ISSN: 2644-2957
DOI: 10.33552/OJCAM.2021.06.000631
Online Journal of
Complementary & Alternative Medicine
Iris Publishers

Research Article

Visualizing Traditional Chinese Medicine and Information Representation and Retrieval: Opportunities and Challenges in a New Era of Big Data

David C Mainenti*

Palmer iSchool of Library, Long Island University, USA

***Corresponding author:** David C Mainenti, Palmer iSchool of Library, Long Island University, USA.

Received Date: December 20, 2020

Published Date: March 23, 2021

Abstract

Computer-based medical diagnostic systems have seen tremendous growth since the 1950s, particularly with the arrival of personal computers, the Internet, portable devices, and big data analytical environments. Such technologies utilize the fundamental principles of information representation and retrieval (IRR) to solve complex questions pertaining to health and disease. However, since inception, such systems have virtually ignored traditional Chinese medicine (TCM) techniques, oftentimes due to their lack of success in randomized controlled trials. To this day, little is known about how TCM works scientifically and, yet, it remains an essential part of the world's healthcare system, particularly in several Asian countries. As disease remains widespread across society, the diagnostic and treatment methods of TCM should be compared alongside Western medical models, in light of modern IRR techniques, to determine if a new, futuristic form of translational medicine can be developed that improves medical outcomes and reduces health care costs worldwide. This study analyzes all published research in SCOPUS relating to TCM and IRR for the period 1985-2020 and employs bibliometric techniques, multiple correspondence analysis, and data visualizations to investigate author productivity, collaborations, and research trends. Opportunities and challenges were discovered that will help identify future directions within the field as we enter a new era of data-intensive scientific discovery in medicine.

Keywords: Bibliometrics; Computer-based medical diagnostic systems; Data visualization; Herbal pharmaceutical technology; Multiple correspondence analysis; Traditional Chinese medicine; Information retrieval

Abbreviations: IRR: Information Representation and Retrieval; TCM: Traditional Chinese Medicine; IoT: Internet of Things; AI: Artificial Intelligence; ISO: International Organization for Standardization

Introduction

Since its creation thousands of years ago [1], scholars of traditional Chinese medicine (TCM) have provided a wide variety of resources for biomedical and health science, comprised of published literature, medicinal materials, herbs, diagnostic matrices, clinical records, medical formularies, and the like. In many ways, the founders of TCM were true pioneers of information science – they identified problems affecting society in their day and consciously searched available data at the time, represented mainly by classical medical literature passed down from the sages of antiquity in chant and song prior to the development of the written record, in addition to plants and other natural substances garnered from the environment [2]. These early medical leaders proceeded to organize and index such information into influential works (e.g., primitive expert knowledge systems) to prevent the spread of epidemic disease so that all those who followed would have retrieval access to a robust decisioning mechanism designed to solve a host of medical issues [3]. These ancient systems, based primarily on natural patterns of disharmony, herbal medical information representation, and treatment models, resemble mathematical thinking based on the presence or absence of human biological information [4] and utilize a binary system of numbers and probabilities that closely identifies with the concept of bits and bytes that form the working principle found in our current computing environment [5].

With the advent of big data analytics, Internet of Things (IoT) technologies, and society's continued pursuit of optimal health and well-being, the idea of developing computer-aided medical decision-making using artificial intelligence (AI) models to simulate the deductive procedure of disease diagnosis is soon to become a common reality that laid roots down over forty years ago [6]. In the future, Samsung [7] predicts sensors embedded all around us monitoring our health in a continuous manner, linked to one giant AI network, picking up signs of illness, automatically nudging users to make healthier choices, acting as a virtual doctor, directing future medical research but also potentially ranking us and shifting health and life insurance premiums to a pay-as-you-live model. Through the use of information representation and retrieval (IRR) technology, medical data gathered from a multitude of sources can now be gleaned and aggregated with appropriate criteria for use in complex TCM algorithms that bridge gaps currently existing in Western diagnostic practices by imitating classical deductive and reasoning procedures for solving medical problems and recommending treatment protocols [8]. Unfortunately, the complexity of medical knowledge creates a number of IRR difficulties [9] including, but not limited to standardized data formats; formulated inputs; and good feature representation with efficient key factors of the target problem. This is further complicated [10] by the fact that there remains, to this day, no single system able to accurately read the many medical manuscripts written in numerous dialects and stored at different locations all over the world, a literary necessity to ground TCM theory with modern science, due to the historical nature of the medicine. In particular, due to thousands of years of medical development and evolution, regional cultural differences and language variations exist which, while providing a richness and diversity to the TCM terminology system, creates standardization problems with the modernization of this classical medical modality, especially when compared to Western medical systems [11], which also experience similar vocabulary issues. On top of this, the use of natural herbal medicinal and other applications such as acupuncture are widely viewed as experimental by the Western scientific community, due, in part, to the individualistic nature in which such modalities are prescribed to human beings. Moving forward, a shift in research ideology will be required, gravitating away from controlled clinical trials to big data based, information-rich experiments with IRR at its core and mobile technologies as a means to collect large volumes of TCM data in a quantitative manner. To identify research opportunities and challenges that may exist in the design and implementation of a TCM-based diagnostic and medical recommendation system, the purpose of this work is to perform a bibliometric analysis with data visualizations related to IRR and TCM, particularly through author, paper, and co-word analysis, so as to better understand the conceptual structure of the field.

Materials and Methods

Scopus® (http://www.scopus.com), considered by some to be the largest abstract and citation database of peer-reviewed literature, including scientific journals, books and conference proceedings, was initially searched on October 9, 2019 and again on November 16, 2019 for all citations with the Boolean string [("Chinese medicine") AND ("information retrieval") OR ("information representation")] located in the article title, abstract, or keywords. The results of the search revealed 192 documents for the period 1985-2020; a BibTeX export file was saved and read into R, a free software environment for statistical computing and graphics (http://www.r-project.org), using bibliometrix [12], a tool for comprehensive science mapping analysis. The function *readFiles* was initially used to create a single large character vector; this object was then converted into a data frame using the function *convert2df*, with cases corresponding to manuscripts and variables to field tags in the original export file, comprising all bibliographic attributes of each document based on Clarivate Analytics WoS Field Tag codified industry standards [13]. During data cleansing, four entries were removed from the data frame due to a lack of author and other document information: three represented conference proceeding introductions and a fourth represented an introductory chapter on semantic grid applications for traditional Chinese medicine.

Results and Discussion

Descriptive analysis

To begin, a descriptive analysis was performed on the bibliographic data frame using the function *biblioAnalysis*; a display of the main results are included in Table 1. A total of 188 documents from 105 sources and 899 author appearances were noted, including a collaboration index of 2.49. Figure 1 illustrates the number of publications per year for the collection period 1985-2020; one large spike occurs in 2006 (28 citations) which continues into 2008, followed by a dip and then another, slower increase cumulating in 2017. This trend mirrors publications on IRR in general, which also peaks in 2006, according to Scopus, representing the maturation of computer browsing and the initial transition to mobile smart devices.

Table 2 contains the top 10 most cited papers in the collection, with Kanehisa M, et al. [14] having been cited over 1,500 times for their work in Japan on computerizing disease information using pathway maps, all Japanese drugs (including every TCM herbal formula), and gene/molecule lists. The second most-cited paper, Tang JL, et al. [15], represents one of the oldest papers in the current collection and is a summary of issues relating to randomized controlled trials in TCM, specifically: lack of blinding; low sample sizes; using another, unproven TCM treatment as the control; not long-term in nature; incompleteness; lack of quantitative data; missing intention to treat; lack of data on baseline characteristics or side effects; short reporting; and presence of publication bias. The third most-cited paper [16], was published in an American Heart Association journal and concludes, in similar fashion, the insufficiency of TCM evidence in using herbal medicinal for stroke patients, due to bias from poor methodology, even though the agents used appeared to be potentially beneficial and nontoxic in nature. The fourth document with the most citations [17] discusses newly published guidelines and technical notes by the European Union, in collaboration with Chinese scientists, to encourage good practice in the collection, assessment, and publication of TCM literature. The fifth most-cited document [18] reviews advances in automated tongue diagnosis, a key requirement for the accurate gathering of quantitative data, while the sixth most-cited document [19] discusses the development of ontology for TCM IRR. Fang YC, et al. [20] and Qiao X, et al. [21] both discuss the creation of TCM databases, while Wojcikowski K, et al. [22] again point to difficulties with randomized controlled trials in TCM, particularly relating to the use of herbal medicinals in the treatment of kidney disorders. The tenth most-cited document [23] concludes that text mining of TCM literature and clinical data carries with it the potential to clarify misunderstandings, but clear operational definitions are first required.

Table 3 lists total citations by country, along with average article citations; Japan leads this metric due to the Kanehisa M, et al. [24] document noted above, with China positioned strongly behind with 865 total citations. As expected, over 50% of the top 10 countries are located in Asia; the United States remains far behind in this research area with only 10 total citations related to one published article. Table 4 illustrates the top author countries in the collection, with China strongly in the lead with 109 articles (a frequency of 0.76224) – additionally, 89% of these articles (97) are considered single country publications. Given the nature of this data, TCM and IRR research in the East has been mainly conducted as single country publications (China, Hong Kong, Korea, and Japan) while Australia, Canada, and Germany research has been more multi-country in nature.

Table 1: Main information regarding the collection.

Description	
Documents	188
Period	1985 – 2020
Sources	105
Average citations per documents	17.9
Authors	710
Author Appearances	899
Authors of single-authored documents	13
Authors of multi-authored documents	697
Documents per Author	0.265
Authors per Document	3.78
Co-Authors per Documents	4.78
Collaboration Index	2.49

Table 2: Top 10 most cited papers.

Paper	Total Citations (TC)	TC per Year
Nucleic Acids Res [14]	1,506	136.91
Br Med J [15]	203	9.67
Stroke [16]	130	10
J Ethnopharmacol [17]	101	12.62
IEEE Trans Med Imaging [18]	88	5.87
Artif Intell Med [19]	80	5
BMC Complement Altern Med [20]	79	6.58
J Chem Inf Comput Sci [21]	76	4.22
J Lab Clin Med [22]	69	4.93
J Biomed Informatics [23]	59	5.9

Citation: David C Mainenti. Visualizing Traditional Chinese Medicine and Information Representation and Retrieval: Opportunities and Challenges in a New Era of Big Data. On J Complement & Alt Med. 6(2): 2021. OJCAM.MS.ID.000631.
DOI: 10.33552/OJCAM.2021.06.000631.

Table 3: Top 10 total citations per country.

Country	TC	Average Article Citations
Japan	1,509	754.5
China	865	7.94
Hong Kong	286	47.67
Australia	254	31.75
Taiwan	79	79
Korea	51	10.2
United Kingdom	32	16
Singapore	21	21
Germany	13	6.5
USA	10	10

Table 5 lists the top 10 most productive authors in the collection, based on both number of published articles (full counting) and number of published articles fractionalized (which assigns co-authored publications a fraction of one to each of the co-authors); studies have illustrated that, oftentimes, fractional counting offers a more useful perspective than full counting, especially as a means to avoid misunderstanding or misinterpretation [25]. Fractionalized counting does not affect the most productive author (Zhang Y) but does shift the order of the others slightly and results in the appearance of one new author (Xiong X) in the top 10.

Figure 2 applies the *authorProdOverTime* function on the collection to calculate and visualize the production of these top 10 authors over time, in terms of number of publications and total citations per year, for the period 1985-2020. This illustration clearly depicts the top producing author (Zhang Y) as covering both a wide period (2005-2019) along with more recent proliferation, oftentimes as a co-author, as noted by the number of articles fractionalized (2.09). Other authors with more recent production include Yu T (8 overall publications), Li J and Wang Y (6 overall publications each), and Liu L (5 overall publications).

Table 6 contains the top 10 most frequent journals, based on number of published articles in the collection – led by the Chinese Journal of Clinical Rehabilitation with 25 articles and followed by Evidence-Based Complementary and Alternative Medicine with 11 publications. However, it is important to also look at this data from the perspective of number of documents published annually; this information for each of the top five sources is visualized in Fig. 3 using the function *sourceGrowth*, which illustrates that the Chinese Journal of Clinical Rehabilitation was only in existence from 2002-2006. Since then, four newer journals have increased their publication rate, particularly Evidence-Based Complementary and Alternative Medicine, which is second in number of articles but clearly the leading publication in this field, particularly as the journal currently holds an h-index of 72 and sits as the sixth ranked journal in complementary and alternative medicine [26].

Table 4: Top 10 corresponding author's countries.

Country	Articles	Frequency	SCP	MCP	MCP Ratio
China	109	0.76224	97	12	0.11
Australia	8	0.05594	4	4	0.5
Hong Kong	6	0.04196	5	1	0.167
Korea	5	0.03497	5	0	0
Canada	2	0.01399	0	2	1
Germany	2	0.01399	1	1	0.5
Japan	2	0.01399	2	0	0
United Kingdom	2	0.01399	2	0	0
Brazil	1	0.00699	1	0	0
Hungary	1	0.00699	1	0	0

Table 5: Top 10 Most productive authors.

Author	No. of Articles	Author	No. of Articles Fractionalized
Zhang Y	10	*Zhang Y*	2.09
Wu Z	9	*Zhou X*	2.05

Yu T	8	*Wu Z*	1.75
Zhou X	8	*Wang Y*	1.45
Chen H	7	*Yu T*	1.35
Chen X	7	*Li J*	1.31
Li J	6	*Chen H*	1.25
Wang Y	6	*Xiong X*	1.25
Cui M	5	*Cui M*	1.19
Liu L	5	*Chen X*	1.11

Table 6: Top 10 Most frequent journals.

Sources	No. of Articles
Chinese Journal of Clinical Rehabilitation	25
Evidence-Based Complementary and Alternative Medicine	11
Journal of Ethnopharmacology	7
Zhongguo Zhongyao Zazhi	7
Journal of Alternative and Complementary Medicine	6
Chinese Journal of Evidence-Based Medicine	5
Chinese Journal of Integrative Medicine	5
Zhongguo Zhongxiyi Jiehe Zazhi	5
Journal of Chinese Integrative Medicine	4
Journal of Traditional Chinese Medicine	4

Table 7: Top 10 Most frequent keywords.

Author Keywords	No. of Articles	Keywords-Plus	No. of Articles
Traditional Chinese Medicine	27	Information Retrieval	173
Systematic Review	13	Chinese Medicine	151
Chinese Medicine	7	Human	120
Information Retrieval	7	Article	86
Meta Analysis	6	Medicine	77
Information Extraction	5	Humans	67
Ontology	5	Review	66
Review	5	Herbaceous Agent	54
TCM	5	Chinese Traditional	45
Chinese Herbal Medicine	4	Priority Journal	42

Table 8: Historiograph legend.

Year	Reference	Local Citations	Global Citations
1999	TANG JL, 1999, BR MED J	3	203
2001	CHANG IM, 2001, ANN NEW YORK ACAD SCI	1	20
2002	BENSOUSSAN A, 2002, J TOXICOL CLIN TOXICOL	2	36
2002	QIAO X, 2002, J CHEM INF COMPUT SCI	2	76
2004	KA WF, 2004, J ALTERN COMPLEMENT MED	1	7
2004	ZHOU X, 2004, ARTIF INTELL MED	5	80
2005	WANG JF, 2005, CLIN PHARMACOL THER	-	21
2006	LI Y, 2006, ZHONGGUO ZHONG XI YI JIE HE ZA ZHI	1	5
2007	FLOWER A, 2007, J ALTERN COMPLEMENT MED	-	31
2007	CHEN H, 2007, BMC BIOINFORM	2	25
2008	FANG YC, 2008, BMC COMPLEMENT ALTERN MED	5	79

2008	TSE HYG, 2008, J BIOMOL SCREEN	-	2
2008	CHEN H, 2008, BMC BIOINFORM	-	5
2009	MAY BH, 2009, BIOGERONTOLOGY	1	18
2010	ZHANG X, 2010, PROC - INT CONF BIOMED ENG INF, BMEI	1	11
2010	ZHOU X, 2010, J BIOMED INFORMATICS	3	59
2010	BOEHM K, 2010, HEALTH INF LIBR J	-	13
2011	SAMPSON M, 2011, EVID -BASED COMPLEMENT ALTERN MED	-	7
2012	JIANG Z, 2012, IEEE INT CONF E-HEALTH NETWORKING	1	11
2012	MAY BH, 2012, J ALTERN COMPLEMENT MED	1	16
2012	JIANG M, 2012, EVID -BASED COMPLEMENT ALTERN MED	-	43
2014	CHEN X, 2014, COMP MATH METHODS MED	-	3
2014	CHEN H, 2014, BIOMED RES INT	-	3
2015	XIONG X, 2015, NAT REV CARDIOL	-	21
2015	YOU M, 2015, SCI WORLD J	-	1
2015	XIONG X, 2015, BMJ OPEN	1	7
2016	MAY BH, 2016, J ALTERN COMPLEMENT MED	-	9
2016	WAN H, 2016, J AM MED INFORMATICS ASSOC	-	4
2016	YU T, 2016, PROC - INT CONF BIOMED ENG INFORMATICS, BMEI	-	1
2017	LIU YQ, 2017, CHIN J INTEGR MED	1	1
2017	WANG L, 2017, EVID -BASED COMPLEMENT ALTERN MED	-	1
2019	YOON SH, 2019, CHIN J INTEGR MED	-	3

Table 7 contains the top 10 most frequent keywords using two keyword variations: authors' keywords, as specifically selected by each author, and keywords-plus, which are those keywords extracted from the publication by Scopus' database algorithms. The results vary, with keywords-plus identifying many more in common across the collection – for example, information retrieval was only selected by seven authors as a keyword but appears 173 times as a keywords-plus. Overall, while keywords-plus is as effective as authors' keywords in terms of bibliometric analysis investigating the knowledge structure of a particular field, it is often less comprehensive in representing an article's specific content [27]. However, within this collection, the use of keywords-plus may lead to a greater understanding than simply using those keywords identified by the authors, due to the increased volume and commonality of terms; this is particularly evident in Figure 4a and Figure 4b, which clearly illustrate greater and more prolific growth of keyword-plus over time, as compared to authors' keywords.

Network visualizations

To summarize the activity of top authors, journals and keywords presented below, Figure 5 employs the *threeFieldsPlot* function to generate a Sankey diagram that visualizes multiple attributes at the same time; top authors on the left, top author keywords in the center, and major cited references on the right. The width of the bands is linearly proportional to frequency, and the size of the boxes correspond to overall production.

Figure 6 visualizes scientific collaboration networks in and across countries using the *biblioNetwork* function to develop each matrix and the *networkPlot* function to illustrate it. A sphere layout is used with the size of the sphere correlating to overall production, lines density relating to collaboration strength, and color relating to the nature of the collaboration. As expected, the majority of research is clustered in and around China, particularly with Western countries, while six additional countries are illustrated as working independently (Taiwan, Brazil, Japan, Spain, Hungary, and Korea). The main collaboration networks, however faint, are represented by Germany-United Kingdom-China, USA-Canada-China, and Singapore-China, based on the density of the lines.

Figure 7 depicts a word co-occurrence network that maps and clusters terms extracted from author abstracts. The *termExtraction* function was first used with word stemming to gather this information from the textual abstract field of each manuscript. In this visualization, TCM stands alone at the bottom left in its relation to the larger clusters of information science (red), herbology (green), and biomedicine (blue). Similarly, Figure 8 illustrates author keyword co-occurrences, which follows a similar pattern as to the extracted words from the author abstracts but, due to each individual author's knowledge of the work during the keyword selection process, this visualization displays in a more orderly fashion, showing a logical progression from TCM, through a main section of information retrieval, to herbology, and ultimately across to Western biomedicine.

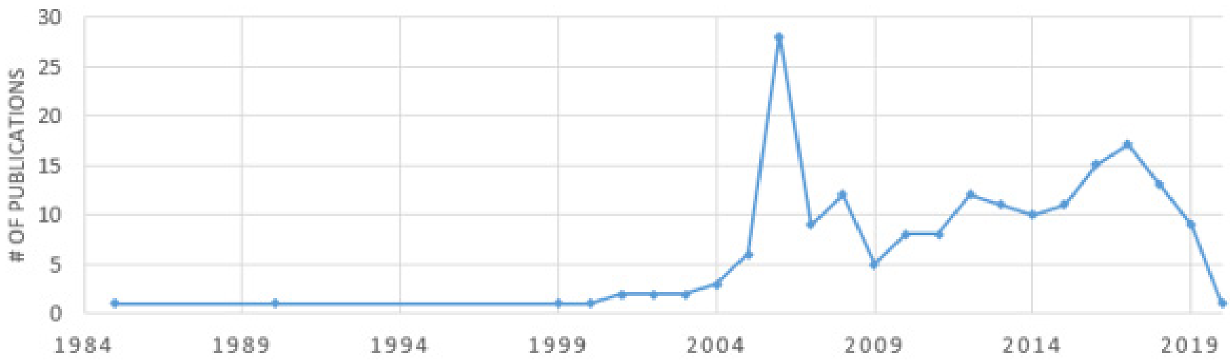

Figure 1: Publications per year 1985-2020 (Scopus).

Figure 2: Top 10 author productivity for the period 1985-2020.

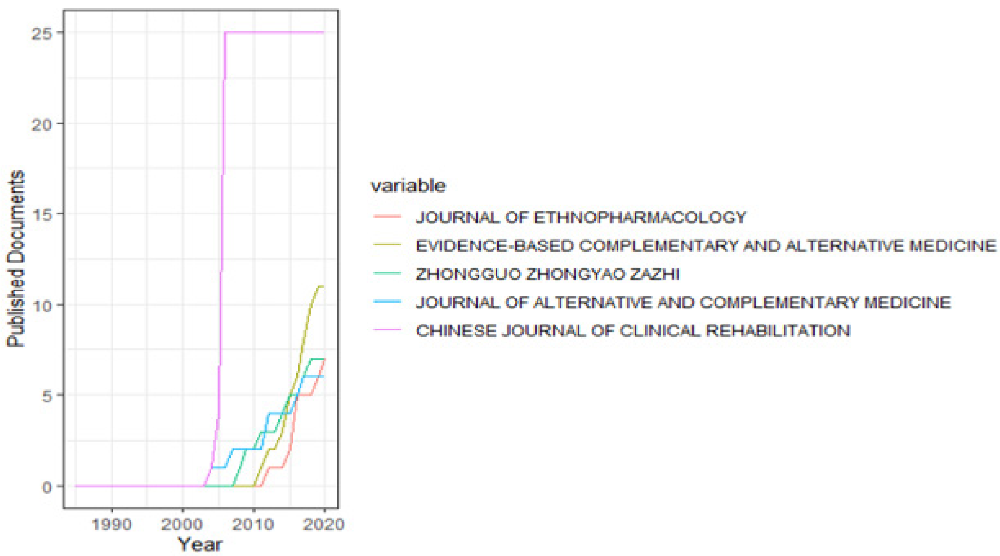

Figure 3: Source Growth, Top Five Journals, 1985-2020.

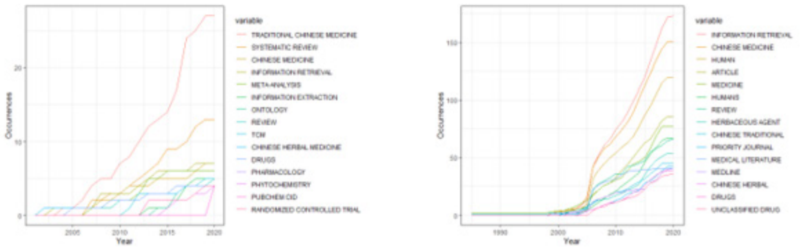

Figure 4a: Author keyword growth per year, 1985-2020. **Figure 4b:** Keyword-plus growth per year, 1985-2020.

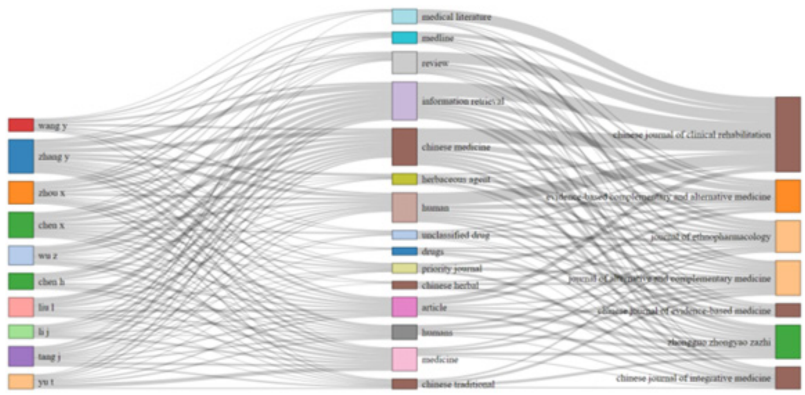

Figure 5: Sankey Diagram of Main Authors, Keywords, and Journals.

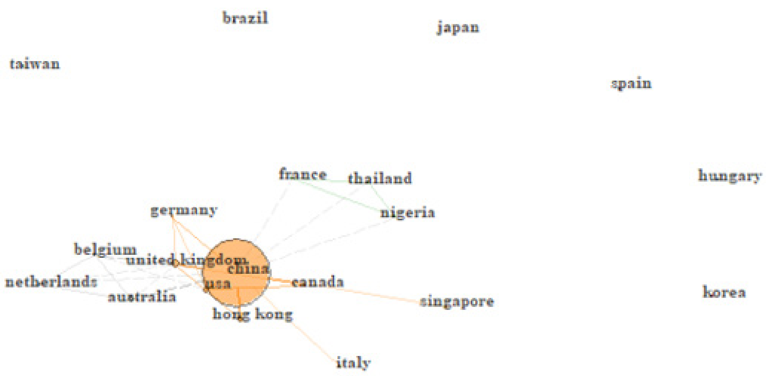

Figure 6: Country collaboration network.

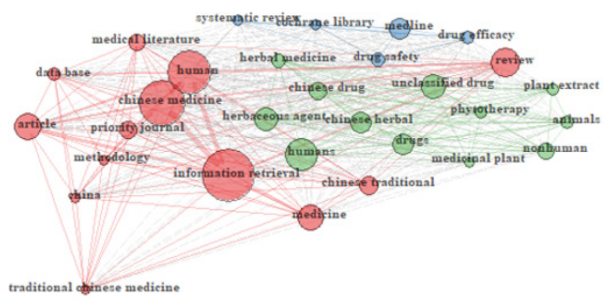

Figure 7: Author abstract co-occurrence network.

Visualizing influential papers

In each and every scientific field, a number of publications play influential roles in its evolution; these articles and their impacts can be accelerating factors in research development [28]. It is therefore important to identify and visualize the most influential articles on TCM and IRR published between 1985 and 2020, in order to better understand the nature and chronology of the field through its key authors, papers, and subjects. Using the *histNetwork* function, we created a historical citation network from the collection, using a minimum number of one global citation for the documents included in the analysis (see Table 8 for legend). We then employed the *histPlot* function to plot the historical co-citation network in the style of Garfield [29], using both local and total citation distributions; nodes displayed in Fig. 9 identify the thirty two specific articles identified in the collection and sort the main bibliography in ascending order by year. Each circle represents a paper, and arrows, pointing from one node to the next (usually to an older paper), indicates the citation relationship between these key works.

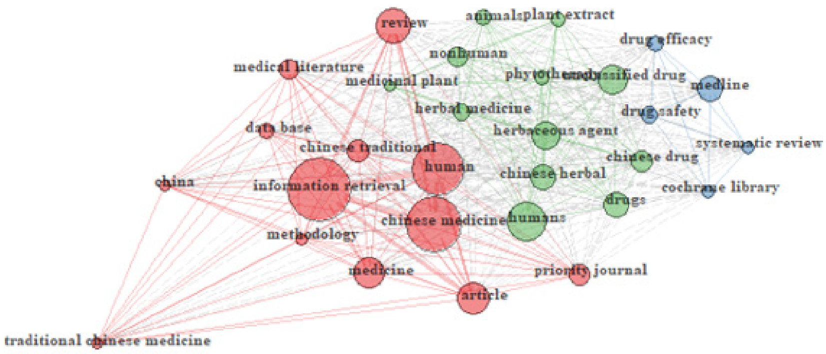

Figure 8: Keywords plus co-occurrence network.

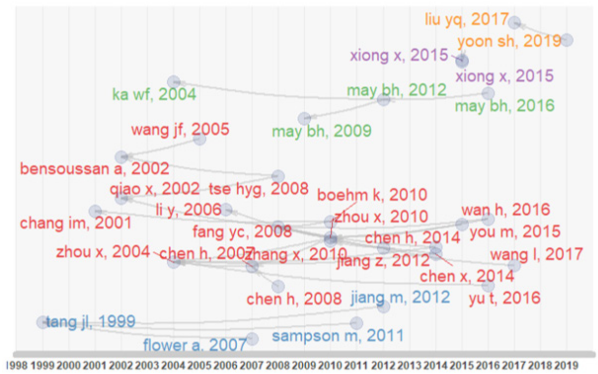

Figure 9: Historiograph of TCM and IRR (minimum one citation).

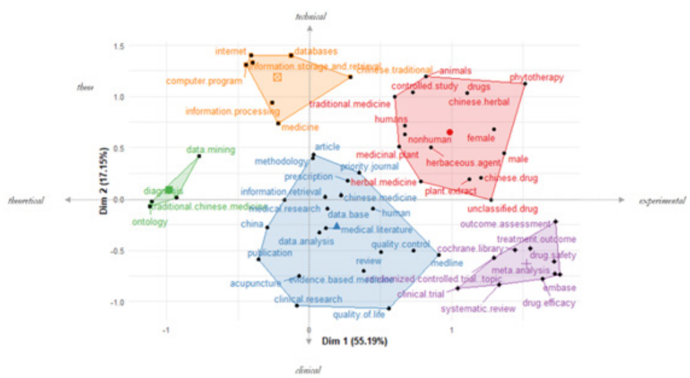

Figure 10: Conceptual map and keyword plus clusters with axis definitions.

In the historiograph for TCM and IRR, the paper by Tang JL, et al. [30] serves as the first work noted and surveyed the efficacy of randomized controlled trials in TCM literature; this spurned a line of research (highlighted in blue) extending to Flower A, et al. [31] who advocated for the Delphi method, Sampson M, et al. [32] who searched for additional databases to identify more successful controlled trials, and Jiang M, et al. [33] who evaluated evidence-based literature for TCM diagnosis and knowledge discovery.

The second and most dominant area of research in Figure 9 (highlighted in red) focuses on TCM information databases and begins with Chang IM [34] who investigated anti-aging and health-promoting elements derived from traditional herbal remedies found in the Traditional Oriental Medicine Database, leading to future research by Boehm K, et al. [35], who provided an overview of 45 published database resources for complementary and alternative medicine. Bensoussan A, et al. [36] established research goals for the search and retrieval of scientific evidence regarding the toxicity of Chinese herbal medicine, which contributed to Wang JF, et al. [37] and the construction of a TCM information database. Qiao X, et al. [38] created a structured database of components extracted from TCM herbs, while Zhou X, et al. [39] wrote an influential paper (5 local citations) that used ontology to construct a unified TCM language system for information retrieval and integration. This led to Zhou X, et al. [40], which investigated research issues regarding TCM text mining, You M, et al. [41], who developed an intelligent system for customized clinical TCM case management and analysis, Wan H, et al. [42], who constructed a heterogeneous factor graph model for extracting relations from TCM literature, and Yu T, et al. [43], who utilized semantic web technologies to build cross-cultural communication between TCM and Western medicine.

Chen H, et al. [44] used semantic and knowledge-based techniques to build e-toolkits that facilitate TCM information sharing; this contributed to Chen H, et al. [45], which introduced state-of-the-art semantic web technologies for biomedicine as a whole, including applications for TCM and translational research. Tse HYG, et al. [46] developed an online TCM herbal medical database built again from the earlier herbal works of Bensoussan A, et al. [47] and Qiao X, et al. [48]. Li Y, et al. [49] focused on utilizing data mining techniques that compared clinical characteristics of TCM and Western medicine in the diagnosis of rheumatoid arthritis, which led to Fang YC, et al. [50], another influential paper, who developed a database to provide information about TCM, genes, diseases, effects, and ingredients from a wide variety of biomedical literature. This was further studied by Zhang X, et al. [51], who created a hierarchical symptom-herb topic model for TCM research in the treatment of diabetes, and Jiang Z, et al. [52], who used link topic models to analyze TCM symptom-herb regularities. Chen X, et al. [53], also used this research to develop a semantic search engine for IRR in modern biology and TCM, along with Chen H, et al. [54], who presented a general web ontology language reasoning framework to study biological entities across TCM and Western medicine. These works ultimately influenced Wang L, et al. [55], who used topic model and multi-label classifiers to predict the function of TCM herbal prescriptions.

A third, smaller research group in Fig. 9 (highlighted in green) begins with Ka WF [56], who introduced journals and other TCM research materials available online and May BH, et al. [57], who searched English and Chinese databases to review the effectiveness and safety of Chinese herbal medicines for use in the treatment of cognitive and memory impairment. These two works led to additional research by May BH, et al. [58], comparing and evaluating published TCM collections for research and drug discovery searches, and May BH, et al. [59], who searched a database of over 1,000 classical and pre-modern TCM texts for the treatment of memory impairment. A fourth research group was also mapped on the historiograph (highlighted in purple) relating to difficulties in drawing clinical conclusions in the treatment of specific Western medicine disorders with TCM: Xiong X [60] reviewed an article on randomized controlled trials for the treatment of cardiovascular disease with TCM, while Xiong X, et al. [61] researched the clinical effects of a TCM herbal decoction in the treatment of hypertension. A fifth and final research group (highlighted in yellow), albeit small, begins late and relates to the standardization of TCM: Liu YQ, et al. [62] focused on standards and proposals established by the International Organization for Standardization (ISO), which Yoon SH, et al. [63] built from this to investigate the pros and cons of proposing standard terminology for acupotomy, a treatment modality which involves the use of both an acupuncture needle and a surgical scalpel. These five pathways of research in TCM and IRR help illustrate both the research difficulties in the field as well as opportunities in the treatment of specific diseases and the construction of modern databases and ontologies for the future use of this medical modality.

Visualizing the conceptual structure of the field

Multiple correspondence analysis is an exploratory multivariate technique that allows for the graphical and numerical analysis of the patters in relationships of categorical dependent variables, such as keywords [64]. We used the *conceptualStructure* function to draw a conceptual map of the field using keywords-plus with a maximum of five clusters and no stemming. Results are interpreted based on the relative position of the points and their distribution across the dimensions; as words are more similar in distribution, the closer they are represented in Figure 10. According to Cuccurullo C, et al. [65], map data can be translated as follows: point size is proportional to the keyword's absolute contribution; proximity between keywords corresponds to shared substance, or lack thereof; and the dimensions of the map reflect characteristic poles of topical orientation within TCM and IRR, with the middle of the map representing the average position of all the articles and thus the center of the research field.

Cluster one (highlighted in red) resides in the top right quadrant and contains keywords from articles that relate to the use of TCM herbs and natural substances as viable treatment options for both acute and chronic diseases. Cluster two (highlighted in blue) is the largest, residing in the middle and bottom center of the map; it contains keywords from articles that relate to the design of databases and retrieval systems for scientific TCM medical literature (including Zhou X, et al. [66]). The third cluster (highlighted in green), resides center left and contains keywords from articles specific to TCM diagnosis, ontology, and data mining. Cluster four (highlighted in purple) resides in the lower right quadrant and contains keywords from articles relating to the success, or lack thereof, of evidence-based research and TCM, particularly as it pertains to clinical trials. Cluster five (highlighted in orange) resides in the top left-center quadrant and contains keywords from articles relating to the representation of TCM information in IRR systems.

From this conceptual structure analysis, one can see that the first dimension of published research extends horizontally from theoretical on the left to experimental on the right. The second dimension, extending vertically, defines published works and their keywords over a spectrum ranging from technical on the top to clinical on the bottom. There are clear research gaps in the upper and lower left quadrants, indicating that more scientific research is required to ground both clinical and technical IRR work in TCM theory – an area we know is sorely lacking at the time, due to the fact that TCM is an individualized medicine, treating each patient according to their specific pattern of disharmony, gleaned from information gathered mainly, at this point, through quantitatively-based examinations performed by human beings. As a result, a great portion of research exists on the right side of the map that is considered, according to Western biomedicine, to be experimental in nature – namely Chinese herbal pharmaceutical medicine, its integration with biomedicine, and the lack of success to date with measuring TCM treatment outcomes and drug efficacy in controlled trials. This map, therefore, charts the course for where future research should be headed as it relates to TCM and IRR – to employ new research paradigms and instruments, namely big data analytics and IoT technologies, to merge Eastern and Western medicine together, in a translational way, so as to better understand the nature of disease progression and its effects on the body.

Conclusion

It may be time for the broad discipline of information studies to come out of its proverbial shell; in the future of both Samsung [67] and Sinclair DA, et al. [68], as our lives become monitored, accentuated and refined as a means to achieve a new state of existence, behind and underneath it all will be a vast, universally-encompassing network of medical IRR systems. Such structures have already begun to take shape in other areas of existence: industry is refining algorithms that learn our preferences, activities, motives, and bioactivities; cars have begun to program themselves to reach destinations; drivers are learning where passengers are headed, before they actually meet them; and music is constantly indexed and streamed around the world across language and culture. These systems all require underlying IRR architecture so that the consumer, the device, the network, and the provider can communicate digitally, behind the scenes, in a grand dance of capitalistic-driven mobile technology. However, by applying these conceptual structures to problems like ancient TCM practitioners faced but with new goals of health and well-being over profit and market share, one could only image if, in the future, through a combination of TCM and IRR, we might be able to one day think out at least half of our medical problems. Such a vision is worth pursuing, particularly in times where many diseases, especially chronic ones like obesity and mental illness, appear to be on the upswing. Further TCM research is therefore warranted, particularly with respect to IRR as it relates to grounding modern research, both clinical and technical, with classical theory. In the style of Jim Gray's vision of scientific discovery, driven by the collection, analysis, and comprehension of digital data by an ever-increasing interdisciplinary community of both professional and citizen-like scientists alike [69], perhaps the future of TCM lies in a big data design and deployment methodology with a new medical IRR at its core that is translational in nature, reducing health care costs and improving medical outcomes worldwide.

Acknowledgements

This paper and the research components related to information representation and retrieval would not have been possible without the support of the author's doctoral advisor, Dr. Heting Chu, professor in the Palmer iSchool of Library and Information Science at Long Island University.

Conflict of Interest

The author has no conflicts of interest to declare.

References

1. Chen H, Wu Z, Huang C, Xu J (2003) TCM-Grid: weaving a medical grid for traditional Chinese medicine. International Conference on Computational Science 2659: 1143-1152.
2. Brashier KE (2014) Public Memory in Early China. Harvard University Asia Center, USA, pp. 526.
3. Pregadio F (2008) Zhang Ji (Zhang Zhongjing) In: Selin H (edr), Encyclopaedia of the History of Science, Technology, and Medicine in Non-Western Cultures. Springer, Netherlands, pp. 2416.
4. Van Wyk BE, Wink M (2004) Medicinal plants of the world. CABI, USA, pp. 520.
5. Yan JF (1991) DNA and the I Ching: The Tao of Life. North Atlantic Books, USA, pp. 193.
6. Weiss SM, Kulikowski CA, Amarel S, Safir A (1978) A model-based method for computer-aided medical decision-making. Artif Intell 11(1-2): 145-172.
7. Samsung (2019) KX50: The Future in Focus.
8. Liang Z, Liu J, Ou A, Zhang H, Li Z, et al. (2019) Deep generative learning for automated EHR diagnosis of traditional Chinese medicine. Computer Methods and Programs in Biomedicine 174: 17-23.

9. Yang H, Lyu MR, King I (2013) Efficient online learning for multitask feature selection. ACM Trans Knowl Discov Data 7(2): 6.
10. Leydier Y, Ouji A, LeBourgeois F, Emptoz H (2009) Towards an omnilingual word retrieval system for ancient manuscripts. Pattern Recognit 42(9): 2089-2105.
11. Long H, Zhu Y, Jia L, Gao B, Liu J, et al. (2019) An ontological framework for the formalization, organization, and usage of TCM-Knowledge. BMC Med Inform Decis Mak 19(2): 53.
12. Aria M, Cuccurullo C (2017) Bibliometrix: An R-tool for comprehensive science mapping analysis. J Informetr 11(4): 959-975.
13. Aria M, Cuccurullo C (2019) Package bibliometrix: an R-tool for comprehensive science mapping analysis.
14. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M (2009) KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res 38: D355-D360.
15. Tang JL, Zhan SY, Ernst E (1999) Review of randomised controlled trials of traditional Chinese medicine. BMJ 319(7203): 160-161.
16. Wu B, Liu M, Liu H, Li W, Tan S, et al. (2007) Meta-analysis of traditional Chinese patent medicine for ischemic stroke. Stroke 38(6): 1973-1979.
17. Chan K, Shaw D, Simmonds MS, Leon CJ, Xu Q, et al. (2012) Good practice in reviewing and publishing studies on herbal medicine, with special emphasis on traditional Chinese medicine and Chinese materia medica. J Ethnopharmacol 140(3): 469-475.
18. Pang B, Zhang D, Wang K (2005) The bi-elliptical deformable contour and its application to automated tongue segmentation in Chinese medicine. IEEE Trans Med Imaging 24(8): 946-956.
19. Zhou X, Wu Z, Yin A, Wu L, Fan W, et al. (2004) Ontology development for unified traditional Chinese medical language system. Artif Intell Med 32(1): 15-27.
20. Fang YC, Huang HC, Chen HH, Juan HF (2008) TCMGeneDIT: a database for associated traditional Chinese medicine, gene and disease information using text mining. BMC Complement Altern Med 8(1): 58.
21. Qiao X, Hou T, Zhang W, Guo S, Xu X (2002) A 3D structure database of components from Chinese traditional medicinal herbs. J Chem Inf Comput Sci 42(3): 481-489.
22. Wojcikowski K, Johnson DW, Gobe G (2006) Herbs or natural substances as complementary therapies for chronic kidney disease: ideas for future studies. J Lab Clin Med 47(4): 160-166.
23. Zhou X, Peng Y, Liu B (2010) Text mining for traditional Chinese medical knowledge discovery: a survey. J Biomed Inform 43(4): 650-660.
24. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M (2009) KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res 38: D355-D360.
25. Perianes Rodriguez A, Waltman L, Van Eck NJ (2016) Constructing bibliometric networks: A comparison between full and fractional counting. J Informetr 10(4): 1178-1195.
26. SCImago (2019) SJR-SCImago Journal & Country Rank [Portal].
27. Zhang J, Yu Q, Zheng F, Long C, Lu Z, et al. (2016) Comparing keywords plus of WOS and author keywords: A case study of patient adherence research. J Assoc Inf Sci Technol 67(4): 967-972.
28. Berry LL, Parasuraman A (1993) Building a new academic field-The case of services marketing. J Retail 69(1): 13-60.
29. Garfield E (2004) Historiographic mapping of knowledge domains literature. J Inf Sci 30(2): 119-145.
30. Tang JL, Zhan SY, Ernst E (1999) Review of randomised controlled trials of traditional Chinese medicine. BMJ 319(7203): 160-161.
31. Flower A, Lewith GT, Little P (2007) Seeking an oracle: using the Delphi process to develop practice guidelines for the treatment of endometriosis with Chinese herbal medicine. J Altern Complement Med 13(9): 969-976.
32. Sampson M, Cogo E, Ajiferuke I, Manheimer E, Campbell K, et al. (2011) Searching for controlled trials of complementary and alternative medicine: a comparison of 15 databases. Evid Based Complement Alternat Med 2011: 1-9.
33. Jiang M, Zhang C, Zheng G, Guo H, Li L, et al. (2012) Traditional Chinese medicine Zheng in the era of evidence-based medicine: a literature analysis. Evid Based Complement Alternat Med 2012: 1-9.
34. Chang IM (2001) Anti-aging and health-promoting constituents derived from traditional oriental herbal remedies: information retrieval using the TradiMed 2000 DB. Ann N Y Acad Sci 928(1): 281-286.
35. Boehm K, Raak C, Vollmar HC, Ostermann T (2010) An overview of 45 published database resources for complementary and alternative medicine. Health Info Libr J 27(2): 93-105.
36. Bensoussan A, Myers SP, Drew AK, Whyte IM, Dawson AH (2002) Development of a Chinese herbal medicine toxicology database. J Toxicol Clin Toxicol 40(2): 159-167.
37. Wang JF, Zhou H, Han LY, Chen X, Chen YZ, et al. (2005) Traditional Chinese medicine information database. Clin Pharmacol Ther 78(1): 92-93.
38. Qiao X, Hou T, Zhang W, Guo S, Xu X (2002) A 3D structure database of components from Chinese traditional medicinal herbs. J Chem Inf Comput Sci 42(3): 481-489.
39. Zhou X, Wu Z, Yin A, Wu L, Fan W, et al. (2004) Ontology development for unified traditional Chinese medical language system. Artif Intell Med 32(1): 15-27.
40. Zhou X, Peng Y, Liu B (2010) Text mining for traditional Chinese medical knowledge discovery: a survey. J Biomed Inform 43(4): 650-660.
41. You M, Chen C, Li GZ, Yan SX, Sun S, et al. (2015) ISMAC: an intelligent system for customized clinical case management and analysis. Scientific World Journal 2015: 1-12.
42. Wan H, Moens MF, Luyten W, Zhou X, Mei Q, et al. (2016) Extracting relations from traditional Chinese medicine literature via heterogeneous entity networks. J Am Med Inform Assoc 23(2): 356-365.
43. Yu T, Liu J, Yang S, Li J, Jia L (2016) Semantic Web for Knowledge Integration between Traditional Chinese Medicine and Biomedicine. 2015 7th International Conference on Information Technology in Medicine and Education (ITME), China.
44. Chen H, Mao Y, Zheng X, Cui M, Feng Y, et al. (2007) Towards semantic e-science for traditional Chinese medicine. BMC Bioinformatics 8(3): S6.
45. Chen H, Wang Y, Wu Z (2008) Introduction to semantic e-Science in biomedicine. BMC Bioinformatics 8(3): 1-3.
46. Tse HYG, Li VWT, Hui MN, Kwok Chan P, Han Cheng S (2008) Data mining for chinese materia medica and pharmacological research. J Biomol Screen 13(5): 390-395.
47. Bensoussan A, Myers SP, Drew AK, Whyte IM, Dawson AH (2002) Development of a Chinese herbal medicine toxicology database. J Toxicol Clin Toxicol 40(2): 159-167.
48. Qiao X, Hou T, Zhang W, Guo S, Xu X (2002) A 3D structure database of components from Chinese traditional medicinal herbs. J Chem Inf Comput Sci 42(3): 481-489.
49. Li Y, Li S, Lu AP (2006) Comparative analysis via data mining on the clinical features of Western medicine and Chinese medicine in diagnosing rheumatoid arthritis. Zhongguo Zhong Xi Yi Jie He Za Zhi 26(11): 988-991.
50. Fang YC, Huang HC, Chen HH, Juan HF (2008) TCMGeneDIT: a database for associated traditional Chinese medicine, gene and disease information using text mining. BMC Complement Altern Med 8(1): 58.
51. Zhang X, Zhou X, Huang H, Chen S, Liu B (2010) A hierarchical symptom-herb topic model for analyzing traditional Chinese medicine clinical diabetic data. 2010 IEEE 3rd International Conference on Biomedical Engineering and Informatics, USA.

52. Jiang Z, Zhou X, Zhang X, Chen S (2012) Using link topic model to analyze traditional Chinese medicine clinical symptom-herb regularities. 2012 IEEE 14th International Conference on e-Health Networking, Applications and Services (Healthcom), China.

53. Chen X, Chen H, Bi X, Gu P, Chen J, et al. (2014) BioTCM-SE: a semantic search engine for the information retrieval of modern biology and traditional Chinese medicine. Comput Math Methods Med 2014: 1-13.

54. Chen H, Chen X, Gu P, Wu Z, Yu T (2014) OWL reasoning framework over big biological knowledge network. Biomed Res Int 2014: 272915.

55. Wang L, Zhang Y, Zhang Y, Xu X, Cao S (2017) Prescription Function Prediction Using Topic Model and Multilabel Classifiers. Evid Based Complement Alternat Med 2017: 8279109.

56. Ka WF (2004) Online research databases and journals of Chinese medicine. J Altern Complement Med 10(6): 1123-1128.

57. May BH, Yang AW, Zhang AL, Owens MD, Bennett L, et al. (2009) Chinese herbal medicine for Mild Cognitive Impairment and Age Associated Memory Impairment: a review of randomised controlled trials. Biogerontology 10(2): 109-123.

58. May BH, Lu C, Xue CC (2012) Collections of traditional Chinese medical literature as resources for systematic searches. J Altern Complement Med 18(12): 1101-1107.

59. May BH, Feng M, Zhou IW, Chang SY, Lu SC, et al. (2016) Memory impairment, dementia, and Alzheimer's disease in classical and contemporary traditional Chinese medicine. J Altern Complement Med 22(9): 695-705.

60. Xiong X (2015) Integrating traditional Chinese medicine into Western cardiovascular medicine: an evidence-based approach. Nat Rev Cardiol 12(6): 374.

61. Xiong X, Wang P, Li X, Zhang Y (2015) The effect of Chinese herbal medicine Jian Ling Decoction for the treatment of essential hypertension: a systematic review. BMJ Open 5(2): e006502.

62. Liu YQ, Wang YX, Shi NN, Han XJ, Lu AP (2017) Current situation of International Organization for Standardization/Technical Committee 249 international standards of traditional Chinese medicine. Chin J Integr Med 23(5): 376-380.

63. Yoon SH, Kim YS, Jo HG, Kwon CY (2019) Current usage of terminologies related to acupotomy: a literature research and standardization suggestion. Chin J Integr Med 25(2): 147-150.

64. Abdi H, Valentin D (2007) Multiple correspondence analysis. In: Salkind NJ (edr), Encyclopedia of Measurement and Statistics. Sage Publishing, USA, pp. 651-657.

65. Cuccurullo C, Aria M, Sarto F (2016) Foundations and trends in performance management. A twenty-five years bibliometric analysis in business and public administration domains. Scientometrics 108(2): 595-611.

66. Zhou X, Wu Z, Yin A, Wu L, Fan W, et al. (2004) Ontology development for unified traditional Chinese medical language system. Artif Intell Med 32(1): 15-27.

67. Samsung (2019) KX50: The Future in Focus.

68. Sinclair DA, LaPlante MD (2019) Lifespan: Why we Age-and Why We Don't Have To. Atria Books, USA, pp. 310.

69. Bourne PE (2013) The reaming of life: based on the 2010 Jim Gray eScience Award Lecture. Concurr Comput 25(4): 445-453.

Citation: David C Mainenti. Visualizing Traditional Chinese Medicine and Information Representation and Retrieval: Opportunities and Challenges in a New Era of Big Data. On J Complement & Alt Med. 6(2): 2021. OJCAM.MS.ID.000631.
DOI: 10.33552/OJCAM.2021.06.000631.