

Krzysztof Narojczyk

"Fontes ex machina: komputerowa analiza źródeł historycznych", Rafał T. Prinke, Poznań 2000 : [recenzja]

Echa Przeszłości 3, 253-258

2002

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

Rafał T. Prinke, *Fontes ex machina. Komputerowa analiza źródeł historycznych*, Poznań 2000, ss. 342.

Idea wykorzystywania metod komputerowych w naukach humanistycznych nikogo już chyba dzisiaj nie dziwi i znajduje prawo pełnego obywatelstwa w świadomości coraz większej liczby badaczy. Dotyczy to także środowiska historycznego. Przy Zarządzie Głównym Polskiego Towarzystwa Historycznego powołano Komisję Metod Komputerowych. Powstał także Polski Oddział Association for History and Computing (AHC) – międzynarodowej organizacji skupiającej historyków wykorzystujących w pracy naukowej metody komputerowe. Począwszy od 1994 r. corocznie odbywają się ogólnopolskie sympozja metod komputerowych w badaniu i nauczaniu historii¹. W 2001 r. Polska była gospodarzem XV Międzynarodowej Konferencji AHC. Owa aktywność środowiska historycznego zaowocowała ponad stu artykułami i opracowaniami naukowymi, publikowanymi w różnych periodykach ogólnohistorycznych i specjalistycznej serii wydawniczej *Historia i Komputery*. Prace te traktowały o różnych, z reguły dość wąsko zakreślonych aspektach stosowania metod komputerowych w nauce historycznej. Wyraźnie brakowało prac o charakterze monograficznym, ujmujących tę problematykę w sposób kompleksowy i wyczerpujący.

Łukę tę w znacznej mierze wypełnia wydana w roku 2000 obszerna, oparta na znakomitej, głównie anglojęzycznej bazie źródłowej, praca Rafała T. Prinke pt. *Fontes ex machina. Komputerowa analiza źródeł historycznych*². Autor jest doświadczonym badaczem od wielu lat zajmującym się metodami komputerowymi tak od strony teoretycznej, jak i praktycznej. Uczestniczył w zakrojonych na bardzo dużą skalę projektach elektronicznej edycji źródeł historycznych prowadzonych przez Bibliotekę Kórnicką PAN i Centrum Elektronicznych Tekstów Humanistycznych PAN. Książka zgodnie ze swoim podtytułem traktuje o samej istocie pracy badawczej historyka – krytyce i analizie źródeł historycznych dokonywanej za pomocą metod komputerowych. Składa się z pięciu zasadniczych części-rozdziałów oraz wprowadzenia i aneksów. We wprowadzeniu – *Historyk i komputery* autor wykazuje celowość stosowania technologii cyfrowej w badaniach historycznych i na umiejętnie

¹ Interesujące uwagi na temat polskiego środowiska historyków posługujących się w pracy badawczej technologią komputerową formułuje Michał Kopeczyński, *Przyszłość technik i metod komputerowych w badaniach historycznych. Uwagi przechodnia*, w: *Pamiętniki XVI Powszechnego Zjazdu Historyków Polskich*, t. III, cz. 1, Toruń 2001, s. 183-191.

² Przegląd stanu badań prowadzonych w Polsce przedstawił w 1999 r. Andrzej Wałkowski, *Zastosowanie komputerów w źródłoznawstwie historycznym*, w: *Pamiętniki XVI Powszechnego Zjazdu Historyków Polskich*, t. III, cz. 1, Toruń 2001, s. 217-233.

dobrych przykładach obala najczęściej powielane wśród przeciwników komputeryzacji nauk humanistycznych mity o bardzo ograniczonej przydatności nowych technologii dla procesu badawczego w tych naukach.

Część pierwsza *Techniki i metody komputerowe* ma charakter wprowadzający. Autor definiuje w nim podstawowe pojęcia wyraźnie rozgraniczając techniki komputerowe – wykorzystanie komputera jedynie jako wydajnego narzędzia w tradycyjnych metodach badawczych od metod komputerowych – nowe algorytmy postępowania badawczego wypracowane w oparciu o elektroniczne możliwości reprezentacji i analizy wiedzy. Do najczęściej wykorzystywanych przez humanistów narzędzi komputerowych należą dwie podstawowe grupy programów biurowych: edytory tekstów i bazy danych wraz z ich specyficzną odmianą – arkuszami kalkulacyjnymi. W wyniku posługiwania się nimi badacz uzyskuje teksty – ciąg symboli (znaków) pisarskich lub kartotekę, katalog, tabelę ew. inne tego typu zestawienie o wysokim stopniu wewnętrznej strukturalizacji. Z punktu widzenia metod komputerowych niezmiernie istotna jest transformacja tekstu źródła w bazę danych. Wiedza zawarta w źródle musi zostać zeń wydobyta i usystematyzowana, aby poddawać się opracowywaniu metodami komputerowymi. Stwarza to szereg trudnych problemów metodologicznych. Wynikają one głównie z konieczności pewnej normalizacji i standaryzacji procedury transformacji źródeł, co jest niezmiernie trudne wobec ich ogromnej różnorodności i nierzadko unikatowego charakteru. Pomyślnie rozwiązanie tych problemów jest warunkiem niezbędnym uzyskania wiarygodnej bazy, zawierającej wiedzę źródłową, której można stawiać pytania badawcze i uzyskiwać poprawne odpowiedzi. Z kolei zastosowanie w trakcie całego procesu transformacji źródeł i postępowania badawczego jednolitych standardów informatycznych gwarantuje powtarzalność uzyskiwanych wyników oraz możliwość wymiany danych pomiędzy badaczami, co z kolei jest niezbędnym warunkiem niezależnej weryfikacji. Podstawowe elementy tak zarysowanego modelu komputerowej analizy źródeł historycznych znajdują odzwierciedlenie w strukturze dalszych części monografii.

Część druga – najobszerniejsza – traktuje o tekście jako o nośniku wiedzy źródłowej. Autor z dużym znanstwem rozważa rozmaite klasyfikacje źródeł, proponując na użytek budowy komputerowych baz danych przyjęcie typologii ich podziału na: źródła informatycznie efektywne i źródła informatycznie potencjalne. Rozszerza także tradycyjne definicje tekstu, wiążąc go z warstwą informacyjną źródła. W tym ujęciu tekst jest informacją, która może przybierać różnorodne formy reprezentacji (dźwiękowe, graficzne czy – w wypadku technologii cyfrowych określonych przebiegów i stanu ładunków prądu elekt-

rycznego). Dla stosowania metod komputerowych istotna jest elektroniczna reprezentacja tekstu, czyli przypisania jednostek podstawowych (atomów) tekstu wyrażonych za pomocą pojedynczych znaków pisarskich określonym kodom komputerowym (układom ładunków elektrycznych w układach elektronicznych). Kod taki może być następnie ponownie przekształcony do postaci tekstu i zaprezentowany za pomocą grafiki lub dźwięku. Powszechnie stosowany jest dziś amerykański standard ASCII. Składał się on początkowo ze 128 kodów, co – jak się szybko okazało – nie wystarczało do obsługi liter diakrytycznych alfabetów narodowych. Liczbę kodów zwiększono wprawdzie do 256, czyli do maksymalnej pojemności stosowanego wówczas 8-bitowego systemu opisu strony kodowej, ale okazało się to niewystarczające. Zaczęto stosować różne strony kodowe dla różnych zestawów znaków narodowych, co utrudnia jednak wymianę danych zapisanych w różnych alfabetach, bowiem część liter z tekstu źródłowego zastępowana jest przez znaki aktualnie używanej przez oprogramowanie strony kodowej. Ostatecznym rozwiązaniem tego problemu ma być opracowanie przez międzynarodowe konsorcjum (The Unicode Consortium), skupiające największe koncerny z branży informatycznej oraz organizacje i towarzystwa naukowe, uniwersalnego 16-bitowego standardu kodowania znaków – Unicode. W założeniu, obejmie on wszystkie znaki pisarskie we wszystkich językach świata³. Alternatywnym sposobem zapisu znaków w tekście elektronicznym jest system mnemoników stosowany np. w standardach opisu stron WWW. Tekst elektroniczny w czystej postaci (wyrażony np. jednym z powyższych standardów) nie zawiera żadnego klucza semantycznego. Treść taką zawiera dopiero tekst adnotowany. Autor wyczerpująco, ilustrując wieloma trafnie dobranymi przykładami, omawia podstawowe standardy adnotowania tekstu zarówno w ujęciu proceduralnym (TeX, PDF, PostScript, RTF), jak i deklaratywnym (np. COCOA, TUSTEP, polski -KORONA). Analizuje także, będący dziś światowym standardem, metajęzyk SGML, określający reguły tworzenia języków adnotowania oraz stanowiący najważniejszą jego aplikację – język adnotowania hipertekstu HTML. Ostatni podrozdział części drugiej przedstawia pomnikowe dzieło interdyscyplinarnej współpracy międzynarodowej – format kodowania tekstów humanistycznych znany jako Text Encoding Initiative (TEI). Tzw. *Wskazówki* TEI oraz polskie instrukcje wydawnicze powinny,

³ W styczniu 2002 r. opublikowano w wersji beta kolejne rozwinięcie tego standardu oznaczone jako wersja 3.2. W stosunku do wersji poprzedniej rozszerzono go m.in. o kody 1016 nowych znaków graficznych: liter – głównie z alfabetów używanych w językach występujących na Filipinach oraz obszernego zestawu symboli matematycznych, <<http://www.unicode.org/versions/beta.html>>, Styczeń 2002.

zdaniem autora, stanowią podstawę wypracowania standardów dla edytorstwa elektronicznego w naszym kraju.

Część trzecia monografii traktuje o zbiorach danych, stanowiących reprezentację wiedzy źródłowej i związanych z tym problemach metodologicznych i praktycznych. Do zbiorów takich autor zalicza omówione w rozdziale poprzednim edycje tekstów źródłowych oraz bazy danych. W pierwszym wypadku wydawca dąży do identyfikacji i wyodrębnienia możliwie największej liczby cech tekstu, odzwierciedlających zawartą w nim wiedzę źródłową, w drugim zaś projektant bazy stara się wyselekcjonować część wiedzy źródłowej zgodnie z określonym wcześniej zapotrzebowaniem badawczym. W ujęciu takim autor opowiada się za używaniem tzw. węższego znaczenia pojęciowego terminu „baza danych”. Rozumie pod nim wyspecjalizowany do określonych zadań badawczych system informacyjny, operujący najczęściej na niepełnym, ściśle wyselekcjonowanym zasobie danych o uproszczonych zarówno strukturze wewnętrznej, jak i wzajemnych związkach i zależnościach między jego poszczególnymi elementami. Podejście takie pozostaje w zgodzie z aktualną sytuacją na rynku oprogramowania i aplikacji bazodanowych. Dominują na nim systemy informacyjne czwartej generacji – relacyjne bazy danych. Informacje w tych systemach przechowywane są w powiązanych relacjami tablicach, które mają strukturę tabelaryczną. Bazy relacyjne operują na zaledwie kilkunastu prostych typach danych, zapisanych w ściśle określonym formacie pól (integer, string, date, general, memo). Taki model danych, nawet przy bardzo zaawansowanym systemie relacji nie zawsze jest w stanie odzwierciedlić pełną wiedzę źródłową tekstu, co zmusza do upraszczania (spłaszczania do postaci dwuwymiarowej tablicy) struktury danych. Nie można wyrazić np. takich pojęć semantycznych jak agregacja, hierarchia, generalizacja. Wydaje się, że rozwój zastosowań systemów informacyjnych piątej generacji – obiektowych baz danych – wymusi w przyszłości ponowne zdefiniowanie podstawowych typów zbiorów, stanowiących reprezentację wiedzy źródłowej. Obiektowy model danych łączy bowiem zalety baz relacyjnych z możliwością deklarowania i operowania na dowolnych typach danych wraz z pojęciami semantycznymi oraz reprezentowania narastających w czasie zmian w bazie, jak np. wersje obiektów. Można przypuszczać, że dadzą się nim opisać i wyrazić także, traktowane dziś nieco odrębnie, genealogiczne modele danych.

Istotą badania historycznego jest analiza informacji źródłowych. Problemy z tym związane omówione zostały w części czwartej. Każdą z podstawowych faz procesu badawczego można opisać jako ciąg składających się nań prostszych czynności, wykonywanych w określonym porządku bądź

cyklach, te zaś można dalej rozpisać na następne czynności itd. aż do poziomu elementarnego. Ogólny opis osiągnięcia określonego rezultatu w skończonej liczbie kroków–czynności nosi nazwę algorytmu. Może być on sporządzony w formie narracyjnej, jednak zdecydowanie większą przejrzystość i czytelność zapewniają diagramy, używające sformalizowanych symboli graficznych dla określenia poszczególnych kroków i elementów postępowania. Autor stawia trafny postulat publikowania algorytmów wykorzystanych przez historyków w analizach komputerowych (z wyjątkiem programów ogólnie znanych i powszechnie stosowanych). Da to możliwość środowiskowej dyskusji nad nimi w celu weryfikacji poprawności logicznej i metodologicznej, oceny wydajności oraz ew. modyfikacji lub wyboru najlepszego rozwiązania z kilku dostępnych. Spośród wielu języków programowania autor koncentruje się na tych mniej znanych, a szczególnie przydatnych do przetwarzania tekstów jak np. Snobol, Icon, Perl. Interesujący jest także przegląd najpopularniejszych algorytmów rozpoznawania nazwisk opartych na metodach porównywania brzmienia (fonetyczne) i zapisu graficznego (ortografii).

Ostatnia, piąta część opracowania zawiera podsumowanie wcześniejszych rozważań nad modelem stosowania metod komputerowych w analizie źródeł historycznych. Ostateczna konkluzja stwierdza, że wyniki postępowania badawczego, prowadzonego za pomocą technologii informatycznej, muszą być: tworzone w kodzie czytelnym dla innych użytkowników, gromadzone w cyfrowych bibliotekach, udostępniane w sieciach o zasięgu globalnym oraz wykorzystywane za pomocą algorytmów.

Niezwykle cennym uzupełnieniem opracowania są dwa aneksy. W pierwszym przedstawiono polskie tłumaczenie *Wskazówek* dla naukowych edycji elektronicznych, opracowanych w 1997 r. przez Modern Language Association of America. Stanowią one doskonałą bazę teoretyczną i punkt odniesienia do rozważań nad stworzeniem podobnych zasad dla polskich wydawców. Aneks drugi to bardzo obszerne i umiejętnie wyselekcjonowane zestawienie dostępnej literatury przedmiotu. Wskazówki bibliograficzne, będące pierwszym tak reprezentatywnym opracowaniem w Polsce, znakomicie ułatwiają orientację w światowym i polskim piśmiennictwie, poświęconym szeroko rozumianej komputeryzacji nauki historycznej.

Monografia *Fontes ex machina* ze względu na poruszaną problematykę badawczą stanowi bezsprzecznie na rynku polskim edycję pionierską. Równocześnie przedstawiona w tytule tematyka postawiła autora przed niezwykle trudnym zadaniem. Dla wielu historyków problemy komputeryzacji ich warsztatu badawczego pozostają jeszcze w sferze abstrakcji. Próba uczynienia monografii zrozumiałą także dla nich stwarza określone implikacje

w zakresie konstrukcji pracy. Konieczne jest umiejętne pogodzenie zasadniczego wątku relacjonującego i podsumowującego określony etap badań naukowych autora z opisami podstawowych standardów, terminów i pojęć informatycznych. Z zadania tego Rafał T. Prinke wywiązał się znakomicie, ilustrując tekst umiejętnie dobranymi przykładami i licznymi diagramami, ułatwiającymi zrozumienie trudnych niejednokrotnie dla niespecjalistów zagadnień. Stąd jego opracowanie – obok zasadniczej funkcji monografii naukowej – z powodzeniem może pełnić dla pewnych grup czytelników także rolę poradnika, specjalistycznego leksykonu czy też podręcznika, systematyzującego określoną gałąź wiedzy. Znajdzie tu coś dla siebie zarówno historyk doświadczony w stosowaniu technologii cyfrowej, pragnący np. przypomnieć sobie podstawy pracy z TEI, jak i początkujący, szukający wyjaśnienia trudniejszych pojęć i terminów. Od technologii cyfrowej nie ma odwrotu. Metody komputerowe odgrywać będą obiektywnie coraz większą rolę w postępie nauki historycznej. Wymaga to już dzisiaj dostosowania do tych procesów systemu kształcenia studentów historii⁴. Dlatego *Fontes ex machina* (ew. wybrane fragmenty) można polecić jako lekturę w ramach realizacji akademickiego przedmiotu „nauki pomocnicze historii”.

Krzysztof Narojczyk

**„Obwód Kaliningradzki. Przegląd faktów, wydarzeń, opinii”.
Biuletyn Towarzystwa Naukowego i Ośrodka Badań Naukowych
im. Wojciecha Kętrzyńskiego w Olsztynie. Olsztyn, wrzesień 1994
– sierpień 2002 (96 numerów).**

Na kształt stosunków i współpracy transgranicznej i przygranicznej województw północno-wschodniej Polski z Obwodem Kaliningradzkim RFSRR po 1990 r. miały wpływ zarówno przekształcenia ustrojowe w Polsce, jak i działania reformatorskie w ZSRR, umownie określane jako „pierestrojka”.

Zanim Prezydium Rady Najwyższej ZSRR anulowało status Obwodu Kaliningradzkiego jako obwodu zamkniętego, niedostępnego dla cudzoziem-

⁴ Jako pierwszy w 1993 r. projekt programu informatycznego kształcenia historyków zgłosił Michał Kopczyński, *O możliwościach i nieodzownej potrzebie zastosowania w naukach historycznych maszyny zwanej komputerem*, w: *Tradycje i perspektywy nauk pomocniczych historii w Polsce*, pod red. M. Rokosza, Kraków 1995, s. 264-271. Po kilku latach przedstawił zmodyfikowaną wersję swoich postulatów programowych, *Przyszłość technik ...*, s. 190-191.