Association for Information Systems
AIS Electronic Library (AISeL)

PACIS 2015 Proceedings

Pacific Asia Conference on Information Systems (PACIS)

2015

The Trend and Intellectual Structure of Digital Archives Research

Fang-Ming Hsu
National Dong Hwa University, fmhsu@mail.ndhu.edu.tw

Chun-Min Lin
National Dong Hwa University, d9832012@ems.ndhu.edu.tw

Chiu-Tsu Fang
Tungnan University, jofan@mail.tnu.edu.tw

THE TREND AND INTELLECTUAL STRUCTURE OF DIGITAL ARCHIVES RESEARCH

Fang-Ming Hsu, Department of Information Management, National Dong Hwa University, Hualien, Taiwan, R.O.C., fmhsu@mail.ndhu.edu.tw

Chun-Min Lin, Center of General Education, Taiwan Hospitality and Tourism College, and Department of Information Management, National Dong Hwa University, Hualien, Taiwan, R.O.C., d9832012@ems.ndhu.edu.tw

Chiu-Tsu Fang, Department of Business Administration, Tungnan University, Taipei, Taiwan, R.O.C., jofan@mail.tnu.edu.tw

Abstract

Archives are an extremely valuable part of cultural heritage since they represent the trace of the activities of a juridical person or organization in the course of their business. Through various information technology (IT), tremendous amount of digital archives (DA) are created. These archives are the basis for providing evidence and knowledge in everlasting memory of human society. The management of digital archives becomes a fast growing field throughout last decade and introduces abundant articles in academia. However, their trend and intellectual structure have remained obscure in the research community.

To map the trend and intellectual structure of DA research, this study identifies the high-impact articles as well as the correlations among these scholar publications. In this study, text mining techniques, such as co-word and cluster analysis, have been deployed to investigate the intellectual pillars of the DA literature. This study exposes researchers to a new way of profiling knowledge networks and their relationships in the research area of DA, thereby helping academia and practitioners better understand up-to-date studies. The results of the mapping can help identify the research direction of DA research, provide a valuable tool for researchers to access DA literature, and act as an exemplary model for future research.

Keywords: Digital Archives, Intellectual Structure, Co-word Analysis, Text Mining

1 INTRODUCTION

Archival records are important form of explicit knowledge in organizations. Since the late 1990s, many government agencies have launched projects with a particular emphasis on using IT to provide electronic information and services to citizens and businesses (Chen et al. 2001). This digital information cannot be considered trustworthy and is easily lost in a self-perpetuating and expensive cycle of obsolescence and incompatibility (Duranti 2001). An archive is an accumulation of historical records and contains primary source of knowledge that has accumulated over the course of an individual or organization's lifetime, and is kept to show the function of that person or organization. Archives should be improved not only in terms of the completeness of planning, but also in terms of its specificity and consistency of services (Buchanan et al. 2012). In the era of information technology (IT), digital archives are critical to knowledge management especially in digital society. Government agencies often organize their services and operations into programs that may be changed in response to a host of factors, including IT implementations (Walker 2001). After the promotion of electronic government, tremendous records are created and preserved through records and archives management system. Through various IT and media, past activities of humanity are preserved as critical memory after filing.

In the sound archive field, a long-term maintenance of the collective human memory in its original form is not sustainable. All physical carriers are subject to degradation and the information stored on such carriers is bound to vanish. Only a re-mediation of the original documents can prevent precious knowledge from being permanently lost (Canazza 2012). Many digital archives (DA) programs regarding human activities have been developed in last decades. Several trends are emerging in the management of digital archives, including a shift from paper-based storage to computer-based systems, from managing information to supporting its access and retrieval, and from cost-reduction to continued process improvement (Stephens 1998). These trends all highlight the need for the management of digital archives in the operation of digitizing artifacts.

Digital archives are created by information systems (IS). Traditionally, the management of digital archives include all aspects of archival science – as more traditionally understood through the life cycle model, as well as all aspects of the creation, preservation, use, and disposition of archives though IT. The concept of DA research is similarly broadly construed and also includes research on archival and recordkeeping topics being undertaken by researchers (Gilliland-Swetland et al. 2004). Many issues are blooming in this field such as authenticity, digital signature, migration, encapsulation, digital certification and social network. For better preservation and management of archives, agencies start to cooperate with IS vendors to meet the regulation of rules. These alliance activities have dependencies in the sense that the DA operations influence the others, and these dependencies need to be managed.

Since 1990, the development of archival research consciousness has unprecedented growth in the academy and in practice, as well as in scholarly awareness that the construct of the archives, and provides a rich locus for research and theorizing (Gilliland-Swetland et al. 2004). This tendency raises questions regarding what are important when their activities are not well known, what does IT consist of, and what is the focus of this management exercise. Research in DA management is unique insofar as it takes place within a multidisciplinary environment encompassing history, management, computer science, and archival science (Couture et al. 2005). Due to various historical differences in organising, documenting and managing information across cultural institutions, cross-domain resource discovery in the cultural heritage sector remains problematic (Chaudhry et al. 2005). One of the most critical problems for research archives is their definition, which is also part of the recognition of their relevance. Except from the small group of specialists involved in documenting, recording, managing and preserving the digital records and heritage of science, the reasons and the solutions for these increasing complexities and the effort to transform the traditional frameworks and tools into efficient and updated proposals have not been fully investigated (Guercio et al. 2015).

Although DA is becoming increasingly common in this information age, our understanding of their operation and management does not reflect their expanding role in organizations. The understanding regarding the intellectual structure of DA about technology, management and social behavior has been limited. The objective of this study is to investigate the trend and intellectual structure of DA among milieu factors of practice, concept, science and technology.

2 LITERATURE REVIEW

The literature on information technology, systems development, records management, and archives provides the theoretical basis for the DA management. Since the 1990s, digital archives have advanced tremendously – not only in terms of the scope of technology it encompasses, but also in the development of its knowledge (Gilliland-Swetland et al. 2004).

2.1 Related Research on the Management of Digital Archives

A number of prior studies have reflected on archives management as an academic discipline. Assessment of DA as a field, from both an historical and a forward-looking perspective, served as one important theme for archive management. Cox (1987) examined academic journals, monographs, yearbooks etc. related to archival studies, library and information science, and history from 1901 to 1987. Cox classified archival literature into 10 categories: arrangement and description; history, organization, and activities of repositories; management of current records; general literature; preservation, restoration, and storage; application of photographic processes; appraisal and disposition; training and professional development; special physical types of records and manuscripts; and historical editing and documentary publication. Cox and Samuels (1988) argued that typology of research fields in DA included: 1. Developing broader education and practical training in the DA management, 2. Employer needs in the DA managements, 3. Archival perspectives on the DA management, 4. DA management programs in organizations, 5. Archivists' attitudes towards technology, 6. Archival profession and technology. A certain amount of effort should be put into records to assure the authenticity, integrity and accessibility of the records and archives (MacNeil 2000), especially under the concept of records continuum. Archival quality is most closely associated through the preservation management of digital surrogates (Conway 2011).

Digital libraries are complex information systems and therefore demand formal foundations lest development efforts diverge and interoperability suffers. In this article, we propose the fundamental abstractions of streams, structures, spaces, scenarios, and societies, which allow us to define digital libraries rigorously and usefully. Streams are sequences of arbitrary items used to describe both static and dynamic (e.g., video) content. Structures can be viewed as labeled directed graphs, which impose organization. Spaces are sets with operations on those sets that obey certain constraints. Scenarios consist of sequences of events or actions that modify states of a computation in order to accomplish a functional requirement. Societies are sets of entities and activities and the relationships among them. Together these abstractions provide a formal foundation to define, relate, and unify concepts - among others, of digital objects, metadata, collections, and services - required to formalize and elucidate "digital libraries." The applicability, versatility, and unifying power of the model are demonstrated through its use in three distinct applications: building and interpretation of a digital library taxonomy, informal and formal analysis of case studies of digital libraries, and utilization as a formal basis for a description language of digital libraries.

Academic disciplines typically seek to articulate the intellectual structures upon which they can cultivate their futures (Hirschheim et al. 1996). Couture and Doucharme (2005) argued that typology of research fields in archival science included: 1. The object and aim of archival science, 2. Archives and society, 3. The history of archives and of archival science, 4. Archival functions, 5. The management of archival programs and services, 6. Technology, 7. Types of media and archives: electronic records, 8. Archival environments, 9. Specific issues related to archives. Kim and Lee (2008) collected articles of archival science from 2001 to 2004, generated pathfinder networks of 43 clusters and grouped them into seven subject categories: digital libraries and digital archiving

technologies, online resources and finding aids, archives and archivists, legal and political issues, electronic records and technical issues, records and information management, and e-mail and information professionals. Finally, these seven subject categories were merged into three sectors: digital library, archives and management. That study describes dynamic change in the 2001-2004 research themes from traditional single-subject areas to emerging, complex subject areas. Digital archives can be conceptualized as a package of standards. It builds on existing technical standards (e.g., with respect to operating systems, databases, and network standards). It embeds procedural and performance standards as well as numerous classification schemes and terminologies (Hanseth et al. 2006). Therefore, we can conclude that there are at least two-fold dimensions of DA management, i.e. technology-management and practice-concept perspective.

The field of DA has grown by relying on the intellectual structures provided in the classic document management works. This study reconsiders the basis on which the field de facto has been structured. In doing so, it is our hope that a conceptual foundation for future theories in the domain of DA may be possible.

2.2 Prior Research on Intellectual Structure and Bibliometric Method

To understand the issues covered by DA along with their possible solutions, governments may find specific and objective references from numerous research topics in academic literature. The way to locate research trends through the classification of a large number of academic articles is to comprehend the intellectual structure of this subject area and its evolvement over time. The discipline of bibliometrics provides tools for the study of intellectual structure of research subjects. The method of co-word analysis, a powerful tool of bibliometrics, is used in this study to recognize the intellectual structure of DA.

Bibliometric methods have been successfully applied to examine the intellectual structure of several disciplines (White et al. 1998). Bibliometrics is a research method commonly used in library and information science. It uses quantitative analysis and statistics to depict patterns of publication within a given field or body of literature. Researchers may use bibliometrics to determine the influence of an article, for example, to describe the relationship between the given article and the other articles. Co-word analysis reflects many papers that have cited any particular pair of terms and it is explained as a measure of similarity of content of the two terms. Co-word analysis has been accepted as a reasonable way to map the relationships among concepts, ideas, and problems (Callon et al. 1991). In co-word analysis, it is assumed that keywords extracted from papers could represent a specific research direction, research topic or subject of a field. If two keywords co-occur within one paper, the two research topics they represent are related. Higher co-word frequency means stronger correlation in keywords pairs, which can further suggest that two keywords are related to a specific research topic. Co-word analysis has the potential of effectively revealing patterns and trends in a specific discipline (Ding et al. 2001; Hu et al. 2013). Indexes based on the co-occurrence frequency of items, such as an inclusion index and a proximity index, are used to measure the strength of relationships between items. Based on these indexes, items are clustered into groups and displayed in network maps (He 1999). This study applied co-word analysis and cluster analysis to gain insights to the research paradigms of DA research field.

3 RESEARCH METHODOLOGY

The purpose of this study is to explore and map the trend and intellectual structure of DA studies. With bibliometric analysis, this study had five phases, each of which required different approaches to examine the evolution of the DA studies. In this study, Google Scholar (GS) is used as the database for DA articles. 100 research articles with DA highly cited were collected each year within 2004 and 2014. Words and terms used in the titles and abstracts of these literature articles were extracted and their numbers of frequency calculated. The co-occurrence of words and terms is used as the grouping measure. According to the literature of co-word analysis, such an approach can detect the distance among words, which is the similarity of themes, and hence create clusters among words. Further

analysis of each cluster of words can help locate the topic they represented, namely the intellectual structure of this study. After identifying the intellectual structure of DA, the correlation between the two clusters is further examined to find the similarities and differences and to results in the intellectual structure of DS in integration.

This study utilized the keyword search strategy which involved search for the keyword *Digital Archives* in GS databases to identify the potential DA research articles. The data used in this study includes journals, publication titles, article names, publication dates, and keywords. Besides the field of traditional DA management, some fields, especially the medical field, also focused on digitizing information and brought about some research literature. After the digitalization of medical matters, some issues need to be addressed urgently, such as the legal issues of digitized medical records (Fernández-Alemán 2013) and the security of information transmission (Vest 2010). To exclude medical-related literature, four high-frequency medical-related terms were located from the literature, i.e. 'health', 'medical', 'nursing' and 'treatment.' These four terms were excluded in the re-queries of articles in GS.

When conducting text analysis, few issues were encountered, which included the selection of language, computer codes, duplication of information, special characters, stemming, abbreviations and conflation, stop-words and compound words, etc. 100 DA articles each year were obtained from 2004 to 2014 annually with a total of 1,100 articles. Besides, since English is the major language for data analysis, the articles not presented in English are eliminated. Furthermore, the articles duplicated from different sources or without abstract are also eliminated. After this condensation step, 886 articles are left. Then the text contents of an article are extracted. Title and keywords reflect the key concept of an article. The title, abstract, and keywords are combined together and hereinafter referred to as an 'article' in this study.

This study extracts key terms to cluster the articles. The next step is to perform the data collection and analysis of terms in the data set. The most frequently used term in total time period were identified as the core terms in the field and further examined with co-word analysis. The co-word count for each pair of terms is retrieved through 2-items set of association rule method. The co-word count represents the similarity of each pair of terms. A series of operations are conducted, through which it is possible to identify the intellectual structure of DA studies (Pilkington et al. 2006).

Cluster analysis is also commonly used program to map the intellectual structure of studies and determine the common links between articles (Tu 2012). The last step of cluster analysis is performed to group these articles according to the similarity of their research themes and focuses. By taking the co-citation matrix and grouping the articles using cluster analysis of the correlations between the entries, this study can determine which articles are grouped together as well as their common shared elements. The closeness of article points on these maps is algorithmically related to their similarity as perceived by citers.

In this study, the academic literature for DA was collected through GS by means of co-word analysis and formed the article bank. The words used in each article were analyzed into units of terms. Terms with representative value were then located through the analysis of occurring frequency. Finally, the relationship between and among terms were located by the use of co-word analysis and cluster analysis to establish the intellectual structure of DA. The process of research steps in this study is illustrated in Figure 1.

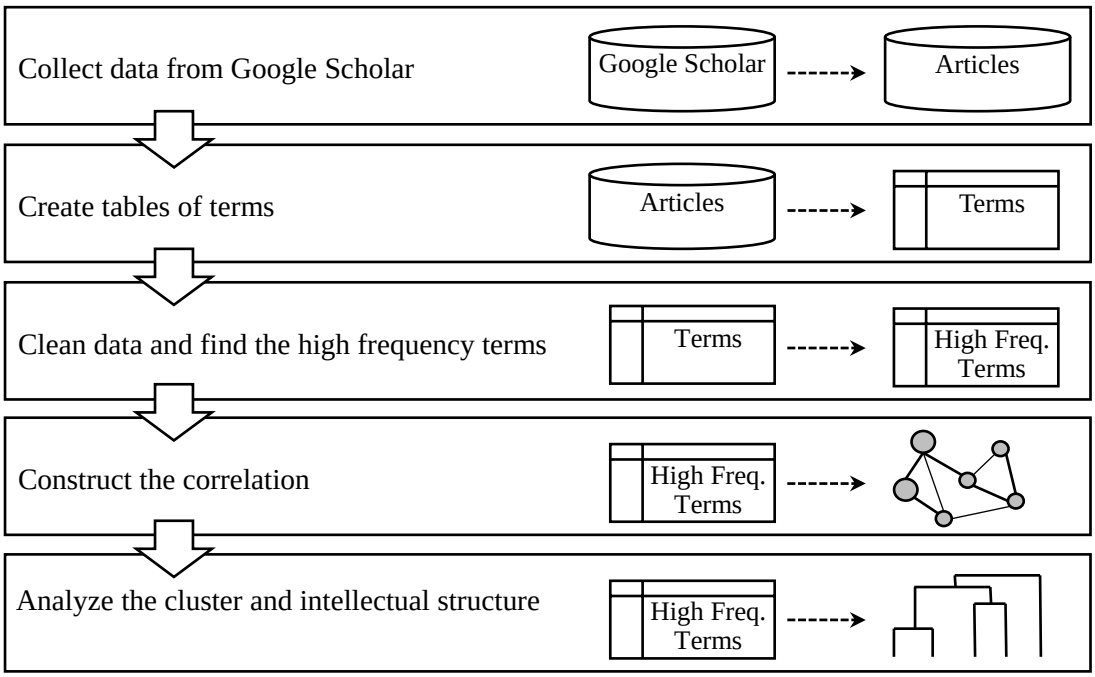

Figure 1. The process of this study

4 RESEACH RESULTS AND FINDINGS

4.1 Highly Frequent Terms

This study used Google Scholar to retrieve the research articles in DA area. The title, keywords and abstract of articles are used to count the frequency of noun terms. The noun terms in articles are isolated and their frequencies are accumulated. Besides the term frequency (TF), the term frequency–inverse document frequency (TFIDF) is also counted to reflect how important the term is to an article in the collection. Only the nouns with high TFIDF and TF frequency are left. After comparison, this study finds that these terms are also included in the glossary or thesaurus proposed by Australia, Canada, China, Taiwan or USA. The top 100 terms with high frequency in articles are described in table 1 and are used for further cluster analysis.

No	DA terms	amount	No	DA terms	amount
1	digital archive	640	51	electronic records management	28
2	data	330	52	xml	28
3	archive	314	53	long-term preservation	27
4	information	312	54	catalog	27
5	system	295	55	classification	26
6	record	237	56	national archives	25
7	management	177	57	dissemination	24
8	access	175	58	authenticity	23
9	electronic record	172	59	capture	23
10	content	160	60	copyright	23
11	application	156	61	ontology	23
12	document	155	62	transfer	23
13	technology	136	63	digital archives system	22
14	web	130	64	integrity	22
15	structure	120	65	open archival information system	21
16	preservation	110	66	digital archives program	21
17	standard	108	67	electronic records archives	21
18	program	106	68	archives management	21
19	network	72	69	digital object	21
20	file	70	70	information retrieval	21
21	context	67	71	protection	20
22	field	67	72	delivery	19
23	internet	65	73	digitalization	19
24	platform	64	74	digital data	17
25	iso	63	75	digital repository	17
26	software	62	76	download	17
27	storage	54	77	interoperability	17
28	electronic records management sys.	51	78	migration	17
29	repository	50	79	digital archives management	16
30	information technology	48	80	electronic document	16
31	archivist	47	81	semantic web	16
32	video	47	82	e-government	16
33	digitization	44	83	international standard	15
34	domain	42	84	arrangement	15
35	retrieval	41	85	e-learning	15
36	digital preservation	39	86	maintenance	15
37	digitize	37	87	visualization	15
38	evaluation	36	88	agency	14
39	description	35	89	hardware	14
40	national digital archives program	34	90	server	14
41	digital content	34	91	taiwan e-learning and digital archives	13
42	management system	34	92	acquisition	13
43	digital information	33	93	archives information	12
44	information system	32	94	digital file	12
45	records management	32	95	information management	12
46	component	32	96	transformation	12
47	index	31	97	cloud computing	11
48	strategy	31	98	digital document	11
49	code	30	99	digital technology	11
50	planning	29	100	appraisal	11

Table 1. Top 100 key terms in DA researches

All terms are converted to singular form. Since the key term *Digital Archives* may appear at any place of a research, it may not appear in title, abstract and keyword. Therefore, not all articles include the term *Digital Archives* in their extracted text. The term *digital archive* occurs with highest frequency. The following terms are *data*, *archive*, *information*, *system*, *record*, and *management*, etc. Furthermore, the amount of articles which include key terms is counted from 2004 to 2014 as depicted

in Table 2. The terms *data*, *information* and *system* keep the important position. The term *record* is increasingly important. The terms *archive* and *content* are less important than before.

Term	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
data	23	38	40	18	34	37	36	15	18	39	32
archive	21	33	34	28	40	42	38	15	19	42	2
information	32	28	27	23	30	39	41	22	25	27	18
system	21	35	32	14	26	25	25	29	24	30	34
record	11	10	13	33	8	11	14	20	34	22	61
management	8	9	9	29	18	14	23	24	18	17	8
access	9	22	20	13	25	16	20	10	13	16	11
electronic record	6	1	4	53	1	2	6	40	37	11	11
content	11	22	27	8	29	18	19	4	9	12	1
application	8	13	20	14	24	18	28	7	6	15	3
document	13	12	15	16	13	13	19	11	12	23	8
technology	7	16	15	9	24	17	15	5	4	13	11
web	12	15	13	4	20	18	23	2	5	16	2
structure	8	13	16	3	12	13	10	11	10	17	7
preservation	11	8	11	17	11	16	5	4	8	16	3
standard	16	10	9	8	7	12	9	8	14	10	5
program	11	12	9	7	16	9	12	3	13	8	6
network	2	5	8	5	7	13	13	3	3	8	5
file	6	10	11	4	4	3	9	8	2	10	3

Table 2. The amount of articles which include top 20 terms from 2004 to 2014

4.2 Co-word Analysis

In co-word analysis, this study used the concept of association rule method to compute the amount of articles which two terms appear simultaneously. The frequency of two terms which appear simultaneously in an article represents their closeness in co-word matrix (Zong, et al. 2013). Therefore, we compute the amount of articles that each two terms exist as the co-word (100 x 100) matrix. The relationship among top 20 frequent terms is depicted in Figure 2.

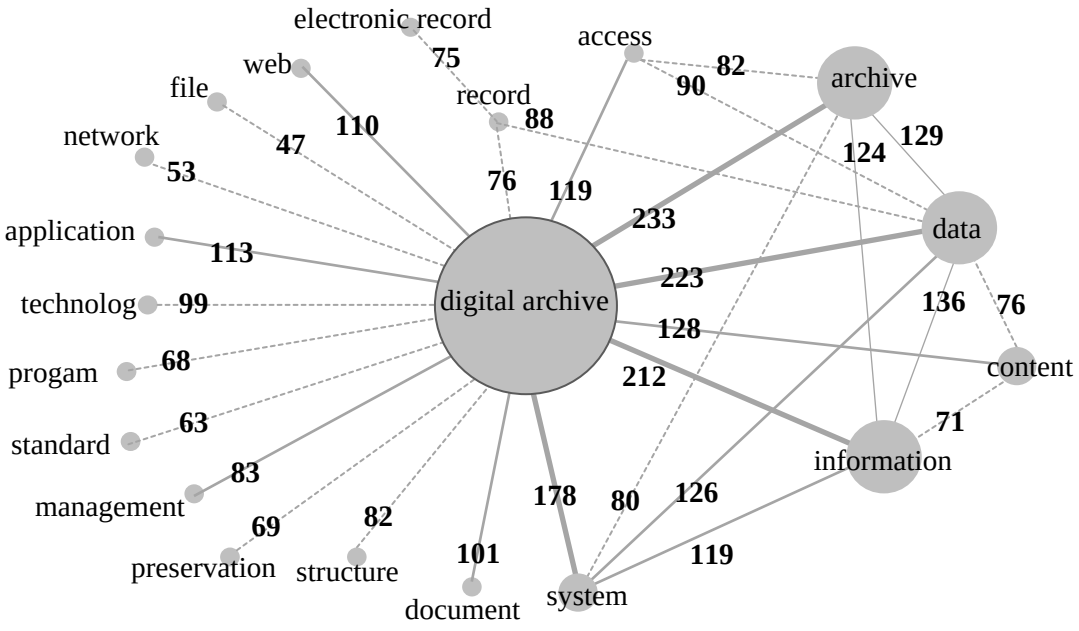

Figure 2. The relationship between two terms

The size of node in Figure 2 represents its TFIDF and the width of edge represents its amount of articles that two terms exist simultaneously. In Figure 2, there are close relationship among archive, data, information and system in digital archives researches. The co-word amounts of two terms are counted for co-word matrix of selected terms. This co-word matrix is then used as the correlation input for hierarchical cluster analysis with Ward's method to evaluate the relationship of terms. Using hierarchical clustering analysis, these 100 selected terms can be aggregated into seven clusters as depicted in Table 3.

Cluster 1 Archival Information	**digital archive**(640), **data**(330), **archive**(314), **information**(312), system(295), **record**(237), management(177), access(175), **electronic record**(172), content(160), application(156), document(155), technology(136), web(130), structure(120), preservation(110), standard(108), program(106), software(62)
Cluster 2 ERMS	network(72), file(70), context(67), field(67), internet(65), platform(64), iso(63), storage(54), **electronic records management system**(51), repository(50), digitization(44), domain(42), digital preservation(39), digitize(37), digital content(34), **information system**(32), catalog(27), digital data(17)
Cluster 3 Lone-term Preservation	information technology(48), archivist(47), retrieval(41), description(35), management system(34), digital information(33), component(32), records management(32), index(31), strategy(31), planning(29), **long-term preservation**(27), national archives(25), dissemination(24), authenticity(23), ontology(23), transfer(23), integrity(22), digital object(21), international standard(15), maintenance(15), digital file(12)
Cluster 4 Interoperability	video(47), code(30), **xml**(28), copyright(23), download(17), **interoperability**(17), server(14)
Cluster 5 Digital Archives Management	evaluation(36), classification(26), digital archives system(22), **archives management**(21), open archival information system(21), information retrieval(21), protection(20), digitalization(19), migration(17), **digital archives management**(16), hardware(14), archives information(12), **information management**(12), transformation(12)
Cluster 6 Digital Archives Projects	**national digital archives program**(34), capture(23), **digital archives program**(21), **electronic records archives**(21), delivery(19), digital repository(17), visualization(15), acquisition(13), Taiwan e-learning and digital archives(13), digital document(11)
Cluster 7 E-government and ERM	**electronic records management**(28), **e-government**(16), electronic document(16), semantic web(16), arrangement(15), e-learning(15), agency(14), appraisal(11), cloud computing(11), digital technology(11)

Table 3. Cluster of selected terms

Seven clusters of selected articles are described as follows:

1. Archival Information (AI)

Cluster 1 includes terms regarding.

Terms such as data, record, archive, information, content, access, web and application are included in cluster 1 regarding traditional research regarding archival information. It is the critical part of DA researches.

2. Electronic Records Management System (ERMS)

Terms such as information system, electronic records management system (ERMS), file, digitization, platform, storage and repository are included in cluster 2 regarding ERMS. A trusted digital repository (TDR) is a set of metrics that are used to certify that a given repository is an appropriate custodian of a collection of digital assets. More than an array of abstract measures, however, a TDR

represents a stable and sustainable organization, a set of policies and procedures for sound management of the digital objects, and a robust and secure technical platform (Johnston 2012). Digitization and repository are important to DA research.

3. Lone-term Preservation (LP)

Terms such as information technology, archivist, description, index, strategy, planning, long-term preservation, national archives, authenticity, ontology, integrity, maintenance, and international standard are included in cluster 3 regarding long-term preservation. Archival quality is closely associated through the preservation management of digital surrogates (Conway 2011).

4. Interoperability (IO)

Terms such as interoperability, xml, code, copyright, video, and server are included in cluster 4 regarding information interchange. Information technologies support the transmission and sharing of DA (Hu, Hsu, Hu and Chen 2010).

5. Digital Archives Management (DAM)

Terms such as digital archives management, archives management, information management, evaluation, classification, and migration are included in cluster 5 regarding digital archives management. The archiving of items is not a mere copying process of bits and bytes from object to another but rather a transformation of a digital object that is made to fit the requirements of provenance and authenticity. This transformational process goes beyond the traditional practices of collection, documentation, and preservation, leading not merely to a change of the context in which the object is embedded but to a change of the object itself (Kallinikos et al. 2013). Digital archives management is the central part of DA researches.

6. Digital Archives Projects (DAP)

Terms such as Electronic Records Archives (ERA) project, national digital archives program, and Taiwan e-learning and digital archives project are included in cluster 6 regarding digital archives projects. Many projects in different countries are issued for building practical DA.

7. Electronic Government and Electronic Records Management (EG&ERM)

Terms such as e-government, agency, electronic records management, e-learning, electronic document, arrangement, appraisal, and cloud computing are included in cluster 7 regarding the electronic records management in electronic government. DA applications used in institutions should be restructured providing an integrated and centralized digital recordkeeping system in order for controlling all records of the organization in all media and form (Külcü et al. 2010). The understanding of electronic records is its analysis of the attributes of a record based on concepts and principles that have evolved over centuries of detailed study of the documentary process (Duranti 2001).

Using multi-dimensional scaling (MDS) analysis, the terms of these seven clusters are scattered roughly as described in Figure 3 with technology-management-system and practice-concept dimensions. The two left clusters are cluster 4 and cluster 3. From table 3, the terms in cluster 4 “Interoperability (IO)” such as interoperability, xml and code are technical terms. The terms in cluster 3 “Long-term Preservation (LP)” such as information technology and authenticity are technical terms. Therefore the left part of horizontal axis is named *Technology.* The two middle clusters in Figure 3 are cluster 5 and cluster 7. The terms in cluster 5 “Digital Archives Management (DAM)” such as archives management, information management and digital archives management are management terms. The terms in cluster 7 “Electronic Government (EG) and Electronic Records Management” such as arrangement, appraisal, electronic government and electronic records management are management terms. Therefore the middle part of horizontal axis is named *Management.* The two right clusters in Figure 3 are cluster 2 and cluster 1. The terms in cluster 2 “Electronic Records Management System (ERMS)” such as information system and electronic records management system are terms regarding systems. Therefore the right part of horizontal axis is named *System.*

Furthermore, we named the horizontal axis as technology-management-system dimension. Besides, since cluster 6 is practical issue, we named the vertical axis as practice-concept dimension.

In Figure 3, from left to right, cluster 4 “Interoperability (IO)” is more focused on technology than management and system. Cluster 3 “Long-term Preservation (LP)” is more focused on technology than management and system. Cluster 6 “Digital Archives Projects (DAP)” which located at the left-up corner is focused on practice and technology domain. Cluster 7 “Electronic Government (EG) and Electronic Records Management” is more focused on management than technology and system. Cluster 5 “Digital Archives Management (DAM)” is more focused on practice-concept dimension than technology-science one. Cluster 1 “Archival Information (AI)” which located at the right-down corner is focused on concept and system domain. Cluster 2 “Electronic Records Management System (ERMS)” is the most system-focused cluster in DA researches. According to the dashed circles in Figure 3, we can find that the clusters which are closer to the central part are more management-focused than technology/system-focused. The clusters 5, 7 and 1 are more management-focused than others. It reflects that the digital archives management, archival information, and E-government and ERM are more management-focused that other clusters.

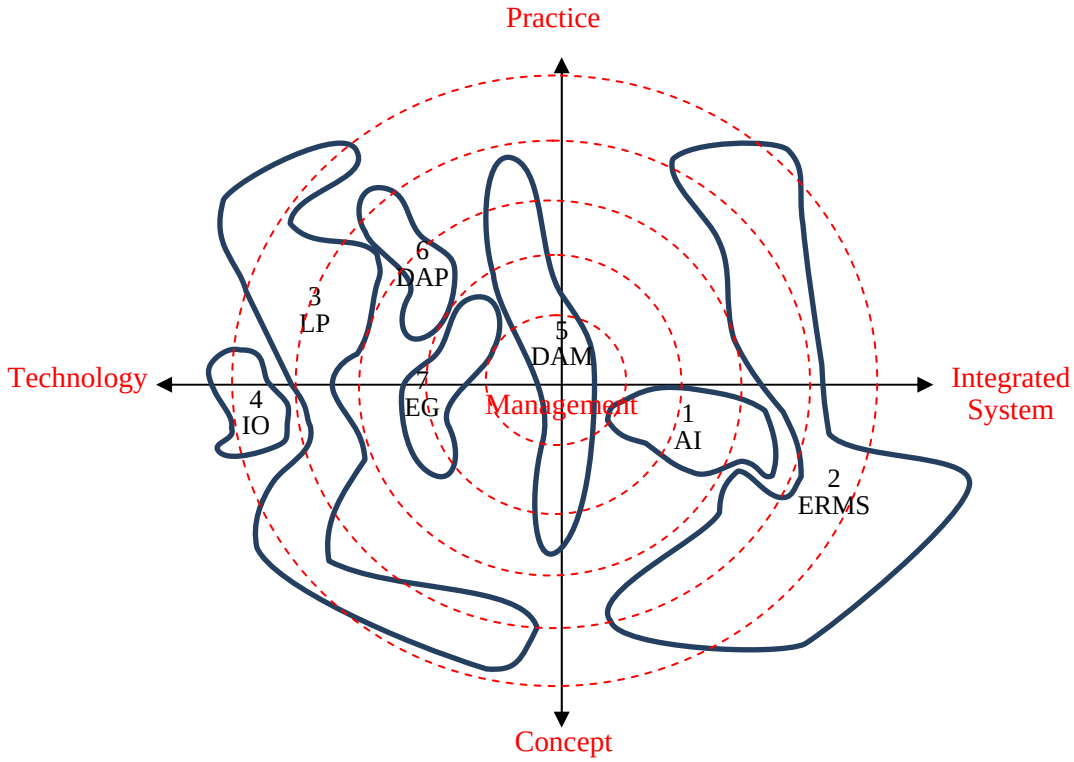

Figure 3. Scatter of clusters

5 CONCLUSIONS AND IMPLICATION

In this century, the volume of digital information is increasing at an extraordinary rate. Owing to the rapid developments of information technology, our intellectual capital of digital objects is increasingly at risk by the volatile character. The preservation of digital archives compromising its authenticity and long-term access are fundamental challenges. The archival bond is defined as “the interrelationships between a record and other records resulting from the same activity.” The International Council on Archives (ICA) Guide notes that “the rapid proliferation of text and data files” made inventorying and preserving digital archives difficult, and turned archivists’ attention to the question of developing policies and practices to ameliorate this decentralized and uncontrolled

situation. With the growth of networking and the development of paperless transactions, archivists have become increasingly concerned about the long-term preservation of digital archives. A number of researches have been done on the topic regarding DA.

Text mining techniques are used in this study. From the trend of key terms between 2004 and 2014, data, information and system play important roles in DA articles. Record is an increasingly important term also. From co-word analysis, we can find the close relationship among archive, data, information and system that these four terms often occur simultaneously in an article. It reflects that a system is needed for suitably organizing data in archive to render information.

Seven clusters are concluded as Archival Information, Electronic Records Management Systems, Lone-term Preservation, Interoperability, Digital Archives Management, Digital Archives Projects, and E-government and ERM. The corresponding mapping of clusters is classified into two dimensions, i.e. practice-concept and technology-management-system perspectives. The central research issue is the Digital Archives Management cluster. The E-government and ERM cluster is closer to the Digital Archives Management cluster than others. Therefore, the management issues are the focused points in DA researches. Although, there are still some researches focused on the technology/system field than on the management field. Further study could investigate the intellectual structure effect on the articles from other publication database, e.g. Web of Science, for more information.

References

Buchanan, S. and Cousins, F. (2012). Evaluating the Strategic Plans of Public Libraries: An Inspection-based Approach. Library & Information Science Research, 34, 125-130.

Callon, M., Courtial, J. P. and Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. Scientometrics, 22(1), 155-205.

Canazza, S. (2012). The digital curation of ethnic music audio archives: from preservation to restoration. International Journal on Digital Libraries, 12(2/3), 121-135.

Chandy, P. R. and Williams, T. G. (1994). The impact of journals and authors on international business research: A citation analyses. Journal of International Business Studies, 25, 715-728.

Chaudhry, Abdus Sattar and Jiun, T. P. (2005) Enhancing access to digital information resources on heritage: A case of development of a taxonomy at the Integrated Museum and Archives System in Singapore. Journal of Documentation, 61(6), 751-776.

Chen, Y.C. and Gant, J. (2001). Transforming local e-government services: the use of application service providers. Government Information Quarterly, 18, 343-355.

Conway, P. (2011). Archival quality and long-term preservation: A research framework for validating the usefulness of digital surrogates. Archival Science, 11(3), 293-309.

Couture, C. and Doucharme, D. (2005). Research in archival science: A status report. Archivaria, 59, 41-67.

Cox, R. J. (1987). American archival literature: Expanding horizons and continuing needs 1901-1987. American Archivist, 50(1), 306-23.

Cox, R. J. and Samuels, H. S. (1988). The archivist's first responsibility: A research agenda to improve the identification and retention of records of enduring value. American Archivist, 51, 28-51.

Cronin, B. (1984). The Citation Process: The Role and Significance of Citations in Scientific Communication, London: Taylor Graham.

Ding, Y., Chowdhury, G. and Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. Information Processing and Management, 37(6), 817-842.

Duranti, L. (2001). Concepts, principles, and methods for the management of electronic records. Information Society, 17(4), 271-293.

Fernández-Alemán, J. L., Señor, I. C., Lozoya, P. Á. O., and Toval, A. (2013). Security and privacy in electronic health records: A systematic literature review. Journal of Biomedical Informatics, 46(3), 541-562.

Ferro, Nicola and Silvello, Gianmaria. (2013). NESTOR: A formal model for digital archives. Information Processing & Management. 49(6), 1206-1240.

Gilliland, A. and McKemmish, S. (2004). Building an infrastructure for archival research. Archival Science, 4(3/4), 143-199.
Gonçalves, M. A., Fox, E. A., Watson, L. T., & Kipp, N. A. (2004). Streams, structures, spaces, scenarios, societies (5S): A formal model for digital libraries. ACM Transactions on Information Systems, 22(2), 270–312.
Guercio, M. and Carloni, C. (2015). The research archives in the digital environment: the Sapienza Digital Library project. Italian Journal of Library & Information Science, 6(1), 1-19.
Hanseth, O., Jacucci, E., Grisot, M. and Aanestad, M. (2006). Reflexive standardization: Side effects and complexity in standard making. MIS Quarterly, 30, 563-581.
He, Qin (1999) Knowledge Discovery through Co-Word Analysis. Library Trends, 48(1), 133-159.
Hirschheim, R., Klein, H. K. and Lyytinen, K. (1996). Exploring the intellectual structures of information systems development: A social action theoretic analysis. Accounting, Management and Information Technologies, 6(1-2), 1-64
Hu, Paul Jen-Hwa, Hsu, Fang-Ming, Hu, Han-Fen and Chen, Hsinchun, (2010). Agency satisfaction with electronic record management systems: A large-scale survey, Journal of the American Society for Information Science and Technology, 61(12), 2559-2574.
Hu, C. P., Hu, J. M., Deng, S. L. and Liu, Y. (2013). A co-word analysis of library and information science in China. Scientometrics, 97, 369-382.
Johnston, W. (2012). Digital preservation initiatives in Ontario: Trusted digital repositories and research data repositories. The Canadian Journal of Library and Information Practice and Research, 7(2), 1-8.
Kallinikos, J., Aaltonen, A. and Marton, A. (2013). The ambivalent ontology of digital artifacts. MIS Quarterly, 37(2), 357-370.
Kim, H. and Lee, J. Y. (2008). Exploring the emerging intellectual structure of archival studies using text mining: 2001-2004. Journal of Information Science, 34(3), 356-369.
Külcü, Ö. and Çakmak, T. (2010). Evaluation of the ERM Application in Turkey within the Framework of InterPARES Project. International Journal of Information Management, 30, 199-211.
MacNeil, H. (2000). Trusting Records: Legal, Historical and Diplomatic Perspectives. Kluwer Academic Publishers.
Pilkington, A. and Teichert, T. (2006). Management of technology: Themes, concepts and relationships. Technovation, 26(3), 288-299.
Stephens, D. O. (1998). Megatrends in records management. ARMA Records Management Quarterly, 32(1), 3-9.
Tu, P. P.N. (2012). Identifying the intellectual structure of organizational psychology studies: Influential contribution. Journal of Human Resource Management, 12(1), 75-112.
Vest, J. R., & Gamm, L. D. (2010). Health information exchange: persistent challenges and new strategies. Journal of the American Medical Informatics Association, 17(3), 288-294.
Walker, R. G. (2001). Reporting on service efforts and accomplishments on a ‘whole of government’ basis. Australian Accounting Review, 11 (3), 4-16.
White, H. D. and McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science. Journal of the American Society for Information Science, 49(4), 327-355.
Zong Qian-Jin, Shen Hong-Zhou, Yuan Qin-Jian, Hu Xiao-Wei, Hou Zhi-Ping, and Deng Shun-Guo, (2013). Doctoral dissertations of library and information science in China: A co-word analysis. Scientometrics, 94(2), 781-799.