

Article

The Penetration of Digital Methods into Historical Scholarship: A Text-Mining Analysis of Russian Publications

Zinaida Sokova ¹, Valery Kruzhinov ¹ and Anna Glazkova ^{2,*}¹ Institute of Social Sciences and Humanities, University of Tyumen, Tyumen 625003, Russia² School of Computer Science, University of Tyumen, Tyumen 625003, Russia

* Correspondence: a.v.glazkova@utmn.ru; Tel.: +7-909-182-6371

Abstract

The integration of digital technologies into historical research is a global trend; however, its manifestation varies across national academic traditions. This study investigates the explicit articulation and terminological adoption of digital methods in Russian historical science by analyzing the prevalence and dynamics of specific technological terms in a large corpus of publications. We first constructed a controlled thesaurus of 166 digital technologies by manually curating keyphrases from Russia's primary specialized journal in the field ("Istoricheskaya Informatika", Historical Informatics). This vocabulary was then used to perform text-mining on two distinct corpora: a broad sample of 95K Russian-language history articles from various journals (2004–2024) and a focused sample of publications on the Great Patriotic War History from the Russian Science Citation Index (RSCI, 2014–2023). Our quantitative analysis reveals the frequency, trends, and thematic context of digital method mentions. The findings highlight a significant disparity between the specialized discourse of "Istoricheskaya Informatika" and the mainstream historical publications, while also identifying specific areas (such as archaeological studies) where certain technologies have gained traction. This research offers a novel, data-driven perspective on the "digital turn" in Russian historiography and contributes to the comparative study of digital humanities' global development.

Keywords: digital history; text mining; bibliometrics; Russia; historical informatics; digital methods adaptation; scientific corpora; research trends

1. Introduction

The integration of digital technologies into historical research has fundamentally reshaped the methodological landscape of the discipline. This transformation is often termed "the digital turn" (Volkaert, 2021). This global shift is part of a wider phenomenon of computational abundance, where the ubiquity of digital technologies blurs their distinctness (Berry, 2024). Concurrently, the very nature of knowledge production is being reshaped by the spread of artificial intelligence, extended reality, and hyper-personalized systems across academic fields (Toktas, 2025). The mentioned shift, encompassing techniques from text mining and geographic information systems (GIS) to network analysis and semantic modeling, has enabled historians to interrogate sources at unprecedented scales and pose new research questions (Joo et al., 2022; Silber-Varod & Geri, 2025). Consequently, a novel international field of digital history has emerged, with its own dedicated journals, conferences, and scholarly discourse.



Academic Editor: Tao Chen

Received: 29 November 2025

Revised: 7 January 2026

Accepted: 16 January 2026

Published: 20 January 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

However, the adoption and trajectory of digital methods are not uniform across national academic traditions. Research infrastructures are increasingly recognized as essential pillars that support innovation, collaboration, and openness in research practices across many domains, including the humanities and cultural heritage (Luzietti et al., 2025). The adoption of such practices is deeply influenced by local research infrastructures, historiographical priorities, and linguistic contexts (Ehrmann et al., 2023; Oberbichler et al., 2022). Nevertheless, the specific manifestations of digital history in some national contexts, including Russia, remain significantly underexplored. This gap limits our understanding of the global diversity of the digital humanities and hinders potential scholarly dialog.

In Russia, digital humanities research on history has developed in a very special way. It is mainly known as “historical informatics”. This field stems from earlier work in quantitative history using numbers and statistics. It became a formal field of study when the Department of Historical Informatics was created at Moscow State University in 2004 (Borodkin, 2024). This department continues research work that started in a special laboratory founded in 1991.

The intellectual foundations and methodological trajectory of Russian historical informatics find their comprehensive exposition in the seminal work of Garskova (Garskova, 2018), which traces the discipline’s emergence as an interdisciplinary synthesis of historical scholarship and computational approaches. Borodkin’s decennial analysis of the field’s institutional development (Borodkin, 2014) documents the consolidation of its core research paradigms during its formative period. Contemporary scholarship reflects a dual trajectory. While maintaining continuity with established methodological traditions, the field has experienced significant revitalization through the integration of data science methodologies, artificial intelligence, and machine learning techniques (Borodkin, 2024).

Russian scholarship has cultivated several distinctive research domains that demonstrate the field’s methodological specialization. Historical cartography and spatial analysis constitute a particularly robust research tradition, inaugurated by Vladimirov’s foundational monograph on historical GIS applications (Vladimirov, 2005) and subsequently advanced through the development of dynamic cartographic representations capable of visualizing historical processes and territorial transformations over time (Frolov, 2017).

The domain of three-dimensional (3D) modeling represents another significant specialization, encompassing both architectural reconstruction and landscape visualization. Zherebyatsev’s pioneering work on detailed digital reconstructions of Moscow’s monastic complexes (Zherebyatsev, 2014) established technical standards that have evolved toward immersive technologies, including virtual and augmented reality applications for historical site presentation (Kovalenko, 2024; Nesterov, 2024; Zherebyatsev et al., 2018).

Computational modeling of historical processes constitutes a third major research vector, with Borodkin’s systematic framework for simulating historical scenarios and analyzing counterfactual developments (Borodkin, 2016) representing a methodological landmark. In economic history, distinctive research traditions have emerged around database construction and statistical analysis, particularly in studies of Russian financial institutions and commercial enterprises (Salomatina, 2004; Salomatina & Frenkel, 2016).

The evolving research agenda of Russian historical informatics finds clear expression in conference proceedings and scholarly meetings, which function as barometers of methodological trends. Current evidence indicates the enduring prominence of spatial analysis and 3D reconstruction alongside the accelerating integration of data science approaches, suggesting both continuity in established research programs and expansion into new computational methodologies (Borodkin, 2024; Vladimirov et al., 2020).

Previous research has established historical informatics in Russia as a well-developed field with its own distinct methodologies and dedicated research centers. Considerable knowledge exists about the advanced digital projects created by specialists in this area. However, there remains a significant gap in understanding how widely these digital methods are adopted by the broader community of historians in Russia. While most studies focus on the innovative methods developed by specialists, a critical question persists: to what extent are these digital methods actually being used by mainstream historians in their everyday work?

Our study addresses this knowledge gap. We aim to measure how frequently digital methods are used in standard historical publications in the Russian language. While other researchers have shown what is possible to do with digital history, we examine what is actually being done in practice. Using computer analysis of numerous of Russian historical articles, we provide the first systematic study of the visibility and terminological adoption of digital methods in mainstream publication metadata. This helps us to understand the degree to which these technologies are explicitly foregrounded as methodological components.

Our research is guided by the following research questions:

- RQ1: What is the overall prevalence and specific composition of explicit digital method terminology in the metadata of mainstream Russian historical publications compared to a focused sub-field?
- RQ2: How has the frequency of key digital methods (identified through our thesaurus) evolved over the last decades, and can we identify emerging trends or stagnating techniques?

Using a combined approach of semi-automatically selected terms and large-scale text analysis of publications, this study maps the terminological adoption and visibility of digital technologies in Russian history research. This approach resonates with the existing challenge of matching computational analysis and human experience in digital humanities research (Bakels et al., 2025). The results help us understand the different ways digital tools are used globally and show the challenges of integrating new methods into established research practices.

This study makes several key contributions to the field of digital humanities. Methodologically, it introduces and validates a novel bipartite taxonomy (Hard-Core vs. Second-Tier digital methods) for tracking explicit digital technology adoption. The proposed taxonomy can be adapted for similar bibliometric analysis in other humanities disciplines and linguistic contexts. Empirically, it provides the first large-scale, quantitative assessment of digital method penetration in Russian historical scholarship, moving beyond specialized digital history outlets to analyze mainstream publications. The findings reveal a significant gap between advanced methodological discourse and mainstream practice, while also identifying the specific technologies that have gained the most traction. Finally, the comparative analysis between a general corpus and a focused sub-field offers nuanced insights into the varying rates of methodological adoption across different historical domains.

The article is structured as follows. Section 2 describes the construction of the thesaurus and corpora used in the study as well as the analysis procedure implemented within the research. Following this, Section 3 details the quantitative findings, including the prevalence of specific digital methods and their evolution over time. Section 4 contextualizes these results, examining what they reveal about the adoption and integration of digital technologies in Russian historical scholarship. The article concludes by summarizing the principal conclusions and their broader significance for the global digital humanities community.

2. Materials and Methods

2.1. Thesaurus Construction

The methodological foundation of this study rests upon a self-automatically constructed thesaurus of digital research methods, organized according to a hierarchical taxonomy that distinguishes between definitive technical approaches and broader methodological concepts. This classification system enables precise identification of digital methodologies while accounting for varying levels of specificity in historical scholarship.

2.1.1. Taxonomic Framework

The thesaurus employs a bipartite classification system that categorizes digital methods into two distinct groups based on diagnostic specificity and contextual ambiguity (see Figure 1):

1. **Hard-Core Digital Methods** encompass unambiguous technical approaches with high diagnostic value, including:
 - Programming & Development (specific languages and frameworks, such as Python, R, SQL (Structured Query Language), TensorFlow, PyTorch);
 - Artificial Intelligence & Machine Learning (neural networks, deep learning, and specialized algorithms);
 - Natural Language Processing & Computational Linguistics (tokenization, named entity recognition, semantic analysis, sentiment analysis, etc.);
 - GIS & Spatial Analysis (georeferencing, spatial analysis, historical GIS applications);
 - Computer Vision & 3D Technologies (photogrammetry, lidar scanning, 3D reconstruction, OCR—Optical Character Recognition, etc.);
 - Visualization & VR/AR, i.e., Virtual Reality, Augmented Reality (interactive dashboards, virtual reconstruction, data visualization platforms);
 - Digital Archives & Infrastructures (datasets, corpora, archives, etc.);
 - Specialized Tools & Platforms (other digital methods).
2. **Secondary Methodological Indicators (Second-Tier)** include broader computational concepts requiring contextual validation:
 - Analytical approaches (data analysis, statistical methods, modeling);
 - Infrastructure concepts (databases, web applications, digital archives);
 - Process characteristics (automation, digitization, computational methods).

The division into Hard-Core and Second-Tier digital methods addresses three fundamental challenges in digital methodology identification:

- **Diagnostic Specificity:** The Hard-Core terms exhibit high contextual stability and minimal semantic ambiguity. Technical proper nouns (“Python,” “QGIS,” “TensorFlow”) and domain-specific techniques (“NER,” “LDA”) maintain consistent referential meaning, making them reliable indicators of digital methodological commitment.
- **Contextual Polysemy Mitigation:** The Second-Tier terms demonstrate higher contextual dependency. Concepts like “modeling” and “visualization” address both computational and traditional methodological traditions, requiring triangulation with other digital indicators for accurate classification. Another notable example is the Russian word “граф” which carries dual meanings of “earl” (a noble title) and “graph” (a mathematical structure). In Russian historical scholarship, the aristocratic meaning occurs significantly more frequently, creating potential ambiguity in automated method identification. Such homonyms require additional human checking to avoid disambiguation between historical concepts and computational methods.

- **Disciplinary Adaptation:** Historical scholarship often employs computational methods using discipline-specific nomenclature rather than technical terminology. The bipartite system accommodates this linguistic diversity while maintaining classification rigor.

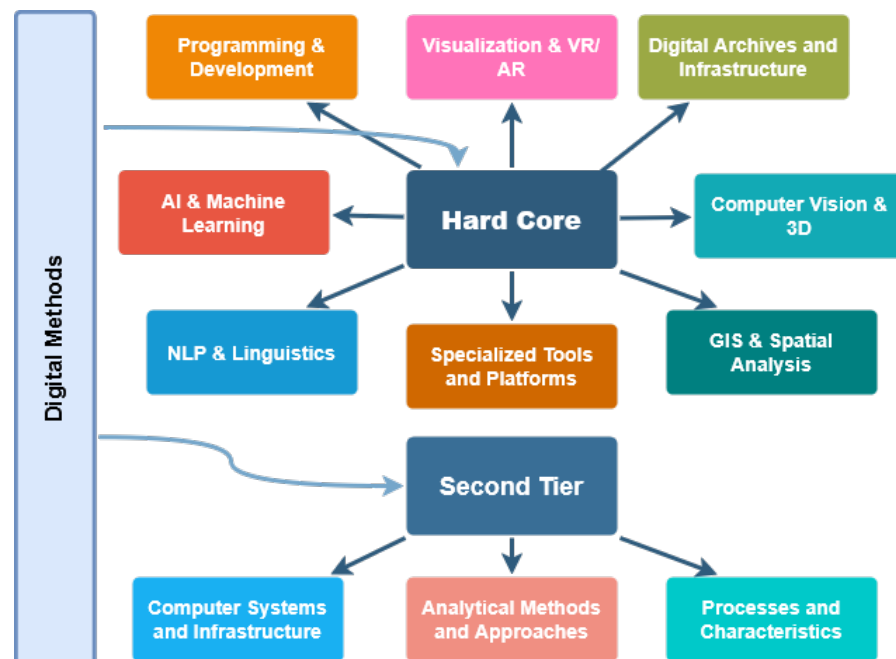


Figure 1. Digital methods taxonomy. The list of the abbreviations used: VR/AR—Virtual Reality, Augmented Reality, AI—Artificial Intelligence, NLP—Natural Language Processing, GIS—Geographic Information Systems.

2.1.2. Compilation and Validation

The thesaurus development followed a rigorous multi-stage process. Initially, we extracted all author-provided keyphrases from the complete publication history of “Istoricheskaya Informatika” (Historical Informatics), representing the most comprehensive source of Russian-language digital history scholarship. In total, 366 articles were included at this stage, yielding 1801 unique author-provided keyphrases prior to expert normalization and curation. The initial set of the articles from “Istoricheskaya Informatika” covered the publications from 2012 to 2025.

Expert historians with digital methodology expertise then applied systematic curation criteria:

- **Core Inclusion:** Specific technologies, software platforms, and technical methods with minimal contextual ambiguity
- **Contextual Inclusion:** Broader methodological concepts that demonstrate digital approaches when combined with core terms
- **Exclusion:** General historical concepts, geographical terms, and traditional methodological approaches

The resulting thesaurus comprises 166 verified terms organized within the taxonomic framework. Each term presented in the form of original Russian spellings was enriched with English equivalents and linguistic variations to accommodate Russian academic writing conventions (for example, “3D-modeling”, “3D modeling”, and “3D”; “GIS” and “geographic information system”; “database” and “DB” for both English and Russian). After adding equivalents, the thesaurus size comprises 406 terms. The original list of 166 terms is provided in Appendix A.

The classification system implements distinct validation protocols for each category. The terms classified as Hard-Core are treated as definitive indicators of digital methodo-

logy. Following the principle that these terms unambiguously identify the technologies mentioned in the text, no additional contextual validation was required. This approach is justified by their technical specificity and limited usage outside digital research contexts. The identification of Second-Tier terms was manually validated by domain experts. During expert review, additional relevant terms were occasionally identified but excluded if they could not be consistently operationalized across the full corpus.

This systematically curated vocabulary provides a domain-specific foundation for detecting digital methodological patterns in historical texts, enabling both precise identification of technical approaches and nuanced recognition of digitally-inflected research methodologies.

2.2. Corpus Preparation

For this study, two distinct corpora of Russian-language scientific publications were compiled and analyzed. Utilizing two different datasets enables a comparative analysis of methodological trends across both a broad disciplinary landscape and a specific, well-defined research area.

- **Corpus 1: General History.** This dataset comprises metadata for approximately 95,000 scholarly articles available in a digital format. The General History corpus was constructed using the CyberLeninka open-access platform¹ as the primary data source, which aggregates Russian-language scholarly journals across multiple indexing tiers, including journals indexed in the Russian Science Citation Index (RSCI), Web of Science, and Scopus. The Cyberleninka is known for its extensive coverage of journals across various disciplines (Gasparyan et al., 2019; Semyachkin et al., 2014).

Since not all journals provide full-text articles in open access, the data collection was focused on metadata that typically encapsulates the core methodological and thematic substance of an article: namely, the title, abstract, and author keyphrases (Zhang et al., 2025). For each entry, these three elements were concatenated into a single text field, forming a representative proxy for the article's primary content. Data were harvested using automated parsing techniques with the BeautifulSoup library² for Python. The data collection was performed in October 2024. Subsequent to the initial collection, a data preprocessing procedure was implemented. This involved the removal of entries not in the Russian language and records lacking a substantive abstract.

The final curated corpus consists of approximately 95,000 text entries published between 2004 and 2023. It is important to acknowledge that the open-access nature of the CyberLeninka platform may introduce certain coverage biases. For instance, prestigious subscription-based journals with limited open-access policies may be underrepresented. Similarly, the corpus likely reflects the output of major research centers more comprehensively than that of smaller regional institutions. Furthermore, authors publishing in open-access venues might exhibit different patterns in methodological self-presentation compared to those in traditional subscription journals. Despite these potential biases, the corpus's substantial size, temporal span, and inclusion of journals from multiple indexing tiers (RSCI, Web of Science, Scopus) provide a robust and extensive sample of mainstream Russian-language historical discourse, suitable for analyzing the visibility of digital methods in scholarly communication.

- **Corpus 2: Great Patriotic War History.** The study of the Great Patriotic War, the term used in Soviet and post-Soviet historiography to denote the period of World War II from the Nazi Germany invasion of the USSR in 1941 to its victory in 1945, represents one of the most prominent and enduring sub-fields within Russian historical science. The Great Patriotic War History corpus was compiled from journals indexed in the Russian Science Citation Index on the Web of Science platform (RSCI WoS). The RSCI WoS database aggregates leading Russian scholarly journals and is integrated into the international WoS citation system, ensuring a selection of high-quality publications. The corpus was constructed by analyzing 13 leading Russian periodicals in history, selected by experts in the field (Sokova et al., 2025). These journals were included based on their status within the RSCI WoS (as of May 2024) and their focus on Russian history. For all articles published in these journals between 2014 and 2023, metadata, which include titles, keyphrases, and abstracts, were collected from their open-access sources on the CyberLeninka platform and official journal websites using the BeautifulSoup library. This metadata were concatenated into a single text per article, similar to the General History corpus. The data collection, performed automatically from May to June 2024, initially yielded over ten thousand articles from the target journals and periods.

To identify the articles specifically related to the Great Patriotic War, a two-stage filtering process was employed. First, based on an empirical analysis of publications in the RSCI database, expert historians compiled two lists of thematic markers: (a) terms that unequivocally identify the topics of the Great Patriotic War and (b) terms that possibly indicate such topics. The collected texts were lemmatized, and a search was conducted for terms and their derivatives from the first (definitive) list. The resulting articles formed the initial core collection. Subsequently, a search for terms from the second (potential) list was conducted. The articles containing these markers were manually reviewed by experts and added to the collection only if they were confirmed to pertain to the Great Patriotic War. Finally, the entire collection underwent a final expert review. The resulting curated corpus consists of 544 texts, providing a focused sample to examine the penetration of digital methods into a well-established research area. Expert validation was conducted by two researchers from the University of Tyumen, Russia, with formal training in digital history. Disagreements were resolved through discussion until consensus was reached.

The following additional filters were applied:

- **Language filter:** only Russian-language publications were retained.
- **Document consistency filter:** articles were included if they contained a title, abstract, and author-provided keyphrases available in open access.

The year distribution for both corpora is presented in Figures 2 and 3. The data statistics are shown in Table 1.

Table 1. Corpora statistics.

Characteristic	General History	Great Patriotic War History
Average text length (symbols)	981.28	1108.66
Standard deviation for avg text length	602.53	728.18
Median text length (symbols)	798	879
Number of texts	95,720	545

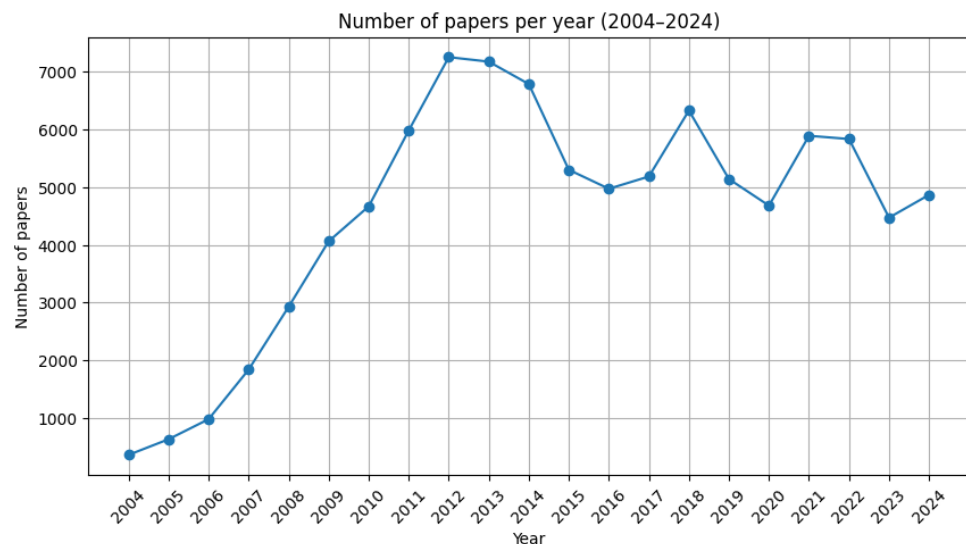


Figure 2. Text distribution per year (General History).

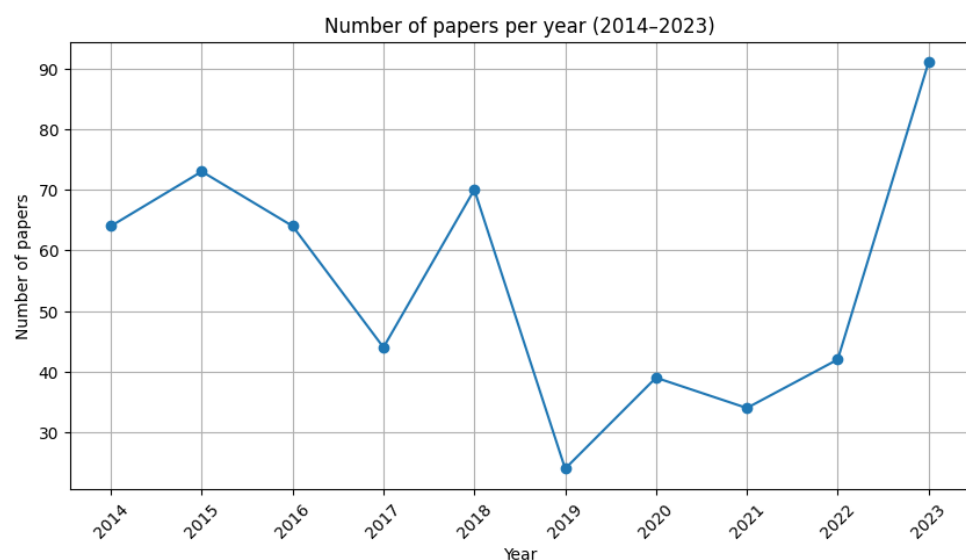


Figure 3. Text distribution per year (Great Patriotic War History).

2.3. Analysis Procedure

Prior to analysis, all textual data from both corpora underwent pre-processing consisting of the following steps: text normalization (lowercasing, punctuation removal), lemmatization using the PyMorph library (Korobov, 2015), and removal of stop-words while preserving technical terminology. The list of stop-words was sourced from the NLTK package (Bird, 2006).

The core quantitative analysis relied on pattern-matching algorithms to identify occurrences of thesaurus terms within the processed texts. For each corpus, the absolute number of publications containing at least one mention of a digital method was calculated to determine the overall prevalence. These counts were further disaggregated into Hard-Core and Second-Tier categories.

All occurrences of Second-Tier terms underwent manual expert verification by two researchers to ensure contextual relevance and mitigate potential polysemy issues. The experts independently assessed whether each term, in its specific context, referred to a computational methodology (e.g., mathematical or simulation “modeling”) rather than a traditional historical concept (e.g., interpretative “modeling” of a historical process).

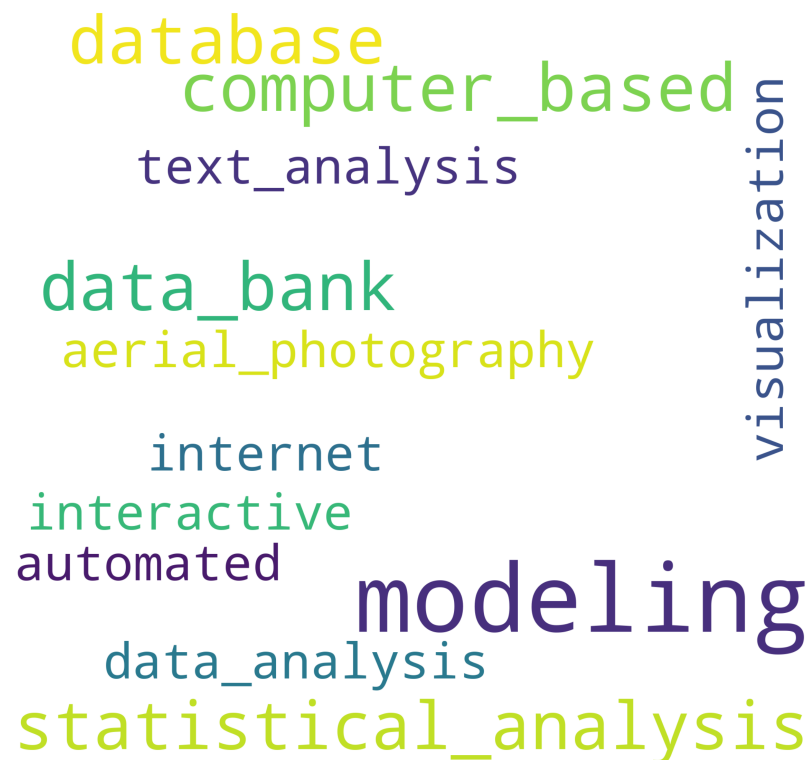


Figure 5. Word cloud of digital methods (Great Patriotic War History).

3.2. Composition and Thematic Context of Digital Terminology

Frequency analysis and word cloud visualizations (Figures 4 and 5) revealed the composition of the most frequently used terms. Thus, in the General History corpus, the most frequent terms were from the Second-Tier: modeling (12.78% of the Second-Tier mentions), database (12.11%), software (11.87%), internet (11.87%), data analysis (10.89%), digital (8.34%), mapping (7.18%). Among the Hard-Core digital methods, the most prevalent were GIS technologies (42.31% of the Hard-Core mentions), cluster analysis (12.02%), text corpus (6.73%), 3D modeling methods (6.73%), photogrammetry (6.25%), and regression analysis (3.37%). Mentions of specialized tools like Leaflet and GPS were less common but present. In the Great Patriotic War corpus, the range of methods was significantly narrower and more basic. General terms predominated: modeling (18.75% of the Second-Tier mentions), database (12.5%), data bank (12.5%), statistical analysis (12.5%), computer-based (12.5%). The only Hard-Core digital methods identified were aerial photography and mentions of GIS technologies.

A thematic analysis of publications containing digital methods showed their concentration in specific areas:

- GIS technologies were applied for spatio-temporal analysis (e.g., modeling migration routes, settlement structures) and the reconstruction of historical landscapes.
- Cluster analysis was used to identify typologies and patterns in data, for example, to classify archaeological artifacts or analyze parliamentary voting.
- 3D modeling and photogrammetry were primarily used in archaeology and architectural history for the virtual reconstruction of cultural heritage objects.
- The Second-Tier terms (modeling, data analysis, databases, etc.) often described the creation of digital research infrastructure (e.g., databases of archival documents) or the application of basic statistical methods to analyze historical sources.

Although a full co-occurrence network visualization is beyond the scope of this paper, a qualitative inspection of frequent joint mentions reveals several stable method clusters. The most common co-occurring terms involve combinations of databases, statistical analysis, and quantitative methods, typically associated with demographic, social, and economic history topics. A second recurring cluster links GIS with spatial analysis and cartographic visualization, predominantly in studies of regional development, settlement patterns, and military operations. Less frequently, 3D modeling co-occurs with visualization and digital reconstruction, primarily in archaeology and material culture research. These patterns suggest that digital methods are not adopted in isolation but tend to appear in recognizable methodological bundles tied to specific thematic domains.

3.3. Dynamics of Mentions over Time

The analysis of time series in the general corpus revealed a slow but steady increase in the absolute number of publications containing digital method terms from 2004 to 2024 (see Figure 6). In absolute terms, the number of publications mentioning digital methods is increasing. However, this is primarily linked to the overall growth in publication volume across the years. Regarding relative measures, both groups of digital methods show an upward trend. Second-Tier methods are more evenly distributed from an empirical standpoint. The trend in the use of Hard-Core methods appears more pronounced. Overall, the total number of publications that mention digital methods remains low throughout the entire period under review, without any significant spikes. In the Great Patriotic War corpus, no significant changes in the frequency of digital method use were identified over the period 2014–2023 due to the small number of digital method mentions.

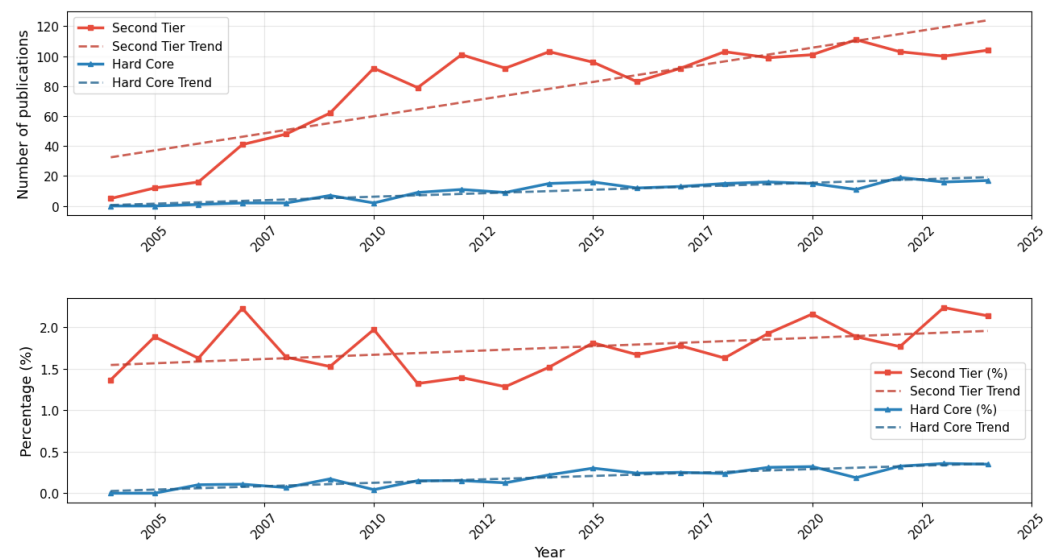


Figure 6. Dynamics of digital method mentions in the General History corpus. Absolute values are shown at the (top), relative values at the (bottom).

To illustrate the individual trajectories of the most prominent Hard-Core technologies, Figure 7 presents the normalized yearly mention counts for GIS and 3D modeling within the General History corpus. For visual comparison, the figure also includes the aggregated trend for all digital method mentions (scaled proportionally to the same vertical range). The plot clearly demonstrates that GIS technologies represent the dominant and most consistently growing trend within the Hard-Core category, significantly outpacing the mention frequency of 3D modeling. This visualization corroborates the findings from the frequency analysis (Section 3.2) and provides concrete evidence that the slow, “confined growth” observed at the aggregate level is primarily driven by the increasing, yet still niche,

application of geospatial methods in historical research. The trajectory of 3D modeling, while positive, remains substantially lower and more volatile.

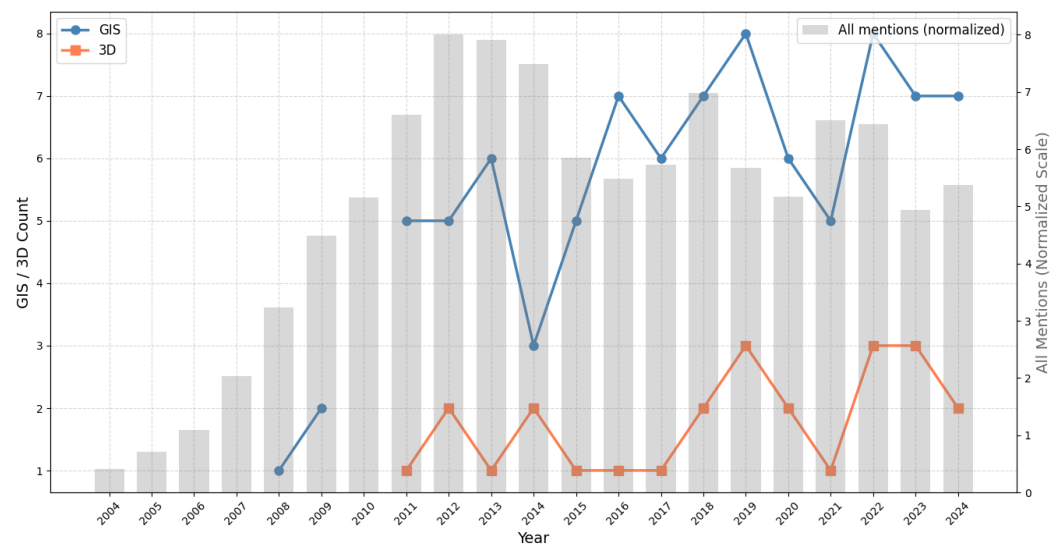


Figure 7. Dynamics of GIS and 3D mentions. The counts of all digital methods mentions were scaled proportionally to fit the same vertical range as GIS/3D mentions for visual comparison. Normalization formula: $\text{Normalized} = (\text{Original} / \text{Max}(\text{All Digital Methods})) \times \text{Max}(\text{GIS}, 3\text{D})$. Gray bars show relative trends, not absolute counts.

4. Discussion

The analysis conducted provides a quantitative assessment of the hypothesis regarding the uneven penetration of the “digital turn” into Russian historiography. The results clearly demonstrate a significant gap between the well-developed research field of “historical informatics” and the practices of mainstream historians.

4.1. Limited Discursive Penetration of Digital Method Terminology

To address the question of overall prevalence and specific composition (RQ1), the finding that only about 2% of historical publications overall and less than 1% of Great Patriotic War publications contain any traces of digital methods is a key outcome of our research. This confirms that, despite the existence of an established school of historical informatics in Russia with its own research centers, journals, and methodologies (Borodkin, 2014, 2024; Garskova, 2018), explicit methodological self-positioning using digital-method terminology remains limited outside specialist communities. The explicit methodological foregrounding of digital methods has not become a “new normal” for the majority of Russian-language historical publications. While specialized literature in historical informatics discusses artificial neural networks, machine learning, and complex natural language processing methods (Borodkin, 2024), mainstream publications are dominated by basic concepts: “database”, “modeling”, “data analysis.” Often, these terms describe not complex computational procedures but the process of creating a digital archive or applying simple statistics.

Importantly, this finding should be interpreted as a measure of discursive diffusion rather than a complete census of research practice. The presence or absence of specific technical terms in titles, abstracts, and keyphrases primarily reflects how historians choose to articulate and legitimize their methodological choices in scholarly communication. It is possible, even likely, that digital tools (e.g., databases, spreadsheets, statistical software) are used at various research stages without being explicitly foregrounded in metadata that prioritizes historical interpretation over methodological exposition. Therefore, our

study primarily maps the rhetorical adoption and visibility of digital methodology within the discipline's communicative ecosystem, revealing a significant gap between advanced methodological discourse and mainstream methodological self-description.

This distinction is analytically important in its own right, as it captures barriers of communication, disciplinary norms, and processes of methodological self-identification. In many cases, historians may rely on digital infrastructures, databases, spreadsheets, or statistical software without foregrounding these tools in abstracts that prioritize substantive historical interpretation over methodological exposition.

The predominance of the Second-Tier digital methods and technologies, such as GIS and 3D modeling from the Hard-Core, indicates that digital methods enter historical scholarship primarily where they solve specific applied tasks: visualization (maps, reconstructions) and the systematization of large data arrays (archaeological artifacts, archival documents). This aligns with the previously described specializations of Russian historical informatics, such as historical cartography and 3D reconstruction (Frolov, 2017; Vladimirov, 2005; Zherebyatev, 2014).

The comparison between the two corpora is particularly revealing. The history of the Great Patriotic War, being one of the most funded and significant areas of Russian historical science, shows an extremely low level of digitalization. This suggests that barriers to adopting new methods are especially high in established, traditional sub-fields with strong, entrenched research paradigms. At the same time, findings for the Great Patriotic War corpus should be interpreted as exploratory, given the small absolute number of publications containing any digital method mentions. The value of this corpus lies primarily in its contrastive function, highlighting differences between a highly institutionalized sub-field and the broader disciplinary landscape.

Despite the identified challenges in mainstream methodological adoption, it is important to acknowledge the distinct and valuable contributions of Russian scholarly schools in the domain of digital history. The field of historical informatics in Russia has cultivated several robust research traditions. Notably, the development of historical GIS applications has provided sophisticated methodologies for spatial analysis of historical processes. Furthermore, significant expertise has been demonstrated in the domain of 3D modeling for cultural heritage reconstruction, allowing for detailed visualization of historical sites. These specialized research vectors, alongside established practices in quantitative history and data analysis, represent a solid scholarly foundation. They demonstrate the capacity for methodological innovation within the national research landscape and offer promising pathways for the future integration of digital tools into broader historical scholarship.

Regarding the temporal evolution of digital methodology mentions (RQ2), our longitudinal analysis reveals a pattern of confined growth rather than widespread adoption. This growth originates from a very low baseline and does not translate into widespread methodological normalization, resulting in sustained marginality despite long-term upward trends. Although the absolute number of publications referencing digital methods demonstrates a gradual increase from 2004 to 2024, this growth remains proportionally insignificant when compared to the total volume of historical publications. This discrepancy suggests that digital approaches are developing within established specialist communities without achieving broad acceptance across the discipline. The complete absence of meaningful growth in the Great Patriotic War corpus further confirms the resilience of traditional methodological paradigms in well-established research domains. These dynamics collectively indicate that while digital methods are gaining limited traction in Russian historiography, they have not yet reached the critical mass necessary to transform mainstream research practices. The absence of any single technology emerging as dominant underscores the fragmented nature of digital adoption in the field.

4.2. Potential Barriers to Adoption

Discussing the reasons for this situation goes beyond the scope of this quantitative study, but our results allow us to suggest several explanations consistent with the literature (Ehrmann et al., 2023; Oberbichler et al., 2022), which can be summarized as follows:

- **Methodological Traditions:** Russian historical science has strong traditions of qualitative, source-based analysis, within which digital methods may be perceived as alien.
- **Infrastructure and Training:** The lack of necessary digital infrastructure and systematic training in digital methods within history departments creates a high barrier to entry for researchers.
- **Linguistic and Cultural Context:** The dominance of English-language interfaces in complex software and a lack of localized educational resources could be an additional barrier.

Furthermore, our findings have direct implications for understanding the evolution of scientific communication in the humanities. The significant disparity in methodological terminology between the specialized journal “Istoricheskaya Informatika” and mainstream historical publications suggests the emergence of two distinct communication ecosystems within Russian historical science. This methodological divide is reflected in and reinforced by publication practices, where specialized venues serve as hubs for methodological innovation while mainstream journals maintain traditional research narratives. This situation creates a challenge for the holistic assessment of historical scholarship, as bibliometric indicators tracking methodological innovation would be largely confined to a narrow segment of the publication landscape. For journal editors and research policymakers, our study highlights the need to foster greater methodological integration, potentially through special issues, interdisciplinary reviewer boards, and explicit guidelines that acknowledge the value of digital methodologies, thereby enriching the overall discourse of historical research.

4.3. Limitations

This research has several limitations. First, the analysis was based on metadata (titles, abstracts, keyphrases) rather than full-text articles, so some instances of method usage might have been missed. While restricting the analysis to explicitly labeled “Methods” sections could reduce background noise, such sections are inconsistently present in Russian-language historical journals. We therefore opted for a uniform metadata-based approach (titles, abstracts, keyphrases) to ensure comparability across time and venues. This choice likely overestimates mentions of general digital discourse and underestimates tacit digital practice, which we explicitly acknowledge. Second, the thesaurus used, although carefully compiled, cannot be exhaustive. Third, the quantitative analysis does not reveal the depth or quality of the use of a particular method but only records its mention. Although large language models offer promising tools for term expansion and semantic disambiguation (for example, in works (Cai et al., 2022; Gousyatskaya & Loukachevitch, 2025)), we deliberately relied on expert-driven curation to ensure interpretability, domain specificity, and reproducibility across time periods.

Furthermore, while this study analyzes a substantial corpus of publications, it does not claim to be a comprehensive account of all developments in Russian historical science over the past decades. Our sampling strategy, focused on mainstream journals and open-access online platforms, provides a specific perspective on methodological adoption. Additionally, we intentionally did not conduct a separate, detailed analysis of technologically intensive sub-fields such as historical reconstruction and archeology. The application of digital methods in these areas is already well-documented in specialized outlets like “Istoricheskaya Informatika” (Historical Informatics), which are precisely the venues designed

to showcase the field's advanced technological capabilities. Our goal was to measure the penetration of these methods beyond their established niches into the broader landscape of historical research.

5. Conclusions

This study has provided a novel, data-driven assessment of the penetration of digital methods into mainstream Russian historical scholarship. By employing a systematically constructed thesaurus and text-mining a large corpus of publications, we have moved beyond the well-documented specialized discourse of “historical informatics” to measure the extent to which the terminology of these methods is explicitly adopted in the metadata of broader research practices.

Our analysis yields two principal conclusions. First, a significant disparity exists between the advanced methodological developments within Russian historical informatics and their application in mainstream historical research. This is evidenced by our finding that only 1.72% of general history publications and a mere 0.7% of specialized Great Patriotic War studies contain explicit mentions of definitive Hard-Core digital methods in their publication metadata. These figures indicate that the explicit articulation of digital methods as central methodological components remains largely confined to specialist venues. Therefore, the foregrounding of such methods in publication metadata continues to be limited outside these specialized circles. Second, the adoption that does occur is qualitatively specific. It is concentrated in applied areas such as spatial analysis (GIS), cultural heritage reconstruction (3D modeling), and the creation of digital research infrastructures (databases). In contrast, advanced computational techniques like artificial intelligence and natural language processing have yet to gain significant traction.

These findings highlight the complex dynamics of methodological innovation in national academic traditions. They suggest that the presence of strong specialized schools is a necessary but insufficient condition for the widespread integration of digital tools, which is also influenced by deeply entrenched research paradigms, educational structures, and infrastructural support.

The main limitations of this study are its reliance on metadata and the inherent constraints of a thesaurus-based approach, which capture the mention of a method but not the depth or sophistication of its application. Future research should involve full-text analysis to gain a more nuanced understanding of methodological integration. Furthermore, qualitative studies, such as surveys and interviews with historians, are needed to explore the perceived barriers, incentives, and training needs that shape their methodological choices. Additionally, extending the analysis to incorporate variables such as institutional affiliation, funding sources, or regional context could reveal important socio-institutional determinants of digital method adoption.

Despite its limited mainstream penetration, the robust development of historical informatics in Russia provides a solid foundation for future growth. Bridging the identified gap will require targeted efforts in methodological education, infrastructure development, and fostering dialog between specialized digital historians and the wider scholarly community. This research contributes to the global comparative study of digital humanities by mapping the unique trajectory of digital history in Russia, underscoring the diversity of its global development.

From the perspective of scientific communication, this study demonstrates how large-scale text mining of publication metadata can serve as a powerful diagnostic tool for research evaluators, journal editors, and science policymakers. The bipartite taxonomy (Hard-Core vs. Second-Tier) developed here provides an adaptable framework for monitoring the terminological adoption of new methodological paradigms across different

disciplines and national contexts. By quantifying the gap between pioneering methodological discourse and mainstream practice, our approach helps identify fields where methodological transfer is lagging. For the editors of mainstream history journals, these findings underscore an opportunity to actively bridge this gap by encouraging submissions that integrate digital methods, thereby fostering a more methodologically diverse and innovative publication ecosystem. Future research could apply this methodology to track the diffusion of other interdisciplinary approaches, further illuminating the dynamic relationship between methodological innovation and scientific publishing. Future work could also combine controlled vocabularies with LLM-based discovery of emerging or implicit methodological terminology.

Author Contributions: Conceptualization, Z.S. and V.K.; methodology, Z.S. and A.G.; software, A.G.; validation, Z.S., V.K. and A.G.; data curation, A.G.; writing—original draft preparation, Z.S.; writing—review and editing, Z.S., V.K. and A.G.; project administration, Z.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: We are grateful to Nadezhda Zhuravleva and Valeria Evdash (Center for Academic Writing “Impulse”, University of Tyumen) for their assistance with the English language.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Digital Methods

This section presents a list of digital methods utilized in this study. The list is translated from the Russian language.

1. Hard-Core Digital Methods:

- Programming & Development: “python”, “r package”, “programming language”, “sql”, “postgresql”, “mysql”, “dbms”, “api”, “json”, “xml”.
- Artificial Intelligence & Machine Learning: “machine learning”, “neural network”, “artificial intelligence”, “deep learning”, “tensorflow”, “pytorch”, “svm”, “arima”, “cluster analysis”, “regression analysis”.
- Natural Language Processing & Computational Linguistics: “nlp”, “natural language processing”, “computational linguistics”, “topic modeling”, “lda”, “large language models”, “semantic analysis”, “named entity recognition”, “ner”, “vectorization”, “word2vec”, “embedding”, “stylometry”, “spacy”, “nltk”, “stanza”, “udpipe”, “treetagger”, “mystem”, “huggingface”, “allennlp”, “gensim”, “transformers”, “sentiment analysis”, “dependency parsing”, “syntactic analysis”, “keyness analysis”, “keyword analysis”, “collocation analysis”, “tokenization”, “lemmatization”, “data annotation”.
- GIS & Spatial Analysis: “qgis”, “arcgis”, “nextgis”, “geoserver”, “postgis”, “geopandas”, “geographic information system”, “web-gis”, “historical gis”, “spatial analysis”, “georeferencing”, “geocoding”, “leaflet”, “openlayers”, “mapbox”, “cartography”, “gps”, “geotracker”, “track navigation”, “tracker”, “aerial photography”.
- Computer Vision & 3D Technologies: “computer vision”, “pattern recognition”, “ocr”, “opencv”, “3d model”, “3d reconstruction”, “3d”, “lidar”, “laser scanning”, “photogrammetry”, “aerial survey”, “point cloud”, “digital elevation model”, “voxel”, “autodesk”, “3ds max”, “sketchup”, “bim”, “blender”, “three.js”.

- Visualization & VR/AR: “data visualization”, “tableau”, “gephi”, “d3.js”, “plotly”, “dashboard”, “interactive map”, “virtual reality”, “vr”, “augmented reality”, “ar”, “unity”, “unreal engine”.
 - Digital Archives & Infrastructures: “digital archive”, “born-digital”, “iiif”, “dataset”, “text corpus”, “machine translation”.
 - Specialized Tools & Platforms: “gpt”, “chatgpt”, “bert”, “deepseek”, “access”, “excel”, “wordpress”, “uav”, “prompt engineering”, “digital humanities”.
2. Second-Tier:
- Analytical approaches: “text analysis”, “data analysis”, “visualization”, “big data”, “mapping”, “quantitative analysis”, “statistical analysis”, “modeling”, “automatic classification”, “mathematical method”, “regression analysis”, “distant reading”.
 - Infrastructure concepts: “database”, “data bank”, “web application”, “web service”, “web resource”, “internet resource”, “network resource”, “internet”, “information system”, “information retrieval system”, “electronic catalog”, “electronic archive”, “electronic resource”, “computing resources”, “cloud computing”, “distributed systems”, “data warehouse”, “database management system”
 - Process characteristics: “data processing”, “software”, “text recognition”, “digitization”, “graph”, “datafication”, “software product”, “automated”, “automatic”, “computer-based”, “digital”, “interactive”, “computer model”, “interoperability”, “scalability”, “visual data”.

Notes

¹ <https://cyberleninka.ru/>, assessed on 11 November 2025.

² <https://beautiful-soup-4.readthedocs.io/en/latest/>, assessed on 11 November 2025.

References

- Bakels, J. H., Grotkopp, M., Scherer, T., Stratil, J., Jiaming, T. Z., & Dongrui, C. (2025). Matching computational analysis and human experience: Performative arts and the digital humanities. *Digital Humanities Research*, 5(2), 59.
- Berry, D. M. (2024). Post-digital humanities: Computation and cultural critique in the arts and humanities. *EDUCAUSE Review*, 49(3), 24–26.
- Bird, S. (2006, July 17–18). *NLTK: The natural language toolkit*. COLING/ACL 2006 Interactive Presentation Sessions (pp. 69–72), Sydney, Australia.
- Borodkin, L. I. (2014). Historical informatics at the Faculty of History of Moscow State University: From computers to supercomputers. In S. P. Karpov (Ed.), *Problems of historiography, source studies and methods of historical research. Proceedings of the 5th scholarly readings in memory of academician I. D. kovalchenko* (pp. 40–47). Moscow University Press. Available online: <https://istina.msu.ru/publications/article/8751589/> (accessed on 8 January 2026). (In Russian)
- Borodkin, L. I. (2016). *Modeling historical processes: From the reconstruction of reality to the analysis of alternatives*. Aletheia. (In Russian)
- Borodkin, L. I. (2024). 20 years of the department of historical informatics of the faculty of history of Moscow State University: New trends in interdisciplinary research. *Historical Informatics*, 3, 16–32. (In Russian). [CrossRef]
- Cai, S., Venugopalan, S., Tomanek, K., Narayanan, A., Morris, M., & Brenner, M. (2022). Context-aware abbreviation expansion using large language models. In *Proceedings of the 2022 conference of the north american chapter of the association for computational linguistics: Human language technologies* (pp. 1261–1275). Association for Computational Linguistics.
- Ehrmann, M., Bunout, E., & Clavert, F. (2023). *Digitised historical newspapers: A changing research landscape* (pp. 1–22). De Gruyter Oldenbourg.
- Frolov, A. A. (2017). A dynamic map as the basis of a historical map in a GIS environment. *Historical Informatics*, 2, 61–73. (In Russian)
- Garskova, I. M. (2018). *Historical informatics: Methodological and historiographical aspects of development* [Doctor dissertation, Russian State University for the Humanities]. (In Russian)
- Gasparyan, A. Y., Yessirkepov, M., Voronov, A. A., Koroleva, A. M., & Kitaz, G. D. (2019). Comprehensive approach to open access publishing: Platforms and tools. *Journal of Korean Medical Science*, 34(27), e184. [CrossRef] [PubMed]
- Gousyatskaya, P., & Loukachevitch, N. (2025). Word sense disambiguation in Russian: A generative LLM approach. In M. Bakaev, R. Bolgov, A. Chizhik, A. Chugunov, V. Demareva, Y. Kabanov, R. Pereira, R. Elakkiya, & W. Zhang (Eds.), *Internet and modern society, IMS 2025, communications in computer and information science* (Vol. 2671). Springer.

- Joo, S., Hootman, J., & Katsurai, M. (2022). Exploring the digital humanities research agenda: A text mining approach. *Journal of Documentation*, 78(4), 853–870. [\[CrossRef\]](#)
- Korobov, M. (2015). Morphological analyzer and generator for Russian and Ukrainian languages. In *International conference on analysis of images, social networks and texts* (pp. 320–332). Springer International Publishing.
- Kovalenko, I. R. (2024). Realistic 3D model of the Albazinsky Fort. *Bulletin of Amur State University. Series: Humanities*, 104, 142–147. Available online: https://vestnik.amursu.ru/wp-content/uploads/2024/03/n104_142-147.pdf (accessed on 8 January 2026). (In Russian)
- Luzietti, R. B., Spadi, A., Giampietro, N., Mancuso, G., Caravale, A., D'Eredità, A., Caradonna, M., Moscati, P., Quochi, V., Monachini, M., & Degl'Innocenti, E. (2025). Digital humanities and heritage science: Moving from landscaping to a dynamic research observatory in an open science cloud. *Umanistica Digitale*, 9(20), 419–439.
- Nesterov, S. P. (2024). Dynamics of structures and historical reconstruction of the Albazin Fort on the Amur River. *Vestnik NSU. Series: History and Philology*, 23(7), 105–115. (In Russian) [\[CrossRef\]](#)
- Oberbichler, S., Boros, E., Doucet, A., Marjanen, J., Pfanzelter, E., Rautiainen, J., Toivonen, H., & Tolonen, M. (2022). Integrated interdisciplinary workflows for research on historical newspapers: Perspectives from humanities scholars, computer scientists, and librarians. *Journal of the Association for Information Science and Technology*, 73(2), 225–239. [\[CrossRef\]](#) [\[PubMed\]](#)
- Salomatina, S. A. (2004). *Commercial banks in Russia: Dynamics and structure of operations, 1864–1917*. ROSSPEN. (In Russian)
- Salomatina, S. A., & Frenkel, O. I. (2016). Regional development of the Russian joint-stock commercial banks in the second half of the 19th century: Statistics and GIS-technologies. *History (Electronic Scientific and Educational Journal)*, 7(51), 17. (In Russian) [\[CrossRef\]](#)
- Semyachkin, D., Kislyak, E., & Sergeev, M. (2014). CyberLeninka: Open access and CRIS trends leading to open science in Russia. *Procedia Computer Science*, 33, 136–139. (In Russian) [\[CrossRef\]](#)
- Silber-Varod, V., & Geri, N. (2025). Winds of generative AI: Research trends of digital humanities in computer science publications. *Online Journal of Applied Knowledge Management (OJAKM)*, 13(1), 1–12. [\[CrossRef\]](#)
- Sokova, Z. N., Kruzhinov, V. M., & Glazkova, A. V. (2025). Topic modeling of scientific texts using BERTopic (Based on scientific abstracts about the Great Patriotic War of 2014–2023). *NSU Vestnik. Series: Linguistics and Intercultural Communication*, 23(3), 107–122. (In Russian) [\[CrossRef\]](#)
- Toktas, E. (2025). Future scenarios of digital humanities and post-humanist education. *Journal of Foresight and Health Governance*, 2(1), 21–31.
- Vladimirov, V. N. (2005). *Historical geoinformatics: Geographic information systems in historical research*. Altai University Press. (In Russian)
- Vladimirov, V. N., Garskova, I. M., & Frolov, A. A. (2020). Historical informatics in a new interdisciplinary field: Academic symposium dedicated to the 15th anniversary of the department of historical informatics of Moscow University. *Historical Informatics*, 1, 158–170. (In Russian) [\[CrossRef\]](#)
- Volkaert, F. (2021). OK computer? The digital turn in legal history: A methodological retrospective. *Tijdschrift voor Rechts-geschiedenis/Revue d'histoire du droit/The Legal History Review*, 89(1–2), 1–46. [\[CrossRef\]](#)
- Zhang, C., Yan, X., Zhao, L., & Zhang, Y. (2025). Enhancing keyphrase extraction from academic articles using section structure information. *Scientometrics*, 130(4), 2311–2343. [\[CrossRef\]](#)
- Zherebyatev, D. I. (2014). *Three-dimensional computer modeling methods in the tasks of historical reconstruction of monastic complexes of Moscow*. MAKSPress. (In Russian)
- Zherebyatev, D. I., Malyshev, A. A., & Moor, V. V. (2018). Gorgippia in the Archaic period: Methods and technologies of 3D reconstruction of an ancient Fortress-City. *Historical Informatics*, 25(3), 33–50. (In Russian). [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.