

© Copyright by Biblioteka Jagiellońska,
Kraków 2013

Opracowanie i redakcja:

Krystyna Sanetra

Korekta i koordynacja prac wydawniczych:

Kaja Stompór-Lesiecka

Korekta tekstów w języku angielskim:

Marek Krośniak

Konsultacja edytorska:

Piotr Myszkowski

Projekt graficzny, skład i łamanie:

Motor Studio

ISBN 978-83-934926-2-6

Publikacja bezpłatna

Nakład: 400 egzemplarzy

Realizacja projektu Jagiellońska Biblioteka Cyfrowa współfinansowana przez Unię Europejską ze środków Europejskiego Funduszu Rozwoju Regionalnego w ramach Programu Operacyjnego Infrastruktura i Środowisko na lata 2007–2013, Priorytet XI „Kultura i Dziedzictwo Kulturowe”, Działanie 11.1 „Ochrona i zachowanie dziedzictwa kulturowego o znaczeniu ponadregionalnym”.

**Uniwersytet
Jagielloński –
Biblioteka
Jagiellońska**

**Jagiellońska
Biblioteka
Cyfrowa**

www.jbc.bj.uj.edu.pl

Introduction**06.** Krystyna Sanetra**Project Management****14.** Małgorzata Mrozek-Buksa, *Jagiellonian Digital Library Project*
30. Adrian Wiecheć, *Operational Aspects of Jagiellonian Digital Library***Collection Preservation****38.** Aleksandra Szalla-Kleemann, *Program of Conservation and Preservation of Special Collections and 19th-20th Century Periodicals*
58. Krzysztof Frankowicz, *Selecting Periodicals for Scanning and Publishing*
66. Zbigniew Koziński, *Repairing and Securing Collection Items before Scanning*
72. Bartłomiej Siedlarz, *Image Acquisition Equipment*
77. Zbigniew Koziński, *Making Copies of Periodicals*
79. Tomasz Zacharski, *Mass Deacidification of Collections***Processing of Digital Objects****84.** Łukasz Mesek, *From Catalog Card through OPAC to Digital Library*
93. Anna Krzak, *File Structure and Naming System*
116. Piotr Myszkowski, *Metadata*
131. Katarzyna Szczepaniec, *Mass Evaluation of Scans*
148. Oktawia Kubas, Mariusz Staniszewski, *Mass Graphic Correction*
159. Leszek Szafranski, *Cropping and Straightening – Choice or Necessity?***Publishing of Digital Objects****176.** Bartłomiej Siedlarz, *Conversion to DjVu Format*
188. Adrian Drabik, *Automated Workflow*
212. Bartłomiej Brawuski, *Hardware Infrastructure***Archivization of Originals and Copies of Protected Items****220.** Adrian Wiecheć, *Digital Archivization*
226. Łukasz Mesek, *Microfilm as an Analogue Backup Copy***Project Promotion****234.** Kaja Stompór-Lesiecka, *Project Promotion***240.** List of Selected Abbreviations**Przedmowa****06.** Krystyna Sanetra**Zarządzanie projektem****14.** Małgorzata Mrozek-Buksa, *Projekt Jagiellońska Biblioteka Cyfrowa*
30. Adrian Wiecheć, *System Obsługi Jagiellońskiej Biblioteki Cyfrowej***Ochrona zbiorów****38.** Aleksandra Szalla-Kleemann, *Program konserwacji oraz zabezpieczenia zbiorów specjalnych i czasopism XIX- i XX-wiecznych*
58. Krzysztof Frankowicz, *Selekcja czasopism do skanowania i publikacji*
66. Zbigniew Koziński, *Naprawa i zabezpieczenie zbiorów przeznaczonych do skanowania*
72. Bartłomiej Siedlarz, *Skanery i skanowanie*
77. Zbigniew Koziński, *Kopiowanie czasopism*
79. Tomasz Zacharski, *Masowe odkwaszanie zbiorów***Opracowanie obiektów cyfrowych****84.** Łukasz Mesek, *Od karty katalogowej poprzez OPAC do biblioteki cyfrowej*
93. Anna Krzak, *Struktura i nazewnictwo plików*
116. Piotr Myszkowski, *Metadane*
131. Katarzyna Szczepaniec, *Masowa ocena skanów*
148. Oktawia Kubas, Mariusz Staniszewski, *Masowa korekta graficzna*
159. Leszek Szafranski, *Kadrowanie i prostowanie – wybór czy konieczność?***Udostępnienie obiektów cyfrowych****176.** Bartłomiej Siedlarz, *Konwersja do DjVu*
188. Adrian Drabik, *Organizacja automatyzacji pracy*
212. Bartłomiej Brawuski, *Infrastruktura sprzętowa***Archiwizacja oryginałów i kopii dzieł zabezpieczonych****220.** Adrian Wiecheć, *Archiwizacja cyfrowa*
226. Łukasz Mesek, *Analogowa kopia zabezpieczająca w postaci mikrofilmu***Promocja projektu****234.** Kaja Stompór-Lesiecka, *Promocja projektu***240.** Wykaz wybranych skrótów

Przedmowa

Przedmiotem Umowy nr POIS.11.01.00-00-057/09-00 między Ministrem Kultury i Dziedzictwa Narodowego a Uniwersytetem Jagiellońskim było udzielenie Beneficjentowi, Bibliotece Jagiellońskiej, dofinansowania na realizację projektu Jagiellońska Biblioteka Cyfrowa. Przedmiotem dofinansowanego projektu było zbudowanie Jagiellońskiej Biblioteki Cyfrowej. Umowę podpisano 15 kwietnia 2010 r., a 19 lipca 2010 r. Biblioteka Jagiellońska uruchomiła pilotażową wersję Jagiellońskiej Biblioteki Cyfrowej, pod adresem <http://jbc.bj.uj.edu.pl/dlibra>. Wtedy, w zestawieniu Federacji Bibliotek Cyfrowych pod względem liczby opublikowanych dokumentów (ponad 500 obiektów cyfrowych) Jagiellońska Biblioteka Cyfrowa zajmowała 39. miejsce, ale już po roku awansowała na 4. pozycję, oferując czytelnikom niemal 56 tys. publikacji. Próg 100 tys. publikacji został przekroczony 2 stycznia 2012 r., a Jagiellońska Biblioteka Cyfrowa usytuowała się na drugim miejscu zestawienia Federacji

Bibliotek Cyfrowych. Obecnie, w dniu 12 grudnia 2012 r. z liczbą opublikowanych obiektów cyfrowych 182 725, Jagiellońska Biblioteka Cyfrowa znalazła się na pierwszym miejscu w kraju.

Na osiągnięcie to miały wpływ m.in. dobra organizacja procesu zabezpieczania zbiorów, racjonalne wykorzystywanie przyznaných środków unijnych, a przede wszystkim zaangażowanie zespołu Jagiellońskiej Biblioteki Cyfrowej, zarówno w bieżącą pracę, jak i w rozwój narzędzi programistycznych wspomagających tę pracę.

Koncepcja tworzenia Jagiellońskiej Biblioteki Cyfrowej została określona w ten sposób, że z jednej strony wbudowana została w istniejącą strukturę organizacyjną Biblioteki Jagiellońskiej i jej zadania, a z drugiej strony – zadania te zostały zdefiniowane na nowo i rozszerzone.

Jagiellońska Biblioteka Cyfrowa nie

została zbudowana w oderwaniu, lecz w odniesieniu do istniejących baz, innych bibliotek cyfrowych i do przyjętych standardów. Do obsługi i przeszukiwania Jagiellońskiej Biblioteki Cyfrowej wybrano oprogramowanie dLibra, wykorzystywane przez większość bibliotek w kraju. Dołączono w ten sposób do rozproszonych bibliotek cyfrowych i repozytoriów widocznych w serwisie pod nazwą Federacja Bibliotek Cyfrowych i przeszukiwanych przez ten serwis. Unikalną cechą serwisu jest możliwość umieszczania i przeszukiwania planów digitalizacji, co pozwala na uniknięcie dublowania już wykonanej digitalizacji danego obiektu. Metadane dotyczące obiektów cyfrowych z Jagiellońskiej Biblioteki Cyfrowej widoczne są też w Europeanie, w czym pośredniczy Federacja Bibliotek Cyfrowych, jako polski agregator metadanych dla Europeany.

W Jagiellońskiej Bibliotece Cyfrowej do tworzenia opisów bibliograficznych dla obiektów cyfrowych wykorzystywany jest międzynarodowy standard Dublin Core. Pośrednio wykorzystywany jest również format MARC 21, gdyż opis kopii cyfrowej tworzony jest na podstawie rekordu bibliograficznego dla oryginału, sporządzanego przez profesjonalistę bibliotekarza, w zależności od rodzaju dokumentu z odpowiedniego oddziału opracowania zbiorów specjalnych lub zbiorów nowych, znajdującego przepisy sztuki katalogowania i tworzenia kartoteki hasel wzorcowych. Re-

kordy bibliograficzne dla oryginałów, zawierające adresy elektroniczne kopii cyfrowych, wprowadzane są w formacie MARC 21 do bazy NUKAT, automatycznie przejmowane do bazy Komputerowego Katalogu Zbiorów Bibliotek UJ, skąd konwertowane są do formatu Dublin Core w Jagiellońskiej Bibliotece Cyfrowej. Takie korelacje między danymi w bazach wzbogacają wyszukiwanie kopii cyfrowej, bo jest ona dostępna za pośrednictwem metadanych w Jagiellońskiej Bibliotece Cyfrowej oraz poprzez odnośniki umieszczone w polu 856 w rekordach bibliograficznych w NUKAT i w KKZBUJ, przy czym pomiędzy metadanymi dotyczącymi danego dokumentu (oryginału i kopii cyfrowej) w JBC i KKZBUJ występuje linkowanie zwrotne.

Jagiellońska Biblioteka Cyfrowa zbudowana została w modelu instytucjonalnym, tzn. tworzona jest przed jedną instytucję, Bibliotekę Jagiellońską, a tworzone i udostępniane w ramach projektu obiekty cyfrowe pochodzą z jej unikatowych i cennych zasobów, dobieranych również pod kątem ich stanu fizycznego, w powiązaniu z zapotrzebowaniem użytkowników. W tym modelu infrastruktura techniczna i posiadanie zaplecza merytoryczno-technicznego leży po stronie instytucji tworzącej bibliotekę cyfrową. Jeżeli chodzi o kadry, Biblioteka Jagiellońska takie zaplecze posiada, ale zadania projektu zostały nakreślone na tyle szeroko, że obok rozszerzenia obowiązków pracowników Bi-

blioteki Jagiellońskiej, konieczne było zatrudnienie dodatkowych osób, które weszły w skład odpowiednich, szczególnie zaangażowanych w realizację projektu oddziałów.

Natomiast Biblioteka Jagiellońska odczuwała mocno brak odpowiedniego zaplecza technologicznego, niezbędnego do konserwacji i bezpiecznej, długofalowej archiwizacji zasobów bibliotecznych, podlegających stopniowej degradacji, brak infrastruktury sprzętowej niezbędnej do digitalizacji, udostępniania i archiwizacji obiektów cyfrowych. Utworzenie więc platformy cyfrowej dla Jagiellońskiej Biblioteki Cyfrowej oraz udostępnienie online i zarchiwizowanie obszernego zasobu zdigitalizowanych zbiorów Biblioteki Jagiellońskiej nie byłoby możliwe bez środków finansowych z projektu.

Niniejsza publikacja zogniskowana jest wokół głównych działań projektu, wokół procesu tworzenia Jagiellońskiej Biblioteki Cyfrowej i składa się z sześciu części.

Pierwsza część dotyczy zarządzania projektem. W artykule autorstwa Małgorzaty Mrożek-Buksy, kierownika Oddziału ds. Projektów, przedstawione zostały główne zadania i realizacja projektu, sposób jego zarządzania, najważniejsze inwestycje i koszty. Od samego początku kreowania wizji Jagiellońskiej Biblioteki Cyfrowej, jeszcze na etapie aplikowania o środki

zdawano sobie sprawę, że ogrom zadań i czynności niezbędnych przy tworzeniu biblioteki cyfrowej, w szczególności poszczególne etapy procesu tworzenia obiektów cyfrowych wymagają obsługi, wspomaganie przez system informatyczny. W Oddziale Komputeryzacji Biblioteki Jagiellońskiej, który odpowiedzialny jest m.in. za obsługę informatyczną Komputerowego Katalogu Zbiorów Bibliotek UJ, prawidłową pracę sprzętu komputerowego i sieci komputerowej BJ, ale także za tworzenie potrzebnego oprogramowania, System Obsługi Jagiellońskiej Biblioteki Cyfrowej rzeczywiście powstał. Powstał na potrzeby zbudowania i obsługi platformy Jagiellońskiej Biblioteki Cyfrowej, wraz z nią się rozwijał oraz był na bieżąco uzupełniany i wzbogacany. Jego zadania i funkcje zarządzające opisuje jego główny twórca, Adrian Wiecheć.

Bez środków z projektu Biblioteka Jagiellońska nie zdołałaby zmodernizować w sposób zasadniczy swojej Pracowni Foto, ani przeprowadzić konserwacji cennych dokumentów. Poczynając od opracowania ankiety konserwatorskiej, oceny stanu zachowania zbiorów specjalnych oraz czasopism, stworzenie programu wykonania konserwacji profilaktyczno-zachowawczej i prowadzenie nadzoru konserwatorskiego przez Sekcję Konserwacji, poprzez nadzorowanie i wykonanie napraw w Sekcji Opraw i Zabezpieczenia Zbiorów, odkwaszenie wytypowanych zbiorów w Sekcji Kliniki Papieru, aż po wykonanie skanów

w Sekcji Reprografii, zadania związane z ochroną zbiorów zrealizowane zostały w Oddziale Zabezpieczania Zbiorów, pod kierownictwem specjalistów z tego Oddziału. Prace te zostały opisane w części drugiej, dotyczącej ochrony zbiorów, przez Zbigniewa Kozińskiego, kierownika Oddziału Zabezpieczania Zbiorów oraz Aleksandrę Szalla-Klemann, kierownika Sekcji Konserwacji i Tomasza Zacharskiego, kierownika Sekcji Kliniki Papieru. Temat ochrony zbiorów dopełnia artykuł Bartłomieja Siedlarza, dotyczący modeli zakupionych skanerów i procesu skanowania oraz artykuł Krzysztofa Frankowicza, kierownika Oddziału Magazynów, przedstawiający cenny zbiór unikatowych czasopism z zasobu Biblioteki Jagiellońskiej, jego historię i stan zachowania, zasady typowania tytułów do skanowania oraz problemy i dylematy z tym związane.

Opracowanie i udostępnienie obiektów cyfrowych to dwa ważne etapy tworzenia biblioteki cyfrowej, oba weszły w skład nowych i stałych zadań Oddziału Dokumentów Audiowizualnych i są wykonywane przez dwóch pracowników Oddziału wraz sześcioma nowozatrudnionymi. Zespół ten zajmował się obróbką plików wynikowych z bieżącej digitalizacji czasopism, jak również plików z digitalizacji zbiorów specjalnych sprzed kilku lat, przechowywanych na dyskach optycznych. Podstawowe rozwiązania związane z ujednocnionym systemem zarządzania plikami przed-

stawiła Anna Krzak, proces masowej oceny skanów zanalizowała Katarzyna Szczepaniec, a potencjalne problemy związane z obróbką graficzną plików omówili Oktawia Kubas i Mariusz Staniszewski, zaś techniki kadrowania i prostowania skanów ze względu na rodzaj edytowanego dokumentu przedstawił Leszek Szafranski. Dwa artykuły w tej części poświęcone są metadansom, Łukasz Mesek, kierownik Oddziału Dokumentów Audiowizualnych, skupił się w szczególności na korzyści, jaką dla cyfrowej „reedycji” czasopism stanowi ich drobiazgowy opis, a Piotr Myszkowski zaprezentował możliwości oprogramowania dLibra w zakresie opisu pojedynczych publikacji oraz grupowania zbiorów w kolekcje. W części poświęconej udostępnianiu obiektów cyfrowych Bartłomiej Siedlarz omówił przygotowywanie wersji prezentacyjnej obiektów cyfrowych, czyli sposoby konfiguracji oprogramowania Document Express, używanego do konwersji obrazów do formatu DjVu, a Adrian Drabik poruszył kwestie dotyczące konieczności automatyzacji powtarzalnych zadań i dostosowywania narzędzi informatycznych do specyficznych potrzeb biblioteki cyfrowej.

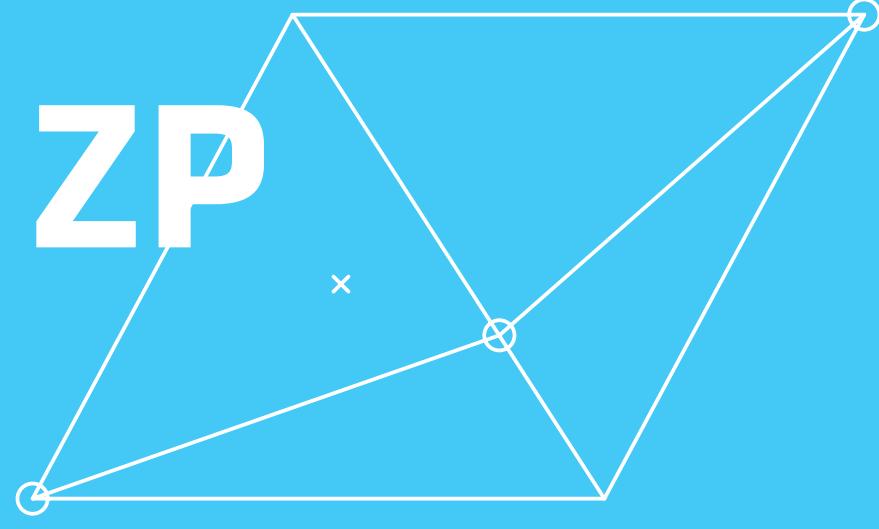
Do udostępnienia i archiwizacji plików niezbędna jest odpowiednia infrastruktura sprzętowa i tę pozyskano w ramach środków z projektu, a obsługa informatyczna utworzonej platformy Jagiellońskiej Biblioteki Cyfrowej weszła na stałe do zadań informatyków z Od-

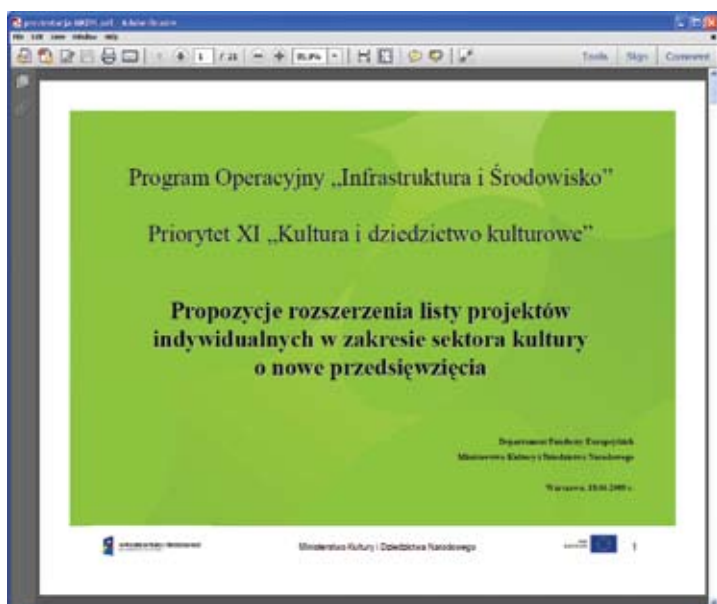
działu Komputeryzacji. Rozmieszczenie komponentów systemu dLibra w Jagiellońskiej Bibliotece Cyfrowej na poszczególne serwery przedstawił Bartłomiej Brawuski.

W części poświęconej archiwizacji kopii dzieł zabezpieczonych Adrian Wiecheć przedstawił sposób archiwizacji plików w formacie TIFF, zaś Łukasz Mesek opisał proces powstawania mikrofilmu z uzyskanej w wyniku digitalizacji kopii cyfrowej. W ostatniej części publikacji Kaja Stompór-Lesiecka przedstawiła promocję projektu.

Mam nadzieję, że szczegółowe, wieloaspektowe omówienie prac związanych z budową Jagiellońskiej Biblioteki Cyfrowej, przedstawienie nabytych doświadczeń i przyjętych rozwiązań jest dobrym podsumowaniem realizowanego projektu i może być interesujące oraz pomocne dla zainteresowanych bibliotekami cyfrowymi. Jednocześnie pragnę podziękować autorom, a zarazem wykonawcom poszczególnych zadań oraz wszystkim współpracownikom. Wiedza i zaangażowanie pracowników Biblioteki Jagiellońskiej umożliwiły realizację głównego celu projektu – zbudowanie Jagiellońskiej Biblioteki Cyfrowej.

Zarządzanie projektem





II. 2a.

II. 2b.



II. 2a.-2b. Prezentacja na stronie MKiDN - informacje dot. projektów indywidualnych.

http://www.funduszeuropejskie.gov.pl/Wiadomosci/Documents/MKiDN_190609_4.pdf
(dostęp: 20.12.2012 r.).

W związku z uznaniem projektu za kluczowy w obrębie działania 11.1 w Bibliotece Jagiellońskiej rozpoczęto opracowywanie dokumentacji podlegającej, wg procedury, ocenie formalnej oraz merytorycznej I i II stopnia, zawierającej dane zaktualizowane na dzień jej składania. Z powodu przedłużającego się terminu wykonania aktualizacji analiz do studium wykonalności, kompletną dokumentację projektu złożono 30 października 2009 r.

W ramach konsultacji społecznych dotyczących listy projektów indywidualnych poparcie dla projektu Jagiellońska Biblioteka Cyfrowa złożyli: Jerzy Miller, wojewoda Małopolski, Jerzy Federowicz, wiceprzewodniczący Sejmowej Komisji Kultury, prof. Jacek Majchrowski, prezydent miasta Krakowa, dr Tomasz Makowski, dyrektor Biblioteki Narodowej.

W następnych miesiącach w trakcie trwania oceny merytorycznej I i II stopnia dokumentacja była dopracowywana

	Finansowanie	Wartość zakładana ogółem	SUMA poniesionych i zaplanowanych wydatków	Oszczędności / Przekroczenia (3-4)
I	Razem wydatki w ramach projektu	6 042 547,81	6 042 47,81	0,00
II	Razem wydatki kwalifikowane	0,00	0,00	0,00
1.	płatności budżet jst	5 136 165,64	5 136 65,64	0,00
2.		0,00	0,00	0,00
3.	Inne krajowe środki publiczne (budżet UJ; dotacja ze środków Ministerstwa Nauki i Szkolnictwa Wyższego)	906 382,17	906 382,17	0,00
4.	prywatne	0,00	0,00	0,00
5.	pożyczki EBI	0,00	0,00	0,00
III	Razem wydatki niekwalifikowane	0,00	0,00	0,00
1.	budżet jst	0,00	0,00	0,00
2.	inne krajowe środki publiczne	0,00	0,00	0,00
3.	prywatne	0,00	0,00	0,00
4.	pożyczki EBI	0,00	0,00	0,00

II. 3. Plan finansowania projektu.

i uzupełniania. Ocena projektu w oparciu o kryteria merytoryczne zakończyła się pozytywnie 5 stycznia 2010 r. W ramach tej oceny projekt otrzymał 43 punkty na 65 możliwych do uzyskania (65%). Z datą 6 stycznia 2010 r. wpłynęła do Biblioteki Jagiellońskiej Decyzja nr 1/2010 ministra B. Zdrojewskiego o przyznaniu dofinansowania z Europejskiego Funduszu Rozwoju Regionalnego dla projektu Jagiellońska Biblioteka Cyfrowa w wysokości 5 136 165, 64 zł.

Mając tak korzystną decyzję, w Bibliotece Jagiellońskiej rozpoczęto przygotowania do zawarcia Umowy o dofinansowanie projektu. Polegały one przede wszystkim na zebraniu wymaganych załączników, stworzeniu harmonogramu projektu, planu wystąpień o środki dotacji celowej i płatności oraz planu finansowania projektu (il. 3).

Kluczowe znaczenie miało wskazanie ostatecznego zakresu przedsięwzięcia, który został ujęty w sześciu głównych działaniach:

Działanie I: Ochrona zbiorów

Konserwacja zachowawcza: działania wstępne – odkurzenie i oczyszczenie obiektów, nadzór konserwatorski nad pracami związanymi z naprawami zniszczonych egzemplarzy, opracowanie ankiet konserwatorskich stanu zachowania obiektów i sporządzenie na ich podstawie dokumentacji konserwatorskiej obiektów objętych projektem, sformuło-

wanie zaleceń konserwatorskich dla prac zabezpieczających fizyczne obiekty, naprawy konserwatorsko-introligatorskie zbiorów, nadzór prac związanych z wykonaniem kopii cyfrowych zbiorów.

Wykonanie kopii cyfrowych i mikrofilmowych: przygotowanie przeznaczonych do skanowania zbiorów XIX-XX-wiecznych, weryfikacja kompletności zasobu czasopism, obsługa urzędzeń skanujących, kontrola jakości wykonanych skanów, obsługa urzędzeń wykonujących mikrofilmy, kontrola jakości wykonanych mikrofilmów.

Działanie II: Opracowanie obiektów cyfrowych do udostępniania

Utworzenie metadanych digitalizowanych obiektów: opracowanie bibliograficzne wydawnictw zwartych, wydawnictw ciągłych, jednostek zbiorów specjalnych oraz sporządzenie rekordów bibliograficznych wraz z hasłami formalnymi i przedmiotowymi w Komputerowym Katalogu Zbiorów Bibliotek UJ.

Przygotowanie obiektów cyfrowych do umieszczenia na platformie dLibra: segregacja i korekta graficzna obrazów cyfrowych, kontrola poprawności i kompletności, konwersja z formatu TIFF do formatu DjVu, wykonanie OCR.

Numer działania	Roboczo-godzina	Środki trwałe	Materiały	Usługi	Razem
	Szacowana wartość brutto	Szacowana wartość brutto	Szacowana wartość brutto	Szacowana wartość brutto	
0	11 000,00	–	–	36 600,00	47 600,00
I	821 375,00	2 291 836,14	411 357,16	141 000,00	3 665 568,30
II	426 568,00	71 447,48	–	–	498 015,48
III	119 851,00	141 457,44	–	12 688,00	273 996,44
IV	–	412 733,32	86 010,00	–	498 743,32
V	328 464,00	89 384,20	31 537,00	–	449 385,20
VI	–	41 858,20	28 670,00	143 404,00	213 932,20
Razem	1 707 258,00	3 048 716,78	557 574,16	333 692,00	5647240,94
Rezerwa inwestycyjna	–	–	–	–	395 306,87
Razem	–	–	–	–	6 042 547,81

Działanie III: Udostępnienie obiektów cyfrowych w JBC

Stworzenie i wdrożenie modułów oprogramowania do zintegrowanego udostępniania produktów projektu, konfiguracja, kontrola funkcjonalności i utrzymanie w ruchu infrastruktury IT SOJBC, obsługa informatyczna udostępniania obiektów cyfrowych w JBC, tworzenie i zarządzanie kolekcjami publikacji cyfrowych, konfiguracja i obsługa baz danych JBC.

Działanie IV: Archiwizacja oryginałów i kopii dzieł zabezpieczonych

Obsługa archiwizacji oryginałów, mikrofilmów i plików cyfrowych, zabezpieczenie fizyczne zbiorów zdigitalizowanych, zabezpieczenie wykonanych plików cyfrowych i mikrofilmów.

Działanie V: Zarządzanie projektem

Koordinacja realizacji zadań projektu, koordynacja rozliczania realizacji rzeczowo-finansowej, przygotowywanie wniosków o płatność, obsługa biurowa i księgową projektu, przygotowywanie przetargów, obsługa ewaluacji projektu.

Działanie VI: Promocja projektu

Działalność promocyjna związana z wizualizacją nabytych środków trwałych wraz z miejscem realizacji projektu i podaniem informacji o współfinansowaniu inwestycji ze środków EFRR. Obsługa

zadań promocyjnych projektu.

Zamieszczona powyżej tabela przedstawia szczegółowe rozłożenie kosztów na poszczególne działania. Wykazano w niej również koszty związane z przygotowaniem projektu (0), które zostały potraktowane jako oddzielne zadanie, niewliczone w kosztorys głównych działań.

Umowa o dofinansowanie projektu Jagiellońska Biblioteka Cyfrowa została zawarta 15.04.2010 r. pomiędzy Ministrem Kultury i Dziedzictwa Narodowego Bogdanem Zdrojewskim, a Uniwersytetem Jagiellońskim, reprezentowanym przez Marię Hulicką, kvestora UJ oraz prof. Adama Juszkiewicza, pełnomocnika rektora ds. Funduszy Strukturalnych (il. 4). Przy podpisaniu umowy obecni byli również przedstawiciele Biblioteki Jagiellońskiej: Krystyna Sanetra, wicedyrektor BJ ds. zbiorów XIX-XXI w. oraz Katarzyna Szklanny, koordynator projektu.

Po podpisaniu umowy, rozpoczęto przygotowania do powołania zespołu, którego celem miała być realizacja zadań wyszczególnionych w założeniach projektu. W trakcie pierwszych rekrutacji wyłoniono dwie osoby odpowiedzialne za obsługę administracyjną projektu, w tym za działania związane z promocją projektu. W lipcu 2010 r. zatrudniono osobę na stanowisko technika-konserwatora książki oraz osobę na stanowisko młodszego bibliotekarza, którego zadaniem było przygotowywanie XIX- i XX-



Il. 4.

wiecznych czasopism z magazynu ogólnego BJ, przeznaczonych do skanowania w ramach projektu. Na przełomie września i października 2010 r. zatrudniono osoby, których zadaniem miało być skanowanie, mikrofilmowanie, korekta graficzna skanów, publikowanie obiektów cyfrowych w Jagiellońskiej Bibliotece Cyfrowej. Łącznie dla potrzeb projektu zatrudniono 20 osób (w ramach 14 etatów):

- 2 osoby – w Zespole ds. Projektów BJ (od stycznia 2011 r. Oddział ds. Projektów),

- 1 osoba – w Oddziale Magazynów,

- 6 osób – w Oddziale Dokumentów Audiowizualnych,

- 11 osób – w Oddziale Zabezpieczenia Zbiorów.

Oprócz 20 osób, które swój czas pracy w 100% poświęcały na realizację zadań projektu, w prace byli zaangażowani również pracownicy Biblioteki Jagiellońskiej, których zakresy obowiązków zostały poszerzone o zadania wynikające z założeń projektu. Prace podzielono na etapy, o różnych terminach realizacji, wynikających z harmonogramu projektu. Łącznie w prace nad projektem zostało zaangażowanych 55 osób.

Ogółem w realizacji projektu wzięły udział następujące oddziały Biblioteki Jagiellońskiej:

- Oddział ds. Projektów, zajmujący się zarządzaniem i koordynacją zadań projektu, w szczególności Działaniem V, tj. wdrożeniem projektu i Działaniem VI, tj. promocją, a więc zarządzaniem, obsługą biurową, obsługą przetargów, ewaluacją oraz promocji projektu (3 osoby);

- Oddział Zabezpieczenia Zbiorów, mający w kompetencjach Działanie I, tj. ocenę stanu zachowania zbiorów, opiekę konserwatorską nad ochroną zbiorów oraz nadzór i koordynację napraw konserwatorskich, wykonywanie kopii cyfrowych i mikrofilmowych (3 osoby);

- Oddział Magazynów realizujący Działanie I, w szczególności przygotowywanie zbiorów do oceny konserwatorskiej i do skanowania, a także Działanie IV – zadanie archiwizacji oryginałów (1 osoba);

- Oddziały Opracowania Zbiorów Zwartych i Ciągłych, Oddziały Zbiorów Specjalnych (rękopisów, starych druków, zbiorów graficznych i kartograficznych, zbiorów muzycznych) realizujące Działanie II, zadanie stworzenia metadanych obiektów przeznaczonych do zdigitalizowania w KKZBUJ (12 osób);

- Oddział Dokumentów Audiowizualnych, realizujący Działanie II, koordynowanie opracowania graficznego kopii cyfrowych i przygotowanie obiektów cyfrowych do umieszczenia na platformie dLibra oraz Działanie IV, w zakre-

sie archiwizacji kopii mikrofilmowych (4 osoby);

- Oddział Komputeryzacji, odpowiedzialny za realizację Działania III i IV, tj. wdrożenie, konfigurację i administrowanie infrastrukturą serwerową (serwery, macierz, biblioteka taśmowa) i sieciową, wdrożenie i rozwój SOJBC, konfigurację i monitorowanie udostępniania obiektów cyfrowych w JBC oraz archiwizację kopii cyfrowych (5 osób).

Oprócz oddziałów BJ w realizacji projektu uczestniczyły:

- Dział Spraw Osobowych UJ (2 osoby);

- Biuro Zamówień Publicznych UJ (1 osoba);

- Dział Rozliczeń Projektów Unijnych UJ (1 osoba).

Kierownikiem projektu Jagiellońska Biblioteka Cyfrowa został dyrektor Biblioteki Jagiellońskiej, prof. dr hab. Zdzisław Pietrzyk, który sprawował nadzór nad poprawnością merytoryczno-finansową realizacji projektu. Rolę wdrożeniowców w projekcie pełnili mgr Krystyna Sanetra, wicedyrektor Biblioteki Jagiellońskiej ds. druków XIX-XXI w., która sprawowała nadzór merytoryczny nad poprawnością realizacji działań projektu oraz mgr Zbigniew Koziński, zajmujący się koordynacją Działania I: Ochrony

Kierownik projektu Dyrektor Naczelny Biblioteki Jagiellońskiej

Nadzór nad poprawnością merytoryczno-finansową realizacji projektu

Oddział ds. Projektów

Zarządzanie realizacją zadań projektu, w szczególności **Działania V:** Wdrożenie projektu i **Działania VI:** Promocja
Zarządzanie obsługą projektu w zakresie:

- obsługi księgowej
- obsługi kadrowej
- obsługi biurowej
- obsługi przetargów
- ewaluacji projektu
- promocji projektu

Nadzór konserwatorski

Kierownik Sekcji Konserwacji w Oddziale Zabezpieczenia Zbiorów BJ, Dyplomowany konserwator papieru i skóry
Zadanie **Działania I:** Ochrona zbiorów w zakresie nadzoru konserwatorskiego nad ochroną zbiorów i nadzoru oceny stanu zachowania zbiorów.

Wdrożeniowiec

Kierownik Oddziału Zabezpieczenia Zbiorów
Koordynacja **Działania I:** Ochrona zbiorów w zakresie realizacji zadań:

- konserwacji zachowawczej: zabezpieczenia fizycznego zbiorów,
- wykonania kopii cyfrowych i mikrofilmów

Oddział Zabezpieczenia Zbiorów

Zadania **Działania I:** Ochrona zbiorów:

- konserwacja zachowawcza,
- wykonanie kopii cyfrowych i mikrofilmów.

Wdrożeniowiec

Wicedyrektor BJ ds. zbiorów XIX-XXI w.
Nadzór merytoryczny nad poprawnością realizacji działań projektu

Oddział Magazynów, oddziały opracowania zbiorów nowych, oddziały zbiorów specjalnych

Zadanie **Działania II:** Opracowanie obiektów cyfrowych do udostępnienia, tj. Stworzenie metadanych obiektów zdigitalizowanych.
Zadania **Działania IV:** Archiwizacja oryginałów i kopii dzieł zabezpieczonych.

Oddział Komputeryzacji

Zadania **Działania III:** Udostępnienie obiektów cyfrowych w JBC i zadań **Działania IV:** Archiwizacja oryginałów i kopii dzieł zabezpieczonych.

Oddział Dokumentacji Audiowizualnych

Zadania **Działania II:** Opracowanie obiektów cyfrowych do udostępniania w zakresie realizacji zadania: Przygotowanie obiektów cyfrowych do umieszczenia w dLibrze oraz **Działania IV:** Archiwizacja oryginałów i kopii dzieł zabezpieczonych.

👁 II. 5.

👁 II. 5. Struktura zarządzania projektem Jagiellońska Biblioteka Cyfrowa.

na zbiorów w zakresie realizacji zadań konserwacji zachowawczej (zabezpieczenia fizycznego zbiorów), a także wykonania kopii cyfrowych i mikrofilmów.

W zarządzaniu i realizacji poszczególnych zadań w obrębie głównych działań projektu została wykorzystana istniejąca struktura organizacyjna Biblioteki Jagiellońskiej, a prace koordynował Zespół Zadaniowy, złożony z kierowników i specjalistów ze wspomnianych wyżej oddziałów oraz wicedyrektora ds. zbiorów XIX-XXI w. Był to ten sam Zespół, który wypracował założenia projektu i czynił starania o pozyskanie funduszy na jego realizację.

Pracownicy Biblioteki, zajmujący odpowiednie stanowiska i posiadający odpowiednie wykształcenie i doświadczenie zawodowe współpracowali w realizacji głównego celu projektu, jakim była „Ochrona i zachowanie zabytkowych zbiorów Biblioteki Jagiellońskiej oraz rozwój bibliotecznych zasobów cyfrowych dla utrwalenia dorobku kulturowego polskiego piśmiennictwa stanowiącego element Narodowego Zasobu Bibliotecznego”.

Do głównych efektów wynikających z realizacji szczegółowych celów inwestycji należy zaliczyć:

- konserwację zachowawczą i archiwizację unikatowych zbiorów bibliotecznych,
- zabezpieczenie zabytkowych zbiorów

bibliotecznych przed utratą i zniszczeniem,

- digitalizację unikatowych w skali kraju zbiorów bibliotecznych, zagrożonych zniszczeniem, dla trwałego zachowania i udostępnienia ich treści,

- utworzenie Jagiellońskiej Biblioteki Cyfrowej, jako podstawowej platformy udostępniania,

- udostępnienie do celów naukowych niedostępnych wcześniej zbiorów.

Głównym produktem projektu, osiągniętym dzięki zaangażowaniu środków z Europejskiego Funduszu Rozwoju Regionalnego, jest konserwacja zachowawcza oraz zabezpieczenie 2 866 woluminów i jednostek zabytkowych zbiorów bibliotecznych poprzez utrwalenie ich treści w formie cyfrowej oraz na kliszach mikrofilmowych. Bezpośrednim rezultatem osiągniętym dzięki realizacji projektu jest przede wszystkim zapewnienie dostępu do nowych zasobów cyfrowych w dziedzinie nauki i kultury. Zbudowana infrastruktura Jagiellońskiej Biblioteki Cyfrowej umożliwia zwiększony poziom dystrybucji zabytkowych zbiorów Biblioteki Jagiellońskiej, szeroką ich popularyzację i efektywne wykorzystanie w pracy naukowej, dydaktycznej, dla celów hobbystycznych i artystycznych. Warto podkreślić, iż zakładano, że po zakończeniu realizacji projektu w roku 2015 liczba wejść do bazy Jagiellońskiej Biblioteki Cyfrowej wyniesie 504 416,

tymczasem już w 2012 r. przekroczyła ona 2 180 360 i wciąż rośnie.

Dla realizacji zadań projektu zakupiono specjalistyczny sprzęt, oprogramowanie i materiały:

- skanery: 1 221 000,00 zł (I kwartał 2010 r.),

- sprzęt komputerowy: 513 390,46 zł (I kwartał 2010 r.),

- linię technologiczną do archiwizacji mikrofilmowej: 1 114 921,00 zł (I kwartał 2010 r.),

- regały przesuwnic i stacjonarne: 122 754,00 zł (III kwartał 2012 r.),

- materiały eksploatacyjne do odkwaszania: 150 000,00 zł (IV kwartał 2011 r.),

- materiały do konserwacji zachowawczej: 103 933,00 zł (IV kwartał 2011 r.),

- meble biurowe: 53 905,00 zł (I kwartał 2010 r.),

- sprzęt konserwatorski (luminometr, miarka, odkurzacz do pyłów niebezpiecznych): 23 252,22 zł (IV kwartał 2010/I kwartał 2011),

- specjalistyczne oprogramowanie do wykonywania OCR: 17 589,00 zł (III kwartał 2012 r.),

- infomat, dla celów promocji projektu: 18 327,00 zł (I kwartał 2012 r.).

Łącznie, kwota poniesiona na zakup środków trwałych sięgnęła 99% ogólnej kwoty zaplanowanej na ten cel w projekcie. Różnica wyniknęła z drobnego przeszacowania planowanych kosztów.

W ramach prac zaplanowanych w projekcie wyłoniono wykonawcę usługi z zakresu konserwacji zachowawczej zbiorów. Koszt wykonania usługi wyniósł 143 830,00 zł. Prace były prowadzone przez zespół doświadczonych konserwatorów papieru, którzy specjalizują się w zabezpieczaniu całych kolekcji historycznych zbiorów bibliotecznych. Koszty zakupów w znacznej części zostały pokryte ze środków otrzymanej zaliczki, pozostałe wydatki zostały przedstawione do refundacji. Wszelkie prace konserwatorskie prowadzone w projekcie były objęte nadzorem konserwatorskim, prowadzonym przez kierownika Sekcji Konserwacji Oddziału Zabezpieczania Zbiorów. Koszt nadzoru konserwatorskiego wyniósł 95% środków zaplanowanych na ten cel w projekcie.

W kategorii kosztów pn. „przygotowanie projektu”, w ramach której przygotowano studium wykonalności, opracowano analizę finansową oraz dokumentację techniczną, a także przygotowywano postępowania przetargowe zanotowano niewielkie, sięgające 1% przekroczenie kosztów w stosunku do ogólnych wydatków zaplanowanych na ten cel w projekcie.

Inaczej przedstawia się sytuacja w odniesieniu do kategorii kosztów „zarządzanie projektem”. Tu wykorzystano zaledwie 60% kosztów zaplanowanych w projekcie. Nie zrealizowano zakupu systemu wspomagającego zarządzanie i monitorowanie projektu. Podczas realizacji projektu wspomagano się dostępnym na Uniwersytecie Jagiellońskim systemem SAP. Na etapie planowania wydatków przeliczono wysokość kosztów osobowych w tej kategorii. Oszczędność pojawiła się również przy zakupie materiałów biurowych oraz kosztów związanych z podróżami służbowymi.

Podczas realizacji projektu pojawiały się problemy, które starano się na bieżąco rozwiązywać. Największą trudnością w realizacji finansowej projektu okazały się „wynagrodzenia dodatkowe” pracowników Biblioteki Jagiellońskiej i jednostek administracji centralnej UJ, realizujących zadania projektu. Trudności wynikały z niejednoznacznej interpretacji Wytocznych w zakresie kwalifikowania wydatków w ramach POIiS, co do sposobu rozliczenia wydatków osobowych. Wydatki, których kwalifikowalność w danym okresie rozliczeniowym została uznana za wątpliwą były tymczasowo wyłączane z wniosków o płatność i przedstawiane do refundacji w późniejszym terminie, zgodnie z Wytocznymi w zakresie sposobu korygowania wydatków nieprawidłowo poniesionych w ramach Programu Operacyjnego Infrastruktura i Środowisko 2007-2013.

Tworzenie Jagiellońskiej Biblioteki Cyfrowej nie kończy się wraz z zakończeniem projektu o tej nazwie. Równoległe z projektem współfinansowanym przez POIiS Biblioteka Jagiellońska realizowała i realizuje nadal projekt „Bezpieczne i kompleksowe udostępnienie zasobów cyfrowych UJ w sieci Internet. Etap I”, współfinansowany w ramach Małopolskiego Regionalnego Programu Operacyjnego. Projekt ten został zaplanowany jako kontynuacja i uzupełnienie projektu Jagiellońska Biblioteka Cyfrowa. Biblioteka Jagiellońska będzie sukcesywnie uzupełniała zbiory Jagiellońskiej Biblioteki Cyfrowej korzystając z zaplecza sprzętowo-informatycznego, pozyskanego dzięki realizacji projektów finansowanych ze środków Unii Europejskiej, a w dłuższej perspektywie jest zainteresowana dalszym pozyskiwaniem środków na jej utrzymanie i rozbudowę, zarówno ze środków krajowych jak funduszy europejskich.

System Obsługi Jagiellońskiej Biblioteki Cyfrowej



Słowa kluczowe: Jagiellońska Biblioteka Cyfrowa, System Obsługi Jagiellońskiej Biblioteki Cyfrowej, digitalizacja, system informatyczny, aplikacja WWW, organizacja pracy.

Keywords: Jagiellonian Digital Library, Jagiellonian Digital Library Workflow System, digitalization, IT system, web application, workflow.

Setki tytułów, tysiące woluminów, miliony plików i... kilkadziesiąt osób nad tym wszystkim pracujących. Rzecz ludzkimi siłami nie do ogarnięcia. Od samego początku powstawania Jagiellońskiej Biblioteki Cyfrowej było zatem wiadomo, że do jej pracy będzie konieczne stworzenie systemu informatycznego. Systemu, dzięki któremu w każdej chwili będzie wiadomo, na jakim etapie przetwarzania jest dany tytuł. Systemu, który podpowie pracownikowi, którym obiektem ma się w danym momencie zająć. Systemu wreszcie, który zapamięta, kto jaką pracę wykonał i kiedy ją zakończył. Jednym słowem: Jagiellońska Biblioteka Cyfro-

wa potrzebowała systemu obsługi.

Taki system, pod nazwą System Obsługi Jagiellońskiej Biblioteki Cyfrowej, istotnie powstał. Zrealizowany został jako aplikacja webowa, do korzystania z której potrzebna jest jedynie przeglądarka WWW. Dzięki temu nie ma problemów z instalacją (różne wersje systemu operacyjnego i bibliotek systemowych) oraz aktualizacją – wystarczy wprowadzić zmiany po stronie serwera i wszyscy użytkownicy mogą korzystać z najnowszej wersji systemu. Praca w SOJBC zaczyna się zatem od wejścia przeglądarką na adres SOJBC i zalogowania na imienne konto. Następnie przechodzi się do odpowiedniego panelu, specyficznego dla rodzaju pracy wykonywanej w ramach SOJBC. Prześledźmy je teraz kolejno.

1. Typowanie

Początek drogi to wprowadzenie do SOJBC tytułów wybranych do digitalizacji. Podaje się tylko podstawowe informa-

cje bibliograficzne (sygnatury oryginału, nazwa autora, tytuł, rok i miejsce wydania, itd.), wystarczające do identyfikacji. Pełne opracowanie bibliograficzne wykonane zostanie w następnym kroku. Publikacja zostaje też zakwalifikowana do odpowiedniego rodzaju, co daje później możliwość wyszukiwania, np. wszystkich grafik czy rękopisów. Ponieważ SOJBC ma obsługiwać wszystkie projekty dotyczące digitalizacji, prowadzone przez Bibliotekę Jagiellońską, na tym etapie przypisuje się też daną publikację do konkretnego projektu.

Obiekty w SOJBC, które w tym kroku powstają, nazywano ogólnie publikacjami. Przyjęto zasadę, że jedną publikację tworzy się dla jednej sygnatury bibliotecznej, co ułatwia współpracę z katalogiem bibliotecznym.

2. Opracowanie

Następnym koniecznym krokiem jest opracowanie publikacji, zgodnie z założeniem, że opisy wszystkich digitalizowanych zbiorów wprowadza się także do Komputerowego Katalogu Zbiorów Bibliotek UJ. Biblioteka Jagiellońska zyskuje w ten sposób większe tempo retrokonwersji, a Jagiellońska Biblioteka Cyfrowa zyskuje metadane opracowane przez specjalistów. Opracowanie bibliograficzne wykonywane jest odrębnie, w Komputerowym Katalogu Zbiorów Bibliotek UJ (obsługiwanym przez oprogramowanie VTLIS/Virtua), przez pracowników oddziałów opracowania zbiorów

nowych oraz pracowników zbiorów specjalnych. Krok ten kończy się jednak wprowadzeniem do SOJBC numeru systemowego z bazy KKZBUJ. Będzie on niezbędny do przeprowadzenia importu metadanych przy publikowaniu obiektów cyfrowych w JBC (patrz pkt 8. Publikacja).

Abstrakt: Artykuł prezentuje System Obsługi Jagiellońskiej Biblioteki Cyfrowej, narzędzie informatyczne służące do organizacji zadań projektu. Omówiona jest zasada jego działania oraz szczegółowe funkcjonalności.

Abstract: The paper presents Jagiellonian Digital Library Workflow System, the IT tool supporting the project's workflow. The basis of its function is described, and the specific functionalities are detailed.



3. Woluminy

Obiekty „publikacji” są tylko częściowo użyteczne w procesie digitalizacji, ponieważ w trakcie konserwacji i skanowania jest istotna pojedyncza fizyczna jednostka. Dla książek chodzi wprawdzie po prostu o pojedynczy egzemplarz i lista „publikacji” byłaby wystarczająca (choć, już w przypadku wydania wielotomowego sytuacja się komplikuje). Jednak w przypadku czasopism (które stanowią większość opracowywanych w ramach projektu zbiorów) rzecz dotyczy oprawnych roczników (półroczników, kwartalników) i dlatego potrzebne są informacje o poszczególnych woluminach. Odwrotna sytuacja zachodzi w przypadku dzieł współprawnych (np. rękopiśmiennych „klocków”), kiedy kilka różnych publikacji (o różnych sygnaturach) zawiera się w jednym fizycznym woluminie - wtedy taką informację trzeba w systemie wprowadzić. Jeszcze bardziej skomplikowana sytuacja występuje w przypadku ikonografii (pojedyncze arkusze bądź teki) oraz kartografii.

Krok „woluminowania” polega więc na wprowadzeniu do SOJBC informacji o podziale publikacji na fizyczne jednostki, które będą przedmiotem prac w dalszych etapach.

4. Ocena konserwatorska

Specyfiką projektu Jagiellońska Biblioteka Cyfrowa jest poświęcanie również dużej uwagi zabezpieczeniu oryginałów

jak samej digitalizacji. W tym celu w kroku oceny konserwatorskiej budowana była szczegółowa baza danych o stanie zachowania objętych projektem woluminów. Do systemu wprowadzany był najpierw opis obiektu (rozmiary, charakterystyka oprawy, kart i bloku), a następnie lista poziomych zniszczeń wywołanych wymienionymi czynnikami – w umownej konwencji od 0 (brak zniszczeń tego rodzaju) do 3 (maksymalne zniszczenie wywołane danym czynnikiem). System pozwala wytypować obiekty do napraw: np. najbardziej zniszczone przez dany czynnik – albo też, po wysumowaniu poziomów, w ogólnie najlepszym czy najgorszym stanie.

Ocena konserwatorska kończy się dla każdego woluminu zaleceniem czynności konserwatorskich oraz przechowywania po skanowaniu.

5. Naprawa

Woluminy po ocenie trafiają do faktycznej konserwacji, gdzie zostają poddane zaleconym procedurom (naprawa, odkwaszanie, dezynfekcja). Wykonane czynności zostają w systemie odnotowane w odniesieniu do każdego woluminu. Naprawa oryginałów była korzystna również dla procesu skanowania, do którego trafił materiał wejściowy w lepszym stanie.

Po naprawie i skanowaniu woluminy zostają ponownie poddane ocenie konserwatorskiej, aby w systemie odnotowany

był ich faktyczny stan zachowania.

Dzięki takiemu podejściu, w momencie zakończenia skanowania i przesłaniu woluminów do długoterminowego przechowywania, będzie znany szczegółowo stan zachowania całego zdigitalizowanego zbioru.

6. Skanowanie

Woluminy po konserwacji trafiają do skanowania. SOJBC pozwala śledzić na bieżąco, którymi woluminami można się już zająć. Dzięki informacji o rozmiarach, wprowadzonej w kroku oceny konserwatorskiej, ułatwiony jest także przydział materiału do skanera odpowiedniej wielkości (A0, A1 oraz A2).

Dostarczone woluminy zostają zeskanowane, a następnie skany muszą trafić do SOJBC. Odpowiednie katalogi na dysku sieciowym zostały już przez system założone w momencie utworzenia woluminu. Uzyskane skany należy wgrać do takiego właśnie katalogu, a następnie potwierdzić w systemie. System przegłąda katalog, zapisuje informacje o plikach i przenosi je do wewnętrznego katalogu, aby uniknąć ich przypadkowego usunięcia czy nadpisania.

7. Kontrola

Wprowadzone skany należy teraz strona po stronie skontrolować. Możliwych błędów jest wiele, zarówno pochodzących z oryginału (błędy drukarskie czy

intrologatorskie), jak i z samego procesu skanowania.

Cały proces kontroli skanów opiera się na statusach. Wolumin, do którego właśnie wgrano świeże skany, otrzymuje status „Sprawdź”. Pracownicy zajmujący się kontrolą przeważnie podzieleni są na zespoły, sprawdzające materiał wpływający z pojedynczego skanera. W panelu kontroli wyszukują zatem woluminy o statusie „Sprawdź”, zeskanowane na „swoim” skanerze. Po wyborze woluminu wybierają akcję „eksportu”, która tworzy dla nich na dysku sieciowym nowy katalog. Zawiera on wszystkie skany danego woluminu, przy czym nazwy plików są usystematyzowane wg numeru systemowego woluminu. Kontrolujący ściąga cały katalog na swój własny dysk i przegląda skany plik po pliku. Jeśli nie znalazł żadnych usterek, zmienia status woluminu na „SprawdzoneOK”, co oznacza, że wolumin jest gotowy do dalszego przetwarzania. Jeśli jednak znalazł jakiś błąd, ustawia status „SprawdzoneBłąd” i opisuje szczegółowo problem w polu „Uwagi o skanach”. Kierując się tymi uwagami, osoba skanująca musi dokonać poprawek, wgrać poprawione skany do systemu i ustawić status „SprawdźPonownie”, po czym wolumin trafi do ponownej kontroli.

8. Publikacja

Po kontroli jakości skanów, można przystąpić do publikacji woluminu. Skany są sortowane (np. rozdzielane na poszcze-

gólne numery czasopisma), konwertowane na format prezentacyjny (DjVu) i publikowane w dLibrze. Metadane opisujące publikację są importowane z Komputerowego Katalogu Zbiorów Bibliotek UJ (VTLS/Virtua) po sprawdzeniu w SOJBC numeru systemowego BIBID (wraz z linkiem do opisu bibliograficznego w KKZBUJ).

Fakt opublikowania zostaje odnotowany w SOJBC przez podanie numeru utworzonej w dLibrze publikacji. Ten numer (adres WWW) z kolei może zostać wprowadzony do KKZBUJ (VTLS/Virtua), przez co dostęp do publikacji jest możliwy z obydwóch katalogów.

9. Archiwizacja cyfrowa

Skany po kontroli zostają zarchiwizowane na taśmach. Panel „Archiwizacja” umożliwia kontrolę, jaką objętość zajmują skany gotowe do archiwizacji. Gdy przekroczy ona objętość taśmy, pliki zostają nagrane na taśmę i usunięte z dysku. System przechowuje informacje o zawartości poszczególnych taśm i umożliwia wydrukowanie etykiety na taśmę. Taśmy (zgodnie z zasadami archiwizowania) nagrane zostają w dwóch kopiach, które przechowuje się w odrębnych lokacjach.

Przy objętościach rzędu 750GB proces nagrywania i weryfikacji pojedynczej taśmy trwa prawie 8 godzin. Dlatego bardzo pomocna jest tutaj możliwość automatyzacji tego procesu, w szczególności zastosowanie napędu taśmowego

z robotem, który pozwala załadować naraz 30 taśm, a potem sterować ich zmianą zdalnie.

W niektórych przypadkach zdarza się, że pliki zarchiwizowane już na taśmach są potrzebne do powtórnej obróbki – np. gdy czujne oczy czytelników dostrzegą jakiś błąd w prezentowanej publikacji. Po znalezieniu w systemie numeru taśmy, na której zawarte są pliki i transferze jej zawartości na macierz dyskową, SOJBC umożliwia odtworzenie katalogu „eksportowego” ze skanami danego woluminu.

10. Archiwizacja mikrofilmowa

Oprócz postaci cyfrowej, projekt zakłada także zachowanie digitalizowanych kopii na mikrofilmach. Mikrofilmy powstają na konwerterze, który wyświetla przygotowane skany i fotografuje je na mikrofilmie.

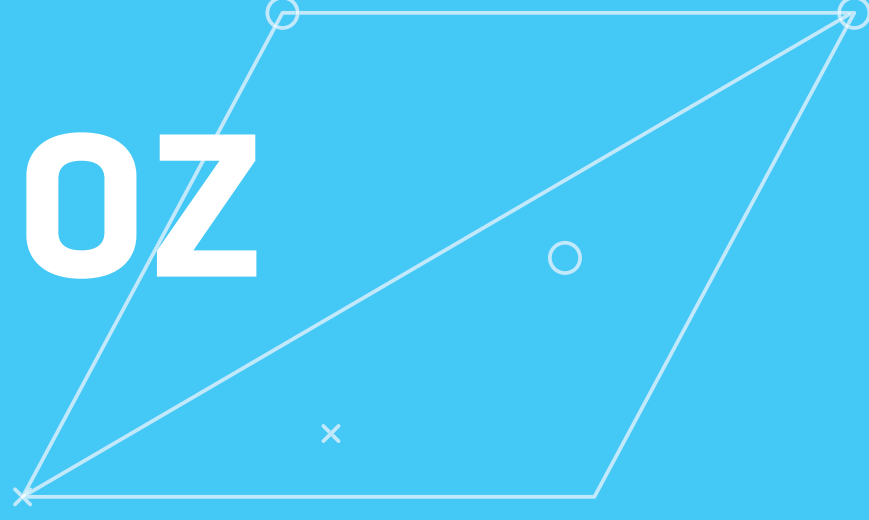
Proces przygotowania zawartości mikrofilmu jest silnie powiązany z przygotowaniem do publikacji, w szczególności z podziałem zeskanowanych woluminów czasopism na poszczególne numery. Trwają prace nad dodaniem do SOJBC wprowadzania informacji o tym podziale. Ułatwi to przygotowanie zawartości dla mikrofilmu i umożliwi prowadzenie rejestru zawartości mikrofilmów na poziomie pojedynczego numeru.

11. Raporty

Oprócz paneli dla osób pracujących przy kolejnych krokach digitalizacji, SOJBC udostępnia również liczne panele raportujące i statystyczne. Statystyki ogólne obrazują w czasie postęp w przetwarzaniu konkretnych woluminów, natomiast statystyki oddziałowe wykazują wkład pracy poszczególnych osób. Ponadto, dla pojedynczej publikacji, można sprawdzić, jak zaawansowany jest proces jej digitalizacji.

Jak zatem widać, System Obsługi z powodzeniem wspiera pracowników projektu na wszystkich etapach, zarówno w procesie digitalizacji, jak i archiwizacji (oryginałów, skanów, mikrofilmów) oraz katalogowania. Z uwagi na specyficzne potrzeby projektu, powstał wewnątrz Biblioteki Jagiellońskiej od podstaw, a jego użytkownicy uznają za spory atut fakt, że system może być na bieżąco poprawiany i uzupełniany o brakujące funkcje.

Ochrona zbiorów



Program konserwacji oraz zabezpieczania zbiorów specjalnych i czasopism XIX- i XX-wiecznych



Słowa kluczowe: konserwacja zachowawcza, program prac konserwatorskich, ocena stanu zachowania kolekcji, ankieta konserwatorska, opakowania ochronne: pudła, koszulki, passe-partout, wykonane z materiałów bezkwasowych

Keywords: preventive conservation, program of conservation work, collection assessment, conservation survey, protective containers: boxes, folders, passe-partout made of nonacidic materials

W latach 2002–2003 w ramach Wieloletniego Programu Rządowego „Kwaśny Papier” w Bibliotece Jagiellońskiej zostały przeprowadzone badania statystyczne oceniające stan zachowania księgozbioru XIX- i XX-wiecznego. Ok. 13% badanych książek i czasopism zostało uznanych za obiekty „w bardzo złym stanie, uzasadniającym konieczność wyłączenia z udostępniania” [Sobucki 2004, s. 182–198]. Zbiór ten stanowią głównie czasopisma, wydawane w II połowie XIX wie-

ku oraz w I połowie wieku XX, a także zbiory specjalne z tego okresu, czyli te, których dotyczy problem zakwaszonego, a co za tym idzie osłabionego mechanicznie papieru. Są to egzemplarze, które Biblioteka Jagiellońska chce obecnie zabezpieczyć oraz zarchiwizować i z których zamierza stworzyć kopie cyfrowe.

Mając świadomość, że problem zakwaszonego papieru z biegiem czasu pogłębia się, ponieważ proces kwasowej degradacji w papierze postępuje, wybrano 225 tytułów najcenniejszych czasopism i gazet XIX- i XX-wiecznych (przechowywanych w Bibliotece Jagiellońskiej jedyne egzemplarze w Polsce). Zdecydowano o przeprowadzeniu ich digitalizacji wraz z częściową naprawą i zabezpieczeniem oryginałów w ramach projektu. Program przewidywał również zabezpieczenie obiektów ze zbiorów specjalnych oznaczonych jako „0” (tzn. jedyne egzemplarze w Polsce), których kopie cyfrowe w postaci plików TIFF, JPG już zostały wykonane w poprzednich latach, a które

zaplanowano udostępnić w Jagiellońskiej Bibliotece Cyfrowej.

Chcąc stworzyć kompleksowy program ochrony wybranych zbiorów zaplanowano szereg zadań będących realizacją Działania I projektu, zatytułowanego „Ochrona zbiorów”. Przewidziano: konserwację zachowawczą, wykonanie obrazów cyfrowych, wykonanie mikrofilmów. Omawiane poniżej prace przeprowadzane były przez Sekcję Konserwacji, Sekcję Opraw i Zabezpieczania Zbiorów, Sekcję Kliniki Papieru, wchodzących w skład Oddziału Zabezpieczenia Zbiorów Biblioteki Jagiellońskiej, realizującego zadanie: konserwacja zachowawcza.

Opracowanie programu prac konserwatorskich

Znając założenia projektu i mając ogólną wiedzę na temat stanu zachowania wybranych do projektu obiektów, opracowano program prac konserwatorskich, zakładający przeprowadzenie szczegółowej oceny stanu zachowania i podstawowych zabiegów z zakresu profilaktyki ochrony zbiorów [Szalla-Kleemann 2008]. Zabiegi konserwatorskie, o różnym zakresie, miały dotyczyć wszystkich obiektów wytypowanych do projektu (ok. 3000 jednostek). Konserwacją zachowawczą miała zostać objęta kolekcja zbiorów unikalnych (zbiory specjalne klasy „0”) i bardzo zniszczonych (czasopisma i gazety).

Biorąc pod uwagę konieczność zabez-

pieczania unikatów przechowywanych w Bibliotece Jagiellońskiej, należących do dziedzictwa kulturowego Polski, postanowiono kompleksowo zabezpieczyć zeskanowane już wcześniej obiekty. Spośród przeszło 1000 egzemplarzy zbiorów specjalnych wchodzących do JBC, około 500 obiektów oznaczonych jako klasy „0” miało zostać oczyszczonych, poddanych zabiegom konserwatorskim i umieszczonych w pudłach lub teczkach ochronnych. Podobnie planowano

Abstrakt: Przedmiotem artykułu jest omówienie realizacji kolejnych punktów kompleksowego programu ochrony wybranego zbioru czasopism XIX- i XX-wiecznych i zbiorów specjalnych, objętych projektem JBC. W ramach Działania I, zatytułowanego „Ochrona zbiorów” zaplanowano szereg zadań takich jak: konserwacja zachowawcza, wykonanie obrazów cyfrowych, wykonanie mikrofilmów. Omawiane poniżej prace przeprowadzane były przez sekcje Oddziału Zabezpieczenia Zbiorów Biblioteki Jagiellońskiej (Sekcje Konserwacji, Sekcje Opraw i Zabezpieczania Zbiorów, Klinikę Papieru), realizujące zadanie: konserwacja zachowawcza.

Abstract: The article discusses the realization aspects of an integrated program of preservation of 19th and 20th century newspaper and special collections, included in the project. Measure I, „Preservation of collections”, involved a number of tasks such as preventive conservation or producing digital images and microfilm copies, carried out by the respective sections of the Preservation Department of the Jagiellonian Library (Conservation, Binding and Preservation of Collections, Paper Clinic) within the general objective of preventive conservation.



zabezpieczyć woluminy czasopism i gazet z XIX i XX wieku.

Tak zabezpieczonych zbiorów nie zamierza się udostępniać w przyszłości czytelnikom. Zastąpią je bowiem obrazy zdigitalizowane, udostępniane w bibliotece cyfrowej. Takie zasady udostępniania zbiorów cennych i zniszczonych przyczyniają się do maksymalizacji ochrony oryginałów zgodnie z międzynarodowymi tendencjami [McIlwaine i in. 2002].

Stworzono osobno programy dla dwóch kategorii zbiorów.

1. Zbiory specjalne klasy „0”

1.1. Stan zachowania zbiorów specjalnych

Ogólny stan zachowania zabytkowych materiałów bibliotecznych, należących do zbiorów specjalnych, był bardzo różnorodny. Wynikało to m.in. z ich dużej różnorodności, szczególnie w przypadku zbiorów graficznych i kartograficznych. Wśród wytypowanych obiektów znajdowały się zarówno egzemplarze dużych formatów: pojedyncze mapy oraz atlasy kartograficzne, jak i małych formatów: pojedyncze grafiki, niekiedy zebrane w teki lub woluminy oprawne. Wśród rękopisów były jednostki oprawne, jak i przechowywane w postaci luźnych kart, zebranych w tzw. fascykułach.

Niektóre z obiektów specjalnych, oznaczonych jako „0” poddane zostały

w przeszłości zabiegom konserwatorskim, niektóre zaś nadal były w złym stanie. Najczęściej objekty te wymagały oczyszczenia, wykonania napraw konserwatorskich w różnym zakresie, przełożenia w nowe opakowania ochronne. Najczęstszymi zniszczeniami ksiąg oprawnych były uszkodzenia mechaniczne opraw (naderwane lub całkowicie oderwane grzbiety i przednie okładziny). Niektóre z obiektów przechowywane były w pudłach lub teczkach ochronnych, w większości jednak wymagały one wymiany na nowe opakowania, wykonane z odpowiednich, trwałych i bezpiecznych materiałów przeznaczonych do archiwizowania zabytków tej klasy. Niezbędne było wykonanie zarówno pudeł ochronnych, jak i teczek, passe-partout na grafiki, kopert i koszulek ochronnych z papieru bezkwasowego.

1.2. Program prac konserwatorskich dla zbiorów specjalnych klasy „0”

Zaplanowano wykonanie prac w zakresie konserwacji zachowawczej:

1. Dokładne odkurzenie księgozbioru (zalecany sprzęt: odkurzacze z filtrami Hepa, które uniemożliwiają rozpylenie zarodników mikroorganizmów, co ma miejsce w przypadku używania zwykłych odkurzaczy).

2. Opracowanie i sporządzenie inwentaryzacji konserwatorskiej, zawierającej krótką charakterystykę techniki wykonania obiektu, użytych materiałów, zdo-

bienia oraz zakres zniszczeń poszczególnych części danego woluminu (oprawy, bloku, papieru); analiza zniszczeń miała pozwolić na przygotowanie szczegółowego programu prac konserwatorskich, dostosowanego do stanu zachowania poszczególnych woluminów, wyodrębnienie obiektów najbardziej zniszczonych i ewentualne przeprowadzenie zabiegów pełnej konserwacji poza projektem.

3. Ocena czystości mikrobiologicznej.

4. Oczyszczenie na sucho opraw oraz bloków książek (gumki typu Wishab, Wallmaster).

5. Przeprowadzenie podstawowych, wymaganych napraw, mieszczących się w zakresie konserwacji zachowawczej, takich jak podklejenie naderwanych grzbietów i okładzin książek, podklejenie rozdarć, odkwaszanie papieru w przypadku papierów maszynowych.

6. Zabezpieczenie wszystkich obiektów poprzez umieszczenie ich w pudłach i teczkach ochronnych z tektury bezkwasowej, wykonanie nowych passe-partout (w przypadku grafik) i koszulek ochronnych z papieru bezkwasowego (w przypadku map, które przechowywane są w mapnikach).

2. Czasopisma XIX- i XX-wieczne

2.1. Stan zachowania czasopism XIX- i XX-wiecznych

Stan zachowania czasopism XIX- i XX-wiecznych był na tyle zły, że ich znaczną część w ostatnich latach wyłączono już z udostępniania. Są one wydrukowane na papierze z dodatkiem ścieru drzewnego, który jest bardzo kruchy i zagrożony zniszczeniem z powodu tzw. kwasowej degradacji celulozy. Większość obiektów to woluminy oprawne dużych formatów (III, IV, V), co sprzyja uszkodzaniu mechanicznemu kruchego papieru podczas użytkowania. Znaczna część woluminów miała silnie uszkodzone grzbiety i szycie, niekiedy przechowywana była też bez zabezpieczenia lub w tzw. fascykułach, które nie stanowią należytej ochrony, zarówno przed dalszymi uszkodzeniami mechanicznymi, jak i przed zanieczyszczeniami atmosferycznymi, przyspieszającymi degradację papieru. Innymi charakterystycznymi zniszczeniami dla tego zbioru były silne uszkodzenia mechaniczne brzegów oraz całych kart kruchego papieru. Z powodu postępującej degradacji kwaśnego papieru wymagały natychmiastowych napraw i odkwaszenia (co pozwoliło na zatrzymanie procesu niszczenia papieru, ale nie cofnęło jego dotychczasowych skutków).

2.2. Program prac konserwatorskich dla czasopism XIX- i XX-wiecznych

1. Wstępne odkurzenie i oczyszczenie na sucho całego zbioru czasopism przeprowadzone w magazynie (zalecany sprzęt: odkurzacze z filtrami Hepa; gumki typu Wishab, Wallmaster).

2. Sporządzenie krótkich opisów stanów zachowania poszczególnych woluminów czasopism, ujęte w tabeli, z uwzględnieniem stanu zachowania oprawy, bloku i stopnia degradacji papieru.

3. Określenie stopnia skażenia mikrobiologicznego woluminów ze śladami wzrostu drobnoustrojów (badanie przy użyciu luminometru, pozwalającego określić aktywność biologiczną mikroorganizmów).

4. Selekcja zainfekowanych obiektów do dezynfekcji w komorze fumigacyjnej (gazem Rotanox).

5. Wykonanie napraw niezbędnych do bezpiecznego wykonania czytelnej kopii cyfrowej czasopisma, takich jak podklejenia uszkodzonych kart lub naprawy zniszczonych grzbietów i opraw woluminów.

6. Przekazanie obiektów do skanowania.

7. Skanowanie.

8. Selekcja obiektów do masowego odkwaszania w instalacji Bookkeeper III w zależności od stopnia degradacji papieru.

9. Odkwaszanie.

10. Wykonanie odpowiednich pudeł ochronnych z trwałych i bezpiecznych materiałów przeznaczonych do archiwizowania.

11. Przepakowanie wszystkich zdigitalizowanych czasopism i gazet do nowych pudeł i teczek ochronnych.

Opracowanie systemu oceny księgozbioru Biblioteki Jagiellońskiej

Kolejnym krokiem było skonstruowanie ankiety konserwatorskiej, przeznaczonej do szczegółowej oceny stanu zachowania obiektów, będącej jednym z elementów narzędzia elektronicznego o nazwie System Obsługi Jagiellońskiej Biblioteki Cyfrowej. Korzystając z wcześniejszych doświadczeń opracowania księgozbioru Biblioteki oo. Kamedułów na Bielanych podczas jego przenosin do Biblioteki Jagiellońskiej¹, wspólnie z Oddziałem Komputeryzacji BJ stworzono ankietę konserwatorską.

Ankieta skonstruowana w formie tabeli, składała się z 4 głównych części i zawierała podstawowe informacje o obiekcie, takie jak krótka identyfikacja i charakterystyka techniki wykonania obiektu, użyte materiały, rodzaje zdobień oraz zakres zniszczeń poszczególnych części danego woluminu (oprawy, bloku, papieru), a także zalecenia konserwatorskie.

W części pierwszej, identyfikacyjnej, umieszczono informacje, czy obiekt jest oprawny czy nie, jego wymiary, jakie posiada opakowanie ochronne, czy był poddawany wcześniej jakimkolwiek interwencjom konserwatorskim, takim jak

¹ Wzorowano się po części na elektronicznej ankiecie konserwatorskiej autorstwa Izabeli Damulewicz i Bartłomieja Brawuskiego, skonstruowanej na potrzeby konserwacji profilaktyczno-zachowawczej księgozbioru oo. Kamedułów, podczas jego przenosin w charakterze depozytu do BJ, w latach 2007-2009.

naprawy, pełna konserwacja, odkwaszanie czy dezynfekcja.

Kolejne kolumny to część druga, opisująca technikę wykonania obiektu i rodzaje użytych materiałów, typ oprawy, rodzaj obleczenia, rodzaj użytego w bloku papieru, techniki zdobienia kart, takie jak malowane inicjały, miniatury, grafiki, mapy itp. Poza tym, każda z kolumn głównych posiadała wolne pole do wpisywania dodatkowych uwag, dotyczących choćby nietypowych rozwiązań technologicznych.

Część trzecia, poświęcona zniszczeniom, jest podzielona na kolumny dotyczące stanu zachowania oprawy, bloku i kart. Przy każdej z tych pozycji wymieniono główne rodzaje zniszczeń występujących na tych elementach książki. W polu opisującym oprawę umieszczono między innymi informacje o tak istotnych zniszczeniach, jak deformacje, uszkodzenia mechaniczne poszczególnych jej elementów, a także stopień zabrudzenia, w polu opisującym blok – informacje o uszkodzeniach jego konstrukcji, o łączeniu bloku z oprawą, w opisie kart – stopień zabrudzenia, zaplamienia, ubytki, rozdarcia itp. Skala poszczególnych zniszczeń została wyrażona w punktacji od 0 do 3, gdzie 0 oznacza brak danego zniszczenia, a 3 najbardziej zaawansowane. Wartości w kolejnych kolumnach sumują się, dając wynik stopnia zniszczenia poszczególnych elementów (oprawy, bloku, kart), jak i całego obiektu. Pozwala to na wyselekcjonowanie według skali

zniszczeń tych obiektów, które wymagają w pierwszej kolejności interwencji konserwatorskiej. Powstała swoistego rodzaju baza danych o różnego rodzaju zniszczeniach, można prześledzić, jakiego rodzaju zniszczenia występują najczęściej lub najrzadziej, można określić ogólny charakter zniszczeń.

Ostatnią część stanowiły zalecenia dotyczące rodzaju opakowania (pudło, teczka, koszulka, passe-partout), rodzaju odkwaszania, zalecanej dezynfekcji, informacje, czy obiekt nadaje się do skanowania czy nie.

Zaletą narzędzia jest niewątpliwie szybkie wpisywanie danych (polegające na odhaczaniu wprowadzonych w systemie skrótów propozycji), tworzących przydatną bazę informacji na temat zniszczeń, ale i charakteru opracowywanego zbioru. Daje to możliwość wyszukiwania obiektów o konkretnych parametrach.

Stworzenie takiej bazy pozwala spojrzeć przekrojowo na całą kolekcję, zaplanować i podjąć kompleksowy program ochrony wybranego księgozbioru.

Ocena stanu zachowania 1810 woluminów czasopism XIX- i XX-wiecznych (przed skanowaniem)

Ocenę czasopism przeprowadził zespół 4 osób, konserwatorów zabytków o specjalizacji papier i skóra², na terenie magazynów czasopism Biblioteki Jagiellońskiej, wprowadzając dane do systemu

² Marek Ciesielczyk, Anna Rychter, Aleksandra Szalla-Kleemann, Anna Żukowska-Zielińska, pracownicy UJ.

przy użyciu komputerów przenośnych. Zasadniczą część zbioru, egzemplarze poprawne i niepoprawne, oceniono w okresie od lipca 2010 r. do stycznia 2011 r., uzupełniając bazę o informacje o pojedynczych egzemplarzach w kolejnych miesiącach do końca 2011 r.

Czasopisma oceniano przed przystąpieniem do skanowania, dlatego ocenę wykonano w pierwszej kolejności, przekazując woluminy sukcesywnie do sprawdzenia kompletności egzemplarza, następnie do napraw i skanowania.

W tej grupie obiektów najistotniejszą informacją było, czy czasopismo kwalifikuje się do skanowania. Głównym kryterium był stan zachowania i wytrzymałość mechaniczna kart, a także czytelność tekstu. Gazety, których karty posiadały rozdarcia, zmarszczenia, zagięcia jak i dawne naprawy, przysłaniające tekst, kierowane były do Sekcji Opaw i Zabezpieczania Zbiorów OZZ BJ, w celu przeprowadzenia zabiegów umożliwiających czytelne skanowanie. Dawne naprawy usuwano, przedarcia mniejsze i większe naprawiano w minimalnym stopniu, niezbędnym do bezpiecznego skanowania.

Kolejnym ważnym elementem oceny w przypadku obiektów na papierze maszynowym była decyzja o skierowaniu do odkwaszania maszynowego o odpowiednim trybie (maszynowy, mieszany, ręczny) w zależności od stanu zachowania kart.

Ocena stanu zachowania 1190 jednostek ze zbiorów specjalnych (po skanowaniu)

Oceny 1190 jednostek ze zbiorów specjalnych dokonał ten sam zespół konserwatorów, w okresie od stycznia do lipca 2011 r. Podobnie jak w przypadku czasopism, zbiory oceniano w magazynach lub czytelnich danego oddziału (ORP, OSD, OZM, OZGiK, OWRiDŹS), z którego pochodziły obiekty. Zasadni-

czą różnicą był fakt, że zbiory specjalne, które włączono do Jagiellońskiej Biblioteki Cyfrowej zostały już wcześniej zeskanowane, dlatego też podczas oceny pominięto kwalifikacje do skanowania, opisano natomiast ówczesny stan zachowania obiektu. Celem oceny było stworzenie bazy danych i wytypowanie obiektów do konserwacji zachowawczej.

Oceniano zeskanowane zbiory specjalne kategorii 0, 0', 1 i W.

Kategorie te oznaczają: 0 – unikatki przechowywane tylko w BJ, kategoria 0' – zbiory rzadkie, cenne ze względu na oprawy, zapiski proveniencyjne itp., (w przypadku rękopisów – obiekty „średniej klasy” mniej znanego twórcy), kategoria 1 – zbiory cenne w małym nakładzie, kategoria W – zbiory cieszące się zainteresowaniem czytelników.

Jedynie tzw. „zerówki” zostały objęte programem konserwacji zachowawczej

Oddział	Ocenił „0” (jednostki i woluminy)	Oprawne (woluminy)	Nieoprawne (jednostki biblioteczne)
ORP	65	45 wol.	20 jedn.
OSD	150 jedn.	134 wol.	16 jedn.
OZGiK	87 jedn.	25 wol.	62 jedn.
OZM	138 jedn.	59 wol. w tym 17 klocków	16 jedn.
OWRiDŹS	52	36	16
Zbiory specjalne	492	299	130

© Tab. 1. Liczba obiektów klasy „0” ze zbiorów specjalnych, ocenionych, wytypowanych do poddania konserwacji profilaktyczno-zachowawczej.

w ramach niniejszego projektu. Dlatego też grupę wybranych, niemalże 500 obiektów, poddano szczegółowej analizie, w celu przeprowadzenia przetargu na wykonanie prac konserwatorskich.

Sporządzenie programu prac, wyłonienie firmy zewnętrznej realizującej konserwację profilaktyczno-zachowawczą

1. Zakwalifikowanie obiektów w czterech kategoriach stanu zachowania, zależnie od stopnia zniszczenia

Po przeprowadzeniu oceny konserwatorskiej 492 woluminów i jednostek klasy „0” ze zbiorów specjalnych, zakwalifikowano wszystkie obiekty do czterech kategorii stanu zachowania, biorąc pod uwagę punktację uzyskaną w ostatecznej sumie wszystkich zniszczeń w ankiecie konserwatorskiej w Systemie Obsługi

Oddział	Kat. I			Kat. II		
	oprawne (0-10)	nie-oprawne (0-5)	razem	oprawne (11-20)	nie-oprawne (6-10)	razem
ORP	6	4	10	20	3	23
OSD	55	8	63	45	8	53
OWRiDŹS	9	2	11	15	1	16
OZGIK	0	27	27	3	28	31
OZM	70	2	72	35	6	41
OGÓŁEM	140	43	183	118	46	164
Oddział	Kat. III			Kat. IV		
	oprawne (21-30)	nie-oprawne (11-15)	razem	oprawne (30+)	nie-oprawne (16+)	razem
ORP	11	10	21	9	2	11
OSD	25	1	26	8	1	9
OWRiDŹS	9	9	18	2	5	7
OZGIK	16	8	24	7	0	7
OZM	8	6	14	3	3	6
OGÓŁEM	69	34	103	29	11	40

☉ Tab. 2. Stan zachowania obiektów oprawnych i nieoprawnych w ujęciu punktacyjnym, z poszczególnych oddziałów zbiorów specjalnych, przed konserwacją.

Jagiellońskiej Biblioteki Cyfrowej. Stworzono odmienną punktację dla obiektów oprawnych i nieoprawnych.

2. Kategorie stanu zachowania i suma punktów w SOJBC

2.1. Kategorie stanu zachowania, suma punktów i zalecenia dla egzemplarzy oprawnych

Kat. I: 0-10 punktów w skali zniszczeń w ankiecie konserwatorskiej SOJBC. Oznacza dobry stan zachowania, niekiedy wymagane oczyszczanie mechaniczne obiektów, niewielkie naprawy.

Kat. II: 11-20 punktów w skali zniszczeń w ankiecie konserwatorskiej SOJBC. Oznacza dość dobry stan zachowania wymagający drobnych miejscowych napraw oprawy, podklejenia drobnych przedarć kart, usunięcia zabrudzeń.

Kat. III: 21-30 punktów w skali zniszczeń w ankiecie konserwatorskiej SOJBC. Oznacza zły stan zachowania, wymagający przeprowadzenia prac bardziej zaawansowanych, takich jak naprawy grzbietów opraw, zerwanych okładzin, uszkodzonych narożników, zerwanych wiązań, napraw zniszczonych kart itp.

Kat. IV: 31-40 punktów w skali zniszczeń w ankiecie konserwatorskiej SOJBC. Oznacza bardzo zły stan zachowania, wymagający przeprowadzenia prac o podobnym zakresie jak w kategorii III, ale na szerszą skalę, w tym także obiekty do pełnej konserwacji.

2.2. Kategorie i suma punktów dla egzemplarzy nieoprawnych

Kat. I: 0-5 dobry stan zachowania.

Kat. II: 6-10 dość dobry stan zachowania, potrzeba miejscowych napraw, zabrudzenia.

Kat. III: 11-15 zły stan zachowania.

Kat. IV: 16-20 bardzo zły stan zachowania, w tym też obiekty do pełnej konserwacji.

3. Program prac i zalecenia

a) Dokładne odkurzenie księgozbioru i oczyszczenie na sucho opraw oraz bloków książek.

b) Przeprowadzenie podstawowych, wymaganych napraw, mieszczących się w zakresie konserwacji zachowawczej, odpowiadającej stanowi zachowania obiektów z danych kategorii, powstałych po przeprowadzeniu ankiet konserwatorskich w SOJBC, opisanych powyżej.

c) Ocena czystości mikrobiologicznej wytypowanych w SOJBC 29 obiektów i ewentualna dezynfekcja w momencie stwierdzonej aktywności drobnoustrojów.

d) Przeprowadzenie odkwaszania ręcznego wytypowanych 79 obiektów do odkwaszania.

e) Zabezpieczenie wszystkich obiektów poprzez umieszczenie ich w pudłach i teczkach ochronnych z tektury bezkwasowej, wykonanie nowych passe-partout (w przypadku grafik) i usztywnianych koszulek ochronnych z papieru bezkwasowego (w przypadku map, które przechowywane są w mapnikach).

Po zakończeniu prac podsumowujących ocenę stanu zachowania ogłoszono przetarg krajowy, nieograniczony. Od oferentów oczekiwano, że wykażą się doświadczeniem w podobnego typu realizacjach konserwacji różnorodnych kolekcji, o podobnej liczbie ok. 500 obiektów, jak i udokumentują wykonanie przynajmniej 3 usług z zakresu konserwacji profilaktyczno-zachowawczej obiektów zabytkowych na papierze i skórze w ciągu ostatnich 3 lat. Ponadto, wykonawca musiał posiadać uprawnienia do wykonywania zawodu konserwatora dzieł sztuki ze specjalizacją: konserwacja papieru i skóry, a także dysponować przynajmniej dwoma osobami zdolnymi do realizacji niniejszego zamówienia, posiadającymi również uprawnienia do wykonywania zawodu konserwatora dzieł sztuki z tą samą specjalizacją.

W wyniku przetargu wyłoniono firmę „Konserwacja Dzieł Sztuki. Izabela Damulewicz” z Torunia, która spełniała powyższe kryteria.

Konserwacja zachowawcza wykonana przez firmę zewnętrzną

Konserwację profilaktyczno-zachowawczą wybranego zbioru przeprowadzono w pracowni Sekcji Konserwacji OZZ BJ. Intensywne prace trwały 5 miesięcy, od połowy września 2011r. do końca lutego 2012 r., przy udziale zespołu złożonego z 3 do 4 osób.

Oddział	pudła	teczki	koszulki	koszulki usztywniane	passe-partout	sugerowana dezynfekcja	odkwaszanie
ORP	20	21	29	1	-	6	7
OSD	44	84	21	-	-	13	-
OZGiK	26	1	34	5	29	7	-
OZM	26	32	9	-	-	2	39
OWRiDŹS	25	25	20	-	-	1	33
	141	163	113	6	29	29	79

Tab. 3. Zalecenia dotyczące wykonania opakowań ochronnych, dezynfekcji i odkwaszania.

Konserwacja wykonywana przez zewnętrzną firmę na terenie BJ była przykładem pierwszej tego rodzaju realizacji, prowadzonej równoległe do normalnego trybu funkcjonowania Sekcji Konserwacji. Wszystkie materiały zakupione zostały ze środków na to zarezerwowanych w projekcie.

Prace konserwatorskie wykonano zgodnie z wcześniej planowanym postępowaniem, niekiedy wykraczając poza przewidziany program prac. Zakres działań przeprowadzanych na wybranym księgozbiornym był dostosowany każdorazowo do wymogów danego obiektu i obejmował następujące prace:

1. Oczyszczanie zbioru z kurzu i nalotu drobnoustrojów:

a) odkurzenie przy pomocy odkurzacza z filtrem typu Hepa,

b) oczyszczanie mechaniczne opraw, brzegów bloków oraz poszczególnych kart przy użyciu specjalistycznych gum typu wallmaster, wishab itp.

2. Konserwacja kart:

a) podklejanie przedarć i miejsc osłabionych w papierze,

b) uzupełnianie ubytków kart,

c) odkwaszenie kart z papieru maszynowego przy użyciu odczynnika Bookkeeper.

3. Konserwacja zabytkowych opraw:

a) oczyszczanie zabytkowych opraw przy użyciu odpowiednich emulsji konserwatorskich,

b) odkwaszanie, dogarbowywanie, natłuszczanie skór opraw,

c) miejscowe naprawy, podklejenia, uzupełnienia ubytków skór,

d) woskowanie skór,

e) wymiana na nowe niektórych opraw współczesnych.

4. Projektowanie oraz wykonanie opakowań ochronnych:

a) projektowanie różnego rodzaju koszulek ochronnych dla przechowywanych pojedynczo kart (OZM, OSD, ÖRP, ÖWRiDŽS),

b) projektowanie passe-partout ochron-

nych dla miniatur na pergaminie (ÖZGiK),

c) projektowanie usztywnianych koszulek dla obiektów o dużych wymiarach: mapy, afisze, plakaty (ÖZGiK, ÖWRiDŽS).

5. Prace dodatkowe, wykraczające poza zakres konserwacji profilaktyczno-zachowawczej. Podczas realizacji zadania zakres niektórych prac został poszerzo-

	Oddział	Kat. I			Kat. II		
		oprawne (0-10)	nie-oprawne (0-5)	razem	oprawne (11-20)	nie-oprawne (5-10)	razem
	ÖRP	39	19	58	7	0	7
	ÖSD	107	18	125	23	0	23
	ÖWRiDŽS	23	0	23	10	8	18
	ÖZGiK	7	63	70	18	0	18
	ÖZM	107	15	122	9	1	10
	ÖGÖŁEM	283	115	398	67	9	76
	Oddział	Kat. III			Kat. IV		
		oprawne (21-30)	nie-oprawne (11-15)	razem	oprawne (30+)	nie-oprawne (16+)	razem
	ÖRP	0	0	0	0	0	0
	ÖSD	3	0	3	0	0	0
	ÖWRiDŽS	2	7	9	0	2	2
	ÖZGiK	1	0	1	0	0	0
	ÖZM	0	1	1	0	0	0
	ÖGÖŁEM	6	8	14	0	2	2

© Tab. 4. Stan zachowania obiektów oprawnych i nieoprawnych w ujęciu punktacyjnym, z poszczególnych oddziałów zbiorów specjalnych, po konserwacji.

ny ze względu na rangę konserwowanych obiektów. Pracami wybiegającymi poza program konserwacji profilaktyczno-zachowawczej były prace związane z naprawami konstrukcji książki (rekonstrukcja zwińców, szycia, brakujących grzbietów książek).

Ze względu na bardzo zły stan niektórych czasopism z Oddziału Wydawnictw Rzadkich i Dokumentów Życia Społecznego, konieczne było przeprowadzenie kąpieli kart, dublaży papieru, czyli czynności również wykraczających poza charakter profilaktyczno-zachowawczy konserwacji.

Informacje o wszystkich działaniach konserwatorskich umieszczone zostały we wspomnianym narzędziu elektronicznym SOJBC, w zakładce Konserwacja. Interwencje zostały ujęte, podobnie jak ocena konserwatorska, w systemie punktacyjnym od 0 do 3 i wprowadzone pod punktacją dla danego zniszczenia. W narzędziu znajdują się również zestawienia sum punktacji dla obiektów przed i po zabiegach konserwatorskich. Można prześledzić jak zmienił się stan zachowania całych grup obiektów z wszystkich działów, z podziałem na oprawne i nieoprawne. Biorąc pod uwagę utworzone wcześniej na podstawie osiągniętej sumy punktów cztery kategorie stanu zachowania widać, że liczba obiektów, które znajdowały się przed konserwacją w kategorii I, czyli najlepszym stanie zachowania, stanowiła 183 sztuki i wzrosła do 398. Natomiast kategoria IV, o najgor-

szym stanie zachowania, zmalała z liczby 40 do 2.

Realizacja pozostałych zadań konserwacji zachowawczej

Od momentu rozpoczęcia projektu trwały prace związane z oczyszczaniem, przygotowaniem czasopism do skanowania, z przepakowywaniem, dezynfekcją i odkwaszaniem masowym papieru, poprzedzone oceną konserwatorską. Prace te były kontynuowane do momentu zakończenia projektu.

Przygotowanie czasopism do bezpiecznego skanowania przeprowadzane jest w Sekcji Opraw i Zabezpieczania Zbiorów OZZ BJ w minimalnym zakresie, tak aby umożliwić czytelne skanowanie. Oznacza to punktowe podklejanie przedarcie papieru, rozprostowywanie zmarszczeń i załamów kart, ale także usuwanie dawnych napraw, zasłaniających tekst, niekiedy poluzowywanie szycia w miejscach, gdzie karty zostały zbyt ciasno zszyte i tekst był niewidoczny.

Po skanowaniu zakwalifikowane do odkwaszania obiekty były odkwaszane w Klinice Papieru OZZ BJ i pakowane do pudeł i teczek bezklejowych, wycinanych na ploterze w Sekcji Opraw i Zabezpieczania Zbiorów.

Zakres nadzoru konserwatorskiego

Ze względu na zabytkowy i unikatowy charakter objętych projektem obiektów, należących do zbiorów specjalnych, a także na wyjątkowo zły stan zachowania wielu spośród czasopism XIX- i XX-wiecznych, niezbędny był stały nadzór konserwatorski podczas wykonywania wszystkich prac zaplanowanych w działaniu „Ochrona zbiorów”.

Nadzór konserwatorski nad całym działaniem „Ochrona zbiorów” powierzony został dyplomowanemu konserwatorowi papieru, kierownikowi Sekcji Konserwacji Oddziału Zabezpieczenia Zbiorów BJ.

Nadzorem konserwatorskim objęto zarówno działania wstępne, takie jak odkurzanie i oczyszczanie obiektów, jak i wszystkie późniejsze prace związane z naprawami zniszczonych egzemplarzy (w tym prace firmy zewnętrznej), przeprowadzenie ankiety konserwatorskiej stanu zachowania obiektów, jak i prac związanych z digitalizacją zbiorów. W przypadku skanowania, ustalono odpowiednie procedury postępowania, tak aby proces przebiegał w sposób bezpieczny i najmniej inwazyjny dla oryginałów.

Bibliografia

Mellwaine, John; Comment, Jean-Marc; Wolf de, Clemens; Peters, Dale; Justrell, Borje; Varlamoff, Marie-Thérèse, Koopman, Sjoerd (2002). Guidelines for Digitization Projects for collections and holdings in the public domain, particularly those held by libraries and archives [online]. IFLA. [Dostęp: 29.08. 2012]. Dostępny w World Wide Web: <http://www.ifla.org/files/assets/preservation-and-conservation/publications/digitization-projects-guidelines.pdf>

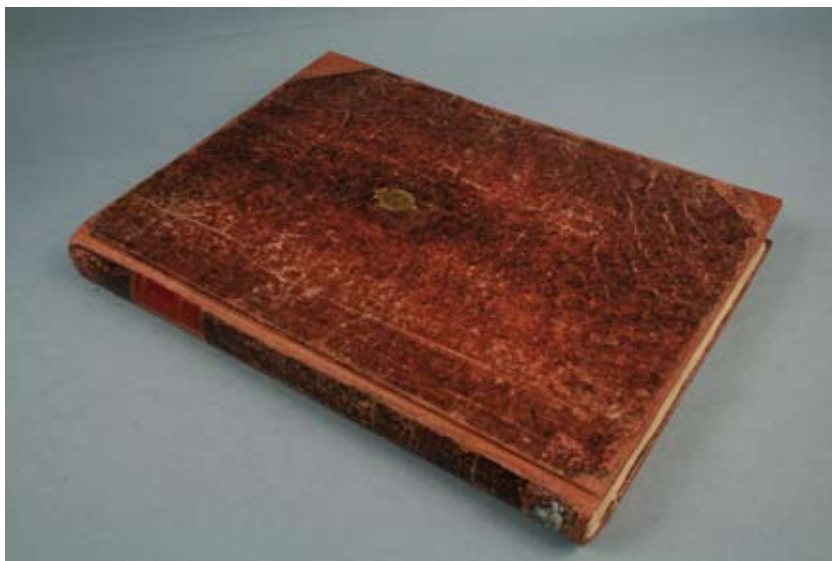
Sobucki, W., Koziński, Z., Drewniewska-Idziak, B. (2004). Stan zachowania księgozbioru z XIX i XX w. w Bibliotece Jagiellońskiej w Krakowie, Notes Konserwatorski, nr 8, s. 182–198.

Szalla-Kleemann, A., Szklanny, K. (2008). Program konserwacji i zabezpieczenia zbiorów specjalnych i czasopism XIX- i XX-wiecznych ze zbiorów Biblioteki Jagiellońskiej w ramach projektu Jagiellońska Biblioteka Cyfrowa. Załącznik nr 30 do wniosku o dofinansowanie realizacji projektu Jagiellońska Biblioteka Cyfrowa, Kraków: Biblioteka Jagiellońska.

© Il. 1a.



© Il. 1b.



© Il. 1a-1b. Druk muzyczny z Oddziału Zbiorów Muzycznych BJ przed i po konserwacji (fot. Izabela Damulewicz).

© Il. 2a.



© Il. 2b.



© Il. 2a-2b. Pergaminowe oprawy kolekcji z Oddziału Zbiorów Graficznych i Kartograficznych BJ, przed i po zabiegach konserwatorskich (fot. Izabela Damulewicz).

© Il. 3a.

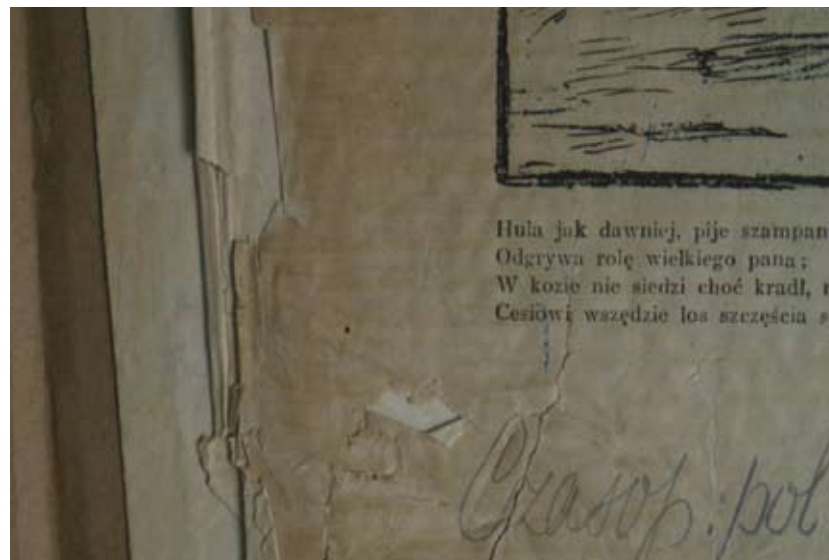


© Il. 3b.

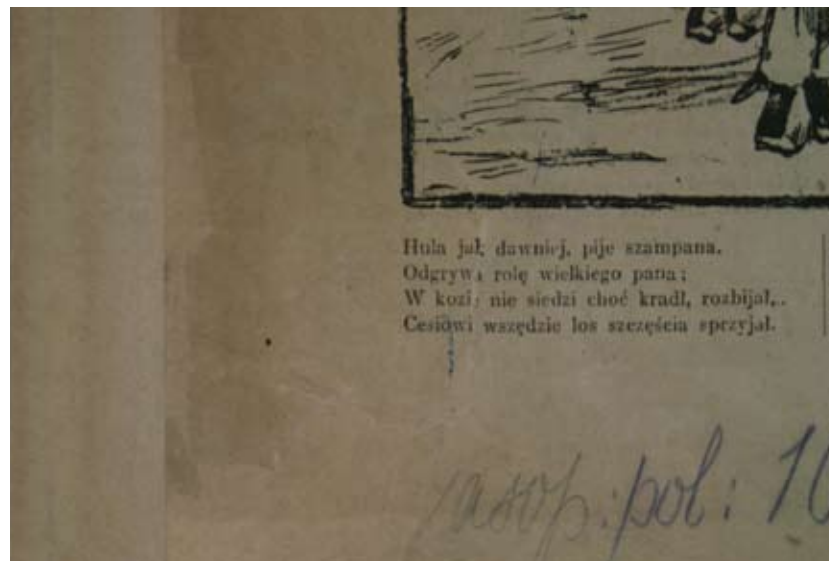


© Il. 3a-3b. Pergaminowa oprawa rękopisu z Oddziału Rękopisów BJ, przed i po usunięciu deformacji (fot. Izabela Damulewicz).

© Il. 4a.



© Il. 4b.



© Il. 4a-4b. Konserwacja czasopisma z Oddziału Wydawnictw Rządskich i Dokumentów Życia Społecznego BJ (fot. Izabela Damulewicz).

Selekcja czasopism do skanowania i publikacji w JBC



Słowa kluczowe: proces selekcji, czasopisma, kryteria selekcji, obiekty drukowane

Keywords: selection process, periodicals, selection criteria, printed items

58

Typowanie spośród całości zasobów tytułów przeznaczonych do mikrofilmowania lub digitalizacji, a także typowanie materiałów wymagających przeprowadzenia prac konserwatorskich, zawsze było dużym problemem dla archiwistów i bibliotekarzy. Świadomość tego, że niewybrana część zbiorów zostaje potraktowana jako mniej cenna, mniej ważna, skazuje wybierającego na duży dyskomfort. Niemniej jednak jest to czynność konieczna i nie da się jej uniknąć. Ty-

pując zbiory do Jagiellońskiej Biblioteki Cyfrowej, przyjęto kilka prostych zasad i od tej pory konsekwentnie się ich trzymano, także przy następnych projektach dotyczących digitalizacji. Z bogatego zbioru czasopism wyselekcjonowano te tytuły, których kopię cyfrową należało utworzyć w pierwszej kolejności. Przede wszystkim zadbano o dzieła najchętniej czytane, a jednocześnie najbardziej zniszczone oraz o jedyne egzemplarze w kraju.

Jest rzeczą powszechnie wiadomą, jak ogromna jest wartość zbiorów Biblioteki Jagiellońskiej. Przez stulecia swojego istnienia Biblioteka Jagiellońska w pełni zasłużyła na tytuł nadany przez Karola Estreichera „Bibliotheca Patria”. On sam w sprawozdaniu z 1901 r. takie podaje powody dla używania tego określenia „Przede wszystkim jest polska, a to z mocy tradycji pięciu stuleci. Wyrosła jako polska biblioteka. Po wtóre, jest biblioteką Uniwersytetu, mającą obowiązek zaspokajać potrzeby naukowe

uniwersyteckie. Po trzecie jest na pół Galicji jedyną publiczną biblioteką, z której czerpią wiedzę zarówno ludzie naukowcy, jak i zawodowi pracownicy, jak na przykład ogół urzędników. Tych ostatnich potrzeby muszą być od czasu do czasu, i to w miarę możliwości, zaspokajane, bo kraj nie posiada drugiego dostępnego źródła do zaspokajania potrzeb umysłowych” [Baumgart 1964]. Zaiste trudno o lepszą argumentację. O ile w powszechnej świadomości niezwykłość, rzadkość zbiorów rękopisów, starych druków czy też grafik funkcjonuje od wieków, o tyle unikalność czasopism jest wyraźnie niedoceniana.

Tymczasem Biblioteka Jagiellońska posiada jeden z najcenniejszych zbiorów gazet i czasopism wydawanych na terenie dawnej Galicji. Już w 1807 r. postanowiono dekretem cesarskim o przekazywaniu egzemplarza obowiązkowego druków krajowych, między innymi do biblioteki Uniwersytetu Krakowskiego. Zawirowania polityczne sprawiły, że przepis ten pozostał martwą literą. Potrzeba było blisko pół wieku, aby wraz z nowym aktem prawnym osiągnąć zamierzony cel. Według ustawy obowiązującej w całym cesarstwie od 27 maja 1852 r. (z późniejszymi niewielkimi zmianami z 17 grudnia 1862 r.) [Gaberle 1928], drukarze i księgarze byli zobowiązani do przekazywania egzemplarza obowiązkowego do bibliotek uniwersyteckich. W przypadku Galicji dokonano podziału – na zachód od Sanu druki trafiały do Biblioteki Jagiellońskiej, na

wschód od Sanu do biblioteki Uniwersytetu Lwowskiego. Rzeczywistość odbiegała od ustanowionego prawa. Biblioteka Jagiellońska miała jednak to szczęście, że w 1868 r. na stanowisku bibliotekarza rozpoczął pracę Karol Estreicher. Nie tylko poprawił on poziom ściągłości egzemplarza obowiązkowego (interweniując nawet u prokuratora), lecz doprowadził nawet do tego, że na mocy rozporządzenia ministra sprawiedliwości z 14 grudnia 1885 r. do Biblioteki

Abstrakt: Autor przedstawia proces selekcji przeznaczonych do digitalizacji czasopism przechowywanych w Bibliotece Jagiellońskiej. Jako najważniejsze kryterium doboru materiału do skanowania wymienia fizyczny stan zachowania zbiorów. Jako pozostałe kryteria autor wymienia unikatowość i kompletność czasopism. W artykule została przedstawiona analiza zadań podjętych w projekcie Jagiellońska Biblioteka Cyfrowa w rozbiu na poszczególne czynności.

Abstract: The author describes the procedure of selecting periodicals from the collections of the Jagiellonian Library for digital scanning. Physical condition of library materials is indicated as the most important factor in the selection process, other criteria being uniqueness and completeness of periodical runs. The respective tasks carried out within in the Jagiellonian Digital Library Project are also reviewed.



59



Il. 1.

trafiały także druki skonfiskowane z terenu Galicji Zachodniej. Dzięki temu Biblioteka Jagiellońska posiada sporo tego typu książek i numerów czasopism z lat 1885-1905, nadsyłanych przez prokuratorę [Baumgart 1964]. Dzięki niestrudzonej energii Karola Estreichera zbiór czasopism w Bibliotece był ustawicznie powiększany o druki z całej Galicji, a także z pozostałych zaborów. Nierzadko zdobywał dary w postaci całych księgozbiorów, zarówno prywatnych, jak i księgozbiorów instytucji [Dybiec 2000].

Karol Estreicher wiele wysiłku poświęcił na kompletowanie właśnie czasopism, uciekając się zarówno do ofiarności społeczeństwa, jak i zwracając się bezpośrednio do redakcji, wydawnictw, autorów. Wykorzystywał swoje koneksje, nie

tylko krakowskie, lecz także lwowskie czy warszawskie, dla ustawicznego powiększania księgozbioru bibliotecznego. To między innymi dzięki jego wieloletnim staraniom i ogromnemu wysiłkowi w Bibliotece Jagiellońskiej znajduje się 935 tytułów czasopism i kalendarzy polskich uznanych za jedyne egzemplarze w naszym kraju. Zbiór gazet i czasopism gromadzony przez dziesiątki lat stał się, niemal automatycznie, unikalny. Z liczby ponad 935 tytułów, które znajdują się tylko w Bibliotece Jagiellońskiej¹, ponad 60% wydane było na terenach Galicji. Lista tych unikalnych czasopism jest zbyt długa, aby przytaczać ją w całości. Są to druki przede wszystkim krakowskie i lwowskie, chociaż nie brakuje wydawnictw z mniejszych ośrodków: Czerniowców, Kołomyży, Nowego Sącza, Przemyśla, Stanisławowa, Tarnopola, Tarnowa

¹ Informacja uzyskana w kwietniu 2008 r. z Biblioteki Narodowej w Warszawie na podstawie Centralnego Katalogu Czasopism Polskich.

Il. 1. Regał w magazynie czasopism dużych formatów Biblioteki Jagiellońskiej.

i innych miast. Szczególnie cenne w tej grupie są poloniki, na przykład „Galicyskie Wiadomości” (wydawane w Brnie), „Nowy Tygodnik Polski” z Wiednia, moskiewskie „Echo Polskie” czy kijowski „Świt”. Niepowtarzalność była zatem jednym z najważniejszych kryteriów wyboru. Część z tego księgozbioru została już skopiowana na mikrofilmach. W niniejszym projekcie przeznaczono do digitalizacji 214 tytułów IV i V formatu. Skupiono się na tych formatach biorąc pod uwagę, że przez swoje rozmiary bardziej były narażone na zniszczenia mechaniczne.

Na wyjątkowość zbioru czasopism w Bibliotece Jagiellońskiej ma wpływ jeszcze jeden bardzo istotny fakt. Mianowicie o jego wartości, w przypadku najważniejszych gazet, świadczy kompletność poszczególnych tytułów. Przykładowo Biblioteka Jagiellońska posiada 281 woluminów „Gazety Lwowskiej”, czasopisma które ukazywało się przez 128 lat. Prawdopodobnie jest to najbardziej kompletny egzemplarz w Polsce. Świadomość tego faktu, spowodowała upublicznienie takich właśnie tytułów, aby potencjalny czytelnik w jednym miejscu uzyskał dostęp do pełnego zasobu. Wiązało się to z różnymi dodatkowymi zadaniami. Przyjęto mianowicie zasadę, że wraz z głównym tytułem publikowane będą wszelkie dodatki, samoistne i niesamoistne, co w fazie projektowania zadań było kłopotliwe, gdyż ilość dodatkowych woluminów znacząco wzrastała. Z tytułów związanych z „Gazetą Lwowską”, na

digitalizację czeka jeszcze tylko „Narodna Casopis” (1891-1912), którego Biblioteka posiada kompletny egzemplarz. Jak już wspomniano wyżej dzięki staraniom Karola Estreichera do Biblioteki trafiały egzemplarze, które ocenizowała policja. Zdarzają się przypadki, że w poszczególnych rocznikach znajdują się wszyte obok siebie, numer przed cenzurą i po, z białymi plamami, co czyni zasoby czasopism jeszcze cenniejsze (np. Głos Narodu, nr 269 z dnia 24 listopada 1900, Nowa Reforma, nr 181 z 7 sierpnia 1884 r.).

Myśląc o naszym dziedzictwie narodowym i stratach spowodowanych w wyniku wojen, nie sposób nie wspomnieć o jeszcze jednej szczególnej przyczynie, dla której podjęto działania zmierzające do jak najlepszej ochrony cennego zbioru czasopism. Zjawisko „kwaśnego papieru” jest dziś powszechnie znane, a jego skutki archiwiści i bibliotekarze odczuwają w swojej pracy codziennie. Problem ten dotyczy wszystkich druków ukazujących się od lat trzydziestych XIX wieku do lat dziewięćdziesiątych XX wieku, ale w przypadku czasopism jest szczególnie dotkliwy. Dzieje się tak dlatego, że do druku wydawnictw ciągłych, a szczególnie gazet, używano papieru kiepskiej jakości. Wiele gazet ukazywało się w dużym formacie. Ma to istotne znaczenie, gdyż druki te są dużo bardziej narażone na uszkodzenia mechaniczne (pęknięcia, rozdarcia, naderwania strony). Wydawnictwa te traktowano jako bardzo szybko dezaktualizujące się – niektóre tytuły

miały dwa wydania dziennie. Dziś wiemy, że nie bardziej błędnego. Najlepszym dowodem na to jest fakt, że trudno sobie wyobrazić pracę naukową, obejmującą dany okres historyczny, bez korzystania z tak bogatego źródła informacji, jakimi są gazety codzienne.

W latach 2002-2003 przeprowadzono w Bibliotece Jagiellońskiej badania oceniające stan zachowania księgozbioru (wyłączono z nich zbiory specjalne). Ok. 13% badanych książek i czasopism zostało uznanych za obiekty „w bardzo złym stanie, uzasadniającym konieczność wyłączenia z udostępniania” [Sobucki 2004]. Zdecydowaną większość z tego zbioru stanowiły właśnie czasopisma wydawane w 2. połowie XIX oraz w 1. połowie wieku XX. Badania przeprowadzone tylko na czasopismach krakowskich dały wyniki jeszcze bardziej niepokojące. Okazało się, że aż 82,5% z nich należy wycofać z udostępniania [Szalla-Kleemann 2001]. Z pewnością na taki stan miał wpływ fakt bardzo częstego udostępniania tytułów wydawanych w Krakowie. W maju 2007 r. dyrekcja Biblioteki Jagiellońskiej podjęła decyzję o wyłączeniu z udostępniania najbardziej zniszczonych gazet i czasopism. Zbiorem tym groziło bowiem niemal kompletne unicestwienie podczas dalszego użytkowania. Decyzja ta odbiła się najbardziej na studentach i pracownikach nauki, zajmujących się badaniami nad 2. połową XIX wieku i 1. połową wieku XX. Analizując zamówienia czytelników z okresu przed wprowadzeniem zarządzenia o wy-

cofaniu zniszczonego zasobu z udostępniania i analogiczny czas po zamknięciu dostępu do części zbiorów, odnotowano wzrost rewersów odmownych z 4.008 (styczeń-czerwiec 2007 r.) do 5.823 (styczeń-czerwiec 2008 r.). Decyzję podjęto, wiedząc jak trudne byłoby zdobycie kolejnych egzemplarzy celem zrobienia kopii mikrofilmowej lub cyfrowej.

Wykonanie kopii jest podstawową metodą zabezpieczania zbiorów, obecnie powszechnie stosowaną. Mając świadomość, że lista wyłączonych z udostępniania zbiorów będzie się zwiększać, w Bibliotece Jagiellońskiej od wielu lat systematycznie tworzone są kopie mikrofilmowe. W ten sposób zabezpieczona jest jedna z najważniejszych gazet polskich krakowski „Czas”², ukazujący się w latach 1848-1939. Dzięki wykonaniu mikrofilmu wyłączony z udostępniania oryginał został poddany procesowi odkwaszania oraz innym zabiegom konserwatorskim. Również inne istotne tytuły, jak choćby „Ilustrowany Kurier Codzienny” (Kraków 1911-1939) czy „Naprzód” (Kraków 1892-1948), zostały skopiowane na mikrofilmy. Popyt na korzystanie z gazet i czasopism nieustannie rośnie, ponieważ wiele prac naukowych powstaje na podstawie informacji w nich zawartych. Dlatego wycofanie z obiegu szeregu tytułów prasowych ogranicza działalność Biblioteki, jako warsztatu pracy intelektualnej. Zapotrzebowanie na informacje z czasopism i gazet wzrasta także i z tego powodu, że coraz większa rzesza regionalistów i hobbystów szuka

² Obecnie zdigitalizowany i udostępniany przez Małopolską Bibliotekę Cyfrową.

w nich choćby krótkich wzmianek o swoich „małych ojczyznach”. Tylko utworzenie cyfrowych kopii umożliwi szerokie udostępnienie tego, jakże cennego dziedzictwa kulturowego, mieszkańcom nie tylko naszego regionu.

Największym problemem podczas selekcji był nadmiar tytułów, które nie tyle kwalifikowały się do digitalizacji, ile wymagały natychmiastowego wykonania kopii. Skupiono się na wydawnictwach ukazujących się w Krakowie i we Lwowie. Ośrodki te były największymi centrami życia politycznego, kulturalnego i społecznego Galicji. Czasopisma te cieszyły się niesłabnącym zainteresowaniem wśród czytelników, co negatywnie odbiło się na ich stanie zachowania. Część woluminów od wielu lat była przechowywana w pozycji poziomej i wyłączona z użytkowania. Stan niektórych można określić jako katastrofalny i ocalenie treści dla potomnych było zadaniem bardzo pilnym. Wybrano siedem tytułów, zakładając, że pozostałe gazety będą typowane do digitalizacji w każdym następnym projekcie. Wśród nich znalazły się „Gazeta Lwowska” (Lwów, 1811-1939), „Nowa Reforma” (Kraków, 1882-1927), „Gazeta Narodowa” (Lwów, 1862-1915), „Dziennik Polski” (Lwów, 1869-1905), „Głos Narodu” (Kraków, 1893-1927), „Słowo Polskie” (Lwów, 1895-1934), „Nowości Ilustrowane” (Kraków, 1904-1925). Nie są to jedyne tytuły czasopism o największym znaczeniu dla kultury polskiej. Skupiono się jednak na gazetach galicyjskich,



Il. 2.

przyjmując, że gazety wydawane na innych terenach Polski, zostaną poddane digitalizacji przez odpowiednie ośrodki.

Druki Sejmu Krajowego Galicyjskiego idealnie wpisują się w zasadę digitalizowania zbiorów najbardziej poczytnych, najbardziej zniszczonych i unikalnych. Wydawnictwa sejmowe ukazujące się w latach 1861-1914 prezentują nie tylko ustawodawczą aktywność reprezentantów kraju, ale również bardzo istotne dane statystyczne. Ukazywały się jako „Sprawozdania Sejmu Krajowego Galicyjskiego”. Ich uzupełnieniem były „Protokoły Sejmu Krajowego Galicyjskiego” oraz „Alegata do Sprawozdań Sejmu Galicyjskiego”. Łącznie tytuły te obejmują 225 tomów i tworzą niezwykłą

Il. 2. Przechowywanie zeskanowanych i odkwaszonych oryginałów w pudłach bezkwasowych.

kolekcję dokumentującą życie polityczne Galicji. W tak kompletnej formie znajdują się one już tylko w Wiedniu. Nawet zbiory lwowskie mimo, iż tam mieściła się siedziba sejmu krajowego, nie obejmują całości tych wydawnictw. Stopniowo digitalizacji poddawane są druki związane z sejmem i działalnością poselską (Provizoryczny regulamin dla Sejmu krajowego Królestwa Galicji i Lodomeryi wraz z..., Wykaz Członków Sejmu krajowego Królestwa Galicji i Lodomeryi tudzież Wielkiego Księstwa Krakowskiego w roku... i in.).

Powstawanie bibliotek cyfrowych, tworzenie kolekcji, przy odpowiednim podejściu, daje bardzo duże możliwości współpracy między ośrodkami. W ramach projektu Jagiellońska Biblioteka Cyfrowa zeskanowano trzy tytuły („Przyjaciół Dzieci”, „Pszczelarz i Ogrodnik”, „Robotnik Polski”), które są dodatkami do „Dziennika Berlińskiego” (Berlin 1897-1939). Tytuł ten jest planowaną publikacją Wielkopolskiej Biblioteki Cyfrowej. Jednak porównując zasoby okazało się, że egzemplarz z Biblioteki Jagiellońskiej jest bardziej kompletny, wobec czego podjęto decyzję o zeskanowaniu całego wydawnictwa. Przykład ten pokazuje, jak dziś można stworzyć cyfrową kolekcję ze zbiorów wielu bibliotek. W przypadku czasopism galicyjskich świetny efekt przyniosłaby współpraca Biblioteki Jagiellońskiej, Austriackiej Biblioteki Narodowej w Wiedniu i Lwowskiej Narodowej Naukowej Biblioteki Ukrainy im. W. Stefanyka.

Przeciwieństwem świadomego tworzenia kolekcji cyfrowych jest podejście bibliotek, które przemyślny wybór tytułów zastąpiły kopiowaniem czasopism według kolejności na półkach. Nie dba się wówczas o kompletność, co jest szczególnie ważne przy dłuższej ukazujących się tytułach. Nie brane są pod uwagę realne potrzeby czytelnika.

Zasady jakimi kierowano się podczas selekcji wydawnictw ciągłych do Jagiellońskiej Biblioteki Cyfrowej można ująć w kilku punktach:

- unikatowość poszczególnych tytułów,
- stan zachowania najbardziej poczytnych tytułów,
- kompletność typowanych czasopism.

Kryteria te są konsekwencją wieloletnich prac bibliotecznych nad wydawnictwami ciągłymi znajdujących się w zasobach Biblioteki Jagiellońskiej. Ponieważ zasady te obowiązują przy wnioskowaniu o dotacje w nowych projektach i przy bieżącej digitalizacji, istnieje duże prawdopodobieństwo stworzenia bardzo ciekawej kolekcji czasopism z terenu dawnej Galicji. Planowana jest również digitalizacja znaczącej ilości poloników, a następnie prasy podziemnej z lat 1939-1945, w tym oczywiście część stanowiącą będą jedyne egzemplarze.

Biblioteka Jagiellońska, przyjmując zasady obowiązujące przy selekcji cza-

sopism do cyfryzacji, dąży świadomie do powstania najciekawszej kolekcji wydawnictw ciągłych z okresu XIX wieku i 1. połowy wieku XX, kontynuując niejako tradycje swoich wielkich poprzedników, dzięki którym zawsze była postrzegana jako Biblioteka wyjątkowa.

Bibliografia

Baumgart, Jan (1964). Karol Estreicher jako bibliotekarz. [W:] Księga pamiątkowa ku czci Karola Estreichera (1827-1908). Kraków: [s. n.], s.133-198.

Dybiec, Julian (2000). Karol Estreicher. [W:] Dybiec, Julian (red.) Złota Księga Wydziału Historycznego. Kraków: Księgarnia Akademicka, s.39-49.

Gaberle, Eustachy (1928). Egzemplarz obowiązkowy na ziemiach polskich dawniej i dziś z opracowanym zbiorem artykułów, ustaw i rozporządzeń. Szkic historyczno-systematyczny. Lwów: Zakład Narodowy im. Ossolińskich.

Sobucki, W.; Koziński, Z.; Drewniewska-Idziak, B. (2004). Stan zachowania księgozbioru z XIX i XX wieku w Bibliotece Jagiellońskiej w Krakowie. „Notes Konserwatorski”, nr 8, s.182-198.

Szalla-Kleemann, Aleksandra; Szyba, Marian (2001). Wstępna ocena stanu zachowania czasopism krakowskich z lat 1860-1890 w zbiorach Biblioteki Jagiellońskiej. [W:] Łojewski, Tomasz; Pietrzyk Zdzisław (red.) Kwaśny papier. Zagrożenie zbiorów bibliotecznych i archiwalnych. Kraków: Biblioteka Jagiellońska, s.169-179.

Naprawa i zabezpieczenie czasopism przeznaczonych do skanowania



Słowa kluczowe: digitalizacja, Biblioteka Jagiellońska, czasopisma, Jagiellońska Biblioteka Cyfrowa, konserwacja, opakowania ochronne, skanowanie, zabezpieczanie zbiorów bibliotecznych

Keywords: digitization, Jagiellonian Library, periodicals, Jagiellonian Digital Library, conservation, protective containers, scanning, preservation of library collections

Cały zbiór czasopism z XIX- i pierwszej połowy XX- wieku przeznaczony w projekcie do digitalizacji został poddany ocenie stanu zachowania przez konserwatorów Biblioteki Jagiellońskiej. Przegląd konserwatorski wykazał, że większość egzemplarzy jest poważnie uszkodzona, a w części z tych woluminów były pojedyncze numery w całości lub w dużej części silnie zniszczone. Uszkodzenia zbioru czasopism spowodowane były przede wszystkim tym, że

papier był tak słaby, że przy próbie przewracania karty pozostawał jej fragment. Dlatego też skierowanie tych czasopism do digitalizacji w projekcie było ostatnim momentem zachowania ich treści. Do najczęściej spotykanych uszkodzeń obiektów, które uniemożliwiałyby ich skanowanie, były:

1. przedarte karty,
2. zasłonięty tekst strony spowodowany przez:
 - a) nieprawidłowe zszywanie podczas oprawy woluminu czasopisma,
 - b) zaklejenie tekstu zwykłym papierem podczas wcześniejszych napraw introliigatorskich (w przeszłości nie było materiałów typu „bibułka japońska”, pozwalających na łączenie rozerwanych kart bez zasłaniania tekstu),
 - c) zagięcia i pofałdowanie papieru,

d) zbyt zwarte zszywanie bloku woluminu (zastosowanie przy oprawie tzw. zszywania bocznego, które uniemożliwiało odpowiednie otwarcie tomu i nie pozwalało na jego prawidłowe skanowanie) i jednocześnie jego nadmierna grubość, co uniemożliwiało odczytanie całości tekstu.

W związku z takim stanem rzeczy, aby skanowanie zaplanowanych do digitalizacji czasopism było możliwe, należało wcześniej poddać je zabiegom konserwatorskim. Ustalono procedury obiegu obiektów wymagających napraw. Po ocenie konserwatorskiej, osoba pracująca na rzecz projektu w Oddziale Magazynów BJ, która była odpowiedzialna za przeglądanie obiektów pod kątem ich kompletności, miała również za zadanie kierowanie uszkodzonych czasopism do konserwacji/naprawy w Sekcji Opaw i Zabezpieczenia Zbiorów. W typowaniu zbiorów do konserwacji kierowała się:

- a) wskazaniami konserwatorów zawartymi w ankietach sporządzonych podczas oceny stanu zachowania poszczególnych woluminów,
 - b) własną obserwacją w trakcie przeglądania zawartości obiektów.
- Każdy wolumin zakwalifikowany do naprawy zaopatrzony był w szczególności uwagi dla introliigatorów, zawierały one wskazówki co do skali i zakresu konserwacji. Przekazane do Sekcji Opaw i Zabezpieczenia Zbiorów obiekty były naprawiane/konserwowane przez pra-

owników Sekcji. Były to dwie osoby opłacane ze środków projektu, kierownik sekcji pracujący dwa i pół roku w wymiarze 25 godzin miesięcznie i technik-konserwator książki, pracująca w przeciągu tego samego okresu, przez dziewięć miesięcy zatrudniona w wymiarze ½ etatu, następnie na pełny etat. Osoby te naprawiały również obiekty, które uległy uszkodzeniu podczas skanowania oraz wykonywały opakowania ochronne (fascykuly) dla zbioru 320 woluminów jedy-

Abstrakt: W artykule opisano prace dot. projektu Jagiellońska Biblioteka Cyfrowa, związane z naprawą i zabezpieczeniem czasopism przeznaczonych do skanowania, które wykonywane były w Sekcji Opaw i Zabezpieczania Zbiorów Oddziału Zabezpieczenia Zbiorów Biblioteki Jagiellońskiej. Na wstępie przedstawiono przykłady najczęściej spotykanych uszkodzeń. Omówiono procedury kwalifikacji obiektów do naprawy. Zajęto się również skalą napraw i jej wpływem na tempo konserwacji i wykonania opakowań ochronnych (pudła, teuczki) na zdigitalizowane zbiory. Na końcu przedstawiono dane statystyczne obrazujące skalę prac wykonanych na rzecz projektu.

Abstract: The article describes the activities of the Jagiellonian Library's Binding and Preservation Section relating to the Jagiellonian Digital Library Project. Examples of the most common kinds of damage to library items are presented in the first place. The procedures of qualifying items for conservation are discussed next. It is indicated that the scope of necessary repairs affects heavily the time needed to complete the entire process and produce protective containers (boxes, folders) for the materials scanned for the Jagiellonian Digital Library Project. The statistical data concerning the total amount of work done within the project are also provided.





II. 1.

II. 1-3. Przykłady zniszczeń.



II. 2.



II. 3.



II. 4.

II. 4. Wspomniany egzemplarz nie nadający się do skanowania.

nych w Polsce egzemplarzy czasopism IV i V formatu. Po pięciu miesiącach intensywnych napraw okazało się, że rzeczywista ilość woluminów kierowanych do konserwacji, a przede wszystkim zakres wykonywanych przy nich prac, był dużo większy niż zakładano w planach. W tej sytuacji stało się konieczne oddelegowanie dodatkowo do tego zadania trzech pracowników Sekcji, mających wieloletnie doświadczenie w pracy intrologatorskiej i naprawie uszkodzonych gazet (jedna z nich pracowała łącznie pełny rok, druga 11 miesięcy, trzecia zaś 2 miesiące). Nie udało się naprawić tylko jednego woluminu Nowej Reformy (zawierającego numery z pierwszej połowy 1904 r.), z powodu zniszczenia spowodowanego postępującą degradacją papieru (kwaśny papier) oraz niewłaściwym użytkowaniem. O skali napraw niech świadczy fakt, że z przewidzianych do skanowania ponad 1220 woluminów czasopism poddano naprawie 1137. Dzięki tym działaniom możliwe stało się wykonanie wszystkich napraw w zaplanowanym terminie. Do najbardziej zniszczonych tytułów należały: „Nowości Ilustrowane”, „Głos Narodu”, „Gazeta Lwowska”, „Reforma/Nowa Reforma”, „Dziennik Polski”, „Słowo Polskie” i tytuły z grupy jedynek w Polsce egzemplarzy czasopism XIX i 1. poł. XX wieku. Wszystkie egzemplarze „Nowości Ilustrowanych” (22 wól.), które skierowane zostały do skanowania, trafiły wcześniej do naprawy. W wymienionych wcześniej tytułach proporcje te kształtowały się następująco: „Głos Narodu” (ponad 96% - 88 wól.

z 91), „Słowo Polskie” (77%), „Reforma/Nowa Reforma” (70%), „Dziennik Polski” (49%) i jedyne w Polsce egzemplarze czasopism (blisko 100%).

W ramach projektu w Sekcji Opaw i Zabezpieczania Zbiorów wykonane zostały opakowania ochronne (pudła, teuczki, koperty) na zbiory specjalne, czasopisma i druki zwarte XIX- i XX- wieku. W produkcji tych opakowań zabezpieczających wykorzystano ploter obsługiwany przez zatrudnionego w sekcji technika. Do naprawy zbiorów i wytwarzania opakowań zabezpieczających zakupiono ze środków projektu materiały najwyższej jakości, spełniające wszelkie wymagania konserwatorskie, potwierdzone odpowiednimi certyfikatami.

Skanery i skanowanie



Słowa kluczowe: pozyskiwanie obrazu, skanery, digitalizacja

Keywords: image acquisition, scanners, digitization

72

W ramach projektu Jagiellońska Biblioteka Cyfrowa zostały zakupione trzy skanery wielkoformatowe Zeutschel Omniscan 14000. Urządzenia z tej serii przystosowane są do skanowania dokumentów zszytych i oprawnych, zbiorów bibliotecznych i archiwalnych, map, atlasów, grafik, starodruków [Digital Center 2012]; posiadają dzielony stół z regulacją wysokości poszczególnych części, dzięki czemu są przystosowane do skanowania nawet bardzo dużych woluminów; szyba

dociskowa, umożliwia rozprostowanie obiektu w celu zminimalizowania zniekształceń geometrycznych. OS 14000 umożliwia dwukierunkowe skanowanie oraz usługę ROI-scan, która pozwala na skanowanie tylko wybranego fragmentu powierzchni roboczej, co znacznie zwiększa szybkość pracy w porównaniu do poprzednich rozwiązań. Według zapewnień producenta urządzenia te nie emitują szkodliwego dla zbiorów promieniowania, zapewniają natomiast wierne odwzorowanie barw [Zeutschel 2012]. Urządzenia Omniscan wykonują skany w rozdzielczościach do 600 ppi (w przypadku wariantu A0: do 400 ppi).

W przypadku sporządzania kopii cyfrowych obiektów materialnych w celu ochrony ich przed zniszczeniem wskazane jest zapewnienie wystarczająco wierne odwzorowania, aby substytut mógł zastąpić oryginał w większości zastosowań. Jedną z najtrudniejszych do przeniesienia cech obiektu jest informacja o kolorze. Oczywiście nie zawsze zachowanie

informacji o barwach obiektu źródłowego jest konieczne: nie wszystkie materiały są równie wartościowe, nie wszystkie wymagają wysokiej dokładności odwzorowania. O stosowaniu ograniczeń w jakości kopii cyfrowych decydują także inne czynniki, takie jak możliwość dostępnego sprzętu, trwałość przetwarzanych materiałów i związany z nią czas na wykonanie kopii. Zachowanie informacji o barwach jest natomiast ważne w przypadku dokumentów, dla których kolor jest istotnym składnikiem ich treści. Wybrane do projektu Jagiellońska Biblioteka Cyfrowa XIX- i XX-wieczne czasopisma nie wymagają wielkiej dokładności odwzorowania, dlatego zakupione do projektu skanery są adekwatne do przeprowadzenia digitalizacji tego materiału, gdyż w tym wypadku najistotniejsza jest szybkość działania.

Proces skanowania rozpoczyna się od przygotowania dokumentu oraz urządzenia do pracy. Dokument jest umieszczony na stole skanera, który dopasowywany jest do kształtu obiektu, w celu zapewnienia wyrównania powierzchni skanowanej; szyba dociskowa natomiast wygładza zagięcia i nierówności powierzchni strony. Takie przygotowanie jest konieczne do uzyskania odwzorowania bez zniekształceń geometrycznych. Stosowane są także ciemne przekładki minimalizujące przebiecia druku z odwrotu strony. Następnym krokiem jest dobór parametrów skanowania i zapisu plików wynikowych, a także ustawienie obszaru skanowania. Generalnie

wykonywane są skany rozkładówkowe (dwie strony dokumentu na jednym skanie) w celu zmniejszenia rozmiaru plików wynikowych oraz zapewnienia lepszego odwzorowania dokumentacyjnego.

Wynikiem skanowania są składki zawierające bitmapy w formacie TIFF bez kompresji lub z kompresją LZW. Po zeskanowaniu składki są przekazywane pracownikom weryfikującym kompletność oraz jakość obrazu, a następnie

Abstrakt: W przypadku sporządzania kopii cyfrowych obiektów materialnych w celu ochrony ich przed zniszczeniem wskazane jest zapewnienie wystarczająco wiernego odwzorowania, aby substytut mógł zastąpić oryginał w większości zastosowań. Artykuł przedstawia sprzęt zakupiony do projektu Jagiellońska Biblioteka Cyfrowa oraz proces skanowania.

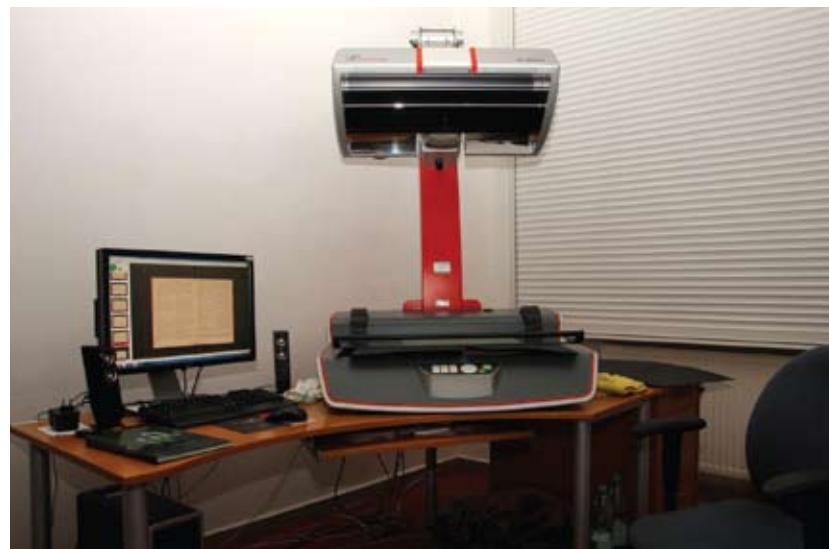
Abstract: Creating digital copies of library items in the purpose of preserving them requires maintaining the quality of scanned images at a level sufficient to allow a digital copy to substitute the original item in most cases. The paper presents the equipment purchased for the Jagiellonian Digital Library Project as well as the relevant scanning procedures.



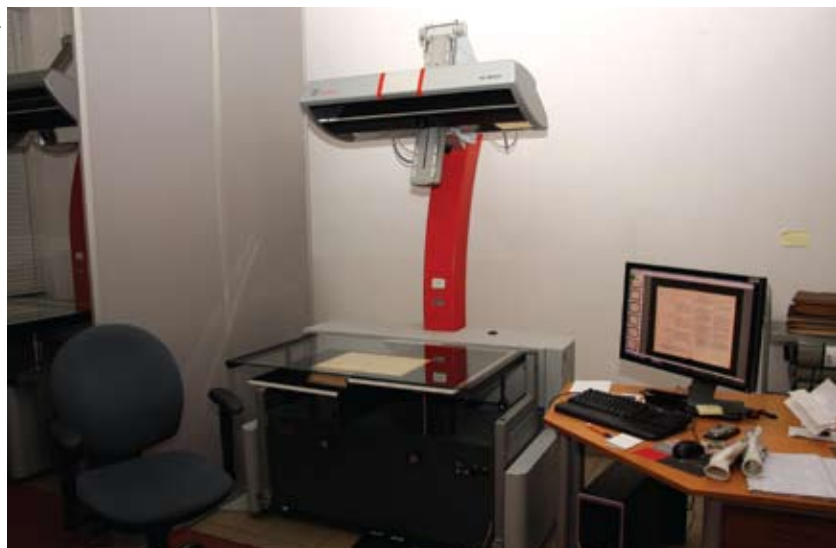
73



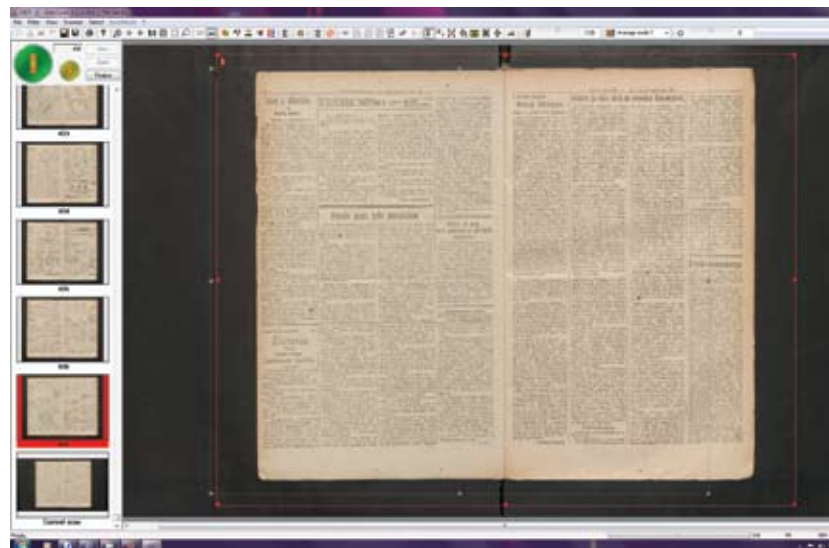
II. 1.



II. 3.



II. 2.



II. 4.

II. 1. Zeutschel OS 14000 A0 (fot. Szymon Kotarski).

II. 2. Zeutschel OS 14000 A1 (fot. Szymon Kotarski).


II. 3. Zeutschel OS 14000 A2 (fot. Hubert Tatarzyński).

II. 4. Zrzut ekranu oprogramowania Omniscan. Skan rozkładówkowy oraz obszar skanowania.

w zależności od wyniku oceny: korygowane i przekazywane ponownie lub w postaci niezminionej archiwizowane na taśmach.

Źródła

Zeutschel [online]. [Dostęp: 20.11.2012]. Dostępny w World Wide Web:

 <http://zeutschel.com>.

Digital-Center [online]. [Dostęp: 20.11.2012]. Dostępny w World Wide Web:

 <http://http://digital-center.pl>.

Kopiowanie czasopism

Jednym z głównych zadań projektu było wykonanie kopii cyfrowych i mikrofilmowych zbiorów przeznaczonych do digitalizacji. Zadania te były realizowane w pracowniach Sekcji Reprografii Oddziału Zabezpieczenia Zbiorów Biblioteki Jagiellońskiej.

Na potrzeby projektu zakupiono trzy profesjonalne skanery (do formatów A0, A1 i A2), konwerter – urządzenie do naświetlania plików cyfrowych na błonę mikrofilmową, wywoływarke, stół montażowo-kontrolny do kontroli jakości, wywołanych filmów i obróbki finalnej. Poza tym zostały zakupione chemikalia do wywoływania filmów oraz pudełka z bezkwasowej tektury do ich konfekcjonowania. Wspomniane skanery i urządzenia do mikrofilmowania umożliwiają szybkie skanowanie i wykonywanie kopii mikrofilmowych.

Słowa kluczowe: digitalizacja, Biblioteka Jagiellońska, czasopisma, Jagiellońska Biblioteka Cyfrowa, kopie cyfrowe, mikrofilmy, mikrofilmowanie, skanowanie, skany.

Keywords: digitization, Jagiellonian Library, periodicals, Jagiellonian Digital Library, digital copies, microfilms, microfilming, scanning, scanned copies.

W ramach projektu do ich obsługi zatrudniono 10 osób, w tym do pracy przy skanowaniu 9 (w wymiarze 4 godzin dziennie) i 1 tworzącą kopie mikrofilmowe (początkowo pół roku pracującą 4, a następnie po 8 godzin dziennie). Zatrudnienie tylu osób w cyklu czterogodzinnym pozwoliło na efektywne wykorzystanie ich czasu pracy, powodując jednocześnie wydłużenie działania jednego skanera do 12- tu godzin dziennie.

Każdy skanowany w ramach projektu obiekt otrzymał zapis cyfrowy w forma-



cie TIFF, w rozdzielczości 300 dpi. Następnie kopia cyfrowa przesyłana była na serwer roboczy projektu, obsługiwany przez bibliotekarzy cyfrowych, którzy mieli za zadanie ocenić jakość i kompletność kopii cyfrowych. Jeżeli bibliotekarze cyfrowi wnosili uwagi, co do jakości i kompletności kopii, skanujący nanosili poprawki i ponownie przesyłali pliki do bibliotekarzy cyfrowych, którzy ostatecznie akceptowali poprawność skanowania. Tempo skanowania zbiorów,

które trafiły do pracowni reprografii uzależnione było od formatu i stanu zachowania oryginału. Daleko posunięta kwasowa degradacja papieru i uszkodzenia mechaniczne, przede wszystkim gazet dużego formatu, wymagały szczególnej ostrożności przy przewracaniu kart w procesie skanowania. Mając na uwadze to, jak i wielkość skanowanego obiektu, kopiowanie dziennie 3000 do 3400 stron należy uznać za bardzo dobry wynik.

W projekcie wykonano również kopie mikrofilmowe z plików cyfrowych zbiorów specjalnych oraz skanowanych czasopism i druków zwartych z XIX- i pierwszej połowy XX-wieku. Przygotowane przez bibliotekarza cyfrowego pliki dostarczano do pracowni reprograficznej, gdzie były naświetlane na błonę mikrofilmową za pomocą wspomnianego wcześniej konwertera. Następnym etapem pracy było jej wywoływanie i utrwalenie. Tak otrzymane mikrofilmy były przeglądane pod względem jakości i kompletności, a następnie umieszczone w pudełkach z tektury bezkwasowej. Z Sekcji Reprografii Oddziału Zabezpieczenia Zbiorów mikrofilmy trafiały do magazynu archiwalnego mikrofilmów, znajdującego się w Oddziale Dokumentów Audiowizualnych Biblioteki Jagiellońskiej.

Abstrakt: W artykule przedstawiono informacje dotyczące realizacji zadań projektowych wykonywanych w pracowniach Sekcji Reprografii Oddziału Zabezpieczenia Zbiorów Biblioteki Jagiellońskiej. Podstawowym zadaniem sekcji było wykonanie kopii cyfrowych z oryginałów i mikrofilmowych z kopii cyfrowych. Na wstępie przedstawiono sprzęt zakupiony do ich wykonania. Potem omówiono poszczególne etapy prac wykonywanych w procesie skanowania i mikrofilmowania. Opisano także wpływ stanu zachowania kopiowanych zbiorów na tempo skanowania. Pokazano współdziałanie między pracownikami wykonującymi kopie cyfrowe i mikrofilmowe a bibliotekarzami cyfrowymi.

Abstract: The article provides information on the contribution of the Jagiellonian Library's Reprographics Section to the Jagiellonian Digital Library Project. The main task of the section was to create digital copies of original items as well as microfilms out of digital files. The equipment purchased to carry out the task is presented. Next the consecutive stages of the scanning and microfilming processes are shown. It is also indicated that the condition of items intended for digitization affects substantially the amount of time and effort required to create copies. The cooperation between the staff working directly with scanners and digital librarians is highlighted.

Masowe odkwaszanie zbiorów

Jednym z etapów realizacji projektu Jagiellońska Biblioteka Cyfrowa było odkwaszanie oryginałów zbiorów Biblioteki Jagiellońskiej zakwalifikowanych do skanowania. Przeprowadzone zabiegi miały na celu powstrzymanie procesu kwasowej degradacji papieru.

Wykonanie tego zadania zostało powierzone Klinice Papieru Biblioteki Jagiellońskiej, która jest halą technologiczną, jedną z nielicznych na świecie, dysponującą jednocześnie dwiema instalacjami do masowego odkwaszania papieru – metodami Bookkeeper i bückeburską (Neschen).

Ze względu na rodzaj i specyfikę zbiorów wytypowanych do konserwacji, podjęto decyzję o wyborze metody Bookkeeper, w której odkwaszanie prowadzone jest w środowisku bezwodnym (perfluoroheptan), za pomocą drobnokrystalicznego tlenku magnezu (MgO). Dzięki temu metoda ta jest bezpieczna dla atramentów, farb, barwników oraz samego papie-

Słowa kluczowe: Biblioteka Jagiellońska, czasopisma, Jagiellońska Biblioteka Cyfrowa, Klinika Papieru Biblioteki Jagiellońskiej, konserwacja masowa, kwaśny papier, masowe odkwaszanie, technologie masowego odkwaszania, zbiory biblioteczne

Keywords: Jagiellonian Library, periodicals, Jagiellonian Digital Library, Paper Clinic, mass preservation, acid paper, mass deacidification, mass deacidification technologies, library collections

ru, co pozwala na zabezpieczenie niemal wszystkich rodzajów zbiorów. Ponadto charakteryzuje się ona również wysoką wydajnością, przy jednoczesnym niskim ryzyku powstawania uszkodzeń mechanicznych w trakcie procesu oraz – co jest szczególnie istotne – pozwala na odkwaszanie zarówno zbiorów luźnych jak i oprawnych.

W typowaniu zbiorów do masowego odkwaszania kierowano się wskazaniami



konserwatorów zawartymi w ankietach sporządzonych podczas oceny stanu zachowania poszczególnych woluminów.

Po przekazaniu zeskanowanych materiałów do Kliniki Papieru okazało się, że ogólny stan zakwalifikowanych zbiorów jest lepszy, niż mogłoby to wynikać z przeprowadzonej wcześniej oceny. Dotyczyło to zwłaszcza części woluminów zakwalifikowanych wstępnie do grupy materiałów, dla których zalecane było

odkwaszanie woluminu mieszane (maszynowe i ręczne). W trakcie przygotowywania ich do odkwaszania oceniono (nadzór konserwatorski i kierownik Kliniki Papieru), że spełniają one wymagania stawiane materiałom kierowanym do konserwacji masowej i mogą być poddane procesowi masowego odkwaszania, przy zachowaniu szczególnych środków bezpieczeństwa podczas przygotowania materiałów do samego procesu, ograniczając ryzyko uszkodzenia obiektów do minimum.

W toku przeprowadzanego odkwaszania pojawiały się również i problemy. Okazało się, że po etapie skanowania część zakwalifikowanych pierwotnie do obróbki mechanicznej zbiorów, znalazła się na granicy spełnienia kryteriów wymaganych dla bezpiecznego przeprowadzenia procesu. Pojawiające się liczne pęknięcia bloku, które destabilizują wolumin jako całość, potęgowały trudności wynikające ze zwiększonej kruchości papieru i znacząco utrudniały, a częstokroć uniemożliwiały odkwaszenie masowe niektórych egzemplarzy. Takie materiały musiały zostać poddane uprzednio żmudnej czynności rozszywania bloku, a następnie były umieszczane w kasetach przeznaczonych do zabezpieczania luźnych dokumentów. Odkwaszanie tym sposobem zapewniało większe bezpieczeństwo konserwowanych zbiorów, równocześnie jednak wpływało na spowolnienie tempa realizacji zadania.

W projekcie procesowi masowego odkwaszania poddanych zostało, przy spełnieniu podanych wyżej warunków, ponad 75% z ogólnej liczby wstępnie zakwalifikowanych obiektów, tj. ponad 1200 woluminów, których waga przekroczyła 2000 kg.



Il. 1.



Il. 2.



Il. 3.

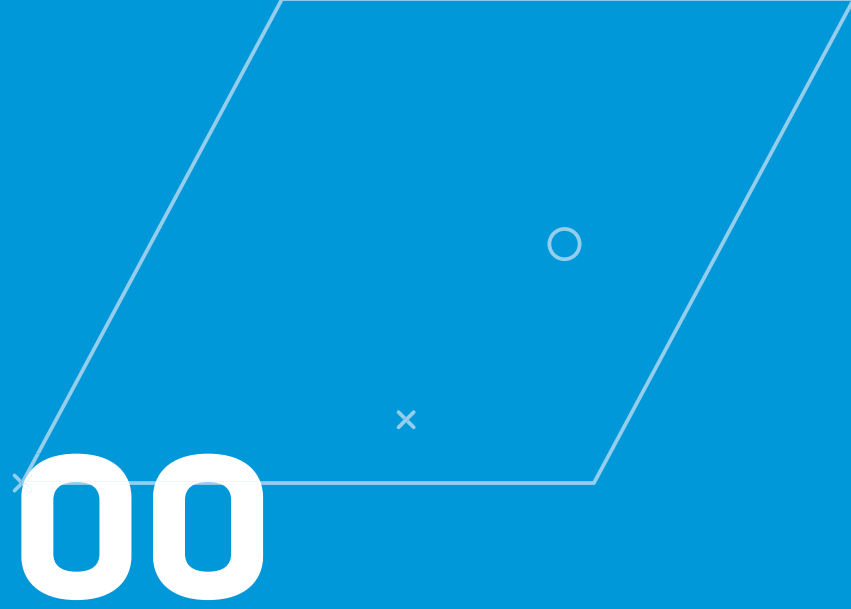
- Il. 1. Odkwaszanie luźnych części w kasetach w instalacji Bookkeeper (fot. Joanna Rościszewska).
- Il. 2. Instalacja do masowego odkwaszania metodą Bookkeeper (fot. Joanna Rościszewska).
- Il. 3. Wnętrze zbiornika z kąpielą odkwaszającą, stosowaną w metodzie Bookkeeper (fot. z archiwum Kliniki Papieru BJ).



Abstrakt: W tekście przedstawiono informacje dotyczące realizacji procesu masowej konserwacji zbiorów przeznaczonych do digitalizacji w ramach projektu Jagiellońska Biblioteka Cyfrowa. Krótko scharakteryzowano metodę wybraną do realizacji zadania oraz podano główne argumenty przemawiające za jej wyborem. Następnie omówiono przebieg realizacji zadania oraz główne problemy, pojawiające się w trakcie. Na zakończenie przedstawiono dane obrazujące efekty realizacji zadania.

Abstract: The text describes the process of mass conservation of collection items selected for digitization within the framework of Jagiellonian Digital Library Project. The method chosen for the task and the main arguments behind that choice are briefly characterized as well as the main issues encountered during the process. At the conclusion, the data on the results are presented.

Opracowanie obiektów cyfrowych



Od karty katalogowej poprzez OPAC do biblioteki cyfrowej



Słowa kluczowe: katalog kartkowy, karta katalogowa, opis bibliograficzny, MARC 21, Dublin Core

Keywords: printed catalog, catalog card, bibliographic description, MARC 21, Dublin Core

Z uwagi na potrzebę sprawnej organizacji pracy na wszystkich poziomach digitalizacji, konieczne było pozyskanie i uporządkowanie informacji o skanowanych obiektach. Zadanie pozyskania tychże informacji rysowało się jako jedno z trudniejszych do wykonania. Uzyskanie wstępnej wiedzy na temat opracowywanego czasopisma gwarantował katalog kartkowy i było to jedyne, oprócz samego czasopisma, źródło informacji. Stąd potrzeba rysy historycznego oraz

krótkiej charakterystyki katalogu alfabetycznego podstawowego czasopism w Bibliotece Jagiellońskiej.

Pierwszy systematyczny spis czasopism wydawanych na ziemiach polskich ukazał się w 1860 r. na łamach „Dodatku Tygodniowego” do „Gazety Lwowskiej”, a ogłoszony został nie przez kogo innego, jak przez Karola Estreichera. Jako późniejszy dyrektor Biblioteki Jagiellońskiej zadbał on, aby zarówno same czasopisma jak i ich tworzący się katalog znalazły się w Bibliotece. Ponownie zestawienie czasopism w porządku alfabetycznym ukazało się w pierwszym tomie „Bibliografii Polskiej XIX. stulecia”, które następnie ukazało się oddzielnie w Krakowie w 1871 r. pt. „1400 pism periodycznych”, gdzie autor przy poszczególnych tytułach zaznaczył, które znajdują się w Bibliotece Jagiellońskiej [Korpałowa 1957, s. 7].

W 1915 r. udostępniony został czytelnikom drukowany wykaz czasopism

w zbiorach Biblioteki Jagiellońskiej pt. „Periodica Bibliothecae C. R. Universitatis Jagiellonicae Cracoviensis” pod red. Fryderyka Papégo, obejmujący 779 tytułów, wśród których znalazła się znaczna ilość czasopism i dzienników polskich.

Kolejnym bibliotekarzem, który przyczynił się do rozwoju katalogu czasopism w Bibliotece Jagiellońskiej był Adam Bar. Od 1927 r. rozpoczął meliorowanie starego zasobu czasopism polskich, co oznaczało zebranie ich w jednym miejscu i na podstawie kart pozostałych po Karolu Estreicherze stworzenie nowych kart katalogowych. W efekcie swoich prac stworzył on katalog alfabetyczny i rzeczowy czasopism polskich obejmujący około 4 000 pozycji. Należy również podkreślić jego zasługi nad zainicjowaniem ogólnopolskich prac, które w efekcie przyczyniły się do powstania Centralnego Katalogu Czasopism Polskich, do którego przekazał materiały zgromadzone podczas prac w Bibliotece Jagiellońskiej [Korpałowa 1957, s. 12].

W 1945 r. katalog obejmował cały zasób przedwojennych czasopism BJ, liczył 11 000 kart formatu 25x20 cm i ułożony był luźno w 48 pudłach [Łagodowa 1975, s. 117]. W roku 1953 wydano okólnik w sprawie melioracji katalogu podstawowego czasopism. Ze względu na plany opracowania informatora „Katalog Czasopism Biblioteki Jagiellońskiej” rozpoczęto porządkowanie katalogu i przeszerogowanie haseł według „Skró-

conych przepisów katalogowania zbiorów” [Grycz, Borkowska 1946]. Praca polegała na przeglądaniu karta po karcie całego zasobu czasopism i podkreślaniu na czerwono właściwego hasła głównego. Dotychczas w katalogu hasłem był pierwszy rzeczownik (w mianowniku) z tytułu. Po przeszerogowaniu hasłem stał się pierwszy znaczący wyraz tytułu i tak przeszerogowane hasła umieszczano w nowym porządku alfabetycznym. Kolejną meliorację pod względem popraw-

Abstrakt: W artykule przedstawiono charakterystykę katalogu czasopism Biblioteki Jagiellońskiej, zwracając szczególną uwagę na różnice pomiędzy metadanymi zawartymi na kartach katalogowych, a stanem magazynowym. Uwypuklona została istota prac prowadzonych w Systemie Obsługi Jagiellońskiej Biblioteki Cyfrowej polegająca na rzetelnym i szczegółowym rozpisaniu każdego woluminu czasopisma oraz korzyściach jakie niesie za sobą tak drobiazgowy opis. Przedstawiono problemy jakie niesie ze sobą retrospektywne katalogowanie czasopism na poziomie ogólnopolskim, jakie korzyści i zagrożenia wynikają z faktu katalogowania niepełnych zasobów czasopism. Zwrócono uwagę na konieczność łączenia różnych narzędzi w celu ułatwienia użytkownikowi dostępu do pełnotekstowych zasobów, zarówno na poziomie lokalnym, ogólnopolskim jak i międzynarodowym.

Abstract: The Jagiellonian Library's card catalog of periodicals is presented briefly. Awareness of possible discrepancies between the metadata provided on printed catalog cards and the actual library items on the storage stacks is particularly stressed. Meticulous entering of content of each periodical volume processed for the Jagiellonian Digital Library is highlighted as well as purposefulness of collecting so detailed data and using them in the publication process in a digital library. Problems and advantages of retrospective bibliographic description of periodicals, particularly those with incomplete runs, in the National Universal Catalog are presented. The need of interlinking bibliographic and full-text online resources on the local, national, and international level is also asserted.



	560 V CZASOP	Lwów Druk. K. Pillera 1863-	
Gazeta	Gazeta narodowa. Red. odp. K. Stupnicki. Główny współprac. Jan Dobrzański.		
	R. I - 1863 nr. 3-268 (1 grudzień)		
	R. II - 1864 " 1-287 ² (30 grudnia)		st. zas. 1927
	" III - 1865 " 1-298		
	" IV - 1866 " 1-299		tekt.
	" V - 1867 " 1-301		
	" VI - 1868 " 1-301 ZN		
	" VII - 1869 " 1-344		wol. 57
	" VIII - 1870 " 1-336 ZN		

II. 1.

ności haseł przeprowadzono w 1957 r., a kolejną przy okazji przepisywania kart do tworzącego się Katalogu Czasopism dla Publiczności w latach 1968-1971. W wyniku akcji tworzenia katalogu dla publiczności liczba kart zwiększyła się do 37 000 i zajęła 180 szufladek w 3 szafach [Łagodowa 1975, s. 120].

Przechodząc do analizy danych zawartych na kartach katalogowych, pierwsze wrażenie było bardzo pozytywne i zapowiadało w miarę kompletny zasób większości czasopism. Wypisane były poszczególne lata w obrębie danego tytułu wraz z uwzględnieniem braków w obrębie danego roku. Stan rzeczy najlepiej został odzwierciedlony na il. 1.

Po sięgnięciu do woluminów okazało się, że często niezanalizowano szczegółowo zawartości każdego z tomów i braki numerów nie zawsze zostały wynotowane na kartach katalogowych. Stąd opieranie się w trakcie procesu retrospektywnego opracowania czasopisma tylko na karcie katalogowej było dalece niewystarczające. Podkreślała to już Franciszka Korpałowa określając to jako „największą bolączkę wszystkich wielkich bibliotek polskich [...], że nie tylko w kompletach czasopism, ale nawet w poszczególnych rocznikach lub tomach mają one wielkie luki, nie zawsze wykazane w kartach katalogowych” [Korpałowa 1957, s. 14]. Dlatego w ramach projektu zdecydowano się na bardzo ważny krok mający na

celu szczegółowe rozpisanie zawartości każdego oprawnego woluminu czasopisma. Dla porównania na il. 2 przedstawiony został zapis powstały w trakcie drobiazgowego przejrzania woluminu i wynotowania wszelkich uwag mogących mieć wpływ na reedycję czasopisma w postaci cyfrowej. Początkowo poszerzony opis poszczególnych woluminów był wprowadzany do rekordu zasobu w Komputerowym Katalogu Zbiorów Bibliotek UJ, jednakże szybko okazało się, że nie jest to odpowiednie miejsce na tego typu rozległe uwagi. Dlatego do wprowadzania takich uszczegółowionych opisów został przystosowany system zarządzania całym procesem digitalizacji, czyli System Obsługi Jagiellońskiej Biblioteki Cyfrowej¹. W rekordzie zasobu ograniczono się tylko do informacji o brakach numerów, odzwierciedlając zapis z karty katalogowej uzupełniony danymi pozyskanymi w trakcie prac nad każdym z woluminów.

Celem, jaki przyświecał tworzonej Jagiellońskiej Bibliotece Cyfrowej, było zapewnienie użytkownikowi przyjaznego środowiska do przeglądania starych czasopism. Aby tego dokonać, przed pracownikami Oddziału Opracowania Czasopism jak i pracownikami Oddziału Magazynów stanęło bardzo niewdzięczne i żmudne zadanie pozyskania możliwie pełnej informacji na temat zawartości każdego woluminu czasopisma. Tylko dzięki rzetelnej informacji proces digitalizacji i „reedycji” czasopisma, w którego skład wchodziły oprócz nume-

rów głównych także: wydania poranne i wieczorne, dodatki samoistne i niesamoistne, przeróżnego rodzaju wkładki, dokumenty życia społecznego itp. był możliwy do przeprowadzenia w takiej skali, jak miało to miejsce w projekcie Jagiellońska Biblioteka Cyfrowa. Niejednokrotnie zdarzało się, że w danym woluminie znajdowały się numery zupełnie innych czasopism, lub błędnie oprawione dodatki do innych czasopism. Same dodatki były bardzo niekonsekwentnie włączane do woluminu. Można było spotkać się z następującymi sytuacjami: dodatek włączony zaraz po numerze głównym, dodatki zebrane razem i oprawione na końcu woluminu, dodatki oprawione w odrębnym woluminie, a zdarzało się, że były one wydzielone w odrębną sygnaturę. Powiązanie wszystkich dokumentów towarzyszących czasopismu w czasie jego publikacji², wymagało pogłębionej analizy czasopisma.

Kolejnym etapem, bezpośrednio wynikającym z gruntownego opracowania poszczególnych tytułów było utworzenie dla nich rekordów bibliograficznych w ogólnopolskiej bazie NUKAT. Szczegółowa analiza historii ukazywania się czasopisma ułatwiała katalogującym zawarcie w jednym rekordzie wszystkich niezbędnych informacji. Spowodowało to rozbudowanie opisów, w których zostali ujęci wszyscy redaktorzy naczelni odpowiedzialni za czasopismo w trakcie jego ukazywania się, wszystkie dodatki samoistne i niesamoistne. Zgodnie z przepisami katalogowania [Nałęcz 2007]

II. 1. Karta katalogowa dla czasopisma z Katalogu Podstawowego (tzw. „Starego”) Biblioteki Jagiellońskiej.

¹ Więcej na temat SOJBC w artykule Adriana Wiechcia „System Obsługi Jagiellońskiej Biblioteki Cyfrowej”.

² Zdigitalizowanym i opublikowanym w JBC czasopismem, które ukazywało się najdłużej była „Gazeta Lwowska”. Ukazywała się w latach 1811-1939, a więc przez 128 lat i w tym czasie miała kilku redaktorów, 10 dodatków samoistnych, nie licząc dodatków niesamoistnych, okazjonalnych wkładek, dokumentów życia spo-

lecznego itp. Późniejsze ujęcie całej historii czasopisma ukazującego się przez tak długi okres siłą rzeczy generuje wydłużenie opisu do nieporównywalnie wielkich rozmiarów w zestawieniu z opisami dla czasopism ukazujących się współcześnie. Dlatego na forum ogólnopolskim zdecydowano się na rozszerzenie opisów dla czasopism XIX-wiecznych o dane niezbędne do prawidłowej reedycji czasopism w bibliotekach cyfrowych.



II.2.

1863	<p>1863 R. 2 nr 2(3 stycznia)-263(31 grudnia); br. nr. 1, 7-8, 22, 35, 114, 135, 164, 174, 176, 194; w nr. 106 br. s. 1-2, w nr. 219 br. s. 5-7, UWAGI: nr 21 z 30 stycz a 23 z 10 marca(w lutym gazeta nie ukazywała się); nr 27 jest z 14 marca i podwójny nr 27-28 jest z 15 marca; nr 97 z 10 czerwca(to faktycznie nr 98, już poprawiony ółówkiem); po nr 119 z 4 lipca wkładka [1s.]- [Odezwa do Narodu] <i>Rząd Narodowy prawdziwie... Dnia 3go Lipca[?]</i> 1863; nr 159 jest z 21 sierp i podwójny 159-160 z 22 sierpnia; nr 172 z 5 września i podwójny 172-173 z 6 września; brak nr. 174-jest tylko dodatek; po nr. 180 z 16 września jest nr 181(z błędem. druk, poprawiony już przez kogoś kredką); nr podwójny 181-182 z 18 września; w nr. 191 s. 3-4 za odwrotnie wklejonymi 6-5; w nr. 201 s. 3-4 chyba nie z tego nr.; w nr. 211 s. 5-6 przed s.tyt.; 214 z 28 paźdz.(to faktycznie nr 215, poprawiony przez kogoś ółówkiem); w nr. 243 s. 3-4 za 6; DODATKI do nr.: o odrębnej od gazety paginacji: "Dodatek do nr ...", w zasadzie o osobnej paginacji, stanowi treściowe poszerzenie rubryk okazyjnego dnia; w l.1863-1869 w każdą niedzielę regularnie zastępuje numer nie ukazujący się generalnie w poniedziałki, dlatego wynotowane wszystkie numery" 19(2s., własna paginacja); 44 [uzup. treści GN]; 80 (4s.) [uzup. treści GN]; po nr 119 z 4 lipca wkładka [1s.]- [Odezwa do Narodu] <i>Rząd Narodowy prawdziwie... Dnia 3go Lipca[?]</i> 1863; 132[1 s.] [uzup. treści GN]; 138 (2s.) [uzup. treści GN]; 149 (2s.) [uzup. treści GN] m.in. <i>Cesarz, Polska i Europa</i>; 155(2s.) - w dod. być może br. s. 3-4, ponieważ brakuje stopki red.; 174(2s.) [uzup. treści GN]; 191 [uzup. treści GN] (2s.); 243[uzup. treści GN] (2s.);</p>
------	---

dla dodatków samoistnych zostały stworzone odrębne rekordy bibliograficzne połączone z rekordami dla czasopism macierzystych za pomocą pól powiązań. Dodatki oraz dobór redaktorów w rekordzie bibliograficznym w formacie MARC 21 zostały przedstawione na il. 3.

Taka rozbudowana struktura czasopisma zaważyła na sposobie prezentacji czasopism w bibliotece cyfrowej, gdzie dodatki samoistne zostały wydzielone

osobno, ale w obrębie tytułu głównego czasopisma, co zostało przedstawione na il. 4.

Z kolei powiązania, jakie istniały między tytułami czasopism, jak tytuły następne, tytuły poprzednie czy mutacje, uwzględnione w rekordach bibliograficznych znalazły swoje odzwierciedlenie w polach powiązań na poziomie opisu w formacie Dublin Core w JBC³. Kryteriami selekcji tytułów czasopism do projektu była

³ Szerzej o formacie Dublin Core i relacjach w tym zakresie zastosowanych w artykule Piotra Myszowskiego „Metadane”.

II. 2. Opis jednego woluminu czasopisma w SOJBC.

700	1	↳ Kratter, Franciszek [d (ca 1760-1838)] 1e Red.
700	1	↳ Łoziński, Władysław [d (1843-1913)] 1e Red.
700	1	↳ Kreczowiecki, Adam [d (1850-1919)] 1e Red.
700	1	↳ Rosowski, Stanisław [d (1861-1940)] 1e Red.
700	1	↳ Benza, Antoni [d (1787-1859)] 1e Red.
700	1	↳ Michałowicz, Mikołaj [d (1792-1846)] 1e Red.
700	1	↳ Kamiński, Jan Nepomucen [d (1777-1855)] 1e Red.
700	1	↳ Sartyni, Mateusz Eustachy [d (1796-1884)] 1e Red.
700	1	↳ Rudyński, Adolf [d (1818-1873)] 1e Red.
740	0 2	↳ Dzieńnik Urzędowy
770	0	↳ Amtsblatt zur Lemberger Zeitung no ... = Dziennik Urzędowy do Gazety Lwowskiej nr ... [g 1848-1866 /w xx001121088
770	0	↳ Dodatek do Gazety Lwowskiej nr ... : doniesienia urzędowe [g 1812, nr 38-1848, nr 68 /w xx001109150
770	0	↳ Dodatek Tygodniowy [g 1851-4.10, nr 52 (1860) /w zz2003915981
770	0	↳ Dodatek Osobny przy Gazecie Lwowskiej [g 1861, nr 1-1862, nr 31 [egz. Kr UJ] /w xx001180532
770	0	↳ Dodatek Nadzwyczajny do nr ... Gazety Lwowskiej [g 1867, [nr 1]-nr 12 /w xx001180640
770	0	↳ Dodatek Tygodniowy do Gazety Lwowskiej [g 1868, nr 1-1869, nr 26 /w xx001180719
770	0	↳ Dodatek Miesięczny do Gazety Lwowskiej [g R. 1, t. 1, z. 1 (stycz. 1873)-R. 2, t. 1, z. 1, 64 s. (stycz. 1873) /w bb2001150913
770	0	↳ Przewodnik Naukowy i Literacki [g 1873, t. 1 [Buty], s. 65-R. 49 (1921) /w da2001050046
770	0	↳ Narodna Časopis [g R. 1 (1891)-1900 /w xx002079801
787	1	↳ Rozmaitości : przydatek do numeru ... gazety Lwowskiej [g 1817-1859 /w zz2002955962



II.3.

- ↳ Gazeta Lwowska 1837
- ↳ Gazeta Lwowska 1838
- ↳ Gazeta Lwowska 1839
- ↳ Dodatki do Gazety Lwowskiej
 - ↳ Rozmaitości
 - ↳ Dodatek do Gazety Lwowskiej - doniesienia urzędowe
 - ↳ Amtsblatt zur Lemberger Zeitung = Dziennik Urzędowy do Gazety Lwowskiej
 - ↳ **Dodatki Tygodniowy przy Gazecie Lwowskiej**
 - ↳ Dodatek Tygodniowy przy Gazecie Lwowskiej 1851
 - ↳ Dodatek Tygodniowy przy Gazecie Lwowskiej 1852
 - ↳ Dodatek Tygodniowy przy Gazecie Lwowskiej 1853
 - ↳ Dodatek Tygodniowy przy Gazecie Lwowskiej 1854
 - ↳ Dodatek Tygodniowy przy Gazecie Lwowskiej 1855
 - ↳ Dodatek Tygodniowy przy Gazecie Lwowskiej 1856
 - ↳ Dodatek Tygodniowy przy Gazecie Lwowskiej 1857
 - ↳ Dodatek Tygodniowy przy Gazecie Lwowskiej 1858
 - ↳ Dodatek Tygodniowy przy Gazecie Lwowskiej 1859
 - ↳ Dodatek Tygodniowy przy Gazecie Lwowskiej 1860
 - ↳ Dodatek Osobny przy Gazecie Lwowskiej
 - ↳ Dodatek Nadzwyczajny do Gazety Lwowskiej
 - ↳ Dodatek Tygodniowy do Gazety Lwowskiej
 - ↳ Dodatek Miesięczny do Gazety Lwowskiej - poświęcony historii, statystyce i historii krajowej
 - ↳ Narodna Časopis - dodatek do Gazety Lwowskiej
 - ↳ Przewodnik Naukowy i Literacki - dodatek do Gazety Lwowskiej



II.4.

ich unikatowość jak i kompletność. Nie oznaczało to jednak zawsze pełnej kompletności numerów na poziomie każdego tytułu. Dlatego stworzenie kompleksowego zespołu opisów bibliograficznych obrazujących całą historię niektórych tytułów było trudne do wykonania na podstawie analizy jedynie oryginałów zachowanych w Bibliotece Jagiellońskiej. Czy to brak kontynuacji pod innym tytułem, czy brak pierwszych lub ostatnich numerów zmuszał katalogujących do

poszukiwań wykraczających poza zasób BJ. Tu pomocne okazały się z jednej strony katalogi innych bibliotek informujące zarówno o drukach jak i zbiorach mikrofilmowych, a z drugiej strony biblioteki cyfrowe.

Dobrym przykładem rozwiązania ułatwiającego dostęp do brakujących w zasobie BJ numerów i współpracy w uzupełnieniu całego zasobu danego tytułu czasopisma jest współpraca

II. 3. Fragment rekordu bibliograficznego w formacie MARC 21 dla Gazety Lwowskiej.

II. 4. Struktura dodatków do czasopisma głównego w JBC.

sługi Jagiellońskiej Biblioteki Cyfrowej objęto wszystkie dokumenty w całości digitalizowane w Bibliotece Jagiellońskiej. Przy tak dużej liczbie skanowanych dokumentów i tworzonych dla nich metadanych, rozwiązanie takie stało się koniecznością.

Opisane rozwiązania przyjęte w Bibliotece Jagiellońskiej dla potrzeb realizacji projektu Jagiellońska Biblioteka Cyfrowa łączą i wzajemnie uzupełniają informacje zawarte w KKZBUJ i zbudowanej bibliotece cyfrowej (JBC). Przenosząc tę ideę na poziom ogólnopolski można wskazać koordynatora dla opisów bibliograficznych w formacie MARC 21, jakim jest katalog NUKAT. Tworzoną bazę rekordów bibliograficznych, dystrybuowaną między wszystkie biblioteki uczestniczące we współkatalogowaniu, przypomina baza Federacji Bibliotek Cyfrowych, gromadząca metadane w formacie Dublin Core z wszystkich należących do FBC bibliotek. Należałoby rozszerzać i uzupełniać współpracę instytucji koordynujących obie te bazy, bo pola ich działań coraz bardziej się zaczębiają.

Bibliografia

Grycz, Józef; Borkowska, Władysława (1946). Skrócone przepisy katalogowania alfabetycznego. Warszawa: Państwowy Zakład Wydawnictw Szkolnych.

Korpałowa, Franciszka (1957). Z dziejów oddziału czasopism Biblioteki Jagiellońskiej. Kraków: Zakład Produkcji Skryptów.

Łagodowa, Zofia (1975). Katalogi alfabetyczne Biblioteki Jagiellońskiej w latach 1945-1972. Biuletyn Biblioteki Jagiellońskiej, R. 25, s. 109-129.

Nałęcz, Barbara (2007). Format MARC 21 rekordu bibliograficznego dla wydawnictwa ciągłego [online]. Warszawa: Centrum NUKAT. [Dostęp: 20.10.2012]. Dostępny w World Wide Web: http://centrum.nukat.edu.pl/images/stories/file/instrukcje_procedury/MARC_21/marc21_ciągle_070717.pdf.

Struktura i nazewnictwo plików

Ogólne założenia dotyczące struktury

Wszystkie prace związane z ujednoczonym systemem zarządzania plikami w Jagiellońskiej Bibliotece Cyfrowej miały na celu uporządkowanie ogromnego zasobu materiałów, następnie ich opracowanie, archiwizowanie, a na końcu opublikowanie na platformie cyfrowej dLibra. W literaturze przedmiotu z zakresu informacji naukowej i bibliotekoznawstwa brak jest opracowania dotyczącego tego typu tematyki, jednakże można znaleźć różnego rodzaju publikacje, które odnoszą się do problematyki systemu organizacji informacji [Agosti, Borbinha, Kapidakis 2009; Borman 2000; Chowdhury, Koo, Hunter 2010]. Podstawą wyróżniania typowych systemów organizacji informacji często bywa sposób uporządkowania wyrażen reprezentujących pole semantyczne pewnej wiedzy dziedzinowej. W literaturze dotyczącej porządkowania cyfrowych zasobów informacyjnych i architektury informacji na ogół wyróżnia

Słowa kluczowe: archiwizacja, biblioteki cyfrowe, digitalizacja, dokumenty cyfrowe

Keywords: archiving, digital libraries, digital documents, digitization

się dwa podstawowe typy organizacji informacji (wiedzy): dokładne schematy organizacyjne i niejednoznaczne schematy organizacyjne. Schematy dokładne bazują na formalnych kryteriach porządkowania wyrażen, schematy niejednoznaczne na kryteriach logiczno-semantycznych. W schematach dokładnych każde wyrażenie wyznacza jednoznacznie zidentyfikowaną klasę obiektów lub treść, a zakresy poszczególnych klas są rozłączne. Schematy dokładne zapewniają szybkie i łatwe odnajdywanie informacji, o ile

użytkownik zna nazwy interesujących go osób, miejsc, dzieł czy innych obiektów informacyjnych; nie umożliwiają jednak całościowego wglądu w zawartość tak zorganizowanych zasobów informacyjnych (inaczej mówiąc, wyszczególniając informację o obiektach jednostkowych, nie zapewniają jej generalizacji). W przypadku schematów niejednoznacznych mamy do czynienia z podziałem na grupy (klasy, kategorie) uporządkowane według kryteriów semantyczno-logicz-

nych. Do schematów niejednoznacznych należą więc różne odmiany struktur hierarchicznych, których klasy reprezentowane są przez wyrażenia naturalne. Ponieważ rzadko się zdarza, żeby wyraz czy wyrażenie miało tylko jedno znaczenie, przy hierarchizacji klas bierze się pod uwagę jedną, arbitralnie przyjętą interpretację znaczeniową reprezentującego ją wyrażenia lub dąży się do objęcia hierarchizacją wielu lub wszystkich możliwych jego znaczeń [Woźniak-Kasperek 2006].

Specyfika środowiska sieciowego oraz oczekiwania odbiorców treści zamieszczanych w Internecie stawiają liczne wyzwania przed twórcami bibliotek czy repozytoriów cyfrowych. Samo wytworzenie plików, często w postaci plików graficznych, jest dopiero początkiem w procesie pełnego, a co ważniejsze, funkcjonalnego udostępniania zasobów. Niezbędnym elementem tego procesu jest sporządzenie odpowiedniego opisu obiektu, który umożliwi, z jednej strony, dotarcie do poszukiwanych informacji, z drugiej pozwoli na efektywne zarządzanie i obsługę dużego zasobu informacyjnego. Dane uporządkowane w standaryzowanej strukturze są jednoznaczne i możliwe do wykorzystania w środowisku sieciowym, chociażby do wymiany czy agregacji przez różne oprogramowania lub system [Potęga; Wróbel 2011, s. 71-72].

Główne problemy, z jakim można się było spotkać w pracy związanej z uporządkowaniem zbiorów już zdigitalizo-

wanych, wynikały z niezorganizowanego systemu nazewnictwa i hierarchii. Przede wszystkim, na samym początku zniwelowano chaotyczny system nazywania plików, aby ujednoczyć go dla wszystkich oryginalnych kopii cyfrowych dokumentów przechowywanych w różnych oddziałach Biblioteki Jagiellońskiej. Prace porządkowe polegały na sprawdzeniu wszystkich zbiorów objętych projektem, pod względem kompletności, jakości obrazu, a także sposobu zapisu plików.

Problematyka kopii cyfrowych

Dariusz Paradowski w swojej książce „Digitalizacja piśmiennictwa” zwrócił szczególną uwagę na to, że często jednym z pierwszych skojarzeń związanych z przechowywaniem obiektów będących częścią dziedzictwa kulturowego, zwłaszcza w odniesieniu do obiektów cyfrowych, jest potrzeba zorganizowania właściwego miejsca ich archiwizacji [Paradowski 2010, s. 41]. Na wielu eta-



Il. 1.

Z uwagi na zróżnicowane zbiory Biblioteki Jagiellońskiej i sposób zapisu folderów przez poszczególne oddziały, przyjęto ujednoczony układ, który zostanie opisany dokładnie w kolejnych podrozdziałach. Takie rozwiązanie pozwoliło na lepszą organizację pracy i czytelność dla wszystkich pracowników (przykładem jest il. 1, który obrazuje strukturę w starym i nowym zapisie nazw folderów).

pach procesu digitalizacji tworzone są informacje. Pod względem objętości jest to przede wszystkim zawartość cyfrowa, będąca celem digitalizacji oraz metadane i informacje wspomagające proces. Konieczne jest dostosowanie metod zapisu i przechowywania tych danych do ich charakteru, ponieważ będzie to czynnik decydujący w uzyskaniu właściwej dostępności tych informacji [Paradowski 2010, s. 129]. Na każdym etapie digitalizacji zbiorów powstają różnego ro-

Il. 1. Zapis starej i nowej struktury nazw folderów na przykładzie kopii cyfrowych rękopisów.



Abstrakt: Przedmiotem artykułu jest struktura i nazewnictwo plików w Jagiellońskiej Bibliotece Cyfrowej. Zwrócono uwagę na zabezpieczenie zbiorów w postaci kopii dokumentów cyfrowych, a także na ich zarządzanie i archiwizację. Omówiono podstawowe rozwiązania związane z ujednoczonym systemem zarządzania plikami w JBC.

Abstract: The topic of the article is the file structure and file naming convention used within the Jagiellonian Digital Library. The issues arising in preserving collections in the form of digital documents are explored, as well as the problems of managing and archiving them. The basic solutions related to the unified file management system in the Jagiellonian Digital Library are discussed.

dzaju dane i sposób ich zapisu niekiedy nastręcza wielu trudności związanych z „rozszyfrowaniem” poszczególnych elementów zapisu plików. Każdy dokument analogowy, który zostaje zakwalifikowany do digitalizacji powinien posiadać niezbędne elementy opisu, które umożliwią jego identyfikację (w przypadku zbiorów Biblioteki Jagiellońskiej jest to sygnatura dokumentu). Aby uniknąć przestoju w procesie digitalizacji, należy zwrócić uwagę, aby dokumenty, które zostały zakwalifikowane do skanowania zostały poprawnie opracowane.

Prace związane z analizowaniem zbiorów zdigitalizowanych wcześniej, a których szacunkowa liczba obrazów cyfrowych zgromadzonych na płytach wynosi około 270 000, miały na celu usystematyzowanie zbiorów, a także wyeliminowanie ewentualnych dubletów.

W ramach prac porządkowych wyznaczono następujące działania:

- segregację w obrębie zbioru kopii cyfrowych (pierwotnie nazywanych DIG), który zawiera: czasopisma, druki zwarte, czasopisma rzadkie, druki zwarte, rzadkie, stare druki, rękopisy, druki muzyczne, dokumenty ikonograficzne i kartograficzne oraz dokumenty życia społecznego,
- przypisywanie odpowiednich dokumentów do właściwych kategorii (na podstawie tych kategorii tworzona była struktura folderów),

- zmianę nazw folderów, jak i plików tak, aby wyeliminować ewentualne nałożenie się kopii cyfrowych w różnych rodzajach dokumentów.

Przeprowadzono także szczegółową kwerendę obejmującą zbiory zdigitalizowane od 2008 roku, zgromadzone na macierzach dyskowych. To przeszukiwanie pozwoliło wyodrębnić dokumenty, które z różnych względów były skanowane ponownie, np. na specjalne zamówienie użytkownika. Można było porównać w ten sposób pliki i wybrać odpowiednie kopie do obróbki cyfrowej, a następnie przygotować do publikacji na platformie dLibra.

Ujednolicony system identyfikacji kopii cyfrowych – nazwy folderów i plików dla różnego rodzaju zbiorów

Podczas prac związanych z ujednoliceniem struktury zbiorów w Jagiellońskiej Bibliotece Cyfrowej powstawało szereg dokumentów wewnętrznych, które pozwoliły zespołowi na wprowadzanie bieżących poprawek i optymalizowanie pracy. Procedura dotycząca nazw katalogów i plików dotyczyła wprowadzenia jednolitego systemu identyfikacji cyfrowych kopii dokumentów. Obiekt cyfrowy musi mieć identyfikator. Identyfikator musi być niepowtarzalny i trwały [Płoszajski 2008, s. 95]. W Bibliotece Jagiellońskiej jako podstawowy identyfikator zastosowano numer DIG, który uzupełniany jest sygnaturą Biblioteki Jagiellońskiej w postaci ujednoliconej na drugim poziomie

szczegółowości, co ma odzwierciedlenie w strukturze folderów. Identyfikator kopii cyfrowej DIG jest dodawany do metadanych każdego dokumentu publikowanego w systemie dLibra. Zakres dozwolonych znaków w sygnaturach DIG ograniczony jest wyłącznie do dużych liter alfabetu łacińskiego (bez polskich znaków) oraz cyfr. Poniżej zaprezentowano budowę identyfikatora.

żadnymi dodatkowymi znakami.

System identyfikacji kopii cyfrowej – DIG – zakłada uszczegółowienie nazw poszczególnych folderów plików składowych poprzez dodanie kolejnych członów identyfikacyjnych.

Poziomy uszczegółowienia DIG:

Poziom 1. Podział ze względu na typ dokumentu. Nazwa folderu zawiera na-

Opis	Nagłówek	Typ dokumentu	Numer porządkowy
	DIG	CZAS	001008
	DIG	GRAF	000033
	DIG	DRUK	000001

Nagłówek w postaci ciągu znaków DIG występuje w każdej sygnaturze, następnie dołączane jest uszczegółowienie na podstawie typu dokumentu, (które może przyjmować wartości: ORP, STDR, MUŻ, GRAF, KART, CZAS, DRUK, DZS, BORZ, BORC) oraz numer porządkowy, odrębny dla każdego typu dokumentu, w postaci ciągu sześciu cyfr (zawsze z wprowadzającymi zerami). Nagłówek, typ dokumentu oraz numer porządkowy nie są oddzielone od siebie

główek DIG oraz dookreślenie typu dokumentu, np. DIGGRAF,

Poziom 2. Podział ze względu na poszczególne jednostki DIG. Nazwa folderu zawiera pełną sygnaturę DIG oraz oddzieloną znakiem podkreślenia (_) ujednoliconą sygnaturę BJ lub zakres sygnatur z sygnaturami krańcowymi rozdzielonymi dywizem „-”,

np. DIGGRAF000001_IR_1837,

Tab. 1. Budowa identyfikatora kopii cyfrowej DIG.

lub otrzymuje ona dookreślenie numeracyjne w postaci ciągu czterech zer,

DIGGRAF000159_IR_2181-2277.

Poziom 3. Podział ze względu na rocznik czasopisma. Nazwa folderu zawiera sygnaturę DIG, nie zawiera już ujednoczonej sygnatury BJ. Dookreślenie w postaci ciągu czterech cyfr oddzielone jest od sygnatury DIG znakiem podkreślenia (_),

np. DIGCZAS001008_1900.

Poziom 4. Podział ze względu na poszczególne numery w obrębie rocznika. Nazwa folderu zawiera pełną sygnaturę DIG, dookreślenie rocznika jak na poziomie 3 oraz oddzielone znakiem podkreślenia (_) dookreślenie numeru w postaci ciągu trzech cyfr,

np. DIGCZAS001008_1900_001.

Poziom 5. Podział ze względu na strony dokumentu oraz elementy dodatkowe. Nazwa każdego pliku składowego zawiera pełną sygnaturę DIG (w przypadku czasopism także dookreślenie rocznika i numeru) oraz numer strony publikacji cyfrowej w postaci ciągu czterech cyfr,

np. DIGGRAF000044_0002.tiff

lub

DIGCZAS001008_1900_001_0001.tiff.

Jeżeli występuje metryczka dokumentu,

np. DIGGRAF000001_0000.tiff.

Numer DIG bez dookreśleń w nazwie oznacza, iż jest to indeks publikacji rozdzielonej lub publikacja scalona, która zawiera wszystkie strony dokumentu źródłowego,

np. DIGGRAF000001.djvu.

Wszystkie poziomy uszczegółowienia występują jedynie w przypadku czasopism, których sygnatura DIG jest wspólna dla wszystkich numerów danego tytułu. Pozostałe dokumenty nie zawierają uszczegółowień 3. i 4. poziomu. Poziomy uszczegółowienia 1.–4. dotyczą folderów, a poziom 5. – plików, które zawierają także rozszerzenia odpowiednie do ich formatu. Umieszczenie na początku nazwy każdego pliku pełnej sygnatury DIG umożliwia identyfikację pojedynczych plików w przypadku zaburzenia struktury i zapobiega sytuacji, w której nazwy plików składowych różnych publikacji są identyczne.

Ujednoczenie struktury folderów

Najistotniejszą kwestią dotyczącą sprawnego zarządzania zasobem kopii cyfrowych dokumentów jest jasne i przejrzyste zdefiniowanie nazw zarówno folderów, jak i plików, tak, aby wyeliminować ewentualne nałożenie się kopii cyfrowych w różnych rodzajach dokumentów.

Odpowiednia struktura zapewnia miejsce na przechowywanie plików, ale w żaden sposób nie określa sposobu nazywania samych plików. Bardzo ważnym jest, aby nadać plikom odpowiednie nazwy [Wortal 2010]. Równie ważna jest też hierarchia, struktury podporządkowane i zależności pomiędzy nimi [Mój BIP 2009]. W kolejnej części artykułu zostanie to dokładnie wyjaśnione na przykładzie struktury plików dla wydawnictwa ciągłego o ustalonej częstotliwości i różnego rodzaju przypadków specjalnych, które pojawiły się w trakcie pracy nad uporządkowaniem zeskanowanych wcześniej zbiorów.

Ze względu na rodzaj dokumentów struktura folderów (oparta na typie dokumentów) wygląda następująco:

- **DIGORP** (rękopisy)
- DIGORP000001_RKP_BJ_5475
- DIGORP000001_0001.tiff
- DIGORP000001_0002.tiff
- ...
- DIGORP000001_0324.tiff
- DIGORP000001_0325.tiff
- DIGORP000735_Ms_Berol_Varnhagen-Sammlung_polonica_Bronikowski

- **DIGSTDR** (starodruki)
- DIGSTDR000001_6810_I
- DIGSTDR000001_0000.tiff
- DIGSTDR000001_0001.tiff
- DIGSTDR000001_0002.tiff
- DIGSTDR000001_0003.tiff
- DIGSTDR000001_0354.tiff
- DIGSTDR001192_1487_I

- **DIGMUZ** (muzykalia)

- DIGMUZ000001 (w tym przypadku brak sygnatury – wszystkie dzieła nieposiadające sygnatury BJ będą występowały tylko pod nazwą i numerem DIG-u)
- DIGMUZ000001_0001.tiff
- DIGMUZ000001_0002.tiff
- ...
- DIGMUZ000001_0051.tiff
- DIGMUZ000290_Chelibini_Aut_20_a_1-2

- **DIGGRAF** (grafika) (strona w obrębie numeru)
- DIGGRAF000001_IR_1837 • ...
- DIGGRAF000001_0001.tiff • DIGCZAS000001_1939_001_0006.tiff
- ... • DIGCZAS000001_1939_002 (numer w obrębie roku)
- DIGGRAF000001_0005.tiff • DIGCZAS000001_1940 (rok w obrębie tytułu)
- DIGGRAF000309_34471_V_Mag_Graf • DIGCZAS000252_434_V_czas

- **DIGKART** (kartografia)
- DIGKART000001_404_III_Atl • **DIGDRUK** (druki zwarte)
- DIGKART000001_0001.tiff • DIGDRUK000001_250_III
- DIGKART000001_0012.tiff • DIGDRUK000001_0001.tiff
- DIGKART000050_M_47_158 • ...

- **DIGCZAS** (czasopisma)
- DIGCZAS000001_7545_IV_czas • DIGDRUK000001_0236.tiff

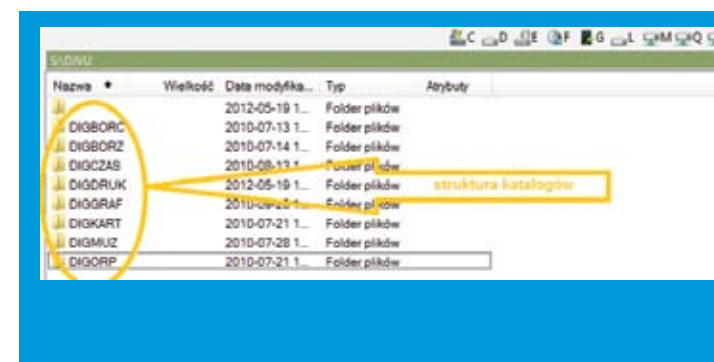
- DIGCZAS000001_1939 (rok w obrębie tytułu) • **DIGDZS** (dokumenty życia społecznego)
- DIGCZAS000001_1939_001_0001.tiff • DIGDZS000001_Dok_Zycia_Spol_XVII_10
- DIG- • DIGDZS000001_0001.tiff
CZAS000001_1939_001_0001.tiff

- ...
- DIGDZS000001_0005.tiff
- DIGDZS000059_73803_II_Dok_Zycia_Spol_XVII_10

Struktura ta ma swoje dokładne odbicie w stosunku do katalogów (il. 2) i plików DjVu. Utrzymanie takiego układu ułatwia dostęp do dokumentów archiwizowanych w postaci plików TIFF jak

Wyjaśnienie struktury plików na przykładzie wydawnictw ciągłych

Strukturę plików w Jagiellońskiej Bibliotece Cyfrowej można najlepiej wyjaśnić na przykładzie wydawnictw ciągłych. Przed samym procesem skanowania, wszystkie czasopisma zostały sprawdzone i opracowane, aby w dalszym etapie wyeliminować wszelkiego rodzaju błędy, które mogły się pojawić przy pracy nad obróbką materiału. Wydawnictwa



Il. 2.

i plików DjVu, do których dodawane są metadane. W przypadku, gdy zajdzie potrzeba rozbudowania tej struktury o dodatkowe kategorie dokumentów jeszcze w niej niezdefiniowane, stworzenie kolejnych podziałów na katalogi nie będzie narażać na żadnych problemów.

ciągłe, które zostały objęte projektem, zawierały często różnego rodzaju dodatki samo- i niesamoistne, różnorodne wydania nadzwyczajne, numery poranne, a także wszelkiego rodzaju dokumenty życia społecznego. Poniższe przykłady szczególnie obrazują, w jaki sposób dokonano zapisu folderów i nazw plików dla poszczególnych przypadków specjalnych.

Il. 2. Struktura katalogów w Jagiellońskiej Bibliotece Cyfrowej.

Struktura dla gazety (bez dodatków w środku)

```
[DIGCZAS0000001].[BJ_V_czas2]\
  [DIGCZAS000000].[ROK].[VOLUMINO3]\
    [DIGCZAS000000].[ROK].[NUMER4]\
```

przykład:

```
DIGCZAS002001_501_V_czas\DIGCZAS002001_1851_v1
\DIGCZAS002001_1851_001\DIGCZAS002001_1851_001_0001.tiff
```

¹ Za każdym razem sześciocyfrowy numer sygnatury DIG.

² Sygnatura Biblioteki Jagiellońskiej.

³ Dodawać identyfikator woluminu (v1, v2 itd.) jeżeli rocznik nie był oprawiony w pojedynczym woluminie.

⁴ Trzycyfrowy numer czasopisma.

Struktura dla gazety wraz z dołączonym dodatkiem przy każdym numerze

```
[DIGCZAS000000].[BJ_V_czas]\
  [DIGCZAS000000].[ROK].[VOLUMINO]\
    [DIGCZAS000000].[ROK].[NUMER]\ +5
      [DIGCZAS000000].[ROK].[NUMER].[DODATEK006]\
```

przykład:

```
DIGCZAS002001_501_V_czas\DIGCZAS002001_1851_v1
\DIGCZAS002001_1851_001\DIGCZAS002001_1851_001_0001.tiff
+
DIGCZAS002001_501_V_czas\DIGCZAS002001_1851_DODATEK06
\DIGCZAS002001_1851_001_DODATEK06
\DIGCZAS002001_1851_001_DODATEK06_0001.tiff
```

W przypadku Gazety Lwowskiej analizowano różnorodny materiał, który w Bibliotece Jagiellońskiej znajduje się na kilku różnych sygnaturach, dlatego systematycznie uzupełniano dodatki samoistne i każdy nowy dodatek miał nada-

wany coraz wyższy numer. W ten sposób zoptymalizowano pracę, aby przygotować również obrobione pliki do wykonania dodatkowej kopii zabezpieczającej w postaci mikrofilmu.

⁵ Obok regularnie tworzonych folderów należy dodać foldery zawierające strony dodatków wg schematu poniżej.

⁶ Tutaj należy wpisać słowo „DODATEK” i dodać na koniec dwucyfrowy numer identyfikujący dany dodatek.

Struktura dla dodatków

```
[DIGCZAS000000]_[B]_V_czas\  
    [DIGCZAS000000]_[ROK]_[DODATEK00]\  
        [DIGCZAS000000]_[ROK]_[NUMER]_[DODATEK00]\  
            \
```

przykład:

```
DIGCZAS002001_501_V_czas\  
DIGCZAS002001_1851_001_DODATEK05\  
DIGCZAS002001_1851_001_DODATEK05_0001.tiff
```

Nazwy plików

```
[DIGCZAS000000]_[ROK]_[NUMER]_[NR0000?].tiff +  
    [DIGCZAS000000]_[ROK]_[NUMER]_[DODATEK00]_[NR0000].tiff
```

przykład:

```
DIGCZAS002001_1851_v1_0001.tiff
```

⁷ Numer pliku zawiera tylko 4 cyfry numerowane kolejno od 0001 dla pierwszej strony czasopisma. Każdy numer ma osobną numerację rozpoczynaną od 0001.

Przygotowanie woluminu do konwersji na mikrofilm

```
[DIGCZAS000000]_[B]_V_czas\
  [DIGCZAS000000]_[ROK]\ +
    [DIGCZAS000000]_[ROK]_[NUMER8].tiff +
    [DIGCZAS000000]_[ROK]_[NUMER]_[NR0000].tiff
  [DIGCZAS000000]_[ROK]_[DODATEK00]\
    [DIGCZAS000000]_[ROK]_[NUMER].tiff +
    [DIGCZAS000000]_[ROK]_[NUMER]_[DODATEK00]_[NR0000].tiff
```

Ze względu na prace związane z archiwizacją materiału również w postaci mikrofilmu, przygotowywano pliki w odpowiedniej strukturze. Do każdego folderu, np. DIGCZAS002001_1909 (jest to rocznik 1909 Gazety Lwowskiej), dodawano pierwszą stronę z pierwszego woluminu rocznika, która była niewykadrowana i miała widoczną linijkę. Taka strona posiadała nazwę wg regularnego schematu z tym, że numer strony jest 000.tiff i nie ma w nazwie numeru gazety. W kolejnym miejscu znajdowała się pierwsza strona tegoż numeru będąca jednocześnie stroną już wykadrowaną. W przypadku czasopism, które posiadały dodatki samoistne, wszystkie te pliki były wyciągane z poszczególnych roczników gazet, a następnie porządkowane w obrębie danego tytułu dodatku, do którego oczywiście dołączano pierwszą niewykadrowaną stronę z danego rocznika.

⁸ Trzycyfrowy numer 000 dla strony pierwszej z linijką.

Bardzo ciekawy przypadek, jeśli chodzi o samą strukturę folderów można zaprezentować na przykładzie Sprawozdań Sejmu Krajowego Galicyjskiego. Ten bogaty materiał nie był wcześniej dokładnie opracowany i przygotowanie odpowiedniej struktury zajęło trochę czasu. Dzięki wspólnej pracy zespołu Jagiellońskiej Biblioteki Cyfrowej udało się stworzyć unikalny układ w podziale na kadencje, sesje i posiedzenia Sejmu.

Struktura nazewnictwa plików Sprawozdań Sejmu Krajowego Galicyjskiego

```
[DIGCZAS000000]_[B]
  [DIGCZAS000000]_[K009]
    [DIGCZAS000000]_[K00]_[S010]
      [DIGCZAS000000]_[K00]_[S0]_[P0011]
        [DIGCZAS000000]_[K00]_[S0]_[P00]_[NR0000].tiff
```

Podczas prac związanych z bezpośrednią obróbką materiału pojawił się szereg innych przypadków specjalnych, gdzie należało zastosować indywidualne rozwiązania dla struktury folderów. Gdy rocznik czasopisma nie pokrywał się z latami kalendarzowymi, należałoby wówczas przyjąć następującą strukturę:

```
[DIGCZAS000000]_[B]
  [DIGCZAS000000]_[R0012]
    [DIGCZAS000000]_[R00]_[NUMER]
      [DIGCZAS000000]_[R00]_[NUMER]_[NR0000].tiff
```

⁹ Dwucyfrowy numer kadencji Sejmu Krajowego.

¹⁰ Jednocyfrowy numer sesji Sejmu Krajowego.

¹¹ Dwucyfrowy numer posiedzenia Sejmu Krajowego.

¹² Dwucyfrowy numer określający kolejny rocznik wydawanego czasopisma.

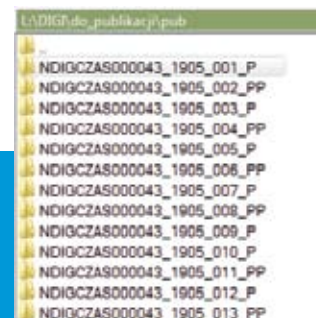
Jeżeli w obrębie wydawanego tytułu czasopisma pojawił się drugi ciąg numerów, struktura dla czasopisma wyglądała następująco:

```
[DIGCZAS000000]_[B]
    [DIGCZAS000000]_[ROK]_[VOLUMINO]
        [DIGCZAS000000]_[ROK]_[NUMER]
            [DIGCZAS000000]_[ROK]_[NUMER]_[NR0000].tiff+
                [DIGCZAS000000]_[ROK]B_[NUMER]
                    [DIGCZAS000000]_[ROK]B_[NUMER]_[NR0000].tiff
```

Zdarzało się również, że wydawnictwa ciągle posiadały w obrębie roczników wszelkiego rodzaju spisy treści, skorowidze. Jeżeli w woluminie czasopisma był spis treści czy indeks dla tego woluminu, to należałoby go umieścić, jako numer czasopisma [000] dla indeksów i spisów znajdujących się na początku woluminu i dla indeksów i spisów znajdujących się na końcu woluminu.

```
[DIGCZAS000000]_[B]
    [DIGCZAS000000]_[ROK]_[VOLUMINO]
        [DIGCZAS000000]_[ROK]_[NUMER]
            [DIGCZAS000000]_[ROK]_[NUMER]_P
                [DIGCZAS000000]_[ROK]_[NUMER]_W
                    [DIGCZAS000000]_[ROK]_[NUMER]_[NR0000].tiff
                        [DIGCZAS000000]_[ROK]_[NUMER]_P_[NR0000].tiff
                            [DIGCZAS000000]_[ROK]_[NUMER]_W_[NR0000].tiff
```

Ostatnim przykładem, są wszelkiego rodzaju wydania poranne, popołudniowe i wieczorne. Jeżeli gazeta posiadała wydania wieczorne, które zawarte były w tej samej strukturze, należy po polu [NUMER] dodać literę „P” (wydanie popołudniowe) i „W” (Wydanie Wieczorne) bez oddzielającego je separatora (przykład il. 3. i il. 4). Wygląda to następująco:



Il. 3.



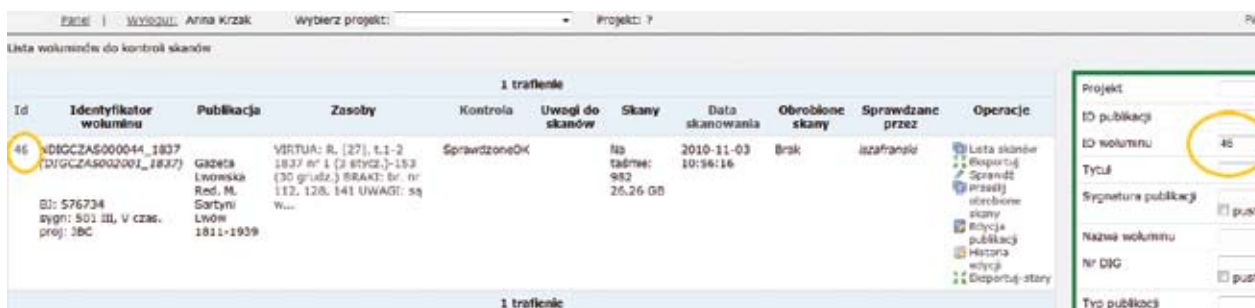
Il. 4.

Il. 3. Struktura katalogów dla wydań porannych i popołudniowych.

Il. 4. Struktura plików dla wydania porannego.



II. 6.



II. 7.

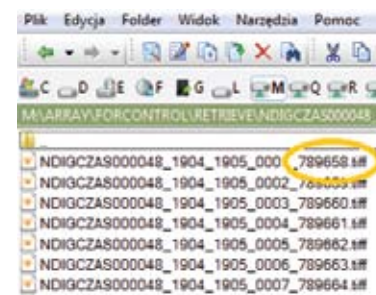
DIGCZAS – czasopisma ze zbiorów magazynowych i RARA

DIGDRUK – druki zwarte ze zbiorów magazynowych i RARA

DIGDZS – druki ulotne i wszelkie inne niekwalifikujące się do żadnej z wyżej wymienionych grup.

W związku z tą sytuacją procedura obejmuje również nanoszenie poprawek

w systemie SOJBC, ponieważ wgrane tam metadane dokumentów sygnowane są starym oznaczeniem DIG. Zmiany w systemie są nanoszone na bieżąco, aby sprawnie wyeliminować wszystkie błędy. Pozostałe kategorie dokumentów mają pierwotnie nadane sygnatury, co przyczynia się do łatwiejszego zarządzania kopiami cyfrowymi, a system dodatkowo generuje nowe sygnatury NDIG. Ujednolicenie nazw plików i folderów w systemie SOJBC i na macierzy pozwoliło na łatwiejsze zaarchiwizowanie już wgranych dokumentów i zwolnienie miejsca na macierzach dyskowych. Kontrola archiwum nośników lub danych w repozytorium jest misją krytyczną całego procesu digitalizacji. Brak zdefiniowanych procedur, niedojrzałe podejście do oceny ryzyka lub niepoprawny optymizm, co do trwałości mediów, mogą zniszczyć efekty często wieloletniej pracy. Trudno oszacować koszty takiej pomyłki [Paradowski 2010, s. 148-149]. Dlatego bardzo ważny jest też system organizacji pracy. Każdy z pracowników zajmujących się bezpośrednią kontrolą zasobów i obróbką cyfrową zwraca uwagę, nie tylko na jakość materiału, ale także na sposób zapisu plików, które są generowane przez SOJBC. Podczas pobierania plików, generowany jest specjalny katalog eksportowy, w którym znajdują się pliki TIFF. Sposób ich zapisu przedstawiono poniżej. Obok sygnatury cyfrowej NDIG wraz z numerem ID publikacji, rocznikiem i numerami kolejnych plików, znajduje się również tak zwany numer systemowy pliku (il. 8). Jest on ważny, gdyż jest utrwalany na taśmach magnetycznych, co pozwala na szybką identyfikację i wyszukiwanie. Również w przypadku, kiedy wystąpią problemy techniczne i pliki źródłowe danego woluminu zostaną usunięte z komputera, wówczas można je odzyskać z taśmy magnetycznej, na której zostały zarchiwizowane pliki master.



II. 8.

twiejsze zaarchiwizowanie już wgranych dokumentów i zwolnienie miejsca na macierzach dyskowych. Kontrola archiwum nośników lub danych w repozytorium jest misją krytyczną całego procesu digitalizacji. Brak zdefiniowanych procedur, niedojrzałe podejście do oceny ryzyka lub niepoprawny optymizm, co do trwałości mediów, mogą zniszczyć efekty często wieloletniej pracy. Trudno oszacować koszty takiej pomyłki [Paradowski 2010, s. 148-149]. Dlatego bardzo ważny jest też system organizacji pracy. Każdy z pracowników zajmujących się bezpośrednią kontrolą zasobów i obróbką cyfrową zwraca uwagę, nie tylko na jakość materiału, ale także na sposób zapisu plików, które są generowane przez SOJBC. Podczas pobierania plików, generowany jest specjalny katalog eksportowy, w którym znajdują się pliki TIFF. Sposób ich zapisu przedstawiono poniżej. Obok sygnatury cyfrowej NDIG wraz z numerem ID publikacji, rocznikiem i numerami kolejnych plików, znajduje się również tak zwany numer systemowy pliku (il. 8). Jest on ważny, gdyż jest utrwalany na taśmach magnetycznych, co pozwala na szybką identyfikację i wyszukiwanie. Również w przypadku, kiedy wystąpią problemy techniczne i pliki źródłowe danego woluminu zostaną usunięte z komputera, wówczas można je odzyskać z taśmy magnetycznej, na której zostały zarchiwizowane pliki master.

II. 8. Numer systemowy pliku.

II. 6. ID publikacji.

II. 7. ID woluminu.

Warto rozważyć wykorzystanie systemów komputerowych wspierających i usprawniających proces digitalizacji i zarządzania wytwarzanym zasobem cyfrowym oraz dających możliwość integracji z programami komputerowymi, przechowującymi informacje o obiektach analogowych i innymi bazami danych. Zastosowanie takiego rozwiązania ułatwia optymalizację pracy, sprawia, że wszystkie informacje o obiekcie analogowym, procesie digitalizacji, parametrach technicznych czy zasobie cyfrowym mogą zostać zgromadzone w jednym miejscu i powiązane ze sobą, co pozwala uniknąć powielania pracy i tworzenia kilku niezależnych opisów jednego obiektu, sporządzanych na potrzeby kilku komórek uczestniczących w procesie digitalizacji [Paradowski 2010, s. 157]. Dlatego na potrzeby projektu Jagiellońskiej Biblioteki Cyfrowej został stworzony System Obsługi Jagiellońskiej Biblioteki Cyfrowej, będący oprogramowaniem służącym do zarządzania zbiorami cyfrowymi. W systemie za pomocą panelu wyszukiwawczego można dokładnie wyszukiwać rekordy publikacji (tytuły) i woluminów. Także przy udziale pracowników zajmujących się kontrolą i obróbką cyfrową materiału, powstało wiele specyficznych rozwiązań w zakresie oprogramowania własnego i organizacji systemu pracy.

Bibliografia

Agosti, Maristella; Borbinha, José; Kapidakis, Sarantos eds. (2009). *Research and Advance Technology for Digital Libraries*. 13th European Conference, ECDL 2009 Corfu, Greece, September 27- October 2, 2009 Proceedings. Berlin Heidelberg: Springer-Verlag, 497 pp.

Borgman, Christine L. (2000). *From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World* Digital Libraries and Electronic Publishing. Cambridge: MIT Press, 321 pp.

Chowdhury, Gobinda; Koo, Chris; Hunter, Jane (red.). (2010). *The Role of Digital Libraries in a Time of Global Change*. 12th International Conference on Asia-Pacific Digital Libraries, ICADL 2010 Gold Coast, Australia, June 21-25, 2010 Proceedings. Berlin Heidelberg: Springer-Verlag, 270 pp.

Lesk, Michael (2004). *Understanding Digital Librarie*. Second Edition. Amsterdam: Morgan Kaufmann Publishers, 455 pp.

Mój BIP (2009). *Nazwy plików i ich struktura*. [online]. [Dostęp: 07.05.2012]. Dostępny w World Wide Web: <http://mojbip.pl/index.php/component/content/article/16-nazwy-plikow-i-ich-struktura>.

Paradowski, Dariusz (oprac. i red). (2010). *Digitalizacja piśmiennictwa*. Warszawa: Biblioteka Narodowa, 212 s.

Płoszajski, Grzegorz (red). (2008). *Standardy w procesie digitalizacji obiektów dziedzictwa kul-*

turowego. Warszawa: Biblioteka Główna Politechniki Warszawskiej, 247 s.

Program digitalizacji 2009-2020 (2009). *Program digitalizacji dóbr kultury oraz gromadzenia, przechowywania i udostępniania obiektów cyfrowych w Polsce 2009-2020* [online]. [Dostęp: 07.05.2012]. Dostępny w World Wide Web: <http://www.kongreskultury.pl/library/File/RaportDigitalizacja/Program%20digitalizacji%202009-2020.pdf>.

Potęga, Joanna; Wróbel, Agnieszka (2010). *The Dublin Core Metadata Element Set, Ver. 1.1 a potrzeby i oczekiwania bibliotekarzy cyfrowych – analiza przypadków*. [W:] Mazurek, Cezary; Stroiński, Maciej; Węglarz, Jan (red.) *Polskie biblioteki cyfrowe 2009*. Materiały z konferencji zorganizowanej dnia 9 grudnia 2009 roku przez Bibliotekę Kórnicką PAN, Poznańską Fundację Bibliotek Naukowych, Poznańskie Centrum Superkomputerowo-Sieciowe. Poznań: Ośrodek Wydawnictw Naukowych, s. 71-72.

Wortal (2010). *Nazewnictwo plików./Projektowanie aplikacji w PHP* [online]. [Dostęp: 07.05.2012]. Dostępny w World Wide Web: <http://php.pl/Wortal/Artykuly/Proces-tworzenia-aplikacji/Projektowanie-aplikacji-w-PHP-Czesc-pierwsza/Nazewnictwo-plikow>.

Woźniak-Kasperek, Jadwiga (2006). *Organizacja informacji w internetowych serwisach kontrolowanej jakości* [online]. [Dostęp: 07.05.2012]. Dostępny w World Wide Web: <http://www.ebib.info/publikacje/matkonf/kaba/wozniak-kasperek.php>.

Metadane



Słowa kluczowe: biblioteki cyfrowe, metadane, dLibra, kolekcje cyfrowe

Keywords: digital libraries, metadata, dLibra, digital collections

Wstęp

Odnalezienie pożądanej informacji w gąszczu zasobów Internetu nie jest łatwym zadaniem. Ten truizm odnosi się również do zawartości bibliotek cyfrowych, stanowiących część globalnej sieci. Warunkiem dotarcia do publikacji udostępnianych w bibliotekach cyfrowych jest istnienie informacji o tych obiektach. Pierwszym i najprostszym skojarzeniem ze słowem „metadane” –

bo tak określa się te „dane o danych” – w kontekście bibliotek cyfrowych jest opis obiektu przygotowany przez redaktora według określonego schematu. Jednak w środowisku sieciowym nie jest to jedyna forma metainformacji, jaka umożliwia odnalezienie właściwego zasobu. Za źródło niezbędnych metadanych uznaje się samą zawartość obiektów (efekt indeksowania pełnotekstowego¹), sygnalizując w ten sposób tendencję do zrównania danych z metadanymi [Nahotko 2010, s. 105]. Jest to jednak model, w który nie do końca można wpisać zasoby, jakie udostępniają polskie biblioteki cyfrowe. Prezentowane w nich obiekty to w większości zdigitalizowane zbiory biblioteczne, których zawartość tekstowa, mimo stosowania coraz sprawniejszych narzędzi do optycznego rozpoznawania znaków, w znacznym stopniu pozostaje niewidoczna dla automatów wyszukiwarek. Powodem niedoskonałego rozpoznania tekstu są specyficzne cechy digitalizowanych materiałów: zły stan zachowania dokumentów źródłowych,

¹ Nie dotyczy to oczywiście zbiorów o charakterze nietekstowym, np. grafiki. Za zbiory nietekstowe (z perspektywy komputera) można uznać również rękopisy, w których rozpoznanie znaków wymaga o wiele bardziej zaawansowanej i niemożliwej do wykorzystania na skalę masową technologii, polegającej na uczeniu maszyny różnych charakterów pisma (ICR – Intelligent Character Recognition). Alternatywą dla użycia technologii

ICR może być umieszczanie uprzednio przygotowanej warstwy tekstowej w pliku z obrazem rękopisu, np. w formacie DjVu [Bednarek 2007]. Jednak nakład pracy, jakiego wymaga przygotowanie tego typu publikacji, sprawia, że takie rozwiązanie można uznać za celowe jedynie w wypadku najcenniejszych zbiorów specjalnych.

czcionki wymagające specjalistycznego, drogiego oprogramowania [Kalota, Raczynski, Rękar 2011, s. 59].

Nie oznacza to, że nie jest możliwe odnalezienie zasobów prezentowanych przez polskie biblioteki cyfrowe przez szukanie bezpośrednio w ich zawartości. W znaczącej części prezentowanych materiałów rozpoznanie tekstu zostało wykonane w stopniu umożliwiającym jego poprawne zaindeksowanie. Wzrasta również liczba publikacji od początku dostępnych w formie elektronicznej (tzw. *born-digital*), których zawartość daje się przeszukać bez żadnych problemów.

Pełnotekstowe przeszukiwanie zasobów bibliotek cyfrowych możliwe jest z dwóch podstawowych poziomów: wyszukiwarki globalnej (np. Google²) oraz lokalnej wyszukiwarki konkretnej biblioteki (Federacja Bibliotek Cyfrowych nie umożliwia szukania w tekstach publikacji). Jednak w obu przypadkach wyniki wyszukiwania prezentowane są jako lista tytułów odnalezionych publikacji. O ile w wypadku Google poza tytułem wyświetlany jest kontekst wyszukiwanej frazy lub słowa, o tyle w wyszukiwarkach bibliotek cyfrowych opartych na systemie dLibra domyślna forma prezentowania wyników wyszukiwania to tylko lista tytułów i określonych dla nich słów kluczowych (w wersjach niższych niż 5.0 tylko w widoku zaawansowanym). A zatem czytelnik, zanim zdecyduje się na wybór proponowanej publikacji, bez konieczności pobierania ocenia jej przydatność

² Wprawdzie większość polskich bibliotek cyfrowych prezentuje zasoby w formacie DjVu, który nie jest indeksowany przez Google, jednak w oprogramowaniu dLibra zaimplementowano mechanizm eksportujący warstwę tekstową plików DjVu i udostępniający ją w sieci jako plik HTML „widziany” przez Google, co umożliwia wyszukiwanie w treści publikacji udostępnianych przez biblioteki cyfrowe (odpowiednie sformułowanie zapy-

na podstawie metadanych określonych przez redaktora [Werla 2011b, s. 126], co pozwala stwierdzić, że prawidłowo opracowane metadane odgrywają zasadniczą rolę w wyborze publikacji, nawet jeżeli nie są brane pod uwagę w pierwotnym zapytaniu.

Metadane w systemie dLibra

System dLibra umożliwia strukturyzację metainformacji (rozumianych jako dane

Abstrakt: W artykule omówiono kwestie związane z metadanymi w Jagiellońskiej Bibliotece Cyfrowej. Zaprezentowano możliwości użytego oprogramowania (dLibra) w zakresie opisu pojedynczych publikacji oraz grupowania większych zbiorów. Opisano przygotowaną na potrzeby JBC polihierarchiczną strukturę kolekcji oraz schemat opisu publikacji. Celowość zastosowanych rozwiązań poparto przykładami ich zastosowań. Pokróćce omówiono również techniczne aspekty powstawania opisów publikacji prezentowanych w JBC.

Abstract: The paper raises the issues related to metadata in the Jagiellonian Digital Library. It presents the capabilities of its underlying software (dLibra framework), which involve handling descriptions of individual publications as well as of larger collections of digital data. The author describes the polyhierarchical collection structure and the description schema created for the purpose of the project. The effectiveness of the implemented solutions is demonstrated through working examples. Some technical aspects of creating publication descriptions are also considered.

tania, np. dodanie nazwy domeny, na której utworzono konkretną bibliotekę cyfrową, pozwala uzyskać skuteczniejsze wyniki).



o obiekcie opracowane przez człowieka) na dwóch poziomach: opisu pojedynczej publikacji oraz całych zbiorów. Do opisu indywidualnego służy schemat metadanych Dublin Core Metadata Element Set³, przy czym oprogramowanie pozwala na jego modyfikację [Parkoła 2008, s. 5]. Informowanie o zawartości większych zbiorów umożliwia mechanizm kolekcji i możliwość ich dowolnego hierarchizowania [Parkoła 2008, s. 2]. Grupowania w kolekcje można dokonywać według dowolnego klucza, np. kryteriów formalnych lub rzeczowych (jak w tradycyjnych katalogach bibliotecznych). Tego typu porządkowanie zasobów wydaje się pożyteczne przede wszystkim ze względu na możliwość przeglądania zbiorów bez konieczności formułowania zapytania. Sytuacje tego typu nie należą bynajmniej do rzadkości – czytelnik, który nie potrafi jasno określić przedmiotu swojego zainteresowania, jest w stanie go sprecyzować, jeśli trafi na dobrze przygotowany system nawigacji, który podpowie mu, co może znaleźć w danym serwisie. Ten tryb poszukiwania, określanany mianem „przeglądania” (*browsing*), umożliwia odnajdywanie informacji, których nie poszukiwało się w sposób świadomy i ukierunkowany [Bates 2002, s. 4–6].

Struktura kolekcji w Jagiellońskiej Bibliotece Cyfrowej

Zasoby Jagiellońskiej Biblioteki Cyfrowej uporządkowano według różnych kryteriów, tworząc polihierarchiczną strukturę [Sosińska-Kalata 2005,

s. 150] opartą na wielokrotnym przyporządkowaniu poszczególnych publikacji lub grup publikacji do kolekcji. Trzy główne kolekcje odzwierciedlają podział według formy dokumentów (kolekcja „Czytelnie”), klasyfikacji systematycznej („Klasyfikacja systematyczna”) oraz proveniencji prezentowanych zasobów (kolekcja „Uniwersytet Jagielloński”). Podział uzupełniają „Kolekcje tematyczne”, przeznaczone dla większych zbiorów, które z różnych względów powinny zostać wyeksponowane, oraz kolekcja „Projekty”, pozwalająca określić finansowanie digitalizacji poszczególnych publikacji (szczegółowe informacje na temat kolekcji przedstawiono w tabeli 1). Taki niejednoznaczny schemat organizacji informacji [Rosenfeld, Morville 2003, s. 77], umożliwiający wieloaspektowe przeglądanie zasobów, zwielokrotnia szanse użytkownika na dotarcie do informacji, której nie potrafi precyzyjnie zdefiniować.

Opis publikacji w Jagiellońskiej Bibliotece Cyfrowej

Schemat metadanych

Opis publikacji prezentowanych w Jagiellońskiej Bibliotece Cyfrowej oparto na wbudowanym domyślnie w oprogramowaniu dLibra schemacie Dublin Core. Jednak na potrzeby serwisu schemat zmodyfikowano, dodając kilka nowych elementów (jako podatrybuty lub elementy samodzielne). Modyfikacji dokonano, uwzględniając analizy sche-

³Najnowsze wersje systemu pozwalają opisywać zasoby w składającym się z 59 elementów profilu aplikacyjnym Dublin Core – PLMET [Werla 2011a], jednak jego użycie na poziomie pojedynczych bibliotek cyfrowych jest do tej pory znikome. Serwisy korzystające z nowego schematu w większości wypadków używają prawie wyłącznie atrybutów znanych z „klasycznego” schematu DC. Sam schemat jest jednak używany do agregowania

wzbogaconych metadanych na poziomie Federacji Bibliotek Cyfrowych.

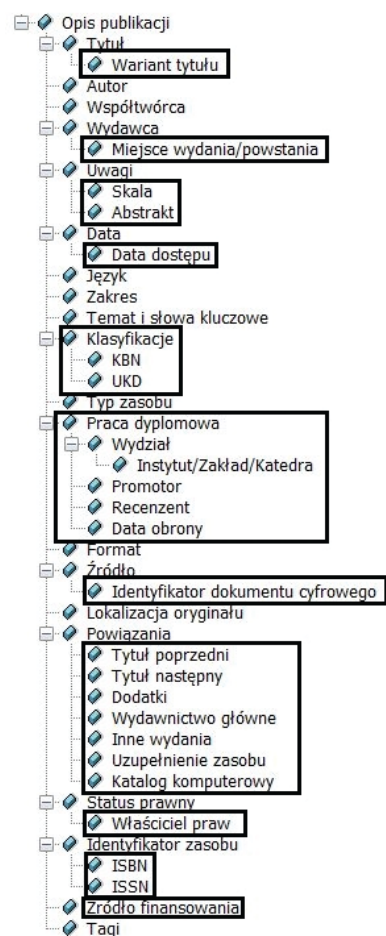
	Kolekcja	Charakterystyka	Podział wewnętrzny lub jego wycinek
	Czytelnie	Podział formalny dokumentów. Nazwy kolekcji i podkolekcji nawiązują do funkcjonujących w Bibliotece Jagiellońskiej czytelni i gabinetów (np. Czytelni Czasopism, Starych Druków, Rękopisów, Gabinetu Zbiorów Graficznych i Kartograficznych).	Czasopisma Zbiory kartograficzne Zbiory ikonograficzne Stare druki Rękopisy Muzykalia Książki Dokumenty życia społecznego Dokumenty dźwiękowe Dokumenty audiowizualne Zasoby online
	Klasyfikacja systematyczna	Klasyfikacja systematyczna oparta na tradycyjnym Katalogu Systematycznym Biblioteki Jagiellońskiej.	Encyklopedie i słowniki Bibliografie Biografie Informatory Księgoznawstwo Bibliotekarstwo Kultura Pedagogika ...
	Uniwersytet Jagielloński	Kolekcja umożliwiająca Jagiellońskiej Bibliotece Cyfrowej pełnienie funkcji repozytorium dla całego uniwersytetu. Grupuje dokumenty, które powstały w Uniwersytecie Jagiellońskim, ze szczegółowym przypisaniem do konkretnej jednostki organizacyjnej. Zawiera także opisy (wraz z odnośnikami) publikacji zamieszczonych na stronach poszczególnych jednostek, tworząc jeden punkt dostępu dla rozproszonych zasobów uczelni.	Wydział Prawa i Administracji Wydział Lekarski Wydział Farmaceutyczny Wydział Nauk o Zdrowiu Wydział Filozoficzny ... Archiwum Uniwersytetu Jagiellońskiego ...
	Kolekcje tematyczne	Kolekcja grupująca zbiory stanowiące całości warte specjalnego wyeksponowania.	Kolekcja berlińska Fryderyk Chopin Jan Paweł II Biblioteka profesora Janusza Skalskiego
	Projekty	Kolekcja odzwierciedla podział publikacji ze względu na źródło finansowania digitalizacji lub samego przygotowania materiału do udostępnienia w sieci.	EFRR POIiŚ 11.1 EFRR MRPO 1.2

Tab. 1. Kolekcje Jagiellońskiej Biblioteki Cyfrowej.

matów metadanych polskich bibliotek cyfrowych [Potęga, Wróbel 2010; Myszowski 2012], wykorzystując efekty prac nad schematem PLMET oraz nowe funkcjonalności programu dLibra (od wersji 5.0). Wszystkie te czynniki złożyły się na ostateczny kształt schematu metadanych używanego w Jagiellońskiej Bibliotece Cyfrowej, który przedstawiono na il.1.

Podstawowy zestaw elementów Dublin Core (w całości zachowany w schemacie JBC) jest szeroko opisany – zarówno od strony teoretycznej [np. Nahotko 2004b, s. 118–131], jak i praktycznej [Domowicz i in. 2008] – zatem charakterystyka poszczególnych atrybutów nie wydaje się konieczna. Nie wymagają jej także elementy dodatkowe znane z innych bibliotek cyfrowych, takie jak: *Wariant tytułu*, *Miejsce wydania/powstania*, *Abstrakt*, *Identyfikator dokumentu cyfrowego*, *ISBN*, *ISSN*, *Lokalizacja oryginału*, *Właściciel praw*, *Źródło finansowania*. Są to atrybuty, którymi twórcy bibliotek bardzo często wzbogacają zestaw wyjściowy, a których nazwy mówią właściwie same za siebie.

Do modyfikacji mniej oczywistych należą elementy zgrupowane w atrybutach *Praca dyplomowa* (*Wydział*, *Instytut/Zakład/Katedra*, *Promotor*, *Recenzent*, *Data obrony*) oraz *Powiązania* (*Tytuł poprzedni*, *Tytuł następny*, *Dodatki*, *Wydawnictwo główne*, *Inne wydania*, *Uzupełnienie zasobu*, *Katalog komputerowy*). Pierwszy z wymienionych zbiorów atrybutów przygotowano z myślą



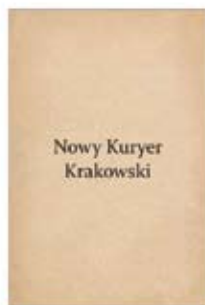
Il. 1.

Il. 1. Schemat metadanych Jagiellońskiej Biblioteki Cyfrowej (atrybuty zaznaczone ramkami to elementy spoza podstawowego schematu).

Il. 2a. Powiązanie opisu publikacji w Jagiellońskiej Bibliotece Cyfrowej z KKZBUJ.

Il. 2b. Powiązanie KKZBUJ z zasobem Jagiellońskiej Biblioteki Cyfrowej.

Il. 3a.



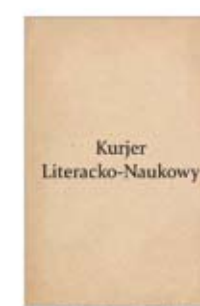
Il. 3b.



Il. 4a.



Il. 4b.



o wykorzystaniu Jagiellońskiej Biblioteki Cyfrowej jako repozytorium prac dyplomowych pisanych w Uniwersytecie Jagiellońskim.

Szerzej warto omówić drugi zbiór, który stworzono, by wzbogacić opisy publikacji prezentowanych w JBC o wygodne interaktywne powiązania. Najprostszym tego rodzaju zabiegiem jest dodanie do opisu odnośnika do katalogu komputerowego tradycyjnej biblioteki (il. 2a). Takie odesłanie może pełnić kilka funkcji. Przede wszystkim stanowi uzupełnienie opisu o informacje, które nie zostały uwzględnione w opisie obiektu cyfrowego⁴. Wskazuje również, że biblioteka cyfrowa nie działa w oderwaniu od biblioteki tradycyjnej, a jest – parafrazując tytuł artykułu Marka Nahotki [Nahotko 2004a] – jej „młodszą siostrą” i obie sta-

Tego rodzaju odnośniki można spotkać w wielu bibliotekach cyfrowych, chociaż do niewielu z nich można znaleźć odnośniki zwrotne w samym katalogu (il. 2b), co przecież w znacznym stopniu ułatwia pracę czytelnikom (może nieświadomym istnienia biblioteki cyfrowej) poszukującym materiałów właśnie w katalogu komputerowym biblioteki⁵.

Powiązania umieszczane w polach *Tytuł poprzedni*, *Tytuł następny*, *Dodatki* oraz *Wydawnictwo główne* dotyczą czasopism – ten typ zasobu stanowi większość prezentowanych w JBC materiałów – i służą usprawnieniu poruszania się po zasobach biblioteki cyfrowej. Rozwiązanie to zostało opracowane na podstawie takich samych pól w komputerowym katalogu bibliotecznym. Jego wdrożenie

⁴ Metadane w serwisie prezentującym pełne teksty służą głównie odnalezieniu i identyfikacji zasobu. Wiele informacji niezbędnych w wypadku opisu katalogowego czytelnik może odnaleźć bezpośrednio w treści publikacji. Możliwość powiązania opisu w bibliotece cyfrowej z katalogiem bibliotecznym pozwala ograniczyć dublowanie informacji, równocześnie dając zainteresowanym użytkownikom możliwość przejścia pełnego opisu.

(z pożądaną funkcjonalnością) w dLibrze było możliwe dopiero od wersji 5.0. W wersjach wcześniejszych wszystkie odnośniki otrzymywały automatyczną etykietę „kliknij tutaj, żeby przejść”, co nie pozwalało na dokładne poinformowanie czytelnika, do czego zostanie odesłany. Od wersji 5.0 redaktor ma możliwość nadawania odnośnikom zamieszczanym w opisie publikacji własnych etykiet. Praktyczne zastosowanie tego rodzaju powiązań pokazują przykłady na il. 3a-4b.

Czasopism dotyczy także pole *Uzupełnienie zasobu* (choć może również służyć np. wydawnictwom wielotomowym), jednak odnośniki w nim umieszczone nie odsyłają do zasobu w JBC, tylko do innych bibliotek cyfrowych. Ma to miejsce wtedy, kiedy zasób opublikowany

w Jagiellońskiej Bibliotece Cyfrowej nie jest kompletny, a inna biblioteka cyfrowa udostępnia brakujące elementy zbioru (uzasadnieniem publikowania w JBC są w takiej sytuacji braki w danym serwisie, których uzupełnienie stanowią właśnie zbiory JBC). Dzięki temu czytelnik nie musi szukać brakujących fragmentów zasobu na własną rękę – takie wyszukiwanie zostało za niego przeprowadzone przez redaktorów JBC. Przykład tego rodzaju powiązania przedstawiono na il. 5.

Pole *Inne wydania*, dedykowane głównie drukom, łączy w sobie cechy poprzednich pól. Zamieszczane w nim odnośniki odsyłają zarówno do zasobów JBC (il. 6), jak i innych bibliotek cyfrowych (il. 7).

Il. 3a. Powiązanie czasopisma z tytułem następnym.

Il. 3b. Powiązanie czasopisma z tytułem poprzednim.

⁵ Łukasz Mesek nazywa tego rodzaju powiązania „sprzężeniem zwrotnym” [Mesek 2012, s. 399–401].

Il. 4a. Powiązanie czasopisma z dodatkiem.

Il. 4b. Powiązanie dodatku do czasopisma z wydawnictwem głównym.

Głos Narodu : dziennik polityczny, społeczny i literacki. 1893

Struktura publikacji:

- ▼ Głos Narodu
 - ▼ Głos Narodu : dziennik polityczny, społeczny i literacki. 1893
 - ▶ Głos Narodu : dziennik polityczny, społeczny i literacki. 1893, nr 1
 - ▶ Głos Narodu : dziennik polityczny, społeczny i literacki. 1893, nr 2
 - ▶ Głos Narodu : dziennik polityczny, społeczny i literacki. 1893, nr 3
 - ▶ Głos Narodu : dziennik polityczny, społeczny i literacki. 1893, nr 4
 - ▶ Głos Narodu : dziennik polityczny, społeczny i literacki. 1893, nr 6
 - ▶ Głos Narodu : dziennik polityczny, społeczny i literacki. 1893, nr 7
 - ▶ Głos Narodu : dziennik polityczny, społeczny i literacki. 1893, nr 8
 - ▶ Głos Narodu : dziennik polityczny, społeczny i literacki. 1893, nr 10
 - ▶ Głos Narodu : dziennik polityczny, społeczny i literacki. 1893, nr 11
 - ▶ Głos Narodu : dziennik polityczny, społeczny i literacki. 1893, nr 12
 - ▶ Głos Narodu : dziennik polityczny, społeczny i literacki. 1893, nr 13
 - ▶ Głos Narodu : dziennik polityczny, społeczny i literacki. 1893, nr 14
 - ▶ Głos Narodu : dziennik polityczny, społeczny i literacki. 1893, nr 15
 - ▶ Głos Narodu : dziennik polityczny, społeczny i literacki. 1893, nr 16
 - ▶ Głos Narodu : dziennik polityczny, społeczny i literacki. 1893, nr 17
 - ▶ Głos Narodu : dziennik polityczny, społeczny i literacki. 1893, nr 18
 - ▶ Głos Narodu : dziennik polityczny, społeczny i literacki. 1893, nr 21
 - ▶ Głos Narodu : dziennik polityczny, społeczny i literacki. 1893, nr 22
 - ▶ Głos Narodu : dziennik polityczny, społeczny i literacki. 1893, nr 23
 - ▶ Głos Narodu : dziennik polityczny, społeczny i literacki. 1893, nr 24
 - ▶ Głos Narodu : dziennik polityczny, społeczny i literacki. 1893, nr 25

II. 8.

bez podawania jego numeru, ale umożliwia również odnalezienie wszystkich czasopism opatrzonych wybraną datą – udogodnienie z pewnością istotne dla historyków⁶, uwzględnienia zmian w tytułach czasopism, a także zmian wydawców i redaktorów⁷.

Przygotowanie tak szczegółowych opisów wymaga dużego nakładu pracy. Znacznym ułatwieniem dla redaktora

biblioteki cyfrowej jest szczegółowy i kompletny opis bibliograficzny w katalogu komputerowym biblioteki. Redaktor JBC, posiłkując się opisem katalogowym, jest w stanie przygotować wstępne dane dla całego tytułu, poszczególnych roczników oraz pojedynczych numerów. Dotyczy to wspomnianych zmian w tytułach oraz zmian redaktorów i wydawców. Przygotowane przez specjalistę od katalogowania informacje są przez redaktora

II. 8. Przykład struktury czasopisma w JBC.

⁶ Kwestie techniczne związane z wpisywaniem dat dziennych zostały poruszone w rozdziale dotyczącym oprogramowania.

⁷ Przykładowy wpis z Instrukcji wypełniania pól opisu w schemacie metadanych JBC (pole Współtwórca): „W każdym roczniku należy wpisywać nazwiska redaktorów właściwych danemu rocznikowi (zasada obowiązuje także w polu wydawca). Jeżeli redaktorzy zmieniają się

JBC weryfikowane w momencie, kiedy przegląda on zasób, którym dysponuje (przygotowując go do publikacji). Taki tryb pracy pozwala wyeliminować ewentualne pomyłki i stworzyć opis w pełni zgodny ze stanem faktycznym.

Proces tworzenia metadanych uzależniony jest od typu prezentowanego zasobu. W wypadku dokumentów zwartych czy grafik opis jest importowany z KKZBUJ (mechanizm importu automatycznie przypisuje odpowiednie pola schematu MARC 21 do schematu metadanych JBC), podlega korekcie dopasowującej zapisy do standardów przyjętych w bibliotece cyfrowej, po czym jest uzupełniany o elementy specyficzne dla dokumentów cyfrowych (np. *Identyfikator dokumentu cyfrowego*, *Status prawny*). Proces komplikuje się w wypadku czasopism. Opisany powyżej import z katalogu umożliwia jedynie stworzenie opisu głównej publikacji dla całego tytułu. Kolejne kroki wymagają poważniejszej interwencji redaktora (choć nadal oparte są na katalogu). Po stworzeniu publikacji głównej powstaje struktura czasopisma, czyli podział na roczniki. Jest on przygotowywany na podstawie opisu bibliograficznego oraz rekordu zasobu. Tak przygotowana struktura może zostać wypełniona zawartością. Redaktorzy przygotowują poszczególne numery czasopisma w postaci katalogów (jeden katalog dla każdego numeru) z dokumentami przedstawiającymi czasopismo, plikami RDF z ich opisami oraz plikami PROPERTIES określają-

cymi m.in. umiejscowienie publikacji w odpowiednim miejscu struktury (szczegóły techniczne przygotowania takich plików omówiono w rozdziale na temat oprogramowania). Dostępny w dLibrze mechanizm masowego ładowania publikacji [Parkoła 2008, s. 52–56] dopełnia dzieła – publikacje zostają przesłane do Jagiellońskiej Biblioteki Cyfrowej równocześnie z opisami.

Podsumowanie

Prawidłowe opracowanie metadanych jest jednym z kluczowych warunków funkcjonowania biblioteki cyfrowej. Dla osiągnięcia pożądanego efektów niezbędne są przeprowadzone w sposób przemyślany prace przygotowawcze nad strukturyzacją zasobów, schematem opisu publikacji i wytycznymi dla jego właściwego stosowania. Prowadząc takie prace, należy mieć na uwadze przyszłą funkcjonalność wdrażanych rozwiązań – zarówno dla użytkowników, jak i administratorów. Warto także brać pod uwagę fakt, że nie wszystko da się przewidzieć, gdy projektuje się tak złożony system. Rosnące możliwości prezentacji różnego rodzaju zasobów w Internecie (może nieprzewidywanych na początku istnienia serwisu) sprawiają, że struktura biblioteki cyfrowej powinna pozostawać otwarta. Wiąże się to na przykład z elastycznością podziału na kolekcje. Poszerzanie oferty biblioteki bardzo często wymaga dodawania nowych kolekcji i rozbudowy już istniejących. Potrzeba zmian może dotyczyć również schematu metadanych.

w ciągu roku, każda publikacja powinna mieć wpisane tylko nazwisko sobie właściwe”.

W tym wypadku warto jednak zachować ostrożność. Wydaje się istotne, by trzon zestawu pozostawał niezmienny (z uwagi na przyzwyczajenia użytkowników), można jednak uwzględnić możliwość poszerzenia schematu o nowe pola (zamiast „upychać” nowe informacje w nieprzeznaczonych dla nich atrybutach). Oczywiście we wszystkim tych kwestiach wskazany jest umiar, który pozwoli zachować równowagę pomiędzy usprawnianiem korzystania z zasobów wzbogacaniem opisu a przeładowaniem go nadmiarem informacji, które zamiast ułatwiać pracę, dają wrażenie szumu informacyjnego.

Bibliografia

Bates, Marcia J. (2002). Toward an Integrated Model of Information Seeking and Searching (Keynote Address, Fourth international Conference on Information Needs, Seeking and Use in Different Contexts, Lisbon, Portugal, September 11, 2002). *New Review of Information Behaviour Research*, vol. 3, s. 1–15.

Bednarek, Grzegorz (2007). Warstwy tekstowe w plikach DjVu [online]. Zabrze: GB Soft. [Dostęp: 11.06.2012]. Dostępny w World Wide Web: http://www.djvu.com.pl/OCR/OCR_in_DjVu.php.

Domowicz, Iwona [i in.] (oprac.) (2008). eProwadnik redaktora zasobów cyfrowych [online]. Wrocław: Biblioteka Uniwersytecka. [Dostęp: 11.06.2012]. Dostępny w World Wide Web: <http://www.bibliotekacyfrowa.pl/dlibra/docmetadata?id=25069&showContent=true>.

Kalota, Tomasz; Raczyński, Rafał; Rękar, Paweł (2011). Przetwarzanie i OCR czasopism drukowanych gotykiem krok po kroku. [W:] Mazurek, Cezary; Stroiński, Maciej; Węglarz, Jan (red.) *Polskie biblioteki cyfrowe 2010. Materiały z konferencji zorganizowanej w dniach 20–21 października 2010 roku przez Bibliotekę Kórnicką PAN, Poznańską Fundację Bibliotek Naukowych, Poznańskie Centrum Superkomputerowo-Sieciowe*. Poznań: Ośrodek Wydawnictw Naukowych, s. 55–62. Publikacja dostępna również – w formie prezentacji – na stronie serwisu digitalizacja.pl: <http://digitalizacja.pl/Tomasz-Kalota/Przetwarzanie-i-OCR-czasopism-drukowanych-gotykiem-krok-po-kroku>.

Mesek, Łukasz (2012). Korelacje między opisami w tradycyjnych i cyfrowych katalogach bibliotek. [W:] Janiak, Małgorzata; Krakowska, Monika; Próchnicka, Maria (red.) *Biblioteki cyfrowe*. Warszawa: Wydawnictwo SBP, s. 396–402.

Myszowski, Piotr; Szafranski, Leszek (2011). Sposób prezentacji czasopisma w bibliotece cyfrowej [online]. Kraków: Biblioteka Jagiellońska [Dostęp: 11.06.2012]. Dostępny w World Wide Web: <http://jbc.bj.uj.edu.pl/dlibra/docmetadata?id=85908&from=publication>.

Myszowski, Piotr (2012). Metadane w polskich bibliotekach cyfrowych. [W:] Janiak, Małgorzata; Krakowska, Monika; Próchnicka, Maria (red.) *Biblioteki cyfrowe*. Warszawa: Wydawnictwo SBP, s. 372–381.

Nahotko, Marek (2004a). Cyfrowa najmłodsza siostra bibliotek. [W:] „Konspekt. Pismo Akademii Pedagogicznej w Krakowie”, nr 19 [online]. Kraków: Akademia Pedagogiczna. [Dostęp: 11.06.2012]. Dostępny w World Wide Web: <http://www.wsp.krakow.pl/konspekt/19/nahotko.html>.

Nahotko, Marek (2004b). Metadane: sposób na uporządkowanie Internetu. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.

Nahotko, Marek (2010). *Komunikacja naukowa w środowisku cyfrowym*. Warszawa: Wydawnictwo SBP.

Parkoła, Tomasz (2008). Podręcznik użytkownika środowiska dLibra (wersja 4.0) [online]. Poznań: PCSS. [Dostęp: 11.06.2012]. Dostępny w World Wide Web: <http://dlibra.psnc.pl/biblio>

teka/dlibra/docmetadata?id=133&from=publication.

Potęga, Joanna; Wróbel, Agnieszka (2010). The Dublin Core Metadata Element Set, Ver. 1.1 a potrzeby i oczekiwania bibliotekarzy cyfrowych – analiza przypadków. [W:] Mazurek, Cezary; Stroiński, Maciej; Węglarz, Jan (red.) Polskie biblioteki cyfrowe 2009. Materiały z konferencji zorganizowanej dnia 9 grudnia 2009 roku przez Bibliotekę Kórnicką PAN, Poznańską Fundację Bibliotek Naukowych, Poznańskie Centrum Superkomputerowo-Sieciowe. Poznań: Ośrodek Wydawnictw Naukowych, s. 71–78.

Rosenfeld, Louis; Morville, Peter (2003). Architektura informacji w serwisach internetowych. Przeł. Krzysztof Masłowski, Tomasz Jarzębowicz. Gliwice: Wydawnictwo Helion.

Sosińska-Kalata, Barbara (2005). Systemy organizacji wiedzy w środowisku sieciowym. [W:] Sosińska-Kalata, Barbara; Przystek-Samokowa, Maria (red.) Od informacji naukowej do technologii społeczeństwa informacyjnego. Warszawa: Wydawnictwo SBP, s. 141–162.

Werla, Marcin (2011a). Dokumentacja schematu metadanych PLMET [online]. Poznań: PCSS [dostęp: 11.06.2012]. Dostępny w World Wide Web: <http://confluence.man.poznan.pl/community/display/FBCMETGUIDE/Dokumentacja+schematu+metadanych+PLMET>.

Werla, Marcin (2011b). Wykorzystanie metadanych z polskich bibliotek cyfrowych. [W:] Mazurek, Cezary; Stroiński, Maciej; Węglarz, Jan (red.) Polskie biblioteki cyfrowe 2010. Materiały z konferencji zorganizowanej w dniach 20–21 paź-

dziernika 2010 roku przez: Bibliotekę Kórnicką PAN, Poznańską Fundację Bibliotek Naukowych, Poznańskie Centrum Superkomputerowo-Sieciowe. Poznań: Ośrodek Wydawnictw Naukowych, s. 125–129.

Masowa ocena skanów

Pierwszym etapem pracy bibliotekarza cyfrowego jest kontrola przeprowadzana w dwóch kierunkach, pod względem kompletności i jakości materiału. Celowość oceny skanów przed przystąpieniem do obróbki graficznej jest uzasadniona kilkoma czynnikami.

Analiza ta, jako część procesu zmierzającego do stworzenia „dobrego” obiektu cyfrowego, wpisuje się w niemal każdy schemat oceny kolekcji cyfrowych. Bibliotekę cyfrową można oceniać na wiele sposobów¹ [Janiak 2010]. Kryteria oceny są zazwyczaj podzielone na kategorie np. w odniesieniu do systemu (funkcjonalność, wydajność) czy też samej kolekcji (zasób). Kluczowym punktem oceny zasobu danej biblioteki cyfrowej jest analiza poszczególnych obiektów pod kątem zakresu, dokładności, kompletności czy przydatności. Droga dedukcji wnioskować należy, że dobrą kolekcję tworzą dobre obiekty, a dobre obiekty tworzą dobre kopie cyfrowe. Właściwie wykonane skany sta-

Słowa kluczowe: ocena skanów, jakość skanów, kompletność skanów

Keywords: digital copy control, quality of scans, completeness of scans



nowią więc podstawę, na której można zbudować (oczywiście w połączeniu z wieloma innymi komponentami takimi jak: odpowiednio dobrane metadane czy skutecznie działający system) przyjazną użytkownikowi kolekcję.

Kontrola i potwierdzenie poprawności kopii cyfrowych pozwala także uniknąć błędów na dalszym etapie pracy. Należy zwrócić uwagę na fakt, że bibliotekarz cyfrowy zapoznaje się z cyfrowymi kopiami oryginalnych dokumentów dopiero w momencie, gdy pobierze je z serwera na

¹ M. Janiak i M. Krakowska omówiły różne modele oceny bibliotek cyfrowych. Zwrócono uwagę na różnice, wynikające z odmiennego podejścia do problemu: systemowego, pragmatycznego oraz psychologicznego.

dysk swojego komputera. Jego fizyczny kontakt z digitalizowanym zasobem jest ograniczony, gdyż proces skanowania bezpośrednio przypada w udziale innym pracownikom. Kolokwialnie mówiąc pobierany folder plików jest jak jajko niespodzianka. Na podstawie numeru DIG² bibliotekarz wie jedynie, jaki typ zasobu pobiera, np. czasopismo, druk zwarty lub muzyczny. Resztę informacji, m.in. o fizycznym stanie zachowania dokumentów, rozmiarze oryginału, informacje



Abstrakt: Analizie poddano proces masowej oceny skanów. Za główne kryteria kontroli kopii cyfrowych przyjęto: jakość i kompletność materiału. Omówiono kluczowe elementy oceny, zwrócono uwagę na pojawiające się problemy oraz zilustrowano wybrane przykłady. Uwzględniono także różnorodność materiału poddanego digitalizacji. Efektem finalnym procesu jest potwierdzenie poprawności kopii cyfrowych.

Abstract: The process of mass evaluation of digital copies is presented. Quality and completeness are the main criteria used in verifying scans. Key elements of evaluation are discussed, considering the arising problems as well as providing instructive examples. The diversity of the material selected for digitization is also taken into account. The final objective of this process is to ensure integrity of digital copies.

o tym, czy obok tekstu pojawiają się fotografie, rysunki i czy są one jedno- czy wielokolorowe, gromadzi przeglądając pliki.

Analiza kopii cyfrowych przed wykadrowaniem jest bardzo ważna również dlatego, że to właśnie te pliki, jako kopie-matki³ [Ober 2005], zostają zabezpieczone w procesie archiwizacji. Są to surowe obrazy, tzw. pliki master, pozyskiwane bezpośrednio ze skanera i niepoddawane żadnej obróbce graficznej. Zgodność z oryginałem i wysoka jakość tych odwzorowań nie podlegają więc dyskusji.

Masowa ocena skanów nabiera znaczenia także w kontekście złożonego procesu, którego jest częścią. Poszczególne etapy pracy zespołu JBC były bardzo od siebie zależne, wręcz się zaciębiały. Bibliotekarz cyfrowy jest odpowiedzialny za pobierany materiał do momentu jego publikacji na platformie dLibra. Zanim jednak pojedyncza publikacja trafia do publicznego udostępnienia na stronie internetowej (co jest ostatnim etapem pracy) trzeba wykonać szereg czynności jakimi są: masowa ocena skanów, kadrowanie, segregowanie, nadanie odpowiednich nazw plików, masowa obróbka graficzna w XnView oraz konwersja do formatu prezyencyjnego. Podczas wstępnej analizy materiału, jakim jest masowa ocena skanów, pozyskiwane są informacje istotne także na dalszym etapie obróbki. Oceniając kopie cyfrowe pracownik gromadzi informacje o jakości i gramaturze papieru, zaplamionych stronach, źle wydrukowanym

² Przejrzyste zdefiniowanie nazw kopii cyfrowych dokumentów odgrywa bardzo istotną kwestię w sprawnym zarządzaniu zasobem kopii cyfrowych. W JBC procedura dotycząca nazw katalogów i plików oparta jest o numery DIG. W zależności od rodzaju dokumentu do przedrostka DIG dodawane są kolejne człony. Przykładowo dla druku muzycznego – DIGMUZ, zaś dla czasopisma – DIGZAS. Szersze omówienie stosowanego

w JBC nazewnictwa plików jest przedmiotem artykułu Anny Krzak.

³ Ober, Krzysztof (2005): „Kopia-matka powinna zaspokajać cały zakres potrzeb użytkowników obejmujący tworzenie dalszych kopii cyfrowych na potrzeby wydruku, wyświetlania i przetwarzania obrazów. Im bogatsza jest kopia-matka tym lepsza pod względem jakości i możliwości przetwarzania może być jakość kolejnych

tekście (słabo odbitym) czy urwanych fragmentach stron. Wszelkie tego typu informacje są bardzo przydatne podczas dalszej obróbki materiału, szczególnie planowania korekty graficznej w XnView i doboru odpowiednich parametrów konwersji do formatu DjVu. Gdyby kolejne z tych czynności były rozdzielone na poszczególnych pracowników w pewnym stopniu dublowaliby oni swoją pracę. Przykładowo podczas segregowania skanów (podziału plików na poszczególne numery czasopism) pracownik, sprawdzający wcześniej kompletność zasobu, dysponuje już wiedzą odnośnie np. braku danych numerów, braku stron, źle wklejonych stron, dodatków niesamoistnych lub samoistnych. Tym samym poszczególne etapy pracy przebiegają o wiele sprawniej. Model pracy przyjęty w Jagiellońskiej Bibliotece Cyfrowej nie jest oczywiście jedynym słusznym podziałem obowiązków. W zależności od wielkości instytucji oraz ilości zbiorów przeznaczonych do digitalizacji zakres obowiązków na poszczególnych stanowiskach pracy może być różny [Paradowski 2010, s. 62].

Trzeba zaznaczyć, że w Jagiellońskiej Bibliotece Cyfrowej ocenie podlegał dwójaki rodzaj materiału. Z jednej strony były to zbiory digitalizowane na bieżąco w ramach projektu Jagiellońska Biblioteka Cyfrowa – od końca 2010 r. odbywało się skanowanie unikatowych w skali kraju kolekcji czasopism i gazet XIX–XX w., których łącznie było ponad tysiąc woluminów (1222). Drugą zaś

grupą dokumentów był obszerny zbiór książek i jednostek zbiorów specjalnych, których digitalizacja dokonywana była w BJ sukcesywnie od 2003 do 2008 r., a które zostały włączone do projektu. Wśród nich znajdowały się stare druki, rękopisy, zbiory muzyczne, graficzne i kartograficzne (łącznie 1644 woluminów książek i jednostek zbiorów specjalnych). Włączenie ich do projektu zostało podyktowane nie tylko troską o zabezpieczenie kopii cyfrowych, zapisanych na mało trwałym nośniku (płyty CD bądź DVD), ale również potrzebą ich udostępnienia. Różnicę pomiędzy dwoma rodzajami materiału widać, nie tylko w typie dokumentów, ale także w sposobie pracy z nimi. Wydawnictwa ciągle zaraz po skanowaniu podlegały analizie. W razie błędów bądź braków wolumin odsyłany był do poprawy i zmiany nanoszone były na bieżąco. Uzupełnienie skanów, bądź ich wymiana następowała szybko, gdyż woluminy znajdowały się w pracowni fototechnicznej do momentu potwierdzenia poprawności skanowania. Wykonanie dodatkowych skanów z woluminu nie stanowiło więc żadnego problemu. W przypadku materiałów zgromadzonych na płytkach procedura poprawy wadliwych skanów, czy też ich uzupełnienie, zajmowało więcej czasu, gdyż konieczne było zamawianie poszczególnych pozycji z magazynu bądź nawet ze skarbca Biblioteki Jagiellońskiej.

Ze względu na to, że w realizowanym projekcie znaczna część to wydawnictwa ciągle ukazujące się przez wiele lat

kopii. Należy wziąć również pod uwagę rosnące potrzeby użytkowników tworząc wystarczająco dobrą kopie-matkę tak, aby mogła zaspokoić ich przyszłe potrzeby. Kopie takie będą pomocne w rozwijaniu zasobów dziedzictwa kulturowego, użytecznych w różnych dziedzinach życia i dla różnych użytkowników i instytucji. Tworzenie wysokiej jakości obrazu cyfrowego może być związane z większym kosztem początkowym, ale będzie mniej

kosztowne niż wytworzenie obrazu niższej jakości, który nie spełni długoterminowych wymagań i będzie musiał być zastąpiony przez lepszą jakościowo kolejną kopię”.

z dużą częstotliwością, procesowi oceny ich kompletności poświęcono najwięcej uwagi. W pierwszej kolejności pliki należało pobrać na dysk lokalny komputera. Eksport danego woluminu z systemu dokonywany był za pośrednictwem SOJBC. W każdym przypadku system zapamiętywał nazwisko pracownika pobierającego dany wolumin.

Jak już wspomniano, kontrola kopii cyfrowych przebiegała dwutorowo. Pliki podlegały ocenie pod względem kompletności oraz jakości.

Ocena kompletności skanów

Kontrola kompletności kopii cyfrowych bywa czasami kłopotliwa. Aby działanie to przebiegło sprawnie i dokładnie, skanowanie w Sekcji Reprografii poprzedzone było dokładną analizą każdego woluminu, dokonywaną przez pracownika Oddziału Magazynów Biblioteki Jagiellońskiej. Przeglądano każdy numer wydawnictwa ciągłego, a uwagi w wersji rozszerzonej notowano w SOJBC i częściowo, dla potrzeb czytelnika, w rekordzie zasobu w KKZBUJ. Uwagi dotyczyły głównie braków numerów, braków stron, pozornych braków stron (gdy strony omyłkowo wklejone są w złym miejscu), dubletów, dodatków samoistnych i niesamoistnych, błędów drukarskich. Informacje tego rodzaju są niezbędne bibliotekarzowi cyfrowemu w ocenie kompletności woluminu. Dzięki nim może dokonać kontroli nie mając nawet w rękach dokumentów oryginalnych.

Kluczowe w masowej ocenie kompletności kopii cyfrowych jest:

- rozróżnienie braków fizycznych od tych wynikających z winy osób skanujących. W tym celu bibliotekarz cyfrowy analizuje paginację, tj. numerację kolejnych stron wydawnictwa. Jeśli zauważony zostanie brak stron, który nie został wyłuszczone przez pracownika magazynu jako brak fizyczny, wolumin kwalifikowany jest do poprawy, tj. doskanowania pominiętych stron. Dotyczy to także stron niezadrukowanych, najczęściej rewersów, które zostały opuszczone, bądź przypadkiem bądź ze względu na niską ocenę ich przydatności. Podobna procedura następuje w przypadku, gdy skanujący zdubluje jakieś strony (wolumin jest odsyłany w celu usunięcia niepotrzebnego skanu).

- zwrócenie uwagi na błędy introligatorskie popełniane podczas oprawy danego tytułu. Najczęściej występującymi błędami jest złe ułożenie kolejnych stron. Skutkiem tego są np.: strony przestawione (strony z numeru 56 są na końcu numeru 57), strony wklejone odwrotnie (np. strona 8 przed 7), a czasami nawet strony pochodzące z innego woluminu (nr 25 z 1901 r. oprawiony jest po nr 24, ale w roczniku 1902). Pomyłki tego rodzaju pozwala ujawnić analiza paginacji czasopisma. Nie zmienia to faktu, że utrudniają one ocenę kompletności kopii cyfrowych i nastęrczają dodatkowych trudności podczas segregacji plików. Brak stron w numerze, a czasem nawet

brak całego numeru, błędnie sugeruje brak fizyczny, bądź błąd osoby skanującej, faktycznie jednak jest to tylko pozorana niezgodność.

- wychwycenie błędów drukarskich. Najczęściej pojawiające się błędy drukarskie to błędy w paginacji, które sugerują braki stron (np. 1, 2, 3, 6) lub suponują błąd introligatora w postaci wklejenia stron w złej kolejności. Tego typu błędy można zweryfikować tylko analizując treść dokumentu i stwierdzając jego logiczną zgodność. Z innych błędów drukarskich, z jakimi można się spotkać, wymienić warto także błędy w numeracji czasopisma bądź dacie dziennej wydania, które można zweryfikować znając ciągłość i częstotliwość ukazywania się czasopisma.

Badanie kompletności kopii cyfrowych czasopisma opiera się głównie na analizie paginacji. Sytuacja, w której mamy do czynienia z całkowitym brakiem paginacji na kolejnych stronach, powoduje znaczne utrudnienia. Choć tytuły wydawnictw ciągłych, w których redaktorzy nie zatroszczyli się o numerację stron, należą do rzadkości, tak samo podlegają one ocenie. Kompletność takiego zasobu można zweryfikować na dwa sposoby. Gdy jest pewność, co do stałej i niezmienniej objętości numerów w danym tytule, można samodzielnie liczyć strony. Gdy nie ma reguły, co do liczby stron i poszczególne numery różnią się objętością, kompletność kopii cyfrowych można sprawdzić tylko na podstawie zgodności

tekstu. Oczywiście w przypadku braku paginacji analiza zajmuje bibliotekarzowi cyfrowemu znacznie więcej czasu.

W przypadku dokumentów innych niż wydawnictwa ciągłe oraz zwarte, gdzie stosowanie paginacji jest niemal normą, sprawdzanie kompletności kopii cyfrowych starodruku, rękopisu czy druku muzycznego jest bardziej kłopotliwe. W dokumentach sprzed kilku wieków nie można mówić o paginacji w dzisiejszym tego słowa znaczeniu. Jednak już w rękopisach i starych drukach pojawiają się jej pierwowzory, o które można oprzeć analizę ich kompletności.

Od V w. zwoje pergaminowe zostały wyparte przez kodeks. Od tego czasu jednostką piśmienniczą stała się składka, będąca częścią kodeksu. Najczęstszą składkę stanowiły cztery arkusze, czyli osiem kart. By oznaczyć właściwą kolejność składek w kodeksie oznaczano je literami lub cyframi, umieszczanymi na ostatniej stronie składki, czyli kustoszami. Od IX w. zachowaniu odpowiedniej kolejności składek podczas oprawy służyły także reklamanty, które pojawiały się w rękopisach już w IX wieku, ale rozpowszechniły się dopiero w XIII w. [Bieńkowska 2005, s. 55]. Były to całe słowa lub sylaby wyrazów składki następnej, pisane na końcu składki poprzedniej. Kiedy pojawiła się książka drukowana reklamanty zaczęto nazywać kustoszami, a kustosze rękopisów sygnaturą [Głombiowski 1979, s. 101-106]. Oznaczanie kolejności poszczególnych

II. 1.



kart w składce upowszechniło się dopiero w XIV w. i nazywane jest foliowaniem. Strony parzyste numerowano cyframi rzymskimi bądź arabskimi, dodając określenia recto (przednia) i verso (tylna strona karty).

Kontrolę kompletności kopii cyfrowych można oprzeć także na informacjach zawartych w rekordzie bibliograficznym analizowanego dzieła. Częścią składową każdego rekordu jest opis fizyczny, który zawiera informacje, m.in. o liczbie stron numerowanych i nienumerowanych. Jeśli zeskanowane dzieło nie posiada żadnej z wyżej wymienionych form paginacji, jedynym sposobem na zweryfikowanie kompletności kopii cyfrowych pozostaje jednoczesna analiza oryginału.

Ocena jakości skanów

W doborze szczegółowych parametrów technicznych skanowania uwzględniany jest m.in. stan fizyczny zachowania materiału oraz typ i rozmiar dokumentu. Trzeba zwrócić uwagę na różnice w digitalizacji czasopism oraz zbiorów specjalnych, tj. rękopisów, starych druków czy grafiki. Głównym celem tworzenia kopii cyfrowych, w przypadku czasopism, jest zachowanie ich treści. W przypadku zaś zbiorów specjalnych priorytetem powinno być wykonanie ich jak najwierniejszej kopii. Ze względu na unikatowość tych zasobów, szczególnego znaczenia nabiera zatem utrwalenie, nie tylko zawartości treściowej, ale także kolorystyki (barw, skali tonalnej czy kontrastu).

II. 1. Nieodpowiednie ułożenie materiału, które spowodowało pofałdowanie strony.

Kontrola jakości kopii cyfrowych, na pierwszym etapie pracy, ma charakter masowy, dlatego też opiera się na ocenie wizualnej. Nie jest więc połączona z pomiarem ani obliczeniem odpowiednich parametrów liczbowych, za pomocą wzorców barwnych bądź innych charakterystyk. W przypadku czasopism taką ocenę uznano za wystarczającą. Sumienność nakazuje jednak kontrolę zbiorów specjalnych z użyciem wzornika barw, określonym przez producenta sprzętu skanującego, przy pomocy którego można zastąpić subiektywną ocenę - odpowiednim pomiarem [Płoszajski 2008, 101-106].

Podczas masowej oceny skanów kontroli podlega technika, dokładność i poprawność wykonania skanów. Spośród setek obrazów selekcionowane są takie, których jakość obniża złe ułożenie materiału. Do poprawy odsyłane są skany, na których widoczne są następujące efekty:

- pofałdowanie obrazu (il.1). Problem taki występuje, gdy skanowany dokument nie jest płasko rozłożony. Najczęściej, choć nie tylko, dochodzi do tego w przypadku pojedynczych stron woluminie, których format jest większy niż reszta kart (składane rozkładówki). Przyczyną błędu może być także nieodpowiedni ucisk szyby skanera;

- niewystarczające rozłożenie i dociśnięcie rozkładówki. Skutkuje to tym, że tekst przy grzbiecie jest niewidoczny. W przypadku, gdy efekt taki spowodo-

wany jest zbyt ciasno zszytą oprawą i zachodzi obawa, że problem nie dotyczy jednostkowego przypadku, ale znacznej części woluminu, skanerzysta powinien zgłosić potrzebę poluzowania oprawy;

- brak tektury pod uszkodzonymi fragmentami stron (il. 2). Niepożądanym efektem w tym przypadku jest nakładanie się druku ze strony znajdującej się pod spodem (poprzedniej lub następnej). Czarna podkładka, w miejscu urwanego fragmentu strony, unaocznia od razu jej defekt;

- nieprecyzyjne dopasowanie fragmentów stron naderwanych lub całkowicie urwanych, w efekcie czego widoczna jest między nimi przerwa (il. 3);

- zagięte fragmenty stron, które zasłaniają druk (il. 4);

- niedokładne odwzorowanie oryginału, gdy fragment obrazu jest rozmazany, lub „przesunięty” (il. 5). Przyczyną takiego błędu może być jakiś niekontrolowany wstrząs lub zwykłe poruszenie materiału w trakcie skanowania.

Odpowiednie ułożenie materiału ma niebagatelny wpływ nie tylko na estetykę i wizualny odbiór obrazu przez użytkownika, ale przede wszystkim na właściwe rozpoznanie tekstu (OCR) przez komputer [Reitz 2004]. Jakość plików wejściowych w znacznym stopniu wpływa na to, jak oprogramowanie poradzi sobie z odczytaniem poszczególnych znaków.



II.2.

Wszelkiego rodzaju zniekształcenia tekstu, spowodowane złym przygotowaniem dokumentu do skanowania, obniżają skuteczność jego rozpoznania.

Omówione wyżej przykłady mieszczą się w kategorii błędów ludzkich. Oprócz nich jakościowa ocena obrazu eliminuje błędy, za które winę ponosi sprzęt. Zarówno przyczyny występowania błędów skanera, jak i ich rodzaje, są różne.

Najczęściej występującym błędem, z jakim zetknięto się podczas realizacji projektu Jagiellońska Biblioteka Cyfrowa, były aberracje chromatyczne. Efekt aberracji polega na rozszczepieniu się na barwy składowej tak, że na obrazie widocznych jest kilka jej odwzorowań, przesuniętych względem siebie. Problem ten widoczny jest na granicach kontra-

stowych obszarów pod postacią kolorowej obwódki, co bardzo obniża jakość skanów (il. 6 - 7).

Inny problem, za który winę ponosi sprzęt, a konkretnie uszkodzona głowica skanera, to zniekształcenie obrazu. Odzworowana treść jest wtedy pofalowana (il. 8 - 9). Głowica czytająca skanera to element ruchomy, który przesuwa się wzdłuż skanowanego materiału rejestrując kolejne jego partie. Jeżeli głowica jest uszkodzona, jej ruch nie jest płynny, w rezultacie czego odwzorowanie obrazu jest nie do zaakceptowania.

Oprócz wspomnianych wyżej problemów, zauważono także pojawianie się na kopiach cyfrowych poziomych i pionowych pasków jedno- lub wielokolorowych (il. 12-14). Trudność w wychwyce-

II.2. Nakładanie się tekstu z sąsiedniej karty.



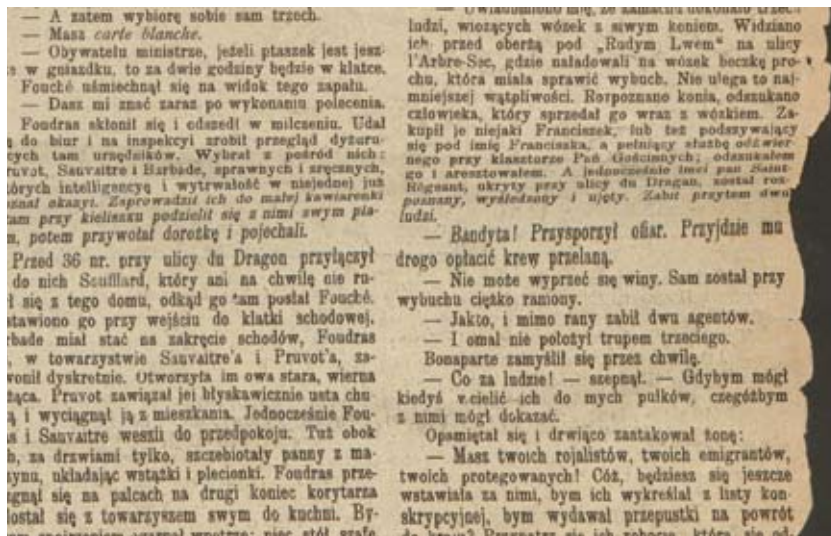
II.3.

II.3. Strona uszkodzona - urwane fragmenty nie przystają do siebie.

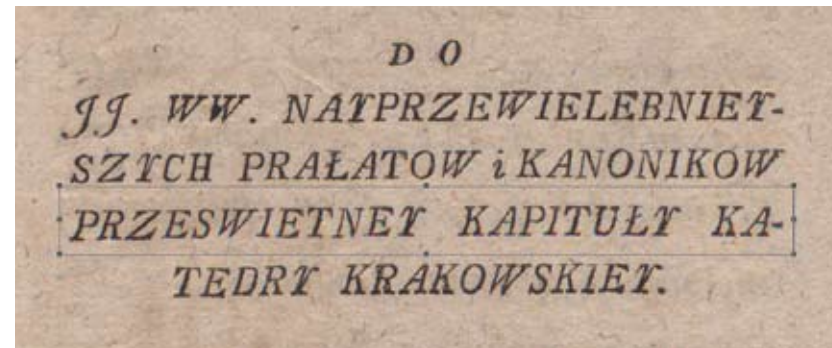


II.4.

II.4. Zagięty fragment strony.

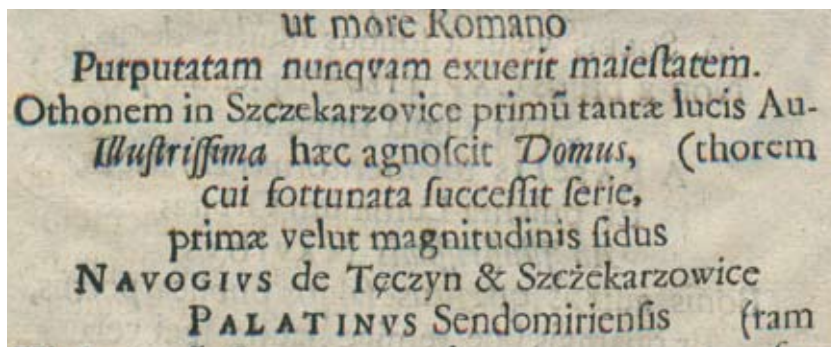


II. 5.



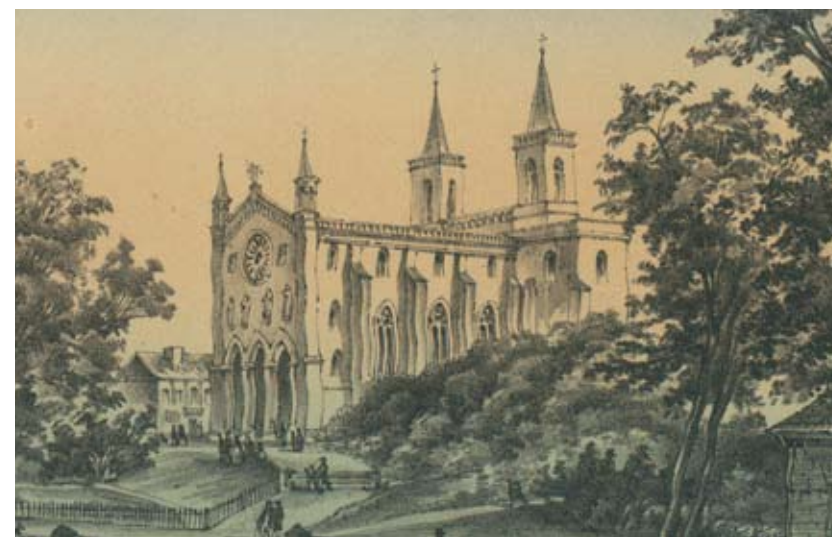
II. 7.

II. 6.



- II. 5. Błąd spowodowany przypadkowym wstrząsem lub poruszeniem skanowanego dokumentu.
- II. 6. Przykład aberracji chromaticznej o dużym natężeniu.

II. 8.



- II. 7. Przykład aberracji chromaticznej o mniejszym natężeniu.
- II. 8. Zniekształcenie obrazu wynikające z uszkodzenia głowicy skanera.

II. 9.

WYCIĄG
z metryk urodzin (chrztu) mężczyzn urodzonych w roku _____

Nr. paragrafu kawy	Imię i nazwisko	Data i miejsce urodzenia	a) Nazwisko i imię ojca b) Nazwisko panieńskie i imię matki	Zawód względnie zatrudnienie ojca	U w a g i
1	2	3	4	5	6

II. 10.



II. 11.

gua Portuguesa, en nuestro vulgar Castellano, vn libro intitulado Mixtica theologia: el qual era muy vtil y prouechoso para la republica Christiana, y en el auia des pasado mucho trabajo: y nos pedistes y suplicastes, os mandamos dar licencia y facultad, para que lo pudiese des hazer imprimir y vender con privilegio nuestro, por el tiempo que nos fuessimos seruido. o como la nuestra mer

- II. 9. Zniekształcenie obrazu wynikające z uszkodzenia głowicy skanera.
- II. 10-11. Przykłady bardzo dużej aberracji chromatycznej przy jednoczesnym wadliwym ruchu głowicy skanera.

II. 12.

— 15 —

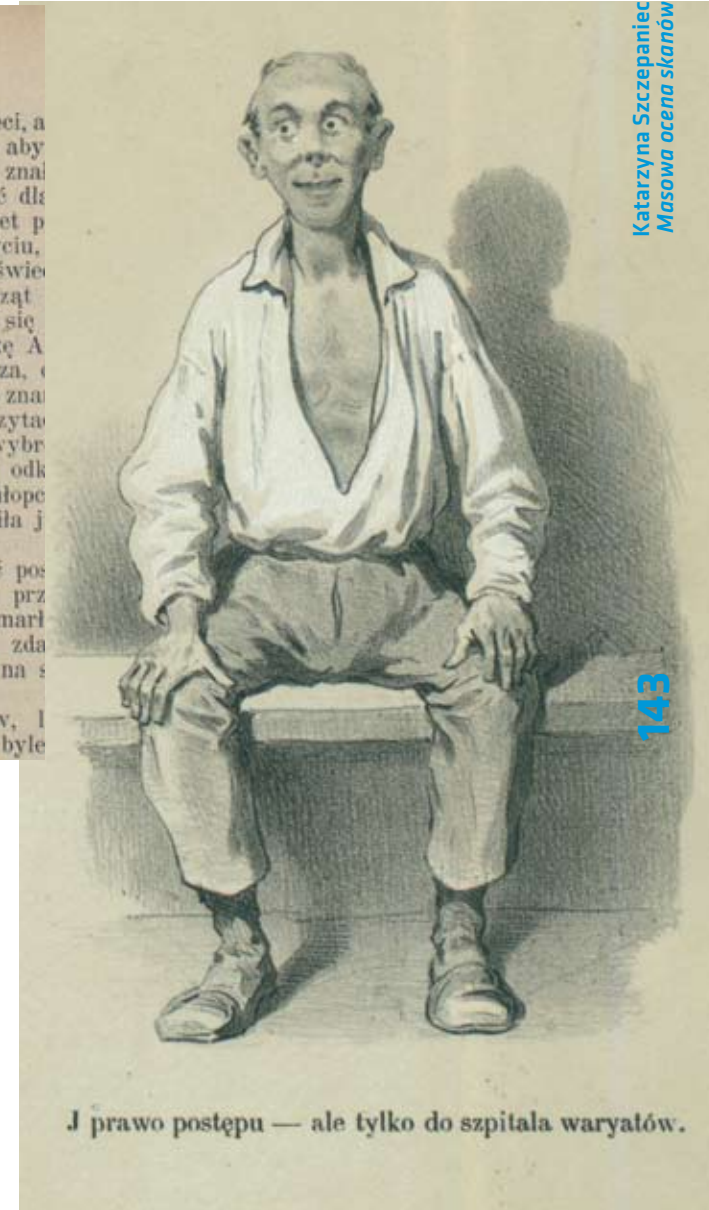
gi na wykształcenie swych dzieci, a na córki, sądząc fałszywie, że aby miała tylko świeży buziaczek i znała ności gospodarskie, to już dosyć dla niedbując często zupełnie nawet p nauki, bez której w dalszem życiu, żonie i matce, ciężko później na świecie.

Wic też choć rój dziewcząt świeżych jak maliny uśmiechał się, on postanowił pojąć za małżonkę A córkę dość zamożnego gospodarza, a nę o rok starszą od siebie, lecz znała ewitości i umiejącą dobrze czytać. Była ona w młodości bardzo wybredna, tego też gdy ten i ów dostał odkórniej jak to zwykle bywa, chłopcy czyli się zupełnie i Anna nosiła ją starej panny.

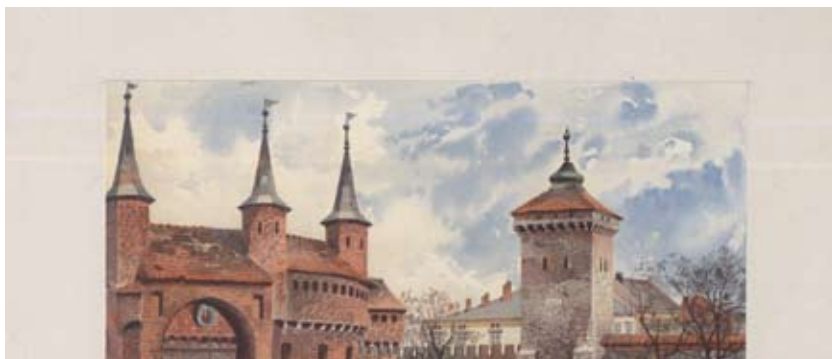
Martwił się tem ojciec, boć posiadał na utrzymanie jedynaczki, przy każdym ojciec, zwłaszcza gdy umarł na, radby był, nie mając syna, zdać oia gospodarstwo, i przy nich na sobie odpocząć.

Próbował różnych sposobów, i wprzód oświadczyła mu, że za byle

II. 13.

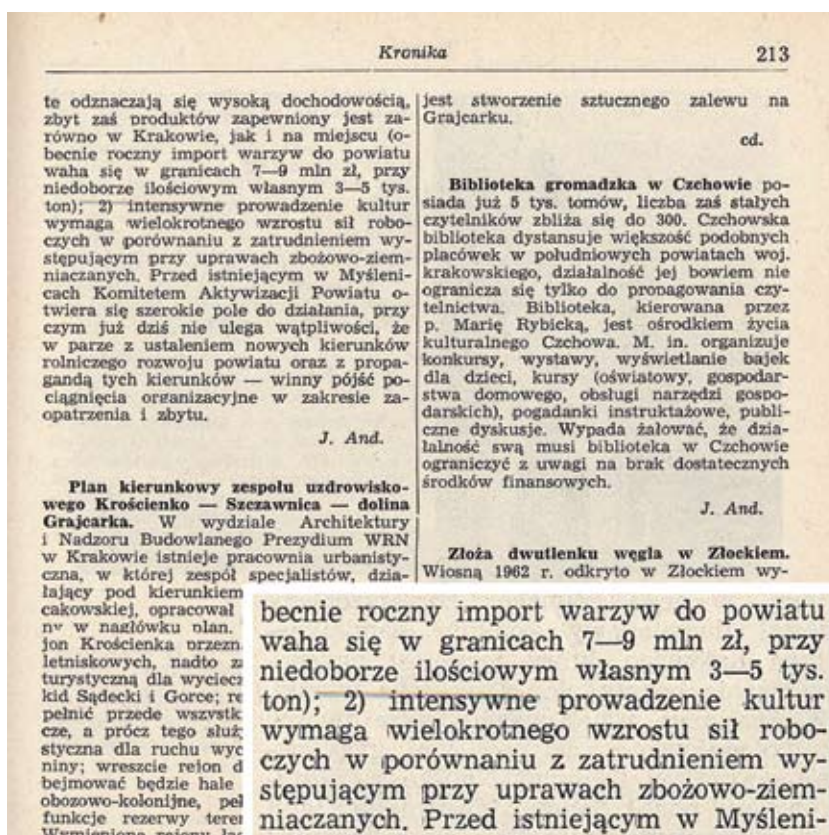


- II. 12-15. Poziome i pionowe paski, jedno i wielokolorowe.



II. 14.

II. 15.

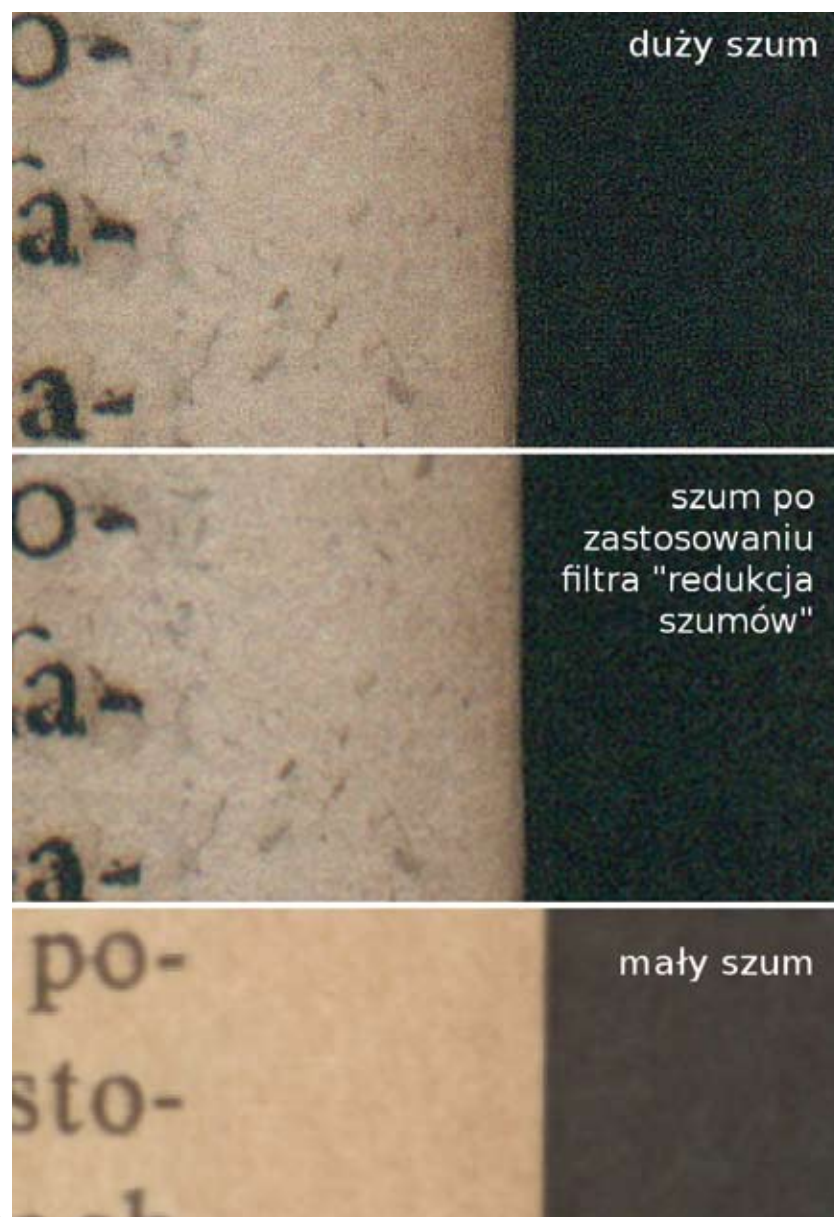


niu takich przypadków jest tym większa, że pasy są na pierwszy rzut oka mało widoczne. Ponadto mają różną szerokość i długość oraz występują w różnych miejscach obrazu. Nie zauważono żadnej prawidłowości lub przyczyny ich występowania. Gdyby problem ten spowodowany był zakłóceniami statycznymi, smugą lub paprochem na szybie skanera, poluzowanym kablem USB, uszkodzeniem żarówki lub innego elementu skanera błąd ten pojawiałby się na wszystkich skanach, wykonanych od momentu wystąpienia problemu. Defekty tego typu pojawiają się jednak z bardzo różną częstotliwością i trudno wyrokować o jakiejś regule.

Problemem, o różnym stopniu nasilenia, są także widoczne szумы na kopiach cyfrowych. Każdy sprzęt generuje szумы, nie jest to więc błąd skanera, ale raczej efekt ograniczonych możliwości sprzętu. W zależności od jakości urządzenia szумы mogą być bardziej lub mniej widoczne. Szum można niwelować odpowiednimi filtrami (np. redukcja szumów XnView), nie można jednak całkowicie się ich pozbyć. Dlatego wybór odpowiedniej klasy urządzenia jest bardzo istotny (il. 15).

W procesie masowej oceny skanów niezwykle istotne są spostrzegawczość oraz dokładność. W wyniku kontroli potwierdzona zostaje poprawność kopii cyfrowych pod względem kompletności i jakości. Pozwala to zmniejszyć do minimum ryzyko wystąpienia błędów na dalszym etapie pracy. Czasami zdarza się, że

mimo wcześniejszej kontroli, zauważony zostaje brak (najczęściej podczas segregowania plików) lub inny błąd (kolorowe paski na skanach dobrze widoczne są dopiero po rozwarstwieniu w DjVu). Każdy błąd jest poprawiany możliwie najszybciej po jego ujawnieniu, niezależnie od tego, na jakim etapie pracy zostanie wychwycony. Warunkiem koniecznym do spełnienia jest dobry efekt finalny, tj. przygotowanie do publikacji obiektu na odpowiednim poziomie.



II. 16.

II. 16. Porównanie szumów występujących na kopiach cyfrowych.

Bibliografia

Bieńkowska, Barbara (2005). Książka na przestrzeni dziejów. Warszawa: Centrum Edukacji Bibliotekarskiej, Informacyjnej i Dokumentacyjnej im. Heleny Radlińskiej.

Głombiowski, Karol; Szejnkowska, Helena (1979). Książka rękopiśmienna i biblioteka w starożytności i średniowieczu. Warszawa: Państwowe Wydaw. Naukowe.

Janiak, Małgorzata; Krakowska, Monika (2010). Ocena bibliotek cyfrowych – kryteria jakości. [W:] Polskie Biblioteki Cyfrowe 2010: materiały z konferencji zorganizowanej w dniach 20-21 października 2010 roku przez Bibliotekę Kórnicką PAN, Poznańską Fundację Bibliotek Naukowych, Poznańskie Centrum Superkomputerowo-Siecioro. Poznań: Ośrodek Wydawnictw Naukowych, s. 101-108.

Ober, Krzysztof (2005). Proces digitalizacji [online]. [Dostęp: 20.05.2012]. Dostępny w World Wide Web: <http://lib.psnc.pl/publication/88>.

Paradowski, Dariusz red. (2010). Digitalizacja piśmiennictwa. Warszawa: Biblioteka Narodowa.

Płoszajski, Grzegorz (2008). Standardy w procesie digitalizacji obiektów dziedzictwa kulturowego. Warszawa: Biblioteka Główna Politechniki Warszawskiej.

Reitz, Joan M. (2004). ODLIS – Online Dictionary for Library and Information Science [online]. [dostęp: 20.05.2012]. Dostępny w World Wide Web: http://lu.com/odlis/odlis_d.cfm.

Masowa korekta graficzna



Słowa kluczowe: biblioteki cyfrowe, konwersja graficzna, masowa korekta graficzna

Keywords: digital libraries, graphic conversion, mass graphic correction

Pliki otrzymane w wyniku procesu skanowania zapisywane są w bezstratnym formacie TIFF. Pozwala on zachować informacje o obiekcie analogowym powstałe w procesie digitalizacji i jest wykorzystywany do archiwizacji kopii cyfrowych [Paradowski 2010, s. 123]. Natomiast w celu opublikowania w Jagiellońskiej Bibliotece Cyfrowej, skany zostają poddane obróbce graficznej, a następnie konwersji do formatu DjVu). Korekta graficzna każdego pliku z osob-

na jest bardzo pracochłonna. Ze względu na skalę prac w JBC oraz charakter digitalizowanych zbiorów, zdecydowano się na obróbkę graficzną wielu plików naraz, przy wykorzystaniu programu XnView.

Wstępna weryfikacja plików otrzymanych ze skanera

Przed rozpoczęciem korekty graficznej otrzymanych skanów, cały obiekt cyfrowy zostaje poddany kontroli. Każdy plik jest sprawdzany pod kątem kompletności i jakości, co pozwala na wyeliminowanie błędów powstałych w procesie skanowania. Zaletą wykonywania kontroli na początku całego procesu jest to, że zanim zacznie się długotrwała i skomplikowana praca, materiał będzie do niej odpowiednio przygotowany.

Kryteria oceny, jakie zostały przyjęte w JBC, są jednoznaczne. Kompletność oznacza wykonanie kopii cyfrowej całego digitalizowanego obiektu. Natomiast ocena jakości obrazów graficznych jest

już problematyczna. W tym celu stosowano jedynie dwustopniową skalę: pliki mogą zostać zakwalifikowane do dalszego procesu albo nie. Czynnikiem dyskwalifikującym są wszelkiego rodzaju błędy skanera (lub osoby skanującej), które zostały szeroko omówione w artykule poświęconym masowej ocenie skanów. Uszkodzone pliki usuwano i zastępowano nowymi, poprawnie wykonanymi, skanami.

Przygotowanie procesu korekty graficznej

Gdy kopia cyfrowa zostanie zakwalifikowana jako poprawna, następuje proces kadrowania plików, stosowany tylko w przypadku czasopism i książek – zbiory specjalne prezentowano w niewykadrowanej formie. Dalsza praca najczęściej odbywa się na mniejszych plikach powstałych w wyniku kadrowania. Pozwala to na zaoszczędzenie czasu i miejsca na przestrzeni dyskowej – masowa korekta graficzna dużych skanów, nawet na wydanym sprzęcie komputerowym, jest bardzo czasochłonna (il. 1).

Dobór filtrów wykorzystywanych do obróbki graficznej zależy zarówno od rodzaju zeskanowanego obiektu, jak również od jego stanu fizycznego. W przypadku czasopism, korekta graficzna miała na celu jak najlepsze przygotowanie plików do dalszej konwersji do formatu DjVu. Zgodność otrzymanego po obróbce skanu z obiektem analogowym nie jest bez znaczenia, jednak schodzi na plan dalszy

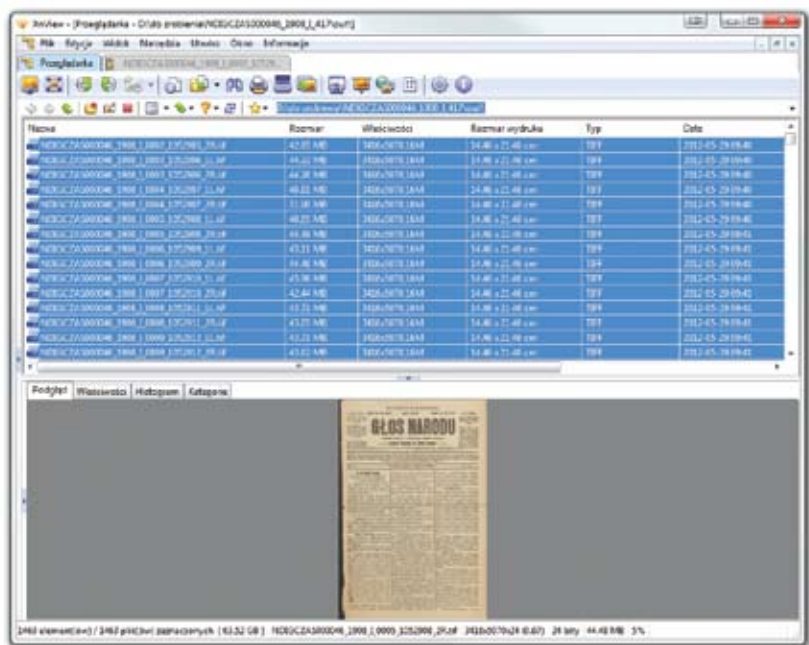
na rzecz poprawnej segmentacji obrazu. Odwrotna sytuacja występuje podczas korekty graficznej zbiorów specjalnych: ich konwersja jest nieco bardziej uproszczona, natomiast nacisk położony jest na efektywne zaprezentowanie rzadkich i cennych obiektów. Ważna jest także zgodność z oryginałem, tzn. kolor karty czy tuszu powinien być zbliżony do fizycznego pierwowzoru. Należy zaznaczyć, że ocena wierności kopii jest kłopotliwa ze względu na różnorodność pro-

Abstrakt: Analizie poddano proces masowej korekty graficznej plików otrzymanych w wyniku skanowania. Zwrócono uwagę na różnorodność materiału poddawanego obróbce graficznej oraz potencjalne problemy, z jakimi można się spotkać na tym etapie pracy w bibliotece cyfrowej. Przedstawiono program XnView oraz szczegółowo omówiono wybrane filtry graficzne. W podsumowaniu zaproponowano gotowy zestaw filtrów do masowej korekty graficznej stosowany w Jagiellońskiej Bibliotece Cyfrowej.

Abstract: The process of mass graphic conversion of files obtained through scanning is thoroughly described. The diversity of items converted to a graphic form and potential problems that may be encountered at this stage of workflow in the digital library are pointed out. The program XnView is presented and selected graphic filters discussed in detail. Finally, a complete set of filters for mass graphic conversion used in the Jagiellonian Digital Library is proposed.



II. 1.



fili kolorów zastosowanych na monitorze i wyświetlarce użytkownika końcowego [Green 2010, s. 53-55].

Materiał pochodzący sprzed kilkadziesiąt lat, jak łatwo sobie wyobrazić, nie jest w „idealnym” stanie. Karty dokumentów bywają zniszczone (il. 2), zakwaszony papier sprawia, że straciły one swoją biel. Tusz, poza swoją degradacją, mógł być źle nałożony – są to częste błędy powstałe w samej drukarni (il. 3).

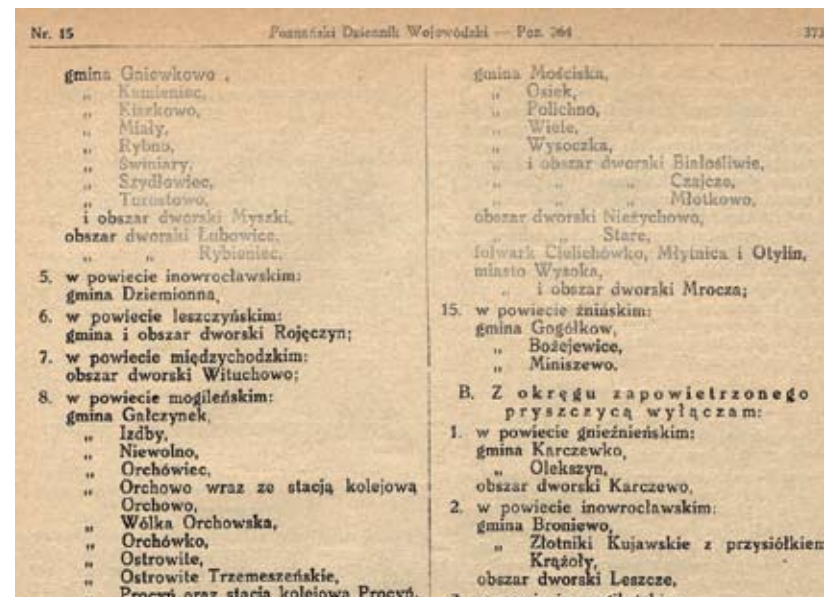
Programem używanym w JBC do masowej korekty jest bezpłatna przeglądarka graficzna XnView (do pobrania ze stro-

ny www.xnview.com). Aplikacja ta jest narzędziem przystosowanym do pracy zarówno na pojedynczych plikach, jak i do pracy masowej (il. 4). Pozwala, między innymi, na wykonywanie takich czynności, jak: kadrowanie, zmiana głębi koloru (odcienie szarości, obraz czarno-biały), korekta jasności, kontrastu, gammy, zmiana rozmiaru i DPI obrazu oraz jego obrót.

II. 1. Wykadowane pliki przygotowane do masowej korekty graficznej.



II. 2.



II. 3.

II. 2. Zniszczona karta „Głosu Narodu” z 1906 r.

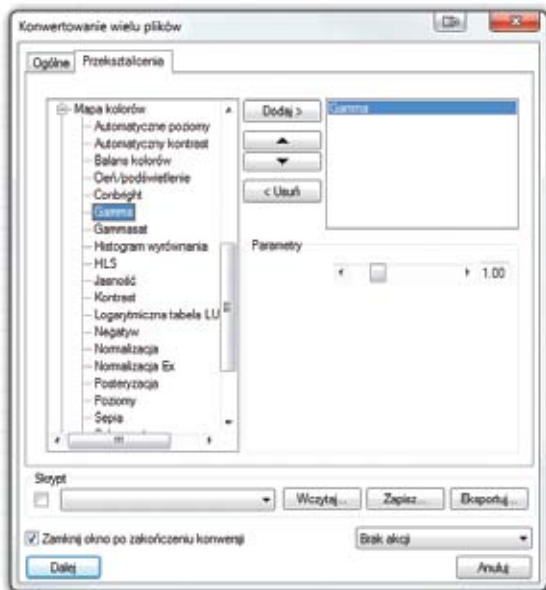
II. 3. Wyblakły tusz w „Poznańskim Dzienniku Wojewódzkim” z 1931 r.

Stosowane filtry graficzne

Materiał poddawany korekcie graficznej w większości przypadków składa się z ciemnego tekstu zapisanego na jasnej karcie. Aby osiągnąć odpowiedni kontrast pomiędzy pikselami przedstawiającymi tekst, a tymi, które obrazują tło, należy uzyskać odpowiednią iluminację poszczególnych pikseli [De Santis, Germani, Jetto 2001, s. 307; Lai, Kuo, Li 1997, s. 898-900].

Zestaw filtrów do masowej obróbki graficznej powinien być tak dobrany, aby można było go zastosować do wielu plików naraz. Z reguły poszczególne strony woluminu wymagają jednego schematu dla poprawy jakości skanu. Jeśli w woluminie znajdują się strony różniące się znacznie od pozostałych, należy je wyodrębnić i ręcznie dokonać obróbki graficznej.

II. 4. ficznej.



II. 4. Konwertowanie wielu plików w programie XnView.



II. 5.

Podstawowy zestaw filtrów wykorzystywany w JBC do masowej korekty graficznej składa się z: gammy, kontrastu, jasności, nasycenia, wzmocnienia ostrości oraz zmiękczenia. Odpowiednie ustawienia powyższych filtrów pozwalają na dostosowanie graficzne większości obiektów prezentowanych w JBC.

Korekcja gamma umożliwi wydobycie szczegółów na mało kontrastowych obrazach w sposób niepowodujący znaczącej zmiany w cieniach i światłach. Wpływa na wszystkie wartości kolorów w obrazie, jednak jego funkcja jest krzywą, dlatego największe zmiany zachodzą w zakresie półcieni. Zwiększenie wartości współczynnika gamma skutkuje rozjaśnieniem obrazu, natomiast zmniejszenie współczynnika poniżej 1.0 powoduje przyciemnienie [Buczyński 2005, s. 64].

Filtr jasność reprezentuje ogólny procentowy udział białego w kolorach obrazu

– maksymalna wartość nadaje pikselom kolor biały. Porównanie działania filtrów gamma i jasność ilustruje zdjęcie (il. 5).

Kontrast to różnica tonów między ciemnymi i jasnymi obszarami obrazu – im wyższa wartość kontrastu, tym te różnice są większe.

Zbyt duże wartości dla któregoś z wymienionych filtrów skutkują zarówno nadmiernymi różnicami w stosunku do oryginału, jak i zmniejszeniem jakości obrazu. Stosując wartości przekraczające przedziały -20-20 (dla jasności i kontrastu) i 0.75-1.35 (dla gammy) trzeba być niezwykle ostrożnym [Sachs 2008, s. 8].

W przypadku wyblakłej karty, na której tekst jest zbyt jasny, by był w pełni czytelny, można zastosować filtr o nazwie „poziomy”. Przedstawia on tony jasne, ciemne i średnie w sposób graficzny (za

pomocą histogramu) i pozwalają korygować ich za pomocą wartości liczbowych. W celu uwydatnienia rozjaśnionych liter można zastosować wyżej wymieniony filtr przyjmując dla niego wartość dolnej granicy ok. 80, a górnej ok. 240-230. Jest to filtr w sposób radykalny zmieniający obraz, dlatego zastosowanie go przy pracy masowej jest raczej niepożądane. W takich przypadkach najlepiej dobrać ustawienia indywidualnie dla każdego pojedynczego obrazu.

HLS to zestaw składający się z trzech filtrów (hue – odcień, lightness – jasność, saturation – nasycenie). HLS jest to jedna z teorii postrzegania kolorów zaakceptowana przez twórców programu XnView. Podczas masowej korekty graficznej stosowany jest tylko jeden filtr z wymienionego zestawu – nasycenie. Jest to stopień czystości lub żywości koloru wyrażony brakiem białego. Kolor o nasyceniu 100 procent nie zawiera białego.

II. 5. Z lewej strony – plik źródłowy; w środku – po zastosowaniu filtra jasność (wartość 50); z prawej strony – po zastosowaniu filtra gamma (wartość 2.0).



II. 6.

II. 7.

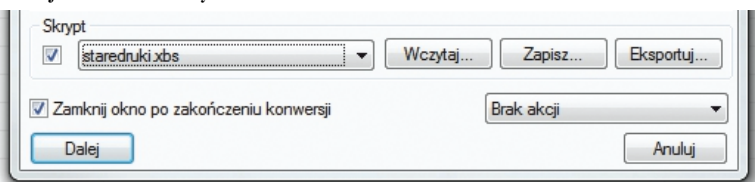


II. 6. Ten sam obraz przed i po zastosowaniu nasycenia (wartość -20).

II. 7. Z lewej strony – plik źródłowy, z prawej strony – plik po zastosowaniu wzmocnienia ostrości.

li, natomiast kolor o nasyceniu 0 procent jest odcieniem szarości. Po zastosowaniu wymienionych wcześniej filtrów (gamma, jasność, kontrast) obraz może zyskać nienaturalnie jaskrawe kolory, dlatego użycie nasycenia z wartością ujemną jest jak najbardziej zasadne (il. 6). Pozwala na zmniejszenie intensywności koloru, jeśli zachodzi taka potrzeba.

Po dobraniu parametrów korygujących kolor obrazu należy poprawić jeszcze jego ostrość. W programie XnView jest do wyboru kilka filtrów: uwydatnienie szczegółów, uwydatnianie krawędzi, wzmocnienie ostrości oraz przywrócenie ostrości. Wyostrażanie obrazu akcentuje krawędzie, ale może również generować szum, co jest efektem niepożądanym. Odpowiedni rezultat można uzyskać po zastosowaniu wzmocnienia ostrości – tekst jest odpowiednio wyostżony, ale nie powoduje to widocznych szumów (il. 7).



Il. 8.

Jeśli jednak kontrast otrzymany po zastosowaniu wzmocnienia ostrości jest zbyt duży, można go częściowo „złagodzić” za pomocą zmiękczenia. Umożliwia ono wygładzanie i tonowanie ostrych krawędzi obrazu bez utraty ważnych szczegółów. Teoretycznie filtry wyostrażające i zmiękczające wykluczają się, jednak w praktyce, zastosowane jednocześnie, pozwalają na uzyskanie obrazu

zblizonego do oryginału.

Opisane wyżej filtry są najczęściej używane do masowej korekty graficznej. W zależności od potrzeb można stosować także inne, np. redukcję szumów czy jaskrawość. Kwestią kluczową jest utworzenie zestawu, który będzie można zastosować wielokrotnie, do różnych woluminów. XnView pozwala na zapisanie

Il. 8. Opcja zapisu i wczytywania skryptu podczas konwersji wielu plików w XnView.



Il. 9.

użytych filtrów (wraz z ich parametrami) w postaci skryptu, który można następnie wybrać z listy i zastosować do poszczególnych plików – bez konieczności ponownego wprowadzania ustawień (il. 8).

Powyżej i poniżej znajdują się przykłady zastosowania omawianych filtrów. Zestaw wykorzystany do obróbki czasopisma składa się z: kontrastu (wartość 27), jasności (-11), nasycenia (-8), zmiękczenia (50) oraz wzmocnienia ostrości (il. 9). Natomiast do korekty skanu XVII-wiecznego woluminu zastosowano kontrast (6), jasność (9), korekcję gammy (0.75), nasycenie (-3), zmiękczenie (50) oraz wzmocnienie ostrości (il. 10).

Bezpłatne oprogramowanie XnView pozwala na wieloaspektowe dostosowanie obrazu na potrzeby publikacji. W Jagiellońskiej Bibliotece Cyfrowej codziennie prace przebiegały na kilkuset plikach. To wymagało rozwiązań, które nie tylko gwarantowały pożądaną jakość, ale zapewniły efektywność pracy. Dzięki opisanym rozwiązaniom masowa korekta graficzna, która jest tylko jednym z etapów przygotowywania obiektu cyfrowego do publikacji, mogła zostać wykonana w odpowiednio krótkim czasie, co zapewniło ciągłość pracy.

Il. 9. Skan czasopisma przed i po obróbce graficznej.



II. 10.

Bibliografia

Buczyński, Ludwik (2005). Skanery i skanowanie. Warszawa: Wydawnictwo Mikom.

De Santis, Alberto; Germani, Alfredo; Jetto, Leopoldo (2001). Space-Variant Two-Dimensional Filtering of Noisy Images. [W:] Aspects of Image Processing and Compression, New York: Academic Press.

Green, Phil (2010). Color Management: Understanding and Using ICC Profiles. Chichester: Wiley.

Lai, Yung-Kai; Kuo, C.-C. Jay; Li, Jin (1997). New Image Compression Artifact Measure using Wavelets. [W:] Visual Communications and Image Processing, Bellingham: Society of Photo-optical Instrumentation Engineers, t. 2.

Paradowski, Dariusz (red.) (2010). Digitalizacja piśmiennictwa. Warszawa: Biblioteka Narodowa.

Sachs, Jonathan (2008). Color Management [online]. [Dostęp: 08.05.2012]. Dostępny w World Wide Web: <http://ftp2.bmtmicro.com/dlc/Color%20Management.pdf>.

Sharma, Abhay (2006). Zrozumieć color management. Warszawa: Ergo BTL.

II. 10. Skan starodruku przed i po obróbkę graficzną.

Kadrowanie i prostowanie - wybór czy konieczność?

Wstęp

Sposób obróbki digitalizowanych materiałów uzależniony jest od ich charakteru. Inaczej będą prezentowane dokumenty i czasopisma XIX- i XX-wieczne, a w inny sposób starodruki i rękopisy. Tomasz Kalota w swoim artykule: *Jak przygotować obiekty do włączenia ich do biblioteki cyfrowej – doświadczenia Biblioteki Uniwersyteckiej we Wrocławiu*, wymienił dwa podstawowe modele bibliotek cyfrowych, dokumentacyjnej i informacyjnej.

Model dokumentacyjny jest stworzony w celu jak najwierniejszego odwzorowania oryginalnego dokumentu. Model informacyjny tworzy się w celu udostępnienia i zachowania warstwy tekstowej i informacyjnej digitalizowanego dokumentu. Wybór modelu biblioteki cyfrowej ma kluczowe znaczenie dla dalszych decyzji dotyczących parametrów technicznych, sprzętu, logistyki i doboru oprogramowania [Kalota, 2007]. W JBC ze względu

Słowa kluczowe: biblioteka cyfrowa, edycja obrazów cyfrowych, kadrowanie, prostowanie, technologia digitalizacji

Keywords: cropping pictures, digital image processing, digital library, digitization technology, straightening

na różnorodność digitalizowanych dokumentów zostały zastosowane oba modele biblioteki cyfrowej. Do archiwizacji plików zastosowany został model dokumentacyjny, a do prezentacji dokumentów w Internecie model dokumentacyjny i informacyjny.

Przygotowanie wersji cyfrowych dokumentów wymaga odpowiedniej obróbki graficznej. W modelu informacyjnym biblioteki cyfrowej podstawową czynnością

przed prezentacją plików w Internecie jest ich wykadrowanie i wyprostowanie. Bardzo często zeskanowane materiały mają dużo niepotrzebnych informacji. Często skanowany oryginał jest wystrzępiony, nierówny na krawędziach lub osoba skanująca krzywo ułożyła go na stole skanera, często również zeskanowany obraz oprócz skanowanego materiału zawiera dodatkowo obszar wokół niego. Dlatego należy przy pomocy odpowiednich programów wykadrować niepo-

trzebne szczegóły i wyprostować obraz.

Według definicji z encyklopedii PWN kadrowanie jest to odrzucenie poza kadr pozytywu nieistotnych lub niepotrzebnych elementów obrazu podyktowane względami kompozycyjnymi lub technicznymi [PWN Encyklopedia 2010]. Według definicji w Wikipedii jest to wycięcie z pierwotnie zarejestrowanego obrazu określonego jego fragmentu w celu uzyskania optymalnego dla widza obrazu, zdjęcia lub pojedynczej klatki fotograficznej lub błony filmowej [Wikipedia 2011].

Natomiast prostowanie jest to obracanie obrazu o niewielki kąt, tak aby prezentowane na nim obiekty lub tekst nie były pochylone [Fotografia 2010].

Prezentacja w Internecie i archiwizacja

Już na etapie skanowania powinno się ustalić parametry dla obrazów wyjściowych. W przypadku skanerów wielkoformatowych należy w trakcie skanowania zostawić marginesy u góry, na dole i po bokach obrazu. Ułatwia to pracę osobie skanującej, bo nie musi zwracać uwagi na to, jak układa skanowany materiał na skanerze i może szybko wyprodukować partie skanów, które w łatwy sposób można kadrować. Natomiast przy wykorzystywaniu skanerów płaskich małoformatowych (maksymalnie A3) należy pamiętać, aby każdorazowo przyciskać skanowany materiał do szyby skanera

oraz uważać, żeby nie wystawał poza obszar roboczy. Obszar skanowania najlepiej ustawić na całą powierzchnię skanera i później odciąć zbędne fragmenty podczas kadrowania.

Przed przystąpieniem do kadrowania należy sprawdzić rozdzielczość zeskanowanego obrazu, rozmiar oraz głębię bitową. W JBC dla obrazów cyfrowych zostały przyjęte następujące parametry: rozdzielczość optyczna 300 dpi i 24-bitowa głębia koloru. Rozmiar dokumentu w wersji cyfrowej powinien być zgodny z oryginałem. W JBC przyjęto zasadę, że pliki do archiwizacji będą skanowane w formacie bezstratnym TIFF (model dokumentacyjny). Natomiast do prezentacji w Internecie zeskanowane obrazy trzeba poddać obróbce graficznej, ponieważ zazwyczaj wymagają między innymi: wyrównania, kadrowania, zmiany rozmiaru i konwersji do formatów skompresowanych, zajmujących mniejszy obszar pamięci. Zaletą kadrowania i prostowania jest lepsza prezentacja i jakość obrazów w bibliotece cyfrowej oraz lepsze rozpoznanie tekstu przez oprogramowanie OCR. Po rozpoznaniu warstwy tekstowej w digitalizowanym dokumencie, tekst jest możliwy do przeszukiwania przez wyszukiwarki oraz można go kopiować do edytorów, np. Microsoft Word, Open Office Writer. Po wyprostowaniu tekstu program OCR ma mniejszy margines błędów i jest w stanie rozpoznać go nawet ze 100% skutecznością. Dlatego bardzo ważne jest przygotowanie plików źródłowych starannie, z uwzględnieniem

wszystkich szczegółów, mających wpływ na jakość wynikowej publikacji cyfrowej [Kalota i Raczyński i Rękar 2011, s. 55-62]. Wadą kadrowania i prostowania jest wydłużenie procesu udostępniania wynikowych plików w bibliotece cyfrowej, ponieważ są to czynności czasochłonne.

W projekcie JBC przyjęto, że dodatkowym zabezpieczeniem zbiorów będzie ich zmikrofilmowanie. Pliki przygotowywane do przetworzenia na mikrofilm są wcześniej kadrowane, dzięki czemu mają estetyczny wygląd oraz zmniejsza się zużycie błony mikrofilmowej. Zaletą mikrofilmu jest jego trwałość. Zwój mikrofilmu o podłożu srebrnym powinien zachować jakość przez około 500 lat. Według danych podanych przez firmę Kodak trwałość mikrofilmu dochodzi nawet do 1600 lat, podczas gdy nośniki cyfrowe mają obecnie trwałość rzędu 10-30 lat [Felińska 2010].

Przystępując do kadrowania należy wybrać edytory obrazów. Są to programy, które zawierają narzędzia pozwalające na przycięcie krawędzi lub części obrazu, tak aby został tylko ten pożądan fragment [Busch 2005, s. 205-206].

W JBC przy wyborze programów do tego typu procesów kierowano się rodzajem zbiorów, które zostaną poddane edycji. Do kadrowania grafik i map używanych jest kilka programów: Gimp, Corel PHOTO-PAINT, XnView. Do edycji pojedynczych plików lub łączenia podzielo-



Abstrakt: Autor opisał metody i techniki kadrowania obrazu cyfrowego przy pomocy wybranych programów do obróbki grafiki. Dobrą praktyką jest stosowanie kilku programów graficznych jednocześnie ze względu na różnorodność opcji w nich dostępnych. Na rynku można wybierać wśród wielu aplikacji płatnych i bezpłatnych dlatego autor artykułu omówił krótko niektóre z nich oraz przedstawił ich zastosowanie w JBC. Opisane zostały również techniki kadrowania i prostowania skanów ze względu na rodzaj edytowanego dokumentu. Autor przedstawił także swoje spostrzeżenia i dobre praktyki w tego typu pracy.

Abstract: The author describes the methods and techniques of digital image processing using the selected software, suggesting parallel use of more than one program on account of their different sets of available editing options. One can choose between a number of applications which are available on the market, both as freeware and as commercial programs. The article presents the actual programs used in the Jagiellonian Digital Library for regular processing of digital files (e. g. cropping, straightening, etc.) as well as for optimisation of hard copy document scans. The author's remarks and hints concerning work streamlining are also included.

nych obrazów, np. mapy składającej się z dwóch części, przydatne są programy Corel PHOTO-PAINT i Gimp.

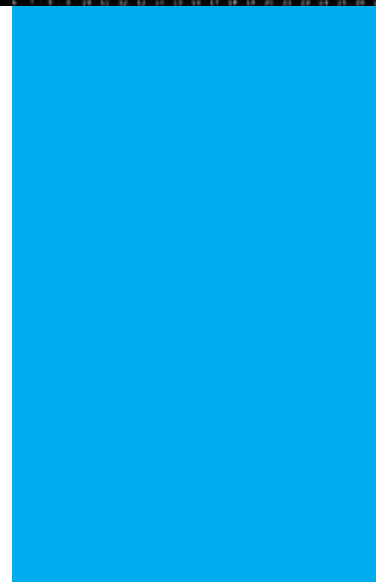
Corel PHOTO-PAINT jest programem przeznaczonym do edycji i tworzenia obrazów rastrowych. Można wykorzystać go między innymi do przygotowania grafiki na strony internetowe, do publikacji elektronicznych, prezentacji multimedialnych, jak również do podstawowej edycji zdjęć i skanów. Jest to program komercyjny [Corel 2012].

Gimp rozpowszechniany jest na licencji GNU (ang. General Public License), czyli jest w pełni bezpłatny. Jest odpowiednikiem programu komercyjnego Adobe Photoshop. Program umożliwia między innymi: wycinanie, łączenie, malowanie obrazów, zmianę tła, operacje na kolorach, usuwanie obiektów ze zdjęć, obracanie, zmianę rozmiaru i wiele innych [Gimp 2012].

Do edycji większej ilości plików jednocześnie bardzo dobrym rozwiązaniem jest skorzystanie z programu XnView. Jest to program służący jako przeglądarka i podstawowy edytor plików w wielu formatach. W XnView można w prosty sposób edytować zdjęcia czy skany, m. in. poprawiać jasność i nasycenie kolorów, przycinać i obracać, wycinać fragmenty, zmieniać wielkość, kolory, poprawiać ostrość, konwertować na inne formaty i wiele innych. Ma on bardzo dużo funkcji do wsadowego przetwarzania plików. Obsługuje interfejs

TWAIN, pozwalający wczytywać obrazy ze skanerów, aparatów fotograficznych i kamer. Program XnView jest darmowy dla zastosowań edukacyjnych i niekomercyjnych [XnView 2012]. W JBC stosowany jest do kadrowania pojedynczych skanów oraz do wsadowej obróbki graficznej większej ilości plików. Przykładowy proces kadrowania przy pomocy programu został opisany w dalszej części artykułu. W JBC do masowego kadrowania skanów dokumentów XIX-XX wiecznych został wybrany program Scan Tailor. Jest to darmowy program dla systemów Unix i Windows, którego kod źródłowy jest dostępny na licencji GNU. W programie można wykonywać operacje podziału strony, obracania, prostowania i kadrowania. Wszystkie operacje wykonywane są automatycznie, ale jest możliwość ręcznego korygowania wyników [Scan Tailor 2010]. Szczegółowo proces kadrowania i prostowania przy pomocy programu został podany w dalszej części artykułu.

W JBC dokumenty rękopiśmienne i starodruki nie są kadrowane po skanowaniu ze względu na ich zabytkowy charakter. Przy prezentacji tego typu zasobów ważne jest wierne odwzorowanie oryginału ze wszystkimi szczegółami.



II.1.
II.2.

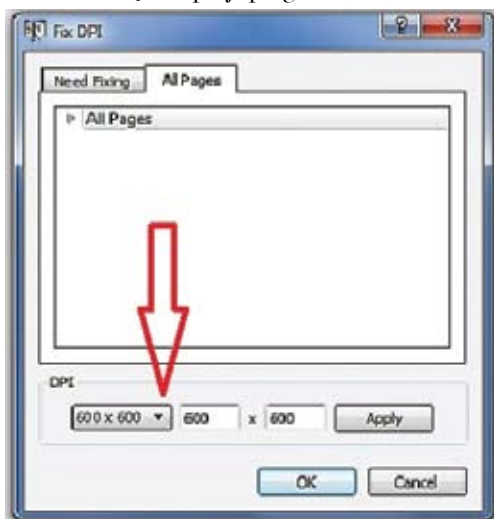
II. 1. Obraz cyfrowy przed obróbką w programie Scan Tailor.

II. 2. Obraz cyfrowy po obróbce w programie Scan Tailor.

Technologia wykonania kadrowania jednego woluminu czasopisma w programie Scan Tailor, wersja 0.9.11.1

Opisaną poniżej metodę można zastosować do masowego kadrowania każdego rodzaju dokumentów. Celem obróbki jest uzyskanie wykadrowanych oraz wyprostowanych pojedynczych stron czasopisma (il. 1 - 2).

Na początku po wybraniu folderu z obrazami do kadrowania program automatycznie otwiera je i tworzy miniatury. Gdy program nie może otworzyć niektórych plików, bo np. mają inne rozmiary od pozostałych lub różnią się rozdzielczością, wyświetla okno z dodatkowymi opcjami o stronach, których nie akceptuje. Należy wtedy za pomocą dostępnych opcji ustawić tymczasowo inną rozdzielczość wszystkich plików w folderze na taką, którą akceptuje program (il. 3).



Il. 3.

Il. 3. Zmiana rozdzielczości dla wszystkich stron dokumentu w programie Scan Tailor.

W pierwszym module programu jest możliwość obracania obrazów (il. 4). Służą do tego dwa przyciski umieszczone w interfejsie programu po lewej stronie, na il. 4 zaznaczone czerwoną strzałką. Użycie ich jest bardzo proste. Po naciśnięciu przycisku obraz obraca się o 90 stopni, w zależności od wybranego przycisku w prawo lub w lewo.

Kolejny moduł programu umożliwia podzielenie rozkładówki na pojedyncze strony (il. 5). Do wyboru są trzy opcje: dzielenie rozkładówki na pojedyncze strony, zaznaczenie całego obszaru skanowania i rozpoznanie pojedynczej strony dokumentu, gdy np. kadrowane są obrazy, które były skanowane jako pojedyncze strony.

W trzecim module program automatycznie prostuje strony (il. 6). Punktem odniesienia przy prostowaniu jest wydrukowany tekst. Gdy strona jest niezadrukowana, nie zawsze jest rozpoznawana przez program i trzeba ją ręcznie poprawić przy pomocy dostępnych opcji (il. 7).

W czwartym module program zaznacza zadrukowane części stron (il. 9). Bardzo ważne jest w tym kroku poprawienie stron, które zostały źle rozpoznane przez program. Przy kadrowaniu starych gazet i czasopism z 24 bitową głębią koloru jest to najbardziej czasochłonny proces, ponieważ trzeba poprawić ręcznie wszystkie strony źle zaznaczone, a ze względu na dużą ilość zabrudzeń papieru program



Il. 4.



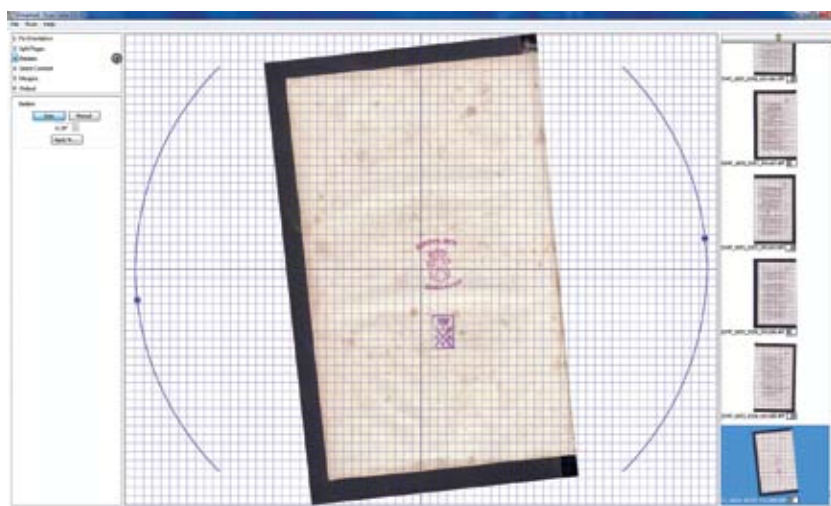
Il. 5.

Il. 4. Ustawienie orientacji skanów w programie Scan Tailor.

Il. 5. Dzielenie rozkładówki na strony w programie Scan Tailor.



Il. 6.



Il. 7.

Il. 6. Moduł do prostowania stron w programie Scan Tailor.

Il. 7. Źle rozpoznana strona przez program Scan Tailor.

ma problemy z właściwym rozpoznaniem składki drukarskiej. Przydatne w tym kroku są dwie opcje: order by increasing height i order by increasing width (il. 8). Po ich zaznaczeniu program ustawia wszystkie strony według wybranej opcji, czyli przy wybraniu „order by increasing height” obrazy sortowane są od najniższego do najwyższego zaznaczenia, a przy „order by increasing width” obrazy sortowane są od najwęższego do najszerszego. Po sortowaniu wystarczy poprawić zaznaczenie. Strony źle zaznaczone, w większości przypadków, będą na górze i na dole miniaturki wyświetlanych po prawej stronie w interfejsie programu.

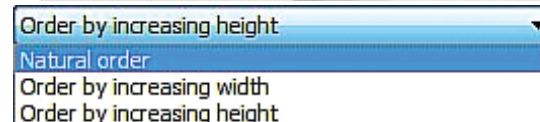
W kolejnym module ustala się marginesy dla wszystkich kadrowanych stron w dokumencie (il. 10). Służą do tego ustawienia dostępne po lewej stronie interfejsu programu. Jest tutaj możliwość

wyłączenia obrazów, które są za duże lub za małe w stosunku do reszty kadrowanych stron. Ułatwia to ustawienie marginesu dla wszystkich kadrowanych obrazów. Należy pamiętać, że strony wyłączone w tym kroku będą mieć inny rozmiar od pozostałych i mogą być też nieodpowiednio skadrowane. Dlatego dobrym rozwiązaniem jest wykadrowanie ich osobno, a później dołączenie do dokumentu.

W ostatnim kroku ustala się rozdzielczość (il. 11) i głęboką bitową (il. 12). Należy pamiętać, że jeżeli rozdzielczość plików była zmieniana przy wczytywaniu ich do programu, należy w tym kroku ustawić tę, która była ustalona w programie. Później będzie możliwość ustawienia rozdzielczości wynikowej w innym edytorze (np. XnView).



Il. 9.



Il. 8.

Il. 8. Opcje sortowania obrazów w programie Scan Tailor.

Il. 9. Zaznaczenie przez program Scan Tailor zadrukowanych części stron.

Na koniec należy wyeksportować pliki i zapisać na dysku komputera. Zaletą programu jest to, że po zapisaniu ustawień można ponownie edytować przetworzony dokument. Program jest łatwy w obsłudze, pozwala na masowe kadrowanie i prostowanie dużej liczby skanów. Wadą Scan Tailor jest to, że edycja skanów bez równej składki drukarskiej i z różną wielkością stron jest bardzo czasochłonna. Programu można użyć również do kadrowania pojedynczych plików, ale szybciej i łatwiej można to zrobić w innych programach do edycji zdjęć.

Technologia wykonania masowego kadrowania przy pomocy programu XnView, wersja 1.97.8

Tę metodę można zastosować do druków nowych lub zeskanowanych po-

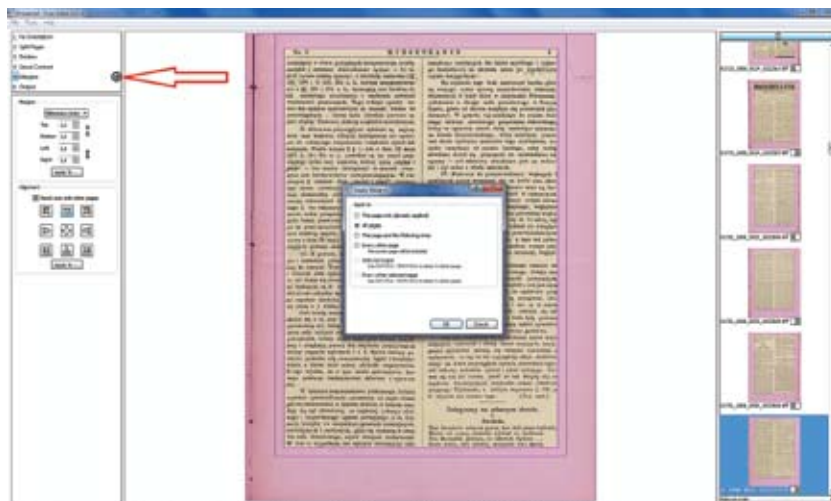
jedynczych stron dokumentu (il. 13). W trakcie skanowania należy każdorazowo przykładać róg strony do krawędzi obszaru roboczego skanera. Obszar skanowania należy ustawić przy pierwszej kartce i potem już go nie zmieniać:

1. przed kadrowaniem należy sprawdzić rozmiar obrazu w pikselach,

2. następnie w programie XnView należy zaznaczyć wszystkie pliki, które mają być wykadrowane i wybrać opcję - konwertuj wiele plików, ustawienia: format: TIFF. Przycinanie dla przykładowego rozmiaru obrazu 1666x2560 przedstawione jest poniżej [Kolasa 2007]:

X	Y	szerokość	wysokość
100	100	1500	2400

II. 10.



II. 10. Ustawianie marginesów w programie Scan Tailor.



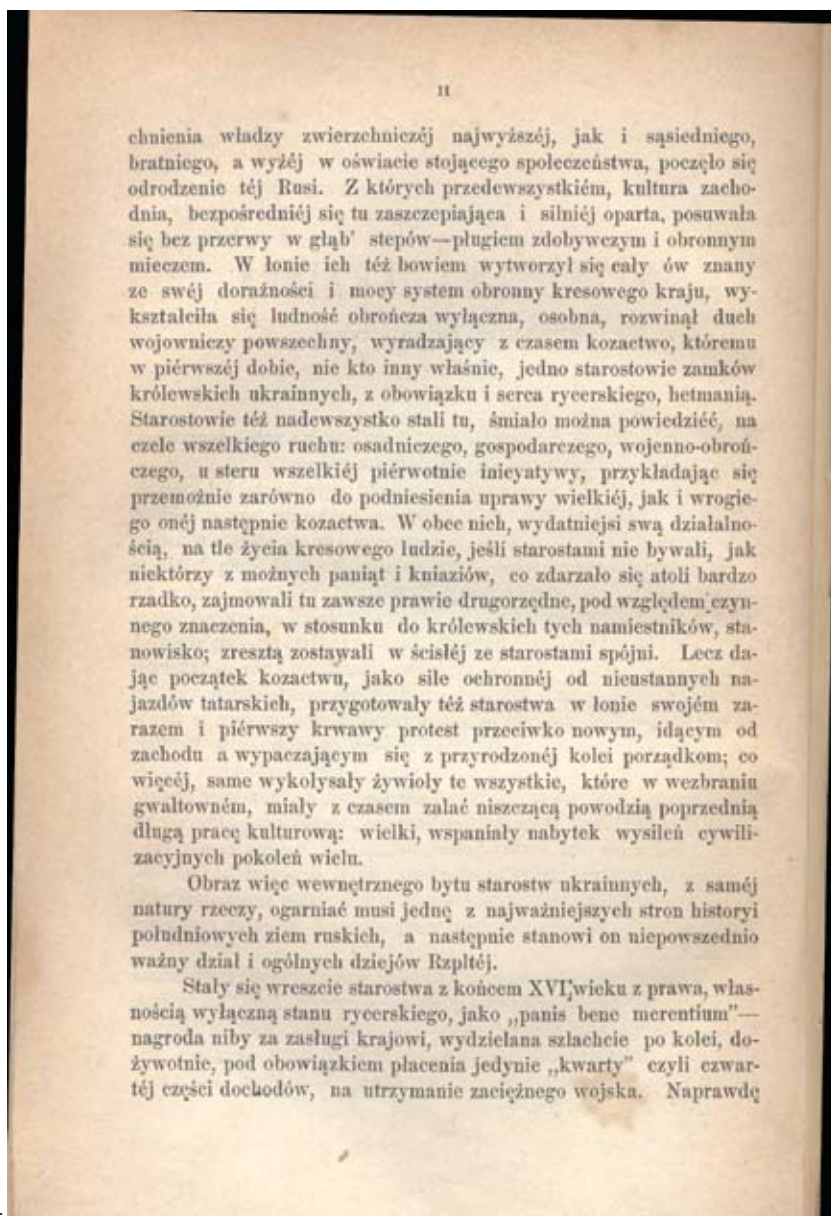
II. 11.



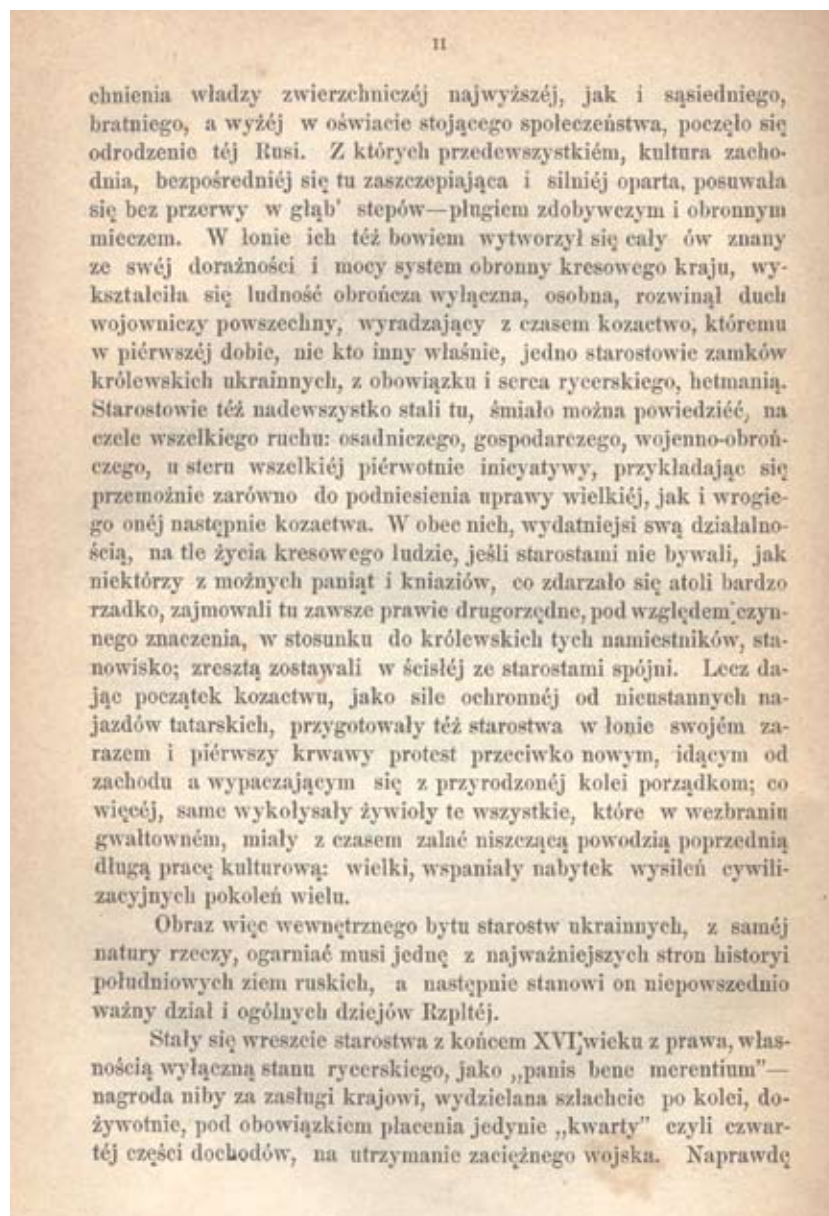
II. 12.

II. 11. Ustawienie rozdzielczości w programie Scan Tailor.

II. 12. Ustawienie głębi bitowej w programie Scan Tailor.



© II. 13a.



© II. 13b.

© II. 13a-13b. Obrazy przed i po kadrowaniu w programie XnView.

Na koniec po kadrowaniu i prostowaniu, niezależnie w jakim programie została wykonana edycja, należy sprawdzić parametry wynikowych plików. Powinny one mieć takie same rozmiary. Ich rozdzielczość i głębia bitowa powinny zgadzać się ze wstępnym założeniem. Szczególnie ważne jest to w przypadku rozdzielczości ponieważ przy ustawieniu złej rozdzielczości wynikowy skan będzie różnił się wymiarem od oryginału.

Podsumowanie

Opisane metody przygotowania publikacji cyfrowych zostały zaprojektowane dla konkretnych ustawień plików. Mogą być wykorzystywane w digitalizacji zbiorów archiwalnych i bibliotecznych. Omówione zostały wybrane programy przydatne w kadrowaniu i prostowaniu obrazów cyfrowych. Jest szereg innych programów, które można wykorzystać przy tego typu procesach.

Podsumowując należy dodać, że kadrowanie i prostowanie obrazów, poza walorami estetycznymi, pozwala w dalszym etapie przygotowania publikacji na uzyskanie lepszego rozpoznania tekstu przy korzystaniu z oprogramowania OCR, w wyniku czego użytkownik biblioteki cyfrowej ma większe możliwości wyszukiwania całych zdań i pojedynczych słów w tekście. Ułatwia to w większości przypadków pracę nad tekstem, ponieważ korzystając z wyszukiwarek szybciej można znaleźć poszukiwane informacje.

Bibliografia

Busch, David D. (2005). Skanowanie cyfrowe dla profesjonalistów. Warszawa: Wydawnictwo RM.

Corel(2012). Corel[online];[dostęp 19.05.2012]. Dostępny w World Wide Web: <http://www.corel.com>.

Felińska, Renata (2010). Od tradycyjnej formy drukowanej do biblioteki cyfrowej na przykładzie „Tygodnika Ilustrowanego” w zbiorach Biblioteki Uniwersytetu Łódzkiego [online]; [dostęp 05.05.2012]. Dostępny w World Wide Web: <http://bcul.lib.uni.lodz.pl/Content/1893/289-300.pdf>.

Fotografia (2010). Edycja kadru [online]; [dostęp 19.05.2012]. Dostępny w World Wide Web: http://pl.wikibooks.org/wiki/Fotografia/Edycja_kadru.

Gimp (2012). Gimp documentation [online]; [dostęp 19.05.2012]. Dostępny w World Wide Web: <http://www.gimp.org>.

Kalota Tomasz (2007). Jak przygotować obiekty do włączenia ich do biblioteki cyfrowej – doświadczenia Biblioteki Uniwersyteckiej we Wrocławiu [online]; [dostęp 05.05.2012]. Dostępny w World Wide Web: http://digital.fides.org.pl/Content/716/Jak_przygotowac_obiekty_do_BC.pdf.

Kalota, Tomasz; Raczyński, Rafał; Rękar, Paweł (2011). Przetwarzanie i OCR czasopism drukowanych gotykiem krok po kroku. [W:] Polskie biblioteki cyfrowe 2010. Materiały z konferencji, red. C. Mazurek, M. Stroiński, J. Węglarz. Poznań:

Ośrodek Wydawnictw Naukowych.

Kolasa Wiesław Marek (2007). DjVu bitonal – skanowanie i kadrowanie [online]; [dostęp 16.04.2012]. Dostępny w World Wide Web: http://fidkar.wbp.krakow.pl/fidkar/mbc/instrukcje_materiały_szkoleniowe/djvu_technologia_bitonal_kadrowanie_mbc.pdf.

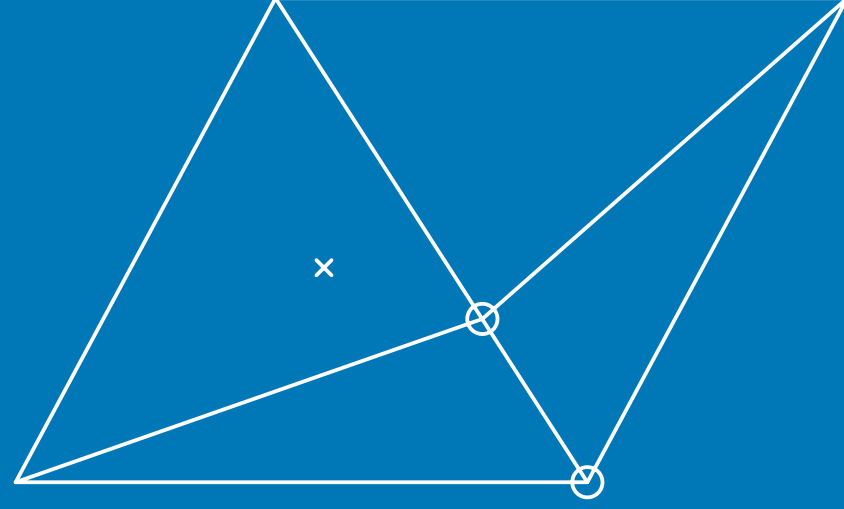
PWN Encyklopedia (2010). Kadrowanie [online]; [dostęp 19.05.2012]. Dostępny w World Wide Web: <http://encyklopedia.pwn.pl/szukaj.htmlmodule=lista&co=&search=kadrowanie&x=0&y=0>.

Scan Tailor (2010). Scan Tailor about [online]; [dostęp 16.04.2012]. Dostępny w World Wide Web: <http://scantailor.sourceforge.net>.

Wikipedia (2011). Kadrowanie [online]; [dostęp 19.05.2012]. Dostępny w World Wide Web: <http://pl.wikipedia.org/wiki/Kadrowanie>

XnView (2012). Info XnView [online]; [dostęp 16.04.2012]. Dostępny w World Wide Web: <http://xnview.com.pl>.

Udostępnienie obiektów cyfrowych



UO

Konwersja do DjVu



Słowa kluczowe: dokumenty zdigitalizowane; DjVu; przetwarzanie obrazów cyfrowych; kompresja stratna obrazu

Keywords: digitized documents; DjVu; digital image processing; lossy image compression

W celu optymalizacji dokumentów do udostępniania w sieci wykonywana jest konwersja do formatu hybrydowego DjVu, który oprócz danych graficznych może zawierać także wynik optycznego rozpoznania tekstu, odnośniki oraz adnotacje.

W Jagiellońskiej Bibliotece Cyfrowej do przeprowadzenia konwersji wykorzystywane jest oprogramowanie Document Express Enterprise oraz pakiet DjVu-

Libre. Dla przygotowanych uprzednio zestawów bitmap sporządzane są profile konwersji, dostosowywane indywidualnie do materiału w celu optymalizacji jakości i wynikowego rozmiaru pliku.

Teoretycznie najlepiej byłoby dobierać ustawienia indywidualnie dla każdego skanu, jednakże ze względu na znaczną ilość przetwarzanego materiału konieczne było sporządzenie kilku bazowych zestawów ustawień w celu przyspieszenia przetwarzania. Wytypowano kilka rodzajów konwertowanego materiału i dobrano odpowiednie metody segmentacji i kompresji. W przypadku dokumentów zawierających w większości tekst oraz nieskomplikowane ilustracje (czyli głównie czasopism i książek) stosowana jest segmentacja i znaczna stratna kompresja w celu uzyskania dobrej czytelności przy niewielkim rozmiarze pliku. W przypadku obrazów ciągłotonalnych, grafik i przypadków, w których segmentacja jest trudno wykonalna, nie stosuje się warstwowania.

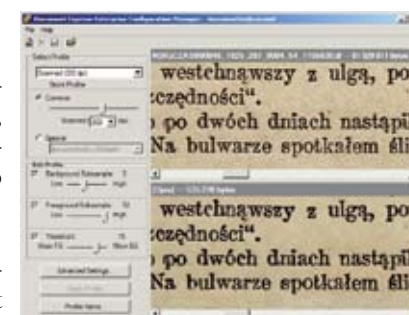
Konfiguracja Document Express

Profile konwersji można tworzyć zarówno w programie Configuration Manager, będącym częścią pakietu, jak i bezpośrednio poprzez wpisywanie wartości do pliku konfiguracyjnego.

Configuration Manager posiada graficzny interfejs, dzięki któremu możliwe jest regulowanie ustawień przez manipulację suwakami oraz zaznaczanie pól wyboru. Niestety program ten nie oferuje dostępu do wszystkich opcji.

Plik konfiguracyjny Document Express zawiera hierarchiczną strukturę, dzięki której w łatwy sposób można tworzyć nowe profile, które mogą dziedziczyć wartości z profili nagłówkowych (zwykle oznaczanych prefiksem H-), a także z innych profili wykonawczych. Producent dostarcza wstępnie przygotowane zestawy profili ogólnego oraz specjalnego przeznaczenia, na podstawie których można utworzyć indywidualne ustawienia dla każdego typu opracowanego materiału. Dzięki dziedziczeniu wartości nowo utworzone profile mogą mieć zaledwie długość kilku linii tekstu. Składnia pliku konfiguracyjnego pozwala na umieszczanie komentarzy, a także unikatowych etykiet profili.

Najważniejszym (oraz jedynym obowiązkowym) elementem w przypadku podanego powyżej profilu jest zawartość trzeciej linii, która składa się z identyfikatora profilu oraz listy profili, z których



Il. 1.

Abstrakt: Prezentacja sposobów konfiguracji oprogramowania Document Express, używanego do konwersji obrazów do formatu DjVu. Opisanie struktury profili oraz pliku konfiguracyjnego documenttodjvu.conf. Zestawienie rodzajów materiału wejściowego oraz innych czynników mających wpływ na konwersję oraz opisanie rodzajów konwersji oraz metod optymalizacji.

Abstract: Ways to configure the Document Express DjVu encoding software. Description of hierarchical structure of profiles and documenttodjvu.conf configuration file. Source item characteristics and other factors influencing conversion process. Types of conversion and methods of optimization.



```
#@displayName:Scanned (300 dpi)
#@readOnly:true
scan300: H-300dpi H-normal-quality H-normal-text
description="Scanned 300 DPI Documents"
```

II. 2.

biejący profil dziedziczy wartości. Linia pierwsza zawiera nazwę wyświetlaną profilu. Jeżeli linia ta występuje, możliwe jest wybranie profilu z listy w programie Workflow Manager, który jest częścią pakietu; brak tych wartości ustawia profil, jako niewidoczny dla Workflow Managera, natomiast nie blokuje możliwości wywołania profilu z linii poleceń. Wszystkie profile nagłówkowe nie są widoczne dla Workflow Managera (zwykle dlatego, iż nie są kompletne i mogą nie dawać oczekiwanych rezultatów). Linia druga ustawia dany profil jako dostępny tylko do

odczytu, co blokuje programowi Configuration Manager dokonywanie jakichkolwiek zmian ustawień w obrębie tak oznaczonego profilu. Nie zabezpiecza to przed wprowadzaniem zmian w wypadku bezpośredniej edycji pliku konfiguracyjnego, w związku z czym podczas dokonywania konfiguracji z użyciem tej metody należy zachować szczególną ostrożność i wykonywać regularne kopie bezpieczeństwa. Configuration Manager jest niezwykle wrażliwy na błędy w składni profilów i w przypadku wystąpienia takich przywraca plik konfiguracyjny

II. 3.

```
#@displayName:Clean (200 dpi)
#@readOnly:true
clean200: H-200dpi H-clean-quality H-normal-text
description="Clean (not scanned) 200 DPI Documents"
pix-filter-level=20
threshold-level=50
inhibit-foreback-level=0
lossless=true
```

II. 2. Przykład profilu bazującego jedynie na profilach nagłówkowych.

II. 3. Przykład profilu ze zmianami w stosunku do dziedziczonych wartości.

do pierwotnej postaci, co może skutkować utratą wcześniej sporządzonych ustawień. Linia czwarta zawiera długi opis profilu, który jest widoczny właściwie jedynie po otwarciu profilu do edycji w programie Configuration Manager.

Powyższy przykład ukazuje sposób dokonywania zmian w wartościach poszczególnych kluczy. Jak widać na przykładzie składnia opiera się na formacie klucz=wartość. Wartości mogą być numeryczne (zwykle są liczbami całkowitymi) oraz binarne (w postaci true lub false). Ukazana powyżej konstrukcja profilu wskazuje na dziedziczenie wartości z profili nagłówkowych, ze zmianami wyszczególnionymi poniżej. Wpisane wartości kluczy zastępują wartości dziedziczone z profili bazowych.

Struktura profili jest hierarchiczna, wszystkie profile dziedziczą z jednego profilu bazowego, który oznaczony jest jako H-base. Domyślnie profil ten nie zawiera żadnych wartości, natomiast jego istnienie pozwala na łatwe ustawienie wartości jakiegoś parametru dla wszystkich następujących profili (o ile wartość ta nie jest wymuszona w profilu docelowym). Domyślne profile nagłówkowe dzielą się na kilka kategorii:

- profile dotyczące optymalizacji dla dokumentów o danej rozdzielczości (np. H-300 dpi, H-150 dpi);
- profile dotyczące jakości kompresji (np. H-normal-quality, H-aggressive-

quality);

- profile dotyczące typu dokumentu wejściowego (np. H-normal-text, H-drawn, H-map).

Na ich podstawie tworzone są kompletne profile użytkowe, które już nadają się do wykorzystania w procesie konwersji. Oprócz profili dla segmentera plik konfiguracyjny zawiera także profile dla konwersji obrazów jednowarstwowych, bitonalnych oraz kolorowych, jednakże z powodu użycia do tych typów kompresji osobnych programów z odrębnymi plikami konfiguracyjnymi, ich wykorzystanie zalecane jest jedynie w przypadku konwersji wymagającej zmiany trybu obrazu graficznego, co wykonywane jest raczej rzadko.

Oprócz wykorzystania przygotowanych przez producenta ustawień, możliwe jest tworzenie profili użytkownika, które zwykle dopisywane są na końcu pliku konfiguracyjnego. Zwykle tworzenie własnego profilu polega jedynie na zmianie jednej lub kilku wartości, rzadko występuje potrzeba stworzenia całego profilu od podstaw, niemniej jest to możliwe i czasem jest jedynym sposobem na osiągnięcie oczekiwanych rezultatów.

Typy konwertowanych dokumentów

Ze względu na specyfikę przetwarzanych w JBC materiałów konieczne było rozszerzenie struktury podziału ustawień bazowych konwersji, gdyż standardowy

schemat (*scanned, drawn, manuscript, map, clean*) nie zapewniał dostatecznej precyzji i konfigurowalności. Większość digitalizowanego materiału to czasopisma i książki XIX- i XX-wieczne, w związku z czym bardzo niewiele dokumentów podpada pod kategorie *map* i *manuscript*, praktycznie nie występują dokumenty z kategorii *clean*. Istniejące kategorie *scanned* i *drawn* stały się podstawą do opracowania bardziej szczegółowego podziału.

Z punktu widzenia digitalizacji nie ma istotnej różnicy pomiędzy czasopismem a książką. W obu przypadkach podstawową jednostką wejściową w procesie przetwarzania jest bitmapa zawierająca obraz strony dokumentu. Zawartość stron czasopism i książek jest jednak zróżnicowana. Można tu wydzielić następujące kategorie według zawartości strony:

- tekst jednobarwny;
- tekst wielobarwny;
- tekst i grafika liniowa jednobarwna;
- tekst i grafika liniowa wielobarwna;
- tekst i ilustracje rastrowane;
- tekst i ilustracje ciągłotonalne;
- ilustracje ciągłotonalne;
- strony niezadrukowane.

Dla każdej z tych kategorii konieczne jest przygotowanie odrębnej rodziny profilów, zoptymalizowanych w celu przeniesienia odpowiednich cech obrazu w wystarczającej jakości, przy zachowaniu niewielkiego rozmiaru pliku wynikowego.

Na jakość konwersji ma wpływ także jakość dokumentu wejściowego. Dla skanów pochodzących z digitalizacji bardzo zniszczonych oryginałów lub wykonanych urządzeniami o niedostatecznych parametrach, konieczne jest dokonanie optymalizacji zwiększających skuteczność segmentacji i czytelność tekstu. Główne czynniki mające wpływ na skuteczność segmentacji to:

- zabrudzenia kart dokumentu;
- przebicie drugiej strony druku;
- niedokładnie odbity druk;
- brakujące fragmenty znaków;
- problemy z ostrością obrazu;
- aberracje chromatyczne i inne zniekształcenia barw;
- szum;
- niedostateczna rozdzielczość obrazu;
- zniekształcenia geometryczne;
- niska dynamika obrazu.

Część z tych problemów może zostać wyeliminowana przez odpowiednie dobranie ustawień profilu konwersji, w większości wypadków wymagane jest także wykonanie dodatkowych operacji na obrazie przed dokonaniem konwersji w celu ułatwienia segmentacji. Najlepsze rezultaty daje połączenie obu metod, czyli wykonanie obróbki wstępnej oraz dopracowanie profilu.

Optymalizacja ze względu na rozmiar oraz jakość

Najistotniejszą cechą formatu DjVu jest układ warstwowy, który umożliwia zastosowanie zoptymalizowanej kompresji dla każdej z warstw dokumentu, co umożliwia uzyskanie niewielkich plików wynikowych przy jednoczesnym zachowaniu znacznej rozdzielczości istotnych elementów obrazu. Mniej istotne fragmenty obrazu zwykle zostają zredukowane, co nie wpływa znacznie na całościowy odbiór wizualny dokumentu.

Typowy dokument DjVu składa się z trzech warstw: bitonalnej maski oraz obrazów pierwszego planu i tła, reprezentujących barwy druku i podkładu, które zwykle są bitmapami ciągłotonalnymi o zredukowanej rozdzielczości. Jeżeli obraz strony dokumentu nie zawiera w obszarze tła znaczących elementów, można znacznie zredukować rozdzielczość tej warstwy, bez wpływu na szczegółowość elementów tekstu, a więc bez większego wpływu na czytelność. W niektórych przypadkach możliwe jest nawet

całkowite wyeliminowanie nieistotnych cech obrazu źródłowego w celu zmniejszenia rozmiaru, co niekiedy poprawia całościowy odbiór dokumentu. Podobnie w przypadku warstwy pierwszego planu, jeżeli barwa druku jest w miarę jednolita na obszarze całej strony możliwa jest znaczna redukcja rozdzielczości warstwy barwnej lub nawet zastąpienie jej jednym kolorem.

Zastosowanie zbyt ostrej redukcji powoduje jednak wrażenie sztuczności obrazu, co jest zwykle niepożądane. Podczas doboru parametrów konwersji zalecane jest zachowanie odpowiedniego balansu pomiędzy rozmiarem i jakością kompresji poszczególnych warstw w celu zapewnienia wystarczającego komfortu odbioru materiału.

Typowym zakresem redukcji obrazów warstw jest podział obrazu pierwszego planu przez 12, a tła przez 3, co dla większości typowych dokumentów zapewnia wystarczające odwzorowanie oryginału. W przypadku czasopism, które zwykle mają większe rozmiary można stosować wyższe dzielniki, gdyż przy całościowym postrzeganiu strony nie ma to większego wpływu na jakość. Natomiast dla dokumentów o niewielkich rozmiarach zalecane jest zmniejszenie wartości współczynników podziału. Ma to związek z zachowaniem użytkownika podczas czytania z ekranu: zaobserwowana została tendencja do pomniejszania większych dokumentów oraz powiększania mniejszych, w związku z czym większość

przeglądarek domyślnie skaluje dokument do rozmiaru okna.

W przypadku dokumentów zawierających ilustracje ciągłotonalne konieczne jest zastosowanie mniejszych dzielników w celu zapewnienia należytej jakości ilustracji, których fragmenty (lub całość) znajdują się w warstwach drugorzędnych dokumentu DjVu.

Generalnie automatyczna segmentacja obrazu poprawia czytelność tekstu, zmniejszając zarazem dokładność odwzorowania ilustracji ciągłotonalnych. Wynika to z niedoskonałości procesu analizy obrazu. Nadal nie jest możliwe uzyskanie idealnej segmentacji obrazów zawierających tekst oraz ilustracje ciągłotonalne (szczególnie pochodzących z dokumentów nie najlepszej jakości). Jedynym gwarantem całkowitej poprawności segmentacji obrazu jest wykonanie procesu ręcznie. Przeprowadzenie ręcznej segmentacji polega na uprzednim przygotowaniu obrazów wszystkich warstw, skompresowaniu ich oddzielnie i połączeniu w całość. W przypadku znacznej ilości materiału jest to uciążliwe i przez to rzadko wykonywane podczas dokonywania masowej digitalizacji.

Optymalizacja pod kątem uzyskania wysokiej jakości w procesie automatycznej segmentacji opiera się generalnie na zmniejszeniu stopnia kompresji oraz redukcji rozdzielczości warstw. W niektórych jednak przypadkach zastosowanie zbyt małych dzielników ma negatywny

wpływ na jakość obrazu wynikowego. Redukcja rozdzielczości oraz kompresja stratna czasami powodują minimalizację lub nawet całkowite usunięcie wad obrazu wejściowego, co może być pożądaną, szczególnie w przypadku pracy z obrazami zawierającymi zniekształcenia barw. Szczególnie narażone na tego typu zniekształcenia są obrazy zawierające ilustracje rastrowane, dla których konieczne jest często wykonanie znacznych korekt w obrazie wejściowym przed dokonaniem kompresji.

Elementy obrazu zakwalifikowane jako elementy istotne (głównie tekst i grafiki liniowe/rastrowe) zapisywane są do bitonalnej maski, która może samodzielnie funkcjonować jako dokument czarno-biały. Jest to jedyna warstwa, na której nie jest dokonywana redukcja rozdzielczości. Bitonalna natura tego obrazu niesie jednak za sobą pewne ograniczenia: krawędzie kształtów nie są gładkie. Możliwa jest jednak kompensacja tego zjawiska po stronie przeglądarki. Do kompresji obrazów bitonalnych wykorzystywany jest w DjVu algorytm JB2, który umożliwia zmniejszenie rozmiaru pliku wynikowego przez zastosowanie redukcji podobnych kształtów do wspólnego prototypu. Zamiast kodowania każdego unikalnego kształtu zapisywany jest jedynie prototyp oraz różnice pomiędzy nim a kształtami pochodnymi. Wykorzystanie tej metody daje bardzo dobre rezultaty w przypadku konwersji dokumentów pochodzących ze składu cyfrowego, gdzie każde wystąpienie znaku pisarskiego jest identyczne.

W połączeniu z użyciem wspólnych dla wielu stron słowników kształtów pozwala to uzyskać bardzo efektywną kompresję.

Dokumenty niesegmentowane

Najprostszym do wykonania typem konwersji jest po prostu wykonanie kompresji obrazu bez zastosowania segmentacji. Dla obrazów ciągłotonalnych możliwe jest dokonanie całościowej redukcji rozdzielczości oraz dobór stopnia kompresji. Wykorzystanie tej metody powoduje jednak znaczne zmniejszenie ostrości tekstu. Obrazy wynikowe nie mogą być skompresowane zbyt mocno, gdyż artefakty kompresji są łatwo dostrzegalne, przekłada się to bezpośrednio na wielkość pliku wynikowego. Metoda ta faworyzuje ilustracje ciągłotonalne o niewielkiej szczegółowości i małej ostrości.

W przypadku obrazów bitonalnych dokonywana jest uprzednio omówiona kompresja z wykorzystaniem prototypów kształtów. Możliwe jest wykonanie operacji odrzucenia pewnych kształtów na etapie kompresji. Pomaga to w przypadku obrazów bitonalnych zawierających odzwierciedlenia zabrudzeń, aczkolwiek nieostrożne użycie może usunąć fragmenty treści dokumentu.

Do przeprowadzenia kompresji obrazów ciągłotonalnych oraz bitonalnych bez wykorzystania segmentacji nie jest wymagane użycie Document Express. Wszystkie konieczne operacje można

przeprowadzić przy pomocy bezpłatnych narzędzi z pakietu DjVuLibre, a w przypadku obrazów bitonalnych także specjalnego enkodera miniDjVu, zoptymalizowanego w celu uzyskania jak najmniejszych plików.

Istnieje jeszcze jeden rodzaj kompresji bez wykonania segmentacji. Obrazy o ograniczonej liczbie kolorów można skompresować także bezstratnie przy pomocy enkodera cpaldjvu z pakietu DjVuLibre. Cpaldjvu wskazuje jeden kolor (domyślnie taki, który dominuje na obrazie), jako kolor tła, na jego podstawie tworzona jest maska, natomiast wszystkie pozostałe barwy zostają przekierowane do pierwszego planu. Metoda ta nadaje się do kompresji, np. zrzutów ekranu.

OCR w Document Express

Przy pomocy Document Express możliwe jest wykonanie optycznego rozpoznania tekstu na obrazach DjVu. Niestety możliwości konfiguracyjne zastosowanego w DE silnika IrisOCR są niewielkie i ograniczają się jedynie do wyboru języka rozpoznania. Na jakość rozpoznania tekstu ma jednak wpływ poprawność segmentacji obrazu, gdyż jako obraz wejściowy do silnika OCR trafia bitonalna maska. Aby zwiększyć skuteczność rozpoznania należy sprawić, aby wszystkie znaki pisarskie zostały poprawnie przydzielone do warstwy maski. Zbyt mała ostrość i słaba dynamika obrazu zmniejszają skuteczność segmentacji,



- II.4. co w rezultacie powoduje, że silnik OCR nie otrzymuje poprawnych kształtów. Taki sposób działania oprogramowania Document Express skłania do dołożenia starań, by wykonać segmentację poprawnie.

Postaci dokumentów DjVu

DjVu umożliwia zapis dokumentów w postaci scalonej i rozdzielonej. Z uwzględnieniem możliwości dostępu do każdej ze stron niezależnie postać rozdzielona dokumentu DjVu ma znaczny wpływ na wygodę użytkownika w środowisku sieciowym o ograniczonej prędkości transmisji. Do przejrzania danej strony nie jest wymagane pobranie całości dokumentu, przeglądarka podczas wyświetlania danej strony automatycznie pobiera następną i poprzednią, co usprawnia płynność przeglądania. Najprostszą postacią rozdzieloną dokument DjVu zawiera obrazy poszczególnych stron oraz indeks łączący je w całość. Jest to zgodne ze strukturą dokumentów HTML, co umożli-

wia wykorzystanie serwerów WWW, jako platformy dystrybucji dokumentów DjVu w sposób bezpośredni.

Postać scalona DjVu zawiera obrazy stron oraz indeks zgromadzone w jednym pliku, co ułatwia lokalne zarządzanie dokumentami.

Zawsze możliwe jest dokonanie konwersji postaci scalonej na rozdzieloną i odwrotnie, nawet przy użyciu jedynie dostępnej bezpłatnie przeglądarki.

Dokumenty DjVu mogą być osadzone bezpośrednio w HTML, co umożliwia integrację z serwisami WWW. Wciąż wymagane jest jednak użycie wtyczek do przeglądarek internetowych, gdyż format nie jest jeszcze natywnie przez nie wspierany.

Istnieje także możliwość dodania znaku wodnego oraz zabezpieczenie dokumentu hasłem. Funkcje te nie są jednak zwykle wykorzystywane w przypadku digita-

- II.4. Fragment czasopisma skonwertowany z wykorzystaniem standardowych ustawień Document Express.

lizacji materiałów z domeny publicznej.

Konwersja obrazów starych czasopism

Dokumenty przetwarzane w ramach projektu Jagiellońska Biblioteka Cyfrowa to głównie wydawnictwa ciągłe z XIX i XX wieku. Znaczną część stanowiły nieilustrowane wydawnictwa wielkoformatowe, dla których należało zastosować zaawansowane metody obróbki i kompresji w celu zapewnienia względnie niewielkich rozmiarów plików wynikowych.

Konwersja materiału zawierającego jedynie tekst jest względnie prosta. Problemy mogą wystąpić jednak w przypadku tytułatury, nagłówków odbitych większą czcionką oraz elementów graficznych, a także fragmentów uszkodzonych lub wydrukowanych niedokładnie.

II.4. przedstawia wynik segmentacji obrazu fragmentu czasopisma z wykorzystaniem standardowych ustawień. Niektóre

znaki lub ich fragmenty zostały przydzielone do nieodpowiednich warstw, co zaburza integralność treści oraz pogarsza aspekt wizualny dokumentu.

Skuteczność segmentacji można zwiększyć dwustopniowo: poprzez dostosowanie profilu oraz optymalizację obrazu wejściowego. Ten sam obraz, skonwertowany przy użyciu specjalnie napisanego profilu prezentuje il. 5.

Taki sposób segmentacji zapewnia skuteczniejsze odwzorowanie treści dokumentu oraz umożliwia generowanie plików wynikowych o mniejszym rozmiarze. Warstwa tła nie zawiera istotnych elementów treści, gdyż wszystkie kształty zostają przydzielone do maski, w związku z czym, możliwe jest zwiększenie stopnia kompresji obrazów warstw bez znacznego pogorszenia czytelności. Niestety użycie tej metody wiąże się z wyłączeniem optymalizacji znaków w Document Express, co skutkuje zwięźleniem kształtów oraz rozjaśnieniem

- II.5.



- II.5. Fragment czasopisma skonwertowany z wykorzystaniem profilu zapewniającego większą kontrolę nad segmentacją.



II. 6. koloru wypełnienia drobnych znaków. Można to jednak skompensować przez dokonanie optymalizacji obrazu wejściowego. Zwiększenie grubości ciemnych obszarów obrazu kompensuje straty spowodowane rezygnacją z tej operacji w Document Express. Optymalizacja taka zwiększa czytelność tekstu oraz poprawia kolor wypełnienia kształtów. II. 6. prezentuje segmentację obrazu zoptymalizowanego przy użyciu profilu z poprzedniego przykładu.

Zwiększenie grubości znaków nie jest wymagane dla każdego dokumentu, stosowane jest jedynie w przypadku stron o słabej jakości druku lub wydrukowanych czcionkami o niewielkich grubościach elementów znaków pisarskich, które są mało widoczne przy skanowaniu z ograniczoną rozdzielczością. Tego rodzaju optymalizacje poprawiają także odzwierciedlenie ilustracji.

Przydział wszystkich elementów grafiki do jednej warstwy zapewnia poprawienie

II. 6. Fragment czasopisma skonwertowany z wykorzystaniem profilu zapewniającego większą kontrolę nad segmentacją wraz z optymalizacją obrazu.

czytelności obrazu przy jednoczesnym zmniejszeniu jego rozmiaru, a pogrubienie kształtów zwiększa ogólny kontrast.

Operacje takie jednak wymagają uzupełnienia procesu digitalizacyjnego o dodatkowe elementy, co wydłuża czas opracowywania materiału, jednakże przygotowane w ten sposób obrazy wejściowe są bardziej przyjazne dla segmentera, co skutkuje zmniejszeniem ilości błędów kompresji.



II. 7.

II. 7. Wpływ optymalizacji na segmentację ilustracji:

1. Ustawienia standardowe;
2. Profil specjalny;
3. Profil specjalny + zoptymalizowany obraz wejściowy.

Organizacja automatyzacji pracy



Słowa kluczowe: digitalizacja, projekt, biblioteka cyfrowa, oprogramowanie, digitalizacja, projekt, biblioteki cyfrowe, DjVu, grafika rastrowa, mieszana zawartość rastrowa.

Keywords: digitization, digital library, software, DjVu, raster graphic, mixed raster content

Wsparcie programowe przy automatyzacji pracy

Cały proces digitalizacji wymagał wielu programów, które używano na różnych etapach cyfryzacji zbiorów. Niemal każdą czynność wykonywano przez inne narzędzie, od zwykłego menedżera plików pozwalającego w szybszy sposób zarządzać materiałem, aż po platformę publikacji dLibra umożliwiającą dostęp do gotowych zbiorów poprzez Internet.

Większość używanych w Jagiellońskiej Bibliotece Cyfrowej programów była przystosowana do pracy masowej (co jest po części kryterium doboru używanych przy digitalizacji w JBC narzędzi). Każdy z nich był jednak pisany z myślą o jednym, konkretnym zadaniu, bez jakiegokolwiek zamiaru łączenia się z innymi programami niezbędnymi przy digitalizacji. Używany w JBC konwerter do formatu DjVu, Document Express Enterprise, nie ma jakiegokolwiek związku z oprogramowaniem dLibra pozwalającym na publikację gotowego materiału.

Z tego względu wiele czynności, będących specyfiką digitalizacji Jagiellońskiej Biblioteki Cyfrowej, wymagało wielokrotnego powtarzania różnych etapów z powodu braku oprogramowania automatyzującego je. Aby umożliwić tak zorganizowaną pracę na powtarzalnych elementach, należało tworzyć dodatkowe narzędzia wykonujące wymagane procesy. W tym celu w zespole realizującym obróbkę cyfrową materiału zapewniono

stałe wsparcie programistyczne. Pozwoliło to stworzyć oprogramowanie nazwane Segreguj_numery i modyfikować je w trakcie pracy, kiedy specjalistyczne potrzeby ulegały zmianom. Dzięki temu dostosowywanie narzędzi mogło odbywać się w sposób ciągły, wychodząc tym samym naprzeciw oczekiwaniom w zmieniających się warunkach, czy to związanych ze specyfikacją zasobu, czy z powstaniem nowych technologii, jak udostępnienie oprogramowania ABBYY Fine Reader w wersji 11 we wrześniu 2011 roku. Podobne praktyki wsparcia informatycznego dla zespołu digitalizującego są obecne w innych bibliotekach Unii Europejskiej [Anderson 2008; Turner i in. 2004, s. 87].

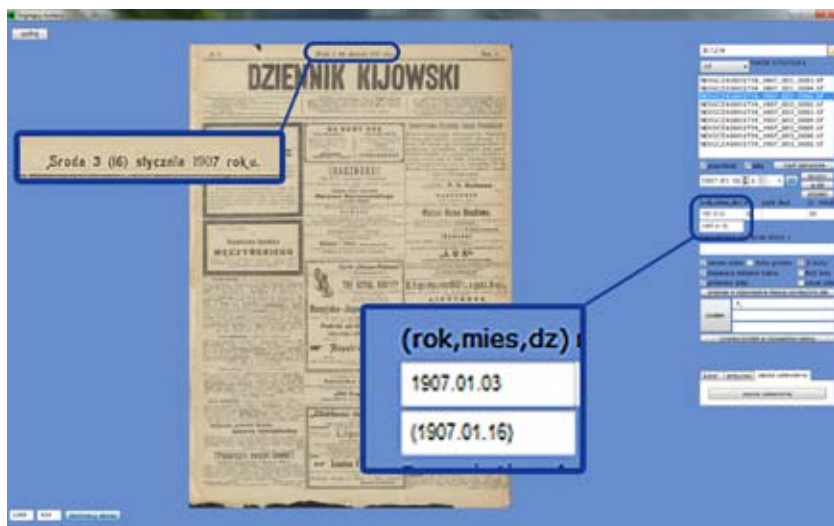
Pierwszym problemem utrudniającym szybkie zarządzanie materiałem było ułożenie kopii cyfrowych zgodnie z stanem faktycznym zbiorów. Każdy numer tytułu czasopisma zamieszczano na serwerze w osobnym pliku, który następnie powiązano z rocznikiem i sygnaturą. Wykonywanie tych czynności ręcznie wymagałoby dużego nakładu pracy. Nie uniknęłoby się przy tym przeglądania każdej strony w gazecie [Allen i in. 2010]. Można jednak zautomatyzować tworzenie folderów i nazewnictwo plików. Utrzymana w ten sposób struktura w wyraźny sposób ułatwiła taką organizację pracy w kolejnych etapach. W celu zautomatyzowanego ustalania kolejności stron w roczniku czasopisma stworzony został program umożliwiający wyświetlanie obrazów zeskanowanego dokumentu

i włączanie go do struktury (wspomniany wyżej Segreguj_numery). Konieczna też była możliwość zarządzania kolejnością stron, gdyż często strony w numerze gazety były poprzestawiane. Chcąc zamieścić prezentację treści, zdecydowano się ułożyć strony dokumentu we właściwej kolejności zamierzonej pierwotnie przez wydawcę czasopisma.

Abstrakt: Artykuł opisuje znaczenie wsparcia programistycznego w procesie digitalizacji. Poruszone są kwestie dotyczące konieczności automatyzacji powtarzalnych zadań i dostosowywania narzędzi do specyficznych potrzeb biblioteki cyfrowej. Opisano również zagadnienia związane z konwersją obrazów grafiki rastrowej do formatu DjVu. Wyjaśnione są podstawowe problemy związane z segmentacją zeskanowanych dokumentów, a także zaprezentowane są przykładowe rozwiązania stosowane przy digitalizacji w JBC.

Abstract: This article discusses the role of software in the digitization process, emphasizing necessity for automation of repeating tasks and importance of software customization to meet specific needs. Some aspects of raster graphic images conversion to DjVu file format as well as the main problems with mass segmentation of scanned documents are also described. It also provides examples of solutions used in the Jagiellonian Digital Library.





© Il. 1.

Ten sam program został wzbogacony o wiele innych funkcji. Jedną z najistotniejszych jest możliwość przygotowywania opisów bibliograficznych pojedynczego numeru czasopisma w sposób zautomatyzowany.

Oprogramowanie dLibra w niewielkim stopniu wspiera masowe ładowanie publikacji. Narzędzia dostępne, które można było wykorzystać, służą jedynie do odczytywania plików opisu bibliograficznego połączonego z kopiami cyfrowymi w formacie DjVu. W żaden sposób nie wytwarzają one gotowych plików zawierających opis publikacji. Dla jednego rocznika czasopisma z reguły wszystkie pola w formacie Dublin-Core są takie same, zmienia się jedynie numer w tytule. Nie ma więc potrzeby wykonywania za każdym razem czyn-

ności uzupełniania tych samych pól z identycznymi wpisami, jeśli można wykonać to automatycznie. Program nie tylko kopiuje wszelkie powtarzalne informacje do każdej następnej przygotowywanej publikacji, ale także odczytuje z nazwy folderu aktualny numer czasopisma i umiejscawia tę informację w odpowiednim polu. W tym momencie w całym procesie rozpoznanie przez bibliotekarza cyfrowego numeru wydawnego tytułu wykonywane jest tylko raz, później ta informacja odczytywana jest z nazwy pliku przy każdym następnym etapie, czy to automatycznie, czy przez człowieka.

Inną potrzebą jest tworzenie specyficznych metadanych przez bibliotekarza w sposób zautomatyzowany. Z reguły

w oprogramowaniach dla bibliotek cyfrowych (np. dLibra) każdą informację należy wpisać ręcznie albo importować z już istniejącej. Tymczasem oprogramowanie Segreguj_numery daje możliwość tworzenia wielu szczegółowych opisów w sposób automatyczny. Większa część odbywa się, jak już wspomniano, przez kopiowanie wartości wspólnych. Postanowiono również, by w zasobie cyfrowym umieszczać datę dzienną publikacji czasopisma zamieszczaną przy każdym numerze. Jeżeli w obrębie jednego tytułu wydano tysiące numerów jest to czynność trudna. Dodanie funkcji w oprogramowaniu Segreguj_numery odpowiedzialnej za tworzenie daty dziennej stanowiła zatem największy problem. Działanie programu opiera się na przewidywaniu kolejnych dni i wpisywaniu ich do siatki. Automatyzacja jest jednak tylko częściowa, gdyż nieregularność kalendarza, pojawianie się świąt, przerw wydawniczych lub braków w zasobie biblioteki sprawia, że każdą wartość należy sprawdzić ze stanem faktycznym. Wykonywane jest to jednak podczas dzielenia digitalizowanego woluminu na poszczególne numery. W tym czasie przegląda się stronę tytułową, odczytuje oznaczenie kolejnego numeru oraz datę. Ustalenie ostatniej wartości odbywa się więc podczas jednego procesu wraz z innymi czynnościami, co znacznie ułatwia i przyspiesza pracę.

Utworzenie przewidującego kalendarza domniemującego daty polegało na wpisaniu konkretnych odstępów dzien-

nych pomiędzy numerami (dziennik, tygodnik, miesięcznik, uwzględnienie wydawania czasopisma dwa razy w tygodniu itp.). W regularnych przypadkach program sugerował daty samodzielnie. Bibliotekarz segregujący materiał czasopisma jedynie zatwierdzał wpis lub go poprawiał. Wszelkie wariacje regularności wydawania czasopisma dostrzeżono w trakcie pracy przez pracowników wykorzystano do dodawania kolejnych opcji wzbogacających domniemywanie daty wydania. Właśnie w trakcie opracowywania zasobu okazało się, że dokument jest datowany w dwóch używanych na terenach Ukrainy kalendarzach. Potrzeba zamieszczenia obu informacji w metadanych zainicjowała dodanie kolejnej funkcji – wpisywanie obu wartości w polu data. Taki przypadek przedstawia il. 1.

Poza wprowadzeniem opisanych wyżej wartości opisu bibliograficznego istnieje możliwość zawarcia w odpowiednich polach dodatkowych informacji opisujących numer czasopisma (np. uwagi dotyczące braku stron). Tak, jak w przypadku ustalania daty dziennej wydania numeru wykonuje się to podczas jednej czynności, w której przegląda się całe czasopismo, kumulując tym samym zadania tak, by praca nad materiałem trwała jak najkrócej i by nie powtarzać tych samych czynności.

Przygotowane w ten sposób metadane program wpisuje w sposób zautomatyzowany do każdej publikacji i tworzy pliki niezbędne do masowego importu przez

© Il. 1. Program do segregacji czasopism. Na środku wyświetlanie aktywnego pliku, z prawej strony menu opcji zarządzania zasobem.

platformę dLibra. Proces publikowania na serwerze jest już w pełni automatyczny, istnieje jedynie konieczność sprawdzenia efektu końcowego gotowej kopii cyfrowej na stronie internetowej JBC, by ustrzec się ewentualnych błędów.

We wrześniu 2011 roku ukazała się nowa wersja oprogramowania ABBYY Fine Reader 11. Od wcześniejszych wersji tego programu różni ją możliwość zapisu wykonywanego projektu do formatu DjVu (wcześniej istniała jedynie ewentualność odczytu takich plików). Zakupione narzędzie nie było przystosowane do pracy masowej na wielu publikacjach. Zorganizowano jednak proces tak, by nie było konieczne dodanie funkcjonalności do programu Fine Reader. Zaadaptowano rozwiązanie z wytworzonego na potrzeby JBC narzędzia. W pierwszej kolejności gotowe do konwersji na format DjVu pliki segmentowano wraz z wykonaniem procesu OCR w programie Fine Reader. Następnie zapisane każde z osobna strony przeglądano pod kątem segregacji materiału, zupełnie tak, jak na plikach przetwarzanych w formacie TIFF. Podczas tego etapu uzupełniano metadane. Dopiero tak przygotowane pliki łączono w pojedyncze publikacje, również masowo, przy pomocy programu DjvuLibre. W ten sposób brak możliwości zautomatyzowania pracy przez oprogramowanie firmy ABBYY jest uzupełniany przez wsparcie programistyczne naszego środowiska. Z tego też względu zakupiono wersje jednostanowiskowe dla każdego z członków zespołu, gdyż wersja serwe-

rowa przy tak intensywnym wykorzystaniu stanowiłaby „wąskie gardło” całego procesu konwersji kopii cyfrowych.

Podczas segregacji materiału bibliotekarz cyfrowy sprawdza też kompletność kopii cyfrowej i jakość wytworzonych wcześniej plików. Podczas segmentacji mogą pojawić się różne błędy, których nie da się przewidzieć. Z tego względu niezbędne jest ponowne sprawdzenie każdej strony dokumentu pod kątem oceny poprawności wytworzonej maski. Wykonywanie tego w trakcie jednego przeglądania, podczas którego nakłada się siatkę wydawniczą, ustala się metadane i tworzy strukturę katalogów jest kolejnym ułatwieniem. Wymagało to jednak wzbogacenia oprogramowania o kolejne funkcje pozwalające na dostosowywanie przeglądania pliku pod kątem oceny poprawności maski, co w tym przypadku sprowadziło się do możliwości wyboru opcji widoku przez wtyczkę dekodującą obraz z formatu DjVu.

Sugerując się próbami nakładania struktury w sposób automatyczny wraz z wpisywaniem metadanych z wiązki OCR rozpoczęto wzbogacanie naszego oprogramowania o podobne funkcje. Jednakże mnogość błędów w OCR nie pozwoliła w tym przypadku osiągnąć większych rezultatów. Podobne próby podejmowano również w innych repozytoriach cyfrowych w USA i Wielkiej Brytanii [Allen i in. 2010, s. 98; Tanner i in. 2009].

Opisany wyżej program Segreguj_nu-

mery stanowił jedną z form wsparcia programistycznego. Innym aspektem ułatwiającym automatyzację pracy było przygotowywanie skryptów batch, współpracujących z linią komend wykorzystywanego oprogramowania (m.in. Document Express Enterprise i DjvuLibre). Większość zadań przy automatyzacji pracy możliwe jest głównie dzięki wspieraniu linii poleceń. Jednakże wpisywanie za każdym razem wszystkich komend nie tylko niewiele przyspiesza pracę, ale może być też powodem wielu błędów, których poprawa w efekcie wydłuży cały proces. Na potrzeby automatyzacji zadań było więc niezbędne napisanie skryptu wysyłającego polecenia do programu według raz ustalonego schematu.

Liczba zadań wymaganych przy takich operacjach jest spora. Co najważniejsze, często za pomocą takiego skryptu można naprawić popełnione wcześniej błędy. Z tego względu wytwarzanie takich narzędzi na bieżąco jest niezbędne. Dodatkowo ciągły rozwój technologiczny wymaga modyfikacji raz utworzonych skryptów tak, by uwzględniały nowości. Podanym wcześniej jaskrawym przykładem jest powstanie nowej wersji programu ABBYY Fine Reader. Oprogramowanie umożliwiające segregację materiału na gotowych plikach DjVu wymagało pewnych modyfikacji ułatwiających sprawdzanie materiału i łatwe zarządzanie nim.

Możliwość obsługi programu za pomocą linii komend jest też pewnym kryterium

doboru narzędzi przy digitalizacji zasobów. Brak wsparcia tego rodzaju często dyskwalifikuje oprogramowanie z użytku w JBC. Ze względu na automatyzację działań ustawianie tych samych funkcji za każdym razem bez możliwości ustalenia schematu w oczywisty sposób utrudnia automatyzację pracy.

Najważniejszym programem, który wymaga ustalenia schematu i wsparcia poleceń z linii komend jest konwerter do formatu DjVu. Ustalenie jednego profilu dla całego digitalizowanego materiału skutkuje niepożądanymi efektami. Często jednak poszczególne numery czasopisma mają podobne wymogi przy segmentacji jednego woluminu, co pozwala na ustawienie schematów wywoływanych za pomocą odpowiednich skryptów. Dla potrzeb masowej konwersji plików do formatu DjVu stworzono jeden skrypt z wieloma funkcjami. Pierwszą z nich jest wykonanie segmentacji, podczas której można dobrać parametry najważniejszych czynników decydujących o zawartości maski. Wybiera się przy tym formę zapisu (*bundle* i *indirect*), ustala się wartość *reshold*, *shapefilter* i *block.size*. Ten ostatni parametr niestety nie ma możliwości wywołania z poziomu linii komend. Ta funkcja nie została uwzględniona przez twórców segmentera, ponieważ uznali ją za zbyt kłopotliwą w modyfikacji, dodatkowo większe bloki utrudniają kodowanie warstwy tekstu [Lin i in. 2005, s. 994; Haffner i in. 2002, s. 33]. Można jednak ustalić profil z tym parametrem i następnie wy-

wołać go podczas segmentacji. Wpisanie jakiegokolwiek profilu konwersji jest i tak konieczne. Nazwy folderów i plików zostały już wcześniej ustalone, a skrypt z nich korzystający nadaje identyczne nazwy wytwarzanym plikom.

Wykonywanie tego typu czynności jest o wiele szybsze za pomocą linii komend nie tylko ze względu na nałożenie schematu na części powtarzalne parametrów konwersji, co w przypadku interfejsu graficznego programu nie jest takie oczywiste. Tego typu rozwiązanie ułatwia przede wszystkim modyfikację profilu segmentacji, gdyż w normalny sposób trzeba za każdym razem korzystać z kreatora przygotowanego przez twórców. W przypadku precyzyjniejszej pracy jest to niemożliwe do wykonania w sposób zautomatyzowany.

Za pomocą tego rodzaju skryptów w łatwy sposób można podmienić też wykonany proces OCR. Zawartą pętlę w tym narzędziu ukazuje il. 3. Najczęściej ta funkcja jest wykorzystywana przy two-

żeniu obrazów niesegmentowanych, które zawierają tekst. Zanim możliwe było wykorzystanie oprogramowania ABBYY Fine Reader tę czynność musieliśmy wykonywać przez podmianę informacji w pliku .xml. Wsparcie przez wytworzony w tym celu skrypt umożliwiło automatyzację pracy przede wszystkim na wielu numerach czasopisma.

Poza skrypcem wykonującym segmentację, innym ważnym narzędziem ułatwiającym czynności powtarzalne jest zamiana sposobu zapisu pliku (z *bundle* na *indirect* i odwrotnie). Sama czynność nie wymaga większej automatyzacji, ale podczas jej wykonywania dodano opcję wykonania miniatur z podanej przez użytkownika strony, która następnie jest przystosowywana do przyjętych w JBC standardów (rozmiar 300x450 pikseli). Wykonywanie tej czynności ręcznie dla każdego numeru czasopisma w oczywisty sposób utrudnia pracę. Dla materiałów specjalnych, takich jak starodruki, rękopisy czy rękopisy muzyczne bibliotekarz cyfrowy i tak musi ustalić

```
for %%a in (*.djvu) do (
  mkdir %%~na
  ddjvu -size=600x900 -format=tiff -page=%~1 %%a %%~na.tif
  nconvert -out jpeg -resize 300 450 %%~na.tif
  del %%~na.tif
  djvmcvt -i %%~na.djvu %%~na %%a
  move %%~na.jpg %%~na/%%~na.jpg
  IF EXIST %%~na.ini (move %%~na.ini %%~na/%%~na.ini)
  del %%a
)
```

© Il. 2.

© Il. 2. Przykładowa pętla wykonująca miniatury i zmieniająca formę zapisu pliku DjVu na indirect.

wygląd miniatury samodzielnie, inaczej tak wykonana miniatura nie przedstawiałaby potrzebnej treści. Miniatury dla numerów czasopisma można wykonać za pomocą pętli w *batch*, którą przedstawia il. 2. Zamieszczony przykład wykorzystuje programy DjvuLibre i XnView (a konkretniej narzędzie odpowiadające za pracę z poziomą linią komend).

Wsparcie programistyczne jest pomocne także podczas naprawy różnego rodzaju błędów. Dobrym przykładem będzie powtarzający się stale błąd silnika OCR, zamieszczonego w oprogramowaniu Document Express Enterprise – Read I.R.I.S [Caminoval]. Wykonując rozpoznanie w języku polskim program często wpada w zapętlenie i tym samym nie może zakończyć procesu z takimi wynikami, jakie uzyskał. Bez możliwości korzystania z linii komend wymagane byłoby powtórzenie całego procesu wraz z segmentacją od nowa i ustawienie innych parametrów konwersji. Tymczasem sam etap dzielenia dokumentu na warstwy mógł zostać wykonany prawidłowo.

```
for %%a in (*.djvu) do (
  djvubundle.exe --ocr=Lang=[Polish][German] "%%a" "%%~na.djvu"
)
```

© Il. 3.

© Il. 3. Przykładowa pętla wykonująca proces OCR na gotowym dokumencie DjVu.

Masowa konwersja cyfrowych kopii do formatu DjVu

Pliki graficzne można przechowywać i udostępniać w różnych formatach. Publikacja zasobów w Internecie ogranicza w pewien sposób dobór zapisu pliku zamieszczonego na serwerze. Przy archiwizowaniu materiału ważne jest przede wszystkim, by zachować możliwie jak najwierniejszą kopię oryginału [Płoszajski i in. 2008, s. 202–204; Paradowski i in. 2010, s. 26–34; Hansen i in. 2006, s. 1–2]. Obraz archiwalny nie jest poddawany jakimkolwiek procesowi korekty. Kopia cyfrowa przeznaczona do publikacji w Internecie musi spełniać inne wymogi [Janiak i in. 2011, s. 103; Breuel i in. 2003, s. 81–82]. W warunkach udostępnienia dużego zasobu na serwerze jako główne cechy formatu uznano: ilość zajmowanego miejsca na dysku przez plik przy możliwie najlepszej jakości przedstawienia wizualnego a także łatwość jego przeglądania wprost z przeglądarki. Jeszcze przed rozpoczęciem tworzenia Jagiellońskiej Biblioteki Cyfrowej zdecydowano się na format DjVu jako podstawową formę udostępniania zasobów. Pliki zapisywane w kolorze są stosunkowo niewielkie i możliwe jest przeglądanie pojedynczych stron bez konieczności ładowania całego dokumentu. Nadto pojedyncza strona najpierw wyświetla się w swojej formie czarno-białej, natomiast pozostała jej treść będąca w kolorze doczytuje się w trakcie umożliwiając tym samym przeglądanie dokumentu przy stosunkowo wolnym przesyłaniu danych

przez Internet. Dodatkowo na każdej stronie zawarte są dane rozpoznanego tekstu, które zamieszczone są pod warstwą obrazu zeskanowanej strony. Możliwe jest też wydzielenie samych informacji z przeprowadzonego procesu OCR i ich modyfikacja [Haffner i in. 2002, s. 2–4; Borgman 2000, s. 82–84; LeCun i in. 2000, s. 8–16]. Poza walorami technologii kompresji ten format został również wybrany przez większość bibliotek cyfrowych w Polsce tworząc obecnie swojego rodzaju standard w procesie digitalizacji.

Sama automatyzacja pracy przy konwersji obrazu grafiki rastrowej do formatu DjVu sprowadza się do kilku zagadnień. Pierwszym jest czas poświęcony na wykonywanie zadań. Im szybciej zrealizuje się zamierzone czynności, tym większa może być wydajność sprzętu i oprogramowania, a co za tym idzie – koszty digitalizacji się zmniejszają. Drugim zagadnieniem jest zautomatyzowanie wielokrotnie powtarzanych czynności w celu już nie tylko przyspieszenia ich realizacji, ale także dla poprawy wygody pracy.

Wyżej wymienione cele przy digitalizacji zbiorów wymagają konkretnych rozwiązań. Przy organizowaniu konwersji kopii cyfrowych do formatu DjVu w przypadku czasopism pojawiają się trudności w postaci specyficznie posegregowanych plików. Wykonywanie wszystkich numerów czasopisma w sposób niezorganizowany wiąże się z prze-

glądaniem i wykonaniem osobnej konwersji dla pojedynczego egzemplarza. W tym przypadku automatyzacja wiąże się z tym, by duże ilości materiału segmentować w jednym procesie [Allen i in., 2010, s. 98]. Takie rozwiązanie eliminuje wymóg nadzoru nad każdą czynnością będącą identyczną dla poszczególnej części woluminu danego tytułu.

Z powyższych rozwiązań wynika, że aby rozpocząć proces masowej konwersji należy ocenić, ile elementów dla jednego tytułu czasopisma (także dla serii zdjęć czy poszczególnych rozdziałów książki) jest identycznych, bądź podobnych. W przypadku gazety powinno się zacząć od nałożenia poprawnej struktury plikom przeznaczonych do konwersji. Gdy każdy z nich będzie w odpowiednim folderze z konkretną, identyfikowalną nazwą, to program będzie mógł wykonać osobny proces na każdym folderze i plikom wynikowym może narzucić ustaloną wcześniej strukturę.

Przy konwersji masowej do formatu DjVu samo ustalenie struktury czasopisma nie wystarczy. Każda strona gazety czy książki może być w sposób indywidualny zniekształcona i przez to wymagać będzie szczególnej uwagi. Sam format ze względu na stosowanie segmentacji też nie jest łatwym rozwiązaniem przy pracy. O wiele prościej jest każdy plik przekonwertować na formę zapisu podstawowej grafiki rastrowej (np. .tiff czy .jpg). Niepoprawna konwersja do formatu DjVu skutkować będzie nie tylko kopią cyfro-

wą gorszej jakości, ale może też mieć decydujący wpływ na przeprowadzony po segmentacji proces OCR.

W kontekście cyfryzacji zbiorów na szerszą skalę ważne jest też, by była możliwość organizacji pracy masowej przy tworzeniu plików w formacie DjVu. Oprogramowanie do tego przeznaczone pozwala na przystosowanie ustawień tak, by duże ilości materiału mogły być konwertowane w jak najkrótszym czasie. Żeby jednak te ustawienia zastosować, należy najpierw odpowiednio zorganizować i ocenić zasób cyfrowy pod kątem segmentacji. Jest to konieczne z tego względu, że przyjęcie takich samych ustawień dla wszystkich pozycji podlegających procesowi digitalizacji zaowocuje efektem niepożądanym. Nawet z pozoru jednolity materiał, jakim może być jeden tytuł czasopisma, wymaga różnych ustawień dla poszczególnych roczników czy woluminów. Gdy w serii numerów jeden z nich okaże się różny od pozostałych, a poprawna konwersja będzie wymagać szczególnego profilu, wyłącza się go z procesu masowej segmentacji.

Aby zdecydować, jakie ustawienia mają być stosowane, należy najpierw ocenić, jak ma wyglądać wynikowy plik w formacie DjVu. Zwykła grafika rastrowa w standardach zapisu obrazu z reguły ogranicza się do pełnej automatyzacji, nawet przy zastosowaniu zaawansowanych przepłotów, jak to ma miejsce np. przy kompresji JPEG2000 [Acharya i in. 2005, s. 82, 140–144; Bell 1990,

s. 6–10; Van der Knijff 2011; Kulovits i in. 2009]. Bardziej zaawansowanym modelem kompresji jest MRC (Mixed Raster Content) [De Queiroz i in. 1999, s. 1106–1107]. Jest to o wiele praktyczniejsza opcja od stosowania technik słownikowej kompresji poza samym przeplotem z pogranicza wektoryzacji obrazu [Führ i in. 2006, s. 179–180; Maheswari i in. 2011, s. 271; Feng i in. 2006]. Jak nazwa wskazuje, celem kompresji podobną metodą jest dostosowanie algorytmu do różnorodnych treści. Na takim podziale można dla każdej sekcji wybrać sposób kompresji. Dla czarnobiałej treści jest wybierany inny format, a dla kolorowego tła inny. Stworzony przez firmę AT&T Labs system segmentacji plików jest typowym rozwiązaniem MRC. Zastosowano w nim trzy różne warstwy, każda kodowana w inny sposób. W tego typu rozwiązaniu, ze względu na bardziej zaawansowane techniki kompresji, konieczne jest definiowanie ustawień całego procesu. Daje to większe możliwości dostosowywania obrazu do aktualnych potrzeb, ale też pociąga za sobą wymogi zaznajomienia się z dokumentacją techniczną oprogramowania [De Queiroz 2006, s. 324–326]. Użycie mieszanej zawartości rastrowej jest też dostępne w programie ABBYY Fine Reader, który tego typu ustawienia przenosi na różne formaty (w tym PDF), ale efekt końcowy ze względu na pełną automatyzację procesu jest często losowy i niekorzystny.

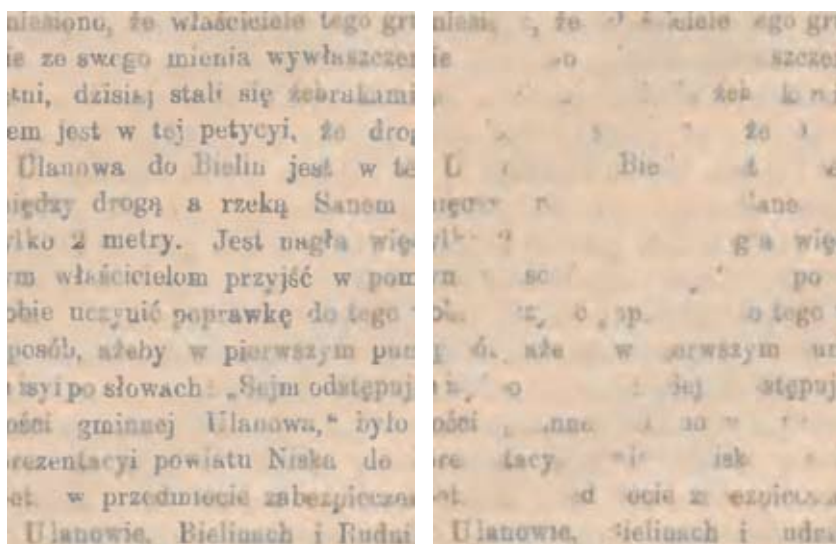
W formacie DjVu w celach największej kompresji stosuje się podzielenie doku-

mentu na warstwy i zapisanie go na czterech różnych płaszczyznach. Pierwszą jest warstwa tła (w specyfikacji nazywana *background*). Podlega ona znacznej modyfikacji i jest tym samym uzależniona od następnej części dokumentu, jaką tworzy maska (*mask*). Ta jest konwersją pliku graficznego do formy czarno-białej, dodatkowo kompresowanej przystosowanym do pracy na zeskanowanych dokumentach różnej treści algorytmem JBIG2 [Salomon 2007, s. 378; de Queiroz 2006, s. 327; Boncelet 2006, s. 290; Bottou i in. 2000]. Tego typu algorytm z góry zakłada przy kompresji stratnej dzielenie dokumentu na bloki, które w zależności od treści na nich zawartej w różnym stopniu podlegają przekształceniom na tryb monochromatycznej palety barw [Tompkins i in. 1999, s. 224–228]. Format JBIG2 został włączony do standardów International Telecommunication Union w lutym 2000 roku [ITU-T.88]. W zamierzeniu czarne sekcje obrazu zawierać powinny jedynie litery. Maską pełni też funkcję informacji o kształtach tekstu czy granicach każdej z liter. Na podstawie tych danych program wycina z tła piksele, które rozpoznano, jako należące do liter. Luki są uzupełniane sąsiednimi kolorami dając wrażenie pustej karty. Wycięte piksele przechodzą z kolei do warstwy tekstu (*foreground*). Tam, zachowując swoje kolory, ograniczone są kształtem do czarnych pikseli z maski. Zarówno *foreground* jak i *background* kodowane są stworzonym przez AT&T Labs formatem IW44. DPI dwóch warstw, tła i tek-

stu, jest dzielone przez taką liczbę, jaka jest podana w profilu i tym samym obraz zajmować może znacznie mniej miejsca. Strata tego typu z reguły w plikach graficznych kończy się rozmyciem. Format DjVu dzięki zastosowaniu maski pozwala na maksymalne wyostrenie tekstu. Dokument jest dzięki temu w pełni czytelny. Cały proces dzielenia obrazu twórcy nazywają segmentacją [LeCun i in. 2001, s. 119–121; Bottou i in. 2000, s. 4; Bottou, Haffner, i in. 1998, s. 412; Haffner i in. 2002, s. 3–8; de Queiroz 2006, s. 327]. Czwartą wspomnianą warstwą jest informacja o rozpoznanym tekście. Na potrzeby zamieszczenia tych danych wykonać należy proces OCR, czyli optyczne rozpoznanie liter. Może być on wykonany przed segmentacją lub na gotowym pliku. Różnica ta ma duże znaczenie dla precyzji rozpoznania tekstu, gdyż pliki w wysokiej jakości dadzą z reguły o wiele lepsze rezultaty niż dokument w formacie DjVu, którego kompresja jest w wyraźny sposób stratna³⁴ [Tompkins i in. 1999, s. 224–228; Holley 2009; Powell i in. 2009; Kalota i in. 2011, s. 56]. Tym bardziej jest to istotne, gdyż rozpoznawanie tekstu odbywa się na czarno-białej warstwie maski, której DPI wprawdzie nie zostało podzielone, ale monochromatyczność barw wyraźnie wpływa na efekt. Dobrze jest więc dobrać takie oprogramowanie, które spełnia podstawowe kryterium jak najbardziej poprawnego procesu OCR [Tanner i in., 2009]. Można również wykorzystać niepoprawnie wykonany proces OCR w celach wyszukiwawczych stosując bardziej

zaawansowane algorytmy wyszukiwawcze [Bieć 2011, s. 11–12].

Chociaż format DjVu daje bardzo niejednolite wyniki i często pliki mogą odbiegać przedstawianą zawartością od pierwowzoru, to zastosowanie odpowiednich ustawień może zminimalizować straty wizualne. Pierwszym i najważniejszym kryterium, jakim należy oceniać gotowe dokumenty w postaci cyfrowej jest czytelność zawartej w nim treści. Dla potrzeb publikacji w Internecie często pozostaje to jedynym wymogiem. Tak więc, aby dokument był czytelny, należy wykonać poprawnie czarno-białą maskę. Jeżeli w tej warstwie litery będą łączyły się z innymi obiektami, lub ze względu np. na stratę pigmentu nie zostaną przy segmentacji uwzględnione, część treści może przestać być czytelna. O ile przy procesie OCR wykonanym na pliku źródłowym, który zawsze jakiejś błędy będzie zawierał, może nie mieć to tak dużego znaczenia, o tyle w przypadku rozpoznawania tekstu na zakodowanej dwubarwnej masce może dać już wysoce niepożądane rezultaty. Pomimo skuteczności procesu rozpoznawania liter źle wykonana maska zaowocuje dodatkowo wizualnie dużą stratą ostrości tekstu. Niejednolicie nałożona czarno-biała warstwa daje efekt różnie wyglądających liter w tym samym wyrazie, co wyraźnie utrudnia czytelność. Taki rezultat przedstawia il. 4. Z kolei dodatkowe obiekty w warstwie tekstu, czasem nakładające się na litery, uniemożliwiają wręcz odczytanie wyrazów. Dzieje się tak dlatego, że



II. 4.

format DjVu w inny sposób koduje warstwę tekstu, a inaczej warstwę tła [Bottou i in., *Efficient...*].

Przy masowej konwersji poprawne utworzenie maski w dokumentach jest możliwe z reguły jedynie w przypadku, gdy karta zawiera czarny tekst na jasnym tle. Podobna sytuacja ma miejsce w przypadku czarno-białych grafik zamieszczonych na stronach dokumentu. Podlegające w zasadzie tym samym ustawieniom, jakie stosuje się przy segmentacji tekstu. Problem może sprawiać kolorowy obraz. Teoretycznie najlepszym rozwiązaniem powinno być zaaplikowanie specjalnych filtrów proponowanych przez segmenter wyszukiwujących odpowiednie kształty, które potencjalnie tworzą litery (tzw. *shapefilter*). Stosowanie tego rodzaju narzędzia pozwala osiągnąć efekt podzie-

lenia dokumentu na maskę zawierającą jedynie litery i na tło, w którym zawarty jest obraz. Taka sytuacja sprawdza się dobrze głównie w plikach z grafikami o niewielkim kontraście bądź jasnych barwach. Często jednak w dokumentach obrazy są różnorodne, zawierają kontrastujące ze sobą elementy, zarówno jasne jak i białe. W takim przypadku filtr kształtów znajduje na obrazie formy odpowiadające literom, które są następnie częścią maski. Poza utrudnieniem czytelności dokumentu tak wykonana segmentacja skutkuje też obniżeniem walorów wizualnych dokumentu. Powodem tego jest stosowanie innego kodowania dla każdej z warstw, co jest typowym rozwiązaniem w modelu kompresji MRC. Jeżeli dodatkowo proces OCR jest wykonywany po segmentacji, elementy tego obrazu są rozpoznawane

II. 4. Z lewej dokument z niepoprawną segmentacją, po prawej tło tego dokumentu.



II. 5.

jako tekst. Nawet twórcy formatu DjVu z AT&T Labs, przedstawiając możliwości segmentacji obrazów na potrzeby publikacji w Internecie, nie wykazują pełnej funkcjonalności [LeCun i in. 1998, s. 222–223; Bottou, Pigeon 1998; Haffner i in. 2000]. Wykonanie natomiast zadania z zadowalającymi rezultatami w sposób zautomatyzowany i szybki jest niemożliwe. Niepożądany rezultat wynika z wykorzystywania wspomnianego algorytmu JB2. Właśnie to narzędzie dzieli obraz na bloki i przydziela dwubarwnej warstwie schemat czy to tekstu, czy obrazu, czy pustej karty, zarówno białej, jak i czarnej, co owocuje w tym przypadku ujednoczeniem całego bloku do jednego z dwóch kolorów [Boncelet, 2006, s. 290–291; Haffner i in. 1998, s. 312]. Wspierający podział kodowania shapefilter nie odpowiada za określanie

sektorów tekstu, tę operację wykonuje algorytm. Optymalnym rozwiązaniem w tej sytuacji byłoby ręczne ustawienie tych bloków. Oprogramowanie jednak nie daje takiej możliwości, choć JB2 tego sposobu nie wyklucza [Taubman 2006, s. 52; Haffner, *DjVu: Analyzing...*].

Kiedy źle przeprowadzona segmentacja w wyraźny sposób zniekształca grafikę zamieszczoną w dokumencie, przy konieczności pracy masowej lepiej jest zrezygnować z dzielenia pliku na warstwy. Oprogramowanie konwertujące pliki do formatu DjVu daje taką możliwość poprzez umieszczenie całej kopii cyfrowej dokumentu w warstwie tła. Sprawia to jednak potem pewne trudności w wykonaniu procesu OCR. Rozpoznanie tekstu nie jest jednak niemożliwe. Gdy proces ten wykonywany jest przed seg-

II. 5. Z lewej poprawna maska dokumentu z czarno-białym obrazem, z prawej tło tego dokumentu.

```
for %%a in (*.tif) do (
documenttodjvu.exe --page-range=1- --profile=scan300 --threshold-level=
40 --shape-filter-level=0 %%a %%~na.djvu
djvubundle.exe --ocr=Lang=polish %%~na.djvu %%~na.djvu
djvutoxml %%~na.djvu %%~na.xml
del %%~na.djvu
documenttodjvu.exe --page-range=1- --profile=photo300 %%a %%~na.djvu
djvuparsexml %%~na.xml
)
```

II. 6.

```
#@displayName:przykładowy_profil      thickening=1
przykładowy_profil: scan300            render-size=0
description="przykładowy_profil"      inhibit-foreback-level=95
dpi=-300                               blurring-size=5
upsample=1                             block-overlap=3
mask-upsample=1                       inversion-level=0
subsample=1                           block-size=500
bg-subsample=3                        edge-size=3
fg-subsample=12                       high-variation-foreground=false
quality=80                             subsample-refine=false
pix-filter-level=15                   pages-per-dict=1
shape-filter-level=0
threshold-level=40
```

II. 7.

mentacją, rezygnacja z dzielenia obrazu nie ogranicza w niczym zamieszczenia danych o zawartych w dokumencie literach. Natomiast w przypadku, gdy oprogramowanie wykonuje tylko i wyłącznie OCR na masce, wówczas można wykonać dwa identyczne pliki, z których pierwszy będzie posiadał rozpoznane litery i zostanie posegmentowany, a drugi pozostanie wierny oryginałowi. Następnie wystarczy wyjąć informację o tekście zawartym w dokumencie i przenieść ją do ostatecznego, niepodzielonego pliku. Używane w JBC oprogramowanie

Document Express Enterprise wykonuje proces OCR jedynie na masce, ale umożliwia podmianę danych poprzez element *djvutoxml.exe* i *djvuparsexml.exe*.

W sytuacji pliku nierozdzielanego na warstwy szczególna dbałość o obraz jest tak samo jak w przypadku zmiany formatu na JPEG, wszystko odbywa się automatycznie (stosowany jest w tym przypadku schemat konwersji IW44). Takie rozwiązanie ma jednak swoją wadę. Waga utworzonych w ten sposób dokumentów jest

o wiele wyższa niż posegmentowanych odpowiedników. Z tego względu ten sposób stosowano rzadko i nie masowo, starając się dokument segmentować, ewentualnie dodając jedynie jedną stronę bez segmentacji.

Jak już wspomniano wyżej, odpowiednia segmentacja sprowadza się do utworzenia poprawnej maski, czyli zamiany zeskanowanej strony dokumentu z kolorowej palety barw (czy względnie ze skali szarości) do dwukolorowego monochromatycznego zapisu. Takie operacje wykonują w zasadzie wszystkie programy graficzne będące obecnie w użyciu. Z reguły funkcjonowanie rozdzielania obrazu na czarne i białe kolory sprowadza się do ustalenia wartości progowej dla barw (tzw. *threshold*) [Shapiro i in. 2001, s. 94]. Program ustala średnią arytmetyczną liczb opisujących barwę poszczególnego piksela (dla RGB będą to trzy liczby, po jednej dla kolorów czerwonego, zielonego i niebieskiego). Na podstawie wybranej wartości progowej program tworzy histogram. Jest on podstawą do oceny, czy liczba opisująca barwę piksela jest większa (w takim przypadku jego kolor zostanie zamieniony na biel), czy mniejsza (odpowiednio kolor zmieni się w czerń) [Gonzalez i in. 2002, s. 568–572, 595–611]. Dla tekstu zapisanego na jasnej karcie znacznie trudności z wykonaniem poprawnego obrazu dwubarwnego nie ma. O wiele większe problemy stwarzają zeskanowane dokumenty z XIX wieku i wcześniejsze. Często zdarza się, że litery są na tyle

jasne, że w wielu miejscach karta jest od nich ciemniejsza. W takim przypadku sama wartość progowa okazuje się co najmniej nieodpowiednia.

Wśród różnych modeli segmentacji obrazów stosuje się odpowiednie rozwiązania. Aby obraz przy zastosowaniu progu odcieni został rozdzielony w prawidłowy sposób, w większym lub mniejszym stopniu musi zostać dokonana jego analiza [Jähne 2005, s. 450–451]. Z tego względu konwerter plików do formatu DjVu posiada o wiele bardziej zaawansowane opcje, które w połączeniu z algorytmem JB2 współdziałają w zamiarze dając o wiele lepsze efekty [De Queiroz 2006, s. 338; Haffner i in. 2002, s. 19; Mello 2007, s. 917; Shafait i in. 2006]. Większość z tych opcji nie wpływa znacząco na efekt końcowy, ale kilka z nich jest kluczowe dla utworzenia poprawnej maski. Można je też w łatwy sposób stosować przy automatyzacji pracy. Podstawowym filtrem jest wspomniany *threshold*. O ile w programach graficznych z reguły wartości progowe zamieszcza się w przedziale 0-255 (na co pozwala 8-mio bitowy zapis na każdy kanał barw), to oprogramowanie segmentujące Document Express stosuje przedział procentowy (0-100). Stosowanie wartości progowej 80 sprawia, że tekst w większości dokumentów jest uwzględniony w treści maski, a dodatkowe elementy zbliżone jasnością pozostają w tle. Taki scenariusz należy jednak do rzadkości i w większości przypadków temu współczynnikowi przypisuje się wartość o wiele niższą,

II. 6. Przykładowa pętla podmieniająca OCR korzystająca z poleceń linii komend oprogramowania Document Express Enterprise.

II. 7. Przykładowy profil wykorzystujący większość ustawień dostępnych w profilach segmentacji DjVu programu Document Express Enterprise.

kiwał godzinami w strasnym zaduchu, w brudnych norach sądu powiatowego karnego. Chwile te niewątpliwie silnie zapisały się w jego

CYGARNICZKA

przez

Artura Gruszeckiego.

79

(Ciąg dalszy).

Dopieroż matka będzie rozpaczała, bo kocha Sobka, i któżby nie kochał takiego syna śmiałego, i ogarnęło ją wielkie współczucie na myśl rozpasy matki.

kiwał godzinami w strasnym zaduchu, w brudnych norach sądu powiatowego karnego. Chwile te niewątpliwie silnie zapisały się w jego

GARNICZKA

przez

Artura Gruszeckiego.

79

(Ciąg dalszy).

Dopieroż matka będzie rozpaczała, bo kocha Sobka, i któżby nie kochał takiego syna śmiałego, i ogarnęło ją wielkie współczucie na myśl rozpasy matki.

© II. 8.

© II. 8. U góry maska z wartością *blocksize* 16, u dołu z wartością 500, pozostałe ustawienia takie same.

co pociąga za sobą konieczność zmiany ustawień innych filtrów w specyficzny sposób.

Drugim znaczącym elementem segmentującym jest wspomniany *shapefilter*. To narzędzie pozwala nie tylko rozróżnić tekst i obraz, ale także znacząco poprawić filtrowanie ciemniejszych sektorów kart przy zastosowaniu niższego progu. Wpisywanie wartości niemniejszej niż 10 z reguły też wyraźnie poprawia proces OCR, wykonywany na rozdzielonym na warstwy dokumencie. W przypadku materiałów starszych stosowanie wysokich wartości *shapefilter* może skutkować odrzuceniem z maski treści i tym samym usunięciem części wyrazów. Tego typu błędy przedstawia il. 4. W JBC przyjęto metodę zamieszczenia w warstwie maski wszystkich liter z tekstu ignorując czasem niepożądane elementy. Założono, że przy tak tworzonych plikach możliwe jest dzielenie DPI tła w większym stopniu, co skutkuje mniejszymi plikami. Gdy maska zawiera każdą literę z segmentowanego obrazu, czytelność dokumentu pozostaje na tyle klarowna, że czytnik nie ma wątpliwości co do zawartej w nim treści. Praktyka pokazała, że realizacja tych założeń jest możliwa w większości przypadków, dzięki wyłączeniu tego narzędzia poprzez wpisanie wartości równej 0.

Trzecim elementem pozwalającym uzyskać lepsze efekty przy segmentacji jest *pixfilter*. Często pojedyncze piksele tła są zamieniane na czarny kolor ze wzglę-

du na zbyt ciemne zabarwienie. Czasami podlegają temu nawet grupy kilku pikseli. *Pixfilter* eliminuje z maski te zabrudzenia, dzięki czemu pojedyncze kropki nie znajdują się w przypadkowo rozpoznawanym tekście. Te ustawienia ograniczają również wielkość pliku wynikowego i poprawiają efekt wizualny prezentowanego materiału. Zadowolające efekty daje stosowanie wartości z przedziału 5-30, z reguły wpisuje się liczbę 15.

Ostatnim często używanym w JBC filtrem jest *blocksize*. Ten element stanowi bezpośrednio odwołanie do algorytmu JB2, odpowiedzialnego za konwersję obrazu do monochromatycznej palety barw. Ustala on wielkość pojedynczego bloku, który jest przekazywany kodownikowi JB2, a ten decyduje o zastosowaniu odpowiedniego schematu konwersji. Tak więc ta funkcja nie ustala w jakikolwiek sposób jakości pliku wynikowego, jak to ma miejsce w przypadku formatów stosujących przeplot [De Queiroz 2006, s. 337]. Podczas stosowania gotowych profili stworzonych przez twórców oprogramowania, wartość tego współczynnika nie jest brana pod uwagę i w związku z tym przyjmuje liczbę domyślną 16 (co odpowiada ilości pikseli w pionie i poziomie przypadających na blok). Stosowanie większych przedziałów pozwala na skuteczniejsze wykorzystanie możliwości algorytmu ograniczania kolorów obrazu do czerni i bieli. Twórcy nie zalecali stosowania tego narzędzia, gdyż rzeczywistość sprawia ono wiele trudności i jest powodem wielu błędów. Niemniej jednak

odpowiednie dobranie bloku sprawia, że maska w wielu sektorach nie posiada dodatkowych zanieczyszczeń. Głównie niezapisane marginesy podlegają redukcji niewłaściwej treści, gdyż szesnasto pikselowy blok w tych rejonach często algorytm JB2 odczytuje jako sektor o pustej zawartości i ciemnym zabarwieniu, a co za tym idzie, podmieniany jest przez czarny kolor. Tak duża ilość dodatkowych elementów maski poza zmniejszeniem walorów wizualnych dokumentu zwiększa też jego objętość. Z tego względu często używano ustawień ustawień pomiędzy 500 a 1000, czasem stosując wartość nawet powyżej dwóch tysięcy. Trzeba zwrócić też uwagę na możliwość błędu występującego podczas segmentacji z zastosowaniem dużego współczynnika *blocksize*. Wielkość bloku nie może przekroczyć liczby pikseli krótszego boku obrazu, inaczej koder JB2 nie wykona konwersji [Lizardtech DjVu Reference]. Trzeba jednak zaznaczyć, że jeden blok podlega uśrednionej segmentacji, tzn. że proces *thresholdingu* oceniając na podstawie histogramu jednego bloku [Popat i in. s. 363], w wartości maski uwzględni tylko najciemniejsze litery. Jaśniejsze elementy tekstu pozostaną w tym przypadku w tle. Tego typu efekt przedstawia il. 8., na której widać segmentację strony dokumentu z przylepioną taśmą z lewej strony. Trójkątny róg białej taśmy wybielił obraz w części dokumentu i tym samym przy dużej wartości *blocksize* treść pozostała nierozpoznana przez segmenter.

Pozostałe filtry nie odgrywają tak znaczącej roli, by przystosowywać je w większym stopniu do pracy masowej. Przy każdym pliku, po utworzeniu poprawnej maski, trzeba też odpowiednio dobrać wartości współczynników decydujących o jakości i wielkości pliku końcowego. Tymi ustawieniami są: *bg-subsample*, *fg-subsample*, *quality* i *fg-quality*. Pierwsze dwa, jak już wspomniałem, dzielą DPI warstw przez liczbę, która jest im przypisana (w JBC stosowano z reguły dzielnik 3-6 dla tła i 12 dla tekstu, co daje ostatecznie DPI 100-50 i 25). Dwa ostatnie decydują o jakości stratnej kompresji kolorowego obrazu [Haffner i in. 2002, s. 20]. W większości materiału używano wartości 80, czasami zmniejszając ją do 50.

Przy segmentacji dobrze zachowanych czasopism, których konwersja nie stwarza większych trudności, stosowanie ustawień według powyższych przykładów pozwala na wykonanie poprawnej konwersji. Obniżenie wartości *threshold* do 40 w wielu przypadkach okazuje się wystarczające, by pozostać przy masowych metodach organizacji pracy z formatem DjVu. Aby zachować jak największą dbałość o walory wizualne kopii cyfrowej przy tak niskiej wartości progów kolorów, należy stosować w profilu zestawienie wyżej wymienionych narzędzi. Powiększony *blocksize* zmniejsza ilość niepożądanego materiału w masce, *pixfilter* eliminuje widoczność zabrudzeń karty, a *shapefilter* poprawia selekcję tekstu i OCR.

Innymi operacjami wykonywanymi masowo przy organizowaniu zasobu zdigitalizowanego w formacie DjVu jest zmiana form zapisu pliku. Kopia cyfrowa wykonana w tym formacie daje możliwość różnych opcji archiwizacji. Wielostronicowa książka może być przechowywana jako pojedynczy plik zawierający wszystkie strony (tzw. *bundle*). Istnieje też opcja rozdzielania każdej strony osobno (zapis pliku w tzw. *indirect*). W takiej formie pojedyncze części dokumentu nie są powiązane i przeglądać można jedynie jedną stronę. Do całego dokumentu z rozdzielonymi elementami można dodać plik indeksu zawierający informacje o wszystkich stronach wchodzących w skład jednostki. Nie zawiera on żadnych danych odnoszących się do treści dokumentu, a jedynie informacje o plikach wchodzących w jego skład. Przeglądanie publikacji przez indeks pozwala na otwarcie wszystkich stron w taki sam sposób, jakby były w zapisane w jednym pliku. Każdy sposób zapisu ma swoje walory. Plik złączony w całość pozwala na łatwiejsze organizowanie archiwizacji materiału. Z kolei przy publikacji w Internecie lepiej jest zamieścić pojedyncze strony z indeksem i skierować odnośnik do pliku indeksu. Pozwala to na otwieranie pojedynczych stron wywoływanych z serwera w przeglądarce bez konieczności ładowania całego dokumentu.

Oprogramowanie w momencie segmentacji pozwala na podjęcie decyzji odnośnie formy zapisu pliku. Zmiana opcji w późniejszym etapie jest możliwa rów-

nież na sposób masowy. Tego typu operacje wykonuje z dobrymi rezultatami program DjVuLibre. Za jego pomocą łatwo można rozdzielić połączony dokument w krótkim czasie. Łączenie plików w jedną całość stwarza już pewne problemy. W przypadku, gdy waga pliku nie przekracza 200MB nie jest zauważalny spadek wydajności. Ze względu na to, że DjVuLibre działa w środowisku 32-bitowym zarządzanie pamięcią powoduje wydłużenie czasu pracy. Może się zdarzyć, że dokument zawierający ponad dwieście stron indeksować się będzie kilkanaście czy nawet kilkadziesiąt godzin. Przy pracy masowej jest to spore utrudnienie, zważywszy na to, że jest to jedynie operacja zmiany formy zapisu. W tym przypadku byłoby to nieopłacalne. Można jednak wykonać to za pomocą programu segmentującego. W takiej sytuacji należy wykonać ponownie całą segmentację i w programie Document Express Enterprise połączyć plik w całość bez większej straty czasu. Po tej operacji zapisane informacje z wcześniejszego procesu OCR zostaną jednak utracone. Można je odzyskać poprzez podmianę w pliku .xml opisaną wyżej, lub poprzez ponowne rozpoznanie tekstu (wykonane jedynie na czarnobiałej masce). W obu przypadkach ta czynność wymaga sporej ilości czasu, niemniej jednak będzie ona szybsza od ponownego zindeksowania dokumentu programem DjVuLibre.

Program Document Express Enterprise jest również napisany w środowisku 32-bitowym. W związku z tym pojawiać

się mogą błędy podczas jego użytkowania. Związane są one z wytwarzaniem miniatur na dużych dokumentach zapisywanych w formie bundle. Pliki powyżej 400MB nie będą mogły być zapisane głównie dlatego, że bufor pamięci RAM zostanie obciążony dekodowanymi wersjami pliku koniecznymi do wykonania miniatury. Z tego względu przy tego typu materiałach zrezygnowano z tworzenia dodatkowych plików, których funkcja ogranicza się jedynie do szybszego podglądu sekcji pliku [Salomon 2007, s. 378, 631].

Bibliografia

Acharya, Tinku; Tsai, Ping-Sing (2005). JPEG2000 Standard for Image Compression: Concepts, Algorithms and VLSI Architectures. Hoboken: Wiley.

Allen, Robert B.; Hall, Catherine (2010). Automated Processing of Digitized Historical Newspapers beyond the Article Level: Sections and Regular Features. [W:] The Role of Digital Libraries in a Time of Global Change. Berlin: Springer, s. 270.

Anderson, Ian G. (2008). Necessary but Not Sufficient: Modelling Online Archive Development in the UK. „D-lib Magazine”, t. 14, nr 1/2 [online]. [Dostęp: 18.02. 2012]. Dostępny w World Wide Web: http://www.dlib.org/dlib/january08/anderson/01_anderson.html.

Bell, Timothy C. (1990). Text compression. Englewood Cliffs: Prentice Hall.

Bień, Janusz (2011). Efficient Search in Hidden Text of Large DjVu Documents. Advanced Language Technologies for Digital Libraries, t. 6699.

Boncelet, Charles (2006). Binary Image Compression. [W:] Barni, Mauro (red.) Document and Image Compression. Boca Raton: Taylor & Francis Ltd., s. 285–298.

Borgman, Christine L. (2000). From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World Digital Libraries and Electronic Publishing. Cambridge: MIT Press.

Bottou, Léon; Haffner, Patrick; Howard, Paul G.; Simard, Patrice; Bengio, Yoshua; LeCun, Yann (1998). High Quality Document Image Compression with DjVu. „Journal of Electronic Imaging”, t. 7, nr 3, s. 410–425.

Bottou, Léon; Haffner, Patrick; LeCun, Yann. Efficient Conversion of Digital Documents to Multi-layer Raster Formats [online]. [Dostęp: 18.02. 2012]. Dostępny w World Wide Web: http://djvu.org/docs/2001_efficient_compression.djvu.

Bottou, Léon; Haffner, Patrick; LeCun, Yann; Howard, Paul G.; Vincent, Pascal; Riemers, Bill (2000). DjVu, Un Système de Compression d'Images pour la Distribution Réticulaire de Documents Numérisés. [W:] Emptoz, Hubert; Vincent, Nicole (red.) CIFED'2000: colloque international francophone sur l'écrit et le document. Lousanne, s. 453–462.

Bottou, Léon; Pigeon, Steven (1998). Lossy Compression of Partially Masked Still Image. [W:] Proceedings of Data Compression Conferen-

ce. Los Alamitos: Institute of Electrical and Electronics Engineers, s. 528–538.

Breuel, Thomas M.; Janssen, William C.; Popat, Kris; Baird, Henry (2003). Reflowable Document Images. [W:] Antonacopoulos, Apostolos; Hu, Jianying (red.) Web Document Analysis: Challenges and Opportunities. New Jersey: World Scientific, s. 81–94.

Caminova [online]. [Dostęp: 2.04. 2012]. Dostępny w World Wide Web: <http://www.caminova.net/en/products/?src=products2.aspx>.

Feng, Guotong; Cheng, Hui; Bouman, Charles (2006). High Quality MRC Document Coding. [W:] IEEE Transactions on Image Processing. New York: Institute of Electrical and Electronics Engineers, s. 3152–3169.

Führ, Hartmut; Demaret, Laurent; Friedrich, Felix (2006). Beyond Wavelets: New Image Representation Paradigms. [W:] Barni, Mauro (red.) Document and Image Compression. Boca Raton: Taylor & Francis Ltd., s. 179–206.

Gonzalez, Rafael C.; Woods, Richard E. (2002). Digital Image Processing. Upper Saddle River: Prentice Hall.

Haffner, Patrick; Bottou, Léon; Howard, Paul G.; LeCun, Yann. DjVu: Analyzing and Compressing Scanned Documents for Internet Distribution [online]. [Dostęp: 5.03. 2012]. Dostępny w World Wide Web: http://djvu.org/docs/1999_analyze_scanned.djvu.

Haffner, Patrick; Bottou, Léon; Howard, Paul G.; Simard, Patrice; Bengio, Yoshua; LeCun, Yann

(1998). Browsing through High Quality Document Images with DjVu. [W:] IEEE International Forum on Research and Technology Advances in Digital Libraries. Santa Barbara, s. 309–318.

Haffner, Patrick; Bottou, Léon; LeCun, Yann; Vincent, Luc (2002). A General segmentation scheme for DjVu Document compression. [W:] Talbot, Hugues; Beare, Richard (red.) Mathematical morphology: proceedings of the VIth International Symposium on Mathematical Morphology (ISMM). Collingwood: CSIRO Publications, s. 17–36.

Haffner, Patrick; LeCun, Yann; Bottou, Léon; Howard, Paul G.; Vincent, Pascal; Riemers, Bill (2000). Color Documents on the Web with DjVu. [W:] Proceedings of International Conference on Image Processing. Piscataway: Institute of Electrical and Electronics Engineers.

Hansen, Christian; Nagy, James G.; O'Leary, Dianne P. (2006). Deblurring Images: Matrices, Spectra, and Filtering. Philadelphia: Society for Industrial and Applied Mathematics.

Holley, Rose (2009). How Good Can It Get? Analysing and Improving OCR Accuracy in Large Scale Historic Newspaper Digitisation Programs. „D-lib Magazine”, t. 15, nr 3/4 [online]. [Dostęp: 5.03. 2012] Dostępny w World Wide Web: <http://www.dlib.org/dlib/march09/holley/03holley.html>.

ITU-T.88: Series T: Terminals for Telematic Services [online]. [Dostęp: 5.03. 2012]. Dostępny w World Wide Web: <http://www.itu.int/rec/T-REC-T.88/en>.

- Jähne, Bernd (2005). *Digital Image Processing*. Berlin: Springer.
- Janiak, Małgorzata; Krakowska, Monika (2011). Ocena bibliotek cyfrowych: kryteria jakości. [W:] Mazurek, Cezary; Stroiński, Maciej; Węglarz, Jan (red.) *Polskie biblioteki cyfrowe 2010*. Poznań: Ośrodek Wydawnictw Naukowych, s. 101–108.
- Kalota, Tomasz; Raczyński, Rafał; Rękar, Paweł (2011). Przetwarzanie i OCR czasopism drukowanych gotykiem krok po kroku. [W:] Mazurek, Cezary; Stroiński, Maciej; Węglarz, Jan (red.) *Polskie biblioteki cyfrowe 2010*. Poznań: Ośrodek Wydawnictw Naukowych, s. 101–108.
- Kulovits, Hannes; Rauber, Andreas; Kugler, Anna; Brantl, Markus; Beinert, Tobias; Schoger, Astrid (2009). From TIFF to JPEG 2000? Preservation Planning at the Bavarian State Library Using a Collection of Digitized 16th Century Printings. „D-lib Magazine”, t. 15, nr 11/12 [online]. [Dostęp: 5.03. 2012]. Dostępny w World Wide Web: <http://www.dlib.org/dlib/november09/kulovits/11kulovits.html>.
- LeCun, Yann; Bottou, Léon; Erofeev, Andrei; Haffner, Patrick; Riemers, Bill (2000). DjVu Document Browsing with on-demand loading and rendering of Image Components. [W:] Beretta, Giordano B.; Schettini, Raimondo (red.) *Internet Imaging*, t. 2. Bellingham: Society of Photo-optical Instrumentation Engineers, s. 7–16.
- LeCun, Yann; Bottou, Léon; Haffner, Patrick; Howard, Paul G. (1998). DjVu: a Compression Method for Distributing Scanned Documents in Color over the Internet. [W:] *Sixth Color Imaging Conference: Color Science, Systems and Applications*. Scottsdale: Society for Imaging Science and Technology, s. 220–223.
- LeCun, Yann; Bottou, Léon; Haffner, Patrick; Triggs, Jeffery (2001). Overview of the DjVu Document Compression Technology. [W:] *Symposium on Document Image Understanding Technologies*. Columbia, s.119–122.
- Lin, Tony; Hao, Pengwei (2005). Compound Image Compression for Real-Time Computer Screen Image Transmission. [W:] *IEEE Transactions on Image Processing*, t. 14. New York: Institute of Electrical and Electronics Engineers, s. 993–1005.
- Lizardtech DjVu Reference [online]. [Dostęp: 8.04. 2012]. Dostępny w World Wide Web: <http://djvu.org/docs/DjVu3Spec.djvu>.
- Maheswari, D.; Radha, V. (2011). Enhanced Hybrid Compression Models for Compound Images. „Trends in Computer Science, Engineering and Information Technology”, t. 204, nr 1, s. 265–274.
- Mello, Carlos (2007). An Algorithm for Foreground-Background Separation in Low Quality Patrimonial Document Image. „Progress in Pattern Recognition, Image Analysis and Applications”, t. 4756, s. 911–920.
- Paradowski, Dariusz; Hnydka, Maciej; Kostynowicz, Jakub; Kwiatek, Filip; Leszyńska, Agnieszka; Lissowska, Agnieszka; Mazur, Paweł; i in. (2010). *Digitalizacja piśmiennictwa*. Warszawa: Biblioteka Narodowa.
- Płoszajski, Grzegorz; Schmidt, Kazimierz; Kalota, Tomasz; Paradowski, Dariusz (2008). Standar-

dy w procesie digitalizacji obiektów dziedzictwa kulturowego. Płoszajski, Grzegorz (red.). Warszawa: Biblioteka Główna Politechniki Warszawskiej.

Popat, Kris; Bloomberg, Dan S. Two-Stage Lossy/Lossless Compression of Grayscale Document Image. [W:] Goutias, John; Vincent, Luc M.; Bloomberg, Dan S. (red.) *Mathematical Morphology and Its Applications to Image and Signal Processing*, s. 361–370.

Powell, Tracy; Paynter, Gordon (2009). Going Grey? Comparing the OCR Accuracy Levels of Bitonal and Greyscale Images. „D-lib Magazine”, t. 15, nr 3/4.

De Queiroz, Ricardo L.; Buckley, Robert; Xu, Ming (1999). Mixed Raster Content (MRC): Model for Compound Image Compression. [W:] Aizawa, Kiyoharu; Stevenson, Robert L.; Zhang, Ya-Qin (red.) *Visual Communications and Image Processing '99*, s. 1106–1117.

De Queiroz, Ricardo L. (2006). *Compressing Compound Documents*. [W:] Barni, Mauro (red.) *Document and Image Compression*. Boca Raton: Taylor & Francis Ltd., s. 323–350.

Sachs, Jonathan. (2003) *Color Management* [online]. [Dostęp: 8.05. 2012]. Dostępny w World Wide Web: <http://ftp2.bmtmicro.com/dlc/Color%20Management.pdf>.

Salomon, David (2007). *Data Compression: the Complete Reference*. London: Springer.

Shafait, Faisal; Keyzers, Daniel; Breuel, Thomas M. (2006). *Performance Comparison of Six Algo-*

rithms for Page Segmentation. „Document Analysis Systems”, t. 3872, nr 7, s. 368–379.

Shapiro, Linda; Stockman, George (2001). *Computer Vision*. Upper Saddle River: Prentice Hall.

Tanner, Simon; Muñoz, Trevor; Ros, Pich Hemy (2009). Measuring Mass Text Digitization Quality and Usefulness: Lessons Learned from Assessing the OCR Accuracy of the British Library's 19th Century Online Newspaper Archive. „D-lib Magazine”, t. 15, nr 7/8 [online]. [Dostęp: 8.05. 2012]. Dostępny w World Wide Web: <http://www.dlib.org/dlib/july09/munoz/07munoz.html>.

Taubman, David (2006). *Advanced Modeling and Coding Techniques for Image Compression*. [W:] Barni, Mauro (red.) *Document and Image Compression*. Boca Raton: Taylor & Francis Ltd., s. 35–68.

Tompkins, D. A.; Kossentini, F. (1999). A Fast Segmentation Algorithm for Bi-Level Image Compression using JBIG2. [W:] *International Conference on Image Processing*, t. 1. Kobe: Institute of Electrical and Electronics Engineers, s. 224–228.

Turner, Thomas P.; Raskin, Howard (2004). *Investing in Staff: Hiring, Training, and Mentoring*. [W:] Barnes, Susan J. (red.) *Becoming a Digital Library*. New York: Marcel Dekker, s. 81–116.

Van der Knijff, Johan (2011). *JPEG 2000 for Long-term Preservation: JP2 as a Preservation Format*. D-lib Magazine, t. 17, nr 5/6 [online]. [Dostęp: 8.05. 2012]. Dostępny w World Wide Web: <http://www.dlib.org/dlib/may11/vanderknijff/05vanderknijff.html>.

Infrastruktura sprzętowa



Słowa kluczowe: infrastruktura serwerowa, serwer, macierz dyskowa, dLibra, komponenty systemu

Keywords: server infra-structure, server, disk array, dLibra, system components

212

Sprzętowa infrastruktura serwerowa obsługująca Jagiellońską Bibliotekę Cyfrową w chwili obecnej jest złożona z:

- dwóch serwerów HP ProLiant DL380 G6
- jednego serwera HP ProLiant DL360 G6
- macierzy dyskowej HP EVA4400
- biblioteki taśmowej HP MSL 6030

dLibra

Początkowo wszystkie komponenty systemu dLibra: wszystkie usługi, baza danych oraz Aplikacja Czytelnika (www) - zostały uruchomione na jednym serwerze fizycznym. Z upływem czasu stopniowo wydzielano następujące komponenty dLibry: Index Server, Aplikacja Czytelnika (www) oraz baza danych, kolejno

przenosząc je na dedykowane serwery. Obecnie rozmieszczenie komponentów systemu dLibra w Jagiellońskiej Bibliotece Cyfrowej jest następujące:

Serwer 1 – serwer fizyczny na HP ProLiant DL360 G6

Komponenty dLibra	Content Server, Metadata Server, Search Server, System Services, Event Manager, Profile Provider, User Server
Processor	1 x Intel Xeon E5520 2,26GHz
RAM	8GB
Przestrzeń na macierzy dyskowej	2 TB

Abstrakt: Artykuł opisuje infrastrukturę serwerową Jagiellońskiej Biblioteki Cyfrowej, oraz sposób realizacji rozdzielania komponentów składowych systemu dLibra pomiędzy zasoby sprzętowe tej infrastruktury.

Abstract: This article describes the server infrastructure of the Jagiellonian Digital Library, and the way of dLibra system components' distribution between available hardware resources.



213

Serwer 2 – serwer wirtualny na HP ProLiant DL380 G6 + VMware ESXi

Komponenty dLibra	Index Server
Procesor	1 core x 2,26GHz Intel Xeon E5520
RAM	4GB
Przestrzeń na macierzy dyskowej	50 GB

Serwer 3 – serwer wirtualny na HP ProLiant DL380 G6 + VMware ESXi

Komponenty dLibra	Serwer bazy danych PostgreSQL 8
Procesor	4 core x 2,26GHz Intel Xeon E5520
RAM	8GB
Przestrzeń na macierzy dyskowej	20 GB

Serwer 4 – serwer wirtualny na HP ProLiant DL380 G6 + VMware ESXi

Komponenty dLibra	Aplikacja Czytnika (www)
Procesor	2 core x 2,26GHz Intel Xeon E5520
RAM	4GB
Przestrzeń na macierzy dyskowej	20 GB

Serwer 5 – serwer wirtualny na HP ProLiant DL380 G6 + VMware ESXi

Komponenty dLibra	Aplikacja Czytnika (www) testowy
Procesor	1 core x 2,26GHz Intel Xeon E5520
RAM	3GB
Przestrzeń na macierzy dyskowej	15 GB

SOJBC

SOJBC obejmuje serwer plików przeznaczony do składowania i obróbki skanów w postaci plików tiff oraz aplikacje www do zarządzania przepływem, obróbką i archiwizacją skanów.

Serwer SOJBC – serwer fizyczny na HP ProLiant DL380 G6

Komponenty	SOJBC
Procesor	2 x Intel Xeon E5520 2,26GHz
RAM	24GB
Przestrzeń na macierzy dyskowej	28,5 TB

Monitoring serwerów

Każdy z serwerów jest objęty monitoringiem wykorzystania zasobów sprzętowych serwera (CPU, RAM, network traffic, disk I/O itp.), na podstawie którego w razie potrzeby dokonywane są odpowiednie korekty przydziału zasobów lub rozdzielanie usług dLibry.

HP ProLiant DL360 G6
Linux

Serwer 1	
Komponenty dLibry	Content Server Metadata Server Search Server System Services Event Manager Profile Provider User Server
Repozytorium plików dla systemu dLibra	pliki DjVu, PDF, HTML, pliki dźwiękowe i inne

HP ProLiant DL360 G6
Vmware ESXi

Serwer 2	
Komponenty dLibry	Index Server
Serwer 3	
Komponenty dLibry	Serwe bazy danych
Serwer 4	
Komponenty dLibry	Aplikacja czytelnika dLibry
Serwer 5	
Komponenty dLibry	Aplikacja czytelnika dLibry

II. 1.

II. 1. Rozmieszczenie komponentów systemu dLibra w Jagiellońskiej Bibliotece Cyfrowej.

Archiwizacja oryginałów i kopii dzieł zabezpieczonych

A0



Archiwizacja cyfrowa



Słowa kluczowe: Jagiellońska Biblioteka Cyfrowa, System Obsługi Jagiellońskiej Biblioteki Cyfrowej, archiwizacja cyfrowa, format archiwalny, taśma magnetyczna, kod paskowy, kopia użytkowa, kopia archiwalna, etykieta

Keywords: Jagiellonian Digital Library, Jagiellonian Digital Library Workflow System, digital archive, archival format, magnetic tape, barcode, functional copy, archival copy, label

Oprócz wersji prezentacyjnych, produktem projektu są także pliki bezpośrednio po skanowaniu. Wraz z oryginałami oraz mikrofilmami stanowią one przedmiot archiwizacji. Pozwoli to uniknąć w przyszłości konieczności powtarzania pracy, w przypadku zaistnienia potrzeby wprowadzenia poprawek bądź zmiany formatu prezentacyjnego. Ponadto, stan wielu oryginałów (zwłaszcza dziewiętnastowiecznych czasopism) zwyczajnie nie pozwala na ich powtórne skanowanie.

Archiwizacji poddane są surowe skany – czyli pliki w formacie TIFF bez żadnej obróbki po uzyskaniu ze skanera. Zapewnia to po pierwsze zgodność archiwum z wymogami archiwizacji.

Trzeba tu oczywiście pamiętać, że proces skanowania kończy się dopiero po zaakceptowaniu plików przez osobę kontrolującą – zarówno pod względem jakości skanów, jak i ich kompletności. Taki materiał jest więc dobrą podstawą do wiernej kopii archiwalnej.

Ponadto wybór nieprzetworzonych TIFF-ów pozwala w dużym stopniu uniezależnić się od rozwoju technik informatycznych: z jednej strony standardowa bitmapa zawsze powinna być dobrym materiałem na konwersję do dowolnego formatu pliku, a z drugiej dobrze będzie się poddawać technikom kompresji czy korekcji obrazu.

Minusem tego rozwiązania jest ogromny rozmiar plików i wynikające z niego

wymagania przestrzeni dyskowej. Na szczęście pliki archiwalne nie muszą być dostępne w trybie ciągłym, więc do ich przechowywania można użyć taśm magnetycznych.

Proces archiwizacji cyfrowej rozpoczyna się w momencie zatwierdzenia poprawności skanów w procesie kontroli (status „SprawdzoneOK”). Panel „Archiwizacja cyfrowa” w Systemie Obsługi Jagiellońskiej Biblioteki Cyfrowej pozwala monitorować sumaryczny rozmiar skanów gotowych do archiwizacji. Gdy osiągnię on maksymalny rozmiar pojedynczej taśmy (ok. 750GB), można rozpocząć nagrywanie.

Najpierw za pomocą systemu przydzielamy poszczególne woluminy do nowotworzonej taśmy. Podajemy także kod paskowy taśmy, na którą wybrane pliki zostaną nagrane (a dokładnie dwóch taśm, ponieważ każde nagranie jest przeprowadzane w dwóch egzemplarzach). Taki kod paskowy pozwoli połączyć fizyczną taśmę

Abstrakt: Powstałe w procesie skanowania surowe skany w formacie TIFF są w JBC jednym z przedmiotów archiwizacji. Artykuł przedstawia szczegółowo zastosowane metody ich długoterminowego przechowywania (przy pomocy Systemu Obsługi Jagiellońskiej Biblioteki Cyfrowej), prezentuje również dane ilościowe dotyczące procesu archiwizacji.

Abstract: Raw scans in TIFF format resulting from the digitization process are a subject of archiving procedure in the Jagiellonian Digital Library. The paper details the methods used for their long-time preservation (using the Jagiellonian Digital Library Workflow System) and presents some quantitative figures pertaining to the process.



ID	Nazwa pliku	Format	Wielkość	Data	Status
7716	Temple systemy : 3 MAGDY : Op. 30	MP3	269.47 MB	December 14, 2011 12:34 PM	Skonieczony
1022	Nowy i Stary Kalendarz, Świąt rocznych y Diegów Świąt niedzieli, y wzorem Głogów, y z Apokalipsy. Ma. S.A. Pawła M. DC. LII [...]	MP3	1.19 GB	December 14, 2011 12:34 PM	Skonieczony
7717	Sana fagon : Polska-polski : Op. 31	MP3	186.55 MB	December 14, 2011 12:35 PM	Skonieczony
1023	Nowy i Stary Kalendarz, Świąt Kaczych y Diegów Kaczych na Kol. Po Zbawieniu Narodzeniu Pana y Zbawienia naszego Jezusa Chrystusa M. DC. LII. [...]	MP3	961.56 MB	December 14, 2011 12:35 PM	Skonieczony
4902	De antiquissima nomina & familie Thurbon[um] origine, & antiquitate [...] Thurboni [...] emendata Paragymis [...]	MP3	457.61 MB	December 14, 2011 12:35 PM	Skonieczony
1024	Nowy i Stary Kalendarz Świąt Kaczych y Diegów Kaczych na Kol. Po Zbawieniu Narodzeniu Pana y Zbawienia naszego Jezusa Chrystusa M. DC. LII. [...]	MP3	979.46 MB	December 14, 2011 12:35 PM	Skonieczony
1029	Albumo Conno Orl. & Saperite, Sub Ecuma	MP3	1.09 GB	December 14, 2011 12:35 PM	Skonieczony

II. 1.

z jej numerem systemowym.

Po zatwierdzeniu system tworzy na macierzy „katalog eksportowy”. Zawiera on softlinki do plików, które mają trafić na daną taśmę, wraz z plikiem opisującym jej zawartość. Zawartość tego katalogu należy teraz nagrać na taśmę (a właściwie obie taśmy) i zweryfikować poprawność zapisu. Przy objętościach rzędu 750GB proces nagrywania i weryfikacji pojedynczej taśmy trwa prawie 8 godzin – dlatego należało te czynności w miarę możliwości zautomatyzować. Jest również korzystne dla zmniejszenia obciążenia macierzy, aby wykonywały się one w miarę możliwości w czasie wolnym od pracy (w nocy i w niedzielę). Konieczność fizycznego umieszczenia taśmy w napędzie streamera byłaby dużym pro-

blemem, dlatego zastosowano bibliotekę taśmową. Pozwala ona na załadowanie jednorazowo 30 taśm magnetycznych, a następnie zdalne przemieszczanie ich między napędem streamera a kasetami. Przy jej użyciu, znając kody paskowe taśm, można cały proces nagrywania i weryfikacji obsłużyć odpowiednim skryptem.



II. 2.

II. 1. Panel archiwizacji cyfrowej SOJBC: lista woluminów.

II. 2. Biblioteka taśmowa.

pisania” taśmy w SOJBC. Wtedy zapisane pliki zostają usunięte z macierzy, a ich status uaktualniony w bazie systemu. W tym momencie dla obydwóch taśm drukuje się etykiety na pudełko, która podaje numer systemowy taśmy oraz w skrócie podsumowuje jej zawartość. Po wyjęciu taśm z biblioteki taśmowej trafiają one do takich pudełek.



II. 4.

134A	<p>Taśma nr 134, kopia A (ZS0101L4), JBC Nagrana: 2012-10-08 09:44:58</p> <p>Zawiera 6051 plików z 27 woluminów: Rozmaitości : tygodniowy dodatek „Dziennika Lwowskiego”: 1928 (1060), Słowo Polskie: 1933/I (668), Głos Narodu: 1923/II (445), 1925/I (448), 1925/II (449), 1926/I (450), Goniec obozowy, Wiadomości dla internowanych: 1940-1944 (926), 1940 (9493), Polska.: 1934 (1038), 1935 (2638), 1937 (2640), 1938 (2641), Ilustrowany Express Codzienny: 1937/I (965), Głogowlanka.: 1921 (942), Ilustrowany Kurjer Czesłowski, Ilustrowany Kurjer Czesłowski: 1932 (967), 1933 (2625), Kraj Głogowski.: 1921 (979), Gazeta : organ Samorządu Państwowego Gimnazjum 358.: 1933-1934 (939), 1935-1936 (2599), 1937 (2600), 1938 (2601), 1939 (2602), W Młodych Oczach : kolumna poezji i sztuki młodego Wilna : dodatek do „Sprawy Otwarte”: 1938 (2646), 1939 (2647), Goniec Pokucki : czasopismo poświęcone polityce i sprawom społecznym Pokucia i okolicy: 1907 (1102), 1908 (2654), Orędownik na Miasto Królewską Hutę i Okolicę : organ miejscowy poświęcony sprawom ludu polskiego.: 1921 (1028).</p>
134B	<p>Taśma nr 134, kopia B (ZS0102L4), JBC Nagrana: 2012-10-08 09:44:58</p> <p>Zawiera 6051 plików z 27 woluminów: Rozmaitości : tygodniowy dodatek „Dziennika Lwowskiego”: 1928 (1060), Słowo Polskie: 1933/I (668), Głos Narodu: 1923/II (445), 1925/I (448), 1925/II (449), 1926/I (450), Goniec obozowy, Wiadomości dla internowanych: 1940-1944 (926), 1940 (9493), Polska.: 1934 (1038), 1935 (2638), 1937 (2640), 1938 (2641), Ilustrowany Express Codzienny: 1937/I (965), Głogowlanka.: 1921 (942), Ilustrowany Kurjer Czesłowski, Ilustrowany Kurjer Czesłowski: 1932 (967), 1933 (2625), Kraj Głogowski.: 1921 (979), Gazeta : organ Samorządu Państwowego Gimnazjum 358.: 1933-1934 (939), 1935-1936 (2599), 1937 (2600), 1938 (2601), 1939 (2602), W Młodych Oczach : kolumna poezji i sztuki młodego Wilna : dodatek do „Sprawy Otwarte”: 1938 (2646), 1939 (2647), Goniec Pokucki : czasopismo poświęcone polityce i sprawom społecznym Pokucia i okolicy: 1907 (1102), 1908 (2654), Orędownik na Miasto Królewską Hutę i Okolicę : organ miejscowy poświęcony sprawom ludu polskiego.: 1921 (1028).</p>

II. 3.



II. 5.

II. 6.

ID	Nazwa 1. taśmy	Nazwa 2. taśmy	L. woluminów	L. plików	Zapisano?	Operacje
21	80V200L4	80V201L4	48	26406	✓	Lista woluminów Zapisz taśmę Etykieta na taśmę
24	80V202L4	80V203L4	52	15722	✓	Lista woluminów Zapisz taśmę Etykieta na taśmę
25	80V204L4	80V205L4	31	14954	✓	Lista woluminów Zapisz taśmę Etykieta na taśmę
26	80V206L4	80V207L4	29	31895	✓	Lista woluminów Zapisz taśmę Etykieta na taśmę
27	80V208L4	80V209L4	17	14951	✓	Lista woluminów Zapisz taśmę Etykieta na taśmę
28	80V210L4	80V211L4	7	10523	✓	Lista woluminów Zapisz taśmę Etykieta na taśmę
29	80V212L4	80V213L4	9	8027	✓	Lista woluminów Zapisz taśmę Etykieta na taśmę

II. 3. Etykieta na taśmę magnetyczną – kopia archiwalna i użytkowa.

II. 4. Taśma magnetyczna przed spakowaniem.

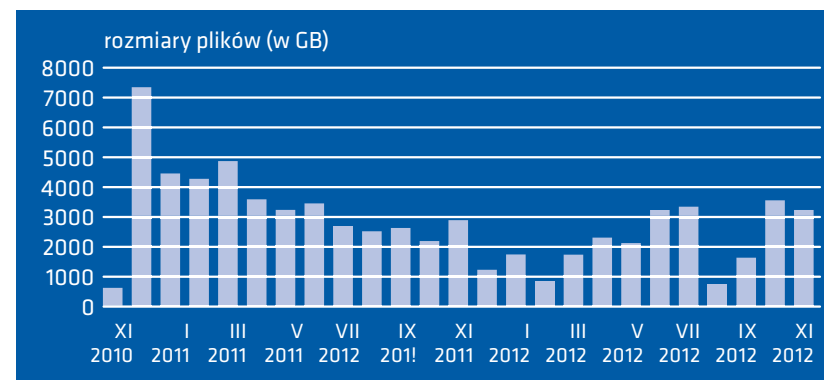
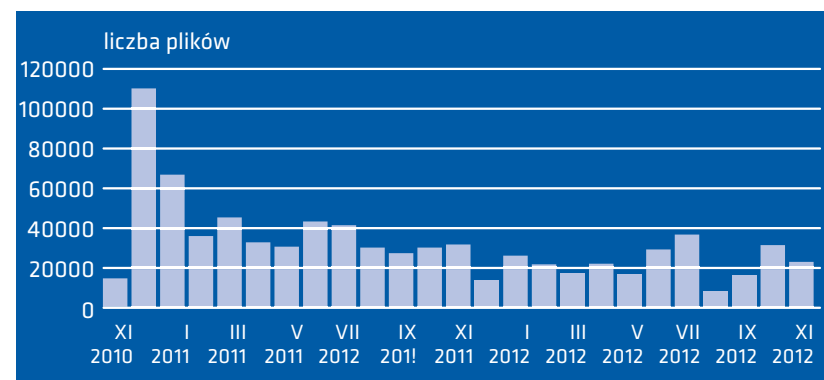
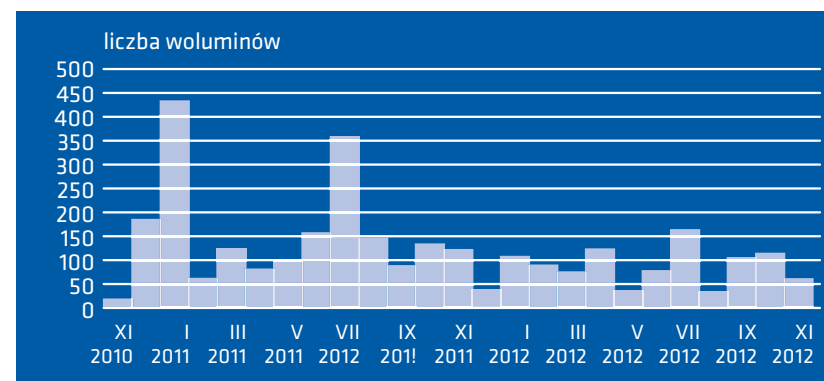
II. 5. Specjalistyczna szafa do przechowywania taśm magnetycznych.

II. 6. Panel archiwizacji cyfrowej SOJBC: lista taśm.

Jak już wspomniano, w procesie archiwizacji każda taśma nagrywana jest w dwóch kopiach, zgodnie z zasadami archiwizacji. Pierwsza kopia, tzw. „A”, trafia do osobnej bezpiecznej lokalizacji. Kopia druga, tzw. „B”, to egzemplarz użytkowy, przechowywany w dedykowanej szafie w Oddziale Komputeryzacji.

Kopia użytkowa jest używana w przypadkach konieczności odtworzenia pliku. Zdarza się to, gdy należy wytworzyć nową wersję pliku prezentacyjnego – np., gdy czujne oczy czytelników dostrzegą jakiś błąd w prezentowanej publikacji. Po znalezieniu w systemie numeru taśmy, na której zawarte są pliki, należy odczytać jej zawartość na macierz dyskową. Następnie dedykowany skrypt SOJBC pozwala na odtworzenie katalogu „eksportowego” ze skanami danego woluminu – dokładnie w formacie, w jakim otrzymują pliki osoby kontrolujące skany.

Na zakończenie prezentujemy wykresy miesięcznego podsumowania opracowanych w projekcie plików wg liczby woluminów, liczby plików oraz ich rozmiaru.



Analogowa kopia zabezpieczająca w postaci mikrofilmu



Słowa kluczowe: digitalizacja, kopia zabezpieczająca, mikrofilm, kopia analogowa, archiwizacja

Keywords: promotion, Jagiellonian Digital Library, conference, promotional materials, press releases

226



Aby spełniać wszelkie standardy zabezpieczenia i archiwizacji dokumentów w projekcie JBC, zdecydowano o tworzeniu kopii analogowej skanowanych materiałów w postaci mikrofilmu. Kopia mikrofilmowa pozwala na odzyskanie obrazu dokumentu zarówno w jego postaci drukowanej jak i cyfrowej [Skinner 2010]. Z racji uszkodzeń nośnika jakim jest papier, jak i wciąż nieokreślonej trwałości dokumentu cyfrowego, forma pośrednia jaką jest mikrofilm idealnie wpisująca się jako rozwiązanie kompletnie chroniące zawartość digitalizowanych dokumentów [Kelly 2005]. Mikrofilmy stosowane w bibliotekach od lat 30-tych [Born 1960] wciąż jeszcze są w nich powszechne i choć powoli wypierane przez nośniki cyfrowe, jako nośnik archiwizacyjny mają nieocenioną wartość. Z jednej strony jest to ich trwałość, z drugiej łatwość konwersji zawartego na nich obrazu z formy analogowej na cyfrową.

Aby zapewnić w pełni zautomatyzowa-

ny proces pozyskiwania mikrofilmów na masową skalę, w ramach projektu został zakupiony ploter mikrofilmowy przedstawiony na il. 1.

Urządzenie to pozwala na przeniesienie na taśmę mikrofilmową obrazów cyfrowych wykonanych z oryginałów w wybranym przez użytkownika formacie. W Bibliotece Jagiellońskiej do stworzenia mikrofilmów wykorzystano obrazy cyfrowe zachowane w formacie JP2.

Proces przygotowania obrazu cyfrowego obejmował prostowanie i kadrowanie stron zdigitalizowanego dokumentu. Po zakończeniu tego etapu następowała konwersja plików do formatu JP2, charakteryzującego się o wiele mniejszymi rozmiarami w stosunku do plików w formacie TIFF [Lowe 2008]. Jakość obrazów okazała się wystarczająco dobra do wytworzenia mikrofilmu, a dzięki kadrowaniu i prostowaniu poszczególnych stron, mikrofilmy uzyskały wysoką jakość. W przypadku cyfrowej katastrofy odzyskanie obrazów za pomocą automatycznych skanerów do mikrofilmów nie powinno przedstawiać większych trudności.

Dla wszystkich wytworzonych mikrofilmów zostały sporządzone rekordy zasobu i rekordy egzemplarza, dołączone do rekordów bibliograficznych [Fitz-Simmons 2008] dla poszczególnych tytułów dokumentów digitalizowanych w ramach projektu Jagiellońska

Biblioteka Cyfrowa. W ten sposób pełna informacja zarówno o oryginale jak i dokumencie wtórnym w postaci mikrofilmu jest zebrana w jednym miejscu w KKZBUJ. Na il. 2 została przedstawiona część zasobu zachowanego w postaci kopii mikrofilmowej.

Na il. 3 przedstawiono opisy poszczególnych egzemplarzy mikrofilmów.

Należy zaznaczyć, że każdy mikrofilm

Abstrakt: W artykule przedstawiono ideę dodatkowego zabezpieczenia czasopism drukowanych w postaci analogowej. Opisany został proces powstawania mikrofilmu z uzyskanej w wyniku digitalizacji kopii cyfrowej. Przedstawiono sposób dobierania ilości obrazów cyfrowych umieszczanych na taśmie mikrofilmowej w zależności od objętości poszczególnych woluminów czasopism oraz innych materiałów bibliotecznych. Zwrócono uwagę na sposób katalogowania mikrofilmów oraz ich składowania w magazynie Oddziału Dokumentów Audiovizualnych.

Abstract: The concept of creating an additional archival copy of printed periodicals on analogue medium is presented briefly. The process of producing a microfilm out of a digital copy is described, particularly determining the number of images it should include in dependence on the size of printed periodical volumes and other library items. Ways of cataloging microfilms in OPAC and their arrangement on the stacks in the storeroom of Audiovisual Materials Department are also presented.



227

II. 2.

ZOBACZ -->

EGZEMPLARZE zasobu(218)

Opis zasobu
Rekord zasobu MARC

Uwagi
Sygnatura
Lokalizacja
Adnotacja
Uwaga dot. reprodukcji
Wykaz woluminów

Słownik Jagielloński- 182 egzemplarzy negatywowe
63 A MF. 020001
63 - Mag. Dok. Audiowiz.
2 sygn. 63 501. III - stan w dniu 31 XII 2013.
Negatym. Frajón : Frajón JBC, 2010-2013.

K. [1] 1811 nr 1 (2 kwiec.)
K. [2] 1812 nr 1 (3 stycz.)-104 (29 grudz.) Uwaga: br. nr 32-33, 52
K. [3] 1813 nr 1 (3 stycz.)-105 (31 grudz.) + doc. do nr 69 (manifest ces.)
Uwaga: br. nr 24, 63, 67; w nr 64 br. s. 541-542; w nr 92 br. s. 793-792
K. [4] 1814 nr 1 (4 stycz.)-102 (23 grudz.) + doc. do nr 27 (Inst. Edukacji), 49
(Do obrębki Australi) Uwaga: br. nr 29, 30, 33-34, 46, 51, 99-98, 92-94,
99-100; w nr 23 br. s. 199-200 i 205-206, w nr 45 br. s. 411-412, w nr 61 br.
s. 557-556
K. [5] 1815 nr 1 (3 stycz.)-104 (29 grudz.)+ wól. o presumencie w nr 46-47,
48, 50 Uwaga: br. nr 3; w nr 34 br. s. 327-310, w nr 130 br. s. 905-906
K. [6] 1.1-2 1814 nr 1 (1 stycz.)-209 (20 grudz.) Uwaga: br. nr 90-91, 94, 97;
w nr 13 br. s. 61-62, w nr 98 br. s. 427-428; w nr 148 br. s. 825-826, w nr 173
br. s. 865-866
K. [7] 1.1-2 1817 nr 1 (1 stycz.)-209 (31 grudz.) Uwaga: br. nr 67-70, 107-110,
124, 127-128, 129-134, 131-132
K. [8] 1.1-2 1818 nr 1 (3 stycz.)-108 (28 grudz.) Uwaga: br. nr 15-16, 26-28,
48-49, 58-59, 127-128, 143-144, 154-155, 164-165; w nr 102 br. s. 411-412
K. [9] 1.1-2 1819 nr 1 (4 stycz.)-149 (31 grudz.) Uwaga: br. nr 44, 47, 62-63,
68, 74, 77-78, 145-146
K. [10] 1.1-2 1820 nr 1 (3 stycz.)-148 (29 grudz.) Uwaga: br. nr 12, 50, 69-69,
74-75
K. [11] 1.1-2 1821 nr 1 (3 stycz.)-150 (31 grudz.) Uwaga: br. nr 41, 45-46, 55,
61-64, 91, 142
K. [12] 1.1-2 1822 nr 1 (2 stycz.)-151 (30 grudz.) Uwaga: br. nr 46, 102-105,
110-117
K. [13] 1.1-2 1823 nr 1 (3 stycz.)-149 (31 grudz.) Uwaga: br. nr 21-45, 49-52,
106
K. [14] 1.1-2 1824 nr 1 (2 stycz.)-150 (31 grudz.)
K. [15] 1.1-2 1825 nr 1 (3 stycz.)-149 (30 grudz.) Uwaga: br. 55, 76
K. [16] 1.1-2 1826 nr 1 (2 stycz.)-147 (29 grudz.)

w obrębie tytułu został opatrzony unikatową sygnaturą. W projekcie JBC czasopisma były skanowane w całości, więc nadano jedną sygnaturę dla całego czasopisma, a w obrębie sygnatury dalszy podział dokonany został na podstawie lat i – na kolejnym poziomie – numerów. Natomiast w obrębie całego zbioru zachowano system *numerus currens*, a więc mikrofilmy z czasopismami przeplatają się z mikrofilmami, na których utrwalało się książki czy rękopisy.

Dla mikrofilmów wykonanych w ramach projektu wyznaczony został odrębny ciąg w ogólnej sygnaturze A MF. Wraz z wyznaczeniem tego ciągu dokonano się przejście do kolejnego etapu w polityce zabezpieczania zbiorów w Bibliotece

II. 3.

Wykaz egzemplarzy zasobu

Szklano - U3 - Tytuł: Gazeta Lwowska.
Tytuł: Gazeta Lwowska.
Adres wydawniczy: Lwów : [i. n.], 1811-1939.
Multimedia: <http://br.jbc.uj.edu.pl/dlibra/publication/132129> Jagiellońska Biblioteka Cyfrowa

Wypozyczeń	0	*	0
Dostępny	1	Zamówionych	0
Do wstępu	0	Inne	217
Łączna egzemplarzy:	218		
Kaznaczej kłótni:	0	Zamówień łączna:	0
Formy 1 - 10 / 218			

Zaznacz, jeśli chcesz zachować opis w Słowniku

Zarezerwowanie egzemplarza
Lokalizacja: 83 - Mag. Dok. Audiowiz.
Sygnatura: 63 A MF. 020001, 1811
Wolumin: 1811, nr 1
Kategoria opregrazera: 009 kopia mikrofilm UDOST. W CZYTELNI
Status udostepnienia: Nie wypozyca sie

Zaznacz, jeśli chcesz zachować opis w Słowniku

Zarezerwowanie egzemplarza
Lokalizacja: 83 - Mag. Dok. Audiowiz.
Sygnatura: 63 A MF. 020001, 1812
Wolumin: 1812, nr 1-104
Kategoria opregrazera: 009 kopia mikrofilm UDOST. W CZYTELNI
Status udostepnienia: Nie wypozyca sie

Zaznacz, jeśli chcesz zachować opis w Słowniku

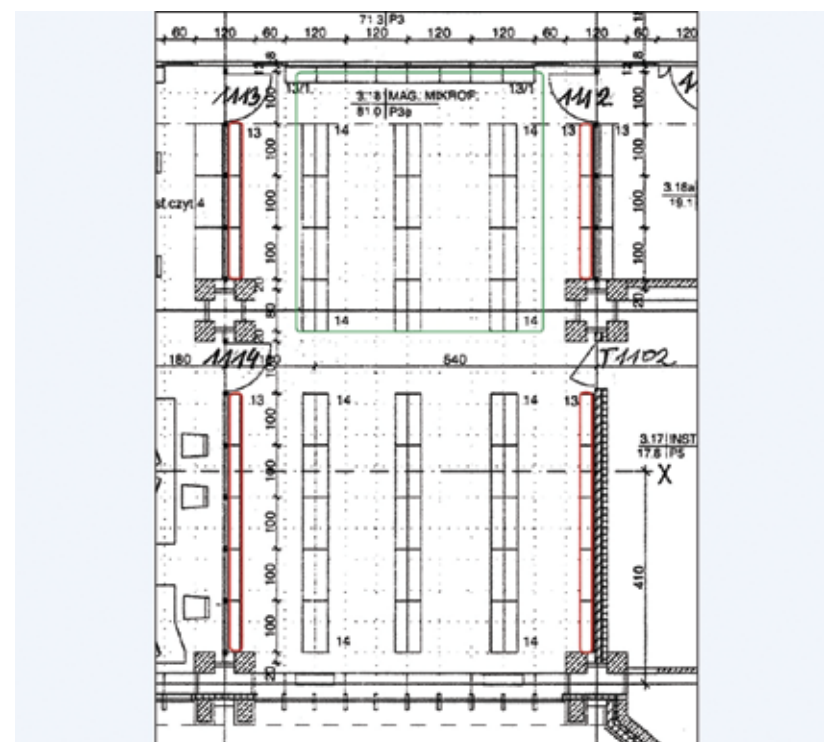
Zarezerwowanie egzemplarza
Lokalizacja: 83 - Mag. Dok. Audiowiz.
Sygnatura: 63 A MF. 020001, 1813
Wolumin: 1813, nr 1-105
Kategoria opregrazera: 009 kopia mikrofilm UDOST. W CZYTELNI
Status udostepnienia: Nie wypozyca sie

Jagiellońskiej. Od tego momentu dla prawie wszystkich zbiorów, jako podstawowe wykonywane są kopie cyfrowe, a kopie mikrofilmowe jako kopie zabezpieczające tylko dla uzupełnienia braków wśród mikrofilmowych kopii dla czasopism. Oczywiście dla najcenniejszych zbiorów oraz w miarę możliwości wykonywane są i będą dodatkowe zabezpieczające kopie mikrofilmowe, ale wykonywane na podstawie uzyskanych, jako podstawowe, kopii cyfrowych.

Według zaleceń konserwatorskich mikrofilmy powinny być przechowywane w pudełkach z tektury beżkowej [Kesse 2003]. Dlatego w projekcie dla wszystkich mikrofilmów zostały przygotowane odpowiednie pudełka, opatrzone

II. 2. Rekord zasobu dla kopii mikrofilmowych w KKZBUJ.

II. 3. Egzemplarze kopii mikrofilmowych w KKZBUJ.



II. 4.

pomarańczową nalepką z sygnaturą oraz kodem kreskowym, zawierającym numer egzemplarza w bazie KKZBUJ.

Mając na uwadze, że mikrofilmy powinny być przewijane co 3 lata [Gibson 1983] sposób ich ustawienia na regałach powinien umożliwiać rotację całego zbioru oraz ewentualne uzupełnienia braków. Był to jeden z powodów, dla których w ramach projektu wymieniono w magazynie Oddziału Dokumentów Audiowizualnych regały stacjonarne na jezdne. Kolejnym powodem racjonalizacji po-

wierzchni magazynowej była spodziewana duża liczba mikrofilmów, uzyskanych jako dodatkowe kopie zabezpieczające w wyniku procesu digitalizacji zasobów bibliotecznych. Dlatego również wąskie regały stacjonarne znajdujące się przy ścianach wymieniono na szersze, przy okazji wykorzystując całą niezagospodarowaną powierzchnię powyżej wcześniej zainstalowanych regałów. Zmiany dokonane w magazynie przedstawione zostały na il. 4.

W ramach prac zainstalowano regały

II. 4. Schemat magazynu z zaznaczonymi regałami i powierzchnią na której zostały zamontowane nowe regały.

przesuwne i stacjonarne o łącznej długości 1642,9 m.b. półek użytkowych, w tym:

- 11 regałów przesuwnych:
2250-2280 mm/5x1000 mm/2x200 mm,
- 1 regał statyczny:
4000 mm/3x900 mm/400 mm,
- 1 regał statyczny:
4150 mm/3x900 mm/400 mm,
- 1 regał statyczny:
4150 mm/4x1200 mm/400 mm,
- 1 regał statyczny:
4150 mm /3x1200+1000 mm/400 mm.

Natomiast elementy istniejących wcześniej regałów stacjonarnych zostały wykorzystane do budowy nowych regałów jezdnych. W ten sposób przestrzeń magazynowa na tej samej powierzchni zwiększyła się ponad 5-krotnie.

Dzięki projektowi Jagiellońska Biblioteka Cyfrowa został otwarty nowy rozdział w polityce zabezpieczenia zbiorów Biblioteki Jagiellońskiej, który w sposób kompleksowy i przy uwzględnieniu standardów światowych pozwala bezpieczniej przechowywać oryginały z jednej strony, a wytwarzać dokumenty wtórne służące zabezpieczeniu i prezentacji szerokiej publiczności, z drugiej.

Bibliografia

Born, Lester K. (1960). History of Microform Activity [online]; [dostęp 05.10.2012]. Dostępny w World Wide Web: https://www.ideals.illinois.edu/bitstream/handle/2142/5884/library-trendsv8i3c_opt.pdf?sequence=1.

FitzSimmons, Anna (2008). Preparing Our Past for the Future: Newspaper Preservation and Digitization at the University of Illinois [online]; [dostęp 17.09.2012]. Dostępny w World Wide Web: <https://www.ideals.illinois.edu/handle/2142/14431>.

Gibson, Gerald D. (1983). Preservation of Non-paper Materials: Present and Future Research and Development in the Preservation of Film, Sound Recordings, Tapes, Computer Records, and Other Nonpaper Materials [online]; [dostęp 17.09.2012]. Dostępny w World Wide Web: <http://hdl.handle.net/2142/452>.

Kelly, Katherine (2005). Microfilm-to-Digital Reformatting and Hybrid Microfilm and Digital Projects: an Overview and Discussion (Web version) [online]; [dostęp 17.09.2012]. Dostępny w World Wide Web: <https://pacer.ischool.utexas.edu/handle/2081/1415>.

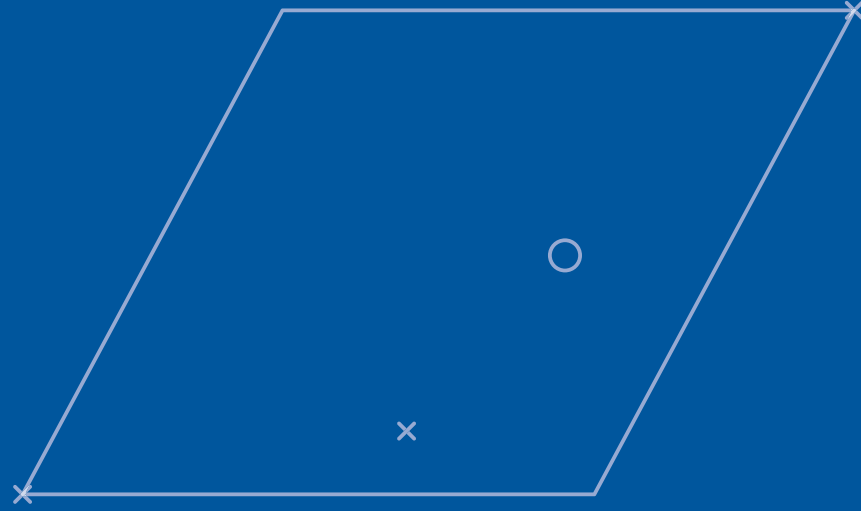
Kesse, Erich (2003). Library collections preservation and digitisation consultancy for the University of Botswana [online]; [dostęp 15.09.2012]. Dostępny w World Wide Web: <http://ufdc.ufl.edu/UF00076777/00001>.

Lowe, David; Bennett, Michael J. (2008). Digital Project Staff Survey of JPEG2000 Implementation in Libraries [online]; [dostęp 9.09.2012]. Dostęp-

ny w World Wide Web: http://digitalcommons.uconn.edu/libr_pubs/16/.

Skinner, Julia (2010). Learning From the Past: Digitization and Information Loss [online]; [dostęp 17.09.2012]. Dostępny w World Wide Web: <http://ir.uiowa.edu/bsides/7/>.

Promocja projektu



PP

Promocja projektu



Słowa kluczowe: promocja, Jagiellońska Biblioteka Cyfrowa, konferencja, materiały promocyjne, artykuły prasowe

Keywords: promotion, Jagiellonian Digital Library, conference, promotional materials, press releases

W trakcie realizacji projektów ze środków Programu Operacyjnego Infrastruktura i Środowisko obowiązują beneficjentów bardzo szczegółowe wytyczne i zasady co do ich promocji. Celem działań informacyjnych i promocyjnych prowadzonych przez beneficjentów jest zwiększenie świadomości opinii publicznej na temat udziału środków Unii Europejskiej w projektach realizowanych w ramach POIiŚ, a także uświadomienie ogółowi społeczeństwa korzyści wynika-

jących z realizacji projektów w obrębie POIiŚ, istotnych dla wzrostu konkurencyjności kraju i regionów.

Działania informacyjno-promocyjne dotyczące danego przedsięwzięcia, tutaj Jagiellońskiej Biblioteki Cyfrowej, stanowią jedno z zadań realizowanych w ramach projektu. Do obowiązków beneficjenta należy stosowanie systemu identyfikacji wizualnej POIiŚ, a więc logotypu Narodowej Strategii Spójności oraz emblematu Unii Europejskiej z odwołaniem słownym do Unii Europejskiej i Europejskiego Funduszu Rozwoju Regionalnego. Unijne logotypy muszą się znajdować na wszystkich dokumentach związanych z realizacją projektu oraz na tablicach informacyjnych, które beneficjent obowiązany jest umieścić w widocznych miejscach. Na wszystkich środkach trwałych zakupionych w ramach projektu musi znajdować się informacja o współfinansowaniu tychże zakupów przez Unię Europejską ze środków EFRR w ramach POIiŚ. Brak takich informacji może skut-

kować uznaniem poniesionych kosztów za niekwalifikowane.

Podczas realizacji projektu Jagiellońska Biblioteka Cyfrowa podjęto szereg działań informacyjno-promocyjnych. Wśród nich należy wymienić artykuły prasowe, informacje w środkach masowego przekazu, akcje promujące Jagiellońską Bibliotekę Cyfrową, prezentacje projektu na krajowych i międzynarodowych konferencjach naukowych, dotyczących zagadnień zachowania i digitalizacji dziedzictwa kulturowego, rozprowadzanie materiałów promocyjnych, mających utrwalić w świadomości potencjalnych użytkowników istnienie Jagiellońskiej Biblioteki Cyfrowej i jej współfinansowanie ze środków Unii Europejskiej.

Aby ułatwić czytelnikom korzystanie z produktów projektu, zakupiono i ustawiono w holu Biblioteki Jagiellońskiej infomat, tj. kiosk informacyjny, za pomocą którego można uzyskać dostęp do platformy Jagiellońskiej Biblioteki Cyfrowej (www.jbc.bj.uj.edu.pl), a także otrzymać wszelkie informacje o projekcie. Na potrzeby kreowania identyfikacji wizualnej Jagiellońskiej Biblioteki Cyfrowej zamówiono w drodze konkursu znak graficzny, którego kształt symbolizuje połączenie tradycji i nowoczesności, a więc oddaje ideę pomysłodawców i twórców projektu, którzy postanowili ocalić bezcenne dziedzictwo kulturowe za pomocą dobrodziejstw cyfryzacji. Dla popularyzacji projektu Jagiellońska Biblioteka Cyfrowa zaprojektowano

i zamówiono szereg materiałów promocyjnych, zaopatrzonych w informacje o źródle współfinansowania (POIiŚ), a wśród nich zakładki do książek, długopisy, nośniki pamięci, teczki, kalendarze i torby bawełniane.

17 listopada 2011 r. Biblioteka Jagiellońska zorganizowała roboczą konferencję pt. *Opracowanie czasopism ukazujących się przed II wojną światową w kontekście ich publikacji w biblio-*

Abstrakt: Artykuł przedstawia działania podjęte na rzecz promocji projektu Jagiellońska Biblioteka Cyfrowa. Wśród nich należy wymienić artykuły prasowe, informacje w środkach masowego przekazu, akcje promujące Jagiellońską Bibliotekę Cyfrową, prezentacje projektu na krajowych i międzynarodowych konferencjach naukowych, dotyczących zachowania i digitalizacji dziedzictwa kulturowego, rozprowadzanie materiałów promocyjnych, mających utrwalić w świadomości potencjalnych użytkowników istnienie Jagiellońskiej Biblioteki Cyfrowej i jej współfinansowanie ze środków Unii Europejskiej.

Abstract: The paper presents the efforts undertaken for the sake of promoting of the Jagiellonian Digital Library Project. These include press releases, information for the media, promotional campaigns, project presentations at national and international conferences on preservation and digitization of cultural heritage, distributing promotional materials, as well as acquainting potential users with the Jagiellonian Digital Library and the fact of it is co-financed from the European Union.





Jagiellońska
Biblioteka
Cyfrowa

II. 1.

tece cyfrowej. Miała ona na celu przedyskutowanie istotnych problemów publikowania cyfrowego czasopism i zaprezentowanie rozwiązań przyjętych w Jagiellońskiej Bibliotece Cyfrowej. Podczas ww. konferencji pracownicy Biblioteki Jagiellońskiej, bezpośrednio zaangażowani w realizację projektu Jagiellońska Biblioteka Cyfrowa, wygłosili szereg referatów. Łukasz Mesek przedstawił dwa wystąpienia: *Odzwierciedlenie zamysłu wydawniczego w rekordzie bibliograficznym w formacie MARC 21* i *Rekord zasobu – praktyczne rozwiązania.*

24-25 stycznia 2013 r. odbyła się międzynarodowa konferencja naukowa, promująca - ale i podsumowująca - projekt Jagiellońska Biblioteka Cyfrowa, pn. *Biblioteka cyfrowa dziś a wyzwania jutra.* Celem konferencji było także omówienie związanych z projektem zagadnień

badawczych, kwestii technicznych oraz wyzwań, jakie niesie przyszłość. Wystąpienia prelegentów zogniskowano wokół trzech głównych zagadnień:

- obiekty oryginalne, w szczególności te należące do dziedzictwa narodowego (dobór, zabezpieczanie, konserwacja)
- obiekty cyfrowe (tworzenie, formaty, jakość)
- metadane (standaryzacja, przetwarzanie).

Działania promocyjne skierowano przede wszystkim do osób i instytucji, zainteresowanych internetowym dostępem do zbiorów Jagiellońskiej Biblioteki Cyfrowej, która gromadzi zdigitalizowane obiekty należące do Narodowego Zasobu Bibliotecznego. W owej grupie

II. 1. Logotyp projektu Jagiellońska Biblioteka Cyfrowa, zaprojektowany przez Motor Studio s.c.



II. 2.

docelowej znajdują się humaniści, pracownicy naukowci, hobbysci z małych ośrodków, badacze „małych ojczyzn”, szeroka rzesza regionalistów, uczelnie nie mające własnych zasobów bibliotecznych, animatorzy kultury, kustosze muzealni, dziennikarze i przedstawiciele mediów, pracownicy organizacji pozarządowych zaangażowanych w działalność kulturalną i naukową, pracownicy samorządowych instytucji kultury, nauczyciele i uczniowie wszystkich szczebli nauczania, w tym studenci. Przykładem ukierunkowanych działań promocyjnych były akcje podejmowane w trakcie Tygodnia Bibliotek, Tygodnia Open Access oraz zapoczątkowanie idei przekazywania przez środowisko uniwersyteckie licencji na publikację w Jagiellońskiej Bibliotece Cyfrowej prac naukowych i wyników badań, objętych prawami autorskimi.

Wysilek propagowania projektu na forum krajowym i zagranicznym podjęli pracownicy Biblioteki Jagiellońskiej, ściśle związani z jego realizacją. 14 kwietnia 2010 r., na dzień przed podpisaniem umowy o dofinansowanie, Łukasz Mesek, Aleksandra Szalla-Klemann, Katarzyna Szklanny wystąpili na konferencji w Centralnej Bibliotece Wojskowej pn. *Ochrona Narodowego Zasobu Bibliotecznego: digitalizacja i konserwacja tradycyjna z prezentacją Konserwacja tradycyjna i digitalizacja w Bibliotece Jagiellońskiej. Kompleksowe zabezpieczanie zbiorów – potrzeby i możliwości*, gdzie przedstawiali wchodzący w etap realizacji projekt. W kwietniu 2011 r., podczas seminarium w Rydze pn. *The complete preservation of the collection of graphic art of the Soviet period in Latvia and provision of the necessary technical-material resources*, Aleksandra Szalla-

II. 2-4. Materiały promocyjne zaprojektowane i zamówione na cele promocji projektu Jagiellońska Biblioteka Cyfrowa.



Il. 3.

Kleemann wystąpiła z prezentacją pt. *Some experiences of Jagiellonian Library* na temat przeprowadzanych w Bibliotece Jagiellońskiej projektów, w tym zwłaszcza Jagiellońskiej Biblioteki Cyfrowej. W dn. 10-14.04.2012 r. uczestniczyła w konferencji *IFLA International Newspaper Conference 2012* w Paryżu z referatem pt. *Deacidification of 19th and 20th centuries rare newspapers as a part of the Digital Jagiellonian Library Preserving Programme*, a w dn. 17-19.10.2012 r. wzięła udział w konferencji *Sztuka, rzemiosło, przemysł z XIX i XX wieku: konserwacja-restauracja*, na Wydziale Sztuk Pięknych (Zakład Konserwacji Papieru i Skóry) Uniwersytetu Mikołaja Kopernika w Toruniu, gdzie przedstawiła prezentację: *Program konserwacji i kompleksowego zabezpieczenia unikatowych zbiorów specjalnych i czasopism XIX- i XX-wiecznych ze zbiorów Biblioteki Jagiellońskiej w ramach projektu Ja-*

giellońska Biblioteka Cyfrowa. Ponadto na przełomie sierpnia i września 2012 r. podczas kursu *Conservation of Japanese Paper* organizowanym przez ICCROM i NRICPT w Tokio, w Japonii przedstawiła m.in. swoją pracę na rzecz projektu Jagiellońska Biblioteka Cyfrowa w Sekcji Konserwacji Biblioteki Jagiellońskiej.

Łukasz Mesek prezentował Jagiellońską Bibliotekę Cyfrową w trakcie sesji posterowej na konferencji *Polskie Biblioteki Cyfrowe 2010* w Poznaniu, mającej miejsce w dn. 18-21.10.2010 r. Na kolejnej edycji tejże konferencji, 13.10.2011 r., wystąpił z referatem pt. *Linkowanie zwrótnie w bibliotekach cyfrowych w aspekcie dostępu do metadanych oraz uzupełniania zasobów*. Promował Jagiellońską Bibliotekę Cyfrową także w bibliotece uniwersyteckiej w Jenie (18.05.2011 r.), na Targach Książki w Krakowie (04.11.2011 r.) oraz podczas obchodów Dnia Domeny Publicznej (20.01.2012 r.) w Krakowie, w klubie Kolanko no 6. Ponadto brał udział w dwóch konferencjach w Jasnej na Słowacji z referatami: *Jagiellonian Digital Library – solutions tailored to readers' needs* (17.04.2012, konferencja Digitalna Kniznica) oraz *E-books in Jagiellonian Digital Library - technological solutions and legal obstacles* (24.10.2012, konferencja *Digital Libraries 2015-2030*).

Adrian Wiecheć, odpowiedzialny za prawidłowe funkcjonowanie od strony technicznej Jagiellońskiej Biblioteki Cyfrowej, w dniach 18-22.10. 2010 r. brał

udział w konferencji *Polskie Biblioteki Cyfrowe* w Poznaniu z referatem *System Obsługi Jagiellońskiej Biblioteki Cyfrowej*. Bartłomiej Siedlarz promował Jagiellońską Bibliotekę Cyfrową podczas szkolenia STT Erasmus (20-28 marca 2012) w Narodowej i Uniwersyteckiej Bibliotece w Zagrzebiu, gdzie wystąpił z prezentacją *Jagiellonian Digital Library*. W dn. 26-27 kwietnia 2012 r. w ramach warsztatów *Cyfryzacja w instytucjach kultury i jednostkach samorządu terytorialnego* wygłosił referat *Wykorzystanie bezpłatnego oprogramowania w procesie digitalizacji*, podczas którego bazował na doświadczeniach Jagiellońskiej Biblioteki Cyfrowej. 24 października 2012 r. podczas Tygodnia Open Access przedstawił w Bibliotece Jagiellońskiej prezentację pn. *Użytkowanie zasobów Jagiellońskiej Biblioteki Cyfrowej*.

Wszystkie przedstawione powyżej działania, jak również niniejsza publikacja, postawiły sobie za cel promowanie wśród jak największej grupy odbiorców realizowany przez Bibliotekę Jagiellońską projekt o nazwie Jagiellońska Biblioteka Cyfrowa, współfinansowany przez Unię Europejską ze środków Europejskiego Funduszu Rozwoju Regionalnego w ramach Programu Operacyjnego Infrastruktura i Środowisko na lata 2007–2013, Priorytet XI „Kultura i Dziedzictwo Kulturowe”, Działanie 11.1. „Ochrona i zachowanie dziedzictwa kulturowego o znaczeniu ponadregionalnym”.



Il. 4.

Wykaz wybranych skrótów

BJ	Biblioteka Jagiellońska	KKZBUJ	Komputerowy Katalog Zbiorów Bibliotek Uniwersytetu Jagiellońskiego
DC	Dublin Core	MKiDN	Ministerstwo Kultury i Dziedzictwa Narodowego
DE	Document Express	NUKAT	Narodowy Uniwersalny Katalog Centralny
EBI	Europejski Bank Inwestycyjny	OCR	Optical Character Recognition
EFRR	Europejski Fundusz Rozwoju Regionalnego	ODLIS	Online Dictionary for Library and Information Science
FBC	Federacja Bibliotek Cyfrowych	OPAC	Online Public Access Catalog
HTML	HyperText Markup Language	ORP	Oddział Rękopisów
ICA	International Council on Archive	OSD	Oddział Starych Druków
ICR	Intelligent Character Recognition	OWRIDŹS	Oddział Wydawnictw Rzadkich i Dokumentów Życia Społecznego
IFLA	International Federation of Library Associations and Institutions	OZGIK	Oddział Zbiorów Graficznych i Kartograficznych
ITU	International Telecommunication Union		
JBC	Jagiellońska Biblioteka Cyfrowa		

OZM	Oddział Zbiorów Muzycznych
OZZ	Oddział Zabezpieczania Zbiorów
PBC	Podkarpacka Biblioteka Cyfrowa
PCSS	Poznańskie Centrum Superkomputerowo-Sieciowe
POIŚ	Program Operacyjny Infrastruktura i Środowisko
ppi	pixels per inch
SOJBC	System Obsługi Jagiellońskiej Biblioteki Cyfrowej
UJ	Uniwersytet Jagielloński
WWW	World Wide Web

