Human Interaction and Emerging Technologies (IHIET-AI 2025), Vol. 161, 2025, 342–348
https://doi.org/10.54941/ahfe1005926

An AI-Driven Ukrainian History Web Platform

Valentyna Kolomiets[1], Pedro F. Oliveira[2], and Paulo Matos[2]

[1]Instituto Politécnico de Bragança, Bragança, Portugal
[2]Research Centre in Digitalization and Intelligent Robotics (CeDRI), Laboratório Associado para a Sustentabilidade e Tecnologia em Regiões de Montanha (SusTEC), Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300–253 Bragança, Portugal

ABSTRACT

The AI-driven Ukrainian History web platform offers an innovative way for users to engage with the nation's rich history. By integrating artificial intelligence, natural language processing (NLP), and geospatial analysis, it presents historical events, significant locations, and notable figures in an interactive and visually engaging format. The platform systematically gathers historical data using tools like Scrapy for web scraping and Tesseract OCR for digitizing scanned documents. While noisy or degraded documents may affect accuracy, the availability of high-quality sources ensures reliable data extraction. Fine-tuned NLP models, including transformers like BERT and RoBERTa, process the data to identify and categorize key entities such as dates, locations, and names of historical figures. Contextual summarization ensures the extracted information is both accurate and easy to understand. Geospatial data is managed with PostGIS, an extension of PostgreSQL, and visualized using Leaflet.js. An interactive map interface enables users to explore events by location and time period, with filters for categories like political milestones or cultural events. The backend, built on PostgreSQL, ensures scalability and performance, while development in Visual Studio Code streamlined integration across components. This platform not only preserves Ukraine's cultural heritage but also demonstrates the potential of modern technology to transform historical education, offering an intuitive way to connect with the past and explore its influence on Ukraine's landscape and culture.

Keywords: AI-driven web scraping, Interactive map, Named entity recognition (NER)

INTRODUCTION

Ukraine's history is a tapestry of profound events, cultural milestones, and influential figures that have shaped not only the nation itself but also the surrounding regions. Preserving this rich heritage is essential, not only for academic purposes but also for fostering a deeper connection to the past. This project is driven by the desire to make Ukraine's historical narrative accessible to everyone, from researchers to curious learners, through a platform that combines modern technology with the timeless value of storytelling.

Historical records often exist in scattered and inaccessible forms—archived manuscripts, digitized collections, and academic publications that are difficult to navigate without specialized knowledge. This platform addresses

that challenge by transforming these diverse materials into a structured, interactive format that brings history to life. By using advanced tools to collect, digitize, and process historical data, the platform makes it possible to uncover stories of key events, important locations, and notable figures that might otherwise remain hidden.

At the core of the platform is an interactive map that allows users to explore Ukraine's history geographically and chronologically. Whether someone is tracing the cultural contributions of a specific region or learning about significant historical conflicts, the map offers an engaging way to visualize and contextualize the past. The ability to filter events by themes or time periods provides users with a personalized and immersive experience that is both educational and intuitive.

More than just a tool for archiving historical data, this platform is a bridge between the past and the present. It reimagines how history can be explored in the digital age, making it engaging, accessible, and relevant. By blending advanced technologies with a user-centered approach, this project highlights the enduring importance of Ukraine's history, ensuring it is preserved and appreciated for generations to come.

Overview of the Project

This project focuses on preserving and sharing Ukraine's rich and diverse history by leveraging modern technologies to create an AI-driven platform. The platform integrates data processing, geospatial visualization, and user-centric design to make historical events, locations, and figures accessible and engaging. By combining Optical Character Recognition (OCR) for digitizing archival documents, Natural Language Processing (NLP) for extracting and summarizing information, and interactive map technology for geospatial exploration, the platform bridges the gap between historical scholarship and public understanding. This innovative approach ensures that Ukraine's cultural heritage is preserved digitally while offering users a dynamic and intuitive way to explore history through space and time.

Objectives

The primary objective of this project is to develop an advanced AI-driven platform that showcases the potential of modern technologies in processing, organizing, and visualizing complex datasets. By integrating Optical Character Recognition, Natural Language Processing, and geospatial analysis, the project demonstrates how fragmented and unstructured data can be transformed into an accessible and interactive digital resource. A key technical goal is to build a scalable system capable of extracting, summarizing, and linking historical information to geospatial and chronological contexts, providing a dynamic and intuitive user experience. The platform highlights the innovative application of AI models for Named Entity Recognition and contextual summarization, combined with advanced database technologies like PostgreSQL with PostGIS for geospatial data handling. While the historical heritage of Ukraine serves as the focal dataset, the project's emphasis is on designing a robust technological framework that

can be adapted to other domains or datasets. Ultimately, this platform not only preserves Ukraine's history but also demonstrates how cutting-edge tools can redefine data accessibility and user interaction in the digital age.

Technologies

This project integrates advanced tools and frameworks to process, visualize, and present historical data efficiently. To digitize historical documents, Optical Character Recognition (OCR) is employed, with Tesseract OCR being a key tool for converting scanned images into machine-readable text. Preprocessing techniques like noise reduction and skew correction improve accuracy, while cloud-based solutions, such as Google Cloud Vision OCR, provide enhanced capabilities for more complex or degraded documents.

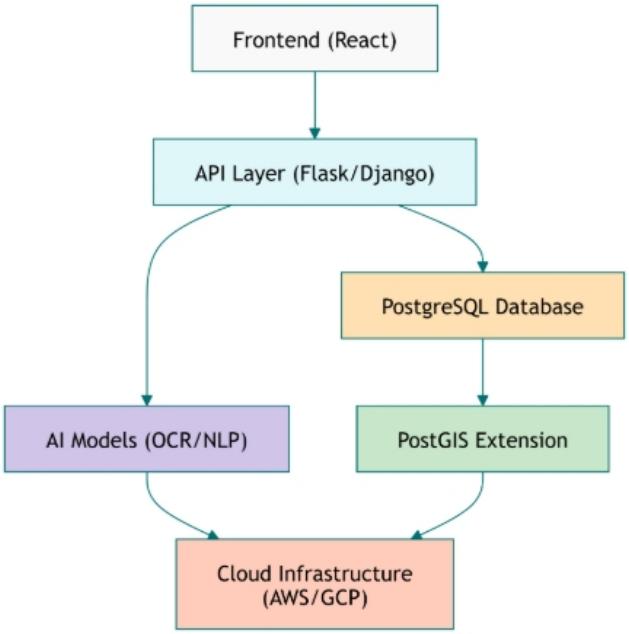

Figure 1: Backend architecture of the platform.

Natural Language Processing (NLP) plays a critical role in extracting meaningful information from unstructured text. Fine-tuned transformer models, such as BERT or RoBERTa, identify dates, locations, and historical figures through Named Entity Recognition (NER). Summarization models like GPT-based systems or T5 create concise descriptions of events, making data accessible to a broader audience. These tools are adapted to handle historical datasets for improved domain-specific performance.

The platform uses geospatial technologies to visualize history spatially. PostGIS, an extension of PostgreSQL, enables the storage and manipulation of geospatial data, such as event locations or regional boundaries. This data

is displayed interactively using Leaflet.js, a lightweight JavaScript library for creating mobile-friendly maps. The integration allows users to explore history geographically and chronologically.

Data collection involves scraping online archives and digitized resources using tools like Scrapy and BeautifulSoup. These datasets are integrated into a cohesive structure within PostgreSQL, enabling efficient storage and retrieval. The frontend, built with modern JavaScript frameworks like React, offers an intuitive and responsive interface, while Leaflet.js powers dynamic map interactions. The backend, developed using Python frameworks such as Flask or Django, connects the database, AI models, and frontend to ensure smooth data handling.

Approach

The process is designed to integrate advanced technologies seamlessly while addressing the challenges associated with processing and visualizing historical data. This approach focuses on transforming fragmented and unstructured information into an engaging and accessible resource.

The first step is the collection and digitization of data from diverse sources, including online repositories, digitized manuscripts, and academic publications. Web scraping tools, such as Scrapy, are used to automate the extraction of data from websites and archives. For scanned documents, Optical Character Recognition (OCR) technologies, including Tesseract OCR and Google Cloud Vision, are employed to convert images into machine-readable text. Preprocessing techniques, such as noise reduction and skew correction, are applied to enhance the accuracy of the extracted text, particularly for older or degraded documents.

Once the raw data is collected, the next phase involves data processing and structuring using Natural Language Processing (NLP) techniques. Transformer-based models like BERT or RoBERTa are fine-tuned for Named Entity Recognition (NER) to identify key elements such as dates, locations, and figures. Contextual summarization models, such as GPT-based systems, are employed to generate concise and user-friendly descriptions of historical events. This processed data is then organized into a structured format and stored in a relational database using PostgreSQL. Geospatial data, such as event locations, is further enhanced with geocoding APIs to generate precise geographic coordinates.

The platform's geospatial visualization capabilities are implemented using PostGIS, a geospatial extension of PostgreSQL, to manage spatial data efficiently. This data is integrated into an interactive map interface powered by Leaflet.js, enabling users to explore historical events by location and time. The map interface includes features such as dynamic filtering by themes, time periods, or regions, providing a personalized and immersive user experience.

The backend of the platform is developed using Python frameworks like Flask or Django, ensuring seamless communication between the database, AI models, and frontend interface. APIs are implemented to handle user requests and deliver real-time data updates to the frontend. The frontend is built with

React, offering a responsive and intuitive interface designed to meet the needs of both researchers and casual users.

This proposed approach integrates advanced technologies in a cohesive workflow, ensuring that the platform meets its objectives of preserving Ukraine's historical heritage while showcasing the potential of modern AI and geospatial tools in digital preservation and education.

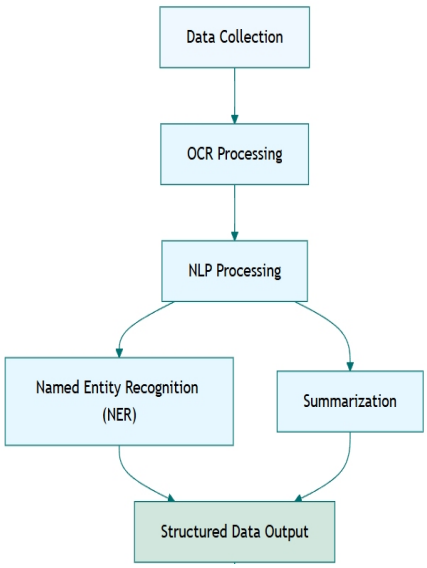

Figure 2: AI workflow for processing historical data.

RESULTS

The development of this project provided valuable insights into the application of AI and geospatial technologies for processing and visualizing historical data. The integration of Optical Character Recognition and Natural Language Processing models proved effective in digitizing and extracting meaningful information from fragmented and unstructured historical records. The process highlighted key challenges, such as handling degraded or diverse datasets, which underscored the importance of robust preprocessing techniques and fine-tuned models.

Through the creation of the platform, we demonstrated how geospatial visualization can contextualize historical events both geographically and chronologically, offering users a deeper understanding of the interconnectedness of historical data. This highlighted the value of mapping technologies in enhancing accessibility and engagement with complex datasets. Additionally, the project emphasized the importance of scalability and system design, ensuring that AI-driven workflows and geospatial tools could be seamlessly integrated and accessed in real time.

Overall, the platform serves as a proof of concept for how advanced technologies can address challenges in data accessibility and engagement,

while also providing a flexible framework that could be adapted to other domains.

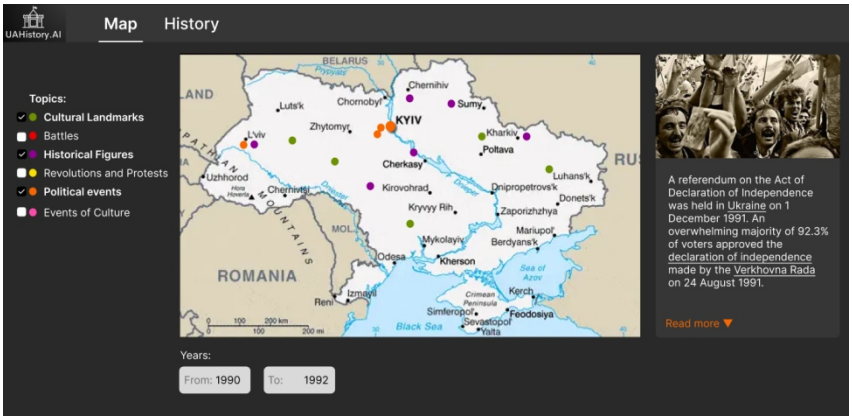

Figure 3: Main page mock up.

CONCLUSION

The development of an AI-driven platform for preserving and exploring Ukraine's historical heritage demonstrates how modern technologies can transform the way historical data is accessed and understood. This project has integrated a powerful combination of Optical Character Recognition, Natural Language Processing, and geospatial visualization to address challenges in digitizing and organizing fragmented historical records. The platform enables users to interact with Ukraine's history through an intuitive map interface, offering a unique way to explore events, figures, and locations geographically and chronologically.

ACKNOWLEDGMENTS

This work was supported by national funds through FCT/ MCTES (PIDDAC): CeDRI, UIDB/05757/2020 (doi: 10.54499/UIDB/05757/2020) and UIDP/05757/2020 (doi: 10.54499/UIDP/05757/2020); and SusTEC, LA/P/0007/2020 (doi: 10.54499/LA/P/0007/2020).

REFERENCES

Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media.

Chang, K.-T., & Tang, G. (2014). Introduction to Geographical Information Systems (8th Edition). McGraw-Hill Education.

Dean, J., & Ghemawat, S. (2004). "MapReduce: Simplified Data Processing on Large Clusters." Proceedings of the 6th Symposium on Operating System Design and Implementation, San Francisco, CA.

Friedl, J. (2006). Mastering Regular Expressions (3rd Edition). O'Reilly Media.

Honnibal, M., & Montani, I. (2020). spaCy 3: Industrial-strength Natural Language Processing in Python. Explosion AI.
PostgreSQL Global Development Group. (2024). PostgreSQL Documentation. Retrieved from https://www.postgresql.org/docs/.
PostGIS Project. (2024). PostGIS Documentation. Retrieved from https://postgis.net/documentation/.
Smith, R. (2007). "An Overview of the Tesseract OCR Engine." Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR).
Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., & Rush, A. M. (2020). "Transformers: State-of-the-Art Natural Language Processing." Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations.
Zandbergen, P. A. (2020). Python Scripting for ArcGIS Pro. Esri Press.