Journal of Library and Information Studies 16:2 (December 2018) pp.1-23
doi: 10.6182/jlis.201812_16(2).001

Investigating Digital Humanities: A Domain Analysis of Conference Proceedings Published in Taiwan, 2009-2016

Hur-Li Lee[1], Shengang Wang[2]

Abstract

The study reports on an analysis of a chosen corpus in the emergent domain of digital humanities (DH). In contrast to other studies of the DH literature that focus on publications in the west, this study examines 129 papers published in the proceedings of the International Conferences of Digital Archives and Digital Humanities that were held in Taiwan between 2009 and 2016. In all, 236 individual authors from 15 countries contributed at least one paper; 50 domains were represented. Three East Asian countries (Taiwan, China, and Japan) show a dominating presence, and top three domains (computer science, history, and Chinese) have the highest numbers of participants and highest numbers of first authors. Unlike their counterparts in the humanities, the papers in the study have a much higher percentage of collaborative works. More than half of the papers that are collaborative works are interdisciplinary, but only one-fifth involve international collaboration. Proportionally, computer scientists' participation rate decreases and humanists' rate increases, however modestly, from 2012 onward. The study also investigates digital technology's impact on DH in various stages of the information lifecycle. More than two-thirds of the papers discuss technology's impact in the area of consuming data from digital collections for various purposes, with the impact on building retrieval systems/online platforms coming in second at 26.6%. Among different years, the first year is exceptional in showing high interest in the impact on building digital collections, building knowledge organization systems, and building retrieval systems but low interest in the impact on data consumption. Humanists in general are more attentive to the impact on consumption than technologists, while the latter lean toward the impact on building retrieval systems. Without any claim to comprehensiveness or representativeness, the study provides a snapshot of the DH literary output.

Keywords: Domain Analysis; Digital Humanities; International Conference of Digital Archives and Digital Humanities (Taipei, Taiwan)

1. Introduction

In library and information science (LIS), domain analysis is an important approach. Domain-analytic studies investigate an individual domain from different perspectives and taking different forms. For example, we may survey information needs or observe the patterns of information-seeking behavior of various user groups in a domain. We may analyze and interpret the literary and/or artistic output of a domain. We may design a knowledge organization or information retrieval system for use in a domain. Or we may plan and evaluate information services directed at a domain. Many LIS researchers build information systems and provide information services with a belief in universality. On the other hand, others are convinced that different

[1,2]School of Information Studies, University of Wisconsin-Milwaukee, Wisconsin, USA
* Corresponding Author: Hur-Li Lee, E-mail: hurli@uwm.edu

臺灣大學學術期刊資料庫

domains hold different needs and require domain-specific systems and services. The study reported in this article takes the latter view and targets a particular body of literature in an attempt to better understand the emergent domain of digital humanities (DH).

DH has seen a dramatic increase in activities, both research oriented and project building, in the past decade or so. As of the moment, DH is young, evolving, and robust. The present study does not intend to engage in the debates about the proper name for the domain or the legitimacy of the domain, nor does it define the domain's scope or content. A simple definition of DH will thus suffice here: The domain in which participants apply digital technology to humanist research and pedagogy. As stated above, the main goal of the study is to contribute to a better understanding of the DH literary output. It will do so in two ways: (1) by examining a body of the DH literature published in Taiwan that represents efforts made predominantly by East Asian DHers, a significant departure from prior research exhibiting a European and North American favoritism; and (2) by examining the broad impact of digital technology on DH, an aspect missing in prior research.

Given that there have been no specialty journals devoted entirely to the DH in Taiwan (the first one will appear in January 2018), proceedings produced for the International Conference of Digital Archives and Digital Humanities between 2009 and 2016 (seven annual meetings, missing 2013, that took place in Taipei, Taiwan) provide a convenient sample. The study takes a quantitative, domain-analytical approach to address three research questions:

1. What are the major characteristics of the target DH literature?
2. In what stages of the information lifecycle did digital technology impact DH during the said years? Did such impacts stay constant or change during the study period, and in what ways?
3. How did the participating domains, or disciplines, differ in terms of the digital technology they apply in their research?

The second and third research questions may take a number of directions: considering technology's impact on various stages of the information lifecycle, analyzing the impact of specific digital tools, etc. It is the first one on which the present study will focus. The coding scheme for these questions and the information lifecycle will be discussed in the Method section below. For the third research question, the study will not consider any trend over time due to the limited numbers of papers in individual years.

2. Literature Review

The term "domain analysis" is not new in the scientific world. For example, Prieto-Díaz (1990) offers a brief introduction to domain analysis from the software engineering perspective. In LIS, Hjørland and Albrechtsen (1995) are the first to describe and theorize domain analysis and assert, "the most fruitful horizon for IS [i.e., information science] is to study the knowledge-domains as thought or discourse communities, which are parts of society's divisions of labor" (p. 400). Like many other essential concepts, the understanding of domain analysis is diverse. Palmer (1999) claims that there are two different conceptions of domain analysis in LIS: One centers around

臺灣大學學術期刊資料庫

theory of knowledge, and the other is based on the study of literatures. Under the umbrella of the domain-analytic paradigm, Hjørland (2002) maintains that 11 specific approaches together define the unique competency for information specialists. These approaches include: producing literature guides and subject gateways; producing special classifications and thesauri; document and genre studies; and bibliometrical studies, among others. In recent years, LIS has also generated a large body of domain-analytic studies providing useful insight into a variety of domains. Some of these studies, mainly literature based, explicitly label themselves as domain analyses (e.g., Graf & Smiraglia, 2013; Guimarães, De Oliveira, & Gracio, 2012). Bibliometric methods are common in those studies that represent one of the branches of domain analysis.

As DH is new, there is considerable attention to the domain itself in recent literature, attempting to identify its origin and scope, to deliberate its theory and methodology, to trace its history and trend, and to critique its limitations and bias (e.g., Burdick, Drucker, Lunenfeld, Presner, & Schnapp, 2012; Clement & Carter, 2017; Liu, 2013; McPherson, 2012; Poole, 2017). Some of these efforts have been spent analyzing the domain's literary output. They represent several of the 11 approaches to domain analysis, as conceptualized and categorized by Hjørland (2002). For example, archaeologist Huggett (2012) examines seven DH journals in English and concludes that not all humanities are well represented in the DH literature. The disciplines of literature, linguistics, and history show the strongest association with DH; a relationship between digital archaeology and the DH is largely absent. Huggett's study is purposefully slanted, aiming to understand digital archaeology and its relationship to DH. DH is not the center of attention. Additionally, Huggett describes the shifting usage of terminology, with decreasing use of the word "computing" (such as in the term "linguistic computing") and the increasing use of the word "digital" (such as in the term "digital history"). We may view the study as taking two approaches described by Hjørland (2002): the study of structures and institutions in scientific communication and terminological studies.

In LIS, Dalbello (2011) constructs a genealogy of DH by closely examining 59 texts (authored and edited books, journal articles, and others) and 14 websites that discuss "specific digital projects and their impact on the ways in which scholarship was practiced, and those in which practice was theorized and identified within the fields of the humanities," (p. 483) all in English. In its findings, the study identifies several evolutionary stages of the domain through the lens of technological development. Starting in the 1950s, digital technology contributed to the building of digital collections and enabling corpora searching; in the 1980s and 1990s, scholars began to critically rethink texts, canonical knowledge, and knowledge creation; roughly in the 1990s and onward, digital collections became institutionalized and consolidated; and, from the mid-2000s, we have seen new forms of digital text creation, extraction, and, most importantly, consumption in the humanities. This research is qualitative in nature and takes the epistemological and critical studies approach described by Hjørland (2002).

In a bibliometric study (another approach in Hjørland's taxonomy of domain analysis), Wang

臺灣大學學術期刊資料庫

and Inaba (2009) perform correspondence analysis (which looks at the association of keywords and publication years) and co-word analysis, both typical of bibliometric techniques. They performed their analysis on 745 works in English from two journals and five sets of conference proceedings in DH between 2005 and 2009. They find the shifting focus in terminology, for example, the continually decreasing popularity of the term "humanities computing" (in both frequency and degree centrality), and the opposite trend of the term "digital humanities" (in frequency alone).

Another study by Tang, Cheng, and Chen (2017) collects 2,115 articles and book chapters in English, the majority from the database Scopus and the rest from six journals published by the members of the Alliance of Digital Humanities Associations between 1989 and 2014. By applying bibliometric analyses on co-authorship, article co-citation, bibliographic coupling networks, and modularity maximization partition, the study examines the degree of topical diversity and intellectual cohesion within the dataset over time. Its findings show that the degree of knowledge diversity is high, but knowledge integration is mixed with some of the indicators, gradually becoming more cohesive with the rest remaining extremely fragmented. Collaboration among authors is mainly limited by language and geographic boundaries.

All four studies look at trends. In the first and third studies, Huggett (2012) and Wang and Inaba (2009) find lexical change. Dalbello (2011) examines the epistemological evolution of the domain, while Tang et al. (2017) analyze intellectual cohesion in DH longitudinally. This shared interest in trends is understandable, as change and shifts over time are a key concern in the study of literary output and the study of a domain's development. When considered together, the studies reviewed display a clear slant toward examining corpora published exclusively in English, mostly from either the United States or the United Kingdom.

3. Method

Similar to many domain-analytic studies, the research reported below analyzed a selected body of literature in the domain of DH—the second type of domain analysis mentioned above. Unlike previous studies of DH that examine corpora published predominantly in the west, the current study chose all the papers from the paper sessions from the proceedings of the first seven meetings of the International Conference of Digital Archives and Digital Humanities (DADH below), co-sponsored by the Research Center for Digital Humanities (RCDH) of the National Taiwan University and the National Taiwan University Library. Founded in 2007, the RCDH was an early leader in DH in Taiwan, responsible for a number of impressive digital libraries projects such as the *Taiwan History Digital Library* and the *Taiwan Ethnological Collections in Overseas Museums*. In addition to large-scale digitization projects and the annual DADH conferences, the center has offered numerous seminars, workshops, and lectures and produced many publications.

With the exception of 2013, the conferences were held annually between 2009 and 2016 in Taipei, Taiwan. The meeting sessions included keynote speeches, papers, panels, and posters. The published proceedings contain abstracts of

臺灣大學學術期刊資料庫

keynote speeches, panels, and posters as well as full-length papers. While the papers are in either Chinese or English, all but one include an abstract and several keywords in both languages supplied by the authors themselves. In all, 129 papers from the paper sessions in the first seven DADH conferences were identified to form the corpus for this study. The majority of the papers (i.e., 128) include the full text, but one of them only has a title, an author name, an abstract in Chinese, and the author's country affiliation (in a different part of the proceedings). After removing this paper, the analysis of digital technology's impact on DH was performed on only 128 papers. The proceedings obtained for the study are digital copies.

The study has three research questions, as stated above. Research question #1 necessitated the use of descriptive statistics including the total number of papers, the total number of authors, the frequency distributions of individual author country and domain affiliations, the frequency distributions of papers by single and multiple authors, by type of collaboration, and by first author domain affiliation, and the frequency distribution of keywords supplied by the authors.

In this study, author information of interest included country affiliation and domain affiliation, both directly derived from the published proceedings. A few of the authors seemed to have different nationality (e.g., an author with a German name but affiliated with an institution in Taiwan) or disciplinary training (e.g., a professor in library and information science trained in computer science) than the information specified in the proceedings, which is not reflected in the data.

Counting domain affiliation was somewhat problematic, for different institutions often name the same field of study differently and have different administrative units. For the purpose of this study, some adjustments were made, sometimes based on our personal knowledge about the institutions in question and a few times verifying with information provided on the Internet by the institutions. The domain computer science is an example; it includes several other variant names such as: information science, computer science and information engineering, information technology, and so forth. All these were coded as computer science. An example in the humanities is Chinese, which includes other variant names such as: Chinese studies, Chinese literature, Chinese language and literature, contemporary Chinese studies, and so forth. All were coded as Chinese. In some cases, an academic department or a research center has in its name more than one discipline, for example, law and psychology, art and humanities, English and applied linguistics, and image arts and science. A center was coded "humanities" if both domains in its name belong to the humanities (e.g., art and humanities), while an academic department was coded "others" if its name includes domains that cannot be easily grouped into one (e.g., image arts and science). The domain "area studies" subsumed China studies, Southern Asia studies, Asia Pacific studies, and Oriental studies. Among all authors, eight had two different affiliations in different years, rendering the sum of author affiliations higher than the sum of authors.

Scientific collaboration among authors is also a major concern of the study. With the help of the visualization and exploration software Gephi (https://gephi.org/), a collaboration network was produced to illustrate the cross-

臺灣大學學術期刊資料庫

country collaboration of authors involved. As an indicator of collaborative scientific activities (Katz & Martin, 1997), measuring multiple authorship is relatively inexpensive and practical, and can be applied to large samples or datasets (Subramanyam, 1983). Collaboration can be classified into various categories from different perspectives. For example, if one paper has three authors from two different countries and different disciplines, then it is considered an international and interdisciplinary collaborative study. Similar to Chang's study (2012), this study uses a researcher's disciplinary affiliation and internationality (i.e., called "geographical distance" by Chang) between co-authors as two indicators to better understand collaboration. To be clear, the information about author domain and country affiliations, which determine types of collaboration, was taken straightly from the proceedings. Authors' nationalities and disciplinary training were of no concern.

For the remaining two research questions, a 6-point coding scheme on the basis of the information lifecycle was developed for categorizing digital technologies applied in individual studies:

(1) N/A, for papers to which none of the following categories apply;
(2) Building/storing digital collections;
(3) Building/enhancing knowledge organization systems (KOSs);
(4) Building retrieval systems/online platforms;
(5) Retrieving/analyzing/evaluating/representing data, including the discovery of new patterns or relationships in data and transformation of concepts (in other words, creating new knowledge); and
(6) Facilitating collaboration.

The information lifecycle is simply defined in terms of four stages for the purpose of this study (see Figure 1): creating a digital collection by collecting digital resources and/or digitizing analog resources; designing and building a KOS for the digital collection; designing and building a retrieval system or an online platform; and taking data from a digital collection for research and pedagogical purposes. The last stage will then lead to production of new resources that become

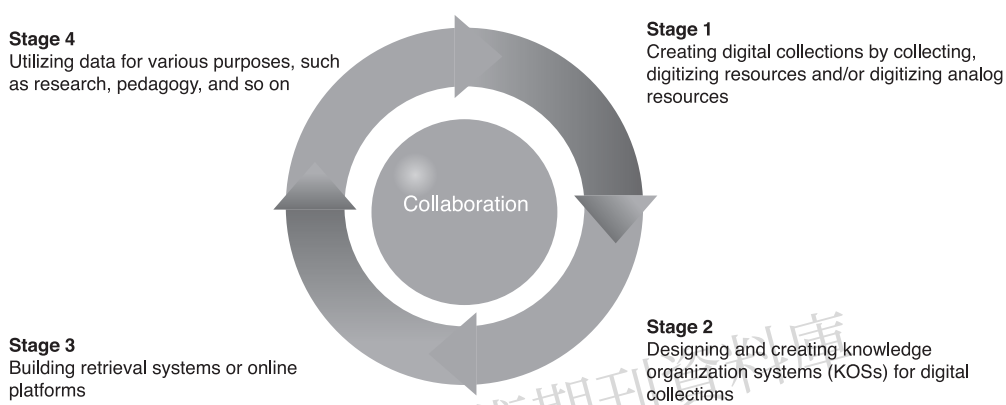


Figure 1. The Information Lifecycle

臺灣大學學術期刊資料庫

candidates for digital collections in stage 1. Lying at the heart of the lifecycle is collaboration, which may transpire at any stage.

Generally speaking, categories 1 to 6 in the coding scheme are clear-cut; one example for the N/A category is a bibliometric study of the DH domain. Categories 2-4 represent digital technology used in the various building stages of digitization projects; and 5 is for coding studies that apply digital technology to the consumption of data from digital collections. In coding, the two authors of the study first applied the scheme to categorizing all the papers from the 2009 proceedings independently by reading through the abstracts, sometimes supplemented with the information in the body of the text. To increase the reliability of coding, the coders then compared the preliminary results and found inter-coder agreement at 82%. Several issues emerged that led to clarification of the categories to resolve the differences in the two sets of results and to assist in further coding.

For example, a principle was established to indicate that 2, 3, and 4 should all be applicable for papers reporting on a digitization project with the exception of any paper having a clear focus on one or two of the three aspects (i.e., forming the collection, organizing the data in the collection, and setting up retrieval mechanisms). For papers reporting on a digitization project, the contributing authors often used examples to demonstrate how the built digital collection is useful in future research. Another guiding principle was established to allow the assignment of 5 in such cases only if the individual papers explained the example in at least a full section. Once the coding was completed, cross-tabulations were run between type of digital technology and author domain affiliations. Because the number of papers in each year's proceedings is comparatively low (between 16 and 24), the results from these cross-tabulations are reported in aggregation and selectively. This limitation of the data source also makes statistical tests infeasible; thus, the study purposefully focuses on summaries of descriptive statistics, skipping statistical tests.

In addition, an analysis was conducted of the keywords, consisting of straightforward frequency counts and visualization, in the conference papers supplied by the authors. The primary purpose of the keyword analysis was to survey the terminology and discern commonly used keywords, which may also reveal popular themes and concerns represented in the dataset. For visualization, an online program WordItOut (https://worditout.com/) was used to generate a word cloud map for the top 100 words.

4. Results

This section is divided into four subsections, the first on authorship, the second on the papers, the third on the terminology, and the last on the impact of digital technology. While the first three subsections address the first research question, the fourth details the cross-tabulations to address the other two research questions.

4.1 Authorship

As mentioned above, author information of interest includes country affiliation and domain affiliation, both directly derived from the published proceedings. The total number of individual authors named in the full-length papers

臺灣大學學術期刊資料庫

included in the study is 236, representing 15 countries/regions, 46 (19.5%) of whom contributed more than one paper. Among the individuals, four had two different country affiliations, with one of them listing two affiliations in the same paper and the others in different papers, thus increasing the total to 240 in Table 1. Understandably, authors from institutions in Taiwan are the largest group, which adds up to 137 (57.1%). China, Japan, and the United States (34/14.2%, 28/11.7%, and 21/8.8%, respectively) round out the top four. Together these four countries account for almost 92% of authors who contributed at least one paper to DADH 1-7.

In all, computer scientists (n = 52) represent the largest group of authors, double of those working at information services (n = 26; see Table 2). The former does not even include those trained in computer science but affiliated with other academic units such as a professor in library and information science and a programmer working at a library. Out of the 26 affiliated with information services—the domain having the second highest number of participants—one was a digital content curator at a research institute, two worked at an information center, and the rest worked at libraries. Not counting the category of "humanities," the first two domains in the humanities with the highest numbers of participants are history (n = 18) and Chinese (n = 15). The Chinese department in the majority, if not all, institutions of higher education cover both Chinese language and literature. In other words, the "Chinese" category may be considered roughly equivalent to "literature" and "linguistics" in Huggett's study (2012), making the top humanities domains participating in DH almost identical in the two studies. The other two domains that had more than 10 participants are general humanities (n = 16) and geography (n = 11). When counting the three in geographic information systems (GIS), the total number of geographers becomes 14.

4.2 The papers

In total, the corpus for the study consists of 129 papers from the paper sessions published in the past seven volumes of the DADH conference proceedings. Table 3 shows the numbers of papers

Table 1. Frequency Distribution of Author Country Affiliations

Country/Region	Count	%	Country/Region	Count	%
Taiwan	137	57.1	South Korea	2	0.8
China	34	14.2	Czech	1	0.4
Japan	28	11.7	Germany	1	0.4
United States	21	8.8	Mongolia	1	0.4
Netherlands	4	1.7	Philippines	1	0.4
United Kingdom	4	1.7	Singapore	1	0.4
France	2	0.8	Sweden	1	0.4
Hong Kong	2	0.8	**Total**	**240**	**100.0**

臺灣大學學術期刊資料庫

Table 2. Authors' Domain Affiliations (frequency ≥ 6)

Domain/Discipline	Frequency
Computer science	52
Information service	26
History	18
Humanities	16
Chinese	15
Geography	11
Digital humanities	8
Bio-industry communication & development	6

Table 3. Distribution of Papers by Authors from Taiwan versus Authors Not from Taiwan

Year	All authors from Taiwan (1)	First author from Taiwan (2)	(1) + (2)	First author not from Taiwan (3)	None from Taiwan (4)	(3) + (4)	In total
2009	13	0	13 (72.2%)	0	5	5 (27.8%)	18
2010	19	0	19 (79.2%)	0	5	5 (20.8%)	24
2011	10	4	14 (82.4%)	1	2	3 (17.6%)	17
2012	10	2	12 (75.0%)	0	4	4 (25.0%)	16
2014	9	4	13 (61.9%)	0	8	8 (38.1%)	21
2015	7	1	8 (47.1%)	1	8	9 (52.9%)	17
2016	5	0	5 (31.3%)	1	10	11 (68.8%)	16
Total	73	11	84 (65.1%)	3	42	45 (34.9%)	129

by year and by country affiliation. The number of papers per year fluctuates slightly between 16 and 18, except in 2010 and 2014, when there are more than 20 papers each. In addition, we use authors' country affiliations to group the papers into: (1) papers by author(s) from Taiwan exclusively; (2) papers of multiple authorship where the first author of each paper was from Taiwan but the other(s) were not; (3) papers of multiple authorship where the first author of each paper was not from Taiwan although at least one of the others was; and (4) papers by author(s) not from Taiwan. Understandably, researchers affiliated with Taiwanese institutions played an active part in this international conference. Papers by a single or multiple authors from Taiwan add up to 73, more than half (56.6%) of all the papers from the paper sessions of DADH 1-7. When we

臺灣大學學術期刊資料庫

also consider a paper's first author, the number of such papers increases to 84. That is, 65.1% of the total full-length papers in the dataset were either written by a single or multiple authors from Taiwan or by a first author from Taiwan. There is a noticeable rising trend in the percentage of papers by authors not from Taiwan (Figure 2). Seen from this view, the internationality of DADH was steadily increasing.

Table 4 shows the frequency distribution of papers by country/region when considering all authors rather than the first authors alone. For example, 88 out of the 129 papers (68.2%) listed at least one author affiliated with an institution in Taiwan.

Figure 3 is a collaboration network produced by the software Gephi to illustrate the cross-country collaboration of authors involved in DADH 1-7. The nodes represent countries or regions from which authors submitted their studies in the paper sessions of the DADH. The size of each node proportionally varies with the number of papers, and the thickness of each link between two of the nodes corresponds to the number of collaborative papers between two connected countries or regions. At the country/region level, Taiwan, without any doubt, lies at the center of this network, occupying the largest node with connection to all the others except Mongolia. In addition, there generally existed a positive relationship between the productiveness of one country/region and its frequency of collaboration with Taiwan. Singapore, South Korea, and Sweden do not appear in the graph because their participants are not in collaboration with authors from other countries.

Of the 129 papers, the number of authors per paper ranges from one to seven, and the average number of authors per paper is 2.6. Figure 4 shows the percentages of papers with varying numbers of authors: single-author papers (35.7%) account for the highest percentage, followed by two-author papers (19.4%) and three-author papers (17.8%); only 20 papers (15.5%) were written by five authors or more. In short, almost two thirds (64.3%) are collaborative works.

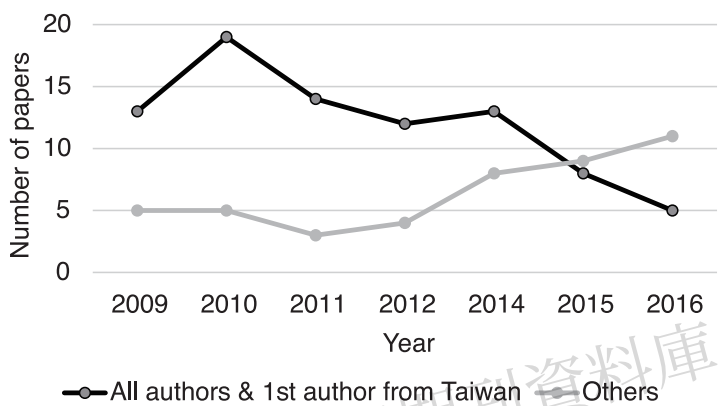


Figure 2. Trends in Counts of Papers from DADH, 2009-2016, Considering Authors' Affiliations with Taiwan

臺灣人文與藝術期刊資料庫

Table 4. Frequency Distribution of Papers by Country/Region

Country/Region	Papers in the paper session, *n* (%)
Taiwan	88 (68.2)
China	18 (14.0)
Japan	13 (10.1)
United States	11 (8.5)
Hong Kong	5 (3.9)
France	2 (1.6)
Germany	2 (1.6)
Netherlands	2 (1.6)
United Kingdom	2 (1.6)
Czech	1 (0.8)
Mongolia	1 (0.8)
Philippines	1 (0.8)
Singapore	1 (0.8)
South Korea	1 (0.8)
Sweden	1 (0.8)

Table 5 shows collaboration by interdisciplinarity and internationality. The number of interdisciplinary papers is larger than that of disciplinary papers (53% versus 47%), which to some extent reflects the interdisciplinary nature of digital humanities as a domain. We can also detect this literature's interdisciplinarity based on author keywords (see Figure 6). In terms of internationality, within-country collaboration is significantly higher than international collaboration (79.5% versus 20.5%).

Under the assumption that the first author makes the most intellectual contribution to a paper, counting papers by the first author's domain across the seven years may be of interest. Table 6 provides such counts for the top 11 domains. Computer science, history, and Chinese remain as the top three, with general humanities tying with Chinese. In Figure 5, these four domains are charted visually. The humanities domains do not seem to exhibit any trend, while papers having computer scientists as the first authors

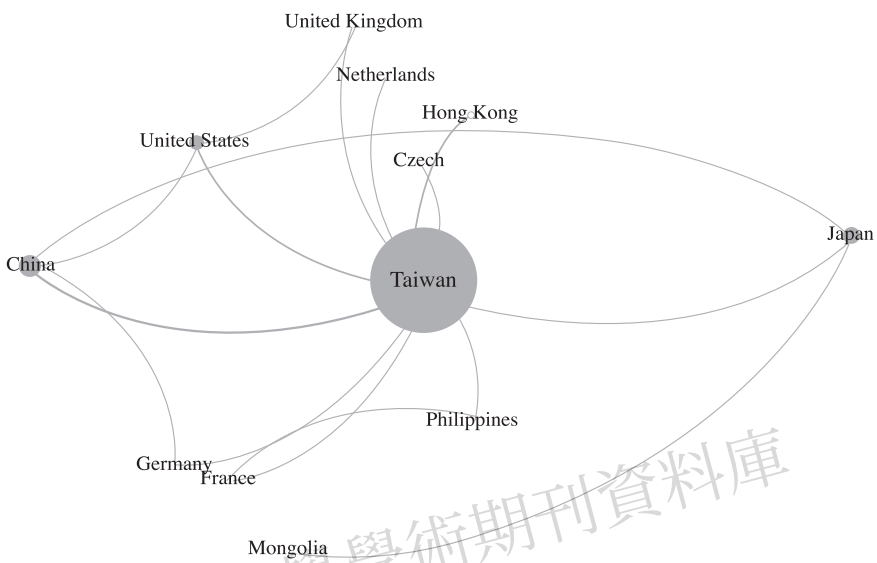


Figure 3. Collaboration Network of Countries/Region Involved in DADH 1-7

臺灣大學學術期刊資料庫

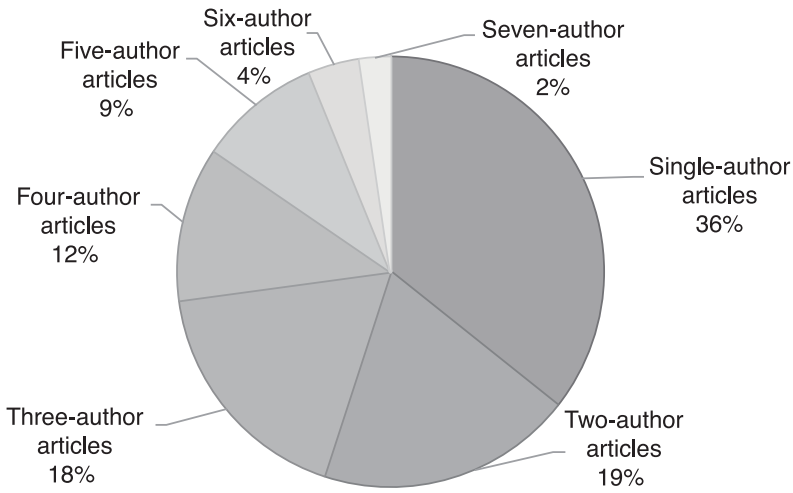


Figure 4. Distribution of Papers by Number of Authors

Table 5. Distribution of Types of Collaboration

Type of collaboration	No. of documents	Percentage
Disciplinary affiliation	83	100.0
Interdisciplinary collaboration	44	53.0
Disciplinary collaboration	39	47.0
Internationality	83	100.0
International collaboration	17	20.5
Non-international collaboration	66	79.5

show a clear decline after 2012. In other words, proportionally more papers had humanist scholars as first authors from 2012 onward. This seems to imply that contributions made by humanists were on the rise.

4.3 The terminology

Often bibliometric studies analyze keywords provided by authors. The common assumption is that author keywords likely represent the core concepts or the foci of the studies. Table 7 lists individual words in the author keywords taken from the 128 papers in the target corpus (with one paper removed due to its lack of keywords) that appear 10 times or more. Not surprisingly, the words "digital" and "humanities" are number 1 and number 3 on the list. For conferences held in Taiwan, the words "Chinese" and "Taiwan" can also be expected. The words "history," "historical," and "archives" most likely result from the heavy

臺灣大學學術期刊資料庫

Table 6. Frequency Distribution of First Author Domain Affiliations by Year (top 10 domains)

Discipline	2009	2010	2011	2012	2014	2015	2016	In total
Computer science	5	5	6	8	6	1	2	33
History	1	3	2	1	2		4	13
Chinese	1	2	1	1	1	4	2	12
Humanities		4		2	2	3	1	12
Geography		3	1	1	1			6
Area studies			1		1	2		4
Bio-industry communication & development		1			1		2	4
Information services	3					1		4
Anthropology	1		1				1	3
Digital humanities	1	2						3
LIS	1			1		1		3

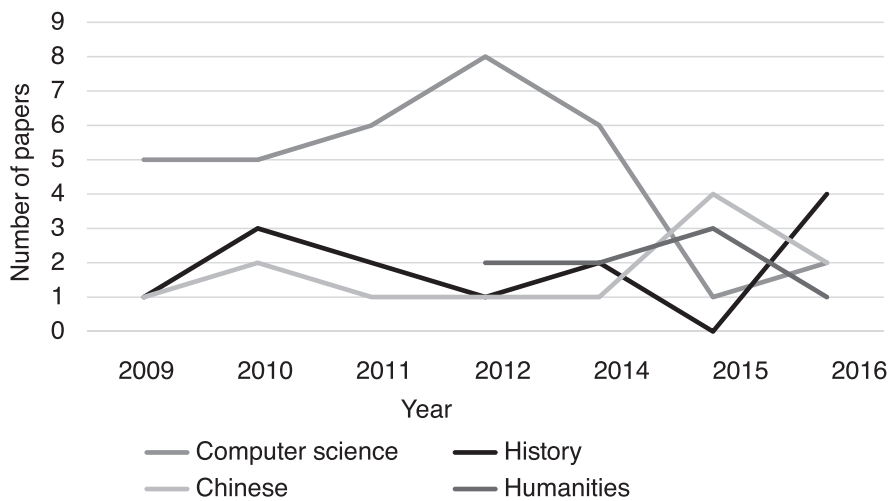


Figure 5. Frequency Distribution of Papers by First Author's Domain (in the top four domains), 2009-2016

participation by historians. Another important branch of topics is related to research methods: "analysis," "data," "text," "database," and "geographic information systems" (GIS).

Figure 6 is a word cloud map of the top 100 individual words, drawn by employing the online word cloud generator WordItOut. Other words relevant to East Asian cultures, such as "Japanese"

臺灣人文學學術期刊資料庫

Table 7. Frequency Distribution of Individual Words from Author Keywords ($n \geq 10$)

Individual word	2009	2010	2011	2012	2014	2015	2016	Total
digital	9	6	8	3	7	5	7	45
analysis	4	4	7	2	9	4	5	35
humanities	3	3	1	1	5	4	5	22
Chinese	1	3	2	4	5	1	2	18
data	3	1	2	2	6	2	1	17
history	1	3	4	1	3	2	1	15
historical	4	3	0	1	0	1	2	11
text	0	4	1	1	2	1	4	13
Taiwan	0	5	2	1	1	2	1	12
database	2	1	4	0	1	2	1	11
GIS	1	3	4	0	1	0	1	10
archives	2	3	1	2	1	0	1	10



Figure 6. Word Cloud Map Generated from Top 100 Author Keywords

and "dynasty," are also quite visible. Furthermore, there is a host of other research methods-related words, such as "social," "network," "extraction," and "term."

In a more useful approach, Table 8 shows high-frequency author key *terms*, rather than individual words. Not surprisingly given the theme of the conferences, keywords with various forms of the

臺灣大學學術期刊資料庫

Table 8. High-frequency Author Keywords (Separated by Semicolon)

Frequency	Keywords
19	digital humanity/ humanities
9	GIS (Geographic Information System)
8	digital archive/ archives
7	text mining
5	digitalization/ digitization; Taiwan History Digital Library (THDL)
4	digital library/ libraries; social network analysis

word "digital" appear the most frequently (five out of eight times). The remaining ones include a particular type of digital technology (i.e., GIS) and two research techniques (i.e., text mining and social network analysis) made possible by advanced digital technology.

4.4 Impact of digital technology

Table 9 shows frequency distribution of various types of impacts on DH by digital technology, allowing multiple codes for each paper; Figure 7 is a visualization of the same data. Clearly the most-coded category is retrieving/ analyzing/evaluating/representing data (n = 88), followed by the category of building retrieval systems/platforms. The first year is exceptional in that it sees the highest numbers of papers in technology's impact on building collections, KOSs, and retrieval systems (n = 8, 10, and 7, respectively), but the lowest number in consumption of data from digital collections (n = 5). With a noticeable peak in 2010, 22 out of the 24 papers addressed issues related to technology's impact on use of data in DH studies. Curiously, only a small number of papers (n = 5) discussed how technology had been used to facilitate collaboration. The figures within each category of technology's impact across the remaining years do not fluctuate as much. Another year exhibiting a significant difference from the others is 2012, when the papers addressed issues concentrating on only half of the categories, with almost nine out of the 10 papers (14 out of 16, or 87.5%) focusing on utilizing data from digital collections.

The lack of standardization in naming individual domains made it necessary to first merge some of the institutional affiliations with somewhat variant names, a procedure mentioned above in the Method section. As a result of the procedure, 50 domains were identified. Due to the fact that many domains have only one contribution in the entire dataset, listing digital technology's impact for all domains is less interesting. The two tables below thus include only aggregated and partial data. Table 10 (also see Figure 8) places some of the domains into two major groups: (1) the humanities, which includes history, Chinese, general humanities, area studies, anthropology, digital humanities, archaeology, art, Buddhist studies, ethnic literature, law, and western languages and literature; and (2) technology-related domains, which includes computer science,

臺灣大學圖書館期刊資料庫

Table 9. Impact of Digital Technology by Year

Type	2009	2010	2011	2012	2014	2015	2016	Total
1. N/A	1	0	1	0	0	2	2	6
2. Building/storing digital collections	8	1	3	0	1	1	1	15
3. Building/enhancing knowledge organization systems	10	1	1	2	4	1	1	20
4. Building retrieval systems/ platforms	7	3	5	4	7	4	4	34
5. Retrieving/analyzing/evaluating/representing data	5	22	12	14	13	11	11	88
6. Facilitating collaboration	2	1	2	0	0	0	0	5

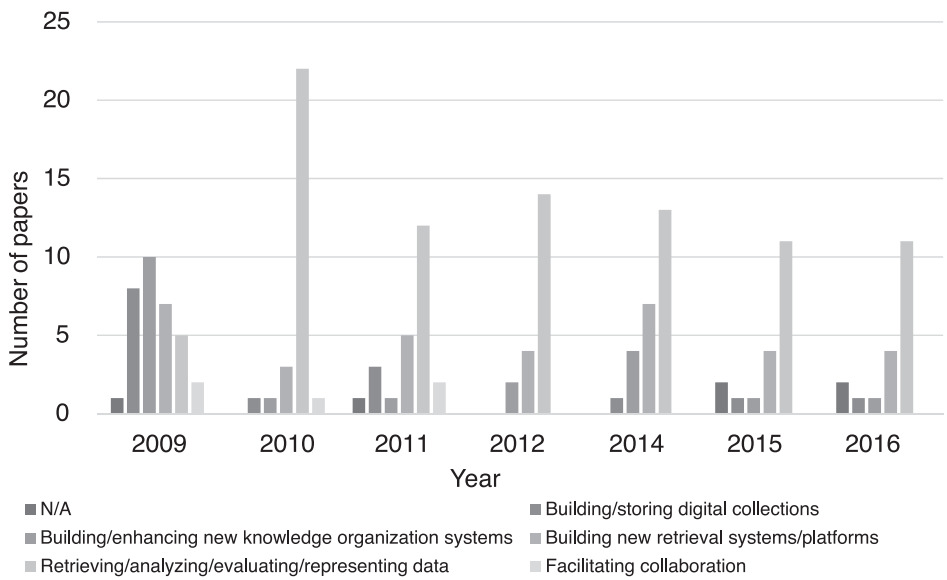


Figure 7. Impact of Digital Technology by Year

information services, LIS, science and engineering, computational linguistics, and creative science and engineering. The only omitted domains are those that do not belong to either camp.

In the second column of Table 10, the number is the total count of papers in all the domains (determined by the first author) involved in each camp. The two types of technology impact that have the highest counts are identical in both camps, i.e., using data from digital collections (n = 41, 25, respectively) and building retrieval systems (n = 14, 17, respectively). However, the humanities group has a higher percentage of papers on using data than the technology domains (77.4% versus 54.3%), while the latter has a higher percentage of papers on building retrieval

臺灣大學學術期刊資料庫

Table 10. Frequency Distributions of Technology's Impact on Humanities versus Technology Domains, According to First Author

Domain	No. of papers	Code 1	Code 2	Code 3	Code 4	Code 5	Code 6
Humanities	53	1	6	7	14	41	4
		1.9%	11.3%	13.2%	26.4%	77.4%	7.5%
Technology	46	4	6	7	17	25	0
		8.7%	13.0%	15.2%	37.0%	54.3%	0.0%

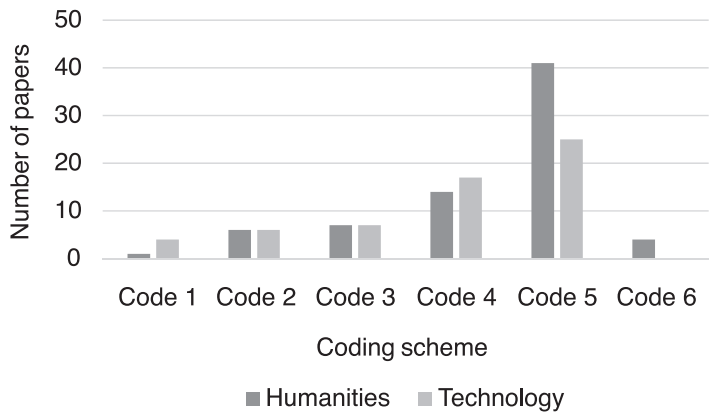


Figure 8. Impact of Digital Technology, the Humanities vs. Technology Group

systems than the former (37% vs. 26.4%). The more intense interest in using data for humanistic research by the former and the more intense interest in building collections (13% versus 11.3%), building KOSs (15.2% versus 13.2%), and building retrieval systems (37% versus 26.4%) by the latter are not at all surprising. It is worth noting that the technology-related domains contributed no papers about technology's impact on facilitating collaboration.

Table 11 shows the frequency distributions of technology's impact for the top four domains (according to the first author) that contributed more than 10 papers each: computer science, history, Chinese, and general humanities. Although 63.6% of the papers contributed by computer scientists as first authors discussed the use of data from digital collections, this figure is still lower than those for the three humanistic domains (69.2% for history, 83.3% for both Chinese and general humanities). Papers with computer scientists as the first authors also tend to address the issues of technology's impact on building digital collections, KOSs, and retrieval systems more frequently. There is one exception: The papers with historians as the first authors discussed behind-the-scenes KO-related issues in a significantly higher percentage than those

臺灣大學學術期刊資料庫

Table 11. Frequency Distributions of Technology's Impact for Top Four Domains, *n* (%) According to First Author

Discipline	No. of papers	Code 1	Code 2	Code 3	Code 4	Code 5	Code 6
Computer science	33	1 (3.0)	3 (9.1)	6 (18.2)	12 (36.4)	21 (63.6)	0
History	13	0	1 (7.7)	3 (23.1)	2 (15.4)	9 (69.2)	0
Chinese	12	0	0	0	3 (25.0)	10 (83.3)	1 (8.3)
Humanities	12	1 (8.3)	1 (8.3)	1 (8.3)	1 (8.3)	10 (83.3)	0

by researchers in other domains (23.1% versus 18.2% for computer science, 8.3% for general humanities, and 0% for Chinese). This seems to suggest that humanities scholars may also be eager to participate in the organization of data in digital collections that will no doubt influence how data in digital collections may be used in future research or teaching. These numbers are not high enough, however, and thus cannot lead to any definitive conclusion.

5. Discussion

Taking a quantitative, domain-analytic approach, this study of a corpus in DH is descriptive, rather than prescriptive, in nature. Its data source is the papers from the paper sessions of the first seven International Conferences of Digital Archives and Digital Humanities between 2009 and 2016. In total, 236 individual authors from 15 countries/regions contributed 129 papers. Less than 20% of the authors contributed more than one paper, suggesting that DH attracted broad interest. The majority of the participating authors (over 83%) came from three East Asian countries (i.e., Taiwan, China, and Japan), with approximately 57% affiliated with institutions in Taiwan—a clear but moderate dominance of the host country. With regard to geographic representation, this study differs from the other studies of DH literary output reviewed earlier that showed favoritism toward the west. Overall, 50 author domain affiliations were identified. The top two among them are technology-related: computer science and information services (52 and 26 authors, respectively); numbers 3-5 are humanity domains: history, general humanities, and Chinese.

As for the 129 papers in the corpus, the host country's dominance was even stronger. Roughly 65% of them were written by a single author from Taiwan, multiple authors all from Taiwan, or a first author from Taiwan; and over 68% of the papers involved Taiwanese participants. A little over one-third of the papers were by single authors, while nearly two-thirds of them (64.3%) were collaborative works, only slightly lower than 66.5% in scientometrics (a social science) (Chang, 2012), for example, and certainly unusual in the humanities (28.9%, according to Ossenblok, Verleysen, & Engels, 2014). Although the study by Ossenblock et al. (2014) includes journal articles and book chapters, thus not a direct comparison to the current study, the difference in the percentage of collaborative works between the two studies is significant. More than half of the collaborations (53%) in DADH were interdisciplinary, but only

臺灣大學學術期刊資料庫

slightly over 20% crossed national boundaries, a finding similar to that in Tang et al. (2017). The interdisciplinary nature of DH is self-evident. On one hand, humanists often need advanced technical assistance from technologists. The latter, on the other hand, may offer fresh ideas to manipulate data, be it textual, graphical, or aural, leading to exciting, novel approaches to humanities research. As Figure 3 reveals, the strongest collaborative link is between Taiwan and China, which share a common cultural heritage and the same official language, Chinese, providing conditions more conducive for collaboration.

Assuming the predominance of the first author, papers could be categorized into various domains by first author's domain affiliation. The largest group of papers belongs to computer science, followed by three humanity domains that had more than 10 papers each—history, Chinese, and general humanities. The top three humanities domains found in the study were very similar to the top domains identified in previous research cited above. Computer science has a clear lead with 33 papers, almost the same as the combined total of 37 in the top three humanity domains. Proportionally, papers by computer scientists as the leading authors are decidedly declining, and those with humanists at the top are increasing in the more recent years. It may not be conclusive at the moment, but DH, at least in the context of DADH, is shifting toward humanistic concerns and away from technological issues.

As for the terminology used in the corpus, the study found the most frequent keywords supplied by the authors to be words associated with the word "digital" (e.g., digital archives and digitization). Such numbers confirmed findings in earlier studies that DH increasingly used "digital" rather than "computing." Other high frequency words of significance included the names of regional places and cultures, words of specific interest to historians (e.g., archives), and words related to research methods.

In terms of the impact by digital technology in the information lifecycle, the category of consuming data from digital collections is of the primary interest in most papers (about 68%), significantly more so than the other categories. An exception to this type of impact is the first year, when only about 28% of the papers are related to technology's impact on using data. For the rest of the years, the percentages of papers on using data range from 68% to almost 92%. The first three stages of the information lifecycle concern the building of systems, each of which sees the highest number of papers in the very first DADH conference, with the rest of the years showing modest fluctuations. Very few papers (n = 5) cover issues of how technology impacts or facilitates collaboration.

Taking domains into consideration, the study examined the top four individual domains (i.e., computer science, history, Chinese, and general humanities) that have more than 10 papers each in the dataset. All of them studied digital technology's impact on data use in the highest percentage, ranging from 63.6% to 83.3%, across all categories. Among the four domains, computer science was the least interested in data use but the most concerned with building retrieval systems; history seemed to be most attentive to knowledge organization; and Chinese and the other general humanities were most enthusiastic about data use. The study also formed two broad

臺灣大學學術期刊資料庫

groups: (1) the humanist group, which includes history, Chinese, general humanities, area studies, anthropology, digital humanities, archaeology, art, Buddhist studies, ethnic literature, law, and western languages and literature; and (2) the technology group, which includes computer science, information services, LIS, science and engineering, computational linguistics, and creative science and engineering. While both groups still showed the most interest in data use, the difference between them is more evident, with more than two-thirds of the humanist papers and only slightly over half of the technologist papers in this category. Technologists, however, consistently had higher percentages of papers than do their humanist colleagues in the three stages of building systems. Generally speaking, humanities scholars gave more attention to the impact of digital technology on retrieving, analyzing, evaluating, and representing data taken from digital collections. Technologists, on the other hand, spread out their interests more evenly.

6. Conclusions

DH is young and continues to evolve at a fast pace. As part of an effort to better understand the domain, the current study examined the basic characteristics of a corpus from a series of DH conferences held in Taipei, Taiwan, between 2009 and 2016. Also of concern was the impact of digital technology on DH in view of the information lifecycle.

The literature being examined in the study shares some common features with others in earlier DH studies. For one thing, the study identified history and Chinese to be the top two contributing domains in the humanities, resembling the findings in another study that shows literature, linguistics, and history to be the top contributors to the DH literature. The term "humanities computing" appeared only once in the keywords provided by the authors, while the word "digital" was in several high-frequency keywords, including "digital humanities" and "digital libraries." Although the term "digital humanities" being part of the conference name might have somewhat influenced authors' word choices, this finding does match the common practice in the recent DH terminology.

Other features of the target corpus include: the high interdisciplinarity, the unusually high rate of multiple authorship as compared to the more common practice of single authorship in the humanities, the high participation of East Asian countries, and the limited international collaboration. Digital technology's impact on DH in all stages of the information lifecycle appeared to be of interest to conference participants, with the use of data from digital collections as the most popular and facilitating collaboration as the least favored. While authors from both the humanities and technology-related domains contributed papers in all areas of technology's impact, humanities scholars were more likely to lead papers on the data use stage and technologists, predominantly computer scientists, contributed more to papers on building systems, especially in the stage of building retrieval systems.

In LIS, the domain-analytic approach has been a major force. As a domain-analytic study, this research is a contribution to an improved understanding of DH, an emerging and dynamic domain. On one hand, it expanded the knowledge

臺灣大學學術期刊資料庫

about DH by examining a body of literature not included in previous research. On the other hand, its focus on digital technology's impact on DH is unique, and such a study, albeit exploratory, is no doubt a valuable addition to research on the domain of DH.

The study is only a snapshot of the DH literature, without any claim of comprehensiveness or representativeness. Its quantitative nature also prevented it from identifying or explaining more in-depth issues. In the future, studies may go in other directions by applying qualitative approaches to discovering and examining critical concepts and major theories as well as the intellectual developments and shifts in DH. Further bibliometric studies are also valuable in that they may target other DH corpora and conduct comparisons with the findings in the current study. The impact of digital technology offers a host of opportunities for further research. For example, the intense interest in technology's impact on the consumption of digital collections suggests a need to devise more detailed studies in this area.

References

Burdick, A., Drucker, J., Lunenfeld, P., Presner, T., & Schnapp, J. (2012). *Digital_Humanities*. Cambridge, MA: MIT Press.

Chang, Y. W. (2012). Tracking scientometric research in Taiwan using bibliometric and content analysis. *Journal of Library and Information Studies, 10*(2), 1-20. doi: 10.6731/TPCC_proceedings.003b-002-R2-00010619

Clement, T. E., & Carter, D. (2017). Connecting theory and practice in digital humanities information work. *Journal of the Association for Information Science and Technology, 68*(6), 1385-1396. doi: 10.1002/asi.23732

Dalbello, M. (2011). A genealogy of digital humanities. *Journal of Documentation, 67*(3), 480-506. doi: 10.1108/00220411111124550

Graf, A., & Smiraglia, R. (2013). Transition in education: Domain analysis from the *Encyclopedia of Milwaukee*. *North American Symposium on Knowledge Organization, 4*(1), 76-87. doi: 10.7152/nasko.v4i1.14647

Guimarães, J. A. C., De Oliveira, E. T., & Gracio, M. C. C. (2012). Theoretical referents in knowledge organization: A domain analysis of the knowledge organization journal. In A. Neelameghan & K. S. Raghavan (Eds.), *Categories, contexts and relations in knowledge organization: Proceedings of the Twelfth International ISKO Conference* (pp. 31-38). Wurzburg, Germany: Ergon.

Hjørland, B. (2002). Domain analysis in information science: Eleven approaches–Traditional as well as innovative. *Journal of documentation, 58*(4), 422-462. doi: 10.1108/00220410210431136

Hjørland, B., & Albrechtsen, H. (1995). Toward a new horizon in information science: Domain-analysis. *Journal of the Association for Information Science and Technology, 46*(6), 400-425. doi: 10.1002/(SICI)1097-4571(199507)46:6<400::AID-ASI2>3.0.CO;2-Y

Huggett, J. (2012). Core or periphery? Digital humanities from an archaeological perspective. *Historical Social Research, 37*(3), 86-105.

臺灣大學藝術期刊資料庫

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy, 26*(1), 1-18. doi: 10.1016/S0048-7333(96)00917-1

Liu, A. (2013). The meaning of the digital humanities. *PMLA, 128*(2), 409-423. doi: 10.1632/pmla.2013.128.2.409

McPherson, T. (2012). Why are the digital humanities so white? or Thinking the histories of race and computation. In M. K. Gold (Ed.), *Debates in the digital humanities* (pp. 139-160). Minneapolis, MN: University of Minnesota Press. doi: 10.5749/minnesota/9780816677948.003.0017

Ossenblok, T. L., Verleysen, F. T., & Engels, T. C. (2014). Coauthorship of journal articles and book chapters in the social sciences and humanities (2000–2010). *Journal of the Association for Information Science and Technology, 65*(5), 882-897. doi: 10.1002/asi.23015

Palmer, C. L. (1999). Aligning studies of information seeking and use with domain analysis. *Journal of the Association for Information Science and Technology, 50*(12), 1139-1140. doi: 10.1002/(SICI)1097-4571(1999)50:12<1139::AID-ASI18>3.0.CO;2-V

Poole, A. H. (2017). The conceptual ecology of digital humanities. *Journal of Documentation, 73*(1), 91-122. doi: 10.1108/JD-05-2016-0065

Prieto-Díaz, R. (1990). Domain analysis: An introduction. *ACM SIGSOFT Software Engineering Notes, 15*(2), 47-54. doi: 10.1145/382296.382703

Subramanyam, K. (1983). Bibliometric studies of research collaboration: A review. *Information Scientist, 6*(1), 33-38. doi: 10.1177/016555158300600105

Tang, M. C., Cheng, Y. J., & Chen, K. H. (2017). A longitudinal study of intellectual cohesion in digital humanities using bibliometric analyses. *Scientometrics, 113*(2), 985-1008. doi: 10.1007/s11192-017-2496-6

Wang, X., & Inaba, M. (2009). Structures and evolution of digital humanities: An empirical research based on correspondence analysis and co-word analysis. *Proceedings of the International Conference of Digital Archives and Digital Humanities, Taiwan, 2009*, 1-16.

(Received: 2018/1/17; Accepted: 2018/5/21)

臺灣大學學術期刊資料庫

圖書資訊學刊　第16卷第2期　（2018.12）　頁1-23
doi: 10.6182/jlis.201812_16(2).001

數位人文的領域分析：以2009-2016臺灣國際會議論文為例

Investigating Digital Humanities: A Domain Analysis of Conference Proceedings Published in Taiwan, 2009-2016

李鶴立[1]　王申罡[2]

Hur-Li Lee[1], Shengang Wang[2]

摘　要

本研究對數位人文領域中的一組論文進行分析。過去類似的研究，均以在西方發表的論文作為對象，本文的研究對象，則是2009至2016年間，在臺灣舉辦的數位典藏與數位人文國際研討會中發表的129篇論文。研究的議題有三：(1)對這組論文的基本特質像是作者人數、代表國家地區、代表領域、常用詞彙等做描述統計；(2)將資訊生命周期分為四階段，分析數位科技對數位人文的影響在各階段的分布；(3)分析數位科技對數位人文造成的階段性影響是否因領域而不同。除了對全組論文進行分析，本研究更將數據分年檢視，探討論文的基本特質和數位科技的影響是否因時間的推進呈現一些趨勢。本研究給數位人文的領域分析，帶入一個新的視角。

關鍵字：領域分析、數位人文、數位典藏與數位人文國際研討會

[1,2]美國威斯康辛州立密爾瓦基大學
School of Information Studies, University of Wisconsin-Milwaukee, Wisconsin, USA
* 通訊作者 Corresponding Author: 李鶴立Hur-Li Lee, E-mail: hurli@uwm.edu

註：本中文摘要由作者提供。
以APA格式引用本文：Lee, H.-L., & Wang, S.-G. (2018). Investigating digital humanities: A domain analysis of conference proceedings published in Taiwan, 2009-2016. *Journal of Library and Information Studies, 16*(2), 01-23. doi: 10.6182/jlis.201812_16(2).001
以Chicago格式引用本文：Hur-Li Lee and Shengang Wang. "Investigating digital humanities: A domain analysis of conference proceedings published in Taiwan, 2009-2016." *Journal of Library and Information Studies* 16, no. 2 (2018): 01-23. doi: 10.6182/jlis.201812_16(2).001

臺灣大學圖書館期刊資料庫