

SKAITMENINIŲ METODŲ TAIKYMAS PRAETIES KELIONIŲ TYRIMUOSE

Rimvydas Laužikas

Vilniaus universiteto Komunikacijos fakultetas
Saulėtekio al. 9, LT-10222, Vilnius, Lietuva
El. paštas rimvydas.lauzikas@kf.vu.lt

Santrauka. *Pastaraisiais dešimtmečiais vykstantis masyvus rašytinių istorijos šaltinių skaitmeninimas, jų tekstų optinis atpažinimas (OCR) bei pateikimas internete sukuria naujas istorinio tyrimo galimybes ir iššūkius. Šiame straipsnyje pristatomas skaitmeninės humanitarikos tyrimo modelis yra grindžiamas informacijos organizavimo paradigmos ir skaitmeninėmis technologijomis grįstų metodų taikymu praeities kelionių tyrimuose. Tyrimo modelis buvo sukurtas ir testuotas naudojant projekto „Homo viator: kelionių erdvė ir keliautojų potyriai ankstyvosios modernybės Lietuvoje“ medžiagą.*

Pagrindinė tyrimo problema yra susijusi su tuo, kad vienas svarbių praeities kelionių pažinimo šaltinių yra egodokumentai (laiškai, dienoraščiai, atsiminimai ir kt.), kuriuose greta kitų konkretaus asmens gyvenime svarbių įvykių aprašomos ir kelionės. Tačiau kelionių aprašymai sudaro tik nedidelę konkretaus egodokumento teksto dalį ir yra netolygiai pasiskirstę skirtinguose egodokumentuose. Todėl, atsižvelgiant į egodokumentų ir jų rinkinių teksto apimtį bei įvairiomis kalbomis publikuotų egodokumentų skaičių, jų – kaip šaltinių – tyrimai tik vienu (kelionių) aspektu reikalauja

Tyrimą finansavo Lietuvos mokslo taryba (LMTLT), sutarties Nr. S-MIP-21-44.

Received: 2023 01 10. Accepted: 2023 09 23

Copyright © 2023 Rimvydas Laužikas. Published by Vilnius University Press. This is an Open Access journal distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

didelių žmogiškųjų ir laiko resursų. Panašių probleminę situaciją turime ir su kitais senųjų kelionių pažinimo šaltiniais: skaitmeniniu pavidalu (įskaitant OCR) publikuotų dokumentų yra daug, jų tekstai didelės apimties, o teksto fragmentai, susiję su kelionėmis, šalių aprašymais, senųjų kelionių maršrutais, kelionių ir mobilumo infrastruktūra, keliautojų potyriais, yra santykinai nedideli ir išsibarstę visame šaltinio tekste. Straipsnyje aprašomą tyrimo modelį sudaro du žingsniai: (i) šaltinių tekstyno su optiškai atpažintu turiniu (OCR) surinkimas; (ii) empirinių duomenų rinkimas taikant žodynu grįstos kompiuterizuotos kokybinės teksto analizės metodą, realizuojant jį su MaxQDA programa. Tekstyno surinkimas yra atliekamas taikant bendruosius mokslinių publikacijų paieškos internete principus bei metodus. Į tekstyną atrenkami autentiški, publikuoti, tyrimui reikalingi šaltiniai (laiškai, dienoraščiai, atsiminimai ir kt.) bei mokslinės publikacijos apie juos, tokiu būdu suformuojant du tekstyno – šaltinių ir literatūros – blokus. Literatūros (mokslinių publikacijų apie šaltinius) blokas naudojamas kaip papildoma medžiaga tikslesnei šaltinių tekstyno atrankai ir interpretavimui. Norint taikyti žodynu grįstos kompiuterizuotos kokybinės teksto analizės metodą svarbiausias elementas yra kokybiškas žodynas, tiksliai apibūdinantis tyrimui aktualius konceptus (kategorijas). Atsižvelgiant į šaltinių specifiką (šaltiniuose ir jų vertimuose vartojamas kalbas) buvo sukurtas daugiakalbis lietuvių–lenkų–anglų–rusų–vokiečių kalbų žodynas. Žodyno struktūrą sudaro šeši su praeities kelionėmis susiję konceptai (kategorijos): (i) kelionė (bendras apibūdinimas), (ii) kelias ir jo infrastruktūra (tiltai, brastos ir kt.), (iii) transporto priemonės, (iv) sustojimo ir nakvynės vietos (miesteliai, kaimai, karčemos, pašto stotys ir kt.), (v) kelyje sutikti žmonės (smuklininkai, plėšikai, vedliai ir kt.), (vi) kelionių maistas. Kiekvienas konceptas yra apibūdinamas reikšminių žodžių ir frazių rinkiniu. Paskutiniame tyrimo etape buvo atliktas tyrimo modelio testavimas. Testavimo metu nustatyta, kad modelis leidžia spręsti paminėtas, projekto įgyvendinimo metu atsiradusias, problemas.

Reikšminiai žodžiai: ankstyvosios modernybės laikų kelionės, skaitmeniniai humanitariniai mokslai, tarpdisciplininiai tyrimai, žodynu grįsta kompiuterizuota kokybinė teksto analizė.

Application of Digital Methods in the Study of Ancient Travel

Summary. *The massive digitisation of written historical sources, optical character recognition (OCR) of texts, and their online availability in recent decades*

have created new opportunities and challenges for historical research. The digital humanities research model presented in this paper is based on the information organisation paradigm and the application of digital technology-based methods in studying ancient travels. The model has been developed and tested using the materials of the project "Homo Viator: Travel Space and Travellers' Experiences in Early Modern Lithuania". The main problem of the research is related to the fact that one of the essential sources of information about ancient travel are egodocuments (letters, diaries, memoirs, etc.) that contain journeys described alongside other important life events of a particular person. However, travel descriptions form only a small part of a text of a given egodocument and are unevenly distributed among different egodocuments. Therefore, given the size of the text of the egodocuments and their collections and the number of egodocuments published in different languages, researching them as sources in only one aspect (travel) requires a significant amount of human and time resources. A similar problematic situation exists with other sources of knowledge on ancient travel: a massive number of documents published in digital form (including OCR), their texts are voluminous, and the text fragments related to travel, country descriptions, ancient travel routes, travel and mobility infrastructure, and travellers' experiences are relatively small and scattered throughout the source text. The research model described in the paper is divided into two steps: (i) collection of the corpus of OCR source texts; (ii) collection of empirical data using a dictionary-based computer-aided [or assisted] qualitative text analysis method implemented with the MaxQDA software. The collection of the source text corpus is carried out by applying the general principles and methods of online search of scientific publications. The corpus comprises authentic, published sources relevant to the study (letters, diaries, memoirs, etc.) and scholarly publications about them, thus forming two blocks of text - sources and literature. The literature block is used as additional material for a more precise selection and interpretation of the source texts. A key element for applying a dictionary-based computer-aided [or assisted] qualitative text analysis method is a high-quality dictionary that accurately describes the concepts (categories) relevant to the research. Considering the specificity of the sources (the languages used in the sources and their translations), a multilingual dictionary (Lithuanian-Polish-English-Russian-German) was compiled. The structure of the dictionary consists of six concepts (categories)

related to ancient travel: (i) journey (general description), (ii) road and its infrastructure (bridges, fords, etc.), (iii) means of transportation, (iv) resting and accommodation places (towns, villages, taverns, post offices, etc.), (v) people encountered on the way (inn-keepers, highwaymen, guides, etc.), and (vi) food of the journey. A set of keywords and phrases describes each concept.

In the last stage of the study, the research model was tested. The testing showed that the model could solve the above problems that arose during the project.

Keywords: *travel in Early Modern Period, digital humanities, interdisciplinary research, dictionary-based computer-aided [or assisted] qualitative text analysis.*

ĮVADAS

Skaitmeninių informacijos valdymo ir komunikacijos technologijų (IT) paplitimas XX a. pab. – XXI a. pirmaisiais dešimtmečiais suteikė naujų galimybių daugelyje veiklos sričių. Taip pat ir humanitarinių mokslų (HM) empirinių duomenų rinkimo (fiksavimo), tyrimo ir komunikacijos procesuose¹. IT poveikį HM galėtume nagrinėti mažiausiai trimis lygmenimis: naujo įrankio, naujo metodo ir naujos paradigmos. Įrankio lygmenyje IT pavertė masiniais, kiekvienam tyrėjui prieinamais kai kuriuos empirinių duomenų rinkimo ir tyrimo metodus bei priemones, kurie HM jau buvo gana seniai žinomi ir taikomi, bet reti, nes reikalavo išskirtinių mokslininko kompetencijų ar specifinės įrangos. Kalbant apie empirinių duomenų rinkimą tokių įrankių pavyzdžiai galėtų būti spalvotoji fotografija, vaizdo filmavimas, garso įrašymas. Kalbant apie tyrimus pavyzdys galėtų būti statistinės duomenų analizės metodai, kurių taikymas HM iki skaitmeninių statistinės analizės programų išpopuliarėjimo buvo gana retas. Iš dalies prie šios grupės galėtume priskirti ir optinio teksto atpažinimo (OCR) programas bei skaitmenines duomenų bazes ir informacines sistemas. Ši – įrankio lygmens – pokytį ir jo paskatintus mokslinius rezultatus galime laikyti labiau kiekybiniais, nes taikomuoju požiūriu, pavyzdžiui, tikrovė, užfiksuota juostiniu fotoaparatu, santykinai nedaug skiriasi nuo tikrovės, užfiksuotos skaitmeniniu

1 ROMEIN, C. Annemieke; KEMMAN, Max; BIRKHOLZ, Julie M.; BAKER, James; DE GRUIJTER, Michel; MEROÑO-PEÑUELA, Albert; RIES, Thorsten; ROS, Ruben; SCAGLIOLA, Stefania. State of the Field: Digital History. *History*, 2020, vol. 105, p. 291–312. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <<https://doi.org/10.1111/1468-229X.12969>>.

fotoaparatu. Žinoma, skaitmeninės technologijos ir čia suteikė tyrėjui naujų galimybių (pavyzdžiui, skaitmeninio vaizdo filtravimas, automatinis vaizdo atpažinimas ir pan.), tačiau jos savo esme buvo susijusios su jau egzistuojančiais ir taikomais konkretais HM metodais, daugiau didindamos mokslininko veiklos efektyvumą nei keisdamos jo turimą „įrankių dėžutę“².

Kur kas didesnis pokytis HM tyrimuose yra susijęs su naujų, IT grįstų empirinių duomenų rinkimo ir analizės metodų atsiradimu (neretai – tarpdisciplininio pobūdžio metodų perkėlimu iš kitų mokslų)³. Kalbant apie empirinių duomenų fiksavimą tokio metodo pavyzdys galėtų būti 3D skenavimas, o tyrimuose (tai yra aktualu šio straipsnio kontekste) – geografinėmis informacijos sistemomis (GIS) ir atitinkama programine įranga (pavyzdžiui, ArcGIS) grįstų metodų pritaikymas. Skaitmeninių ir lazerinių technologijų sankirtoje atsiradęs trimatis skeneris įgalina ne tik fiksuoti maksimaliai tikslų tikrovės vaizdą, bet ir pritaikyti iš esmės naujus, anksčiau neegzistavusius duomenų tyrimo metodus⁴. Geografinės informacinės sistemos (GIS) apibrėžiamos kaip „<...>geografinių objektų, jų charakteristikų ir kitos su Žeme susijusios informacijos kaupimo, tvarkymo, apdorojimo, saugojimo, paieškos ir pateikimo kompiuterizuota informacinė sistema, skirta projektavimo, modeliavimo, analizės, mokslo ir kitiems geografinės erdvės uždaviniams spręsti <...>“⁵. Iš esmės tai – techninės bei programinės įrangos, geografinių duomenų ir dirbančio personalo

- 2 HOEKSTRA, Rik; KOOLEN, Marijn. Data Scopes for Digital History Research. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*. 2019, vol. 52 (2), p. 79–94. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <10.1080/01615440.2018.1484676>.
- 3 BLANEY, Jonathan; MILLIGAN, Sarah; STEER, Marty; WINTERS, Jane. *Doing Digital History*. Manchester, England: Manchester University Press, 2021. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <https://doi.org/10.7765/9781526157713>; MOILANEN, Inka. Virtues, Vices, and Vectors : Digital Tools and the Study of Medieval Sermons. *Digitala Vetenskapliga Arkivet*. 2019. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-167353>.
- 4 ŽIŽIŪNAS, Tadas. *Technologinis veiksnys kultūros paveldo tyrimuose: 3D vaizdo ir spektroskopijos taikymo metodologinis modelis*. Daktaro disertacija. Vilnius: Vilniaus universitetas, 2019; ŽIŽIŪNAS, Tadas; AMILEVIČIUS, Darius. 3D and AI Technologies for the Development of Automated Monitoring of Urban Cultural Heritage. *Digital Humanities in the Nordic Countries 2020: Proceedings of the Digital Humanities in the Nordic Countries: 5th Conference (DHN 2020)*. 2020, p. 364–372. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <https://ceur-ws.org/Vol-2612/short23.pdf>.
- 5 GULBINAS, Zenonas. Geografinė informacinė sistema. Iš *Visuotinė lietuvių enciklopedija*. Vilnius: Mokslo ir enciklopedijų leidybos centras, 2022. [žiūrėta 2022 m. gruodžio 31 d.]. Prieiga per internetą: <https://www.vle.lt/straipsnis/geografine-informacine-sistema/>.

visuma, sukurta efektyviai dirbti su visomis erdvėje orientuotos informacijos formomis. Pagrindinis GIS pranašumas – operavimas erdvine (koordinuota, erdvėje orientuota) informacija. Informacija apie tam tikrą erdvinį objektą ar jo dalį gali būti pateikiama kaip keletas konkrečios teritorijos teminių žemėlapių, atspindinčių tam tikrą vietovės bruožą kartu su galimybe analizuoti tą informaciją įvairiais skirtingais pjūviais. Šios galimybės itin svarbios istorijos tyrimuose, kur erdvė yra viena pagrindinių tyrimo dimensijų⁶ ir ypač tada, kuomet nagrinėjame praeities keliones bei jų aprašymus⁷.

IT pritaikymas HM erdvėje skatina ir dar gilesnius – paradigmų⁸ lygmens – pokyčius. Šio straipsnio kontekste yra svarbus IT paskatintas istorijos (kaip žmonijos praeities) tyrimų poslinkis informacijos mokslų paradigmos linkme⁹,

- 6 LOZANO, Jorge Sebastián. Mapping Art History in the Digital Era. *The Art Bulletin*, 2021, vol. 103 (3), p. 6–16. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <10.1080/00043079.2021.1882819>.
- 7 MURRIETA-FLORES, Patricia; DONALDSON, Christopher; GREGORY, Ian. GIS and Literary History: Advancing Digital Humanities Research through the Spatial Analysis of Historical Travel Writing and Topographical Literature. *Digital Humanities Quarterly*, 2017, vol. 11 (1). [žiūrėta 2023 m. sausio 10 d.]. Prieiga per internetą: <https://chesterrep.openrepository.com/handle/10034/620256>; BENCH, Harmony; ELSWIT, Kate. Mapping Movement on the Move: Dance Touring and Digital Methods. *Theatre Journal*, 2016, vol. 68, no. 4, p. 575–596. [žiūrėta 2023 m. sausio 10 d.]. Prieiga per internetą: <http://www.jstor.org/stable/26367451>; BENCH, Harmony; ELSWIT, Kate. Dance Touring and Embodied Data Some Approaches to Katherine Dunham's Movement on the Move. *Current Research in Digital History*, 2019, vol. 2. [žiūrėta 2023 m. sausio 10 d.]. Prieiga per internetą: <https://doi.org/10.31835/crdh.2019.12>; BÜRGERMEISTER, Martina. Measuring Urban Change in Travel Texts Using the Example of the City of Graz in the Long Nineteenth Century. *Magazén*, 2022, vol. 3 (1), p. 61–90. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <https://edizionicafoscari.unive.it/media/pdf/article/magazen/2022/1/art-10.30687-mag-2724-3923-2022-05-003.pdf>; ANDERSON, Carrie; CESERANI, Giovanna; DONALDSON, Christopher; GREGORY, Ian N.; HALL, Melanie; ROSENBAUM, Adam T.; TAYLOR, Joanna E. Digital Humanities and Tourism History. *Journal of Tourism History*, 2017, vol. 9 (2–3), p. 246–269. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <10.1080/1755182X.2017.1419455>; MIDURA, Rachel. Itinerating Europe: Early Modern Spatial Networks in Printed Itineraries, 1545–1700. *Journal of Social History*, 2021, vol. 54, iss. 4, p. 1023–1063. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <https://doi.org/10.1093/jsh/shab011>.
- 8 Paradigmą suprantant kaip konceptualią pasaulėžiūrą, sudarytą iš formalių teorijų, klasikinių eksperimentų ir patikimų metodų (pagal: KUHN, Thomas S. *Mokslo revoliucijų struktūra*. Vilnius: Pradai, 2003).
- 9 HUVILA, Isto. *The Ecology of Information Work: A Case of Bridging Archaeological Work and Virtual Reality Based Knowledge Organisation*. Åbo: Åbo Akademi University Press, 2006. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <http://urn.fi/URN:ISBN:951-765-337-9>.

tyrimui pritaikant informacinės veiklos (*information work*) ir žinių veiklos (*knowledge work*) konceptus. Abiejų konceptų bendrieji apibūdinimai mokslinėje literatūroje yra skirtingi ir gana neapibrėžti¹⁰. Isto Huvila apibūdina informacijos veiklas kaip dvejopus tikrovę aprašančius ir kritiškai vertinančius informacinius mechanizmus, kurie daro įtaką socialiniams veiksams, konceptualizuojamiems kaip „veiklos“¹¹, o žinių veiklas – kaip žmonių grupių ir visuomenių, įgyvendinančių žinių reikalaujančius tikslus, veiklas¹². Istorijos tyrimų požiūriu svarbūs taikomieji informacijos ir žinių veiklų aspektai, sietini su šaltinių duomenų ir informacijos organizavimu, istorijos mokslo žinių organizavimu bei skaitmeninių technologijų panaudojimu šiose veiklose. Atsižvelgiant į šį kontekstą, straipsnyje pristatomas modelis gali būti vertinamas kaip tarpdisciplininis, informacijos organizavimo paradigma ir skaitmeninėmis technologijomis grindžiamas skaitmeninės humanitarikos tyrimo modelis. Jis patenka į platesnį tyrimų lauką – skaitmeninius istorijos tyrimus, klasifikuojant šią lauko dalį kaip metodus ir įrankius, kurie skatina „pokyčius fiziniame tyrimo procese“. Detalizuojant juos atkreipiamas dėmesys į plačiai vykdomą archyvų rinkinių katalogų ir pačių dokumentų skaitmeninimą, istoriko galimybes naudoti skaitmeninę fotografiją archyve ir dalintis šiais vaizdais su kitais bendruomenės nariais¹³. Tačiau kalbant tiek apie dabartį, tiek apie perspektyvas akcentuojama dokumentų prieiga per skaitmeninimą, bet ne galimybės taikyti automatinius ar pusiau automatinius metodus bei įrankius, įgalinančius rinkti

- 10 HILL, Michael W. *The Impact of Information on Society*. London: Bowker-Saur, 1999; KUHLTHAU, Carol C. Towards Collaboration between Information Seeking and Information Retrieval. *Information Research*, 2005, vol. 10 (2); JACOBS, Ronald L. Knowledge Work and Human Resource Development. *Human Resource Development Review*, 2017, vol. 16 (2), p. 176–202. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <doi:10.1177/1534484317704293>; PYÖRIÄ, Pasi. The Concept of Knowledge Work Revisited. *Journal of Knowledge Management*, 2005, vol. 9 (3), p. L116–L127. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <10.1108/13673270510602818>.
- 11 HUVILA, Isto. How a Museum Knows? Structures, Work Roles, and Infrastructures of Information Work. *Journal of the American Society for Information Science and Technology*, 2013, vol. 64, p. 1375–1387. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <https://doi.org/10.1002/asi.22852>.
- 12 HUVILA, Isto. *The Ecology of Information Work: A Case of Bridging Archaeological Work and Virtual Reality Based Knowledge Organisation*. Åbo: Åbo Akademi University Press, 2006. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <http://urn.fi/URN:ISBN:951-765-337-9>.
- 13 MILLIGAN, Ian. *The Transformation of Historical Research in the Digital Age* (Elements in Historical Theory and Practice). Cambridge: Cambridge University Press, 2022.

duomenis iš pačių suskaitmenintų archyvinių dokumentų¹⁴. Šios perspektyvos aptiriamos ne tiek istorijos mokslo tyrimų, kiek informacijos ir žinių organizavimo priegios taikymo bendresniame skaitmeninės humanitarikos kontekste¹⁵. Isto Huvilo tyrimuose susiejama skaitmeninėmis technologijomis grįstas informacijos ir žinių organizavimas su archeologijos mokslo žiniomis¹⁶.

Taikomuoju požiūriu, šiame straipsnyje pristatomas tyrimų modelis skaitmeninės humanitarikos erdvėje yra artimiausias kompiuterizuotų teksto analizės¹⁷ ir teksto temų automatinio indeksavimo bei klasifikavimo metodų¹⁸ grupėms. Pagrindinis skirtumas yra tai, jog straipsnyje pristatomo tyrimo modelio tikslas yra ne surasti ir paimti pavienius žodžius, siekiant kurti ar turtinti metaduomenų rinkinį, bet identifikuoti teksto pastraipą, kuri yra potencialiai naudinga istoriko tyrimui. Kitas svarbus skirtumas yra tas, jog straipsnyje pristatomam tyrimo modeliui yra naudojami visiškai standartizuoti ir daugeliui mokslininkų žinomi kompiuteriniai tyrimo įrankiai (pavyzdžiui, kokybinės turinio analizės programa MaxQDA).

Informacijos mokslų paradigmos pritaikymas leidžia ieškoti optimalių istorijos tyrimo metodologinių sprendimų, pritaikant informacijos vertės konceptą ir jam realizuoti panaudojant skaitmeninius IT įrankius. Šiuo požiūriu mokslo tyrimo metu sukuriama vertė suprantama kaip numatoma (tikėtina) naujų žinių vertė, o informacija – kaip mokslinė informacija, leidžianti generuoti tą naują žinių vertę. Mokslinė informacija gali būti apibrėžiama kaip mokslo tyrimui svarbūs organizuoti (struktūruoti) empiriniai duomenys ir jų tarpusavio ryšiai, apibūdinantys tam tikrą tiriamą tikrovės reiškinį ar situaciją.

- 14 SIEBOLD, Anna; VALLERIANI, Matteo. Digital Perspectives in History. *Histories*, 2022, vol. 2 (2), p. 170–177.
- 15 GOLUB, Koraljka; LIU, Ying-Hsang. *Information and Knowledge Organisation in Digital Humanities Global Perspectives*. New York: Routledge, 2022.
- 16 HUVILA, Isto; HUGGETT, Jeremy. Archaeological Practices, Knowledge Work and Digitalisation. *Journal of Computer Applications in Archaeology*, 2018, vol. 1 (1), p. 88–100. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <10.5334/jcaa.6>.
- 17 EASTON, Evan; ALLEN, Calabria; ALLEN, William. Computational Text Analysis and Archival Methods. In *International Organizations and Research Methods*. An Arbor: University of Michigan Press, 2023, p. 285–291.
- 18 GOLUB, Koraljka. Automatic Subject Indexing of Text. *Knowledge Organization*, 2019, vol. 46 (2), p. 104–121; GOLUB, Koraljka. Automatic Identification of Topics: Applications and Challenges. In *Doing Digital Humanities: Concepts, Approaches, Cases*. Växjö: Linnaeus University Press, 2022, p. 5–26; GOLUB, Koraljka; SOERGEL, Dagobert; BUCHANAN, George; TUDHOPE, Douglas; LYKKE, Marianne; HIOM, Debra. A Framework for Evaluating Automatic Indexing or Classification in the Context of Retrieval. *Journal of the Association for Information Science and Technology*, 2016, vol. 67 (1), p. 3–16.

Ekonominiu (taip pat ir mokslo) požiūriu, viena svarbių informacijos savybių yra potencialas ją panaudoti kuriant pridedamąją naujų žinių vertę. Ši tikėtina vertė gali būti apibrėžiama dviem būdais: (i) kiekybiniu naujų mokslui dar nežinomų informacijos ir žinių kiekio lyginimu su jau turima informacija ir žiniomis bei (ii) sukauptos naujos informacijos ir žinių potencialu kurti pridedamąją inovacijų, kultūrinio, socialinio ar kito visuomenei svarbaus kapitalo vertę¹⁹. Pritaikydami šį požiūrį praeities kelionių tyrimuose galime kalbėti apie priklausomybę tarp empirinių duomenų rinkimo bei tyrimo kainos ir tikėtinos naujos mokslinės informacijos bei žinių vertės. Pažymėtina, kad istorinių tyrimų kontekste kaina gali būti suprantama kaip tiesioginė (finansinė) metodo taikymo kaina ar (ir) laiko bei kitų resursų sąnaudos, reikalingos metodui pritaikyti. Šiuo požiūriu galime skirti brangius (resursams imlius) ir pigesnius metodus bei daugiau naujos informacijos ir žinių suteikiančius (informatyvius) ir mažiau informatyvius empirinių duomenų rinkimo ir tyrimų metodus.

Tokio požiūrio ir su juo susijusių skaitmeninių informacijos organizavimo sprendimų poreikį grindžia projekte „*Homo viator: kelionių erdvė ir keliautojų potyriai ankstyvosios modernybės Lietuvoje*“ išryškėjusi problema. Vienas svarbių praeities kelionių pažinimo šaltinių yra egodokumentai (laiškai, dieno-raščiai, atsiminimai ir kt.), kuriuose greta kitų konkretaus asmens gyvenime svarbių įvykių aprašomos ir kelionės. Tačiau kelionių aprašymai sudaro tik nedidelę konkretaus egodokumento teksto dalį ir yra netolygiai pasiskirstę skirtinguose egodokumentuose. Dėmesys kelionės aprašymui priklausė nuo egodokumentą sukūrusio asmens: vienais atvejais apie keliones neužsimenama ar užsimenama fragmentiškai, o kitais – pateikiamas pakankamai detalus jų aprašymas. Atsižvelgiant į egodokumentų ir jų rinkinių teksto apimtį bei įvairiomis kalbomis publikuotų egodokumentų skaičių, pasakytina, kad jų – kaip šaltinių – tyrimai tik vienu (kelionių) aspektu reikalauja didelių žmogiškųjų ir laiko resursų. Panašią probleminę situaciją turime ir su kitais senųjų kelionių pažinimo šaltiniais: skaitmeniniu pavidalu (įskaitant OCR) publikuotų dokumentų yra daug, jų tekstai didelės apimties, o teksto fragmentai, susiję su kelionėmis, šalių aprašymais, senųjų kelionių maršrutais, kelionių ir mobilumo infrastruktūra, keliautojų potyriais, yra santykinai nedideli ir išsibarstę visame šaltinio tekste. Šiame straipsnyje pristatomo tyrimo tikslas yra parengti ir testuoti skaitmeninės humanitarikos tyrimo modelį, skirtą empiriniams duomenims apie senąsias keliones rinkti ir analizuoti. Modelis yra grindžiamas sąvei-

19 KUNCEVIČIUS, Albinas; LAUŽIKAS, Rimvydas; JANKAUSKAS, Rimantas; AUGUSTINAVIČIUS, Renaldas; ŠMIGELSKAS, Ramūnas. *Dubingių mikroregionas ir Lietuvos valstybės ištakos*. Vilnius: Petro ofsetas, 2015.

ka tarp informacijos mokslams būdingo informacijos organizavimo koncepto bei skaitmeninių IT įrankių pritaikymo istorijos mokslo tyrimuose. Modeliui testuoti panaudoti realūs, projekte „*Homo viator*: kelionių erdvė ir keliautojų potyriai ankstyvosios modernybės Lietuvoje“ naudoti šaltiniai bei projektą vykdančių mokslininkų Arvydo Pacevičiaus, Tomo Čelkio, Axelio Walterio, Tomášo Božerockio ir straipsnio autoriaus istorijos šaltinių tyrimo bei tarpdisciplininių tyrimų patirtys.

Straipsnyje pristatomas tyrimas atliktas įgyvendinant projektą „*Homo viator*: kelionių erdvė ir keliautojų potyriai ankstyvosios modernybės Lietuvoje“. Tyrimą finansuoja Lietuvos mokslo taryba (LMTLT), sutarties Nr. S-MIP-21-44.

TYRIMO MODELIS

Straipsnyje pristatomas tyrimo modelis yra suformuotas per tarpdisciplininę sąveiką tarp humanitarinių ir socialinių mokslų, pritaikant socialiniams mokslams būdingą kokybinės turinio analizės metodą praeities kelionių tyrimams. Tyrimo modelį sudaro du žingsniai: (i) šaltinių tekstyno su optiškai atpažintu turiniu (OCR) surinkimas; (ii) empirinių duomenų rinkimas taikant žodynu grįstos kompiuterizuotos teksto analizės (*Dictionary-based Computer-aided [or assisted] Text Analysis*) metodą, realizuojant jį su MaxQDA programa.

ŠALTINIŲ TEKSTYNO SUDARYMAS. Pirmasis tyrimo modelio žingsnis (tekstyno sudarymas) yra atliekamas taikant bendruosius mokslinių publikacijų paieškos internete principus bei metodus. Į tekstyną atrenkami autentiški, publikuoti, tyrimui reikalingi šaltiniai (laiškai, dienoraščiai, atsiminimai ir kt.) bei mokslinės publikacijos apie juos, tokiu būdu suformuojant du tekstyno – šaltinių ir literatūros – blokus. Literatūros (mokslinių publikacijų apie šaltinius) blokas naudojamas kaip papildoma medžiaga tikslesnei šaltinių tekstyno atrankai ir interpretavimui bei tyrimų įrankio – žodyno – rengimui. Paprastai panašios atrankos procedūros atliekamos kuriant kitus HM tekstynus²⁰ ar vykdant mokslinės literatūros sistemines analizes bei metaanalizes. Literatūros analizių atveju medžiagos rinkimo dizainas vykdomas laikantis kurios nors standartizuotos metodikos (pavyzdžiui, PRISMA-P)²¹. Atrinkant šal-

20 MCGILLIVRAY, Barbara; TÓTH, Gábor Mihály. *Applying Language Technology in Humanities Research: Design, Application, and the Underlying Logic*. Berlin: Springer Nature, 2020.

21 MOHER, David; SHAMSEER, Larissa; CLARKE, Mike; GHERSI, Davina; LIBERATI, Alessandro; PETTICREW, Mark; SHEKELLE, Paul; STEWART, Lesley A.; PRISMA-P Group. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-

tinius atliekamos tyrimų užklauskos skirtingose duomenų bazėse (pavyzdžiui, JSTOR, Scopus, Web of Science, Google Scholar), skaitmeninėse bibliotekose (pavyzdžiui, Google Books, Polona, Gallica, Oxford Text Archive, Internet Archive, ePaveldas) ir bendrai internete. Atliekant paieškas naudojami reikšminiai žodžiai ar iš jų sudarytos užklauskos (žodžių sekos, pritaikant loginius operatorius ar be jų). Reikšminių žodžių pasirinkimą ir jungimą į užklauskas lemia tyrimų tematika. Sudarant tekstyną gali būti panaudota ir specializuota bibliografinių nuorodų rinkimo programinė įranga (pavyzdžiui, Publish or Perish). Bibliografinės nuorodos ir visateksčiai šaltiniai bei mokslinių publikacijų tekstai gali būti išsaugomi naudojant nuorodų valdymo programinę įrangą (pavyzdžiui, Zotero). Tokiu būdu yra sudaromas pradinis tekstynas, kuris paskui turėtų būti peržiūrėtas atrenkant tinkamiausius šaltinius bei literatūrą. Tai gali būti vykdoma pagal tą pačią PRISMA-P metodiką, analizuojant šaltinių pavadinimus, aprašymus ir metaduomenis (pavyzdžiui, geografinę, chronologinę, teminę šaltinio aprėptį) bei mokslinių publikacijų pavadinimus, santraukas ir pačių mokslinių publikacijų tekstus. Siekiant užtikrinti kuo platesnę šaltinių aprėptį, naudojamas „sniego gniūžtės“ metodas, peržiūrint bibliografines nuorodas (*citation hopping*) jau atrinktuose šaltiniuose ir mokslinėse publikacijose. Tam galima panaudoti ir mokslinių publikacijų duomenų bazių (pavyzdžiui, Google Scholar) atvirkštinio citavimo (*reverse citation*) funkciją. Galutiniame etape patikrinama, ar visi atrinkti šaltiniai yra optiškai atpažinti (OCR) dokumentai, nes tik tokie tinkami tyrimui kitame metodo taikymo etape. Tokiu būdu yra parengiamas galutinis šaltinių tekstynas, kuris bus importuotas į MaxQDA kompiuterinės programos aplinką ir naudojamas tolesniame tyrime.

Sudarant tekstyną yra galimybė tyrimams panaudoti ir optiškai neatpažintus (ne OCR) dokumentus, tačiau tokiu atveju prisideda du papildomi žingsniai: (i) teksto atpažinimas, naudojant kurią nors OCR programą (pavyzdžiui, Online OCR, <https://www.onlineocr.net> ar kurią kitą) ir (ii) optiškai atpažinto teksto kokybės tikrinimas. Tais atvejais, kai suskaitmenintas šaltinis yra spausdintas įmantresniu (pavyzdžiui, gotišku) šriftu, optinio atpažinimo metu gauto teksto kokybė gali būti labai prasta ir reikalauti daug darbo sutvarkant tekstą. Žinoma, tokiu atveju galima bandyti ir specializuotas OCR programas, skirtas gotiškam ar kitam specifiniam šriftui atpažinti (pavyzdžiui, Transcribus, <https://readcoop.eu/transcribus>). Skirtingų tekstų optinio atpažinimo su programa Online OCR rezultatai pateikiami 1 lentelėje.

1 šaltinis. Stefan Batory. Sprawy Wojenne Króla Stefana Batorego: Dyjaryjusze, Relacje, Listy I Akta Z Lat 1576–1586. Acta Historica Res Gestas Poloniae Illustrantia Ab Anno 1507 Ad Annum 1795 T. 11. Kraków: Akad. Umiejętności, 1887, p. 3.

Originalus tekstas

III. List Jana Hlebowicza ¹⁾ do Jana Chodkiewicza Kasztelana Wileńskiego.

Z Zaslawia 28 Stycznia 1576. [Archiw. Sapiel. Nr. 23, fol. 371.]

Co mię doszło wiedzieć, W. Mci dają znać, W. Mśc z listu będziesz raczył wyrozumieć, iż Moskiewski gońce posyła. Res bene agitur, otóż nam do kupy potrzeba dla Boga/ dla Boga!

Co mię za list od P. Wojewody Wileńskiego ²⁾ potkał W. Mci go posyłam, quid hoc ignoro ale podobno probat me na co cautissime respons uczynię cum gratiarum actione a też uniwersał W. Mci posyłam, ztąd W. Mśc porozumiesz że oni sibi auctoritatem non officio W. Mci przypisują gdzie wmięszali Pana Stolnika ob formam jeszcze, kryjąc gubernacyą.

Pisać więcej czas nie znosi jedno służby me w Mciwą łaskę mego Mciwego Pana pilnie zalecam.

Nesutvarkytas OCR teksto variantas

III. List Jana Hlebowicza ¹⁾ do Jana Chodkiewicza Kasztelana Wileńskiego.

Z Zaslawta 28 Stycznia 1576. lArchiw. Sapieli. Nr. 23, jol. 371.1

Co mię doszło wiedzieć, W. Mci dają znać, W. Mśc z listu będziesz raczył wyrozumieć, iż Moskiewski gońce posyła. Res b e n e agitur, otóż nam do kupy potrzeba dla Boga/ dla Boga 1 Co mię za list od P. Wojewody Wileńskiego ¹⁾ potkał W. Mci go posyłam, quid hoc ignoro ale podobno probat me na co cautissime r esp o ns uczynię cum gratiar um ac tione a też uniwersał W. Mci posyłam, ztąd W. Mśc porozumiesz że oni sibi auc t oritatem non o f f i c i o W. Mci przypisują gdzie wmięszali Pana Stolnika ob for mam jeszcze, kryjąc gubernacya. Pisać więcej czas nic znosi jedno służby me w Mciwą łaskę mego Mciwego Pana pilnie zalecam.

2 šaltinis. Schulz, Friedrich. Reise Eines Liefländers Von Riga Nach Warschau, Durch Südpreußen, über Breslau, Dresden, Karlsbad, Bayreuth, Nürnberg, Regensburg, München, Salzburg, Linz, Wien Und Klagenfurt, Nach Botzen in Tyrol. Berlin: Bei Friedrich Vieweg Dem ältern, 1795, p. 60.

Originalus tekstas

Prinz J** P** war einmal in einem ähnlichen Falle. Er hielt sich ein Mädchen, an die er sehr hing. Um ihr diesen Triumph zu entreißen, verbanden sich drey der reichsten Weiber in Warschau, ihn, koste es was es wolle, zu erobern. Da ihnen dieß auf dem gewöhnlichen Wege nicht gelang, spielten sie ihre Angriffe bis in sein Schlafzimmer. Beruflicherisch, in einem Grade, wie nur ein polnisches Weib der hohen Klasse es seyn kann, erwarteten sie ihn eines Abends in demselben, die schönste in seinem Bette, und die beyden andern, wie Nymphen gekleidet, vor demselben. Diese flogen ihm entgegen, als er hereintrat, und führten ihn zu dem wollüstigen Lager hin; aber sie hatten die Beschämung, den Prinzen umkehren und sich in die Arme seines Mädchens retten zu sehen. Ein Heldennuth, den man ihm lange nicht hat verzeihen können.

Nesutvarkytas OCR teksto variantas

3* tp.. mar einmal in einem ablieen Ralle. er biett Wein VIZAbcben, an bit er fetg hing. Um ibt bieten Zriutppb in entreigen, verbanben fid) brev ber rdbegib: ften Meiber in Saftbau, ihn, toffe es matt ed in erobern. Za ihnen biegt auf bem gerobbnlichen Wege nicht gelang, fpietten rye lige ?ingriffe bis in fein eittafimmer. Zero fübererifd), in einem grabe, mie nur ein pot/ nifcbes See ber hoben Stlalre es reyn tann, ermatteten sie ibn eines 21,benbs in bemfelben, bie in feinem 'Zetre, unb bie bepben anbern, tvie Olvmpben getteibet, vor bemfets ben. Ziff(' flogen ihm entgegen, als er ber: eintritt; unb führten ihn bu Dem rooffiltigen Vager igitt; aber sie bauen bie £13efrginung• ben ‚prinien umfegren unb fict) in bie ?Urne feines Qabitens retten iu • feben. Zin .Ve[s benmatb, ben man ihm lange nicht bat vers feil en tütinen. •

ŽODYNU GRĮSTA KOMPIUTERIZUOTA TEKSTO TURINIO ANALIZĖ. Šis metodas yra didesnės analitinių metodų ir su jais susijusių programinių priemonių grupės – kompiuterizuotos teksto turinio analizės (*Computer-aided [or assisted] Text Analysis, CATA*) dalis²². Žodynu grįsta kompiuterizuota teksto turinio analizė (ŽGKTTA) gali būti kiekybinė ir kokybinė. Techniškai ji yra kiek panaši į paiešką, naudojant specifinius reikšminius žodžius, kuriuos nustato tyrėjas, atsižvelgdamas į tyrimo tikslus. Pagrindinis skirtumas tarp šio metodo ir paieškos pagal reikšminius žodžius yra tas, kad ŽGKTTA tyrime naudojamas žodynas yra sistemingas reikšminių žodžių rinkinys, grindžiamas ankstesniais tyrimais, ir jį sudarantys žodžiai yra aprobuoti kaip reprezentuojantys tam tikrą konceptą²³. Kokybinio tyrimo metu metodas leidžia tą konceptą (ne pavienius žodžius) identifikuoti šaltinio tekste. O tai įgalina sistemingai koduoti ir analizuoti tekstus, siekiant iširti plačias mokslines problemas²⁴. Tyrimas vadinamas kokybiniu dėl to, kad jo taikymo tikslas nėra kiekybiškai, statistiškai ar skaitinėmis reikšmėmis apibendrinti ir analizuoti vienų ar kitų šaltinio teksto turinio elementų (pavyzdžiui, pasirinktų žodžių dažnių tekste). Taikant šį metodą yra imituojamas istoriko darbas, atpažįstant ir pažymint tyrimo problemos ir klausimų požiūriu svarbias teksto vietas (pastraipas), kurios – ne pavieniai žodžiai – ir yra tyrimo analitinis vienetas. Šios pastraipos vėliau yra analizuojamos ir interpretuojamos tyrėjo, naudojant istorijos mokslui būdingus metodus.

Pagrindinis ŽGKTTA įrankis yra žodynas. Vienais atvejais gali būti naudojami jau egzistuojantys žodynai, kuriuos galima pritaikyti konkretaus tyrimo reikmėms²⁵. Kitais atvejais žodynai yra kuriami pačių mokslininkų. Tokie žodynai

- 22 NEUENDORF, Kimberly A. *The Content Analysis Guidebook*. London: SAGE, 2017.
- 23 CAMPION, Emily D.; CAMPION, Michael A. Using Computer-assisted Text Analysis (CATA) to Inform Employment Decisions: Approaches, Software, and Findings. *Research in Personnel and Human Resources Management*, 2020, vol. 38, p. 285–325.
- 24 SHORT, Jeremy C.; MCKENNY, Aaron F.; REID, Shane W. More Than Words? Computer-Aided Text Analysis in Organizational Behavior and Psychology Research. *Annual Review of Organizational Psychology and Organizational Behavior*, 2018, vol. 5, p. 415–435. [žiūrėta 2022 m. gruodžio 31 d.]. Prieiga per internetą: <<https://www.annualreviews.org/doi/10.1146/annurev-orgpsych-032117-104622>>.
- 25 MAERZ, Seraphine F.; PUSCHMANN, Cornelius. Text as Data for Conflict Research: A Literature Survey. In DEUTSCHMANN, E., LORENZ, J., NARDIN, L., NATALINI, D., WILHELM, A. (eds.). *Computational Conflict Research. Computational Social Sciences*. Springer, 2020. [žiūrėta 2022 m. gruodžio 31 d.]. Prieiga per internetą: <https://doi.org/10.1007/978-3-030-29333-8_3>; GRIMMER, Justin; STEWART, Brandon M. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, 2013, vol. 21 (3), p. 267–297. [žiūrėta 2022 m. gruodžio 31 d.]. Prieiga per internetą: <doi:10.1093/pan/mps028>.

formuojami remiantis teorija, ankstesniais tyrimais, mokslininko patirtimi ar bandomuoju numatomo analizuoti tekstyno dalies tyrimu²⁶. Atsižvelgiant į tai, kad praeities kelionių tyrimams skirto žodyno nėra, jis buvo kuriamas projekto „*Homo viator*: kelionių erdvė ir keliautojų potyriai ankstyvosios modernybės Lietuvoje“ įgyvendinimo metu. Straipsnio autoriaus vykdytame keliautojų gastronominių patirčių tyrime „*Homo viator*: kelionių erdvė ir keliautojų potyriai ankstyvosios modernybės Lietuvoje“ projekte naudotą šaltinių rinkinį (tekstyną) sudarė iš viso 112 dokumentų (kelionių dienoraščių, atsiminimų, reliacijų, šalių ir miestų aprašymų). Pagrindiniai jų atrankos kriterijai buvo teminis ir geografinis. Šaltiniuose turi būti pateikiami kelionių per / po Lietuvą aprašymai ar į potencialius keliautojus orientuoti Lietuvos, kaip šalies, aprašymai. Papildomi kriterijai buvo chronologinis (šaltiniai turi būti ne vėlesni kaip XIX a. pradžios) ir medijos (šaltinio visatekstė skaitmeninė versija turi būti pasiekama internete). Tokie atrankos kriterijai nulėmė tiriamų tekstų rinkinio sudėtį ir jo kalbinę įvairovę. Atsižvelgiant į šaltinių specifiką (šaltiniuose ir jų vertimuose vartojamas kalbas) buvo sukurtas daugiakalbis lietuvių–lenkų–anglų–rusų–vokiečių kalbų žodynas.

Bandomajam tyrimui skirto žodyno struktūrą sudaro šeši su praeities kelionėmis susiję konceptai (kategorijos): (i) kelionė (bendras apibūdinimas), (ii) kelias ir jo infrastruktūra (tiltai, brastos ir kt.), (iii) transporto priemonės, (iv) sustojimo ir nakvynės vietos (miesteliai, kaimai, karčemos, pašto stotys ir kt.), (v) kelyje sutikti žmonės (smuklininkai, plėšikai, vedliai ir kt.), (vi) kelionių maistas. Konceptų skaičius, apibrėžimai ir juos sudarantys žodžiai buvo išskirti ir parinkti remiantis ankstesniais tyrimais ir tikslinami asinchroniškai (Google Drive dokumento rengimo formatu) vykdytų diskusijų su projektą „*Homo viator*: kelionių erdvė ir keliautojų potyriai ankstyvosios modernybės Lietuvoje“ įgyvendinančiais mokslininkais Arvydu Pacevičiumi, Tomu Čelkiu, Axeliu Walteriu, Tomašu Božerockiu metu. Kiekvienas konceptas yra apibūdinamas reikšminių žodžių ir frazių rinkiniu. Nors bandomajam tyrimui skirtame žodyne dominuoja daiktavardžiai, tokio apribojimo žodyno sudarymui nėra. Į žodyną gali būti įtraukiami bet kokie tyrimo požiūriu svarbūs žodžiai ar frazės. Žodžių atitikmenys kitomis kalbomis buvo taip pat parenkami remiantis ankstesniais tyrimais ir tikslinami diskutuojant su projektą vykdančiais mokslininkais. Kaip papildomi žodyno daugiakalbystės šaltiniai buvo naudojami kal-

26 NEUENDORF, Kimberly A. *The Content Analysis Guidebook*. London: SAGE, 2017; WATANABE, Kohei. Computer-aided Dictionary Making: An Efficient Dictionary Construction Technique for Content Analysis. *Proceedings of the International Conference on the Advances in Computational Analysis of Political Text (PolText 2016)*. Dubrovnik, 2016, p. 81–87.

bū žodynai, atspindintys šaltinių kalbą (pavyzdžiui, „Słownik Polszczyzny XVI wieku“, <http://spxvi.edu.pl/indeks/ar> „Elektroniczny słownik języka polskiego XVII i XVIII wieku“, <https://sxvii.pl/>). Kelionių maisto konceptą apibūdinančių reikšminių žodžių rinkinys pateikiamas 2 lentelėje.

2 LENTELE. Kelionių maisto konceptą (kategoriją) apibūdinantys reikšminiai žodžiai penkiomis kalbomis

lietuvių	lenkų	rusų	anglų	vokiečių
alus	piwo	пиво	beer	Bier
duona	chleb	хлеб	bread	Brot
kava	kawa	кофе	coffee	Kaffee
mėsa	mięso	мясо	meat	Fleisch
patiekalas	potrawa	блюдо	dish	Gericht
patiekalas	danie	блюдо	meal	Speise
pietūs	obiad	обед	dinner	Mittagessen
pietūs	obiad	обед	dinner	Festmahl
pietūs	obiad	обед	dinner	Gastmahl
pusryčiai	śniadanie	завтрак	breakfast	Frühstück
sūris	ser	сыр	cheese	Käse
vakarienė	kolacja	ужин	lunch	Abendessen
vakarienė	wieczera	ужин	supper	Abendbrot
valgis	posiłek	пища	food	Essen
valgis	strawa	пища	mess	Mahl
valgis	strawa	пища	mess	Mähler
vynas	wino	вино	wine	Wein

Svarbu pažymėti, kad, taikant šį tyrimų metodą, žodyno struktūra iš esmės yra atvira tiek pildant žodyną naujais konceptais, tiek grupuojant terminus ir įvedant naujus subkonceptus (subkategorijas) žodyno viduje (pavyzdžiui, maisto koncepte išskirti ir sugrupuoti valgius ir gėrimus), tiek pildant konceptus naujais reikšminiais žodžiais ar frazėmis, tiek plečiant žodyną daugiakalbystės požiūriu. Straipsnyje aprašytas žodynas buvo perkeltas ir tyrimas atliekamas su MaxQDA programa, skirta kokybinei turinio analizei. Perkeltam žodyną į MaxQDA, kiekvienam reikšminiam žodžiui ar frazei galima priskirti paieškos

savybes, leidžiančias ieškoti tekste pagal tikslų terminą, pagal tikslų terminą atsižvelgiant į didžiąsias ir mažąsias raides ar naudojant paieškai tik reikšminio žodžio dalį.

TYRIMO MODELIO TESTAVIMAS

Aprašytas tyrimo modelis buvo testuojamas panaudojant tekstyną, surinktą įgyvendinant projektą „*Homo viator*: kelionių erdvė ir keliautojų potyriai ankstyvosios modernybės Lietuvoje“. Šaltinių tyrimui pasirinktas teminis laukas buvo keliautojų gastronomicinės patirtys, kurioms identifikuoti naudotas žodyno kelionių maisto konceptas (kategorija). Toks teminis laukas testavimui pasirinktas atsižvelgiant į straipsnio autoriaus tyrimų tematiką projekte (keliautojų gastronomicinės patirtys). Modelio testavimo tikslais buvo atrinktas vienas šaltinis anglų kalba (3 lentelė). Šio šaltinio pasirinkimą lėmė keli veiksniai: (i) šaltinio pobūdis (dokumentas priklauso didelei grupei panašių dokumentų, kuriuose aprašomos tranzitinės kelionės per Lietuvą į tolimesnius kraštus); (ii) turinio reprezentatyvumas projekto tematikos prasme (dokumente yra informacijos apie visus anksčiau minėtus šešis projektui aktualius teminius konceptus); (iii) turinio reprezentatyvumas gastronomicinių patirčių požiūriu (šaltinyje yra pakankamai informacijos apie gastronomicines keliautojų patirtis); (iv) šaltinio ištirtumas (dokumentas yra išanalizuotas netaikant skaitmeninių metodų, tad galimas rezultatų, gautų įprastiniu (viso šaltinio skaitymo *de visu*) metodu ir gautų skaitmeniniais metodais, palyginimas); (v) anglų kalba dėl savo struktūros ir žodžių vartosenos yra patogi tokio pobūdžio testavimui; (vi) šaltinio OCR teksto kokybė yra gera, raidės ir ženklai, lyginant su originaliu spausdintu tekstu, yra kokybiškai atpažinti. Tyrimo modelio kūrimas yra tęstinis, todėl straipsnyje aprašomas testavimas su vienu šaltiniu yra tik pirmas žingsnis, siekiantis patikrinti, kiek iš esmės yra perspektyvus šis modelis. Identifikavus modelio (iš esmės) tinkamumą šaltinių tyrimui, jis toliau tobulinamas dirbant su tipologiškai kitokiais ir kitomis kalbomis parašytais šaltiniais.

Testavimas buvo atliekamas naudojant MaxQDA programos funkciją „Auto-code with dictionary“, renkantis žymėjimą (kodavimą) tik to sakinio, kuriame aptinkamas žodyno reikšminis žodis ar frazė. Teksto kodavimo rezultatų pavyzdžiai (surasti su keliautojų maistu susiję fragmentai) pateikiami 3 lentelėje. Cituojami ištraukų fragmentai yra pateikiami taip, kaip jie buvo fiksuoti OCR dokumente. Kai kuriose teksto vietose matomos klaidos yra atsiradusios dėl nepakankamai tiksliai atlikto OCR teksto atpažinimo.

3 LENTELĖ. Tekstų kodavimo su programa MaxQDA rezultatai. Analizuotas šaltinis: Korb, Johann Georg. *Diary of an Austrian Secretary of Legation at the Court of Czar Peter the Great*. Translated from the original Latin and edited by the count Mac Donnell. London, 1863, Vol. I, p. 40–60. Iš viso 329 puslapių apimties tekste buvo surasti 347 fragmentai, susiję su keliautojų maistu.

Tema	Koduoto teksto pradžia	Koduoto teksto pabaiga	Koduotas teksto segmentas
Maistas	44: 1091	44: 1166	He offered us beer, brandy, and bread; and for our horfes oats and hay, of which there was an exceeding dearth in Poland.
Maistas	51: 418	51: 570	After dinner we went to fee the handsome triumphal arch fet up by the magiftrates of Dantzick, for the reception of the newly- elected King of Poland.
Maistas	59: 241	59: 371	The presents wherewith they honoured the Lord Envoy confited of fifh of marvellous size and a dozen of Rhenifh and Spanifh wine.
Maistas	60: 54	60: 186	We had ftill another river, called the Nagat, to crofs in like condition, after which we reached the village of Rofpach to dinner.

Automatiniu būdu analizuoto teksto lyginimas su to paties teksto analize *de visu* parodė, kad taikant šį metodą buvo rasti pagrindiniai teksto fragmentai, susiję su keliautojų maistu. Fragmentus, kurie nebuvo rasti, galima surasti tobulinant (pildant) žodyną (4 lentelė).

4 LENTELĖ. ŽGKTTA naudojamo žodyno pildymo pavyzdžiai po teksto analizės *de visu*

Papildomi reikšminiai žodžiai žodyne	Taikant papildomus žodžius surasti teksto fragmentai
„water“; „beverage“	Here we found acidulous waters , an agreeable and wholefome beverage for perfons heated with travelling.
„dined“	At early down we again fet out, passing Tornaft and Scheuzen and dined at the village of Deutfchammer, after getting over a bad , ftony, and hilly road.

„liqueur“	Next morning the Staroft vifited us in perfon to wifh us a happy journey, not confining himfelf to words, but moft civilly inviting each of us to accept a glafs of a rare and fuperior kind of liqueur;
„brandy“	The intenfe cold of the day induced fome of our party to indulge too copiously in brandy , thinking in the fun's defpite to turn winter into fummer by plentiful potations . But the cowards were fairly overcome by the brandy and the cold together.

Automatinio kodavimo metu nurodytame tekste nebuvo fiksuota su tema (keliautojų maistas) nesusijusių teksto fragmentų. Keli tokie fiksuoti kitose šaltinio vietose. Pavyzdžiui, p. 66 buvo koduotas toks teksto fragmentas: „Having taken a guide to fhow us the way, we reached Kalapina at dinner-time.“ Šiame fragmente žodis „dinner“ yra nuoroda į laiką („pietų metas“), bet ne į patį valgyimo procesą.

Išvados ir diskusija

Pastaraisiais dešimtmečiais vykstantis masyvus rašytinių istorijos šaltinių skaitmeninimas, jų tekstų optinis atpažinimas (OCR) bei pateikimas internete sukuria naujas istorinio tyrimo galimybes ir iššūkius. Tais atvejais, kai: (i) tyrimo tema yra specifinė ir gana siaura; (ii) internete yra daug tyrimui naudingų šaltinių ir (iii) tyrimui reikalinga nedaug, bet per visą šaltinio tekstą išbarstytos informacijos, istoriniame tyrime galima pritaikyti informacijos mokslams būdingą informacijos organizavimo paradigmą, skaitmeninius informacijos organizavimo ir tyrimų metodus bei standartizuotus kompiuterinius programinius įrankius.

Šiame straipsnyje aptartas tyrimo modelis jungia šaltinių bei mokslinių publikacijų tekstyno formavimą su automatine žodynu grįsta kompiuterizuota kokybine teksto turinio analize (naudojant programą MaxQDA). Modelio tikslas yra imituoti istoriko darbą, surandant bei pažymint tyrimui svarbius teksto fragmentus skaitmeniniame OCR šaltinyje. Šie fragmentai yra naudojami tolesniame istoriko tyrime.

Modelis buvo testuotas su projekto „*Homo viator*: kelionių erdvė ir keliautojų potyriai ankstyvosios modernybės Lietuvoje“ medžiaga ir pasiteisino. Straipsnyje pristatomas tyrimo modelis leidžia greitai ir efektyviai surinkti empirinius duomenis, reikalingus praeities kelionių tyrimams, ir sumažinti istoriko darbo laiką, skirtą tyrimui, sąnaudas.

Testavimo metu išryškėjo svarbiausi tyrimo modelio ribotumai ir jo taikymo iššūkiai. Korektiškam modelio taikymui būtinas elementas yra kokybiškas žodynas, tiksliai apibūdinantis tyrimui aktualius konceptus. Žodyno neišsamumas yra pagrindinė modelio neefektyvaus veikimo priežastis. Kitas svarbus ribojimas – žodyne vartojamų terminų kaitymas giminėmis, linksniavimas, asmenavimas konkrečioje kalboje. Automatinėje turinio analizėje naudojama žodžio forma lemia paieškos rezultatus, kurie kai kada (pavyzdžiui, naudojant tik žodžio šaknį) gali būti per daug platūs, apimantys per daug tyrėjui nereikalingo teksto elementų ir taip mažinantys mokslininko darbo efektyvumą. Dar vienas svarbus tyrimo ribojimas – dokumentų OCR kokybė, kuri kai kuriais atvejais yra per prasta. Dėl šios priežasties nemaža dalis tyrėjui reikalingų teksto elementų gali būti neidentifikuoti.

Tai, kad tyrimo modelis testavimo metu iš esmės pasiteisino, paskatino jį taikyti visame projekto „*Homo viator: kelionių erdvė ir keliautojų potyriai anksstyvosios modernybės Lietuvoje*“ įgyvendinimo metu atliktame keliautojų gastronominių patirčių tyrime. Taikant modelį buvo analizuoti visi 112 tekстыne esantys šaltiniai anglų, lenkų, lietuvių, rusų ir vokiečių kalbomis. Tekstynas buvo papildytas 17 šaltinių prancūzų kalba ir taip pat atlikta jų analizė. Šio tyrimo metu nebuvo fiksuota modelio taikymo skirtumų, priklausomai nuo šaltinio tipologijos. Šaltinių skirtumai pagal kalbas išryškėjo tik sąsajose su anksčiau paminėtais modelio apribojimais. Prastesni rezultatai buvo gauti analizuojant šaltinius kalbomis, kurioms didelę įtaką turi kaitymas linksniais, asmenimis ar giminėmis (pavyzdžiui, lietuvių kalba), bei kalbose, kurios dėl rašto specifikos prasčiau perskaitomos OCR (pavyzdžiui, vokiški tekstai, parašyti gotišku šriftu).

Dar viena tyrimo metu išryškėjusi problema yra turinio geografinis indeksavimas. Tyrimo modelis įgalina neblogai surinkti informaciją, susijusią su gastronominėmis keliautojų patirtimis, tačiau jis neskiria patirčių pagal geografinę erdvę. Šis trūkumas ypač ryškus atvejais, kai tiriamas šaltinis apima tranzitinę kelionę, kurios tik nedidelė dalis vyko Lietuvos teritorijoje. Šiai problemai spręsti ateityje galėtų būti taikomi GIS skaitmeniniai metodai bei įrankiai. Vienas iš galimų metodų galėtų būti koduotuose teksto fragmentuose esančių vietovardžių automatinis atpažinimas ir susiejimas su geografinėmis koordinatėmis taikant geografinių vardų indeksavimo tekste (*geoparsing*) programinę įrangą (pavyzdžiui, *The Edinburgh Geoparser*). Naudojant šį metodą (anglų kalba sinonimiškai vadinamą *geocoding*, *georecognition*, *geotagging*, *toponym recognition*) nestruktūruotame tekste surandami ir išskiriami vietovardžiai ar į konkrečią lokaciją nurodančios frazės (pavyzdžiui, „trys kilometrai

šiauriau Vilniaus“) bei jiems yra priskiriami geografiniai identifikatoriai (kodai arba geografinės koordinatės)²⁷. Šis metodas ir jam skirti skaitmeniniai įrankiai yra sukurti orientuojantis į dabartinę geografiją. Tačiau kuomet kalbame apie skaitmeninių technologijų taikymą istorinės geografijos (praeities kelionių) tyrimams, susiduriame su kai kuriomis problemomis, kurios yra specifiškai būdingos istorinei geografijai²⁸.

Geografinių vardų indeksavimo metu sukurtas vietovardžių indeksas (*gazeteer*) galėtų būti eksportuotas ir analizuojamas taikant GIS metodus su ArcGIS programine įranga. Tačiau gali būti, kad viso to nereikės. Stiprėjantis dirbtinis intelektas jau dabar perima dalį rutininių ir standartizuotų mokslininko veiklų. Spausdintų OCR šaltinių tyrimo bandymai su ChatGPT 4.0 versija rodo, jog ši programa jau dabar visai neblogai analizuoja dokumentus.

Literatūra

1. ANDERSON, Carrie; CESERANI, Giovanna; DONALDSON, Christopher; GREGORY, Ian N.; HALL, Melanie; ROSENBAUM, Adam T.; TAYLOR, Joanna E. Digital Humanities and Tourism History. *Journal of Tourism History*, 2017, vol. 9 (2–3), p. 246–269. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <10.1080/1755182X.2017.1419455>.
2. BENCH, Harmony; ELSWIT, Kate. Dance Touring and Embodied Data Some Approaches to Katherine Dunham’s Movement on the Move. *Current Research in Digital History*, 2019, vol. 2. [žiūrėta 2023 m. sausio 10 d.]. Prieiga per internetą: <https://doi.org/10.31835/crdh.2019.12>.
3. BENCH, Harmony; ELSWIT, Kate. Mapping Movement on the Move: Dance Touring and Digital Methods. *Theatre Journal*, 2016, vol. 68, no. 4, p. 575–596. [žiūrėta 2023 m. sausio 10 d.]. Prieiga per internetą: <http://www.jstor.org/stable/26367451>.
4. BLANEY, Jonathan; MILLIGAN, Sarah; STEER, Marty; WINTERS, Jane. *Doing Digital History*. Manchester, England: Manchester University Press, 2021. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <https://doi.org/10.7765/9781526157713>.
5. BÜRGERMEISTER, Martina. Measuring Urban Change in Travel Texts Using the Example of the City of Graz in the Long Nineteenth Century. *Magazén*, 2022, vol. 3 (1), p. 61–90. [žiū-

27 CAMPELO, Claudio E. C. Geographically-Aware Information Retrieval on the Web. In *Encyclopedia of Information Science and Technology*. Hrshey: IGI Global, 2015. [žiūrėta 2022 m. gruodžio 31 d.]. Prieiga per internetą: <https://www.igi-global.com/chapter/geographically-aware-information-retrieval-on-the-web/112830>.

28 LAUŽIKAS, Rimvydas; VOSYLIŪTĖ, Ingrida; JARONIS, Justinas. Beyond the Space: the LoCloud Historical Place Names Micro-service. *CAA2015. Keep The Revolution Going: Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology*. Archaeopress Publishing Ltd, 2016, p. 235–240. [žiūrėta 2022 m. gruodžio 31 d.]. Prieiga per internetą: <https://books.google.lt/books?id=2nwwEAAAQBAJ&lpg=PP1&hl=lt&pg=PP1#v=onepage&q&f=false>.

- rėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <<https://edizionicafoscarini.unive.it/media/pdf/article/magazen/2022/1/art-10.30687-mag-2724-3923-2022-05-003.pdf>>.
6. CAMPELO, Claudio E. C. Geographically-Aware Information Retrieval on the Web. In *Encyclopedia of Information Science and Technology*. Hrshey: IGI Global, 2015. [žiūrėta 2022 m. gruodžio 31 d.]. Prieiga per internetą: <<https://www.igi-global.com/chapter/geographically-aware-information-retrieval-on-the-web/112830>>.
 7. CAMPION, Emily D.; CAMPION, Michael A. Using Computer-assisted Text Analysis (CATA) to Inform Employment Decisions: Approaches, Software, and Findings. *Research in Personnel and Human Resources Management*, 2020, vol. 38, p. 285–325.
 8. EASTON, Evan; ALLEN, Calabria; ALLEN, William. Computational Text Analysis and Archival Methods. In *International Organizations and Research Methods*. An Arbor: University of Michigan Press, 2023, p. 285–291.
 9. GOLUB, Koraljka. Automatic Identification of Topics: Applications and Challenges. In *Doing Digital Humanities: Concepts, Approaches, Cases*. Växjö: Linnaeus University Press, 2022, p. 5–26.
 10. GOLUB, Koraljka. Automatic Subject Indexing of Text. *Knowledge Organization*, 2019, vol. 46 (2), p. 104–121. <https://doi.org/10.5771/0943-7444-2019-2-104>.
 11. GOLUB, Koraljka; LIU, Ying-Hsang. *Information and Knowledge Organisation in Digital Humanities Global Perspectives*. New York: Routledge, 2022.
 12. GOLUB, Koraljka; SOERGEL, Dagobert; BUCHANAN, George; TUDHOPE, Douglas; LYKKE, Marianne; HIOM, Debra. A Framework for Evaluating Automatic Indexing or Classification in the Context of Retrieval. *Journal of the Association for Information Science and Technology*, 2016, vol. 67 (1), p. 3–16. <https://doi.org/10.1002/asi.23600>.
 13. GRIMMER, Justin; STEWART, Brandon M. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, 2013, vol. 21 (3), p. 267–297. [žiūrėta 2022 m. gruodžio 31 d.]. Prieiga per internetą: <[doi:10.1093/pan/mps028](https://doi.org/10.1093/pan/mps028)>.
 14. GULBINAS, Zenonas. Geografinė informacinė sistema. Iš *Visuotinė lietuvių enciklopedija*. Vilnius: Mokslo ir enciklopedijų leidybos centras, 2022. [žiūrėta 2022 m. gruodžio 31 d.]. Prieiga per internetą: <<https://www.vle.lt/straipsnis/geografine-informacine-sistema/>>.
 15. HILL, Michael W. *The Impact of Information on Society*. London: Bowker-Saur, 1999.
 16. HOEKSTRA, Rik; KOOLEN, Marijn. Data Scopes for Digital History Research. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 2019, vol. 52 (2), p. 79–94. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <[10.1080/01615440.2018.1484676](https://doi.org/10.1080/01615440.2018.1484676)>.
 17. HUVILA, Isto. How a Museum Knows? Structures, Work Roles, and Infrastructures of Information Work. *Journal of the American Society for Information Science and Technology*, 2013, vol. 64, p. 1375–1387. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <<https://doi.org/10.1002/asi.22852>>.
 18. HUVILA, Isto. *The Ecology of Information Work: A Case of Bridging Archaeological Work and Virtual Reality Based Knowledge Organisation*. Åbo: Åbo Akademi University Press, 2006. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <<http://urn.fi/URN:ISBN:951-765-337-9>>.
 19. HUVILA, Isto; HUGGETT, Jeremy. Archaeological Practices, Knowledge Work and Digitalisation. *Journal of Computer Applications in Archaeology*, 2018, vol. 1 (1), p. 88–100. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <[10.5334/jcaa.6](https://doi.org/10.5334/jcaa.6)>.

20. JACOBS, Ronald L. Knowledge Work and Human Resource Development. *Human Resource Development Review*, 2017, vol. 16 (2), p. 176–202. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <doi:10.1177/1534484317704293>.
21. KUHLTHAU, Carol C. Towards Collaboration between Information Seeking and Information Retrieval. *Information Research*, 2005, vol. 10 (2).
22. KUHN, Thomas S. *Mokslo revoliucijų struktūra*. Vilnius: Prada, 2003.
23. KUNCEVIČIUS, Albinas; LAUŽIKAS, Rimvydas; JANKAUSKAS, Rimantas; AUGUSTINAVIČIUS, Renaldas; ŠMIGELSKAS, Ramūnas. *Dubingių mikroregionas ir Lietuvos valstybės ištakos*. Vilnius: Petro ofsetas, 2015.
24. LAUŽIKAS, Rimvydas; VOSYLIŪTĖ, Ingrida; JARONIS, Justinas. Beyond the Space: the LoCloud Historical Place Names Micro-service. *CAA2015. Keep The Revolution Going: Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology*. Archaeopress Publishing Ltd, 2016, p. 235–240. [žiūrėta 2022 m. gruodžio 31 d.]. Prieiga per internetą: <https://books.google.lt/books?id=2nwwEAAAQBAJ&lpq=PP1&hl=lt&pg=PP1#v=onepage&q&f=false>.
25. LOZANO, Jorge Sebastián. Mapping Art History in the Digital Era. *The Art Bulletin*, 2021, vol. 103 (3), p. 6–16. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <10.1080/00043079.2021.1882819>.
26. MAERZ, Seraphine F; PUSCHMANN, Cornelius. Text as Data for Conflict Research: A Literature Survey. In DEUTSCHMANN, E., LORENZ, J., NARDIN, L., NATALINI, D., WILHELM, A. (eds.). *Computational Conflict Research. Computational Social Sciences*. Springer, 2020. [žiūrėta 2022 m. gruodžio 31 d.]. Prieiga per internetą: <https://doi.org/10.1007/978-3-030-29333-8_3>.
27. MCGILLIVRAY, Barbara; TÓTH, Gábor Mihály. *Applying Language Technology in Humanities Research: Design, Application, and the Underlying Logic*. Berlin: Springer Nature, 2020.
28. MIDURA, Rachel. Itinerating Europe: Early Modern Spatial Networks in Printed Itineraries, 1545–1700. *Journal of Social History*, 2021, vol. 54, iss. 4, p. 1023–1063. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <https://doi.org/10.1093/jsh/shab011>.
29. MILLIGAN, Ian. *The Transformation of Historical Research in the Digital Age* (Elements in Historical Theory and Practice). Cambridge: Cambridge University Press, 2022. doi:10.1017/9781009026055.
30. MOHER, David; SHAMSEER, Larissa; CLARKE, Mike; GHERSI, Davina; LIBERATI, Alessandro; PETTICREW, Mark; SHEKELLE, Paul; STEWART, Lesley A.; PRISMA-P Group. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 Statement. *Systematic Reviews*, 2015, vol. 4, no. 1 (1). [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <https://doi.org/10.1186/2046-4053-4-1>.
31. MOILANEN, Inka. Virtues, Vices, and Vectors : Digital Tools and the Study of Medieval Sermons. *Digitala Vetenskapliga Arkivet*, 2019. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-167353>.
32. MURRIETA-FLORES, Patricia; DONALDSON, Christopher; GREGORY, Ian. GIS and Literary History: Advancing Digital Humanities Research through the Spatial Analysis of Historical Travel Writing and Topographical Literature. *Digital Humanities Quarterly*, 2017, vol. 11 (1). [žiūrėta 2023 m. sausio 10 d.]. Prieiga per internetą: <https://chesterrep.openrepository.com/handle/10034/620256>.
33. NEUENDORF, Kimberly A. *The Content Analysis Guidebook*. London: SAGE, 2017.

34. PYÖRIÄ, Pasi. The Concept of Knowledge Work Revisited. *Journal of Knowledge Management*, 2005, vol. 9 (3), p. L116–L127. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <10.1108/13673270510602818>.
35. ROMEIN, C. Annemieke; KEMMAN, Max; BIRKHOLZ, Julie M.; BAKER, James; DE GRUIJTER, Michel; MEROÑO-PENUELA, Albert; RIES, Thorsten; ROS, Ruben; SCAGLIOLA, Stefania. State of the Field: Digital History. *History*, 2020, vol. 105, p. 291–312. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <<https://doi.org/10.1111/1468-229X.12969>>.
36. SHORT, Jeremy C.; MCKENNY, Aaron F., REID, Shane W. More Than Words? Computer-Aided Text Analysis in Organizational Behavior and Psychology Research. *Annual Review of Organizational Psychology and Organizational Behavior*, 2018, vol. 5, p. 415–435. [žiūrėta 2022 m. gruodžio 31 d.]. Prieiga per internetą: <<https://www.annualreviews.org/doi/10.1146/annurev-orgpsych-032117-104622>>.
37. SIEBOLD, Anna; VALLERIANI, Matteo. Digital Perspectives in History. *Histories*, 2022, vol.2, no. 2, p. 170–177. <https://doi.org/10.3390/histories2020013>.
38. WATANABE, Kohei. Computer-aided Dictionary Making: An Efficient Dictionary Construction Technique for Content Analysis. *Proceedings of the International Conference on the Advances in Computational Analysis of Political Text (PolText 2016)*. Dubrovnik, 2016, p. 81–87.
39. ŽIŽIŪNAS, Tadas. *Technologinis veiksnyų kultūros paveldo tyrimuose: 3D vaizdo ir spektroskopijos taikymo metodologinis modelis*. Daktaro disertacija. Vilnius: Vilniaus universitetas, 2019.
40. ŽIŽIŪNAS, Tadas; AMILEVIČIUS, Darius. 3D and AI Technologies for the Development of Automated Monitoring of Urban Cultural Heritage. *Digital Humanities in the Nordic Countries 2020: Proceedings of the Digital Humanities in the Nordic Countries: 5th Conference (DHN 2020)*, 2020, p. 364–372. [žiūrėta 2022 m. gruodžio 30 d.]. Prieiga per internetą: <<https://ceur-ws.org/Vol-2612/short23.pdf>>.