



Zyta Szymańska
Beata Domośławska
Pracownia Bibliografii Bieżącej w Poznaniu Instytut Badań Literackich PAN
Maciej Matysiak
ADVIS

Polska Bibliografia Literacka – ujęcie realistyczne



Dr Zyta Szymańska - absolwentka filologii polskiej i psychologii Uniwersytetu im. A. Mickiewicza w Poznaniu, starszy dokumentalista dyplomowany, doktor nauk humanistycznych w zakresie bibliologii i informatologii (Uniwersytet Wrocławski). Praca doktorska *Funkcje bibliografii literackich* jest przygotowywana do opublikowania. Zatrudniona od roku 1989 w zespole opracowującym Polską Bibliografię Literacką — współautorka roczników PBL za lata 1985-2000 i redaktor roczników PBL za lata 1987, 1999-2000. W Polskiej Bibliografii Literackiej zajmuje się działami recepcji

literatur obcych w Polsce i teorią literatury. Zainteresowaniami obejmuje sposoby organizacji wiedzy, struktury poznawcze, literaturoznawstwo.



Beata Domośławska - absolwentka filologii polskiej Uniwersytetu im. A. Mickiewicza w Poznaniu. Od roku 1989 zatrudniona w Pracowni Bibliografii Bieżącej IBL PAN. Współautorka Polskiej Bibliografii Literackiej od rocznika 1986. Redaktorka roczników PBL za lata 1991-1995.



Maciej Matysiak - prywatny przedsiębiorca, absolwent Politechniki Poznańskiej, były pracownik naukowo-dydaktyczny Instytutu Informatyki Politechniki Poznańskiej i Loyola University New Orleans. Specjalista w zakresie baz danych i technologii internetowych. Architekt, projektant i wykonawca cyfrowej postaci Polskiej Bibliografii Literackiej. Główne dziedziny zainteresowań to: systemy baz danych, optymalizacja zapytań, bibliografia, filozofia.

Streszczenie: Piętnaście lat funkcjonowania w Internecie bazy danych Polska Bibliografia Literacka (PBL) i kilkadziesiąt lat doświadczeń w tworzeniu tej bibliografii pozwalają jej twórcom przedstawić problematykę bibliograficznych baz danych zarówno od strony uniwersalnej metodologii, jak i rozwiązań technologicznych. W referacie zaakcentowano właściwy punkt wyjścia badań bibliograficznych, jakim jest filozofia realistyczna i konieczność zdefiniowania zakresu badanej rzeczywistości. Wychodząc od wskazania badanych bytów literackich i pochodnej twórczości, pokazano rolę bibliografów i to, czym jest wynik ich pracy zapisany w postaci PBL.

Obecnie baza danych zawiera około 500 tys. opisów bibliograficznych. Planowana digitalizacja archiwalnych roczników PBL wraz z ich analizą semantyczną zwielokrotni tę liczbę, czyniąc z PBL dzieło unikatowe w skali światowej. Do rozwoju PBL przyczyni się także dokumentowanie twórczości literackiej w Internecie, które jest jeszcze przedmiotem badań.

W referacie zaprezentowano przegląd aktualnych problemów twórców systemu, zgłaszane potrzeby użytkowników oraz plany rozwojowe PBL.

Słowa kluczowe: metodologia badań, filozofia realistyczna, bibliografia realistyczna, wymiana danych, FRBR, digitalizacja, analiza semantyczna, zasoby sieciowe, portale naukowo-społecznościowe, Polska Bibliografia Literacka.

Abstract: The bibliographic database of Polish Literary Bibliography (PBL) has been available on the Internet for 15 years but the bibliography has been developed for decades. This experience allows the authors of PBL to present the field of bibliography either from the side of universal methodology or technological issues. The paper highlights the proper starting point in bibliographic research that is



Bibliograficzne bazy danych i ich rola w rozwoju nauki

II Konferencja naukowa Konsorcjum BazTech

Poznań, 17-19 kwietnia 2013



realistic philosophy, the need to define entities and the scope of the reality documented. Starting from the indications of literary entities, subjects, and derivative works we discuss the role of bibliographers and what the result of their work is presented in the form of PBL. Currently, the database contains about 500,000 bibliographic descriptions. It will be multiplied by the planned digitalization of archive annual volumes of PBL, including their semantic analysis and data integration, making PBL unique in the world. We also plan to develop the PBL database by methodological documenting of dynamic web literature which is still the subject of research. The paper also presents an overview of current issues encountered by the authors, escalation of user requirements and development plans for the PBL.

Keywords: *research methodology, realistic philosophy, realistic bibliography, data exchange, FRBR, digitization, semantic analysis, web literature, science and social portals, the Polish Literary Bibliography*

[Prezentacja](#)

Wprowadzenie

Bibliografia jest pojęciem szerszym nie tylko od wykazu literatury dołączonego do jakiegoś dzieła, lecz także od katalogu zasobów biblioteki czy spisu zawartości wybranej grupy czasopism. Jest dziedziną wiedzy i działalnością obejmującą problemy identyfikacji i opisywania wszelkich dokumentów, zapisanych nie tylko w języku naturalnym. W dobie dynamicznie zmieniających się metod tworzenia i nowych postaci dokumentów, powszechnego użycia narzędzi informatycznych i Internetu, rodzą się w naturalny sposób pytania: co, jak i dlaczego należy dokumentować. Problemy te dotyczą także Polskiej Bibliografii Literackiej (PBL), która jest tworzona od 65 lat, od 16 lat w postaci bibliograficznej bazy danych, a od 15 lat udostępniana w Internecie.

Zanim jednak omówimy specyfikę PBL i problemy badawcze związane z tekstem literackim, zatrzymajmy się na chwilę na problemach wspólnych dla wszystkich bibliografii i bibliograficznych baz danych. Wiele z nich bierze się z niewłaściwego punktu wyjścia w rozważaniach nad bibliografią, czego skutkiem jest tworzenie pseudoproblemów, a potem szukanie dla nich pseudorozwiązań, które rodzą kolejne pseudoproblemy. Następuje wikłanie się w sztuczne problemy językowe lub stosowanie przesadnie skomplikowanych i uciążliwych standardów. Dla zilustrowania tego zjawiska rozważmy klasyczny już dylemat: co jest książką, a co nią nie jest? Czy płyta CD z tekstem i animowanymi obrazkami trzymana w lewej ręce jest tak samo książką jak jej odpowiednik w twardej oprawie trzymany w prawej ręce? Czy pobrany z Internetu plik PDF, DjVu lub TXT (bez podziału na strony) to też książka? Teoretycy bibliografii sięgają w tym momencie po definicję książki i rozciągają ją jakby była z gumy, obejmując nią wszystko — od inskrypcji naskalnych w jaskiniach, po efemeryczne teksty chwilowo dostępne w otchłaniach Internetu. Ucieczka od pojęcia książki w stronę dokumentu, jednostki informacyjnej, obiektu, encji itp. nie rozwiązuje problemu, co jest przedmiotem bibliografii. Badacze poruszający się po powierzchni problemów językowych rozwiązują je, tworząc nowe znaki, nowe definicje, nowe problemy, nowe standardy, np. polska norma (PN-ISO 10324) tłumacząca, czym jest dokument, wyjaśnia: *Dokument — jednostka bibliograficzna lub zespół jednostek [...], które [...] stanowią podstawę pojedynczego opisu bibliograficznego.* To przecież tautologia. Jeśli przedmiotem opisów bibliografii są dokumenty, a dokument jest jednostką stanowiącą podstawę opisu bibliograficznego, to znaczy tyle, że przedmiotem



bibliografii jest jej przedmiot. Podobnego rodzaju anomalie występują w tak popularnych i skomplikowanych standardach zapisu danych bibliograficznych, jak MARC (również kolejnych — MARC 21) i ciągną się od lat. Ten format z punktu widzenia projektowania baz danych nie spełnia warunków nawet drugiej postaci normalnej. Implementacje, które niwelują wynikające z tego anomalie, dokonują dekompozycji formatu MARC, zapisując, *de facto*, dane bibliograficzne już w innym formacie, czyli teoretycznie używa się standardu, którego, *de facto*, się nie używa. Dzisiejsze postulaty katalogowania zasobów i tworzenia bibliografii zgodnych ze standardem FRBR (niezgodny z MARC) są potwierdzeniem błędów koncepcyjnych i projektowych poczynionych przed laty.

Tych i wielu innych błędów można unikać, stosując podejście realistyczne. Właściwym punktem wyjścia jest zawsze rzeczywistość, a nie język i terminologia. Również dlatego, że wiele problemów dotyczących rzeczywistości, badania własności bytu, zostało już rozwiązanych tysiące lat temu przez filozofię. Przyjrzyjmy się temu filozoficznemu fundamentowi, wspólnemu dla wszystkich dziedzin nauki. Następnie zobaczmy, czym jest bibliografia, aby określić, czym jest Polska Bibliografia Literacka i jaka rzeczywistość w tym przypadku jest przedmiotem badań bibliografów. Krótko powiemy o początkach PBL, stanie obecnym, napotykanym problemach, zgłaszanych potrzebach odbiorców i wymaganych zmian oraz problemie dokumentowania twórczości literackiej w Internecie. Całość zamknie prezentacja planów rozwojowych PBL.

Filozofia realistyczna i bibliografia realistyczna

Przedmiotem badań wszystkich dziedzin nauki powinna być zawsze rzeczywistość i pragnienie poznania prawdy o niej. U samego źródła tych badań jest realnie istniejący byt jako taki, a dyscyplinę tych badań od czasów Arystotelesa nazywamy *metafizyką*. Współcześnie cały krąg myśli filozoficznej, przyjmującej za swój punkt wyjścia byt, określamy terminem *filozofia realistyczna*. Pierwsze zasady ludzkiego rozumowania, jak zasada niesprzeczności, wywodzą się właśnie z analizy bytu — byt jest, a niebytu nie ma, więc nie można głosić, że byt jednocześnie jest i go nie ma.

Poznanie rzeczywistości dokonuje się w procesie odbierania informacji i formułowania myśli. Owocem myślenia jest rozumienie rzeczy, jej pojęcie i sąd. Poznajemy prawdę wtedy, kiedy nasza myśl jest zgodna z rzeczywistością. Prawdziwe pojęcia i sądy porządkujemy, powiększając ogół ludzkiej wiedzy. Nauka, mając za cel poznanie prawdy o rzeczywistości, doskonali metody i narzędzia jej poznawania. Wiedza, która nie jest prawdą, nie jest wiedzą. Naukowiec, który nie poświęca się poznaniu prawdy, nie jest naukowcem (może być celebrytą lub ideologiem).

Każdą myśl, wynik poznania rzeczywistości, jeśli chcemy przekazać drugiemu człowiekowi, musimy wyrazić w jakimś języku. Każdy język to systemu znaków (liter, słów, obrazów, gestów). Nie da się zapisać myśli ani żadnej informacji bez użycia języka. Dopiero na tym etapie mamy do czynienia ze znakowaniem, czyli tworzeniem terminologii i standardów.



Bibliograficzne bazy danych i ich rola w rozwoju nauki

II Konferencja naukowa Konsorcjum BazTech

Poznań, 17-19 kwietnia 2013



Najważniejsza w tym podejściu jest kolejność: wyjście od bytu — rzeczywistości, poznanie prawdy o niej przez naukowca (bibliografa), a na końcu zapisanie wyniku badań w jakimś zrozumiałym języku. Uczy tego historia nauki, filozofowie bowiem wychodzący od poznającego podmiotu (człowieka) i procesu poznania, ogłaszali niepoznawalność bytu. Filozofowie wychodzący z poziomu analizy języka coraz bardziej wnikali się w problemy językowe.

Bibliografia jako dyscyplina nauki powinna stosować to samo podejście, wychodząc od bytu, który jest przedmiotem badań. Dokumentować można wszystkie rodzaje bytów: małe molekuly, wielkie galaktyki, zwierzęta, zjawiska społeczne, dzieła literackie, ludzie, książki w bibliotece, artykuły w czasopiśmie, utwory muzyczne, realnie zaistniałe fakty, wydarzenia, konkursy lub konferencje. Zakres badanej rzeczywistości może być określony m.in. tematycznie, geograficznie, selektywnie lub poziomem szczegółowości.

W przypadku bibliografii podmiotem, który poznaje rzeczywistość, jest bibliograf — odbiera informację, identyfikuje byty, orzeka o nich, kategoryzuje je, rozpoznaje także potrzeby odbiorców. Jak każdy naukowiec doskonali swoje metody badawcze i stosowane narzędzia. To dostosowanie jest konieczne, zwłaszcza wtedy, kiedy badana rzeczywistość zmienia się (rozwój technologii, Internet, nowe sposoby publikowania, nowe postacie dokumentów).

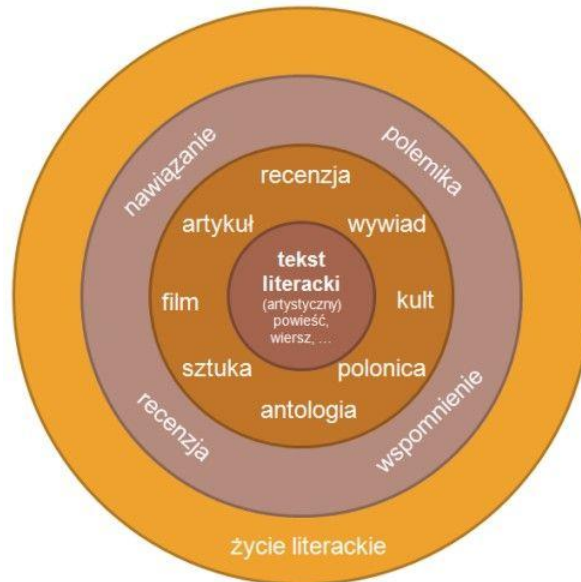
Na koniec wynik swojej pracy, wynik poznania badawczego bibliograf zapisuje w wybranym języku. Jak każdy język jest on żywym zestawem znaków – słów, symboli, schematów, standardów (np. MARC, Dublin Core, XML, PBL), gdzie zastosowana jest przyjęta terminologia i forma prezentacji informacji. Ten etap jest też źródłem problemów z przekazaniem pozyskanej wiedzy odbiorcy, który, znając dobrze język polski, może nie rozumieć języka bibliografii.

Polska Bibliografia Literacka

Czym jest Polska Bibliografia Literacka? W uproszczeniu, przedmiotem jej badań jest literatura polska i recepcja literatur obcych w Polsce. Jest to dokumentowanie dziedzictwa kulturowego polskiego narodu, począwszy od roku 1944 (choć planowane jest także włączenie materiałów od roku 1939), dlatego ta bibliografia jest tak ważna dla Polaków. Jedną z podstawowych różnic między PBL a popularnymi bibliografiami zasobów bibliotecznych lub bibliografiami zawartości czasopism jest jej zasięg geograficzny. Jest nim twórczość literacka i jej pochodne obejmujące nie tylko Polskę, ale cały świat.

Pierwszym, prymarnym bytem, który interesuje bibliografa opracowującego PBL, jest tekst literacki (artystyczny) o ściśle określonych cechach (przypadłościach) identyfikacyjnych, np. powieści, teksty dramatyczne, wiersze. Wokół tej podstawowej twórczości literackiej występuje w rzeczywistości wiele bytów mających charakter pochodnej twórczości, np.: recenzje książek, artykuły o twórcach i ich utworach, filmy nakręcone na podstawie książek i ich recenzje, sztuki teatralne wystawione na podstawie zaadaptowanych tekstów literackich i ich recenzje. Poziomów pochodnej

twórczości może być wiele, np. książka pisarza inspiruje do napisania scenariusza filmowego, według którego zostaje nakręcony film, następnie ktoś pisze książkę o tym filmie i ukazuje się jej recenzja, a potem może pojawić się polemika z recenzją itd. Schematycznie pokazano te zależności na rysunku 1.



Rys. 1. Przedmiot badań twórców PBL — byt literacki i pochodna twórczość
Źródło: Opracowanie własne.

W zakresie rzeczywistości dokumentowanej w PBL znajdują się także byty będące podmiotami twórczości: pisarze, wydawnictwa, instytucje, teatry. Całość badanego zakresu literackiej rzeczywistości spaja życie literackie, w ramach którego odnotowywane są wydarzenia, m.in. takie jak: konferencje, konkursy, nagrody, zgony twórców. W przedstawionym ujęciu realistycznym mamy zdefiniowany fragment rzeczywistości, który podlega identyfikacji, analizie, opracowaniu i dokumentowaniu. Podmiotem prowadzącym badania jest zespół bibliografów Pracowni Bibliografii Bieżącej IBL PAN w Poznaniu, a wynikiem pracy jest właśnie Polska Bibliografia Literacka zapisana w języku opracowanym przez jej twórców.

PBL jest wydawana w formie drukowanej od roku 1954. Jej twórcą był prof. Stefan Vrtel-Wierczyński, który opracował metodologię badań, sposób zapisu i wyszkolił pierwszy zespół bibliografów. Wykonali oni badania dokumentacyjne literatury polskiej wydawanej od roku 1944. Postać cyfrowa PBL pojawiła się w roku 1997. Architektem i projektantem tego systemu był Maciej Matysiak. Pierwotnie system informacyjny miał służyć do automatycznego redagowania postaci drukowanej na podstawie zebranych materiałów źródłowych, dlatego konieczna była kontynuacja stosowanej metodologii. Zbierane opisy bibliograficzne były udostępniane także w Internecie. Według naszej wiedzy była to wtedy pierwsza bibliograficzna baza danych w Polsce dostępna w Internecie. Wysokie koszty druku i ograniczona dostępność drukowanych roczników PBL, wobec niskich kosztów udostępniania bibliografii w sieci, spowodowały wycofanie się z wydawania PBL drukiem. Pozostała tylko wersja internetowa.



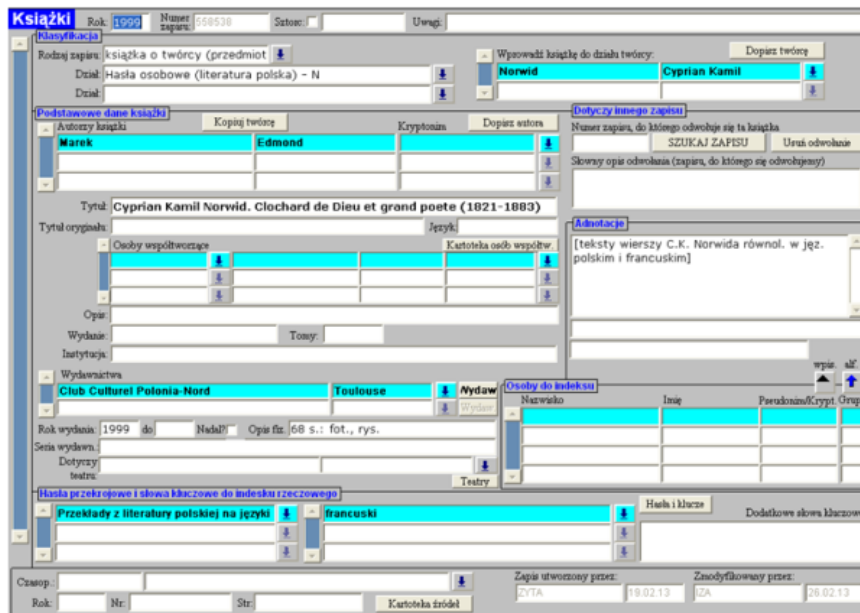
Baza danych i aplikacje

Obecnie, po 16 latach PBL liczy sobie około 500 tys. opisów bibliograficznych, czyli tekstów literackich i wszelkich bytów pochodnej twórczości literackiej, które wchodzi w dokumentowany zakres rzeczywistości. Materiał ten obejmuje tylko piśmiennictwo za lata 1988–2000. Materiały porządkowane są w 2 tys. działów tematycznych, ułożonych hierarchicznie. Oprócz działów tematycznych dostęp do materiałów jest możliwy dzięki 11 tys. haseł indeksu rzeczowego. Kartoteki osobowe zawierają około 100 tys. zidentyfikowanych twórców i współtwórców. Dokładna liczba jest trudna do ustalenia z uwagi na problemy z identyfikacją poszczególnych osób (ta sama osoba używa czasami kilku różnych pseudonimów). Do tego w osobnych kartotekach zgromadzono opisy dotyczące m.in. 12 tys. wydawnictw i 1200 teatrów z Polski i całego świata (np. gdy na podstawie polskiego utworu literackiego wystawiana jest sztuka na deskach jakiegoś teatru w Australii, PBL odnotowuje ten fakt ze wskazaniem, o jaki teatr chodzi).

Struktura bazy danych jest zaskakująco prosta i wynika z wykonanej przed laty analizy danych i modelowania ich za pomocą diagramów związków encji (EDR), które ukazały rzeczywiste byty, o których informacje będą przechowywane w systemie. Schemat bazy danych do zapisu bibliografii nie wzoruje się zatem na sztucznych twórcach jak MARC i nie powiela zawartych w nim błędów, ale jest wiernym odwzorowaniem świata rzeczywistego w zakresie, jaki jest przedmiotem badań bibliografów PBL.

Niejako w centrum mamy struktury do gromadzenia głównych zapisów bibliograficznych o interesujących nas bytach literackich. Poszczególne atrybuty (części zapisu) odpowiadają cechom opisywanego bytu literackiego (np. książki, filmu, sztuki, recenzji). Tymi cechami mogą być: tytuł, autor, język, opis fizyczny itd. czy też adnotacje bibliografa dotyczące treści, opisy bibliograficzne są bowiem tworzone z autopsji.

Wokół tych centralnych struktur do gromadzenia opisów bibliograficznych mamy zbudowanych szereg kartotek, w których znajdują się opisy bytów biorących udział w powstaniu tekstu literackiego: opisy twórców, wydawnictw, teatrów, czasopism itp., a także kartoteki orzekające o przynależności bytu do danej kategorii: rodzaje zapisów, działy tematyczne, roczniki. Istnieją też struktury pomocnicze służące do redagowania materiału zgodnie z przyjętą metodologią i prezentowania na stronie internetowej w wymaganym porządku.



*Rys. 2. Aplikacja do wprowadzania opisów bibliograficznych
 Źródło: Opracowanie własne na podstawie aplikacji do wprowadzania danych.*

Na rysunku 2 przedstawiono przykładowy formularz, za pomocą którego są wprowadzane i edytowane opisy bibliograficzne książek. W przykładzie widać opis książki przedmiotowej o Cyprianie Kamile Norwidzie napisanej przez Marka Esmonda, ze wskazaniem wydawnictwa, które ją wydało, opisem fizycznym książki, podaniem haseł indeksu rzeczowego (przekład na język francuski) oraz autorskimi adnotacjami bibliografa. Na rysunku 3 pokazano dostęp do PBL przez Internet w formie przykładowej listy wyszukanych zapisów bibliograficznych dotyczących literatury bułgarskiej – książek i powiązanych z nimi recenzji.

WYSZUKANE ZAPISY

Zastosowane kryterium: **Dział:** Literatura obce / Literatura bułgarska

Możesz zmienić zakres lat: 1988 do 2000 **SZUKAJ**

Literatury obce / Literatura bułgarska / Część ogólna (bułgarska)
 książki (alfabet autorów)

1. książka: **Od Wisły do Maricy = Ot Visla do Marica**. 1985 [szczegóły](#)
 ● recenzja: Sujecka Jolanta: *Pamiętnik Słowiński 1988 t. 35 (1985) s. 270-274* [szczegóły](#)

Literatury obce / Literatura bułgarska / Historia literatury (bułgarska)
 książki (alfabet autorów)

2. książka: Gałązka Wojciech: **Oswajanie skorpionów. Szkice o literaturze bułgarskiej**. 1992 ([Zawiera:] Filozofia dziejów bułgarskich. Mit niewoli w literaturze i ...) [szczegóły](#)
 ● recenzja: Juda Celina: *Pamiętnik Słowiński 1994 t. 42 s. 120-121* [szczegóły](#)

3. książka: **Kształtowanie się nowobułgarskiego języka literackiego (do roku 1878)**. 1990 (Tadeusz Szymański: Przedmowa. - Julija Baitova: Slovoobrazovatelnii ino...) [szczegóły](#)

4. książka: Sujecka Jolanta: **Obraz wielkiej przemiany w poezji bułgarskiej (1918-1925)**. 1996 [szczegóły](#)
 ● recenzja: Poźniak Telesfor: *Pamiętnik Słowiński 1997 t. 45/46 (1995/1996) s. 195-197* [szczegóły](#)

materiały czasopiśmiennicze (alfabet autorów)

5. artykuł: Czajka Henryka: **Z bułgarskich utopii literacko-folklorystycznych**. *Prace Naukowe Uniwersytetu Śląskiego 1997 nr 1600 s. 18-25* [szczegóły](#)

*Rys. 3. Dostęp do PBL przez Internet
 Źródło: Opracowanie własne na podstawie bazy PBL.*



Aktualne problemy PBL i oczekiwania użytkowników

W tworzeniu PBL i jego obecnym używaniu jednym z poważniejszych problemów jest dychotomia wynikająca z tego, że materiał opracowany za lata 1944–1988 jest dostępny w formie nieporęcznych i trudno dzisiaj dostępnych drukowanych roczników, natomiast materiał od roku 1988 do roku 2000 jest dostępny wyłącznie w Internecie. Nie ma zatem możliwości prostej nawigacji między tymi częściami lub kierowania do systemu zapytań obejmujących materiał zachodzący na obie części.

Kolejnym problemem jest kilkuletnie opóźnienie w opracowywaniu bibliografii powstałe w latach 80., przez co nie ma możliwości z pomocą PBL prowadzenia poszukiwań i analiz dotyczących teraźniejszości. Na problem ten nałożyła się dodatkowo eksplozja piśmiennictwa drukowanego, jaka miała miejsce po roku 1989. Koniecznością okazało się ograniczenie zakresu bibliografii do starannie dobranej listy czasopism. Na dzień dzisiejszy opóźnienie wynosi około dziesięć lat.

Oprócz problemu kompletności PBL w sensie przedziału lat (brak dostępu przez Internet do archiwalnych roczników oraz opóźnienie w opracowaniu bieżących materiałów) problem nabiera wagi wobec nowego i bardzo dynamicznie zmieniającego się fragmentu rzeczywistości literackiej, jakim jest piśmiennictwo internetowe. Istnieją wydawnictwa internetowe, portale literackie czy nawet indywidualne blogi pisarzy i poetów, które na chwilę obecną żyją spontanicznie swoim życiem i nie są przedmiotem systematycznego opisu w ramach PBL.

Dostępność PBL w Internecie, w połączeniu ze zwiększającą się masowo ilością dostępnych w sieci dokumentów, zeskanowanych książek, pełnych tekstów utworów i przyzwyczajenie się odbiorcy do tego stanu rzeczy, powoduje eskalację wymagań użytkowników wobec PBL. Na przykład kiedyś czytelnik PBL zadowolony był odesłaniem do danego czasopisma z podaniem numeru i strony. Dzisiaj wymaga, aby dodatkowym kliknięciem mógł pobrać cyfrową wersję interesującego go dokumentu lub zostać odesłany do miejsca w sieci, gdzie może zapoznać się z jego pełną treścią. Użytkownicy oprócz chęci pobierania informacji z PBL są także zainteresowani uzupełnianiem jej o swoje uwagi i materiały uzupełniające. Zwiększone wymagania użytkowników dotyczą także chęci uzupełniania w PBL swojego dorobku literackiego przez samych autorów, którzy zaglądając tam, widzą braki w dokumentacji swojej twórczości. Jeszcze innym polem niezaspokojonych potrzeb jest kontaktowanie czytelników i badaczy literatury z autorami.

Niestety, z drugiej strony cięcie kosztów inwestycyjnych, kosztów osobowych oraz coraz gorsze finansowanie podejmowanych przez IBL projektów związanych z Polską Bibliografią Literacką utrudniają rozwiązanie wspomnianych problemów i rozwój funkcjonalności PBL, jej metod i narzędzi, modernizacji przestarzałej już dzisiaj infrastruktury. Recenzenci wniosków o granty badawcze zdają się nie rozumieć roli bibliografii dziedzinowych i bibliograficznych baz danych. Na przykład anonimowy recenzent wniosku o grant złożony w 2012 r. spowodował odrzucenie wniosku opiniami typu: [...] *tworzenie od podstaw zapisów bibliograficznych nie ma sensu, podobnie jak pod znakiem zapytania staje sensowność tworzenia odrębnych baz tema-*



tycznych [...], czym podważył automatycznie sensowność istnienia dziedzinowych bibliograficznych baz danych. Przecież takich recenzji nie piszą studenci, którym wystarczy Google i Wikipedia.

Digitalizacja archiwalnych roczników PBL

Przez digitalizację najczęściej rozumie się skanowanie dzieła wydanego w formie książki, opisanie go za pomocą metadanych i umieszczenie wynikowego pliku DjVu lub PDF w odpowiedniej „szufladzie” cyfrowej biblioteki. Ten proces bywa uzupełniany rozpoznawaniem pisma (OCR) i wtedy powstaje elektroniczna kopia pełnotekstowa. Trzeba jednak pamiętać, że digitalizacji podlegać mogą różne obiekty składające się na nasze dziedzictwo kulturowe (np. trójwymiarowe obiekty muzealne, mikrofilmy), a wynikiem tego procesu mogą być różne inne cyfrowe postaci tych obiektów. Dlatego, aby powiedzieć, co rozumiemy przez daną digitalizację, trzeba nie tylko wskazać, co jest przedmiotem digitalizacji (oryginał), lecz także zdefiniować, co dokładnie będzie wynikiem tego procesu (jaka cyfrowa kopia) i czemu to będzie służyć.

W przypadku Polskiej Bibliografii Literackiej przedmiotem digitalizacji będą książki w miękkiej oprawie, grube tomy formatu A4, tzw. roczniki, dokumentujące kolejne lata polskiej twórczości literackiej i recepcję obcych literatur w Polsce. Wynikiem tego procesu, czyli cyfrową kopią tych roczników, nie mają być jednak ani pliki DjVu zawierające skanowane strony, ani też postać pełnotekstowa. Z punktu widzenia użytkowników PBL taka postać cyfrowa byłaby za mało funkcjonalna. Planujemy oprócz skanowania wszystkich stron i przejścia procesu OCR dokonać analizy semantycznej tekstu i jego dekompozycji na poszczególne elementy składowe, wprowadzić je w struktury istniejącej bazy danych, aby odtworzyć maksymalnie wiernie obraz świata rzeczywistego, który zapisany jest specyficznym językiem archiwalnych roczników PBL. Ten proces wymaga analizy składni, analizy semantyki, rozwiązania wielu problemów identyfikacji poszczególnych bytów literackich. Dla przykładu, ze względu na oszczędność miejsca w druku stosowano skróty tytułów czasopism, a stosowane skróty ulegały czasami zmianie, czyli pod różnymi skrótami na przestrzeni kolejnych roczników PBL kryje się to samo czasopismo. Jeśli to czasopismo jest wydawane jeszcze obecnie, to jest także opisane w istniejącej już bazie danych. Wszystkie opisy bibliograficzne artykułów, recenzji, wierszy itp. muszą zostać zatem powiązane z jednym rekordem w kartotece czasopism. Takie same problemy identyfikacji dotyczą też innych bytów, jak autorzy tekstów (piszący pod pseudonimami), wydawnictw (które zmieniały nazwy) itp.

Celem takiego podejścia jest zupełne usunięcie dychotomii między rocznikami archiwalnymi i opracowywaną na bieżąco bibliografią literacką, dzięki czemu będzie możliwe swobodne nawigowanie między dowolnymi latami oraz łączne i bardzo selektywne przeszukiwanie Polskiej Bibliografii Literackiej w pełnym zakresie lat od 1944 do chwili obecnej. Tego typu bibliografia będzie absolutnie unikatowym dziełem w skali światowej. Uzyskaliśmy także zgodę twórców bibliografii literatury polskiej i teatru okresu wojennego na włączenie ich dzieła do tego projektu, więc kolejnym etapem będzie dodatkowe rozszerzenie bazy danych PBL wstecz, tak aby obejmowała całą twórczość literacką od roku 1939 do dzisiaj.



Problem dokumentowania twórczości literackiej o dużej dynamice

Jednym z ciekawszych i trudniejszych problemów, z jakimi mierzą się badacze literatury, jest wielka dynamika piśmiennictwa, szczególnie publikowanego w Internecie. Powstają, jak grzyby po deszczu, nowe wydawnictwa (niektóre działają tylko w Internecie), portale literackie czy indywidualne witryny prezentujące twórczość literacką na wysokim poziomie. Chcąc utrzymać wysoką wartość PBL jako narzędzia wspomagającego studentów, badaczy i sympatyków literatury, planujemy dążyć do kompletności PBL również w zakresie dokumentowania tego dynamicznego zakresu literackiej rzeczywistości. Dla jasności trzeba powiedzieć, że wyszukiwarki internetowe (np. Google), biblioteki cyfrowe, bibliografie biblioteczne czy dziedzinowe bibliografie zawartości czasopism w żaden sposób nie przyczyniają się do rozwiązania tego problemu użytkowników.

Zasadniczym problemem jest dostęp do źródła informacji, pozyskanie tej informacji, analiza, opracowanie i wprowadzenie opisu bibliograficznego. O ile dla dużych wydawnictw sprawa jest stosunkowo łatwa, jeśli chodzi o źródło informacji, o tyle w przypadku małego wydawnictwa bibliograf może nie mieć w ogóle świadomości, że ono istnieje. A jeśli dowie się o nim, takie wydawnictwo nie musi nawet publikować wykazu tego, co zostało wydane. Ten problem potęguje się w przypadku wydawnictw funkcjonujących w Internecie, portali literackich lub indywidualnych blogów. Skąd bibliograf ma wiedzieć o danej publikacji? Jak ma uzyskać komplet informacji? Jak opracowywać i dokumentować różnorodną twórczość internetową? Wychodzimy z założenia, że małe wydawnictwa i indywidualni twórcy literatury muszą aktywnie włączyć się w dokumentowanie swojej twórczości. Generalnie planujemy zastosowanie trzech metod w odniesieniu do:

- dużych wydawnictw – automatyczne pobieranie opisów książek z systemu wydawnictwa (aktywny jest bibliograf),
- małych wydawnictw – udostępnienie po stronie PBL narzędzi służących do przesyłania informacji w zadanym formacie o wydanych książkach (aktywne jest wydawnictwo, a potem bibliograf),
- indywidualnych twórców – udostępnienie narzędzi do samodzielnego dokumentowania swojej twórczości w PBL (aktywny jest głównie autor).

Oczywiście opisy bibliograficzne wprowadzane dowolnym z przedstawionych wyżej sposobów będą zawsze weryfikowane i korygowane pod kątem zgodności z naukową metodologią PBL (choćby kategoryzacja, indeksy rzeczowe). Wdrożeniu tych rozwiązań i nawiązaniu aktywnej współpracy z twórcami (podmiotami twórczości) sprzyja zgodność interesu małych wydawnictw, indywidualnych twórców z interesem bibliografów dbających o jak największą kompletność bibliografii. Małe wydawnictwa są zainteresowane promocją swoich publikacji tak samo jak indywidualny twórca jest zainteresowany propagowaniem swojej twórczości oraz dokumentowaniem swojego dorobku, a przez to docieraniem ze swoją twórczością do coraz większej grupy czytelników. Należy dostarczyć narzędzi, które umożliwią tę współpracę.



Plany rozwojowe

Oprócz wspomnianego procesu digitalizacji archiwalnych roczników PBL i rozszerzenia zakresu badanej rzeczywistości na piśmiennictwo internetowe, w planach rozwojowych uwzględniamy m.in.: modernizację przestarzałej już infrastruktury technicznej, modernizację aplikacji dedykowanych pracownikom i zewnętrznym użytkownikom, integrację danych z zewnętrznymi źródłami (automatyczne dokumentowanie tego co możliwe), dostosowanie modelu danych do pełnej zgodności ze standardem FRBR, udostępnienie zapisów źródłowych z naszej bazy danych w innych formatach, integrację na zasadzie odsyłaczy z zewnętrznymi bazami danych (np. BazHum), bibliotekami i archiwami zawierającymi pełne teksty dokumentów, aby umożliwić czytelnikom szybkie sięgnięcie po dokument, do którego odsyła PBL. Planujemy także uzupełnienie istniejących aplikacji o narzędzia do eksploracji danych (*data mining*) i odkrywania wiedzy, które będą wspomagały badaczy literatury w ich analizach.

Docelowym modelem bibliograficznej bazy danych PBL, do którego dążymy, jest portal naukowo-społecznościowy zbudowany na bazie systemu informatycznego tworzonego zgodnie z zaprezentowanym w niniejszym artykule ujęciem realistycznym. Pozwoli to także zaspokoić takie wymagania zgłaszane przez użytkowników, jak uzupełnianie przez autorów własnego dorobku i biografii, kontaktowanie się czytelników z autorami i czytelników z czytelnikami. Portal ten pozwoli także na włączenie użytkowników PBL w proces tworzenia bibliografii, zamieszczanie przez nich materiałów uzupełniających (ilustracje, linki, komentarze itp.), a także prowadzenie dyskusji i integrację środowiska użytkowników PBL z Polski i całego świata.

Bibliografia:

1. GOGACZ, M. *Elementarz metafizyki*. Warszawa: Oficyna Wydawnicza Navo, 2008. ISBN 978-83-89920-98-0.
2. KRAPIEC, M. A. *Język i świat realny*. Lublin: Redakcja Wydawnictw KUL, 1985. ISBN 83-00-00436-X.
3. GRABOWSKA, M. Bibliografia u progu XXI wieku. *Biuletyn EBIB* [on-line]. 2003, nr 10 [Dostęp 12.04.2013]. Dostępny w World Wide Web: <http://ebib.oss.wroc.pl/2003/50/grabowska.php>.
4. NAROJCZYK, K. *Dokument elektroniczny i jego opis bibliograficzny w publikacjach humanistycznych*. Olsztyn: Wydaw. Uniwersytetu Warmińsko-Mazurskiego, 2005. ISBN 83-7299-390-4.
5. PACEK, J. *Bibliografia w zmieniającym się środowisku informacyjnym*. Warszawa: Wydaw. SBP, 2010. ISBN 978-83-61464-44-0.
6. PŁOSZAJSKI, G. *Standardy w procesie digitalizacji obiektów dziedzictwa kulturowego*. Warszawa: Biblioteka Główna Politechniki Warszawskiej, 2008. ISBN 978-83-7207-797-4.
7. *IFLA Functional Requirements for Bibliographic Records (FRBR)* [on-line]. [Dostęp 12.04.2013]. Dostępny w World Wide Web: <http://www.ifla.org/>.

Szymańska, Z., Domosławska, B., Matysiak, M. Polska Bibliografia Literacka – ujęcie realistyczne. W: Bibliograficzne bazy danych i ich rola w rozwoju nauki. II Konferencja naukowa Konsorcjum BazTech, Poznań, 17-19 kwietnia 2013 [on-line]. Stowarzyszenie EBIB, 2013 [Dostęp: 30.08.2013]. Materiały konferencyjne EBIB, nr 24, Dostępny w World Wide Web: http://open.ebib.pl/ojs/index.php/Mat_konf/article/view/31. ISBN 978-83-63458-06-5.