Agata Błoch[1]
Instytut Historii im. Tadeusza Manteuffla PAN
ORCID 0000-0002-2070-2750

Clodomir Santana[2]
Instytut Historii im. Tadeusza Manteuffla PAN
ORCID 0000-0001-7869-7184

Od historii cyfrowej do obliczeniowej. Studium przypadku cyfrowej historii kolonialnego imperium portugalskiego

Słowa kluczowe: humanistyka cyfrowa, historia obliczeniowa, historia komputacyjna, analiza sentymentu, modelowanie tematyczne, imperium portugalskie

Różnorodne oblicza *digital humanities*

Na pytanie, czym jest *digital humanities*, istnieje wiele różnych odpowiedzi. Na stronie *What Is Digital Humanities?* można znaleźć ponad osiemset unikalnych interpretacji tego zagadnienia. Jedni skupiają się na dziedzinie humanistyki, w której przy pomocy mediów cyfrowych możliwe jest dekodowanie znaczeń w produktach kulturowych oraz analizie życia społecznego (Alex Sevigny)[3]. W centrum ich zainteresowań leży poszukiwanie wiedzy i dociekanie w różnych obszarach tematycznych, takich jak literatura, języki, lingwistyka, historia, klasyka, antropologia czy archeologia, z wykorzystaniem narzędzi cyfrowych (Elizabeth Lisa McAulay)[4]. Drudzy natomiast koncentrują się na rozwijaniu narzędzi komputerowych, które mają potencjał zmienić oblicze humanistyki. Ich

[1] Artykuł został sfinansowany w ramach grantu „Człowiek twórcą historii".

[2] Artykuł został sfinansowany w ramach grantu NCN OPUS „Kolonialni mieszkańcy Brazylii i Afryki Zachodniej (1640–1822): historia globalna z perspektywy sieci korespondencji", OPUS, nr 2022/45/B/HS3/00473, 2023–2027.

[3] Wersja oryginalna: *Using digital media to explore, create, analyse and decode meanings in cultural products, current affairs and social life*, wypowiedź Alexa Sevigny'ego, https://www.whatisdigitalhumanities.com/ [dostęp: 10 IX 2023].

[4] Wersja oryginalna: *The pursuit of knowledge and inquiry in a cluster of subject areas (literatures, languages, linguistics, history, classics, anthropology, archaeology) with the use of some digital method*, wypowiedź Elizabeth McAulay, https://www.whatisdigitalhumanities.com/, [dostęp: 10 IX 2023].

celem jest stworzenie narzędzi obliczeniowych, które umożliwią bardziej zaawansowane badania humanistyczne (Travis Brown). Dodatkowo starają się wykorzystywać informatykę w badaniach z zakresu nauk humanistycznych, patrząc na nią z perspektywy tradycji tej dziedziny (Geoffrey Rockwell)[5].

W niniejszym artykule skupiamy się na tematyce historii cyfrowej, która jest wynikiem intensywnych badań w ramach humanistyki cyfrowej. Początki badań sięgają lat pięćdziesiątych i sześćdziesiątych XX w., kiedy to metody cyfrowe zaczęły coraz częściej być wykorzystywane w dziedzinach takich jak lingwistyka i studia literackie – i to one dominowały aż do lat osiemdziesiątych[6]. W miarę postępu technologii cyfrowych narzędzia te stały się coraz bardziej dostępne także dla badaczy zajmujących się historią.

U progu XXI w. humanistyka cyfrowa zaczęła jednak wywołać wiele kontrowersji. Dyskusję tę rozpoczął John Usworths w 2002 r., zadając prowokacyjne pytanie, czym właściwie jest, a czym nie jest humanistyka cyfrowa. Wyraził on mocne stwierdzenie, że możemy mówić o humanistyce cyfrowej tylko wtedy, gdy komputer jest używany jako narzędzie do modelowania danych humanistycznych i wspomaga naszą ich interpretację. Wiele osób podzielało tę opinię, m.in. Kathleen Fitzpatrick, która również podkreśliła, że nie wszystkie badania wykorzystujące komputery można zakwalifikować jako należące do humanistyki cyfrowej[7]. Niestety zamiast łączyć, humanistyka cyfrowa zaczęła dzielić swoją społeczność.

W naszych założeniach, skupiając się na dziedzinie historii cyfrowej, dążymy do bardziej wszechstronnego i otwartego podejścia. Nie bagatelizujemy znaczenia żadnej metody cyfrowej, która ma na celu przekształcenie historycznych źródeł w formę dostosowaną do dalszej analizy komputerowej. Naszym wspólnym celem powinno być sprawienie, aby dane były czytelne dla maszyn i oprogramowania. W tym kontekście nie ma znaczenia, jakie narzędzia są używane w tym procesie.

Warto zastanowić się nad pytaniem, czy proces digitalizacji źródeł historycznych można uznać za integralną część dziedziny historii cyfrowej. Nasza odpowiedź brzmi: tak. Jak zauważa Hannu Salmi, od dziesięcioleci biblioteki narodowe w różnych krajach, takich jak Australia, Finlandia, Meksyk, Brazylia, Stany Zjednoczone czy Wielka Brytania, podejmują trud digitalizacji swoich zbiorów, co umożliwiło dostęp do setek milionów dokumentów[8]. Coraz więcej materiałów historycznych jest także przetwarzanych za pomocą technologii OCR (optical character recognition), konwertującej obrazy na tekst, który może być następnie przetwarzany przez komputery[9].

---

5 Wersja oryginalna: *I don't, I use and develop computational tools for humanities research*, wypowiedź Travisa Browna, https://www.whatisdigitalhumanities.com/ [dostęp: 10 IX 2023].
Wersja oryginalna: *The thoughtful use of computing in humanistic inquiry and the thinking through of computing from the perspective of the traditions of the humanities*, wypowiedź Geoffreya Rockwella, https://www.whatisdigitalhumanities.com/ [dostęp: 10 IX 2023].

6 S. Antonijević, *Amongst Digital Humanists. An ethnographic study of digital knowledge production*, New York 2015, s. 13–14.

7 Por. J. Unsworth, *What Is Humanities Computing and What Is Not?*, „Jahrbuch für Computerphilologie" 2002, nr 4, s. 71–83; K. Fitzpatrick, *The Humanities Done Digitally*, [w:] *Debates in the Digital Humanities*, red. M.K. Gold, Minneapolis 2012, s. 12–15; S. Antonijević, op. cit., s. 20–22.

8 H. Salmi, *What Is Digital History?*, Cambridge 2020, s. 44–45.

9 Ibidem, s. 32–33.

Na podstawie naszych doświadczeń wszystkie te kroki w ramach historii cyfrowej są niezbędne, abyśmy mogli przekroczyć próg historii obliczeniowej, zwanej również komputacyjną. Dostęp do dokumentów historycznych, bez względu na to, jak bardzo są one ustrukturyzowane[10], otwiera przed nami szeroki zakres możliwości pracy z algorytmami. Historia obliczeniowa idzie o krok dalej. Jest to bowiem dziedzina, która wykorzystuje narzędzia takie jak uczenie maszynowe, algorytmy przetwarzania języka naturalnego oraz metody modelowania statystycznego i matematycznego do badania danych historycznych.

Naszym celem jest więc nie tylko zrozumienie samej natury historii obliczeniowej, ale przede wszystkim podkreślenie konieczności podejścia interdyscyplinarnego do tego zagadnienia. Nie dotyczy to jedynie historyków, którzy wykorzystują narzędzia informatyczne w swoich badaniach, ale także naukowców z dziedziny informatyki, którzy korzystają ze źródeł historycznych do doskonalenia swoich modeli i algorytmów. W związku z tym historia sama w sobie może być elementem kształtującym historię cyfrową, ale także służyć jako instrument w rękach ekspertów z dziedziny nauk komputerowych.

Rewolucja cyfrowa w analizie danych historycznych

Zainteresowanie dziedzinami historycznymi, które korzystają z wiedzy informatycznej, prowadzi do powstawania coraz bardziej interdyscyplinarnych grup badawczych. Przykładem takiej inicjatywy jest zespół Helsinki Computational History Group (COMHIS) ze stolicy Finlandii[11]. Dzięki wielodyscyplinarnemu zespołowi badawczemu grupa ta opiera swoją pracę na różnorodnych metodach, czerpiąc inspirację zarówno z historii, jak i językoznawstwa oraz uczenia maszynowego. Historię obliczeniową badacze ci definiują jako „wykorzystanie metod mieszanych, w których podejście *big data* łączy się z wiedzą ekspercką w dziedzinie historii intelektualnej i historii książki". Z kolei The Luxembourg Centre for Contemporary and Digital History (C^2DH) jest jedną z najbardziej prężnych jednostek naukowych, które wykorzystują podejście interdyscyplinarne, ze szczególnym uwzględnieniem nowych metod i narzędzi cyfrowych w dziedzinie współczesnej historii Luksemburga i Europy[12]. Przykłady takie można by mnożyć, ponieważ w większości krajów Unii Europejskiej istnieją instytuty badawcze, które wprowadzają metody cyfrowe do badań historycznych. W Polsce również znajduje się kilka takich placówek, a wśród nich można wymienić Pracownię Historii Cyfrowej na Wydziale Historii Uniwersytetu Warszawskiego, Pracownię Historii Cyfrowej w Instytucie Historii Polskiej Akademii Nauk oraz Katedrę Humanistyki Cyfrowej i Metodologii Historii na Uniwersytecie Marii Curie-Skłodowskiej.

Skupmy się teraz na obszarach, w których pytania badawcze są skierowane w stronę historii. W jakich jej dyscyplinach można zastosować metody matematyczne i informatyczne? Okazuje się, że wyobraźnia historyków nie zna granic. Coraz odważniej wykorzystują oni narzędzia cyfrowe

[10] Określenie „dokumenty ustrukturyzowane" odnosi się do plików cyfrowych lub treści, które są zorganizowane i sformatowane w taki sposób, aby ich podstawowa struktura była łatwo identyfikowalna i czytelna dla maszyn. Więcej na temat *unstructured text* i *structured text* – zob. rozdz. IV i V pozycji H. Salmi. op. cit., s. 58–105.

[11] Helsinki Computational History Group (COMHIS), https://www.helsinki.fi/en/researchgroups/computational-history [dostęp: 1 XI 2023].

[12] The Luxembourg Centre for Contemporary and Digital History (C^2DH), https://www.c2dh.uni.lu [dostęp: 1 XI 2023].

w różnych aspektach badań nad przeszłością. Przykładem może być historia starożytna w kontekście rekonstrukcji struktury elity republikańskiego Rzymu[13]. Mediewiści z kolei aktywnie używają narzędzi cyfrowych w badaniach nad inkwizycją, dążąc do stworzenia systematycznego podejścia do analizy wzorców relacyjnych odnajdywanych w archiwach inkwizycji[14]. Historia gospodarcza, historia polityczna oraz badania nad władzą i elitami stanowią przedmiot zainteresowania nie tylko mediewistów, ale również badaczy epoki nowożytnej i okresów późniejszych[15]. Warto zaznaczyć, że to właśnie od historii gospodarczej rozpoczęła się swoista rewolucja cyfrowa w tej dziedzinie, ponieważ tego rodzaju dane można z łatwością wykorzystać do analizy numerycznej. Przykłady danych w historii gospodarczej to m.in. informacje handlowe, obejmujące zarówno produkty importowane, jak i eksportowane wraz z ich liczbą oraz cenami. Szczególnie ciekawym aspektem było zrozumienie przepływu towarów i produktów między regionami poprzez analizę sieci, układu powiązań handlowych oraz kontaktów między handlarzami[16].

Rewolucja ta otrzymała nawet swoją własną nazwę – rewolucja kliometryczna. Jej początek, choć skromny, miał miejsce w 1957 r., podczas spotkania Conference on Income and Wealth organizacji Economic History Association i National Bureau of Economic Research w mieście Williamstown w stanie Massachusetts. Ten ruch dążył do unowocześnienia tradycyjnej historiografii ekonomicznej dzięki wykorzystaniu bardziej zaawansowanych metod ilościowych[17].

Oprócz historii gospodarczej metody cyfrowe znajdują dzisiaj również zastosowanie w badaniach nad historią prawa, historią powiązań rodzinnych, historią intelektualną, historią przestrzeni i sieci geograficznych, międzynarodowymi sieciami intelektualnymi czy historią szpiegostwa oraz religii[18]. Widzimy zatem, że narzędzia cyfrowe można wykorzystywać w niemal każdym obszarze historii.

---

13 C. Rollinger, *Networking the Res Publica. Social network analysis and Republican Rome*, [w:] *The Power of Networks. Prospects of historical network research*, red. F. Kerschbaumer et al., Abingdon 2020, s. 13–36.

14 D. Zbíral, R.L.J. Shaw, *Hearing Voices. Reapproaching medieval inquisition records*, „Religions" 2022, 12(13), s. 1175.

15 R. Gramsch-Stehfest, *Network Analytical Modelling of Political Structures and Actions in the Middle Ages*, [w:] *The Power of Networks…*, s. 37–55; R. Gramsch-Stehfest, *Entangled Powers. Network analytical approaches to the history of the Holy Roman Empire during the late Staufer period*, „German History" 2018, 3(36), s. 365–380; B. Wurpts, *The Value of Network Analysis in Historical Sociology. Economic and social relations in medieval Lübeck*, [w:] *The Power of Networks…*, s. 56–84; J. Haggerty, S. Haggerty, *The Life Cycle of a Metropolitan Business Network. Liverpool 1750–1810*, „Explorations in Economic History" 2011, 2(48), s. 189–206.

16 A. Pereira Antunes, *Social Network Analysis in the History of Sciences. Visualising sociability in scientific expeditions with Gephi*, „Humanidades Digitales en tiempos convulsos AAHD" 2021, 1(2), s. 15–16.

17 M. Tamm, P. Burke (red.), *Debating New Approaches to History*, London 2018, s. 277–278.

18 C. Petz, J. Pfeffer, *Configuration to Conviction. Network structures of political judiciary in the Austrian Corporate State,* „Social Networks" 2021, nr 66, s. 185–201; C. Fertig, *Kinship Networks in North Western German Rural Society (18th/19th Centuries)*, [w:] *The Power of Networks…*, s. 110–124; C. Verbruggen, H. Blomme, T. D'haeninck, *Mobility and Movements in Intellectual History. A social network approach*, [w:] *The Power of Networks…*, s. 125–150.

Zróżnicowane podejścia do analizy cyfrowej materiałów historycznych. Od ręcznego kodowania do zaawansowanych narzędzi komputerowych

Warto następnie zastanowić się, jakie rodzaje źródeł historycznych nadają się do analizy przy użyciu narzędzi informatycznych. Okazuje się, że i tu wyobraźnia historyków nie zna granic, ponieważ nie istnieją jednoznaczne kryteria, które określiłyby, jakie źródła historyczne można przekształcić w formę cyfrową. Wszystkie rodzaje dokumentów źródłowych mogą być poddane analizie historycznej za pomocą obliczeń komputerowych. Mogą to być pamiętniki z podróży, czasopisma, listy członkostwa w partii, akty sądowe, korespondencja, a także wiersze, dokumenty cywilne, spisy ludności, biografie i wiele innych, w tym także święte księgi. Widzimy zatem, że historycy korzystający z tradycyjnych źródeł – czyli tych, które nie zostały jeszcze zdigitalizowane – w kontekście historii cyfrowej nie powinni się martwić. Każdy dokument historyczny może stać się *digital*. W kolejnych akapitach wyjaśnimy, w jaki sposób to osiągnąć.

Ręczne opracowywanie metadanych z dostępnych źródeł

Pierwszym podejściem jest wykorzystanie dostępnych materiałów źródłowych do ręcznego przygotowania metadanych. W przypadku analizy sieci polega to na manualnym wydobyciu relacji pomiędzy poszczególnymi jednostkami. Na przykład Anderson Pereira Antunes opracował w ten sposób relacje pomiędzy 168 jednostkami na podstawie dziewiętnastowiecznej książki podróżniczej zatytułowanej *A Journey in Brazil*, którą napisał amerykańsko-szwajcarski zoolog Louis Agassiz podczas swojej naukowej ekspedycji do Brazylii w latach 1865–1866[19]. Michelle Jia Ye zdecydowała się na ręczne wyciągnięcie danych pochodzących z czterech czasopism wydanych między rokiem 1913 a 1923 przez komercyjnego wydawcę China Book Company w Szanghaju[20]. Christian Rollinger również sięgnął po niezdigitalizowane źródła, a mianowicie po akta notarialne wybranych urzędów hiszpańskich z lat 1580–1650. W swoich badaniach wyróżnił 1696 aktów, które obejmowały 3488 osób. Na tej podstawie samodzielnie stworzył listę obcokrajowców przebywających w hiszpańskiej Sewilli[21]. Manualna ekstrakcja danych z dokumentów historycznych, jak widzimy, jest dość powszechną praktyką. Niestety istnieje poważna wada tego podejścia – ograniczenie związane z pracą wykonywaną ręcznie oraz trudnością w jej zastosowaniu przy dużych zbiorach tekstu, czyli *big data*, obejmujących tysiące, a czasem i setki tysięcy dokumentów. Jednak to właśnie dzięki takim pracom manualnym powstają cyfrowe bazy danych, które stanowią źródło badań dla kolejnych historyków, o czym opowiemy w omówieniu podejścia trzeciego.

Kodowanie treści dokumentów historycznych

Drugim krokiem jest kodowanie informacji zawartych w dokumentach historycznych. Aby to zrobić, należy przekonwertować tekst dokumentu (np. w programie Word) do wybranego oprogramowania do kodowania, takiego jak MAXQDA. Na stronie MAXQDA dowiadujemy się, że „kod

[19] A. Pereira Antunes, op. cit.

[20] M. Jia Ye, *A History from Below. Translators in the publication network of four magazines issued by the China Book Company, 1913–1923*, „Translation Studies" 2022, 1(15), s. 37–53.

[21] C. Rollinger, op. cit.

jest […] narzędziem do identyfikacji treści dokumentu, być może jego klasyfikacji i ułatwienia jego ponownego znalezienia. Z technicznego punktu widzenia kod w MAXQDA to tekst składający się z maksymalnie 63 znaków. Kody są jak pudełka indeksowe, które zawierają karty indeksowe, do których dołączone są fragmenty tekstu, części obrazów lub segmenty wideo, a na górze znajduje się nazwa kategorii"[22].

Taką procedurę zastosowaliśmy w przypadku kodowania dokumentów będących petycjami wysłanymi przez społeczności afrykańskie i autochtoniczne do portugalskiego monarchy. Na początku ręcznie przepisaliśmy treść tych dokumentów, a następnie podzieliliśmy je na sześć kategorii: autochtoniczne kobiety, autochtoniczni mężczyźni, wolne kobiety, zniewolone kobiety, wolni mężczyźni i zniewoleni mężczyźni. W kolejnym kroku stworzyliśmy drzewo kodowe, grupując słowa według dwóch głównych kategorii – języka tożsamości, czyli sposobu, w jaki te społeczności przedstawiały siebie, oraz kodów zachowań, które odnoszą się do argumentów dotyczących ich działań. Kolejnym krokiem było ręczne kodowanie każdego z tekstów[23]. Kodowanie polega na przypisaniu przynajmniej jednego kodu do wybranego fragmentu tekstu. Ten proces pozwolił na przeprowadzenie systematycznej analizy jakościowej treści i kategoryzowanie informacji[24].

Christian Rollinger, analizując relacje typu *amicitia* w starożytnym Rzymie, również zdecydował się na kodowanie tych relacji pomiędzy aktorem A i B. Rollinger podzielił proces kodowania na dwa etapy. Na początek gromadził ogólne informacje, takie jak imiona i nazwiska, status społeczny oraz charakter powiązań, i przedstawiał je w formie tabelarycznej. Następnie przystąpił do zakodowania tych danych w formacie języka UCINET (DL). Wprowadzone informacje zostały zintegrowane z programem UCINET[25].

Wracając do tematu istoty wcześniejszych ręcznych ekstrakcji danych, można zauważyć, że służą one do tworzenia dostępnych zasobów cyfrowych, co otwiera przed historykami cyfrowymi nowe możliwości. Cindarella Petz i Juergen Pfeffer skorzystali bowiem z danych, które zostały już wcześniej zebrane przez zespół badawczy Wenningera w 2017 r. Tamten projekt umożliwił zgromadzenie 1800 akt spraw sprzed wiedeńskiego sądu prowincjonalnego z 1935 r. Zebrane wówczas akta ręcznie przepisano i nadano im metadane, takie jak identyfikator sprawy czy czas trwania kary, a także informacje dotyczące osób związanych ze sprawą. Dzięki tym danym Petz i Pfeffer mogli przeprowadzić badania obliczeniowe w celu analizy procedur karnych oraz identyfikacji wzorców działań prokuratury na przykładzie wybranych grup opozycyjnych w okresie austriackiego państwa korporacyjnego[26].

Pamiętajmy, że udostępnianie materiałów leży w gestii historyków cyfrowych. Zebrane dane powinny być dostępne, co umożliwia ich dalszą analizę przez kolejne grupy badawcze. Wśród takich rozwiązań znajdują się np. manuskrypty przygotowane zgodnie z wytycznymi IIIF (International Image Interoperability Framework). Głównym celem tego standardu jest umożliwienie

---

22 Program MAXQDA dostępny jest na stronie https://www.maxqda.com/ [dostęp: 1 XI 2023].

23 A. Błoch, *The 'Miserable Vassals' of the Empire. The androgynous codes of behaviour of black and indigenous peoples in late colonial Brazil (1775–1808)*, „Journal of History" 2022, 3(57), s. 420–457.

24 *Codes and Coding in MAXQDA*, https://www.maxqda.com/help-mx20/codes-2/information-codes-coding-maxqda [dostęp: 1 X 2023].

25 C. Rollinger, op. cit., s. 18.

26 C. Petz, J. Pfeffer, op. cit.

łatwego zintegrowania materiałów z różnymi aplikacjami i platformami. Dzięki temu historycy unikają konieczności ręcznej ekstrakcji danych i automatycznie uzyskują dostęp do obrazów, anotacji oraz innych metadanych zawartych w tych dokumentach[27]. Równie użytecznym narzędziem jest INDXR, którego twórcy podkreślają, że nie tylko umożliwia ono transkrypcję oraz publikację źródeł historycznych, ale przede wszystkim zachowuje połączenia między wpisami w bazie danych a konkretnymi miejscami na skanach rękopisów[28].

Rola narzędzi informatycznych w analizie i strukturyzacji danych historycznych

Ostatnią metodą, naszym zdaniem najbardziej kompleksową, wymagającą znacznie większych zasobów ludzkich i finansowych, jest wykorzystanie narzędzi informatycznych już na samym początku pracy nad dokumentami historycznymi. Oznacza to, że unikamy ręcznej ekstrakcji danych oraz ręcznego opracowywania metadanych i od samego początku rozważamy, jakie narzędzia cyfrowe są niezbędne do pracy nad wybranymi źródłami historycznymi. Takie podejście możemy zaobserwować m.in. w pracach grupy badawczej, która skupia się wokół czeskiego projektu „Networks of Dissent: Computational Modelling of Dissident and Inquisitorial Cultures in Medieval Europe" (DISSINET), finansowanego przez Europejską Radę ds. Badań Naukowych (ERBN) w ramach grantu Consolidator[29]. Ta grupa oferuje podejście oparte na technologii komputerowej do analizy źródeł historycznych, tworząc systematyczną strukturę do badania wzorców relacyjnych w dokumentacji inkwizycyjnej. Badacze wykorzystują przy tym modelowanie semantyczne, modele relacji społecznych oraz cechy dyskursywno-tekstowe[30]. W ramach swojej pracy stworzyli oni model o nazwie CASTEMO, co jest skrótem od „Computer-Assisted Semantic Text Modelling". Dzięki temu modelowi możliwe jest śledzenie naturalnej struktury składni tekstu pisemnego i kompleksowe reprezentowanie zawartych w nim treści[31].

Podobne wyzwania związane z przetwarzaniem danych napotkaliśmy w ramach naszego projektu, któremu poświęcimy w dalszej części artykułu studium przypadku, „Mapping the Atlantic Portuguese Empire" (MAPE). W projekcie tym zajmujemy się analizą około 170 000 rejestrów korespondencji z okresu nowożytnego pomiędzy Portugalią a jej koloniami w Afryce i Ameryce. Chociaż mieliśmy dostęp do łącznie 31 katalogów cyfrowych, dane te nadal były nieustrukturyzowane. W rozumieniu cyfrowym oznacza to, że nie były one odpowiednio sformatowane do przeprowadzenia analizy komputerowej, ponieważ brakowało im odpowiedniej struktury. Tego rodzaju dane nie są również *reproducible*, a więc nie można ich łatwo odtworzyć w identyczny sposób.

Warto zaznaczyć, że przetwarzanie wstępne danych jest nieodłącznym elementem pracy z dużą ilością zróżnicowanego tekstu. Projekt „Living with Machines", jeden z największych

---

[27] Manuskrypty IIIF dostępne są do pobrania na portalu Biblissima, https://iiif.biblissima.fr/collections/ [dostęp: 1 XI 2023].

[28] Narzędzie INDXR jest dostępne na stronie https://atlas.ihpan.edu.pl/indxr/index.php [dostęp: 12 XI 2023].

[29] Więcej na temat projektu „Networks of Dissent: Computational Modelling of Dissident and Inquisitorial Cultures in Medieval Europe" (DISSINET) – zob. https://dissinet.cz/ [dostęp: 1 XI 2023].

[30] D. Zbíral, R.L.J. Shaw, op. cit., s. 1175.

[31] Więcej na temat modelu CASTEMO – zob. *Structured Data vs. Contextual Complexity of Texts. An unnecessary dilemma?*, https://dissinet.cz/news/articles/castemo-structured-data-vs-contextual-complexity-of-texts [dostęp: 1 XI 2023].

projektów z dziedziny humanistyki cyfrowej, finansowany przez brytyjską instytucję UK Research and Innovation (UKRI), również podkreśla tę kwestię[32]. W ramach tego projektu zajmowano się różnymi rodzajami danych, takimi jak teksty gazetowe, dane tabelaryczne ze spisów ludności, dane przestrzenne zawierające metadane z map i gazet, obrazy, np. zeskanowane mapy i artykuły prasowe, oraz dane GIS, które obejmowały dane wektorowe pobrane z map. W ramach projektu stawiano sobie za cel zrozumienie nieścisłości występujących w każdej z tych kategorii danych oraz zastosowanie odpowiednich strategii przetwarzania danych. Na przykład wykorzystywano różne formaty dla danych tekstowych: JSON lub XML były preferowane do tworzenia baz danych, podczas gdy TEI sprawdzało się lepiej w przypadku rękopisów, a DocBook było bardziej odpowiednie do książek[33].

Wstępne przetwarzanie danych jest również powszechne w przypadku mniejszych projektów badawczych. Na przykład Kimmo Elo rozpoczął od zeskanowania wszystkich dokumentów graficznych przy użyciu wcześniej wspomnianej metody OCR za pomocą narzędzia Tesseract. Następnie opracował program w języku Python 2.0, aby przetworzyć pliki tekstowe i wyodrębnić ich zawartość. W ostatnim etapie dokonał również weryfikacji danych pod kątem błędów. Dzięki temu automatycznie wydobyto następujące informacje z dokumentów dotyczących spraw nordyckich, które zostały sporządzone w latach 1975–1989 przez wschodnioniemiecką służbę wywiadowczą: data raportu, słowa kluczowe odnoszące się do treści, referencje krajowe, informacje dotyczące obiektów (partie, uniwersytety) oraz informacje o osobach, których raporty dotyczyły[34].

Krótki przewodnik po narzędziach cyfrowych dla badaczy humanistycznych

Zanim przejdziemy do naszego studium przypadku, pozostaje teraz pytanie dotyczące dostępnych narzędzi w omawianych powyżej zagadnieniach. Warto dokonać rozróżnienia pomiędzy narzędziami *user-friendly*, których można używać bez potrzeby programowania, a tymi, które wymagają umiejętności kodowania, co pokażemy w kolejnej części. Te informacje są istotne szczególnie na etapie planowania projektów w dziedzinie humanistyki. Z własnego doświadczenia wiemy, że historyk może często pracować samodzielnie, szczególnie jeśli bardziej interesuje go sam rezultat przetwarzania danych niż tworzenie algorytmów i nowych modeli. Jednak w przypadku drugiego rozwiązania konieczne jest uwzględnienie w planie projektu badawczego miejsca i kosztów zatrudnienia specjalisty z dziedziny informatyki.

Jednym z popularniejszych narzędzi do pracy z analizą sieci społecznych jest Gephi – oprogramowanie typu *open source*, co oznacza, że jest ono dostępne bezpłatnie, a jego instalacja na własnym komputerze jest łatwa i wygodna. Mathieu Bastian, Sebastien Heymann oraz Mathieu Jacomy udowadniają, że Gephi doskonale sprawdza się w eksploracji sieci społecznych i manipulacji nimi, nawet w przypadku dużych modeli, które zawierają ponad 20 000 węzłów[35]. Chociaż, jak

[32] Więcej na temat projektu „Living with Machines" – zob. https://livingwithmachines.ac.uk/ [dostęp: 15 XI 2023].

[33] R. Ahnert et al., *Collaborative Historical Research in the Age of Big Data. Lessons from an interdisciplinary project. Elements in historical theory and practice*, Cambridge 2023, s. 43.

[34] K. Elo, *Utilizing Historical Network Analysis on Meta-Data to Model East German Foreign Intelligence Cycle in the Baltic Sea Region 1975–89*, [w:] *The Power of Networks…*, s. 153–171.

[35] M. Bastian, S. Heymann, M. Jacomy, *Gephi. An open source software for exploring and manipulating networks*, „Proceedings of the International AAAI Conference on Web and Social Media" 2009, 1(3).

twierdzą założyciele oprogramowania Gephi, do jego obsługi nie są potrzebne umiejętności programistyczne, warto jednak pamiętać, że samodzielne opanowanie tego, jak i innych narzędzi zajmie trochę czasu. Gephi może się pochwalić szybkim silnikiem umożliwiającym wizualizację w czasie rzeczywistym, interaktywnością oraz zdolnością do tworzenia sieci o rozmiarze nawet 100 000 węzłów i 100 000 krawędzi. Użytkownik ma dużą swobodę w dostosowywaniu układu sieci, filtrowaniu jej elementów oraz wyborze krawędzi i węzłów. Interfejs jest intuicyjny i interaktywny, co pozwala użytkownikowi na śledzenie zmian w czasie rzeczywistym. Ponadto Gephi oferuje szeroki zakres dostępnych statystyk, które umożliwiają automatyczną analizę sieci, w tym miary takie jak *betweenness centrality*, *closeness*, *diameter* czy *clustering coefficient*[36]. Devyanshu Pal i Rahul Johari polecają trzy layouty: *ForceAtlas* do poprawy wizualizacji połączonych i niepołączonych węzłów, *Label Adjust* do efektywnego zarządzania etykietami oraz *Expansion* do powiększania grafów, co pozwala na lepsze zobaczenie wszystkich węzłów[37]. Gephi jest narzędziem chętnie wykorzystywanym w nauce, dziennikarstwie, a także przy analizie mediów społecznościowych[38].

Kolejnym narzędziem, podobnym do wcześniej omówionego, jest UCINET. Ten pakiet został stworzony przez trzech guru w dziedzinie analizy sieci społecznych: Lina Freemana, Martina Everetta i Steve'a Borgattiego. W skład pakietu UCINET wchodzi również narzędzie o nazwie NetDraw, które umożliwia wizualizację. Odpowiednie tutoriale można znaleźć na stronie internetowej produktu[39]. Wyzwaniem dla historyków korzystających z obu programów jest właściwe przygotowanie danych w programie Excel, zgodnie z wymaganiami każdego z tych narzędzi. W przypadku UCINET proponowane są dwie formy przygotowania danych: *nodelist* oraz *edgelist*. Pierwsza z nich jest rekomendowana do danych ankietowych, natomiast druga jest przydatna, gdy dane pochodzą ze źródeł archiwalnych[40].

Social Network Visualizer (SocNetV) to równie przyjazne dla użytkowników narzędzie, dostępne w ramach oprogramowania typu *open source*, do pobrania za darmo. Pozwala na importowanie różnego rodzaju plików sieciowych, np. w formacie UCINET[41]. Oprócz standardowych obliczeń SocNetV posiada również moduł do tworzenia sieci interakcji na podstawie podanych adresów URL[42].

Warto również wspomnieć o narzędziu o nazwie Nodegoat, które dynamicznie się rozwija i doskonale sprawdza się w indywidualnych projektach badawczych. Twórcy tego narzędzia

---

36 Program Gephi można bezpłatnie ściągnąć ze strony https://gephi.org/ [dostęp: 1 XI 2023].

37 P. Rani, J. Shokeen, *A Survey of Tools for Social Network Analysis*, „International Journal of Web Engineering and Technology" 2021, 3(16), s. 189–216.

38 J.P. Cherian, J.J. Kizhakkethottam, A.A. Alexander, *A Comparative Review on Different Social Network Analytical Tools*, „Information and Communication Technologies: 8th Conference, TICEC 2020, Guayaquil, Ecuador, November 25–27, 2020, Proceedings" 2020, s. 194.

39 Instrukcje korzystania z UCINET i NetDraw: https://docs.google.com/document/d/1UO_hnn5jIl8cV0WDpFo36dtNQ9Bo8NsVshw7qAtBEl0/edit#heading=h.b8fgaui5u4ou [dostęp: 15 XI 2023].

40 Więcej na ten temat w instrukcji korzystania z UCINET i NetDraw: https://docs.google.com/document/d/1UO_hnn5jIl8cV0WDpFo36dtNQ9Bo8NsVshw7qAtBEl0/edit#heading=h.b8fgaui5u4ou [dostęp: 15 XI 2023]. Więcej o tym, jak importować dane do UCINET – zob. https://docs.google.com/document/d/1Qq6EIypPrv-np2xYS1etKftPh8BDlhSf_OBzI7Qog4U/edit [dostęp: 15 XI 2023].

41 Program SocNetV można bezpłatnie ściągnąć ze strony https://socnetv.org/ [dostęp: 15 XI 2023].

42 M.A.M. Faysal, S. Arifuzzaman, *A Comparative Analysis of Large-Scale Network Visualization Tools*, „IEEE International Conference on Big Data (Big Data)" 2018.

podkreślają, że pozwala ono na tworzenie kolekcji danych opartych na własnych materiałach, traktując różnorodne elementy, takie jak ludzie, wydarzenia, artefakty i źródła, jak obiekty. Dzięki takiemu podejściu jesteśmy w stanie tworzyć sieci oparte na relacjach między tymi obiektami, które można uzupełnić o dodatkowe atrybuty, związane z lokalizacją i czasem. To z kolei umożliwia tworzenie wizualizacji geograficznych i chronologicznych[43]. Grupa badawcza zrzeszona wokół Nodegoat oferuje wiele warsztatów z zakresu wykorzystywania tego narzędzia, zwłaszcza wśród humanistów cyfrowych[44].

Wspomniany wcześniej program do kodowania MAXQDA cieszy się dużą popularnością w analizie danych jakościowych. Pozwala nie tylko na kategoryzację danych poprzez tworzenie kodów, ale także na analizę treści – poprzez identyfikowanie wzorców w tekście i wizualizowanie danych za pomocą wykresów, grafik i map konceptualnych. MAXQDA obsługuje szeroki zakres rodzajów danych. Obejmuje to m.in. teksty, takie jak notatki czy raporty badawcze w formatach DOC/DOCX, ODT, TXT, RTF i RTFD. Program umożliwia także pracę z dokumentami, np. artykułami z czasopism w formacie PDF. MAXQDA świetnie radzi sobie również z analizą nagrań dźwiękowych (MP3, WAV) oraz wideo (MP4, MOV i inne). Ponadto umożliwia import arkuszy kalkulacyjnych z programu Excel oraz obsługuje zdjęcia i grafiki[45].

Znaczącą wadą MAXQDA, a także innych programów często wykorzystywanych do modelowania tematycznego, jest niestety ich komercjalizacja i, co za tym idzie, wysokie koszty subskrypcji. Oprogramowanie tego rodzaju nie zawsze jest tworzone z myślą o rozwoju humanistyki cyfrowej, lecz skupia się na potrzebach korporacyjnych. W związku z tym korzystanie z podobnych narzędzi może generować znaczne koszty. Dlatego warto poszukać platform stworzonych specjalnie w celu wspierania infrastruktury naukowej. Doskonałym przykładem jest CLARIN-PL, polskie konsorcjum naukowe będące częścią Europejskiej Infrastruktury Badawczej CLARIN[46], które opracowuje narzędzia do przetwarzania tekstów, takie jak Morpho – analizator morfologiczny, czy Serel – identyfikator relacji między anotacjami w tekście. Ponadto CLARIN-PL oferuje narzędzia do przetwarzania tekstu wielojęzycznego, ekstrakcji informacji z tekstu oraz eksploracji tekstu, w tym narzędzie Topic, służące do identyfikowania grup tematycznych, oraz narzędzia do normalizacji tekstu[47].

Zaprezentowane narzędzia umożliwiają historykowi samodzielne rozpoczęcie badań z zakresu historii cyfrowej, szczególnie jeśli priorytetem są wyniki badań, a nie eksperymentowanie z algorytmami czy tworzenie nowych modeli. W przypadku gotowości do eksploracji danych za pomocą kodowania zachęcamy do zapoznania się z naszym studium przypadku – MAPE („Mapping the Atlantic Portuguese Empire")[48], przedstawionym poniżej.

---

43 Program Nodegoat można bezpłatnie ściągnąć ze strony https://nodegoat.net/about [dostęp: 15 XI 2023].

44 Więcej na temat dostępnych narzędzi do analizowania sieci – zob. tab. 1 i tab. 2 pozycji M.A.M. Faysal, S. Arifuzzaman, op. cit., s. 4838.

45 U. Kuckartz, S. Rädiker, *Analyzing Qualitative Data with MAXQDA. Text, audio, and video*, Switzerland 2019, s. 1–11.

46 Więcej o CLARIN: https://clarin-pl.eu/index.php/o-nas/ [dostęp: 15 XI 2023].

47 Wspomniane narzędzia cyfrowe znajdują się na stronie https://ws.clarin-pl.eu/topicml [dostęp: 15 XI 2023].

48 Więcej na temat naszego projektu MAPE – „Mapping the Atlantic Portuguese Empire": https://www.projectmape.org/ [dostęp: 11 XII 2023].

Studium przypadku. Historia cyfrowa kolonialnego imperium portugalskiego

Nasz projekt MAPE koncentruje się na rekonstrukcji sieci korespondencji administracyjnej z wykorzystaniem obszernego zbioru danych pochodzących z Historycznego Archiwum Zamorskiego w Lizbonie. Zbiór ten obejmuje prawie 170 000 korespondencji z okresu od 1610 do 1833 r. wymienianych między Portugalią a jej dawnymi koloniami atlantyckimi[49]. Składa się on z 30 repozytoriów, z czego aż 26 to dawne stany Brazylii, a pozostałe cztery to Angola, Wyspy Zielonego Przylądka i Gwinea, Mozambik oraz Wyspy Świętego Tomasza i Książęca.

Celem naszego projektu jest połączenie *digital humanities* z historią, kluczowe dla naszego interdyscyplinarnego zespołu, w którym krzyżują się potrzeby badawcze humanistów i specjalistów cyfrowych. Zastosowanie podejścia cyfrowego oznacza testowanie, eksperymentowanie, dostosowywanie i tworzenie algorytmów do analizy obliczeniowej, które najlepiej sprawdzą się przy badaniu naszego zbioru danych. Z perspektywy humanistyki chcemy wykorzystać te analizy do lepszego zrozumienia dynamiki wzorców komunikacyjnych w obrębie kolonialnego imperium portugalskiego. Nasze zainteresowania skupiają się na tym, w jaki sposób wybrani aktorzy konstruowali swoje narracje i strategie myślenia po obu stronach Atlantyku. Warto dodać, że analiza tych wzorców pozwoli nam zrozumieć zachowania ludzi w różnych częściach portugalskiego imperium oraz porównać i powiązać ich doświadczenia kolonialne. Oficjalna korespondencja stanowi bowiem doskonały materiał badawczy do studiów nad historią społeczną, umożliwiając odtworzenie historii różnych grup społecznych.

W tym studium przypadku wybraliśmy jezuitów jako przykład aktorów społecznych, którzy aktywnie prowadzili swoją działalność misyjną w koloniach. Analiza obliczeniowa ma nam pomóc zrozumieć, w jaki sposób korespondencja dotycząca jezuitów pozycjonowała ich w przestrzeni kolonialnej. Szczególnie interesuje nas zmiana narracji o jezuitach w kontekście czasowym, zwłaszcza od roku 1759, kiedy to zostali oni wydaleni ze wszystkich terytoriów portugalskich na kontynencie i za oceanem. Czy jesteśmy w stanie zbadać ogólną dynamikę narracji jezuitów zarówno w koloniach afrykańskich, jak i brazylijskich? Czy analiza obliczeniowa ukaże nam różnice w ich postrzeganiu przez władze kolonialne i w kontekście, w jakim się pojawiają?

W kolejnych sekcjach przedstawimy zatem, jak wygląda historia cyfrowa od kuchni. Pokażemy również przykłady kodów, wybór algorytmów oraz pytania, jakie mogą być stawiane przez historyków. Statystyki i wykresy umieszczone poniżej zostały stworzone do celów tego artykułu i nie powinny być jeszcze uznawane za punkt odniesienia w historii Towarzystwa Jezusowego.

[49] A. Błoch, D. Vasques Filho, M. Bojanowski, *Networks from Archives. Reconstructing networks of official correspondence in the early modern Portuguese empire*, „Social Networks" 2022, nr 69, s. 123–135.

Część pierwsza: Ładowanie korpusu danych i przygotowanie do analizy

W fazie przygotowawczej naszego korpusu do dalszej analizy musieliśmy zmierzyć się z problemem nieustrukturyzowanego charakteru tekstów, na których pracujemy. Skuteczna identyfikacja nadawców i odbiorców każdego listu była rezultatem poszukiwania wzorców w tekście. Wyrażenia regularne, znane jako *regular expressions*, okazały się niezwykle przydatne do podziału tekstu na trzy części: informacje o nadawcy, informacje o odbiorcy i treść korespondencji. Dodatkowe atrybuty były identyfikowane za pomocą naszego własnego modelu NER (*named-entity recognition*) do rozpoznawania nazwanych jednostek. Te jednostki obejmują osoby (zarówno mężczyzn, jak i kobiety), tytuły szlacheckie, organizacje (zarówno instytucje cywilne, jak i świeckie), instytucje wojskowe i religijne, zawody oraz lokalizacje geograficzne. W tym celu przeprowadziliśmy ręczną identyfikację i ekstrakcję jednostek, takich jak osoby, lokalizacje, organizacje i tytuły, z próbki 4230 zapisów. Podczas tego procesu korzystaliśmy z programów takich jak MAXQDA oraz Prodigy. Od samego początku wykorzystaliśmy również bibliotekę spaCy[50], która stanowi narzędzie do przetwarzania języka naturalnego (NLP) i jest obsługiwana w języku Python[51].

Poniżej (il. 1) przedstawiamy w formie kodu i tabeli, jak wyglądają nasze ustrukturyzowane źródła historyczne, a więc takie, które są czytelne dla komputera. Plik *full_text.tab* zawiera pierwotne dane, które stanowią podstawę wszystkich dotychczasowych przeprowadzonych przez nas badań. Jest także bazą dla innych plików w repozytorium, takich jak *full_texts_filtered.csv*. Plik *tab* składa się z trzech kolumn: *id_document*, *source_file* i *full_text*. Pierwsze dwie kolumny dotyczą źródła dokumentów i nie są używane w naszych analizach obliczeniowych. Natomiast ostatnia kolumna, jak wskazuje nazwa, zawiera pełną transkrypcję tekstu dokumentów.

```
file_ = codecs.open("./datasets/full_text.tab","r","UTF-8")
root_data = pd.read_csv(file_, sep="\t")
root_data.head()
```

	id_document	source_file	full_text
0	164421	CU-Bahia.pdf	11857- [Ant. 1765, Novembro, 9] REQUERIMENTO ...
1	281830	CU-RioJaneiro.pdf	17228- [post. 1808] REQUERIMENTO do soldado ...
2	195142	CU-Guine-Parcial.pdf	721 [ca. 1753] RELAÇÃO (traslado) das fazend...
3	188871	CU-CaboVerde-Parcial.pdf	1549. 1744, Março, 30, Ribeira Grande CERTID...
4	168427	CU-Bahia.pdf	15875- [post. 1803, Agosto, 18] REQUERIMENTO ...

Il. 1. Ustrukturyzowane źródła historyczne (kod i tabela)

Część druga: Przygotowanie do analizy sentymentu i modelowania tematycznego

Poza identyfikacją nadawcy, daty oraz tematu listu naszym celem jest także zbadanie ewolucji społeczno-politycznej na poziomie makro. Chcemy wyciągnąć dane semantyczne w celu

[50] Strony internetowe wymienionych oprogramowań: MAXQDA – https://www.maxqda.com/, Prodigy – https://prodi.gy/, spaCy – https://spacy.io/ [dostęp: 11 XII 2023].

[51] A. Błoch, D. Vasques Filho, M. Bojanowski, op. cit.

zidentyfikowania poruszanych w korespondencji tematów, obejmujące kwestie polityczne, administracyjne, ekonomiczne czy religijne.

Tym samym przygotowaliśmy nasz korpus do dwóch rodzajów analiz: modelowania tematycznego oraz analizy sentymentu. Pierwsza metoda ma na celu wykrycie głównych struktur semantycznych w obszernych zbiorach dokumentów poprzez ich grupowanie w kategorie tematyczne. Pozwala to na identyfikację wiodącego tematu w tekście. Druga metoda, analiza sentymentu, umożliwia klasyfikację emocji zawartych w dokumentach jako negatywne, pozytywne lub neutralne[52].

W celu realizacji tych analiz ustaliliśmy liczbę tematów, które powinny zostać zidentyfikowane przez algorytmy. Wybraliśmy tę liczbę z zakresu od 2 do 10 na podstawie najwyższego wyniku metryki koherencji (il. 2). Dodatkowo przygotowaliśmy słownik zawierający nazwy tematów, co poprawia czytelność prezentowanych wyników.

```
n_topics = 20
topic_names = [f"Topic {i+1}" for i in range(n_topics)]
topic_dict = {i:topic_names[i] for i in range(n_topics)}
print(topic_dict)
```

```
{0: 'Topic 1', 1: 'Topic 2', 2: 'Topic 3', 3: 'Topic 4', 4: 'Topic 5', 5: 'Topic 6', 6: 'Topic 7', 7: 'Topic
8', 8: 'Topic 9', 9: 'Topic 10', 10: 'Topic 11', 11: 'Topic 12', 12: 'Topic 13', 13: 'Topic 14', 14: 'Topic
15', 15: 'Topic 16', 16: 'Topic 17', 17: 'Topic 18', 18: 'Topic 19', 19: 'Topic 20'}
```

Il. 2. Modelowanie tematyczne – wybór liczby tematów

Ładując dane w celu przeprowadzenia analizy sentymentu (SA) na bazie wszystkich tekstów w naszym korpusie, zastosowaliśmy algorytm, który wymagał długiego przetwarzania danych (il. 3). Wykonaliśmy to *a priori* i zapisaliśmy wynik jako plik XLSX. Podobnie jak w oryginalnych danych, pierwsze dwie kolumny odnoszą się do źródła dokumentów i nie są uwzględniane w naszych analizach. Ostatnia kolumna, zgodnie z nazwą, zawiera opis korespondencji.

```
In [2]:
root_data = pd.read_excel("other_files/df_sentiment.xlsx")
root_data.head(4)
```

Out[2]:

	Unnamed: 0	id_document	source_file	full_text	year	sentiment
0	0	164421	CU-Bahia.pdf	11857- [Ant. 1765, Novembro, 9] REQUERIMENTO ...	1765	NEGATIVE
1	1	281830	CU-RioJaneiro.pdf	17228- [post. 1808] REQUERIMENTO do soldado ...	1722	NEGATIVE
2	2	195142	CU-Guine-Parcial.pdf	721 [ca. 1753] RELAÇÃO (traslado) das fazend...	1753	NEUTRAL
3	3	188871	CU-CaboVerde-Parcial.pdf	1549. 1744, Março, 30, Ribeira Grande CERTID...	1744	NEUTRAL

Il. 3. Analiza sentymentu (kod i tabela)

Część trzecia: Filtrowanie danych

O ile wcześniejsze operacje obejmowały cały korpus danych, naszym kolejnym celem jest przefiltrowanie zbioru danych, tak aby zawierał on dane obejmujące przynajmniej jedno ze słów kluczowych związanych z jezuitami (il. 4).

[52] Więcej na temat modelowania tematycznego – zob. K. Pooja, P. Bansal, *Topic Modeling. A comprehensive review*, „EAI Endorsed Transactions on Scalable Information Systems" 2019, 7(24), s. 1–16. Więcej na temat analizy sentymentu – zob. K. Tomanek, *Analiza sentymentu – metoda analizy danych jakościowych. Przykład zastosowania oraz ewaluacja słownika RID i metody klasyfikacji Bayesa w analizie danych jakościowych*, „Przegląd Socjologii Jakościowej" 2014, 2(10), s. 118–136.

```
In [5]:
filter_df_keywords(root_data, keywords=['companhia de jesus', 'jesuita', 'inaciano', 'jesuitico'])
root_data = root_data[root_data.any_term==True]
root_data['source_name'] = root_data.source_file.apply(lambda x: x.split('-')[1].split('.')[0])
root_data.to_excel("other_files/jesuits_references.xlsx")
root_data.head(4)
```

```
100%|██████████| 4/4 [01:06<00:00, 16.54s/it]
Out[5]:
```

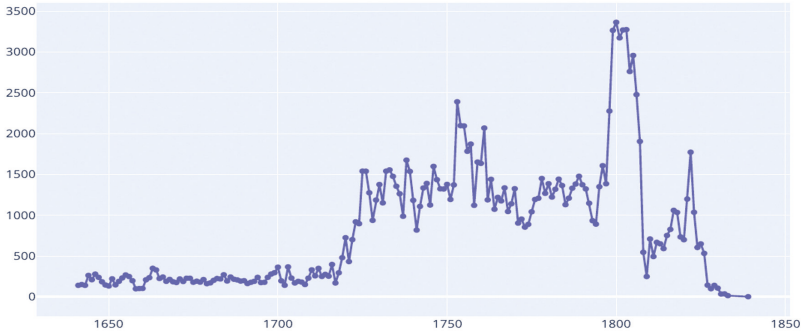
	Unnamed: 0	id_document	source_file	full_text	year	sentiment	companhia de jesus	jesuita	inaciano	jesuitico	any_term
130	130	239762	CU-Paraiba.pdf	1883- 1770, abril, 25, Paraíba OFÍCIO do gove...	1770	NEGATIVE	False	True	False	False	True
181	181	197199	CU-Maranhao.pdf	1758- 1729, Agosto, 8, São Luís do Maranhão C...	1758	NEGATIVE	False	True	False	False	True
221	221	270274	CU-RioJaneiro.pdf	5658- 1760, Março, 11, Rio de Janeiro OFÍCIO ...	1760	NEGATIVE	True	False	False	False	True
245	245	276557	CU-RioJaneiro.pdf	11947- ant. 1796, Setembro, 6 REQUERIMENTO de...	1796	NEUTRAL	False	True	False	False	True

Il. 4. Filtrowanie danych (kod i tabela)

Część czwarta: Analiza sentymentu na przykładzie wybranego tematu – „jezuici" – dla całego korpusu danych

Poniżej (il. 5) znajduje się ogólny zarys liczby korespondencji portugalskiej, uwzględniający poszczególne lata. Przedstawiony wykres pozwala zrozumieć ogólną dynamikę wymiany korespondencji między atlantyckimi peryferiami a centrum politycznym w Lizbonie. Jest również użyteczny jako wzorzec porównawczy do dalszej analizy dynamiki korespondencji wybranych aktorów społecznych, w tym jezuitów.

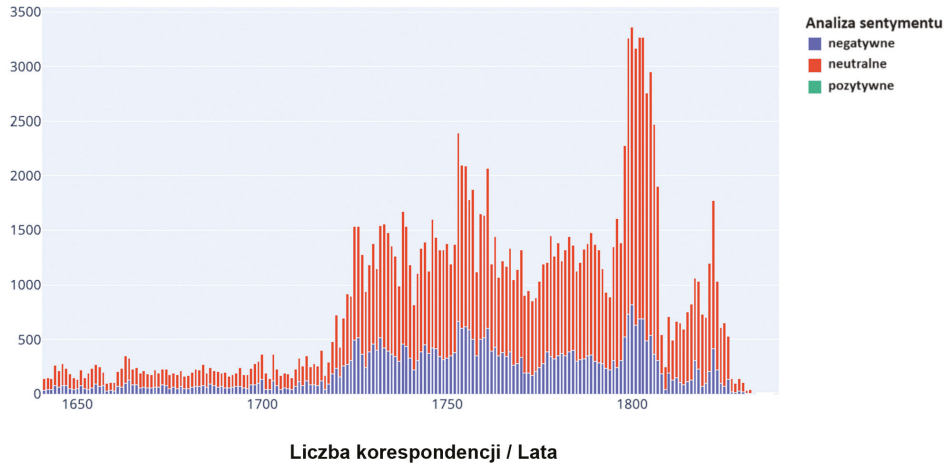
```
In [3]:
df_grouped = root_data.groupby(['year']).count().reset_index()
fig = px.line(df_grouped[(df_grouped.year>1640) & (df_grouped.year<1840)], x="year", y="id_document", markers=True,
              title='<b>Fig. 1:</b> Number of records per year',  labels={"year": "Year","id_document": "Number of Refe
fig.update_traces(line={'width': 2})
fig.show()
```



Il. 5. Liczba całej korespondencji portugalskiej z dokładnością do roku (kod i wykres)

Na podstawie ogólnego wykresu korespondencji (zob. il. 5) przygotowaliśmy wykres (zob. il. 6) prezentujący analizę sentymentu w danym okresie. Przeprowadziliśmy to badanie również w celu zbadania potencjalnej stronniczości algorytmu w kierunku negatywnych rezultatów. Otrzymany wynik potwierdza nasze przypuszczenia, że oczekiwany ton w tych korespondencjach powinien być neutralny. Nasz zbiór danych koncentruje się głównie na materiałach administracyjnych, takich jak decyzje urzędnicze, korespondencja między instytucjami czy petycje przesłane przez różne grupy społeczne.

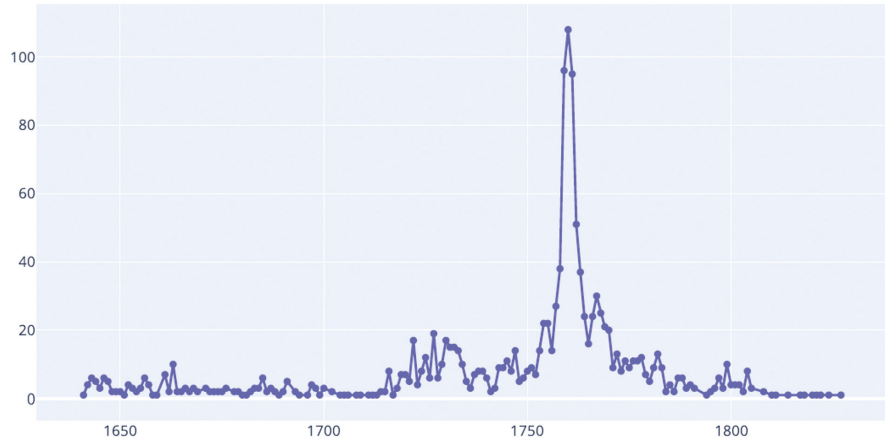
```
In [4]:
df_grouped = root_data.groupby(['year', 'sentiment']).count().reset_index()
fig = px.bar(df_grouped[(df_grouped.year>1640) & (df_grouped.year<1840)], x="year", y="id_document", color='sentiment',
             title='<b>Fig. 2:</b> Sentiment analysis of the entire data',  labels={"year": "Year","id_document": "Num
fig.show()
```



Il. 6. Analiza sentymentu całej korespondencji portugalskiej z dokładnością do roku (kod i tabela)

W kolejnym etapie porównamy ogólną liczbę korespondencji (zob. il. 5 i il. 6) z dokumentami skupionymi wyłącznie na jezuitach (zob. il. 7). Dodatkowo przeprowadzimy analizę sentymentu w korespondencji dotyczącej jezuitów (zob. il. 8). Na poniższym wykresie (il. 7) można zauważyć, że liczba odniesień do jezuitów w dokumentach osiągnęła szczyt ok. roku 1760, a nie roku 1800, co jest odmienne od ogólnej tendencji korespondencji przedstawionej na il. 6. Prawdopodobnie taka zmiana dynamiki wynikała z kontrowersji, które wzbudziło wypędzenie jezuitów z całego imperium portugalskiego w 1759 r.

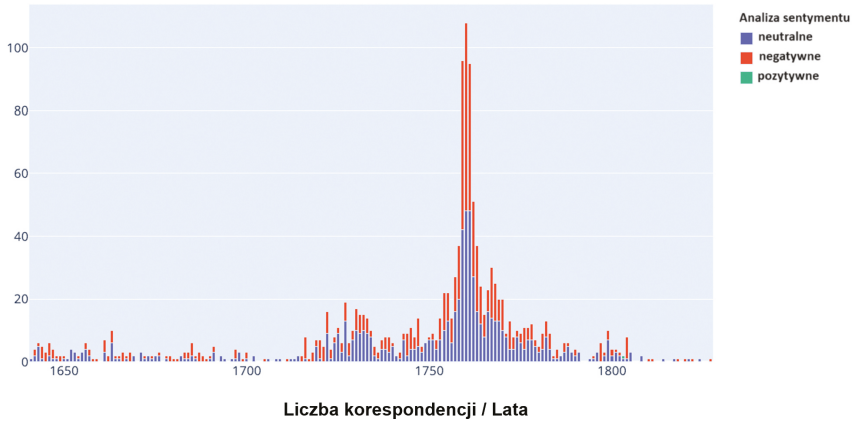
```
In [6]:
df_grouped = root_data.groupby(['year']).count().reset_index()
fig = px.line(df_grouped[(df_grouped.year>1640) & (df_grouped.year<1840)], x="year", y="id_document", markers=True,
              title='<b>Fig. 3:</b> Number of records per year with refferences to Jesuits',  labels={"year": "Year","i
fig.update_traces(line={'width': 2})
fig.show()
```



Il. 7. Liczba rocznej korespondencji zawierającej odniesienia do jezuitów (kod i wykres)

Analizując korespondencję dotyczącą jezuitów (zob. il. 7) oraz przeprowadzając analizę sentymentu (zob. il. 8), zauważamy, że kontekst, w którym pojawiają się wzmianki o jezuitach, ma wydźwięk głównie negatywny. Ten wynik różni się od ogólnego trendu w danych, który obrazuje przewagę sentymentu neutralnego (zob. il. 6). Pozwala to na dogłębne zbadanie przez historyków przypuszczalnych negatywnych emocji wobec jezuitów, które mogły stopniowo narastać i przyczynić się do ich wypędzenia w 1759 r.

```
In [7]:
root_data.to_excel("other_files/only_jesuits_references.xlsx", index=False)
df_grouped = root_data.groupby(['year', 'sentiment']).count().reset_index()
fig = px.bar(df_grouped[(df_grouped.year>1640) & (df_grouped.year<1840)], x="year", y="id_document", color='sentiment',
             title='<b>Fig. 4:</b> Sentiment assessment of records with refferences to Jesuits',
             labels={"year": "Year","id_document": "Number of References"})
fig.show()
```



Il. 8. Analiza sentymentu korespondencji dotyczącej jezuitów (kod i wykres)

Część piąta: Analiza sentymentu na przykładzie wybranego tematu – „jezuici" – dla korpusu dotyczącego kolonii afrykańskich

Widzimy, że dotąd ogólny trend dotyczący jezuitów miał charakter negatywny, osiągając apogeum w latach pięćdziesiątych i sześćdziesiątych XVIII w. Pojawia się jednak pytanie, czy był on podobny po obu stronach Atlantyku – zarówno w koloniach afrykańskich, jak i w Brazylii. W celu udzielenia odpowiedzi na to pytanie w pierwszym kroku przeprowadziliśmy analizę skupiającą się jedynie na Afryce. Poniższy kod filtruje nasz zbiór danych, usuwając wszelką korespondencję spoza tego obszaru. Termin „kolonie afrykańskie" odnosi się tu do dawnych obszarów zajmowanych przez Portugalczyków w Afryce Zachodniej – Wysp Zielonego Przylądka wraz z Gwineą Bissau, Angoli oraz Wysp Świętego Tomasza i Książęcej. Wyłączyliśmy tutaj również Mozambik, uznając go za strefę wpływów Portugalskiego Państwa Indii.

In [8]:

```
keep_africa = ['CU-CaboVerde-Parcial.pdf', 'CU-Angola-Parcial.pdf', 'CU-SaoTome-Parcial.pdf']
root_data_af = root_data[root_data.source_file.isin(keep_africa)]
root_data_af.head(4)
```

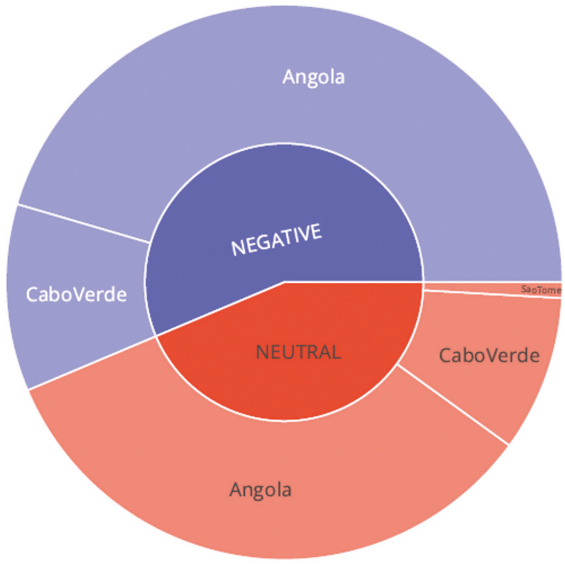
Out[8]:

	Unnamed: 0	id_document	source_file	full_text	year	sentiment	companhia de jesus	jesuita	inaciano	jesuitico	any_term	sour
458	458	148357	CU-Angola-Parcial.pdf	380. post. 1645 CARTA sobre quatro capuchos c...	1645	NEGATIVE	True	False	False	False	True	
965	965	148359	CU-Angola-Parcial.pdf	382. 1646, Fevereiro, 10, Lisboa CONSULTA do ...	1646	NEGATIVE	False	True	False	False	True	
2496	2496	152365	CU-Angola-Parcial.pdf	4398. 1762, Abril, 7, São Paulo da Assunção	1762	NEUTRAL	True	False	False	False	True	

Il. 9. Selekcja danych dotyczących tematu „jezuici" w koloniach afrykańskich (kod i tabela)

Po uprzednim przefiltrowaniu i wyborze dokumentów dotyczących kolonii afrykańskich skoncentrujemy się teraz na analizie sentymentu związanej z obecnością jezuitów w poszczególnych regionach Afryki (zob. il. 10). Naszym celem jest zrozumienie ewentualnych istotnych różnic w stosunku do ogólnej tendencji. Należy jednak zachować ostrożność w interpretacji tych wyników ze względu na ograniczoną dostępność pochodzących z tych źródeł danych na temat jezuitów.

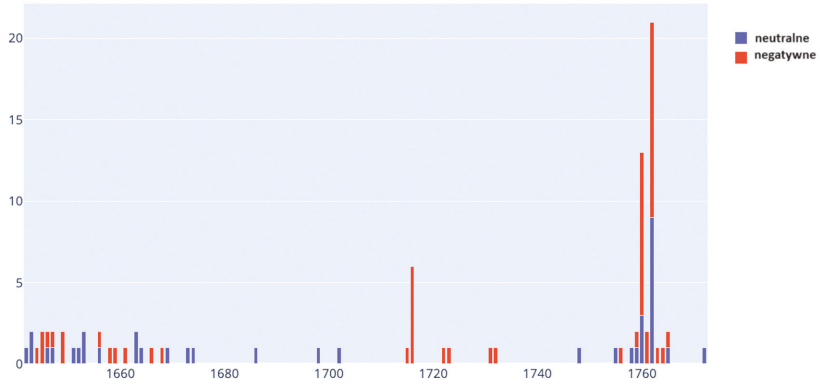
```
In [9]:
title = "<b>Fig. 5:</b> Distribution of documents from African sources"
f_df_af = root_data_af.groupby(['sentiment', 'source_name']).count().reset_index()
fig = px.sunburst(f_df_af, values='id_document', path=['sentiment', 'source_name'], title=title,
                  labels={"labels": "Name", "parent": "Sentiment","id_document": "Value"})
fig.show()
```



Il. 10. Analiza sentymentu dotycząca jezuitów z kolonii afrykańskich (kod i wykres)

Powyższy wykres (il. 10) ukazuje, że ogólnie niekorzystny kontekst, w którym jezuici pojawiali się w korespondencji kolonialnej, utrzymuje się również w analizie pod względem regionalnym. Szczególnie wyraźne jest to w przypadku Angoli, gdzie jezuici byli opisywani w negatywnym świetle. Dodatkowo analizując przebieg czasowy (zob. il. 11), można zauważyć kolejny szczyt w okolicach roku 1760, charakteryzujący się przeważającą negatywną oceną. Ponadto dostrzega się, że wzmianki o jezuitach w XVII w. były głównie neutralne, natomiast w XVIII w. – negatywne. To otwiera przed nami nowe obszary badawcze dotyczące narastających konfliktów między jezuitami a *sobas* (lokalnymi władcami) oraz jezuitami a administracją kolonialną.

```
In [10]:
df_grouped = root_data_af.groupby(['year', 'sentiment']).count().reset_index()
fig = px.bar(df_grouped[(df_grouped.year>1640) & (df_grouped.year<1840)], x="year", y="id_document", color='sentiment',
             title='<b>Fig. 6:</b> Sentiment assessment of records with refferences to Jesuits',
             labels={"year": "Year","id_document": "Number of References"})
fig.show()
```



Il. 11. Analiza sentymentu korespondencji dotyczącej jezuitów (kod i wykres)

Część szósta: Analiza sentymentu na przykładzie wybranego tematu – „jezuici" – dla całego korpusu danych pochodzących z kolonialnej Brazylii

Podobnie jak przy analizie opartej na materiałach źródłowych z kolonii afrykańskich, teraz pragniemy zgłębić sytuację w Brazylii, gdzie jezuici odgrywali rolę najbardziej aktywnego zakonu religijnego zaangażowanego w proces przymusowej chrystianizacji rdzennych ludów. Poniższy kod służy do przefiltrowania zbioru danych poprzez wyłączenie całej korespondencji z kolonii afrykańskich oraz archiwów związanych z Timorem Wschodnim, co pozostawia jedynie dokumenty dotyczące wszystkich regionów Brazylii.

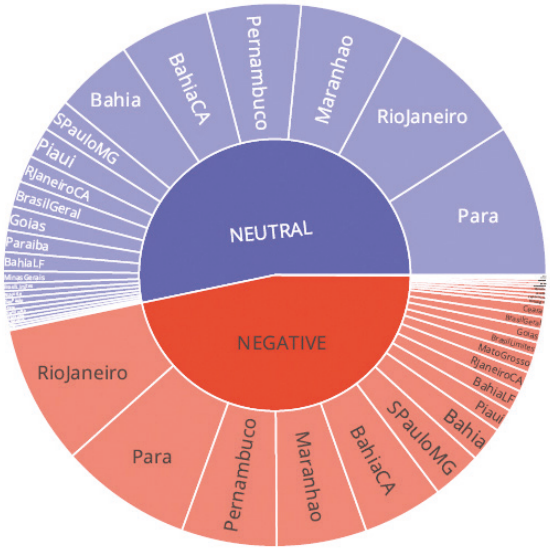
```
In [11]:
remove_list = ['CU-CaboVerde-Parcial.pdf', 'CU-Mocambique-Parcial.pdf', 'CU-Angola-Parcial.pdf', 'CU-SaoTome-Parcial.pd
root_data_br = root_data[~root_data.source_file.isin(remove_list)]
root_data_br.head(4)
```

Out[11]:

	Unnamed: 0	id_document	source_file	full_text	year	sentiment	companhia de jesus	jesuita	inaciano	jesuitico	any_term
130	130	239762	CU-Paraiba.pdf	1883- 1770, abril, 25, Paraíba OFÍCIO do gove...	1770	NEGATIVE	False	True	False	False	True
181	181	197199	CU-Maranhao.pdf	1758- 1729, Agosto, 8, São Luís do Maranhão C...	1758	NEGATIVE	False	True	False	False	True
221	221	270274	CU-RioJaneiro.pdf	5658- 1760, Março, 11, Rio de Janeiro OFÍCIO ...	1760	NEGATIVE	True	False	False	False	True
245	245	276557	CU-RioJaneiro.pdf	11947- ant. 1796, Setembro, 6 REQUERIMENTO de...	1796	NEUTRAL	False	True	False	False	True

Il. 12. Selekcja danych dotyczących tematu „jezuici" w Brazylii (kod i tabela)

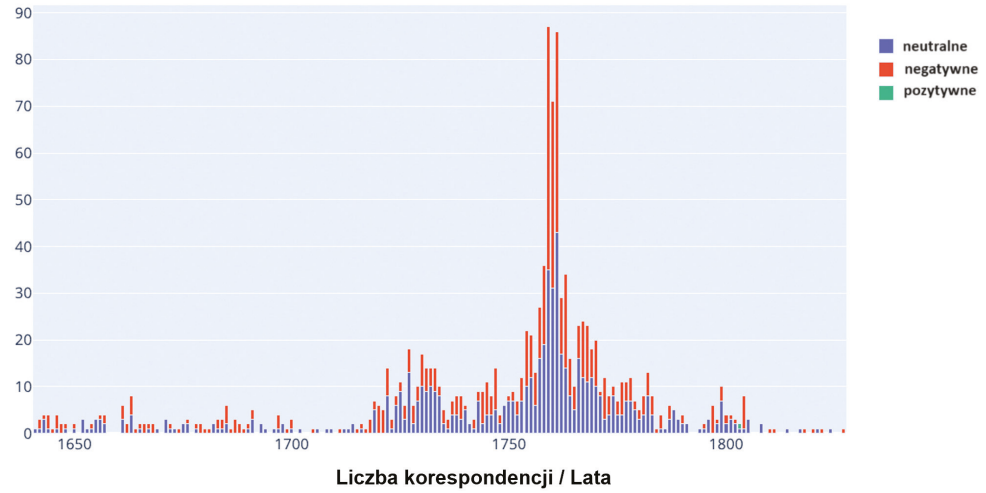
Wynik analizy sentymentu korespondencji dotyczącej jezuitów w różnych prowincjach Brazylii (zob. il. 13) nie wpisuje się w ogólną tendencję dotyczącą całego zbioru danych (zob. il. 8) oraz kolonii afrykańskich (zob. il. 10), w której przeważa negatywny wydźwięk. W przypadku Brazylii dominuje tendencja neutralna, chociaż w stanach, w których aktywność jezuitów była największa – Rio de Janeiro, Pará, Pernambuco, Maranhão i Bahia – wartości negatywne i neutralne są bardzo zbliżone.



Il. 13. Analiza sentymentu korespondencji dotyczącej jezuitów z podziałem na prowincje Brazylii

Jeśli przyjrzymy się analizie sentymentu korespondencji dotyczącej jezuitów w Brazylii, uwzględniając szczegóły z poszczególnych lat (zob. il. 14), zauważymy utrzymującą się tendencję negatywną w XVII w. oraz w okresie zbliżającym się do 1759 r. – daty wydalenia jezuitów z imperium portugalskiego. Ciekawym spostrzeżeniem jest fakt, że w pierwszej połowie XVIII w. moglibyśmy oczekiwać głównie negatywnego kontekstu, jednak obserwujemy przewagę sentymentu neutralnego, z okazjonalnymi wystąpieniami negatywnymi.

```
In [13]:
df_grouped = root_data_br.groupby(['year', 'sentiment']).count().reset_index()
fig = px.bar(df_grouped[(df_grouped.year>1640) & (df_grouped.year<1840)], x="year", y="id_document", color='sentiment',
             title='<b>Fig. 8:</b> Sentiment assessment of records with refferences to Jesuits',
             labels={"year": "Year","id_document": "Number of References"})
fig.show()
```

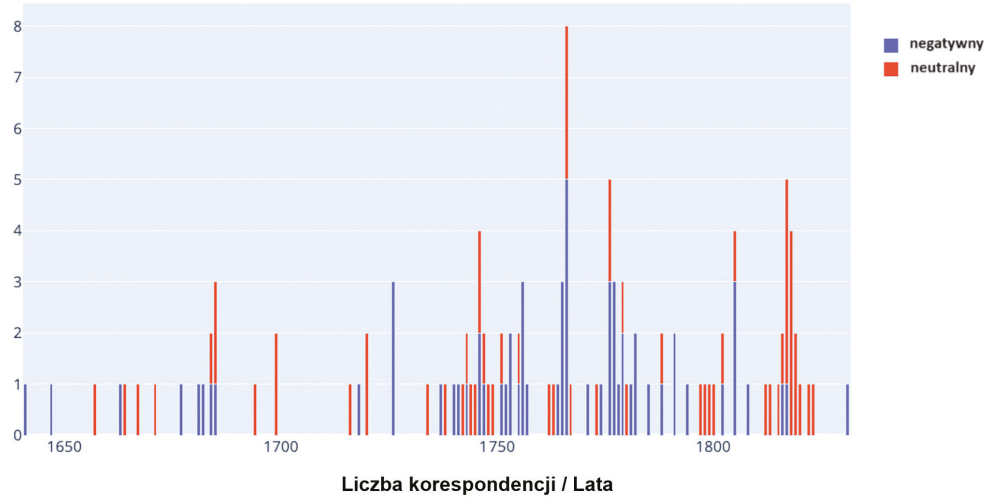


Il. 14. Analiza sentymentu korespondencji dotyczącej jezuitów w Brazylii z dokładnością do roku (kod i wykres)

Można byłoby zadać pytanie, czy ogólna dynamika negatywnego i pozytywnego sentymentu wobec jezuitów będzie również przekładać się na inne zakony, np. franciszkanów. Wróćmy jeszcze do il. 8 i przyjrzyjmy się latom poprzedzającym rok 1759 i po nim następującym. Tuż przed wydaleniem jezuitów i zaraz po tym wydarzeniu utrzymywał się sentyment negatywny. W następnych latach liczba korespondencji związanych z tym zakonem stopniowo spadała, a stosunek do niego stawał się coraz bardziej neutralny, aż do początków XIX w.

W przypadku franciszkanów (zob. il. 15) ogólna tendencja analizy sentymentu nieco różni się od tej dotyczącej jezuitów. W latach poprzedzających 1759 r. negatywne i neutralne emocje w kontekście, w którym pojawiają się franciszkanie, są bardzo zbliżone. Mogłoby się wydawać, że po wydaleniu jezuitów franciszkanie zdominują rzeczywistość kolonialną. Jednak aż do początku XIX w., gdy jezuici pojawiają się w kontekście coraz bardziej neutralnym, dzieje się na odwrót – w stosunku do franciszkanów utrzymuje się tendencja negatywna aż do pierwszych dwóch dekad XIX w. Neutralnie postrzegani będą oni dopiero w okolicach 1822 r. – czyli w czasie ogłoszenia przez Brazylię niepodległości i, co za tym idzie, zerwania tradycyjnych relacji z Portugalią.

```
root_data.to_excel("other_files/only_franciscans_references.xlsx", index=False)
df_grouped = root_data.groupby(['year', 'sentiment']).count().reset_index()
fig = px.bar(df_grouped[(df_grouped.year>1640) & (df_grouped.year<1840)], x="year", y="id_document", color='sentiment',
             title='<b>Fig. 4:</b> Sentiment assessment of records with refferences to Franciscans',
             labels={"year": "Year","id_document": "Number of References"})
fig.show()
```



Il. 15. Analiza sentymentu korespondencji dotyczącej franciszkanów (kod i wykres)

Część siódma: Modelowanie tematyczne

Do tej pory, poprzez analizę sentymentu, ujawniliśmy trwające tendencje negatywne związane z obecnością jezuitów w korespondencji kolonialnej. Dostrzegliśmy też różnice ze względu na czas i region. Przed wydaleniem jezuitów z imperium portugalskiego w Brazylii dominował sentyment neutralny, natomiast w Afryce – sentyment negatywny. Kolejnym krokiem w celu lepszej analizy obecności jezuitów w imperium portugalskim jest modelowanie tematyczne, które grupuje korespondencję ze względu na tematykę. W jaki sposób może to być użyteczne? Przykładowo modelowanie pozwala na analizę dynamiki tematów poruszanych przez jezuitów w obrębie imperium. Pomaga też zrozumieć, skąd mogły wynikać negatywne opinie w Afryce oraz co mogło przyczynić się do dominującego neutralnego nastawienia w Brazylii.

Dostępne algorytmy do modelowania tematycznego obejmują *Latent Dirichlet Allocation* (LDA), *Latent Semantic Indexing* (LSI) oraz *Gibbs Sampling Dirichlet Multinomial Mixture* (GSDMM). Zaobserwowaliśmy, że zwiększenie liczby grup tematycznych nie wpływa na zdolność modeli LDA i LSI do generowania lepszych wyników, co utrudnia identyfikację i ocenę dominujących tematów w grupach. Natomiast zauważaliśmy poprawę wyników uzyskanych dzięki modelowi GSDMM, co umożliwiło identyfikację głównych tematów związanych z terminami religijnymi, wojskowymi, administracyjnymi i gospodarczymi. Ponieważ algorytm GSDMM był jedyną metodą skupioną na analizie krótkich tekstów, aktualnie prowadzimy dodatkowe eksperymenty, wykorzystując cztery inne nowoczesne modele do analizy tego rodzaju tekstów.

Nasze wyniki wskazują na zróżnicowany udział źródeł dokumentów w poszczególnych tematach, co sugeruje istnienie powiązań między lokalizacjami a analizowanymi kwestiami. Jednak z uwagi na nierównomierność w zbiorze danych konieczne są dalsze eksperymenty. Być może

warto rozważyć podejście proporcjonalne, aby lepiej ocenić wpływ poszczególnych lokalizacji na omawiane tematy.

Model LSI

Algorytm LSI jest jednym z dostępnych podejść do modelowania tematów. Został wybrany ze względu na jego zdolność do radzenia sobie z synonimami i polisemią podczas grupowania. Poniższy zbiór danych prezentuje kilka przykładów słów w wykrytych klastrach (zob. il. 16).

```
In [91]: model_list, coherence_values = calc_coherence_values(dictionary=id2word, corpus=corpus_tfidf, texts=lemma_tex
                                                               gensim_model=LsiModel, start=start, limit=limit, step=st
         index = n_topics-2#coherence_values.index(max(coherence_values))
         result_lsi = format_topics_sentences(model_list[index], corpus_tfidf, list(root_data.full_text), lemma_text,
         result_lsi.head()
```

Il. 16. Modelowanie tematyczne na bazie algorytmu LSI (kod i rysunek)

Model GSDMM

Model GSDMM został stworzony z myślą o pracy z krótkimi dokumentami tekstowymi, takimi jak nasze. Podczas testowania zauważyliśmy, że algorytm ten uzyskał znacznie lepsze wyniki niż LDA i LSI. W poniższym zbiorze danych prezentujemy kilka przykładów słów w wykrytych klastrach dla kolonii afrykańskich i Brazylii:

```
In [53]: create_WordClouds(result_gsdmm, topic_names, n_topics, [3, 8])
```

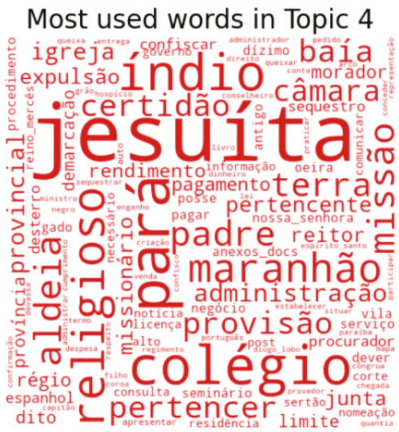
Il. 17. Modelowanie tematyczne na bazie algorytmu GSDMM dla grupy „jezuici w koloniach afrykańskich" – tematyka religijna (kod i rysunek)

Ilustracja 17 ukazuje modelowanie tematyczne w korpusie danych związanym z jezuitami. W pierwszej kolejności pojawiają się tematy związane z religią. Ta grupa zawiera terminy głównie związane z Kościołem, jego hierarchiczną organizacją oraz zakonami religijnymi. W tym przypadku dokumenty z Angoli, Republiki Zielonego Przylądka i Mozambiku stanowią odpowiednio 29,1%, 23,6% i 29,5% dokumentów.

```
In [55]: create_WordClouds(result_gsdmm, topic_names, n_topics, [13, 0, 11, 14, 17])
```

Il. 18. Modelowanie tematyczne na bazie algorytmu GSDMM dla grupy „jezuici w koloniach afrykańskich" – tematyka gospodarcza (kod i rysunek)

Ilustracja 18 przedstawia tematy wiążące jezuitów z handlem i finansami w Afryce. Ta grupa zawiera terminy związane z różnymi działalnościami gospodarczymi, takimi jak handel, w tym także handel niewolników, transakcje handlowe oraz działalność górnicza i rolnicza. Obejmuje również terminy dotyczące różnego rodzaju transakcji i płatności. Dokumenty pochodzące z Mozambiku stanowią większość w grupie kolonii afrykańskich (ok. 73% dokumentów).



Il. 19. Modelowanie tematyczne na bazie algorytmu GSDMM dla grupy „jezuici w Brazylii" (rysunek)

Ilustracja 19 przedstawia tematy wskazujące na powiązania jezuitów z ludnością autochtoniczną w Brazylii. Ta kategoria obejmuje terminy dotyczące przymusowej chrystianizacji ludności tubylczej w Brazylii, struktury misji jezuickich oraz *aldeias*, czyli wiosek zamieszkałych przez autochtonów. Modelowanie tematyczne wskazuje głównie na obecność jezuitów w rzeczywistości brazylijskich autochtonów oraz na kwestie podziału ziem. W tej grupie pojawiają się również tematy związane z władzą kolonialną, administracją i osadnictwem, które współtworzyły rzeczywistość północno-wschodniej Brazylii.

Potencjał narzędzi cyfrowych w analizie historycznej

Przedstawione przykłady zastosowania narzędzi cyfrowych na wybranym fragmencie naszych badań miały dwa główne cele. Po pierwsze, chcieliśmy zobrazować proces prowadzenia badań w dziedzinie historii cyfrowej od podszewki – jak bardzo kluczowe jest uporządkowanie źródeł oraz dobór odpowiednich modeli do analizy obszernego zbioru dokumentów. Po drugie, dążyliśmy do ukazania potencjału metod cyfrowych w stymulowaniu wyobraźni historyka oraz odkrywaniu aspektów, które mogłyby pozostać niezauważone przy tradycyjnych metodach badawczych. Szczególnie dotyczy to tak obszernego zbioru danych jak nasz, obejmującego prawie 170 tys. zapisów.

Nie oczekujemy, że narzędzia cyfrowe dostarczą gotowych odpowiedzi, ale mogą ujawnić procesy, które byłyby trudne do zauważenia w inny sposób. Jako studium przypadku skupiliśmy się na obecności jezuitów w imperium portugalskim – starając się krok po kroku wykazać, w jaki sposób narzędzia cyfrowe odsłaniały różne oblicza ich działalności w przestrzeni kolonialnej.

Analiza sentymentu pozwoliła nam wizualizować zmienność ogólnego kontekstu dotyczącego jezuitów – zarówno neutralnego, jak i negatywnego – w różnych okresach i lokalizacjach. Natomiast modelowanie tematyczne ukazało różnice w podejściu jezuitów do różnych kolonii – kwestie ekonomiczne i handlowe dominowały w Afryce, podczas gdy sprawy religijno-społeczne były bardziej znaczące w Brazylii.

Historia cyfrowa historią przyszłości?

W niniejszym rozdziale omawialiśmy zróżnicowane podejścia do historii cyfrowej, ze szczególnym uwzględnieniem perspektywy obliczeniowej. Staraliśmy się udowodnić, że rozpoczęcie eksploracji tego obszaru historii jest możliwe dla badaczy każdej epoki oraz specjalistów zajmujących się różnorodnymi źródłami historycznymi. Kluczowe pytanie, które wielokrotnie podkreślaliśmy w naszym tekście, dotyczyło tego, czy celem jest uzyskanie cyfrowych wyników przetwarzania danych czy też eksperymentowanie z algorytmami. Każda droga, którą wybierze historyk, jest słuszna w kontekście tworzenia historii cyfrowej. W przypadku pierwszego podejścia możemy pracować samodzielnie, natomiast w drugim potrzebujemy wsparcia interdyscyplinarnego zespołu.

Niniejszy tekst jest efektem naszych doświadczeń badawczych w tej dziedzinie. Staraliśmy się zawrzeć informacje, które sami chcielibyśmy posiadać na początku naszej przygody z badaniami nad historią obliczeniową. Dużo czasu spędziliśmy na zrozumieniu naszych źródeł i określeniu, czy wybrany przez nas materiał był wystarczający, a także co go wyróżniało i jakie były jego mankamenty pod kątem analizy komputerowej. Kolejne kroki obejmowały rozważania na temat

równorzędnych metod przetwarzania języka naturalnego na naszych tekstach, a następnie eksperymentowanie z algorytmami. Eksperymenty te wciąż trwają.

Co do pytania, czy historia cyfrowa jest historią przyszłości, nie możemy udzielić jednoznacznej odpowiedzi. To zależy od wielu czynników. Rozwój technologii, postęp w dziedzinie analizy danych oraz sposób, w jaki historycy będą chcieli integrować i interpretować informacje historyczne, będą miały istotny wpływ na kształtowanie historii w przyszłości.

Bibliografia

Ahnert R. et al., *Collaborative Historical Research in the Age of Big Data. Lessons from an interdisciplinary project. Elements in historical theory and practice*, Cambridge 2023.

Antonijević S., *Amongst Digital Humanists. An ethnographic study of digital knowledge production*, New York 2015.

Bastian M., Heymann S., Jacomy M., *Gephi. An open source software for exploring and manipulating networks*, „Proceedings of the International AAAI Conference on Web and Social Media" 2009, 1(3).

Błoch A., *The 'Miserable Vassals' of the Empire. The androgynous codes of behaviour of black and indigenous peoples in late colonial Brazil (1775–1808)*, „Journal of History" 2022, 3(57).

Błoch A., Vasques Filho D., Bojanowski M., *Networks from Archives. Reconstructing networks of official correspondence in the early modern Portuguese empire*, „Social Networks" 2022, nr 69.

Cherian J.P., Kizhakkethottam J.J., Alexander A. A., *A Comparative Review on Different Social Network Analytical Tools*, „Information and Communication Technologies: 8th Conference, TICEC 2020, Guayaquil, Ecuador, November 25–27, 2020, Proceedings" 2020.

Elo K., *Utilizing Historical Network Analysis on Meta-Data to Model East German Foreign Intelligence Cycle in the Baltic Sea Region 1975–89*, [w:] *The Power of Networks. Prospects of historical network research*, red. F. Kerschbaumer et al., Abingdon 2020.

Faysal M.A.M., Arifuzzaman S., *A Comparative Analysis of Large-Scale Network Visualization Tools*, „IEEE International Conference on Big Data (Big Data)" 2018.

Fertig C., *Kinship Networks in North Western German Rural Society (18th/19th Centuries)*, [w:] *The Power of Networks. Prospects of historical network research*, red. F. Kerschbaumer et al., Abingdon 2020.

Gramsch-Stehfest R., *Entangled Powers. Network analytical approaches to the history of the Holy Roman Empire during the late Staufer period*, „German History" 2018, 3(36).

Gramsch-Stehfest R., *Network Analytical Modelling of Political Structures and Actions in the Middle Ages*, [w:] *The Power of Networks. Prospects of historical network research*, red. F. Kerschbaumer et al., Abingdon 2020.

Haggerty J., Haggerty S., *The Life Cycle of a Metropolitan Business Network. Liverpool 1750–1810*, „Explorations in Economic History" 2011, 2(48).

Jia Ye M., *A History from Below. Translators in the publication network of four magazines issued by the China Book Company, 1913–1923*, „Translation Studies" 2022, 1(15).

Kuckartz U., Rädiker S., *Analyzing Qualitative Data with MAXQDA. Text, audio, and video*, Switzerland 2019.

Pereira Antunes A., *Social Network Analysis in the History of Sciences. Visualising sociability in scientific expeditions with Gephi*, „Humanidades Digitales en tiempos convulsos AAHD" 2021, 1(2), s. 15–16.

Petz C., Pfeffer J., *Configuration to Conviction. Network structures of political judiciary in the Austrian Corporate State*, „Social Networks" 2021, nr 66.

Rani P., Shokeen J., *A Survey of Tools for Social Network Analysis*, „International Journal of Web Engineering and Technology" 2021, 3(16).

Rollinger C., *Networking the Res Publica. Social network analysis and Republican Rome*, [w:] *The Power of Networks. Prospects of historical network research*, red. F. Kerschbaumer et al., Abingdon 2020.

Salmi H., *What Is Digital History?*, Cambridge 2020.

Tamm M., Burke P. (red.), *Debating New Approaches to History*, London 2018.

Wurpts B., *The Value of Network Analysis in Historical Sociology. Economic and social relations in medieval Lübeck*, [w:] *The Power of Networks. Prospects of historical network research*, red. F. Kerschbaumer et al., Abingdon 2020.

Zbíral D., Shaw R.L.J., *Hearing Voices. Reapproaching medieval inquisition records*, „Religions" 2022, 12(13).

dr Agata Błoch, Instytut Historii im. Tadeusza Manteuffla Polskiej Akademii Nauk, e-mail: abloch@ihpan.edu.pl

dr Clodomir Santana, Instytut Historii im. Tadeusza Manteuffla Polskiej Akademii Nauk, e-mail: csantana@ihpan.edu.pl

Summary

From Digital History to Computational History: A Case Study of the Colonial Portuguese Empire

Our chapter focuses on various approaches to digital history, emphasizing a computational perspective. The aim is to demonstrate that it is accessible to researchers specializing in different historical periods and working with various sources. The key question we repeatedly address concerns the research intent: is the goal to obtain digital results from data processing or to experiment with algorithms? Each path chosen by the historian has its justification in the context of creating digital history. In the first approach, independent work is possible, whereas in the second, support from an interdisciplinary team is required. In the subsequent section of the article, we present examples of computational history based on our case study of the colonial Portuguese Empire. We demonstrate the functionality of two models: sentiment analysis and topic modeling, presenting graphs, codes, and their practical application for both digital research practice and historians.

Translated by Olgierd Dróżdż