

УДК 930.85:004.8(477)

DOI <https://doi.org/10.24919/2308-4863/85-3-4>

**Олександр ЯКУБА,**

[orcid.org/0009-0002-8546-7576](https://orcid.org/0009-0002-8546-7576)

аспірант кафедри історії України

Харківського національного університету імені В. Н. Каразіна

(Харків, Україна) [oleksandr.yakuba@student.karazin.ua](mailto:oleksandr.yakuba@student.karazin.ua)

## ШТУЧНИЙ ІНТЕЛЕКТ У ДОСЛІДЖЕННІ ТОПОНІМІВ: ПЕРСПЕКТИВИ ТА ПОРІВНЯННЯ З ТРАДИЦІЙНИМИ МЕТОДАМИ

У статті досліджується потенціал застосування штучного інтелекту (ШІ) для вивчення топонімічних змін, зокрема в контексті аналізу географічних назв. Традиційно вивчення топонімів відбувається за допомогою баз даних, архівних документів та історичних карт. Такий підхід дозволяє структурувати дані та зберігати історичні відомості, однак він потребує великих ресурсів, часу на ручну обробку та обмежений у своїй здатності виявляти глибші закономірності. У зв'язку з цим, застосування штучного інтелекту, включаючи алгоритми обробки природної мови (NLP), геокодування, глибокого картування (deep mapping) та машинного навчання, відкриває нові можливості для автоматизованого аналізу великих масивів інформації. Ці технології дозволяють швидко обробляти значні обсяги даних, виявляти приховані зв'язки та закономірності у зміні географічних назв, що значно знижує витрати часу та підвищує ефективність досліджень.

У статті також розглядаються переваги і виклики застосування ШІ в топонімічних дослідженнях, зокрема питання необхідності великих обсягів навчальних даних, труднощі формування моделей для роботи з історичними джерелами, а також адаптації алгоритмів до специфіки топонімічної інформації. Показано, що інтеграція традиційних методів з методами штучного інтелекту відкриває нові горизонти для досліджень топонімічних змін, зокрема дозволяє ефективно обробляти великі архівні текстові та картографічні дані, що раніше були недоступні для масового аналізу.

Аналізуються також критерії вибору між традиційними методами та ШІ, що залежать від специфіки дослідження, масштабів роботи та доступних ресурсів. Стаття наголошує на важливості комплексного підходу в дослідженнях топонімічних змін, де бази даних служать основою для навчання моделей, а ШІ допомагає відкривати нові зв'язки, тенденції та прогнози змін топонімічного ландшафту. У підсумку зроблено висновок про перспективність і необхідність інтеграції технологій штучного інтелекту в дослідження історичних змін топонімів, зокрема для аналізу великих даних та прогнозування майбутніх тенденцій у сфері перейменувань.

**Ключові слова:** топоніміка, штучний інтелект, обробка природної мови, геокодування, глибоке картування, бази даних, топонімічні зміни.

**Oleksandr YAKUBA,**

[orcid.org/0009-0002-8546-7576](https://orcid.org/0009-0002-8546-7576)

Postgraduate student at the Department of History of Ukraine

V. N. Karazin Kharkiv National University

(Kharkiv, Ukraine) [oleksandr.yakuba@student.karazin.ua](mailto:oleksandr.yakuba@student.karazin.ua)

## ARTIFICIAL INTELLIGENCE IN TOPONYMIC RESEARCH: PROSPECTS AND COMPARISON WITH TRADITIONAL METHODS

The article explores the potential of artificial intelligence (AI) for studying toponymic changes, particularly in the context of analyzing geographic names. Traditionally, toponymic studies rely on databases, archival documents, and historical maps. This approach helps structure data and preserve historical information, but it requires significant resources, manual processing time, and is limited in its ability to reveal deeper patterns. Therefore, the application of AI, including natural language processing (NLP) algorithms, geocoding, deep mapping, and machine learning, opens new opportunities for automating the analysis of large datasets. These technologies enable the rapid processing of large volumes of data, uncovering hidden links and patterns in the transformation of geographic names, thus reducing time costs and increasing research efficiency.

The article also discusses the advantages and challenges of using AI in toponymic studies, particularly the need for large volumes of training data, the challenges of creating models for working with historical sources, and adapting algorithms to the specifics of toponymic information. It shows that integrating traditional methods with AI opens new possibilities for studying toponymic changes, allowing researchers to efficiently process large archival text and cartographic data that were previously inaccessible for mass analysis.

The article also analyzes the criteria for choosing between traditional methods and AI, depending on the specifics of the research, the scale of the work, and available resources. It emphasizes the importance of a comprehensive approach in studying toponymic changes, where databases serve as the foundation for training models, and AI helps uncover new connections, trends, and forecasts in the transformation of the toponymic landscape. In conclusion, the article stresses the potential and necessity of integrating AI technologies into historical toponymic studies, particularly for analyzing large data sets and predicting future trends in renaming.

**Key words:** toponymy, artificial intelligence, natural language processing, geocoding, deep mapping, databases, toponymic changes.

**Постановка проблеми.** Топоніміка як наукова дисципліна відіграє ключову роль у вивченні історичних, географічних та культурних процесів, оскільки дозволяє простежити еволюцію назв географічних об'єктів та їх зв'язок із соціальними трансформаціями. Традиційні методи дослідження топонімів базуються на аналізі архівних документів, історичних карт та текстових джерел, однак вони мають обмеження щодо швидкості та ефективності обробки великих масивів даних. Впровадження технологій штучного інтелекту (ШІ) відкриває нові можливості для автоматизації процесу аналізу топонімічних змін, що дозволяє швидко та точно визначати закономірності у зміні географічних назв. Проте застосування ШІ у топоніміці супроводжується низкою викликів, серед яких необхідність великих обсягів навчальних даних, адаптація алгоритмів до історичних контекстів та проблема амбівалентності топонімів.

**Аналіз досліджень.** Використання цифрових технологій та штучного інтелекту у вивченні топонімічних змін є відносно новою сферою досліджень, яка поєднує методи обробки природної мови (NLP), машинного навчання, глибокого картування (deep mapping) та геокодування. Дослідження західних науковців, таких як Бодехаммер (2015) та Ху і Адамс (2020), показують, що автоматизований аналіз текстових джерел може значно полегшити виявлення змін у географічних назвах та сприяти створенню інтерактивних картографічних систем. Вей, Лі та Чжоу (2024) представили підхід до розпізнавання топонімічних сутностей за допомогою попередньо натренованих моделей, що дозволяє значно підвищити точність визначення місцеположень у текстах. Водночас деякі дослідники, зокрема Монкла (2014) та Гровер (2010), акцентують увагу на проблемах геокодування історичних топонімів, оскільки багато назв змінилися або більше не використовуються в сучасних адміністративних межах.

**Метою** цієї статті є аналіз перспектив та викликів застосування штучного інтелекту у дослідженні топонімічних змін, а також порівняння ефективності традиційних методів із сучасними алгоритмічними підходами. Особлива увага приділяється інтеграції баз даних із моделями машинного навчання та можливостям автоматизованого аналізу архівних текстів та картографічних матеріалів. Стаття також розглядає потенціал використання глибокого картування для реконструкції історичних змін топонімічного ландшафту та визначає ключові критерії вибору між традиційними та ШІ-методами залежно від масштабу дослідження.

**Виклад основного матеріалу.** Топоніміка як наукова дисципліна має велике значення для історичної географії, культурології та лінгвістики, оскільки досліджує походження, еволюцію та значення географічних назв. Вивчення топонімів допомагає зрозуміти історичні процеси, міграційні шляхи, зміни політичних кордонів та культурні трансформації регіонів. Актуальність дослідження топонімів зростає, оскільки вони є важливими свідченнями як матеріальної, так і нематеріальної культурної спадщини.

Традиційно вивчення топонімів базувалося на архівних джерелах, польових дослідженнях та картографічних матеріалах. Ці методи, хоча і дають глибокі результати, мають обмеження в масштабах досліджень та швидкості обробки інформації, особливо коли йдеться про великі масиви даних. З появою цифрових технологій та архівів зростає потреба в автоматизації аналізу, щоб обробляти тисячі документів за короткий час.

Сучасні підходи, засновані на штучному інтелекті (AI) та машинному навчанні (ML), відкривають нові можливості для автоматизованого аналізу топонімів у великих масивах історичних текстів і картографічних матеріалів. Технології, такі як текстовий майнінг, deep mapping і геокодування, дозволяють виявляти зміни в назвах місць, автоматично ідентифікувати їх у текстах і навіть відновлювати їхнє місце на картах.

Основні проблеми дослідження топонімів

Топоніміка це розділ ономастики, який вивчає топоніми (грец. τόπος «місце; місцевість, область, країна, простір» й ὄνομα, ὄνυμα «ім'я, назва»), що об'єднують власні назви будь-яких географічних об'єктів. Топоніміка є міждисциплінарною галуззю, яка досліджує походження, етимологію, еволюцію та значення топонімів. Вона поєднує в собі елементи лінгвістики, географії, історії, культурології та навіть соціології, оскільки вивчає, як географічні назви відображають соціальні, політичні та культурні зміни в суспільстві. Вивчення топонімів дозволяє реконструювати історичну картину певної місцевості, зрозуміти міграційні процеси, економічні зв'язки, політичні впливи та культурні взаємодії між народами.

Назви міст, річок, гір та інших географічних об'єктів часто зберігають інформацію про історичні події, персоналії, релігійні та міфологічні мотиви. Наприклад, багато топонімів походять від імен великих історичних постатей, легендарних героїв чи місцевих божеств, а також можуть відображати природні особливості місцевості, такі як рельєф, флора і фауна. Таким чином, топоніми є своєрідними «історичними маркерами», які дозво-

ляють дослідникам краще зрозуміти процеси, що вплинули на формування ідентичностей населення регіону.

Історичні дослідження топонімів можуть виявити значні культурні трансформації, що відбувалися в минулому. Наприклад, зміни в назвах населених пунктів часто були пов'язані з політичними подіями, такими як зміна державної влади, колонізація або адміністративні реформи. У деяких випадках назви змінювалися через політичну ідеологію або як маркер імперської присутності на цих територіях – перейменування у радянський час та під час окупації українських міст Росією є яскравим прикладом такої тенденції. Тому вивчення топонімів надає цінну інформацію про те, як змінювались культурні та політичні орієнтири в різні періоди історії.

Один із викликів, з яким стикаються топонімісти, – це багатозначність та варіативність назв. Часто одна й та сама назва може стосуватися різних місць, наприклад, назви типу «Вишневе» чи «Степове» (Перелік сіл та селищ, 2018) можуть зустрічатися десятки разів на території одного регіону, і для точного розуміння, про яке поселення йдеться, потрібно використовувати контекст та додаткові історичні джерела. Також назва географічного об'єкта може змінювати свою форму в залежності від мови, на якій вона вимовляється. Це явище, відоме як топонімічна омонімія, ускладнює дослідження, оскільки без належного контексту важко визначити, про яке місце йдеться. Також топоніми можуть мати кілька історичних назв, що змінювались із часом внаслідок політичних, культурних або демографічних змін. Наприклад, місто Львів в Україні відоме як Lwów у польській та Lemberg в часи Австро-Угорської імперії.

Іншим суттєвим викликом є зміни топонімів з часом. Географічні назви можуть зазнавати змін через політичні, соціальні або культурні трансформації. Наприклад, під час колонізації, деколонізації чи зміни режимів багато місць отримували нові назви, витісняючи попередні топоніми. Для прикладу відповідно до постанови Верховної ради України №3984-IX «Про перейменування окремих населених пунктів та районів» тільки в Харківській області було перейменовано десять населених пунктів з назвою «Першотравневе» (Постанова ВРУ №1377-VIII). У таких випадках дослідники мають враховувати контекст цих змін і намагатися відстежити первісні назви, щоб повноцінно зрозуміти історичні процеси.

Ще однією складністю є зникнення історичних топонімів через урбанізацію, міграцію населення та інші зміни. Це все призводить до того, що старі

назви витісняються новими або і просто зникають з карт, що тягне за собою втрату частини історичного контексту.

Топонімічні дослідження можуть допомогти відновити ці забуті назви та зберегти історичну пам'ять про місця, які були втрачені або перейменовані. В останні роки дослідники почали використовувати новітні технології, зокрема штучний інтелект та геоінформаційні системи (GIS), для аналізу великих масивів даних. Інтеграція методів обробки природної мови (NLP), машинного навчання та *deep mapping* дозволяє автоматизувати процеси збору, аналізу та візуалізації топонімічних даних, що значно розширює можливості дослідження історичних і сучасних змін у топонімах. Це особливо важливо в контексті цифрової гуманітаристики, де поєднання історичних архівів з інтерактивними картами сприяє створенню нових, більш комплексних досліджень, які враховують як географічні, так і соціокультурні аспекти розвитку регіонів.

Застосування штучного інтелекту у вивченні топонімів

Як ми бачимо, вищезгадані виклики роблять топоніміку складною, але надзвичайно цікавою галуззю, яка дозволяє глибше зрозуміти не лише історичні події, але й культурні та соціальні процеси, що впливали на формування географічних назв. Зі свого боку сучасні методи на основі штучного інтелекту можуть значно полегшити процес дослідження, допомагаючи автоматично аналізувати масиви даних і знаходити відповідності між історичними та сучасними назвами.

Текстовий майнінг та обробка природної мови (NLP) надають нові можливості для автоматизації процесу ідентифікації та аналізу топонімів у великих текстових архівах. Обробка природної мови – це здатність комп'ютерної програми розуміти людську мову так, як вона говорить та пишеться, тобто природну мову (What is natural language processing (NLP)?). Використання NLP дозволяє обробляти величезні обсяги історичних текстів та документів, що раніше вимагало ручної праці дослідників. Алгоритми NLP здатні знаходити та класифікувати географічні назви у контексті різних історичних джерел, таких як листи, судові записи, газети, літописи тощо. Проте одним із основних викликів у цій сфері є багатозначність топонімів – коли одна й та сама назва може стосуватися кількох різних місць. Наприклад, назва «Paris» може стосуватися як столиці Франції, так і маленького міста в штаті Техас. Використовуючи інформацію про сусідні топоніми або адміністративну належність, алгоритми можуть точно іден-



тифікувати конкретне місце. Наприклад, геокодування «Вишневе» в контексті Київської області допоможе відрізнити його від «Вишневого», яке може знаходитися в Харківському регіоні.

Один з ключових методів NLP також є розпізнання іменних сутностей (NER), який використовується для автоматичної ідентифікації топонімів серед інших типів власних назв (імена людей, організацій). NER-моделі здатні визначати назви місць та класифікувати їх як географічні об'єкти, допомагаючи виділяти важливі топоніми з тексту для подальшого аналізу (What is named entity recognition?). У статті «A Study on Toponymic Entity Recognition Based on Pre-Trained Models Fused with Local Features for Genglubu in the South China Sea», автори досліджують модель розпізнавання топонімів, яка поєднує попередньо натреновані мовні моделі з локальними ознаками для підвищення точності ідентифікації географічних назв. Модель враховує специфіку топонімів, зокрема їхню відмінності у граматичних структурах стародавніх і сучасних мов (Wei, Y., 2024).

Застосування NER у контексті топонімів дозволяє не лише автоматизувати процес ідентифікації, але й підвищити якість аналізу завдяки глибшому розумінню контексту вживання назв. Це особливо важливо для дослідження в історичних і культурних контекстах, де точність ідентифікації має вирішальне значення для розуміння географічних змін і їхнього впливу на розвиток суспільства. Дослідження показують, що впровадження NER може суттєво покращити результати в аналізі історичних текстів, забезпечуючи науковців ефективними інструментами для обробки та аналізу великих обсягів даних.

Цей підхід є особливо актуальним для України, де існує багато топонімічних аналогій і схожих назв, що ускладнює точне визначення конкретних географічних об'єктів. Використання NER-моделей може допомогти в розрізненні подібних назв та забезпечити більш глибоке розуміння історичних і культурних зв'язків.

Ще одним значущим прикладом є дослідження, проведені з використанням ШІ для геолокації у великих масивах тексту, які поєднують сучасні моделі обробки природної мови з географічною інформаційною системою (GIS). У роботі, опублікованій на платформі SpringerLink, дослідники описують, як NLP та геоінформаційні технології використовуються для автоматичного прив'язування топонімів до їхніх координат на сучасних картах. Це особливо корисно для аналізу старих текстів, де топоніми вже могли змінитися або зникнути.

У дослідженні проведеному в університеті Баффало, були представлені нові методи для автоматичного збору географічної інформації з текстових даних. Зокрема, автори зазначають, що застосування NLP разом з методами машинного навчання дозволяє виявляти не лише топоніми, а й інші важливі елементи, такі як соціально-економічні дані, що можуть бути прив'язані до конкретних географічних об'єктів (Yingjie Hu, 2020).

Ще одним важливим інструментом є Deep Mapping, або глибоке картування. Його називають «важливим наступним кроком» після просторового повороту, який перетворив гуманітарні науки на рубежі століть (David J. Bodenhamer, 2015). Власне глибоке картування – це метод, який об'єднує географічні, історичні та культурні дані для створення багатовимірних візуальних моделей, що дозволяють краще розуміти взаємозв'язки між просторовими та соціальними змінами. Замість того, щоб просто відображати географічні місця, глибоке картування прагне включити наративи, тексти, зображення, звук та інші медіа, що додають контексту й створюють більш цілісне розуміння певного регіону. Це дозволяє дослідникам не тільки візуалізувати зміни в ландшафті, але й вивчати, як ці зміни відображають історичні події, соціальні структури та культурні перетворення.

Одним з прикладів успішного використання глибокого картування є проект Lancaster University Deep Mapping Project, де дослідники використовують ШІ для інтеграції картографічних даних з історичними текстами та архівами (Deep mapping). Проект має на меті створення інтерактивних карт, які дозволяють вивчати зміни у ландшафті через багатовимірний підхід, що включає елементи економічної, соціальної та культурної історії. Завдяки вищезгаданій технології обробки природної мови та машинного навчання, дослідники можуть автоматично аналізувати старовинні документи та карти, виділяти важливі топоніми й з'єднувати їх з сучасними географічними об'єктами.

Сучасні підходи до глибокого картування активно використовують алгоритми машинного навчання для автоматизації складних картографічних процесів, таких як розпізнавання шаблонів, генералізація карт і стилізація (Lars Harier, 2024). Зокрема, глибинне навчання, включаючи нейронні мережі та графічні структури, дозволяє ефективно обробляти великі обсяги просторових даних, підвищуючи точність карт і знижуючи потребу в ручній праці. Важливою складовою є інтеграція картографічних знань для покращення ефективності моделей, що відкриває нові можливості для створення багатошарових та адаптивних карт.

Застосування глибокого картування в українському контексті може дати новий поштовх у дослідженні історичних топонімів, особливо для регіонів, які зазнали значних соціально-економічних змін або постраждали від бойових дій, політичних перетворень, тощо. Використання цього підходу дозволяє відслідковувати еволюцію топонімів через час, досліджувати, як зміни адміністративних кордонів або демографічних показників вплинули на назви населених пунктів, та як вони змінювалися в контексті історичних подій.

Інтеграція алгоритмів текстового майнінгу та NLP з іншими технологіями, такими як *deep mapping*, дозволяє створювати багатовимірні моделі історичних географічних даних, що допомагає краще зрозуміти не лише фізичні зміни у ландшафті, але й культурні зміни, що відбувалися разом із трансформацією назв місць. Зокрема, використання автоматизованих систем для аналізу архівних документів в українському контексті може сприяти відновленню втрачених або забутих назв, що, в свою чергу, поглиблює розуміння регіональної ідентичності та її еволюцію в часі.

Інструменти ШІ для аналізу топонімів

Геопарсинг та геокодування (*geoparsing and geocoding*) є ключовими технологіями, які забезпечують автоматичну ідентифікацію та координатне прив'язування географічних об'єктів у текстах. Геопарсинг дозволяє знаходити топоніми у великих текстових масивах, а геокодування переводить ці топоніми у конкретні географічні координати. (Ludovic Moncla, 2014) Геопарсинг включає розпізнавання топонімів і поділяється на методи, що базуються на навчанні та методи, які використовують правила синтаксично-семантичного аналізу. Геокодування, або вирішення топонімічних невизначеностей, має на меті зіставлення топоніму з географічними координатами, вирішуючи проблеми амбівалентності. Це особливо корисно для обробки історичних документів, де багато місць можуть бути зазначені не в сучасних назвах, а в старих чи архаїчних формах.

У статті Клер Гровер описано використання геокодування для автоматичного визначення координат топонімів, знайдених у старих текстових документах та картографічних матеріалах. Це дозволило оцифровувати історичні колекції та прив'язувати старі назви до сучасних карт (Claire Grover, 2010). Такий підхід допомагає дослідникам отримувати точні географічні дані зі старих карт, наприклад, для аналізу історичних маршрутів та подій.

Також цікавим для огляду, в контексті геокодування, є проект GeoTxt. Дослідники розробили

веб-API, що автоматично розпізнає та геокодує топоніми у текстах (Morteza Karimzadeh, 2013). Цей інструмент призначений для роботи з даними, що містять просторову інформацію, зокрема з короткими текстами. GeoTxt API працює як веб-сервіс, який приймає текстові дані та визначає назви місць, присвоюючи їм географічні координати. Крім того підтримується розпізнавання місць, людей та організацій з використанням різних механізмів розпізнавання іменованих сутностей. Логіка цього інструменту враховує просторові зв'язки та ієрархію географічних об'єктів. Наприклад, у випадках, коли в досліджуваному тексті згадується декілька однакових топонімів, система враховує контекст для правильного визначення місця.

Ще одним цікавим інструментом є генерація географічних моделей за допомогою штучного інтелекту та глибинного навчання, який відкриває нові можливості для аналізу і візуалізації історичних змін топонімів. Власне, глибинне навчання є підмножиною машинного навчання та штучного інтелекту, яке імітує роботу людського мозку, дозволяючи комп'ютерам вирішувати складні завдання, що створюють нові ідеї та рішення (What Are Deep Learning Models?). Використовуючи нейронні мережі, ШІ може автоматично ідентифікувати та відновлювати деталі старих карт, накладаючи їх на сучасні географічні контури. Це дозволяє дослідникам відслідковувати зміни у назвах місць, міграційні шляхи та адміністративні трансформації, які відбувалися протягом часу.

Глибинне навчання допомагає зібрати дані з різних джерел, таких як старі документи, рукописи та карти, для створення комплексних географічних моделей. Завдяки аналізу великих масивів даних, ШІ здатний автоматично виявляти закономірності та тренди у зміні топонімів, що раніше вимагало значних зусиль ручної роботи. Наприклад, це може бути корисним для дослідження міських районів, які зазнали суттєвих змін, або для відстеження старих торгових шляхів та їхніх впливів на локальну географію.

У дослідженні Йоахіма Дайбера показано, як поєднання методів геоаналітики та історичних даних може бути використано для створення багатовимірних моделей, що дозволяють вивчати зміни в топонімах через час (Joachim Daiber, 2013). Використання ШІ для аналізу даних дозволяє з'єднувати інформацію з різних періодів і прив'язувати її до сучасних географічних координат. Це дозволяє виявляти не лише фізичні зміни у ландшафті, але й соціально-економічні та культурні зміни, що впливали на еволюцію топонімів.

Поєднання цих технологій дозволяє створювати інтерактивні карти, які можуть візуалізувати складні історичні процеси та показувати, як змінювалися кордони, адміністративні одиниці або назви місць під впливом різних факторів. Наприклад, використання ШІ для автоматизованого створення географічних моделей може допомогти виявити, як змінювалися назви українських сіл і міст через політичні події, колонізацію, міграцію, війни та інші фактори.

Перспективи автоматизованого аналізу змін топонімів: традиційні методи vs. штучний інтелект

У дослідженні змін топонімів традиційно використовуються бази даних та аналітичні методи, що базуються на архівних документах, історичних картах і адміністративних актах. Бази даних дозволяють зберігати, структурувати та швидко обробляти інформацію про назви населених пунктів, дати їх перейменування та причини змін. Найбільш поширеними джерелами для побудови таких баз є реєстри адміністративно-територіального устрою, архівні документи та історичні атласи, а також геоінформаційні системи, які дозволяють візуалізувати просторові зміни топонімів.

Для роботи з базами даних існує широкий спектр інструментів, які можуть бути використані залежно від складності проекту та потреб дослідників. Серед найбільш поширених:

- PostgreSQL (зокрема, розширення PostGIS) – потужна реляційна база даних, що дозволяє здійснювати географічний аналіз і працювати з просторовими даними.

- MySQL – зручна база даних, що добре підходить для управління текстовими і числовими даними.

- Microsoft SQL Server – інструмент, що часто використовується в державних установах для роботи з адміністративними даними.

- MongoDB – нереляційна база даних, що добре підходить для обробки неструктурованих історичних даних і документів.

- QGIS та ArcGIS – програмне забезпечення для просторового аналізу, яке може інтегрувати бази даних з картографічною інформацією, що особливо корисно у топонімічних дослідженнях.

- Google BigQuery – хмарне рішення для аналізу великих обсягів даних, що може бути корисним при обробці масивних історичних записів.

Перевагою цього методу є чітка структурованість даних та висока точність, оскільки інформація вводиться вручну. Однак він має суттєві обмеження: бази даних ефективні для невеликих обсягів інформації, але їх створення та оновлення потребує значних людських ресурсів. Крім того, аналіз таких даних переважно є описовим і не виявляє прихова-

них закономірностей. Використання традиційних методів передбачає залучення експертів-істориків, які мають ґрунтовні знання про регіональні особливості та політичний контекст перейменувань. Це дає змогу глибоко аналізувати зміни топонімів, але вимагає значного часу та людських ресурсів.

На відміну від баз даних, штучний інтелект дозволяє обробляти великі масиви текстової та картографічної інформації без необхідності ручного введення кожного запису. Основні методи штучного інтелекту, що можуть застосовуватися у дослідженні змін топонімів, включають обробку природної мови, яка дозволяє автоматично знаходити топоніми в історичних текстах, співвідносити їх із сучасними назвами та виявляти тенденції у перейменуванні. Також важливу роль відіграють геокодування та геопарсинг, що прив'язують знайдені у текстах топоніми до географічних координат, спрощуючи їх аналіз та візуалізацію. Крім того, технологія *deep mapping* дозволяє інтегрувати картографічні дані із історичними джерелами, створюючи динамічні моделі змін назв місць у часі. Використання машинного навчання допомагає аналізувати великі обсяги топонімічних даних і автоматично визначати, які фактори найчастіше впливали на перейменування, серед яких політичні, соціальні та географічні аспекти.

Застосування штучного інтелекту дає змогу отримати комплексний аналіз топонімічних трансформацій у регіонах із високою динамікою змін назв населених пунктів. Наприклад, у регіонах, що зазнали суттєвих змін унаслідок політичних реформ чи військових конфліктів, моделі машинного навчання можуть допомогти відстежити не лише дати та причини перейменувань, а й прогнозувати можливі майбутні зміни, ґрунтуючись на історичних закономірностях. Штучний інтелект також дозволяє автоматично інтегрувати дані з різних джерел, таких як історичні тексти, газети, адміністративні укази та топографічні карти, що дає можливість значно підвищити точність аналізу.

Незважаючи на значні переваги, використання штучного інтелекту в аналізі змін топонімів має також і низку проблем та викликів. Першим з них є необхідність великих обсягів якісних даних. Для ефективного навчання моделей машинного навчання необхідно мати тисячі або навіть мільйони записів, що включають історичні згадки топонімів, їхні варіації та контекст використання. Збір таких даних є складним завданням, оскільки вони можуть бути розпорошені по різних архівах, рукописах та картах, багато з яких не оцифровані або містять помилки.

Ще однією важливою проблемою є складність створення ефективної моделі. Вибір архітектури



нейронної мережі або алгоритму навчання, налаштування гіперпараметрів та підбір відповідних вхідних даних вимагають значних знань у сфері штучного інтелекту та машинного навчання. Неправильно підібрана модель може давати неточні або нерелевантні результати, що знижує ефективність аналізу.

Також варто враховувати складність правильної генерації запитів (prompt engineering). Для моделей обробки природної мови важливо правильно формулювати запити, щоб отримати релевантні результати. Наприклад, якщо модель аналізує історичні документи, її можна навчити розпізнавати специфічні патерни змін назв, але формулювання некоректного запиту може призвести до пропуску важливих даних або генерації помилкових висновків.

Іншим викликом є проблема неоднозначності топонімів. Одна й та сама назва може використовуватися для різних населених пунктів або змінювати форму в різні історичні періоди. Це ускладнює автоматичний аналіз та потребує додаткових алгоритмів для контекстуального визначення правильного топоніму.

Також існує проблема адаптації моделей до різних мовних та діалектичних варіацій. Багато історичних джерел можуть бути написані застарілими мовами або містити регіональні особливості, які стандартні NLP-моделі не завжди можуть правильно інтерпретувати. Це потребує створення спеціальних корпусів текстів для навчання моделей та постійного оновлення даних.

Нарешті, використання штучного інтелекту пов'язане з великими обчислювальними витратами. Обробка текстових та картографічних даних вимагає значних ресурсів, особливо якщо потрібно аналізувати великі обсяги інформації. Це може стати бар'єром для дослідників, які не мають доступу до потужних серверів або хмарних обчислювальних сервісів.

| Критерії        | База даних                                                     | Штучний інтелект                                                       |
|-----------------|----------------------------------------------------------------|------------------------------------------------------------------------|
| Обсяг даних     | Ефективно працює з малими обсягами (до десятків тисяч записів) | Потребує великих обсягів даних (мільйони записів для навчання моделей) |
| Автоматизація   | Введення даних і аналіз проводяться вручну                     | Аналізує великі масиви даних автоматично                               |
| Глибина аналізу | Виявляє лише зафіксовані зміни                                 | Може визначити приховані закономірності в змінах топонімів             |

|                   |                                             |                                                                |
|-------------------|---------------------------------------------|----------------------------------------------------------------|
| Гнучкість         | Вимагає ручного оновлення                   | Може працювати з динамічними наборами даних                    |
| Візуалізація      | Обмежена простими таблицями та картами      | Використовує GIS, deer mapping і аналітику трендів             |
| Швидкість обробки | Обмежена ручним введенням даних та аналізом | Швидке опрацювання великих текстових та картографічних масивів |
| Адаптивність      | Потребує ручного доповнення нових даних     | Може адаптуватися до нових даних та коригувати результати      |

Попри ці виклики, штучний інтелект залишається потужним інструментом для дослідження змін топонімів, особливо у поєднанні з традиційними методами та базами даних. Впровадження комплексного підходу, що поєднує ручний аналіз, алгоритми машинного навчання та геоінформаційні системи, дозволить отримати більш точні результати та розширити можливості дослідження історичних трансформацій топонімів.

При виборі між використанням баз даних та штучного інтелекту необхідно враховувати обсяг дослідження, його цілі, доступні ресурси та бажану швидкість аналізу. Умовно можна розділити випадки використання цих методів на два типи: локальні дослідження та масштабні аналітичні проекти.

Якщо дослідження охоплює невеликий регіон або обмежену кількість топонімів, а дослідник вже має чітко сформульовані питання, на які хоче отримати відповіді, ефективним інструментом є бази даних. Вони дозволяють зберігати інформацію у структурованому вигляді, легко здійснювати пошук і фільтрацію, а також інтегрувати дані з картографічними ресурсами. Бази даних добре підходять для аналізу адміністративних змін у межах одного регіону, короткострокових досліджень або створення каталогу змінених топонімів.

Однак якщо дослідження є масштабним та вимагає аналізу великої кількості джерел, таких як архівні документи, історичні карти, літописи та географічні довідники, ефективнішим рішенням буде штучний інтелект. Алгоритми машинного навчання можуть автоматично ідентифікувати топоніми у текстах, виявляти зв'язки між ними та визначати приховані закономірності, що залишаються непомітними при традиційному аналізі. Крім того, ШІ дозволяє працювати з неструктурованими даними, що значно розширює можливості дослідників.

Основними критеріями вибору між базами даних та штучним інтелектом є:

Обсяг даних: невеликі структуровані набори краще обробляти за допомогою баз даних, великі та неструктуровані – за допомогою штучного інтелекту.

Цілі дослідження: якщо потрібно отримати відповіді на конкретні питання, бази є ефективнішими, якщо ж необхідно знайти нові зв'язки або тенденції – краще використовувати ШІ.

Часові витрати: бази даних вимагають більше ручної роботи при введенні даних, але штучний інтелект потребує часу на навчання моделей та їхню адаптацію.

Ресурси: робота з ШІ потребує значних обчислювальних потужностей, тоді як БД можуть працювати на стандартних серверних потужностях або навіть локальних комп'ютерах.

Разом з тим застосування баз даних і штучного інтелекту не слід розглядати як два протилежні підходи. Навпаки, бази даних є важливим джерелом структурованої інформації, яка може слугувати основою для алгоритмів машинного навчання та аналітичних моделей. Ефективне дослідження топонімічних змін потребує взаємодії цих методів, оскільки саме традиційні бази даних містять велику кількість інформації, необхідної для навчання моделей штучного інтелекту.

Бази даних дозволяють впорядковувати історичні записи про топоніми, фіксуючи інформацію про їхні зміни у різні періоди. Однак самостійно вони не можуть виявляти складні закономірності або прогнозувати майбутні тенденції. У цьому аспекті штучний інтелект стає ключовим інструментом, здатним аналізувати величезні обсяги текстів, карт та архівних матеріалів.

Таким чином, бази даних є фундаментом для досліджень, а штучний інтелект – інструментом, який дозволяє підняти аналіз на новий рівень. Поєднання цих методів відкриває можливості для масштабного вивчення топонімічних процесів,

спрощуючи обробку великих обсягів інформації та підвищуючи точність висновків дослідників.

**Висновки.** Вивчення топонімів є важливою складовою історичної географії та культурологічних досліджень. Традиційні методи аналізу, такі як використання баз даних та архівних джерел, дозволяють глибоко досліджувати окремі випадки зміни назв, але мають значні обмеження у швидкості та масштабі обробки даних.

Запровадження штучного інтелекту у сферу топонімічних досліджень відкриває нові можливості для аналізу, дозволяючи автоматично ідентифікувати топоніми у великих текстових архівах, співвідносити їх із географічними координатами та відстежувати закономірності їхніх змін. Технології, такі як NLP, deep mapping та геокодування, дозволяють значно розширити обсяги аналізованих джерел і знаходити приховані зв'язки між змінами топонімів та історичними подіями.

Водночас використання ШІ потребує значних обчислювальних ресурсів і великих обсягів якісних даних для навчання моделей. Також залишається низка викликів, пов'язаних із неоднозначністю топонімів, мовними особливостями історичних джерел і коректним формулюванням запитів для аналізу.

Оптимальним підходом до дослідження топонімів є поєднання традиційних методів із можливостями штучного інтелекту. Бази даних залишаються ефективними для структурованого аналізу та короткострокових досліджень, тоді як алгоритми машинного навчання дозволяють виявляти нові закономірності та автоматизувати обробку великих масивів інформації.

Подальший розвиток методів автоматизованого аналізу топонімів сприятиме розширенню можливостей історичних та географічних досліджень, дозволяючи глибше розуміти соціокультурні процеси, що впливали на формування та трансформацію географічних назв.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Перелік сіл та селищ, перейменованих відповідно до Закону України «Про засудження комуністичного та націонал-соціалістичного (нацистського) тоталітарних режимів в Україні та заборону пропаганди їхньої символіки». URL: [https://land.gov.ua/wp-content/uploads/2018/08/Перейменування-сіл\\_ВРУ\\_28\\_08.pdf](https://land.gov.ua/wp-content/uploads/2018/08/Перейменування-сіл_ВРУ_28_08.pdf).
2. Про перейменування окремих населених пунктів та районів: Постанова ВРУ від 19 травня 2016 року № 1377-VIII. URL: <https://zakon.rada.gov.ua/laws/show/1377-19>.
3. Claire Grover, Richard Tobin, Kate Byrne, Matthew Woollard, James Reid, Stuart Dunn, Julian Ball. Use of the Edinburgh geoparser for georeferencing digitized historical collections. *Philosophical transactions of the Royal Society A*. 2010. С. 3875-3889. <https://doi.org/10.1098/rsta.2010.0149>
4. David J. Bodenhamer, John Corrigan, Trevor M. Harris. *Deep Maps and Spatial Narratives*. Indiana University Press. 2015. 256 с.
5. Deep mapping. *Geospatial Innovation in the Digital Humanities*. URL: <https://wp.lancs.ac.uk/lakesdeepmap/the-project/gis-deep-mapping/>
6. Joachim Daiber, Max Jakob, Chris Hokamp, Pablo N. Mendes. Improving efficiency and accuracy in multilingual entity extraction. *I-SEMANTICS '13: Proceedings of the 9th International Conference on Semantic Systems*. 2013. С. 121-124. <https://doi.org/10.1145/2506182.2506198>



7. Lars Harrie, Guillaume Touya, Rachid Oucheikh, Tinghua Ai, Azelle Courtial & Kai-Florian Richter. Machine learning in cartography. *Cartography and Geographic Information Science*. 2024. С. 1-19. <https://doi.org/10.1080/15230406.2023.2295948>
8. Ludovic Moncla, Walter Renteria-Agualimpia, Javier Nogueras-Iso, Mauro Gaio. Geocoding for texts with fine-grain toponyms: an experiment on a geoparsed hiking descriptions corpus. *International Conference on Advances in Geographic Information Systems*. 2014. 11 с.
9. Morteza Karimzadeh, Wenyi Huang, Siddhartha Banerjee. GeoTxt: a web API to leverage place references in text. *SIGSPATIAL'13: 21st SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 2013. С. 72-73. <https://doi.org/10.1145/2533888.25339>
10. What Are Deep Learning Models? Types, Uses, and More. *Coursera*. URL: <https://www.coursera.org/articles/deep-learning-models>
11. What is named entity recognition? *IBM*. URL: <https://www.ibm.com/think/topics/named-entity-recognition>
12. What is natural language processing (NLP)? *TechTarget Network*. URL: <https://www.techtarget.com/searchenterpriseai/definition/natural-language-processing-NLP>
13. Wei, Y.; Li, Y.; Zhou, X. A Study on Toponymic Entity Recognition Based on Pre-Trained Models Fused with Local Features for Genglubu in the South China Sea. *Electronics*. 2024. 16 с. <https://doi.org/10.3390/electronics13010004>
14. Yingjie Hu & Benjamin Adams. Harvesting Big Geospatial Data from Natural Language Texts. *Handbook of Big Geospatial Data*. 2020. С. 487-507.

## REFERENCES

1. Claire Grover, Richard Tobin, Kate Byrne, Matthew Woollard, James Reid, Stuart Dunn, Julian Ball. (2010) Use of the Edinburgh geoparser for georeferencing digitized historical collections. *Philosophical transactions of the Royal Society A*. pp. 3875-3889. <https://doi.org/10.1098/rsta.2010.0149>.
2. David J. Bodenhamer, John Corrigan, Trevor M. Harris. (2015) *Deep Maps and Spatial Narratives*. Indiana University Press. 256 p.
3. Deep mapping. *Geospatial Innovation in the Digital Humanities*. URL: <https://wp.lancs.ac.uk/lakesdeepmap/the-project/gis-deep-mapping/>.
4. Joachim Daiber, Max Jakob, Chris Hokamp, Pablo N. Mendes. (2013) Improving efficiency and accuracy in multilingual entity extraction. *I-SEMANTICS '13: Proceedings of the 9th International Conference on Semantic Systems*. pp. 121-124. <https://doi.org/10.1145/2506182.2506198>.
5. Lars Harrie, Guillaume Touya, Rachid Oucheikh, Tinghua Ai, Azelle Courtial & Kai-Florian Richter. (2024) Machine learning in cartography. *Cartography and Geographic Information Science*. pp. 1-19. <https://doi.org/10.1080/15230406.2023.2295948>.
6. Ludovic Moncla, Walter Renteria-Agualimpia, Javier Nogueras-Iso, Mauro Gaio. (2014) Geocoding for texts with fine-grain toponyms: an experiment on a geoparsed hiking descriptions corpus. *International Conference on Advances in Geographic Information Systems*. 11 p.
7. Morteza Karimzadeh, Wenyi Huang, Siddhartha Banerjee. (2013) GeoTxt: a web API to leverage place references in text. *SIGSPATIAL'13: 21st SIGSPATIAL International Conference on Advances in Geographic Information Systems*. pp. 72-73. <https://doi.org/10.1145/2533888.25339>.
8. Perelik sil ta selyshch, pereimenovanykh vidpovidno do Zakonu Ukrainy «Pro zasudzhenia komunistychnoho ta natsional-sotsialistychnoho (natsystskoho) totalitarnykh rezhymiv v Ukraini ta zaboronu propahandy yikhnoi symboliky». [List of villages and settlements renamed in accordance with the Law of Ukraine «On the Condemnation of Communist and National Socialist (Nazi) Totalitarian Regimes in Ukraine and the Prohibition of Propaganda of Their Symbols.»] URL: [https://land.gov.ua/wp-content/uploads/2018/08/Pereimenuvannia-sil\\_VRU\\_28\\_08.pdf](https://land.gov.ua/wp-content/uploads/2018/08/Pereimenuvannia-sil_VRU_28_08.pdf). [in Ukrainian].
9. Pro pereimenuvannia okremykh naselenykh punktiv ta raioniv: Postanova VRU vid 19 travnia 2016 roku №1377-VIII. [On the Renaming of Certain Settlements and Districts: Resolution of the Verkhovna Rada of Ukraine dated May 19, 2016, No. 1377-VIII.] URL: <https://zakon.rada.gov.ua/laws/show/1377-19>. [in Ukrainian].
10. What Are Deep Learning Models? Types, Uses, and More. *Coursera*. URL: <https://www.coursera.org/articles/deep-learning-models>.
11. What is named entity recognition? *IBM*. URL: <https://www.ibm.com/think/topics/named-entity-recognition>.
12. What is natural language processing (NLP)? *TechTarget Network*. URL: <https://www.techtarget.com/searchenterpriseai/definition/natural-language-processing-NLP>.
13. Wei, Y.; Li, Y.; Zhou, X. (2024) A Study on Toponymic Entity Recognition Based on Pre-Trained Models Fused with Local Features for Genglubu in the South China Sea. *Electronics*. 16 p. <https://doi.org/10.3390/electronics13010004>.
14. Yingjie Hu & Benjamin Adams. (2020) Harvesting Big Geospatial Data from Natural Language Texts. *Handbook of Big Geospatial Data*. pp. 487-507. [in English].