Water Related Information Science Area: A Topic Clustering Model for information retrieval

Youlin Zhao[1, a, *], Yahui Liu[2, b], Xiaotao Li[3, c], Xiangming Mu[4, d], Huiping Chen[5, e]

[1]*Business School of Hohai University, #8 Focheng West Road, Jiangning District, Nanjing city, Jiangsu Province, PR China*

[2]*Business School of Hohai University, #8 Focheng West Road, Jiangning District, Nanjing city, Jiangsu Province, PR China*

[3]*Library of Nanjing University of Aeronautics and Astronautics, Jiangjun Road, Jiangning District, Nanjing city, Jiangsu Province, PR China*

[4]*School of Information Studies of University of Wisconsin-Milwaukee, MI, Wisconsin, USA*

[5]*School of Economics and Management of Hubei University of Technology,Wuchang District, Wuhan City, Hubei Province, PR China*

[a]*sobzyl@hhu.edu.cn,* [b]*lyh201516@126.com,* [c]*oshoxiaot@163.com,* [d]*mux@uwm.edu,* [e]*chenghuiping @hbut.edu.cn*

**corresponding author:Youlin Zhao*

Keywords: water and water resource, topic clustering, topic terminology, water related information science, interdisciplinary study, information science & library science

Abstract: Understanding the direction and magnitude of water related information science area is crucial in formulating research priorities and topic identification. There is no consensus on the current state of water related information science area, thus this study was conducted to elucidate the trend in the water and water resource (W&WR) publications. We conducted an in-depth analysis on the total number of publications and the total number of citations of water related information science publication in the information science & library science area from Thomson ISI database. Ways to cluster water related information science publication for the topic terminology identification was proposed. Results revealed a main research trend in water related information science publication from 1987 to September 2016 did indicate that this trend is distributed interdisciplinary. Nine topic terminology categories were clustered and 31 corresponding sub-topic terminologies reflected the focus researches in water related information science area. Results further revealed that the water related information science publication is biased towards "water resource", "hydrology science", "water information" and "water quality" with a few studies were published in International Journal of Geographical Information Science, Scientometrics, Library Journal and Scientist for a coverage of 70.89% in totally. The cluster identification based on present study highlights that "hydrology science" topic focused on GIS and remote sensing technology, "water information" models mainly are spatial-temporal data and environmental information, "water conservancy" topic might be tend to engineer research. Increases in the number of publication with "water rights/law" and "water crisis" are consistent with the development and utilization of water resources by humans around the world. Finally, since only the water related information science area boundary and topic cluster terminology have been investigated, thesaurus-ontology can be stressed on water related information science area to achieve a better information retrieval

on computer science technologies.

1. Introduction

As the utilization of W&WR area continues growing to be one of the closet activities in human life, it becomes increasingly important to understand what W&WR researchers are focusing on.

"Hydrodynamics, Water acoustics, Cloud and Precipitation Physics, Hydrogeology, Hydrology, Aquatic biology, Aquatic ecology and lake ecology, Forest Hydrology, Soil and Water Conservation, Aquatic science, Diving medicine, Engineering Hydrology, Sea water resources, Farmland water conservancy, Soil and water conservation, Water conservancy Engineering Survey, Hydropower project, Water energy, Urban water supply and drainage works, Water Conservancy Project, Waterway transportation, Water Environment, Water Pollution Control Project, Ecological ecology of waters, Fisheries Economics, Waterway Economics" are the main contents for the water related topic classification standard in China .   In USA, Classification of Instructional Programs (2000 edition) provides a taxonomic scheme that will support the accurate tracking, assessment, and reporting of fields of study and program completions activity.   This updated taxonomy descriptions defines the water related areas such as Geology and Earth Science - Hydrology and Water Resources; Science, Ecology, evolution, classification and population biology - aquatic biology / lake science; Civil Engineering - Water Conservancy Project; Agricultural production and management – aquaculture; Natural Resource Management and Policy - Water and Marine Resources Management; Transportation and Transport Services - Water Transportation Services.

Taxonomic scheme on the water related topic standard in above represents that is indeed an interdisciplinary study, approximately including Geography Science, Information Science, Management Science and Economics etc. However, when considering water related researches, interdisciplinary integration is still relatively uncommon, this has been attributed to the boundaries among different disciplines, a qualification that refers to mutual misunderstandings between disciplinary cultures, and to a lack of effective communication [11]. Interdisciplinary efforts tend to be perceived as being more complex for participants than traditional intra-disciplinary collaborations because participants have different paradigms and approaches [6] [1] [17]. There is widespread recognition amongst scholars in the research and management of water systems that cross-disciplinary efforts are necessary to increase our understanding of complex water issues, the bridging of disciplinary perspectives figures prominently on the schedule, as evidenced by the promotion of fields like hydro ecology, ecohydrology, eco-hydro morphology and eco-geomorphology. It has been suggested that these fields extend beyond ecology, geomorphology and hydrology into other contributing fields such as civil engineering, economics and social sciences [2] [8] [13] [24]. Following the approach outlined in the previous studies, the landscape of water related research should be mapped using a combination of methods, as especially water-related issues call for the development of cross-disciplinary approaches to understand the systemic nature of the water landscape with its ecological, social, political, economic and cultural dimensions [10,22][26]. This suggests that traditional disciplinary divisions between the biological, environmental and physical dimensions of W&WR researches are still prominent.   The complex societal content of water related issues not only demands understanding from the natural science but also from the social sciences including psychology, sociology, geography, political science, economics and policy studies [3,16,20]. Meanwhile, Vugteveen has done a research on rivers to verify that over half of the social are related to the field of hydrology and water resources and are

published in the more general water resources (management) oriented journals such as Water Resources Management and Water Policy, other societal are within environmental management and in water sciences and technology [25].

For instance, the earliest Web of Science-recorded water related publications was published in the area of information science & library science, which was written by Fabietti in 1996 to explain a logic of water distribution system [7], this research area is a combination between Information Science and Hydrology-Water Resources Science. Such similar phenomenon appeared as ontology related research on the water conservation [28]. The problem of information management has become more apparent with the very rapid development of the interdisciplinary information classification and storage while some predominantly belongs to a single research field whereas most are researched by a variety of fields, namely:

(1) Water related issues were published in a variety of fields. Without the related background knowledge, the analysis on this problem seems not so vivid in a particular field.

(2) Water related issues, published on each journal with different field terminologies, scattered information distributed without systematic integration, making the information storage and retrieval inconvenienced.

(3) Knowledge organization on technologies is hard to achieve due to the lack of field terminology identification capacity.

The above discussion indicates that there is a clear need to improve comprehensive and systematic topic classification capacities in order to make the most of knowledge organization on technologies affordances today. This paper reports on a survey regarding to water and water resource (W&WR) research in Information Science and Library Science (named water related information science area) and deals with very specific topic classification process, terminological and representational aspects (control vocabulary) and evaluation process on water related information science area, aimed at integrating water related information science on filed ontology to achieve information retrieval on computer science technologies, and other metadata enabled applications. There is a disconnection between theoretical ontology building and retrieval technology systems using topic terminology in water related information science research, which makes it very often difficult to assess some of the theoretical propositions. This paper offers both theoretical propositions and their empirical evaluation with an academic journal corpus.

2. Data Sources

Most journals of these topics in Information Science and Library Science are included in the SCI-Expanded database and Social Sciences Citation Index, such as MIS Quarterly, Journal of Information technology, Journal of the American Medical informatics association et al., there are totally 81 journals in this research area.

The database of this paper was built based on the online version of the Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI) via the Web of Science. In this study, the topic= "water or river* or lake* or basin* or hydro*" was used as the main topic retrieval key words to traverse the 81 journals in the research area of Information Science and Library Science from 1987 to September 2016. Totally 1052 book review and 146 articles were indexed related to the W&WR in the research area of Information Science and Library Science. To avoid underestimating data, the effect of topic relation degree on the data is assessed before selecting appropriate filters. In this study, we used artificial and expert selection to refine the valid data that takes more relationship with the W&WR research. People with the background of information science & library science are divided into two groups, classified as group one "environment, water, river and basin" and group two "ecology, population and energy". Then main contents of abstracts

related with W&WR area used as standard to select the useful data sources by artificial identification. The details of the standard as follows:

(1) The topic in the abstracts are related with river/water/lake or river/water/lake resources/ or river/water/lake environment. Otherwise, some environment-related topics should be excluded, such as "research collaboration environments" [18,19,9] and "Knowledge sharing and transfer environments"[14] are not really related with water environment.

(2) The topic in the abstracts are related with some factors influenced by rivers, such as soil, weather, population and project. For example, plant nutrition and soil chemistry towards soil water [15], information needs of engineers of Volta River [21] are selected as our research materials.

Thus, the results after filtering-selection process of the most related topic may contribute a more accurate conclusion on this research. 77 publications included in the totally publication papers were assessed to have a strong association with W&WR area.

3. Methods for Construction of the Topic Clustering Model

Table 1 Top 24-word frequency results from top two authority journals in water resources area.

Keywords	Word Frequency Count	Keywords	Word Frequency Count	Keywords	Word Frequency Count
wastewater	239	surface water	58	wastewater reuse	28
drinking water	212	risk assessment	42	soil moisture	27
wastewater treatment	211	transport	41	hydrodynamics	20
water quality	138	drinking water treatment	40	sustainability	20
water treatment	136	water reuse	38	environmental fate	16
groundwater	136	remote sensing	33	water supply	12
modeling	97	transformation products	32	water purification	11
climate change	63	solute transport	31	water security	11

At the beginning of this study, we calculated the word frequency in the field of water resources from the Journal "water research" and "water resource research" between the year 2007-2016, top two journals with an Impact Factor (IF) journals on the research area of water resources area, which could reflect the trends in this field and belong to authority journals in this field. Table 1 is results for word frequency from these two journals which illustrated a guide for summarizing topic terminologies during the later work.

3.1 Literature Co-Citation and Literature Coupling

In the first step, from the co-citation and coupling results in Figure 1 and Figure 2, publication number 13 (Yi H, Jie W. (2011)) , number 1 (Gao W, Chen Y, Liu Y, Guo H. (2015).) and number 3 (Gao W, Guo H. (2014)) represented as owning relationships, the same relationship appeared with number 38 (Yu C, Cheng X(2012)) and number 67(Van Der Knijff JM (2010)). "eco-environment" and "hydrologic model/system" were titled as the initial topic terminologies based on their topic contents in Table 2.
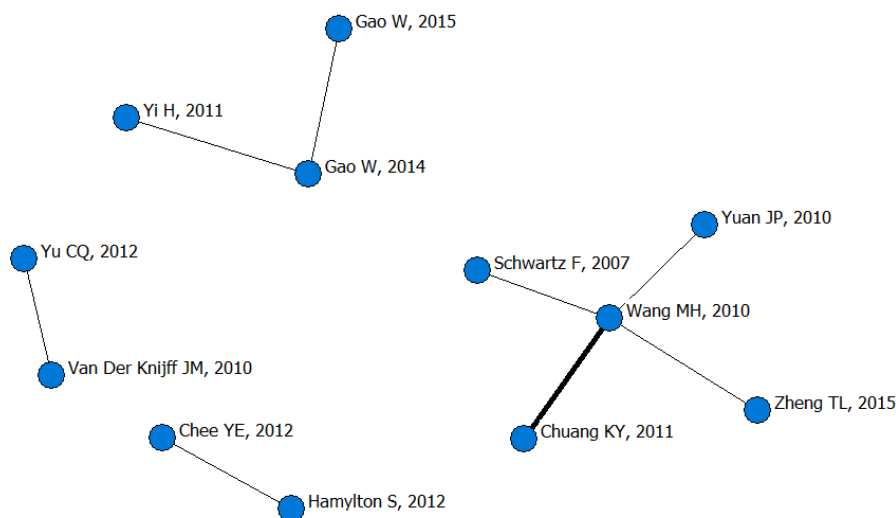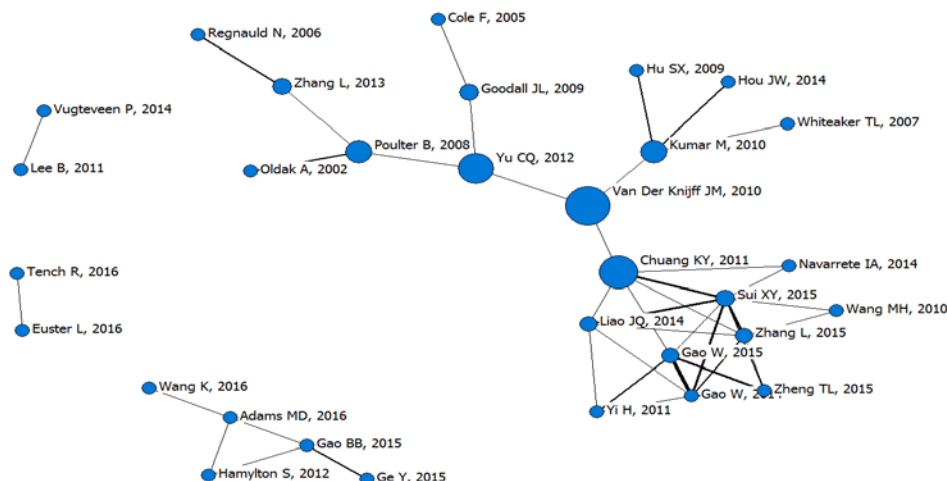
Figure 1 Results of literature co-citation.

Figure 2 Results of literature coupling.

Table 2 Initial topic terminologies in the first step

Topic terminology	Number of corresponding publication
eco-environment	No.1,3,13
hydrologic model/system	No.38,67

3.2 Intersections of the Remaining Results on Literature Co-Citation and Coupling

In the second step, by using the results from step one, intersections of literature co-citation and coupling methods was considered as the sources to obtain the clustering results.

Based on the results of Figure 1 and Figure 2, Table 3 is a further topic categories results named with topic terminology. In this process, the detailed explanations on achieved results are reported as follows:

(1) The publication number 16(Zhang L (2015)) is associated with both publications number 3(Gao W, Guo H. (2014)), number 31(Liao J, Huang Y. (2014)) and number 37(Sui X (2015)), and the correlation coefficients are 2,1 and 3 respectively. Compared on this result, publication number 16 might be much suitable to the group which includes publication number 37. Almost the same

relevance results happened with publication number 28 (Zheng T(2015)), this publication might belong to the group which including publication number 1 and 31.

(2) Publication number 26 (Chuang K, Wang M (2011)) is associated both with publications number 1,31 and publications number 16, 37, and the correlation coefficients are 1, 1, 1 and 3 respectively; also, publication number 26 has a correlation coefficient 1 with publication number 67(Van Der Knijff JM,(2011)). Considering the topic of publication number 16, it is better to make publication number 26 into a new group different with the topic group "eco-environment" and "hydrologic model/system".

(3) Although publication number 30 (Hamylton S. (2012)) and publication number 42 (Chee YE, Elith J. (2012)) are not associated with the publication number 31, we believed that both of the two publications are suitable for topic "ecosystems" from their research contents. Publication number 20 (Gao B, Wang J, (2015)) and number 44 (Adams MD, Kanaroglou PS (2016)) both have relationships with publication 30 from the result of literature coupling matrix, therefore publications number 30, number 42, 20, 44, number 52 (Ge Y, Wang JH, (2015)), number 53 (Wang K, Chen N (2016)) can be clustered.

(4) Publication number 24 (Wang M, Yu T, Ho Y. (2010)), 36 (Yuan JP, Yue WP (2010)),76 (Schwartz F, Fang YC. (2007)) are clustered into group "resource management and plan" with publication number 16,26 and 37. Number 64 (Navarrete IA, Asio VB. (2014)) has a relationship with number 26 and 37 whilst regarding to a content named "soil science". Number 12 (Whiteaker TL, Maidment DR (2007).), 22 (Hou J, Mi W, Sun J. (2014)), 23(Kumar M, Bhatt G, Duffy CJ. (2010)) and 48 (Hu S, Bian L. (2009)) are associated with number 67 as well as number 8 (Poulter B, Halpin PN. (2008)),15(Oldak A, Jackson TJ (2002)),14(Cole F. (2005)),39 (Regnauld N, Mackaness WA. (2006)), 51 (Zhang L, Guilbert E. (2013)) and 65 (Goodall JL, Maidment DR. (2009)) are related with number 38, based on the above phenomenon, these publications should be clustered into one group, the same situation reaches with number 2 (L. E. (2016)) and 18 (R T. (2016).), 68 (Vugteveen P, (2014)) and 73 (Lee B, Kwon O (2011)).

Table 3 Intersections results on topic clustering.

Cluster of topic	Topic terminology	Number of corresponding publication
1	eco-environment	No.1,3,13
	water pollution	No.28
	ecosystems	No.31,30,42
	water monitoring networks	No.44,20,52,53
2	resource management and plan	No.16,24,26,37,76,36
	soil science	No.64
3	hydrologic model/system	No.38,67,23,48,12,65,22
	hydrologic environment	No.8,15,51,39,14,
4	environmental issues	No.2,18
5	interdisciplinary research fields/water culture	No.68,73

3.3 Coupling Network Analysis Based on Artificial Labeling Keywords

In the third step, the remaining publications collected by keywords coupling network analysis.

The definition of the artificial labeling keywords is that if the word appeared in title or keywords once or in abstract three times, these words would be named by new artificial labeling keywords from a keyword plus participle. Considering the distance dimension of the publications, the shorter distance used as a standard to make the uncluttered publications into a related group, the result can be showed in Figure 3 and Figure 4.
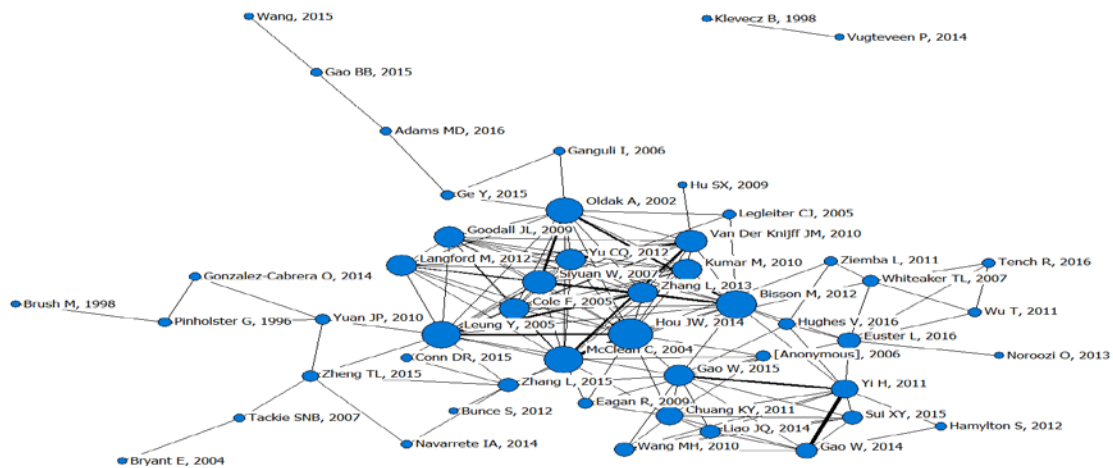
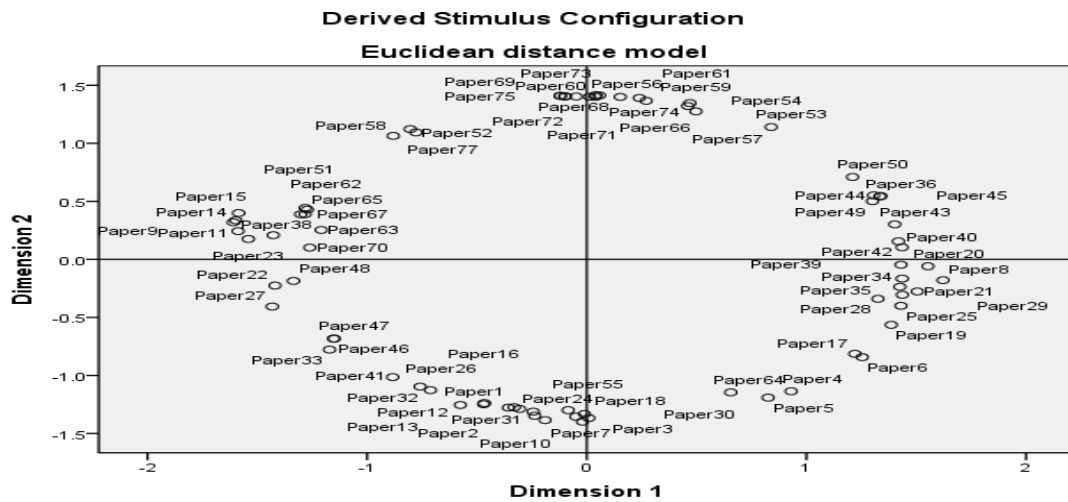Figure 3 Results of keywords coupling network analysis.

Figure 4 Euclidean distance model based on key words coupling relationship.

From the results of coupling network analysis, Table 4 showed an updated topic clustering results.

Table 4 Results from keywords coupling network.

Cluster of topic	Topic terminology	Number of corresponding publication
1	eco-environment	No.1,3,13,25
	water pollution	No.28
	ecosystems	No.31,30,42
	water monitoring networks	No.44,20,52,53,34
2	resource management and plan	No.16,24,26,37,76,36
	soil science	No.64
3	hydrologic model/system	No.38,67,23,48,12,65,22,9,63,58
	hydrologic environment	No.8,15,51,39,27,11,14,70
4	environmental issues	No.2,18,55
5	interdisciplinary research fields/water culture	No.68,73,56
6	water treatment	No.71,57,74
7	conservancy engineer	No.59,66
8	water resources use	No.46,47
9		No.7
10	classification schemes	No.41
11	prediction of water	No.32
12	model/system	No.77

Number 34 (Gao B (2015)) was coupled with number 20, 44 and 52. Number 9 (McClean C. (2004)), 58(Bunce S (2012)),70 (Bisson M, Bini M. (2012)), 11(Siyuan W (2007)), 14 (Cole F. (2005)) had a relative strong relationship with number 22,23, 38 and 67, the same as number 63 and 51, 27(Ganguli I. (2006)) and 15. Number 55(Wu T. (2011)) both related with number 2, 18 and number 12 whilst distrusted into group "environmental issues" instead of "hydrologic model/system" for the topic content. Number 56 (Klevecz B. (1998)) was clustered into group number 68 as the coupling result. The other publications number 71(Hellerstein N. (2015)), 57(Pinholster G. (1996)), 74 (Gonzalez-Cabrera O (2014)) or 59(Bryant E. (2004)),66 (Tackie SNB, Adams M. (2007)) or 46, 47 (Eagan R. (2009)) or 41(Ziemba L, Cornejo C, Beck HW. (2011)) or 32(Hughes V. (2016)) or 77(Legleiter CJ, Goodchild MF. (2005)) coupled in Figure 3 were clustered into their corresponding topic terminology regarding to the detail contents. For example, since number 41 and number 32 paid attention on different contents, they were divided into two groups. Especially, number 7 (Noroozi O (2013)) seemed to be associated with number 2, actually, the topic of the two papers were different in a certain degree.

Table 5 Results of the last remaining publications based on Euclidean distance model.

More related number of publication	Number of Publication distance <0.5	Number of publication distance >0.5
No.43,45, 40	No.44,42,20	
No.72,75, 60,61,69,54,53	No.71,57,74	
No.21,	No.8,29	No.44
No.19	No.8,21	
No.29	No.8,21,19	
No.4,5,6,17	No.64	No.19
No.10	No.3,13,3	
No.33,35	No.46,47	
No.50,49	No.36	

From Euclidean distance model result appeared in Table 5, the publications number 4(Smet E D NP. (1997).), number 5(Francke T (2008)), number 6, number 17(J B. (2016)), number 19 (Thelwall M, Vann K (2006)), number 21(Manafy M. (2007)), number 29 (Daviss B. (2004)) can be clustered into the same topic group.

3.4 Delphi Method of Expert Evaluation Mechanism

In the fourth step, based on the mechanism method result, the expert evaluation should be done to verify and update the topic clustering. Delphi method of expert evaluation mechanism [23] was adopted to improve the evolution function of topic clustering being developed by this study. In this research, the experts are in the field of water resource management. And the e-mail is used as the main review way to send message to the experts. The expert review is generally divided into two stages: discussion and voting. Voting should be based on full discussion. For the classification between topic terminologies, the expert review must follow the rules:

(1) Elected two highest authority experts in water & water resources field, the two authoritative experts have the final voting rights and are entitled to exercise a veto with only one vote on all different opinions from other experts. If someone does not agree with this decision, the review should start recruiting.

(2) A discussion among the submission of the topics, the evolution of the whole scheme could be achieved.

(3) Bearer scoring. The strict form is to distribute all the schemes to the experts, secret to each selected score. Each expert has a personal attention to the private voting space.

(4) No abstained.

(5) Collect all the schemes to the two experts. By the judgment of the two experts, the final determination scheme would be confirmed.

The results can be measured from the arithmetic mean of the optional schemes, as illustrated in Table 6.

Table 6 Score chart.

Experts	Optional schemes					
	1	2	…	j	…	n
1	C_{11}	C_{12}	…	C_{1j}	…	C_{1n}
2	C_{21}	C_{22}	…	C_{2j}	…	C_{2n}
…	…	…	…	…	…	…
I	C_{i1}	C_{i2}	…	C_{ij}	…	C_{in}
…	…	…	…	…	…	…
M	C_{m1}	C_{m2}	…	C_{mj}	…	C_{mn}

From Table 6, the arithmetic mean value of all schemes can be calculated according to equation 1:

$$\square\, M_j \;=\; 1\,/\,m_j \sum C_{ij} \tag{1}$$

M_j is the arithmetic mean value of j scheme ($j = 1, 2 \ldots, n$). m_j stands for the number of experts who participated in f scheme. C_{ij} stands for the value that i expert set for j scheme ($i = 1, 2 \ldots, m$). From the result, the scheme which has the maximum arithmetic mean value can be chosen as the recommended scheme.

If the arithmetic mean value of a publication is close to two schemes, it should be divided to the two different topic terminologies. For example, publication number 12 belongs to topic 1 and topic 3, number 16 is summarized into group 1 and 2, the same status as number 27 (topic 2 and 3), number 44 (topic 1 and 5) and number 65 (topic 3 and 4). During the evaluation step, one part is to definite the sub-topic, a second part it to verify the main topic of these publications. All the two parts should be complied with the standard of the arithmetic mean value scheme.

4. Results and Discussion

4.1 Topic Clustering Terminologies on W& WR Area

By collating with mechanism method and expert evaluation, we get a final topic clustering result. The comparison results are in Table 7.

There are two topics named "water chemistry" (number 29) and "population/ climate/ pollution/ wastewater" (number 33) clustered extra exception the other similarly topic level. Based on the expert's evaluation, we put these two publications into their corresponding thematic category.

In Table 7, the similarity of topic terminology recognition rate (number of topic clustering model over number of expert evaluation in sub-topic level) reaches 67.74%, and the specific publication classification similarity reaches 72.73% (21 publications exceptional). There are almost 6 publications belong to two categories whilst they are own one topic level tag without the expert's evaluation (publications such as number 12, 16, 19, 21, 44 and 65), excluding these obscure publications, the similarity ratio reaches a much higher value as 80.52%.

There are nine topic terminologies in topic level in the final result after the experts' verification. They are "water management", "water resource", "hydrology science", "water information", "water quality", "water rights/law", "water crisis", "water engineer" and "interdisciplinary research

fields/water culture". Furthermore, almost 31 sub-topic terminologies were summarized according to each topic level. These topic terminologies might be utilized as a thesaurus-ontology with a further study, for that some equivalent and hierarchical relationships appeared among "topic" and "sub-topic".

Table 7 Combined compared results of topic clustering model and experts' evaluation method.

Results of topic clustering model			Results of after added the experts' evaluation		
Topic	**Sub-Topic**	**Publication Number**	**Topic**	**Sub-Topic**	**Publication Number**
water management	water treatment	No.57,71,74,54,60,61,69,72,75,53	water management	water treatment	No.57,71,74,72
	water resources use	No.46,47,		water resources use	No.46,47
				water management strategy	No.43,45
				water quality management	No.44
				economic	No.19,21,75
				water resources allocation	No.12,16
water resource	resource management and plan	No.9,16,24,26,36,37,76,49,50	water resource	resource management and plan	No.9,16,22,24,26,36,37,58,63,76
	soil science	No.64		soil science	No.27,64
hydrology science	hydrologic model/system	No.12,23,38,48,65,67,22,9,63,58	hydrology science	hydrologic model/system	No.12,23,38,48,65,67
	hydrologic environment	No.8,11,15,27,70,14,51,39,		hydrologic environment	No.8,11,15,27,62,70
				water chemistry	No.29,49
	water chemistry	No.29		hydrologic monitor	No.52,53
water information	environmental information model	No.4,5,6	water information	spatio-temporal data model	No.39,51,65
	water resource integrated	No.17,19,64		terminology vocabularies	No.14,72
	water economic management model	No.21,		environmental information model	No.4,5,6
				water resource integrated	No.17,19
				water economic management model	No.21
water quality	eco-environment	No.1,3,10,13,25	water quality	eco-environment	No.1,3,10,13
	water pollution	No.28		ecosystems	No.30,31,42
	ecosystems	No.31,30,42,		water monitoring networks	No.20,34,44
	water monitoring networks	No.44,20,52,53,34,40,42,43,45		model/system	No.77
	model/system	No.77		water security	No.60,61
				water pollution	No.28
water rights/law	environmental issues	No.2,18,55	water rights/law	environmental issues	No.2,18,55,69
				water share agreement	No.45
water crisis	prediction of water	No.32	water crisis	population/ climate/ pollution/ wastewater	No.33,35
	population/ climate/ pollution/ wastewater	No.33,35		fish	No.54
				prediction of water	No.32
water engineer	conservancy engineer	No.59,66	water engineer	conservancy engineer	No.40,59,66
interdisciplinary research fields/water culture	classification schemes	No.41	interdisciplinary research fields/water culture	classification schemes	No.41
	interdisciplinary research fields/water culture	No.68,73,56		interdisciplinary research fields/water culture	No.7,50,56,68,73

4.2 Distribution Details on W&WR Area

The purpose of this research is to focus on the identification of the topic terminology on science area. Each journal published particular topic publications, keywords and journal belonging are two

important factors on the features of topic. In Table 9, TN is short for total number. From Table 8 and Table 9, the representative analysis among topics identification are reported as following:

(1) Topic 1 almost distributes in 14 publications, and it seems this topic has some relationships with topic 2, 3, 4, 5 and 6, especially a much stronger relationship with topic 4 with the ratio of 42.86%. The corresponding terminology tags are "water treatment" and "terminology vocabularies", "economic" and "water resource integrated" "water economic management model". And the rest relationships corresponding terminology tags are "water resource allocation" and "hydrologic model/system, resource management and plan", "water quality management" and "water monitoring networks", "water management strategy" and "water share agreement".

(2) Topic 2 distributes in 12 publications, half of them are published in the journal of Scientometrics. Topic 3 has 81.25% papers published in the journal of International Journal of Geographical Information Science, "hydrology science" always focused on the GIS and remote sensing technology in the aspects such as water rights, water supply, land cover and water balance.

(3) Topic 4 "water information" means the related data in the water research area, especially in Journal of Information Science, partly of them mainly pay close attention on the data-related vocabularies by representational practice, partly of them follow closely the semantic integrations of government data on water quality management. In this data research, the main models are spatial-temporal data, environmental information, water resource and water economic management.

(4) More than 73.33% publications of topic 5 are published in the journal of International Journal of Geographical Information Science and Scientometrics. The remaining publications are average distributed in the journal of Scientist and Library Journal, both of two classifications of the research points are "eco-environment" and "water security".

(5) Topic 6 and 7 seem to have particular area research goals, and the results of the distributed journals reflect the evidence. From the main key words of these articles, the "rights/law" is in the direction of "environment", "management", "social-economic', "allocation" and "supply safety", the "crisis" also follows with interest as "population growth", "environment", "climate change" "conservation" and "humans". 88.89% publications are published from 2011 with an increasing trend, specifically beginning with 2014, and the annual increased growth rate reached 66.67%. This status demonstrates that the water rights/law and water crisis play two significant roles along with the development and utilization of water resources by humans around the world, such as south Africa[5], south western Pakistani Baluchistan[7], Israel[4] and United States[28].

(6) Topic 8 focuses on "water conservancy" and emerged on 2004, 2007 and 2015, it seems that in water related information science   research area, this topic is seldom researched, this might be tended to engineering research.

(7) Topic 9 is the interdisciplinary research field about water culture, from the key words of these publications, we can find that in water related information science research area, some cooperation researches can be further improved. In the current research achievements, some inspiration about interdisciplinary research topics appeared, such as cluster analysis (GIS and spatial-temporal analysis), knowledge sharing and knowledge transfer (computer supported collaborative learning), knowledge retrieval and knowledge organization (ontology classification, digital data and information monitoring), water shortage (water science and environment).

(8) Taking everything into account, International Journal of Geographical Information Science, Scientometrics, Library Journal and Scientist covered almost 70.89% publications in totally, among them, the publications numbers from topic 1 to topic 5 cover 86.89% ratio, especially topic 2, 3 and 5 are the most focused area. This phenomenon is sufficient to show that these four journals can be used as an interdisciplinary research journal in the area of "water resource, hydrology science, water information and water quality".

Table 8 Top frequency of key words statistics for topic in the abstract.

Topic Number	Topic	The main key words
1	water management	GIS, water resources, water quality, water rights, water supply, bibliometric analysis, water treatment, river network
2	water resources	GIS, bibliometric approach, water resources, water quality, remote sensing, bibliometric analysis, soil moisture
3	hydrology science	GIS, soil moisture, hydrological process, flood, water rights, remote sensing, water supply, land cover, environmental data, water balance
4	water information	GIS, environment, remote sensing, river network, semantic relationship, public transport network, Simulated data
5	water quality	bibliometric approach, water pollution, ecosystems, environment, water quality, water treatment, water monitoring networks, water purification system
6	water rights/law	water rights, environment, water management, socio-economic rights, shareholders, optimize water allocation, property, water law, water supply safety
7	water crisis	population growth, GIS, environment, climate change, water conservation, water crisis, water for humans
8	water engineer	water conservancy engineer, water conservation, water predictions,
9	interdisciplinary research fields/water culture	interdisciplinary research, water management, river research, river runs, water shortage, cluster analysis, knowledge sharing, knowledge transfer, digital library, classification schemes, knowledge retrieval, ontology,

Table 9 Publication number for each topic in every Journal.

Journal name	Topic 1 (TN:14)	Topic 2 (TN:12)	Topic 3 (TN:16)	Topic 4 (TN:11)	Topic 5 (TN:15)	Topic 6 (TN:5)	Topic 7 (TN:4)	Topic 8 (TN:3)	Topic 9 (TN:6)	Total number
International Journal of Geographical Information Science	2	3	13	4	6	0	0	0	0	25
Scientometrics	1	6	0	0	5	0	0	1	1	13
Scientist	3	1	3	1	2	0	0	1	1	11
Library Journal	1	0	0	0	2	1	3	0	0	7
Total Number	*7*	*10*	*16*	*5*	*15*	*1*	*3*	*2*	*2*	*56*
Journal of The American Society for Information Science and Technology	1	1	0	1	0	0	0	0	0	2
Econtent	1	0	0	1	0	1	0	0	0	2
Journal of Information Technology	1	0	0	0	0	0	0	0	0	1
Social Science Information Sur Les Sciences Sociales	1	0	0	0	0	1	0	0	0	1
Australian Library Journal	0	1	0	0	0	0	0	0	0	1
Government Information Quarterly	2	0	0	1	0	0	0	0	0	2
Journal of Information Science	0	0	0	2	0	0	0	0	1	3
Transinformacao	1	0	0	0	0	0	0	0	0	1
Australian Library Journal	0	1	0	0	0	0	0	0	0	1
Social Science Computer Review	0	0	0	1	0	0	0	0	0	1
Reference & User Services Quarterly	0	0	0	0	0	1	0	0	1	2
Social Science Information Sur Les Sciences Sociales	0	0	0	0	0	1	0	0	0	1
Ethics and Information Technology	0	0	0	0	0	1	0	0	0	1
Law Library Journal	0	0	0	0	0	0	1	0	0	1
Electronic Library	0	0	0	0	0	0	0	0	1	1
African Journal of Library Archives and Information Science	0	0	0	0	0	0	0	1	0	1
International Journal of Computer-Supported Collaborative Learning	0	0	0	0	0	0	0	0	1	1

4.3 Caveats of Evaluation on Water Related Information Science Area

There are a number of caveats to this study. First, the naming of terminology inevitably involves some subjective decisions about the appropriate category in which to place a particular clustering term. Although in the long term the contribution of each site or search topic terminology to the whole is small, there is a margin of imprecision to the classifications. In interpreting the clustering information, one should focus on the overall pattern, rather than the precise size of each category if someone belongs to more than one category.

Second, from the development history and knowledge boundary, although 9 topics' terminologies might be used as water related information science temporary research boundary and the corresponding 31 sub-topics' terminologies recognized as hot points from Table 7, there may be some relationship bias, while the direction of which is impossible to detect. In particular, the semantic relationship of each terminology is not clear for a further conceptual retrieval on instructional technology. For example, definition of the relationship between sub-topic term "economic" in "water management" and "water economic management model" in "water information" is a critical role in construction on the control vocabulary.

Third, the neologism of the water information related area and the co-occurrence frequency of the word in Table 8 might enjoy some relationships with identification of association rules[12]. For example, "environmental issues" and "water share agreement" in topic "water rights/law" might have association relationship with "water rights, environment, water management, socio-economic rights, shareholders, optimize water allocation, property, water law, water supply safety. In a certain domain, within syntax rules, it is possible to distinguish the terminologies hierarchical relationship from equivalence relationship [27].

Lastly, it should also be borne in mind that a thesaurus-ontology on the water information related area might be achieved a better information retrieval on computer science technologies. That is, however, unlikely, given the large size of the set of contents from which the sample was drawn.

5. Conclusions

A total of 146 publications in 5 topic research were indexed in Web of Science during 1987-September 2016, in which 77 publications had a strong relationship with W&WR area. Three stages of topic clustering were observed on 77 publications. Nine topic categories were identified on W&WR research in the discipline of information science and library science. Through a verification of the Delphi method of expert evaluation mechanism, 31 further sub-topic categories reflected the focus of W&WR research.

The similarity of terminology recognition rates between topic clustering model and experts' evaluation method reached a much higher credibly value as 80.52%. Taking everything into account, Journal International Journal of Geographical Information Science, Scientometrics, Library Journal and Scientist covered almost 70.89% articles in total, among them, topic one to topic five cover 86.89% ratio, especially topic two, three and five are the most concentrated area. This phenomenon is sufficient to show that these four journals can be used as an interdisciplinary research journal in the area of "water resource, hydrology science, water information and water quality". The cluster identification based on present study highlights that "hydrology science" topic focused on GIS and remote sensing technology, "water information" models mainly are spatial-temporal data and environmental information, and "Water conservancy" topic might pertain to engineering research. Increases of the number of publications with "water rights/law" and "water crisis" are consistent with the development and utilization of water resources by humans around the world.

Water information science related area is a valuable disciplinary combined with the research area in information science & library science and water resources. Topic clustering results on W&WR

can be used as a guide to reveal the interdisciplinary research focus, as well as prove a basic ontology-thesaurus for a further summarize on water information science related knowledge.

Acknowledgments

References

[1] Benda LE, Poff NL, Tague C, Palmer MA, Pizzuto J, Cooper S, Stanley E, Moglen G. (2002). How to avoid train wrecks when using science in environmental problem solving. BIOSCIENCE, 52(12), 1127-1136.

[2] Bond B. (2003). Hydrology and ecology meet - and the meeting is good. HYDROL PROCESS, 17(10), 2087-2089.

[3] Brierley G J FKA. (2008). Moves toward an era of river repair. River futures: an integrative scientific approach to river repair, 3-15

[4] Conn D. (2015). Let There Be Water: Israel's Solution for a Water-Starved World. Library Journal

[5] Cooper N, Swan A, Townend D. (2014). A confluence of new technology and the right to water: experience and potential from South Africa's constitution and commons. Ethics and Information Technology, 16(2), 119-134.

[6] CULLEN P. (1990). THE TURBULENT BOUNDARY BETWEEN WATER SCIENCE AND WATER MANAGEMENT. FRESHWATER BIOL, 24(1), 201-209.

[7] Fabietti U. (1996). Water-sharing systems in south-western Pakistani Baluchistan: Between synchronization and desynchronization. Social Science Information Sur Les Sciences Sociales, 35(4), 721-734.

[8] Hannah DM, Wood PJ, Sadler JP. (2004). Ecohydrology and hydroecology: A new paradigm'? HYDROL PROCESS, 18(17), 3439-3445.

[9] Kawakami AK. (2000). "LOEX" of the west: Collaboration and instructional design in a virtual environment. Library Quarterly, 70(4), 506-508

[10] Lenders HJR, Knippenberg L. (2005). The temporal and social dimensions of river rehabilitation: towards a multi-dimensional research perspective. Archiv fuer Hydrobiologie Supplement, 155(1-4), 119-131

[11] M. H. (2009). Integrating Knowledge: The Key Challenge for a New Paradigm in River Management. Geography compass, 3(6), 1988-2010.

[12] Maedche A. (2002). Ontology Learning for the Semantic Web (252). Boston: Springer. (Reprinted

[13] Merkx F, van den Besselaar P. (2008). Positioning indicators for cross-disciplinary challenges: the Dutch coastal defense research case. Research evaluation, 17(1), 4-16.

[14] Nakano D, Jr. Muniz J, Jr. Batista ED. (2013). Engaging environments: tacit knowledge sharing on the shop floor. Journal of Knowledge Management, 17(2), 290-306.

[15] Navarrete IA, Asio VB. (2014). Research productivity in soil science in the Philippines. SCIENTOMETRICS, 100(1), 261-272.

[16] Pahl-Wostl C, Craps M, Dewulf A, Mostert E, Tabara D, Taillieu T. (2007). Social learning and water resources management. ECOL SOC, 12(52)

[17] Petts GE, Nestler J, Kennedy R. (2006). Advancing science for water resources management. HYDROBIOLOGIA, 565, 277-288.

[18] Rudzioniene J. (2014). Collaboration in Libraries and Learning Environments. Electronic Library, 32(6), 923-924.

[19] Siqin T, van Aalst J, Chu SKW. (2015). Fixed group and opportunistic collaboration in a CSCL environment. International Journal of Computer-Supported Collaborative Learning, 10(2), 161-181.

[20] Surridge B HB. (2007). Science-driven integrated river basin management: a mirage. INTERDISCIPL SCI REV, 32(2), 298-312.

[21] Tackie SNB, Adams M. (2007). Information needs and seeking behaviour of engineers in Ghana: A case study of the Volta River Authority. African Journal of Library Archives and Information Science, 17(2), 69-78

[22] Thorp J H SJAT. (2007). EDITORIAL GLOBAL PARTNERSHIPS AND THE NEW INTERNATIONAL SOCIETY FOR RIVER SCIENCE (ISRS. RIVER RES APPL, 23(1), 1-5.

[23] van Beeck E. (2017). The Delphi method: A tool to support injury control and trauma care policy. INJURY, 48(1), 3-4.

[24] Vaughan IP, Diamond M, Gurnell AM, Hall KA, Jenkins A, Milner NJ, Naylor LA, Sear DA, Woodward G, Ormerod SJ. (2009). integrating ecology with hydromorphology: a priority for river science and management. AQUAT

CONSERV, 19(1), 113-125.

[25] Vugteveen P, Lenders R, Van den Besselaar P. (2014). The dynamics of interdisciplinary research fields: the case of river research. SCIENTOMETRICS, 100(1), 73-96.

[26] Vugteveen P, Leuven RSEW, Huijbregts MAJ, Lenders HJR. (2006). Redefinition and elaboration of river ecosystem health: perspective for river management. HYDROBIOLOGIA, 565, 289-308.

[27] Zhao Y, Nunes JMB, Deng Z. (2014). Construction and Evolution of a Chinese Information Science and Information Service (CIS&IS) Onto-Thesaurus. Knowledge Organization, 41(2), 131-144

[28] Ziemba L, Cornejo C, Beck HW. (2011). A water conservation digital, library using ontologies. Electronic Library, 29(2), 200-211.