

# Distribution features and intellectual structures of digital humanities

## A bibliometric analysis

223

Qing Wang  
*Business School, Nankai University, Tianjin, China*

Received 24 May 2017  
Revised 14 August 2017  
Accepted 17 September 2017

### Abstract

**Purpose** – The purpose of this paper is to conduct a retrospective bibliometric analysis of documents about digital humanities, an emerging but interdisciplinary movement. It examines the distribution of research outputs and languages, identifies the active journals and institutions, dissects the network of categories and cited references, and interprets the hot research topics.

**Design/methodology/approach** – The source data are derived from the Web of Science (WoS) core collection. To reveal the holistic landscape of this field, VOSviewer and CiteSpace as popular visualization tools are employed to process the bibliographic data including author, category, reference, and keyword. Furthermore, the parameter design of the visualization tools follows the general procedures and methods for bibliometric analysis.

**Findings** – There is an obviously rapid growth in digital humanities research. English is still the leading academic language in this field. The most influential authors all come from or have scientific relationships with Europe and North America, and two leading countries of which are the UK and USA. Digital humanities is the result of a dynamic dialogue between humanistic exploration and digital means. This research field is closely associated with history, literary and cultural heritage, and information and library science.

**Research limitations/implications** – This analysis relies on the metadata information extracted from the WoS database; however, some valuable literatures in the field of digital humanities may not be retrieved from the database owing to the inherent challenge of topic search. This study is also restricted by the scope of publications, the limitation regarding the source of data is that WoS database may have underrepresented publications in this domain.

**Originality/value** – The output of this paper could be a valuable reference for researchers and practitioners interesting in the knowledge domain of digital humanities. Moreover, the conclusions of this retrospective analysis can be deemed as the comparable foundation for future study.

**Keywords** Library and information services, Digital humanities, VOSviewer, CiteSpace, Digital cultural heritage, Digital history, Digital literary, Humanities computing, Visualization analysis

**Paper type** Research paper

### Introduction

When one nowadays expects to understand one's own research field, literary analysis or otherwise, it is sensible to make use of the novel computational methods and tools and converge them with traditional research paradigm within that field (Naukkarinen and Bragge, 2016). Contemporary humanities research has expanded beyond anything that could be considered traditional, digital humanities is the umbrella term that describes much of this kind of work (Varner and Hswe, 2016). Digital humanities first emerged in the late 1940s as "humanities computing," when it formed the basis for such projects as the Index Thomisticus conceived by an Italian priest named Father Roberto Busa. Today, digital humanities are applying advanced computational tools to more diverse disciplines, ranging from history and literature to cultural studies (Mone, 2016), and even computational landscape. However, digital humanities must be understood in the context of the history,

The author is grateful to the anonymous reviewers for their constructive comments and suggestions on this study.



methodologies, and perceptions which its practitioners bring to the table (Jennifer and Kevin, 2013). Grounded in the interactive values of humanities and techniques, the digital humanities attempts to lead the humanities community into new domain by promoting experimentation, collaboration, and openness (Spiro, 2012).

As an interdisciplinary movement, digital humanities is the result of a dynamic dialogue between humanistic exploration and digital means. Digital technologies will keep coming, although digital humanities has been growing rapidly around the world, it will remain essential that uptake of technologies is a modern but necessary part of the humanities and arts, and even as broad a sense as possible (Terras, 2016). At the same time, digital humanist still need to push harder to get humanistic approaches and methods into the digital humanities (Berdan, 2013).

Digital humanities is not limited to any one discipline or field, in the digital era it has become a catch-all term for anyone who is engaging in the discovery, preservation, and interpretation of humanities materials (documents, images, sound) to enable a better and deeper understanding of current society. To some extent it is difficult to understand and interpret what digital humanities actually is, but a core feature is that it encourages researchers and practitioners to think about application probability of digital methods in traditional humanities disciplines.

The rest of this study is organized as follows. A literature review is briefly described in "Literature review" section. In "Methodology" section, the source of data, visualization techniques, and parameter design are explained. Then, the results of the bibliometric analysis are provided and interpreted in section "Results and discussion". In subsection "Yearly research output" and "Language of documents," the distribution features of publication and language are revealed through essential statistical analysis. Moreover, this paper identifies the most active contributors (research strength) at the country, institution, and author levels. The distribution of active journals is also illuminated simply. In subsection "Network configuration of category," disciplinary distribution can be detected. In "Highly cited references" subsection, the co-cited references are analyzed so that the most influential references and their distribution can be identified. In "Keywords co-occurrence network" subsection, the results of co-keyword analysis are presented. The co-occurrence network of keywords, providing link-based information on their relationship can be used to determine hot research topics. The final section of this paper presents the conclusions and discusses implication and limitations of this research work.

### Literature review

Bibliometric analysis is not new. A quick search in the Web of Science (WoS) core collection database using topic query "bibliometric analysis" resulted in more than 4,000 documents, with nearly three-quarters of those being academic articles. Bibliometric analysis has been adopted in various forms for more than a century (Pritchard and Wittig, 1981; Hood and Wilson, 2001). Since Eugene Garfield founded the Information Sciences Institute in 1958 and introduced the impact factor (IF), an increasing number of bibliometric indicators and tools have been developed (Liu *et al.*, 2015). Bibliometric analysis has been adopted in various disciplines to visualize the patterns or intellectual structures of a scientific topic and to assess the scientific outputs in a given field (Huang *et al.*, 2016; Li *et al.*, 2017).

Some studies use bibliometric tools to explore a specific field. For example, J.A. Pratt *et al.* (2012) employed bibliometric analysis to define the intellectual structures within information systems (IS) and between IS and other College of Business disciplines. N. Sinkovics (2016) identified the clusters and the major themes of 410 articles connected to ethics in marketing research with VOSviewer and NVivo. X. Wang *et al.* (2016) developed a visualization of the research trends in energy policy studies over the past 50 years using the bibliometric methods. H. Chen *et al.* (2017) discussed the state of the art of research on food

waste, based on a bibliometrics. Bibliometric analysis has also been applied to assess research productivity. For example, Zhang and Feng (2014) analyzed the growth and development of research productivity concerning artificial blood vessels with HistCite. Work has been also done on visualizing the global pattern of scientific publications by analyzing aggregated Science Citation Index data (Chen and Chen, 2016). All of these achievements can be shown as examples which can contribute to bibliometric analysis.

As the awareness of digital humanities grows, an increasing number of people from different countries, disciplines, and institutions are showing interest in this field. The interdiscipline of digital humanities has been subject to a number of bibliometric analyses. For example, M. Dalbello (2011) reconstructed the genealogy of “digital humanities” as an approach to revealing knowledge production in the humanities. A. Svensson (2016) presented a bibliographic review that seeks to investigate the extent to which Swedish research is represented in the digital humanities. G.V. Mozhaeva *et al.* (2016) described the interactional communication structures in the field of digital humanities with the help of the bibliometric analysis. Although many digital humanities studies have been conducted, none of them interpreted the spatio-temporal distribution and co-citation network of the knowledge domain in digital humanities. Moreover, it is difficult for newly entered researchers and practitioners to identify the fundamental documents and essential themes from among the massive available resources in this academic community. Thus, a holistic study of the academic publications is urgently needed for us to comprehensively review and interpret the progress in digital humanities research.

Unlike previous reviews of digital humanities research, this study conducted a bibliometric-based analysis that investigates different aspects of intellectual attributes in this fast-growing field between January 1968 and March 2017. In a nutshell, this paper addresses the questions “what has done in the field of digital humanities until now and what is next?”. Such a bibliometric analysis is beneficial for the self-understanding of this interdisciplinary field. Distribution of publications will be conducive to deeper interpretation of the spatial-temporal features of research strength and the tendency for inclusion of the growing number of countries, institutes, authors, etc. in this field. The conclusion gives insights into theories and practices that are developing and shaping potential modes of humanities research in the digital era.

One of the potential values of using visualization method for presenting the distribution features and intellectual structures of digital humanities is that we can explore changes in the digital humanities over time. This means that the differences of research strength, spatio-temporal distribution, topics, and themes in the field of digital humanities would be seen if the same research project as this is launched every decade or more. In other words, this report can be deemed as the comparable foundation for future study.

## Methodology

### *Analysis schema*

In this study, a bibliometric analysis was conducted from two aspects: distribution features and intellectual structures. Distribution features were identified as properties irrelevant to textual content of documents (Zhu *et al.*, 2015), including yearly research output, publishing languages, institutions and authors performance, active journals, author co-citation network, and categories. Intellectual structures analysis comprised knowledge base (frequently cited documents) and hot research topics. This study aims to undertake a bibliometric survey of the digital humanities research. To achieve this aim, several research questions were addressed and traced to specific bibliometric indicators (Eito-Brun and Rodriguez, 2016). The bibliometric data analysis schema is shown in Table I.

After defining the analytical perspectives and identifying analytical units, the rest steps were taken as follows:

- selection of database;
- search definition and execution;
- data collection and processing;
- visualization and obtaining the bibliometric indicators; and
- analyzing results and discussion.

*Data sources*

The definition of the databases is an important concern for conducting a bibliometric study (Landström *et al.*, 2012; Martens *et al.*, 2016). The database chosen for this study was the WoS core collection, which has about 18,000 journals carefully and objectively selected for quality, with 1.3 billion cited references going back to 1900, and provides researchers with quick, powerful access to the world’s leading citation databases. As a citation index, the WoS core collection exposes the connective tissue in the sciences, social sciences, and art and humanities, letting researchers explore the connections that scholars establish as they document their research (Clarivate Analytics, 2017).

The source data were derived from the WoS core collection for timespan “all years.” The citation database was set as “SCI-EXPANED, SSCI, A&HCI, CPCI-S, CPCI-SSH, Emerging Sources Citation Index (ESCI).” After pre-analysis and comparison, “digital humanities” and “humanities computing” and “ehumanities” and “e-humanities” were used as search subject terms. The initial retrieved results were defined by document types, including “ARTICLE,” “PROCEEDINGS PAPERS,” and “REVIEW.” To ensure the accuracy of the sample, the title and abstract of each paper was screened to confirm the result matching up with the theme or not, irrelevant results were excluded. Then the selected results were sent to file of plain text by setting the “record content” as “full record and cited references.” Data set last updated on March 17, 2017.

The repeated results were removed from the selection by checking the metadata information of publications. It was found that among all the 805 of obtained bibliographic records, two repeated papers were detected, they should be excluded from the data set. After eliminating invalid records, 803 valid bibliographic records were obtained, of which articles accounted for 66.50 percent (534) of all documents, proceedings papers accounted for 28.14 percent (226), and reviews accounted for 1.37 percent (11); 31 documents were both articles and proceedings papers.

*Visualization tools*

Bibliometricians usually do not develop their own mapping techniques, but instead they employ existing visualization tools to perform meaningful analyses in a bibliometric

Analytical perspectives	Analytical units	Major bibliometric methods
Distribution features	Yearly research output	Time-trend analysis
	Publishing languages	Straightforward counting
	Productive institutions	Spatial analysis
	Active journals	Research output analysis
	Prolific authors	Academic evaluation
	Author co-citation network	Co-citation analysis, co-authorship analysis
Intellectual structures	Frequently cited documents	Academic evaluation
	Author keyword	Co-word analysis, cluster analysis, topic analysis

**Table I.**  
Bibliometric data  
analysis schema

context (Van Eck and Waltman, 2017). The visualization tools used in this study are VOSviewer and CiteSpace.

The first software selected for visualization was VOSviewer, a freely available computer program for bibliometric mapping. It pays special attention to the graphical representation of scientific maps. In addition, It is especially useful for constructing and displaying large scientific maps in an easy-to-understand way (Van Eck and Waltman, 2010). In particular, VOSviewer employs the visualization of similarities mapping technique and the main advantage of the software is the high quality of its visual representation (Sinkovics, 2016).

CiteSpace is a Java-based scientific visualization software developed by Chaomei Chen. It would be potentially used by researchers and practitioners in different academic communities, to detect and visualize abrupt changes, emerging trends, and dynamics in scientific domains (Chen, 2006). The ideas, principles, and algorithm applied to CiteSpace are as follows: divide-and-conquer, success breeds success, pathfinder network scaling, minimum spanning trees (MST), and expectation maximization clustering method (Feng *et al.*, 2015).

In addition to the above softwares, SATI, a tool-box developed by Liu and Ye (2012) was applied to processing tabbed data and mapping the distributions of yearly research output, languages, institutions, and journals. It is worth mentioning that the provision of practical tool such as CiteSpace, VOSviewer, and SATI enables us to interpret bibliometric study of subject domains more vividly, as well as to understand an emerging research field more deeply.

### Parameter design

In this study, the parameter design of the visualization tools followed the general procedures and methods for bibliometric analysis. During processing and implementation, there are various types of nodes, such as authors, institutions, articles, and keywords, and the constructed nodes are represented variously in different networks (Song *et al.*, 2016). The well-constructed links represent co-occurrence or co-citation relationships between people, organizations, concepts or other entities in a more specific sense. Furthermore, the network node types should be befittingly selected in accordance with the actual situation of the research objectives and purport.

The higher co-occurrence frequency of the two nodes reveals their closer relationship, which is represented by a thicker line linking the two nodes in the network (Najmi *et al.*, 2017). The size of the disc and the label is proportional to co-occurrence frequency. The strength among links is measured by the Cosine metric:

$$\text{Cosine}(x, y) = \frac{C_x C_y}{\|C_x\| \|C_y\|}$$

where  $C_x C_y$  represents the co-citation counts between paper  $x$  and paper  $y$ , and  $\|C_x\|$  and  $\|C_y\|$  represent the times cited of paper  $x$  and paper  $y$ , respectively (Ruan *et al.*, 2016).

The size of a cluster has no uniform interpretation, that is due to many factors including the co-occurrence frequency of the terms, the number of terms in the cluster, the link strength which they are related to each other (Pinto *et al.*, 2015).

According to "Price Law" (Price, 1965), a classical algorithm can be used as filtering strategy to determine the candidates for core authors in a given research field, the threshold is calculated using the following formula:

$$M_c = 0.749\sqrt{N_{\max}}$$

In the above formula,  $N_{\max}$  represents the total number of articles published by the most prolific author in a given research field, and  $M_c$  means the lowest frequency of articles published by the candidate who can be identified as a core author (Zhong, 2012).

“Price Law” has wide-ranging extended applications in many fields including scientific productivity, cooperative network (Liu and Wang, 2016), core keywords detecting (Wang, 2015), etc.

Since keywords contain information regarding the core content of articles, the co-occurrence analysis of keywords (i.e. the appearance of keywords in an academic document) can be used to identify information about the formation of multidisciplinary evolving research frontiers of a particular knowledge domain (He, 1999; Su and Lee, 2010; Zhu and Hua, 2017). Besides, an analysis of keywords provides an approach to finding the hot topics (Xie, 2015).

In recent years, keyword co-occurrence networks (KCNs) are exploited for knowledge mapping. The KCN represents cumulative knowledge domain and helps to reveal meaningful knowledge components (Radhakrishnan *et al.*, 2017). In addition, a KCN-based analysis with VOSviewer provided a knowledge map and some additional insights that were not covered by the traditional systematic review and the content analysis (Mariano and Walter, 2015).

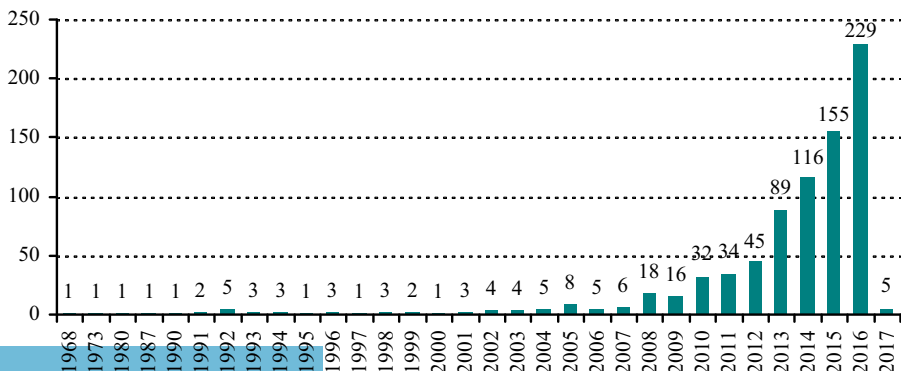
## Results and discussion

### *Yearly research output*

From the period of 1968 to 2017, the year-wise distribution of documents indexed in the WoS database is shown in Figure 1. In 1968, only one paper on digital humanities research was published, the number of publications was zero in the period of 1969-1972, the same as the period of 1974-1979, 1981-1986, 1988-1989; and in 2016, the number of publications increased to 228. This map clearly shows that the growth rate of publication in the period of 1992-2007 was slow, but since 2008, the output of documents has been growing rapidly. This forward movement indicates that the research on digital humanities was consistently the focus of scholars during the past decade, particularly in 2016.

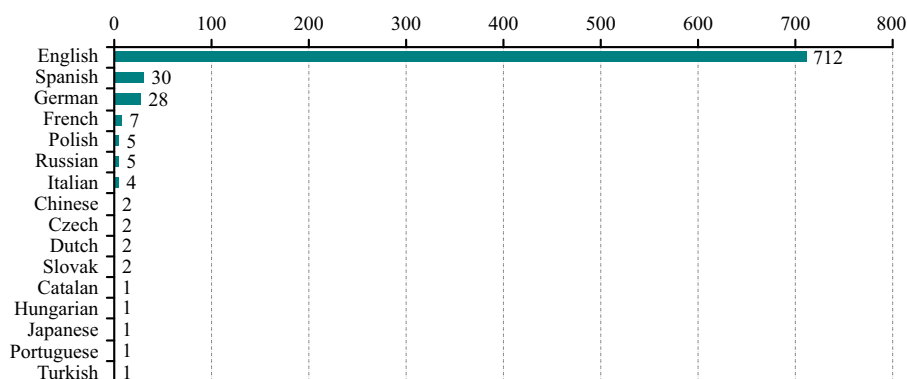
### *Publishing languages*

According to the statistical results, the scientific literatures published on digital humanities for the whole period had been written in 16 languages. The distribution of language usage is shown in Figure 2. As expected, English is the dominant used language, according to data gathered from the WoS database. Among the 16 languages, 712 articles (88.67 percent) are published in English, followed by Spanish (30, 3.74 percent), German (28, 3.49 percent), French (7, 0.87 percent), Polish (5, 0.62 percent), Russian (5, 0.62 percent), Italian (4, 0.50 percent); the number of articles published in Dutch (2, 0.25 percent) is the same as Chinese, Czech, and Slovak; a similar distribution is reflected in next ranked five languages which are Hungarian, Japanese, Catalan, Turkish and Portuguese (1, 0.13 percent). This pattern meets



**Figure 1.**  
Yearly publications of digital humanities research output from 1968 to 2017

**Figure 2.**  
Distribution of language use in digital humanities research



the expectations. The possible explanations for this might be that English is widely used in the world, functioning as an international language. In addition, some academic journals published in non-English countries/regions was not indexed in the WoS database.

#### *Productive institutions*

In terms of productive institutions, there are 681 different institutions worldwide showing research interest in this field. To identify the active and leading institutions in digital humanities research, we can analyze institution-based distribution of output to understand the productivity level of institutions around the world. Table II shows the distribution of top nine most productive institutions ranked by the total number of publications. The nine most productive institutions published 147 papers in the area of digital humanities research, which contributed to 18.31 percent of the total output. Among the top nine institutions, four are in the USA, two each in the UK and the Netherlands, and one in Germany. As shown in Table II, institutions from the UK (King's College London, University College London) ranks in top two, followed by three institutions from the USA (Indiana University, University of Illinois and Texas A&M University). It shows the dominance of English-speaking countries, it also indicates that higher education institutions are important innovation systems and play an active role in promoting the sustainable development of digital humanities research.

#### *Active journals*

A total number of 502 journals were involved in digital humanities research, the top ten active journals are shown in Table III. This top ten most productive journals had published

Rank	Institution	Country	TP <sup>a</sup> (%)
1	King's College London	UK	24 (2.99)
2	University College London	UK	23 (2.86)
3	Indiana University	USA	22 (2.74)
4	University of Illinois	USA	17 (2.12)
5	Texas A&M University	USA	14 (1.74)
6	Universiteit Utrecht	Netherlands	13 (1.62)
7	Vrije Universiteit Amsterdam	Netherlands	12 (1.49)
8	Universität Leipzig	Germany	11 (1.37)
8	University of California, Los Angeles	USA	11 (1.37)

**Table II.**  
Top nine productive institutions contributing to digital humanities research

Note: <sup>a</sup>TP means the number of total articles

**Table III.**  
Top ten active  
journals of digital  
humanities research

Rank	Journal	TP (%)	Country
1	<i>Literary and Linguistic Computing</i>	48 (5.98)	UK
2	<i>Computers and the Humanities (C&amp;H)</i>	26 (3.24)	The Netherlands
3	<i>Digital Humanities Quarterly (DHQ)<sup>a</sup></i>	23 (2.86)	ADHO
4	<i>Digital Scholarship in the Humanities</i>	20 (2.49)	UK
5	<i>Evaluation</i>	19 (2.37)	USA
6	<i>Bibliothek Forschung und Praxis</i>	14 (1.74)	Germany
7	<i>Historical Social Research Historische Sozialforschung</i>	13 (1.62)	Germany
8	<i>International Journal of Humanities and Arts Computing</i>	10 (1.25)	UK
8	<i>LREC 2014-Ninth International Conference on Language Resources and Evaluation</i>	10 (1.25)	Iceland
8	<i>Technologies for E-learning and Digital Entertainment, Proceedings</i>	10 (1.25)	Germany

**Note:** <sup>a</sup>DHQ is funded and published by the Alliance of Digital Humanities Organizations

193 papers, contributing to 24.03 percent of the research findings. The top most productive journal was *Literary and Linguistic Computing (LLC)*, which is the former title of the journal now known as *Digital Scholarship in the Humanities (DSH)*, LLC was renamed as *DSH* from the beginning of 2015. This is followed by the *Computers and the Humanities (C&H)*, *Digital Humanities Quarterly (DHQ)* and *DSH*, which published 26, 23, and 20 articles, accounting for 3.24, 2.86 and 2.49 percent of the total output, respectively. Table III shows that among the top ten journals, three each in the UK and Germany, one each in the Netherlands, the USA, and Iceland. It also indicates that there are three categories (classified by *JCR*) that maintain their dominance within this field, which are linguistics, computer science, and interdisciplinary applications.

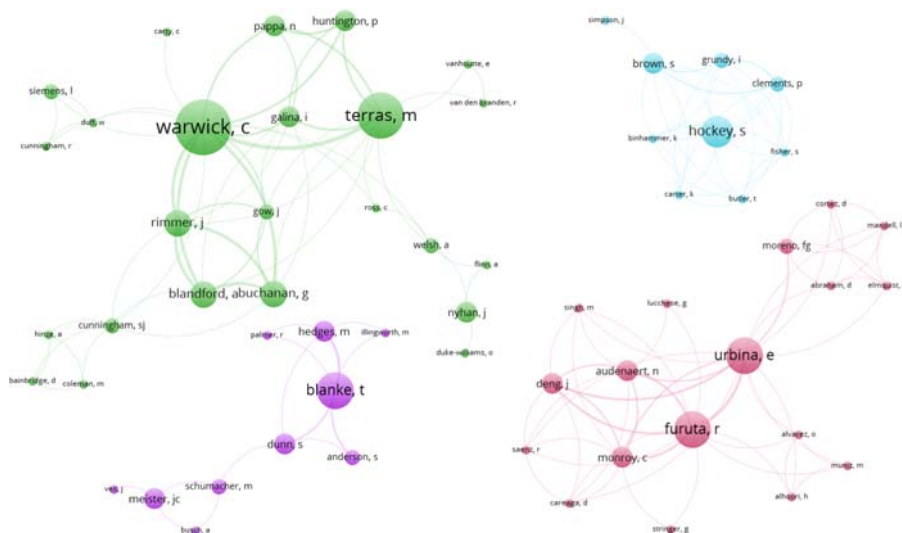
It should be mentioned that *DHQ* is a digital journal published by ADHO, with a commitment to co-publishing articles with *LLC*. However, all of the articles published in *DHQ* do not have human-assigned keywords. The WoS assigns KeyWords Plus to these articles when they are indexed in ESCI database. ESCI was launched in late 2015 as a new database within the WoS. Therefore, *DSH* is included from issue 10 (2016). *DSH* is the successor to and the continuation of *LLC*. After almost 30 years of publishing scholarly papers in this field, the journal's name did not cover the subject anymore. In addition, from 2005 onwards, the *C&H* appeared as *Language Resources and Evaluation*, the official journal of the European Language Resources Association.

#### *Prolific authors*

The analysis of co-authorship can help us understand collaborative network of different authors and detect the influential researchers on digital humanities around the world. By selecting the unit of analysis and setting the appropriate thresholds, the collaborative network of the most productive authors generated, as Figure 3 shows.

According to the results, there are 1,547 authors showing interest in this field. Table IV lists the top five prolific authors ranked by their productivity of digital humanities literature. As rank list shows, among the five top authors, three are from the UK, two are from the USA. Claire Warwick tops this list, as she produced ten papers during the period covered by this survey. The highest citation counts within this group of five prolific authors were also received by Claire Warwick. Melissa Terras who ranked second on the basis of article productivity also ranked second in the total citations criteria. In terms of authorship patterns, it shows that about 46.82 percent (376 out of 803) documents are created by a single author. In other words, the remaining 53.18 percent documents are produced collaboratively, which reveals that collaboration in digital humanities research is clearly noticed. It is further observed that fewer documents are created as a collaborative effort





**Figure 3.**  
The collaborative network of prolific authors

Rank	Author	Country	Institution	TP <sup>a</sup> (%)	Citations
1	Warwick, Claire	UK	University College London	10 (1.25)	103
2	Terras, Melissa	UK	University College London	8 (1.00)	71
3	Blanke, Tobias	UK	King's College London	6 (0.75)	21
4	Furuta, Richard	USA	Texas A&M University	6 (0.75)	4
5	Urbina, Eduardo	USA	Texas A&M University	6 (0.75)	3

**Note:** <sup>a</sup>TP means the number of articles published by this author

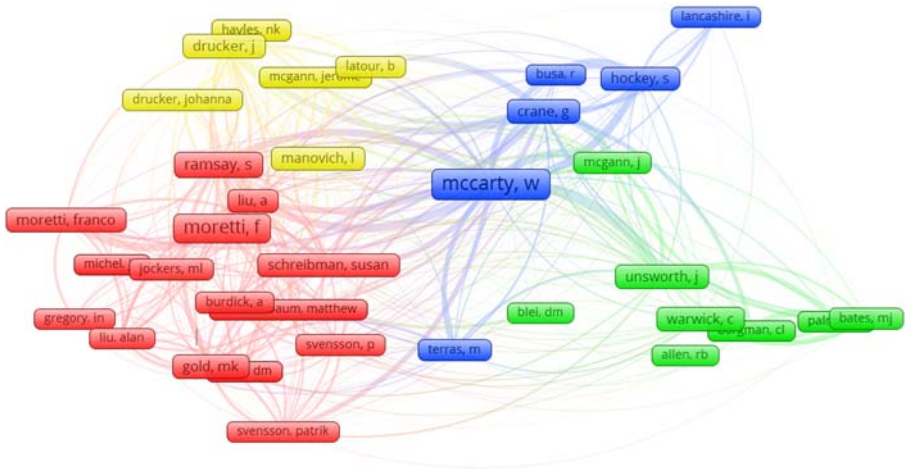
**Table IV.**  
The top five most productive author

between five authors or more. There are four documents produced by more than ten authors in collaboration. However, the authorship patterns further indicated that transnational collaboration was almost non-existent in the field of digital humanities research.

### Author co-citation network

The collaboration network of scientific communities can be identified and interpreted by author co-citation analysis which focuses on interrelationships among individual authors. This method offers an functional tool for representing a landscape of author co-citation network. Moreover, some other “invisible” information (i.e. institution, country and co-citation frequency) about these influential authors can also be detected out. Figure 4 plots the author co-citation network, which is comprised of 35 most cited contributors and 431 links, in which the size of each node corresponds to the weight of citations of the author it represents. The links reflect a cooperative relationship between two authors.

According to the number of co-citation frequency, the most highly cited authors and relevant information can be extracted from the author co-citation network. As depicted in Figure 4, Franco Moretti (with 102 citations, Stanford University, USA), McCarty Willard (79, King’s College London, UK; University of Western Sydney, Australia) and Stephen Ramsay (46, University of Nebraska-Lincoln, USA) occupied the top three positions regarding the total co-citation frequencies, followed by Matthew K. Gold (43, City University of New York, USA),



**Figure 4.**  
Author co-citation network

Johanna Drucker (41, University of California at Los Angeles, USA), Gregory Crane (41, Tufts University, USA), John Unsworth (40, University of Virginia, USA). It should be pointed out that the co-citation network includes two identical people (Moretti, Franco and Moretti, F) who ranked on the top of the list.

Notably, the influential authors' research covered various majors and disciplines including literary history, distant reading, literary criticism, text analysis, and philosophical issues related to the use of technology in digital humanities. This analysis suggests that these authors' publications had an important impact on digital humanities research field and promoted the evolutionary development of research theme. From a geographical distribution, the result proved that this research field is overwhelmingly dominated by Anglo-American scientific communities.

*Network configuration of category*

Disciplinary distribution can help us detect to what extent a given field is shaped by the convergence of disciplines and their respective role (Liu *et al.*, 2015). According to the journal subject classification rules of the WoS database, each paper published by that journal is assigned to more than one category. For example, all papers published in the *Interdisciplinary Science Reviews* are assigned to two categories: multidisciplinary science and social sciences, and interdisciplinary. Co-occurrence analysis of category is an effective way to identify the disciplines involved in a given research domain. Figure 5 shows



**Figure 5.**  
Network configuration of categories involved in digital humanities research

the co-occurrence network of categories involved in digital humanities after pruning by MST algorithm in CiteSpace.

According to the results, over 50 subject categories were related to this research field. The most common category is computer science, which is the largest node with a frequency of 254, followed by information science and library science (131), literature (119), and linguistics (108). Other influential categories with frequency of more than 50 but less than 100 include arts and humanities, other topics (84), language and linguistics (84), humanities (75), and history (54). The results indicate that the study of digital humanities is an interdisciplinary subject. Moreover, from the diverse names and publication trends of top journals, it can be confirmed again that the phenomenon of interdiscipline is existed in this field, it prompted current researchers and potential entrants to expand the relevant topics from multiple angles.

### Highly cited references

The references analysis is one key aspect of bibliometrics, which can identify the influence of the authors and the documents, and it is also useful in detecting the dynamics, structure, and patterns of a given research field. The intellectual structure of highly cited documents can be developed by means of CiteSpace, the top ten most cited references and other mentionable documents in the co-citation network are selected for specific analysis. Although the operation of CiteSpace is a simple process, some filtering strategies should be considered in key setting for the co-citation analysis and simplification. Considering the number of documents, the “Top N per slice” is set as  $N=50$ , that means 50 most cited or occurred items are selected from each slice. Figure 6 shows the salient intellectual structure of co-cited references using pruning algorithm based on MST. With the help of CiteSpace, the document co-citation network is obtained, as Figure 6 shows. This network consists of 102 nodes and 127 links.

From the network summary table exported by CiteSpace, the top ten cited references can be found according to their co-citation frequency, as Table V lists. From the perspective of co-authorship, five multi-authored publications are included in the list of the ten most co-cited references, and the other documents were single authored. These single authored but highly cited publications are all monographs, among these works, two are created by Italian literary scholar Franco Moretti who was named to the American Academy of Arts and Sciences in 2006. From a geographical perspective, the top ten most-cited authors were all serving as researchers (instructors) at universities or academic institutions. They are distributed as follows: seven from the USA, two from the UK, and one from Ireland. Evidence indicate that this field is overwhelmingly dominated by North America and Europe. By tracing the curriculum vitae, it can be found that most of these scholars have different experiences of scientific research, and the affiliation of them changed over the past decade, especially for senior experts.



Figure 6.  
Map of highly cited  
references

Rank	Author/Editor	The title of document	PY <sup>a</sup>	FC <sup>b</sup>	TS <sup>c</sup>
1	Matthew K. Gold	<i>Debates in the Digital Humanities</i>	2012	39	Edited book
2	Susan Schreibman <i>et al.</i>	<i>A Companion to Digital Humanities</i>	2004	30	Edited book
3	Anne Burdick <i>et al.</i>	<i>Digital Humanities</i>	2012	28	Edited book
4	Franco Moretti	<i>Distant Reading</i>	2013	26	Monograph
5	Franco Moretti	<i>Graphs, Maps, Trees: Abstract Models for Literary History</i>	2005	24	Monograph
6	Willard McCarty	<i>Humanities Computing</i>	2005	22	Monograph
7	Jean-Baptiste Michel <i>et al.</i>	Quantitative analysis of culture using millions of digitized books	2011	22	Journal article
8	Matthew L. Jockers	<i>Macroanalysis: Digital Methods and Literary History (Topics in the Digital Humanities)</i>	2013	22	Monograph
9	David M. Berry	<i>Understanding Digital Humanities</i>	2012	21	Edited book
10	Stephen Ramsay	<i>Reading Machines: Toward an Algorithmic Criticism (Topics in the Digital Humanities)</i>	2011	15	Monograph

**Table V.**  
Ranking of top ten  
most co-cited  
references

**Notes:** <sup>a</sup>PY indicates the publication year of co-cited documents; <sup>b</sup>FC means co-citation frequency of co-cited documents; <sup>c</sup>TS indicates the type of co-cited documents

Of the references, five publications with high co-citation counts are also worth a mention. These publications focus on describing latent Dirichlet allocation, a generative probabilistic model for collections of discrete data (Blei *et al.*, 2003); introducing a humanities approach to the graphical expression of interpretation (Drucker, 2011); reviewing the history of humanities computing (Hockey, 2004); providing an overview of how the term “humanities computing” developed into the term “digital humanities” (Terras *et al.*, 2013); and re-orienting humanities scholarship by engaging the technology and specifically directing it to the subject matter of the humanities (Bodenhamer *et al.*, 2010).

According to the issues discussed in each publication, the top ten most co-cited references can be divided into two subject areas: the exploration of multiple possibilities and tensions (SA1) and the paradigm revolution in particular disciplines (SA2). SA1 includes the 2012 Matthew K. Gold edited book, the 2004 Susan Schreibman edited book, the 2012 Anne Burdick monograph, the 2005 Willard McCarty monograph, and the 2012 David M. Berry edited book. SA2 includes the other publications listed in Table V.

#### *SA1: the exploration of multiple possibilities and tensions*

In all the references, the 2012 book edited by Matthew K. Gold (2012) is the most frequently cited work. *Debates in the Digital Humanities* provided not only a valuable primer for those newer to this field, but also a detailed survey of six of the troupes: defining, theorizing, critiquing, practising, teaching, and imagining (Kahn, 2013). Moreover, several essays pointed out that the practitioners should embrace new ideas, collaboration, and knowledge sharing (Gold, 2012). When it comes to academic contribution, *Understanding Digital Humanities*, edited by David M. Berry (2012), also provided an introduction to important debates surrounding issues raised when using algorithmic techniques in the field of arts and humanities. It also discussed the new forms of collaboration in an interdisciplinary way, such as research terms, new organizational structures, fresh techniques, and methodologies (Berry, 2012).

As can be seen, another most influential publication is a 2004 book edited by Susan Schreibman *et al.* (2004). The book entitled *A Companion to Digital Humanities* offered a concise, thorough overview of this emerging field. This work consisted of 37 original papers created by leaders in this field. It focused on specific applications, methods, and basic

principles of humanities computing. Other related issues were also discussed, such as production, dissemination, and archiving of digital works (Schreibman *et al.*, 2004). Similarly, *Digital Humanities* provided an in-depth examination of this emerging field. The central concern of this report is that contemporary culture is related to the methodologies and techniques unfamiliar to traditional modes of humanistic inquiry. As a game-changing report, *Digital Humanities* is an critical tool for rethinking the formation of new academic ecosystem, and it is also a vision statement for the future (Burdick *et al.*, 2012).

The 2005 monograph written by Willard McCarty (2005) is widely regarded as a foundational text. *Humanities Computing* helps us understand the digital humanities scholarship from perspectives of philosophy, history, ethnography, and criticism. The contribution of this book is that it provides a rationale for a computing practice in the field of humanities, strengthens current computing practice in almost all disciplines, and advocates cross-border cooperation. In addition, this landmark book introduced new approaches to conceptualizing computing practices and fulfilling the basic mandate of humanities computing. At the same time, it provided those with whom they interact a new way of collaboration, outlined some useful suggestions for a productive relationship, as well as an agenda for individual contribution (McCarty, 2005).

#### *SA2: the paradigm revolution in particular disciplines*

The application of digital techniques and computational analysis in the humanities may result in new changes for the research on literary history and literary interpretation, although these emerging paradigms may be unfamiliar to traditional modes of humanistic inquiry. In the 2005 book entitled *Graphs, Maps, Trees: Abstract Models for Literary History*, Franco Moretti (2005) provides charts, maps, and time lines for the further analysis of literary history. He argued that the concept of esthetic form could be redefined (Moretti, 2005). These ideas are reflected in his 2013 monograph (Moretti, 2013). The essays in *Distant Reading* led to a new and often contested paradigm of literary analysis. This book followed two decades of conceptual development of “distant reading,” constructed a growing field of unorthodox literary studies, and this contested paradigm of literary analysis looked significantly different from others (Moretti, 2013).

The application of digital methods in literary interpretation is a new proposition. However, it is also controversial. *Macroanalysis: Digital Methods and Literary History*, published by the University of Illinois Press (UIP), introduced a new approach to literary interpretation, the author of this book suggested that researchers can draw conclusions by using digital methods. Quantifiable evidence is an important source of reference for analyzing literary trends across periods, and this approach can help us to better understand the linguistic patterns of individual works (Jockers, 2013). *Reading Machines: Toward an Algorithmic Criticism*, also published by UIP, discussed the importance of computational text analysis used as new approach to exploring possibilities for critical interpretation. The author of this book argued that this new form of text analysis could be employed for literary critics to penetrate the core activity of literary studies (Ramsay, 2011).

The study entitled “Quantitative analysis of culture using millions of digitized books” by Michel *et al.* (2011) is the only highly co-cited article in Table V. This paper was published on a top journal in multidisciplinary science area, that is *Science*. In this quantitative investigation, the researchers created a corpus of digitized texts containing about 4 percent of all books ever published, focused on linguistic and cultural phenomena, and investigated cultural trends by computational analysis. It showed that quantitative analysis could provide insights about diverse fields, such as lexicography, the evolution of grammar, and historical epidemiology. This approach extends the boundaries of rigorous quantitative inquiry to a wide array of new phenomena spanning the social sciences and the humanities (Michel *et al.*, 2011).

*SA3: the information services for humanities scholars*

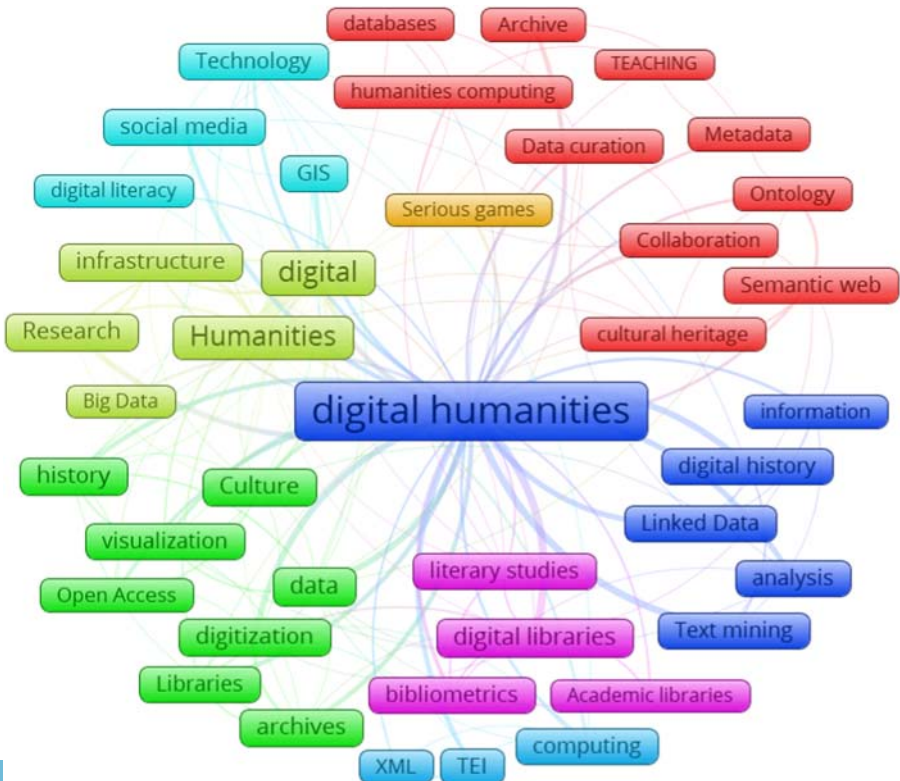
It is worth mentioning that a non-negligible subject area was detected in the network, as shown in the upper right corner of Figure 6. According to the topics discussed in the highly cited documents, this subject area focused on the information services for humanities scholars (SA3), which is a result of the study to understand the needs and uses of humanities scholars.

As Figure 6 shows, the most influential work in the subject area of SA3 is a 1996 article created by Stone Sue (1982). This study entitled “Humanities scholars: information needs and uses” reflects the fact that it is concerned almost exclusively with humanities scholars in universities and on the role of academic libraries in meeting their needs. This review does not provide librarians with clear guidelines as to how they should proceed in terms of meeting the needs of humanities scholars, and some of it seems more likely to confuse than elucidate (Stone, 1982).

*Keywords co-occurrence network*

To reveal and illustrate the hot research topics better, top 41 keywords are selected as core terms with reference to the central idea of “Price Law” (Price, 1965; Zhong, 2012). By running VOSviewer, a clusters view of keywords co-occurrence network is generated in this manner, as Figure 7 shows.

The co-occurrence network presents descriptors related to the topic of digital humanities, while also providing link-based information on their relationship. Of the 41 keywords selected, that of largest centrality (the most frequently used keyword) was “digital humanities,”



**Figure 7.**  
Map of keywords  
co-occurrence network

which occurred 261 times and with a strong relationship with other keywords, especially “digital libraries,” “text mining,” “digital history,” “digitization,” “data,” “linked data,” and “analysis.” It should be noted that some terms arguably belong to more than one cluster, serving as the general noun phrases such as “data” and “technology.” From the analysis of both meaning of the terms and their relationship with each other, one can infer the inherent dynamism and complexity of the digital humanities field, in which the clustering of general terms stands for.

The top 40 keywords from Figure 7 are extracted according to the co-occurrence frequency, as shown in Table VI, which lists the top 40 keywords having a co-occurrence frequency of more than 5, except for the descriptor “digital humanities” being retrieved as the subject term. After an integrated analysis of associated information in Table VI and Figure 7, the top 40 keywords can be logically and semantically grouped into four main hot research topics.

*Library and information services for digital humanities project.* Digital humanities is an emerging field of theory and practice. It has already made an important impact on the traditional humanities disciplines as well as the information and library science (ILS) professions (and vice versa). In terms of similarity relationship, ILS and digital humanities are both interested in studying recorded information and often share institutional frameworks (Koltay, 2016). The ILS community has proven especially receptive to the field, many librarians and information science scholars have noted that libraries serve as one of the core contributors to the development of digital humanities (Wong, 2016). Moreover, much scholarly work looks at the ways in which ILS professions bring new approaches to exploring intellectual services for digital humanities research (Poole, 2017).

With an exponential growth in the availability of digital libraries, electric databases and digitalized resources are no longer restricted to particular research field as they also encompass the humanities, arts, and social science (Ferrer, 2011). New technologies and methodologies bring humanities researchers the advantages as well as the challenges. Collaboration with the library is the only realistic option for long-term sustainability of digital humanities research projects in the e-science environment (Kretschmar and Potter, 2010). For patrons, they can obtain new digital literacy skills by integrating digital humanities projects into instructional services programs in libraries, and the library also benefits from this engagement (Pun, 2015).

Creating and sharing humanities data are important aspects of digital humanities research projects. However, as a critical component of the data ecology, greater scrutiny of data tools are often neglected. With the support of viable free and open source software, librarians can use these low-barrier-to-entry data tools to address data access gaps while

Keywords	CF <sup>a</sup>	Keywords	CF	Keywords	CF	Keywords	CF
Humanities	47	GIS	10	Big Data	8	literary studies	6
digital	38	history	9	Open Access	7	Academic libraries	6
digital libraries	18	humanities computing	9	Culture	7	bibliometrics	6
digital history	14	Linked Data	9	Collaboration	7	Serious games	6
Research	13	infrastructure	9	computing	7	Metadata	6
Text mining	12	analysis	8	archives	7	digital literacy	6
Technology	12	Libraries	8	information	7	Ontology	6
data	12	Semantic web	8	visualization	7	XML	6
social media	10	Archive	8	databases	6	TEACHING	6
digitization	10	TEI	8	Data curation	6	cultural heritage	6

Note: <sup>a</sup>CF indicates co-occurrence frequency of the keyword

**Table VI.**  
The top 40  
co-occurrence  
frequency out of 2,269  
different keywords

promoting relevant participation in digital humanities scholarship (Rath, 2016). As librarians increasingly support digital humanities publishing platform, they must also understand the user experience of these tools and pay attention to the lessons for digital literacy instruction (Tracy, 2016). In sum, librarians take an active role in serving digital humanities community as critical mediators, content curators, data and information scientists, instructors, consultants, and reliable collaborators.

*Digital history.* The internet has been an essential medium in academia since the mid-1990s, and digitization has been regarded as a promising development trend in field of humanities research. Digital humanities seem to be omnipresent in recent years and the discipline of history is no exception. Within a broader historical context of development of the digital humanities, it argues that there is too much emphasis on digital data and tool, while little attention is being paid to how historical science is changing as a result of computing turn (Zaagsma, 2013). Furthermore, digital history can offer contributions to the future of the history of education and give new meaning to existing concepts within the related disciplines (Van Ruyskensvelde, 2014).

The present status of digital history studies shows that the history project benefits from applying digital humanities tools to the study, obtaining new answers to old questions without denying the benefits of digital archives and methods, according to enthusiastic supporters and practitioners. Some researchers discussed the fundamental question what this methodological innovation means for historical science, as well as how does this affect the historian's use of web archives and digital materials (Brugger, 2012; Piersma and Ribbens, 2013). Although the practice of digital history will have to be based on critical analysis of digital data, this does not mean that the hermeneutic tradition of humanities is obsolete and the application of historical contextualization and classical source criticism is indispensable for digital historians (Fickers, 2013).

Being a result of the digital technology converged with historical science, digital history methods and techniques have shown broad application in humanities science, especially the historical review, the historical geography, the history of psychology, and the journalism (digital analysis of the contents of old journals and newspapers). The digital methods applied in the above disciplines are various, it is a common phenomenon that an array of novel methods has been applied in one digital history project (Brotton, 2014; Baena-Sanchez *et al.*, 2014; Green, 2016). In addition, the special attention should be drawn to the integrated collocation and best practices in the international scholarly community.

*Digital literary.* The application of methodologies and research practices from digital technology to social media is having a great impact on the way literary studies is being conducted. Inspired by the increasing availability of online text corpora and large document collections, on the one hand, literary scholars can adopt digital tools and computational approaches to explore questions in the field of literature from new perspectives (Grayson *et al.*, 2016); and on the other hand, the digital literacy analysis will also lead to the development of new methods and data models for related fields, such as information science (Adams and Gahegan, 2016).

As a typical interdisciplinary field, literary studies with digital methods is always an emphasis which is concerned by literary researchers and enthusiasts. Information visualization, which rely on geo-information system (GIS) and linked data analysis, has been the most commonly used tool for literary researchers to explore untapped literary collections (Hinrichs *et al.*, 2016), and so on. Other frequently used methods include TEI, XML, and text analysis. Situated at the intersection of the arts and sciences, visual geo-literary and historical analysis are contributing to new knowledge systems emerging in the digital humanities (Travis, 2015). In addition, the researchers also need to reflect on the challenges and the response measures when conducting visualization application in literary studies (McCurdy *et al.*, 2016).



Quantitative methods for analyzing literary works have existed for many years and digital technology can also facilitate collaboration and data sharing within and outside the domain of literary study (Kaltenbrunner, 2015). Many applications have been developed for automatic analysis and some systems are more user friendly. However, they are often fail to meet the needs of a specific research community because of their inherent complexity: technical skills are required to use them and to check up on the results (Moretti *et al.*, 2016). The study shows that more attention should be paid on cultivating digital literacy of researchers and practitioners who have demonstrated or expressed interest in digital literary domain.

*Digital cultural heritage.* The protection of cultural heritage is growing rapidly in response to a rise in Web use. What humanists mostly work on, and which forms much of the contents of institutional repositories, are surrogates of originally analog artifacts (Schmidt, 2012). Digitalization is a process that allows the reproduction of cultural heritage by digital methods and techniques (Giordano, 2016). The popular topics and related issues surrounding digital cultural heritage studies were discussed from diverse dimensions and perspectives, which comprised of ontology, metadata, GIS, mobile app, visual analysis, semantic web, digital archive, digital preservation, digital curation, digital edition, digital library, digital exhibition, etc.

As a precious resource with multiple values, cultural heritage can be presented to the public in a variety of forms by means of modern techniques. It is never too late to make everyone aware of cultural heritage, especially present new applications in a way attractive for the new generations (Ruttkey, 2014). And the facts have shown that sustaining cultural heritage is the common responsibility shared by the international community as well as individual level. For instance, global media represents and transmits the intangible cultural heritage of nation states officially safeguarded by UNESCO (Pietrobruno, 2014).

Digital humanities methods, both quantitative and interpretive can be employed to analyze and present the data in digital archives of cultural heritage (Varela, 2015). Under the impetus of open access movement, increasing availability of digital cultural heritage content provides humanities researchers more opportunities to share and reuse of digital data, although issues of licensing persist (Terras, 2015). Another practical challenge is that some digital tools, even if advanced, are not always able to transfer the different levels of information about cultural heritage (Zaggia *et al.*, 2015). In this context, new methods and collaborative models need to be explored for cultural heritage studies with the aim of sharing collective knowledge and values.

### Conclusion and implication

Digital methods and techniques are rapidly having an important impact on the development of humanities, arts and social sciences (HASS). A steady increase in focus on digital humanities can be found in this study, which indicates that digital humanities has been one of the more popular research domains in the digital age. This study aimed at providing insights about the progress of digital humanities research based on 803 records collected from the WoS database. Basic situation can help people get a general understanding about digital humanities research from different perspectives: yearly research output, language of documents, productive institutions, active journals, prolific authors, and categories. The interpretation of highly cited references along with keywords distribution analysis were used to represent intellectual structures, including influential studies and hot research topics. VOSviewer and CiteSpace as popular visualization tools were applied to process the bibliographic data including authors, categories, references, and keywords.

According to the bibliometric analysis of publication outputs, the results show that while the digital humanities is not a relatively “young” field, there is a rapid growth in

productivity of researchers in this domain. Significant growth of scientific outputs means that digital humanities research has progressively accelerated in the last decade. It can be predicted that the number of publications on this topic will continue to grow in the future. In terms of languages distribution, English was still the leading academic language.

The publication ownership interpreted in this paper consists of three levels that comprise different perspectives. At the micro (authors) level, the productive authors were identified in the authorship analyses, Warwick, Claire from University College London was the most prolific author, and Terras, Melissa from the same institute produced more high articles than other authors, among the top five active authors. At the meso (institutional) level, King's College London, University College London, and Indiana University were the most productive institutions in digital humanities research. At the macro (countries) level, the co-citation analysis indicated that the ten most influential authors all come from or have scientific relationships with Europe and the USA. In particular, the UK and the USA were two leading countries in this domain, and the inter-institutional collaboration was more prevalent than the international cooperation.

In addition, the comparative analysis showed that not all highly active authors had the same academic impact on digital humanities research. In other words, certain authors who did not publish larger number of papers but did obtain more co-citation counts.

Among journals in the selected database, the top three active journals were *LLC*, *C&H*, and *DHJ*. It needs to be explained, several journals have changed their names and some other journals relating to this research topic were not indexed in the WoS database.

When it comes to disciplines distribution, over 50 subject categories involved in this field, the most common categories, to some extent, were computer science, information science and library science, literature and linguistics, reflecting in part the institutional setting for convergence of digital technologies and HASS.

Regarding highly cited references, co-cited publication analysis identified the importance of edited book, monograph, and journal article. The result shows that the edited books and monographs have played important roles in improving the evolution of digital humanities research. The intellectual bases of digital humanities consist of three aspects: the exploration of multiple possibilities and tensions, the paradigm revolution in particular disciplines, and the information services for humanities scholars. The references co-citation analysis also provided insights about the fundamental issues and potential trends of digital humanities research.

Furthermore, the co-occurrence analysis of keywords, providing a cluster view of hot research topics during these years, is the valuable contribution provided in this study. To summarize, the subject field of digital humanities, located at a disciplinary crossroads, is grouped into four main hot research topics: library and information services for digital humanities project, digital history, digital literary, and digital cultural heritage. It should be noted that these four sub-categories of digital humanities all appeared in the context of digital environment, and the related themes were deeply influenced by the revolution and application of modern information technology. In other words, digital technology is likely to be a pivotal factor in the evolution of humanities research in future. These four concepts could be an excellent starting point for further investigation into how digital methods and techniques are affecting the innovation process of HASS. Another potential interesting research avenue could be the exploration of the relationship between the application of novel techniques and humanities education, that is, the extent to which digital humanities is understood by academic community.

Digital humanities provides us an interdisciplinary perspective to observe and reconsider the relationship between humanities and technology. This paper explores the distribution features of digital humanities and its intellectual structures. This can be helpful for established researchers to identify white spots, as well as for new entrants to gain a

holistic overview of existing research in the digital humanities domain. More concretely, this study provides insights regarding prolific institutions and scholars, core contents of focus, the status and potential direction of this domain. Moreover, for both researchers and practitioners, this study raises awareness of the benefits of digital methods and techniques for humanities research, the analysis offers valuable information regarding the approach to conducting future project, the reported results permitted the identification of significant researchers in this field and promoted collaborative development between the different social organizations and academic communities.

### Limitations

Although the view provided valuable insights into the current status and intellectual structure of digital humanities, it has to be understood that some deficiencies still exist, while some research biases or limitations are inevitable.

The analysis relied on the information provided in the publications. However, it is still likely that thematically relevant literature cannot be retrieved from the database if a set of terms used for the topic search do not appear in the titles, abstracts, and keywords (Kim *et al.*, 2014). The inherent challenge of keyword-based search lies in the vocabulary mismatch problem (Deerwester *et al.*, 1990). Furthermore, there is no standard approach to select keywords or determine hot research topics, some processes have to be manually handled according to specific conditions.

This paper set out to explore a visualization method to illustrate the intellectual structure of digital humanities. However, any type of visualization is only an approach and will never be perfect because a real intellectual structure is very complicated (Su and Lee, 2010). The study is also restricted by the scope of publications. The limitation regarding source of data is that the WoS database may have underrepresented publications in this domain, while some other document types such as edited book and monograph may be valuable for analysis. For example, citation counts might also include other publication categories, such as monographs, edited books, reports, and conference proceedings (Nederhof, 2006). In addition to the problem of books and citations, the coverage of languages also causes intractable problems for bibliometrics. The dominance of English grants *de facto* advantage to English-language journals and researchers of English-speaking countries (Bouchard *et al.*, 2015). Furthermore, the social sciences are more nationally oriented than the natural sciences, but the related bibliometric problems always exist (Hicks, 1999). It is difficult to achieve full coverage of international social science and humanities literature and bibliometric consequences. It is desirable to take more types of publications into account in the future with a view to offering a more comprehensive description of digital humanities.

### References

- Adams, B. and Gahegan, M. (2016), "Exploratory chronotopic data analysis", in Miller, J.A., Osullivan, D. and Wiegand, N. (Eds), *Geographic Information Science (GIScience)*, Springer, Montreal, pp. 243-258.
- Baena-Sanchez, F., Fernandez-Travieso, C., Espejo-Cala, C. and Diaz-Noci, J. (2014), "Encoding and mapping news. Applying digital humanities methods to research on journalism in the modern age", *Profesional de la Informacion*, Vol. 23 No. 5, pp. 519-526.
- Berdan, J. (2013), "The emerging field of digital humanities: an interview with Johanna Drucker", *InterActions: UCLA Journal of Education and Information Studies*, Vol. 9 No. 2, pp. 1-8.
- Berry, D.M. (2012), *Understanding Digital Humanities*, Palgrave Macmillan, London.
- Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003), "Latent dirichlet allocation", *Journal of Machine Learning Research*, Vol. 3 Nos 4/5, pp. 993-1022.

- Bodenhamer, D.J., Corrigan, J. and Harris, T.M. (2010), *The Spatial Humanities: GIS and the Future of Humanities Scholarship*, Indiana University Press, Bloomington, IN.
- Bouchard, L., Albertini, M., Batista, R. and de Montigny, J. (2015), "Research on health inequalities: a bibliometric analysis (1966-2014)", *Social Science & Medicine*, Vol. 141, September, pp. 100-108.
- Brotton, J. (2014), "Maps online: digital historical geographies", *Journal of Historical Geography*, Vol. 43 No. 1, pp. 169-174.
- Brugger, N. (2012), "When the present web is later the past: web historiography, digital history, and internet studies", *Historical Social Research – Historische Sozialforschung*, Vol. 37 No. 4, pp. 102-117.
- Burdick, A., Drucker, J. and Lunenfeld, P. (2012), *Digital humanities*, MIT Press, Cambridge.
- Chen, C. (2006), "CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature", *Journal of the American Society for Information Science and Technology*, Vol. 57 No. 3, pp. 359-377.
- Chen, H., Jiang, W., Yang, Y., Yang, Y. and Man, X. (2017), "State of the art on food waste research: a bibliometrics study from 1997 to 2014", *Journal of Cleaner Production*, Vol. 140 No. 2, pp. 840-846.
- Chen, R.H.G. and Chen, C.M. (2016), "Visualizing the world's scientific publications", *Journal of the Association for Information Science and Technology*, Vol. 67 No. 10, pp. 2477-2488.
- Clarivate Analytics (2017), "Web of Science core collection", available at: <https://clarivate.com/products/web-of-science/web-science-form/web-science-core-collection/> (accessed August 12, 2017).
- Dalbello, M. (2011), "A genealogy of digital humanities", *Journal of Documentation*, Vol. 67 No. 3, pp. 480-506.
- Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K. and Harshman, R. (1990), "Indexing by latent semantic analysis", *Journal of the American Society for Information Science*, Vol. 41 No. 6, pp. 391-407.
- Drucker, J. (2011), "Humanities approaches to graphical display", *Digital Humanities Quarterly*, Vol. 5 No. 1, available at: [www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html](http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html)
- Eito-Brun, R. and Rodríguez, M.L. (2016), "50 years of space research in Europe: a bibliometric profile of the European Space Agency (ESA)", *Scientometrics*, Vol. 109 No. 1, pp. 551-576.
- Feng, F., Zhang, L., Du, Y. and Wang, W. (2015), "Visualization and quantitative study in bibliographic databases: a case in the field of university-industry cooperation", *Journal of Informetrics*, Vol. 9 No. 1, pp. 118-134.
- Ferrer, C. (2011), "Challenges and opportunities of e-humanists: the case of Latin American literary studies", in Chova, L.G., Belenguer, D.M. and Martínez, A.L. (Eds), *3rd International Conference on Education and New Learning Technologies*, Edulearn, Barcelona, pp. 4210-4217.
- Fickers, A. (2013), "Veins billed with the muted sap of rationality: a critical reply to Rens Bod", *BMGN – Low Countries Historical Review*, Vol. 128 No. 4, pp. 155-163.
- Giordano, M. (2016), "Digitalization as prevention of degradation and contribution for the knowledge of the libraries' cultural heritage", in Corniello, L. (Ed.), *World Heritage and Degradation: Smart Design, Planning and Technologies*, Scuola Pitagora Editrice, Naples, pp. 192-201.
- Gold, M.K. (2012), *Debates in the Digital Humanities*, University of Minnesota Press, Minneapolis, MN.
- Grayson, S., Wade, K., Meaney, G., Rothwell, J., Mulvany, M. and Greene, D. (2016), "Discovering structure in social networks of 19th century fiction", *Proceedings of the 2016 ACM Web Science Conference (WebSci'16)*, ACM, Hannover, pp. 325-326.
- Green, C.D. (2016), "A digital future for the history of psychology?", *History of Psychology*, Vol. 19 No. 3, pp. 209-219.
- He, Q. (1999), "Knowledge discovery through co-word analysis", *Library Trends*, Vol. 48 No. 1, pp. 133-159.
- Hicks, D. (1999), "The difficulty of achieving full coverage of international social science literature and the bibliometric consequences", *Scientometrics*, Vol. 44 No. 2, pp. 193-215.

- Hinrichs, U., Forlini, S. and Moynihan, B. (2016), "Speculative practices: utilizing InfoVis to explore untapped literary collections", *IEEE Transactions on Visualization and Computer Graphics*, Vol. 22 No. 1, pp. 429-438.
- Hockey, S. (2004), "The history of humanities computing", in Schreibman, S., Siemens, R. and Unsworth, J. (Eds), *A Companion to Digital Humanities*, Blackwell Publishing, Oxford, pp. 3-19.
- Hood, W.W. and Wilson, C.S. (2001), "The literature of bibliometrics, scientometrics, and informetrics", *Scientometrics*, Vol. 52 No. 2, pp. 291-314.
- Huang, Y., Huang, Q., Ali, S., Zhai, X., Bi, X. and Liu, R. (2016), "Rehabilitation using virtual reality technology: a bibliometric analysis, 1996-2015", *Scientometrics*, Vol. 109 No. 3, pp. 1-13.
- Jennifer, L.A. and Kevin, B.G. (2013), "Keeping up with ... digital humanities", available at: [www.ala.org/acrl/publications/keeping\\_up\\_with/digital\\_humanities](http://www.ala.org/acrl/publications/keeping_up_with/digital_humanities) (accessed May 17, 2016).
- Jockers, M.L. (2013), *Macroanalysis: Digital Methods and Literary History (Topics in the Digital Humanities)*, University of Illinois Press, Champaign, IL.
- Kahn, R. (2013), "Debates in the digital humanities", *Information Communication & Society*, Vol. 16 No. 6, pp. 1017-1018.
- Kaltenbrunner, W. (2015), "Scholarly labour and digital collaboration in literary studies", *Social Epistemology*, Vol. 29 No. 2, pp. 207-233.
- Kim, M.C., Zhu, Y. and Chen, C. (2014), "How are they different? A quantitative domain comparison of information visualization and data visualization (2000-2014)", *Scientometrics*, Vol. 107 No. 1, pp. 123-165.
- Koltay, T. (2016), "Library and information science and the digital humanities: perceived and real strengths and weaknesses", *Journal of Documentation*, Vol. 72 No. 4, pp. 781-792.
- Kretzschmar, W.A.J. and Potter, W.G. (2010), "Library collaboration with large digital humanities projects", *Literary and Linguistic Computing*, Vol. 25 No. 4, pp. 439-445.
- Landström, H., Harirchi, G. and Åström, F. (2012), "Entrepreneurship: exploring the knowledge base", *Research Policy*, Vol. 41 No. 7, pp. 1154-1181.
- Li, X., Qiao, H. and Wang, S. (2017), "Exploring evolution and emerging trends in business model study: a co-citation analysis", *Scientometrics*, Vol. 111 No. 2, pp. 869-887.
- Liu, L.M. and Wang, Q. (2016), "Visualization analysis of domestic competitive intelligence genre, core idea and hotspots", *Journal of Intelligence*, Vol. 35 No. 7, pp. 111-116.
- Liu, Q. and Ye, Y. (2012), "A study on mining bibliographic records by designed software SATI: case study on library and information science", *Journal of Information Resources Management*, Vol. 2 No. 1, pp. 50-58.
- Liu, Z., Yin, Y., Liu, W. and Dunford, M. (2015), "Visualizing the intellectual structure and evolution of innovation systems research: a bibliometric analysis", *Scientometrics*, Vol. 103 No. 1, pp. 135-158.
- McCarty, W. (2005), *Humanities Computing*, Palgrave Macmillan, London.
- McCurdy, N., Lein, J., Coles, K. and Meyer, M. (2016), "Poemage: visualizing the sonic topology of a poem", *IEEE Transactions on Visualization and Computer Graphics*, Vol. 22 No. 1, pp. 439-448.
- Mariano, S. and Walter, C. (2015), "The construct of absorptive capacity in knowledge management and intellectual capital research: content and text analyses", *Journal of Knowledge Management*, Vol. 19 No. 2, pp. 372-400.
- Martens, C.D.P., Lacerda, F.M., Belfort, A.C. and de Freitas, H.M.R. (2016), "Research on entrepreneurial orientation: current status and future agenda", *International Journal of Entrepreneurial Behavior & Research*, Vol. 22 No. 4, pp. 556-583.
- Michel, J.B., Shen, Y.K., Aiden, A.P., Veres, A., Gray, M.K., Google Books Team, Pickett, J.P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M.A. and Aiden, E.L. (2011), "Quantitative analysis of culture using millions of digitized books", *Science*, Vol. 331 No. 6014, pp. 176-182.

- Mone, G. (2016), "What's next for digital humanities? New computational tools spur advances in an evolving field", *Communications of the ACM*, Vol. 59 No. 6, pp. 20-21.
- Moretti, F. (2005), *Graphs, Maps, Trees: Abstract Models for Literary History*, Verso Books, London.
- Moretti, F. (2013), *Distant Reading*, Verso Books, London.
- Moretti, G., Sprugnoli, R., Menini, S. and Tonelli, S. (2016), "ALCIDE: extracting and visualising content from large document collections to support humanities studies", *Knowledge-Based Systems*, Vol. 2016 No. 111, pp. 100-112.
- Mozhaeva, G.V., Renha, P.N.M. and Zakharova, U.S. (2016), "Information environment of digital humanities: analysis of information interactions", *Journal of Siberian Federal University: Humanities & Social Sciences*, Vol. 9 No. 7, pp. 1572-1585.
- Najmi, A., Rashidi, T.H., Abbasi, A. and Waller, S.T. (2017), "Reviewing the transport domain: an evolutionary bibliometrics and network analysis", *Scientometrics*, Vol. 110 No. 2, pp. 843-865.
- Naukkarinen, O. and Bragge, J. (2016), "Aesthetics in the age of digital humanities", *Journal of Aesthetics & Culture*, Vol. 8 No. 1, pp. 1-18.
- Nederhof, A.J. (2006), "Bibliometric monitoring of research performance in the social sciences and the humanities: a review", *Scientometrics*, Vol. 66 No. 1, pp. 81-100.
- Piersma, H. and Ribbens, K. (2013), "Digital historical research context, concepts and the need for reflection", *BMGN – Low Countries Historical Review*, Vol. 128 No. 4, pp. 78-102.
- Pietrobruno, S. (2014), "Between narratives and lists: performing digital intangible heritage through global media", *International Journal of Heritage Studies*, Vol. 20 Nos 7/8, pp. 742-759.
- Pinto, M., Escalona, M.I., Pulgarin, A. and Uribe-Tirado, A. (2015), "The scientific production of Ibero-American authors on information literacy (1985-2013)", *Scientometrics*, Vol. 102 No. 2, pp. 1555-1576.
- Poole, A.H. (2017), "The conceptual ecology of digital humanities", *Journal of Documentation*, Vol. 73 No. 1, pp. 91-122.
- Pratt, J.A., Hauser, K. and Sugimoto, C.R. (2012), "Defining the intellectual structure of information systems and related college of business disciplines: a bibliometric analysis", *Scientometrics*, Vol. 93 No. 2, pp. 279-304.
- Price, D.J.D.S. (1965), *Little Science, Big Science and Beyond*, Columbia University Press, New York, NY.
- Pritchard, A. and Wittig, G.R. (1981), *Bibliometrics: A Bibliography and Index*, Vol. 1, Allm Books, Watford Hertfordshire.
- Pun, R. (2015), "Conceptualizing the integration of digital humanities in instructional services: possibilities to enhance digital literacy in the 21st century", *Library Hi Tech*, Vol. 33 No. 1, pp. 134-142.
- Radhakrishnan, S., Erbis, S., Isaacs, J.A. and Kamarthi, S. (2017), "Novel keyword co-occurrence network-based methods to foster systematic reviews of scientific literature", *PLoS One*, Vol. 12 No. 3, pp. e0172778.
- Ramsay, S. (2011), *Reading Machines: Toward an Algorithmic Criticism (Topics in the Digital Humanities)*, University of Illinois Press, Champaign, IL.
- Rath, L.L. (2016), "Low-barrier-to-entry data tools: creating and sharing humanities data", *Library Hi Tech*, Vol. 34 No. 2, pp. 268-285.
- Ruan, J., Chan, F.T.S., Zhu, F., Wang, X. and Yang, J. (2016), "A visualization review of cloud computing algorithms in the last decade", *Sustainability*, Vol. 8 No. 10, p. 1008, doi: 10.3390/su8101008.
- Ruttikay, Z. (2014), "Pervasive computing for sustaining cultural heritage", *IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, IEEE, Budapest, pp. 344-344.
- Schmidt, D. (2012), "The role of markup in the digital humanities", *Historical Social Research – Historische Sozialforschung*, Vol. 37 No. 3, pp. 125-146.

- Schreibman, S., Siemens, R. and Unsworth, J. (2004), *A companion to Digital Humanities*, Wiley-Blackwell, Hoboken, NJ.
- Sinkovics, N. (2016), "Enhancing the foundations for theorising through bibliometric mapping", *International Marketing Review*, Vol. 33 No. 3, pp. 327-350.
- Song, J., Zhang, H. and Dong, W. (2016), "A review of emerging trends in global PPP research: analysis and visualization", *Scientometrics*, Vol. 107 No. 3, pp. 1111-1147.
- Spiro, L. (2012), "'This is why we fight': Defining the values of the digital humanities", available at: [http://kompetenzzentrum.uni-trier.de/files/9113/9696/2499/Values\\_in\\_the\\_DH\\_Lisa\\_Spiro.pdf](http://kompetenzzentrum.uni-trier.de/files/9113/9696/2499/Values_in_the_DH_Lisa_Spiro.pdf) (accessed May 10, 2016).
- Stone, S. (1982), "Humanities scholars: information needs and uses", *Journal of Documentation*, Vol. 38 No. 4, pp. 292-313.
- Su, H.N. and Lee, P.C. (2010), "Mapping knowledge structure by keyword co-occurrence: a first look at journal papers in technology foresight", *Scientometrics*, Vol. 85 No. 1, pp. 65-79.
- Svensson, A. (2016), "Digital humanities in Swedish research: a bibliographic review", available at: [www.hf.uio.no/iln/english/research/networks/digital-humanities/news-and-events/events/2016/pdf/abstracts/posters/poster\\_svensson.pdf](http://www.hf.uio.no/iln/english/research/networks/digital-humanities/news-and-events/events/2016/pdf/abstracts/posters/poster_svensson.pdf) (accessed November 25, 2016).
- Terras, M. (2015), "Opening access to collections: the making and using of open digitised cultural content", *Online Information Review*, Vol. 39 No. 5, pp. 733-752.
- Terras, M. (2016), "A decade in digital humanities", *Journal of Siberian Federal University: Humanities & Social Sciences*, Vol. 9 No. 7, pp. 1637-1650.
- Terras, M., Nyhan, J. and Vanhoutte, E. (2013), *Defining Digital Humanities: A Reader*, Ashgate Publishing Company, Farnham.
- Tracy, D.G. (2016), "Assessing digital humanities tools: use of Scalar at a research university", *Portal: Libraries and the Academy*, Vol. 16 No. 1, pp. 163-189.
- Travis, C. (2015), "Visual geo-literary and historical analysis, Tweetflickrtubing, and James Joyce's Ulysses (1922)", *Annals of the Association of American Geographers*, Vol. 105 No. 5, pp. 927-950.
- Van Eck, N.J. and Waltman, L. (2010), "Software survey: VOSviewer, a computer program for bibliometric mapping", *Scientometrics*, Vol. 84 No. 2, pp. 523-538.
- Van Eck, N.J. and Waltman, L. (2017), "Citation-based clustering of publications using CitNetExplorer and VOSviewer", *Scientometrics*, Vol. 111 No. 2, pp. 1053-1070.
- Van Ruyskensvelde, S. (2014), "Towards a history of e-ducation? Exploring the possibilities of digital humanities for the history of education", *Paedagogica Historica*, Vol. 50 No. 6, pp. 861-870.
- Varela, M.E. (2015), "Digital archives and digital methods: a indonesian case study", in Allen, R.B., Hunter, J. and Zeng, M.L. (Eds), *Digital Libraries: Providing Quality Information*, Vol. 9469, Springer, Seoul, pp. 340-341.
- Varner, S. and Hswe, P. (2016), "Special report: digital humanities in libraries", available at: <http://americanlibrariesmagazine.org/2016/01/04/special-report-digital-humanities-libraries/> (accessed May 20, 2016).
- Wang, Q. (2015), "An overview of domestic library corporate governance research", *Public Library Journal*, Vol. 3, September, pp. 19-25.
- Wang, X., Nathwani, J. and Wu, C. (2016), "Visualization of international energy policy research", *Energies*, Vol. 9 No. 2, p. 72.
- Wong, S.H.R. (2016), "Digital humanities: what can libraries offer?", *Portal: Libraries and the Academy*, Vol. 16 No. 4, pp. 669-690.
- Xie, P. (2015), "Study of international anticancer research trends via co-word and document co-citation visualization analysis", *Scientometrics*, Vol. 105 No. 1, pp. 611-622.
- Zaagsma, G. (2013), "On digital history?", *BMGN – Low Countries Historical Review*, Vol. 128 No. 4, pp. 3-29.

- Zaggia, S., Bertolazzi, A. and Panarotto, F. (2015), "Cultural heritage 2.0. Toward innovative tools for the communication of cultural and historical asset", in Piscitelli, M. (Ed.), *Heritage and Technology: Mind Knowledge Experience*, Vol. 56, Scuola Pitagora Editrice, Aversa, pp. 1180-1187.
- Zhang, L. and Feng, Y. (2014), "Bibliometrics and visualization analysis of artificial blood vessel research", *Current Science*, Vol. 106 No. 6, pp. 816-822.
- Zhong, W.J. (2012), "Evaluation about the core authors based on price law and comprehensive index method", *Science and Technology Management Research*, Vol. 2, September, pp. 57-60.
- Zhu, J. and Hua, W. (2017), "Visualizing the knowledge domain of sustainable development research between 1987 and 2015: a bibliometric analysis", *Scientometrics*, Vol. 110 No. 2, pp. 893-914.
- Zhu, Q., Kong, X., Hong, S., Li, J. and He, Z. (2015), "Global ontology research progress: a bibliometric analysis", *Aslib Journal of Information Management*, Vol. 67 No. 1, pp. 27-54.

#### **About the author**

Qing Wang is a PhD Student in the Department of Information Resource Management at the Business School of Nankai University, China. He received the Bachelor's Degree from the School of Finance & Economics, Jiangsu University, China in 2009 and the Master's Degree in Information Science from Shanxi University, China in 2014. His main research interests focus on modern library management, digital humanities, information science, and bibliometric studies. Qing Wang can be contacted at: nku\_wq@163.com

---

For instructions on how to order reprints of this article, please visit our website:

[www.emeraldgroupublishing.com/licensing/reprints.htm](http://www.emeraldgroupublishing.com/licensing/reprints.htm)

Or contact us for further details: [permissions@emeraldinsight.com](mailto:permissions@emeraldinsight.com)



Reproduced with permission of copyright owner. Further reproduction prohibited without permission.