

CiteSpace 적용을 통한 디지털 보존 지식영역 비주얼화 연구

A Study on Visualization of Digital Preservation Knowledge Domain Using CiteSpace

김 희 정(Hee-Jung Kim)*

목 차

- | | |
|-----------------------------|------------------------------|
| 1. 서론 | 3. 2. 2 연관성 중심분석 |
| 1. 1 연구의 목적 및 필요성 | 3. 3 문헌 동시인용분석 |
| 1. 2 연구의 방법 및 범위 | 3. 3. 1 빈도수 중심분석 |
| 2. 지식영역 비주얼화와 CiteSpace | 3. 3. 2 연관성 중심분석 |
| 3. 디지털 보존 영역에의 CiteSpace 적용 | 4. 디지털 보존 지식영역 비주얼화 결과
분석 |
| 3. 1 데이터 수집 | 5. 결론 |
| 3. 2 저자 동시인용분석 | |
| 3. 2. 1 빈도수 중심분석 | |

초 록

디지털 보존 주제분야를 중심으로 지식영역 비주얼화 (knowledge domain visualization)를 시도하였다. 분석을 위한 데이터는 1990년부터 2005년까지의 기간 동안의 Web of Science DB를 중심으로 총 74건의 문헌을 추출하여 활용하였다. 지식영역 비주얼화를 위하여 사용한 툴은 서지 DB를 중심으로 비주얼 데이터마이닝 결과를 제공하는 Java 어플리케이션인 CiteSpace이다. 분석 결과, 디지털 보존 분야의 핵심적인 지식 영역은 최신정보기술을 중심으로 한 디지털 보존전략, 정보네트워크와 보존시스템, 전자정부와 지식관리의 세 영역인 것으로 나타났다.

ABSTRACT

This article identifies an emerging research paradigm and monitors the changes in digital preservation area using CiteSpace, a Java application which supports visual exploration with knowledge discovery in bibliographic databases. 74 articles on digital preservation field covering the time period from 1990-2005 were extracted from Web of Science. According to the result of analysis, core knowledge domains in digital preservation are technical preservation strategies, information network and preservation system, knowledge management and electronic government.

키워드: 디지털 보존, 지식영역비주얼화, 비주얼 데이터마이닝
Digital Preservation, Knowledge Domain Visualization, Visual Data Mining Information
Visualization, Knowledge Visualization

* 미국 University of North Carolina at Chapel Hill 박사후연구과정(방문학자, visiting scholar)
(hjkim@ils.unc.edu)
논문접수일자 2005년 11월 15일
게재확정일자 2005년 12월 15일

1. 서론

지식영역 비주얼화 (knowledge domain visualization)는 시간의 경과에 따라서 지식 영역이 어떻게 동적으로 변화하였는가를 나타내는 데에 그 목적이 있다. 또한 분석 대상 데이터 객체와 데이터 객체들간의 관계를 시각적으로 표현함으로써 새롭게 출현하는 지식 영역을 식별하고 기존 연구 패러다임의 변화를 파악하는 기능을 한다(Chen 2005).

전통적인 다차원축척기법 (Multidimensional Scaling)의 경우에는 설정된 큰 군집에 대한 대략적인 파악은 쉽게 이루어지지만 해당 군집 내의 세부 영역이나 개별 요소에 대한 구조는 표현되지 않는다는 제한점이 있었다. 이를 보완하기 위해서는 다단계매핑기법(multi-level mapping) (Ding, et al. 2001; Noyons and Van Raan 1998.)이나 네트워크 분석기법 (Chen 1999; White 2003)을 이용할 수 있다.

본 연구에서는 디지털 보존 주제분야를 중심으로 지식영역 비주얼화를 시도하였다. 비주얼화를 위한 툴로는 미국 Drexel 대학 교수인 Chaomei Chen이 개발한 CiteSpace를 활용하

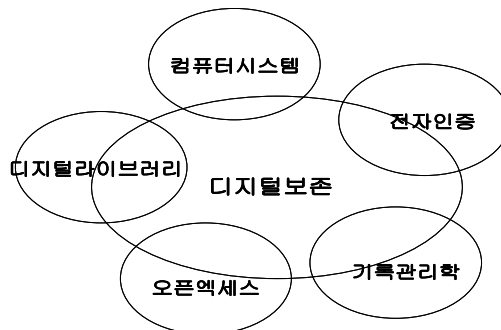
였다. CiteSpace는 Java로 개발된 동시인용분석도구로서 Pathfinder 네트워크 분석과 군집화, hub와 pivot 식별 기능 등을 수행하며, 비영리 학문적 목적으로 사용할 수 있도록 제공하고 있는 어플리케이션이다.

1.1 연구의 목적 및 필요성

본 연구는 인용분석을 기반으로 하는 지식영역 비주얼 툴인 CiteSpace를 통하여 디지털 보존 영역 중 핵심적인 세부 지식영역에 대한 파악을 시도하였다.

디지털 보존 분야는 어느 한 학문분야의 하위 영역으로 제한되는 주제 영역이 아니라, 관련되는 인접 분야에서 다양한 관점에서 접근하여 연구되고 있는 다학문적 영역이다 (그림 1 참조). 이에 따라 디지털 보존 분야의 연구 주제와 범위는 매우 다양하고 광범위하다.

본 연구에서는 이와 같이 광범위하게 수행되어 온 디지털 보존의 주제 영역 중 그간 중점적으로 수행되어 온 세부적인 핵심 지식영역이 어떠한 분야인지를 계량정보학적인 연구방법을 통하여 확인해 보고자 한다.



〈그림 1〉 디지털 보존 인접 주제영역

1. 2 연구의 방법 및 범위

디지털 보존 영역의 핵심적인 지식영역을 확인하기 위하여 CiteSpace를 활용하였다.

분석 대상 데이터베이스는 Web of Science를 중심으로 하였으며 기간은 1990년부터 2005년까지를 대상으로 하였다.

Web of Science에서 문헌들을 검색하기 위한 키워드는 디지털 보존 프로젝트들을 통하여 추출하였다.

디지털 보존 관련 대표적인 프로젝트 7개를 선정한 후 이 프로젝트들로부터 산출되는 결과물들의 표제와 초록을 분석하여 고빈도 키워드를 추출하였다. 선정된 7개의 프로젝트들은 CAMiLEON, CEDARS, DSPACE, InterPARES, LOCKSS, NEDLIB, PRISM이다.

이 프로젝트들의 산출물 분석을 통하여 최종적으로 추출된 8개의 키워드를 Web of Science에서 키워드로 활용하여 74건의 문헌들을 검색하였다. 이 74건의 문헌들의 서지사항들을 입력 데이터 파일로 활용하여 CiteSpace를 통하여 비주얼맵을 산출하였다.

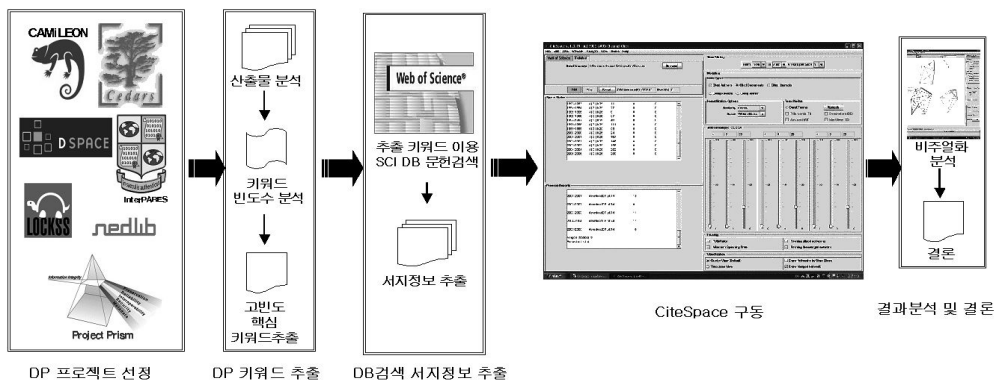
산출된 비주얼맵과 인용히스토리 정보들을 종합하여 결론을 도출하였다. 이상의 연구개요에 대한 내용을 그림으로 나타내면 다음의 <그림 2>와 같다.

2. 지식영역 비주얼화와 CiteSpace

지식영역 비주얼화 (knowledge domain visualization)의 모태는 Kuhn(1962)의 *The Structure of Scientific Revolutions*로부터 비롯되었다.

이 책에서 Kuhn은 과학은 평이한 과학적 특성(normal science)으로부터 복잡한 과학적 특성(scientific crisis)으로 복잡한 과학적 특성으로부터 더욱 혁신적인 과학적 특성(scientific revolution)으로 이동되어 가는 경향이 있음을 주장하였는데, 이 이론이 동시인용 네트워크 분석을 통하여 과학적 연구의 패러다임이 이동하는 것을 감지하고 관찰하는 계량정보학 분야의 연구 확산을 위한 기반이 되었다.

한편 정량적 연구의 대표적인 방법론인 계량



<그림 2> 연구의 개요

정보학 결과의 입체적 표현을 위하여 시각화 기법이 적극적으로 도입되어 오고 있다. 이는 디지털 패러다임에서의 연구 영역들 간의 특성에 기이인한 것으로 짐작된다. 즉, 연구 영역들의 특성들이 매우 동적으로 변화되어 감에 따라 인접 학문분야와의 학제적 (interdisciplinary) 성향이 두드러지고, 이에 따라 새로운 주제들이 빠르게 출현하고 있으며, 그러한 하다면 이전 주제들이 다른 영역과 합쳐지기도 하고, 또는 세분화되기도 한다. 이와 같이 복잡하게 상호 연결된 연구물들간의 관계와 동향을 보다 쉽게 제시하고 이해하고자 하는 의도에서 시각화 표현 기법 연구들이 활성화되고 있는 것으로 추측된다.

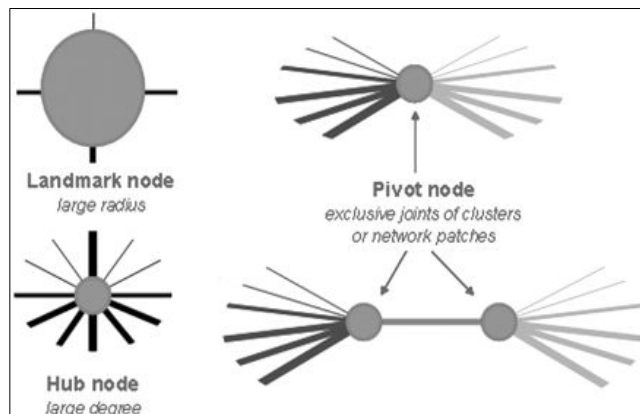
지식영역의 시각화 기법으로는 요인분석(Factor Analysis), 다차원축척기법(Multidimensional Scaling), 패스파인더 네트워크(Pathfinder Network Scaling, SOMs(Self-Organizing Maps) 등이 있다.

또한 지식영역의 시각화 기법은 정보과학 학문분야의 성장하는 연구영역을 구조적으로 매

핑하는 데에만 그 유용성이 있는 것이 아니라 정보 검색과 조직 영역까지도 지원해주는 보조적인 도구가 될 수 있다는 데에 그 의미가 있다 (Börner 2003).

본 연구에서 활용한 CiteSpace는 Drexel 대학 교수인 Chaomei Chen 이 Java 로 개발한 동시인용 분석 도구로서 패스파인더 네트워크 분석, 군집화, hub node와 pivot node 의 식별 등의 기능을 갖고 있다. 여기에서 hub node란 주변 노드들에 폭넓게 인용되는 노드를 의미하며, pivot node는 서로 다른 두 클러스터링을 중간에서 연결해주는 게이트웨이 기능을 하는 노드를 의미한다. 이 외에도 landmark node는 다인용된 노드로서 가장 중점적인 역할을 하는 노드이다(Chen 2004). 이들은 비주얼 맵 상에서 다음 <그림 3>과 같이 표현된다.

CiteSpace를 이용하여서는 여러 가지 관점에서의 인용분석을 시도할 수 있는데, 본 연구에서는 이 중 저자 동시인용분석과 문헌 동시인용분석의 두 가지 요소를 중심으로 분석하였다.



<그림 3> 동시인용 네트워크상에서의 세가지 유형의 주도적 노드들

3. 디지털 보존 영역에의 CiteSpace 적용

CiteSpace 를 이용하여 디지털 보존 지식영역 비주얼화 도출을 시도하였다. 이를 위하여 디지털 보존 영역의 프로젝트들을 일차적으로 분석하였고, 그 다음 단계로 Web of Science 를 통하여 서지정보를 추출하였다. 수집된 서지정보를 CiteSpace 에 입력하여 동시인용저자 및 동시인용문헌을 중심으로 한 디지털 보존영역 비주얼화 결과를 산출하였다.

3.1 데이터 수집

계량정보학 연구에 있어서 무엇보다 중요한 것은 분석 대상이 되는 데이터범위를 설정하는 일이다. 테스트 컬렉션의 범주와 성격에 따라서 결과가 상이하게 달라지기 때문이다.

본 연구에서는 디지털 보존 프로젝트들을 1차적인 데이터 수집 대상으로 선정하였다. 일반적으로는 계량분석 연구를 위한 데이터 대상으로 전문학술지를 선정한다. 그러나 디지털 보존 영역 관련 최신 연구들은 전문 학술지들을 중심으로 하는 개인연구가 아닌 국가기관이나 국제기구가 주도하는 프로젝트들을 중심으로 이루어지고 있다. 이렇게 프로젝트들을 중심으로 수행된 연구들은 결과보고서로 발표되고, 디지털 보존 분야 연구자들의 다양한 연구물들(프로젝트 결과를 기반으로 한 연구들을 포함한)은 다시 학술 데이터베이스에 수록되는 양상을 나타낸다.

디지털 보존 관련 연구물들을 학술 데이터베이스를 통하여 검색해 보면 대체적으로 학제적(interdisciplinary) 성향의 학회지 등에 다양하게 수록되고 있는 것으로 나타난다. 디지털 보존 프로젝트들을 중심으로 수행된 연구 결과물들도 정보학, 컴퓨터과학 등의 학제적 성향의 학회지들을 중심으로 출현하고 있다.

이에 따라 본 연구에서는 디지털 보존 분야에 있어서의 7개의 대표적인 프로젝트들을 먼저 선정하였다. 프로젝트 선정에 있어서는 연구주제 분야, 연구 산출물 건수, 프로젝트 규모 등을 고려하였으며, 전문가와의 면담을 참조로 하여 수행하였다.¹⁾

선정된 7개의 프로젝트들을 중심으로 연구 결과물들을 수집하였다. 프로젝트들에서 최종적으로 산출되는 연구결과물들은 학술논문과 같이 일관된 형식이 아니며, 각각의 특성에 따라 형태와 내용의 차이가 폭이 매우 크다.

따라서 최소한의 공통된 요소를 중심으로 결과물들을 분석하였다. 연구결과물들의 표제와 초록으로부터 디지털 보존 영역의 핵심 키워드들을 수작업으로 추출하였다.

그 다음 단계로 각 프로젝트에서 추출된 핵심 키워드들의 빈도수를 산출하였다. 각각의 개별 프로젝트들을 중심으로 고빈도 키워드들을 산출한 후에 7개의 프로젝트 전체를 중심으로 공통적으로 출현하는 8개의 고빈도 키워드들을 최종 산출하였다.

선정된 프로젝트 내용 및 최종 선정 키워드들에 관한 내용을 도표로 정리하면 다음 <표 1>과 <표 2>와 같다.

1) 프로젝트 선정은 University of North Carolina at Chapel Hill 문헌정보학과 교수(기록관리학 전공)인 Dr. Helen Tibbo와의 면담(2005. 9. 29) 내용을 참조하였다.

〈표 1〉 데이터 수집을 위하여 선정한 디지털 보존 프로젝트들

프로젝트명	결과물 건수	추출 keyword 수	3회이상 출현 keyword 수	고빈도 출현 keyword
CAMILEON	9	27	8	digital preservation, migration
CEDARS	14	19	3	digital preservation, long term preservation
Dspace	11	27	7	digital repository, open sources
InterPARES	18	12	4	electronic record, digital preservation
LOCKSS	16	34	10	digital preservation, access
NEDLIB	10	37	5	library, long term preservation
PRISM	8	24	1	digital preservation, archiving

〈표 2〉 디지털보존 프로젝트 핵심 keywords

최종선정 keywords	어형통계
authenticity	authentication 포함
digital preservation	-
digital repository	digital repositories 포함
electronic record	electronic records 포함
emulation	-
long term preservation	long-term preservation 포함
migration	-
preservation strategy	preservation strategies 포함

〈표 3〉 Web of Science 탐색 범주 및 검색 문헌

포함 DB	기간	탐색범위	최종문헌
Science Citation Index Expanded(SCI-Expanded)	1990~2005	선정된 8개 최종키워드들의 TS(Topic) 와 TI(Title) field tag 포함여부	74건
Social Sciences Citation Index(SSCI)			
Arts & Humanities Citation Index(A&HCI)			

선정된 8개의 키워드들을 이용하여 Web of Science DB를 중심으로 관련 문헌들을 검색하였다. Web of Science DB의 Citation DB 중 SCI, SSCI, A&HCI의 DB를 모두 대상으로 하였으며, 기간은 1990년부터 2005년까지를 대상으로 하였다.

8개의 키워드를 Web of Science DB에서 제공하는 TS(Topic) 태그(field tag) 와 TI(Title) 태그에 포함되는 문헌들을 중심으로 탐색한 결과 최종 74건의 문헌을 검색하였다

Web of Science DB에서의 탐색 범주 및 최종 검색문헌에 대한 내용을 정리하면 다음 〈표 3〉과

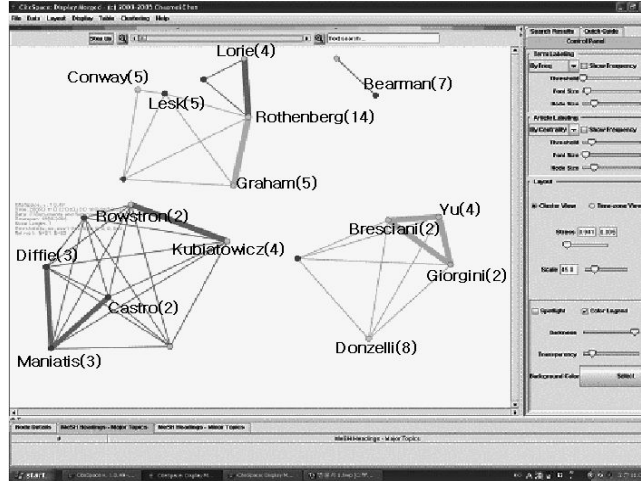
같다.

3. 2 저자 동시인용분석

Web of Science를 검색하여 추출한 최종 74건 문헌들의 서지정보를 CiteSpace를 통하여 분석한 결과 저자 동시인용 관련 결과가 다음의 〈그림 4〉와 같이 나타났다.

그림에서의 노드들은 각 저자를 나타내며 괄호 안의 수치는 인용횟수를 의미한다. 또한 연결선의 두께는 인접 저자와의 연관성 정도를 나타낸다.

CiteSpace를 통한 저자 동시인용분석 비주



〈그림 4〉 저자 동시인용분석을 통한 지식영역 비주얼화

얼화 결과에 인용빈도수가 5회 이상인 경우와 연결선이 굵은 노드를 중심으로 노드명을 명시하였다.

결과를 보면, 저자들 중 Rothenberg는 빈도수와 연관성 결합도가 모두 높은 수치로 나타났다. 또한 Graham도 빈도수와 연관성이 모두 높은 편이다.

최종 74건의 문헌들을 중심으로 한 서지사항들을 입력 데이터파일로 변환하여 CiteSpace를 구동하였다.

한편, (Donzelli, Bearman, Conway, Lesk)의 경우에는 빈도수는 높은 편이지만 타저자와의 연관성은 약한 편으로 나타나고 있다.

반면(Lorie, Yu, Bresciani, Giorgini, Maniatis, Diffie, Castro, Rowstron, Kubiawicz)의 경우에는 빈도수는 적지만 상호 연관성이 강하여 굵은 선으로 네트워크화 되어 표현되고 있다.

Rothenberg의 경우에는 land mark의 노드인 동시에 서로 다른 색상의 Lorie와 Graham을 연결함으로써 pivot node의 역할도 수행하

고 있는 것으로 보인다.

Maniatis가 속한 클러스터에는 다수의 hub node들이 나타나고 있다.

〈그림 4〉의 비주얼화 결과에서는 빈도수와 상호연관성이라는 두 가지 요소가 함께 표현되고 있으므로, 보다 명료하게 분석하고 정리하기 위하여 두 가지 요인을 나누어서 살펴보았다.

3. 2. 1 빈도수 중심 분석

5회 이상의 인용빈도를 나타내고 있는 저자는 모두 7명으로 나타나고 있다. 이를 도표로 정리하면 다음 〈표 4〉와 같다.

〈표 4〉 5회 이상 인용 저자

저자명	빈도수
Rothenberg	14
Donzelli	8
Bearman	7
Conway	5
Graham	5
Lesk	5

〈표 5〉 각 저자의 인용된 문헌 중 고빈도 인용문헌

저자	인용된 문헌	인용횟수
Rothenberg	“Ensuring The Longevity of Digital Documents” (<i>Scientific American</i> 272(1):42-47 Jan 1995)	9
	<i>Avoiding Techonological Quicksand</i> (Washington, DC: Council on Library and Information Resources, c1999)	9
Bearman	“Reality and Chimeras in the Preservation of Electronic Records” (<i>D-Lib Magazine</i> 5(4) Apr 1999)	3
Graham	“Long-term Intellectual Preservation” In <i>Digital imaging technology for preservation</i> , (ed. N.E. Elkington, Mountain View, Calif.:Research Libraries Group, pp.41-57, 1994)	2
Lesk	<i>Preservation of New Technology</i> (Report of the commission on Preservation and Access Oct 1992)	4

그런데 CiteSpace에서 비주얼화 도출 과정에서 분석된 인용문헌 히스토리를 점검한 결과, Donzelli와 Conway의 인용건수에는 자기인용건수가 각각 4회와 1회 포함되어 있었다.

자기인용 건수를 제외하면 기준으로 설정한 5회 동시인용빈도 기준에 미달하므로 두 저자를 제외하였다. 따라서 높은 빈도로 인용된 저자들은 (Rothenberg, Bearman, Graham, Lesk) 등의 4명으로 정리할 수 있다.

각 저자를 동시인용한 문헌들을 조사한 결과, Rothenberg의 경우에는 14번의 인용횟수 대부분이 2편의 논문들에 집중되어 있는 것으로 나타났다. 한편, Bearman의 경우에는 Bearman을 인용한 저자들의 인용문헌들에는 총 9편이라는 많은 논문들이 출현하고 있으나 이 중 한 편만이 3번 동시인용되고 있다. Graham의 경우에는 5편이 인용되었으나 이 중 한 편만이 2번 동시인용되었다. Lesk의 경우에는 인용된 문헌은 3편이었으나 이 중 한 편이 4번 동시인용됨으로써 동시인용빈도가 높은 것으로 나타났다. 2번 이상 동시인용된 문헌들의 내용은 다음 〈표 5〉와 같다.

3. 2. 2 연관성 중심 분석

연관성을 중심으로 네트워크화된 저자들을 영역별로 그룹화하여 도표로 정리하면 다음 〈표 6〉과 같다. 빈도수를 함께 제시하였다.

〈표 6〉 상호연관성을 중심으로 한 저자 그룹

그룹	빈도수	빈도수
1	Rothenberg	14
	Graham	5
	Lorie	4
2	Yu	4
	Bresciani	2
	Giorgini	2
3	Maniatis	3
	Diffie	3
	Castro	2
4	Rowstron	2
	Kubiatowicz	2

〈그림 4〉에서는 그룹 1의 Rothenberg와 Graham은 녹색으로 연결이 되어있으며, Rothenberg와 Lorie는 적색으로 연결이 되어있다 연결하고 있는 색상은 인용문헌의 시계열적 특성을 반영한다.

즉, 녹색으로 갈수록 이전 연도의 문헌들을 나타내고 적색으로 갈수록 최근연도를 반영한

다. Rothenberg는 Graham과 Lorie의 양자를 모두 연결하고 있으나 Graham과 Lorie는 상호 간의 직접적인 연결관계가 없다.

이 외의 그룹 2, 그룹 3, 그룹 4의 저자 네트워크들은 모두 동일 색상의 연결선으로서 동시대의 동시인용으로서 연결성을 갖는다. 각 저자 네트워크를 연결하고 있는 동시인용 문헌들을 <표 7>에서 정리하였다.

표를 통하여서 확인해 보면, 그룹 1의 Rothenberg와 Graham이 4개의 문헌으로 연결되어 있으므로 연결네트워크가 가장 강한 것으로 나타나고 있다. 이 외의 모든 저자 쌍들은 각각 2건의 문헌들로만 상호 연결되고 있다.

3. 3 문헌 동시인용분석

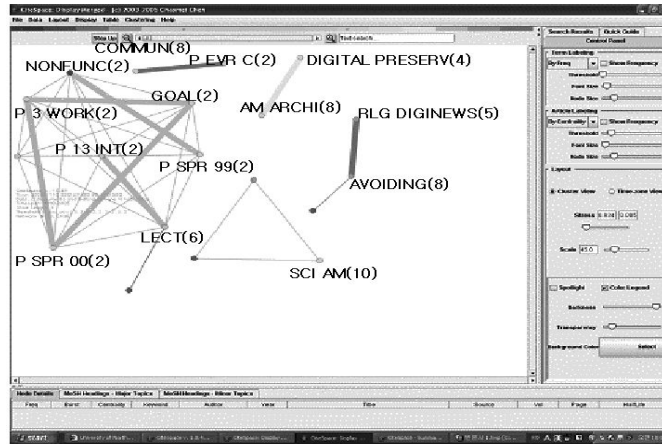
CiteSpace 를 통하여 나타난 문헌 동시인용 분석 결과는 다음 <그림 5>와 같다.

각 노드는 문헌을 대표하며, 괄호안의 숫자는 동시인용된 횟수를 나타낸다. 연결선의 두께는 문헌 간의 연관성을 나타내며 연결선의 색상은 시계열적인 특성을 반영하는데, 색상의 스펙트럼에 따라서 녹색과 청색으로 갈수록 과거의 데이터를 의미하고, 황색에서 주황색, 적색으로 갈수록 최근의 데이터를 의미한다.

마찬가지로 그림에서는 네트워크와 동시인용횟수를 모두 나타내고 있으므로 각각의 요소를 중심으로 정리해 보았다.

<표 7> 네트워크 형성 저자간 동시인용된 문헌 리스트

그룹	저자간 관계	동시인용문헌
1	Lorie ↔ Rothenberg (네트워크 색상 : 적색)	1. Gladney HM, 2005. "Trustworthy 100-year Digital Objects: Durable Encoding for When it's Too Late To Ask." <i>ACM Transactions on Information Systems</i> , 23(3): 299-324. 2. van der Hoeven JR, 2005. "Development of a Universal Virtual Computer (UVC) for Long-term Preservation of Digital Objects." <i>Journal of Information Science</i> , 31(3): 196-208.
	Rothenberg ↔ Graham (네트워크 색상 : 녹색)	3. Day MW, 1998. "Online Serials: Preservation Issues." <i>Serials Librarian</i> , 33(3-4): 199-221. 4. Hedstrom M, 1997. "Digital Preservation: A Time Bomb for Digital Libraries." <i>Computers and the Humanities</i> , 31(3): 189-202. 5. Kranch DA, 1998. "Beyond Migration: Preserving Electronic Documents with Digital Tablets." <i>Information Technology and Libraries</i> , 17(3): 138-148. 6. Lee KH, 2002. "The State of the Art and Practice in Digital Preservation." <i>Journal of Research of the National Institute of Standards and Technology</i> , 107(1): 93-106.
2	Yu ↔ Bresciani ↔ Giorgini (네트워크 색상 : 주황)	7. Bresciani P, 2003. "Requirements Engineering for Knowledge Management in eGovernment." <i>Lecture Notes in Artificial Intelligence</i> , 2645: 48-59. 8. Donzelli P, 2003. "Goal-oriented Requirements Engineering: A Case Study in e-Government." <i>Lecture Notes in Computer Science</i> , 2681: 601-616.
3	Maniatis ↔ Diffie ↔ Castro (네트워크 색상 : 주황)	9. Cooper BF, 2005. "Authenticity and Availability in PIPE Networks." <i>Future Generation Computer Systems</i> , 21(3): 391-400.
4	Rowstron ↔ Kubiatowicz (네트워크 색상 : 주황)	10. Maniatis P, 2005. "The LOCKSS Peer-to-peer Digital Preservation System.." <i>ACM Transactions on Computer Systems</i> , 23(1): 2-50.



〈그림 5〉 문헌 동시인용분석을 통한 지식영역 비주얼화

3. 3. 1 빈도수 중심 분석

5회 이상의 인용빈도를 나타내고 있는 문헌은 모두 6종으로 나타났다. 문헌명과 인용횟수를 도표로 정리하면 다음 〈표 8〉과 같다.

문헌 중에서는 Scientific American이 가장 동시인용빈도수가 높았다. 그 다음으로는 American Archivist, Avoiding Technological Quicksand, Communications of the ACM이 모두 각각 8회로 높은 동시인용빈도수를 나타내고 있다.

Scientific American 저널은 인용빈도가 높음에도 불구하고 타 저널과의 동시인용 빈도는 낮아서 네트워크를 활발하게 형성되지 않는 것으로 나타나고 있다.

3. 2. 2 연관성 중심 분석

비주얼화 매핑에 나타난 문헌 간 네트워크를 중심으로 한 내용을 다음의 〈표 9〉에 정리하였다. 각 네트워크화된 문헌들을 단위로 하여 모두 여섯 개의 그룹으로 정리되었다.

그런데, 그룹 1의 DIGITAL PRESERV의 경우는 Lazinger S, Lee KH, Beagrie N, Marcum 등의 네 저자의 동시인용에 의하여 네트워크화가 형성된 것인데, 중심 문헌들을 분석하는 과정에서 서로 다른 문헌들을 동일 그룹으로 묶은 오류가 확인되었다.

Lazinger S가 인용한 Hedstrom(2003)의 “Digital Preservation”과 Lee KH가 인용한

〈표 8〉 5회 이상 인용 문헌

약어	원 문헌명	빈도수
SCI AM	Scientific American	10
AM ARCHI	American Archivist	8
AVOIDING	Avoiding Technological Quicksand	8
COMMUN	Communications of the ACM	8
LECT	Lecture Notes in Computer Science	6
RLG DIGINEWS	RLG DigiNews	5

〈표 9〉 상호연관성을 중심으로 한 문헌 그룹

그룹	문헌약어	문헌원어	빈도수
1	AM ARCHI	<i>American Archivist</i>	8
	DIGITAL PRESERV	-	4
2	LECT	<i>Lecture Notes in Computer Science</i>	6
	P 13 INT	<i>Proceedings of the Thirteenth International Conference on Software Engineering and Knowledge Engineering</i>	2
3	P SPR 99	<i>Proceedings of the Spring 99 Simulation Interoperability Workshop</i>	2
	NONFUNC	<i>Non-Functional Requirements in Software Engineering</i>	2
4	P SPR 00	<i>Proceedings of the Spring 2000 Simulation Interoperability Workshop</i>	2
	P 3 WORK	<i>Proceedings of 3rd Workshop on Requirements Engineering For Software Quality</i>	2
	GOAL	<i>The Goal Question Metric Approach</i>	2
5	AVOIDING	<i>Avoiding Technological Quicksand</i>	8
	RLG DIGINEWS	<i>RLG DigiNews</i>	5
6	COMMUN	<i>Communications of the ACM</i>	8
	P EUR C	<i>Proceedings of the European Conference on Digital Libraries</i>	2

〈표 10〉 상호연관성을 중심으로 한 문헌 그룹

그룹	문헌간 관계	동시인용문헌
2	LECT ↔ P 13 INT (네트워크 색상: 주황)	1. Bresciani P, 2003. "Requirements Engineering for Knowledge Management in eGovernment." <i>Lecture Notes in Artificial Intelligence</i> , 2645: 48-59. 2. Donzelli P, 2003. "Goal-oriented Requirements Engineering: A Case Study in e-Government." <i>Lecture Notes in Computer Science</i> , 2681: 601-616. 3. Gladney HM, 2005. "Trustworthy 100-year Digital Objects: Durable Encoding for When it's Too Late To Ask." <i>ACM Transactions on Information Systems</i> , 23(3): 299-324. 4. Lee KH, 2002. "The State of the Art and Practice in Digital Preservation." <i>Journal of Research of the National Institute of Standards and Technology</i> , 107(1): 93-106. 5. Muir A, 2004. "Digital Preservation: Awareness, Responsibility and Rights Issues." <i>Journal of Information Science</i> , 30(1): 73-92. 6. van der Hoeven JR, 2005. "Development of a Universal Virtual Computer (UVC) for Long-term Preservation of Digital Objects." <i>Journal of Information Science</i> , 31(3): 196-208.
3	P SPR 99 ↔ NONFUNC (네트워크 색상: 주황)	
4	P SPR 00 ↔ P 3 WORK ↔ GOAL (네트워크 색상: 주황)	
5	AVOIDING ↔ RLG DIGINEWS (네트워크 색상: 적색)	
6	COMMUN ↔ P EUR C (네트워크 색상: 적색)	7. Cooper BF, 2005. "Authenticity and Availability in PIPE Networks." <i>Future Generation Computer Systems</i> , 21(3): 391-400. 8. Gladney HM, 2005. "Trustworthy 100-year Digital Objects: Durable Encoding for When it's Too Late To Ask." <i>ACM Transactions on Information Systems</i> , 23(3): 299-324.

Fresko(1998)의 "Digital Preservation Guidelines" 등을 동일문헌 "Digital Preservation"으로 설정하고 있는 것이다. 이는 CiteSpace의 오류가 아닌 Web of Science에서의 서지데이터 생성 과정에서의 오류로 생각된다.

이에 따라서 그룹 1을 제외하고 그룹 2에서부터 그룹 6까지를 동시인용된 문헌들을 중심으로 〈표 10〉에서 정리하였다.

〈표 10〉을 살펴보면 대부분의 그룹들이 두 저자에 의하여 동시인용된 것을 확인할 수

있다.

그룹 5의 AVOIDING과 RLG DIGINEWS는 네 명의 저자들에 의하여 동시인용됨으로서 타 문헌들 보다 상대적으로 더 활발히 인용되고 있는 것으로 나타난다.

4. 디지털 보존 지식영역 비주얼화 결과 분석

CiteSpace를 통하여 저자 동시인용분석과 문헌 동시인용분석을 수행하였다. 비주얼 맵을 통하여 인용빈도가 높은 노드와 상호 관계가 높은 노드들을 확인할 수 있었다.

먼저 저자 동시인용분석의 결과를 통해서

Rothenberg와 Graham이 높은 빈도를 가진 저자들로 확인되었으며, 동시에 상호 관계도 높은 저자들로 확인되었다.

문헌 동시인용분석의 결과를 통해서 AVOIDING과 RLG DIGINEWS 두 문헌들이 인용 빈도도 높고 상호 연관성도 높은 것으로 나타났다.

이상의 저자 동시인용분석과 문헌 동시인용 분석 결과를 중심으로 높은 수치를 나타낸 노드들을 대상으로 내용을 정리하면 다음 <표 11>과 같다.

최종결과로 산출한 두 개의 디지털 보존 지식 영역 비주얼 맵들에 이상의 분석 내용들을 반영하여 클러스터링화 하였다.

인용빈도와 네트워크 정도, 구성노드수, 시

<표 11> 디지털 보존 지식영역 비주얼 결과 중 고빈도 데이터

빈도/결합도 높은 노드	고빈도 인용문헌 및 동시인용 문헌
Rothenberg	"Ensuring The Longevity of Digital Documents" (<i>Scientific American</i> 272(1):42-47 Jan 1995) <i>Avoiding Techonological Quicksand</i> (Washington, DC: Council on Library and Information Resources, c1999)
Graham	"Long-term Intellectual Preservation" In <i>Digital imaging technology for preservation</i> ,(ed. N.E. Elkington, Mountain View, Calif.:Research Libraries Group, pp.41-57. 1994)
Rothenberg ↔ Graham	Day MW, 1998. "Online Serials: Preservation Issues." <i>Serials Librarian</i> , 33(3-4): 199-221. Hedstrom M, 1997. "Digital Preservation: A Time Bomb for Digital Libraries." <i>Computers and the Humanities</i> , 31(3): 189-202. Kranck DA, 1998. "Beyond Migration: Preserving Electronic Documents with Digital Tablets." <i>Information Technology and Libraries</i> , 17(3): 138-148. Lee KH, 2002. "The State of the Art and Practice in Digital Preservation." <i>Journal of Research of the National Institute of Standards and Technology</i> , 107(1): 93-106.
AVOIDING ↔ RLG DIGINEWS	Gladney HM, 2005. "Trustworthy 100-year Digital Objects: Durable Encoding for When it's Too Late To Ask." <i>ACM Transactions on Information Systems</i> , 23(3): 299-324. Lee KH, 2002. "The State of the Art and Practice in Digital Preservation." <i>Journal of Research of the National Institute of Standards and Technology</i> , 107(1): 93-106. Muir A, 2004. "Digital Preservation: Awareness, Responsibility and Rights Issues." <i>Journal of Information Science</i> , 30(1): 73-92. van der Hoeven JR, 2005. "Development of a Universal Virtual Computer(UVC) for Long-term Preservation of Digital Objects." <i>Journal of Information Science</i> , 31(3): 196-208.
-	총 10 건

계열 색상 등을 고려하여 보다 핵심적이고 영향력을 가질 수 있는 클러스터를 중심으로 순위를 부여하였다.

또한, 각각의 클러스터링을 중심으로 <표 11>에서 제시한 배경 문헌 및 각 클러스터를 이루고 있는 문헌들의 내용, 해당 문헌들과 관련 DB에서 제공하는 키워드 등을 참조하여 클러스터링 이름을 부여하였다.

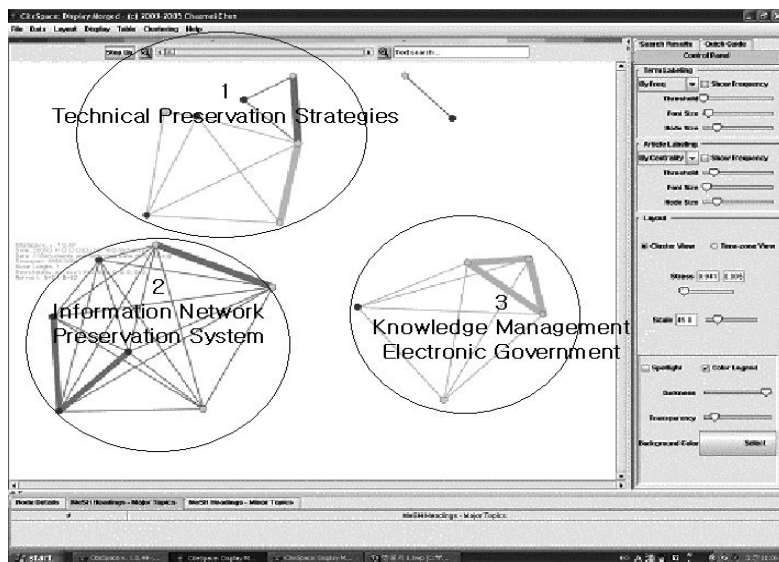
다음 <그림 6>은 <그림 4>의 저자 동시인용 분석을 중심으로 한 비주얼 맵에 순위와 클러스터링 이름을 부여한 것이다. 저자 동시인용 분석 비주얼맵을 순위화한 결과, 정보기술적인 디지털 보존기술에 대한 영역이 가장 중점적인 위치를 차지하는 것으로 나타났다. 그 다음으로는 디지털 보존과 연관된 정보 네트워크와 보존 시스템을 중심으로 하는 영역이 주요한 위치를 차지하고 있다. 그 다음 순위는 전자정부와 지식관리에 관한 영역인 것으로 나타났다.

즉 활발한 hub node 중심의 클러스터 2 보다는 land mark 및 pivot node 중심의 클러스터 1이 더 핵심적인 영역으로 나타났다.

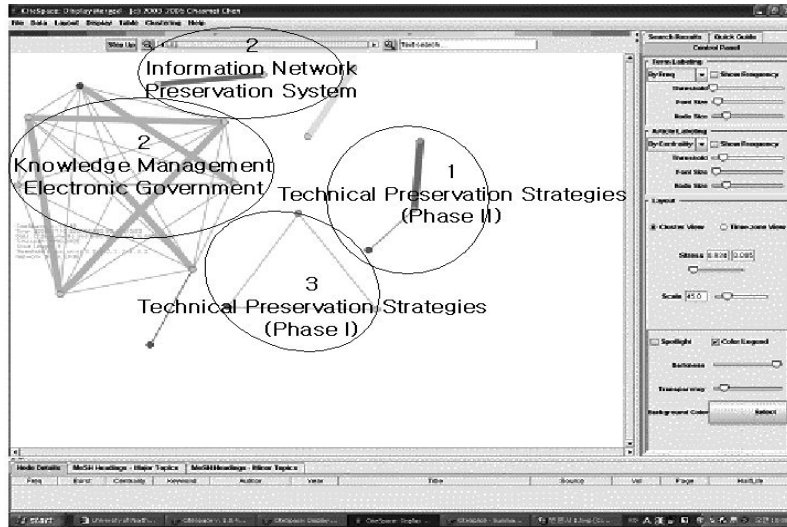
마찬가지로 <그림 7>은 <그림 5>의 문헌 동시인용분석을 중심으로 한 비주얼 맵에 순위와 클러스터링 이름을 부여한 것이다.

문헌 동시인용 분석 비주얼맵을 순위화한 결과, 두 클러스터링은 그 규모와 인용정도가 비슷하여 동일 순위를 부여하였다.

정보기술 중심의 디지털 보존기술에 대한 영역이 1순위를 차지하였고, 정보 네트워크와 보존 시스템 영역, 그리고 지식관리와 전자정부 영역이 2순위로 나타났다. 3순위의 클러스터링은 1순위의 클러스터링 명과 동일하게 기술적 보존기술로 나타나고 있는데 여기에서의 디지털 보존기술은 1990년대 후반에 주요 연구된 초창기 디지털 보존기술을 의미하는 것이다. 즉, 메체전환, 마이그레이션, 에뮬레이션, 인캡



<그림 6> 순위와 클러스터명을 부여한 저자 동시인용분석 비주얼맵



〈그림 7〉 순위와 클러스터명을 부여한 문헌 동시인용분석 비주얼맵

슬레이션 등의 전통적인 디지털 보존 전략에 대한 연구들이 중심이 된 영역이다.

반면, 1순위의 디지털 보존 전략 영역은 2000년 이후 최근의 연구흐름을 반영한다. Gladney (2005)와 van der Hoeven(2005)에 따르면 디지털 객체 및 매체의 보존 전략에만 중점을 두었던 이전과 달리 최근에는 디지털 객체를 둘러싼 운영체제 및 시스템, 경영적인 측면에까지 확대되어 보존 전략에 대한 논의가 진행되는 추세이다. 이와 관련하여 PIPE(Peer to peer Information Preservation and Exchange Network) 등의 새로운 보존 전략 지식 영역들도 출현하고 있다. 이와 같은 배경을 반영하여 3순위의 보존전략은 1세대(Phase I) 보존전략으로, 1순위의 보존전략은 2세대(Phase II) 보존전략으로 구분하여 클러스터링 명을 부여하였다.

이상을 종합하여, 디지털 보존 영역에서의 중점 지식영역은 최신정보기술을 중심으로 한 디지털 보존전략, 정보네트워크와 보존시스템,

전자정부와 지식관리 등의 세 영역이 중점적인 지식 영역인 것으로 비주얼맵을 통하여 확인할 수 있었다.

5. 결론

CiteSpace를 이용하여 디지털 보존 영역에서의 핵심 지식영역을 비주얼 맵을 통하여 확인하였다. 이를 위하여서는 1차적으로 7개의 디지털 보존 프로젝트들의 결과물들을 중심으로 추출된 8개 키워드를 활용하여 Web of Science를 검색하였으며, 검색결과 산출된 74건의 문헌들의 서지정보들을 CiteSpace 구동을 위한 입력 데이터로 활용하였다.

실험 결과 저자 동시인용 분석 및 문헌 동시인용 분석 비주얼 맵이 도출되었으며, 각 노드간 네트워크를 형성하는 인용 히스토리를 분석한 결과, 디지털 보존 영역에서는 최신정보기술을

중심으로 한 디지털 보존전략, 정보네트워크와 보존시스템, 전자정부와 지식관리 등의 세 영역이 핵심적인 지식 영역인 것으로 확인되었다.

그런데, 이 지식 영역 중 전자정부와 지식관리는 CiteSpace 결과에서 노랑색으로 그 네트워크가 표현되었고, 정보네트워크와 보존시스템은 적색으로, 최신정보기술을 중심으로 한 디지털 보존전략은 녹색과 적색으로 표현되고 있으므로, 전자정부와 지식관리의 주제와 함께 최근에는 최신정보기술을 중심으로 한 디지털 보존전략과 정보네트워크와 보존시스템에 관한 연구가 중점수행되고 있는 것으로 파악된다. 특히, 최신정보기술을 중심으로 한 디지털 보존전략과 정보네트워크는 녹색과 적색 두 가지 색상으로 네트워크가 표현되는 만큼 오랜 시간의 폭 동안 관심 있는 분야로 연구가 수행되어 오고 있음을 확인할 수 있다.

본 연구에서는 인용분석 기법을 기반으로 하는 지식영역 비주얼 툴을 활용한 만큼, 인용결

과가 어느 정도 누적되어 있는, 1990년대 후반부터 2000년도 초반에 종료된 디지털 보존 분야의 대표적인 7개 프로젝트들을 1차적인 데이터로 확보하여 연구를 수행하였다.

그러나, 2005년 현재 기록의 장기적 보존과 관련하여 관심을 끌고 있는 중요한 프로젝트로는 미국 국립기록청(NARA)의 ERA(Electronic Records Archive) 프로젝트 영국 국립기록보존소(TNA: The National Archives)의 Seamless Workflow Programme, 호주 빅토리아주 기록보존소(PROV: Public Records Office Victoris)의 VERS(Victoria Electronic Records Strategy), 네덜란드의 국립기록보존소가 주축이 되어 진행 중인 DPT(Digital Preservation Testbed) 등을 간과할 수 없다. 이상의 최신 프로젝트를 중심으로 한 동향분석을 위하여서는 인용분석 기법이 아닌 다른 연구 기법이 적용되어야 할 것이다. 이와 관련된 연구는 다음 과제로 수행해 나가고자 한다.

참 고 문 헌

- Börner, K. et al, 2003. "Visualizing Knowledge Domain." *Annual Review of Information Science & Technology*, 37.
- CAMiLEON <http://www.si.umich.edu/CAMiLEON/>
- CEDARS <http://www.leeds.ac.uk/cedars>
- Chen, C. 2005. "Measuring the Movement of a Research Paradigm." *Proceeding of SPIE-IS&T: Visualization and Data Analysis 2005*, 5669: 63-76.
- Chen, C. 2004. "Searching for Intellectual Turning Points: Progressive Knowledge Domain Visualization." *Proceedings of the National Academy of Sciences of the United States of America* 2004, 101: 5303-5310.
- CiteSpace. <http://cluster.cis.drexel.edu/~chen/citespace/>

- Cox, R. J. 2000. "Searching for Authority: Archivists and Electronic Records in the New World at the Fin-de-Siecle." *First Monday*, 5(1).
- Ding, Y., G. G. Chowdhury, et al. 2001. "Bibliometric cartography of information retrieval research by using co-word analysis." *Information Processing & Management*, 37(6): 817-842.
- DSPACE <http://www.dspace.org/>
- Fresko, M. and K. Tombs, 1998. *Digital Preservation Guidelines* European Commission DG XIII/E, Luxembourg.
- Gilliland-Swetland, A. J. 1992. "Archivy and the computer: a citation analysis of North American archival periodical literature." *Archival Issues*, 17(2): 95-112.
- Gladney HM, 2005. "Trustworthy 100-year Digital Objects: Durable Encoding for When it's Too Late To Ask." *ACM Transactions on Information Systems*, 23(3): 299-324.
- Hedstrom, M. 1997. "Digital Preservation: A Timebomb for Digital Libraries." *Computers and the Humanities*, 31(3): 189-202.
- InterPARES <http://www.interpares.org/>
- Kuhn, T. S. 1962. *The Structure of Scientific Revolutions*. University of Chicago press.
- Lee KH, 2002. "The State of the Art and Practice in Digital Preservation." *Journal of Research of the National Institute of Standards and Technology*, 107(1): 93-106.
- LOCKSS <http://lockss.stanford.edu/>
- NEDLIB <http://www.kb.nl/coop/nedlib/>
- PRISM <http://prism.cornell.edu/>
- Noyons, E. C. M. and A. F. J. Van Raan. 1998. "Advanced mapping of science and technology." *Scientometrics*, 41(1-2): 61-67.
- van der Hoeven JR, 2005. "Development of a Universal Virtual Computer(UVC) for Long-term Preservation of Digital Objects." *Journal of Information Science*, 31(3): 196-208.
- White, H. D. 2003. "Author Cocitation Analysis and Pearson's r." *Journal of American Society of for Information Science and Technology*, 54(13): 1250-1259.