. do 31 grudnia 2009 r., na podstawie umowy nr 6395/09/FPK/DDK, zawartej w dniu 27.04.2009 r., pomiędzy Ministrem Kultury Narodowej i Dziedzictwa Narodowego a Uniwersytetem Warszawskim¹.

¹ Faktycznie, jak wynika to z przytoczonych dat realizacja projektu trwała tylko osiem miesięcy (od początku maja do końca grudnia 2009 r.).

1.1. Historia

Celem projektu „Cyfrowa Biblioteka Druków Ulotnych Polskich i Polski Dotyczących z XVI, XVII i XVIII w.” było udostępnienie w Internecie cyfrowej postaci możliwie wszystkich tego typu druków² zachowanych i opisanych w literaturze przedmiotu. Powody, które sprawiły, że zainteresowaliśmy się właśnie takimi starodrukami, miały dwojaki charakter. Po pierwsze, już od dość dawna zgłaszane było zapotrzebowanie na ułatwienie dostępu do tych rzadkich i rozproszonych druków przez następujące środowiska naukowe:

- 1) zespoły opracowujące słowniki historyczne języka polskiego:
 - *Słownik języka polskiego XVII i 1. poł. XVIII wieku* w Instytucie Języka Polskiego PAN,
 - *Słownik polszczyzny XVI wieku* w Instytucie Badań Literackich PAN,
- 2) historycy prasy zarówno polskiej, jak i niemieckiej, m.in. w Instytucie Dziennikarstwa UW,
- 3) historycy przekładu, m.in. w następujących ośrodkach:
 - Instytut Lingwistyki Stosowanej UW,
 - Instytut Germanistyki UW.

Po drugie, istnieje pełna, niezwykle dokładnie zestawiona bibliografia interesujących nas druków, mianowicie trzytomowe dzieło Konrada Zawadzkiego „Gazety ulotne polskie i Polski dotyczące od XVI-XVIII wieku”³. Była więc gotowa podstawa biblioteki, którą postanowiliśmy uzupełnić podobiznami druków. Przy okazji planowaliśmy scalenie, poprawienie i ewentualne uzupełnienie bibliografii Konrada Zawadzkiego. Bibliografia ta, zestawiana przez kilkanaście lat, ma układ chronologiczny, a jej ostatni tom stanowi swego rodzaju suplement dwu pierwszych tomów, co nieuchronnie narusza założony przez autora układ chronologiczny.

Od początku było jasne, że w bibliotece będziemy chcieli zawrzeć więcej danych o zamieszczanych tam dokumentach, niż robi się to w bibliotekach czy to standardowych, czy cyfrowych. W szczególności – idąc za Zawadzkim – chcieliśmy poszerzyć informacje historyczne dotyczące faktów przedstawianych w danym dokumencie oraz uzupełnić i wzbogacić informacje o powiązaniach pomiędzy poszczególnymi dokumentami (zwłaszcza wtedy, gdy są one przekładami). Zupełną nowością stały się słowniczki trudniejszych wyrazów i zwrotów dołączane do wybranych dokumentów. Tym samym sposób opracowania zamieszczanych w bibliotece dokumentów miał przypominać w jakimś stopniu sposób opracowania charakterystyczny dla naukowych (lub dydaktycznych) edycji starodruków. Jak już wspomnieliśmy, planowane było przez nas również uzupełnienie bibliografii Zawadzkiego, czemu miało służyć dokonanie kwerendy bibliograficznej piętnastu najważniejszych polskich czasopism bibliotecznych z lat 1983-2008.

Osiągnięcie tak wyznaczonych celów możliwe było tylko pod warunkiem skompletowania zespołu składającego się ze specjalistów reprezentujących różne dziedziny wiedzy. Ostatecznie w jego skład weszły następujące osoby:

- 1) autor projektu: dr hab. Włodzimierz Gruszczyński, prof. SWPS i IJP PAN; językoznawca polonista, medjoznawca⁴,
- 2) informatycy:
 - dr Maciej Ogrodniczuk, IPI PAN, informatyk i językoznawca, specjalista z zakresu zarządzania dokumentami i zarządzania wiedzą,
 - mgr Jakub Wilk, programista,
- 3) historyk: dr Adam Koźuchowski, Instytut Historii PAN,
- 4) filologowie:
 - dr Ewa Gruszczyńska, polonistka i przekładoznawca, Wydział Lingwistyki Stosowanej UW,
 - dr Anna Just, germanistka, historyk języka niemieckiego, Instytut Germanistyki UW,

² W zasadzie chodzi o starodruki, które można uznać za gazety ulotne lub tzw. gazety seryjne (z okresu poprzedzającego wydawanie regularnej prasy).

³ Por. bibliografia.

⁴ W okresie realizacji projektu był także zatrudniony w Instytucie Dziennikarstwa UW.

- dr Dorota Lewandowska-Jaros, polonistka i medioznawca, Instytut Dziennikarstwa UW,
 - mgr Katarzyna Jasińska-Zdun, filolog klasyczny, Instytutu Badań Interdyscyplinarnych „Artes Liberales”.
- 5) bibliotekarze: mgr Maria Piber, kustosz Biblioteki WDiNP UW wraz z zespołem,
- 6) koordynator projektu: mgr Grażyna Oblas, prezes Fundacji na rzecz Rozwoju Szkolnictwa Dziennikarskiego, działającej przy Instytucie Dziennikarstwa UW.

Instytucją, która udostępniła zasoby pamięci na swoim serwerze, był Instytut Dziennikarstwa Uniwersytetu Warszawskiego.

1.2. *Technologia*

Prace techniczne rozpoczęto od zeskanowania bibliografii Zawadzkiego w formacie wystarczającym do dokonania automatycznego rozpoznania tekstu (OCR) – w skali szarości, z rozdzielczością 200 dpi. Ich wynikiem były pliki w formacie TIFF z kompresją bezstratną zawierające graficzną reprezentację stron bibliografii. Następnie dokonano rozpoznania tekstu programem ABBYY FineReader w wersji 9 z uwzględnieniem zawartości językowej przetwarzanych tekstów (polski, niemiecki, francuski, łacina), w wyniku czego otrzymano zapis zawartości bibliografii (opisów materiałów, ale także treści dodatkowej – wstępów, indeksów, errat, streszczeń) w formacie czysto tekstowym. Skany materiałów przykładowych zamieszczone na końcu tomów oryginału wyodrębniono i pozostawiono w formacie graficznym na potrzeby składu elektronicznej wersji bibliografii.

Zeskanowane i przetworzone do postaci tekstowej opisy podzielono na rekordy danych i wyodrębniono w nich pola odpowiadające poszczególnym elementom opisu (tytuł pełny i skrócony, informacja o autorze, wydawcy, miejscu wydania itd.) Następnie rekordy zweryfikowano pod względem poprawności podziału oraz spodziewanej zawartości poszczególnych pól, zaś zestawy wartości z pól niezbędnych na dalszych etapach prac (sygnatury mikrofilmów, informacja o miejscu przechowywania oryginału) wyekstrahowano i dodatkowo sprawdzono. Zestaw pól posłużył do stworzenia modelu metadanych do docelowego opisu przechowywanych materiałów w systemie informatycznym; uwzględnił on zarówno informacje oryginalne (pochodzące z bibliografii Zawadzkiego), jak też dodane na potrzeby projektu (komentarze historyczne i językowe, słowniczki terminów).

Ze względu na dużą liczbę dobrej jakości mikrofilmów druków w Bibliotece Narodowej (której Zawadzki był wieloletnim i zasłużonym pracownikiem) skanowanie materiałów postanowiono ograniczyć tylko do tego jednego źródła. Wykonanie skanów zlecono Zakładowi Reprografii i Digitalizacji BN, dysponującemu profesjonalnym sprzętem i bogatym doświadczeniem. Skany z mikrofilmów zostały wykonane w rozdzielczości 300 dpi, w skali szarości i zapisane w formacie TIFF. Większość skanów otrzymanych z BN wymagała dalszej ręcznej obróbki, polegającej na eliminacji zbędnych klatek (w szczególności zeskanowanych kart katalogowych), korekty (zwiększanie ostrości i/lub jasności, obracanie obrazów tak, aby strony dokumentu były w pionie itp.), a zwłaszcza rozdzielaniu dwóch sąsiadujących stron (w CBDU prezentowane są zawsze pojedyncze strony dokumentów). Po korektach skany zostały skonwertowane do formatu DjVu oraz powiązane z opisami obiektów.

Do przechowywania materiałów w bibliotece cyfrowej (zarówno na potrzeby ich docelowej publikacji w Internecie, jak też na czas prac projektowych – redakcyjnych i uzupełniających) został wybrany system EPrints – darmowe oprogramowanie do udostępniania repozytoriów informacji autorstwa Uniwersytetu Southampton, wieloplatformowe, rozwijane od 2000 r. i dostępne na licencji GPL. System ten został skonfigurowany, dostosowany i rozszerzony zgodnie z założeniami projektu, a także w dużej części spolonizowany.

W związku z liczną reprezentacją materiałów zależnych, system EPrints rozbudowano o funkcję otwierania powiązanych dokumentów w nowym oknie, umożliwiającą ich wizualne porównanie (nie tylko metadanych, ale także skanów). Jednocześnie powstało także wiele innych pożytecznych rozszerzeń systemu EPrints, takich jak obsługa linków śródtekstowych zapisanych na bazie numerów obiektów, nowe metody przeglądania repozytorium (wg roku wydania, języka, miejsca wydania), szybka wyszukiwarka wg numeru obiektu w bibliografii Zawadzkiego, lista ostatnio zmienionych obiektów itp. Same powiązania między tekstami pochodnymi wykrywano na dwa sposoby – półautomatycznie, na podstawie zapisów o numerze wydania oraz ręcznie, z wykorzystaniem informacji w opisach źródeł i zawartych na skanach materiałów.

Ostatnia faza projektu miała na celu stworzenie nowego, jednotomowego, elektronicznego wydania zbiorczego publikacji K. Zawadzkiego, uzupełnionego i poprawionego na podstawie materiałów wypracowanych w ramach projektu. W trakcie łączenia w pojedynczy tom danych zawartych w tomach 1-3 skomasowano wstępy i ilustracje ze wszystkich tomów, zaś opisy bibliograficzne zamieszczono łącznie z bezpośrednim uwzględnieniem suplementów i errat. Ujednolicono pisownię oznaczeń, a wykaz tytułów utworzono od nowa na podstawie krótkiej wersji tytułów druków. Indeks nazwisk, nazw geograficznych i etnicznych oraz spis drukarzy, drukarni i wydawców przejęto z wersji oryginalnej i scalono. Zachowano układ według numeracji oryginalnej. Opisy bibliograficzne jednostek przejęto z biblioteki cyfrowej, dzięki czemu odpowiadają najbardziej aktualnej wersji materiałów. Powstał także program komputerowy do automatycznego generowania części publikacji zawierającej noty bibliograficzne, dzięki czemu jest możliwe ponowne automatyczne wykonanie składu w razie przyszłych korekt lub uzupełnień zawartości biblioteki.

1.3. Stan obecny

Cyfrowa Biblioteka Druków Ulotnych dostępna jest (na zasadzie wolnego dostępu) pod adresami <http://cbdu.id.uw.edu.pl/> i www.cbdu.pl. Weszła ona w skład Federacji Bibliotek Cyfrowych, co umożliwi wyszukiwanie zawartych w niej publikacji także bez znajomości jej adresu. W bibliotece zostały zebrane, tzn. opisane i w miarę możliwości udostępnione w postaci skanów, wszystkie znane obecnie druki, które można uznać za gazety ulotne. Znajdują się w niej opisy bibliograficzne 2009 starodruków (a więc nieco więcej niż w bibliografii Zawadzkiego), spośród których w postaci skanów udostępniono 1404. Łączna objętość materiału udostępnionego w postaci skanów wynosi 11585 stron. Publikacji polskojęzycznych jest 325 (zeskanowanych zostało 315 spośród nich, o łącznej objętości 2271 stron). Poza drukami polskojęzycznymi w CBDU znajdują się publikacje w następujących językach: niemieckim (797), włoskim (180), łacinie (69), szwedzkim (49), francuskim (32), hiszpańskim (14), angielskim (5), holenderskim (5), czeskim (2) i duńskim (2). Najstarsze opisane teksty pochodzą z 1501 r., a najstarsze udostępnione – z 1514 r. Do blisko stu polskich tekstów dołączono słowniczki trudniejszych wyrazów i form językowych. Słowniczkami takimi opatrzone także ponad pięćdziesiąt druków niemieckojęzycznych. Do kilkunastu tekstów polskich dołączono także słowniczki występujących w nich łacińskich słów i zwrotów.

Interfejs użytkownika umożliwi wyszukiwanie druków według różnych kryteriów, zarówno takich, które standardowo wykorzystywane są w opisach bibliograficznych, jak i mniej typowych, np. językowych, tematycznych i innych. Druki odnoszące się do tych samych faktów, a zwłaszcza te, które są (prawdopodobnie lub na pewno) przekładami lub przeróbkami innych obecnych w bibliotece tekstów, zostały ze sobą połączone linkami. Relacje zachodzące między takimi tekstami zostały skomentowane w polu „Powiązania”, tak by ułatwić porównywanie różnojęzycznych wersji informacji dotyczących tego samego wydarzenia, a wywodzących się – zapewne – z jednego źródła. W wypadku dokumentów będących prawdopodobnie tłumaczeniami tekstów niezachowanych (lub nieznanymi nam) podano informacje o tym, z jakiego języka były dokonane przekłady. Blisko 100 druków (zwłaszcza polskojęzycznych, ale też kilka niemieckojęzycznych) zostało dodatkowo opatrzone komentarzami historycznymi dotyczącymi faktów i osób, o których mowa w danym dokumencie, komentarzami historyczno-medialnymi, niektóre także historyczno-językowymi. Do większości druków polskojęzycznych i do wielu niemieckojęzycznych dołączono słowniczki objaśniające znaczenie nieużywanych dziś wyrazów czy związków frazeologicznych (w słowniczkach tych objaśnienia podawane są odpowiednio we współczesnej polszczyźnie i współczesnej niemczyźnie). Do niektórych tekstów polskich dołączono także słowniczki wyrażeń łacińskich zawartych w tych tekstach. Tego rodzaju komentarze i słowniczki nie są obecne w dotychczasowych bibliotekach cyfrowych i stanowią novum tego projektu.

1.4. Perspektywy

Pożądana jest modyfikacja i rozbudowa CBDU. Planowane są omówione poniżej działania, oczywiście pod warunkiem uzyskania wsparcia finansowego. Po pierwsze, aby ułatwić lekturę tekstów studentom i uczniom, powinno się wprowadzić transliterację (a może także transkrypcję) wybranych tekstów polskojęzycznych (zwłaszcza tych, które w oryginale złożone są czcionką gotycką – szwabachą lub frakturą) oraz połączyć słowniczki z tekstami

transliterowanymi (za pomocą hipertekstu). Po drugie, należy uzupełnić słowniczki. Po trzecie, powinno się kontynuować tłumaczenia wtrętów łacińskich w tekstach polskich (i ewentualnie innych) na język polski. Po czwarte, należy dążyć do przybliżenia polskiemu odbiorcy tekstów obcojęzycznych – dokonać transliteracji i ewentualnie przetłumaczyć wybrane teksty z niemieckiego, włoskiego, łaciny, szwedzkiego, francuskiego, hiszpańskiego, angielskiego, holenderskiego, czeskiego, duńskiego. Po piąte, konieczne jest zweryfikowanie informacji bibliograficznych dotyczących druków odnalezionych w pierwszym etapie realizacji projektu a niewykazanych w bibliografii K. Zawadzkiego i wprowadzenie tych danych do repozytorium. Oczywiście pożądanym byłoby również zeskanowanie tych druków i umieszczenie ich w repozytorium, jednak warunkiem tego jest wyrażenie zgody przez biblioteki będące ich właścicielami. Można rozpatrywać także umieszczanie w CBDU linków do obiektów opisanych w CBDU, a wystawionych w innych bibliotekach cyfrowych. Odrębnym zadaniem będzie poszukiwanie kolejnych druków niewykazanych w bibliografii Zawadzkiego, zwłaszcza w katalogach (opublikowanych drukiem lub w Internecie) bibliotek zwłaszcza: rosyjskich, ukraińskich, białoruskich, a także litewskich i łotewskich.

Kolejnym zadaniem projektu będzie modyfikacja interfejsu użytkownika repozytorium, a zwłaszcza wprowadzenie możliwości jednoczesnego wyświetlania skanu i tekstu transliterowanego lub tłumaczonego (w miarę możliwości przewijanie obu tekstów powinno być zsynchronizowane, co oznacza konieczność specjalnego oznakowania skanów i fragmentów tekstów transliterowanych i tłumaczonych).

2. Zastosowania naukowe

Starodruki udostępnione w CBDU budzą niewątpliwie zainteresowanie badaczy zarówno polskich, jak i zagranicznych. Świadczy o tym m.in. statystyka i „geografia” wejść do naszej biblioteki. Spoza Polski są to wejścia głównie z terenu Niemiec, co nie może dziwić zważywszy to, że ponad jedna trzecia druków w CBDU to druki niemieckojęzyczne. Wiadomo też, że trwają badania naukowe opierające się na materiałach prezentowanych w CBDU. Dotyczą one m.in. dawnego polskiego przekładu prasowego (badania Ewy Gruszczyńskiej z Instytutu Lingwistyki Stosowanej UW) oraz cech gramatycznych i stylistycznych języka niemieckiego w najdawniejszej prasie (badania Anny Just z Instytutu Germanistyki UW).

Materiały z CBDU wykorzystywane są też bardzo intensywnie w pracach nad Słownikiem języka polskiego XVII i 1. poł. XVIII wieku⁵, a ostatnio także w międzynarodowym projekcie IMPACT (IMProving Access to Text), który dotyczy problematyki digitalizacji dziedzictwa kulturowego, w szczególności tekstów historycznych.

3. Potencjalne zastosowania dydaktyczne

Opisana powyżej biblioteka cyfrowa może służyć jako pomoc dydaktyczna w nauczaniu historii prasy i dziennikarstwa na studiach dziennikarskich. O jej istnieniu powiadomieni zostali historycy prasy prowadzący zajęcia m.in. w Instytucie Dziennikarstwa Uniwersytetu Warszawskiego. Okazuje się jednak, że podobizny druków w niej zawarte służyć mogą jako pomoc dydaktyczna także w innych sytuacjach. O ile wiemy, wybrane z CBDU teksty szwedzkie posłużyły jako materiał do zajęć z tłumaczenia dawnych tekstów szwedzkich na archaizowany język polski na skandynawistyce w Szkole Wyższej Psychologii Społecznej w Warszawie. Nie ulega kwestii, że w podobny sposób można wyzyskiwać teksty w innych językach na zajęciach ze studentami innych filologii obcych. Również dla studentów polonistyki teksty zawarte w CBDU stanowić mogą świetny materiał do studiowania historii stylu dziennikarskiego. Nie ulega też wątpliwości, że CBDU to kopalnia tekstów źródłowych dla studentów historii.

Piśmiennictwo

- [1] Konrad Zawadzki, *Gazety ulotne polskie i Polski dotyczące*, tom 1: 1514-1661, tom 2: 1662-1728, tom 3: 1501-1725, Zakład Narodowy im. Ossolińskich, Wydawnictwo Polskiej Akademii Nauk, Wrocław 1977, 1984, 1990.

⁵ Dostępny w Internecie pod adresami: http://xvii-wiek.ijp-pan.krakow.pl/pan_klient lub <http://sxvii.pl>