

텍스트 마이닝을 이용한 국내 기록관리학 분야 지적구조 분석

Examining the Intellectual Structure of Records Management & Archival Science in Korea with Text Mining

이 재 윤(Jae-Yun Lee)*

문 주 영(Ju-Young Moon)**

김 희 정(Hee-Jung Kim)***

목 차

1. 서 론	3.3 군집단위 지적구조 분석
2. 관련 연구	3.4 문헌단위 지적구조 분석
3. 자료 수집 및 분석	3.5 시기별 분석
3.1 자료 수집 및 처리	4. 결 론
3.2 주제영역 파악	

초 록

이 연구에서는 텍스트 마이닝의 주요 기법인 문헌 클러스터링과 문헌 유사도 네트워크 분석을 적용하여 기록관리학 연구의 지적구조를 분석하였다. 대상 데이터는 2001년부터 2006년까지 발간된 국내 문헌정보학 영역의 대표적인 저널 5종에서 선정된 기록관리학 관련 논문 145건을 중심으로 분석하였다. 군집단위 지적구조 분석 결과, 국내에서 수행된 기록관리학 영역의 핵심적인 주제 영역은 <전자기록관리·디지털보존>, <기록관리정책·제도>, <기록물 기술/목록>, <기록관리학 영역/교육>이었으며, 문헌단위 지적구조 분석을 통하여서는 <디지털 아카이빙> 주제 영역이 중심을 이루고 있음을 확인할 수 있었다. 또한 시기별 분석을 통해서서는 <기록정보서비스> 영역이 새롭게 등장하고 있음이 드러났다.

ABSTRACT

In this study, the intellectual structure of Records Management & Archival Science in Korea was analyzed using document clustering, a widely used method of text mining, and document similarity network analysis. The data used in this study were 145 articles written on the subject of Records Management & Archival Science selected from five major representative journals in the field of Library & Information Science in Korea, published from 2001 to 2006. The results of cluster analysis show that the core subject areas are "electronic records management and digital preservation," "records management policy and institution," "records description and catalogues," and "records management domain and education." The results of document analysis, which is more detailed than cluster analysis, show that "digital archiving," a specialized subject in digital preservation, plays a central role. The results of serial analysis, which proceeds according to a timeline, show the emergence of "archival services" as a new subject area.

키워드: 텍스트 마이닝, 문헌 클러스터링, 네트워크 분석, 패스파인더 네트워크, 기록관리학, 지적구조
Text Mining, Document Clustering, Network Analysis, Pathfinder Network,
Records Management & Archival Science, Intellectual Structure

* 경기대학교 문헌정보학전공 조교수(memexlee@kgu.ac.kr)

** 연세대학교 대학원, 이화여자대학교 국제사무학과 강사(jymoonglory@hanmail.net)

*** 연세대학교 문헌정보학과 강사(joyheejung@gmail.com), 교신저자.

논문접수일자 2007년 2월 20일

게재확정일자 2007년 3월 12일

1. 서론

Menne-Haritz(2004)는 기록관리학을 ‘오래된 전통을 가진 새로운 학문(Archival science is a rather young discipline with old roots)’이라고 언급하고 있다. 즉, 기록관리학의 모태와 근간을 이루는 고문서학(diplomatics)의 기원과 역사는 오래되었지만, 디지털기록물의 급증으로 인한 관리연구의 필요성이 새롭게 제기됨에 따라 개척해나가야 할 신학문분야로 주목받기 시작한 것은 그리 오래 되지 않았음을 시사한 표현일 것이다. 이와 더불어 디지털 환경의 보편화와 학제성 및 지식의 융합 등의 키워드로 표현되는 최근 연구 환경의 변화는 복합학으로서의 기록관리학의 연구영역을 더욱 확대시키고 있다.

그러나 국내에서 기록관리 관련 학·협회가 출범한지도 6년여의 세월이 흐른 반면, 기록관리학 지적구조에 대한 연구는 그다지 많이 수행되어 오지 않은 가운데, 기록관리학 연구영역의 세부 주제영역 및 그동안의 연구 흐름에 대한 파악은 쉽지 않은 실정이다.

그동안 국내에서 수행된 기록관리학 지적구조에 관한 연구로서는 정연경(2003), 김희정(2005; 2006)의 연구가 있다. 정연경(2003)은 SAA의 교육지침서(2002)와 ACA의 Handbook for Archival Certification(2003)을 분석하여 기록관리학의 지식영역을 조명하였으며, 김희정(2005)은 2000년부터 2004년까지 발간된 국내와 미국의 기록관리 분야의 대표적인 학술지를 중심으로 한 국내외 논문 총 333편으로부터 선정된 핵심 저자들간의 동시인용빈도를 적용하여 다차원축척기법을 통하여 국내의 기록관리학 지

적구조를 비교분석하였다. 뒤이어 수행된 김희정(2006)의 연구에서는 1980년대부터 2000년도까지 수행된 국내외 기록관리학 지적구조 연구들을 시계열적으로 분석하여 기록관리학 연구영역의 변천과 확대과정을 제시하였다.

국외연구로서 Cox(1987)는 1901년부터 1987년까지의 기간동안 미국 내에서 수행된 기록관리학 연구 동향을 분석하고 평가하였는데, 이 연구를 위하여 Cox는 기록관리학, 문헌정보학, 역사학 등의 관련 학술지널과 단행본, 연감 등을 망라적으로 조사하고 분석함으로써, 각 시기별 기록관리학 연구영역의 주제 및 경향을 파악하였다.

Brichford(1988)는 *American Archivist* 저널을 대상으로 103개의 논문들을 선정해서 분석하여 기록관리학 영역의 대표 주제 클러스터를 여섯 개로 제시하였으며, Gilliland-Swetland(1992)는 기록관리학 분야의 여섯 개의 대표적인 학술지를 중심으로 136개의 논문들을 선정하여 인용분석기법을 적용하여 지적구조를 분석하였다. 뒤이어 Gilliland-Swetland(1995)는 1972년부터 1994년 기간동안에 발행된 기록관리학 분야 연구물들을 대상으로 인용분석을 수행하여 해당기간동안의 대표적인 저자군을 산출하였다. 또한 Cox(2000)는 기록관리학 분야의 대표적인 학술지 4종을 선정하여 전자기록관리(Electronic Records Management) 영역에 관한 인용분석을 수행하였다.

이상의 연구를 통해서 확인할 수 있듯이 지금까지 수행된 기록관리학 영역의 지적구조에 대한 연구는 문헌분석을 중심으로 수행되어 온 연구(김희정 2006; 정연경 2003; Brichford 1988; Cox 1987)와 인용분석을 중심으로 한 계량서

지학적 기법을 적용한 연구(김희정 2005; Cox 2000; Gilliland-Swetland 1992; 1995)로 나눌 수 있다.

본 연구에서는 텍스트 마이닝의 문헌 클러스터링 기법과 문헌 유사도 네트워크 분석을 적용하여, 국내에서 수행된 기록관리학 영역의 지적구조를 다각적이고도 면밀하게 분석해 보고자 한다. 데이터 범위는 2001년부터 2006년까지 발간된 국내 5종의 학술지를 대상으로 한 기록관리학 주제의 145편의 논문들을 중심으로 하였다. 또한 연구시기 중 2003년도까지를 1기, 2004년 이후를 2기로 설정하여 전반기와 후반기의 연구 경향의 추이도 함께 분석하였다.

2. 관련 연구

최근 텍스트 마이닝에 관한 이론서를 편찬한 Kao와 Poteet(2007)은 자유롭거나 비구조화된 텍스트로부터 흥미롭고 일상적이지 않은 지식을 발견, 추출하는 것이 텍스트 마이닝이라고 정의하였다. 이런 견해는 텍스트 마이닝을 데이터 마이닝(data mining)의 연장선 상에 있는 것으로 보고 텍스트를 대상으로 하는 데이터 마이닝이라고 보는 입장이다. 이에 따라 한때는 'textual data mining'이라는 표현도 사용되었다(Losiewicz, Oard, & Kostoff 2000). 2005년에 발간된 *ASIS&T Thesaurus of Information Science, Technology, and Librarianship*(3rd ed.)에서는 텍스트 마이닝을 지식 발견(knowledge discovery)의 하위어로 포함시키면서 스크리포트에서 "특정한 목적으로 정보를 추출하기 위해 자연어 텍스트로부터 패턴을 찾는 것"이

라고 정의하고 있다(Redmond-Neal & Hlava 2005).

텍스트를 대상으로 하기 때문에 텍스트 마이닝에서 사용되는 기법은 전통적으로 정보검색 분야에서 사용되어온 것이 많다. 2003년에 발간된 *Survey of Text Mining*에서는 주요 기법으로 자동분류(classification, clustering)와 정보검색 기법을 들고 있다(Berry 2003). 전통적으로 자동분류도 정보검색의 한 분야로 간주되어 왔음을 감안하면 사실상 기법 면에서는 텍스트 마이닝을 정보검색과 구분하는 것이 어려운 셈이다.

최근에는 정보검색 이외에 텍스트 마이닝의 주요 기법으로 네트워크 분석 기법을 포함하는 경향이 나타나고 있다. 2007년에 발간된 *Handbook of Text Mining*에서는 텍스트 마이닝의 주요 기법으로 자동분류와 정보추출, 링크 분석(네트워크 분석)을 들고 있다(Feldman & Sanger 2007). 물론 이런 주요 기법의 바탕이 되는 지식으로는 통계이론과 자연언어처리 이론을 언급하고 있다. 이와 같이 텍스트 마이닝은 정보검색의 기법을 계승하되 네트워크 분석과 같은 새로운 기법까지 폭넓게 수용하면서 발전해나가고 있다.

최근 들어 이와 같은 텍스트 마이닝이 활발히 논의되고 있는 것은 분석의 대상이 되는 텍스트가 급증하고 있으며 디지털 텍스트의 입수가 매우 용이하게 되었기 때문이다. 거의 자동화하여 일상적으로 수집되는 엄청난 텍스트 집합으로부터 가치가 있는 정보를 분석, 발굴하는 일이 새로운 과제가 된 것이다.

텍스트 마이닝이 적용되는 응용 분야와 대상 자료는 매우 광범위하다. 특히 비즈니스 분야

에 적용하는 연구(Miller 2004; Sullivan 2001; Zanasi 2005)와 생명과학 분야 문헌을 대상으로 하는 연구(Ananiadou & Mcnaught 2005; Chen et al. 2005)가 활발한 편이다. 특별히 과학기술 동향 분석을 위해서 사용하는 경우를 '과학기술 텍스트 마이닝(S&T text mining)'이라고 부르며(Kostoff 2003), 생명과학분야 논문에 적용하는 경우를 'biomedical textual mining'이라고 표현하기도 한다(Chen et al. 2005).

이 연구와 같이 계량정보학 분야에서 텍스트 마이닝을 적용하는 연구는 데이터베이스 단층 분석법(database tomography)이 대표적이다. R. N. Kostoff가 1990년대 초에 제안한 데이터베이스 단층분석법은 논문의 원문으로부터 명사구를 자동으로 추출한 다음 이를 이용하여 동시출현단어분석을 수행하는 기법이다(Kostoff 1993; Kostoff et al. 2001a). 물론 일반적인 동시출현단어분석(co-word analysis; Callon, Law, & Rip 1986)도 텍스트를 대상으로 하는 것이므로 텍스트 마이닝을 적용한 것이라고 보기도 한다. 그러나 동시출현단어분석을 위해서는 흔히 색인어 필드와 통제어휘를 이용하므로 비구조화된 텍스트를 이용한다는 텍스트 마이닝의 개념과는 다소 거리가 있다. R. N. Kostoff의 데이터베이스 단층분석법은 비구조화된 텍스트인 제목, 초록, 본문으로부터 명사구를 추출하여 분석 단위로 삼으므로 텍스트 마이닝의 적용이라고 볼 수 있는 것이다. R. N. Kostoff는 여러 분야에 대해서 데이터베이스 단층분석법을 적용하여 자동추출한 명사구의 빈도 분석과 클러스터링 분석을 수행한 결과를 연속해서 발표하였다(Kostoff, Eberhart, & Toothman 1998; Kostoff et al. 2000; Kostoff et al. 2001b;

Kostoff, Shlesinger, & Malpohl 2004).

최근에는 텍스트 마이닝의 핵심 기법인 문헌 클러스터링을 이용해서 지적 구조를 분석하려는 시도가 등장하고 있다. R. N. Kostoff는 동시출현 단어분석 이외에 문헌 클러스터링을 이용해서 주제영역의 계층 구조를 살펴보는 시도를 추가하였다(Kostoff et al. 2005). Glenisson 등(2005)도 문헌 클러스터링을 통해서 *Scientometrics*의 특집호에 실린 논문들을 시험적으로 분석해보았다. 이들은 초록을 이용한 것보다 본문을 이용한 문헌 클러스터링을 이용하는 것이 분석에 더 도움이 된다고 보고하였다.

이 연구에서는 이와 같이 텍스트 마이닝의 주요 기법인 문헌 클러스터링과 문헌 유사도 네트워크 분석을 병행하여 국내 문헌정보학계에서 진행되고 있는 기록관리학 연구의 지적 구조를 분석하고자 한다.

3. 자료 수집 및 분석

3.1 자료 수집 및 처리

문헌을 수집하기 위한 대상 학술지는 국내의 기록관리학 분야의 대표적인 학술지인 기록관리학회지를 기본으로 하고, 이외에 학술진흥재단 등재 및 등재후보 학술지인 한국문헌정보학회지, 정보관리학회지, 한국도서관정보학회지, 한국비블리아학회지로 제한하였다. 기록학회지는 초록이 포함되어 있지 않으므로 제외하였다. 결과적으로 수집된 문헌은 문헌정보학계에서 이루어진 기록관리학 분야 연구를 대표한다고 볼 수 있다.

기록관리학회지가 2001년부터 발간되었으므로 수집 대상 논문의 게재 시기는 2001년부터 현재(2006년 10월)까지를 범위로 하였다. 기록관리학회지의 수록 논문은 모두 포함하고, 여타 문헌정보학 분야 학술지 중에서 기록관리학 관련 논문을 파악하여 추가하였다.¹⁾ 수집 결과 전체 문헌은 145개였으며 그 구성은 <표 1>과 같다. 전체 목록은 부록으로 제시하였다. 시기별 분석을 위해서 대상 논문의 절반이 발간된 2003년까지를 1기로 하고 2004년 이후를 2기로 설정하였다.

문헌 145건의 제목과 초록을 한국어 형태소 분석기 HAM 5.0버전에 포함된 색인 프로그램을 사용하여 자동색인하였다. 색인 옵션은 복합명사 자체와 복합명사 구성 단어를 모두 포함하도록 하였고, 1음절 색인어는 제외하도록 설정하였다. 색인 결과 중에서 고빈도 일반 어휘와 숫자는 불용어로 간주하여 제외하였다. 고빈도 일반 어휘로는 문헌빈도가 30 이상인 단어 중에서 주제를 전혀 나타낼 수 없는 '연구', '대한', '관련', '목적', '제시', '중심', '필요', '결과'의 8개를 선정하였다. 문헌빈도 30 미만

의 단어 중에서도 주제성이 없는 단어가 상당수 포함되어 있지만, 분석 결과에는 큰 영향을 미치지 않으리라고 판단하여 모두 문헌간 유사도 산출에 사용하였다. 이 연구의 목적이 문헌 클러스터링을 이용한 지적 구조 분석인 만큼, 정보검색에서의 문헌 클러스터링과 마찬가지로 가급적 모든 색인어를 다 사용하는 방법을 채택한 것이다. 불용어 선정 과정의 최소화는 전체 분석과정의 자동화에도 유리하다.

색인 결과 최고빈도어는 문헌빈도 112인 '연구'였지만, 불용어가 아닌 단어로는 문헌빈도 98인 '기록'이 가장 빈도가 높았다. 불용어를 제외한 색인어 종수는 4,128종이었으며 문헌빈도 1인 색인어가 71.2%를 차지하였다. 한 문헌당 색인어 종수는 평균 67.6종이었다.

각 문헌을 벡터로 표현하기 위해서 색인어에 가중치를 부여하였다. 전체문헌이 N개일 때 문헌 d_x 에서 색인어 t_i 의 가중치 $w(t_i, d_x)$ 는 다음과 같은 로그 TFIDF 공식을 사용하였다.

$$w(t_i, d_x) = (1 + \ln(tf(t_i, d_x))) \times \log_2\left(\frac{N}{df(t_i)}\right)$$

<표 1> 분석 대상 문헌의 게재 학술지와 게재 시기

게재 학술지	게재 시기								
	'01	'02	'03	'04	'05	'06	1기('01-'03)	2기('04-'06)	합계
기록관리학회지	18	16	16	16	17	5	50	38	88
한국문헌정보학회지	4	1	6	5	6	1	11	12	23
정보관리학회지	2	0	4	2	7	3	6	12	18
한국비블리아학회지	2	0	0	3	4	0	2	7	9
한국도서관·정보학회지	1	1	1	3	0	1	3	4	7
합계	27	18	27	29	34	10	72	73	145

1) 학술지에 수록된 논문들 중 기록관리학 관련 논문 선정의 근거는 표제에 포함된 키워드를 중심으로 판단하였으며, 판단의 근거는 『기록관리학사전』(2005)을 중심으로 하였다.

· 각 단어의 가중치로 구성된 문헌-단어 행렬은 145 × 4,128의 크기를 가지게 되며 문헌간 유사도는 다음과 같이 두 문헌벡터 사이의 코사인 유사도로 산출하였다.

$$\cos(d_x, d_y) = \frac{\sum_{i=1}^n (w(t_i, d_x) \times w(t_i, d_y))}{\sqrt{\sum_{i=1}^n (w(t_i, d_x))^2 \times \sum_{i=1}^n (w(t_i, d_y))^2}}$$

이와 같은 방법으로 크기가 145×145인 문헌간 코사인 유사도 행렬이 <표 2>와 같이 산출되었다.

3.2 주제영역 파악

3.2.1 문헌 클러스터링

주제영역 파악을 위해서 문헌 클러스터링을 수행하였다. 문헌 클러스터링에는 문헌간 유사도 행렬을 이용하여 Ward 기법을 사용하였다.

통상적인 지적 구조 분석 연구에서는 특정 클러스터링 기법을 적용한 다음 적절한 군집수를 채택하는 것으로 클러스터링을 마친다. 그러나 이 연구에서는 클러스터링 결과의 품질을 높이고 지적 구조 해석을 돕기 위해서 2단계로 클러스터링을 수행하였다.

2단계 클러스터링을 제안한 이유는, 이 연구의 문헌 클러스터링은 일반적인 지적 구조 분석과 달리 분석 대상 단위의 수가 100개 이상으로 상당히 많아 통상적인 계층적 응집식 클러스터링 기법으로는 클러스터링 결과의 품질 저하가 우려되었기 때문이다. 개별 문헌으로부터 군집을 뭉쳐나가는 계층적 응집식 클러스터링 기법은 대상 문헌집단을 소수의 큰 군집으로 묶을 때에는 불리하다고 알려져 있으며 응집식 기법은 상향식이므로 군집의 크기가 작은 소분류 상황이 대부분 상황보다 판단의 횟수가 적어서 오류가 적게 누적되기 때문이다(이재윤, 정

<표 2> 문헌간 코사인 유사도 행렬(일부)

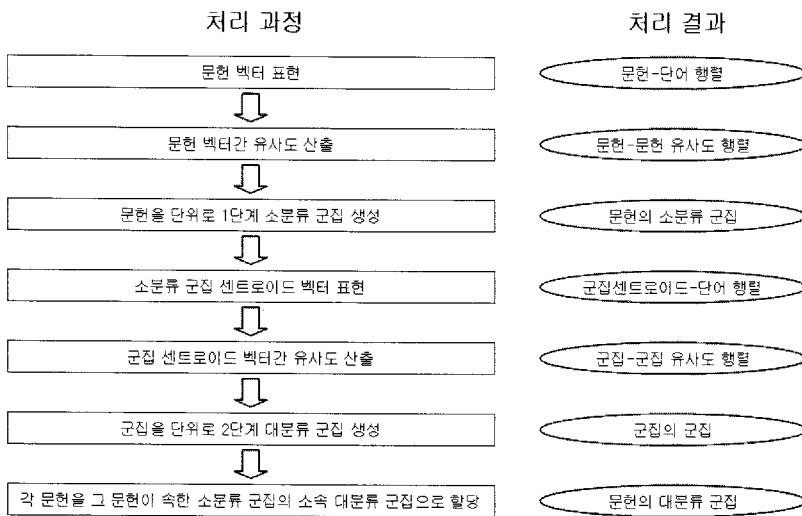
DID	D001	D002	D003	D004	D005	D006	D007	D008	D009	D010	D011	D012	D013	D014	D015	D016	D017	D018	D019	D020
D001	1.000	0.103	0.073	0.023	0.106	0.023	0.001	0.001	0.089	0.019	0.017	0.031	0.006	0.036	0.068	0.009	0.009	0.039	0.018	0.054
D002	0.103	1.000	0.067	0.060	0.038	0.012	0.009	0.000	0.089	0.009	0.001	0.025	0.032	0.023	0.019	0.000	0.022	0.008	0.029	0.031
D003	0.073	0.067	1.000	0.028	0.022	0.009	0.010	0.013	0.076	0.023	0.011	0.030	0.034	0.066	0.019	0.027	0.012	0.015	0.041	0.060
D004	0.023	0.060	0.028	1.000	0.025	0.019	0.003	0.012	0.037	0.007	0.044	0.057	0.002	0.012	0.029	0.007	0.001	0.002	0.036	0.019
D005	0.106	0.038	0.022	0.025	1.000	0.030	0.011	0.015	0.038	0.031	0.001	0.082	0.000	0.094	0.082	0.000	0.010	0.005	0.025	0.023
D006	0.023	0.012	0.009	0.019	0.030	1.000	0.033	0.026	0.017	0.020	0.040	0.067	0.012	0.014	0.028	0.025	0.033	0.024	0.033	0.015
D007	0.001	0.009	0.010	0.003	0.011	0.033	1.000	0.075	0.015	0.039	0.042	0.022	0.019	0.019	0.008	0.009	0.033	0.029	0.061	0.021
D008	0.001	0.000	0.013	0.012	0.015	0.026	0.075	1.000	0.028	0.010	0.031	0.025	0.007	0.035	0.012	0.000	0.033	0.026	0.092	0.020
D009	0.089	0.089	0.076	0.037	0.038	0.017	0.015	0.028	1.000	0.027	0.042	0.016	0.009	0.029	0.033	0.010	0.036	0.011	0.085	0.073
D010	0.019	0.009	0.023	0.007	0.031	0.020	0.039	0.010	0.027	1.000	0.018	0.024	0.020	0.029	0.035	0.022	0.014	0.037	0.024	0.017
D011	0.017	0.001	0.011	0.044	0.001	0.040	0.042	0.031	0.042	0.018	1.000	0.063	0.006	0.034	0.027	0.040	0.002	0.048	0.050	0.067
D012	0.031	0.025	0.030	0.057	0.082	0.067	0.022	0.025	0.016	0.024	0.063	1.000	0.018	0.052	0.038	0.000	0.013	0.011	0.020	0.031
D013	0.006	0.032	0.034	0.002	0.000	0.012	0.019	0.007	0.009	0.020	0.006	0.018	1.000	0.032	0.009	0.005	0.015	0.010	0.030	0.015
D014	0.036	0.023	0.066	0.012	0.094	0.014	0.019	0.035	0.029	0.029	0.034	0.052	0.032	1.000	0.101	0.000	0.008	0.004	0.029	0.011
D015	0.068	0.019	0.019	0.029	0.082	0.028	0.008	0.012	0.033	0.035	0.027	0.038	0.009	0.101	1.000	0.025	0.010	0.018	0.020	0.038
D016	0.009	0.000	0.027	0.007	0.000	0.025	0.009	0.000	0.010	0.022	0.040	0.000	0.005	0.000	0.025	1.000	0.035	0.019	0.037	0.018
D017	0.009	0.022	0.012	0.001	0.010	0.033	0.033	0.033	0.036	0.014	0.002	0.013	0.015	0.008	0.010	0.035	1.000	0.020	0.033	0.016
D018	0.039	0.008	0.015	0.002	0.005	0.024	0.029	0.026	0.011	0.037	0.048	0.011	0.010	0.004	0.018	0.019	0.020	1.000	0.052	0.090
D019	0.018	0.029	0.041	0.036	0.025	0.033	0.061	0.092	0.085	0.024	0.050	0.020	0.030	0.029	0.020	0.037	0.033	0.052	1.000	0.025
D020	0.054	0.031	0.060	0.019	0.023	0.015	0.021	0.020	0.073	0.017	0.067	0.031	0.015	0.011	0.038	0.018	0.016	0.090	0.025	1.000

진아 2005). 통상적인 저자동시인용 분석에서는 수십 명의 저자를 대여섯 개의 군집으로 묶기 때문에 한 군집 당 저자의 수가 7~8명 정도에 그친다. 이러한 상대적 소분류 상황에서는 응집식 기법의 단점이 그다지 문제가 되지 않는다. 그러나 이 연구에서처럼 100개 이상의 문헌을 대상으로 하는 경우에는 군집 병합이 여러 차례에 걸쳐서 이루어지는 대분류 상황이 되어 응집식 기법의 단점이 문제가 될 가능성이 있다. 이는 분석 대상의 수가 많아질수록 심각하게 고려해야할 부분이다.

이런 점을 감안하여 본 연구에서는 <그림 1>과 같은 2단계 문헌 클러스터링을 수행하였다. 1단계로 대상 문헌집단으로부터 다수의 작은 군집(상대적으로 소분류 군집이라고 할 수 있음)을 생성한 다음, 생성된 소분류 군집을 각각 센트로이드 벡터로 압축 표현하고 이에 대해서 다시 클러스터링 기법을 적용하여 대분류 군집을 생성하였다. 1단계와 2단계 군집에서는 모

두 계층적 응집식 클러스터링 기법 중에서 군집의 크기가 비교적 고르게 생성되는 Ward 기법을 적용하였다.

145개 문헌을 대상으로 Ward기법으로 1단계 클러스터링을 수행한 결과 각 군집의 소속 문헌이 나타내는 주제가 비교적 균일하게 유지되는 소분류 수준을 21개 군집이 생성되었을 때로 판단하였다. 군집의 수를 이보다 적게 줄이기 시작하면 군집 내 문헌의 주제가 동질적이지 못한 경우가 눈에 띄기 시작하였기 때문이다. 21개 군집별로 소속 문헌의 색인어 가중치를 평균하여 군집 센트로이드 벡터를 만든 다음, 21개 군집 센트로이드를 대상으로 다시 Ward 기법으로 2단계 클러스터링을 수행하였다. 2단계 클러스터링 결과에서 군집 핵심주제의 동질성이 유지되는 수준을 찾아서 12개 군집을 최종으로 확정하였다. 12개 군집의 주제 분야 명칭은 연구자 3인의 의견을 종합하고 『기록관리학사전』(2005)을 참고하여 결정하였다.



<그림 1> 2단계 문헌 클러스터링 과정

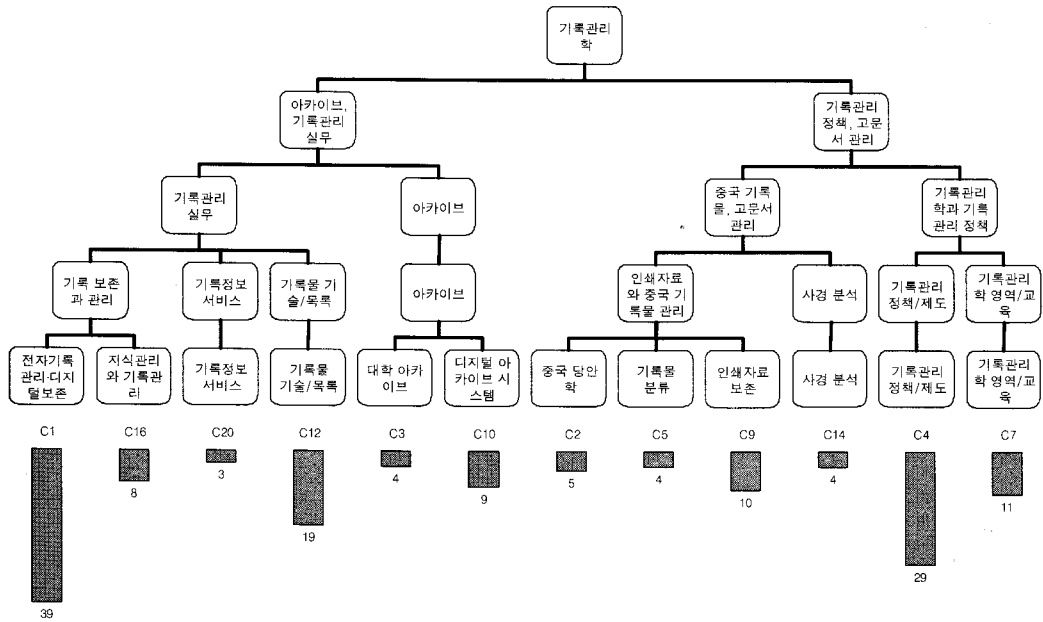
1단계와 2단계 클러스터링 결과는 <표 3>과 같다. 이 표를 보면 1단계 클러스터링만으로 군집의 수를 12개로 줄인 경우보다 1단계 21개 군집을 거쳐서 2단계에서 12개로 합친 결과가 분석에 더 도움이 된다는 것을 알 수 있다. 예를 들어 1단계 클러스터링에서는 21개보다 더 클러스터링을 진행하여 12개 군집으로 줄여도 '기록관리 정책' 군집과 '기록관리 제도' 군집은 동일 군집으로 모이지 않으나, 2단계 클러스터링에서는 이들이 하나로 합쳐진다. 또한 기록관리학 교육에 관련된 군집들(C7, C21)도 1단계 클러스터링에서는 12개 군집으로 줄여도 분리되어 있지만, 2단계 클러스터링에서는 하나로 합쳐지는 것으로 나타난다. 반대로 1단계 클

러스터링의 12군집 결과에서는 '기록정보서비스' 군집과 '기록물 정리/기술' 군집이 하나로 병합되어 동질적이지 않은 군집을 형성하는데 반해서 2단계 클러스터링에서는 '기록정보서비스' 군집은 독립 군집을 유지하고 '기록물 정리/기술' 군집은 유사한 주제인 '기록물 목록' 군집, 그리고 '기록물 메타데이터' 군집과 병합되어 더 합리적인 결과를 나타내었다. 따라서 이 분석 결과를 기초로 비교적 많은 수를 대상으로 지적 구조 분석을 수행하는데 있어서 2단계 클러스터링이 기존 방식보다 더 유리하다고 말할 수 있다.

12군집을 계속 결합해나가면 <그림 2>와 같이 주제 계층을 형성할 수 있다. 이 주제 계층에

<표 3> 1단계와 2단계 클러스터링 결과

1단계 클러스터링 결과			2단계 클러스터링 결과 - 12군집	
군집의 핵심 주제	21군집번호	12군집번호	12군집번호	군집의 핵심 주제
디지털 보존	C1	C1	C1	전자기록관리·디지털보존
보존 메타데이터	C8	C7		
전자기록물 보존	C6	C5		
전자문서 관리	C17	C5		
지식관리와 기록관리	C16	C10	C11	지식관리와 기록관리
기록정보서비스	C20	C3	C12	기록정보서비스
기록물 정리/기술	C12	C3	C9	기록물 기술/목록
기록물 목록	C18	C11		
기록물 메타데이터	C19	C7		
대학 아카이브	C3	C2	C3	대학 아카이브
디지털 아카이브	C10	C2	C8	디지털 아카이브 시스템
OAIS, 전자저널	C15	C5		
중국 당안학	C2	C2	C2	중국 당안학
기록물 분류	C5	C4	C5	기록물 분류
인쇄자료 보존	C9	C1	C7	인쇄자료 보존
사경 분석	C14	C9	C10	사경 분석
기록관리 정책	C4	C3	C4	기록관리 정책/제도
기록관리 제도	C13	C8		
국의 기록관리 제도	C11	C8		
기록관리학, 기록관리 교육	C7	C6	C6	기록관리학 영역/교육
기록관리 대학원 교육	C21	C12		



〈그림 2〉 군집분석에 의한 국내 기록관리학 연구 영역의 계층적 분류와 소속 문헌 수

서 2군집 수준과 8군집 수준, 12군집 수준은 덴드로그램의 2군집, 8군집, 12군집 수준을 그대로 제시하였으며, 주제 계층의 4군집 수준은 계층적 분류의 균형이 맞도록 덴드로그램 상의 구분 위치를 왼쪽과 오른쪽을 달리하여 제시하였다. 실제 덴드로그램에서는 왼쪽의 '기록관리 실무' 군집과 '아카이브' 군집은 5계층 수준에서 분리되며, 오른쪽의 '중국 기록물, 고문서 관리' 군집과 '기록관리학과 기록관리 정책' 군집은 3계층 수준에서 분리된다. 이 그림을 통하여 국내 기록관리학 연구 영역의 계층적 분류를 살펴 볼 수 있다.

〈그림 2〉에서 나타난 기록관리학 영역의 계층적 분류를 보면 크게 두 개의 주제 영역으로 구분할 수 있다. 두 영역에 실제 포함되는 문헌들을 살펴보면, 우측의 〈기록관리정책, 고문서 관리〉를 중심으로 한 주제영역은 1기에 해당되

는 2001년도부터 2003년도에 주로 수행된 문헌들로서 기록관리학회지를 중심으로 발표된 연구들을 확인할 수 있었다. 반면, 〈그림 2〉 좌측의 〈아카이브, 기록관리실무〉를 중심으로 한 주제영역은 2기에 해당되는 2004년도 이후에 수행된 문헌들로서 기록관리학회지 외의 문헌정보학 영역의 일반 학술지를 중심으로 발표된 연구들이었다.

또한, 계층적 분류표 하단에 표시된 문헌 수 막대의 크기를 기준으로 살펴보면, 우측의 1기에 수행된 대표적인 기록관리학 영역의 연구주체는 〈기록관리정책 및 제도〉임을 확인할 수 있고, 좌측의 2기에 수행된 대표적인 연구주체는 〈전자기록관리, 디지털보존〉임을 확인할 수 있다.

실제로 2000년도부터 2004년도까지 동일기간 동안 수행되었던 미국과 국내의 기록관리학

연구영역에 대한 연구(김희정 2005)에서는 국내의 주력 연구영역이 <기록관리정책 및 제도> 영역으로 나타났고, 미국의 주력 연구영역은 <전자기록관리 및 디지털보존>으로 나타난 바 있다. 이를 감안하면 국내의 연구동향이 기록관리학 연구 역사가 오래된 미국의 연구 동향과 점차 유사해지는 추세라고 볼 수 있다. 이는 국내 기록관리학 연구가 정책 연구라는 도입기의 특수성을 점차 벗어나서 정착기로 접어들고 있음을 시사하는 결과이다.

3.2.2 군집별 대표 문헌과 대표 단어 선정

각 군집의 주제 해석에 객관성을 높이기 위하여 군집의 대표 문헌과 대표 단어를 자동으로 파악하는 방법을 모색하였다. 이를 통해서 다수의 문헌을 클러스터링하여 지적 구조를 분석하는 경우, 필요한 수작업 분석량을 상당히 줄일 수 있다. 특히 분석 영역에 대해서 익숙하지 않은 상황에서는 대표 문헌과 대표 단어가 각 군집의 주제를 수작업 분석에 가깝게 설명해줄 수 있다면 주제 파악과 해석에 큰 도움이 될 것이다.

군집별 대표 문헌을 파악하기 위해서 중심성 분석을 수행해보았다. 중심성은 한 개체가 네트워크 상에서 구조적으로 중심에 위치하는 정도를 나타내는 성질이며, 이를 측정하는 지표를 중심성 척도라고 부른다. 여러 가지 중심성 척도가 존재하지만 이 연구에서는 계량서지적 네트워크에 적합하도록 개발된 삼각매개중심성 척도(이재운 2006)를 사용하였다.

삼각매개중심성(triangle betweenness centrality; C_{TB} 로 표기)은 한 노드가 다른 노드들 사이를 결속시켜주는 능력을 측정하는 척도로

서 측정은 노드 셋으로 구성된 삼각형을 기본 분석 단위로 한다. 분석 기준이 되는 노드를 a 라고 할 때 a 를 제외한 다른 노드들의 조합쌍 중에서 다음 조건을 만족하는 노드 쌍(b, c)의 비율로 측정한다(이재운 2006).

$$sim(b, c) < MIN(sim(a, b), sim(a, c))$$

삼각매개중심성은 소수와 집중된 관계를 맺는 노드보다는, 강도가 약하더라도 폭넓은 관계를 가진 노드에게 유리한 척도이다(이재운 2006). 따라서 한 문헌 군집에 속한 문헌들을 대상으로 측정하면 그 문헌 군집에 속한 문헌들을 고르게 반영하는 대표 문헌을 파악할 수 있다. 이와 같이 중심성 측정의 범위를 군집 내로 제한한 것을 군집내 삼각매개중심성(C_{TB-C} 로 표기)이라고 부른다. 12개 군집별로 군집내 삼각매개중심성을 측정하여 가장 중심성이 높게 나타난 문헌을 제시하면 <표 4>와 같다.

군집별 대표 단어 선정을 위해서는 두 가지 방법을 검토하였다. 주요 단어를 선정하는 기존의 방법으로는 K-L 다이버전스(Kullback-Leibler Divergence; 이하 KLD로 약칭)를 들 수 있다. KLD는 확률분포의 차이를 측정하는 방법으로서, 정보검색 분야의 적합성 피드백 연구에서 1차 검색 후 적합 문헌으로부터 초기 질의에 추가할 중요한 단어를 선정하는 기준으로 사용된 바가 있다(Carpineto et al. 2001). 동일한 원리를 이용하여 적합 문헌 집단이 아닌 군집에서 중요한 단어를 선정하는데 KLD를 이용할 수 있다.

군집 c 내 단어 t_i 의 출현확률 $p(t_i|c)$ 와 전체 출현확률 $p(t_i)$ 사이의 차이 $KLD(t_i, c)$ 는 다음

〈표 4〉 군집내 삼각매개중심성(CTB-C) 기준으로 선정한 군집별 대표 문헌

군집		군집 대표 문헌	
번호	주제	번호	제목
C1	전자기록관리·디지털 보존	72	전자기록물의 장기보존을 위한 방안 연구 : 개념을 중심으로
C2	중국 당안학	117	중국의 도시건설 기록물 관리사업에 대한 고찰
C3	대학 아카이브	108	대학도서관과 아카이브즈에 대한 비교 연구
C4	기록관리 정책/제도	60	한국의 기록관리 현황 및 발전방향에 관한 연구
C5	기록물 분류	132	중국의 기록물 정리분류의 원칙과 사례 분석
C6	기록관리학 영역/교육	39	미국의 기록관리학 지식 범주에 관한 연구
C7	인쇄자료 보존	13	조선시대 고문서에 사용된 종이 분석
C8	디지털 아카이브 시스템	118	디지털 아카이브의 영구적 보존을 위한 개념적 모형 설계에 관한 연구
C9	기록물 기술/목록	19	국가기록원의 기록물 정리·기술의 현황과 개선방안
C10	사경 분석	68	조선초 익안대군발원사경에 관한 연구
C11	지식관리와 기록관리	128	지식경영 관점에서 본 기록관리와 지식경영의 연관 관계
C12	기록정보서비스	88	기록정보서비스 활성화를 위한 전자적 질의/응답 기능 개선방안

과 같이 산출한다. 이때 출현확률은 대상 문헌 중에서 해당 단어가 출현한 문헌의 비율로 산출한다. 즉 전체 출현확률은 단어의 문헌빈도를 전체 문헌 수로 나눈 값이고, 군집내 출현확률은 군집 소속 문헌 중에서 단어가 출현한 문헌의 수를 군집 소속 문헌의 수로 나눈 값이다.

$$KLD(t_i, c_j) = p(t_i | c_j) \log \frac{p(t_i | c_j)}{p(t_i)}$$

그런데 KLD는 문헌빈도를 이용하여 산출하기 때문에 한 문헌 안에서의 출현빈도는 고려하지 않는다. 이를 감안하여 이 연구에서는 문헌 안에서 단어의 출현빈도가 높고 낮음을 고려하는 새로운 기준을 다음과 같이 고안해보았다.

제안하는 방식은 특정 단어가 군집 내에서는 문헌별 평균 빈도가 높고, 전체적으로는 문헌별 평균 빈도가 낮을수록 중요한 단어라는 판단에 근거한 것이다. 군집 내에서의 평균 빈도와 전체에서의 평균 빈도 사이의 차이를 구하는 새로운 기준을 평균 빈도차 DAF(Difference in Average Frequency)라고 부르기로 하며 공식

으로 표현하면 다음과 같다.

$$DAF(t_i, c_j) = AF(t_i, c_j) - \frac{1}{C} \sum_{k=1}^C AF(t_i, c_k)$$

여기서 $AF(t_i, c_j)$ 는 군집 c_j 에서 단어 t_i 의 문헌별 빈도의 평균이고 C 는 군집의 수이다.

DAF를 기준으로 군집별 상위 10위 이내 단어를 제시하면 〈표 5〉와 같다. 10위 이내 중에서 출현한 문헌이 한 개 뿐인 단어는 제거하였으므로 대표 단어가 10개보다 적은 경우도 있다. 군집 C3이나 군집 C12는 소속 문헌 수가 적기 때문에 군집내 문헌빈도가 1에 불과해도 상위에 자리잡은 단어가 많아서 이를 제외하면 소수 단어만 포함되었다. 군집별 상위 10위 이내 단어 중에서 복합명사 분해로 인해 생성된 부분 단어(예: '태타')를 제거하고 의미가 겹치는 단어는 하나만 남기며 주제성이 적은 단어는 제외하여 군집별 최대 네 개까지의 단어만 남기면 주제를 적절히 표현하는 단어를 얻을 수 있다. KLD를 기준으로 산출한 결과와 DAF를 기준으로 산출한 결과는 군집별로 상당히

〈표 5〉 DAF 기준으로 선정한 군집별 대표 단어

번호	군집 주제	군집 대표 단어	
		상위 10위 이내 대표 단어	선정된 대표 단어
C1	전자기록관리·디지털 보존	메타, 디지털, 전자, 보존, 메타데이터, 전자문서, 진본, 장기, 체계, 자원	디지털, 보존, 메타데이터, 전자문서
C2	중국 당안학	당안, 건설, 중국, 당안관, 도시, 응용, 사업, 당안관리	당안, 중국, 건설
C3	대학 아카이브	아카이브즈, 이브즈, 아카, 대학	아카이브즈, 대학
C4	기록관리 정책/제도	기록관리, 중앙, 제도,물관리, 정부기록보존소, 국가, 공공, 관리법, 대통령, 공개	기록관리, 제도, 국가, 공공
C5	기록물 분류	분류, 원칙, 정리분류, 조선총독부, 총독부, 검색분류, 기능분류체계, 체계, 기능분류, 공문서	분류, 원칙, 체계
C6	기록관리학 영역/교육	기록관리학, 관리학, 교육, 과목, 과정, 프로그램, 실습, 교육과정, 전문가, 대학원	기록관리학, 교육
C7	인쇄자료 보존	조선, 종이, 조선시대, 섬유, 미생물, 시대, 사용, 보존성, 종류	종이, 섬유, 미생물, 보존성
C8	디지털 아카이브 시스템	아카이브, 이브, 아카, 저널, 전자저널, 참조모, 참조모형, OAIS, 모형, 원문	아카이브, 전자저널, 참조모형, OAIS
C9	기록물 기술/목록	기술, 규칙, EAD, ISAD, 기술규칙, 데이터, 전자, 기준표, 불기술, 기술요소	기술규칙, EAD, ISAD, 전자
C10	사경 분석	사경, 발원, 표지화, 발원문, 연화, 묘법, 표지, 양식, 고려사, 고려	사경, 발원문
C11	지식관리와 기록관리	지식, 경영, 지식경영, 기업, 민간, 지식공유, 과학	지식경영, 기업, 민간
C12	기록정보서비스	기록정보서비스, 이용자, 서비스, 요구	기록정보서비스, 이용자

유사하였으므로 여기에는 KLD 기준에 의한 결과만 제시하였다.

3.3 군집단위 지적구조 분석

3.3.1 다차원척도 분석

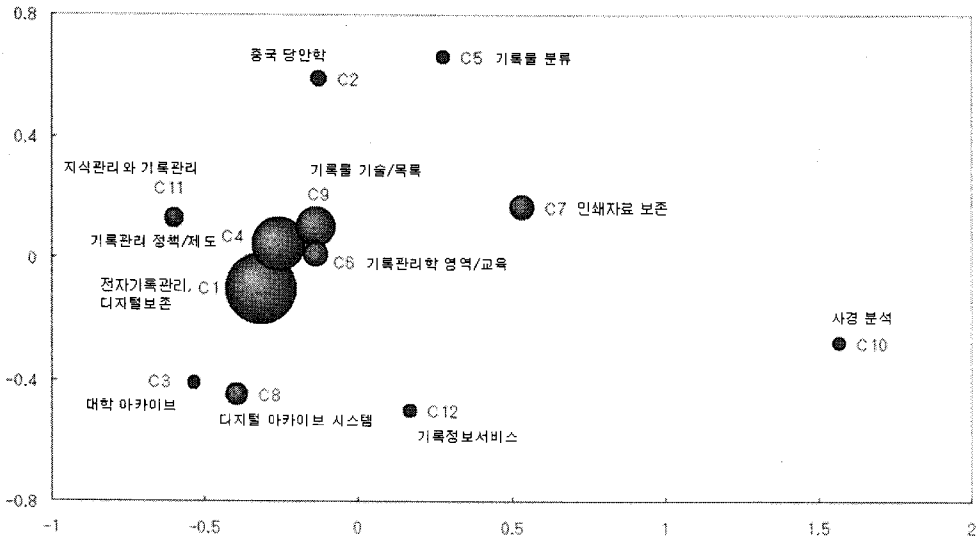
문헌 군집으로 구성된 기록관리학 분야의 지적 구조를 시각적으로 분석하기 위해서 다차원척도 분석을 수행하였다. 전체 문헌 145개는 SPSS의 다차원분석 프로시저에서 처리가 불가능할 뿐만 아니라 설사 표현되더라도 개별 문헌을 시각적으로 구분하기도 어려우므로 앞에서 구성한 12개 군집으로 2차원 지도를 생성하였다.

군집간 코사인 유사도 행렬을 입력데이터로 삼아서 SPSS의 PROXSCAL 명령을 사용하여 〈그림 3〉과 같은 2차원 지도를 생성하였다.

PROXSCAL 명령을 수행할 때의 옵션 설정은 ‘데이터형식: 데이터로부터 근접행렬 작성’, ‘모형: 순서척도’, ‘측도: 유클리디안 거리’, ‘값 변환-표준화: 변수기준 z점수’를 선택하였고, 나머지는 SPSS의 초기 설정대로 두었다. 이 그림을 보면 군집 C10이 혼자 오른쪽에 떨어져 있는데, 이 군집에 속한 문헌들은 서지학에 가까운 주제로서 기록관리학의 일반 주제와는 다소 이질적이라고 할 수 있다. 군집 C10을 제외하고 살펴보면 중앙의 네 군집(C1, C4, C6, C9)이 핵심 영역에 해당하고 바깥의 일곱 군집(C2, C3, C5, C7, C8, C11, C12)이 주변 영역에 해당하는 동심원 구조임을 알 수 있다.

3.3.2 네트워크 분석

앞에서는 Ward 기법으로 클러스터링한 21개 주제군집을 2단계 클러스터링을 통해서 12개로



〈그림 3〉 12개 문헌군집의 MDS 지도

축소하고 주제를 분석하였다. 여기서는 의미상의 단위주제군집인 21개 군집의 네트워크 분석을 통해서 기록관리학 분야의 지적 구조를 파악해보기로 한다. 군집 간 유사도를 군집 간 링크의 가중치로 간주하고 21개 군집의 네트워크에 대해서 지적 구조 분석을 위한 중심성 척도 4가지(이재운 2006)를 적용해보면 〈표 6〉과 같은 결과를 얻는다.

역시 동일한 21개 군집 간 유사도 행렬을 입력하여 패스파인더 네트워크 PFNet($q=n-1, r=\infty$)를 구축하고 주제 영역을 분석한 결과는 〈그림 4〉와 같다.

〈그림 4〉에서는 〈C4, 기록관리 정책〉과 〈C1, 디지털 보존〉의 두 군집이 중심을 차지하며 〈기록관리 정책/제도〉, 〈고문서〉, 〈메타데이터〉, 〈디지털 아카이빙〉, 〈기록물 분류/목록〉의 5개 영역이 주요한 가지로 나타났다. 이는 기록관리학 연구 영역이 위와 같은 5개 영역으로 세분하여 나타남을 설명한다고 할 수 있다. 즉, 기록관리학 영역

의 핵심적인 두 주제영역은 〈기록관리정책〉과 〈디지털보존〉 영역이며, 〈디지털 보존〉 영역은, 전자기록물을 중심으로 한 메타데이터, 디지털 아카이빙, 고문서 분야와의 연관성이 높은 것을 확인할 수 있다. 한편 〈기록관리 정책〉 영역은 기록물 분류/목록, 기록관리 정책/제도 분야와의 연관성이 높은 것으로 나타났다.

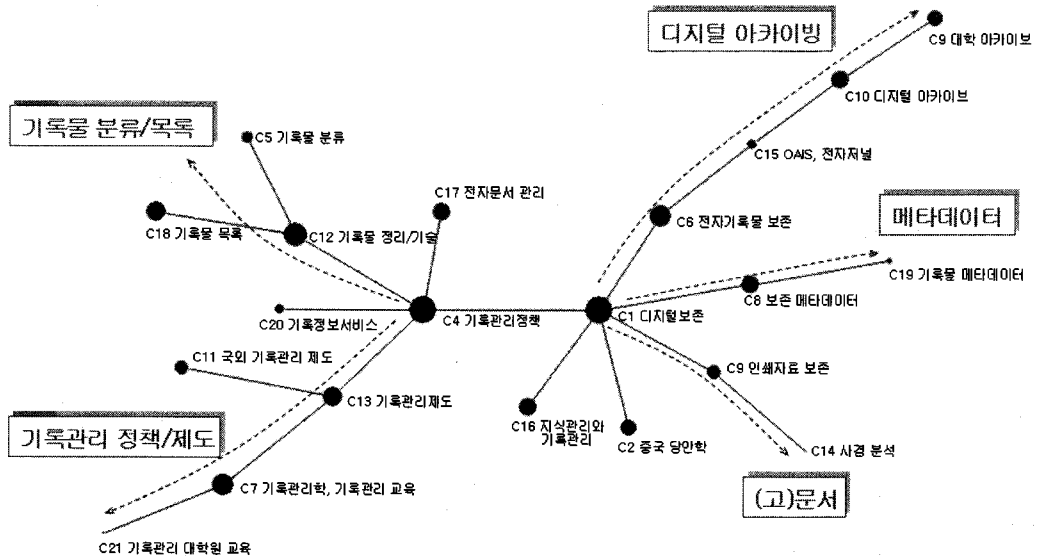
3.4 문헌단위 지적구조 분석

3.4.1 문헌 유사도 네트워크 분석

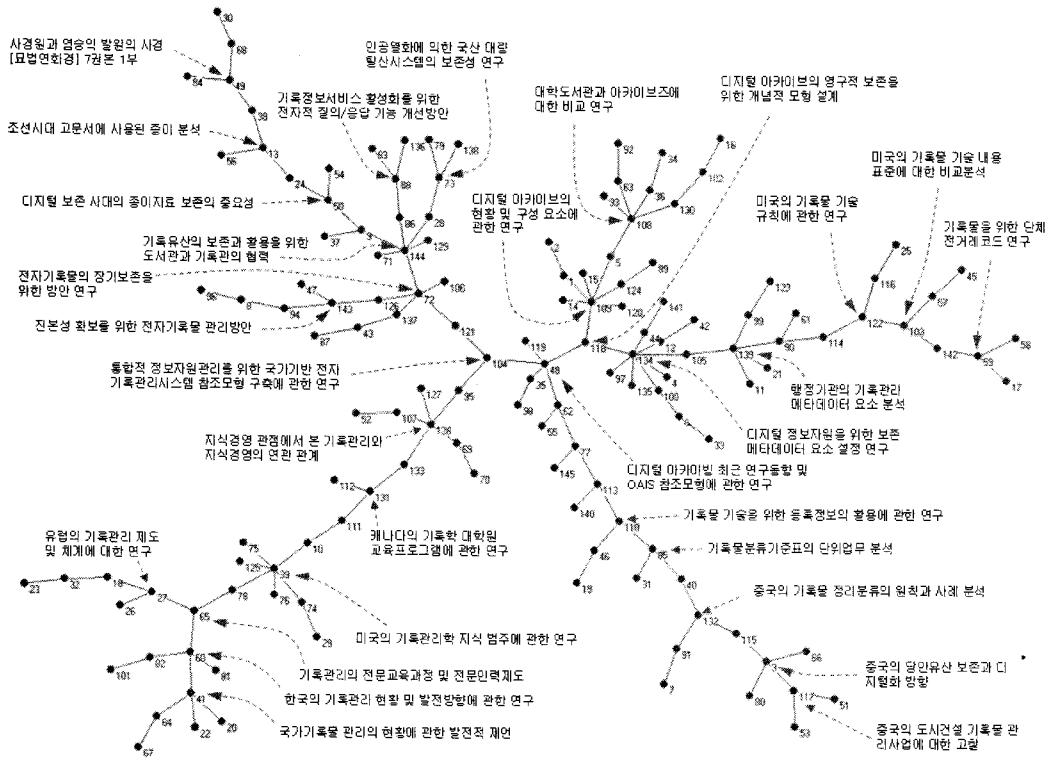
문헌 군집을 단위로 하는 분석에서는 개별 문헌 단위의 세밀한 분석은 불가능하다. 이를 보완하기 위해서 개별 문헌을 단위로 하는 문헌 유사도 네트워크를 분석해보았다. 145개 문헌간 코사인 유사도를 입력하여 생성한 패스파인더 네트워크는 〈그림 5〉와 같다. 가지별로 주제를 파악하기 위해서 네트워크 상에서 연결된 문헌이 많은 경우는 제목을 표시하였다.

〈표 6〉 21개 문헌군집의 중심성 측정 결과

NODE	군집의 핵심 주제	상대적 삼각매개 중심성	평균 연관성	최근접 중심성	평균 프로파일 연관성	군집크기
C1	디지털 보존	0.826	0.139	0.250	0.041	20
C2	중국 당안학	0.279	0.075	0.000	-0.049	5
C3	대학 아카이브	0.268	0.083	0.000	-0.030	4
C4	기록관리 정책	0.837	0.149	0.250	0.053	18
C5	기록물 분류	0.137	0.065	0.000	-0.062	4
C6	전자기록물 보존	0.516	0.110	0.000	0.004	9
C7	기록관리학, 기록관리 교육	0.558	0.113	0.050	-0.007	8
C8	보존 메타데이터	0.368	0.107	0.050	0.001	6
C9	인쇄자료 보존	0.216	0.064	0.050	-0.064	10
C10	디지털 아카이브	0.358	0.095	0.100	-0.020	5
C11	국의 기록관리 제도	0.211	0.076	0.000	-0.042	5
C12	기록물 정리/기술	0.637	0.120	0.100	0.018	8
C13	기록관리 제도	0.405	0.100	0.100	-0.003	6
C14	사경 분석	0.000	0.013	0.000	-0.136	4
C15	OAIS, 전자저널	0.100	0.079	0.050	-0.036	4
C16	지식관리와 기록관리	0.326	0.087	0.000	-0.027	8
C17	전자문서 관리	0.326	0.099	0.000	-0.006	4
C18	기록물 목록	0.474	0.104	0.000	-0.005	9
C19	기록물 메타데이터	0.037	0.065	0.000	-0.056	2
C20	기록정보서비스	0.116	0.065	0.000	-0.055	3
C21	기록관리 대학원 교육	0.005	0.052	0.050	-0.079	3



〈그림 4〉 21군집의 PFNet(군집의 삼각매개중심성에 따른 크기 표시)



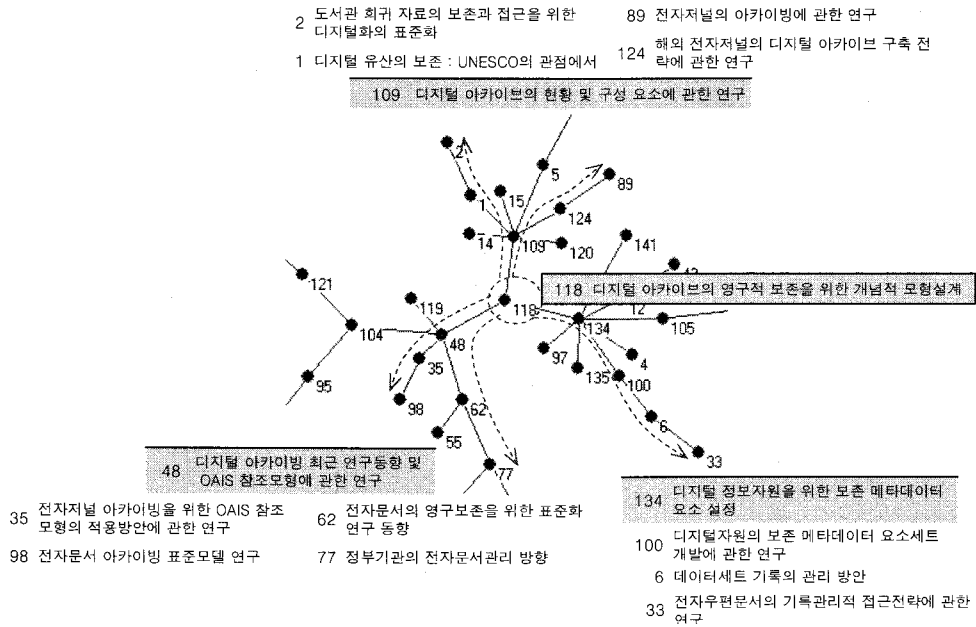
〈그림 5〉 145개 문헌의 PFNet($q = n-1, r = \infty$)

〈그림 5〉에서 타 문헌과의 연결가치가 많은 문헌, 즉 중심문헌은 연구 범위가 세부적이고 특정적이기보다는 포괄적이면서 복합적인 문헌인 것으로 나타났다. 또한 발표 시기 상으로 보았을 때 중심문헌이 반드시 선행하는 논문은 아니었다. 이는 이 그림에 나타난 문헌간의 관계가 인용이 아닌 색인어 공유에 근거했기 때문이다. 따라서 중심문헌이 가지 끝에 위치한 문헌보다 상대적으로 중요한 문헌이라고 판단하거나, 선행연구문헌이라고 판단할 수는 없다. 문헌 유사도 네트워크에서 중심문헌의 의미는 연결된 주변 문헌의 주제를 포괄적으로 대변하는 문헌으로 제한하여 해석해야 한다.

〈그림 5〉와 같이 문헌 유사도 PFNet을 산출한 결과, 중심에 나타나는 문헌들은 전자기록관리시스템 및 디지털 아카이빙에 관한 논문들임을 확인할 수 있다.

PFNet에서는 개별 문헌간의 관계가 드러나므로 특정 영역에 대한 세부 구성을 파악할 수 있는데, 전체적인 패스파인더 네트워크는 그림에서 나타나는 대로 좌측 상단, 하단과 우측 상단, 하단으로 구분할 수 있다. 이 중에서 좌측 상단, 중앙, 좌측 하단의 세 부분을 확대하여 문헌간의 주제적 연계 구조를 파악해보기로 한다.

우선 전자기록 관리에 관한 좌측 상단 가지를 확대하여 분석하면 〈그림 6〉과 같다.



〈그림 7〉 PFNet의 중앙에 위치한 디지털 아카이빙 영역을 확대한 결과

마지막으로 패스파인더 네트워크에서 기록 관리 교육 및 정책을 중심으로 한 좌측 하단 가지의 연구들은 〈그림 8〉과 같은 형태로 나타나고 있다. 기록관리 교육·정책 영역에서의 중심 문헌은 ‘기록관리의 전문교육과정 및 전문인력 제도’인데, 이는 네 개의 주제 영역으로 연결된다. 네 개의 주제 영역은 〈기록관리전문가 양성에 관한 영역〉 및 〈국가기록물 관리 현황 영역〉, 〈한국기록관리 현황 및 발전방향 영역〉, 〈유럽의 기록관리 제도 및 체계 영역〉 등이다.

이상의 개별문헌 단위의 문헌 유사도 네트워크 그림에 나타난 주요 문헌을 참고하여 각 가지에 주제 영역을 부여하면 〈그림 9〉와 같다.

〈그림 9〉에서 알 수 있듯이 개별 문헌단위 유사도 분석에 의한 패스파인더 네트워크 분석 결과 중심 주제는 〈디지털 아카이빙〉인 것으로 나

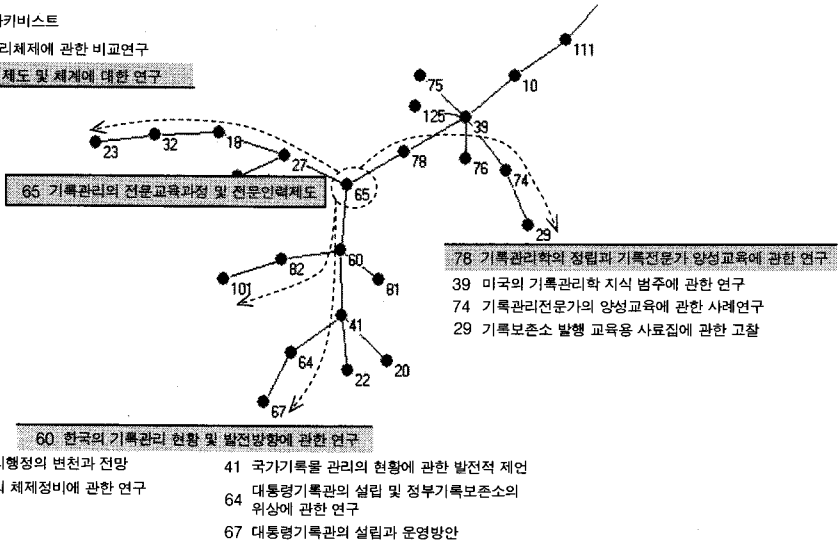
타났고, 이는 다시 〈전자기록관리〉, 〈기록보존 실무〉 및 〈고문서〉 가지영역, 〈기록관리 정책〉 영역, 〈도서관과 아카이브〉 및 〈기록물 목록〉 영역, 〈기록물 분류〉 및 〈당안관리학〉 영역으로 연결되어 확장되고 있음을 확인할 수 있다.

3.5 시기별 분석

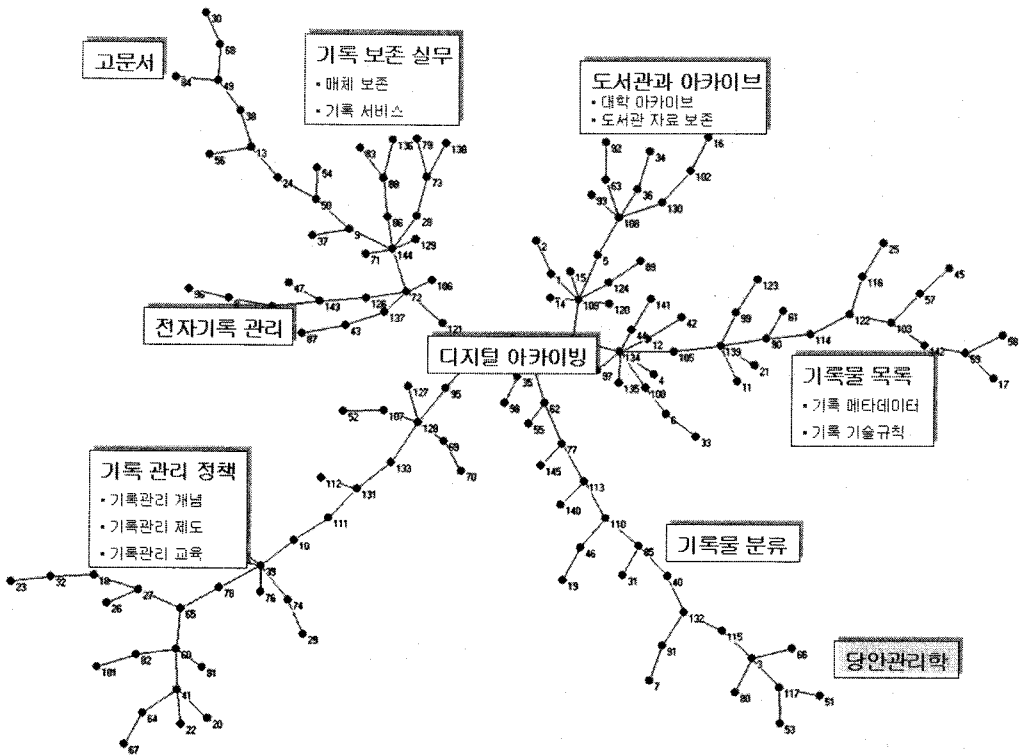
각 군집 소속 문헌의 게재 시기별 분포를 확인하기 위하여 시기별 분석을 수행하였다. 1기에 비하여 2기에 문헌이 증가하여 군집이 성장한 정도를 수치로 나타내어 비교하기 위해서 군집 성장 지수 CGI(Cluster Growth Index)를 다음과 같이 고안하였다.

$$CGI = \frac{2기\ 문헌\ 수 - 1기\ 문헌\ 수}{2기\ 문헌\ 수 + 1기\ 문헌\ 수}$$

- 23 러시아혁명 이후부터 1930년대까지의 소련의 기록관리제도
- 32 러시아 혁명과 아카이비스트
- 18 북미지역 기록관리체제에 관한 비교연구
- 27 유럽의 기록관리 체제 및 체계에 대한 연구



<그림 8> PFNet의 좌측 하단에 위치한 기록관리 교육·정책 영역을 확대한 결과



<그림 9> 145개 문헌의 패스파인더 네트워크에 표시한 주제 영역

이 값은 0을 기준으로 2기에 문헌이 증가하였으면 양수가 되고 감소하였으면 음수가 된다. 범위는 -1에서 1사이로서 1기에 있다가 2기에 전혀 없는 경우는 -1이 되며, 1기에 전혀 없다가 2기에 처음으로 문헌이 발표되었다면 1이 된다. 한편 CGI에 각 군집의 성장 규모를 반영하는 지수로 다음과 같은 가중 군집 성장 지수 WCGI(Weighted CGI)를 산출할 수 있다.

$$WCGI = |2기\ 문헌수 - 1기\ 문헌수| \times CGI$$

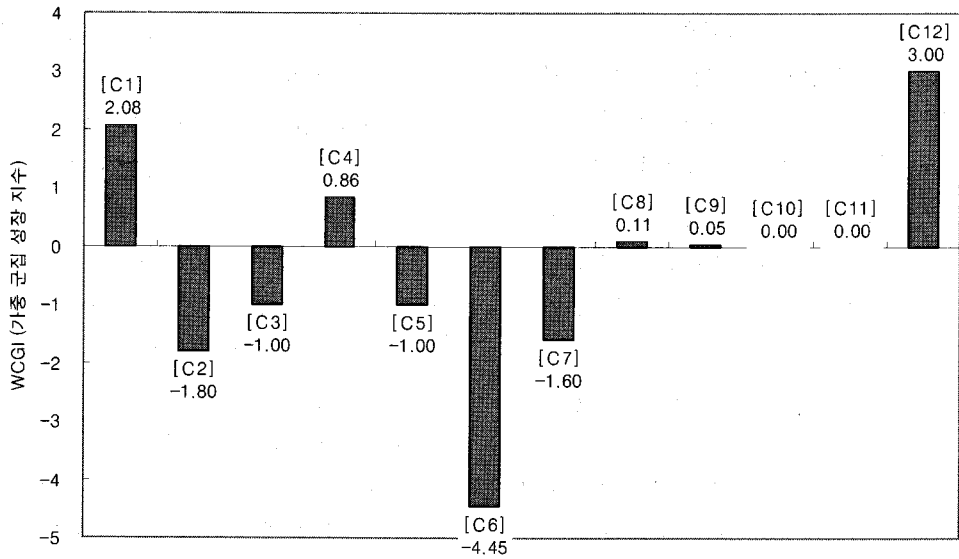
CGI가 상대적인 비율 위주의 지표인데 반해서 WCGI는 절대적인 문헌 수의 증가도 어느 정도 반영하는 지표가 된다. 예를 들어 1기에는 전혀 문헌이 없다가 2기에 문헌이 새롭게 발표된 경우에는 몇 건이 발표되었더라도 CGI는 항상 1이 되지만, WCGI는 증가한 문헌의 수에 비례해서 값이 커진다. 따라서 WCGI가 더 분별력이 있는 지표라고 할 수 있으므로 여기서는 WCGI로 군집의 시기별 성장을 측정하였다.

<그림 10>에서 확인할 수 있듯이, 1기에 비하여 2기에 두드러지게 성장한 군집은 <기록정보서비스 영역(C12)>이다. 실제 문헌 수는 3편에 불과하지만, 이 주제 영역은 1기에는 없었던 주제 영역이 2기 시기에 새로 등장함으로써 가장 성장한 군집으로서 나타나고 있다. 그 다음으로는 C1의 <전자기록관리·디지털보존> 영역으로서, 성장지수 뿐만 아니라 양적인 군집 내 문헌 수에 있어서도 주도적인 영역을 차지하고 있음을 확인할 수 있다. 한편, 1기에 비하여 2기에 뚜렷하게 축소된 군집은 C6으로서 <기록관리학 영역 및 교육 주제> 영역이다.

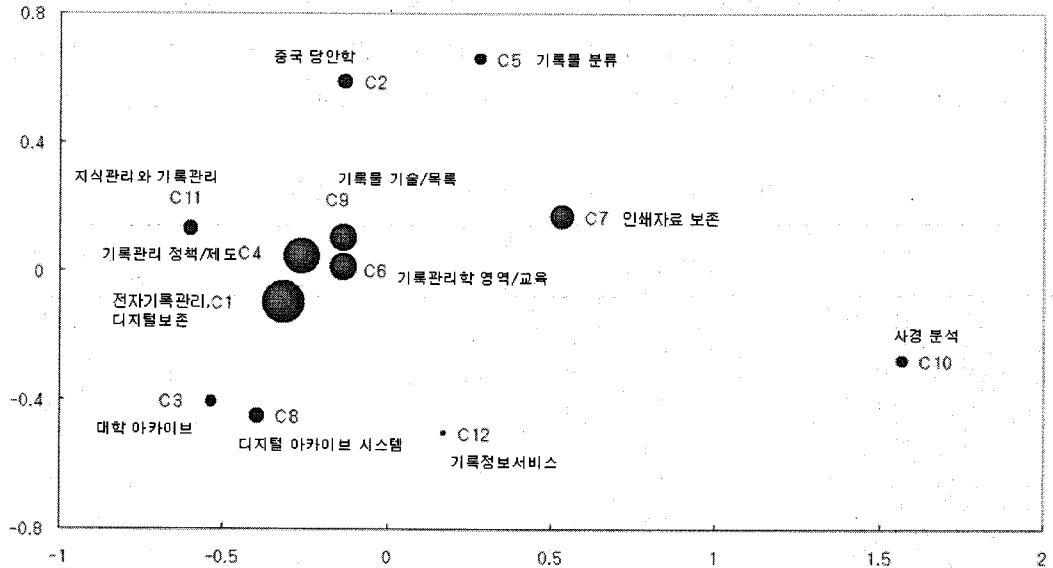
1기 문헌 수 및 2기 문헌 수를 군집의 크기를 반영하여 MDS 지도를 중심으로 표현해보면 <그림 11> 및 <그림 12>와 같다. 두 그림을 비교해서 살펴보면 군집 성장 지수로 확인하였던 주제 영역별 부침을 시각적으로 확인할 수 있다. 대체적으로 지도에서 위쪽에 위치한 군집인 C2, C5, C7이 2기에 감소하였고, 상대적으로 아래쪽에 위치한 군집인 C1, C8, C12가 2기

<표 7> 각 군집 소속 문헌 수의 시기별 분석

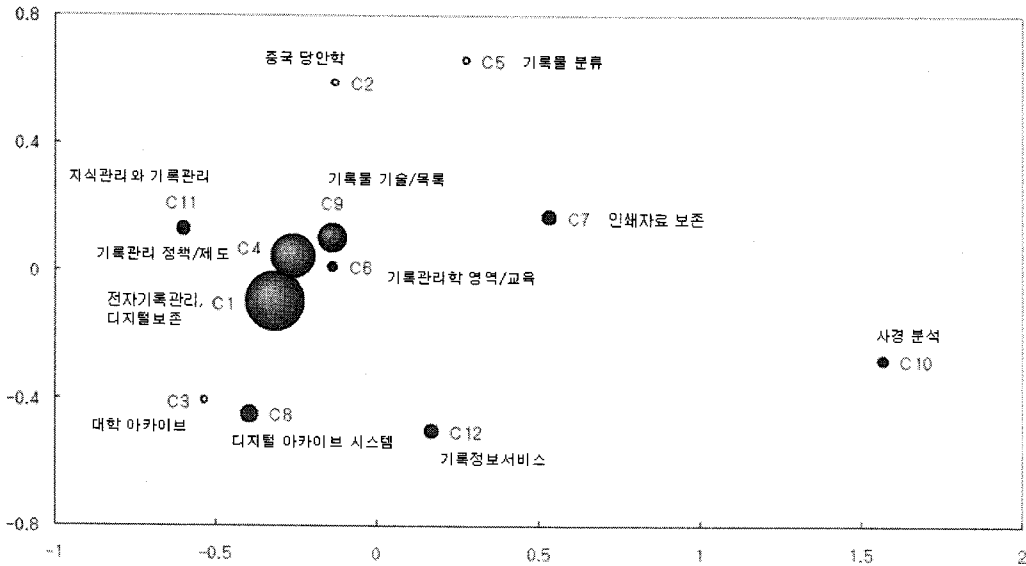
군 집	시기별 소속 문헌 수									WCGI
	'01	'02	'03	'04	'05	'06	합계	1기('01-'03)	2기('04-'06)	
C1 전자기록관리·디지털 보존	2	5	8	8	12	4	39	15	24	2.08
C2 중국 당안학	1	2	1	0	1	0	5	4	1	-1.80
C3 대학 아카이브	2	0	1	0	1	0	4	3	1	-1.00
C4 기록관리 정책/제도	6	2	4	10	6	1	29	12	17	0.86
C5 기록물 분류	0	0	3	0	1	0	4	3	1	-1.00
C6 기록관리학 영역/교육	4	3	2	0	2	0	11	9	2	-4.45
C7 인쇄자료 보존	4	1	2	1	2	0	10	7	3	-1.60
C8 디지털 아카이브 시스템	2	0	2	2	3	0	9	4	5	0.11
C9 기록물 기술/목록	2	4	3	5	4	1	19	9	10	0.05
C10 사경 분석	1	0	1	1	0	1	4	2	2	0.00
C11 지식관리와 기록관리	3	1	0	2	2	0	8	4	4	0.00
C12 기록정보서비스	0	0	0	0	0	3	3	0	3	3.00
합계	27	18	27	29	34	10	145	72	73	



<그림 10> 군집별 WCGI(가중 군집 성장 지수)



<그림 11> 1기 문헌 수를 군집의 크기로 반영한 MDS 지도



〈그림 12〉 2기 문헌 수를 군집의 크기로 반영한 MDS 지도

에 성장한 것으로 나타났다.

4. 결론

이 연구에서는 텍스트 마이닝의 문헌 클러스터링 기법과 문헌 유사도 네트워크 분석 기법을 적용하여 국내에서 수행된 기록관리학 분야의 지적구조를 파악하였다. 분석을 위한 데이터는 2001년부터 2006년까지 발간된 국내 문헌정보학 분야에서의 대표적인 5종 학술지를 중심으로 선정된 145편의 기록관리학 관련 주제 논문들을 그 대상으로 하였다.

문헌 클러스터링과 문헌 유사도 네트워크 분석을 통해 파악된 기록관리학 연구분야 지적구조에서 확인할 수 있는 결론 및 의의는 다음과 같다.

첫째, 주제영역 분석을 통하여서, 상위 단계

에서는 기록관리학 영역을 크게 두 개의 주제영역인 <기록관리 정책, 고문서 관리>와 <아카이브, 기록관리 실무>로 구분할 수 있었으며, 하위 단계에서는 총 12개의 군집을 산출하였다. 이 중에서 <기록관리 정책, 고문서 관리> 영역은 2001년도부터 2003년도까지에 해당되는 1기에 주로 연구되었으며, 이 영역 내에서 가장 활발히 수행된 세부 연구 주제는 ‘기록관리 정책/제도’인 것으로 나타났다. <아카이브, 기록관리 실무> 영역은 2004년도부터 2006년도에 해당되는 2기에 주로 수행되었으며, 이 영역 내에서 가장 활발히 수행된 세부 연구주제는 ‘전자기록관리·디지털보존’인 것으로 나타났다.

둘째, 다차원척도 및 네트워크 분석을 통하여 군집단위 지적구조 분석을 수행하였다. 군집단위 지적구조 분석 결과, 기록관리학 영역의 가장 핵심적인 주제 영역은 <전자기록관리·디지털보존>, <기록관리정책·제도>, <기록물 기술

/목록), <기록관리학 영역/연구>의 순서인 것으로 나타났다. 또한 PFNet 분석 결과 이 중 가장 중심을 차지하는 주제 영역인 <전자기록관리·디지털보존>영역은 '디지털 아카이빙', '메타데이터', '고문서' 주제 영역으로 연결되며, <기록관리정책·제도> 영역은 '기록물 분류/목록', '기록관리 정책/제도' 로 연결되는 것을 확인할 수 있었다.

셋째, 개별 문헌단위의 세밀한 지적구조를 반영하기 위한 네트워크 분석 결과, 145개 문헌의 중심 주제영역은 <디지털 아카이빙>인 것으로 나타났으며, 이 영역으로부터 <전자기록관리>, <도서관과 아카이브>, <기록물 목록>, <기록물 분류>, <기록물 보존 실무>, <기록관리 정책>, <당안관리학>, <고문서> 등의 영역으로 연결되고 있음을 확인할 수 있었다. 또한 구체적으로는 유사한 문헌들의 연결구조를 파악할 수 있었다.

넷째, 시기별 분석을 수행한 결과 1기(2001~2003년)에서 2기(2004~2006년)로 넘어오면서 가장 성장한 주제 영역은 <기록정보서비스>와 <전자기록관리·디지털보존>으로 나타났고, 반대로 두드러지게 연구비중이 감소한 주제 영역은 <기록관리학 영역/교육>과 <인쇄자료 보존>인 것으로 나타났다.

이상의 결과를 종합하여 보면, 2001년부터 2006년까지 수행된 국내의 기록관리학 영역의 핵심적인 주제 영역은 <전자기록관리·디지털보존>, <기록관리정책·제도>, <기록물 기술/목록>, <기록관리학 영역/연구>이며, 전반기보다 후반기에 들어서면서 <전자기록관리·디지털보존> 영역에 대한 관심과 연구가 증가하고 있음을 확인할 수 있다. 이를 통하여 국내 기록

관리학 연구가 이제는 태동기를 벗어나 기록관리학 형성기를 향하여 나아가고 있으며, 앞으로는 그 비중이 더욱 증가할 전자기록관리 및 디지털 보존 영역을 중심으로 확장될 것을 기대할 수 있다.

또한 문헌 분석을 통하여서는 <디지털 아카이빙> 주제 영역이 중심을 이루고 있으며, 최근 들어서는 <기록정보서비스> 영역이 새롭게 등장하고 있음을 확인할 수 있었다.

이 연구에서 사용된 분석 기법은 다음과 같은 장점이 있는 것으로 나타났다.

첫째, 비구조화된 본문을 이용하는 텍스트 마이닝에서의 문헌 클러스터링 기법과 유사도 네트워크 분석 기법을 적용함으로써 인용 DB를 구축하지 않고서도 지적 구조를 분석할 수 있었다.

둘째, 분석대상 단위의 수가 100개 이상인 관계로 통상적인 계층적 응집식 클러스터링 기법 대신 2단계 문헌 클러스터링을 수행하였다. 이를 통하여 지적구조를 더 정확히 파악할 수 있었다.

이상의 분석 결과는 분석 대상 논문에 나타난 국내 연구동향을 반영한 것이므로 기록관리학 분야의 전체적인 연구 영역을 나타내는 것으로 확대 해석하는 것은 무리일 것이다. 추후 학위 논문과 단행본까지 그 분석 범위를 넓힌다면 연구 결과의 보편성을 더 획득할 수 있을 것으로 기대된다. 또한 본 연구에서 수행한 기본 틀에 덧붙여 저자별 분석, 국외 데이터들을 중심으로 한 분석 등을 추가 수행할 수 있을 것이며, 일정 기간이 흐른 이후 3기의 기록관리학 연구영역을 수행한다면 국내 기록관리학 연구 영역의 확장 과정 및 변화 추이도 파악할 수 있을 것이다.

참 고 문 헌

- 김희정. 2005. 저자 동시인용분석에 의한 국내의 기록관리학 분야의 지적구조 비교에 관한 연구. 『문헌정보학회지』, 39(3): 207-224.
- 김희정. 2006. 정보기술 관점을 기반으로 한 기록관리학 연구영역 확장성 연구. 『한국 기록관리학회 2006년도 추계학술발표논문집』, pp.7-25.
- 이재윤. 2006. 계량서지적 네트워크 분석을 위한 중심성 척도에 관한 연구. 『한국문헌정보학회지』, 40(3): 191-214.
- 이재윤, 정진아. 2005. 계층적 문서 클러스터링을 위한 응집식 기법과 분할식 기법의 비교 연구. 『제12회 한국정보관리학회 학술대회 논문집』, pp.65-70.
- 정연경. 2003. 미국의 기록관리학 지식범주에 관한 연구. 『한국기록관리학회지』, 3(2): 33-50.
- 최정태 외. 2005. 『기록관리학사전』. 한올아카데미.
- Ananiadou, S., and J. Mcnaught, eds. 2005. *Text Mining for Biology and Biomedicine*. Artech House Publishers.
- Berry, M. W. 2003. *Survey of Text Mining: Clustering, Classification, and Retrieval*. London: Springer-Verlag.
- Brichford, M. 1988. "Who are the archivists and what do they do?" *American Archivist*, 51:106-110.
- Callon, M., J. Law, and A. Rip, eds. 1986. *Mapping the Dynamics of Science and Technology: Sociology of Science in the Real World*. London: The Macmillan Press Ltd.
- Carpineto, C., R. de Mori, G. Romano, and B. Bigi. 2001. "An information-theoretic approach to automatic query expansion." *ACM Transactions on Information Systems*, 19(1): 1-27.
- Chen, H., S. S. Fuller, C. Friedman, and W. Hersh, eds. 2005. *Medical Informatics: Knowledge Management and Data Mining in Biomedicine*. London: Springer-Verlag.
- Cox, R. J. 1987. "American archival literature: Expanding horizons and continuing needs, 1901-1987." *American Archivist*, 50(1): 306-323.
- Cox, R. J. 2000. "Searching for authority: Archivists and electronic records in the new world at the Fin-de-Siecle." *First Monday*, 5(1).
- Feldman, R., and J. Sanger. 2007. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. New York, NY: Cambridge University Press.
- Gilliland-Swetland, A. J. 1992. "Archivy and the computer: a citation analysis of North American archival periodical literature." *Archival Issues*, 17(2): 95-

- 112.
- Gilliland-Swetland, A. J. 1995. *Development of an expert assistant for archival appraisal of electronic communications : an exploratory study*. Ph.D dissertation, University of Michigan.
- Glenisson, P., W. Glänzel, and O. Persson. 2005. "Combining full-text analysis and bibliometric indicators." *Scientometrics*, 63(1): 163-180.
- Kao, A, and S. R. Poteet. 2007. "Overview." In A. Kao and S. R. Poteet, eds., *Natural Language Processing and Text Mining*. London: Springer-Verlag, pp.1-7.
- Konchady, M. 2006. *Text Mining Application Programming*. Charles River Media.
- Kostoff, R. N. 1993. "Database tomography for technical intelligence." *Competitive Intelligence Review*, 4(1): 38-43.
- Kostoff, R. N. 2003. "Text mining for global technology watch." In: Drake M, editor. *Encyclopedia of library and information science*, vol. 4, 2nd ed. New York: Marcel Dekker 2003. p. 2789 - 99.
- Kostoff, R. N., D. R. Toothman, H. J. Eberhart, and J. A. Humenik. 2001a. "Text mining using database tomography and bibliometrics: A review." *Technological Forecasting & Social Change*, 68(3): 223-253.
- Kostoff, R. N., H. J. Eberhart, and D. R. Toothman. 1998. "Database tomography for technical intelligence: A roadmap of the near-earth space science and technology literature." *Information Processing & Management*, 34(1): 69-85.
- Kostoff, R. N., J. A. del Rio, J. A. Humenik, E. O. Garcia, and A. M. Ramirez. 2001b. "Citation mining: Integrating text mining and bibliometrics for research user profiling." *Journal of the American Society for Information Science and Technology*, 52(13): 1148-1156.
- Kostoff, R. N., M. F. Shlesinger, and G. Malpohl. 2004. "Fractals text mining using bibliometrics and database tomography." *Fractals*, 12(1): 1-16.
- Kostoff, R. N., R. Tshiteya, K. M. Pfeil, J. A. Humenik, and G. Karypis. 2005. "Power source roadmaps using bibliometrics and database tomography." *Energy*, 30(5): 709-730.
- Kostoff, R. N., T. Braun, A. Schubert, D. R. Toothman, and J. A. Humenik. 2000. "Fullerene data mining using bibliometrics and database tomography." *Journal of Chemical Information and Modeling*, 40(1): 19-39.
- Losiewicz, P., D. W. Oard, and R. N. Kostoff. 2000. "Textual data mining to support science and technology management." *Journal of Intelligent Information Systems*, 15(2): 99-119.
- Menne-Haritz, A. 2004. *Business Processes :*

- An Archival Science Approach to Collaborative Decision Making, Records, and Knowledge Management.* Dordrecht ; Boston : Kluwer Academic Publishers.
- Miller, Thomas W. 2004. *Data and Text Mining: A Business Applications Approach.* Prentice Hall.
- Redmond-Meal, A., and M. M. K. Hlava, eds. 2005. *ASIS&T Thesaurus of Information Science, Technology, and Librarianship.* Medford, NJ: Information Today, Inc.
- Sullivan, Dan. 2001. *Document Warehousing and Text Mining: Techniques for Improving Business Operations, Marketing, and Sales.* John Wiley & Sons.
- Zanasi, A., ed. 2005. *Text Mining and Its Applications to Intelligence, CRM and Knowledge Management.* WIT Press.

〈부록〉 분석 대상 기록관리학 관련 논문 145건과 12군집 형성 결과

문헌번호	제목	군집
D001	디지털 유산의 보존 : UNESCO의 관점에서	1
D002	도서관 희귀 자료의 보존과 접근을 위한 디지털화의 표준화	
D004	디지털라이제이션의 디지털 이미지와 메타데이터 표준화작업	
D009	한국의 세계기록유산 보존 현황 및 과제	
D028	도서관 및 기록관 자료보존프로그램에 관한 고찰	
D037	미디어에 있어서의 자막기록의 의미와 전달성 - 공중파방송과 비디오 아트에서의 자막기록을 중심으로 -	
D044	디지털 도서관의 웹자원 기반 장서개발	
D045	The Concept of Finding Aids	
D050	디지털 보존 시대의 종이자료 보존의 중요성	
D054	종이기록물의 복구와 보존	
D063	국립중앙도서관 자료보존 환경분석에 관한 연구	
D092	디지털자료의 납본과 보존을 위한 각 국가의 노력	
D093	디지털 정보자원 보존의 위험관리 분석 : 대학도서관 전자정보실 중심으로	
D097	디지털 보존의 비용요소에 관한 연구	
D106	장기보존기록물 선별을 위한 업무분석적 평가방안	
D119	디지털 아카이빙의 경제성 분석 연구	
D129	기록물 재평가 및 처분을 통한 보존관리 전략에 관한 연구	
D135	디지털 아카이빙의 워크플로우와 보존처리 기술에 관한 연구	
D141	웹 아카이빙의 성과와 과제	
D144	기록유산의 보존과 활용을 위한 도서관과 기록관의 협력	
D012	전자정부 메타데이터 표준	
D042	전자정부의 의의와 추진실적 및 향후 과제	
D100	디지털자료의 보존 메타데이터 요소세트 개발에 관한 연구	
D105	기록 관리 메타데이터의 개념 모델링	
D134	디지털 정보자원을 위한 보존 메타데이터 요소 설정 연구	
D139	행정기관의 기록관리 메타데이터 요소 분석 - ISO 15489를 기준으로 -	
D008	전자기록철의 구조와 관리방안 - 영국 ERMS 표준을 중심으로 -	
D047	기록관리에 있어서 '진본성'의 이해: 실무자와 이용자 여론 조사	
D072	전자기록물의 장기보존을 위한 방안 연구 : 개념을 중심으로	
D094	전자기록의 진본성 유지를 위한 전략	
D096	전자기록관리시스템(ERMS) 설계표준의 기능요건 분석 - ISO 15489를 기준으로 -	
D104	통합적 정보자원관리를 위한 국가기반 전자기록관리시스템 참조모형 구축에 관한 연구	
D121	신뢰성 있는 전자기록관리를 위한 법적 기반에 관한 연구	
D126	전자기록물을 위한 보존매체의 관리	
D143	진본성 확보를 위한 전자기록물 관리방안	
D055	종이문서의 디지털 기록의로의 전환-보존	
D062	전자문서의 영구보존을 위한 표준화 연구 동향	
D077	정부기관의 전자문서관리 방향	
D098	전자문서 아카이빙 표준모델 연구	
D003	중국의 당안유산 보존과 디지털화 방향	2
D051	퇴번 당안자적의 회복과 보호	
D053	미래중국당안관리연구-수자화 당안관 건설 초기 탐색	
D080	중국 당안학의 현황 및 발전추세	
D117	중국의 도시건설 기록물 관리사업에 대한 고찰	
D005	디지털 아카이브즈와 보존을 위한 실무 기술	3
D036	대학 아카이브즈의 사진자료 관리에 관한 연구 - 부산대학교 도서관의 사례를 중심으로 -	
D108	대학도서관과 아카이브즈에 대한 비교 연구	
D130	대학 Archives란 무엇인가 : Archives의 개념과 내용	

D006	데이터세트 기록의 관리 방안	4
D011	국가 기록관리 표준 정비의 방향	
D016	대학기록물 수집정책에 관한 연구 : 한신대학교를 중심으로	
D017	여성사 관련 자료의 보존현황과 관리방안: 해방이후 ~ 1980년대 여성정책 및 여성단체 자료를 중심으로	
D025	가톨릭 교회 기록물 관리에 관한 연구 - 인천교구 기록관을 중심으로 -	
D033	전자우편문서의 기록관리적 접근전략에 관한 연구	
D043	한국 정부기록보존소의 역사기록물 공개에 관한 검토	
D071	우리나라 기록관리정책의 발전전략에 관한 고찰	
D081	한국 공공기관 기록보존 관리의 현황조사에 관한 연구	
D082	한국 기록관리행정의 변천과 전망	
D083	한국의 국가기록관리기관 정보조사제공 활성화를 위한 정책 연구	
D087	우리나라의 비밀기록 관리현황에 관한 연구	
D101	국회 기록관리 체제정비에 관한 연구	
D102	국내 대학기록물 관리의 현황 분석	
D113	한국 자료관 제도의 역할과 특징에 관한 연구	
D116	기록관리 용어 사용 실태 분석과 표준화 방안 연구	
D137	공개활성화를 위한 전자기록물 관리정책 연구	
D140	특성화에 기반한 시·군 자료관 설립 방안 연구 - 경기도를 사례로 -	
D020	중국·일본의 기록관리 제도에 관한 연구	
D022	기록관리법의 개정과 관련한 제문제 연구	
D041	국가기록물 관리의 현황에 관한 발전적 제언	
D060	한국의 기록관리 현황 및 발전방향에 관한 연구	
D064	대통령기록관의 설립 및 정부기록보존소의 위상에 관한 연구	
D067	대통령기록관의 설립과 운영방안: 미국 대통령 기록관의 사례와 교훈	
D018	북미지역 기록관리체제에 관한 비교 연구	
D023	러시아혁명 이후부터 1930년대까지의 소련의 기록관리제도	
D026	아시아의 기록관리 제도 및 체계에 대한 연구 -- 인도, 싱가포르, 베트남을 중심으로 -	
D027	유럽의 기록관리 제도 및 체계에 대한 연구 - 영국·프랑스·독일을 중심으로 -	
D032	러시아 혁명과 아키비스트, 1917-1920	
D007	시민단체 기록 분류방안 연구 : 환경연합을 중심으로	5
D091	조선총독부 기록물을 위한 기능분류체계 개발 연구	
D115	중국의 기록물 검색분류 - 『中國檔案分類法』(1997)을 중심으로 -	
D132	중국의 기록물 정리분류의 원칙과 사례 분석	
D010	한국과 미국의 대학원 기록관리학 교과과정의 비교 연구	6
D039	미국의 기록관리학 지식 범주에 관한 연구	
D065	기록관리의 전문교육과정 및 전문인력제도	
D074	기록관리전문가의 양성교육에 관한 사례연구 - 이탈리아의 기록관리학 전공과 교육과정을 중심으로 -	
D075	기록관리학의 발전을 위한 교육과정연구 - 스루가다이대학의 경우를 중심으로 -	
D076	일본에 있어서 기록관리학의 발전상황과 기록관리학회의 역할	
D078	기록관리학의 정립과 기록전문가 양성교육에 관한 연구	
D125	저자 동시인용분석에 의한 국내외 기록관리학 분야의 지적구조 비교에 관한 연구	
D111	미국 문헌정보학 대학원의 기록관리학 교육과정에 관한 연구	
D112	기록학 대학원 교육 프로그램에 관한 비교 연구 - 한국과 북미를 중심으로 -	
D131	캐나다의 기록학 대학원 교육프로그램에 관한 연구	
D013	조선시대 고문서에 사용된 종이 분석	7
D024	삼지의 보존성에 관한 연구	
D034	사진용 필름, 인화지의 감도측정에 따른 보존특성	
D038	고문서의 보존처리에 관한 사례연구 - 장흥보림사사천왕상에서 발견된 월인태보(권25)를 중심으로 -	
D056	하버드 엔칭도서관 소장 사마방목에 관한 고찰	
D066	[삼국유사]의 환웅과 열제신능의 기록학적 비교고찰	
D073	인공열화에 의한 국산 대량탈산시스템의 보존성 연구	
D079	기록물에서 분리된 미생물에 대한 훈증소독의 효능검증과 독성조사	
D138	자료매체의 보존성	
D145	문서기록물의 파일링시스템에 관한 연구	

D014	전자문서관리와 시스템 설계의 동향과 전망	8
D015	대중 매체에 비친 아카이브의 공공 인식: 아카이브와 관련된 직업군, 기관, 기록물에 대한 2003-2004년간의 국내보도	
D109	디지털 아카이브의 현황 및 구성 요소에 관한 연구	
D118	디지털 아카이브의 영구적 보존을 위한 개념적 모형 설계에 관한 연구	
D120	아카이브의 효율적인 운영과 정리에 관한 연구	
D035	전자저널 아카이빙을 위한 OAIS 참조모형의 적용방안에 관한 연구	
D048	디지털 아카이빙 최근 연구동향 및 OAIS 참조모형에 관한 연구	
D089	전자저널의 아카이빙에 관한 연구	
D124	해외 전자저널의 디지털 아카이브 구축 전략에 관한 연구	
D019	국가기록원의 기록물 정리·기술의 현황과 개선방안	9
D021	호주의 레코드키퍼링 시스템에 대한 연구	
D029	기록보존소 발행 교육용 사료집에 관한 고찰 - 호주와 미국, 영국 국립기록보존소를 중심으로 -	
D031	기록물처리일정표의 작성과 유지 관리	
D040	문서분류의 이론과 변천에 관한 연구 - 조선조이후 현행 '정부공문서분류'까지 -	
D046	한국의 기록물 등록 및 기술에 대한 기록관리적 접근	
D085	기록물분류기준표의 단위업무 분석	
D110	기록물 기술을 위한 등록정보의 활용에 관한 연구	
D057	EAD를 이용한 기록자료 자동화방안 - Online Archive of California 프로젝트를 중심으로 -	
D058	출처주의의 새로운 경향과 적용에 관한 연구	
D059	기록물을 위한 단체 전거레코드 연구	
D099	기록물용 KORMARC 데이터필드 개발을 위한 메타데이터 요소에 관한 연구	
D103	미국의 기록물 기술 내용표준에 대한 비교분석 - APPM2와 DACS를 중심으로 -	
D114	기록물 기술을 위한 MARC AMC에 관한 연구	
D122	미국의 기록물 기술규칙에 관한 연구	
D123	특수형태 기록물 관리를 위한 기술요소에 관한 연구	
D142	기록물 생산자 전거제어를 통한 맥락정보의 구축 및 교환 - ISAAR(CPF) 2판과 EAC를 중심으로 -	
D061	디지털 도서관 콘텐츠 관리를 위한 KORMARC(EAD)통합 시스템 구현	
D090	기록물정보 관리를 위한 메타데이터 설계와 구현에 관한 연구	
D030	일본에 현존하는 4권의 주본화엄경의 변상서에 관한 연구	10
D049	사경원과 염승익 발원의 사경 [묘법연화경] 7권본 1부	
D068	조선초 익안대군발원사경에 관한 연구	
D084	고려 개인발원사경	
D052	조직체에 있어서 지혜와 경험의 전승	11
D069	지식공유에 영향을 미치는 요인에 관한 연구	
D070	과학기술 종합지식 경영시스템(R&D-KMS)구축	
D095	기록관리를 기반으로 한 통합형 지식관리시스템 구축 방향 연구	
D107	한국의 민간기업 기록보존시스템에 관한 연구	
D127	CiteSpace 적용을 통한 디지털 보존 지식영역 비주얼화 연구	
D128	지식경영 관점에서 본 기록관리와 지식경영의 연관 관계 - 지식의 정의와 지식관리 프로세스를 중심으로 -	
D133	도서관·정보학 그리고 기록학	
D086	이용자 중심의 기록정보 활용 및 서비스 활성화에 관한 연구	12
D088	기록정보서비스 활성화를 위한 전자적 질의/응답 기능 개선방안 - 국가기록원을 중심으로	
D136	국가기록원 인터넷 게시판 이용자의 정보이용행태 분석	