

Krzysztof Narojczyk, *Budowa badawczych baz danych w oparciu o historyczne źródła statystyczne*, Olsztyn 2005, Wydawnictwo LITTERA, s. 127 + CD-ROM SUND. Pakiet badawczy, v. 1.0

Praca Krzysztofa Narojczyka poświęcona budowie badawczych baz danych na podstawie historycznych źródeł statystycznych jest interesująca z wielu względów. W polskiej literaturze historycznej jest w zasadzie pierwszym tego typu studium. Autor stawia sobie niezwykle ambitne założenie: „W niniejszym opracowaniu podjęto próbę sformułowania na podstawie badań empirycznych uogólnionych podstaw teoretycznych budowy badawczych baz danych w oparciu o statystyczne źródła historyczne oraz określenia minimalnych standardów poprawności stosowania poszczególnych narzędzi i metod komputerowych” (s. 9); „Zasadniczym celem projektu było zbadanie komputerowego przetwarzania i udostępniania źródeł historycznych w postaci badawczych baz danych” (s. 119). Cytuję oba fragmenty, ponieważ definiują one temat pracy. Opublikowane rozważania dotyczą wyłącznie budowy baz danych w oparciu o źródła historyczne. Czytelnicy, którzy spodziewają się tam szerszych rozważań historycznych w oparciu o materiał użyty do budowy bazy danych, pozostaną zawiedzeni. Autor ogranicza się wyłącznie do podania postulatów, co można byłoby zrobić na podstawie uporządkowanych w ten sposób danych, niż do konkretnej ich analizy. Otrzymujemy w ten sposób bazę danych, dokładniej aplikację bazodanową, która ma być wykorzystana przez innych historyków w ich badaniach. Dlatego też interesujące jest szersze usytuowanie pracy Krzysztofa Narojczyka w polskim kontekście badawczym, poświęconym zagadnieniom zastosowania komputerów w badaniach historycznych.

W opublikowanej w 2000 r. książce poświęconej komputerowej analizie źródeł historycznych Rafał T. Prinke proponował transformację tekstu źródłowego w bazę danych¹. Etapem koniecznym takiego przekształcenia, jego zdaniem, powinna być edycja elektroniczna oryginału. U podstaw takiego przekonania leżał fakt, iż baza danych jest tylko swego rodzaju spojrzeniem na dane źródło i że spojrzeń takich może być wiele. Baza danych uzależniona jest od problemu, który stawia sobie badacz. W takiej koncepcji edycji źródła, „tekst staje się hybrydą – połączeniem tekstu tradycyjnego i wielu potencjalnych baz danych”². Takie podejście daje możliwości produkcji różnych instancji tego samego źró-

¹ R.T. Prinke, *Font ex machina. Komputerowa analiza źródeł historycznych*, Poznań 2000.

² Tamże, s. 22.

dła, przeznaczonych dla bardzo odmiennych odbiorców: naukowych, popularno-naukowych czy też popularnych.

Założenia przyjęte przez autora recenzowanej pracy są odmienne. Jego celem było stworzenie bazy danych z pominięciem edycji elektronicznej źródła. W takim ujęciu wydobywanie danych ze źródła dokonane jest w sposób tradycyjny, przez „ręczną” konwersję do postaci cyfrowej. Otrzymano w ten sposób arkusze kalkulacyjne Microsoft Excel, a nie bazę danych. Transformacja zawartych w nich danych została następnie dokonana przez początkową unifikację struktur danych z poszczególnych arkuszy, które później przeniesiono już automatycznie do zaprojektowanej struktury bazy.

Podstawą realizacji zadania są tabele cła sundzkiego (TCS) – źródło o charakterze masowym, będące rezultatem pracy duńskiej komory celnej w latach 1429–1857. Księgi opublikowane w pierwszej połowie XX w. stały się podstawą wielu prac poświęconych handlowi bałtyckiemu i europejskiemu w XVI–XVIII w.³ Wiele dyskutowano nad ich wiarygodnością i możliwościami interpretacji zawartych w nich danych⁴. Oceny tego rodzaju dokonywane były zdaniem autora z punktu widzenia statycznego i ich dynamicznie ujęcie może przynieść nowe spojrzenie na długofalowe procesy mające podłoże w stałych tendencjach gospodarczych. Ogrom podjętego zadania zobrazować może fakt, iż opublikowane w formie zestawień tabelarycznych księgi stanowią łącznie siedem tomów liczących ponad 3750 stron. Fakt ten usprawiedliwić może pominięcie edycji elektronicznej źródła, gdzie na jej sporządzenie trzeba byłoby poświęcić bardzo dużo energii.

Recenzowane opracowanie składa się ze wstępu, pięciu rozdziałów i zakończenia. Pierwszy z nich dotyczy rozwoju zastosowań elektronicznych baz danych w badaniach historycznych, drugi przedstawia założenia teoretyczne projektu badawczego. Następnie dokonana jest wstępna analiza tabel cła sundzkiego jako materiału źródłowego do badań nad budową badawczych baz danych. Rozdział czwarty przedstawia budowę modelu logiczno-konceptualnego, gdzie omówiony jest charakter źródła, jego kompletność i wewnętrzna struktura. Dużo uwagi autor poświęca tu również metodom rekonstrukcji brakujących danych, analizie kwestionariusza badawczego historyka, opracowaniu standardu opisu rekordu i określeniu charakterystyki wyszukiwawczej bazy danych. Rozdział piąty to etap praktyczny, budowa bazy w oparciu o poczynione założenia. Omówione są tam zagadnienia tworzenia struktury danych i interfejsu użytkownika, jak też budowa formularza bazy danych, projektowanie kwerend i obiektów dokumentacyjnych oraz narzędzia analizy danych.

Baza danych w pracy historyka jest narzędziem stosunkowo nowym. Dla tego celowe wydaje się przełożenie przedstawionych powyżej założeń na język

³ Na przykład M. Bogucka, *Handel zagraniczny Gdańska w pierwszej połowie XVII wieku*, Wrocław 1970; A. Mączak, *Miedzy Gdańskiem a Sundem. Studia nad handlem bałtyckim od połowy XVI do połowy XVII wieku*, Warszawa 1972.

⁴ Zob. A. Mączak, *Die Sundzollregister als eine preisgeschichtliche Quelle, 1557 bis 1647*, „Jahrbuch für Wirtschaftsgeschichte” 3, 1970, nr 3, s. 179–220.

zrozumiały dla wszystkich. Dążeniem autora było stworzenie aplikacji bazodanowej odzwierciedlającej źródło, jakim są tabele cła sundzkiego. Dane tam się znajdujące zostały przeniesione do arkuszy kalkulacyjnych, a następnie już automatycznie do tabel, z zachowaniem możliwie najdokładniej ich oryginalnej struktury. Powstałe tabele, powiązane przez system relacji, pozwoliły stworzyć strukturę hierarchiczną i nadać danym spójność. Pozostawienie konstrukcji bazy na tym etapie, zdaniem twórcy bazy, jest jednak nieuzasadnione, ponieważ skorzystać z niej potrafiłby jedynie specjalista znający zasady języka SQL. Dlatego też wyposażono ją w interfejs użytkownika, zawierający narzędzia pozwalające na konsultację, edycję i analizę zawartości bazy. Przyjęto zasadę, że interfejs tego typu powinien zapewnić możliwość pracy zarówno dla badacza specjalisty, jak też dla okazjonalnego użytkownika. Chęć udostępnienia tworzonej bazy każdemu wiązała się też z faktem zastosowania aplikacji umożliwiającej wygenerowanie niezależnego pliku typu exe, który mógłby być uruchamiany samodzielnie, już bez obecności programu użytego do jego stworzenia. Zastosowano w tym celu program Visual FoxPro w wersji 8.0 firmy Microsoft. Innym powodem użycia tej aplikacji był także fakt, iż dawał on możliwości automatycznej konwersji arkuszy kalkulacyjnych w formacie Excel do formatu bazy danych. Otrzymujemy w ten sposób aplikację składającą się z systemu danych, posiadający nadbudowę w postaci interfejsu użytkownika. Elementy te znajdują swoje połączenie w wygenerowanym pliku exe aplikacji. Związek między nimi stanowi siłę, jak też zasadniczą słabość programu. Zaletą będzie tu jego dostępność dla wszystkich, możliwość zadawania pytań i edycji danych. Wadą jest na pewno brak przejrzystości, nie możemy zajrzeć do środka i przyrzec się dokładniej jego funkcjonowaniu. Najpoważniejszym problemem wydaje się być jego brak elastyczności, wynikający z wymogów, jakie postawiono interfejsowi. Powinien on być dostępny każdemu użytkownikowi, nieposiadającemu praktycznie żadnej wiedzy informatycznej. Wbudowanie narzędzi analizy, chęć stworzenia możliwości porównania bardzo długich szeregów statystycznych, wpłynęły w bardzo dużym stopniu na przygotowanie podstawy aplikacji, jakimi są dane. W tym też chyba dopatrywać się trzeba roli zagadnień związanych z rekonstrukcją brakujących danych, którym w pracy poświęcono bardzo wiele miejsca.

Spójrzmy skróto na proces konstrukcji bazy danych. Przed jej stworzeniem powinniśmy wprowadzić rozróżnienie między danymi a informacją. Pomocze to nam odpowiedzieć na pytanie, co powinno się w niej znaleźć. **Dane** są pewnymi wartościami statycznymi, niezmiennymi, zebranymi w tabelach bazy, **informacja** to dane już wybrane i posegregowane, w taki sposób, aby coś znaczyły dla ich użytkownika. W tabelach gromadzimy dane, na podstawie których otrzymujemy informację, czyli odpowiedzi na stawiane pytania. Dane to wartości elementarne⁵, niezmiennicze, zgromadzone i usystematyzowane w tabelach bazy. Do informacji zaliczymy wszystko to, co otrzymamy na podstawie analizy danych. Jeżeli zsumujemy dane elementarne, będzie to już

⁵ R.T. Prinke, dz. cyt., s. 21.

informacja. Informacje otrzymane w ten sposób nie powinny znaleźć się na nowo w bazie danych. Jeżeli mamy dane niepełne, które się urywają, nie pozwalają na stworzenie ciągłych szeregów statystycznych, to pozostawiamy je w tej formie, informując o tym przyszłego użytkownika. Ich ewentualne uzupełnienie dokonać się może według różnych metod, w zależności od potrzeb. Mówiąc to, dotykamy niezwykle istotnego problemu dla naszego autora bazy danych, ponieważ rozróżnienie między danymi a informacją nie zostało wprowadzone. Autor pisze o przekonwertowaniu przestrzeni informacyjnej źródła, jak też o wzbogaceniu jej przez nowe, niewystępujące dotychczas elementy (s. 20.) Dalej w rozdziale poświęconym opracowaniu standardu opisu rekordu znajdujemy wzmianki o tym, że dane jednostkowe wchodzące w skład poszczególnych terytoriów zostały zsumowane i tak zagregowane wprowadzone do tabel. Zabiegi te spowodowały zwiększenie o ponad 55% liczby informacji jednostkowych przechowywanych w tabelach. Dzięki temu otrzymano możliwość ich bezpośredniego analizowania i porównywania (s. 67). Wprowadzenie danych zagregowanych na takim poziomie zmodyfikowało na pewno w tym samym stopniu strukturę danych, ponieważ ich transformacja była wynikiem pewnej ich analizy i interpretacji. Mamy więc tu już do czynienia z informacją, która została wprowadzona do bazy jako element podstawowy. Jednym z problemów jest uzupełnienie brakujących danych dla zachowania wewnętrznej ciągłości struktury danych. Nie mamy żadnych informacji o tym, które dane w nich zostały dodane i w jaki sposób. Agregacja i uzupełnienie brakujących danych było niewątpliwie konieczne dla prawidłowego funkcjonowania aplikacji, autor publikacji nie udziela nam jednak tutaj wystarczających odpowiedzi.

Na pewno trzeba dążyć do zachowania całej przestrzeni danych źródła pierwotnego. Czy trzeba jednak wzbogacić przestrzeń informacyjną źródła o nowe, niewystępujące dotychczas elementy? Jak też czy trzeba koniecznie ją uzupełnić o dane brakujące? Jakie są konsekwencje takiego uzupełnienia? Odpowiedź na to pytanie znajduje się w arkuszach kalkulacyjnych sporządzonych na podstawie źródła. Dlatego też trzeba zaapelować do autora o udostępnienie powstałych arkuszy kalkulacyjnych, służących za podstawę aplikacji. Należałoby opublikować arkusze, które posłużyły do unifikacji danych, ich normalizacji, jak też te, które otrzymano po przeprowadzeniu tychże działań i które były podstawą do przeniesienia danych już do bazy. Dałoby to okazję każdemu, posiadającemu takie zdolności, powtórzenia dokonanych czynności. Bardziej interesująca byłaby jednak możliwość modyfikacji tego procesu, co dawałoby tym samym inne spojrzenie, jak też propozycje rozwiązania napotykanych problemów. Zasady takiego postępowania znane są pod nazwą programów typu *open source* i licencji z nimi związanych. Udostępniają one kod źródłowy aplikacji, dając użytkownikowi pozwolenie jego zmiany na własny użytek. Arkusze kalkulacyjne rozumiem tutaj jako kod źródłowy bazy danych. Modyfikacja według potrzeb przez inne osoby nie odbiera praw autorskich ich twórcy. Aplikacja bazodanowa to bardzo dobry pomysł, ale jeszcze lepiej, aby można było pracować ze wszystkimi źródłami służącymi do jej powstania.

Dlatego też w przypadku pracy trzeba mówić raczej o aplikacji bazodanowej niż o bazie danych, co w niczym nie odbiera jej wartości poznawczych. Wygenerowanie bazy w formie tego typu, dostępnej wszystkim, miało jednak wpływ na jej przejrzystość i transformację danych, bez której nie mogłaby ona prawidłowo funkcjonować.

Historykom posiadającym już pewne doświadczenie informatyczne na pewno potrzebne jest coś więcej niż skompilowany plik programowy. W chwili obecnej powtarzalność rezultatów otrzymanych w danej dziedzinie badań jest bardzo istotna. Nie wystarczy tutaj możliwość eksportu danych do innych programów, trzeba jeszcze wiedzieć, co się eksportuje i jak te dane się mają do danych źródłowych, czy są one wynikiem jakiejś analizy, interpolacji itp. Nie znajdziemy w pracy szerszej refleksji na temat tworzenia baz danych współczesnych i opartych na źródłach historycznych. Aplikacje bazodanowe stworzone dzisiaj rozwiązują zadania, na które chcemy uzyskać odpowiedź w danej chwili. Dokonuje się to przez zaplanowanie wszystkich zadań, które mają być rozwiązane, a wszystkie elementy, które do tego posłużą, są nam dostępne. Źródło historyczne nie daje nam takiej okazji. Czym więc jest baza danych sporządzona na tej podstawie? Jakie są różnice przy tworzeniu i jej wykorzystaniu? Jakie są też ich cechy wspólne wynikające z teorii baz danych?

Autor ogranicza się również wyłącznie do konstrukcji, nie proponując żadnej analizy danych zawartych w aplikacji. Mamy tutaj raczej listę tego, co można byłoby zrobić na jej podstawie. Wielka szkoda, bo czas i energia poświęcona na analizę tego bardzo interesującego źródła mogłaby tu niewątpliwie być bardzo pomocna w konkretnym uzasadnieniu tej pracy. Chodziłoby o wybór chociażby jednego skromnego tematu, który dogłębniej zobrazowałby przedstawione postulaty. Przykład taki byłby doskonały dla tych, którzy zastanawiają się, czy metody komputerowe mogą być pomocne w badaniach historycznych.

W tym miejscu trzeba dodać, iż do opracowania dołączony jest CD-ROM, zawierający rezultaty przyjętych założeń w postaci skompilowanego pliku aplikacji bazodanowej, mogącej funkcjonować samodzielnie na każdym komputerze, nieposiadającym programu użytego do jego stworzenia. Jest to na pewno bardzo cenny dodatek do tej niezwykle interesującej pracy nad bazami danych i zastosowaniem komputerów w badaniach historycznych.

Marek Mrówczyński