

Krzysztof KOPÍŃSKI
Toruń

„Księgi pruskie w Internecie” — projekt finansowany przez Fundację na rzecz Nauki Polskiej

W ostatnich latach coraz większym zainteresowaniem cieszą się edycje źródeł historycznych w postaci różnego rodzaju nośników elektronicznych oraz w Internecie. Jednym z przykładów tej ostatniej formy udostępnienia źródeł jest projekt „Księgi pruskie w Internecie” realizowany od 2007 r. przez Towarzystwo Naukowe w Toruniu w ramach programu PARTNERZY Fundacji na rzecz Nauki Polskiej. W omawianym projekcie zdecydowano się na **opracowanie modułu informatycznego**, w którym można byłoby gromadzić, przetwarzać i udostępniać przygotowaną zgodnie z przyjętymi zasadami edytorskimi zawartość ksiąg proveniencji krzyżackiej oraz miejskich. Jest faktem bezspornym, iż w erze wydawnictw multimedialnych i Internetu szczególnie ważnym zadaniem będzie zaprojektowanie aplikacji, która pozwoli z czasem na złamanie monopolu drukowanego papieru w edytorstwie źródłowym. W tym kontekście w związku z realizowanymi projektami edytorskimi rozpoczęto w ośrodku toruńskim starania mające na celu przygotowanie bazy danych z opracowanymi księgami.

W zrealizowanych w latach 2004–2007 w Toruniu trzech projektach edytorskich zamierzeniem było opracowanie i przygotowanie czterech ksiąg miejskich do druku oraz do elektronicznego udostępnienia¹. Powstały wówczas, częściowo zrealizowane, plany zbudowania prototypu aplikacji bazodanowej zawierającej wszystkie zapiski źródłowe znajdujące się w księgach wraz z indeksem osobowym, geograficznym i rzeczowym z możliwością swobodnego sortowania materiału przy uwzględnianiu porządku alfabetycznego i elementów chronologicznych. W ramach rekordu, jeżeli tylko byłoby to możliwe, miały być podane dodatkowe dane na temat osób wywodzących się z kręgu elit pruskich. Założono także, że baza danych powinna być dostosowana do pracy w Internecie na stronach www.

¹ 1. Projekt badawczy finansowany przez Komitet Badań Naukowych nr 2 H01G 045 24 „Księgi miejskie Młodego Miasta Gdańska z lat 1400–1455–opracowanie edytorskie i analiza treści”, zakończony 31 XII 2004 r., realizowany na Uniwersytecie Mikołaja Kopernika w Toruniu pod kierunkiem Piotra Olińskiego przy udziale Krzysztofa Kopińskiego; 2. Projekt badawczy finansowany przez Ministerstwo Nauki i Szkolnictwa Wyższego nr 1 H01G 037 29 „Księga ławnicza Starego Miasta Torunia (1456–1479)–opracowanie edytorskie i analiza treści” zakończony 28 II 2007 r., realizowany w Towarzystwie Naukowym w Toruniu pod kierunkiem Janusza Tandeckiego przy udziale K. Kopińskiego; 3. Projekt badawczy finansowany przez Ministerstwo Nauki i Szkolnictwa Wyższego nr 1 H01G 003 29 „Księga rachunkowa kamlarii miasta Torunia (1453–1495)–opracowanie edytorskie i analiza treści” zakończony 28 II 2007 r., realizowany w Polskim Towarzystwie Historycznym pod kierunkiem K. Kopińskiego przy udziale J. Tandeckiego i Krzysztofa Mikulskiego.

W przygotowaniu tekstu ksiąg do druku i do elektronicznego udostępnienia wykorzystano prace Adama Wolffa² i wskazówki opracowane przez Waltera Heinmeyera³, oparto się też na zasadach wypracowanych przez Karola Górskiego i Mariana Biskupa, a stosowanych przy edycji *Akt Stanów Prus Królewskich*⁴. Skorzystano również z doświadczenia zdobytego przy przygotowywaniu do druku księgi ławniczej Starego Miasta Torunia z lat 1428–1456⁵. Należy podkreślić, że pragnąc elektronicznie opracować i udostępniać treść ksiąg miejskich nie planowano zmiany wypracowanej i sprawdzonej w ośrodku toruńskim metody wydawniczej. Poziom obecnej komputeryzacji stwarza na tyle duże możliwości, że nie powinno być takiej konieczności⁶.

Na potrzeby w/w aplikacji zaplanowano wówczas stosunkowo niewielkie środki — 3800 zł⁷. Ta skromna suma pozyskana z grantów byłego Komitetu Badań Naukowych pozwoliła jednak wykażać się pewnymi osiągnięciami w staraniach o dalsze wsparcie finansowe pomysłu. Zaprojektowana wówczas baza o nazwie „Pruskie księgi miejskie” składała się ze strony głównej, na której znajdowało się krótkie omówienie projektu. W zakładce „Źródła” zaplanowano możliwość wybrania przez użytkownika interesującej go księgi za pomocą linku zawierającego tytuł, nazwiska wydawców i miejsce oraz rok przygotowania edycji. Następnie stworzono okno, z poziomu którego możemy przejść do wstępu do księgi, który składa się z rozdziałów zawierających:

- wprowadzenie
- informacje o urzędzie, który wytworzył księgę,
- opis i dzieje źródła,
- informacje o wykorzystaniu księgi w badaniach historycznych,
- informacje o zawartości księgi,
- opis metody wydawniczej,
- informacje o przypisach i skorowidzach sporządzonych dla księgi.
- podziękowania wydawców

Do wstępu dołączony jest także wykaz skrótów zastosowanych w księdze.

Z poziomu omawianego okna, w przypadku księgi ławniczej z lat 1456–1479, użytkownik może również przejść do zapisek za poszczególne lata i następnie wybrać numer właściwej wzmianki źródłowej. Wyświetla się wówczas oprócz numeru zapiski jej treść, numer strony w oryginale oraz poniżej przypisy edytorskie oraz przypisy merytoryczne. Można wówczas przejść do następnej wzmianki źródłowej w ramach tego samego roku, lub powrócić do poprzedniego okna i wybrać kolejne lata wraz z ich numerami zapisek.

Dla bazy danych stworzone zostały również wyszukiwarki:

- pełnotekstowa, która pozwala na wyszukiwanie słów lub fraz wpisanych przez użytkownika. Dodatkową zaletą jest możliwość ograniczenia wyników wyszukiwania do określonych lat. Wyszukane przez nas słowo lub fraza jest w odnalezionych wzmiankach uwypuklone kolorem żółtym oraz za pomocą podkreślenia. Jednocześnie użytkownik jest informowany, z jakich lat pochodzą zapiski z poszukiwanymi wyrazami,
- haseł rzeczowych, która pozwala na odnalezienie wzmianek źródłowych poprzez indeksy rzeczowe przygotowane przez wydawców poszczególnych źródeł. Pewną wadą może być jednak

² A. Wolff, *Projekt instrukcji wydawniczej dla pisanych źródeł historycznych do połowy XVI wieku*, St. Źródł., 1, 1957, s. 155–184.

³ Walter Heinmeyer, *Richtlinien für die Edition mittelalterlichen Amtsbücher*, w: *Richtlinien für die Edition Landesgeschichtlicher Quellen*, hrsg. Walter Heinmeyer, Marburg–Köln 1978, s. 18–23.

⁴ Zob. *Akta Stanów Prus Królewskich*, wyd. Karol Górski, Marian Biskup, t. 1–2, Toruń 1955–1957 (=TNT, Fontes 41 i 43).

⁵ *Księga ławnicza Starego Miasta Torunia (1428–1456)*, wyd. Karola Ciesielska, Janusz Tandecki, Część pierwsza (1428–1443), Toruń 1992; Część druga (1444–1456), Toruń 1993..

⁶ K. Kopiński, *Eine Datenbank zu den Stadtbüchern von Thorn und Danzig*, in: *Editionswissenschaftliche Kolloquien 2005/2007. Methodik–Amtsbücher. Digitale Edition–Projekte*, hrsg. M. Thumser, J. Tandecki, unter Mitarbeit A. Thumser, Toruń 2008, s. 65–72.

⁷ Towarzystwo Naukowe w Toruniu, Registratura własna.

fakt, iż indeksy do opracowanych ksiąg zostały sporządzone w języku polskim i na dzień dzisiejszy nie planuje się ich tłumaczenia. Warto nadmienić jednak, że terminy niemieckie i łacińskie podawano w brzmieniu źródłowym z odesłaniem do hasła głównego. Przy hasłach w języku polskim podano w nawiasie wersję lub wersje oryginalne. Terminy, których nie zdołano rozwiązać pozostawiono w wersji oryginalnej. Hasła treściowo bogatsze zostały rozbudowane za pomocą podhaseł. Dla uzasadnienia takich działań warto przypomnieć, że większość indeksów rzeczowych do źródeł wydawanych w toruńskim ośrodku naukowym była w podobny sposób budowana.

Wyszukiwanie rozpoczynamy od wybrania litery, na którą zaczyna się poszukiwane hasło rzeczowe. Następnie na wyświetlonej liście przy odnalezionym hasle wybieramy cyfrę, która oznacza numer zapiski. Wynik zapytania wyświetla się wraz z informacją o roku, z którego pochodzi wzmianka, jej numerem i treścią oraz danymi na temat numeru strony w oryginale. Pod wzmianką mogą znajdować się także przypisy edytorskie oraz przypisy merytoryczne. Nad zapiską zostało umieszczone wybrane przez użytkownika wcześniej hasło indeksowe. Z tego poziomu możemy powrócić do listy z wybraną przez nas wcześniej literą lub od razu możemy wybrać kolejną literę i przejść tym samym do następnej listy,

— hasel osobowych i geograficznych, który obejmuje osoby i nazwy miejscowe występujące w tekście źródłowym. W tym miejscu należy przypomnieć, że w projekcie wykorzystano źródła wydawane w latach 1936–2008 (w dalszej części artykułu zestawienie). Indeksy osobowe i geograficzne do tych źródeł budowano niekiedy bardzo niekonsekwentnie. W nowszych publikacjach źródłowych w hasłach z reguły umieszczono nazwiska oraz nazwy miejscowości w formie przyjętej w literaturze polskiej lub najczęściej występującej w źródle (wszystkie pozostałe formy umieszczono obok w nawiasach). Imiona wymienione przy nazwiskach podawano, o ile było to możliwe, w brzmieniu polskim oraz obok w nawiasach w brzmieniu źródłowym. Władców umieszczono pod imieniem. Hasła rozpoczynające się na literę C wymawianą jako K umieszczono pod literą K. Indeksy znajdujące się z kolei w starszych wydawnictwach źródłowych nie były dostosowywane w omawianym projekcie do obecnie stosowanych zasad budowania hasel. Wyszukiwanie hasel osobowych i geograficznych w aplikacji odbywa się analogicznie, jak w przypadku hasel rzeczowych.

Zaplanowano również, że stworzony zostanie w aplikacji panel administracyjny, który będzie ułatwiał włączanie do bazy danych wybranych ksiąg. Jako przykład w tym miejscu niech posłuży nam Księga ławnicza z lat 1456–1479. W jej opracowaniu edytorskim pozostawiono w zasadzie bez zmian istniejący układ wpisów. Jedynym odstępstwem było uwzględnianie znajdujących się w tekście źródłowym znaków pisarskich, które informowały wydawców o konieczności przesunięcia niektórych słów, zwrotów, zdań czy nawet wzmianek. Wytworzona sytuacja pozwalała na nadanie wszystkim zapiskom numerów. Wspomniana numeracja została wykorzystana także przy sporządzaniu przypisów edytorskich i rzeczowych oraz indeksów. Nie można w tym miejscu nie zauważyć, iż ten stan rzeczy był bardzo korzystny z punktu widzenia technicznego. Ponadto księga bardzo łatwo dała się również podzielić na jednoroczne przedziały chronologiczne, co również dawało pewne możliwości z punktu widzenia informatycznego. Proces budowy bazy danych ulegał więc przyspieszeniu. Nie było również większych problemów z włączeniem opracowania źródła do przygotowywanej aplikacji ze względu na znaczną przejrzystość struktury księgi, której dodatkowo jeszcze wydawcy nadali ostateczny kształt. Nabyte doświadczenie w trakcie prac informatycznych z opracowaniem edytorskim księgi ławniczej z lat 1456–1479 miało być wykorzystane przy włączaniu do bazy danych następnych ksiąg. Należy jednak wyraźnie zaznaczyć, że proces automatyzacji włączania ksiąg za pomocą panelu administracyjnego mógł dotyczyć jedynie źródeł o przejrzystej strukturze wewnętrznej. Przed wszystkim relatywnie sprawnie można było włączyć jednorodne pod względem budowy pozostałe księgi ławnicze uwzględnione w projekcie.

Tak przygotowany projekt bazy danych nie uległ w zasadzie do teraz żadnym większym przeobrażeniom. Należy zaznaczyć, że nie zmienialiśmy w posiadanej aplikacji utworzonych relacji pomiędzy poszczególnymi tabelami.

W lipcu 2007 r. Towarzystwo Naukowe w Toruniu zdecydowało się na wystąpienie z projektem „Księgi pruskie w Internecie. Studium nad metodą elektronicznej publikacji źródeł” w ramach programu PARTNERZY Fundacji na rzecz Nauki Polskiej. Program PARTNERZY został powołany właśnie w 2007 r. dla zwiększenia dotychczasowego zakresu współpracy Fundacji z ogólnopolskimi, interdyscyplinarnymi towarzystwami naukowymi oraz innymi organizacjami pozarządowymi o uznanym prestiżu, działającymi w obszarze nauki lub na rzecz wspierania nauki i rozwoju kadry naukowej w Polsce i na świecie. Celem programu PARTNERZY jest pobudzenie aktywności i rozwój działalności towarzystw naukowych i innych organizacji pozarządowych, działających w obszarze nauki, poprzez wsparcie finansowe Fundacji. Za służące realizacji tych celów Fundacja uznaje przede wszystkim te inicjatywy i projekty, które bezpośrednio służą rozwojowi potencjału młodej kadry naukowej, poprawie współpracy między naukowcami, zwiększeniu znaczenia organizacji pozarządowych działających w obszarze nauki, a także wspierają rozwój instytucjonalny tych organizacji⁸. Na marginesie warto dodać, że w 2008 r. Fundacja zakończyła realizację programu PARTNERZY, co zostało uzasadnione gwałtownym pogorszeniem sytuacji na rynkach finansowych w drugiej połowie 2008 r.⁹

We wniosku Towarzystwo Naukowe poinformowało o dotychczasowych pracach nad bazą danych. Mianowicie w aplikacji znajdowały się już księgi opublikowane w latach 2007–2008:

1. *Księga ławnicza Starego Miasta Torunia (1456–1479)*, wyd. K. Kopiński, J. Tandecki, Toruń 2007 (=Fontes TNT 99).
2. *Księga kamlarii miasta Torunia z lat 1453–1495*, wyd. K. Kopiński, K. Mikulski, J. Tandecki, Toruń 2007 (=Źródła do dziejów średniowiecznego Torunia, cz. 3).
3. *Księgi Młodego Miasta Gdańska 1400–1455 [1458–1459]*, wyd. K. Kopiński, P. Oliński, Toruń 2008 (=Fontes TNT 100).

Towarzystwo Naukowe zauważyło również, że na obecnym etapie szczególnie cenne byłoby nadanie właściwego rozmachu i nowego kierunku dla skromnej dotąd inicjatywy. Cele projektu miały być osiągnięte poprzez wykorzystanie struktury posiadanej bazy danych, do której włączonoby część dotychczas opublikowanych źródeł w serii „Fontes” Towarzystwa Naukowego w Toruniu¹⁰. Umowa pomiędzy Towarzystwem Naukowym a Fundacją została podpisana w grudniu 2007 r. Wykonawcami projektu miało być sześć osób: prof. dr hab. Janusz Tandecki; mgr Janusz Bonczkowski, pracownik Archiwum Państwowego w Toruniu; mgr Agnieszka Rosa, doktorantka w Instytucie Historii i Archiwistyki Uniwersytetu Mikołaja Kopernika; mgr Paweł Jeziorski, pracownik Zakładu Historii Pomorza i Krajów Bałtyckich Polskiej Akademii Nauk w Toruniu; informatyk mgr inż. Bolesław Rassalski, pracownik Archiwum Państwowego w Bydgoszczy i piszący te słowa.

W ramach projektu zaplanowano zeskanowanie średniowiecznych i nowożytnych ksiąg miejskich i ksiąg proveniencji krzyżackiej wydanych w minionych latach przez Towarzystwo Naukowe (11 wydawnictw źródłowych, w tym dwa dwutomowe, opublikowane w latach 1936–1997 — łącznie 13 „Fontesów”):

1. *Księga czynszów fary chełmińskiej (1435–1496)*, wyd. Z.H. Nowak, J. Tandecki (=Fontes TNT 78), Toruń 1994;
2. *Księga długów miasta Torunia z okresu wojny trzynastoletniej*, wyd. K. Ciesielska, I. Janosz-Biskupowa (=Fontes TNT 55), Toruń 1964;

⁸ Regulamin programu PARTNERZY Fundacji na rzecz Nauki Polskiej. Tekst jednolity przyjęty uchwałą nr 73/2007 Zarządu Fundacji na rzecz Nauki Polskiej z dnia 8 V 2007 r., s. 1 (<http://www.fnp.org.pl/files/programy/25/regulamin_partnerzy_10052007.pdf>).

⁹ <http://www.fnp.org.pl/programy/programy_zakonczone/inne_inicjatywy/program_partnerzy>; <http://www.fnp.org.pl/aktualnosci/aktualnosci_52>.

¹⁰ Towarzystwo Naukowe w Toruniu, Registratura własna.

3. *Księga komturstwa gdańskiego*, wyd. K. Ciesielska, I. Janosz–Biskupowa (=Fontes TNT 70), Toruń 1985;
4. *Księga ławnicza Nowego Miasta Torunia 1387–1450*, wyd. K. Ciesielska (=Fontes TNT 63), Toruń 1973;
5. *Księga ławnicza sądu przedmiejskiego Chełmna 1480–1559 (1567)*, wyd. Z.H. Nowak, J. Tandecki (=Fontes TNT 74), Toruń 1990;
6. *Księga ławnicza Starego Miasta Torunia*, wyd. K. Ciesielska, J. Tandecki,
Cz. 1: (1428–1443). — Toruń 1992. — XXV, 203 s. — (=Fontes TNT 75),
Cz. 2: (1444–1456). — Toruń 1993. — 272 s. — (=Fontes TNT 76);
7. *Księga rachunkowa urzędów rybickich komturstw malborskiego i dzierzgońskiego 1440–1445*, wyd. Z.H. Nowak, J. Tandecki (=Fontes TNT 82), Toruń 1997;
8. *Księga Theudenkusa*, wyd. L. Koczy (=Fontes TNT 33), Toruń 1937;
9. *Księga żoldu Związku Pruskiego z okresu wojny trzynastoletniej 1454–1466*, wyd. A. Czarowski (=Fontes TNT 61), Toruń 1969;
10. *Liber scabinorum Veteris Civitatis Thoruniensis 1363–1428*, wyd. K. Kaczmarczyk (=Fontes TNT 29), Toruń 1936;
11. *Nowa księga rachunkowa Starego Miasta Elbląga: 1404–1414*,
Cz. 1: (1404–1410), wyd. M. Pelech (=Fontes TNT 72), Toruń 1987,
Cz. 2: (1411–1414), wyd. M. Pelech (=Fontes TNT 73), Toruń 1989.

Następnie przewidziano dokonanie OCR (ang. Optical Character Recognition — w największym skrócie rozpoznanie tekstu w zeskanowanym dokumencie), po którym powinna nastąpić niezbędna weryfikacja i korekta plików tekstowych związana z możliwością wystąpienia błędów (literówek) po obróbce (około 5%). Przyjęto założenie, że każde wydawnictwo będzie sprawdzane i poprawiane przez dwie różne osoby. Ponadto osoby sprawdzające miały dostosowywać posiadane pliki tekstowe do włączenia do bazy danych. Korektorzy wskazywali, że dużo czasu zajmowało im usuwanie dzielenia wyrazów, podwójnych spacji pomiędzy wyrazami, poprawianie różnego rodzaju literówek, szczególnie w przypadku liter podobnych graficznie (l = (równało się) i, i=l, nn=m, in=m, m=in, m=ln), sprawdzanie i ewentualnie dodawanie brakujących niemieckich i polskich znaków diakrytycznych, usuwanie niepotrzebnego formatowania tekstu, wprowadzenie zgubionych znaków interpunkcyjnych, poprawianie znaków interpunkcyjnych (zamiast kropki pojawiały się dwie kropki lub zamiast przecinka pojawiała się kropka i przecinek). Osoby sprawdzające informowały także, że występowały problemy, gdy w wydanej drukarni wydawnictwie źródłowym umieszczano numer strony lub karty w rękopisie na marginesie obok tekstu źródłowego. Po OCR brakowało takiego numeru strony lub karty, łączył się z tekstem źródłowym w nieprawidłowym miejscu, bądź uzyskiwał formatowanie różne od formatowania tekstu źródłowego. Dużo kłopotów sprawiały poprawiającym również znaki graficzne występujące w tekstach źródłowych. Po obróbce OCR wspomniane znaki były zastępowane przez zbędne symbole lub litery. Ponadto litery oznaczające przypisy edytorskie zamieniały się w liczby. Zarówno przypisy merytoryczne jak i rzeczowe mogły przemieszczać się na kolejne strony. Wszystkie zasygnalizowane problemy wymagały interwencji sprawdzających. Korektorzy zgodnie twierdzili, że pierwsze poprawiane przez nich księgi miały zdecydowanie więcej błędów po OCR niż ostatnie oddane im do weryfikacji źródła. Program do rozpoznawania zeskanowanego tekstu miał możliwość zapamiętywania nowych nieznanych mu znaków występujących w wydawnictwach źródłowych Towarzystwa Naukowego. Przy niewielkim nakładzie pracy ze strony informatyka kolejne księgi przekazywane do poprawiania zawierały zdecydowanie mniej błędów. Wspomniana weryfikacja teksów ksiąg została już zakończona. W pracach tych mieli uczestniczyć: mgr Janusz Bonczkowski; mgr Agnieszka Rosa; mgr Paweł Jeziorski. Nie obyło się bez pewnych problemów. Z udziału w projekcie ostatecznie zrezygnował P. Jeziorski w związku z finalizowaniem prac nad swoją dysertacją. P. Jeziorski został ostatecznie zastąpiony przez mgrów Dorotę i Bartosza Drzewieckich, doktorantów w IHiA UMK.

Równocześnie cały czas prowadzone były prace nad wspomnianą bazą danych. Po rozszerzeniu jej zakresu merytorycznego zaplanowano zmianę nazwy aplikacji z „Pruskie księgi miejskie” na „Księgi pruskie w Internecie”. Jednocześnie baza danych została nieznacznie zmodyfikowana i dostosowana do nowych potrzeb, zadbano również o jej nowy design. Prace te w przeważającej mierze były zaplanowane na trzeci rok trwania projektu (XII 2009–XII 2010). Obecnie koncentrujemy się na sukcesywnym włączaniu ksiąg po korekcie do bazy danych. Do tego celu wykorzystujemy strukturę aplikacji przygotowaną w ramach grantów byłego Komitetu Badań Naukowych i Ministerstwa Nauki i Szkolnictwa Wyższego. Wspomnianą strukturę uznajemy za już sprawdzoną i przetestowaną, i nie planujemy żadnych zasadniczych zmian dotychczasowego projektu bazy danych. Doświadczenia nabyte w trakcie budowy pierwszej bazy danych są teraz dla nas bardzo cenne. Być może z punktu widzenia metodycznego należałoby zaczynać właśnie od niewielkich projektów elektronicznego opracowywania i udostępniania źródeł, aby następnie można było podjąć się próby realizacji kolejnych większych pomysłów w tym zakresie.

Aplikacja została przygotowana w bazie danych SQL (ang. Structured Query Language — strukturalny język zapytań używany do tworzenia, modyfikowania baz danych oraz do umieszczania i pobierania danych z baz danych) w powiązaniu z językiem PHP. Zastosowane oprogramowanie to najnowsza wersja MySQL (wolnodostępny system zarządzania relacyjnymi bazami danych). Ważniejszy jest jednak fakt, że zdecydowano się na przygotowanie aplikacji z możliwością uwidaczniania w XML. Przypomnijmy w tym miejscu, że HTML to bardzo sztywna struktura, która umożliwia jedynie zapisywanie typowych dokumentów hipertekstowych. XML (ang. Extensible Markup Language, w wolnym tłumaczeniu Rozszerzalny Język Znaczników) z kolei jest uniwersalnym językiem bez predefiniowanego zbioru znaczników. XML pozwala także na wymianę danych między bazami, co oznacza, że nie jesteśmy uzależnieni od oprogramowania i w przyszłości będziemy mogli ewentualnie przeprowadzić migrację danych.

Po przeprowadzonych pracach technicznych powinna być przygotowana dokumentacja bazy. Planuje się utrzymanie oprócz interfejsu w języku polskim, interfejsu w języku niemieckim. Z każdą księgą w bazie danych zaplanowano zintegrowanie dołączonego do niej indeksu geograficznego i osobowego oraz indeksu rzeczowego. Opisane działania przeprowadzono już dla części ksiąg, co wiązało się ze znacznym nakładem pracy w związku z dostosowywaniem indeksów do potrzeb bazy danych. Prace polegały m. in. na usuwaniu dwukolumnowego formatowania tekstu indeksów, usuwaniu łamania indeksu wykonanego w wydawnictwie oraz pracochłonnym, ale i zarazem czysto technicznym dodawaniu haseł głównych dla podhaseł (zob. przykład 1–2). O ten ostatni zabieg poprosił informatyk. Mianowicie bez tych działań indeksy mogłyby być, szczególnie w przypadku haseł rozbudowanych, nieczytelne dla użytkowników baz danych.

Przykład 1: Hasło „Agnieszka” wraz z podhasłami w Księdze ławniczej Starego Miasta Torunia z lat 1428–1456.

- Agnieszka (Agnite, Agnithe) 409
- bracia Burkard, Zygfryd i krawiec Antoni 1220
- brat Herman Czenmark 1036, 1037
- córka Kunegundy Czenmark 1036, 1037
- Mikołaja Porren 597
- siostra Tomasza Scholtzchcin 604
- szwagier Urban 1220
- wdowa po Ludwiku Schindeler 1381
- żona Agustyna Bogener 616
- Jana Czimmerman 1561
- Kacpra Kzensze 1562

Przykład 2: Hasło „Agnieszka” wraz z podhasłami po dostosowaniu do potrzeb bazy danych.

Agnieszka (Agnite, Agnithe) 409

Agnieszka, bracia Burkard, Zygryd i krawiec Antoni 1220

Agnieszka, brat Herman Czenmark 1036, 1037

Agnieszka, córka Kunegundy Czenmark 1036, 1037

Agnieszka, córka Mikołaja Porren 597

Agnieszka, siostra Tomasza Scholtzchin 604

Agnieszka, szwagier Urban 1220

Agnieszka, wdowa po Ludwiku Schindeler 1381

Agnieszka, żona Agustyna Bogener 616

Agnieszka, żona Jana Czimmerman 1561

Agnieszka, żona Kacpra Kzensze 1562

Należy w tym miejscu zaznaczyć, że B. Rassalski wskazywał, jako najbardziej absorbujące zadanie, ręczne łączenie (ustanawianie relacji) numerów stron lub zapisek w indeksach z tekstem źródłowym. Pomimo znacznego zaawansowania obecnych technologii nie udało się tej kwestii rozwiązać w inny sposób i zautomatyzować procesu powiązania numerów (liczb) w indeksach ze źródłem. Informatykowi problemy sprawiało również łączenie wzmianek źródłowych z przypisami edytorskimi i przypisami merytorycznymi. Dodatkowa trudność polegała na tym, że w przygotowanym do druku wydawnictwie zarówno wzmianki jak i przypisy mogły znajdować się na kilku stronach. Wówczas trzeba było scalać te elementy i umieszczać we właściwych rekordach bazy danych. Przygotowano do tego celu specjalne tabele, które ułatwiały włączanie przypisów edytorskich i merytorycznych do aplikacji (zob. tab. 1). Podobne zestawienie tabelaryczne stworzono dla numerów zapisek i stron w rękopisie. Dzięki temu zabiegowi pod każdą zapiską jest informacja, na jakiej stronie znajduje się ona w oryginale. Kolejną sygnalizowaną trudnością było usuwanie zbędnych akapitów (tabulacji) oraz wymuszonych przenosin tekstu do następnego wersu (enterów). Być może najlepiej będzie oddawać te kwestie wypowiedź informatyka, który stwierdził, że niniejszy projekt jest dla niego „walką z formatowaniem tekstu” wprowadzonym przy łamaniu źródła do druku i powstałym częściowo także po obróbce OCR.

Tabela 1. Przypisy edytorskie i merytoryczne dla pierwszych zapisek w Księdze Theudenkusa umieszczone w tabeli pozwalającej na sprawne włączenie ich treści do aplikacji

NR ZAPISKI	SYMBOL PRZYPISU	TREŚĆ PRZYPISU
1	1	Brak monografii o udziale Torunia w wojnie trzynastoletniej, choć materiał jest nader obfity i dość skupiony. Źródła współczesne, jak recesy Stanów Pruskich, materiał dyplomatyczny, wreszcie kroniki pozwalają poznać wyjątkową rolę tego miasta w dziejach wojny, zwłaszcza w latach 1453–1455. Toruń był z całą pewnością bardziej bezinteresowny w ofiarach na rzecz Związku od Gdańska, który za każdy wydany grosz wyłudzał przywileje, a wobec Polski prowadził politykę, która mu umożliwiała powrót pod rządzą Zakonu. Najobszerniej w tym przedmiocie traktuje praca J.E. Wernicke’go: <i>Geschichte Thorns aus Urkunden, Dokumenten und Handschriften</i> , t. I (1842), str. 201–284, która atoli nie może się mierzyć z analogiczną rozprawą Simsona o udziale Gdańska w wojnie trzynastoletniej (P. Simson, <i>Danzig im dreizehnjährigen Kriege 1454–1466</i> . <i>Ztschr.d. Wpr. Gesch. Ver.</i> , zes. 29 (1891)). O wiele więcej daje krótka rozprawa E. Kestnera: <i>Thorns Abfall vom Deutschen Orden (Beiträge zur Geschichte d. Stadt Thorn. Thorn 1883</i> , str. 67–145). Zwięźle, ale na podstawie nowych materiałów, omówił stosunek Torunia do Zakonu w czasie wojny trzynastoletniej Karol Górski w <i>Historji politycznej Torunia do roku 1793 (w Dziejach Torunia. Praca zbiorowa z okazji 700–lecia Miasta pod redakcją Kaz. Tymienieckiego. Toruń 1933</i> , str. 49–65) i w »Pomorzu w dobie wojny trzynastoletniej« (Odbitka z VII tomu <i>Prac Komisji Historycznej Pozn. Tow. Przyj. Nauk. 1932</i> , str. 239 i n.).

NR ZAPISKI	SYMBOL PRZYPISU	TREŚĆ PRZYPISU
1	2	Co do szosu, które miasta pruskie pobierały na cele Związku Pruskiego, wbrew woli Zakonu a za zgodą cesarza Fryderyka III, porównaj przedmowę i M. Toeppena: Acten d. Stdtage, III, nr. 95, 137, 144, 203, 275, 315, 337, 341 i uwaga 1 na str. 706. Już na zjeździe w Kwidzynie 25 lutego 1453, a więc w dwa miesiące po wyroku cesarskim (22 grudnia 1452 roku), »haben landt und stedte voramet ein gemein geschos obir lande und stedte in der vereinunge wesenden, in sulcher weise, das man ½ fr. gutten geldes zu vorgeschosze undt 2 den. von der geringen mark geben sal, iglicher bey seinem eyde und die cleinen stedte sullen allewege binnen 14 tagen der grossen stat bey in gelegen semlich geld, was von dem geschosse mit in wirt gefallen, brengen...« (Toeppen: Acten d. Stdtage III, nr. 315). W Toruniu, jak o tem donosił komtur tamtejszy, szos istotnie wybrano już przed 8 marca 1453 (tamże III, nr. 330).
9	a	Nieodgadnione.
10	b	Przekreślone.

Przykładowe zestawienie numerów stron w rękopisie z kilkoma pierwszymi zapisami w *Liber scabinorum* przedstawiało się następująco:

- 1 — s. 1
- 2 — s. 1
- 3 — s. 1–1v
- 4 — s. 1v
- 5 — s. 1v
- 6 — s. 1v
- 7 — s. 1v–2
- 8 — s. 2
- 9 — s. 2
- 10 — s. 2
- 11 — s. 2
- 12 — s. 2–2v

Oprócz tego zdecydowano się także na realizację jeszcze jednego bardzo istotnego zamierzenia (prof. J. Tandecki i piszący te słowa), którym zarówno w kontekście technicznym, jak i merytorycznym będzie przygotowanie na bazie indeksów rzeczowych przygotowywanych dla ksiąg przez różnych wydawców jednego centralnego wykazu słów rzeczowych. Pozwoli to na stworzenie glosarium wspólnego dla wszystkich ksiąg, co w efekcie znacznie ułatwi przeszukiwanie aplikacji. We wspomnianym glosarium usunięte zostaną dotychczas zauważone błędy występujące w indeksach rzeczowych oraz wykonane zostaną niezbędne ujednolicenia. W trakcie prac uwzględnimy także uwagi krytyczne do indeksów rzeczowych znajdujące się w recenzjach naukowych wydanych źródeł. Wydaje się, że dzięki glosarium badacze różnych dziedzin (historycy, historycy sztuki, językoznawcy, archeolodzy, etnologzy) uzyskają nowe możliwości wyszukiwania potrzebnych im informacji¹¹. Jednocześnie zwiększy się jakość tych informacji.

Po zakończeniu przedstawionego projektu w grudniu 2010 r. planuje się w dalszej perspektywie wystąpienie do Ministerstwa Nauki i Szkolnictwa Wyższego z wnioskiem o sfinansowanie włączenia do bazy danych tekstów wielotomowej serii wydawniczej „Akt Stanów Prus Królewskich” oraz „Protokołów Sejmiku Generalnego Prus Królewskich”¹². Należy sądzić, że byłoby to szczególnie istot-

¹¹ Wspomniane glosarium będzie nawiązywało do podobnych pomysłów realizowanych już w ośrodku toruńskim: *Słownik trudniejszych wyrazów niemieckich w Aktach Stanów Prus Królewskich*, t. 1–2 (1479–1492), oprac. A. Bzdęga przy współpracy H. Zawadzkiej, Toruń 1959.

¹² *Akta Stanów Prus Królewskich*, wyd. K. Górski, M. Biskup, I. Janosz–Biskupowa, t. 1–8, (=Towarzystwo Naukowe w Toruniu, Fontes 41, 43, 50, 54, 57, 59, 64, 65, 66, 68, 71, 77, 1955–1993); *Protokoły sejmiku Generalnego Prus Królew-*

ne uzupełnienie aplikacji, które pozwoliłoby zarazem na sprawdzenie i weryfikację wypracowanej metody elektronicznej publikacji źródeł. Jednocześnie świat nauki uzyskałby nieskrępowany dostęp online do najcenniejszych edycji źródłowych wydawanych w ramach serii „Fontes” Towarzystwa Naukowego w Toruniu.

Projekt koncentruje się obecnie na źródłach, które zostały już wydane drukiem w minionych latach. Być może już w niedalekiej przyszłości należałoby przetestować posiadaną aplikację pod jeszcze jednym kątem. Mianowicie należałoby sprawdzić przydatność bazy danych w procesie opracowania i przygotowania źródła do druku. W nowoczesnym (skomputeryzowanym) edytorstwie źródłowym droga winna być bowiem następująca: od bazy danych do drukowanej książki, względnie CD lub prezentacji w Internecie, a nie odwrotnie¹³.

Z przygotowaniem elektronicznej publikacji źródeł wiąże się również wypracowanie i przetestowanie nowych metod pracy oraz rozwiązanie wielu problemów metodycznych. Już w trakcie prac nad prototypem aplikacji wystąpiły trudności związane z włączeniem do bazy danych księgi rachunkowej kamlarii miasta Torunia. Tekst księgi kamlarii nie był jednorodny. Wspomniana księga rachunkowa nie dawała się również w przejrzysty sposób podzielić chronologicznie. Ponadto tylko niewielka część wyodrębniających się z całości zapisów miała oryginalne tytuły. Dla zachowania przejrzystości opracowania edytorskiego wydawcy podjęli decyzję o podzieleniu tekstu źródła na ponumerowane od 1 do 40 grupy rzeczowe. Do każdej z tych grup dołączono krótki opis. Naturalnie bez problemu można będzie włączyć tekst źródła do bazy danych. Jednakże wersja drukowana opracowania posiada indeksy odwołujące się do przełamanych stron druku. Do czego z kolei miałyby odwoływać się indeksy w bazie danych? Już wtedy w początkowym etapie prac wiedzieliśmy, że problemem będzie powiązanie indeksów z czterdziestoma dużymi grupami rzeczowymi. Pewną możliwością byłoby także powiązanie indeksów ze stronami w rękopisie¹⁴. W jednym i drugim przypadku indeksy musiałyby zostać poprawione, co wiązałoby się ze znacznym nakładem pracy. Należy zaznaczyć, że dla księgi kamlarii początkowo chcieliśmy uruchomić tylko wyszukiwanie pełnotekstowe.

Podział przekazu źródłowego na czterdzieści grup rzeczowych przy jednoczesnym braku możliwości wykorzystania elementów chronologicznych wymusił również dostosowanie interfejsu w bazie danych do specyfiki księgi kamlarii. W tym celu wykorzystano opisy wzmiankowanych grup rzeczowych.

Zasygnalizowany problem musiał szybko znaleźć rozwiązanie, gdyż w projekcie zgłoszonym do Fundacji na 11 wydawnictw źródłowych jeszcze cztery księgi posiadały indeksy odwołujące się do numerów stron druku: Księga komturstwa gdańskiego; Księga długów miasta Torunia z okresu wojny trzynastoletniej; Księga rachunkowa urzędów rybickich komturstw malborskiego i dzierzgońskiego 1440–1445; Księga żołdu Związku Pruskiego z okresu wojny trzynastoletniej 1454–1466. Z wymienionych pierwsze dwie księgi były najobszerniejsze. Dodatkowo Księga komturstwa gdańskiego była podzielona przez wydawczynię na 235 wyodrębniających się, w zasadzie kancelaryjnie, wpisów poprzedzonych numerem, polską nazwą miejscowości, której dotyczył tekst, miejscem i datą wystawienia dokumentu, rejestrem, adnotacjami o miejscu przechowywania oryginału, kopii, transumptu oraz miejscu druku, także przy samych tylko rejestrach¹⁵. Z kolei Księga długów miasta Torunia z okresu wojny trzynastoletniej przy zachowaniu oryginalnego układu również została podzielona na 177 zapisów, które zostały opatrzone w krótkie rejestry. Oprócz tego wydawczynię sporządziły chrono-

skich, wyd. M. Biskup, B. Dybaś, J. Tandecki, t. 1–2, Toruń 2001–2005 (=Towarzystwo Naukowe w Toruniu, Fontes 89, 95); *Protokoły Sejmiku Generalnego Prus Królewskich*, wyd. M. Biskup, K. Kopiński, P. Oliński, J. Tandecki, t. 3, Toruń 2010 (=Fontes 102).

¹³ Por. S. Waldhoff, *Amtsbuchedition zwischen bewährter editorischer Praxis, gestiegenen Ansprüchen und neuen technischen Möglichkeiten*, w: *Editionswissenschaftliche Kolloquien 2005/2007*, s. 160.

¹⁴ K. Kopiński, *Eine Datenbank*, s. 69–70; S. Waldhoff, *Amtsbuchedition*, s. 158.

¹⁵ *Księga komturstwa gdańskiego*, wyd. K. Ciesielska, I. Janosz–Biskupowa, Toruń 1985, s. XXII, XXV–XLVI.

giczne zestawienie zapisek¹⁶. Należy wskazać, że były to elementy, które pozwalałyby na odejście od powiązania numerów w indeksach ze stronami druku wydanych ksiąg. Nad uczestnikami projektu ciążyłoby jednak pracochłonne poprawianie zestawionych już wcześniej indeksów dla relatywnie obszernych wydawnictw źródłowych. Wydanie decyzji w tej sprawie było szczególnie problematyczne. Wykorzystywane w projekcie księgi cytowane są już w literaturze naukowej wraz z numerami stron do interesujących badaczy zapisów. Trudno jednoznacznie wskazać, czy zmiana tego stanu rzeczy (tj. ewentualne wykorzystanie numerów zapisów w indeksach) byłaby korzystnym rozwiązaniem. Jednakże postanowiono, że w bazie danych w przypadku Księgi komturstwa gdańskiego i Księgi długów miasta Torunia z okresu wojny trzynastoletniej stworzona zostanie możliwość zapoznania użytkownika z podziałami na wpisy, które zaproponowały wydawczynie. Dodatkowo dla Księgi długów przygotowana zostanie możliwość przeszukiwania z wykorzystaniem elementów chronologicznych.

Wykorzystanie numerów stron w indeksach wiązało się także z decyzją, że każda strona drukowana z tekstem źródłowym będzie stanowiła w aplikacji jeden rekord. Z tego powodu łącznie pięć ksiąg wymagało w bazie danych specjalnego potraktowania. Od razu też pojawiłby się do rozwiązania problem techniczny związany z dzieleniem wyrazów w wydanych księgach. Zachowanie przeniesienia części wyrazu na następną stronę z tekstu źródłowego lub przypisów nie pozwalałoby znaleźć danego słowa w wyszukiwaniu pełnotekstowym. Uczestniczący w projekcie informatyk otrzymałby więc kolejne zadanie do rozstrzygnięcia. Wspomniane problemy pokazują, jak wiele zagadnień jest do przeanalizowania w projektach elektronicznego udostępniania źródeł. W wielu zaś przypadkach przyjęte rozwiązania metodyczne i programistyczne nie będą potem w ogóle dostrzegalne dla użytkownika, którego będzie interesowała tylko konkretna informacja.

Warto w tym miejscu zwrócić uwagę na jeszcze jedną kwestię. W bazie danych, nad którą prace zakończyliśmy w 2007 r., została umieszczona publikacja „Księgi Młodego Miasta Gdańska 1400–1455 [1458–1459]”. Na wspomnianą edycję źródłową składały się dwie księgi — księga rady i księga przyjęć do prawa miejskiego. Każda z tych ksiąg miała swoje indeksy geograficzne i osobowe oraz rzeczowe. Wspomniane wydawnictwo źródłowe udało się w zasadzie bezproblemowo umieścić w zaprojektowanej aplikacji. Jedynie w przypadku księgi rady w związku z niemożnością ustalenia chronologii wszystkich zapisek trzeba było dostosować interfejs w bazie danych do specyfiki źródła¹⁷. Z podobnym problemem zetknęliśmy się w Księdze Theudenkusa, która jest księgą rachunkową dzielącą się na dwa działy: dochodów i wydatków. Wydawca zdecydował się na wprowadzenie dla każdego z tych działów odrębnej numeracji zapisów. Indeksy są jednak dla całej księgi wspólne. Po hasle indeksowym mogą występować numery zapisów dotyczące dochodów, a następnie po dużej literce B. podawane są numerów zapisów związane z rozchodami¹⁸. Po raz kolejny więc mamy do czynienia z przypadkiem, gdzie musimy posiadać przez nas źródło potraktować indywidualnie. W naszej aplikacji będzie więc występowała Księga Theudenkusa — dochody i Księga Theudenkusa — rozchody. Dalszego i głębszego zaangażowania strony technicznej projektu będzie zaś wymagało zintegrowanie indeksów z tak rozbitą w bazie księgą.

W podsumowaniu niniejszego artykułu powinna paść odpowiedź na pytanie, dlaczego decydujemy się na unifikację w jednej bazie danych tak różnorodnych ksiąg? Naturalnie najkorzystniej byłoby w przyszłości opracowywać bazy danych dla określonych serii ksiąg. W ten sposób unikniemy dostosowywania interfejsu posiadanego modułu informatycznego do specyfiki różnych źródeł. Tworzona aplikacja będzie wówczas bardziej jednolita i przystępna¹⁹. Podtrzymując wyrażony pogląd należy także wskazać na dwie zalety realizowanego teraz projektu. Po pierwsze, wprowadzanie do bazy danych i udostępnianie różnych ksiąg przynosi cenne doświadczenia, które zostaną wykorzystane

¹⁶ *Księga długów miasta Torunia z okresu wojny trzynastoletniej*, wyd. K. Ciesielska, I. Janosz–Biskupowa, Toruń 1964, s. XVI–XVII, XX–XXIX.

¹⁷ K. Kopiński, *Eine Datenbank*, s. 70.

¹⁸ *Księga Theudenkusa*, hrsg. L. Koczy, Toruń 1937, s. XVIII, passim.

¹⁹ K. Kopiński, *Eine Datenbank*, s. 71.

w przyszłości. Po drugie, wspomniana unifikacja przynosi nową jakość. W jednej bazie danych będzie 15 ksiąg z określonego terenu, dzięki czemu uzyskujemy wiele możliwości wyszukiwawczych. Trud pracy włożony w rozwiązania metodyczne i programistyczne związane z prezentowanym projektem przynosi więc badaczom nowe możliwości analizy.

**“Prussian Books on the Internet” — An Electronic Publication of Sources Financed
by the Foundation for Polish Science**

Summary

The intention of three editorial projects realised in 2004-2007 in Torun was to prepare four town books for printing and electronic access. This was the period of an emergence of partly realised plans for building a prototype of a database application with all the source records contained in the books. It was also assumed that the database should be adapted to work on the Internet. In 2007 it was noticed that it would be particularly valuable to grant suitable impetus and a new direction to the up to then rather modest initiative. The Scientific Society in Torun successfully filed a motion within the PARTNERS Foundation for Polish Science programme. The project entitled “Prussian books on the Internet. A study on the methods of an electronic publication of sources” decided to work out an information module in which it would be possible to gather, process and render available the contents of mediaeval town books and books of Teutonic Order origin in accordance with accepted editorial principles. Within the scope of the project it was planned to scan 11 source publications issued in past years by the Scientific Society. Then, it was foreseen to carry out Optical Character Recognition, to be followed by indispensable verification and proofreading of the texts due to the possible occurrence of errors. The application was prepared in a Structured Query Language database with an opportunity for showing in Extensible Markup Language (XML). XML allows an exchange of data between bases, which means that we are not dependent on software and in the future could eventually perform data migration.