

Przemysław Tadla

Obecność polskich utworów literackich w zagranicznych bibliotekach cyfrowych

Wprowadzenie

Biblioteka cyfrowa to relatywnie nowy cyfrowy twór umożliwiający przechowywanie i udostępnianie wirtualnych treści, niezależnie od miejsca i czasu. Ich funkcjonowanie opiera się w całości na sieci Internet i rozwija równoległe z nim. Nie powinno więc dziwić, iż prototyp wirtualnej biblioteki – „Projekt Gutenberg”, powstał już w 1971 roku. Zasady jego działania stały się wzorcem dla licznych naśladowców. Krok naprzód możliwy był dopiero po unowocześnieniu języka programowania HTML i popularyzacji samej technologii komputerowej. Spełnianie podstawowych funkcji bibliotecznych stało się możliwe od 1995 roku. Proces przemian trwał do ok. 2002 r., kiedy to rozpoczęła się epoka bibliotek cyfrowych II generacji¹. Od zarania historii bibliotek cyfrowych, główny trzon ich zasobu stanowiły teksty reprezentujące dziedzictwo kulturowe, w szczególności zaś europejski i amerykański kanon literacki.

W niniejszym artykule podjęto próbę zbadania tego zjawiska w odniesieniu do polskiego kręgu kulturowego, czyli obecności polskiego kanonu literackiego w międzynarodowych bibliotekach cyfrowych. Główną ideą było zbadanie ich ilościowego rozkładu w głównych zagranicznych agregatorach treści oraz jakościowa analiza wytypowanych kolekcji.

Istotnym problemem napotkanym przez autora na etapie założeń badawczych okazał się fakt niespójności i płynności pojęcia kanonu literackiego. Temat ten podjął Piotr Wilczek² który zauważył, że kanon zależy od doświadczeń osoby decydującej o jego kształcie. Inne ujęcie problemu przedstawił Stefan Chwin³. Wątek

¹ W.M. Kolasa, *Biblioteki cyfrowe na świecie – powstanie i rozwój* [online] http://eprints.rclis.org/16141/1/kolasa_cyrowe.pdf [dostęp 29.04.2016].

² P. Wilczek, *Kanon jako problem kultury współczesnej*, Gazeta Uniwersytecka [online] <http://gazeta.us.edu.pl/node/223361> [dostęp 20.07.2016]; tenże, *Czy istnieje kanon literatury polskiej?* [online] http://www.studiapolskie.us.edu.pl/wirtualna_katedra/lit_pol_w_swiecie_t1/02Wilczek.pdf [dostęp 20.07.2016].

³ S. Chwin, *Europa i kanon dla dorosłych*, [w:] *Europejski kanon literacki. Dylematy XXI wieku*, red. nauk. E. Wichrowska, Warszawa 2012, s. 57.

podejmuje też Csaba Gy. Kiss⁴, który spostrzegł, że w kulturze środkowoeuropejskiej kanon literacki służy również do budowania tożsamości narodowej. Proces zmian w polskim kanonie lektur szkolnych w latach 1976 do 2007 zaprezentował Krzysztof Mrowcewicz⁵. Roger Lundin⁶, zwraca natomiast uwagę na wiele istotnych różnic pomiędzy pojęciem klasyki oraz kanonu. Dyskusję podjął również czeski badacz Pavel Janousek⁷.

Analizowane biblioteki cyfrowe nie doczekały się szerszych omówień. Kluczowe osiągnięcia na polu cyfryzacji dziedzictwa kulturowego prezentuje jedynie Ivan Ronai⁸. Polskie biblioteki cyfrowe w kontekście europejskim prezentują Marcin Werla, Tomasz Parkoła oraz Cezary Mazurek⁹. Problem obecności poloniców w wybranych bibliotekach cyfrowych podjęła Katarzyna Szczepaniec¹⁰. Amerykański projekt Internet Archive opisano głównie z perspektywy usługi archiwizacji stron internetowych. Wayback Machine przedstawiła szczegółowo Katarzyna Gmerek¹¹. Niepełnym źródłem informacji wydaje się być również portal samego projektu, gdzie zamieszczono jedynie ogólne dane¹². Znacznie lepiej scharakteryzowano jednak projekt Google Books. Wybór dostępnych dzieł oraz zasadę działania tej biblioteki opisał Renata Matusiak¹³. Porównanie sytuacji prawnej Google Books w Stanach Zjednoczonych oraz Zjednoczonym Królestwie, opisuje Paul Ganley¹⁴.

⁴ C. Gy. Kiss, *Gdzie literatura tworzyła naród (Dylematy kultur środkowoeuropejskich)*, [w:] *Europejski kanon literacki. Dylematy XXI wieku*, red. nauk. E. Wichrowska, Warszawa 2012, s. 85–89.

⁵ K. Mrowcewicz, *Inteligencja polska w świetle kanonu lektur szkolnych. Zmierzch klasyki literackiej czy zmierzch inteligencji?*, [w:] *Inteligencja w Polsce. Specjaliści, twórcy, klerkowie, klasa średnia?*, red. H. Domański, Warszawa 2008, s. 423–439

⁶ R. Lauter, *Canons and Contexts*, New York 1991, s. ix. [cyt. za] P. Wilczek, *Kanon jako problem kultury...*

⁷ P. Janousek, *O nowym historycznoliterackim paradygmacie, historii, kanonie i badaczach literatury* [online] <https://repozytorium.amu.edu.pl/handle/10593/1098> [dostęp 20.07.2016].“

⁸ I. Ronai, *Kulturalis digitalizalas-felúton az Europeana Fele*, „Tudományos es Muszaki Tajekoztatás (Scientific and Technical Information; Hungarian journal of library and information science)” 8.56 (2009): 371-377 [online] <https://www.infona.pl/resource/bwmeta1.element.62628895-e26e-3631-84af-d01e75a52db5> [dostęp 30.06.2017]

⁹ C. Mazurek, T. Parkoła, M. Werla. *Rola bibliotek cyfrowych w zachowaniu regionalnej kultury i tradycji*, „Studia Oeconomica Posnaniensia” 1.5 (254) (2013) [online] http://cejsh.icm.edu.pl/cejsh/element/bwmeta1.element.desklight-d2aa40c0-c141-4be8-957d-b083f-61cc923/c/mazurek_parkola_werla.pdf [dostęp 30.06.2017]

¹⁰ K. Szczepaniec, *Polonica w wybranych międzynarodowych bibliotekach cyfrowych* [online] <http://jbc.bj.uj.edu.pl/Content/180403/praca%20magisterska.pdf> [dostęp 30.06.2017]

¹¹ K. Gmerek, *Archiwa internetowe po obu stronach Atlantyku – Internet Archive, Wayback Machine oraz UK Web Archive*, „Biuletyn EBIB” 1 (2012): 128 [online] http://www.ebib.pl/images/stories/numery/128/128_gmerek.pdf [dostęp 30.06.2017]

¹² <https://archive.org/about/>

¹³ R. Matusiak, *Dostęp do pełnych tekstów książek w Google Book Search*, [w:] „Biuletyn EBIB”, 2008, nr 3, [online] <http://www.ebib.pl/2008/94/a.php?mateusiak>, [dostęp 30.06.2017]

¹⁴ P. Ganley, *Google Book Search: Fair use, fair dealing and the case for intermediary copying* (2006) [online] <https://papers.ssrn.com/sol3/papers.cfm?abstract-id=875384> [dostęp 30.06.2017]

Podobne uwagi podnosi Corinna Baksik¹⁵. Z kolei Marcin Malinowski w swoim artykule¹⁶ sygnalizuje problem możliwej nieautorytatywności informacji dostarczanych przez wyszukiwarę.

Kanon literatury polskiej

Pierwszym krokiem na drodze do empirycznego zbadania obecności polskiego kanonu literackiego w zasobach bibliotek cyfrowych jest ustalenie roboczej listy dzieł doń należących. Nie jest to jednak czynność prosta, gdyż jak ustalił P. Wilczek kanon literacki jest pojęciem względnym, zależnym często od doświadczeń szkolnych osoby odpowiedzialnej za jego przygotowanie. Ponadto ukształtowany obecnie konsensus lekturowy jest nadal związany z uwarunkowaniami poprzedniego systemu politycznego i przez to ma inny charakter niż jego odpowiedniki na Zachodzie. Mniejsze więc znaczenie ma pochodzenie czy płeć twórcy (bądź też jego poglądy), a ciężar doboru lektur dla uczniów szkół podstawowych i średnich skupia się raczej na treści i okolicznościach powstania dzieła. Co więcej, należało zauważyć niebezpieczne zjawisko uznawania za klasyczne dzieł z różnych powodów dobrze się sprzedających, bez względu na ich wartość literacką. Zjawisko to potęgują reprodukcje filmowe danego utworu. W związku ze specyfiką kina, opierające się nieraz bardzo luźno na fabule pierwowzoru, filmy zdobywają często ogromną popularność, co skutkuje jednocześnie wzrostem popularności danej książki. W takim przypadku dana pozycja zdobywa niejednokrotnie miano klasycznej, a to w ostateczności może doprowadzić do jej wprowadzenia na listę lektur. Wydaje się więc, iż rozwój kinematografii wpływa w pewnym sensie na kształt kanonów literackich. Dlatego też omawiając ten problem nie sposób nie uwzględnić również reprodukcji dzieł.

Rozpatrując – na potrzeby badania – kształt kanonu literackiego postanowiono również o zawężeniu zakresu tego pojęcia do twórców wyłącznie rodzimych. Ponadto wnioskując z niewielkich jedynie przeobrażeń kanonu lektur szkolnych w wieku ubiegłym można stwierdzić, iż mają one charakter do pewnego stopnia niezmienny. Stanowisko takie przyjęto również w niniejszym studium, gdyż okazało się ono użyteczne w trakcie badań. Umożliwiło bowiem operacjonalizację hipotez i selekcję materiału.

Dużą pomocą w ustaleniu wykazu dzieł tworzących korpus polskich utworów klasycznych odegrały dwa źródła: „Wirtualna Biblioteka Literatury Polskiej” założona przez Marka Adamca¹⁷, zawierająca najważniejsze zdaniem założyciela pozycje czytelnicze kultury polskiej, a także lista lektur obowiązkowych zawarta w rozporządzeniu Ministra Edukacji Narodowej z dnia 3 lipca 2007 roku¹⁸, określa-

¹⁵ C. Baksik, *Fair use or exploitation? The google book search controversy*, portal: Libraries and the Academy 6.4 (2006): 399-415 [online] <https://muse.jhu.edu/article/203348/summary> [dostęp 30.06.2017]

¹⁶ M. Malinowski, *Google Books w badaniach naukowych – wcale nie tak różowo*, *Historia i Media*, [online] <http://historiaimedia.org/2010/09/02/google-books-w-badaniach-naukowych/> [dostęp 8.07.2016]

¹⁷ M. Adamiec, „Wirtualna Biblioteka Literatury Polskiej” [online] <http://literat.ug.edu.pl/books.htm> [dostęp 20.07.2016]

¹⁸ Rozporządzenie Ministra Edukacji Narodowej z dnia 3 lipca 2007 r. zmieniające rozporządzenie w sprawie podstawy programowej wychowania przedszkolnego oraz kształcenia ogólnego w poszczególnych typach szkół, Dz.U. 2007 nr 123 poz. 853

jąca kształt ostatniego przed zmianami kanonu szkolnego. Należy nadmienić, iż w trakcie procesu badawczego ogłoszono pracę nad nowym kanonem lektur szkolnych. Jednakże ponieważ przedstawiono jedynie projekt nowego rozporządzenia, postanowiono posłużyć się jego starszą wersją. Ponadto z kanonu lektur szkolnych odrzucono wszystkie utwory zagranicznych autorów oraz dzieła religijne. Co więcej, w procesie kwalifikacji dokumentów brano pod uwagę również dzieła wzorowane na danym utworze. Uwzględniano zatem, także dokumenty ikonograficzne (fotografie, kadry filmowe, pocztówki, albumy), kartograficzne (mapy), muzyczne (nagrania audio, druki muzyczne) czy filmowe. Przygotowując wstępną listę utworów starano się zachować podobną liczbę dzieł dla każdej z epok literackich. Celem zachowania reprezentatywności próby badawczej konieczne było więc odrzucenie niektórych tytułów, w tym również lektur szkolnych. W tym celu skorelowano wstępnie wybraną listę utworów z „Wirtualną Biblioteką Literatury Polskiej”¹⁹. Wobec tego, chociaż wyłoniony kanon literacki bazuje na kanonie lektur szkolnych, został on jednak ograniczony ze względu na konieczność selekcji materiału. Wybraną listę zakwalifikowanych do badania utworów przedstawiono w tabeli 1.

Na wyłoniętej liście [tab. 1] okres średniowiecza okazał się być reprezentowany jedynie w minimalnym stopniu. Po przeanalizowaniu dostępnych źródeł postanowiono o zakwalifikowaniu zaledwie dwóch utworów: *Bogurodzicy* oraz *Kroniki i czynów książąt czyli władców polskich* Galla Anonima. Do epoki renesansu zaliczyć można dzieła sześciu autorów: *Treny* Jana Kochanowskiego, *Krótką rozprawę między trzema osobami Panem, Wójtem a Plebanem* Mikołaja Reja, *Kazania Sejmowe* Piotra Skargi, *Sonet* Mikołaja Sępa Szarzyńskiego oraz *Sielanki* Szymona Szymonowica. W przypadku dzieła Andrzeja Frycza Modrzewskiego zastosowano pisownię polską oraz łacińską (*O poprawie Rzeczypospolitej* oraz *Commentariorum de Republica emendanda libri quinque*). Barok reprezentują jedynie trzy utwory: *Lutnia* Jana Andrzeja Morsztyna, *Cuda Miłości* Zbigniewa Morsztyna tudzież *Ogród ale nieplewiony* Wacława Potockiego. Utwory te okazały się być najczęściej obecnymi na branych pod uwagę zestawieniach lektur czy dzieł uznanych za klasyczne. Analogicznie do baroku, niewiele zakwalifikowano też utworów z epoki oświecenia. Na liście znalazły się *Mikołaja Doświadczyńskiego przypadki* Ignacego Krasickiego, *Do wąsów* Franciszka Dionizego Książnina, *Uwagi śmierci niechybnej* Józefa Baki oraz *Laura i Filion* Franciszka Karpińskiego.

Tab. 1. Tab. 1. Kanon literatury polskiej (układ według epok)

Epoka literacka	Autor
Średniowiecze	Gall Anonim
Renesans	J. Kochanowski, A. Modrzewski, M. Rej, Skarga, M. Szarzyński, S. Szymonowic
Barok	J. Morsztyn, Z. Morsztyn, W. Potocki
Oświecenie	J. Baka, F. Karpiński, I. Krasicki, Książnin
Romantyzm	A. Fredro, Z. Krasiński, A. Mickiewicz, C.K. Norwid, J. Słowacki
Pozytywizm	E. Orzeszkowa, B. Prus, H. Sienkiewicz

¹⁹ M. Adamiec, „Wirtualna Biblioteka Literatury Polskiej” [online] <http://literat.ug.edu.pl/books.htm> [dostęp 20.07.2016]

Młoda Polska	J. Kasprówic, W. Reymont, L. Staff, K. Tetmajer, S. Wyspiański
XX-lecie międzywojenne	M. Dąbrowska, W. Gombrowicz, J. Lechoń, B. Leśmian, Z. Nałkowska, J. Przyboś, B. Schulz, A. Słonimski, J. Tuwim, K. Wierzyński, S.I. Witkiewicz, S. Żeromski
Współczesność	S. Barańczak, M. Białoszewski, T. Borowski, J. Korczak, W. Broniewski, A. Czerniawski, M. Głowiński, Z. Herbert, A. Hertz, I. Jurgielewiczowa, R. Kapuściński, H. Krall, S. Lem, C. Miłosz, S. Mrożek, T. Różewicz, W. Szyborska, A. Zagajewski,

Źródło: opracowanie własne

Spośród dzieł powstałych w epoce romantyzmu uwzględniono dwa dzieła Adama Mickiewicza (*Dziady* oraz *Pana Tadeusza*), dwa utwory Juliusza Słowackiego (*Balladynę* oraz *Kordiana*), *Zemstę* Aleksandra Fredry, *Nie-Boską Komedię* Zygmunta Krasińskiego oraz *Vademecum* Cypriana Kamila Norwida. Z dzieł przynależnych do epoki pozytywizmu uwzględniono trylogię (*Ogniem i mieczem*, *Pana Wołodyjowskiego* oraz *Potop*) Henryka Sienkiewicza, *Lalkę* Bolesława Prusa oraz *Nad Niemnem* Elizy Orzeszkowej. Epoka Młodej Polski, reprezentowana była przez 6 autorów. Korpus ten tworzyli: *Chłopi* Władysława Reymonta, *Krzak dzikiej róży w Ciemnych Smreczynach* Jana Kasprówicza, *Koniec wieku XIX* Kazimierza Przerwy-Tetmajera, *Wesele* Stanisława Wyspiańskiego oraz *Kowal* Leopolda Staffa.

Do najnowszych epok polskiej literatury (XX-lecia międzywojennego oraz współczesności – datowanej na lata od 1939 do współczesności) zaliczyć można największą spośród zakwalifikowanych liczbę autorów. Do klasycznych dzieł przynależnych do pierwszej epoki przyporządkować można: *Mieszkańców* Juliana Tuwima, *Ferdydurke* Witolda Gombrowicza, *Przedwiośnie* Stefana Żeromskiego, *Granicę* Zofii Nałkowskiej, *Noce i dni* Marii Dąbrowskiej, *Karmazynowy poemat* Jana Lechonia, *Lękę* Bolesława Leśmiana, *Póki my żyjemy* Juliana Przybosia, *Sanatorium pod klepsydrą* i *Sklepy cynamonowe* Brunona Schulza, *Wieżę Babel* Antoniego Słonimskiego, *Kufer na plecach* Kazimierza Wierzyńskiego oraz *Szewców* Stanisława Ignacego Witkiewicza. Do literackiej współczesności: *Pożegnanie z Marią* Tadeusza Borowskiego, *Tango* Sławomira Mrożka, *Zdążyć przed Panem Bogiem* Hanny Krall, *Ocalonego* Tadeusza Różewicza, *Jakieś Ty* Stanisława Barańczaka, *Dziennik/Pamiętnik* Janusza Korczaka, *Ankę* Władysława Broniewskiego, *Pamiętnik z powstania warszawskiego* Mirona Białoszewskiego oraz dzieł dalszych utworów: *Fragmenty niespokojnego dzieciństwa* Adama Czerniawskiego, *Czarne sezony* Michała Głowińskiego, *Raport z oblężonego miasta* Zbigniewa Herberta, *Żydzi w kulturze polskiej* Aleksandra Hertza, *Ten obcy* Ireny Jurgielewiczowej, *Cesarz* Ryszarda Kapuścińskiego, *Bajki robotów* Stanisława Lema, *Traktat poetycki* Czesława Miłosza, *Nic dwa razy* Wisławy Szymborskiej oraz wiersz *Prawda* Adama Zagajewskiego. Ostatecznie do badania zakwalifikowano 63 różne dzieła autorstwa 57 pisarzy.

Cyfrowe agregatory treści

Aby zweryfikować założone na wstępie cele badawcze zbadano zasoby trzech największych światowych bibliotek cyfrowych (agregatorów treści), które gromadzą dzieła należące do kręgu kultury zachodniej, a więc serwisy: Europeana, Internet

Archive oraz Google Books²⁰. Wybór taki został podyktowany faktem, że usługi te scalają większość narodowych bibliotek tego typu i są wobec siebie komplementarne. Pierwszą z nich jest Europeana [<http://www.europeana.eu>], który scala pod wspólnym interfejsem niemal wszystkie biblioteki z krajów Unii Europejskiej i udostępnia 54 358 074 jednostek²¹ (stan na maj 2017). Drugą – Internet Archive [dalej: IA] – „biblioteka internetowa”²², będąca częścią Open Content Alliance, do jej celów należy oferowanie stałego dostępu do kolekcji historycznych, istniejących w formacie cyfrowym, badaczom, historykom, uczonym, osobom niepełnosprawnym i szeroko pojętej publiczności²³. Obecnie portal oferuje dostęp do m.in. 13 milionów książek, 298 miliardów stron www, 3 milionów plików wideo czy też 3,5 miliona plików audio²⁴ (łącznie około 300 miliardów różnego rodzaju zbiorów). Ostatnią z instytucji biorących udział w badaniu jest Google Books. Ta komercyjna platforma jest częścią firmy Google, która oprócz wspomnianej usługi oferuje również dostęp do wielu innych – często unikalnych – narzędzi i witryn. Należy jednak wspomnieć, iż jedynie część spośród zgromadzonych przez bibliotekę danych dostępna jest w domenie publicznej. W badaniu uwzględniono jedynie publikacje oferowane w otwartym dostępie. Kwalifikując Google Books do badania brano pod uwagę zasięg i międzynarodową pozycję tej organizacji, jednocześnie uwzględniając przyjętą w jego ramach zasadę pełnej otwartości. Badanie realizowano w pierwszym kwartale 2017 roku. Głównym kryterium wyszukiwawczym było pochodzenie obiektu. Za prawidłowe uznawano również dzieła tłumaczone na języki obce. Celem ograniczenia liczby obiektów badawczych odfiltrowano fałszywe trafienia, wynikające ze zbieżności nazwisk bądź błędnie zaklasyfikowane. Za istotne uznano nie tylko poszczególne dzieła, ale również ich autorów.

Wykres 1 prezentuje liczbę wszystkich dzieł udostępnianych przez analizowane biblioteki cyfrowe. Największą liczbę zbiorów odnotowano w Europeanie (ok. 21,9 milionów). Europejski agregator przechowuje też największą liczbę dzieł w języku polskim (2,1 miliona tj. ok. 10% wszystkich rekordów stanowią utwory polskojęzyczne). 12 milionów tekstów udostępnia natomiast Internet Archive. Jest to więc najmniejszy z uwzględnionych księgozbiorów. Co więcej zbiorów polskojęzycznych biblioteka przechowuje jedynie 6441 (mniej niż 1%). Dane dotyczące Google Books zaczerpnięto z oficjalnych wypowiedzi pracowników korporacji²⁵. Poinformowali oni o przekroczeniu kamienia milowego 20 milionów książek w roku 2012. Ponadto używając wyszukiwania zaawansowanego oraz maski „*”, udało się ustalić orientacyjną liczbę wszystkich udostępnianych dzieł w języku polskim na ok. 1,1 miliona (5%). Należy jednak nadmienić, iż zarówno Europeana, jak i Internet Archive przechowują w swoich katalogach pliki multimedialne, a nawet – w przypadku IA – oprogramowanie komputerowe.

²⁰ W.M. Kolasa, *Biblioteki cyfrowe na świecie – powstanie i rozwój*, [w:] *Biblioteki cyfrowe*, red. M. Próchnicka, M. Janiak, M. Krakowska, Warszawa 2012, s. 66–95.

²¹ Link do oficjalnych statystyk Europeany: <http://pro.europeana.eu/about-us/europeana-statistics/2017-q1>

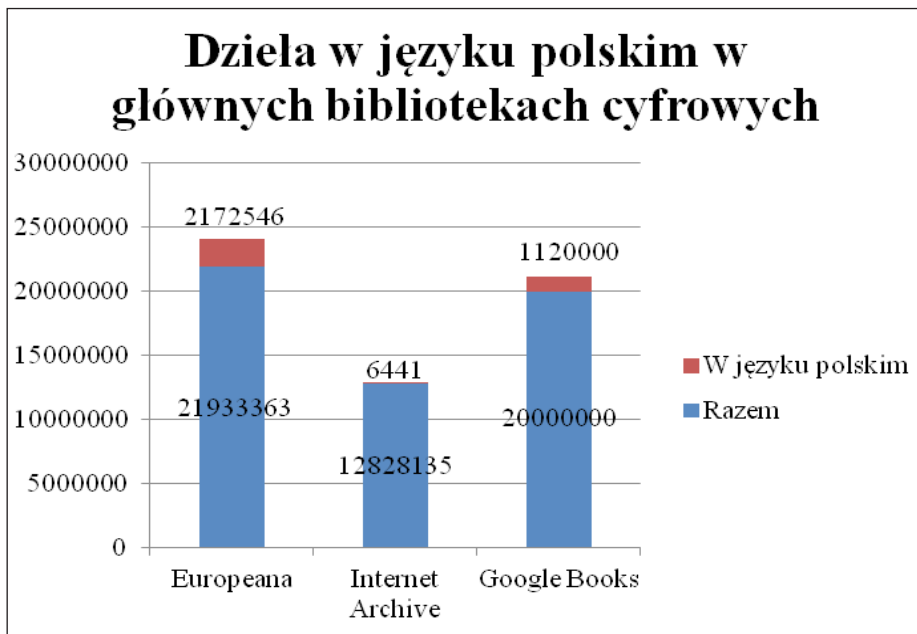
²² K. Gmerek, *Archiwa internetowe po obu stronach Atlantyku...*

²³ Ibidem.

²⁴ Link do oficjalnych danych Internet Archive: <https://archive.org/about/>

²⁵ J. Howard, *Google Begins to Scale Back Its Scanning of Books From University Libraries*, *The Chronicle of Higher Education*, 2012 [online] <http://www.chronicle.com/article/Google-Begins-to-Scale-Back/131109/> [dostęp 01.07.2017].

Wyk. 1. Dzieła w języku polskim w badanych bibliotekach cyfrowych (stan na 06.09.2016)



Źródło: opracowanie własne.

Rezultaty

W ramach prowadzonych badań przeprowadzono wyszukiwanie zbioru przygotowanych utworów w wybranych 3 bibliotekach międzynarodowych – Europeana, Internet Archive oraz Google Books. Na podstawie uzyskanych wyników opracowano zestawienie kanonu literatury polskiej obecnego w zagranicznych księżnicach wirtualnych, które zaprezentowano w tab. 2.

Tab. 2. Wybrane polskie utwory literackie w międzynarodowych bibliotekach cyfrowych

Tytuł/Autor	W języku polskim			W innych językach			Razem			łącznie
	EU	GB	IA	EU	GB	IA	EU	GB	IA	
<i>Dziady</i> – Mickiewicz	444	26	1	13	18	1	457	44	2	503
<i>Tango</i> – Mrozek	4	5	0	249	11	0	253	16	0	269
<i>Ogniem i mieczem</i> – Sienkiewicz	146	28	65	1	5	2	147	33	67	247
<i>Pan Tadeusz</i> – Mickiewicz	121	25	3	11	26	5	132	51	8	191
<i>Potop</i> – Sienkiewicz	66	19	64	1	6	7	67	25	71	163
<i>Wesele</i> – Wyspiański	104	18	1	12	10	0	116	28	1	145

<i>Pan Wołodyjowski – Sienkiewicz</i>	21	31	64	0	13	1	21	44	65	130
<i>Kordian – Słowacki</i>	75	17	0	11	7	0	86	24	0	110
<i>Chłopi – Reymont</i>	46	29	14	1	18	0	47	47	14	108
<i>Bogurodzica</i>	60	18	1	1	9	0	61	27	1	89
<i>Balladyna – Słowacki</i>	54	18	0	1	9	0	55	27	1	83
<i>Treny – Kochanowski</i>	40	12	1	4	5	0	44	17	1	62
<i>Lalka – Prus</i>	26	12	13	1	7	0	27	19	13	59
<i>Nie-Boska Komedia – Krasieński</i>	46	6	0	2	2	0	48	8	0	56
<i>Vade-mecum – Norwid</i>	22	18	0	1	6	0	23	24	0	47
<i>Nad Niemnem – Orzeszkowa</i>	23	14	2	0	5	0	23	19	2	44
<i>Zemsta – Fredro</i>	19	19	0	0	5	0	19	24	0	43
<i>Kazania sejmowe – Skarga</i>	11	16	0	4	5	0	15	21	0	36
<i>Ferdydurke – Gombrowicz</i>	4	9	0	9	10	1	13	19	1	33
<i>Sielanki – Szymanowicz</i>	16	8	1	4	4	0	20	12	1	33
<i>Przedwiośnie – Żeromski</i>	18	9	0	2	4	0	20	13	0	33
<i>Sklepy cynamonowe – Schulz</i>	11	5	2	1	6	0	12	11	2	25
<i>O popr. Rzeczyp. – Modrzewski</i>	6	4	0	5	9	0	11	13	0	24
<i>Mikołaja Dośw. Przyp. – Krasicki</i>	11	8	0	0	3	0	11	11	0	22
<i>Traktat poetycki – Miłosz</i>	1	14	0	0	5	0	1	19	0	20
Inne utwory	39	120	3	7	48	1	46	168	4	218
Razem	1434	508	235	341	256	19	1775	764	254	2793

Źródło: opracowanie własne. Wyniki w poszczególnych komórkach tabeli oznaczają kolejno wyniki dla: EU – Europeana, GB – Google Books, IA – Internet Archive.

Zgodnie z założeniami odnaleziony zasób powinien być oferowany w wolnym dostępie. Zastosowana metoda badawcza to analiza zawartości medium, dokonana za pomocą pytań wyszukiwawczych zadawanych wyszukiwarkom – zarówno z użyciem przeszukiwania prostego, jak i zaawansowanego. Przeszukiwanie rozpoczęto każdorazowo od wpisania instrukcji wyszukiwawczej w polu wyszukiwania

prostego celem wstępnego oszacowania liczby dostępnych zasobów. W drugiej kolejności wyniki filtrowano według autora i języka. Wykorzystano w tym celu przygotowaną listę pisarzy polskich, których praca wywarła trwały wpływ na kulturę polską. Każdorazowo przy charakteryzowaniu trafień starano się ustalić typ odnalezionej jednostki oraz możliwości jej dalszego wykorzystania przez czytelnika. Po zakończeniu procesu wyszukiwania przystępowano do opracowania wyników badania. W trakcie analizy zwracano też uwagę na inne obiekty odpowiadające zapytaniu wyszukiwawczemu, na przykład obiekty audiowizualne.

Kolejnym etapem badania było przygotowanie zestawienia zbiorczego ukazującego liczbę odnalezionych pozycji w każdej kategorii przewidzianej przez interfejs biblioteki. Wyniki wyszukiwania scharakteryzowano na podstawie: autorstwa obiektu, typu zasobu oraz jego języka. Uzyskane wyszukiwania charakteryzowano też pod względem ilościowym. W przypadku nazwisk zawierających litery diakrytyczne lub też osób posługujących się pseudonimem artystycznym przeprowadzono dwukrotne wyszukiwanie. Instrukcje dla wyszukiwarek były formułowane z użyciem cudzysłowu, tj. w następujący sposób – „Nazwisko Imię”. Jeżeli wyniki wyszukiwania wydawały się wskazywać na błąd mechanizmu – a było tak w przypadku niewielkiej ilości rezultatów dla nazwisk znanych twórców – badanie powtarzano bez użycia cudzysłowu, tj. używając składni – Nazwisko Imię. Odnalezienie w zasobach danej wirtualnej ksiąźnicy jednostek pasujących do zapytania wyszukiwawczego odnotowywano jako trafienie(/a) i umieszczano w odpowiednich tabelach.

Biorące udział w badaniu dzieła literatury polskiej są najliczniej reprezentowane przez **Europeane**. Łączna liczba wszystkich odnalezionych jednostek wyniosła 1775. Najwięcej trafień dotyczyło *Dziadów* Adama Mickiewicza (457). Dużą część całkowitego rezultatu Europeany stanowiły też trafienia dla *Tanga* Sławomira Mrożka (253). Pozostałe utwory cieszyły się znacznie mniejszą popularnością. Wspomnieć należy chociażby o 147 trafieniach dla *Ogniem i mieczem* Henryka Sienkiewicza czy 67 jednostkach *Potopu* tego samego autora, jak również o 132 rekordach odnoszących się do *Pana Tadeusza* Adama Mickiewicza. Aż 17 biorących udział w tej części badania utworów było reprezentowanych w minimalnym jedynie stopniu – liczba trafień nie przekroczyła 10, natomiast 18 tytułów nie było w ogóle dostępnych w bazach omawianej biblioteki cyfrowej. Rezultaty badania pozwalają wnioskować, iż w omawianej bibliotece cyfrowej tłumaczenia wybranych tekstów literatury polskiej nie są licznie reprezentowane.

Najwięcej trafień w języku polskim odnotowano – podobnie, jak w przypadku wyników ogólnych – dla *Dziadów* Adama Mickiewicza (444). Tak duża liczba odnalezionych jednostek okazała się być wyjątkiem ponieważ drugi najliczniejszy tytuł – *Ogniem i mieczem* Henryka Sienkiewicza – był reprezentowany jedynie przez 146 trafień. Niewiele mniej zasobów odnaleziono wyszukując *Pana Tadeusza* Adama Mickiewicza (121). W przypadku 18 autorów nie odnotowano więcej niż 10 jednostek. Okazało się, że 23 tytuły nie zostały w ogóle w bazach Europeany odnalezione.

Poza wyjątkiem obecności w bazach Europeany aż 167 jednostek w języku czeskim odnoszących się do *Tanga* Sławomira Mrożka (dla języka węgierskiego odnotowano 72 pozycje), łączna liczba wyników dla pozostałych języków obcych nie przekroczyła 100 rekordów. Duża liczba zasobów w innych językach, wynikała ze sporej ilości dokumentów wielojęzycznych (w wielu przypadkach nie pozostawiono żadnej informacji umożliwiającej jednoznaczny ich identyfikację).

W zasobach **Google Books** odnaleziono jedynie 764 wydania szukanych tytułów. Jest to więc rezultat lepszy niż w przypadku Internet Archive, ale gorszy niż w Europeanie. Największą liczbę jednostek odnotowano w przypadku *Pana Tadeusza* Adama Mickiewicza (51). Niewiele gorsze wyniki osiągnęły *Dziady* (44) tego samego autora, 47 jednostek odnaleziono dla *Chłopów* Władysława Reymonta, a 44 dla *Pana Wołodyjowskiego* Henryka Sienkiewicza. W przebiegu badania udało się w Google Books odnaleźć niemal wszystkie zakładane utwory, za wyjątkiem czterech; natomiast 32 tytuły odznaczyły się jedynie minimalną obecnością. Spośród nich aż trzy – tj. *Uwagi śmierci niechybnej* Józefa Baki, *Ogród ale nieplewiony* Wacława Potockiego oraz *Sonety* Mikołaja Sępa Szarzyńskiego – nie były w ogóle obecne w żadnej z branych pod uwagę bibliotek cyfrowych. *Dziennik* Janusza Korczaka obecny był tylko w jednej kopii i to tylko w Internet Archive. Biorąc to pod uwagę można stwierdzić, iż Google Books osiągnęło najlepsze ogólne rezultaty.

Łączna liczba odnotowanych dokumentów w języku polskim wyniosła 508 jednostek. Warto dodać, że jeden z utworów – *Raport z oblężonego miasta* Zbigniewa Herberta – dostępny był w Google Books tylko w języku angielskim. Najwięcej wyników odnotowano dla dzieł Henryka Sienkiewicza – *Pan Wołodyjowski* (31) i *Ogniem i mieczem* (28) oraz Adama Mickiewicza – *Dziady* (26) i *Pan Tadeusz* (25). Odnaleziono również 29 rekordów dzieła *Chłopi* Władysława Reymonta. Natomiast aż 39 tytułów (59%) obecne było jedynie w minimalnych ilościach (nie więcej niż 10 jednostek). W opisywanej bibliotece cyfrowej siedem utworów w ogóle nie było obecnych.

Za wyjątkiem języka angielskiego – w przypadku którego zanotowano aż 210 jednostek – Google Books zawarło w swoich bazach jedynie jednostkowe wyniki dla pozostałych języków obcych. Omawiana wirtualna księżnica najlepiej ze wszystkich branych pod uwagę w badaniu poradziła sobie z językiem niemieckim – odnotowano 25 jednostek (patrz tab. 2). Również wynik dla języka francuskiego (8) stawia Google Books na pierwszej pozycji, aczkolwiek w skali całego badania odnotowano jedynie kilkanaście francuskojęzycznych rekordów. Opisywana biblioteka cyfrowa jako jedyna zwróciła też wynik dla języka włoskiego – był to pojedynczy egzemplarz tłumaczenia *Pana Tadeusza* Adama Mickiewicza.

Internet Archive osiągnął wyniki kilkakrotnie gorsze od pozostałych biorących udział w badaniu bibliotek cyfrowych. Łącznie udało się odnaleźć jedynie 254 dokumenty. Największą popularnością cieszyły się dzieła Henryka Sienkiewicza. Odnaleziono 71 kopii *Potopu*, 67 jednostek *Ogniem i mieczem* oraz 65 *Pana Wołodyjowskiego*. Ponadto zanotowano też 14 trafień wydania *Chłopów* Władysława Reymonta oraz 13 jednostek *Lalki* Bolesława Prusa. Pozostałe tytuły zwróciły jedynie pojedyncze rezultaty. Spośród 63 utworów, aż 41 (tj. 65%) nie zostało w bazie Internet Archive odnalezionych. Jest to wynik najniższy w skali całego badania.

Wyniki dla języka polskiego są dość podobne do łącznych rezultatów ogólnych dla tej wirtualnej księżnicy. Łącznie odnaleziono 235 rekordów. Najliczniejsze wyniki zanotowano kolejno dla wszystkich trzech biorących udział w badaniu utworów Henryka Sienkiewicza, *Ogniem i mieczem* reprezentowane było przez 65 pozycji. Sporo bo aż 64 jednostek odnaleziono ex aequo dla *Pana Wołodyjowskiego* oraz *Potopu*. W przypadku *Chłopów* Władysława Reymonta (14) oraz *Lalki* Bolesława Prusa (13) odnotowane zasoby dostępne były tylko w języku polskim. *Dziennik* Janusza

Korczaka oraz *Ferdydurke* Witolda Gombrowicza osiągalne były tylko w językach obcych. W języku polskim nie były dostępne aż 43 utwory.

Internet Archive odnotował w zakresie języków obcych zdecydowanie najgorsze w badaniu wyniki. Najwięcej jednostek odnaleziono dla języka angielskiego (13). Oprócz tego odnotowano jedynie po dwa rekordy dla języków: francuskiego, oraz niemieckiego. Odnaleziono też dwie jednostki oznaczone jako wielojęzyczne, lub wydane w języku występującym w badaniu jedynie akcydentalnie (jidysz) – były to *Ogniem i mieczem* Henryka Sienkiewicza oraz *Dziennik* Janusza Korczaka.

Po analizie uzyskanych wyników stwierdzono relatywnie niewielką liczbę zbiorów w części europejskich języków obcych. Celem właściwego opracowania materiału zdecydowano o konieczności wydzielenia zbiorów anglo- i niemieckojęzycznych. Zdecydowała o tym, zarówno popularność tych języków, jak również wyróżniająca się liczba takich zasobów na tle całego badania. Pełną listę wyników dla języków angielskiego i niemieckiego przedstawiono w tabeli nr 3.

Tab. 3. Lista wyników dla języka angielskiego i niemieckiego w międzynarodowych bibliotekach cyfrowych

Tytuł/Autor	W języku angielskim			W języku niemieckim			łącznie
	EU	GB	IA	EU	GB	IA	
<i>Pan Tadeusz</i> – Mickiewicz	0	21	4	1	1	1	24
<i>Chłopi</i> – Reymont	0	18	0	1	0	0	19
<i>Dziady</i> – Mickiewicz	0	14	0	1	1	0	16
<i>Pan Wołodyjowski</i> – Sienkiewicz	0	7	1	0	6	0	14
<i>Potop</i> – Sienkiewicz	0	6	7	0	0	0	13
<i>Tango</i> – Mrozek	0	8	0	1	3	0	12
<i>Balladyna</i> – Słowacki	0	9	0	0	0	1	10
<i>Wesele</i> – Wyspiański	0	9	0	0	1	0	10
<i>Ferdydurke</i> – Gombrowicz	0	8	0	0	1	0	9
<i>Bogurodzica</i>	0	8	0	0	0	0	8
<i>Kazania sejmowe</i> – Skarga	1	5	0	2	0	0	8
<i>Ogniem i mieczem</i> – Sienkiewicz	1	3	1	0	2	0	7
<i>Sielanki</i> – Szymonowicz	0	2	0	3	2	0	7
<i>Noce i dnie</i> – Dąbrowska	0	5	0	0	1	0	6
<i>Sklepy cytronowe</i> – Schulz	0	5	0	0	1	0	6
<i>Kordian</i> – Słowacki	0	6	0	0	0	0	6
<i>Treny</i> – Kochanowski	0	4	0	0	1	0	5
<i>Traktat poetycki</i> – Miłosz	0	5	0	0	0	0	5
<i>Vade-mecum</i> – Norwid	0	5	0	0	0	0	5
<i>Zemsta</i> – Fredro	0	4	0	0	0	0	4

<i>Laura i Filon</i> – Karpiński	0	4	0	0	0	0	4
<i>Bajki robotów</i> – Lem	0	4	0	0	0	0	4
<i>Nad Niemnem</i> – Orzeszkowa	0	4	0	0	0	0	4
<i>Lalka</i> – Prus	0	4	0	0	0	0	4
<i>Przedwiośnie</i> – Żeromski	0	4	0	0	0	0	4
Inni	2	38	0	2	5	0	47
Razem	4	172	13	11	25	2	227

Źródło: opracowanie własne. Wyniki w poszczególnych komórkach tabeli oznaczają kolejno wyniki dla: EU – Europeana, GB – Google Books, IA – Internet Archive.

Największą łączną liczbę trafień dla języków angielskiego i niemieckiego zanotowano w Google Books, natomiast najmniejszą – w Internet Archive. Rozpatrując osobno poszczególne utwory, stwierdzić należy, iż najwięcej trafień w języku obcym odnotowano w Europeanie dla *Pana Tadeusza* Adama Mickiewicza (24). W przypadku pozostałych utworów liczba odnalezionych jednostek była jednak minimalna. Odnotowano więc 13 trafień *Dziadów* Adama Mickiewicza, 12 *Wesela* Wyspiańskiego oraz po 11 jednostek *Pana Tadeusza* i *Kordiana* Słowackiego. Warto również wspomnieć o 3 niemieckojęzycznych trafieniach dla *Sielanek* Szymona Szymonowicza. Analogicznie niewiele zbiorów odnotowano w Google Books. Odnaleziono jedynie 7 anglojęzycznych kopii *Potopu* Sienkiewicza, 4 jednostki *Pana Tadeusza* Adama Mickiewicza w języku angielskim i 1 w języku niemieckim. Do grupy zbiorów niemieckojęzycznych zaliczyć należy też *Balladynę* Słowackiego. Z kolei na korzyść Internet Archive przemawia fakt, iż w agregatorze tym znaleziono najwięcej zbiorów obcojęzycznych, ze wszystkich bibliotek cyfrowych biorących udział w badaniu. Tak więc, dla języka angielskiego zarejestrowano 21 jednostek *Pana Tadeusza* Adama Mickiewicza oraz 14 pozycji *Dziadów* tego samego autora. Ponadto w bazach IA odnotowano też 18 trafień dla *Chłopów* Reymonta, a także komplet trylogii Sienkiewicza w tym języku (3 dla *Ogniem i mieczem*, 6 dla *Potopu* oraz 7 dla *Pana Wołodyjowskiego*). Ponadto odnotowano po 9 jednostek *Balladyny* Słowackiego oraz *Wesela* Wyspiańskiego. Co więcej odnaleziono też jednostkowe trafienia dla kilkunastu pozostałych utworów.

Rezultaty wyszukiwania według typu obiektów

W trakcie badania wykazano istnienie problemów analizowanych bibliotek cyfrowych z kategoryzacją obiektów multimedialnych. Jedynie część dokumentów ikonograficznych charakteryzuje się przyporządkowanym językiem ich wykonania. Decyduje o tym zwłaszcza obecność adnotacji pod obrazem. W przypadku braku podpisu uznawano, że kategoria metadanych: „język dokumentu” nie znajduje zastosowania. Oznacza to, iż przeszukując bazy bibliotek pod kątem polskojęzycznych dokumentów związanych z danym pojęciem (choćby utworem, bądź twórcą), należy liczyć się z przekłamaniami związanymi z podwójną klasyfikacją danej pozycji. Konieczne jest więc odfiltrowanie wyników w zależności od przyjętego kryterium.

Problemem typowym dla Internet Archive było kategoryzowanie niektórych dokumentów jako zarówno dźwiękowych, jak i audiowizualnych. Ponadto w przypadku większości odnalezionych jednostek, nie wypełniono prawidłowo pola oznaczenia odpowiedzialności, co oznaczało, iż części wykazanych tutaj zasobów nie udało się odnaleźć w wyszukiwaniu autorskim.

Tab. 4. Wyniki wyszukiwania zakwalifikowanych utworów uszeregowane według typów obiektów

Typ obiektu	Europeana	Internet Archive	Google Books
Dok. piśmienniczy	1122	218	764
Dok. ikonograficzny	636	0	0
Dok. dźwiękowy	4	9	0
Dok. audiowizualny	13	27	0
Razem	1775	254	764

Źródło: opracowanie własne

Spośród rezultatów odnalezionych przez Europeana, aż 63% to dokumenty piśmiennicze. Możliwość pobrania, czy też nawet odczytu danego dokumentu dotyczy tylko niewielkiej grupy tekstów. W wielu wypadkach udostępnione rekordy miały charakter fragmentu danego dzieła. Istotną ułomnością Europeany jest nieprawidłowa obsługa komendy Pobierz. Zamiast udostępnienia użytkownikowi danego pliku, księżnica wysłała jedynie plik zastępczy będący częścią strony głównej biblioteki docelowej. Warto dodać, iż tylko część z tych księżnic umożliwia zapoznanie się z pełną wersją tekstu i to jedynie pod pewnymi warunkami. Europeana nie podaje jednak szczegółowych informacji w tym zakresie. Drugą grupę rekordów stanowiły dokumenty ikonograficzne. Stanowiły one 35% wszystkich odnalezionych obiektów. Należy jednak zauważyć, że duża część tych zasobów to materiały bliźniaczo do siebie podobne (przykładowo następujące po sobie kadry z filmu). Dokumenty dźwiękowe oraz audiowizualne stanowiły najmniej liczną grupę rekordów (łącznie mniej niż 1%). Należy nadmienić, że w badaniu nie uwzględniono kategorii obiektów „3D”. Patrząc na dość różnorodną zawartość kolekcji (choćby modele lamp ulicznych czy parkowych ławek), można odnieść wrażenie, iż dobór zasobów był losowy. Internet Archive wyróżnia się działem oprogramowania komputerowego (w tym programów szkodliwych – wirusów), natomiast usługa Google Books kataloguje tylko i wyłącznie dokumenty piśmiennicze.

Internet Archive udostępnił łącznie niemal 7 razy mniej zasobów niż Europeana. Wydawać by się mogło jednak, że zasoby te są dla potencjalnego czytelnika znacznie bardziej użyteczne, ponieważ składają się w większości z pełnych wydań danego dzieła. Aż 85% z nich stanowiły dokumenty piśmiennicze. Każdorazowo istniała możliwość pobrania zasobu w jednym z wybranych formatów zapisu. Dokumenty audiowizualne składały się na 10% odnalezionych rekordów. Pozostała część trafiła to dokumenty ikonograficzne i dźwiękowe. Można więc zauważyć, iż oferta Europeany i Internet Archive wzajemnie uzupełnia się w zakresie obiektów multimedialnych. Zauważono charakterystyczny dla Internet Archive brak moderacji przechowywanych dokumentów. Część odnalezionych rekordów w żaden sposób nie łączyła się z zadaniem pytaniem wyszukiwawczym. Ponadto kilkakrotnie

odnotowano sytuację, w której dany plik był falsyfikatem tzn. jego treść nie pasowała do podanego tytułu. Zgłoszenie nieprawidłowości możliwe jest jedynie po zarejestrowaniu się na stronie Internet Archive.

Google Books skatalogowało 764 dokumentów piśmienniczych, co było drugim najlepszym rezultatem w teście. W porównaniu z 1122 jednostkami Europeany oraz 218 dokumentami Internet Archive, można uznać taki rezultat za zadawalający. Wśród odnalezionych rekordów znalazł się nie tylko kanon literatury, ale również opracowania tych tekstów, ich recenzje czy też dzieła do nich nawiązujące. Część spośród trafień dostępna była w wersji demonstracyjnej (fragment tekstu). Licznie dostępne były również kolejne wydania tego samego dzieła. Wyjątkiem w skali badania była możliwość wyszukiwania danej frazy w przeglądany tekście. Funkcja ta teoretycznie jest dostępna również w Internet Archive, chociaż w przeciwieństwie do Google Books, nie działa z językiem polskim, a także nie pokazuje wyszukiwanej frazy na przeszukiwanym tekście. Żadnych narzędzi opracowania tekstu nie udostępniła natomiast Europeana.

Rezultaty wyszukiwania według dostawcy obiektów

Wszystkie biorące udział w badaniu biblioteki cyfrowe akumulują zasoby pochodzące od zewnętrznych dostawców. Sposób współpracy z partnerami zależy od zasad przyjętych przez daną bibliotekę. Można więc stwierdzić, iż każda z analizowanych księżnic jest w pewnym sensie unikalna. Przykładem niech będzie Internet Archive, który gromadzi m.in. zasoby pochodzące od indywidualnych użytkowników. Kontrowersje budzi działanie usługi Google Books, która informacje o wydawnictwach pobiera bezpośrednio z katalogu wyszukiwarki Google Search. Jak na stan obecny, firmie Alphabet udaje się skutecznie bronić prawa dozwolonego użytku.

Tab. 5. Wyniki wyszukiwania utworów według kryterium pochodzenia zasobu

Lp.	Europeana	Liczba
1	National Library of Poland	489
2	Federacja Bibliotek Cyfrowych	472
3	ECLAP,e-library for Performing Arts	343
4	The European Library	191
5	Linked Heritage	181
6	LoCloud	58
7	Slovenian National E-content Aggregator	12
8	EUscreenXL	8
9	AthenaPlus	7
10	Inne	14
	Razem	1775
Lp.	Internet Archive	Liczba

1	Canadian Libraries	184
2	Community Video	17
3	Prywatne zbiory	12
4	Community Media	10
5	American Libraries	8
6	European Libraries	7
7	The Library of Congress	6
8	The LibriCox Free Audiobook Collection LibriVox	5
9	Internet Archive	3
10	Community Audio	2
	Razem	254

Źródło: opracowanie własne

Największym dostawcą treści okazała się Biblioteka Narodowa, odpowiedzialna za aż 27% odnalezionych zasobów (489 jednostek). Instytucja współpracuje z Europeana od stycznia 2007 roku²⁶. Od września 2007 roku księżnice współpracują ze sobą w celu budowy (i rozwoju) portalu europeana.eu oraz wypełniania go treścią. Odnalezione zasoby pochodzą z Cyfrowej Biblioteki Narodowej Polona i zostały udostępnione na zasadzie domeny publicznej²⁷.

Federacja Bibliotek Cyfrowych [dalej: FBC], która udostępniła aż 26% wszystkich analizowanych zasobów, była też jedynym dostawcą pełnotekstowych zasobów w języku polskim. Warto dodać, iż rozpoczęcie współpracy pomiędzy FBC a Europeana datuje się na 2009 rok. Oznacza to, że rozwój oferty polskojęzycznych zbiorów przebiega dość szybko. Z drugiej strony nie wyróżnia się innych dostawców rodzimych treści. Ponownie daje się zauważyć, iż odpowiedzialność za digitalizację księgozbioru narodowego ciąży na bibliotekach narodowych, a Europeana stanowi jedynie agregat tych treści. Należy też nadmienić o bardzo słabej reprezentacji tłumaczeń polskich dzieł na najważniejsze języki międzynarodowe. Związane jest to ze specyfiką działania FBC, która ukierunkowana jest na współpracę z polskimi bibliotekami cyfrowymi. Godnym uwagi jest również projekt ECLAP, czy biblioteka sztuk teatralnych, która odpowiada za 19% odszukanych rekordów. Jej zaletą jest charakter udostępnianych przez nią zbiorów. Odnotowano tu znaczną liczbę pełnych nieograniczonych treści. Należy dodać, iż spośród 343 odnalezionych jednostek większość dotyczyła znanych lektur szkolnych bądź uznanych przez krytyków adaptacji. Procentowy udział dostawców treści odnalezionych w trakcie tej części badania prezentuje wykres 2.

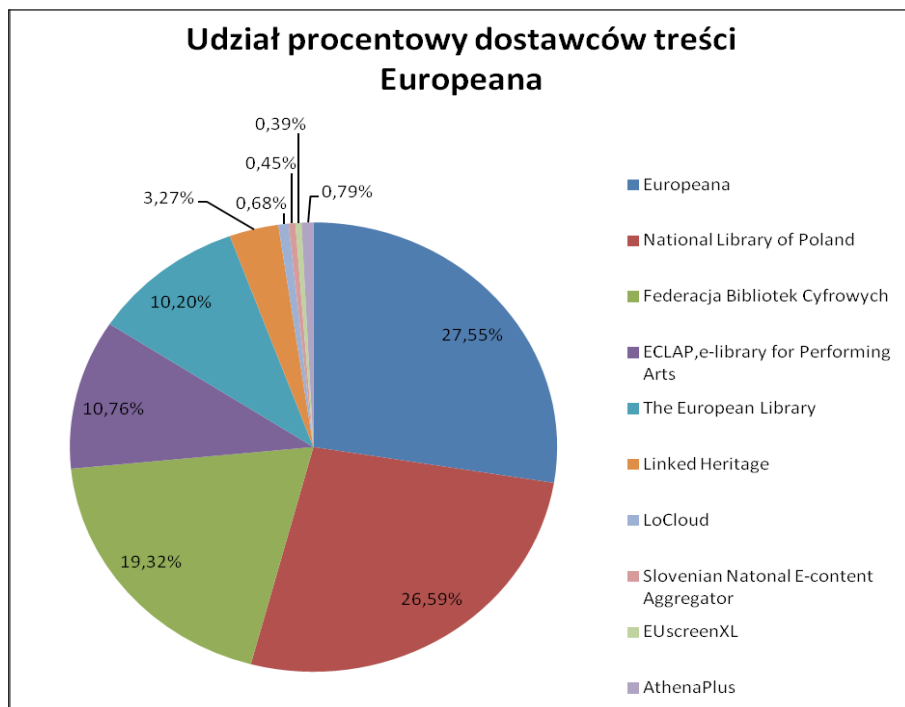
Zupełnie inaczej sytuacja wygląda w przypadku Internet Archive. „Zespół biblioteki” odpowiadał jedynie za 3 spośród odnalezionych dokumentów. Należy pamiętać, iż podpisy udostępniających nie są weryfikowane. Oznacza to, że dowolny prywatny użytkownik może przywłaszczyć sobie tożsamość obsługi witryny. Co więcej Internet Archive nie podaje konkretnych danych dostawców ograniczając

²⁶ Europeana, Biblioteka Narodowa, [online] <http://www.bn.org.pl/zasoby-cyfrowe-i-linki/europeana> [dostęp 16.02.2017]

²⁷ Ibidem.

informacje do uogólnień (np. „amerykańskie biblioteki”). Chociaż co prawda ich listę można sprawdzić, przeglądając udostępnione kolekcje²⁸, to jednak nie jest to w pełni użyteczne źródło informacji. Brakuje chociażby pełnej listy wszystkich ujętych pod tą nazwą instytucji. Uznano to za ułomność omawianej witryny.

Wyk. 2. Udział dostawców treści w wyszukiwaniu tytułowym – wyniki dla Europeana



Źródło: opracowanie własne.

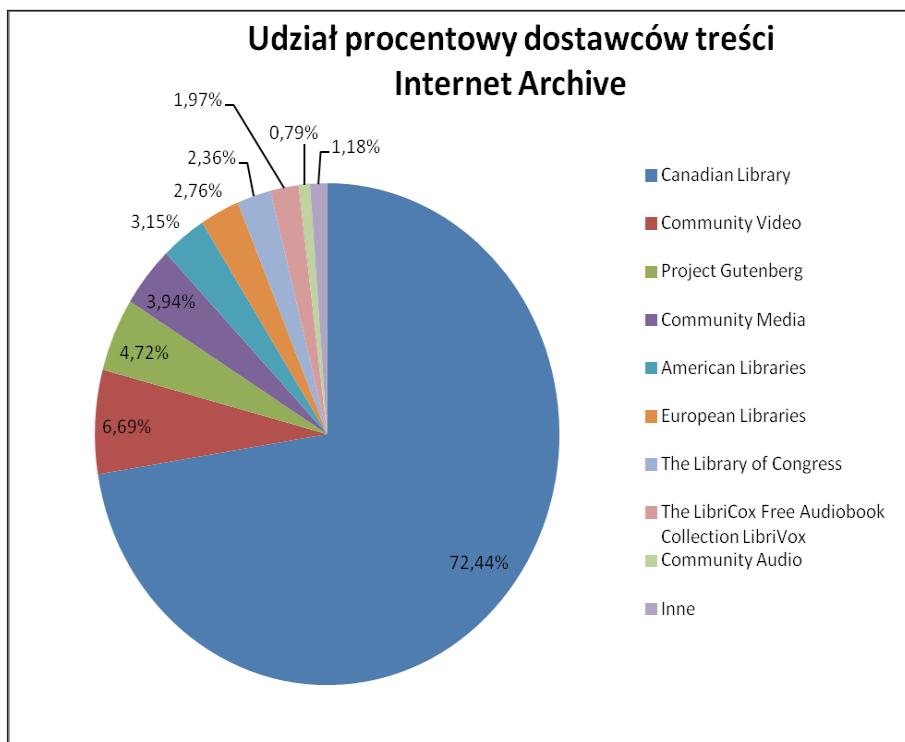
Aż 72% wszystkich zasobów dostarczyła biblioteka Uniwersytetu w Toronto (184) Ustanowione w 2004 roku centrum digitalizacji stanowi jeden z mecenasów serwisu Internet Archive. Zbiory oznaczone jako Community Media, Community Audio, Community Video stanowiły natomiast niewielką część zbiorów – jedynie 16% (41 jednostek). Należy nadmienić, że część zbiorów pochodziła z zasobów usługi Google Books bądź została witrynie przekazana przez jeden ze współpracujących instytutów. Biorąc pod uwagę wszystkie powyższe czynniki można wywnioskować, iż rzeczywista całkowita liczba rekordów różni się od danych prezentowanych przez bibliotekę. Wykres 3 prezentuje procentowy udział dostawców odnalezionych w tej części badania jednostek.

Google Books nie udostępnia statystyk użytkowania swojej usługi, co można wytłumaczyć tajemnicą firmową. Warto jednak wspomnieć, iż przechowywane zbiory podlegają retrokonwersji, dzięki czemu można przeszukiwać ich treść pod

²⁸ American Libraries, Internet Archive [online] <https://archive.org/details/americana> [dostęp 16.02.2017]

kontem np. częstotliwości występowania danego słowa w tekście. Istnieje również możliwość sprawdzenia jak często użytkownicy wyszukiwarki Google przeszukują sieć Internet w poszukiwaniu danego słowa (takiego jak tytuł książki). Jest to więc niezwykle użyteczne narzędzie mogące stanowić pomoc w uzupełnieniu badań nad czytelnictwem. Należy jednak wspomnieć o przysługującym obywatelom Unii Europejskiej prawie do bycia zapomnianym. Na jego podstawie można domagać się usunięcia danego dzieła z indeksu usługi. Aby to osiągnąć, konieczne jest podanie ważnej wiążącej prawnie przyczyny. Można więc stwierdzić, że wspomniana usługa nie wyczerpuje wszystkich potrzeb badań nad czytelnictwem.

Wyk. 3. Udział dostawców treści w wyszukiwaniu tytułowym – wyniki dla Internet Archive



Źródło: opracowanie własne

Wnioski

Przeprowadzone badania dowiodły, że dzieła należące do polskiego kanonu literackiego są szeroko obecne w międzynarodowych bibliotekach cyfrowych. Najwięcej trafień odnaleziono w Europeanie. Portal udostępnił też minimum po jednej jednostce z niemal każdego z europejskich języków obcych odnotowanych w badaniu. Drugą największą liczbę jednostek odnaleziono w zasobach Google Books. Najgorzej w badaniu wypadł natomiast Internet Archive. Zaletą Europeany (gdzie

stwierdzono 1434 dzieł polskich klasyków literackich, tj. 65%) jest dbałość o poszanowanie praw autorskich. Każdy zasób jest oznaczony odpowiednią notą prawną. Możliwe jest również posegregowanie wyników według tych zasobów, które można ponownie wykorzystać. Istnieje więc prawdopodobieństwo graniczące z pewnością, że wszystkie prawa twórców zostały odpowiednio zabezpieczone. Niestety takiej samej pewności nie może mieć użytkownik Internet Archive (235 dzieł, tj. 10%). Pomijając niewielką liczbę odnalezionych jednostek, sama biblioteka nie zawsze udostępnia informacje o możliwości ponownego wykorzystania przechowywanych zasobów. Ponadto portal nie jest konsekwentny w kwestii opracowania zbiorów. Część odnalezionych jednostek opatrzono jedynie minimalną ilością metadanych.

Patrząc na stronę edytorsko-techniczną zdecydowanie najbardziej wszechstronna okazała się bezpłatna usługa Google Books. Co prawda odnotowano tu jedynie 508 (tj. 23%) dzieł stanowiących polski kanon literacki, jednak portal nadrabia te ułomności licznymi zaletami. Projekt jest samowystarczalny, ponieważ korzysta z mechanizmów wyszukiwarki Google. Oznacza to samoczynne i automatyczne przeszukiwanie Internetu w poszukiwaniu nowych pozycji. Rekordy te są również nalezycie opisywane. Pozycje zabezpieczone mechanizmem ochrony praw autorskich, oznaczane są odnośnikiem umożliwiającym łatwe nabycie danego dzieła. Dodatkowo pierwszą pozycją na liście wyników jest notka encyklopedyczna o wyszukiwanej frazie. Oferowane zbiory pochodzą nie tylko z bibliotek narodowych, ale również z księgarni i bibliotek cyfrowych.

Biorąc pod uwagę rezultaty badania, stwierdzić należy, iż analizowany kanon literacki jest obecny w międzynarodowych bibliotekach cyfrowych w stopniu umiarkowanym. Obecność danego utworu w zasobach cyfrowych agregatorów uzależniona jest od jego udostępnienia przez krajowego dostawcę. Podobną sytuację odnotowano w przypadku adaptacji danego dzieła. Niewielką obecnością cieszyły się też tłumaczenia tekstów na języki obce. Wszystko to pozwala sądzić, iż główną zaletą omawianych usług jest przede wszystkim możliwość uzyskania dostępu do ogromnej ilości rekordów, oferowanych w ramach jednego interfejsu. Wydaje się, że użytkownicy szukający bogatszej niż narodowa oferty czytelniczej, nie będą usatysfakcjonowani.

Z badania wnioskować należy, że biorące w nim udział biblioteki cyfrowe, charakteryzują się odmiennymi zasadami działania. Europeana wydaje się być typowym międzynarodowym agregatorem treści. Nie bierze ona jednak szerszego udziału w procesie digitalizacji (a więc powstawania) swoich zbiorów. Zadaniem Internet Archive jest zbudowanie takiej instytucji, która zaoferuje „uniwersalny dostęp do całej znanej ludzkości wiedzy”. W tym celu organizacja aktywnie poszukuje nowych partnerów, prowadzi prelekcje i wykłady, a także udziela mecenatu projektom cyfryzacyjnym. Z kolei Google Books postawiło sobie za zadanie skatalogowanie wszystkich książek świata. W tym celu właściciel portalu wykorzystuje szereg zaawansowanych narzędzi, które w dużej mierze automatyzują cały proces. Ponadto projekt korporacji Google skupia się wyłącznie na samych wydawnictwach, a nie jak w pozostałych przypadkach – również ich multimedialnych aranżacjach. Nawet biorąc pod uwagę powyższe wnioski wciąż trudno jednoznacznie stwierdzić, który z portali osiągnie swój cel i okaże się pełnym sukcesem. Mając jednak w pamięci liczne wskazane niedostatki omawianych bibliotek cyfrowych można wysnuć wniosek, iż nie dojdzie raczej do rewolucji w zakresie dostępu do wiedzy – przeznaczeniem analizowanych

instytucji wydaje się być raczej uzupełnianie – a nie zastępowanie roli tradycyjnych bibliotek w nowoczesnym społeczeństwie.

Bibliografia

- Rozporządzenie Ministra Edukacji Narodowej z dnia 3 lipca 2007 r. zmieniające rozporządzenie w sprawie podstawy programowej wychowania przedszkolnego oraz kształcenia ogólnego w poszczególnych typach szkół, Dz.U. 2007 nr 123 poz. 853
- Adamiec M., „Wirtualna Biblioteka Literatury Polskiej” [online] <http://literat.ug.edu.pl/books.htm> [dostęp 20.07.2016]
- Baksik C., *Fair use or exploitation? The google book search controversy*, portal: Libraries and the Academy 6.4 (2006): 399-415 [online] <https://muse.jhu.edu/article/203348/summary> [dostęp 30.06.2017]
- Chwin S., *Europa i kanon dla dorosłych*, [w:] *Europejski kanon literacki. Dylematy XXI wieku*, red. nauk. E. Wichrowska, Warszawa 2012, s. 57.
- Europeana, Biblioteka Narodowa, [online] <http://www.bn.org.pl/zasoby-cyfrowe-i-linki/europeana> [dostęp 16.02.2017]
- Howard J., *Google Begins to Scale Back Its Scanning of Books From University Libraries*, The Chronicle of Higher Education, 2012 [online] <http://www.chronicle.com/article/Google-Begins-to-Scale-Back/131109/> [dostęp 01.07.2017].
- Ganley P., *Google Book Search: Fair use, fair dealing and the case for intermediary copying*, (2006) [online] <https://papers.ssrn.com/sol3/papers.cfm?abstract-id=875384> [dostęp 30.06.2017]
- Gmerek K., *Archiwa internetowe po obu stronach Atlantyku – Internet Archive, Wayback Machine oraz UK Web Archive*, Biuletyn EBIB 1 (2012): 128 [online] http://www.ebib.pl/images/stories/numery/128/128_gmerek.pdf [dostęp 30.06.2017]
- Janousek P., *O nowym historycznoliterackim paradygmacie, historii, kanonie i badaczach literatury*, „Porównania” 2007, nr 4, [online] <https://repozytorium.amu.edu.pl/handle/10593/1098> [dostęp 20.07.2016]
- Kiss C. Gy., *Gdzie literatura tworzyła naród (Dylematy kultur środkowoeuropejskich)*, [w:] *Europejski kanon literacki. Dylematy XXI wieku*, red. nauk. E. Wichrowska, Warszawa 2012, s. 85–89.
- Kolasa W. M., *Biblioteki cyfrowe na świecie – powstanie i rozwój*, [w:] *Biblioteki cyfrowe*, red. M. Próchnicka, M. Janiak, M. Krakowska, Warszawa 2012, s. 66–95, [toż online] http://eprints.rclis.org/16141/1/kolasa_cyrowe.pdf [dostęp 29.04.2016]
- Malinowski M., *Google Books w badaniach naukowych – wcale nie tak różowo*, „Historia i Media”, [online] <http://historiaimedia.org/2010/09/02/google-books-w-badaniach-naukowych/> [dostęp 8.07.2016]
- Matusiak R., *Dostęp do pełnych tekstów książek w Google Book Search*, „Biuletyn EBIB” 2008, nr 3, [online] <http://www.ebib.info/2010/94/a.php?mateusiak>, [dostęp 8.07.2016]
- Mazurek C., Parkoła T., Werla M., *Rola bibliotek cyfrowych w zachowaniu regionalnej kultury i tradycji*, „Studia Oeconomica Posnaniensia” 1.5 (254) (2013) [online] http://cejsh.icm.edu.pl/cejsh/element/bwmeta1.element.desklight-d2aa40c0-c141-4be8-957d-b083f61cc923/c/mazurek_parkola_werla.pdf [dostęp 30.06.2017]

- Mrowcewicz K., *Inteligencja polska w świetle kanonu lektur szkolnych. Zmierzch klasyki literackiej czy zmierzch inteligencji?*, [w:] *Inteligencja w Polsce. Specjaliści, twórcy, klerkowie, klasa średnia?*, red. H. Domański, Warszawa 2008, s. 423–439
- Ronai I., *Kulturalis digitalizálás-felúton az Europeana Fele*, „Tudományos és Muszaki Tajekoztatás (Scientific and Technical Information; Hungarian journal of library and information science)” 8.56 (2009): 371-377 [online] <https://www.infona.pl/resource/bwmeta1.element.62628895-e26e-3631-84af-d01e75a52db5> [dostęp 30.06.2017]
- Szczepaniec K., *Polonica w wybranych międzynarodowych bibliotekach cyfrowych*, [online] <http://jbc.bj.uj.edu.pl/Content/180403/praca%20magisterska.pdf> [dostęp 30.06.2017]
- Wilczek P., *Kanon jako problem kultury współczesnej*, *Gazeta Uniwersytecka* [online] <http://gazeta.us.edu.pl/node/223361> [dostęp 20.07.2016]

Presence of Polish literary collections in foreign digital libraries

Abstract

The article contains an analysis of the presence of Polish literary canon in international digital libraries. The main aim of the elaboration is to examine the quantitative dissemination of literary works comprising the canon of Polish literature in main foreign content aggregators and qualitative analysis of the typed collections. The data from the analysis of contents of selected digital libraries comprised the source of the study. As a result of the research, it was established that Europeana provides 1775 units, of which 1434 in Polish and 341 in other languages, Google books provides 764 units, of which 508 in Polish and 256 in other languages, and Internet Archive provides 254 units, of which 235 in Polish and 19 in other languages. The type of analyzed resources was also taken into consideration. Europeana comprised mainly literary and iconographic works, while Internet Access – literary and audiovisual. Google Books provides only text files. The work is concluded by a short analysis of the content providers. For Europeana it was mainly the National Library and the Federation of Digital Libraries, while for Internet Archive - Canadian Libraries. Foreign-language documents come from national libraries of the respective countries.

Key words: Digital libraries, Polish literature canon, statistics

Przemysław Tadla
Institute of Information Sciences
Pedagogical University of Cracow