

Dr hab. Bartłomiej Szleszyński, prof. IBL PAN

Recenzja pracy Jakuba Eichlera *Dane Polskiej Bibliografii Literackiej jako dane badawcze. Perspektywa humanistyki cyfrowej i studium przypadku*.

### Kwestie ogólne

Praca Jakuba Eichlera zatytułowana jest *Dane Polskiej Bibliografii Literackiej jako dane badawcze. Perspektywa humanistyki cyfrowej i studium przypadku*. I już w przypadku tytułu napotykamy pierwszy problem tej w ogóle bardzo problematycznej pracy.

Otóż w istocie praca nie jest w większości poświęcona ani Polskiej Bibliografii Literackiej (nie wnosi także nic nowego do wiedzy o niej), ani zagadnieniu danych badawczych, nie rozwija też ani satysfakcjonująco nie referuje istotnych zagadnień humanistyki cyfrowej - ani nie zdaje satysfakcjonująco sprawy z aktualnego stanu badań, ani też nie pokazuje aktualnego rozwoju cyfrowych projektów, narzędzi czy infrastruktur.

Niełatwo się zorientować, o czym w istocie jest (czy ma być) ta praca, co stanowi jej temat przewodni. Nie tylko tytuł jest w tej kwestii zwodniczy - nie zostało to także jasno wyłożone we wstępie pracy, a deklaracje, które można gdzieś odnaleźć nie znajdują potwierdzenia w dalszym jej toku. Nieprawdziwe jest bowiem stwierdzenie otwierające wstęp, iż „Praca koncentruje się na przedstawieniu zasobów Polskiej Bibliografii Literackiej – jako źródła danych badawczych.” – na ten temat jest maksymalnie półtora rozdziału, zresztą odnosząc się do tematu w wysoce niesatysfakcjonujący sposób.

Praca nie spełnia zatem kryterium oryginalnego rozwiązania problemu naukowego, bo w istocie nie odnosi się do problemu, który deklaratywnie sama sobie stawia

Drugim problemem pracy jest jej struktura narracyjna, a właściwie jej brak. Wywód prowadzony przez doktoranta jest chaotyczny i nieuporządkowany, tytuły podrozdziałów często nie odpowiadają zawartości. Kolejne rozdziały nie łączą się ze sobą, podrozdziały następują po sobie często nie tworząc wyraźniej narracji. Autor dorzuca także w wielu miejscach dygresyjne fragmenty o zagadnieniach, które z bibliograficznymi danymi badawczymi się niespecjalnie łączą (jak kwestie topograficzne czy edycje cyfrowe), lub w ogóle nie dotyczą tego tematu (jak obszernie streszczane publikacje o przyszłości Internetu z lat 90 i 00). W efekcie zaproponowana w pracy narracja nie zachowuje wymogów czytelności i spójności oraz nie spełnia kryterium naukowości.

Trzecim problemem jest warstwa językowa. To oczywiste, że w każdej pracy zdarzają się potknięcia, jednak to co pod tym względem dzieje się w pracy p. Jakuba Eichlera przekracza granice akceptowalności. Jest ona napisana w sposób niezwykle wręcz niechlujny, obfituje w stwierdzenia potoczne, anglicyzmy błędy, literówki. Często występują w pracy sformułowania odwołujące się do subiektywnych odczuć lub (braku) wiedzy autora, takie jak: „W moim

osobistym odczuciu” (s. 16), „sądzę” (s. 57, 121, 205, 245), „Moim zdaniem” (s. 162), „o ile mi wiadomo” (s. 118), które nie powinny znaleźć się w rozprawie naukowej. Ogólne złe wrażenie jeśli chodzi o stronę formalno-językową niejako wieńczy podanie w przypisie 425 błędnego linku do zdeponowanych na Zenodo danych, których wytworzenie opisane jest w rozdziale czwartym i które stanowią właściwie jedyny wartościowy efekt pracy. Problem ten kieruje uwagę czytelnika na kolejny, **czwarty** z problemów pracy, którego waga jest już o wiele większa.

Nie sposób bowiem nie zauważyć problemów z operowaniem w pracy słowem cudzym. Praca obfituje w absurdalnie obszerne streszczenia książek i artykułów, które nie są w pracy zinstrumentalizowane dla wsparcia wyводу prowadzonego przez autora, lecz połączone ze sobą na zasadzie kolażu, bez wyraźnej myśli przewodniej.

Wyrazisty przykład stanowią fragmenty rozdziału 1. Strony od 27 do 37 to w większości bezrefleksyjne streszczenie prac Manovitcha (z początku lat 2000), przechodzące w trzystronicowe streszczenie dokumentu *The Promise of Digital Humanities* z 2009 roku, potem czterostronicowe streszczenie tekstu Filimonowicza, by po krótkiej przerwie, tym razem od strony 46 zacząć streszczać (klasyczne i powszechnie dziś znane) prace Morettiego, co trwa nieprzerwanie do strony 53. To 25 stron pod rząd wypełnianych w zdecydowanej większości zupełnie niesfunkcjonalizowanym streszczeniem innych tekstów, niebyt zresztą aktualnych. Podobne zagęszczone fragmenty „streszczeniowe” pojawiają się w pracy wielokrotnie.

Podobnie bywa z niezwykle obszernymi cytatami, przywoływanymi bez komentarza autora, przeważnie też bez wyciągnięcia z nich wniosków czy wyrazistego zestawienia z innymi cytatami na podobny temat.

Dokładniejsza analiza problemu czwartego prowadzi jeszcze głębiej. Doktorant nie tylko nie potrafi się umiejętnie posługiwać słowem cudzym ale też błędnie używa aparatu naukowego – to problem **numer pięć**.

W niezmiernie wielu fragmentach błędnie i nieumiejętnie użyto odesłań do tekstów naukowych.

W pracy znajdują się liczne przykłady niepoprawnych przypisów - przywołajmy te o numerze 188 i 355. Pierwszy brzmi „Tolonen, Mikko i Umerle, Tomasz” i odnosi się do cytatu, drugi „Por.: Drabek” i odnosi się do stwierdzenia autorskiego - domniemywać można, że chodzi o tekst Tolonena i Umerlego w przypadku pierwszym, a Drabek a przypadku drugim, ale wtedy należałoby podać jego tytuł i lokalizację cytatu w tekście. Tego typu uchybienia, powtarzające się wielokrotnie, są zasadniczo nieakceptowalne na poziomie pracy doktorskiej – tym bardziej, traktującej o danych bibliograficznych, które wymagają niezwyklej precyzji.

Natrafimy także na liczne fragmenty omówień cudzych tekstów, gdzie odniesienie do konkretnej pracy dostajemy jedynie raz, na samym początku fragmentu, a następnie, choć autor omawia cudze teksty, zaś odniesień do konkretnych fragmentów i cytatów brak – jak na s. 95:

„zawiera dobre praktyki w kontekście metadanych jako danych badawczych. Główne wskazówki dotyczą: stałych identyfikatorów, słów kluczowych, abstraktów i cytowań. Pod

względem stałych identyfikatorów wskazuje na różnorodność źródeł i potrzebę konsolidacji właśnie za ich pomocą (muszą być możliwe do odczytania przez maszyny). Kolejnym aspektem jest wielojęzyczność (obecna w międzynarodowym środowisku naukowym), a także różnorodność, nazwana bibliodiversity: różne organizacje tworzą i przetwarzają dane, robiąc to według własnych metodologii i kryteriów. Stąd powstające ontologie powinny uwzględniać łączliwość z tymi systemami. Występuje także potrzeba agregacji mniejszych źródeł i kompleksowe podejście do całości. Wiąże się to także z rozproszeniem tych „miejsc” produkcji i przechowywania danych produkowanych przez humanistyczne projekty: poziom analizy powinien uwzględniać tę wielość.”

W bibliografii pojawia się wiele ważnych tekstów które nie są omówione ani nawet wspomniane w pracy – sprawa jest o tyle istotna, że części owych publikacji w istocie stanowczo powinno w pracy zostać omówione ze względów merytorycznych (tak jak *Open Data for Humanists, A Pragmatic Guide*, Edmond i Tóth-Czifry) czy artykuł z „Tekstów Drugich” *Dane badawcze w literaturoznawstwie* Maryla, Błaszczynskiej, Szleszyńskiego i Umerlego. Podobnych przypadków znalazłem kilkanaście, przywołuję je w przypisie<sup>1</sup>. W najbardziej pozytywnej dla doktoranta wersji, można tę kwestię interpretować jako potwierdzenie braku znajomości reguł tworzenia aparatu naukowego, w wersji negatywnej, jako desperacką próbę pokazania znajomości tekstów, do których nie odwołał się w pracy, a powinien. Niezależnie od wyjaśnienia, tego rodzaju uchybienie jest w dwójnasób kompromitujące ze względu na temat pracy – wszakże traktuje ona o danych bibliograficznych!

Wymienione powyżej pięć problemów pracy poddaje w wątpliwość umiejętności pisania tekstów naukowych przez Jakuba Eichlera, co na poziomie doktoratu jest zarówno zaskakujące, jak i zasadniczo nieakceptowalne – tym bardziej w świetle kolejnych kryteriów oceny pracy doktorskiej jaki stanowi umiejętność prowadzenia samodzielnej pracy naukowej. Pracę uratować mogłaby w tej sytuacji jedynie świetna warstwa merytoryczna. Niestety – najdelikatniej powiedzieć można, także ona pozostawia wiele do życzenia.

<sup>1</sup> Umerle, Tomasz, Błaszczynska, Marta, Wnuk, Madgalena, Franczak, Mateusz, Stojanovski, Jadranka, Rosiński, Cezary, Wołczuk, Nikodem, i in. „TRIPLE Deliverable 8.5: Guidelines on the Research Data in the Humanities”, 2023-01-20.

<https://doi.org/10.5281/ZENODO.7568873>.

Jasiewicz, Justyna, i Michalina Stopnicka. „Data literacy. Kompetencje w zakresie korzystania z danych w odniesieniu do przemian komunikacji naukowej”. *Kultura Popularna* 4, nr 50 (2017-05-31): 36–49. <https://doi.org/10.5604/01.3001.0010.0042>.

Kulecka, Alicja, red. *Archiwistyka cyfrowa i nauki pomocnicze historii w edukacji archiwalnej. Problemy dydaktyki archiwistyki*. Warsaw University Press, 2016.

<https://doi.org/10.31338/uw.9788323521259>.

Bachmann, Christian A., i Stephanie Heimgartner. „Book - Material - Text (Download Book)”. *Book - Material - Text*, 2017-01-09, 1-78 Pages.

<https://doi.org/10.13154/BMT.1.2016.1-78>.

Bail, Christopher A. „The Cultural Environment: Measuring Culture with Big Data”. *Theory and Society* 43, nr 3–4 (lipiec 2014): 465–82. <https://doi.org/10.1007/s11186-014-9216-5>.

Borgman, Christine L. *Big Data, Little Data, No Data: Scholarship in the Networked World*. First MIT Press paperback edition. Cambridge, Massachusetts London, England: The MIT Press, 2016.

Filiciak, Mirosław. *Media, wersja beta: film i telewizja w czasach gier komputerowych i internetu*. Gdańsk: Wydawnictwo Naukowe Katedra, 2013.

Problemem **szóstym** - merytorycznym jest przyjmowana w pracy perspektywa, **siódmym** - łącząca się z szóstym - zasadnicza ignorancja jeżeli chodzi o dokonania humanistyki cyfrowej z ostatniej dekady w omawianych obszarach zagadnień.

W swojej perspektywie doktorant zdaje się spoglądać na temat, tak jakby humanistyka cyfrowa czy literaturoznawstwo cyfrowe dopiero raczkowały, jak gdyby nie powstały bardzo zaawansowane projekty i infrastruktury z dziedziny badań literackich i naukowe opracowania na ich temat, jak gdyby temat danych badawczych był dziewiczym terenem, gdzie trzeba ustalać podstawowe pojęcia. Jak gdyby w pisaniu o PBL-u mogło wystarczać streszczanie wyników badań z przed niemal dekady, przy jednoczesnym ignorowaniu publikacji i prac na ten temat z lat ostatnich.

Tymczasem - ani nie jesteśmy w momencie pionierskim dla tematu danych badawczych, ani też pionierskim dla pisania o cyfrowym literaturoznawstwie w Polsce, ani (już zupełnie) na pionierskim etapie tworzenia projektów literaturoznawczych. Na temat danych badawczych powstało w ostatnich latach wiele opracowań, które pominięte są w pracy. Dane badawcze to już bardzo dobrze wyodrębniona dziedzina Humanistyki Cyfrowej, ze swoją specyfiką praktyką, licznymi grupami badawczymi (takimi jak grupa robocza konsorcjum DARIAH Research Data Management), i tworzonymi przez nie szczegółowymi raportami, takimi jak choćby *Sustainable and FAIR Data Sharing in the Humanities: Recommendations of the ALLEA Working Group E-Humanities* z 2020 (autorzy: Harrower Natalie, Maryl Maciej, Timea Biro, Beat Immenhauser); *OPERAS-P Deliverable D6.5: Report on the Future of Scholarly Writing in SSH* (autorzy Maryl, Maciej, Marta Błaszczynska, Agnieszka Szulińska, Anna Buchner, Piotr Wciślik, Iva Melinščak Zlodi, Jadranka Stojanovski, i in.) z 2021, *Research Data Management for Arts and Humanities: Integrating Voices of the Community* z 2023 r. (autorzy: Erzsébet Tóth-Czifra, Marta Błaszczynska, Francesco Gelati, Femmy Admiraal, Mirjam Blümm, Erik Buelinckx, Vera Chiquet, i in.), czy teksty takie jak *Seeing Shapes in the Cloud: Perspectives from the Humanities* Michelle Doran z 2022, *The Trouble With Big Data* Jennifer Edmond i in. Z 2022 czy wspomniany już, a praktycznie zignorowany w pracy artykuł *Dane badawcze w literaturoznawstwie* z 2021

Niemalże w każdym obszarze, który wzmiankowany jest w pracy, uderza brak wiedzy autora, brak odniesień do oczywistych w tym kontekście tekstów i projektów, dojmujący tym bardziej, że powołuje się on w pracy na studia „Humanistyka Cyfrowa” w Instytucie Badań Literackich PAN, gdzie wiele z tych projektów i naukowych opracowań w ciągu ostatniej dekady powstało, a w ramach których osoby doktorskie stworzyły wiele dobrych, bardzo dobrych i wybitnych prac, dotyczących zagadnień, o których doktorant mówi.

Doktorant przywołuje, często dosyć przypadkowe teksty z lat 90' oraz 00', z których jedynie część dotyczy humanistyki cyfrowej, zaś wiele Internetu w ogóle. Znikoma zaś liczba tematu danych badawczych. Pomija przy tym kluczowe pozycje (zarówno zagraniczne, jak i polskie) z lat ubiegłych.

Przekleństwem pracy jest przywoływanie wiekowych zagranicznych tekstów, często teoretycznie projektujących rozwój badań cyfrowych, podczas gdy pod ręką są polskie publikacje dotyczące aktualnej sytuacji cyfrowego literaturoznawstwa, uwzględniające

rodzimy kontekst rozwoju dyscypliny oraz infrastruktur badawczych. Powrót do prehistorii Humanistyki Cyfrowej mógłby mieć ewentualnie sens dla zarysowania dziejów dyscypliny, ale połączony z pominięciem publikacji dokumentujących jej dynamiczny rozwój w ostatnich latach tegoż sensu jest pozbawiony. Można na przykład zapytać doktoranta, dlaczego, skoro tak obszernie i licznie przywoływane są teksty Macieja Maryla sprzed dekady, zaś pominięto najnowsze jego publikacje, oddające aktualny stan rozwoju Humanistyki Cyfrowej w Polsce? Pytanie tym bardziej zasadne, iż część z nich dotyczy interesującej deklaratywnie autora pracy tematu danych badawczych.

Najlepszą ilustracją powyższego problemu jest fakt, że fragment „Ku definicji humanistyki cyfrowej” zamykający rozdział 1, opiera się na tekstach Macieja Maryla i Marcina Werli sprzed dziesięciu lat. Dodajmy – to akurat te dziesięć lat, przez które Maryl budował imponujących rozmiarów Centrum Humanistyki Cyfrowej IBL PAN – publikując zresztą na ten temat kolejne teksty, podsumowujące zmianę jego perspektywy po kolejnych etapach działalności Centrum.

Powyższe uwagi prowadzą do negatywnej wersyfikacji kolejnego z kryteriów oceny pracy doktorskiej, jakim jest posiadanie przez doktoranta ogólnej wiedzy teoretycznej w dyscyplinie.

### **Uwagi szczegółowe**

W dalszej części recenzji prześlę kolejne fragmenty pracy, aby pokazać przykłady obrazujące wymienne powyżej problemy ogólne. Jest to wybór z fragmentów i stwierdzeń budzących szczególnie recenzenckie wątpliwości.

Problemy numer sześć i siedem dobrze obrazuje początek fragmentu *Digitalizacja literaturoznawstwa*. Autor stwierdza iż

„Aby przybliżyć, jak mogą funkcjonować metody cyfrowe w literaturoznawstwie, chciałbym zaproponować osobom czytelniczym opis tekstu naukowego, który nigdy nie powstał”. I dalej przedstawia pomysł na artykuł dotyczący potencjalnej cyfrowej analizy wiersza Marcina Świetlickiego.

Pomysł ten trudno ocenić inaczej, niż jako absolutnie chybiony.

Przede wszystkim należało w tym miejscu przywołać teksty, które powstały. Literaturoznawstwo cyfrowe to zawansowaną dziedziną wiedzy z kilkoma dekadami doświadczeń i kilkoma innymi, które rozpoznawały jego możliwości nie znajdując uzasadnienia do pisania o nim jako o projekcie przyszłości. Rodzi to zresztą pytanie, na ile autor pracy jest świadom światowego i krajowego stanu wiedzy w cyfrowym literaturoznawstwie.

Oczywiście w kompetentnej pracy naukowej wypadałoby, żeby posiłkować się znajomością międzynarodowych projektów omawianych w literaturze przedmiotu (co np. zbiorczo robi zob. Martin Paul Eve w wydanej w 2022 książce *The Digital Humanities and Literary Studies*). W tym temacie przecież wystarczyło zorientować się, co dzieje się w np. w temacie literaturoznawstwa cyfrowego i wykorzystania w nim metod Natural Language Processing w niedalekiej odległości od doktoranta – w IBL PAN, gdzie studiował on przez cztery lata na studiach Humanistyka Cyfrowa. Żeby daleko nie szukać Maciej Maryl, który prowadził seminarium na tych studiach, uczestniczył w tworzeniu narzędzi takich jak LEM i GoLEM i z

pewnością chętnie opowiedziałby, jak są one wykorzystywane w praktyce. Trzymając wątku tych studiów - Anna Mędrzecka-Stefańska w ich obrębie napisała wyróżnioną pracę o ironii w poematach Juliusza Słowackiego opierając się dostępnym w otwartym dostępie narzędziach do komputerowej analizy tekstu oraz analizy stylometrycznej – prezentując przecież swoje ustalenia podczas seminariów, na które i doktorant uczęszczał.

Ale odchodząc od tematu studiów, a wracając do analizy matematycznej utworów literackich – przecież to nie współczesne humanistyka cyfrowa tego typu kierunki badawcze zaproponowała – wszak *Poetyka i matematyka* Marii Renaty Mayenowej wyszła w roku 1965. Z innej strony można przywołać ‘analogowe’ badania stylu – choćby pracę Jadwigi Sambor *Badania statystyczne nad słownictwem. Na materiale „Pana Tadeusza”* z 1969 roku czy ważną książkę *Lalka: konteksty stylu* Tadeusza Budrewicza z 1990.

Jeżeli jednak autor chciałby skupić na tekście Świetlickiego, czemu zamiast opisywać, co potencjalnie można zrobić, nie zrobił tego sam i nie włączył do pracy? Zamiast streszczać artykuły Próchniaka Stali i Bonowicza, należało ten „projektowany” artykuł napisać i umieścić jako rozdział pokazujący kompetencje doktoranta w humanistyce cyfrowej. (Pomijam tu kwestię, jaki sens mają wycieczki w tę stronę skoro praca miała być odnosić się do PBL-u i danych badawczych).

W powyższe uwagi wpisuje się także oznajmienie, iż „Przeprowadza się też na tekstach badania stylometryczne, badając podobieństwo jednych utworów do drugich”, po którym nie podaje on żadnych konkretnych przykładów, czy też fragment:

„Uważam, że nurt humanistyki cyfrowej oferuje kontynuację badań, zgodnie z założeniami metod literaturoznawstwa, ale posługuje się przy tym unowocześnionymi narzędziami. Co więcej, iż tradycyjne literaturoznawstwo, zwłaszcza takie, które nie ukrywa strukturalistycznego rodowodu, już od dawna zadaje pytania, na które humanistyka cyfrowa chce obecnie odpowiadać – z tą różnicą, że możemy udzielać na nie odpowiedzi z dużo większą dozą precyzji.”

Humanistyka cyfrowa (literaturoznawstwo cyfrowe) odpowiada na te pytania od (bardzo) dawna przywołajmy *Computation into criticism*, J.F. Burrows, z 1987 r, tylko doktorant zdaje się o tym nie wiedzieć.

### *Infrastruktury i laboratoria humanistyki*

Kwestia infrastruktury badawczych, tak jak właściwie wszystkie kwestie w pracy potraktowana jest powierzchownie i bez znajomości stanu badań. Jednak jeden z fragmentów szczególnie zwraca uwagę. Na jakiej podstawie p. Eichler twierdzi, iż: „Narzędzia cyfrowych humanistów są dostarczane przez globalne przedsiębiorstwa nastawione na maksymalizację zysku”?

Większość (rodzimy) narzędzi i infrastruktury badawczych stworzonych jest z pieniędzy publicznych i tworzy akademicką infrastrukturę naukową – dotyczy to licznych narzędzi i infrastruktury stworzonych przez IBL PAN (wśród nich przecież PBL-owych), infrastruktury DARIAH\_LAB, narzędzi stylometrycznych stworzonych przez CLARIN, czy platformy naukowych edycji cyfrowych TEI Panorama. Wymienione infrastruktury otwarcie informują o źródłach finansowania i są w większości dla naukowców bezpłatne. Podobnie jest w przypadku

europejskich narzędzi i infrastruktur, które powstają w przypadku projektów Horizon. Albo więc doktorant jest w stanie podać znaczącą przeciwwagę komercyjnych infrastruktur tworzonych dla polskiej i Europejskiej humanistyki cyfrowej, albo jest to stwierdzenie fałszywe.

Przy okazji powtórzyć można refren o braku aktualnej literatury przedmiotu choćby niezwykle ważnym artykule *Sustaining Knowledge and Governing Its Infrastructure in the Digital Age: An Integrated View* współtworzonym przez Pierre'a Mouniera, postać nie tylko ważną dla europejskiej Humanistyki Cyfrowej, ale także od lat współpracującą z IBL PAN.

#### *Topografia literatury (czy sportu)*

Bardzo charakterystyczny jest fragment zatytułowany *Dane geograficzne i dotyczące położenia*. Oczekiwalibyśmy w nim analizy wykorzystania danych topograficznych w badaniach literackich ze szczególnym uwzględnieniem licznych projektów cyfrowych, w tym bardzo zaawansowanych przedsięwzięć krajowych. Niestety, nic bardziej mylnego.

Na wstępie fragmentu możemy przeczytać:

„Sport stanowi dziedzinę, w której użycie danych wspiera istniejące już metody pracy. Nie jest niczym nowym zastosowanie dietetyki czy indywidualnych planów treningowych opartych na danych. Współczesne wysiłki w tej dziedzinie rozwijają się jednak również na inne sposoby, na przykład z użyciem danych geolokalizacji w dyscyplinach drużynowych”

I dalej, przez kolejną stronę czytamy o wykorzystaniu danych przestrzennych w sporcie, by przejść do rozważań w stylu „Wyobraźmy sobie teraz humanistę cyfrowego jako trenera drużyny” czy „Tego rodzaju informacje, poza swoim typowym zastosowaniem geodezyjnym, militarnym czy planistycznym, mogą służyć do innych celów”

Tym czasem w tej dziedzinie danych przestrzennych nie musimy sobie nic wyobrażać ani w humanistyce cyfrowej, ani w cyfrowych badaniach literackich. Cyfrowe topografie danowe mają w humanistyce cyfrowej swoją długą tradycję i ogromne zasługi w wielu dyscyplinach, podobnie jak topograficzne podejście do badań literackich. Jeżeli doktorant postanowił nie wspominać o dziesiątkach projektów analogowych oraz cyfrowych powstałych za granicą, przynajmniej rodzime osiągnięcia w tej dziedzinie wypadałoby przywołać.

Wspomnieć należałoby choćby o niezwykle zaawansowanych pracach rodzimych historyków i tworzonej przez nich cyfrowej wersji atlasu historii Polski (spora porcja ich imponujących dokonań dostępna jest w wolnym dostępie -na platformie <https://atlasfontium.pl/> wraz z najważniejszą zakładką na platformie zatytułowaną „dane”). Trudno na poważnie rozważać kwestię topografii i danych w humanistyce cyfrowej, ignorując te dokonania.

Wypada także zauważyć wbrew sugestiom doktoranta, rozważania na ten temat nie zaczęły się w 2014 wraz z publikacją Elżbiety Rybickiej - klasyczne prace powstały w IBL PAN przed ponad dwoma dekadami, jak *Warszawa – dziwne miasto* Marty Zielińskiej z 1995 roku czy *Getto warszawskie – przewodnik po nieistniejącym mieście* Barbary Engelking i Jacka Leociaka z 2001 roku. Do dziś dzień wspomniani badacze rozwijają swoje koncepcje i publikują na ten temat – także w ramach projektów cyfrowych, tworzonych w IBL PAN – zresztą na platformie

Nowa Panorama Literatury Polskiej, którą doktorant przywołuje, natomiast najwyraźniej nie zapoznał się z nią zbyt uważnie, skoro zignorował opublikowany tam *Atlas Literatury Romantyzmu* tworzony we współpracy z Martą Zielińską i *Atlas Literatury Zagłady* tworzony wraz z Jackiem Leociakiem.

#### *Dygresja o naukowych edycjach cyfrowych*

Nie możemy się nie zatrzymać przy fragmencie dotyczącym edycji tekstu. Doktorant wchodząc na ten obszar ponownie pokazuje ogromne braki w wiedzy jeżeli chodzi o stan badań nad naukowym edytorstwem cyfrowym oraz jeśli chodzi o projekty oraz infrastruktury związane z tym zagadnieniem.

Potencjalne edycje wielowersjowe i pokazujące genetykę tekstów powstają na świecie od ponad dekady (polecić można choćby wspaniałą edycję Fausta), a od kilku lat także w IBL PAN.

Niezrozumiałe w tym przypadku jest powołanie się na *Alchemię rękopisu* Marka Troszyńskiego, a zignorowanie dalszych kroków, przekształcających edycję Troszyńskiego w płynną, całkowicie cyfrową edycję dramatu *Samuel Zborowski* stworzoną przez Annę Mędrzecką-Stefańską oraz Ewę Mirkowską (o której zresztą opowiadały na konferencjach i o której publikowały artykuły), a która w praktyce pokazuje, jak funkcjonują edycje cyfrowe. Znowu jako o potencjalnej możliwości pisze doktorant o czymś, co realizowane jest w praktyce od lat.

Nie sposób przy okazji nie wspomnieć, że IBL PAN, w którym p. Eichler studiował, jest ważnym ośrodkiem naukowego edytorstwa cyfrowego, tam powstała platforma TEI Panorama, zawierająca liczne cyfrowe edycje (w tym genetyczne) w otwartym dostępie, zaś wspomniana edycja dramatu Słowackiego powstała w ramach projektu infrastrukturalnego DARIAH LAB, na który w swojej pracy doktorant się powołuje. Stan wiedzy doktoranta jest tym bardziej rozczarowujący, że istnieją wszakże liczne publikacje po polsku, wraz z całym numerem „Sztuki Edycji” z 2023 r poświęconym praktykom edycji cyfrowych.

#### **Kilka uwag o poszczególnych rozdziałach.**

**Rozdział pierwszy** wedle deklaracji mający pokazywać aktualny stan rodzimej zasadniczo ignoruje projekty, przedsięwzięcia i teksty rodzimej i światowej Humanistyki Cyfrowej, zwłaszcza tej związanej z literaturoznawstwem z ostatniej dekady. Jest zasadniczo niepotrzebny, a także niezwiązany w deklarowanym temacie pracy.

**Rozdział drugi** powiela problematyczne struktury – zaczyna się od obszernych streszczeń (m.in. artykułu o przyszłości sieci z 1995 roku), które nie wnoszą niczego do tematu danych. Pomimo iż jeden z podrozdziałów zatytułowany jest *Stan badań i dobre praktyki w obszarze otwartych danych* nie dowiadujemy się wiele na ten temat – stan badań pozostaje właściwie nieopisany.

W dalszej części rozdziału zaczynają się pojawiać definicje, jednak zupełnie nie wiadomo, na jakiej podstawie budowane. Przy ich opisywaniu nie jest przywołana literatura przedmiotu, ani różne koncepcje dotyczące tego, czym są dane w literaturoznawstwie.

Nieprzerwanie napotykamy na ogromne problemy z budowaniem poprawnego tekstu naukowego i oparciem go w stanie badań –jak już wielokrotnie powiedziano w tej dziedzinie



bardzo zaawansowany. Powoduje to, że nawet tam, gdzie zasadniczo z podanymi definicjami można się zgodzić, nie obcujemy z rozprawą naukową.

W dalszej części rozdziału dochodzimy do tematu danych - jednak wciąż streszczane są artykuły z dosyć odległych czasów, których przywołanie miałoby sens, gdyby doktorant odwołał się do kolejnych tekstów naukowych i dokumentów, które powstały później, znacznie rozbudowując stan badań i opis zjawiska. Raczej dziwne jest także przywołanie filmu z YT dotyczącego zarządzania danymi badawczymi podczas gdy dysponujemy już naukowymi artykułami i obszernymi raportami, które przywoływałem we wcześniejszej części recenzji.

Obszerny fragment o internetowych zasobach dotyczących Dworkin mógłby służyć jako króciutkie, esencjonalne studium przypadku i katalog zasobów/ danych, które na jej temat można zgromadzić – zamienia się jednak w rozwlekłą opowieść o korzystaniu z wyszukiwarki, z szczegółowymi opisami tego, jak wyglądają strony *Filmwebu* czy *Lubimyczytać*, (co dla każdego korzystającego z Internetu wydaje się oczywiste), a której relacja z deklarowanym tematem pracy pozostaje bardzo niejasna – tym bardziej, że dygresje od tej dygresji idą dosyć daleko, np. opisując zjawisko *review bombing*. Autor umieszcza także fragment o wideoesejach naukowych, ale ignoruje nim niemałe dokonania instytutu badań literackich i wideoesjeje na profilach IBL. Nie odnosi się także do wpływu pandemii na rozpowszechnienie tej formy

Wciąż niepoprawny jest aparat naukowy – brak przypisów do przywoływanych tekstów, a gdy już się pojawiają, są niepoprawne.

**Rozdział trzeci** – który wedle tytułu i deklaracji ze wstępu miałby być „głównym daniem” pracy to w większości dosyć bezrefleksyjne i pozbawione wniosków streszczenia: opisów ze stron Internetowych (jak w przypadku projektu NPRH zakończonego w 2018 roku) i artykułów z lat 2016-2018.

Gdy autor przywołuje dane o użytkownikach za artykułem Marcina Werli sprzed 8 lat, a następnie stwierdza:

„Warto mieć na uwadze kilkumiesięczny zakres czasowy badań Google Analytics, od których minęło 7 lat. Nie tylko pożądanym byłoby powtórzyć analizę tego typu, ale też uruchomić ją na dłuższy okres (badanie pełnego roku kalendarzowego, badanie krótkich okresów np. kilkudniowe odczyty raz miesiąc, ale przynajmniej przez rok ze wskazaniem różnic miesiąc do miesiąca itp.)”

można tylko przyklasnąć – tak należałoby to właśnie zrobić. Tym bardziej, że przez czas, który upłynął od uzyskania danych przez Marcina Werlę obfitował w intensywną pracę nad platformami PBL-u ale też zmienił nawyki użytkowników.

Dlaczego więc doktorant tego nie zrobił? Dlaczego przywołuje w pracy archiwalne dane bez żadnego porównania ze stanem aktualnym dla pisania pracy? Co wnosi rozdział polegający na streszczeniu dobrze znanych artykułów, opisujących minione etapy tworzenia PBL-u? Dlaczego nie uzyskujemy nowej wiedzy o PBL-u od doktoranta, który miał wszakże dostęp do członków zespołu PBL-u?

Ta kwestia jest tym bardziej wyrazista, jeżeli wspomnimy np. artykuły o PBL-u z monografii *Panorama Literaturoznawstwa Cyfrowego* czy obronioną niedawno w IBL-u znakomitą pracę Karoliny Przysieckiej.

**Rozdział czwarty**, choć zdecydowanie najbardziej udany z całej pracy - jedyny raz, kiedy przedstawione są konkretne działania które doktorant podjął i ukończył – uznać można za satysfakcjonujący jedynie w kontekście innych rozdziałów pracy. Niestety - wciąż obarczony jest wszystkimi jej słabościami, jak kuriozalnie długie cytaty czy nieumiejętność operowania aparatem naukowym. – dodatkowo trzeba wspomnieć, że czym innym jest opowieść o metodzie i kolejnych działaniach (to ciekawe i potrzebne), czym innym opowieść o nabywaniu kompetencji przez doktoranta – tego typu „opowieści biograficzne”, podobnie jak inne osobiste wtręty w pracy nie spełniają kryteriów naukowości.

### **Zakończenie**

#### *Pro domo sua*

Kończąc tą recenzję nie mogę pozostawić bez komentarza faktu, że doktorant postanowił powołać się na studia „Humanistyka Cyfrowa” w IBL PAN. Przedstawić pracę tak nieudaną, powołując się przy tym na instytucję, która od dekady jest krajowym liderem wielu obszarów Humanistyki Cyfrowej, jednocześnie w większości pominąć w pracy tej instytucji dokonania – to co najmniej duża niezręczność.

#### *Konkluzja*

Jeżeli rozważymy wymagania stawiane pracom doktorskim: Czy rozprawa zawiera oryginalne rozwiązanie problemu naukowego? Czy kandydat posiada ogólną wiedzę teoretyczną w dyscyplinie? Czy kandydat posiada umiejętność samodzielnego prowadzenia pracy naukowej? Odpowiedź brzmi – trzy razy nie. Wynika z tego, że praca nie spełnia wymogów stawianych pracom doktorskim. Wnioskuje o to, by nie kierować jej do dalszych etapów przewodu.

Warszawa 07.10.2024

Dr hab. Bartłomiej Szleszyński, prof. IBL PAN

