

UDC 004.811.581+94(510)+39(510)“1984/2024”

HISTORIOGRAPHICAL OVERVIEW OF THE DEVELOPMENT OF DIGITAL HUMANITIES IN SINOLOGY (1984–2024)

Victor Kiktenko

DSc (Philosophy), Senior Research Fellow

A. Krymskyi Institute of Oriental Studies, NAS of Ukraine

4, Mykhailo Hrushevskyi St., Kyiv, 01001, Ukraine

kiktenko@gmail.com

ORCID: 0000-0002-0078-4285

The study presents a comprehensive historiographical analysis of the formation and development of digital humanities within Sinology over the course of forty years, from 1984 to 2024. The work traces the evolution of the scientific sphere – from the earliest experiments in computational processing of Chinese texts to the establishment of a modern digital infrastructure for complex Sinological research. The methodological foundation is a chronological analysis combined with a systematic classification of key projects, initiatives, and technological solutions. More than thirty reference projects and initiatives have been examined, including encoding standards (GB 2312, Big5, Unicode), technological platforms (TEI, XML), and large-scale digital repositories. Particular emphasis is made on the emergence of digital Sinology as an autonomous academic discipline. It is demonstrated that this focus area has developed along three main trajectories: technological (the standardization of Chinese character encoding through Unicode, the implementation of XML and TEI standards for text markup); methodological (the development of principles for structured text encoding, the creation of relational databases for biographical and textual information); and institutional (the establishment of international scholarly networks and research centers). Eight foundational principles of digital Sinology are outlined: encoding standardization, structural markup, relational databases, international cooperation, open access, metadata preservation, scalability, and format compatibility. It is shown that these principles continue to define the current development of this scientific sphere. The study highlights the successful integration of digital technologies into Sinological research, which has resulted in the establishment of a robust technological foundation for future academic projects. The geographical scope – Mainland China, Taiwan, Hong Kong, and the United States – illustrates the international character of the emergence of digital Sinology. The contributions of leading institutions are systematized, including Academia Sinica, the Ministry of Culture of the PRC, the Harvard-Yenching Institute, the National Library of China, and international technological consortia. The originality of this research consists in its first holistic review of the forty-year development of the scientific sphere, detailing key stages, systematizing projects, and formulating the fundamental principles that have shaped the modern digital infrastructure of Sinological studies. The practical significance of the results consists in the potential for their application in strategic planning of future digital initiatives in the scholarly study of China.

Keywords: artificial intelligence; big data analytics; digital humanities; institutionalization; Sinology; standardization; text digitization

ІСТОРИОГРАФІЧНИЙ ОГЛЯД РОЗВИТКУ ЦИФРОВОЇ ГУМАНІТАРИСТИКИ В КИТАЄЗНАВЧИХ СТУДІЯХ (1984–2024)

В. О. Кіктенко

Вступ

У XXI столітті цифрові інструменти та аналітика великих даних дедалі глибше інтегруються в гуманітарні дослідження, радикально трансформуючи усталені парадигми наукового пізнання. Особливо виразно ці зміни спостерігаються у сфері китаєзнавства: сучасні дослідники здобули безпрецедентний доступ до масштабних електронних корпусів класичних текстів та археографічних джерел, водночас стикаючись із викликами теоретичного та методичного осмислення нових способів їхньої обробки та інтерпретації. Попри проривні зміни останнього десятиліття, цифрова гуманітаристика в китаєзнавчих студіях протягом тривалого часу залишалася поза фокусом глобальних академічних дискурсів, лише нині поступово виходячи на передній план наукових інновацій [Kurzynski 2025]. Саме тому системне дослідження стану, здобутків та перспектив розвитку цифрового китаєзнавства є вкрай актуальним, з огляду на високий темп упровадження технологій штучного інтелекту, машинного навчання й аналітики великих даних у вивчення історії, мови та культури Китаю. Комплексний аналіз етапів становлення, методологічних викликів та потенціалу цифрової гуманітаристики дає змогу не лише визначити нові підходи до дослідження китайських джерел, а й осмислити глибші трансформації в теорії та практиці гуманітарних наук. Отже, це дослідження спрямоване на критичне переосмислення впливу технологічних інновацій на традиційні методи китаєзнавчих студій, окреслення ключових напрямів та перспектив подальшого розвитку галузі.

Попри те що окремі аспекти розвитку цифрової гуманітаристики в китаєзнавстві неодноразово привертали увагу дослідників, досі бракувало комплексного огляду історії впровадження цифрових методів у вивчення Китаю. Зокрема, Ген Юаньлі здійснив широкий аналіз цифровізації стародавніх китайських текстів протягом тридцяти років [耿元骊 2009]; Ян Гуйхуа, Вей Лю, Сунь Вей, Чжень Сіхуей, Вей Давей та Сунь Їган дослідили динаміку розвитку цифрових бібліотек у Китаї [Yang Guihua 2002; Wei Liu 2003; Sun Wei 2005; Zhen Xihui 2010; Wei Dawei, Sun Yigang 2010]; Лі Южень та Сун Їнчунь у своїй спільній статті охарактеризували процес формування інфраструктури західного цифрового китаєзнавства [李友仁, 宋迎春 2018], а Д. Стерджен детально розглянув еволюцію “Китайського текстового проекту” [Sturgeon 2019]. Окремі аспекти найвагоміших змін у цифрових гуманітарних дослідженнях Китаю проаналізувала Чжао Вей [赵薇 2021], а Чень Цзін окреслив витоки й основні етапи становлення цифрового підходу до китаєзнавчих студій, особливо акцентуючи на методологічних та інституційних чинниках [Chen Jing 2022]. Вагоме узагальнення розвитку цифрової гуманітаристики в КНР у 1980–2020-х роках здійснили Лю Ши, Чжао Вей та Лі Фейюе [Liu Shi, Zhao Wei, Li Feiyue 2024].

Отже, метою цієї статті є здійснення комплексного дослідження еволюції цифрової гуманітаристики в китаєзнавстві протягом 1984–2024 років з акцентом на аналізі найважливіших проектів, методологічних підходів та інституційних ініціатив. Для реалізації поставленої мети визначено такі завдання: 1) провести ретроспективний аналіз ключових подій, проектів та організацій, які визначили становлення цифрової гуманітаристики в китаєзнавстві за чотири десятиліття, і виокремити основні фази її розвитку; 2) здійснити систематизацію та класифікацію головних напрямів цифрових досліджень у китаєзнавчих студіях – від текстології й епіграфіки до археології, лінгвістичного аналізу, соціальних мереж, картографування та обробки великих даних.

Методологія цього дослідження ґрунтується на комплексному підході до аналізу розвитку цифрової гуманітаристики в китаєзнавстві на основі поєднання традиційного інструментарію гуманітарних наук із сучасними методами, що є характерними для самої цифрової гуманітаристики. Основним дослідницьким інструментом є систематичний огляд наукової літератури, що дає змогу ідентифікувати, критично оцінювати та синтезувати наявні знання з обраної проблематики. Для збору інформації було здійснено пошук та аналіз публікацій у рецензованих фахових журналах (зокрема, “Digital Scholarship in the Humanities”, “Digital Humanities Quarterly”), матеріалів наукових конференцій, звітів про наукові проекти, а також офіційних веб-ресурсів дослідницьких інституцій. Відбір релевантних джерел здійснювався із застосуванням пошукових запитів англійською (“Digital Methods in Sinology”) та китайською мовою (“汉学数字方法”). Для організації та осмислення зібраної інформації використовувалися тематичне та хронологічне впорядкування, а також порівняльний аналіз, який забезпечує зіставлення різних підходів і методологій у межах цифрової гуманітаристики. Історіографічний підхід дав змогу простежити динаміку розвитку ключових концепцій, ідей та практик цифрових досліджень у галузі китаєзнавства. Крім того, у роботі застосовано елементи кейс-стаді для глибокого аналізу репрезентативних проектів та ініціатив, що ілюструють еволюцію цифрових технологій у китаєзнавстві. Такий вибір методів відображає міждисциплінарний характер дослідження та сприяє комплексному розумінню взаємодії традиційних і цифрових стратегій у сучасному китаєзнавстві.

1. Генезис цифрової гуманітаристики в китаєзнавстві: початковий етап формування дисципліни (1984–1995)

Перші експерименти із застосування комп’ютерних технологій у вивченні китайських джерел припадають на середину 1970-х років, хоча тоді ще не існувало поняття “цифрова гуманітаристика” в сучасному значенні. Зокрема, у 1975 році професор Гамбурзького університету У Юнтун здійснив новаторське дослідження з комп’ютерного індексування англійського перекладу “Книги пісень” (“*Ши цзін*”), уперше використавши обчислювальні засоби для аналізу класичного тексту. Того ж року в Сичуанському університеті (Ченду, КНР) було реалізовано одну з найперших спроб цифрової реконструкції фрагментованих написів на ворожилних кістках із давніми китайськими знаками (甲骨文 *цзягувень*) [耿元骊 2009, 17–28], що стало значущим прецедентом використання обчислювальних методів в епіграфічних дослідженнях Китаю. Попри ці ізольовані ініціативи, системний перехід до цифрових підходів у китаєзнавстві розпочався лише в 1980-х роках, коли такі практики набули інституційного й методологічного оформлення.

Одним із визначальних складників розвитку цифрової гуманітаристики в китаєзнавстві став розвиток машинного перекладу. Система KY-1 для англо-китайського перекладу, розроблена Академією військових наук КНР у 1982–1987 роках, стала першою успішною реалізацією машинного перекладу в Китаї. Вона ґрунтувалася на *rule-based* підході, спираючись на розвинену лінгвістичну теорію з потужним інтерпретатором, а також на логічну, запропоновану відомим лінгвістом Дун Чженьдуном (董振东, 1937–2019) ще в 1978 році. Ключовими аспектами цієї семантики були поділ між процесами аналізу й генерації мови, а також здійснення перекладу через семантичні ролі, а не синтаксичні структури. У 1987 році було започатковано амбітний багатомовний проект машинного перекладу за участі Японії, Китаю, Малайзії, Індонезії та Таїланду, який тривав шість років і мав на меті створення високоякісної системи автоматичного перекладу між п’ятьма азійськими мовами [董振东 1984, 25–45; Qian 2005].

У 1988 році система KY-1 була трансформована в англо-китайську систему машинного перекладу Transtar, що стала першим комерційним продуктом такого типу на базі персонального комп’ютера в Китаї. Transtar містив загальний лексичний

запас у 50 тисяч одиниць та три спеціалізовані словники, кожен із яких налічував понад 30 тисяч записів. У 1989 році ця система була відзначена Національною премією КНР за досягнення в галузі науки і техніки. Transtar презентували на численних комп'ютерних виставках і форумах, зокрема в Гуанчжоу, Сінгапурі, Гонконзі та Ганновері, що сприяло підвищенню міжнародного інтересу до розробки. Вже в 1992 році компанії CAS-ICT та GSL Co. Ltd. (Гонконг) вивели на ринок перший електронний словник для машинного перекладу між китайською та англійською мовами – Instant-Dict EC863, який набув широкої популярності та комерційного успіху [Liu 2011].

Визначальним кроком у становленні цифрового китаєзнавства стало впровадження надійних стандартів комп'ютерного кодування китайських ієрогліфів. У 1980 році був затверджений стандарт GB 2312-1980, який охоплював 6763 спрощені китайські знаки та 682 неханські символи, загалом – 7445 одиниць. Запровадження цієї системи кодування зробило можливим масове та послідовне опрацювання китайських текстів з використанням комп'ютерних технологій. У 1981 році стандарт GB 2312 був офіційно впроваджений у Китаї, заклавши підґрунтя для цифрової обробки текстів і забезпечивши уніфіковане кодування спрощених ієрогліфів. Через дев'ять років, у 1990-му, з'явилося доповнення GB/T 12345, яке включило підтримку традиційних символів і суттєво розширило можливості роботи з історичними джерелами. Паралельно із встановленням стандартів кодування в середині 1980-х активно розвивалися нові методи введення китайських ієрогліфів. Серед них одним із перших став *метод цанцзе* (仓颉输入法) – розробка 1976 року видатного ученого й винахідника Чжу Банфу (朱邦复, нар. 1937), що здобув титул “батько китайського комп'ютера” і дав можливість користуватися китайською графікою на стандартній клавіатурі QWERTY.

У 1985 році в Китаї було представлено персональний комп'ютер Great Wall 0520CH – перший комп'ютер, здатний повноцінно обробляти китайські ієрогліфи й підтримувати роботу з національною графікою. Водночас із поступом комп'ютерної техніки активно вдосконалювалися й засоби введення китайських ієрогліфів. Ван Юнмін (王永民, нар. 1948), після п'яти років наполегливої роботи, запропонував метод введення “п'ять штрихів” (五笔 у бі), що давав змогу набирати будь-який ієрогліф лише за п'ять натискань клавіш. Упровадження цього підходу стало справжнім проривом: він кардинально вирішив проблему ефективного введення китайських текстів на стандартній клавіатурі, істотно спростивши й прискоривши роботу з цифровими текстами для широкого кола користувачів.

У 1984 році один із найвидатніших китайських інтелектуалів ХХ століття, письменник і літературознавець Цянь Чжуншу (錢鍾書, 1910–1998) вперше запропонував використання комп'ютерів для аналізу класичних китайських текстів, що ознаменувало новий етап у розвитку цифрового китаєзнавства. Адже опрацювання великих обсягів класичної літератури потребувало не лише ґрунтовної фахової підготовки, а й значних часових і когнітивних ресурсів. Під керівництвом його асистента Луань Гуйміна (栾贵明, 1940–2022) була створена низка знакових оцифрованих колекцій, зокрема “Повна система китайських символів”, “База даних поезії династії Тан” і “Компендіум філософів”. У 1994 році команда Цянь Чжуншу вперше висунула ідею створення глобальної бази даних на основі мікрокомп'ютерів, що стало початковим етапом упровадження мережових технологій для збереження й пошуку джерел. Цей проєкт заклав підвалини майбутньої універсальної електронної платформи для доступу до текстів, яка згодом стала реальністю з поширенням Інтернету. Паралельно з ініціативами в материковому Китаї в 1984 році в Інституті історії та філології Академії Сініка (Тайвань) було започатковано проєкт “Скрипта Сініка” (“Scripta Sinica”, “漢籍電子文獻資料庫”) – розробку структурованої електронної повнотекстової бази даних китайської класичної спадщини за чотирма головними

категоріями: “经” (класика), “史” (історія), “子” (філософія, наука), “集” (літературні збірники) [中央研究院數位文化中心 2014; 中央研究院歷史語言研究所 2025a].

У 1986 році при Шанхайському університеті Цзяотун було створено першу в Китаї лабораторію комп’ютерних досліджень музики, що сконцентрувала свої зусилля на вивченні традиційного китайського інструмента *гуцінь* (古琴). Цей проєкт став важливою віхою в розвитку міждисциплінарних досліджень, де методи комп’ютерних наук почали поєднуватися з музикознавством. Визначальну роль у цьому процесі відіграв професор Чень Чанлін (陈长林, нар. 1932), який ще в 1980-х роках розробив комп’ютерну систему обробки символів *цзяньцзипу* (减字谱) – традиційної системи нотації для *гуціня*. Запровадження цієї інновації дало змогу автоматизувати обробку стародавніх музичних текстів і стало одним із новаторських прикладів застосування обчислювальних методів у китайському музикознавстві. Дослідники лабораторії особливу увагу зосередили на розв’язанні ключової проблеми – браку ритмічної інформації в системі *цзяньцзипу* попри її детальне документування виконавських технік, що ускладнювало інтерпретацію й реконструкцію музики *гуціня*. Заснування цієї лабораторії стало імпульсом для подальшого розвитку аналогічних міждисциплінарних досліджень: вже наступного року, у 1987-му, подібні лабораторії постали й в інших музичних інститутах, зокрема в Уханському музичному коледжі [Yu Hui 2025].

1988 рік відзначився започаткуванням одного з наймасштабніших проєктів цифровізації у сфері китаєзнавства – створення бази даних “Стародавні китайські тексти” (CHANT, CHinese ANcient Texts) у Китайському університеті Гонконгу під керівництвом видатного мовознавця та філософа професора Лю Дяньцзюе (刘殿爵, D. C. Lau, 1921–2010). Ініціатива виникла з усвідомлення надзвичайної складності роботи з великим масивом стародавніх китайських текстів, що традиційно створювало великі виклики для синологічних досліджень. Проєкт реалізовувався під егідою Інституту китайських досліджень університету, та в підсумку було сформовано комплексну електронну базу даних, яка охоплює понад два тисячоліття літературної спадщини Китаю – від XV століття до н. е. до приблизно VI століття н. е. У структурі CHANT виокремлено сім основних колекцій: література доцїньського та ханського періодів (“先秦兩漢文獻”), твори епох Вей, Цзінь і Шістнадцяти царств (“魏晉南北朝文獻”), тексти на бамбукових дощечках та шовкових тканинах (“竹簡帛書文獻”), написи на ворожильних кістках (“甲骨文”), надписи на бронзі (“金文”), традиційні енциклопедичні зібрання (“傳統類書”) та корпус стародавньої лексики (“古代詞彙”), що репрезентують розмаїття писемної та культурної спадщини Китаю. Первісна версія проєкту реалізована на системі Eten DOS, що була популярною платформою для роботи з китайськими текстами в 1980-х роках; у 1990-х роках здійснено перехід до середовища Windows із метою підвищення функціональності та підтримки мультимедійних можливостей. Одним із найбільших технологічних викликів для команди проєкту була потреба працювати з рідкісними та застарілими китайськими ієрогліфами: у стародавніх текстах налічується до 60 тисяч знаків та їхніх варіантів, тоді як стандартні комп’ютерні шрифти охоплювали лише близько 13 тисяч традиційних ієрогліфів [漢達文庫 2025].

У 1994 році Британська бібліотека започаткувала “Міжнародний проєкт Дуньхуан” (International Dunhuang Project, IDP) – новаторську ініціативу міжнародної співпраці, спрямовану на збереження, каталогізацію та цифровізацію рукописів, друкованих текстів, художніх творів, текстилю й артефактів із печер Могао (莫高窟) біля міста Дуньхуана, а також інших археологічних пам’яток східної частини Великого Шовкового шляху. Заснування цього проєкту стало визначною віхою для цифрового китаєзнавства та міжнародного наукового партнерства, слугуючи взірцем для подальших масштабних ініціатив з оцифрування культурної спадщини в усьому світі [International Dunhuang Programme 2025].

Корпусна лінгвістика стала одним із провідних напрямів розвитку цифрової гуманітаристики в китаєзнавстві. Зокрема, у Пекінському університеті була створена система електронного пошуку для повного зібрання поезії епохи Тан, яка завдяки алгоритмам порівняння текстів дала змогу ідентифікувати майже п'ять тисяч повторюваних поезій. Китайські дослідники також застосували стилOMETричні й статистичні методи та комп'ютеризовані підходи для аналізу однорідності перших 80 й останніх 40 глав класичного роману “Сон у червоному теремі”. Одним із перших прикладів кількісної методології в літературознавстві стало дослідження, у якому було проаналізовано географічне розташування поетів династії Тан та їхні міжпоколінні зв'язки і взаємодії [Liu Shi, Zhao Wei, Li Feiyue 2024].

Всі ці проекти стали визначними віхами в історії розвитку цифрової гуманітаристики в китаєзнавстві, адже вони не лише продемонстрували величезний потенціал цифрових технологій для збереження та вивчення культурної спадщини, а й заклали системну основу для подальшого масштабного оцифрування величезного корпусу історико-культурних джерел Китаю. Отже, цифрові гуманітарні практики в китаєзнавстві розглядаються вже не просто як технічне нововведення, а як відповідь на давно усвідомлену потребу гуманітарної науки – працювати з текстами ефективніше, глибше й у значно ширших масштабах. Розвитку цієї галузі значною мірою сприяло активне міжнародне співробітництво науковців Китаю, Тайваню, Гонконгу, США, Великої Британії, що дало імпульс формуванню міждисциплінарної, трансрегіональної, міжкультурної й багатомовної дослідницької спільноти [Tsui, Chen 2019; Liu Shi, Zhao Wei, Li Feiyue 2024, /].

Отже, у 1984–1995 роках у китаєзнавчих дослідженнях настав справжній переломний момент: до традиційного гуманітарного інструментарію були залучені комп'ютерні технології. За цей час було закладено фундаментальні технічні й методологічні підвалини, що стали основою для широкого впровадження обчислювальних підходів у сферу вивчення китайської культури та літератури. Цей період сформував базу для подальшої систематизації і стандартизації цифрових інструментів у гуманітарних науках, зробивши можливим масштабне та організоване використання новітніх технологічних рішень. Однак розвиток цифрових гуманітарних практик у китаєзнавстві супроводжувався низкою унікальних технічних викликів, передусім пов'язаних зі специфікою китайської мови. Як класичний, так і сучасний варіанти китайської мови створювали складні завдання під час процесу оцифрування, організації й аналізу текстів, особливо щодо таких аспектів, як сегментація слів; додаткову складність становили багатство та розмаїття китайських ієрогліфів.

2. Становлення цифрової інфраструктури китаєзнавства: створення архівних ресурсів та методологічних засад (1996–2005)

Потужним імпульсом для динамічного розвитку цифрового китаєзнавства у другій половині 1990-х років стали вибуховий розвиток інтернет-технологій та масштабне нарощування обчислювальних потужностей. Як зазначає Чжень Сіхуей, “після 1995 року Національна бібліотека Китаю уважно стежила за появою електронних бібліотек у світі, заклавши підвалини для формування власної цифрової платформи” [Zhen Xihui 2010]. Вже наприкінці десятиліття це дало змогу запустити перші експериментальні проекти з оцифрування класичних китайських текстів і створення цифрових бібліотек [Yang Guihua 2002; Sun Wei 2005]. Фахівці з різних галузей – лінгвісти, IT-інженери, бібліотекарі – у спільній роботі випробовували найсучасніші на той час сканери та спеціалізоване програмне забезпечення для обробки ієрогліфів. Цей етап вирізнявся динамікою спроб і помилок, проте з таких експериментів і народжувалися перші інструменти для цифрового пошуку й індексації китайських рукописів у мережі. Попри всі труднощі й недосконалості, саме ця дослідницька співпраця заклала підґрунтя для створення масштабних цифрових архівів, які стали сьогодні невіддільною частиною дослідницької інфраструктури [Yang Guihua 2002; Sun Wei 2005].

У середині 1990-х років стало очевидно, що подальший поступ цифрового китаєзнавства неможливий без чіткого, уніфікованого підходу до кодування та обробки китайських ієрогліфів у цифровому середовищі. Відповіддю на цю проблему стало ухвалення в 1996 році документа RFC 1922, який уперше унормував робочі методи кодування та “передачі” китайських текстів у таких сервісах Інтернету, як електронна пошта (RFC 822), мережеві новини (RFC 1036), Telnet (RFC 854), а також на вебсторінках (RFC 1866). До появи цього стандарту надсилати електронні листи китайською було неможливо: ієрогліфи перетворювалися на хаотичні набори знаків. З ухваленням RFC 1922 ця ситуація докорінно змінилася, відкривши нові перспективи для транснаціональної комунікації та розвитку цифрового китаєзнавства [Crispin 2022].

Паралельно з розробленням мережевих протоколів вирішувалися й фундаментальні питання – як “упакувати” сотні тисяч китайських ієрогліфів у єдину систему кодування, що забезпечуватиме їхнє коректне відображення на будь-яких пристроях і в усьому світі. Спочатку існувало кілька взаємонепоєднаних національних стандартів, але справжній технологічний прорив був здійснений із впровадженням стандарту Unicode. Для порівняння: китайський стандарт GB 2312 (1980) міщував лише 6763 знаки, тоді як Unicode 3.2 (2002) вже підтримував понад 95 тисяч ієрогліфів. Особливим переломним моментом став вихід Unicode 3.1 у 2001 році, коли було додано розширення CJK Extension A, що вперше дало змогу повноцінно відображати рідкісні та історичні знаки, потрібні для дослідження класичних текстів та унікальних манускриптів. Не менш важливим досягненням було запровадження стандарту GBK (1993), який уперше забезпечив сумісність з ідеограмами Unicode 1.1, – при цьому відпала потреба у складних режимах кодування на кшталт ISO-2022 (див. табл. 1). Це означало, що китайські національні схеми кодування остаточно збіглися з міжнародним стандартом, і китайські ієрогліфи з того моменту можна було вільно використовувати в будь-яких поєднаннях і на всіх сучасних пристроях [Cook 2002].

Табл. 1. Еволюція стандартів кодування китайських символів (1980–2002)

Стандарт	Рік прийняття	Кількість символів	Тип символів	Географічне використання	Значення для цифрового китаєзнавства
GB 2312	1980	6763	Спрощені китайські ієрогліфи	Материковий Китай	Перший стандартизований набір для комп'ютерів
Big5	1984	13060	Традиційні китайські ієрогліфи	Тайвань, Гонконг	Стандарт де-факто для традиційних символів
GB/T 12345	1990	6825	Традиційні китайські ієрогліфи	Материковий Китай	Підтримка традиційних символів
Unicode 1.0	1991	20902	CJK базовий набір	Глобальний	Початок уніфікації CJK
GBK	1993	21886	Спрощені та традиційні китайські ієрогліфи	Материковий Китай	Сумісність із Unicode 1.1
Unicode 2.0	1996	38885	CJK + розширення	Глобальний	Розширення до історичних варіантів

Unicode 3.0	1999	49194	CJK уніфіковані	Глобальний	Стабільна основа для цифрових проєктів
Unicode 3.1	2001	94140	CJK Extension A	Глобальний	Підтримка рідкісних символів
Unicode 3.2	2002	95156	CJK Extension B	Глобальний	Масштабне розширення для історичних текстів

Паралельно з упровадженням стандартів кодування тривали інтенсивні роботи над уніфікацією правил структурної розмітки текстів. Поява мови XML у 1998 році зробила очевидним, що стандарт TEI (Text Encoding Initiative) потребує адаптації до нового формату даних. Саме тому у 2002 році за редакцією Лу Бернарда та Сіда Баумана було оприлюднено версію TEI P4, яка вперше забезпечила повну сумісність TEI із XML. Це стало визначною подією, що відкрила шлях до автоматизованого й масштабного опрацювання китайських текстів у різних цифрових системах [Sun, Xue, Qu, Wang, Sun].

Вирішення фундаментальних технічних питань підготувало підґрунтя для реалізації в цей період кількох справді амбітних проєктів, спрямованих на створення потужної цифрової інфраструктури для китаєзнавчих досліджень. Зокрема, у 1996 році тайванські вчені започаткували дві провідні ініціативи. По-перше, Інститут історії та філології Академії Сініка (Тайвань) розпочав проєкт “Архівні студії династій Мін та Цін” (“明清檔案工作室”), головною метою якого стали збереження, оцифрування й дослідження унікальної колекції документів цих історичних періодів. Особливий акцент був зроблений на цифровій обробці текстів і створенні докладних метаданих для кожного документа. Починаючи з 2001 року, каталоги та матеріали з метаданими стали доступними онлайн, а зареєстровані користувачі дістали можливість безплатно переглядати оцифровані зображення архівних документів [中央研究院歷史語言研究所 2025b]. По-друге, задля підтримки міжнародних досліджень буддизму була створена цифрова платформа “Цифрова бібліотека буддології Національного Тайванського університету” (“臺大佛學數位圖書館”), яка забезпечує безплатний доступ до великої колекції наукових матеріалів, цифрових текстів та різноманітних довідково-інформаційних інструментів. Цей унікальний ресурс охоплює понад 490 тисяч бібліографічних записів, близько 140 тисяч повнотекстових дослідницьких статей із буддизму 45 мовами (зокрема, санскрит, палі, тибетська, китайська, японська, англійська). База також містить понад 120 тисяч авторитетних імен буддійських авторів, предметний класифікатор, рубрикатор журналів, п’ятимовний тезаурус основних ключових слів, буддійський онлайн-музей, освітні матеріали та різноманітний мультимедійний контент [國立臺灣大學圖書館 2025].

У континентальному Китаї 1997 року за ініціативи Міністерства культури КНР було розпочато “Національний пілотний проєкт цифрової бібліотеки”, що знаменував початок епохи масового оцифрування книжкових і документальних фондів завдяки скоординованій співпраці бібліотекарів, інженерів та науковців. Підсумком цієї успішної ініціативи стало створення у 2005 році масштабного “Національного проєкту цифрової бібліотеки” [Yang Guihua 2002; Wei Liu 2003]. Водночас якісно нового рівня розвитку досягла “Біографічна база даних Китаю” (China Biographical Database, CBDB). Її витoki сягають колекції Роберта Гартвелла, сформованої ще в 1970–1980-х роках, але справжнього розквіту проєкт набув у 2004–2005 роках, коли завдяки зусиллям Майкла Фуллера відбулися його повна реструктуризація та переведення на сучасні бази даних. Саме тоді “Біографічна база даних Китаю” набула сучасний вебінтерфейс і стала незамінним ресурсом для біографічних та джерелознавчих досліджень [China Biographical Database Project 2025].

Важливим кроком у розвитку цифрового китаєзнавства стало започаткування під керівництвом професора Пітера К. Бола (Peter K. Bol) у 2001 році “Історико-географічної інформаційної системи Китаю” (China Historical Geographic Information System, CHGIS) як спільного проєкту Гарвардського університету (Harvard University, США) та Фуданського університету (复旦大学, КНР). Основною метою цього проєкту було створення деталізованої просторової бази даних, яка охоплює як сучасні, так і історичні географічні об’єкти – міста, поселення, адміністративні межі, що відповідають різним періодам історії Китаю. CHGIS дає змогу будувати інтерактивні історико-географічні карти, аналізувати динаміку зміни адміністративних кордонів, ідентифікувати давні місцевості, досліджувати історичну топоніміку, а також інтегрувати просторові дані з біографічними та іншими архівними джерелами. Платформа відкрита для всіх науковців і надає безкоштовний доступ до інструментів аналізу та візуалізації даних про Китай [China Historical GIS 2025].

Значущим внеском у розвиток цифрових гуманітарних ресурсів стала поява у 2001 році “Бази даних старих фотографій Тайваню” (“臺灣舊照片資料庫”), започаткованої Національним Тайванським університетом з метою збереження та репрезентації візуальної спадщини періоду японського управління островом. Відібрані для проєкту фотографії ретельно оцифровувалися, супроводжувалися докладними метаданими та пояснювальними підписами. На сьогодні база містить понад 40 тисяч одиниць – фотографій, ілюстрацій, альбомів та інших матеріалів. Особливу цінність ресурсу надає широкий тематичний спектр: освіта, охорона здоров’я, політика, економіка, географія, флора і фауна, соціум, культура, мова, література, подорожі тощо. Для користувачів передбачені можливості пошуку за анотованими даними, онлайн-перегляд зображень, завантаження метаданих і самих фотографій для наукових та освітніх завдань у межах академічного використання. База функціонує на сучасній платформі Omeka S, обладнана потужними інструментами пошуку й навігації. Ця база даних є ключовим ресурсом для дослідження історичних візуальних матеріалів Тайваню та широкого використання серед науковців, викладачів, студентів і всіх, хто цікавиться культурною спадщиною острова [國立臺灣大學圖書館數位典藏館 2025].

У 2004 році Академія Сініка (Тайвань) ініціювала створення “Цифрового архіву ворожильних кісток” (“甲骨文數位典藏”) – масштабного дослідницького проєкту, спрямованого на збереження та систематизацію унікальної колекції давньокитайських епіграфічних пам’яток. Архів містить понад 45 тисяч цифрових відбитків ворожильних кісток, переважна більшість яких походить з археологічних розкопок 1928–1937 років на території Іньсюй (殷墟) – руїн стародавньої столиці династії Шан у провінції Хенань. Колекцію доповнюють артефакти з приватних зібрань. Функціональні можливості архіву включають комплексну систему каталогізації з детальними метаданими кожного експоната: морфологічні характеристики, архівну нумерацію, високоякісні зображення. Особливо цінною є функція зіставлення інскрипцій з їхніми візуальними репрезентаціями, що значно полегшує компаративний аналіз. Для вирішення технічних проблем, пов’язаних із відображенням архаїчних ієрогліфічних форм, було розроблено спеціалізовану “Базу даних графем китайських ієрогліфів”, яка підтримує множинні варіанти написання та забезпечує коректну візуалізацію давніх текстів. Наукове значення проєкту полягає в структурованій цифровізації унікального сегмента китайської культурної спадщини, що створює неоціненний ресурс для міждисциплінарних досліджень. Архів сприяє інтеграції синології, історії, лінгвістики та археології, забезпечуючи дослідникам ефективні інструменти для встановлення кореляцій між зображеннями, текстами, архівними записами та фізичними артефактами. Проєкт являє собою один із найбільш системно організованих і технологічно досконалих цифрових

архівів ворожильних кісток у світі, що активно сприяє розвитку міжнародних наукових колаборацій та популяризації китайської епіграфіки [中央研究院歷史語言研究所 2004].

У 2006 році Дональд Стерджен започаткував “Китайський текстовий проєкт” (“Chinese Text Project”, “中國哲學書電子化計劃”) – відкриту онлайн-бібліотеку, яка акумулює класичні китайські тексти від доцінської епохи до республіканського періоду включно. Проєкт охоплює широкий спектр жанрів: філософські трактати, історичні твори, художню літературу, медичні й математичні тексти, словники та коментарі. Завдяки потужним технологічним рішенням платформа пропонує сучасні інструменти для роботи з джерелами: зручний багаторівневий пошук, інтерактивність, гіперпосилання в тексті й словникові ресурси, а також спеціалізовані функції для порівняльного аналізу й дослідження варіантів текстів. Користувачам доступні англо-китайський інтерфейс, підтримка традиційної та спрощеної китайської графіки, гнучкий вибір шрифтів і велика база рідковживаних ієрогліфів завдяки повній сумісності з Unicode (зокрема, CJK Extensions). База даних постійно оновлюється шляхом поповнення новими матеріалами, скануванням і розпізнаванням рукописів (OCR), а також завдяки внеску користувачів в уточнення перекладів та виправлення текстових і технічних помилок. Сьогодні “Китайський текстовий проєкт” є незамінним ресурсом для науковців, викладачів, студентів і всіх, хто займається класичною китайською літературою та гуманітарними дослідженнями [Chinese Text Project 2016; Sturgeon 2019].

Аналіз цифрових ініціатив цього періоду дає змогу виокремити низку ключових принципів, що лягли в основу сучасного цифрового Китаєзнавства. Ці принципи сформувалися як відповідь на комплекс технологічних та методологічних викликів, із якими зіткнулися дослідники під час опрацювання китайських текстів у цифровому середовищі. Провідним серед них став принцип стандартизації кодування, що забезпечив узгодженість та сумісність між різними системами та проєктами. Широке впровадження уніфікованих стандартів – насамперед Unicode та TEI – дало змогу створити універсальні підходи до кодування китайських символів і текстових структур, відкривши можливість безперешкодної передачі, обробки й використання китайських джерел на будь-яких цифрових платформах. Значну увагу було приділено принципу структурної розмітки тексту (XML і TEI), що дало змогу створювати семантично насичені цифрові репрезентації: кожен елемент тексту дістає власне значення і контекст, що істотно підвищує якість і глибину аналізу культурної спадщини. Створення реляційних баз даних забезпечило зручне зберігання великих масивів біографічної і текстової інформації, ефективний пошук та аналітичну обробку даних. Не менш фундаментальною стала ідея міжнародної співпраці: мультиінституційний та транснаціональний обмін цифровими ресурсами й експертизою сприяв формуванню потужної глобальної дослідницької мережі. Нарешті, принцип відкритого доступу, втілений у таких проєктах, як “Біографічна база даних Китаю” та “Китайський текстовий проєкт”, справив революційний вплив: уперше у світовій практиці був забезпечений вільний онлайн-доступ до унікальних китайських джерел для дослідників та всіх зацікавлених з усього світу, що раніше було майже неможливо [Chen Jing 2022, 25; De Weerdт 2020].

Формування цифрових архівів китайських текстів супроводжувалося цілим спектром як технічних, так і методологічних викликів. Однією з найскладніших проблем стала адаптація західних стандартів цифрової обробки тексту до специфіки китайської мови й письма: китайські ієрогліфи виявилися набагато складнішими для кодування, обробки й пошуку, ніж алфавітні системи, що потребувало створення нових технологічних підходів і розширення наявних стандартів [Chen Yongsheng, Su Huanning 2017]. Особливо гостро ці труднощі відчувалися на початку

2000-х років, коли обчислювальні ресурси були обмежені, що унеможливлювало масове оцифрування великих текстових корпусів. Серйозною проблемою залишався і брак уніфікованих стандартів для метаданих: різні архіви й колекції покладалися на власні схеми опису, ускладнюючи процеси інтеграції, пошуку та кросколекційного аналізу. Додаткові виклики створювали мовні й культурні бар'єри між дослідницькими центрами Сходу та Заходу, що виявлялося в різних підходах до організації даних, аналітики, а часто й у самому розумінні, яким має бути “правильний” цифровий архів. Це гальмувало ефективний науковий обмін та доступ до матеріалів. Не менш важливою залишалася і проблема інтелектуальної власності та авторського права – особливо в умовах, коли проекти прагнули забезпечити відкритий доступ до оцифрованих джерел. Саме тому розробка чітких, зрозумілих і технологічно універсальних стандартів для цифрових архівів залишалася одним із найбільш актуальних завдань галузі впродовж тривалого часу.

Розвиток цифрового китаєзнавства супроводжувався формуванням складної екосистеми взаємодії між численними установами різних типів. Визначну роль у цьому процесі відіграла Академія Сініка на Тайвані, яка у 2002 році приєдналася до Програми електронного навчання та цифрових архівів Тайваню (Taiwan e-Learning and Digital Archives Program), ставши провідним центром створення й збереження цифрових знань і культурної спадщини [Digital Taiwan... 2011; 中央研究院數位文化中心 2018]. Прикладом ефективною міжнародної співпраці стала партнерська взаємодія між Гарвардським університетом та Інститутом історії та філології Академії Сініка в рамках розробки “Біографічної бази даних Китаю”. Це стало зразком успішної транстихоокеанської кооперації й продемонструвало унікальний потенціал об'єднання інституційних ресурсів та експертизи для створення масштабних цифрових платформ у сфері китаєзнавчих досліджень [Tsui, Wang 2020; 李友仁, 宋迎春 2018, 54–58, 63].

Національна бібліотека Китаю започаткувала низку масштабних державних ініціатив, кульмінацією яких стало створення у 2005 році “Національного проєкту цифрової бібліотеки”. Це викликало значний суспільний резонанс, заклавши технологічну й інституційну основу для масового збереження і доступу до китайської культурної спадщини в цифровому форматі [Wei Dawei, Sun Yigang 2010]. Своєю чергою “Китайський текстовий проєкт” запровадив інноваційну модель краудсорсингу для корекції й анотування текстів, спираючись на підходи, вже відпрацьовані в проєктах на зразок електронної енциклопедії Вікіпедія (Wikipedia) та бібліотеки “Проект Гутенберг” (“Project Gutenberg”). Такий підхід суттєво прискорив створення якісних цифрових джерел, забезпечуючи їхню відкриту доступність і високу точність обробки. Важливим технічним проривом того часу стала розробка спеціалізованих систем оптичного розпізнавання символів (OCR) для китайських текстів, які враховували особливості каліграфії та широкий спектр історичних графічних варіантів, що суттєво підвищило якість та релевантність отриманих даних. Не менш значуще місце в розвитку цифрового китаєзнавства посідає створення уніфікованих стандартів метаданих для забезпечення міжсистемної сумісності та інтегрованого пошуку матеріалів. У 2002 році Національна бібліотека Китаю, Інститут науково-технічної інформації Китаю та Національна наукова бібліотека Китайської академії наук об'єднали зусилля у впровадженні “Проекту національних стандартів цифрових бібліотек”. Новостворена система була розроблена на основі міжнародних рекомендацій Dublin Core та MARC, водночас інтегруючи традиції й вимоги китайської бібліографічної практики. Встановлені стандарти стали фундаментом для подальших розробок і гармонізації цифрових архівів Китаю [Liu Shi, Zhao Wei, Li Feiyue 2024; Yue, Me, Yimin, Rahmat 2025].

Табл. 2. Проекти / ініціативи в цифровому китаєзнавстві (1996–2005)

Рік	Проект / ініціатива	Організація	Тип ініціативи
1996	RFC 1922 – кодування китайських ієрогліфів для інтернет-повідомлень	IETF/RFC	Стандартизація
1996	“Архівні студії династій Мін та Цін”	Інститут історії та філології Академії Сініка (Тайвань)	Інституційна ініціатива
1997–2001	Національний пілотний проект цифрової бібліотеки	Міністерство культури КНР	Державний проект
1998	XML-стандарт – основа для TEI	W3C	Технологічний стандарт
1999	Китайський експериментальний проект цифрової бібліотеки	Міністерство культури КНР	Державний проект
2001	“Історико-географічна інформаційна система Китаю”	Гарвардський університет (США), Фуданський університет (КНР).	Академічний проект
2001	“База даних старих фотографій Тайваню”	Національний Тайванський університет (Тайвань)	Академічний проект
2002	TEI P4 – XML-версія для китайських текстів	TEI Consortium	Стандартизація
2002	Академія Сініка – приєднання до “Програми електронного навчання та цифрових архівів Тайваню”	Академія Сініка (Тайвань)	Інституційна ініціатива
2002	Національна бібліотека Китаю, Інститут науково-технічної інформації Китаю, Національна наукова бібліотека Китайської академії наук	“Проект національних стандартів цифрових бібліотек”	Стандартизація
2004	“Цифровий архів ворожильних кісток”	Академія Сініка (Тайвань)	Академічний проект
2004–2005	Реструктуризація “Біографічної бази даних Китаю”	Гарвард-Яньцзін інститут (США/КНР)	Академічний проект
2005	Національний проект цифрової бібліотеки	Національна бібліотека Китаю	Державний проект
2006	“Китайський текстовий проект”	Дональд Стерджен	Приватна ініціатива

Отже, період 1996–2005 років став визначальним етапом у становленні цифрового китаєзнавства як окремої дисципліни. Саме тоді були закладені основоположні технологічні, методологічні та інституційні засади, які зумовили подальший динамічний розвиток галузі. Серед ключових досягнень цього десятиліття – стандартизація кодування китайських символів шляхом упровадження Unicode, створення XML-базованих стандартів текстової розмітки, а також запуск перших масштабних цифрових архівів. Важливим здобутком стало формування міжнародної мережі співпраці між дослідницькими центрами, що значно прискорило інтеграцію знань та технологій. Саме в цей час були вироблені і впроваджені базові принципи сучасного цифрового китаєзнавства: стандартизація, структурна розмітка текстів, реляційні бази даних, міжнародна співпраця, відкритий доступ, збереження метаданих,

масштабованість та сумісність форматів. Ці принципи залишаються надзвичайно актуальними й сьогодні. На їхній основі в наступні роки стрімко розвивалися цифрові ресурси, що залучили до співпраці академічні центри, технологічні компанії та урядові інституції різних країн для створення сучасної цифрової інфраструктури вивчення китайської мови та культури. Попри численні виклики, було доведено: цифрові технології можуть органічно інтегруватися в традиційне китаєзнавство, суттєво розширюючи його аналітичні та комунікаційні можливості.

Табл. 3. Базові принципи цифрового китаєзнавства

Принцип	Опис	Ключові проєкти	Технічні рішення	Виклики
Стандартизація кодування	Використання уніфікованих стандартів (Unicode, TEI) для забезпечення сумісності	RFC 1922, GBK, Unicode CJK	Unicode UTF-8, GB18030	Конфлікти між національними стандартами
Структурна розмітка тексту	Застосування XML та TEI для семантичної розмітки китайських текстів	TEI P4, “Китайський текстовий проєкт”	XML, XSLT, DTD схеми	Складність китайської граматики та семантики
Реляційні бази даних	Створення структурованих баз даних для біографічної та текстової інформації	“Біографічна база даних Китаю”, “Китайський текстовий проєкт”	SQL, Microsoft Access, dBase	Обмежені обчислювальні ресурси
Міжнародна співпраця	Співпраця між установами різних країн для обміну ресурсами	Гарвардський університет – Академія Сініка, “Біографічна база даних Китаю”	Інтернет-протоколи, HTTP, FTP	Мовні та культурні бар’єри
Відкритий доступ	Забезпечення вільного доступу до цифрових ресурсів для наукової спільноти	“Біографічна база даних Китаю”, “Китайський текстовий проєкт”	Web interfaces, API	Питання інтелектуальної власності
Збереження метаданих	Систематичне збереження описових даних про тексти та їхній контекст	Національні стандарти електронної бібліотеки	Dublin Core, MARC, кастомні схеми	Брак стандартизованих схем
Масштабованість	Розробка систем, здатних працювати з великими обсягами даних	Національний проєкт цифрової бібліотеки Китаю	Розподілені системи, клієнт-сервер	Технічні обмеження апаратного забезпечення
Сумісність форматів	Забезпечення сумісності між різними технічними платформами	Перехід на Unicode, рекомендації TEI	Крос-платформні бібліотеки	Швидкі зміни технологічних стандартів

3. Інституційна консолідація цифрового китаєзнавства: великі дослідницькі ініціативи та міжнародна співпраця (2006–2015)

Передусім варто підкреслити, що у 2006–2015 роках продовжували активно розвиватися низка ключових цифрових проєктів попереднього періоду. Одним із найуспішніших стала “Біографічна база даних Китаю”, що на 2015 рік переросла в найбільший у світі цифровий репозиторій біографічної інформації про історичних діячів Китаю (понад 400 тисяч персоналій). База “Стародавні китайські тексти” (CHANT) стала найбільшою колекцією цього типу, досягнувши понад 80 мільйонів записів. Важливий прецедент міжнародної співпраці у сфері цифрового збереження культурної спадщини задав “Міжнародний проєкт Дуньхуан”, який об’єднав ресурси

22 провідних установ із 12 країн світу та забезпечив відкритий доступ до інформації й зображень манускриптів, картин, текстилю й артефактів з Дуньхуана і Шовкового шляху. Ще суттєвішого розвитку галузь набула завдяки масштабній інституціоналізації цифрового китаєзнавства й ініціюванню значних міжнародних проєктів із цифровізації культурної спадщини Китаю в цей період.

Важливою віхою в інституалізації цифрового китаєзнавства стало проведення Тайванським університетом у 2009 році першого “Міжнародного семінару з цифрових колекцій та цифрової гуманітаристики” (“数字典藏与数字人文国际研讨会”). Захід охоплював такі теми, як створення баз даних, інтелектуальний аналіз тексту (*text mining*) і застосування соціальних мереж у гуманітарних дослідженнях. Станом на сьогодні цей семінар відбувся уже тринадцять разів і став найголовнішою міжнародною платформою цифрової гуманітаристики у Східній Азії [项洁主编 2011, 9–28]. У тому ж 2009 році в КНР відбулися значущі події: професор Уханського університету Ван Сяогун (王晓光) ґрунтовно представив вітчизняним науковцям концепцію та історію розвитку цифрової гуманітаристики, її сучасний стан і найкращі практики [王晓光 2010, 207–221]. Цього ж року була проведена перша конференція з цифрової гуманітаристики в КНР. У 2011 році при Уханському університеті відкрито перший дослідницький центр цифрової гуманітаристики в КНР – цей центр і дотепер залишається єдиним представником материкового Китаю в міжнародній мережі відповідних центрів. Того ж року аналогічний центр з’явився в Сіанському університеті, а вже у 2012 році – при Тайванському, Пекінському, Нанкінському університетах та Університеті Цінхуа. Пекінський університет також активно займається організацією наукових заходів у цій сфері, зокрема з 2016 року проводить щорічний форум “Цифрова гуманітаристика”, що сприяє подальшій консолідації цієї дослідницької галузі.

Серед нових ініціатив першого десятиліття XXI століття ключове місце належить “Базі даних місцевих історичних документів Китаю” (“地方历史文献数据库”). Її створення стартувало у 2006 році з першої фази польових досліджень і збору матеріалів командою історичного факультету Шанхайського транспортного університету – насамперед у селищі Шицан, провінція Чжецзян (石仓村, 浙江省松阳县). Протягом 2006–2013 років фахівці Центру місцевої документації не лише поширили географію пошуку, а й активно викупували, копіювали, реставрували та оцифровували рідкісні архіви сільських громад і родинних колекцій із провінцій Аньхой, Чжецзян, Цзянсі, Фуцзянь, Цзянсу, Хунань, Гуандун, Юньнань, Ляонін, Шаньсі, Хубей і Хебей. Архів охоплює найрізноманітніші типи джерел: договори, бухгалтерські книги, податкові документи, адміністративні й судові матеріали, медичні протоколи, освітні та релігійні джерела, родові генеалогії, ділову кореспонденцію, локальні енциклопедії, спеціалізовані колекції (серед яких – унікальні судові справи з офіційними печатками). Загальний обсяг ресурсів перевищує 350 тисяч одиниць (понад 1,5 мільйона сторінок). Усі документи зберігають високу якість завдяки реставрації, ретельному скануванню, перевірці тексту й докладному структуруванню. База даних підтримує багатовимірний повнотекстовий пошук з анотованими та точно промаркованими текстами. Повнотекстові розшифровки додатково перевірялися науковцями-спеціалістами для забезпечення академічної надійності ресурсу. “База даних місцевих історичних документів Китаю” є найбільшим у країні повнотекстовим архівом оцифрованих рукописних джерел, зібраних з різних регіонів, та найбільш комплексною колекцією локальних матеріалів для вивчення історії, економіки, права, сімейних структур і повсякдення періоду Мін, Цін, Китайської Республіки й КНР. Вона стала унікальним ресурсом для дослідників соціально-економічних процесів та традиційного життя Китаю [地方历史文献数据库 2025].

У 2007 році під керівництвом професора Ху Цзяюя (胡家瑜) з кафедри антропології Національного Тайванського університету, у партнерстві з Британським музеєм,

а також низкою провідних європейських та американських музеїв, було започатковано масштабний проєкт “Тайванські етнологічні колекції в закордонних музеях” (“海外博物馆台湾民族学藏品资料”). Його мета – створення повного електронного каталогу артефактів, пов’язаних із матеріальною культурою острова середини XIX – початку XX ст., які зберігаються у США, Великій Британії, Канаді, Німеччині та інших країнах. На цифровому ресурсі реалізовано пошукову систему за колекціями та типами предметів, інтегровано базу фотографій, детальних описів та оцифрованих архівів. Це значно підвищило доступність і наукову видимість розпорошених по світу артефактів, сприяючи етнологічним, історичним та музейним дослідженням [海外博物馆台湾民族学藏品资料 2025]. Ще одним важливим проєктом став запуск у 2009 році “Цифрової бібліотеки історії Тайваню” (“台湾历史数位图书馆”) командою Центру цифрових гуманітарних досліджень Національного Тайванського університету. Цей ресурс акумулює понад 100 тисяч історичних документів, зокрема адміністративні й династичні архіви (“淡新檔案”), документи Мін і Цін (“明清臺灣行政檔案”), давні договори (“古契書”) та інші джерела. Бібліотека підтримує повнотекстовий пошук, аналіз даних і метаданих, а також візуалізацію зв’язків між джерелами. Окремі цифрові інструменти дають змогу працювати з хронологіями, посадами, здійснювати наукову бібліографічну обробку й автоматичне формування цитат. Загалом проєкт істотно покращив доступ дослідників до джерел з історії Тайваню, сприяв розвитку цифрових гуманітарних технологій і підтримці міждисциплінарних досліджень [台湾历史数位图书馆 2025].

У 2010 році Нанкінський університет започаткував створення “Бази даних комерційної реклами Китаю” (“中国商业广告数据库”) – унікального спеціалізованого онлайн-архіву для збереження та комплексного вивчення історичної комерційної реклами Китаю кінця XIX – першої половини XX століття. Цей ресурс акумулює понад 18 700 високоякісних оцифрованих зображень рекламних оголошень, кожне з яких супроводжується масивом із 44 метаданих (зокрема, інформацією про місце, час, тип реклами та інші параметри). Всі зображення були отримані й оцифровані з провідних комерційних газет п’яти ключових договірних портових міст: “Шеньбао” (“申报”, Шанхай), “Дагунбао” (“大公报”, Тяньцзінь), “Шенцзін шибао” (“盛京时报”, Шеньян), “Ханькоу чжунсібао” (“汉口中西报”, Ханькоу), “Юехуабао” (“越华报”, Гуанчжоу). Оригінали документів зберігаються в найбільших бібліотеках Китаю (Національна бібліотека Китаю, Шанхайська бібліотека, Гуандунська бібліотека Сунь Ятсена, Бібліотека Університету Цінхуа) та в Азійському мовознавчому відділі бібліотеки Університету Вашингтона (Сієтл, США). Зображення були отримані з мікрофільмів, оцифровані, детально структуровані й відкриті для комп’ютерного та міждисциплінарного аналізу. Архів дає змогу переглядати, аналізувати й порівнювати рекламу різних періодів, тематик і регіонів. Проєкт істотно розширив доступ дослідників – істориків, культурологів, соціологів, лінгвістів і фахівців цифрової гуманітаристики – до великої колекції джерел з історії матеріальної, візуальної й соціальної культури Китаю тієї епохи [南京大学中国商业广告数据库 2025].

У 2016 році Нанкінський університет розпочав проєкт просторової історії – “Історична географія та інформатизація міста Цзянькан за часів Шести династій” (“六朝建康历史地理信息化建设”). Його мета – створення комплексної спеціалізованої бази даних і WebGIS-системи для дослідження історії Нанкіна в епоху Шести династій (IV–VI ст.). Проєкт охоплює введення, редагування, аналіз і візуалізацію просторових і текстових даних, застосовуючи передові цифрові й комп’ютерні технології. Система містить: повнотекстовий пошук по “Правдивих записах про Цзянькан” (“建康实录”); інформаційну підсистему з 1401 історичним топонімом; базу археологічних об’єктів (253 пам’ятки, 320 наукових публікацій, 822 мультимедійних матеріали); колекцію зображень артефактів (понад 1366 фотографій і зображень археологічних об’єктів); 3D-модулі демонстрації кам’яних скульптур із могил.

Для створення цієї системи були задіяні такі технології, як WebGIS, мультимедійна інтеграція, сканування документів, Microsoft Visual Studio 2013, ArcGIS Server 10.2, Silverlight, архітектура MVC, Baidu Maps API, B/S-архітектура, JavaScript та ActiveX. Реалізований програмний комплекс дає змогу використовувати сучасні методи просторового аналізу, інтерактивної візуалізації й міждисциплінарної комунікації, що відкриває нові можливості для історико-картографічних досліджень і цифрової гуманітаристики в контексті Китаю [南京大学艺术与文化创意平台 2018].

У 2018 році Пекінський університет ініціював інноваційний проєкт – “Картографування систематизованих записів про вчених та їхні вчення періоду династій Сун і Юань” (“宋元学案知识图谱系统”). Ця вебсистема перетворила 2,4 мільйона ієрогліфів фундаментальної праці з історії китайської філософії – “Розгляд шкіл династій Сун та Юань” (“宋元学案”) – на масштабний семантичний граф знань із розвиненими можливостями візуалізації, семантичного пошуку та “дистанційного читання” класичного тексту. Проєкт, очолюваний професором Ван Цзюнем (王军, нар. 1968) та командою студентів і випускників лабораторії KVLab (Лі Сяюй, Лю Шужань, Бянь Еньхуа та ін.), розкрив основні сутності (діячі, дати, географічні локуси, твори та складні розгалужені зв'язки між ними). Побудований граф знань дає змогу візуалізувати багатовимірні інтелектуальні та історичні зв'язки, здійснювати інтерактивний аналіз за параметрами (час, місце, особистість, твір), виконувати семантичний пошук та аналітику. Система містить інформацію про 87 інтелектуальних шкіл Китаю та понад 2 тисячі видатних учених, дає змогу простежити рівень популярності шкіл у різні періоди, механізми інтелектуального “спадкування” – еволюцію, саморозвиток, розгалуження шкіл тощо [宋元学案知识图谱 2025]. Продовження цих досліджень вийшло на рівень міжнародної співпраці: у 2020 році професор Ван Цзюнь очолив спільний проєкт Пекінського та Гарвардського університетів “Дослідження наукової спадщини династії Сун за допомогою графіка знань” (“用知识图谱探索宋代士人师承关系”) [KVision Research Group 2018]. Ще одним результатом цієї роботи стала розробка окремої “Китайської платформи візуалізації академічного наставництва” (“中国学术师承可视化平台”), яка дає змогу детально простежувати мережі академічної спадкоємності [北京大学 KVision 数字图书馆实验室 2025].

Аналіз ключових цифрових проєктів у сучасному китаєзнавстві свідчить про різноманіття моделей міжнародної співпраці, що використовуються для створення й розвитку відповідних ресурсів. До основних типів кооперації належать *трилатеральна модель* (наприклад, “Біографічна база даних Китаю” – США, Тайвань, КНР), *мультинаціональні консорціуми* (“Міжнародний проєкт Дуньхуан”), *європейські платформи* (“Дослідницька інфраструктура для вивчення Євразії” [RISE Project 2019]), а також *двостороння* канадсько-тайванська співпраця (“База даних жіночих творів епохи Мін-Цін” [Ming Qing Women's Writings 2005]). Транснаціональна інтеграція наукових, технічних та організаційних зусиль супроводжувалася істотними технологічними інноваціями, особливо у сфері обробки, анотації та аналізу китайськомовних текстів. Порівняльний аналіз обсягів провідних баз даних демонструє значну варіативність масштабів: від 15 тисяч записів у “Базі даних жіночих творів епохи Мін-Цін” до понад 80 мільйонів у репозитарії “Китайський текстовий проєкт”. Це спектральне різноманіття свідчить про галузеву спеціалізацію окремих цифрових ініціатив та орієнтацію кожного проєкту на конкретні дослідницькі завдання й аудиторії.

Окремої уваги заслуговує поява спеціалізованих цифрових інструментів для аналізу та обробки китайських текстів. Зокрема, у 2012–2014 роках було створено інструмент MARKUS – платформу для автоматичного маркування широкого спектра сутностей у китайських та корейських текстах. MARKUS виконує автоматичне

розпізнавання і тегування персональних імен, географічних назв, дат, бюрократичних посад, а також інших іменованих сутностей; підтримує перевірку посилань на довідкові інформаційні ресурси під час читання документа; дає змогу порівнювати фрагменти, де згадуються ті самі імена чи ключові слова; забезпечує експорт промаркованих уривків тексту разом із посиланнями на біографічні, топонімічні та інші бази даних для подальшої роботи в зовнішніх аналітичних інструментах і платформах. Розробка MARKUS здійснювалася в рамках проєкту “Комунікація та імперія: китайські імперії в порівняльній перспективі”, що фінансувався Європейською дослідницькою радою (European Research Council, ERC), а також іншими науковими грантами і фондами [Chinese Empires Digital Humanities 2025].

Розвиток цифрового китаєзнавства у 2006–2015 роках позначився як визначальний етап становлення галузі, що охопив низку ключових процесів. Відбулася стандартизація термінології, закладено інституційну й технічну інфраструктуру, а також розпочато реалізацію нових освітніх програм із цифрової гуманітаристики. Активно провадилось оцифрування текстів, удосконалювалися стандарти метаданих, а стрімке зростання числа наукових публікацій свідчило про розширення спектра дослідницьких напрямів у цій царині. Накопичений у цей період досвід створив міцне підґрунтя для сучасного цифрового китаєзнавства, продемонструвавши вагомий позитивний потенціал міжнародної співпраці у сфері збереження і вивчення культурної спадщини Китаю за допомогою новітніх технологій. Водночас етап 2006–2015 років був позначений низкою проблем і викликів: мовні бар’єри, що обмежували ефективність комунікації у цифрових гуманітарних проєктах, а також питання мультилінгвальності і культурної інклюзивності у глобальному контексті. Це зумовило нагальну потребу подальшого розвитку технологій семантичної обробки даних та їхньої інтеграції в єдину інформаційну інфраструктуру галузі.

4. Сучасний етап розвитку цифрової гуманітаристики в китаєзнавстві: нові технології та дослідницькі парадигми (2016–2024)

На сучасному етапі цифрова гуманітаристика в китаєзнавстві зазнала суттєвої трансформації, що привело до глибоких змін у методології дослідження китайської культури, історії та літератури. Оцінка еволюції цієї галузі засвідчує стрімке зростання – як у кількісному, так і в якісному вимірі, що докорінно змінює наукові підходи до роботи з інформаційними ресурсами та аналізу корпусів даних. Із 2016 року спостерігається постійне і стабільне збільшення кількості наукових публікацій, цифрових проєктів і спеціалізованих конференцій. Це вказує не лише на активний процес інституціоналізації цифрового китаєзнавства, а й на його поступове визнання як окремої академічної дисципліни в наукових колах. Особливо показовим є те, що за цей період відбувся якісний стрибок: сфера досліджень у Китаї увійшла у фазу прискореного розвитку та комплексної інституціональної інтеграції [赵薇 2021].

Період із 2016-го до 2024 року став визначальним для цифрової гуманітаристики в китаєзнавстві, адже саме в цей час були створені численні спеціалізовані центри й інститути, що займаються цією дисципліною. За спостереженнями Чжао Вей, із 2016 року цифрова гуманітаристика в Китаї увійшла у фазу стрімкого розвитку та інституціоналізації: у цій сфері суттєво зросла участь дослідників-гуманітаріїв, сформувалася власна “методологічна спільнота”, а розмежування між розбудовою мережевої інфраструктури та власне цифровими гуманітарними дослідженнями стало ще більш чітким і структурованим [赵薇 2021, 191–206, 232–233]. Значною віхою став початок активного розвитку цифрової гуманітаристики в Уханському університеті, який із 2016 року відіграє провідну роль у галузі. У травні того ж року Пекінський університет організував перший національний форум із цифрової гуманітаристики, що зібрав близько двохсот учасників [Tsui 2016]. Окремо слід відзначити внесок Університету Цінхуа, який заснував перший у материковому Китаї

спеціалізований журнал з цифрової гуманітаристики (“數字人文” / “Digital Humanities”). Видання виходить із 2019 року щоквартально, китайською та англійською мовами, ставши головною академічною платформою для обміну результатами та інноваціями в цій сфері [清华大学 2025].

Сучасний етап розвитку цифрової гуманітаристики в китаєзнавстві характеризується значною диверсифікацією технологічних підходів та зростанням їхньої зрілості. Серед найбільш розвинених інструментів слід виділити оптичне розпізнавання символів (OCR) для китайських текстів, а також методи машинного навчання, які забезпечили автоматизацію і вдосконалення аналізу великих текстових корпусів. Водночас розвиток технологій є нерівномірним: традиційні підходи на кшталт OCR та алгоритмів машинного навчання досягли високого рівня технологічної зрілості, тоді як новітні рішення на основі штучного інтелекту, як-от ChatGPT, лише почали впроваджуватись у цифрове китаєзнавство з 2023 року [Henriot 2024]. Особливої уваги заслуговує прогрес у сфері методів обробки природної мови для китайських текстів: саме рівень і масштаби автоматизованої комп'ютерної обробки лінгвістичної інформації стали визначальним критерієм переходу до нового витка інформаційного суспільства та підвищення якості цифрової гуманітаристики в китаєзнавстві.

Формуються багаторівневі екосистеми спеціалізованих платформ і цифрових інструментів, кожен із яких має свою унікальну функціональну специфіку та орієнтований на різні дослідницькі аудиторії. Це дає змогу комплексно охоплювати і підтримувати різні напрямки галузі – від текстологічного аналізу до геоінформаційних досліджень і корпусної лінгвістики. Аналіз динаміки розвитку цифрових платформ демонструє їхню поступову еволюцію з помітним прискоренням після 2016 року. Однією з найпопулярніших і найбільш відвідуваних платформ сучасного китаєзнавства є “Китайський текстовий проєкт” (Chinese Text Project), який наразі обслуговує понад 25 тисяч активних користувачів та пропонує альтернативну модель створення й модерування повнотекстових матеріалів класичного китайського канону [Sturgeon 2019]. Водночас наймасштабнішою повнотекстовою базою китайських історичних джерел залишається тайванська платформа “Скрипта Сініка” [中央研究院數位文化中心 2014; 中央研究院歷史語言研究所 2025a], яка є головним інституційним центром для збереження, оцифрування та аналізу класичних текстів Китаю.

Розвиток технологій оптичного розпізнавання символів (OCR) для китайських текстів демонструє значну різноманітність, що залежить насамперед від типу й складності оброблюваного матеріалу. Найвищих показників точності OCR-системи досягають при роботі із сучасними друкованими китайськими текстами, однак автоматичне розпізнавання історичних документів і надалі залишається серйозною науково-технічною проблемою. Серед найскладніших надзавдань – обробка числової системи сучжоуських цифр (蘇州碼子 *сучжоу мацзи*), яка широко застосовувалася в комерційній практиці домодерного Китаю. Попри її значну культурну й історичну цінність, сучжоуські цифри й досі не мають достатньої підтримки в сучасних OCR-стандартах, що суттєво ускладнює цифровізацію архівних торгово-економічних документів. Важливим кроком уперед у галузі стала розробка системи визначення порядку читання (Reading Order Detection, ROD), здатної автоматично відновлювати послідовність розташування фрагментів стародавніх текстів. ROD використовує принципи, натхнені людським зоровим сприйняттям, і аналізує візуальні підказки в цифрових зображеннях для реконструкції логіки прочитання джерела, що значно поліпшило якість і повноту досліджень історичних китайських документів [Ma, Huang, Liu 2024, 23808–23810].

Важливим напрямком сучасної цифрової гуманітаристики в китаєзнавстві стало формування просторової історії Китаю, у межах якої активно розвиваються низка

масштабних проєктів із залученням найновіших ГІС-технологій і міждисциплінарної співпраці:

1) “Історико-географічна інформаційна система Китаю”. У 2016 році ця система досягла вже шостої версії. Вона забезпечує інституційну основу для просторового аналізу історичних подій, територіального розвитку, адмінподілу Китаю і створює платформу для інтеграції різних історико-географічних ресурсів.

2) “Китайська цивілізація в часі та просторі”. Проєкт Академії Сініка (Тайвань), орієнтований на підтримку міждисциплінарних досліджень та інтеграцію історико-просторових даних із найдавніших часів до епохи Цін. Система охоплює: базові просторові дані (основна карта розроблена на основі історичного атласу, створеного за редакцією китайського історика, одного з основоположників китайської історичної географії Таня Цісяна (譚其驤, 1911–1992), та різноманітних картографічних і супутникових матеріалів); інтегроване WebGIS-середовище для роботи з картографічними шарами; тематичну інформацію – дослідницькі матеріали, відкриті дані, історичні атласи. Важливо, що ресурс інтегрований із провідними цифровими репозитаріями (“Скрипта Сініка”, система електронних документів китайської літератури (汉籍电子文献系统), база кон’юнктури цін та каталог місцевих хронік) і слугує розвинутою платформою для просторово-часового аналізу дослідників, викладачів та широкої аудиторії [中央研究院 2025].

3) “Відкрита платформа з історії та географії Шовкового шляху” (Пекінський університет). Цифровий ресурс, який збирає дані про Шовковий шлях протягом понад двох тисячоліть – від династій Хань і Тан до сучасності. Платформа акумулює інформацію про середовище, торгівлю, транспорт, етноси, релігії, культурну взаємодію тощо. Реалізовано засоби аналізу, WebGIS-візуалізацію, моделювання, завантаження даних; платформа орієнтована на дослідників та освітян, але наразі фрагментарно відкрита для зовнішніх користувачів.

4) “Геоінформаційна система Шовкового шляху” (Фуданський університет, під керівництвом професора Хоу Янфана (侯杨方)). Проєкт став першим у світі прикладом комплексної експедиції з точного просторового відтворення маршрутів історичного Шовкового шляху (маршрути подорожей буддійських монахів, експедиції давніх і сучасних дослідників тощо) на основі низки джерел – історичних документів, польових досліджень, археології тощо. Ресурс дає змогу користувачам досліджувати маршрути, порівнювати їх із сучасними комунікаціями, а також долучатися до віртуальних реконструкцій стародавніх локацій [复旦大学].

Отже, просторово-часові ГІС-рішення стали одним із найперспективніших векторів цифрового китаєзнавства, значно розширивши аналітичні, дослідницькі й освітні можливості для різних груп користувачів.

Визначним інституційним центром формування та просування цифрової гуманітаристики у сфері китаєзнавства став Національний Тайванський університет, де було ініційовано понад 28 спеціалізованих проєктів у цій царині. Ключовим досягненням університету є створення та запуск у 2018 році потужної цифрової платформи DocuSky (“數位人文學術研究平台”), що стала інноваційним рішенням для підтримки академічних та міждисциплінарних досліджень у галузі цифрової гуманітаристики. DocuSky орієнтована на потреби науковців у зручному й ефективному зборі, організації, аналізі та зберіганні індивідуальних дослідницьких матеріалів. Платформа надає інструменти для форматування, маркування та конвертації текстів, а також для завантаження їх до власної бази даних із подальшим використанням у дослідженнях. Важливими функціональними можливостями DocuSky є текстовий аналіз, побудова різноманітних візуалізацій, застосування ГІС-інструментів для просторового аналізу, а також підтримка роботи з багатьма форматами даних. Платформа підтримує багатомовність (китайська, англійська, корейська мови), що значно підвищує рівень міжнародної співпраці й відкритість для дослідників у

різних країнах. У 2025 році була анонсована нова, розширена версія – DocuSky-Plus із моделлю платного доступу й академічними ліцензіями для груп та індивідуальних користувачів. Отже, DocuSky істотно підвищує рівень цифрових досліджень текстів та адаптована до різних методів і дослідницьких завдань сучасного китаєзнавства [數位人文學術研究平台 2025].

Корпорація “Tencent”, що належить до найбільших інвестиційних компаній КНР і світу, здійснює інтенсивний розвиток власних рішень у галузі штучного інтелекту та надає системну підтримку проектам цифровізації культурної спадщини. Серед таких ініціатив варто відзначити запущену у 2024 році “Платформу для співпраці дослідження ворожильних кісток за допомогою штучного інтелекту” (“甲骨文AI协同平台”) – спеціалізований інструмент для дослідження давньокитайської писемності. Застосування технологій штучного інтелекту забезпечує суттєве підвищення ефективності й точності розпізнавання знаків, що становить одну з ключових методологічних проблем у дослідженні цих епіграфічних пам’яток. Платформа пропонує комплексний інструментарій для наукових досліджень: 1) *високоякісна візуалізація матеріалів*: багатошарові зображення, компаративний аналіз, покращення деталізації для наукових досліджень та реконструкцій; 2) *пошукові та аналітичні функції*: швидкий пошук потрібних знаків з отриманням інформації про їхню репрезентацію в корпусі написів, стилістичні варіанти, фонетичні та семантичні характеристики, діахронічну еволюцію написання; *міжнародна інтеграція*: об’єднання наукових напрацювань музеїв, університетів та архівних ініціатив у сфері китайської писемності, створених дослідниками з КНР, Республіки Корея та США [数字甲骨共创中心 2024].

Період 2020–2024 років став ключовим, переломним етапом у розвитку цифрової гуманітаристики в китаєзнавстві, головною особливістю якого стало широке впровадження технологій штучного інтелекту. Надзвичайно помітним є зростання застосування сучасних мовних моделей, зокрема Bidirectional Encoder Representations from Transformers (BERT), а з 2023 року – й ChatGPT як платформи для генерації тексту, аналізу, автоматичного тегування і смислової інтерпретації китайських текстових масивів. Один із провідних напрямків – використання глибокого навчання для проведення сентиментального аналізу китайської літератури. Для цього була розроблена модель BERT-LLSTM-DL, яка поєднує мовну репрезентацію на основі BERT, послідовне навчання на базі LSTM (Long Short-Term Memory) та глибоке навчання для виділення основних ознак тексту, що забезпечує високу точність автоматизованого аналізу [Shen Xiaohui 2024, 1565–1574]. Ще одним перспективним підходом у галузі є концепція Lean AI, що передбачає впровадження відповідальних та демократичних алгоритмів штучного інтелекту в цифровому китаєзнавстві, з акцентом на зниженні залежності від великих технологічних компаній. Lean AI має особливе значення для роботи з історичними чи маловживаними мовами, які традиційно висвітлені недостатньо в комерційних модельних рішеннях, – цей напрям відкриває нові можливості для розширення й інклюзії в цифрових дослідженнях Китаю [Henriot 2024].

Краудсорсинг утвердився як один із найефективніших інструментів для цифровізації великих колекцій китайських текстів, даючи змогу залучати широкую спільноту волонтерів і дослідників до обробки історичних документів та розшифровки рукописних джерел. Одним із найяскравіших прикладів успішної реалізації цієї методології є проєкт “Транскрибувати Шена” (“Transcribing Sheng”), що став першою масштабною краудсорсинговою ініціативою у сфері цифрової культурної спадщини Китаю. Мета цього проєкту полягає у транскрибуванні архіву записів відомого підприємця, адміністратора й педагога Шена Сюаньхуая (盛宣懷, 1844–1916), а також оцифруванні близько 100 мільйонів слів історичних документів, що охоплюють

період з 1850-го до 1936 року. Високу ефективність і сталість застосування краудсорсингових моделей демонструє й “Китайський текстовий проєкт”, учасники якого здійснюють до 100 редагувань щодоби, що сприяє постійному зростанню індексу оцифрованих, відкритих для науковців та широкої публіки текстів китайського класичного корпусу [Sturgeon 2019].

Розвиток цифрової гуманітаристики у китаєзнавстві значною мірою зумовлений розгалуженим міжнародним співробітництвом, оскільки численні інноваційні проєкти мають глобальний і комплексний характер, об’єднуючи дослідників, інституції й лабораторії з різних країн та континентів. Особливу роль у цьому процесі відіграє партнерство між китайськими та західними науковими центрами, завдяки якому значно активізуються обмін знаннями, кооперація у розробці цифрових методологій, а також реалізація масштабних міждисциплінарних і мультиінституційних ініціатив. Характерним прикладом успішної глобальної співпраці є проєкт “Дослідницька інфраструктура для вивчення Євразії” (“Research Infrastructure for the Study of Eurasia”), що пропонує інноваційний підхід до формування гнучких, розподілених цифрових інфраструктур з метою забезпечення ефективної взаємодії, інтеграції даних та спільного наукового прогресу в царині гуманітаристики [Wang, Belouin, Chen, Ho].

Одним із ключових викликів для сучасної цифрової гуманітаристики в китаєзнавстві залишається обмежений доступ до даних. Навіть порівняно “нескладні” інформаційно-культурні ресурси – книги й архівні матеріали доцінського Китаю – нерідко є частково або повністю недоступними через обмеження в доступі, проприєтарність баз даних та поширення платних моделей підписки. Вказані бар’єри суттєво ускладнюють повноцінну реалізацію принципу відкритої науки. Попри помітний прогрес у цій галузі, серйозною проблемою залишаються й технічні труднощі OCR для історичних документів. До найскладніших викликів належать варіативність рукописних почерків і стилів друку, нерівний контраст між текстом і папером, перетікання тексту на зворотні сторінки, а також складні чи незвичні макети документів, що значно знижує якість автоматичного розпізнавання й потребує додаткових досліджень для їхнього подолання [Ma, Huang, Liu 2024].

Майбутній розвиток цифрової гуманітаристики в китаєзнавстві тісно пов’язаний з інтеграцією великих мовних моделей і технологій штучного інтелекту, які відкривають нові можливості для аналізу й інтерпретації китайських текстів [Henriot 2024]. Одним із перспективних напрямків є створення спеціалізованих інструментів, орієнтованих на обробку різних типів китайських текстів і документів з урахуванням їхньої форматної та історико-культурної специфіки. Особливої уваги заслуговує розробка суттєвих наборів даних для глибокого навчання, зокрема DeepJiandu – інноваційний корпус для розпізнавання символів на бамбукових та дерев’яних планках, який став важливим кроком у дослідженні та цифровізації стародавніх китайських текстів, що має суттєвий вплив на розширення можливостей автоматичного аналізу таких унікальних культурних артефактів [Liu, Zhang, Qi 2025; Wu, Qi, Li, Wang 2018, 3245].

Аналіз розвитку цифрової гуманітаристики в китаєзнавстві за період 2016–2024 років засвідчує глибоку трансформацію галузі – від поодиноких емпіричних експериментів до становлення як повноцінної академічної дисципліни з інституалізованою інфраструктурою та згуртованою міжнародною спільнотою. До визначальних ознак цього періоду належать експоненціальне зростання кількості проєктів та публікацій, диверсифікація технологічних підходів, а також формування потужних мереж транснаціонального співробітництва між дослідниками, лабораторіями та науковими центрами. На 2024 рік найрозвиненішими технологіями залишаються OCR для китайських текстів та класичні методи машинного навчання, проте новітні

підходи на основі штучного інтелекту та великих мовних моделей лише починають набирати обертів і мають величезний потенціал для галузі. Водночас важливою тенденцією є активізація краудсорсингу і громадської участі в цифровізації культурної спадщини, що прискорює темпи оцифрування і розширює доступ до джерел. Подальший прогрес у цій царині безпосередньо пов'язаний із впровадженням великих мовних моделей, розробкою спеціалізованих інструментів для роботи з різними типами китайських текстів і розв'язання проблем відкритості й доступу до даних. Сучасні дослідження підкреслюють, що цифрова гуманітаристика є потужним каталізатором розвитку китаєзнавства, відкриваючи принципово нові можливості для застосування інноваційних методів, формулювання нових дослідницьких питань і здійснення важливих наукових відкриттів.

У травні 2022 року Канцелярія ЦК Комуністичної партії Китаю та Канцелярія Державної ради КНР опублікували “Положення про просування реалізації національної стратегії цифровізації культури” (“关于 推进实施国家文化数字化战略的意见”). Документ окреслює цілі до кінця 14-ї п'ятирічки, серед яких – створення базової інфраструктури та сервісних платформ для цифровізації культури, а також формування системи надання культурних послуг, що інтегрує онлайн та офлайн формати. До 2035 року планується розгорнути національну систему великих культурних даних, яка забезпечить механізми фізичного розподілу, логічних зв'язків, швидкого доступу, ефективного пошуку, повного обміну та інтеграції ключових культурних даних. Головна мета цієї системи – представити панораму китайської культури та зробити її цифрові здобутки повністю доступними для всіх громадян. Вказані заходи створять сприятливі умови для подальшого розвитку цифрового китаєзнавства, стимулюючи впровадження нових технологічних рішень і розширення цифрових ресурсів у цій науковій сфері [经济形势报告网 2022].

Проблематика розвитку цифрової гуманітаристики в китаєзнавстві залишається надзвичайно актуальною і багатовимірною. Попри наявний прогрес, ця сфера й досі характеризується комплексом викликів, які умовно поділяються на технологічні, лінгвістичні, етичні та методологічні. Найбільшою перешкодою є складність роботи з китайськими текстами – специфіка ієрогліфічної писемності, де немає явних меж між словами, обумовлює потребу попереднього формування корпусів і словників. Без цього сучасні алгоритми аналізу часто виявляються неефективними. Ситуація ще більше ускладнюється в разі роботи з давніми джерелами: брак уніфікованих схем кодування символів суттєво гальмує процес їхньої цифрової обробки. Навіть найсучасніші системи розпізнавання ієрогліфів демонструють недостатній рівень точності сегментації (близько 84,7 %), що для академічних досліджень є критично низьким показником [Mo Henghui, Wei Linjing 2024, 3480]. Додатковою проблемою є брак достатньо анотованих корпусів, які б охоплювали різні історичні періоди та діалекти китайської мови, що обмежує застосування сучасних методів обробки природної мови для аналізу стародавніх текстів. Недосконалість або брак адаптованого програмного забезпечення для пошуку, аналізу й візуалізації китайських культурно-історичних матеріалів значно ускладнюють проведення міждисциплінарних і транс-регіональних досліджень [Diao Xiaolei, Zhou Zhihan, Shi Lida 2025]. Особливої уваги заслуговують питання доступу до архівних джерел, що часто обмежується цензурою, політичними заборонами й політикою закритості даних у КНР. Це призводить до суттєвої нерівності між китайськими та закордонними дослідниками. Етичний аспект охоплює питання приватності, дотримання авторських прав та особливості роботи з персональними архівами, що створює додаткові моральні дилеми для дослідників. Варто також зазначити, що значний розвиток інфраструктури та створення баз даних не завжди супроводжуються належною якістю змістовного аналізу. Це може призводити до поверхового осмислення складних культурних чи літературних феноменів. Надмірна орієнтація на цифрові інструменти без критичного осмислення

їхніх обмежень потенційно нівелює цінність класичних методів гуманітарного аналізу [Armand 2024]. З огляду на викладене подолання зазначених проблем потребує інтеграції класичних і цифрових підходів, розширення доступу до ресурсів, вдосконалення лінгвістичних та етичних стандартів, а також активізації міжнародної співпраці в межах наукової спільноти.

Висновки

Аналіз розвитку цифрової гуманітаристики в китаєзнавстві за останні 40 років (1984–2024) чітко засвідчує еволюцію цієї галузі від початкових комп'ютерних експериментів до повноцінної зрілої академічної дисципліни з розвиненою методологією, стійкою інституційною інфраструктурою й сучасними технологічними інструментами.

– Перший етап (1984–1995) визначається роллю першопрохідців: почали застосовуватись комп'ютерні технології для обробки китайських текстів, розроблялися стандарти кодування ієрогліфів та створювалися перші електронні бази даних, що заклало фундамент для подальшої еволюції галузі.

– Другий етап (1996–2005) характеризується формуванням ключових принципів цифрового китаєзнавства й стандартизацією підходів до цифрової обробки джерел. З'явилися масштабні проекти на зразок “Біографічної бази даних Китаю” та “Історико-географічної інформаційної системи Китаю”, що збіглися з інституційною інтеграцією і створенням перших спеціалізованих дослідницьких центрів.

– Третій етап (2006–2015) – це період інституціоналізації та стрімкого розвитку: відбувається стандартизація термінології, закладається інституційна й технічна інфраструктура, розширюється міжнародна співпраця. Впроваджуються масштабні міжнародні проекти цифровізації, що довели ефективність цифрових методів для вивчення китайської історії та культури.

– Сучасний етап (2016–2024) відзначається значною технологічною диверсифікацією: активно впроваджуються технології штучного інтелекту та машинного навчання, розвивається краудсорсинг, з'являються спеціалізовані цифрові платформи, а новітні мовні моделі й технології глибокого навчання значно спрощують аналіз китайських текстів та відкривають нові горизонти для дослідників.

Подальші перспективи пов'язані з інтеграцією великих мовних моделей і штучного інтелекту для аналізу класичних текстів, розробкою спеціалізованих інструментів для оцифрування рукописних та епіграфічних матеріалів. Особливого значення набуває розвиток відкритих цифрових репозиторіїв і метаданих, що розширить відкритість і повторне використання даних. Упровадження концепції *Lean AI* забезпечить відповідальніший і демократичний підхід до штучного інтелекту, критично важливий для роботи з історичними мовами й культурними текстами. Також розвиток цифрової гуманітаристики у сфері китаєзнавства супроводжується низкою комплексних викликів. Зокрема, значними є технологічні труднощі, пов'язані з обробкою китайськомовних текстів, а також потреба у впровадженні новітніх підходів до аналізу культурної спадщини Китаю. Крім того, актуальними залишаються етичні та політико-правові питання, що потребують окремої уваги. Для ефективного подолання зазначених проблем потрібна інтеграція сучасних технологічних інновацій із класичними методологіями китаєзнавчих досліджень, з акцентом на міждисциплінарному підході. У цьому контексті стратегічними завданнями є розробка спеціалізованих інструментів для гуманітарного аналізу, формування анотованих корпусів китайськомовних текстів, а також підготовка фахівців, здатних працювати на перетині традиційного китаєзнавства та цифрових технологій. Реалізація таких заходів можлива за умови консолідованих зусиль академічної спільноти. Це дасть змогу подолати наявні обмеження й забезпечити сталий розвиток цифрової гуманітаристики в китаєзнавчих студіях.

Отже, цифрова гуманітаристика в Китаєзнавстві за чотири десятиліття перетворилася з експериментальної галузі на невіддільну частину дослідницької діяльності. Вона не лише поглиблює вивчення культури та історії Китаю, а й упроваджує нові методологічні парадигми, відкриваючи якісно нові можливості для розвитку гуманітарних наук у цифрову епоху.

ЛІТЕРАТУРА

- Armand C. (2024), “ReOrienting Digital History: New Frontiers in Chinese Studies”, *Journal of Digital History*, Vol. 3, No. 3. DOI: <https://doi.org/10.1515/jdh-2024-0001>
- Chen Jing (2022), “One Among Many: Digital Humanities in China”, *IAS Newsletter*, No. 92, p. 25.
- Chen Yongsheng and Su Huanning (2017), “The Value and Problems of Digital Preservation for Historical Documents in China”, *Proceedings from the Document Academy*, Vol. 4, Iss. 2, Article 14. DOI: <https://doi.org/10.35492/docam/4/2/14> (accessed June 21, 2025).
- China Biographical Database Project* (2025), [online], available at: <https://projects.iq.harvard.edu/cbdb/home> (accessed June 21, 2025).
- China Historical GIS* (2025), [online], available at: <https://chgis.fas.harvard.edu/> (accessed June 23, 2025).
- Chinese Empires Digital Humanities* (2025), “MARKUS – a Free Online MARKUP System”, [online], available at: <https://dh.chinese-empires.eu/markus/beta/index.html> (accessed July 22, 2025).
- Chinese Text Project* (2016), [online], available at: <https://ctext.org/> (accessed June 21, 2025).
- Cook R. (2002), “CJKV Unified Ideographs Extension C”, in *22nd International Unicode Conference (IUC22) Unicode and the Web: Evolution or Revolution? September 9–13, 2002, San Jose, California* <http://www.unicode.org/iuc/iuc22/>, [online], available at: <https://linguistics.berkeley.edu/~rscook/pdf/IUC22-328.pdf> (accessed June 21, 2025).
- Crispin M. (2022), “RFC 1922: Chinese Character Encoding for Internet Messages”, in *IETF Datatracker*, [online], available at: <https://datatracker.ietf.org/doc/rfc1922/> (accessed June 21, 2025).
- De Weerd H. (2020), “Creating, Linking, and Analyzing Chinese and Korean Datasets: Digital Text Annotation in MARKUS and COMPARATIVUS”, *Journal of Chinese History*, Vol. 4, Issue 2, pp. 519–527. DOI: <https://doi.org/10.1017/jch.2020.23>
- Diao Xiaolei, Zhou Zhihan, Shi Lida et al. (2025), “InteChar: A Unified Oracle Bone Character List for Ancient Chinese Language Modeling”, in *Semanticscholar*, available at: <https://www.semanticscholar.org/reader/9fe6376eb0bc7f91b3f8733ade39e6b93397cd97> (accessed June 12, 2025).
- Digital Taiwan – Culture & Nature* (2011), [online], available at: <https://culture.teldap.tw/culture/index.php> (accessed June 22, 2025).
- Henriot C. (2024), “Digital Humanities, AI, and Asian Studies”, in *Elites, Networks and Power in Modern China*, [online]. DOI: <https://doi.org/10.58079/12rhi>
- International Dunhuang Programme* (2025), [online], available at: <https://idp.bl.uk/> (accessed June 20, 2025).
- Kurzynski M. (2025), “Global DH vs. Area Studies: Rethinking ‘China and the West (Rest)’ Beyond Lenticular Logic”, in *The Digital Orientalist*, 23 May, [online], available at: <https://digitalorientalist.com/2025/05/23/global-dh-vs-area-studies-questioning-the-china-and-the-west-rest-paradigm/> (accessed June 5, 2025).
- KVision Research Group* (2018), “KVision CBDB_KG: Explore Academic Lineage in Song Dynasty with Knowledge Graph”, [online], available at: <https://cbdb-refinder.pkudh.org/index-en.html> (accessed June 22, 2025).
- Liu Q. (2011), *Machine Translation in China*, [online], available at: [https://liuquncn.github.io/html_public_casict/research/publications/Machine%20Translation%20in%20China%20\(Qun%20Liu\)%202011116final.pdf](https://liuquncn.github.io/html_public_casict/research/publications/Machine%20Translation%20in%20China%20(Qun%20Liu)%202011116final.pdf) (accessed July 19, 2025).
- Liu Shi, Zhao Wei and Li Feiyue (2024), “Digital Humanities in China, 1980–2020”, in *World Humanities Report*, CHCI 2024, pp. 1–12, [online], available at: https://worldhumanitiesreport.org/wp-content/uploads/2024/07/Chi_8.3_Digital-Humanities-in-China-1980–2020.pdf (accessed May 28, 2025).

Liu Y., Zhang Q., Qi Y. et al. (2025), “DeepJiandu Dataset for Character Detection and Recognition on Jiandu Manuscript”, *Scientific Data*, Vol. 12, Iss. 398, [online]. DOI: <https://doi.org/10.1038/s41597-025-04716-3>

Ma H.-Y., Huang H.-H. and Liu C.-L. (2024), “Reading between the Lines: Image-Based Order Detection in OCR for Chinese Historical Documents”, *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38, Iss. 21, pp. 23808–23810. DOI: <https://doi.org/10.1609/aaai.v38i21.30572>

MARKUS (2025), [online], available at: <https://dh.chinese-empires.eu/markus/beta/index.html> (accessed June 23, 2025).

Ming Qing Women's Writings (2005), [online], available at: <https://digital.library.mcgill.ca/mingqing/english/index.php> (accessed June 22, 2025).

Mo Henghui and Wei Linjing (2024), “Fine Segmentation of Chinese Character Strokes Based on Coordinate Awareness and Enhanced BiFPN”, *Sensors*, No. 11, p. 3480. DOI: <https://doi.org/10.3390/s24113480>

Qian D. (2005), “Prospects of Machine Translation in the Chinese Context”, *Meta*, Vol. 50, No. 4, [online]. DOI: <https://doi.org/10.7202/019854ar>

RISE Project (2019), [online], available at: <https://rise.mpiwg-berlin.mpg.de/> (accessed June 22, 2025).

Shen Xiaohui (2024), “Sentiment Analysis of Modern Chinese Literature Based on Deep Learning”, *Journal of Electrical Systems*, Vol. 20 No. 6s, pp. 1565–1574. DOI: <https://doi.org/10.52783/jes.3075>

Sturgeon D. (2019), “Chinese Text Project: A Dynamic Digital Library of Premodern Chinese”, *Digital Scholarship in the Humanities*, Vol. 34, Issue Supplement 1, pp. 794–801. DOI: <https://doi.org/10.1093/llc/fqz046>

Sun L., Xue S., Qu W., Wang X. and Sun Y. (n.d.), “Constructing of a Large-Scale Chinese-English Parallel Corpus”, in *The ACL Anthology*, [online], available at: <https://aclanthology.org/W02-1211.pdf> (accessed June 21, 2025).

Sun Wei (2005), “The Development of China Digital Library and Its Influence on China and the World”, *Chinese Librarianship: An International Electronic Journal*, No. 20, [online], available at: <https://www.white-clouds.com/iclc/cliej/cl20sun.htm> (accessed June 21, 2025).

Tsui L. H. (2016), “The Digital Humanities as an Emerging Field in China”, in *Harvard University*, [online], available at: <https://projects.iq.harvard.edu/chinesecbdb/digital-humanities-emerging-field-china> (accessed June 23, 2025).

Tsui L. H. and Chen J. (2019), “Defining and Debating Digital Humanities in China: New or Old?”, in *The Index of Digital Humanities Conferences*, [online], available at: <https://dh-abstracts.library.virginia.edu/works/10098> (accessed June 10, 2025).

Tsui L. H. and Wang H. (2020), “Harvesting Big Biographical Data for Chinese History: The China Biographical Database (CBDB)”, *Journal of Chinese History*, Vol. 4, Iss. 2, pp. 505–511, DOI: <https://doi.org/10.1017/jch.2020.21>

Wang S., Belouin P., Chen, S.-P. and Ho I. B. H., *Research Infrastructure for the Study of Eurasia (RISE): Towards a flexible and distributed digital infrastructure for resource access via standardized APIs and metadata*, [online], available at: https://pure.mpg.de/rest/items/item_3033461_9/component/file_3033494/content (accessed June 23, 2025).

Wei Dawei and Sun Yigang (2010), “The National Digital Library Project”, *D-Lib Magazine*, Vol. 16, No. 5/6, [online]. DOI: <https://doi.org/10.1045/may2010-dawei>

Wei Liu (2003), “The New Development of Digital Libraries in China”, in *CiteSeerx*, [online], available at: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=43b703c2669a87fb9387b89eac15c9098b21c5ff> (accessed June 21, 2025).

Wu T., Qi G., Li C. and Wang M. (2018), “A Survey of Techniques for Constructing Chinese Knowledge Graphs and Their Applications”, *Sustainability*, Vol. 10, No. 9, p. 3245. DOI: <https://doi.org/10.3390/su10093245>

Yang Guihua (2002), “The Development of the China Digital Library”, *Electronic Journal of Academic and Special Librarianship*, Vol. 3, No. 3, [online], available at: https://southernlibrarianship.icaap.org/content/v03n03/Yang_g01.htm (accessed June 21, 2025).

Yu Hui (2025), “A Retrospective View of Chinese Contributions to Digital Humanities in Computational Guqin Music Research”, *Computer Music Journal*, [online]. DOI: <https://doi.org/10.1162/comj.a.4>

Yue M., Me R. C., Yimin L. and Rahmat R. W. O. K. (2025), “Developing a Metadata Elements Set for Digital Archives of Chinese Restoration Techniques: A Scoping Literature Review”, *Multidisciplinary Reviews*, Vol. 8, Iss. 8, p. 2025243. DOI: <https://doi.org/10.31893/multi-rev.2025243>

Zhen Xihui (2010), “Overview of Digital Library Development in China”, *D-Lib Magazine*, Vol. 16, No. 5/6, [online]. DOI: <https://doi.org/10.1045/may2010-zhen>

北京大学KVision数字图书馆实验室 (2025), “中国学术师承可视化平台”, [在线], 可获于: <https://nav.pkudh.org/project/rio-furniture-4/> (浏览日期: 2025年7月22日)。

地方历史文献数据库 (2025), [在线], 可获于: <https://ndfwx.datahistory.cn/> (浏览日期: 2025年7月18日)。

董振东 (1984), “逻辑语义及其在机译中的应用”, 在 *中国的机器翻译*, 刘涌泉编, 知识出版社, 北京, 第25–45页。

冯志伟 (1987), *自动翻译*, 上海知识出版社, 上海。

复旦大学 (无日期), “丝绸之路地理信息系统”, [线上], 可获于: <https://silkroad.fudan.edu.cn/project.html> (浏览日期: 2025年7月18日)。

耿元骊 (2009), “三十年来中国古籍数字化研究综述 (1979–2009)”, 第二届中国古籍数字化国际学术研讨会论文集, 五洲传播出版社, 北京, 第17–33页。

國立臺灣大學圖書館 (2025), “臺大佛學數位圖書館”, [線上], 可獲於: <https://buddhism.lib.ntu.edu.tw/> (瀏覽日期: 2025年7月23日)。

國立臺灣大學圖書館數位典藏館 (2025), “臺灣舊照片資料庫” [線上], 可獲於: <https://dl.lib.ntu.edu.tw/s/photo/page/Home> (瀏覽日期: 2025年7月23日)。

海外博物馆台湾民族学藏品资料 (2025), [線上], 可獲於: <http://tcbfm.digital.ntu.edu.tw/main.php> (瀏覽日期: 2025年7月23日)。

漢達文庫 (2025), “漢達文庫 – 個人帳號 – 登入”, [線上], 可獲於: <https://www.chant.org/clogin> (瀏覽日期: 2025年7月18日)。

经济形势报告网 (2022), “关于推进实施国家文化数字化战略的意见” (全文), 5月23日, [線上], 可获于: <http://www.china-cer.com.cn/zhengcefagui/2022052318622.html> (浏览日期: 2025年7月18日)。

李友仁, 宋迎春 (2018), “北美与西欧的数字人文中国研究状况论析”, *山东社会科学*, 第7号, 第54–58+63页. DOI: 10.14112/j.cnki.37-1053/c.2018.07.007

刘涌泉 (1984), *中国的机器翻译*, 知识出版社, 北京。

南京大学艺术与文化创意平台 (2018), “六朝建康历史地理信息化建设”, [線上], 可获于: <https://digitalhumanities.nju.edu.cn/en/projects/5d9d9cb39b3b5c8f52/> (浏览日期: 2025年7月23日)。

南京大学中国商业广告数据库 (2025), [線上], 可获于: <https://ccaa.nju.edu.cn/html/index.html> (浏览日期: 2025年7月22日)。

清华大学 (2025), “数字人文 [Digital Humanities]”, [線上], 可获于: <https://szrw.cbpt.cnki.net/WKC/WebPublication/index.aspx?mid=SZRW> (浏览日期: 2025年7月22日)。

數位人文學術研究平台 (2025), [線上], 可獲於 <https://docusky.org.tw/DocuSky/home/> (瀏覽日期: 2025年7月19日)。

数字甲骨共创中心 (2024), “甲骨文AI协同平台”, [線上], 可获于: <https://www.jgwlbq.org.cn/home> (浏览日期: 2025年7月23日)。

宋元学案知识图谱 (2025), [線上], 可获于: <https://syxa.pkudh.org/> (浏览日期: 2025年7月22日)。

台灣歷史數位圖書館 (2025), [線上], 可獲於: <http://thdl.ntu.edu.tw/>, (瀏覽日期: 2025年7月19日)。

王晓光 (2010), ““数字人文”的产生、发展与前沿”, 在 *方法创新与哲学社会科学* 发展, 武汉大学出版社, 武汉, 第207–221页。

项洁主编 (2011), *从保存到创造: 开启数字人文研究*, 台湾大学出版中心, 台北。

赵薇 (2021), “数字时代人文学研究的变革与超越——数字人文在中国”, *探索与争鸣*, 第6号, 第191–206+232–233页。

中央研究院 (2025), “中華文明之時空基礎架構--計畫簡介”, [線上], 可獲於: <https://ccts.sinica.edu.tw/intro.html> (瀏覽日期: 2025年7月18日)。

中央研究院歷史語言研究所 (2004), “甲骨文數位典藏”, [線上], 可獲於: <https://rub.ihp.sinica.edu.tw/~oracle/home.htm> (瀏覽日期: 2025年7月23日)。

中央研究院歷史語言研究所 (2025a), “漢籍全文資料庫”, [線上], 可獲於: <https://hanchi.ihp.sinica.edu.tw/ihp/hanji.htm> (瀏覽日期: 2025年7月18日)。

中央研究院歷史語言研究所 (2025b), “明清檔案工作室”, [線上], 可獲於: <https://newarchive.ihp.sinica.edu.tw/> (瀏覽日期: 2025年7月18日)。

中央研究院數位文化中心 (2014), “臺灣漢學數位典藏資源導覽”, [線上], 可獲於: https://sinology.ascdc.sinica.edu.tw/index_en.html (瀏覽日期: 2025年7月18日)。

中央研究院數位文化中心 (2018), “中央研究院數位典藏”, [線上], 可獲於: <https://sinica.digitalarchives.tw/> (瀏覽日期: 2025年7月18日)。

REFERENCES

Armand C. (2024), “ReOrienting Digital History: New Frontiers in Chinese Studies”, *Journal of Digital History*, Vol. 3, No. 3. DOI: <https://doi.org/10.1515/jdh-2024-0001>

Chen Jing (2022), “One Among Many: Digital Humanities in China”, *IIAS Newsletter*, No. 92, p. 25.

Chen Yongsheng and Su Huanning (2017), “The Value and Problems of Digital Preservation for Historical Documents in China”, *Proceedings from the Document Academy*, Vol. 4, Iss. 2, Article 14. DOI: <https://doi.org/10.35492/docam/4/2/14> (accessed June 21, 2025).

China Biographical Database Project (2025), [online], available at: <https://projects.iq.harvard.edu/cbdb/home> (accessed June 21, 2025).

China Historical GIS (2025), [online], available at: <https://chgis.fas.harvard.edu/> (accessed June 23, 2025).

Chinese Empires Digital Humanities (2025), “MARKUS – a Free Online MARKUP System”, [online], available at: <https://dh.chinese-empires.eu/markus/beta/index.html> (accessed July 22, 2025).

Chinese Text Project (2016), [online], available at: <https://ctext.org/> (accessed June 21, 2025).

Cook R. (2002), “CJKV Unified Ideographs Extension C”, in *22nd International Unicode Conference (IUC22) Unicode and the Web: Evolution or Revolution? September 9–13, 2002, San Jose, California* <http://www.unicode.org/iuc/iuc22/>, [online], available at: <https://linguistics.berkeley.edu/~rscCook/pdf/IUC22-328.pdf> (accessed June 21, 2025).

Crispin M. (2022), “RFC 1922: Chinese Character Encoding for Internet Messages”, in *IETF Datatracker*, [online], available at: <https://datatracker.ietf.org/doc/rfc1922/> (accessed June 21, 2025).

De Weerdt H. (2020), “Creating, Linking, and Analyzing Chinese and Korean Datasets: Digital Text Annotation in MARKUS and COMPARATIVUS”, *Journal of Chinese History*, Vol. 4, Issue 2, pp. 519–527. DOI: <https://doi.org/10.1017/jch.2020.23>

Diao Xiaolei, Zhou Zhihan, Shi Lida et al. (2025), “InteChar: A Unified Oracle Bone Character List for Ancient Chinese Language Modeling”, in *Semanticscholar*, available at: <https://www.semanticscholar.org/reader/9fe6376eb0bc7f91b3f8733ade39e6b93397cd97> (accessed June 12, 2025).

Digital Taiwan – Culture & Nature (2011), [online], available at: <https://culture.teldap.tw/culture/index.php> (accessed June 22, 2025).

Henriot C. (2024), “Digital Humanities, AI, and Asian Studies”, in *Elites, Networks and Power in Modern China*, [online]. DOI: <https://doi.org/10.58079/12rhi>

International Dunhuang Programme (2025), [online], available at: <https://idp.bl.uk/> (accessed June 20, 2025).

Kurzynski M. (2025), “Global DH vs. Area Studies: Rethinking ‘China and the West (Rest)’ Beyond Lenticular Logic”, in *The Digital Orientalist*, 23 May, [online], available at: <https://digitalorientalist.com/2025/05/23/global-dh-vs-area-studies-questioning-the-china-and-the-west-rest-paradigm/> (accessed June 5, 2025).

KVision Research Group (2018), “KVision CBDB_KG: Explore Academic Lineage in Song Dynasty with Knowledge Graph”, [online], available at: <https://cbdb-relfinder.pkudh.org/index-en.html> (accessed June 22, 2025).

Liu Q. (2011), *Machine Translation in China*, [online], available at: [https://liuquncn.github.io/html_public_casict/research/publications/Machine%20Translation%20in%20China%20\(Qun%20Liu\)%2020111116final.pdf](https://liuquncn.github.io/html_public_casict/research/publications/Machine%20Translation%20in%20China%20(Qun%20Liu)%2020111116final.pdf) (accessed July 19, 2025).

Liu Shi, Zhao Wei and Li Feiyue (2024), “Digital Humanities in China, 1980–2020”, in *World Humanities Report*, CHCI 2024, pp. 1–12, [online], available at: https://worldhumanitiesreport.org/wp-content/uploads/2024/07/Chi_8.3_Digital-Humanities-in-China-1980–2020.pdf (accessed May 28, 2025).

Liu Y., Zhang Q., Qi Y. et al. (2025), “DeepJiandu Dataset for Character Detection and Recognition on Jiandu Manuscript”, *Scientific Data*, Vol. 12, Iss. 398, [online]. DOI: <https://doi.org/10.1038/s41597-025-04716-3>

Ma H.-Y., Huang H.-H. and Liu C.-L. (2024), “Reading between the Lines: Image-Based Order Detection in OCR for Chinese Historical Documents”, *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38, Iss. 21, pp. 23808–23810. DOI: <https://doi.org/10.1609/aaai.v38i21.30572>

MARKUS (2025), [online], available at: <https://dh.chinese-empires.eu/markus/beta/index.html> (accessed June 23, 2025).

Ming Qing Women's Writings (2005), [online], available at: <https://digital.library.mcgill.ca/mingqing/english/index.php> (accessed June 22, 2025).

Mo Henghui and Wei Linjing (2024), “Fine Segmentation of Chinese Character Strokes Based on Coordinate Awareness and Enhanced BiFPN”, *Sensors*, No. 11, p. 3480. DOI: <https://doi.org/10.3390/s24113480>

Qian D. (2005), “Prospects of Machine Translation in the Chinese Context”, *Meta*, Vol. 50, No. 4, [online]. DOI: <https://doi.org/10.7202/019854ar>

RISE Project (2019), [online], available at: <https://rise.mpiwg-berlin.mpg.de/> (accessed June 22, 2025).

Shen Xiaohui (2024), “Sentiment Analysis of Modern Chinese Literature Based on Deep Learning”, *Journal of Electrical Systems*, Vol. 20 No. 6s, pp. 1565–1574. DOI: <https://doi.org/10.52783/jes.3075>

Sturgeon D. (2019), “Chinese Text Project: A Dynamic Digital Library of Premodern Chinese”, *Digital Scholarship in the Humanities*, Vol. 34, Issue Supplement 1, pp. 794–801. DOI: <https://doi.org/10.1093/lc/fqz046>

Sun L., Xue S., Qu W., Wang X. and Sun Y. (n.d.), “Constructing of a Large-Scale Chinese-English Parallel Corpus”, in *The ACL Anthology*, [online], available at: <https://aclanthology.org/W02-1211.pdf> (accessed June 21, 2025).

Sun Wei (2005), “The Development of China Digital Library and Its Influence on China and the World”, *Chinese Librarianship: An International Electronic Journal*, No. 20, [online], available at: <https://www.white-clouds.com/iclc/cliej/cl20sun.htm> (accessed June 21, 2025).

Tsui L. H. (2016), “The Digital Humanities as an Emerging Field in China”, in *Harvard University*, [online], available at: <https://projects.iq.harvard.edu/chinesebdb/digital-humanities-emerging-field-china> (accessed June 23, 2025).

Tsui L. H. and Chen J. (2019), “Defining and Debating Digital Humanities in China: New or Old?”, in *The Index of Digital Humanities Conferences*, [online], available at: <https://dh-abstracts.library.virginia.edu/works/10098> (accessed June 10, 2025).

Tsui L. H. and Wang H. (2020), “Harvesting Big Biographical Data for Chinese History: The China Biographical Database (CBDB)”, *Journal of Chinese History*, Vol. 4, Iss. 2, pp. 505–511, DOI: <https://doi.org/10.1017/jch.2020.21>

Wang S., Belouin P., Chen, S.-P. and Ho I. B. H., *Research Infrastructure for the Study of Eurasia (RISE): Towards a flexible and distributed digital infrastructure for resource access via standardized APIs and metadata*, [online], available at: https://pure.mpg.de/rest/items/item_3033461_9/component/file_3033494/content (accessed June 23, 2025).

Wei Dawei and Sun Yigang (2010), “The National Digital Library Project”, *D-Lib Magazine*, Vol. 16, No. 5/6, [online]. DOI: <https://doi.org/10.1045/may2010-dawei>

Wei Liu (2003), “The New Development of Digital Libraries in China”, in *CiteSeerx*, [online], available at: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=43b703c2669a87fb9387b89eac15c9098b21c5ff> (accessed June 21, 2025).

Wu T., Qi G., Li C. and Wang M. (2018), “A Survey of Techniques for Constructing Chinese Knowledge Graphs and Their Applications”, *Sustainability*, Vol. 10, No. 9, p. 3245. DOI: <https://doi.org/10.3390/su10093245>

Yang Guihua (2002), “The Development of the China Digital Library”, *Electronic Journal of Academic and Special Librarianship*, Vol. 3, No. 3, [online], available at: https://southernlibrarianship.icaap.org/content/v03n03/Yang_g01.htm (accessed June 21, 2025).

Yu Hui (2025), “A Retrospective View of Chinese Contributions to Digital Humanities in Computational Guqin Music Research”, *Computer Music Journal*, [online]. DOI: <https://doi.org/10.1162/comj.a.4>

- Yue M., Me R. C., Yimin L. and Rahmat R. W. O. K. (2025), "Developing a Metadata Elements Set for Digital Archives of Chinese Restoration Techniques: A Scoping Literature Review", *Multidisciplinary Reviews*, Vol. 8, Iss. 8, p. 2025243. DOI: <https://doi.org/10.31893/multi-rev.2025243>
- Zhen Xihui (2010), "Overview of Digital Library Development in China", *D-Lib Magazine*, Vol. 16, No. 5/6, [online]. DOI: <https://doi.org/10.1045/may2010-zhen>
- Beijing daxue KVision shuzi tushuguan shiyanshi* (2025), "Zhongguo xueshu shicheng keshi-hua pingtai", [online], available at: <https://nav.pkudh.org/project/rio-furniture-4/> (accessed July 22, 2025). (In Chinese).
- Difang lishi wenxian shujuku* (2025), [online], available at: <https://ndfwx.datahistory.cn/> (accessed July 18, 2025). (In Chinese).
- Dong Zhendong (1984), "Luoji yuyi ji qi zai jiyizhong de yingyong", in *Zhongguo de jiqi fanyi*, Liu Yongquan (ed.), Zhishi chubanshe, Beijing, pp. 25–45. (In Chinese).
- Feng Zhiwei (1987), *Zidong fanyi*, Shanghai zhishi chubanshe, Shanghai. (In Chinese).
- Fudan daxue* (s.a.), "Sichou zhi lu dili xinxi xitong", [online], available at: <https://silkroad.fudan.edu.cn/project.html> (accessed July 18, 2025). (In Chinese).
- Geng Yuanli (2009), "Sanshi nianlai zhongguo guji shuzihua yanjiu zongshu (1979–2009)", in *Di er jie zhongguo guji shuzihua guoji xueshu yantao hui lunwen ji*, Wuzhou chuanbo chubanshe, Beijing, pp. 17–33. (In Chinese).
- Guoli taiwan daxue tushuguan* (2025), "Taida foxue shuwei tushuguan", [online], available at: <https://buddhism.lib.ntu.edu.tw/> (accessed July 23, 2025). (In Chinese).
- Guoli Taiwan daxue tushuguan shuwei diancangguan* (2025), "Taiwan jiu zhaopian ziliaoku" [online], available at: <https://dl.lib.ntu.edu.tw/s/photo/page/home> (accessed July 23, 2025). (In Chinese).
- Haiwai bowuguan Taiwan minzu xue cangpin ziliao* (2025), [online], available at: <http://tcbfm.digital.ntu.edu.tw/main.php> (accessed July 23, 2025). (In Chinese).
- Han da wenku* (2025), "Han da wenku – geren zhanghao – dengru", [online], available at: <https://www.chant.org/clogin> (accessed July 18, 2025). (In Chinese).
- Jingji xingshi baogao wang* (2022), "Guanyu tuijin shishi guojia wenhua shuzihua zhanlue de yijian", May 23, [online], available at: <http://www.china-cer.com.cn/zhengcefagui/2022052318622.html> (accessed July 18, 2025). (In Chinese).
- Li Youren and Song Yingchun (2018), "Beimei yu xi'ou de shuzi renwen zhongguo yanjiu zhuangkuang lunxi", *Shandong shehui kexue*, No. 7, pp. 54–58+63. (In Chinese). DOI: 10.14112/j.cnki.37-1053/c.2018.07.007
- Liu Yongquan (1984), *Zhongguo de jiqi fanyi*, Zhishi chubanshe, Beijing. (In Chinese).
- Nanjing daxue yishu yu wenhua chuangxin pingtai* (2018), "Liuchao jian kang lishi dili xinxi hua jianshe", [online], available at: <https://digitalhumanities.nju.edu.cn/en/projects/5d9d9cb39b3b5c8f52/> (accessed July 23, 2025). (In Chinese).
- Nanjing daxue zhongguo shangye guanggao shujuku* (2025), [online], available at: <https://ccaa.nju.edu.cn/html/index.html> (accessed July 22, 2025). (In Chinese).
- Qinghua daxue* (2025), "Shuzi renwen [Digital Humanities]", [online], available at: <https://szrw.cbpt.cnki.net/WKC/WebPublication/index.aspx?mid=SZRW> (accessed July 22, 2025). (In Chinese).
- Shuwei renwen xueshu yanjiu pingtai* (2025), [online], available at: <https://docusky.org.tw/DocuSky/home/> (accessed July 19, 2025). (In Chinese).
- Shuzi jiagu gongchuang zhongxin* (2024), "Jiaguwen AI xietong pingtai", [online], available at: <https://www.jgwlbq.org.cn/home> (accessed July 23, 2025). (In Chinese).
- Songyuanxue an zhishi tupu* (2025), [online], available at: <https://syxa.pkudh.org/> (accessed July 22, 2025). (In Chinese).
- Taiwan lishi shuwei tushuguan* (2025), [online], available at: <http://thdl.ntu.edu.tw/> (accessed July 19, 2025). (In Chinese).
- Wang Xiaoguang (2010), "'Shuzi renwen' de chansheng, fazhan yu qianyan", in *Fangfa chuangxin yu zhexue shehui kexue fazhan*, Wuhan daxue chubanshe, Wuhan, pp. 207–221. (In Chinese).
- Xiang Jiezu bian (2011), *Cong baocun dao chuangzao: kaiqi shuzi renwen yanjiu*, Taiwan daxue chuban zhongxin, Taipei. (In Chinese).

Zhao Wei (2021), “Shuzi shidai renwenxue yanjiu de biange yu chaoyue—shuzi renwen zai Zhongguo”, *Tansuo yu zhengming*, No. 6, pp. 191–206+232–233. (In Chinese).

Zhongyang yanjiuyuan (2025), “Zhonghua wenming zhi shikong jichu jiagou – ji hua jianjie”, [online], available at: <https://ccts.sinica.edu.tw/intro.html> (accessed July 28, 2025). (In Chinese).

Zhongyang yanjiuyuan lishi yuyan yanjiusuo (2004), “Jiaguwen shuwei diancang”, [online], available at: <https://rub.ihp.sinica.edu.tw/~oracle/home.htm> (accessed July 23, 2025). (In Chinese).

Zhongyang yanjiuyuan lishi yuyan yanjiusuo (2025a), “Hanji quanwen ziliao ku”, [online], available at: <https://hanchi.ihp.sinica.edu.tw/ihp/hanji.htm> (accessed July 18, 2025). (In Chinese).

Zhongyang yanjiuyuan lishi yuyan yanjiusuo (2025b), “Ming Qing dang’an gongzuoshi”, [online], available at: <https://newarchive.ihp.sinica.edu.tw/> (accessed July 18, 2025). (In Chinese).

Zhongyang yanjiuyuan shuwei wenhua zhongxin (2014), “Taiwan hanxue shuwei diancang ziyuan daolan”, [online], available at: https://sinology.ascdc.sinica.edu.tw/index_en.html (accessed July 18, 2025). (In Chinese).

Zhongyang yanjiuyuan shuwei wenhua zhongxin (2018), “Zhongyang yanjiuyuan shuwei diancang”, [online], available at: <https://sinica.digitalarchives.tw/> (accessed July 18, 2025). (In Chinese).

В. О. Кіктенко

Історіографічний огляд розвитку цифрової гуманітаристики в китаєзнавчих студіях (1984–2024)

Дослідження репрезентує всебічний історіографічний аналіз становлення та розвитку цифрової гуманітаристики в китаєзнавстві за сорок років – від 1984-го до 2024 року. Праця простежує еволюцію галузі: від перших експериментів комп’ютерної обробки китайських текстів до створення сучасної цифрової інфраструктури для комплексних китаєзнавчих досліджень. Методологічною основою слугує хронологічний аналіз у поєднанні із системною класифікацією ключових проєктів, ініціатив та технологічних рішень. Опрацьовано понад тридцять знакових проєктів та ініціатив, серед яких – стандарти кодування (GB 2312, Big5, Unicode), технологічні платформи (TEI, XML) та масштабні цифрові репозиторії. Особливий акцент зроблено на формуванні цифрового китаєзнавства як окремої академічної дисципліни. Показано, що розвиток галузі відбувався у трьох основних напрямках: *технологічному* (стандартизація кодування китайських ієрогліфів через Unicode; впровадження XML-та TEI-стандартів розмітки тексту); *методологічному* (розробка принципів структурованої розмітки текстів, формування реляційних баз даних для біографічної та текстової інформації); *інституційному* (створення міжнародних наукових мереж і дослідницьких центрів). Окреслено вісім основоположних принципів цифрового китаєзнавства: стандартизація кодування, структурна розмітка, реляційні бази даних, міжнародна співпраця, відкритий доступ, збереження метаданих, масштабованість, сумісність форматів. Доведено, що ці принципи визначають і сьогоdnішній розвиток галузі. У роботі висвітлено успіхи інтеграції цифрових технологій у китаєзнавчі дослідження, що привело до створення міцної технологічної бази для майбутніх академічних проєктів. Географічне охоплення – материковий Китай, Тайвань, Гонконг, США – ілюструє міжнародний характер становлення цифрового китаєзнавства. Систематизовано внесок провідних інституцій: Academia Sinica, Міністерства культури КНР, Гарвард-Янцзін інституту, Національної бібліотеки Китаю, міжнародних технологічних консорціумів. Оригінальність дослідження полягає в першому цілісному огляді сорокарічного розвитку цієї галузі з деталізацією ключових етапів, систематизацією проєктів і формулюванням фундаментальних принципів, що визначили сучасну цифрову інфраструктуру китаєзнавчих студій. Практична значущість результатів полягає в можливості їхнього застосування для стратегічного планування наступних цифрових ініціатив у науковому вивченні Китаю.

Ключові слова: аналітика великих даних; інституціоналізація; китаєзнавство; оцифрування текстів; стандартизація; цифрова гуманітаристика; штучний інтелект

Стаття надійшла до видання 26.07.2025
Прийнято до друку після рецензування 9.09.2025
Опубліковано 7.10.2025