

Andrzej Radomski*

ORCID: 0000-0002-1735-605X

Uniwersytet Marii Curie-Skłodowskiej w Lublinie

ANALITYKA KULTUROWA, CZYLI JAK NARZĘDZIA *DATA SCIENCE* ZMIENIŁY HUMANISTYKĘ

W artykule zostały przedstawione paradygmaty badawcze, które radykalnie zmieniły współczesną humanistykę. Najważniejszym z nich jest analityka kulturowa. Jest ona oparta na metodach *data science*. Autor prezentuje założenia *data science*, a następnie cechy charakterystyczne humanistyki cyfrowej i analityki kulturowej. W drugiej części artykułu zostały przedstawione przykłady badań i projektów prowadzonych w ramach analityki kulturowej. Są to projekty realizowane w DH Lab Uniwersytetu Yale, Software Studies Initiative oraz Medialab Katowice. Badania prowadzone w tych instytucjach przeobraziły humanistykę. Jej cechy charakterystyczne to badanie dużych kolekcji danych, automatyzacja badań, wykorzystanie uczenia maszynowego i wizualizacja wiedzy. Nowa humanistyka, twierdzi autor, stała się nauką ścisłą.

Słowa kluczowe: wizualizacja, *big data*, uczenie maszynowe, humanistyka cyfrowa, humanistyka, analityka kulturowa, *data science*

WPROWADZENIE

Jest już banałem stwierdzenie, że świat współczesny jest w decydującym stopniu kształtowany przez różnego rodzaju technologie ICT¹. Zmieniły one oblicze wszystkich praktyk społecznych i powołały do życia nowe. Nauka także uległa daleko idącym transformacjom będącym konsekwencją rewolucji cyfrowej. Zmieniły się: przedmiot/-y badań, metodologia, infrastruktura badawcza, organizacja badań, sposoby współpracy, praktyki wydawnicze, a także formy komunikacji wyników badań.

Humanistyka pozostawała nieco na uboczu tych przemian. Co prawda, także badacze historii, literatury czy dziejów sztuki obficie korzystają z narzędzi cyfrowych, cyfrowych archiwów i źródeł czy możliwości, jakie dają wirtualne światy, jeśli chodzi o współpracę,

* Uniwersytet Marii Curie-Skłodowskiej w Lublinie, Instytut Nauk o Kulturze, pl. Marii Curie-Skłodowskiej 4, 20-031 Lublin; e-mail: andrzejradomski64@gmail.com

¹ Skrót od angielskiego: *information and communication technology*.

organizację konferencji i wydarzeń czy dydaktykę akademicką, to w wielu innych aspektach uprawiania badań nadal preferują dotychczasowe standardy. Chodzi tu przede wszystkim o przedmiot badań, metody badawcze i sposoby prezentacji wiedzy.

W artykule zostanie przedstawiony nowy nurt badań zwany analityką kulturową, który wraz z innym prężnie rozwijającym się paradygmatem określanym jako humanistyka cyfrowa zaproponowały inny model prowadzenia badań na gruncie różnych dyscyplin humanistycznych. Nie będzie to oczywiście jakaś dogłębna i szczegółowa charakterystyka tego podejścia (ramy krótkiego artykułu na to po prostu nie pozwalają), tylko bardziej ogólna prezentacja tego, co do praktyki badań humanistycznych wniosła analityka kulturowa (a także, choć bardzo skrótowo, humanistyka cyfrowa) i szersze konsekwencje tej rewolucji dla całej humanistyki.

Po wyczerpaniu się idei postmodernistycznych, co nastąpiło w końcu XX wieku, w szybkim tempie zaczęły pojawiać się nowe nurty, szkoły czy paradygmaty na gruncie różnych dyscyplin humanistycznych czy społecznych.

Analityka kulturowa wraz z humanistyką cyfrową postawiły na szerokie wykorzystanie narzędzi ICT i metod badawczych wypracowanych na gruncie innych dyscyplin, a zwłaszcza *data science*. Toteż czytelnik otrzyma także garść informacji temu poświęconych, aby mieć ogólne przynajmniej wyobrażenie o specyfice podejścia do badania różnych zjawisk preferowanych przez *data scientist*'ów. Wszystko to zostanie zilustrowane przykładami z tych nowych praktyk badawczych. Plan „wypowiedzi” będzie składał się z następujących zagadnień: (a) humanistyka wobec wyzwań cyfrowego świata, (b) metody i narzędzia *data science*, (c) analityka kulturowa, (d) przykłady badań charakterystyczne dla analityki kulturowej, (e) najważniejsze zmiany, które wprowadziły do praktyki humanistycznej nowe nurty badawcze.

CYFROWY ŚWIAT I WYZWANIA DLA HUMANISTYKI

Jedną z najbardziej charakterystycznych cech współczesnego świata jest zdecydowana dominacja różnego rodzaju technologii cyfrowych. To one, jak się powszechnie sądzi, mają decydujący wpływ na „umeblowanie” większości praktyk społecznych. Technologie elektroniczne, cybernetyczne, telekomunikacyjne czy bioinformatyczne powołały też do życia nowe praktyki – od gospodarki począwszy, a na artystycznych skończywszy. U podstaw tego cyfrowego technopolu leżą algorytmy napisane za pomocą któregoś z kilkuset języków programowania². W związku z tym mówi się wręcz o kulturze algorytmów (Szpunar, 2019, s. 23). Coraz większą rolę zaczynają też odgrywać algorytmy sztucznej inteligencji i tzw. internet rzeczy, które wkraczają w kolejne obszary życia współczesnych społeczeństw.

Z punktu widzenia badacza reprezentującego humanistykę są dwie (co najmniej) konsekwencje wspomnianych wyżej zjawisk. Pierwsza to pojawienie się nowych form aktywności, które mają przede wszystkim postać praktyk wirtualnych (np. portale społecznościowe).

² Można iść jeszcze „głębiej” i sprowadzić kod do układu zerojedynkowego, a następnie do impulsów elektrycznych zrozumiałych dla maszyny (np. dla komputera).

Druga to pojawienie się ogromnej liczby danych – generowanych przez wspomniane praktyki wirtualne, a także sztuczną inteligencję czy internet rzeczy.

Wśród najważniejszych form aktywności charakterystycznych dla praktyk wirtualnych wymienia się: kulturę daru, produkcję partnerską, ruch *open source* i *open access*, naukę obywatelską, hakytywizm społeczny, grywalizację, ruch prosumencki, współprodukcję mediów i innych wartości artystycznych. Jednym słowem, możemy tu mówić o społeczeństwie współpracy (Jemielniak i Przegalińska, 2020, s. 11).

Drugim niezwykle istotnym zjawiskiem współczesności jest gwałtowny przyrost różnego typu danych. Aby uświadomić sobie skalę tego zjawiska, przytoczmy parę przykładów. Każdej minuty 1,7 miliona osób loguje się do Facebooka – najpopularniejszego portalu internetowego (ponad miliard użytkowników). Zostawiają oni w tym czasie tysiące postów, polubień czy zdjęć. W tym samym czasie na YouTube pojawia się 4,7 miliona filmów, 194 tysięcy tweetów, 2,5 miliona snapów (Snapchat), 59 milionów komunikatów w Messengerze i jest wysyłanych 19 milionów SMS-ów³.

Przytoczone powyżej dane są przykładem naszej codziennej działalności. Składają się więc na nowy typ źródeł czy szerzej: danych empirycznych, za pomocą których możemy śledzić nasze poczynania zarówno w tzw. świecie fizycznym, jak i przede wszystkim wirtualnym. Zaczynają także być najważniejszym świadectwem do poznania współczesnej kultury (swoistym „cyfrowym śladem”). Ten zbiór jest również sukcesywnie uzupełniany o zdigitalizowane dane z przeszłości. Całość określa się mianem: *big data*. *Big data* stają się z jednej strony podstawowym materiałem do badań także prowadzonych na gruncie humanistyki. Z drugiej strony stanowią one duże wyzwanie dla tej grupy dyscyplin.

Wspomniana sytuacja jest spowodowana tym, że dotychczasowy przedmiot badań humanistyki i metody badawcze były/są dostosowane do innego świata (przedmiotu badań), innego typu źródeł i innych pytań generowanych przez klasyczne teorie poszczególnych dyscyplin humanistycznych.

Od początku istnienia dyscyplin humanistycznych zawsze ważnym (jeśli nie najważniejszym) przedmiotem badania był język (mowa, pismo, teksty). Humanistów, a badaczy kultury, literatury i historii w szczególności, interesowały kody i dyskursy, wartości i idee, metafory i symbole, artefakty i znaczenia, gramatyka kultury i jej struktury czy wreszcie komunikacja kulturowa i międzykulturowa. Główne nurty XX-wiecznych badań nad kulturą i historią w centrum stawały zwykle źródła o charakterze językowym. Podstawową metodą była interpretacja – tekstów i zachowań ludzkich (działań, czynności o charakterze kulturowym). Czytając teksty czy obserwując różne czynności wykonywane w poszczególnych praktykach społecznych, zdobywaliśmy wiedzę o interesujących nas zjawiskach kulturowych czy historycznych. Oglądając obrazy (zdjęcia, filmy czy animacje), szukano przede wszystkim treści, które one niosły. Interesował nas zazwyczaj jakiś przekaz (ideologiczny, światopoglądowy, etyczny czy estetyczny), który staraliśmy się zwerbalizować. Sytuacja nie ulegała również zmianie, kiedy przedmiotem zainteresowań antropologów, kulturoznawców czy historyków stawały się wizualne aspekty kultury (włączając w to także nowe media).

³ Dane przygotowane przez Lori Lewis i dostępne pod następującym linkiem: <https://www.allaccess.com/merge/archive/31294/infographic-what-happens-in-an-internet-minute> [10.10.2021].

Metodologia badań humanistycznych również była skoncentrowana na sposobach interpretacji tekstu/-ów. W wieku XX pojawiło się wiele paradygmatów opartych na określonym podejściu do interpretacji danych językowo-tekstualnych. Najważniejsze to: strukturalizm, semiotyka, hermeneutyka i poststrukturalizm. Ich wspólnym mianownikiem było wyodrębnienie kanonu dzieł literackich, filozoficznych czy naukowych (i oczywiście innych najważniejszych źródeł pisanych dla danej epoki), a następnie próba wyinterpretowania z nich obrazu świata, człowieka czy procesów historyczno-kulturowych.

Jak wspomniano we wstępie, od kilkudziesięciu lat jesteśmy świadkami tworzenia się wielu nurtów i zwrotów w humanistyce czy naukach społecznych. Są one reakcją na ciągle zmieniającą się rzeczywistość. Humanistyka wchodzi też w nowe obszary badań, takie jak na przykład kwestie klimatyczne, biopolityka czy kulturowe uwarunkowania i konsekwencje rozwoju nowych technologii. Do najbardziej znaczących nowych paradygmatów, które pojawiły się na mapie naukowej współczesnej humanistyki zaliczyć można: humanistykę ekologiczną, kognitywną, performatywną, artystyczną, studia postkolonialne, postgenderowe, studia nad kulturą wizualną, badania pamięci czy posthumanistykę (Nycz, 2017, s. 44–48). Mówi się też o zwrotach: posthumanistycznym, relacyjnym, przestrzennym, postsekularnym, zwrocie ku materialności, ku sprawczości, zwrocie afektywnym, gatunkowym, ku zwierzętom czy neurobiologicznym (Filip, 2019). Analityka kulturowa oraz humanistyka cyfrowa też są odpowiedzią na nowe problemy i wyzwania, jakie niesie ze sobą świat XXI wieku.

Jak wskazano wcześniej, cyfrowy świat i społeczeństwo współpracy generuje inne typy danych, co rodzi nowe przedmiot/-y badań, nowe pytania i potrzebę stosowania nowych narzędzi badawczych. Początkową reakcją na ten stan rzeczy była próba aplikacji metod wypracowanych do badania świata fizycznego (czy jak chcieli postmoderniści: tekstualnego⁴). Jednym z takich najbardziej znaczących przykładów była/jest netnografia. Nurt ten pojawił się w ostatniej dekadzie XX wieku i za główny przedmiot swego zainteresowania obrał społeczność wirtualne realizujące złożone praktyki kulturowe w internecie, a także w innych światach – przede wszystkim gier komputerowych. Jeden z głównych jej przedstawicieli tak ją definiuje: „netnografia jest badaniem posługującym się obserwacją uczestniczącą do pracy w wirtualnym terenie. Wykorzystuje także inne techniki badań, takie jak: wywiady, statystyki opisowe, studia przypadków, gromadzenie danych archiwalnych, wideografia, kolaż czy analiza semiotyczna” (Kozinets, 2012, s. 93). Są to więc prace oparte przede wszystkim na tzw. metodologii badań jakościowych stworzonej z myślą o badaniu tradycyjnego (fizycznego) świata kultury i społeczeństwa. Poza ich zasięgiem poznawczym jest świat cyfrowych danych – zwłaszcza *big data*. I to jest jedno z największych wyzwań dla współczesnej humanistyki.

Kolejnym, zresztą powiązanim z tym pierwszym, jest wizualizacja. Świat współczesny jest światem opartym na obrazach (oczywiście cyfrowych). Do jego badania używa się w dalszym ciągu metod tekstualnych. Z kolei badacze reprezentujący takie nurty, jak: etnografia wizualna czy historia wizualna też opierają się na metodologii badań jakościowych i są w stanie przebadać wybrane tylko kolekcje filmów, obrazów czy innej grafiki. Wielkie

⁴ Postmoderniści uważali bowiem, że świat kultury/kultur można pojmować jako tekst/teksty.

zbiory materiałów wizualnych zgromadzone na portalach typu YouTube, Vimeo, Flickr czy Instagram są z reguły poza możliwościami poznawczymi tych nurtów.

DATA SCIENCE

Wielkie zbiory danych określane jako *big data* są ważnym przedmiotem zainteresowań współczesnej nauki i oczywiście świata biznesu. Stały się one głównym obiektem badań nowej dyscypliny – *data science*. Ta nowa dziedzina wiedzy dostarcza również narzędzi i metod do badania różnego typu danych.

Historycy kultury czy badacze dziejów, badając wcześniejsze epoki, mieli zazwyczaj do czynienia z niedoborem danych (w terminologii historiograficznej: źródła). W XX i XXI stuleciu mamy zupełnie odmienną sytuację, czyli nadmiar danych czy źródeł. Te duże zasoby danych stwarzają z jednej strony nowe możliwości badawcze, z drugiej są ogromnym wyzwaniem dla nauki – w tym i dla humanistyki. Taka sytuacja wymusza konieczność przededefiniowania celów badawczych, metod i narzędzi.

Jak pisał (na początku obecnego stulecia) nestor badań w internecie Barry Wellman, do pisania artykułów naukowych na temat zachowań online wystarczyło po prostu mieć ciekawe przemyślenia i intuicje (Jemieliński, 2019, s. 40). Inną wszechobecną strategią było opieranie się na szeroko rozumianych próbach losowych. Teraz możemy i wręcz musimy analizować dużo więcej danych. Od kiedy społeczeństwa musiały mierzyć się z wielkimi liczbami, zauważają Mayer-Schonberger i Cukier, to były one początkowo zdane na korzystanie z prób losowych. Było to konsekwencją naturalnych ograniczeń epoki analogowej. Zanim zaczęły dominować wysoko wydajne technologie cyfrowe, nie postrzegaliśmy korzystania z próby losowej jako czegoś sztucznego. Przyjmowano wręcz taki stan jako rzecz oczywistą (Mayer-Schonberger i Cukier, 2014, s. 28). Taki stan rzeczy dominował (i do pewnego stopnia nadal dominuje) w humanistyce i częściowo w naukach społecznych.

W definicji Gartnera z 2012 roku (ośrodek analityczno-badawczy w USA) *big data* określa się jako: zbiory informacji o dużej objętości, dużej zmienności lub dużej różnorodności, które wymagają nowych form przetwarzania w celu wspomaganego podejmowania decyzji, odkrywania nowych zjawisk oraz optymalizacji procesów⁵.

Big data są wytwarzane nie tylko w internecie. Różnego typu dane, i to w coraz większej ilości, generowane są przez urządzenia mobilne – na czele z telefonami komórkowymi, przez tzw. internet rzeczy – czyli różnego rodzaju maszyny (np. sprzęt AGD czy samochody autonomiczne) podłączone do sieci i wymieniające się informacjami bez udziału człowieka czy urządzenia monitorujące (np. kamery uliczne, czujniki, GPS, opaski zdrowia). Zatem mamy tu do czynienia z nowym zespołem danych i informacji, który staje się nowym przedmiotem badań, przekraczającym granice dotychczasowych dyscyplin naukowych. Co więcej, bez uwzględnienia tego zasobu nowych cyfrowych danych nie zrozumiemy

⁵ Dane z Wikipedii: https://pl.wikipedia.org/wiki/Big_data [30.09.2021]. Czytelnik chcący uzyskać bardziej pogłębioną charakterystykę proszony jest o wejście na stronę internetową ośrodka.

kluczowych i fundamentalnych zjawisk, trendów czy faktów, które zachodzą w kulturze współczesnej, a i w wielu przypadkach będziemy zmuszeni zweryfikować dotychczasowe wizje czy narracje odnoszące się do przeszłości – kiedy przeanalizujemy zdigitalizowane, historyczne *big data*.

Podstawą *data science* są więc wielkie zbiory danych. Aby móc je eksplorować poznawczo, potrzebne są też wielkie moce obliczeniowe („chmury” obliczeniowe), które na przykład oferują takie usługi, jak: Google Cloud, Amazon Services czy Microsoft Azure. Potrzebne są też specjalne algorytmy uczenia – zwane uczeniem maszynowym. *Data science* można zatem określić jako interdyscyplinarną dziedzinę, traktującą o procesie pozyskania, obróbki, wizualizacji i wnioskowania na podstawie danych ustrukturyzowanych i nieustrukturyzowanych, z użyciem metod statystycznych, eksploracji danych, uczenia maszynowego i analizy predykcyjnej⁶.

Gwałtowny rozwój nauki o danych nastąpił po roku 2010, co wiązało się ze wzrostem mocy obliczeniowej komputerów, nowymi możliwościami pobierania, przechowywania i analizowania danych oraz rozwojem sztucznej inteligencji (pojawieniem się konwolucyjnych sieci neuronowych).

W ramach *data science* analizuje się głównie duże kolekcje danych. Taka analiza różnych rodzajów danych jest oparta na algorytmach tzw. uczenia maszynowego. Uczenie maszynowe występuje w dwóch zasadniczych odmianach, które można określić jako: klasyczne (oparte na statystyce) oraz tzw. głębokie (*deep learning*), oparte na sieciach neuronowych (np. CNN)⁷. W tym pierwszym modelu tworzy się programy, do których następnie wprowadza się dane, które mają je przetwarzać – zgodnie z napisanymi instrukcjami i określonym rezultatem (Flasiński, 2018, s. 23). Klasyczne uczenie maszynowe jest zazwyczaj wykorzystywane do tworzenia modeli regresji liniowej, klasyfikacji czy drzew decyzyjnych, które mają zastosowanie w biznesie, zwłaszcza w marketingu, reklamie, ubezpieczeniach czy bankowości i oczywiście w nauce (przede wszystkim w przyrodznawstwie i medycynie).

Uczenie „głębokie” jest oparte na złożonych sieciach neuronowych. Tego rodzaju sieć składa się z trzech podstawowych elementów: pewnej liczby neuronów, określonej liczby warstw i odpowiednich rodzajów połączeń między tymi warstwami. Uczenie tego typu składa się z wielu cykli i kończy się, gdy otrzymamy automatyczną informację, że nasz model z dużym stopniem prawdopodobieństwa rozpoznaje dane obiekty (Patterson i Gibson, 2017, s. 66).

Uczenie maszynowe (niezależnie od typu) pracuje na dwóch rodzajach danych. Mogą one być ustrukturyzowane (w postaci tabelarycznej, np. w formacie csv) i nieustrukturyzowane, które stanowią większość istniejących obecnie danych i są trudniejsze do analizy. Filmy i obrazy zalicza się do danych nieustrukturyzowanych. W uczeniu maszynowym dane dzieli się na dwie grupy: treningowe i testowe. Model uczy się na danych treningowych, a na testowych sprawdza się skuteczność tej nauki.

Uczenie maszynowe może mieć postać nadzorowanego i nienadzorowanego. W przypadku tego pierwszego, tworząc model ML (*machine learning*), wprowadzamy dane z etykietami

⁶ <https://mateuszgrzyb.pl/czym-jest-data-science/> [10.07.2021]. Jest to najczęściej spotykane rozumienie *data science*.

⁷ Konwolucyjne sieci neuronowe. Potrafią one wyodrębnić cechy odróżniające od siebie różne obrazy.

(np. człowiek czy samochód) i model uczy się rozpoznawać nowe obiekty (np. samochody), ucząc się na danych z naszymi etykietami. W przypadku uczenia nienadzorowanego zbiór treningowy nie zawiera etykiet i model sam musi zidentyfikować określone obiekty bądź występujące w tym zbiorze wzorce.

Wśród narzędzi, którymi posługują się badacze i analitycy w ramach *data science*, można wyróżnić cztery zasadnicze grupy: narzędzia do analizy i wizualizacji danych tabelarycznych (dane ustrukturyzowane), narzędzia do przetwarzania języka naturalnego (NLP, *natural language processing*), narzędzia do przetwarzania obrazów (*computer vision*) oraz do analizy sieci społecznych (NSA, *social network analysis*).

Mimo pewnych dość zasadniczych różnic, jakie istnieją między nimi, da się wyróżnić cechy wspólne. Do tych najważniejszych zaliczyć można następujące: (a) dążenie do przebadania w miarę możliwości wszystkich informacji na dany temat, a nie tylko próbek czy wybranych źródeł; (b) automatyzacja większości czynności badawczo-analitycznych – począwszy od etapu gromadzenia danych, a na prezentacji wyników skończywszy; (c) wizualizacja – rozumiana jako metoda badawcza i jako forma prezentacji rezultatów badawczych; (d) szerokie wykorzystanie tzw. głębokiego uczenia maszynowego (popularnie zwanego sztuczną inteligencją).

ANALITYKA KULTUROWA – NOWY NURT BADAŃ W HUMANISTYCE

Reakcją na powstanie cyfrowego świata i cyfrowej kultury było pojawienie się nowych nurtów badawczych. Niektóre z nich miały/mają charakter bardziej teoretyczny i koncentrują się na refleksji wokół zmian, jakie wprowadziły do ludzkich praktyk technologie ICT. Można tu wymienić między innymi: cyberkulturoznawstwo, netnografię, *software studies*, *visual studies*, *sound studies* czy *code studies*. Niezależnie od nich pojawiły się nurty bardziej empiryczne, które zaproponowały nowe podejście do badania kultury czy historii oparte na narzędziach cyfrowych – w tym i na tych używanych w ramach *data science*. Owe dwa nurty (w kolejności chronologicznej) to humanistyka cyfrowa i tytułowa analityka kulturowa.

Humanistyka cyfrowa to, mówiąc najogólniej, nurt badawczy polegający na wykorzystaniu narzędzi ICT do badania świata kultury i historii. Wykorzystanie komputerów do analizy danych badawczych w dyscyplinach artystycznych i humanistycznych, takich jak literatura i historia, sięga końca lat czterdziestych XX wieku. Za ojca założyciela humanistyki cyfrowej uchodzi włoski jezuita ks. Roberto Busa – pionier zastosowania informatyki do badania literatury⁸.

Humanistyka cyfrowa, określana również jako informatyka humanistyczna, może być dziedziną badań, nauczania i wynalazków na przecięciu informatyki i dyscyplin humanistycznych. Są to więc z natury badania interdyscyplinarne (Oza, 2020).

⁸ Więcej informacji o Roberto Busie i jego działalności znajdzie czytelnik w Wikipedii: https://pl.wikipedia.org/wiki/Roberto_Busa [11.10.2021].

Swym zakresem humanistyka cyfrowa obejmuje: (1) działania digitalizacyjne (tworzenie cyfrowych archiwów, repozytoriów i innych baz danych), (2) tworzenie cyfrowych kopii zabytków historyczno-archeologicznych (np. modeli 3D) i animacji poświęconych określonym wydarzeniom, (3) badania zdigitalizowanych artefaktów (np. literatury, danych geoarcheologicznych i innych źródeł).

Obecnie punkt ciężkości w działalności cyfrowych humanistów przenosi się z fazy digitalizacji, gromadzenia i udostępniania danych do fazy ich przetwarzania, interpretowania (czyli badania) i wizualizowania – rozumianego jako metoda analizy, a także prezentacji wyników badań. Stąd i u nich daje się zauważyć wzrastające zainteresowanie *data science* (Drucker, 2021, s. 112).

Analityka kulturowa jest drugim wyróżniającym się nurtem współczesnej humanistyki szeroko wykorzystującym technologie ICT. Jest to już paradygmat zrodzony w XXI wieku i od początku jego funkcjonowania wspiera się on/ona na narzędziach i metodach *data science*.

Za jego twórcę uchodzi Lev Manovich, który w 2007 roku powołał do życia laboratorium badawcze Software Studies Initiative i jednocześnie zaproponował nazwę: analityka kulturowa na oznaczenie tego nowego nurtu badań, który zaczął praktykować we wspomnianej instytucji.

Analityka kulturowa sytuuje się na przecięciu: *data science*, *media studies* i *digital culture studies*⁹, a więc stosunkowo młodych dyscyplin, które zajmują się badaniem i wizualizacją współczesnej cyfrowej kultury z użyciem metod komputerowych (Manovich, 2021, s. 7).

Analityka kulturowa jest postrzegana jako inny nurt w stosunku np. do zaprezentowanej „przed chwilą” humanistyki cyfrowej. Głównym bowiem celem cyfrowych humanistów była/jest analiza tekstów i źródeł historycznych (np. z literatury), a także dygitalizacja i modelowanie różnych obiektów i zdarzeń. Tymczasem dla analityka kulturowego bardziej interesującymi zagadnieniami były/są: (a) jak globalizacja i powstanie nowych technologii komunikacyjnych zmieniły geografie kultury, (b) czy możemy znaleźć określone kulturowe trendy w małych miastach, leżących z dala od wielkich metropolii, (c) jak współczesna kultura różnicuje rozwój, czy ulega on przyspieszeniu, czy spowolnieniu. Tego typu pytania stały się swoistym bodźcem do poszukiwań innych metod do badania kultury niż te, które oferowała humanistyka cyfrowa (Manovich, 2021, s. 3).

Przedmiotem zainteresowania analityki kulturowej są przede wszystkim następujące zbiory danych:

- zapisy ludzkiej aktywności w sieci (cyfrowe ślady), typu: odwiedzane strony, posty czy udostępnione materiały,
- zapisy fizycznych zachowań, ich lokalizacja geograficzna, data i czas zamieszczenia postów w sieci,
- zawartość mediów tworzona przez koncerty medialne: piosenki, wideo, książki i filmy,
- zawartość mediów tworzona przez użytkowników: posty, dyskusje, zdjęcia i wideo (Manovich, 2021, s. 57).

Do analizy zarówno powyższych, jak i innych problemów z obszaru *big data* analitycy kulturowi używają nowych metod i narzędzi: *data mining*, uczenia maszynowego i sztucznej

⁹ Jak dotychczas nie ma polskich odpowiedników tych anglojęzycznych nazw.

inteligencji (głębokiego uczenia maszynowego), przetwarzania obrazu (*computer vision*), przetwarzania języka naturalnego i lingwistykę obliczeniową.

Analitycy kulturowi, w przeciwieństwie do cyfrowych humanistów, preferują pracę na wielkich zbiorach danych – czyli naszych *big data*. Sądzą bowiem, że bez uwzględnienia dużych zbiorów danych nie zrozumiemy współczesnej kultury i to nie tylko cyfrowej (tzw. próbki losowe już tu nie wystarczą). Szukają oni w kolekcjach językowych czy medialnych określonych wzorów i prawidłowości, relacji i korelacji. Szczegółne znaczenie przypisują wizualizacji. Uważają bowiem, że posługiwanie się metodami opartymi na wizualizacji pozwala szybko wykryć wszystkie interesujące nas trendy i wzory, i to w sposób zautomatyzowany. W przypadku badania produkcji medialnej (filmów, zdjęć czy innej grafiki) wizualizacja może nam ukazać określone cechy tego typu wytworów, takich jak: jasność kolorów, odcienie barw, nasycenie czy kształty na obrazie, a także statystykę (Manovich, 2021, s. 228).

PRZYKŁADY

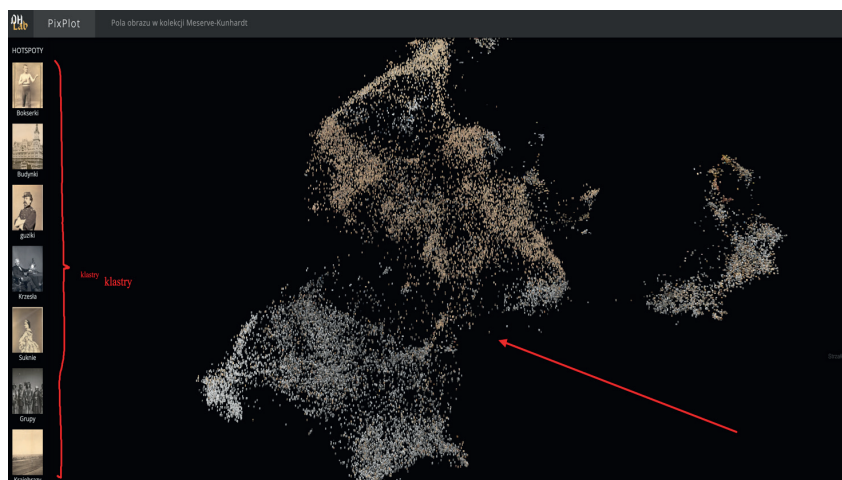
W tej części zostaną zaprezentowane przykłady badań zjawisk kulturowych, które są prowadzone w ramach analityki kulturowej. Będą one stanowić empiryczną egzemplifikację założeń teoretycznych tego nurtu zaprezentowanych powyżej.

Zacznijmy od projektu zrealizowanego przez Uniwersytet w Yale (USA). Projekt ten jest dostępny na stronach Digital Humanities Laboratory (DH Lab) wspomnianego uniwersytetu. Badacze z DH Lab zgromadzili kolekcję ponad 27 tysięcy zdjęć o różnej tematyce i postanowili je posegregować do różnych grup tematycznych. Do tego celu zastosowali głębokie uczenie maszynowe. Sieci neuronowe (w tym wypadku CNN) wytrenowano tak, aby na dostępnym materiale nauczyły się rozpoznawać, co jest na zdjęciu, a następnie mogły zaklasyfikować je do konkretnej kategorii. Jest to niezwykle ważne przy dużych zestawach danych – czyli *big data*, w których przypadku badacz/badacze nie są w stanie przejrzeć całego zasobu. Potrzebna jest więc zautomatyzowana analiza.

Oprogramowanie użyte do analizy tego materiału nosi nazwę PixPlot. Jest to rozwiązanie typu *open source*¹⁰ i służy do wizualizacji wielkoskalowych zbiorów obrazów na interaktywnej scenie WebGL¹¹. Wizualizacja oparta została na technologii WebGL i składa się z dwuwymiarowej projekcji, w której gromadzą się podobne obrazy. Użytkownicy mogą poruszać się po przestrzeni, przesuwać, powiększać i pomniejszać grupy będące przedmiotem zainteresowania lub mogą przeskakiwać do wyznaczonych „punktów aktywnych”, które zawierają reprezentatywny obraz z każdego klastra, zidentyfikowany przez komputer. Użytkownicy mogą również wybierać własne klastry. Rysunki 1 i 2 przedstawiają efekty pracy sztucznej inteligencji.

¹⁰ Kod źródłowy jest dostępny w repozytorium Git hub: <https://github.com/yaledhlab/pix-plot> [17.10.2021].

¹¹ WebGL jest rozszerzeniem języka JavaScript, zapewniającym dostęp do trójwymiarowego API w przeglądarce internetowej. Dostarcza ona programistom interfejs grafiki 3D. Korzysta z występującego w HTML 5 elementu Canvas. Aktualnie zaimplementowane jest w przeglądarkach: Mozilla Firefox, Google Chrome, Safari, Opera. Microsoft Edge oraz Internet Explorer (<https://pl.wikipedia.org/wiki/WebGL>).

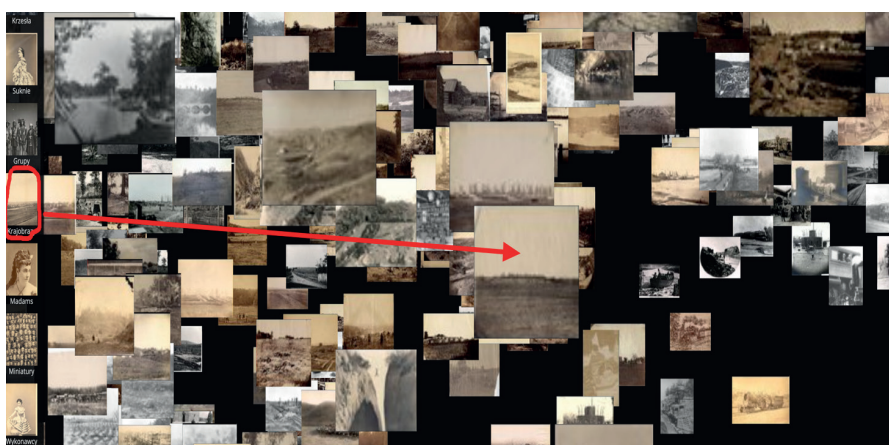


Rysunek 1. Wizualizacja kolekcji zdjęć w DH Lab Yale

Źródło: <https://s3-us-west-2.amazonaws.com/lab-apps/pix-plot/index.html>

W części centralnej mamy zgromadzone wszystkie zdjęcia będące przedmiotem uczenia maszynowego. Po lewej stronie są usytuowane zakładki zawierające najważniejsze kategorie, do których sztuczna inteligencja zakwalifikowała wszystkie zdjęcia. Wszystkie fotografie można powiększać. Co więcej, klikając na daną kategorię (po lewej stronie), od razu otrzymujemy cały zestaw fotografii do niej zakwalifikowanych.

Przykładowo klikając w zakładkę „krajobrazy”, otrzymujemy zbiór zdjęć o tej tematyce. Możemy je następnie powiększyć, aby być w stanie dokładniej obejrzeć dany krajobraz i inne interesujące nas cechy.

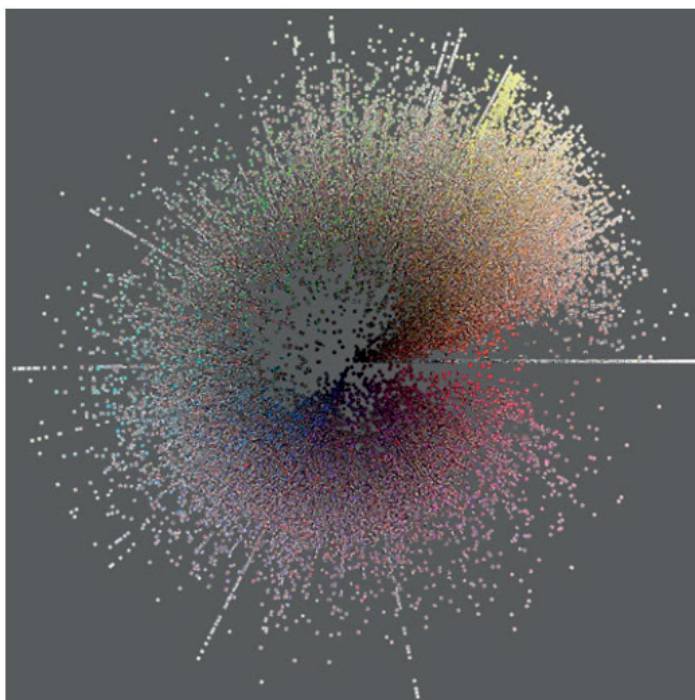


Rysunek 2. Zdjęcia zawierające kategorię krajobrazy z kolekcji DH Lab w Yale

Źródło: <https://s3-us-west-2.amazonaws.com/lab-apps/pix-plot/index.html>

Taki wstępnie wytrenowany i poklasyfikowany zbiór może oczywiście być poddawany dalszej analizie i badaniom bardziej jakościowym.

Drugi przykład jest związany z działalnością Software Studies Initiative Lva Manovicha. W tej instytucji badawczej zrealizowano kilkanaście dużych projektów, z których większość stanowią te tworzone w nurcie analityki kulturowej. Jednym z nich jest Phototrails. Jest on przykładem wizualizacji mediów. Zawiera analizę i wizualizację zdjęć z trzynastu metropolii zrobionych przez „zwykłe” osoby i umieszczonych na Instagramie. Technika opracowana przez rosyjskiego badacza umożliwia (z jednej strony) zamieszczenie na jednej zbiorczej wizualizacji wszystkich zdjęć z danej kolekcji, a (z drugiej strony) zoomowanie konkretnej fotografii. Każde wykorzystane w projekcie zdjęcie zawierało podstawowe metadane, typu miejsce i czas zrobienia, zastosowane filtry i czasami dane o autorze/autorce. Dzięki temu można było zwizualizować pod określonym kątem i porównywać cechy fotografii z różnych miast. Rysunek 3 przedstawia tego typu analizę wizualacyjną.



Rysunek 3. Wizualizacja fotografii Bangkoku w ramach programu Phototrails

Źródło: <https://phototrails.net/exhibition/>

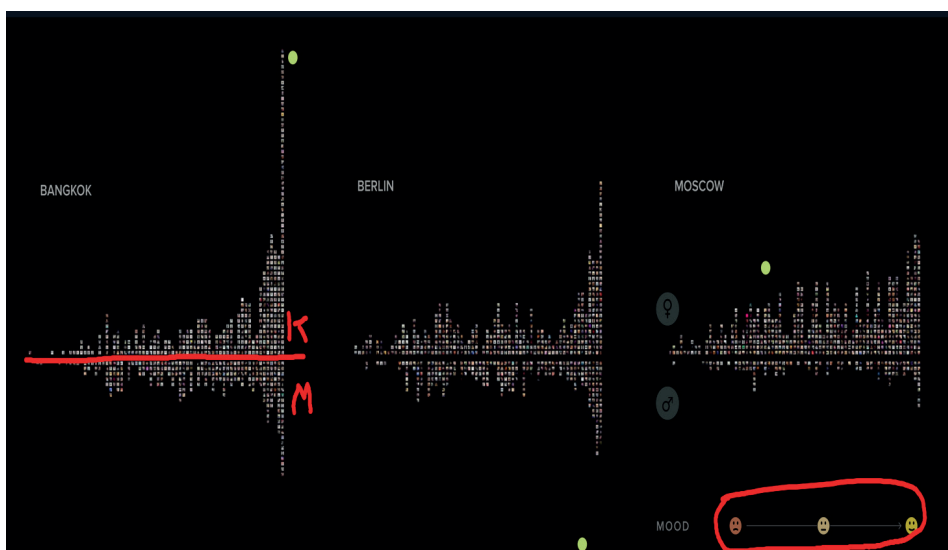
Na tej wizualizacji zostało umieszczonych 50 tysięcy fotografii stolicy Tajlandii. Zdjęcia za pomocą programu Image Plot i ImageJ zostały rozmieszczone na okręgu – w taki sposób, że skorelowano promień okręgu z jasnością, tak że te bliżej środka zawierają ciemniejsze fotografie, te oddalone – jaśniejsze. Natomiast odcienie kolorów skorelowano z obwodem

koła. Ten sam zabieg zastosowano także do innych badanych miast. Dzięki temu można przeprowadzić analizę porównawczą zwyczajów estetycznych autorów zdjęć (np. jakie preferują kolory bądź jak są one nasycone) w zależności od tego, jaką metropolię reprezentują.

Fotografii zarówno z Instagrama, jak i z innych portali można jednakże użyć do bardziej pogłębionych badań, niezwiązanych z cechami samego zdjęcia (kolor, odcienie czy nasycenie daną barwą).

Przykładem takich badań może być projekt Selfiecity, który również został zrealizowany w laboratorium Manovicha. W jego ramach także wykorzystano zdjęcia z Instagrama. Zostały one wykonane w pięciu miastach: Bangkoku, Berlinie, Moskwie, Nowym Jorku i Sao Paulo. Tym razem badaczy zainteresowały dwa zagadnienia: (1) w których ze wspomnianych miast ludzie uśmiechają się najczęściej?, (2) przedstawiciele której płci uśmiechają się częściej? Ponieważ były to fotografie typu selfie, to odpowiedź na to pytanie nie była zbyt skomplikowana. Jednakże aby była to w miarę precyzyjna odpowiedź, to przygotowano odpowiednią skalę, za pomocą której program szacował „natężenie” uśmiechu.

Rezultaty analiz przedstawia rysunek 4.



Rysunek 4. Wizualizacja pokazująca rozkład uśmiechu w wybranych miastach na podstawie zdjęć z Instagrama

Źródło: <http://selfiecity.net/>

Mieszkańcy pięciu miast zostali podzieleni na dwie grupy: mężczyźni (poniżej czerwonej linii – podkreślenie autora) i kobiety (powyżej). Natężenie poziomu uśmiechu wzrastało od strony lewej ku prawej (według skali: *mood* – podkreślenie autora).

Z zaprezentowanej powyżej wizualizacji wynika, że najczęściej uśmiechniętymi osobami byli mieszkańcy Bangkoku, najrzadziej uśmiechali się zaś moskwianie. Z kolei kobiety częściej uśmiechały się od mężczyzn. Ta dysproporcja najbardziej była widoczna w przypadku Moskwy.

Tego typu badania (tak jak w przypadku projektu Phototrails) mogą służyć dalszym, np. mierzącym poziom nastrojów społecznych czy innych aspektów życia codziennego w danym regionie metropolitalnym, a nawet w skali całego kraju.

I ostatni przykład. Będzie on pochodził z badań realizowanych w Medialab Katowice. Przedmiotem zainteresowania pracowników tego labu stały się parki miejskie w Bratysławie. Podeszli oni jednak do tego tematu inaczej, niż to było praktykowane do tej pory. Zrezygnowali bowiem z tradycyjnych metod dominujących przy tego rodzaju badaniach (takich jak badania terenowe, obserwacje czy wywiady), tylko postanowili opracować temat za pomocą danych z Instagrama. W mniejszym stopniu chodziło jednak o zdjęcia, tylko podstawowym materiałem empirycznym stały się posty, a mówiąc ściślej: hasztagi. Badaczy zainteresowało: co ludzie robią podczas pobytu w pięciu największych parkach w Bratysławie? Aby odpowiedzieć na powyższe pytanie, przeanalizowano kilkanaście tysięcy hasztagów, którymi opisywano fotografie.

Badanie składało się z trzech zasadniczych etapów: (1) pobieranie postów za pomocą API Instagrama, (2) wstępne przetwarzanie i „czyszczenie” danych, (3) właściwa analiza i wizualizacja. Na każdym z etapów wykorzystano odpowiednie algorytmy napisane w Pythonie (główny język używany przez *data scientist*’ów). Jak wskazują autorzy raportu z tych badań¹², odeszli oni z narzucania odgórnie kategorii (bardziej teoretycznych) do poklasyfikowania głównych tematów postów, tylko użyli nienadzorowanego uczenia maszynowego. Sztuczna inteligencja sama zidentyfikowała najważniejsze klastry tematyczne. Podobny zabieg zastosowano również przy analizie zdjęć. Algorytmy Google Vision API automatycznie rozpoznawały treść obrazów i oznaczały je odpowiednią kategorią.

Wyniki badań wykazały, że były różne wzorce korzystania z parków i że zmieniły się one w czasie pandemii koronawirusa. Autorzy badań przeprowadzonych za pomocą danych z portalu społecznościowego, a dotyczących studiów miejskich uważają, że tego typu metody oparte na *big data* i uczeniu maszynowym dają lepsze wyobrażenie o zachowaniach mieszkańców miast. Wynika to stąd, że media społecznościowe są coraz popularniejszym sposobem (i to dla wszystkich grup wiekowych) wyrażania swoich preferencji i emocji w postaci postów, zdjęć czy filmów. Ponadto użytkownicy dostarczają (często nieświadomie) różnych danych o swoich wzorcach użytkowania przestrzeni miejskiej, niż wynika to z bezpośrednich rozmów z nimi. I oczywiście ważna tu jest także skala tych danych – sprawiająca, że są one bardziej reprezentatywne niż te zdobyte klasycznymi metodami (np. próbki losowe bądź jakaś forma obserwacji). Wreszcie można też prowadzić badania porównawcze i rejestrować zmiany w czasie.

DWA „INTERESY” HUMANISTYKI

Jurgen Habermas, „późny” przedstawiciel szkoły frankfurckiej – w swojej teorii krytycznej wyróżnił trzy „interesy” kierujące poznaniem w ramach różnych dyscyplin wiedzy. Odpowiednio były to: (a) interes panowania, (b) wolnej od zakłóceń komunikacji oraz (c) emancypacyjno-krytyczny. Interesem „panowania” kierują się przede wszystkim przedstawiciele nauk analityczno-empirycznych, których zasadniczym celem jest formułowanie twierdzeń

¹² Jest on dostępny pod następującym linkiem: <https://karolpiekar.ski/exploring-city-parks-with-instagram/> [21.10.2021]

czy budowanie modeli służących kontrolowaniu zjawisk przyrodniczych (przewidywanie, wyjaśnianie etc.). Z kolei dwa pozostałe interesy są charakterystyczne dla nauk hermeneutyczno-krytycznych – czyli głównie humanistycznych i społecznych. Albowiem przedstawiciele tych ostatnich często zajmują się krytyką istniejących stosunków politycznych, społecznych czy kulturowych, ujawniają założenia stojące za poszczególnymi poglądami – umożliwiając w ten sposób emancypację różnych grup społecznych czy kultur od dominujących przekonań i ideologii na temat świata, człowieka czy historii (podstawą jest tutaj rozumienie i dekonstrukcja tekstów czy, powiedzmy, mediów).

Wymienione w części pierwszej niniejszego artykułu nurty i zwroty w humanistyce, które pojawiły się w ostatnich dwóch dekadach, realizują przede wszystkim interes emancypacyjno-krytyczny – w rozumieniu habermasowskim. Analitykę kulturową (a także humanistykę cyfrową) można z kolei postrzegać jako realizujące habermasowski interes panowania – w tym przypadku nad społeczeństwem i kulturą. Do najważniejszych cech odróżniających obecnie ten model praktykowania humanistyki od modelu emancypacyjno-krytycznego można zaliczyć omówione poniżej zagadnienia.

- a) **Tworzenie wiedzy humanistycznej przenosi się do laboratorium.** Odchodzi się tu od wzorca badacza humanisty, który siedzi w swoim gabinecie albo archiwum i pracuje z reprezentatywnymi tekstami/źródłami – interpretując je za pomocą określonych kategorii teoretycznych/filozoficznych/artystycznych. Analityk kulturowy pracuje z danymi, do których analizy używa przede wszystkim programów komputerowych, aplikacji i maszyn obliczeniowych o dużej mocy (coraz częściej „chmurowych”, np. Spark czy Hadoop).
- b) **Większość zadań badawczych jest zautomatyzowana.** Wszystkie główne etapy badania są wykonywane przez komputery – na czele z analizą dużych korpusów danych.
- c) **Duża część procesu badawczego jest określona przez algorytmy programów komputerowych.** Wiedza teoretyczna przestaje być najważniejszym „przewodnikiem” w badaniu świata kultury i historii. Jej miejsce zajmują algorytmy programów komputerowych (w tym sztucznej inteligencji), które w „morzu” danych odnajdują odpowiednie obiekty, klasyfikują zbiory, odnajdują różnego typu relacje, korelacje czy trendy.
- d) **Humanistyka staje się transdyscyplinarna.** Punktem wyjścia jest realizacja konkretnego projektu (dotyczącego np. analizy zjawisk historyczno-kulturowych), do realizacji którego włącza się przedstawiciele różnych specjalności: badaczy poszczególnych dyscyplin humanistycznych, a także programistów, web developerów, grafików czy specjalistów od wizualizacji.
- e) **Pracuje się na wielkich zespołach danych.** W miarę możliwości dąży się do zebrania i przeanalizowania wszystkich danych związanych z określonym tematem – w myśl formuły: $N = \text{all}$. Zmienia to wyniki badań, które czasem mogą znacząco się różnić od tych opartych na kanonie czy próbach losowych.
- f) **Wyróżniona rola wizualizacji.** Współcześnie badacze często pracują z różnymi rodzajami danych, które są zestawione w postaci tabelarycznej. Zawierają one kolumny liczb bądź dane tekstualne zapisane np. w formacie csv. Takie zestawy nie komunikują wiedzy. Interesujące rezultaty otrzymujemy, gdy pokazują one jakieś trendy czy zależności. I z reguły techniki wizualizacyjne umożliwiają wydobycie ze zbioru danych konkretnej wiedzy (np. za pomocą diagramów, wykresów, linii trendu itp.).

- g) **Wiedza humanistyczna uzyskana za pomocą np. metod *data science* staje się w pewnych przypadkach intersubiektywnie sprawdzalna i komunikowalna.** Powoduje to, że procedury badawcze są możliwe do powtórzenia przez inne zespoły badawcze – gdyż są one takie same (algorytmy, oprogramowanie i dane). Uzyskuje się też (czasami) takie same bądź podobne rezultaty.
- h) **Zanika ostre przeciwstawienie humanistyki i przyrodoznawstwa.** Humanistyka w tym modelu staje się nauką bardziej „ściśłą” czy „zobiektywizowaną”. Wyniki badań produkowane przez analityków kultury czy cyfrowych humanistów nie odbiegają pod względem pewności, dokładności i ścisłości od tych kreowanych przez przyrodoznawców. Cyfrowe narzędzia i związana z nimi metodologia umożliwiają również postawienie nowych pytań badawczych. Teorie i idee z obszaru humanistyki można teraz sprawdzać dokładniej i szybciej niż kiedykolwiek przedtem, przy czym znajduje się schematy, których tradycyjnymi metodami („na oko”) nigdy by nie znaleziono (Bod, 2013, s. 412–413).

W stosunku do badaczy praktykujących analitykę kulturową i/lub humanistykę cyfrową wysuwane są też pewne uwagi krytyczne. Zarzuca się więc, że wiele dyskusji na gruncie analityki kulturowej koncentruje się albo na ogólnych i nadrzędnych perspektywach albo sprowadza się często do wąskich technicznych zagadnień. Tymczasem jako dziedzina naukowa humanistyka cyfrowa będzie musiała lepiej wyrazić, czym jest, skąd pochodzi i jak jej praca przyczynia się do wzrostu naszej zbiorowej wiedzy. Głównym wyzwaniem dla humanistyki cyfrowej i analityki kulturowej jako nowych paradygmatów jest wykazanie głębi, innowacyjności i jakości pracy, które wnoszą do gmachu wiedzy naukowej. A z tym na razie mają problem (Svennson, 2016).

Na jeszcze inną kwestię zwraca uwagę Alan Liu, który zauważa, że po roku 1968 jedną z wiodących cech humanistyki stała się krytyka kulturowa, obejmująca zarówno interpretacyjne studia kulturowe, jak i pogłębioną krytykę różnych zjawisk. Niestety analitycy kulturowi czy cyfrowi humaniści jak na razie stronią od takiego zaangażowania się. Zaangażowanie w krytykę kulturową powinno więc stać się, postuluje, jednym z istotniejszych zadań cyfrowych badaczy – chcących włączyć się do głównego nurtu badań humanistycznych. W przeciwnym razie, przewiduje, może to już na starcie zablokować rozwój tych nowych nurtów (Liu, 2011).

Niezależnie od głosów krytycznych analityka kulturowa wraz z humanistyką cyfrową są jednymi z nielicznych paradygmatów we współczesnej humanistyce, które próbują zmierzyć się z problemem *big data* i w ten sposób poszerzają nasze możliwości poznawcze – zwłaszcza w stosunku do świata cyfrowego i wirtualnego.

BIBLIOGRAFIA

- Bod, R. (2013). *Historia humanistyki. Zapomniane nauki*. Warszawa: Alethea.
- Drucker, J. (2021). *The Digital Humanities Course Book. An Introduction to Digital Method for Research and Scholarship*. London – New York: Routledge.
- Flasiński, M. (2018). *Wstęp do sztucznej inteligencji*. Warszawa: Wydawnictwo Naukowe PWN.
- Filip, G. (2019). *Humanistyka rozmnożona*. Pobrano z: <https://prenumeruj.forumakademickie.pl/fa/2019/01/humanistyka-rozmnozona> [12.12.2021].

- Jemielniak, D., Przegalińska A. (2020). *Spoleczeństwo współpracy*. Warszawa: Wydawnictwo Naukowe Scholar.
- Jemielniak, D. (2019). *Socjologia Internetu*. Warszawa: Wydawnictwo Naukowe Scholar.
- Kozinets, R. (2012). *Netnografia. Badania etnograficzne online*. Warszawa: Wydawnictwo Naukowe PWN.
- Liu, A. (2011). *Where Is Cultural Criticism in the Digital Humanities?*. Pobrano z: <https://dhdebates.gc.cuny.edu/read/untitled-88c11800-9446-469b-a3be-3fdb36bfbd1e/section/896742e7-5218-42c5-89b0-0c3c75682a2f> [13.12.2021].
- Manovich, L. (2020). *Cultural Analytics*. Cambridge MA: The MIT Press.
- Mayer-Schonberger, V., Cukier, K. (2014). *Big Data. Rewolucja, która zmieni nasze myślenie, pracę i życie*. Warszawa: MT Biznes.
- Oza, P. (2020). *Digital humanities. An introduction*. Pobrano z: https://www.researchgate.net/publication/342692665_Digital_Humanities-An_Introduction [13.12.2021].
- Nycz, R. (2019). *Kultura jako czasownik. Sondowanie nowej humanistyki*. Warszawa: Instytut Badań Literackich PAN.
- Patterson, J., Gibson, A. (2017). *Deep learning. A Practitioner's Approach*. Sebastopol: O'Reilly Media.
- Svensson, P. (2016). *Big Digital Humanities: Imagining a Meeting Place for the Humanities and the Digital*. Pobrano z: <https://www.jstor.org/stable/j.ctv65sx0t> [13.12.2021].
- Szpunar, M. (2019). *Kultura algorytmów*. Kraków: Instytut Dziennikarstwa, Mediów i Komunikacji Społecznej Uniwersytetu Jagiellońskiego – Wydawnictwo ToC .

CULTURAL ANALYTICS OR HOW DATA SCIENCE TOOLS CHANGED THE HUMANITIES

The article presents research paradigms that have radically changed the contemporary humanities. The most important of these is cultural analytics. It is based on Data Science methods. The author presents the assumptions of data science, and then the characteristics of digital humanities and cultural analytics. The second part of the article presents examples of research and projects conducted as part of cultural analysis. These are projects implemented at the DH Lab at Yale University, Software Studies Initiative, and Media Lab Katowice. Research conducted in these institutions transformed the humanities. Its characteristic features are the study of large data collections, research automation, the use of machine learning and knowledge visualization. The new humanities, the author claims, has become an exact science.

Keywords: machine learning, big data, visualization, humanities, digital humanities, cultural analytics, data science

Zgłoszenie artykułu: 3.11.2021

Recenzje: 5.12.2021

Rewizja: 16.12.2021

Akceptacja: 16.12.2021

Publikacja online: 31.03.2022