

동시링크를 이용한 사회학 분야 웹 정보원의 지적구조 분석:

A Study on the Intellectual Structure in Web Information of Sociology Using the Co-links Analysis

김원진, 연세대학교 대학원 문헌정보학과, wonk2005@yonsei.ac.kr

Won-Jin Kim, Dept. of Library and Information Science, Graduate School of Yonsei University

본 연구에서는 사회학 분야 웹 정보원을 대상으로 동시링크분석을 실시하여 특정 학문분야의 지적구조를 분석하고, 검색엔진별 지적 구조의 차이를 분석함으로써 웹정보원의 특성을 살펴보았다. 세 개의 검색엔진을 대상으로 지적구조의 차이를 비교한 결과, 웹 정보원 지도에서 전체적인 지적구조는 비슷하게 나타났지만, 몇 개의 웹 정보원의 경우에는 소속군집이 다르게 나타난 경우도 있었다. 그리고 AltaVista와 AlltheWeb은 지도상의 웹 정보원 분포에 있어서 거의 유사한 구조를 보여주었다.

1. 서론

1.1 연구의 목적

웹의 힘의 원천은 링크, 즉 한 페이지에서 다른 페이지로 이동할 수 있도록 해주는 URL에 있다. 그것들은 우리가 정보들을 찾고, 옮겨 다니고, 하나로 엮을 수 있도록 해 준다. 이 링크들은 개별적인 문서들의 단순한 집합을 하나의 거대한 네트워크로 만들어 준다. 인터넷은 서로간의 링크들을 가진 많은 수의 웹 페이지들로 구성된다.

최근 인터넷의 급속한 성장으로 계량정보학의 연구대상이 웹 정보원으로 확산되면서, "Cybermetrics" 또는 "Webometrics"라는 용어가 등장하였다. 웹 정보원이 계량정보학의 연구방법에 있어서 적용될 수 있는 것은 웹 정보원의 '하이퍼링크'라는 개념을 인쇄문헌의 인용에 적용시킬 수 있기 때문이다. 즉, 사이트들과 페이지들 간의 웹 링크는 출판된 문헌들에서의 인용과 비슷하다고 할 수 있다. 웹을

통한 정보원은 크기와 범위면에서 높은 평가를 받고 있으며, 이용하는데 있어서도 편리성과 다양성을 인정받고 있다.

본 연구는 동시링크를 이용한 커뮤니케이션 분야 웹 정보원의 지적구조를 분석한 이전 연구에 대하여 다른 분야에 있어서도 웹 정보원의 적용가능여부를 살펴보고자 하였으며, 여러 가지 검색엔진별 차이를 분석함으로써 웹 정보원의 특성을 연구하고자 하였다.

1.2 연구의 범위와 방법

인쇄문헌이 출판된 데이터를 이용하여 계량정보학적 분석을 하는 것과는 달리, 웹 정보원을 이용한 계량정보학적 분석에서는 직접 웹 정보원을 검색하여 데이터를 수집한다. 일반적으로, 계량정보학에서는 ISI의 인용 데이터베이스를 이용하여 연구를 수행하지만, 웹 정보원을 이용하기 위해서는 Altavista와 같은 검색엔진을 이용하여 연구를 수행할 수 있다.

본 연구에서는 웹 정보원에 대한 계량정보

학적 분석을 위하여 사회학 분야를 선정하였다. 실험 대상 선정 방법은 사회과학 정보 게이트웨이인 SOSIG(Social Science Information Gateway)(<http://www.sosig.ac.uk>) 웹 사이트를 방문하여 사회학 분야 1위부터 50위(2006년도 5월 11일-6월 11일)까지의 웹 정보원을 조사하여, 각 정보원의 링크 수, 동시링크 빈도를 고려하여 27개의 웹 정보원을 최종적으로 선정하였다.

본 연구에서는 사회학 분야의 지적구조를 분석함과 동시에 검색엔진간의 차이를 분석하기 위하여, 선정된 웹 정보원의 동시링크빈도를 AltaVista, AlltheWeb, MSN에서 각각 조사하여 3개의 동시링크 행렬을 작성하였다. 동시링크 행렬은 SPSSWIN 12.0을 사용하여 상관관계 행렬을 도출하였으며, 구체적인 상관관계 분석을 위하여 다차원축적과 군집분석을 수행하였다.

2. 이론적 배경

Mckiernan(1996)은 인터넷 웹 사이트들 안에서 인용 현상을 설명하는데 있어서 "citation"이라는 용어를 최초로 사용하였다. Ingwersen(1998)은 도메인이나 웹사이트의 전체 온라인 임팩트(overall online impact)를 측정하기 위하여 WIF의 개념을 제시하였다. Rousseau(1997)는 Lotka 법칙을 이용하여 웹 사이트 상에서 도메인 이름들과 링크들의 분포에 대한 설명을 가능하게 하였다. 즉 인터넷 사이트들 간의 링크수를 측정하기 위하여 웹 사이트 상에서 도메인의 수와 도메인별 인용의 수를 조사한 결과, 웹 문헌의 분포가 Lotka 법칙과 일치함을 논증하였다. Ingwersen(1998)은 저널의 영향력 평가를 위하여 ISI에서 실시하는 IF (Impact Factor)가 웹 정보원에서 적용가능한지를 조사하였으며,

대규모 웹 사이트에 대한 IF가 소규모 웹사이트의 경우보다 신뢰성이 높으며, 인쇄매체보다 역사가 짧고 한 잡지당 기사수가 적은 전자정보원에 더 적합한 것으로 발표되었다.

국내 연구로는 정동열과 최윤미(1999)가 커뮤니케이션 분야를 대상으로 웹 정보원에 있어서 동시인용분석 기법을 적용하여 커뮤니케이션의 지적구조를 파악하였다. 이성숙(2005)은 동시링크 분석을 이용하여 6년간의 시간차에 따른 커뮤니케이션 분야 웹 정보원의 지적구조의 변화를 파악하고, 지적구조의 검색엔진별 차이를 분석함으로써 웹 정보원의 특성을 연구하였다.

3. 웹 정보원의 동시링크분석

3.1 분석대상의 선정

사회학 분야 웹 정보원을 선정하기 위하여 SOSIG(Social Science Information Gateway)에서 제시하는 "Top 50 sites in Sociology"를 우선적으로 선정하였다. 그리고 AltaVista의 "link:URL"이라는 검색식을 이용하여 각 웹사이트의 링크 수를 산출하였다. 2006년 5월11일에서 6월11일 한달 동안 가장 많이 방문한 사회학 분야 사이트 상위 50개 정보원을 대상으로 각 링크수를 조사하여, 링크된 수가 80개 이상인 27개의 웹 정보원을 최종적으로 선정하였다.<표1> 선정된 웹 정보원의 주제와 성격은 SOSIG에서 제시한 주제명과 하위주제명을 주로 참고했으며, Yahoo 디렉토리에서 제시하는 사회학 분야를 참고했다.

검색엔진 별로 웹 정보원의 지적구조를 비교하기 위하여, 현재 웹 정보원들의 동시링크를 분석할 수 있는 검색엔진 중에서 검색결과 면에서 비교적 안정적인 AltaVista, AlltheWeb, MSN을 선정하였다. 27개의 분석 대상 웹 정보

<표 1> 사회학 분야 분석대상 웹 정보원

웹정보원 고유번호	URL	링크수	주제분야	성격
S1	Social Psychology Network http://www.socialpsychology.org/	17,000	Sociology	Resource Guides
S2	Global Site: Critical Gateway to World Politics, Society and Culture http://www.theglobalsite.ac.uk/index.html	1,140	Culture and Identity	Papers/Reports/Articles (collections)
S3	Sociology Central http://www.sociology.org.uk/	702	Sociology	Resource Guides
S4	Sociology of Sport Journal http://www.humankinetics.com/SSJ/journalAbout.cfm	91	Sport	Journals (contents and abstracts)
S5	What is Culture? http://www.wsu.edu:8001/vcwsu/commons/topics/culture/culture-index.html	741	Culture and Identity	Educational Materials
S6	Contemporary Issues in Early Childhood http://www.wwords.co.uk/ciecc/	474	Sport	Journals (full text)
S7	Sociological Research Online http://www.socresonline.org.uk/home.html	306	Sociology	Journals (full text)
S8	Sociology of Knowledge http://www.trinity.edu/~mkearl/knowledg.html	140	Sociology	Educational Materials
S9	Social Problems http://www.ucpress.edu/journals/sp/	147	Social Welfare	Journals (contents and abstracts)
S10	British Journal of Sociology Online http://www.lse.ac.uk/serials/Bjs/HOMEPAGE.HTM	308	Sociology	Journals (contents and abstracts)
S11	Economic Sociology: The European Website http://econsoc.mpifg.de/	164	Sociology	Journals (full text)
S12	Durkheim Pages http://www.relst.uiuc.edu/durkheim/	574	Sociological Theory	Papers/Reports/Articles (collections)
S13	Centre for Research in Ethnic Relations (CRER) http://www.warwick.ac.uk/CRER/	147	Culture and Identity	Research Projects
S14	Sosol - Sociology of Sport Online http://physed.otago.ac.nz/sosol/home.html	207	Sport	Journals (full text)
S15	Gender and Education http://www.tandf.co.uk/journals/titles/09540253.asp	87	Gender and Sexuality	Journals (contents and abstracts)
S16	American Journal of Sociology http://www.journals.uchicago.edu/AJS/home.html	678	Sociology	Journals (contents and abstracts)
S17	Marxism Made Simple http://www.marxismmadesimple.esmartweb.com/mainpage.htm	124	Culture and Identity	Educational Materials
S18	Sociology of Health and Illness http://www.blackwellpublishing.com/journal.asp?ref=0141-9889	85	Social Welfare	Journals (contents and abstracts)
S19	Postmodern Thought http://carbon.cudenver.edu/~mryder/itc_data/postmodern.html	2,260	Sociological Theory	Resource Guides
S20	Dead Sociologists' Society http://www2.pfeiffer.edu/~Iridener/DSS/DEADSOC.HTML	598	Sociological Theory	Resource Guides
S21	Sociological Tour Through Cyberspace (A) http://www.trinity.edu/~mkearl/index.html	1,980	Sociology	Resource Guides
S22	Teaching Sociology http://www.lemoyne.edu/ts/tsmain.html	490	Sociology	Educational Materials
S23	Electronic Journal of Sociology http://www.sociology.org/	2,430	Sociology	Journals (full text)
S24	CyberCulture, Identity and Gender Resources http://fragment.nl/resources/	845	Gender and Sexuality	Bibliographic Databases
S25	Sociology of Knowledge http://carbon.cudenver.edu/~mryder/itc_data/soc_knowledge.html	126	Sociological Theory	Resource Guides
S26	Women's Studies Database http://www.mith2.umd.edu/WomensStudies/	7,520	Sociology	Bibliographic Databases
S27	Social Research Update http://www.soc.surrey.ac.uk/sru/	405	Sociology	Journals (full text)

원을 위의 3가지 검색엔진에서 각각 동시링크 빈도를 조사하여 동시링크빈도행렬을 작성하였다.

3.2 분석대상 웹 정보원들의 동시링크 행렬 생성

최종적으로 선정된 분석대상 27개의 웹 정보원을 대상으로 실험을 실시하였다. 2개의 웹 정보원을 공통으로 링크하고 있는 웹 정보원을 조사하기 위하여 AltaVista, AlltheWeb, MSN 검색엔진에서 다음과 같은 검색식 “link:URL AND link:URL” 을 사용하였다. 이러한 방법으로 웹 정보원들의 쌍을 동시에 링크하고 있는 정보원들의 검색건수를 구하여 동시링크빈도 행렬을 작성하였다. 통계 프로그램의 분석방법으로 동시링크빈도 행렬의 대각선 값을 상위 수치 3개를 합하여 2로 나눈 값으로 하향처리 하였고, 정보원들 간의 상호관련성 존재여부를 조사하기 위하여 웹 정보원의 동시링크빈도를 SPSS 12.0을 이용하여 상관계수행렬로 변환하였다. 웹 정보원들 간의 상관관계와 유사성을 더 자세히 알아보기 위하여 다차원 축적기법과 군집분석을 이용하였다. 다차원축적은 데이터 속에 숨어있는 특정 구조를 파악할 수 있는 기법으로 들쭉 쌍을 이룬 개체간의 유사성을 거리의 개념으로 다차원 공간 속에 보여준다. 군집분석은 모든 분석대상들이 각 하나의 독립된 군집을 이루고 있는 단계로부터 시작하여 서로 유사한 대상들을 묶어감으로 최종적으로는 하나의 군집이 되게 하는 방식이다.

4. 웹 정보원의 지적구조 분석

4.1 웹 정보원의 동시링크빈도

분석대상 27개의 웹 정보원으로부터 각 검색엔진별로 동시링크 행렬이 작성되었다. 가장

높은 동시링크빈도 쌍을 살펴보면, AltaVista의 경우 가장 높은 동시링크빈도를 나타낸 웹 정보원 쌍은 S22번(Teaching Sociology)과 S23번(Electronic Journal of Sociology)으로 220회이다. AlltheWeb의 경우 가장 높은 동시링크빈도를 나타낸 쌍은 S22번(Teaching Sociology)과 S23번(Electronic Journal of Sociology)으로 207회이다. MSN의 경우 가장 높은 동시링크빈도를 나타낸 쌍은 S21번(Sociological Tour Through Cyberspace)과 S23번(Electronic Journal of Sociology)으로 288회이다.

다른 정보원과 가장 많이 동시링크된 정보원을 파악하기 위하여 27개의 정보원과 동시링크된 총 건수를 조사결과, 총 건수가 가장 많은 웹 정보원은 AltaVista의 경우 S23번(Electronic Journal of Sociology)이 1,027회로 가장 많고, 그 다음으로 S21번(Sociological Tour Through Cyberspace)이 932회로 많았고, S1번(Social Psychology Network)이 401회로 나타났다. AlltheWeb의 경우 S23번(Electronic Journal of Sociology)이 1000회로 가장 많았고, S21번(Sociological Tour Through Cyberspace)이 823회, S1번(Social Psychology Network)이 431회 순으로 나타났다. MSN의 경우 S23번(Electronic Journal of Sociology)이 1119회, S21번(Sociological Tour Through Cyberspace)이 1033회로 나타났으며, S20번(Dead Sociologists' Society)이 534회 순으로 나타났다.

반면에 다른 웹 정보원과 가장 적게 동시링크된 정보원은 AltaVista의 경우 S24번(CyberCulture, Identity and Gender Resources)과 S18번(Sociology of Health and Illness)이 각각 28회, 35회였고, AlltheWeb의 경우를 살펴보면, S24번(CyberCulture, Identity and Gender Resources)과 S18번(Sociology of Health and Illness)이 각각 29

회, 35회로 나타났다. MSN의 경우 S24번(CyberCulture, Identity and Gender Resources)과 S11번(Economic Sociology)이 각각 30회, 54회 순으로 나타났다.

각 검색엔진별로 동시링크의 총건수를 살펴보면, AltaVista는 5769회, Alltheweb은 5325회, MSN는 6136회로 나타났다. 세 개의 검색엔진은 검색건수에 있어서 약간의 차이를 나타냈지만(일반적으로 MSN이 가장 높았고, AltaVista, AlltheWeb순으로), 검색결과에 있어서 비슷한 양상을 보여주고 있다.

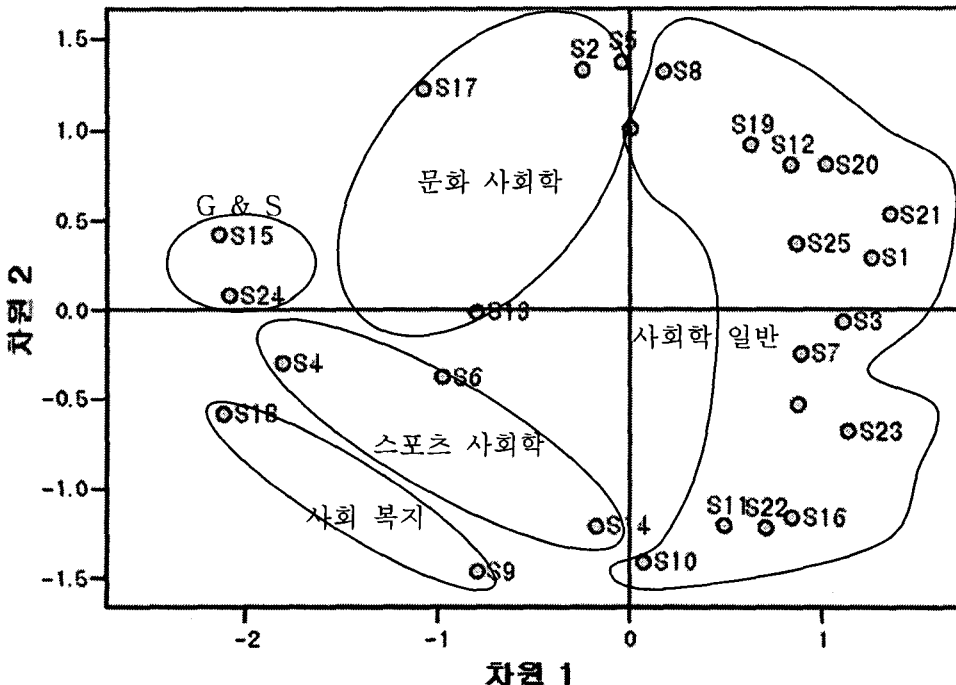
4.2 웹 정보원의 다변량 분석

27개의 사회학 분야 웹 정보원들 간의 상관관계를 보다 자세히 분석하기 위하여 다차원 척도 프로그램 ALSCAL(Alternating Least Squares Scaling)을 이용하여 웹 정보원들 간의 위치를 2차원 공간상에 점으로 표시한 2차원 지도를 출력하였다. 27개 웹 정보원에 대한

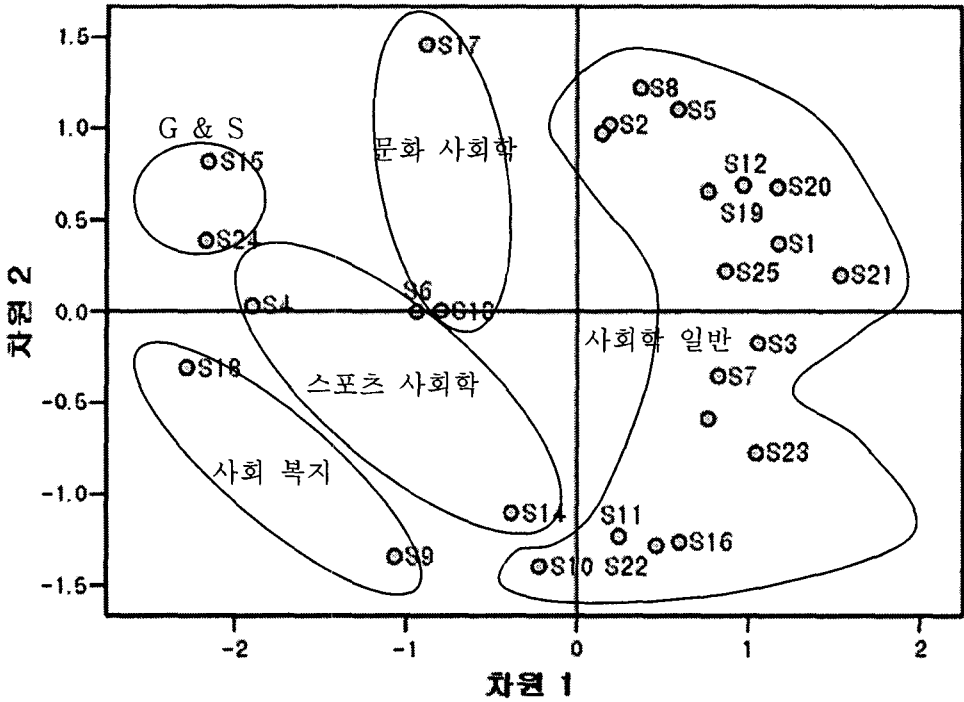
다차원 축적 분석 결과, AltaVista의 경우 스트레스 값은 0.15이며, RSQ값은 0.90으로, AlltheWeb의 경우 스트레스 값은 0.12, RSQ값은 0.91로 나타났다. MSN의 경우 스트레스 값은 0.12이며, RSQ값은 0.93으로 나타났다.

덴드로그램을 이용해서 웹 정보원들을 유사한 구성원끼리 묶어서 각 검색엔진별로 5개의 군집으로 나누고, 이미 부여된 웹 정보원의 주제 분야를 근거로 형성된 군집에 따라 정보원 지도에 군집간의 경계선을 그리면 다음과 같다. <그림1><그림2><그림3>

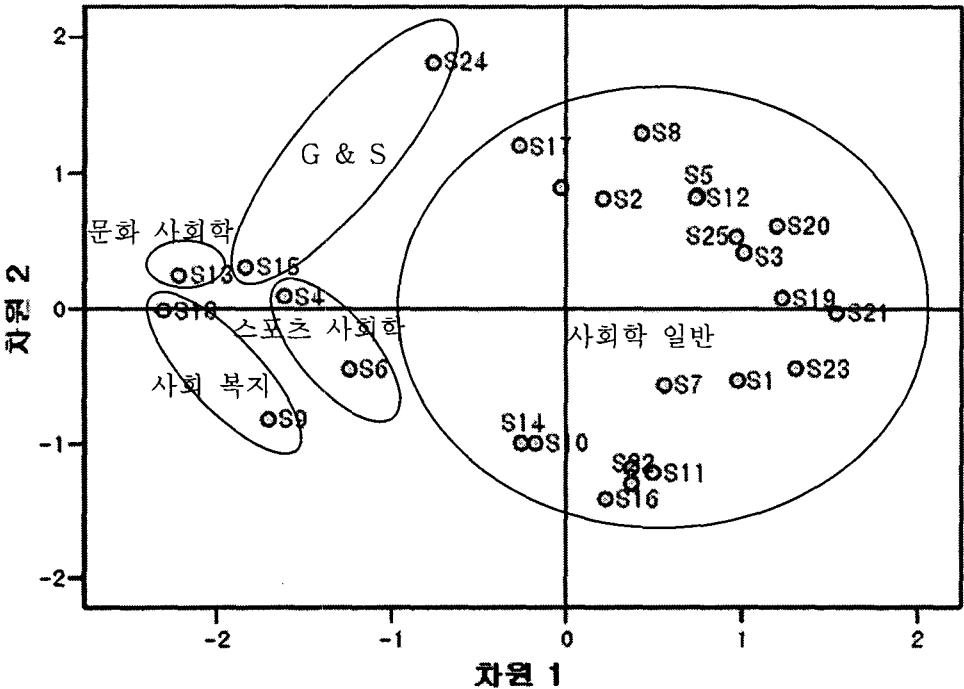
AltaVista의 경우, 제 1군집은 지도의 중앙상단부분에 분포하고 있으며 사회학에 있어서 문화와 관련된 정보원을 다루고 있으므로 '문화 사회학'이라는 군집명이 부여되었다. 제 2군집은 지도의 오른쪽에 분포하고 있으며 사회학에 관한 일반적인 정보를 제공하고 있어 '사회학 일반'이라는 군집명이 부여되었다. 이 군집에서 사회학 이론에 관한 정보를 제공하는 S12번, S19번, S20번, S25번이 포함되어



<그림 1> 군집분석에 의한 사회학 분야의 웹 정보원 지도 (AltaVista)



<그림 2> 군집분석에 의한 사회학 분야의 웹 정보원 지도 (AlltheWeb)



<그림 3> 군집분석에 의한 사회학 분야의 웹 정보원 지도(MSN)

있다. 제 3군집은 'G & S' (Gender & Sexuality) 라는 군집명을 가지며, S15번과 S24번이 포함되어 있다. 제 4군집은 지도의 왼쪽 하단부분에 분포하고 있으며, 2개의 정보원이 사회복지에 관련된 정보원으로써 '사회복지' 라는 군집명이 부여되었다. 제 5군집은 '스포츠 사회학' 이라는 군집명을 가지며, 사회학과 연관된 스포츠 관련 정보원들로 구성되어 있으며, S4번, S6번, S14번이 포함되어 있다.

AlltheWeb의 경우, 제 1군집은 S13번과 S17번으로 사회학과 문화와 관련된 정보를 제공하고 있어 '문화 사회학'이라는 군집명이 부여되었다. 제 2군집은 지도의 오른쪽에 위치하고 있으며, '사회학 일반' 이라는 군집명을 가진다. 사회학의 일반적인 정보를 제공하고 있으며, 사회학 이론에 관한 정보를 제공하는 S12번, S19번, S20번, S25번이 포함되어 있으며, 문화 사회학에 관한 정보를 제공하는 S2번, S5번도 포함되어 있다. 제 3군집은 지도의 왼쪽에 분포하고 있으며, 'G & S' (Gender & Sexuality) 라는 군집명이 부여되었다. 제 4군집은 S9번과 S18번으로 '사회 복지'라는 군집명이 부여되었다. 제 5군집은 S4번, S6번, S14번으로 '스포츠 사회학' 이라는 군집명이 부여되었다.

MSN의 경우, 제 1군집은 지도의 왼쪽에 분포하고 있으며, '문화 사회학' 이라는 군집명이 부여되었다. 제 2군집은 지도의 오른쪽에 위치하고 있으며, '사회학 일반' 이라는 군집명이 부여되었다. 이 군집에 포함되는 S12번, S19번, S20번, S25번은 사회학 이론에 관한 정보를 제공하고 있으며, S2번, S5번, S17번은 문화 사회학에 관한 정보를 제공한다. 제 3군집은 S15번과 S24번으로 'G & S' (Gender & Sexuality)라는 군집명이 부여되었다. 제 4군집은 '사회복지'라는 군집명이 부여되었으며, S9번과 S18번이 포함되어 있다. 제 5군집은

'스포츠 사회학' 이라는 군집명이 부여되었으며, S4번과 S6번이 포함되어 있다.

AltaVista, AlltheWeb, MSN 웹 정보원 지도를 살펴보면, 큰 주제 범위에서 전체적으로 비슷한 구조를 보여주고 있다. 지도의 오른쪽 부분은 사회학 일반이 있고, 왼쪽 상단에 문화 사회학과 G & S (Gender & Sexuality)가 있으며, 왼쪽 하단에 사회복지와 스포츠 사회학이 위치하고 있다. 세부 주제에 속하는 각각의 웹 정보원은 대부분 일치했지만, 몇 개의 웹 정보원은 소속 군집이 달라지는 경우도 있다. 예를 들면 S17번이 AltaVista와 AlltheWeb에서는 '문화 사회학'에 포함되어 있으나, MSN에서는 '사회학 일반'에 포함되어 있다. S14번도 AltaVista와 AlltheWeb에서는 '스포츠 사회학'에 포함되어 있으나, MSN에서는 '사회학 일반'에 포함되어 있다. S2번과 S5번은 AltaVista에서는 '문화 사회학'에 소속되어 있으나, AlltheWeb과 MSN에서는 '사회학 일반'에 포함되어 있다. 이와 같은 경우를 제외하고는, 세 개의 검색엔진의 웹 정보원 지도에서 사회학 분야의 지적 구조는 비슷하게 나타나고 있다.

웹 정보원의 동시링크 분석에 있어서는 AltaVista, AlltheWeb, MSN 중 어떤 검색엔진을 사용하여도 특정 주제의 지적구조는 비슷하다고 할 수 있다. 특히 AltaVista와 AlltheWeb의 경우에는 지도상의 웹 정보원의 분포면에서도 거의 유사한 모양을 보여주었다.

5. 결론

본 연구에서는 사회학 분야 웹 정보원을 대상으로 지적구조를 분석하고, 검색 엔진별 지적 구조의 차이를 분석함으로써 웹 정보원의 특성을 연구하고자 하였다. 이를 위해서 27개의 웹 정보원을 대상으로 각 검색엔진별로 동

시링크회수를 산출하여, SPSS 통계의 피어슨 상관계수를 구하였고, 다차원 축적 분석을 통한 2차원 공간에서의 정보원들 간의 맵핑을 시도하였다. 연구결과를 요약하면 다음과 같다.

첫째, 사회학 분야 웹 정보원을 대상으로 군집분석을 한 결과 정보원들 간에 같은 주제 분야로 군집되어 있음을 알 수 있다. 둘째, AltaVista, AlltheWeb, MSN 검색엔진간의 지적구조의 차이를 살펴본 결과, 웹 정보원 지도에 나타난 전체적인 지적구조는 비슷하게 나타났지만, 몇 개의 웹 정보원의 경우에는 소속 군집이 다르게 나타난 경우도 있었다. 그리고, MSN보다 AltaVista와 Alltheweb은 지도상의 웹 정보원 분포면에서 더 비슷한 구조를 나타냈다. 따라서 동시링크를 분석할 때에는 한 개의 검색엔진보다는 두 개 이상의 검색엔진을 활용하는 것이 바람직하다.

이전 연구에서 동시링크 분석을 커뮤니케이션 분야에 적용시켜 웹 정보원의 지적구조 분석이 가능하다는 전제 아래, 다른 분야에서도 적용할 수 있는지 확인해 보기 위하여 사회학 분야에 적용시켜 보았다. 그 결과 적용 가능하다는 실험결과가 나왔으며, 더 나아가 커뮤니케이션과 사회학 이외의 다른 분야에도 적용할 수 있을 것이다. 아직까지 webometrics에 대한 연구가 활발하지 않지만, 웹이 가지고 있는 풍부하고 다양한 정보원을 활용할 수 있는 지속적인 연구가 요구된다.

참고문헌

- 이성숙. 2005. 동시링크분석을 이용한 웹 정보원의 지적구조 변화에 관한 연구. 『정보관리학회지』, 22(2): 205-226.
- 정동열, 최윤미. 1999. 웹 정보원의 동시인용분석에 관한 실험적 연구. 『정보관리학회지』, 16(3): 7-25.
- 조현양. 2003. 인용분석을 통한 웹 자원의 가치 평가에 관한 연구. 『정보관리학회지』, 20(3): 225-240.
- Ingwersen, P. 1998. "The calculation of Web Impact factor." *Journal of Documentation*, 54(2): 236-243.
- Mckiernan, G. 1996. *CitedSites(sm): Citation Indexing of Web resources*. 「cited 2006. 6. 2」. <<http://www.public.iastate.edu/~CYBERSTACKS/Cited.htm>>.
- Rousseau, R. 1997. "Sitations: An exploratory study." *Cybermetrics*, 1(1). 「cited 2006. 6. 2」. <<http://www.cindoc.csic.es/cybermetrics/articles/v2i1p2.html>>.
- Thelwall, M., Vaughan, L. andBjorneborn, L. 2005. "Webometrics" *ARIST*, 39: 81-135.