

# 네트워크 분석을 기반으로 한 웹 아카이빙 주제영역 연구

## A Study on Web Archiving Subject Analysis Based on Network Analysis

김 희 정(Heejung Kim)\*

### 초 록

본 연구에서는 Web of Science DB를 대상으로 주제어(topic)가 web archiving에 해당하는 288건의 논문을 대상으로 동시출현단어 네트워크 분석을 수행하였다. 분석 결과 웹 아카이빙 주제 영역에서는 의학영역 정보기술 및 시스템과 관련된 이미지 아카이빙 관련연구들이 가장 중점적으로 수행되어 왔다. 문헌정보학 및 기록관리학 영역에서의 웹 아카이빙 연구는 크게 웹 아카이빙 및 디지털 보존 프로젝트 주제와 웹 아카이빙 툴과 방법론 주제를 중심으로 수행되어왔으며, 향후 웹 아카이빙 소프트웨어 및 툴 관련 연구가 활성화될 수 있을 것으로 예측된다.

### ABSTRACT

In this study, co-word occurrence analysis was performed on 288 articles retrieved from the Web of Science DB with the topic of web archiving. Results showed that research on image archiving information technology and system were most frequently carried out especially in medical area. Within library and information science and records management & archives areas, web archiving/digital preservation project subject and web archiving tools and methodology subject were studied mostly. It is expected that research related to web archiving software and tools will be increased in near future.

키워드: 웹 아카이빙, 디지털 보존, 디지털 아카이빙, 동시출현용어분석, 계량정보학, 기록관리학  
Web Archiving, Digital Preservation, Digital Archiving, Co-word Occurrence Analysis,  
Bibliometrics, Records Management and Archives

---

\* 국제백신연구소 정보자료실장(heejung@ivi.int)

논문접수일자 : 2011년 6월 8일    논문심사일자 : 2011년 6월 10일    게재확정일자 : 2011년 6월 15일

## 1. 서론

웹 정보자원에 대한 의존도가 높아짐에 따라 서 가치 있는 웹 자원의 장기보존 및 아카이빙에 대한 필요성이 증대하고 있다. 해외 주요 국가들의 웹 아카이빙 국제 협력은 대략 2003년부터 시작되었는데, 주로 국가도서관을 중심으로 이루어져왔다.

웹 아카이빙 사업이 국가 프로젝트를 중심으로 진행되어 옴에 따라 국가 사례를 중심으로 한 연구를 비롯하여, 웹 아카이빙 수행 절차 및 방법론 연구에 이르기까지 다양한 연구가 진행되고 있다.

웹 아카이빙 영역은 대상 자원의 특성상 정보 기술 및 시스템, 워크플로우 및 메타데이터 영역 외에도 사업 운영과 관련된 법 제도 및 정책 영역에 이르기까지 광범위한 주제 영역이 연관된다.

본 연구에서는 웹 아카이빙과 관련된 연구 데이터들을 광범위하게 수집하여 계량분석함으로써, 실제 그동안 수행되어온 누적연구들 속에서 웹 아카이빙의 다양한 연구 주제간의 관련성과 세부 주제영역을 확인하고자 한다.

분석을 위한 데이터베이스는 Web of Science를 선정하였고, 계량서지용 공개 소프트웨어인 Network Workbench를 전처리 및 행렬도출을 위한 도구로 사용하였다. 시각화 도구로는 역시 공개 소프트웨어인 NodeXL을 활용하였다.

또한, 웹 아카이빙 주제 영역은 비교적 최근에 출현한 주제 영역으로서, 인용정보를 수집하고 분석하기에는 누적된 정보가 부족하다. 즉, 인용 정보 누적 역사가 짧고, 최근에 새롭게 출현한 영역일수록 인용 정보 분석 보다는, 현 시

점에서의 주제 분포를 보다 직접적으로 제시해주는 동시출현단어 분석 기법이 보다 더 적절하다. 따라서 본 연구에서는 동시출현단어 분석 기법을 선정하여 연구를 수행하였다.

## 2. 선행 연구

웹 아카이빙과 관련된 연구는 이제윤과 조현주(1998) 및 조현주(1999)의 초기 연구가 있는데, 인터넷 자원을 국가 유산으로 간주하여 국가 차원에서 개발, 보존, 이용시키는 정책의 필요성을 강조하고 있다. 외국의 사례 정리와 함께 우리나라에 맞는 모형을 개발하였다. 여기에서는 웹 아카이빙이라는 용어가 아닌 인터넷 보존, 또는 웹 보존이라는 용어가 사용되었다.

웹 아카이빙이라는 용어가 출현하면서 수행된 국내 연구는 2004년부터 시작하였는데, 서혜란(2004)은 Kulturarw3, PANDORA 그리고 Internet Archive의 사례소개와 함께 웹 아카이빙의 기술적 과제와 정책적 과제를 제시하였다.

김유승(2007)은 저작권, 납본과 수집 및 진본성과 관련된 법 제도적 문제 고찰과 함께, 선택적 웹 아카이빙과 포괄적 웹 아카이빙 정책을 결합한 복합적 웹 아카이빙 정책을 제안하였다.

Heejung Kim and Hyewon Lee(2007)는 선택적 웹 아카이빙에 있어서 필요한 관리적 메타데이터 요소를 제안하였다.

오상훈 외(2008)는 웹 아카이빙을 위한 디지털 납본 프로세스의 단계별 정의와 기능을 제시

하였다.

차승준 외(2009)는 공공기관 심층 웹 기록물의 아카이빙을 위하여 장기보존 포맷과 더블린 코어를 매핑한 KoDeWeb 메타데이터 요소들을 제안하였다.

김희정(2010)은 해외 국가 웹 아카이빙 사례들을 인터페이스 유형 및 특성들을 중심으로 항목별로 분석하고 살펴보았다. 분석 결과 국내의 경우에는 수집 대상 웹 콘텐츠 주제 영역의 다양화, 웹 자원 수집의 분담화, 체계적인 아카이빙 전략 수행 등의 개선 방안이 제시되었다.

해외 연구로는 Masanès(2005)가 웹 아카이빙 방법론과 접근법에 관한 비교연구를 수행하였다. 포괄적 웹 아카이빙 및 선택적 웹 아카이빙에 따른 방법론과 해석을 제시하고 있다.

Guenthera and Myrickb(2007)의 연구에서는 MINERVA 프로젝트를 중심으로 웹 아카이빙 관련 기술정보 및 보존정보를 제공하는 메타데이터 전략을 제시하고 있다. 더불어 METS 및 PREMIS에 대한 분석도 수행하였다.

Murray and Hsieha(2008)는 웹 아카이빙 프로젝트인 Web-at-Risk 프로젝트를 수행함으로써, 수집한 웹 콘텐츠에 대한 이용자 평가를 수행하였다. 사서, 연구자, 큐레이터 및 아키비스트를 대상으로 웹 아카이빙 콘텐츠에 대한 유용성 및 정보요구를 분석하였다.

Spaniol et al.(2009)은 웹 아카이빙 대상 콘텐츠의 일관성 있고 효율적인 수집을 위한 전략을 제시하였다. Internet Archive의 WARC을 중심으로 기술하고 있다.

Denev et al.(2011)은 웹 아카이빙의 데이터 질적 평가를 위하여 SHARC라는 프레임워크를 제시하고 있다. 질적으로 우수하고 검증된 웹 콘텐츠들을 아카이빙 하기 위한 전략과 방법론을 제시하고 있으며, 웹 아카이빙 수행시에 적용할 수 있는 척도를 개발하였다.

이상에서 살펴볼 수 있듯이, 웹 아카이빙과 관련된 연구들은 프로젝트를 중심으로 한 연구와 법제도 정책적 관점에서의 연구, 그리고 메타데이터 관련 연구 등 다양한 연구들이 수행되어 왔지만, 웹 아카이빙 주제를 중심으로 수집한 광범위한 데이터를 계량 분석한 사례는 확인되지 않는다.

계량적 기법을 적용한 웹 아카이빙 관련 연구의 분석을 통해서 실제로 수행되어 온 웹 아카이빙 주제 연구의 하위 영역의 분포 및 시계열적 흐름도 확인해보고자 한다.

### 3. 데이터 수집 및 분석 과정

#### 3.1 데이터 수집

웹 아카이빙과 관련된 연구데이터를 수집하기 위하여서 Web of Science 데이터베이스<sup>1)</sup>를 선정하였다. 검색조건에 있어서는 주제(topic)이 web archiving인 경우로 한정하였고, 대상은 SSCI, SCI-Expanded, A&HCI 데이터베이스, 그리고 대상기간은 모든 기간으로 설정하였다. 검색결과 288건의 데이터가 수집되었다(<표 1> 참조).

1) Web of Science는 Journal Citation Reports 및 기타 다른 DB들과 함께 Web of Knowledge의 하위범주에 포함된다.

〈표 1〉 분석 대상 데이터 조건 및 검색 결과

분석대상 DB		검색조건	기간	검색결과
Web of Science	SSCI	topic = web archiving	1900~2011년	288 articles
	SCI-Expanded			
	A&HCI			

288건의 검색논문들 중 동시출현단어 분석에 필요한 관련 데이터를 다운로드받았다. 이에는 인용정보를 제외한 분석에 필요한 초록, 표제, 그리고 키워드 정보가 해당되며, 분석을 위하여 텍스트파일 형식을 선택하였다. 관련 내용은 〈그림 1〉과 같다.

### 3.2 분석 과정

검색된 288건의 논문들은 계량분석 도구인 Network Workbench를 활용하여 수집 데이터의 전처리 과정을 진행하였다. 전처리 과정을 추가하여 설명하면 다음과 같다. 먼저, 먼저 다운로드 받은 텍스트파일의 원 데이터를 cleaing 과정을 통하여 필드별로 분리하고, 공저자의 경우 delimiter로 구분한다. 그 다음 문서정규화(text normalization) 통하여 초록, 표제(title), 그리고 키워드의 stemming 과정을 수행한다. 전처리 과정이 수행된 전체 데이터의 동시출현 용어 행렬을 1차적으로 산출하였다. 관련 그림

은 〈그림 2〉와 같다.

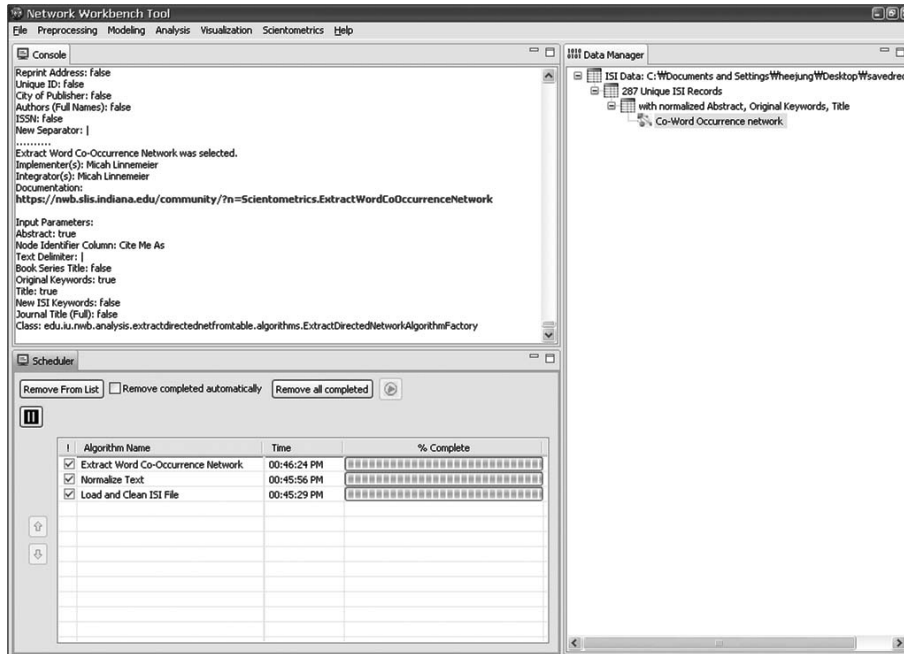
Network Workbench는 인디애나 대학의 NWB(Network Workbench) team을 중심으로 개발된 자바 기반의 네트워크 분석 및 계량 서지 분석 전용의 공개 소프트웨어이다. 본 논문에서 Network Workbench를 활용한 이유는, ISI 형식 파일 등을 읽어서 다양한 네트워크 분석을 효율적으로 수행하고, 동시인용을 비롯한 여러 분석 기능을 구현하는 최신 계량 분석 도구이기 때문이다.

Network Workbench를 활용하여 산출한 1차적인 동시출현용어 행렬의 일부 데이터를 제시하면 〈표 2〉와 같다.

〈표 3〉에서 확인할 수 있듯이 1차적으로 산출된 행렬은 제거해야 할 불용어도 많고, 단어가 총 4,405종으로서 그 규모가 너무 방대하므로, 분석과 해석의 용이성을 위하여 규모를 축소할 필요가 있다. 이에 따라 빈도수 30회 이상의 용어를 재선정하였으며, 대상 용어 중 수작업으로 불용어(stop word)를 제거하였다. 그



〈그림 1〉 동시출현단어 분석을 위한 관련 데이터 수집



〈그림 2〉 Network Workbench에 의한 문서정규화 과정 및 동시출현용어행렬 산출

〈표 2〉 1차 행렬 산출 word list 데이터 예(일부)

*Nodes		
id*int	label*string	references*int
1	Miguel S, 2011, V62, P1130, DOI	0
2	year	26
3	seen	1
4	emerg	19
5	open	46
6	access	123
7	oa	6
8	option	7
9	scholar	14
10	commun	128
	이하 증략	
4396	strict	1
4397	Kieley JM, 1996, V28, P300	0
4398	tip	1
4399	get	1
4400	Shah JJ, 1996, V28, P347	0
4401	interrog	1
4402	drl	1
4403	ddl	1
4404	Atwell R, 1995, V19, P966	0
4405	iv	1

〈표 3〉 1차 행렬 산출 matrix list 데이터 예(일부)

*UndirectedEdges		
source*int	target*int	weight*int
2	3	1
2	4	1
2	5	3
3	4	1
3	5	1
2	6	9
4	5	2
3	6	1
2	7	2
4	6	10
이하 생략		

결과 최종적으로 동시출현단어 분석 대상으로 93개의 핵심용어군을 선정하였다.

93개의 핵심용어군을 대상으로 Network Workbench를 활용하여 동시출현단어 네트워크를 위한 행렬 데이터를 재산출하였다. 지도 작성 방식은 패스파인더 네트워크 방식을 선택하였다.

산출된 행렬 결과는 여러 가지 시각화 도구를 활용할 수 있는데, 본 논문에서는 NodeXL을 활용하였다. NodeXL은 네트워크 분석 결과를 시각화하여 제시해주는 오픈소스 프로그램으로서, 특히 엑셀 기반의 데이터 처리를 효율적으로 수행할 수 있다.

#### 4. 웹 아카이빙 주제영역 분석

웹 아카이빙 주제영역 분석은 앞에서 언급한 분석 과정을 거쳐서 도출된 동시출현단어 네트워크 지도 분석과 함께 수집된 기본 데이터 분석을 중심으로 한 빈도 및 범주 분석으로 수행

되었다.

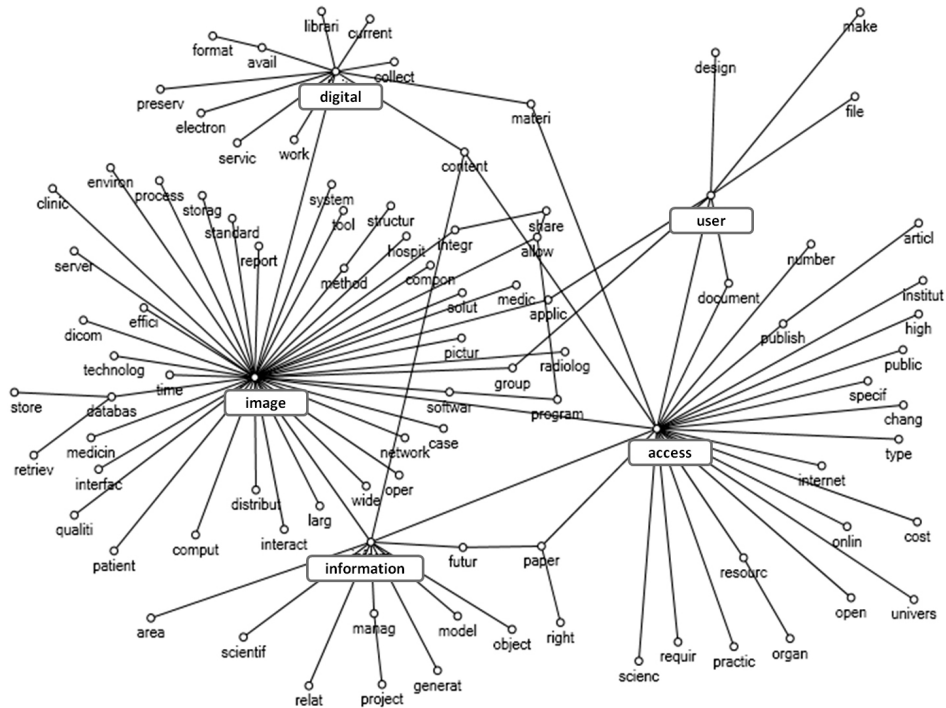
#### 4.1 주제영역 네트워크 분석

NodeXL을 활용하여 93개의 핵심용어군을 대상으로 한 동시출현단어 네트워크를 시각화하였다. 시각화 결과는 〈그림 3〉과 같다.

〈그림 3〉은 패스파인더 네트워크 방식으로 작성한 것인데, 웹 아카이빙과 관련된 핵심용어 93개의 동시출현 단어 네트워크에서는 크게 다섯 개의 중심 용어를 확인할 수 있다.

가장 많은 sub branch를 갖는 중심 용어는 image로 나타났고, 그 다음 규모로는 access, 그리고 그 다음으로는 information, digital, 그리고 user로 연결되고 있다.

〈그림 3〉에서는 웹 아카이빙 주제 영역의 핵심 키워드가 되는 주제어를 중심으로 한 네트워크 구조를 전체적인 관점에서의 조망은 가능하지만, 네트워크 내에서 각 용어의 비중, 즉, 각 용어의 빈도수를 기반으로 한 비중은 확인하기 어렵다.



〈그림 3〉 웹 아카이빙 주제 영역의 동시출현단어 네트워크

이에 따라, 빈도수를 각 노드의 크기에 반영시켜 네트워크를 재구성하였다. 이에 대한 결과는 〈그림 4〉와 같다.

〈그림 4〉에서 확인할 수 있듯이, 다섯 개의 주요 네트워크 중 먼저 image를 중심으로 한 네트워크에서는 image 용어와 함께 고빈도로 출현한 용어들은 technology, system, software, database, tool, 그리고 picture 등으로 나타나고 있다.

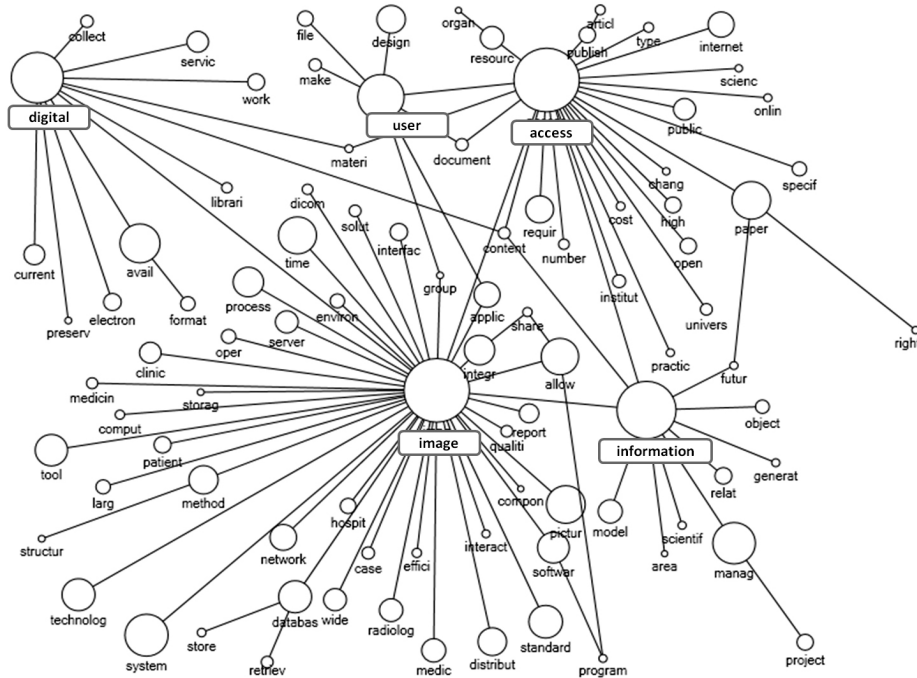
두 번째 규모인 access 네트워크에서는 internet, public, paper 등의 고빈도 용어에 연결되고 있으며, 그 다음 information 네트워크는 management와 project, 그리고 model 등의 용어가 고빈도 용어로 나타난다.

digital 네트워크는 availability와 format, 그

리고 user 네트워크는 design과 access 네트워크와 연결되고 있다.

즉, 웹 아카이빙 주제영역에서 가장 중점적으로 수행된 주제영역은 image 아카이빙을 중심으로 한 연구로서, 정보기술 및 시스템, 도구 등이 연관된 연구들이다. 또한 그 외 주변 노드들의 주제어인 medical, radiology, picture 등을 고려해보면, 의학영역과 밀접한 관련성을 갖고 있다.

두 번째로는 internet, resource, public 등의 주제와 연관된 access 관점에서의 웹 아카이빙 주제영역 연구가 이루어져왔으며, 그 다음으로 information 주제 영역에서는 웹 아카이빙과 관련된 management나 model 등의 연구가 수행되어 왔다.



〈그림 4〉 웹 아카이빙 주제영역 각 용어 빈도수를 반영한 동시출현 네트워크

웹 아카이빙 연구는 또한 digital availability 나 format과 관련된 연구 영역도 확인이 되며, user design, user access 등의 주제 연구도 수행되어 왔다.

#### 4.2 주제영역 빈도 및 범주 분석

웹 아카이빙 주제영역의 특성 확인 및 분석을 위하여 네트워크를 기반으로 한 분석에 이어서, 본 논문에서 수집한 최초의 288건의 데이터를 중심으로 빈도 및 범주 분석을 수행하였다.

먼저, 웹 아카이빙 관련 논문이 게재된 학술지를 파악하였다. 3번 이상 출현한 학술지는 총 21종이었으며, 빈도순 학술지 리스트는 〈표 4〉와 같다.

〈표 4〉에서 확인할 수 있듯이, 의학분야의 학

술지가 상당히 많이 포함되어 있다. 특히, Journal of Digital Imaging은 가장 고빈도를 보이고 있는데, 이 학술지 역시 주제영역이 ‘medical imaging (의학이미지)’과 ‘radiology(방사선학)’로 설정하고 있는 의학영역 관련 학술지이다.

이어서 웹 아카이빙 관련 논문들의 연도별 출간 빈도를 확인해 보았다. 관련 그림은 〈그림 5〉와 같다.

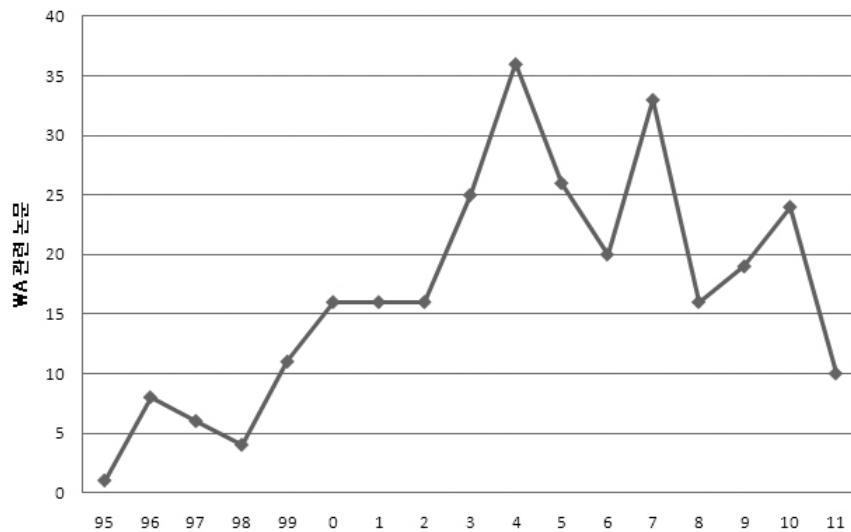
〈그림 5〉에서 살펴볼 수 있듯이, 웹 아카이빙 관련 논문은 꾸준히 증가세를 보이다가 특히 2003년부터 급속히 증가하고 있음을 확인할 수 있다.

이에 따라, 보다 핵심적인 웹 아카이빙 영역에서 수행된 연구 내용을 살펴보기 위하여, 〈표 5〉에서 선정된 상위 10개의 학술지 게재 논문 중 2003년 이후 출간된 연구논문들의 주제 영역을



〈표 4〉 웹 아카이빙 관련 논문 게재 빈도순 학술지 리스트

	학술지명	빈도수
1	JOURNAL OF DIGITAL IMAGING	38
2	RADIOGRAPHICS	15
3	IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE	6
4	LIBRARY TRENDS	6
5	EUROPEAN RADIOLOGY	5
6	PROGRAM-ELECTRONIC LIBRARY AND INFORMATION SYSTEMS	5
7	BIOINFORMATICS	4
8	COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE	4
9	COMPUTERIZED MEDICAL IMAGING AND GRAPHICS	4
10	ROFO-FORTSCHRITTE AUF DEM GEBIET DER RONTGENSTRAHLEN UND DER BILDGEBENDEN VERFAHREN	4
11	BIOTECHNIQUES	3
12	COMPUTERS & GEOSCIENCES	3
13	ELECTRONIC LIBRARY	3
14	JOURNAL OF INFORMATION SCIENCE	3
15	JOURNAL OF MEDICAL INTERNET RESEARCH	3
16	JOURNAL OF TELEMEDICINE AND TELECare	3
17	JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY	3
18	MALAYSIAN JOURNAL OF LIBRARY & INFORMATION SCIENCE	3
19	METHODS OF INFORMATION IN MEDICINE	3
20	NUCLEIC ACIDS RESEARCH	3
21	ONLINE INFORMATION REVIEW	3



〈그림 5〉 연도별 웹 아카이빙(WA) 관련 논문 수

〈표 5〉 상위 10개 저널의 2003년 이후 발간 논문 65편의 주제 범주

	주제범주	빈도수
1	PACS(Picture Archiving and Communication System)	20
2	Web PACS	9
3	Clinical Tools	9
4	Web Archiving(WA) / Digital Preservation(DP) project	6
5	DICOM	5
6	Medical Teaching	4
7	Web Archiving Tools and Methods	4
8	Other Image / Information System	3
9	EMR(Electronic Medica Record)	2
10	etc	3

살펴보았다.

상위 10개 학술지의 발표 논문들 중 2003년 이후에 출간된 연구논문들은 총 65편이었다. 65편의 논문들의 각 초록 및 논문 표제들을 내용 분석을 하여 주제 범주를 부여한 결과, 고빈도 주제 범주 및 빈도수는 다음과 같이 나타났다.

주제 범주 중 PACS는 전체 65편의 논문 중 20편이 해당되어 나타나고 있다. Web PACS 까지 감안하면, 전체 논문의 절반에 가까운 논문이 PACS 주제 범주에 해당된다. PACS는 의료영상저장전송장치로 불리는데, 임상결과 및 엑스레이 사진을 비롯한 의료영상이미지들을 필름 없이 컴퓨터에 저장하고 전송하는 시스템이다. 5번에 나타난 DICOM은 의료영상 및 디지털 이미지를 위한 파일 포맷으로서 국제의료영상표준으로 해석할 수 있다.

전체 주제 범주 리스트를 통해서 확인할 수 있듯이, 웹 아카이빙이라는 주제어와 관련된 논문들의 상당 부분이 의학영역의 논문들이다.

Other Image/Information System에 해당되는 7번 범주 역시 Bioinformatics 저널 및 Journal of Digital Imaging에서 게재한 논문들로서 의학영역에서의 임상진료 및 의료연구와 관련

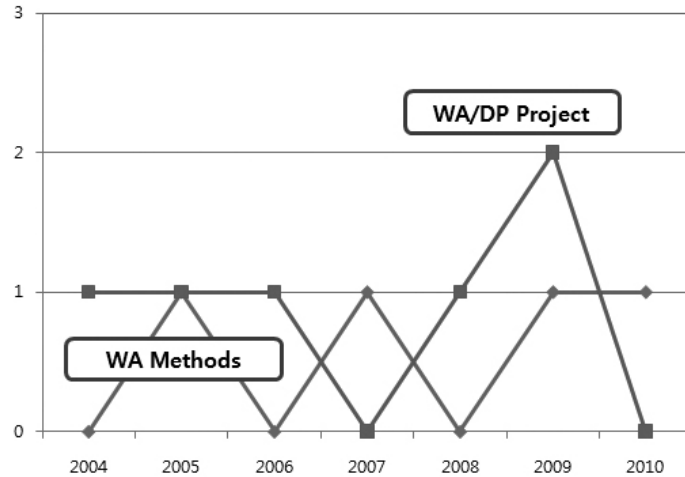
된 이미지/정보시스템에 관련된 내용들이다.

문헌정보학 및 기록관리학 영역 주제 범주에 해당되는 4번과 7번 주제의 연구들의 연도별 논문빈도를 확인해 보았다. 저빈도이기는 하지만 2004년 이후 꾸준히 연구를 이어오고 있었다. web archiving과 주제적으로 연관되는 digital preservation 관련 연구도 수행되어 왔다. 관련 그림은 〈그림 6〉과 같다.

앞의 표와 그림에서 살펴볼 수 있듯이, 웹 아카이빙과 관련된 문헌정보학 및 기록관리학 연구는 주로 웹 아카이빙 및 디지털 보존 프로젝트, 그리고 웹 아카이빙 툴과 방법론 등이 중점적으로 이루어져왔다.

국제적으로 1990년대 중반 이후부터 국가적 차원에서의 웹 아카이빙 또는 디지털 보존 프로젝트가 적극적으로 수행되어 왔는데, 완료된 사업에 대한 정리와 소개가 2000년도 중반 이후 주로 이루어져 왔다. 또한, 웹 아카이빙과 관련해서는 새롭게 출현하는 정보기술을 반영한 다각적인 방법론 및 도구 등이 개발되고 있는데, 추후에는 웹 아카이빙 방법론 및 새로운 툴 개발 관련 연구주제가 증가할 것으로 예측된다.

한편, 본 연구에서 도모하고자 하였던 'web



〈그림 6〉 주제범주 4번과 7번의 연도별 분포

〈표 6〉 주제어(topic) “web archiving”과 대상 DB SSCI 한정검색 결과

	학술지명	주제범주	세부주제영역	출간연도
1	Program-Electronic Library and Information System	WA / DP Project	NEDLIB	2004
2	Journal of Information Science	WA Policy	moral perspective	2004
3	Library Trends	WA Tools and Methods	Internet Archive와 BNF	2005
4	Journal of Information Science	WA / DP Project	South Africa 국가 Project	2006
5	Online Information Review	WA Others	WA Review	2007
6	Government Information Quarterly	WA / DP Project	NDIIPP	2008
7	Program-Electronic Library and Information System	WA / DP Project	크로아티아 National Project	2008
8	New Media & Society	WA Theory	WA Theory and History	2009
9	Library Trends	WA / DP Project	NDIIPP	2009
10	Library Trends	WA Tools and Methods	Web Archives Workbench	2009
11	Library Trends	WA / DP Project	일리노이대학 프로젝트	2009
12	Journal of Librarianship and Information Science	WA / DP Project	National Library와 WA	2009
13	Electronic Library	WA / DP Project	PANDORA and Web2.0	2009
14	Australian Library Journal	WA / DP Project	KB and PANDORA	2010

archiving’ 핵심 연구결과를 재확인하고자 다시 한번 Web of Science를 조건을 다르게 하여 재검색해보았다.

검색 주제어는 web archiving이 아닌 “web archiving”으로 재선정하여 용어를 고정하였

고, 대상 DB도 SSCI만 선택하였다. 대상연도는 마찬가지로 ‘all years’로 하였다. 그 결과 단지 14건만이 검색되었다. 2004년에 최초로 출현하였고, 이후 최근에 올수록 증가하고 있었다. 관련 내용은 〈표 6〉과 같다.

〈표 6〉에서 확인할 수 있듯이 web archiving 과 관련된 연구는 2004년부터 출현하기 시작하였으며, 2009년에 가장 활발하게 이루어져왔다. 연구 내용은 web archiving 관련 프로젝트 주제가 가장 많았고, 그 다음으로는 web archiving tools와 방법론, 그리고 그 외 web archiving 정책 및 이론 관련 연구가 소수 있었다.

1990년대 중반부터 국제적으로 사업이 시작되었던 web archiving 프로젝트의 완료 시점인 2000년대 초반 관련 연구가 많이 산출이 된 것으로 보인다. 최근에 게재되는 프로젝트 관련 내용들은 주로 부족한 영역의 보완이라든지 확대 발전을 위한 내용 등이 중심을 이루고 있다.

## 5. 결 론

본 논문에서는 동시출현단어 네트워크를 활용하여 웹 아카이빙 주제 영역을 분석하였다. 분석을 위하여 Web of Science 데이터베이스를 선정하였으며 검색 결과 선정된 288건의 논문을 대상으로 하였다.

웹 아카이빙 주제 영역의 논문은 1995년도부터 출현하기 시작하였고, 2003년부터 급속히 증가해 왔다.

또한, 분석 결과 웹 아카이빙 주제영역에서는 의학영역 정보기술 및 시스템과 관련된 이미지 아카이빙 관련 연구들이 가장 중점적으로 수행된 것을 확인할 수 있었다.

문헌정보학 및 기록관리학 영역에서의 웹 아카이빙 연구는 2004년부터 출현하고 있으며, 2009년에 가장 활발하게 이루어져 왔다. 크게 주제 범주를 웹 아카이빙 및 디지털 보존 프로젝트 영역과 웹 아카이빙 툴과 방법론 영역으로 구분할 수 있었다. 이 외에도 web archiving 정책 및 이론 관련 연구도 소수 수행되었다.

웹 아카이빙 영역은 비교적 최근 출현하기 시작한 연구인만큼 누적된 인용정보에 의존하는 계량기법이 아닌 동시출현단어 기법을 적용하여 세부 주제영역을 확인해보고자 시도하였다. 분석 결과 도출된 네트워크는, 서지데이터 원정보를 빈도 및 범주 분석함으로써 보완적인 해석이 가능하였다.

웹 아카이빙 영역은 주로 국제적인 대규모 프로젝트가 핵심 사업으로 수행되어 온 만큼, 기대와 달리 다각적인 세부주제 영역은 확인할 수 없었다. 그러나, 최근의 연구 흐름이 마무리된 국가 프로젝트의 보완과 개선을 위한 정보기술과 시스템에 관심을 갖고 있는 것으로 확인된 만큼 향후에는 웹 아카이빙 관련 소프트웨어와 툴 분석 관련 연구가 활성화될 수 있을 것으로 전망된다.

아울러, 본 연구에서는 PACS나 DICOM 관련 논문에 의하여 image 허브의 비중이 강조되었는데, 추후 두 주제를 제외한 데이터 범주를 중심으로 한 연구 및 초기에 사용되었던 'internet archiving' 용어 관련 연구데이터를 포함한 분석은 후속 연구로 남겨두고자 한다.

## 참 고 문 헌

- 김유승. 2007. 웹 아카이빙의 법·제도적 문제에 대한 고찰. 『한국문헌정보학회지』, 41(3): 5-24.
- \_\_\_\_\_. 2008. 복합적 웹 아카이빙 정책에 관한 고찰. 『한국문헌정보학회지』, 42(4): 159-179.
- 김희정. 2010. 웹 아카이빙 인터페이스 유형 및 특성 분석. 『한국기록관리학회지』, 10(2): 147-170.
- 서혜란. 2004. 웹 아카이빙의 성과와 과제. 『한국비블리아학회지』, 15(1): 5-22.
- 오상훈 외. 2008. 디지털 자원의 웹 아카이빙을 위한 납본 프로세스 개발 및 기능 설계. 『정보관리학회지』, 25(4): 5-23.
- 이재운, 조현주. 1998. 인터넷 자원의 장기적 보존 및 이용을 위한 국가프로젝트. 『정보관리연구』, 29(3): 27-42.
- 조현주. 1999. 『웹 자원의 장기 보존을 위한 국가인터넷아카이브 모형 구축에 관한 연구』. 석사학위논문. 연세대학교 문헌정보학과 대학원.
- 차승준 외. 2008. 웹 기록물 아카이빙 기반기술 연구 개발. 『지식정보산업연합회 창립 기념 학술대회』, 12: 359-368.
- \_\_\_\_\_. 2009. 공공기관 심층 웹 기록물 아카이빙을 위한 메타데이터 설계. 『한국전자거래학회지』, 14(4): 181-193.
- Australia's Web Archive. PANDORA. [cited 2011.6.5]. <<http://pandora.nla.gov.au/about.html>>.
- Brown, Adrian. 2006. *Archiving Websites*. facet publishing.
- Denev, Dimitar. et al. 2011. "The SHARC Framework for Data Quality in Web Archiving." *The VLDB Journal*, 20(2): 183-207.
- Gladney, Henry. 2007. *Preserving Digital Information*. Berlin: Springer.
- Guenthera, Rebecca and Leslie Myrickb. 2007. "Archiving Web Sites for Preservation and Access: MODS, METS and MINERVA." *Journal of Archival Organization*, 4(1): 141-166.
- Heejung Kim and Hyewon Lee. 2007. "Development of Metadata Elements for Intensive Web Archiving." *Journal of the Korean Society for Information Management*, 24(2): 143-160.
- Library of Congress Web Archives. [cited 2011.6.5]. <<http://lcweb2.loc.gov/diglib/lcwa/html/lcwa-home.html>>.
- Masanès, Julien. 2002. "Towards Continuous Web Archiving." *D-Lib Magazine*, 8(12).
- \_\_\_\_\_. 2005. "Web Archiving Methods and Approaches: A Comparative Study." *Library Trends*, 54(1): 72-90.
- \_\_\_\_\_. 2006. *Web Archiving*. Springer.

Murray, Kathleen and Inga Hsieha. 2008. "Archiving Web-published materials: A needs assessment of librarians, researchers, and content providers." *Government Information Quarterly*, 25(1): 66-89.

NodeXL. *Network Overview, Discovery and Exploration for Excel*. [cited 2011.6.5]. <<http://nodeexl.codeplex.com>>.

NWB Team. 2006. *Network Workbench Tool*.

Indiana University, Northeastern University, and University of Michigan. [cited 2011.6.5]. <<http://nwb.slis.indiana.edu>>.

Spaniol, Marc. et al. 2009. *Catch me if you can: Visual Analysis of Coherence Defects in Web Archiving*. [cited 2011.6.5]. <[http://halma.mpi-inf.mpg.de/intranet/ag5/ag5publ.nsf/e127ff338913b2a3c12565f4005ef860/741b5f3149cc5897c12576b1004acf76/\\$FILE/iwaw.pdf](http://halma.mpi-inf.mpg.de/intranet/ag5/ag5publ.nsf/e127ff338913b2a3c12565f4005ef860/741b5f3149cc5897c12576b1004acf76/$FILE/iwaw.pdf)>.