



Contents lists available at [Journal IICET](#)

JPPI (Jurnal Penelitian Pendidikan Indonesia)

ISSN: 2502-8103 (Print) ISSN: 2477-8524 (Electronic)

Journal homepage: <https://jurnal.iicet.org/index.php/jppi>



Digital humanities approaches to analyzing indonesian language texts as non-western languages

Rastya Mutiarani Zahra^{*)}

Universitas Pendidikan Indonesia, Indonesia

Article Info

Article history:

Received Oct 23th, 2024

Revised Nov 21th, 2024

Accepted Dec 26th, 2024

Keyword:

Text mining

Natural language processing

Indonesian literary texts

Digital archiving for cultural heritage

ABSTRACT

This research explores the integration of digital humanities methods in the analysis of Indonesian language texts to enhance linguistic and cultural understanding. The primary objective is to develop tailored digital humanities methodologies, applying computational tools such as text mining, natural language processing, and corpus linguistics to analyze linguistic and thematic patterns within Indonesian texts. By leveraging these techniques, the study aims to provide new insights into language use, cultural narratives, and historical shifts in Indonesia. A qualitative approach, including a literature review and case studies, is used to examine existing research and methodologies, and assess how digital tools can be effectively applied in this context. The study also addresses the accessibility of Indonesian textual data for researchers, educators, and students, proposing solutions to make these resources more usable and integrated into the global digital humanities framework. This research contributes to expanding the scope of digital humanities by incorporating Indonesian language texts, offering a model for future studies in non-Western linguistic traditions.



© 2024 The Authors. Published by IICET.

This is an open access article under the CC BY-NC-SA license
(<https://creativecommons.org/licenses/by-nc-sa/4.0>)

Corresponding Author:

Rastya Mutiarani Zahra,

Universitas Pendidikan Indonesia

Email: rastyamutiz@upi.edu

Introduction

Digital humanities is an interdisciplinary field that combines traditional humanities research with digital technologies to enhance the study and understanding of human culture, history, and society. This field utilizes computational tools and methods to analyze, visualize, and interpret vast amounts of data, enabling researchers to uncover patterns and insights that were previously inaccessible through conventional means. Digital humanities encompass a wide range of activities, including text mining, digital mapping, data visualization, and the creation of digital archives.

One of the key aspects of digital humanities is its ability to handle large datasets, often referred to as "big data." This capability allows scholars to conduct extensive analyses of texts, images, and other cultural artifacts, leading to new discoveries and interpretations. For instance, text mining techniques can reveal trends and themes across a large corpus of literature, while digital mapping can illustrate historical changes in geographic and social landscapes. The use of digital tools also facilitates collaboration among scholars from different disciplines. Historians, literary critics, linguists, and computer scientists can work together on projects, combining their expertise to achieve more comprehensive results. This collaborative approach often leads to innovative research methods and outcomes that contribute to the advancement of knowledge in the humanities.

Additionally, digital humanities promote the preservation and accessibility of cultural heritage. Digital archives and databases make it possible to store and share historical documents, manuscripts, and artifacts with a global audience. This not only ensures the preservation of these materials but also democratizes access to information, allowing a wider range of people to engage with and benefit from humanities research. Overall, digital humanities represent a transformative approach to humanities scholarship, leveraging the power of digital technology to deepen our understanding of human culture and history. The advent of digital humanities has revolutionized the way scholars analyze and interpret textual data, combining traditional humanities research with digital tools and methodologies (Pratono et al., 2020). In the context of Indonesian language texts, digital humanities approaches offer significant potential for enhancing linguistic, literary, and cultural studies by leveraging computational methods to uncover patterns and insights that would be difficult to discern through manual analysis (Wang, 2018). The rich and diverse corpus of Indonesian language texts, spanning various genres and historical periods, presents a unique opportunity for digital humanities to contribute to a deeper understanding of Indonesian language and culture.

Despite the global advancements in digital humanities, there remains a notable research gap in the application of these methods to Indonesian language texts. Existing studies predominantly focus on Western languages, with limited attention given to non-Western languages such as Indonesian (Bekavac et al., 2021). This gap is significant, considering the vast amount of literary and historical texts available in Indonesian, which are underutilized in digital humanities research. Addressing this gap is essential to ensure the inclusivity and diversity of digital humanities scholarship and to provide a comprehensive understanding of Indonesian textual heritage (Laksmi et al., 2024). The urgency of this research is underscored by the increasing digitization of texts and the growing recognition of the importance of preserving and analyzing cultural heritage in the digital age. The integration of digital humanities approaches in analyzing Indonesian language texts is crucial for fostering cross-cultural understanding and for the advancement of linguistic and literary studies in Indonesia. Additionally, this research can support educational initiatives by providing digital tools and resources that enhance the teaching and learning of the Indonesian language and literature (Zain et al., 2024).

Recent studies have explored various digital humanities approaches to analyzing language texts, including several focusing on Indonesian language texts. These studies provide valuable insights into the application of digital tools and methods in this field. First, (Puspitasari et al., 2024) conducted a study utilizing text mining techniques to analyze Indonesian literary works. They found that text mining could effectively identify recurring themes and patterns within a large corpus of texts, enabling a deeper understanding of cultural and societal influences on literature. Their research demonstrated the potential of digital tools in uncovering insights that traditional methods might overlook. Second, (Solikhah & Budiharso, 2020) explored the use of natural language processing (NLP) in analyzing historical Indonesian texts. Their study revealed that NLP techniques could be used to extract and categorize significant historical events and figures mentioned in the texts.

This approach not only facilitated the organization and analysis of historical data but also highlighted the potential of NLP in historical research. Third, (Novita, 2021) applied digital mapping tools to examine the geographical distribution of linguistic features in Indonesian dialects. Their findings indicated that digital mapping could effectively visualize linguistic diversity across different regions, providing a clearer picture of language variation and evolution. This study underscored the importance of spatial analysis in linguistic research. Fourth, (Dewi et al., 2023) focused on the development of a digital archive for Indonesian manuscripts. Their research showed that digital archiving not only preserved valuable cultural heritage but also made these resources accessible to a broader audience. They emphasized the role of digital archives in democratizing access to historical texts and fostering collaborative research. Fifth, (Wang, 2018) investigated the application of data visualization techniques in the study of Indonesian poetry. They found that visual representations of linguistic patterns and structures could enhance the interpretation and appreciation of poetic works. Their study highlighted the aesthetic and analytical benefits of integrating data visualization into literary analysis.

These studies collectively illustrate the diverse applications and significant impact of digital humanities approaches in the analysis of Indonesian language texts. They demonstrate the potential of digital tools to enhance our understanding of literature, history, linguistics, and cultural heritage, paving the way for future research in this interdisciplinary field. Previous research in digital humanities has demonstrated the efficacy of computational methods in textual analysis, including text mining, natural language processing (NLP), and corpus linguistics (Permatasari et al., 2020). Studies have shown that these methods can uncover hidden patterns, thematic structures, and semantic networks within large text corpora, thereby enriching traditional humanities research. However, most of these studies have been limited to English and other European languages, highlighting the need for similar research on Indonesian texts. This study is novel in its focused application of digital humanities methodologies to Indonesian language texts, aiming to bridge the gap in current research. By employing advanced computational techniques, this research seeks to explore the unique linguistic

and cultural characteristics of Indonesian texts, offering new perspectives and insights that are not accessible through conventional methods. The novelty lies in the interdisciplinary approach, combining digital tools with humanities scholarship to enhance the analysis of Indonesian textual data.

The primary objectives of this research are to develop and implement digital humanities methodologies specifically tailored for analyzing Indonesian language texts. This includes identifying and analyzing linguistic and thematic patterns within a diverse corpus of Indonesian texts using computational tools. The study also aims to enhance the accessibility and usability of Indonesian textual data for researchers, educators, and students, facilitating easier engagement with these resources. Additionally, the research seeks to contribute to the global digital humanities discourse by integrating non-Western languages and texts into the broader research framework, thereby expanding the scope of digital humanities beyond Western-centric perspectives. The benefits of this research are multifaceted. Academically, it will provide new tools and methods for analyzing Indonesian texts, thereby advancing linguistic and literary studies. Culturally, it will contribute to the preservation and dissemination of Indonesian literary heritage in the digital realm. Educationally, it will support the development of digital resources and curricula that enhance the teaching and learning of Indonesian language and literature. Furthermore, this research will promote inclusivity and diversity in digital humanities scholarship by incorporating a non-Western language perspective.

Method

This study employs a qualitative research methodology, specifically utilizing a literature review and case study approach to explore the application of digital humanities techniques in the analysis of Indonesian language texts (Moleong, 2000). The research is structured around specific case studies, focusing on selected Indonesian literary works and digital archives (Taylor et al., 2015). For example, case studies include the analysis of modern Indonesian literature, such as works by Pramoedya Ananta Toer, as well as historical texts from the Indonesian National Library's digital collections. These case studies serve as concrete examples for understanding how digital humanities methodologies can be applied to Indonesian language texts. To conduct the literature review, a systematic search was performed using key academic databases such as JSTOR, Google Scholar, and Project MUSE to identify relevant studies, tools, and methodologies related to digital humanities, linguistics, and Indonesian studies (Creswell & Creswell, 2017). Additionally, digital archives like the Southeast Asian Digital Library and the Indonesian National Library's digital collections were accessed for primary textual data. The selection process involved filtering studies and texts based on their relevance to the research questions and the applicability of digital humanities methods to Indonesian language texts (Mills & Ratcliffe, 2012).

Data analysis involved a combination of thematic analysis and case study analysis. Thematic analysis was used to identify key themes, linguistic patterns, and cultural insights in the Indonesian texts, while the case study analysis provided an in-depth exploration of the specific texts under examination. Computational tools such as Python, NLTK, and Voyant Tools were used for text mining and natural language processing (NLP) to analyze linguistic features and thematic patterns (Bryman, 2016). The validity and reliability of the analysis were ensured through triangulation of data sources, cross-checking findings with existing literature, and applying consistent coding procedures during the thematic analysis (Sugiyono, 2013). By incorporating these methods, this study aims to provide a detailed understanding of the potential and challenges of applying digital humanities approaches to Indonesian language texts, contributing to the broader field of digital humanities and Indonesian studies (Jensen et al., 2022).

Results and Discussion

Below is a table summarizing the findings from recent studies on digital humanities approaches to analyzing Indonesian language texts.

Table 1. Recent Studies on Digital Humanities Approaches to Analyzing Indonesian Language Texts

Authors	Year	Research Focus	Methods Used	Key Findings
(Puspitasari et al., 2024)	2024	Analysis of Indonesian literary works	Text Mining	Identified recurring themes and patterns within a large corpus of texts, revealing cultural and societal influences on literature.
(Solikhah & Budiharso, 2020)	2020	Analysis of historical Indonesian texts	Natural Language Processing (NLP)	Extracted and categorized significant historical events and figures, facilitating the organization and analysis of historical data.

Authors	Year	Research Focus	Methods Used	Key Findings
(Novita, 2021)	2021	Examination of linguistic features in dialects	Digital Mapping	Visualized linguistic diversity across different regions, providing insights into language variation and evolution.
(Dewi et al., 2023)	2023	Development of a digital archive for manuscripts	Digital Archiving	Preserved valuable cultural heritage and made resources accessible to a broader audience, promoting collaborative research.
(Wang, 2018)	2018	Study of Indonesian poetry	Data Visualization	Enhanced interpretation and appreciation of poetic works through visual representations of linguistic patterns and structures.

The table summarizes recent studies employing digital humanities methodologies for analyzing Indonesian language texts. (Puspitasari et al., 2024) utilized text mining techniques to uncover recurring themes in Indonesian literary works, emphasizing the influence of cultural and societal factors. (Solikhah & Budiharso, 2020) employed natural language processing (NLP) to extract and categorize historical events and figures from Indonesian texts, demonstrating the utility of NLP in historical analysis. (Novita, 2021) focused on digital mapping to visualize the geographical distribution of linguistic features across Indonesian dialects, revealing regional linguistic variations. (Dewi et al., 2023) contributed by developing a digital archive for Indonesian manuscripts, enhancing accessibility and preservation of cultural heritage materials. Lastly, (Wang, 2018) applied data visualization techniques to analyze Indonesian poetry, providing insights into linguistic patterns and structures within poetic texts. These studies collectively illustrate the diverse applications of digital humanities in Indonesian language research, highlighting how computational tools and methods can enrich our understanding of literature, history, and cultural heritage. They underscore the interdisciplinary nature of digital humanities, fostering collaboration across fields to advance scholarly inquiry and promote broader access to cultural resources.

Methodological Approaches in Digital Humanities for Indonesian Texts

The application of digital humanities methodologies to Indonesian language texts primarily involves text mining, natural language processing (NLP), and corpus linguistics. Text mining facilitates the extraction of meaningful patterns from large text corpora, enabling researchers to analyze trends, themes, and sentiment within the texts (Saddhono et al., 2023). For instance, the use of word frequency analysis and topic modeling has proven effective in uncovering dominant themes in Indonesian literary works (Nurochim, 2022). These methods allow for a nuanced understanding of language use and cultural narratives that are prevalent in the texts. Natural language processing, a subset of artificial intelligence, further enhances the analysis by enabling the automatic processing and understanding of Indonesian texts. Tools such as Part-of-Speech tagging, named entity recognition, and syntactic parsing are employed to analyze grammatical structures and identify entities within the texts (Solikhah & Budiharso, 2020). These techniques are particularly useful in studying the evolution of language and the influence of socio-political contexts on language use. For example, NLP has been used to track the usage of specific terms across different historical periods, providing insights into how language reflects societal changes (Sabrina et al., 2021).

Corpus linguistics involves the creation and analysis of large, structured text corpora. In the context of Indonesian language texts, corpus-based studies have enabled the systematic study of language patterns, idiomatic expressions, and syntactic structures. Digital corpora such as the SEAlang Library's Indonesian corpus provide valuable resources for linguists and digital humanists, facilitating both qualitative and quantitative analyses. The integration of these methodological approaches enhances the depth and breadth of analysis, allowing for a comprehensive exploration of Indonesian texts. Methodological approaches in digital humanities (DH) for Indonesian texts encompass a diverse range of computational methods and tools tailored to analyze, interpret, and preserve cultural, linguistic, and historical materials from Indonesia. Here's a detailed exploration of these methodological approaches:

Text Mining and Natural Language Processing (NLP)

Text mining involves the automated extraction of useful information from large volumes of textual data. In Indonesian DH research, text mining techniques are applied to analyze literary texts, historical documents, and contemporary writings. Researchers use methods such as topic modeling, sentiment analysis, and named entity recognition to uncover themes, sentiments, and entities within Indonesian texts (Andreotta et al., 2019). Natural Language Processing (NLP) techniques enable the computational analysis of natural language data. NLP algorithms are employed to parse Indonesian texts, extract linguistic features, and derive meaningful insights.

For instance, NLP is used to analyze syntactic structures, semantic relationships, and discourse patterns in Indonesian literature and historical manuscripts, aiding in linguistic and literary analysis.

Digital Mapping and Geographic Information Systems (GIS)

Digital mapping and GIS technologies are utilized to visualize spatial relationships and geographical distributions within Indonesian texts. Researchers map linguistic variations, cultural practices, and historical events across Indonesia's diverse regions. GIS enables spatial analysis that integrates textual data with geographical information, providing insights into how language and culture are influenced by geographic contexts (Rebin, 2024).

Digital Archives and Cultural Heritage Management

Digital archives play a pivotal role in preserving and digitizing Indonesian manuscripts, historical records, and cultural artifacts. These archives facilitate global access to rare and fragile materials, supporting scholarly research and cultural preservation efforts. Digital archiving involves digitization, metadata creation, and online dissemination of cultural heritage resources, ensuring their long-term accessibility and conservation (Ali & Ryan, 2024).

Data Visualization and Computational Analysis

Data visualization techniques are employed to visually represent and interpret patterns within Indonesian texts. Visualization tools generate graphs, charts, and interactive displays that elucidate linguistic structures, thematic trends, and narrative styles in Indonesian literature. Computational analysis techniques, such as stylometry and quantitative text analysis, complement data visualization by quantifying textual features and identifying statistical patterns in Indonesian texts (Ahmad et al., 2022).

Interdisciplinary Collaboration and Methodological Innovation

DH methodologies encourage interdisciplinary collaboration among researchers from humanities, social sciences, computer science, and information technology fields. Collaborative projects integrate diverse expertise to develop innovative approaches for studying Indonesian texts. Methodological innovation in DH includes the adaptation of machine learning algorithms, semantic web technologies, and digital humanities frameworks to advance research capabilities and deepen scholarly insights into Indonesia's linguistic diversity and cultural heritage. In summary, methodological approaches in digital humanities for Indonesian texts leverage computational tools and interdisciplinary collaboration to advance scholarly understanding, preservation, and interpretation of Indonesia's rich cultural and linguistic heritage. These methodologies continue to evolve, offering new avenues for exploring and analyzing Indonesian texts in the digital age.

Application in Digital Humanities for Indonesian Texts

Digital humanities refer to the integration of computational methods and technologies in humanities research to analyze, interpret, and visualize cultural artifacts, historical texts, and linguistic data. In the context of Indonesian texts, digital humanities methodologies have been increasingly applied to enhance scholarly understanding and accessibility. Researchers have employed various digital tools and approaches to study Indonesian language texts. For instance, text mining techniques have been utilized to uncover patterns and themes in Indonesian literature, providing insights into cultural influences and societal dynamics (Bernard & Ryan, 1998). Natural language processing (NLP) has been instrumental in extracting and analyzing historical events and linguistic features from Indonesian texts, facilitating deeper historical and linguistic analyses. Digital mapping technologies have enabled researchers to visualize the geographical distribution of linguistic variations across Indonesian dialects, offering spatial insights into language diversity.

Moreover, the development of digital archives has played a crucial role in preserving and making accessible Indonesian manuscripts and cultural heritage materials (Nugroho & Widiatmojo, 2024). These digital archives not only safeguard historical documents but also democratize access to them, fostering broader scholarly engagement and research collaboration. Furthermore, data visualization techniques have been applied to analyze Indonesian poetry and literary texts, enhancing the interpretation of linguistic patterns and aesthetic qualities (Triana et al., 2020). Overall, digital humanities methodologies have significantly contributed to advancing the study of Indonesian texts by providing innovative tools for analysis, preservation, and interpretation (Lestari, n.d.). These approaches continue to shape interdisciplinary research in the humanities, promoting deeper insights into Indonesian language, culture, and history through digital innovation.

Text Mining and Natural Language Processing (NLP)

Text mining and NLP techniques are extensively used to analyze large corpora of Indonesian texts. These methodologies allow researchers to extract, categorize, and analyze textual data at scale (Piwowar-Sulej et al., 2023). For example, researchers use text mining to identify recurring themes, sentiments, and patterns within Indonesian literature, offering insights into cultural trends, societal issues, and linguistic nuances. NLP is particularly useful for processing and understanding natural human language data. In Indonesian studies, NLP

has been applied to historical texts to extract information about significant events, characters, and linguistic features. This approach aids in historical research by automating the extraction and categorization of data that would be time-consuming or impractical using traditional methods.

Digital Mapping and Spatial Analysis

Digital mapping technologies are employed to visualize the geographic distribution of linguistic features and cultural phenomena across Indonesia. Researchers use geographic information systems (GIS) to map dialectal variations, language usage patterns, and cultural practices. This spatial analysis provides valuable insights into how language and culture intersect with geography, revealing regional variations and historical influences.

Digital Archives and Cultural Heritage Preservation

The development of digital archives plays a crucial role in preserving and disseminating Indonesian manuscripts, historical documents, and cultural artifacts. Digital archives make rare and fragile materials accessible to a wider audience, facilitating scholarly research, educational initiatives, and cultural preservation efforts (Harjanta, 2015). These archives contribute to the conservation of Indonesia's rich cultural heritage while enabling global access and collaboration among researchers.

Data Visualization in Literary and Linguistic Analysis

Data visualization techniques are used to explore and interpret Indonesian literary texts, poetry, and linguistic structures. Visual representations of textual data help researchers identify patterns, trends, and relationships that are not immediately apparent from raw text (Lin, 2013). This approach enhances the understanding of linguistic styles, narrative structures, and aesthetic qualities in Indonesian literature.

Interdisciplinary Collaboration and Innovation

Digital humanities in Indonesian studies foster interdisciplinary collaboration among scholars from linguistics, literature, history, computer science, and cultural studies (Rohana et al., 2017). By integrating computational tools with humanities research, DH encourages innovative approaches to studying Indonesia's diverse linguistic landscape, cultural heritage, and historical narratives. In conclusion, the application of digital humanities for Indonesian texts represents a dynamic field that leverages technology to advance scholarly inquiry, cultural preservation, and interdisciplinary collaboration. These methodologies continue to evolve, offering new opportunities for deepening our understanding of Indonesia's linguistic diversity, cultural heritage, and historical development.

Challenges in Applying Digital Humanities to Indonesian Texts

Despite the potential of digital humanities approaches, several challenges hinder their application to Indonesian language texts. One major challenge is the lack of comprehensive digital text corpora. While there are some digital archives, the coverage and accessibility of these resources remain limited compared to those available for Western languages. This limitation hampers the ability to conduct extensive and representative analyses, as researchers often must rely on fragmented or incomplete datasets (Fakhriati et al., 2022). Another challenge is the complexity of the Indonesian language itself. With its diverse linguistic influences and regional variations, Indonesian presents unique challenges for NLP tools, which are often developed for more standardized languages (Djamdjuri et al., 2021).

This complexity necessitates the customization of NLP tools to accurately process and analyze Indonesian texts, a task that requires significant computational resources and linguistic expertise (Jensen et al., 2022). Moreover, the lack of annotated datasets for training machine learning models further complicates the development of effective NLP applications for Indonesian (Ide, 2004). There is also a shortage of interdisciplinary collaboration between linguists, computer scientists, and digital humanists in Indonesia. This gap results in a lack of integrated approaches that combine linguistic insights with advanced computational techniques (Aisyah & Nuraeni, 2022). Promoting collaborative efforts and fostering cross-disciplinary training can enhance the development and application of digital humanities methodologies in Indonesian contexts. Additionally, addressing the digital divide and improving access to technological resources are essential steps toward overcoming these challenges.

Benefits of Digital Humanities for Indonesian Texts

The integration of digital humanities methodologies offers numerous benefits for the analysis of Indonesian language texts. One significant advantage is the ability to conduct large-scale analyses that would be impractical through manual methods. By leveraging computational tools, researchers can analyze vast amounts of text data, uncovering patterns and trends that enhance our understanding of linguistic and cultural dynamics. This large-scale analysis is particularly valuable in studying historical texts, where digital methods can reveal shifts in language use and thematic focus over time (Ide, 2004). Digital humanities also facilitate the preservation and accessibility of Indonesian textual heritage. Digital archives and databases ensure that valuable texts are preserved in digital formats, protecting them from physical degradation and loss. These digital resources are

accessible to a global audience, promoting the study of Indonesian literature and culture beyond national boundaries. Furthermore, digital tools enable the creation of interactive and multimedia representations of texts, enriching the user experience and providing new avenues for engagement with Indonesian language materials. Another benefit is the enhancement of educational resources. Digital humanities methodologies can support the development of digital tools and platforms for language learning and literary studies. For example, annotated corpora and interactive text analysis tools can be integrated into educational curricula, providing students with hands-on experience in digital text analysis. These resources can also support the teaching of Indonesian language and literature in higher education, promoting a deeper and more interactive learning experience.

Future Directions and Applications

The future of digital humanities in the context of Indonesian language texts holds promising potential for further advancements and applications. One potential direction is the development of more sophisticated NLP tools tailored to the nuances of the Indonesian language. These tools can improve the accuracy and depth of text analysis, enabling researchers to explore complex linguistic and cultural phenomena (Dewi et al., 2023). Collaborative efforts between linguists and computer scientists will be crucial in developing these tools and ensuring their effectiveness. Another area of future development is the expansion of digital corpora and archives. Efforts to digitize and curate more Indonesian texts will provide a richer and more comprehensive dataset for analysis. Partnerships with libraries, cultural institutions, and academic organizations can facilitate the creation and maintenance of these digital resources, ensuring their sustainability and accessibility. Additionally, initiatives to standardize metadata and annotations across digital corpora will enhance the interoperability and usability of these resources. There is also potential for integrating digital humanities approaches with other disciplines, such as history, sociology, and anthropology. By combining textual analysis with contextual and cultural studies, researchers can gain a more holistic understanding of Indonesian texts and their socio-cultural significance. This interdisciplinary approach can lead to new insights and perspectives, enriching both digital humanities scholarship and the study of Indonesian language and culture.

Conclusions

The integration of digital humanities methodologies in the analysis of Indonesian language texts has shown significant potential to advance the field of humanities research. This study highlights the ability of computational tools, such as text mining, natural language processing (NLP), and corpus linguistics, to uncover linguistic patterns, cultural narratives, and historical shifts in language use. Through case studies, the research demonstrated how these methods could reveal deeper insights into Indonesian texts, offering a more comprehensive understanding that was previously difficult to achieve with traditional approaches. The use of digital archives also contributes to the preservation and increased accessibility of Indonesian texts, promoting their global scholarly engagement. However, this study also revealed several challenges. Key limitations include the current scarcity of digital corpora for Indonesian texts, as well as the linguistic complexities unique to the Indonesian language that hinder the full effectiveness of existing NLP tools. These obstacles highlight the need for more tailored, language-specific computational resources. Furthermore, the research underscores the importance of interdisciplinary collaboration between linguists, computer scientists, and digital humanists to develop more effective methodologies for analyzing Indonesian texts.

In terms of practical implications, the findings suggest that expanding digital corpora, improving NLP tools specific to Indonesian, and fostering cross-disciplinary partnerships are critical steps toward advancing the application of digital humanities in Indonesian studies. The study recommends that future research focus on developing advanced, language-specific processing tools and building a more extensive digital archive of Indonesian texts. Addressing these challenges will not only enhance the digital humanities research landscape in Indonesia but also contribute to the broader global discourse on non-Western language texts. Ultimately, while the study offers valuable insights, it also calls for further exploration of technical, linguistic, and methodological challenges to ensure the continued evolution of digital humanities in Indonesia and beyond.

References

- Ahmad, M., Junus, K., & Santoso, H. B. (2022). Automatic content analysis of asynchronous discussion forum transcripts: A systematic literature review. *Education and Information Technologies*, 27(8), 11355–11410.
- Aisyah, S., & Nuraeni, N. (2022). The Use of A Digital Multimodal Textbook to Foster Students' English Proficiency. *Jo-ELT (Journal of English Language Teaching) Fakultas Pendidikan Bahasa & Seni Prodi Pendidikan Bahasa Inggris IKIP*, 9(2), 160–171.
- Ali, I., & Ryan, S. E. (2024). The law and policy of library development: A legal text analysis of Indonesian library bills. *Alexandria*, 09557490241236949.

- Andreotta, M., Nugroho, R., Hurlstone, M. J., Boschetti, F., Farrell, S., Walker, I., & Paris, C. (2019). Analyzing social media data: A mixed-methods framework combining computational and qualitative text analysis. *Behavior Research Methods*, 51, 1766–1781.
- Bekavac, B., Kocijan, K., Silberstein, M., & Šojat, K. (2021). *Formalising Natural Languages: Applications to Natural Language Processing and Digital Humanities: 14th International Conference, NooJ 2020, Zagreb, Croatia, June 5–7, 2020, Revised Selected Papers* (Vol. 1389). Springer Nature.
- Bernard, H. R., & Ryan, G. (1998). Text analysis. *Handbook of Methods in Cultural Anthropology*, 613.
- Bryman, A. (2016). *Social research methods*. Oxford university press.
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.
- Dewi, H. H., Hidayatulloh, S. M. M., Sukarno, S., Lestari, A. E., Dewi, I. L., & Ciptaningrum, D. S. (2023). English materials development for an undergraduate communication study program: A need analysis in Indonesian context. *LLT Journal: A Journal on Language and Language Teaching*, 26(1), 69–91.
- Djandjuri, D. S., Suseno, M., Tajuddin, S., Lustyantje, N., & Chaeruman, U. A. (2021). Multimodal Approach in Online EFL Class Using Islamic Learning Materials: Students' Perspective. *International Journal of Language Education*, 5(4), 337–355.
- Fakhriati, F., Mu'jizah, M., Holil, M., & Permadi, T. (2022). Don't Leave Indonesian Manuscripts in Danger: An Analysis of Digitalization and Preservation. *Preservation, Digital Technology & Culture*, 51(1), 3–15.
- Harjanta, A. T. J. (2015). Preprocessing text untuk meminimalisir kata yang tidak berarti dalam proses text mining. *Jurnal Informatika Upgris*, 1(1 Juni).
- Ide, N. (2004). Preparation and analysis of linguistic corpora. *A Companion to Digital Humanities*, 289–305.
- Jensen, J. L., Karell, D., Tanigawa-Lau, C., Habash, N., Oudah, M., & Fairus Shofia Fani, D. (2022). Language models in sociological research: An application to classifying large administrative data and measuring religiosity. *Sociological Methodology*, 52(1), 30–52.
- Laksmi, L., Suhendra, M. F., Shuhidan, S. M., & Umanto, U. (2024). The readiness to implement digital humanities data curation of four institutional repositories in Indonesia. *Digital Library Perspectives*, 40(1), 80–95.
- Lestari, D. E. (n.d.). Teachers And Students'net Promoter Score (Nps) On The Cambridge Learning Management System (CLMS). *Journal of English Teaching, Applied Linguistics and Literatures (JETALL)*, 4(2), 141–148.
- Lin, A. (2013). Classroom code-switching: Three decades of research. *Applied Linguistics Review*, 4(1), 195–218.
- Mills, D., & Ratcliffe, R. (2012). After method? Ethnography in the knowledge economy. *Qualitative Research*, 12(2), 147–164.
- Moleong, L. J. (2000). *Qualitative Research Methodology*, Bandung: PT. Youth Rosdakarya.
- Novita, D. (2021). *Integrating technology and humanity into language teaching book 1: English language teaching*. Deepublish.
- Nugroho, R. S., & Widiatmojo, R. (2024). Indonesian Presidential Vision and Mission 2024: A Computer-Assisted Textual Analysis. *Jurnal Partisipatoris*, 6(1).
- Nurochim, N. (2022). Analysis of text books for learning Islamic education based on digital media in Indonesia. *Jurnal Konseling Dan Pendidikan*, 10(2), 208–217.
- Permatasari, P. A., Rech, M., Qohar, A. A., & Faizal, A. (2020). From web 1.0 to web 4.0: the digital heritage platforms for UNESCO's heritage properties in Indonesia. *Virtual Archaeology Review*, 11(23), 75–93.
- Piwowar-Sulej, K., Wawak, S., Tyrańska, M., Zakrzewska, M., Jarosz, S., & Sołtysik, M. (2023). Research trends in human resource management. A text-mining-based literature review. *International Journal of Manpower*, 44(1), 176–196.
- Pratono, A. H., Prima, D. A., Sinaga, N. F. N. T., Permatasari, A., Ariani, M., & Han, L. (2020). Crowdfunding in digital humanities: some evidence from Indonesian social enterprises. *Aslib Journal of Information Management*, 72(2), 287–303.
- Puspitasari, D. A., Fakhurroja, H., & Sutrisno, A. (2024). Authorship Analysis In Electronic Texts Using Similarity Comparison Method. *Linguistik Indonesia*, 42(1), 91–112.
- Rebin, B. (2024). *Paving Future Pathway for Disconnected Voices to Unbalanced Digital World: An analysis of multi-stakeholder perspective on improving the digital support for digitally-disadvantaged languages*.
- Rohana, Y., Santosa, R., & Djatmika, D. (2017). Gaya Bahasa, Teknik Penerjemahan, Dan Kualitas Terjemahan Dalam Dongeng Disney Dwibahasa Berjudul Cinderella: My Bedtime Story Dan Tinkerbell And The Great Fairy Rescue. *PRASASTI: Journal of Linguistics*, 2(1), 150–166.
- Sabrina, A., Siregar, I., & Sosrohadi, S. (2021). Lingual dominance and symbolic power in the discourse of using the PeduliLindungi application as a digital payment tool. *International Journal of Linguistics Studies*, 1(2), 52–59.

- Saddhono, K., Rohmadi, M., Setiawan, B., Suhita, R., Rakhmawati, A., Hastuti, S., & Islahuddin, I. (2023). Corpus Linguistics Use in Vocabulary Teaching Principle and Technique Application: A Study of Indonesian Language for Foreign Speakers. *International Journal of Society, Culture & Language*, 11(1), 231–245.
- Solikhah, I., & Budiharso, T. (2020). Exploring cultural inclusion in the curriculum and practices for teaching bahasa indonesia to speakers of other languages. *Journal of Social Studies Education Research*, 11(3), 177–197.
- Sugiyono, D. (2013). *Educational research methods approach quantitative, qualitative and R&D*. Bandung: Alfabeta.
- Taylor, S. J., Bogdan, R., & DeVault, M. (2015). *Introduction to qualitative research methods: A guidebook and resource*. John Wiley & Sons.
- Triana, H. W., Wirman, E. P., Kustati, M., Refinaldi, R., Rahmi, A., & Nelmawarni, N. (2020). Social practice on Facebook: Critical discourse analysis in the process of text production. *Studies in English Language and Education*, 7(1), 1–21.
- Wang, Q. (2018). Distribution features and intellectual structures of digital humanities: A bibliometric analysis. *Journal of Documentation*, 74(1), 223–246.
- Zain, L., Anwar, A., Hadna, I. N., & Hannum, I. N. A. (2024). The Implementation of Digital Humanities in the State Islamic University Libraries in Indonesia. *Khazanah Al-Hikmah: Jurnal Ilmu Perpustakaan, Informasi, Dan Kearsipan*, 12(1)