

웹 정보원의 동시인용분석에 관한 실험적 연구

An Experimental Study of Cocitation Analysis on Web Information

정동열(Dong-Youl Jeong)*, 최윤미(Yoon-Mee Choi)**

목 차

- | | |
|------------------------|--------------------|
| 1 서 론 | 3 웹 정보원의 동시인용분석 |
| 1.1 연구목적 | 3.1 분석대상의 선정 |
| 1.2 연구의 범위와 방법 | 3.2 동시인용분석 실험설계 |
| 2 웹 정보원의 계량적 연구 | 3.3 웹 정보원의 지적구조 분석 |
| 2.1 웹 정보원의 특성 | 4 결 론 |
| 2.2 웹 정보원 인용연구의 이론적 배경 | |

초 록

본 연구는 계량정보학적 분석방법을 웹 정보원에 적용하여 커뮤니케이션 분야의 웹 정보원에 대한 동시인용분석을 실험적으로 실시하고 있다. 특정 학문분야의 지적구조를 규명하기 위하여 링크 건수를 기준으로 선정된 웹 정보원을 대상으로 동시인용빈도, 상관계수행렬, 다차원축적기법과 군집분석 등 동시인용분석 기법을 적용하고 있다. 분석대상이 인쇄문헌이 아닌 기관, 개인의 홈페이지와 웹 색인을 이용한 동시인용분석에서 하위주제분야가 뚜렷이 나타남에 따라 기존 인쇄문헌에 적용되어 온 분석기법이 웹 정보원에 적용될 수 있음을 규명하고 있다. 아울러 동시인용분석에서 인쇄문헌과 웹 정보원이 갖는 특성과 연구방법상의 차별성 등을 분석한다.

ABSTRACT

This experimental study examines informetric analysis of World Wide Web based upon cocitation analysis of Web pages and features of Web resources in the field of communication studies. Cocitation analysis is basically performed to examine the intellectual structure of the communication studies in reflecting link count on the Web. The selected Web resources in the field are mapped in two dimensions based upon the similarities of cocitation frequency, correlation matrix, multidimensional scale and cluster analysis. Cocitation analysis methods using organizational homepage, personal homepage, or Web index, to Web produced clustering of Web resources that had topical similarities. So far, although informetric analysis of Web resources is in the preliminary stage, it shows that Web can be a new tool for indicating the intellectual structure of a specific research field. In addition, this study analyzes characteristics of printing resources and Web resources, and differences of research methods in applying cocitation analysis.

키워드: 웹 정보원, 동시인용분석, WWW, 인터넷 계량정보

* 이화여자대학교 사회과학대학 정보학 전공 교수

** (주)ITA Contents Business 사업부

■ 논문 접수일 : 1999년 4월 8일

1 서론

1.1 연구목적

인용연구는 특정분야의 연구자들이 생산한 문헌이나 기사 및 저자들간의 연관성을 계량적으로 분석함으로써 그 주제분야의 지적구조를 파악하려는 것이 주된 목적이다. ISI에서 인쇄문헌을 대상으로 SCI, SSCI, A&HCI를 간행하면서 다양한 연구방법을 인용분석에 적용할 수 있었다. 그러나 1980년대 말부터 인터넷의 등장과 급속한 성장은 정보생성과 검색에 획기적인 변화를 초래하게 되었다. 전자적 기술을 이용한 출판물이 대량으로 생성되어 네트워크를 활용한 학술적 커뮤니케이션에 변화를 가져오고 있다. 인쇄본의 존속여부와 관계없이 단행본이나 저널의 형태로 이루어지고 있는 텍스트 중에서 상당부분이 전자화된 데이터베이스의 형태로 변화되고 있다. 이러한 전자화된 텍스트는 인쇄본보다 더욱 다양한 정보를 제공함은 물론 신속하게 통신수단을 활용함으로써 이용자들의 정보요구에 대한 만족도를 향상시킬 것으로 예상된다.

인터넷의 등장 이전에 주로 인쇄매체를 대상으로 활발히 연구된 계량정보학의 대상이 인터넷, 특히 WWW(이하 웹)과 같은 전자 정보원의 텍스트를 대상으로 분석이 가능할 것이라는 예측이 제기되고 있다. 웹은 각종 정보원들이 링크를 통해서 연결되어 지속적으로 변화되고 있어 현재 진행중인 연구와 관심분야를 파악하는데 가장 적합한 정보원이며, 이러한 구조는 디지털 도서관 구조의 원형이라고 볼 수 있다. 웹 정보원의 특징 분석을 통하여 계량정보학적 방법론의 적용 가능성에 대한 연구의 필요성이 대두되고 있다.

따라서 본 연구의 목적은 계량정보학의 연구대

상으로 웹 정보원의 특징과 이론적 근거를 제시하며, 웹 정보원에 실제로 계량정보학적 방법 중 동시인용분석을 적용하여 인쇄문헌을 대상으로 한 연구와의 관련성을 실험적으로 분석하고 있다. 특히 동시인용분석에서 인쇄문헌과 웹 정보원의 특성은 물론 연구방법상의 차별성을 제시하고 있다.

1.2 연구의 범위와 방법

본 연구의 범위는 웹 정보원에 대한 계량정보학적 분석의 가능성을 살펴보기 위하여 웹 정보원의 특징과 이론적 근거, 웹 정보원을 이용한 실험적 적용으로 되어 있다. 인쇄문헌이 출판된 데이터(예: SCI, SSCI, A&HCI)를 이용하여 계량정보학적 분석을 실시한 것과는 달리 웹 정보원을 이용한 계량정보학적 분석에서는 그러한 데이터가 존재하지 않기 때문에 직접 웹 정보원을 검색하여 데이터를 수집하여야 하는 것이 큰 차이점이다. 웹에서 자신이 원하는 정보를 검색하거나 분석대상을 선정하기 위해서는 검색엔진을 통해 원하는 주제어를 직접 검색하거나 Yahoo, WWW Virtual Library 등과 같은 주제별 색인을 사용하여 대상자료를 수집하는 것이 일반적인 방법이다.

웹에서 이용가능한 정보원을 활용하는 방법으로는 개인 홈페이지, 기관 홈페이지, contents 페이지, 색인 페이지 등 다양한 유형의 정보원을 이용할 수 있다. 특정 저자가 생산한 개별적인 웹 문헌인 contents 페이지를 분석함으로써 대상이 되는 주제분야 정보원의 성격과 특징이 뚜렷이 나타나며, 그 웹 페이지와 링크하고 있는 정보원들의 의도도 분명하게 알 수 있어 매우 이상적이다. 그러나 현재까지의 웹 구조에서는 contents

페이지에 대한 링크의 수가 매우 제한적이기 때문에 동시인용분석을 수행하는 데에는 어려움이 있어 개별 웹 뿐만 아니라 기관 홈페이지와 색인 페이지 등 모든 유형의 웹 정보원을 포함하고 있다.

본 연구에서는 웹 정보원의 동시인용분석 실험 대상으로 커뮤니케이션 분야를 선정하여 주제분야의 지적구조를 분석하고 있다. 커뮤니케이션 분야의 대표적인 웹 사이트를 선정하기 위하여 여러 개의 웹 색인을 통해 얻어진 자료를 링크 순으로 배열하여 많은 링크를 포함하고 있는 웹 정보원을 대상으로 한 것은 인쇄문헌의 계량정보학적 분석에서 인용빈도가 높은 문헌을 대상으로 하는 것과 동일한 원리이다.

웹 정보원의 동시인용분석을 위한 실험 대상 선정방법은 기 출간된 웹 정보원 참고문헌이나 주제별 전자저널 사이트 (예: <http://www.coalliance.org/index/subject.html>, <http://www.edoc.com/jrl-bin/wilma/hpr>, <http://www.arl.org/scomm/edir/6th/i.html>), 그리고 Yahoo 등의 주제별 디렉토리를 이용할 수 있다. 본 연구에서는 실험적 차원에서 기 출간된 웹 정보원 참고문헌인 Kitao(1988)의 "Internet Resources: ELT, Linguistics, and Communication"에서 제공하는 커뮤니케이션 분야에 대한 대표적인 웹 정보원을 참조로 하여 그 중에서 서지정보를 제공하는 웹 정보원들을 대상으로 검색을 실시하여 선정후보 목록을 작성하였다. 선정후보 웹 정보원을 대상으로 AltaVista 검색엔진을 사용하여 각각의 링크 수를 순위별로 나열한 후 100회 이상의 링크를 가지고 있는 32개의 웹 정보원을 최종 분석대상으로 선정하였다. 선정된 분석대상을 AltaVista 검색엔진으로 웹 정보원들 간의 동시인용빈도를 추출하여 각각의 동시인용빈도행렬을 작성하고, 이를 기초로 유사성 검증

을 위한 상관계수행렬을 도출하며, 상관관계를 분석하기 위하여 다차원축적기법과 군집분석을 실시하였다.

2 웹 정보원의 계량적 연구

2.1 웹 정보원의 특성

웹은 대규모의 컴퓨터 데이터베이스로 인터넷 상에서 문헌과 소프트웨어, 사진, 영화 등 각종 정보자원을 제공하는 하이퍼텍스트 형태의 정보 검색 도구이며, 인터넷상에서 사용되는 여러 가지 프로토콜 및 데이터 형식 등에 대해 통합적인 접속도구 역할을 한다. 정보원으로서의 웹은 인쇄매체와 구별되는 특징을 가지고 있다.

첫째, 웹에서 구할 수 있는 정보량은 매우 다양하고 방대하여 크기와 범위면에서 가장 우수한 정보원으로 인정받고 있다. 둘째, 웹 정보원은 하이퍼텍스트의 개념으로 구현되어 있기 때문에 링크를 통하여 비연속적으로 접근할 수 있으며 계속적으로 관심있는 분야의 정보에 접근할 수 있다. 특히 링크를 따라 이동하면서 자신이 의도하지 않았던 정보를 발견할 수 있어 연구활동과 관심분야를 확대시키는 요인이 될 수 있다. 셋째, 웹을 통해 서버와 클라이언트의 상호작용이 가능하며 텍스트는 물론 그래픽, 동화상, 소리정보 등 다양한 멀티미디어 정보를 검색할 수 있다. 넷째, 메뉴를 클릭하거나 북마크를 이용하여 정보원간의 이동이나 접근이 신속하며, 제공되는 정보원에 대한 복제, 저장, 인쇄, 주문, 배포 등이 매우 용이하다.

이처럼 웹 정보원은 과학적 커뮤니케이션의 새로운 도구로 부상되고 있으며, 최근의 많은 학술

정보와 공공정보가 웹을 통해서 제공되고 있어 정보원의 크기와 범위면에서 우수한 정보원으로 인정받게 되었다. 뿐만 아니라 이용의 편리성과 이용 가능한 정보가 다양하다는 점에서 다양한 계층의 연구자들에게 이용되고 있으며, 대부분의 정보가 무료로 제공되기 때문에 연구자들의 비용과 시간을 절약할 수 있다. Berghel(1995)은 이러한 웹 정보원의 역할을 'Digital Village'로 표현할 정도로 그 자체가 하나의 사회로서 기능을 하고 있다.

2.2 웹 정보원 인용연구의 이론적 배경

인쇄매체를 대상으로 한 계량정보학에서 주로 이용하는 데이터는 학술문헌의 참고문헌을 색인한 인용데이터에 근거를 두고 있다. 이는 과거의 문헌에 대한 분석으로 현재의 학문구조를 반영하지 못하는 점과 발전 가능성이 있는 소수의 연구분야에 대한 분석이 불가능하다는 한계를 지니고 있다. 그러나 웹을 대상으로 한 인용분석은 웹 상에 게재되어 있는 다양한 정보원을 통하여 학자들의 관심분야와 현재 진행되고 있는 연구를 신속하게 파악할 수 있으며 미래의 발전분야를 예측하는 것이 가능하다. 또한 웹 인용연구는 종래의 계량정보학적 분석방법을 그대로 적용하기에 적합한 구조로 되어 있으며, 웹 정보원들 사이의 링크를 통해 쉽게 접근할 수 있어 그 자체가 하나의 정보원으로 활용되고 있다. 동시에 접근점은 HTML코드의 태그를 사용하여 개별 문서를 구분할 수 있어 ISI의 디스크립터와 동일하게 구분할 수 있다.

링크는 인쇄문헌에서 인용과 같은 개념으로 웹 페이지를 만든 기관이나 저자가 특정한 웹 페이지를 자신과 링크시키는 것은 자신의 관심분야를

반영하며 다른 웹 페이지에 대한 인정의 의미로 볼 수 있다. 웹 상에서 링크가 형성되면 검색과정에 있는 이용자들은 링크된 웹 페이지에 접근할 가능성이 높아지게 되므로 많은 링크를 가지고 있는 웹 문헌의 인지도는 높아지며 다른 정보원과 링크되지 않은 웹 정보원들은 웹 상에서 고립된다고 볼 수 있다.

웹을 과학적 커뮤니케이션이 이루어지는 새로운 정보원으로 파악하여 계량정보학적 방법을 처음으로 적용한 Bossy(1995)와 Cameron(1995)의 연구에서 출판물에 한정된 계량정보학을 웹 정보원으로 확장하게 되었다. 이들은 네트워크의 발전에 따라 전자정보원을 과학적 커뮤니케이션의 연구에 포함시켜야 한다는 의미에서 'Netometrics'라는 용어를 제시하였다. 또한 Almind와 Ingwersen(1997)은 'webometrics'라는 용어를 사용하여 웹을 이용한 계량적 연구의 개념과 방법 및 연구대상의 확대를 주장하였다.

인용이라는 개념이 웹에서 사용될 때는 인터넷 사이트들간의 관계를 나타내는 것으로 볼 수 있으며, 인용색인의 개념은 웹을 하이퍼링크로 연결된 하나의 코퍼스 혹은 인용된 사이트들의 집합체로 볼 수 있다. Rousseau(1997)는 인터넷 사이트들간의 링크수를 측정하기 위하여 웹 페이지를 대상으로 도메인의 수와 도메인별 인용의 수 등을 조사한 결과 웹 문헌의 분포가 Lotka의 법칙과 일치함을 밝혀냈다. 또한 Ingwersen(1998)은 저널의 영향력을 평가하기 위하여 ISI에서 실시중인 영향도(impact factor)가 웹 정보원에서 적용가능 여부를 연구한 결과, 대규모 웹 사이트에 대한 영향도가 소규모 웹 사이트의 경우보다 신뢰성이 높으며, 오히려 인쇄매체보다 역사가 짧고 한 잡지당 기사수가 적은 전자정보원에 더욱 적합한 것으로 나타났다. Rousseau와 Ingwersen의 두

실험적 연구는 인쇄매체와 전자정보원이 동일한 원리에서 인용연구가 가능하다는 사례를 제시하고 있다.

웹 정보원의 계량적 연구에 있어 해결되어야 할 문제점으로는 첫째, 시간에 따라 수시로 변화하는 역동적인 정보공간으로 통시적인 분석에 어려움이 생길 수 있다. 둘째, 웹 파일의 구조상 조사하고자 하는 웹 페이지의 전체를 규명하기 곤란하며, 웹의 무수한 데이터 중에서 조사자가 원하는 정보를 정확히 찾아내기 힘들다. 셋째, 웹을 통한 인용연구시 저자가 제시되지 않는 경우가 많아 저자를 규명하기 어렵다. 넷째, 웹 정보원 이용에 있어 저작권 문제는 디지털 자료의 출현과 더불어 항상 존재한 문제점으로 앞으로 많은 연구가 필요한 분야로 남아있다. 이러한 제반 문제점은 웹 색인이 보다 세련되고 정교한 방향으로 발전되고 있어 점차적으로 줄어들 것으로 예상된다.

3.1 분석대상의 선정

커뮤니케이션 분야에 대한 대표적인 웹 정보원을 선정하기 위해 Kitao(1988)의 "Internet Resources: ELT, Linguistics, and Communication"에서 제시하는 웹 사이트 중에서 커뮤니케이션 분야의 세부 주제별 정보원을 제공하는 대표적인 서지 사이트 7개를 우선적으로 선정하였다(표 1) 참조).

7개의 웹 사이트를 출발점으로 하여 이들이 링크하고 있는 웹 정보원들 중 커뮤니케이션 분야에 대한 웹 정보원들을 추출하여 1차적으로 선정 후보로 한다(부록 참조). 1차 선정후보 웹 정보원을 대상으로 웹 상에서 다른 정보원에 의해 링크된 건수를 조사하기 위하여 검색엔진 AltaVista가 사용되었다. AltaVista의 Advanced Queries에서 특정 웹 페이지의 링크 수는 "link: URL"와 같은 형식의 검색식을 입력하여 산출하게 된다.

예: link:www.americancomm.org/

3 웹 정보원의 동시인용분석

〈표 1〉 웹 정보원 선정을 위한 검색도구로 사용된 정보원 목록

정 보 원	URL
American Communication Association	http://www.americancomm.org
Communication Links(Boston College)	http://www.bc.edu/bc_org/arp/cas/comm/comm.html
WWW Virtual Library: Communication	http://galaxy.einet.net./galaxy/social-Sciences/Comm.html
Advertizing World	http://advweb.cocomm.utexas.edu/world/
Communication Resources(Vanderbilt Univ.)	http://vanderbilt.edu/~parker/comm/comm.html
Yahoo's List: Communication	http://www.yahoo.comsocial_science/communication
Communication Bookmarks	http://www1.arcade.uiowa.edu/gw/comm

1999년 2월말 현재 커뮤니케이션 분야의 7개 사이트로부터 선정된 1차 후보 83개 웹 정보원을

대상으로 각각의 링크 수를 조사하여, 선정후보 중 링크된 건수가 100건 이상인 총 32개의 웹 정

〈표 2〉 커뮤니케이션 분야 분석대상 웹 정보원

번호	정보원	링크수	주제분야	성격
1	American Communication Association http://www.americanscomm.org/ http://cavern.uark.edu/comminfo/www/ACA.html	551	GC	단체
2	Association for Educational Communication and Technology http://www.aect.org/	1034	GC	단체
3	Communication Institute for Online Scholarship http://www.cios.org/	742	GC	단체
4	ICA International Communication Association http://www.icahdq.org . 또는 http://www.io.com/~icahdq/ica/	316	GC	단체
5	The Public Relations Society of America http://www.prsa.org/	1014	PR	단체
6	Southern States Communication Association http://ssca.net/	163	GC	단체
7	Analysis Telecoms Virtual Library http://www.analysys.com/vlib/	475	TEC	색인
8	Computer Mediated Communication Magazines http://www.december.com/cmc/mag	2183	CMC	전자
9	Journal of Computer-Mediated Communication http://jcmc.huji.ac.il/ 또는 http://www.ascusc.org/jcmc http://shum.huji.ac.il/jcmc/jcmc.html	997	CMC	전자
10	Advertizing World http://advweb.cocomm.utexas.edu/world	294	GC	색인
11	Vanderbilt Television News Archive http://tvnews.vandebilt.edu/	1682	MC	색인
12	Debate Central http://beluga.uvm.edu/	294	SC	색인
13	Debate.net http://debate.net/	311	SC	색인
14	Center for Media Education http://www.cme.org/cme	147	MC	단체
15	Communication & Education on the Web http://alnilam.usc.indiana.edu:1027/	390	GC	색인
16	CompuServe's Journalism Forum http://www.jforum.org/	199	JR	단체

〈표 2〉 커뮤니케이션 분야 분석대상 웹 정보원(계속)

번호	정 보 원	링크수	주제분야	성격
17	FACSNET http://www.facsnet.org/	815	JR	단체
18	Media and Communication Studies Site http://www.aber.ac.uk/~dgc/media/html	794	MC	단체
19	Media Studies Center http://www.nediastudies.org/	1135	MC	색인
20	Megasources http://www.acs.rterson.ca/~journal/measources.html	340	JR	색인
21	Reproter.org http://www.reporter.org/	942	JR	단체
22	WWW Virtual Library: Journalism http://www.cais.com/makulow/vlj.html	757	JR	색인
23	Information Sources: The Internet and Computer-Meduated Communication http://www.december.com/cmc/info/index.html	334	CMC	색인
24	Department of Speech Communication: Univ. of Illinois http://www.spcomm.uiuc.edu/	738	SC	색인
25	Department of Advertising & Public Relations http://www.apr.ua.edu/	123	AD,PR	색인
26	European Speech Communication Association http://ophale.icp.grenet.fr/esca/esca.html	529	SC	단체
27	NCA Undergraduate Student Organization http://cotton.uamont.edu/~roiger/scclub/star.htmlx http://coton.uamont.edu/~scauso/	174	GC	단체
28	Sarah Zupko's Cultural Studies Center http://www.mcs.net/~zupko/popcult.htm	368	M&C	단체
29	Electronic Journ@list, The http://www.spj.org/	468	JR	단체
30	Online Journalism Review http://www.ojr.org/	163	JR	단체
31	Ryerson Review of Journalism Online http://www.ryerson.ca/rj/	115	JR	전자
32	JournalismNet http://www.journalismnet.com/	159	JR	단체

보원을 결정하였다(〈표 2〉 참조). 선정된 웹 정보원을 성격과 주제에 따라 구분하기 위하여

American Communication Association의 주제명과 ALA의 Directory of Electronic Journals

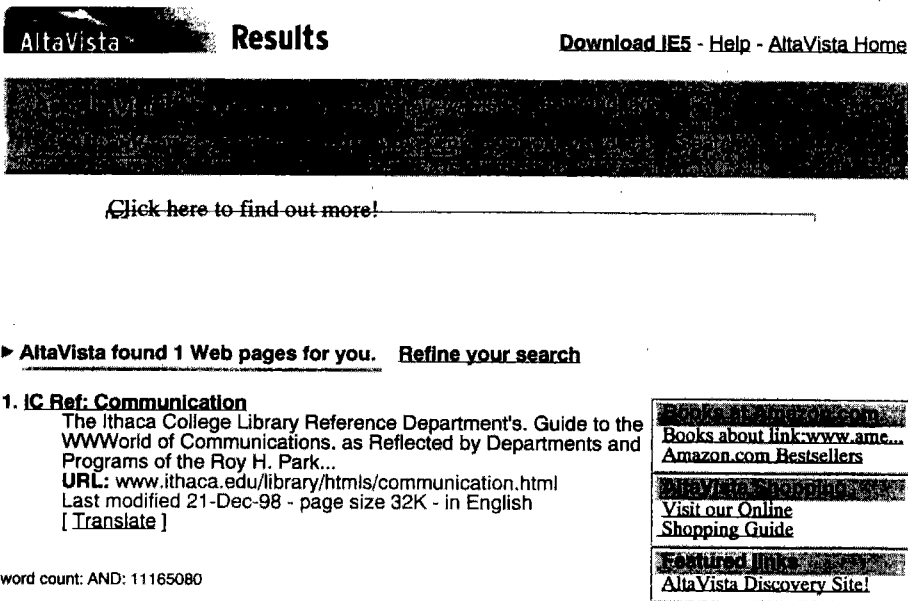
Newsletters and Academic List Serve에서 제시한 커뮤니케이션 분야의 하위주제명을 참고로 하였다. 하위주제명으로는 General Communication[GC], Computer-mediated Communication [CMC], Speech Communication [SC], Telecommunication[TEC], Mass Communication [MC], Cultural Study[CS], Media and Culture[M&C], Media and Technology [M&T], Journalism[JR], Advertising[AD], Public Relations[PR] 등으로 부여하였다.

웹 정보원에 수록된 웹 페이지의 기능과 성격은 헤드(head) 페이지와 레퍼런스(reference) 페이지의 내용으로 식별된다. 헤드 페이지는 여러 개의 웹 페이지를 대표하고 다른 웹 페이지를 방문할 수 있도록 연결하는 역할을 하는 페이지로 기관(혹은 단체)이나 개인의 홈페이지, 색인 페이지, 색인에 대한 색인을 제공하는 원색인(source index) 등이 속한다. 레퍼런스 페이지는 그 자체

가 정보원의 기능을 담당하는 페이지로 destination 페이지와 contents 페이지로 구분된다. Destination 페이지는 각종 서지정보를 제공하며 다른 웹 페이지로부터 많은 링크를 받지만 자신은 다른 웹 페이지들을 링크하지는 않는다. Contents 페이지는 서지정보가 아닌 실제정보를 제공하는 웹 페이지로 개별 저자의 저작 등이 이에 속한다. 전자잡지의 경우 실제 정보원인 기사를 연결하는 역할을 하기 때문에 색인 페이지와는 성격면에서 차이가 있어 별도로 구분하였다.

3.2 동시인용분석 실험설계

분석대상 32개의 웹 정보원을 대상으로 2개의 웹 정보원을 공통으로 링크하고 있는 웹 정보원을 조사하기 위하여 AltaVista의 Advanced Queries에 "link:URL AND link:URL"라는 검색식을 이용한다. 예를 들면,



<그림 1> 동시인용빈도 추출화면

link:www.americancomm.org/

AND link:tvnews.vanderbilt.edu/

이는 동시인용빈도를 추출하는 과정으로 (그림 1)은 위의 예에 대한 동시인용빈도 추출화면이다. 이러한 방법으로 웹 정보원들의 쌍을 동시에 링크하고 있는 정보원들의 검색건수를 구하여 동시인용빈도 행렬을 작성하게 된다.

동시인용된 웹 정보원의 인용빈도를 통계 프로그램(본 연구에서는 SPSSWIN 7.5)을 사용하여 상관계수행렬로 변환함으로써 정보원들간의 상대적인 유사성과 비유사성을 나타낼 수 있다. 상관계수행렬을 계산하는 이유는 상관계수가 정보원들간의 상대적인 유사성과 비유사성을 표현함으로써 산술적인 빈도수의 차이에서 오는 수치효과를 없앨 수 있고 상관계수가 내포하는 보다 많은 정보를 사용할 수 있기 때문이다.

상관계수를 도출하기 위하여 동시인용빈도 행렬의 대각선 값, 즉 자기인용을 포함한 값을 수정할 필요가 있다. 일반적으로 웹 페이지를 평가하는데 있어 링크의 수가 중요한 평가요소이지만 기관이나 단체의 홈페이지일수록 많은 수의 자기링크를 포함하고 있다. 인용분석에서 대각선 값을 조정하는 방법으로 대각원소의 값을 상실데이터로 처리하거나 다른 데이터와 동시인용된 빈도 중 상위 수치 3개를 합하여 2로 나눈 값으로 하향조절하는 방법이 사용된다. 본 연구에서는 하향조절한 값으로 처리하여 계산되었다.

웹 정보원들간의 상관관계와 유사성을 더욱 자세히 알아보기 위하여 다차원축척과 군집분석이 이용된다. 다차원축척은 데이터 속에 숨어있는 특정한 구조를 분석할 수 있는 기법으로 둘씩 쌍을 이룬 개체간의 유사성과 상이성을 토대로 다차원 공간 속에 좌표화하여 거리의 개념으로 표

현한 것이다. 즉 유사한 개체일수록 지도상에 가깝게 위치하며, 다른 개체와 많이 연결되어 있을수록 중심부에 위치하게 된다. 군집분석은 일정한 속성에 따라 몇 개의 집단으로 분류하여 집단 내 대상물의 유사성과 집단간의 차이점을 거리로 계산하여 나타내는 방법으로 일반적으로 평균 연결법을 사용한다.

3.3 웹 정보원의 지적구조 분석

3.3.1 웹 정보원의 동시인용빈도

분석대상 32개의 커뮤니케이션 분야 웹 정보원에서 496개의 웹 정보원 쌍이 형성되어 동시인용빈도 행렬이 작성되었다. <표 3>에서 각 정보원들이 동시인용된 빈도 수의 범위는 0에서 238까지로 가장 높은 동시인용빈도를 가지는 쌍은 8번(Computer Mediated Communication Magazines)과 9번(Journal of Computer-Mediated Communication)이고, 다음으로 191회의 동시인용빈도를 가지는 쌍은 20번(Megasources)과 22번(WWW Virtual Library: Journalism)으로 나타났다.

가장 많이 동시인용된 정보원을 파악하기 위하여 31개 정보원과 동시인용된 총 건수를 조사한 결과, 총 건수가 가장 많은 웹 정보원은 22번(WWW Virtual Library: Journalism)이 622건으로 가장 많고, 다음으로 9번(Journal of Computer-Mediated Communication) 462건, 8번(Computer Mediated Communication Magazines) 393건 순으로 나타났다. 반면에 다른 웹 정보원과 동시인용된 총 건수가 적은 웹 정보원은 14번(Center for Media Education) 4건, 24번(Department of Speech Communication) 12건, 31번(Ryerson Review of Journalism Online) 12

적 높기 때문이다.

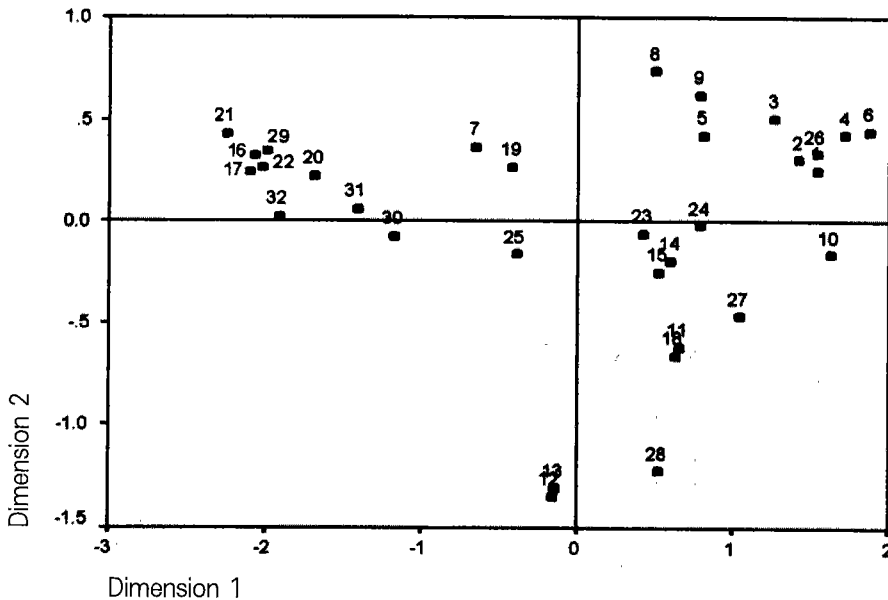
3.3.3 웹 정보원의 다변량분석

32개의 커뮤니케이션 분야 웹 정보원들의 상관 관계를 보다 자세히 분석하기 위하여 다차원축적 프로그램인 ALSCAL을 이용하여 웹 정보원들의 위치를 2차원 공간상에 점으로 표시한 정보원 지도를 출력할 수 있다. 개별 웹 정보원의 좌표값에 따라 작성된 정보원 지도가 <그림 2>이다. 정보원 지도상에 표시된 32개의 정보원 위치는 상관관계 행렬에서 높은 수치를 나타낸 8번과 9번, 17번과 21번, 4번과 6번, 12번과 13번 등의 위치는 가깝게 나타나고, 낮은 수치를 나타낸 10번과 29번, 10번과 16번 등은 지도상에서 멀리 떨어져서 위치하고 있다.

32개 웹 정보원에 대한 다차원축적 분석 결과, 2차원에서의 최종 스트레스 값은 0.01이며, RSQ

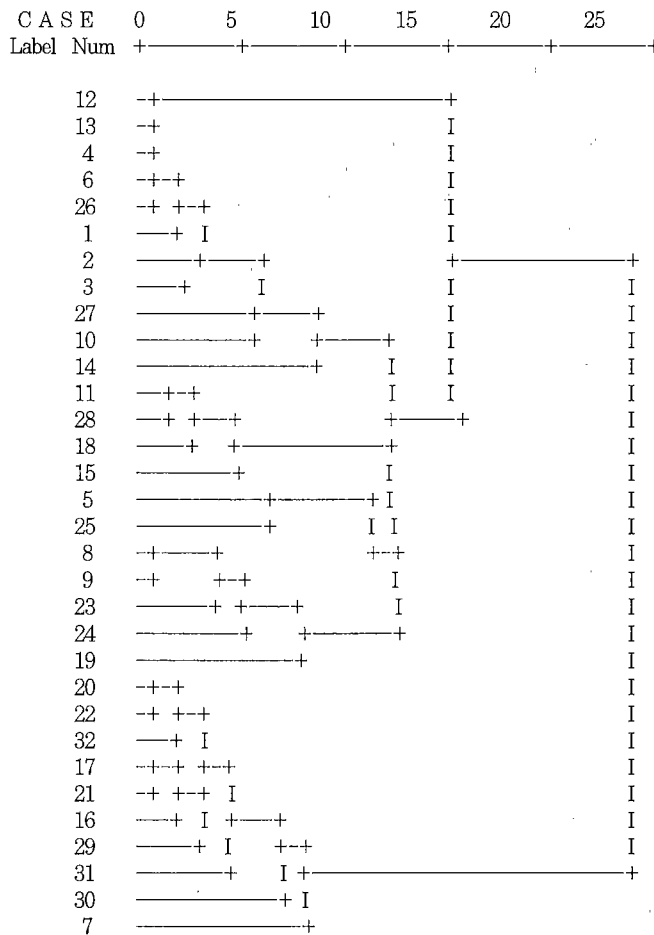
값은 0.95로 나타났다. 스트레스 값은 주어진 유사행렬로부터 축적된 지도가 본래의 자료를 어느 정도 재현하였는가를 나타내는 값으로 그 값이 2.5% 정도이면 매우 잘 재현되었다고 할 수 있다. 통계적으로 스트레스 값이 낮고 표현된 변량(RSQ)의 비율이 높을수록 지도의 적합도가 높다고 해석되므로 본 연구의 스트레스 값과 RSQ는 매우 만족할만한 수준으로 해석된다.

웹 정보원들을 유사한 구성원끼리 묶어서 군집의 형성과정을 나타낸 덴드로그램에서 5개의 군집으로 나누어 볼 수 있다(<그림 3> 참조). 왼쪽의 번호에 따라 12번과 13번 정보원이 제1군집, 4번에서 14번까지 9개 정보원이 제2군집, 11번에서 15번까지 4개 정보원이 제3군집, 5번에서 19번까지 7개 정보원이 제4군집, 그리고 20번에서 7번까지 10개 정보원이 제5군집을 형성하고 있다.



<그림 2> 다차원축적에 의한 커뮤니케이션 분야의 웹 정보원지도

Dendrogram using Average Linkage (Between Groups)
Rescaled Distance Cluster Combine



〈그림 3〉 커뮤니케이션 분야의 웹 정보원의 군집형성 과정

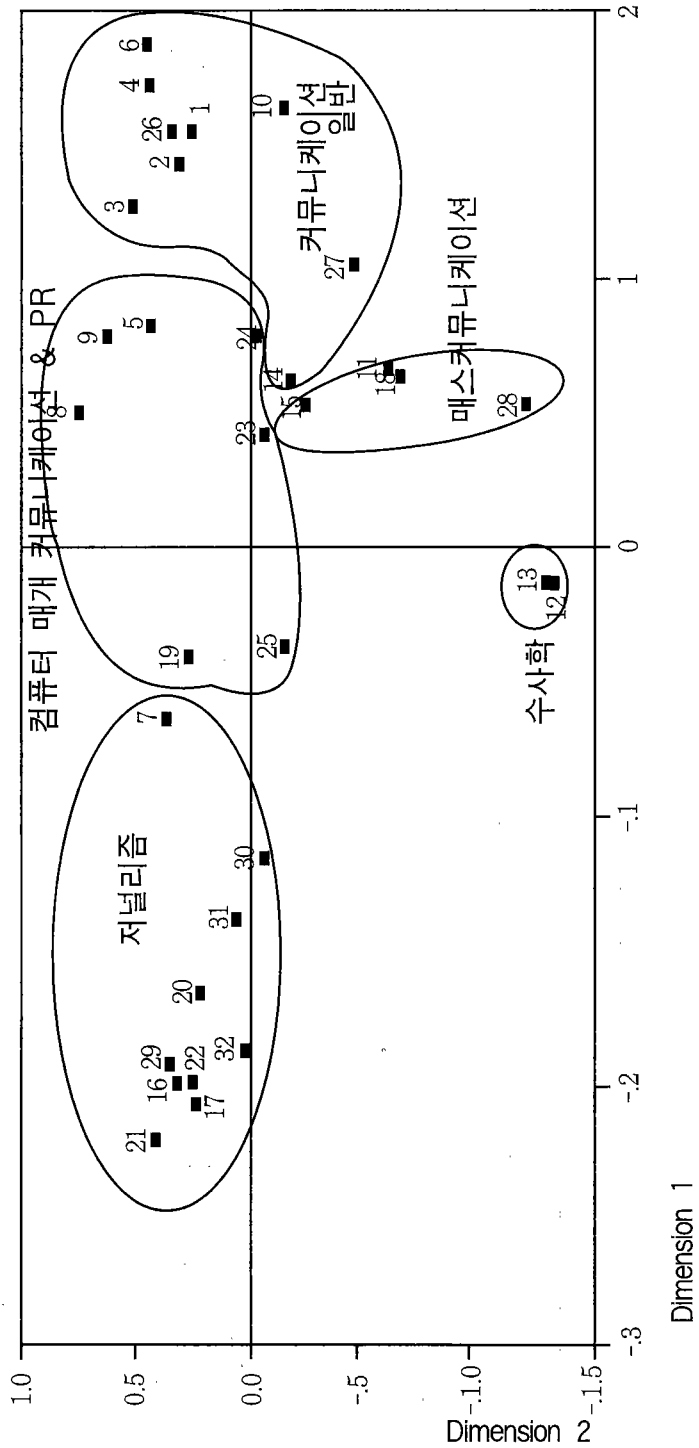
이미 부여된 웹 정보원의 주제분야를 근거로 형성된 군집에 따라 정보원지도에 군집간의 경계선을 그리면 〈그림 4〉와 같다.

제1군집은 12번과 13번 정보원으로 언어적 커뮤니케이션 중에서 토론(debate)에 관한 정보원들로 커뮤니케이션에 관한 일반적인 정보를 제공하는 정보원들과 구별되기에 '수사학'이라는 군집명을 부여하였다.

제2군집은 지도의 오른쪽에 위치하며 커뮤니케이션의 일반적인 정보를 제공하고 있어 '커뮤니케이션 일반'이라는 군집명을 부여하였다.

제3군집은 '매스 커뮤니케이션'이라는 군집명으로 매스 커뮤니케이션에 관한 정보원인 11번, 15번, 18번이 포함되어 있으며, 문화 커뮤니케이션을 다루고 있는 28번도 포함되어 있다.

제4군집은 지도의 중앙부분에 위치하고 있으며



<그림 4> 군집분석에 의한 커뮤니케이션 분야의 웹 정보원지도

컴퓨터 매개 커뮤니케이션에 관한 정보원들과 광고, PR에 관한 정보원들로 구성되어 있어 '컴퓨터 매개 커뮤니케이션 & PR'로 군집화하였다. 커뮤니케이션에 대한 일반적인 정보원보다 컴퓨터를 매개로 한 커뮤니케이션이 지도상에서 중심 부분을 차지하는 것은 이 분야의 연구가 컴퓨터를 대상으로 하는 커뮤니케이션을 다루는 만큼 학자들이 자신들의 연구결과를 네트워크 상에 보다 활발히 게재하고 있으며, 최근 커뮤니케이션 분야의 많은 연구가 컴퓨터를 활용하고 있다는 증거가 되기도 한다. 또한 커뮤니케이션 분야 내에서 컴퓨터 매개 커뮤니케이션을 다루는 전자잡지인 8번(Computer Mediated Communication Magazines)과 9번(Journal of Computer-Mediated Communication)의 인지도가 높게 나타났을 뿐만 아니라 일반적인 색인 정보원들에 비해 컴퓨터 매개 커뮤니케이션 분야의 정보원 수는 한정되어 있으므로 많은 색인 페이지들이 이 두 정보원을 링크하고 있기 때문으로 해석된다.

제5군집은 지도상의 왼쪽 상단 부분에 분포하며 10개의 정보원이 모두 저널리즘에 관련된 정보원으로 '저널리즘'이라는 군집명이 부여되었다. 저널리즘에 관련된 정보원들은 일반 커뮤니케이션 분야와 뚜렷하게 식별되어 하나의 군집을 형성하고 있음을 알 수 있다.

4 결 론

웹 정보원을 대상으로 지적 구조를 규명하기 위하여 커뮤니케이션 분야의 대표적인 웹 정보원과 전자잡지에 대한 동시인용분석을 실시한 결과를 요약하면 다음과 같다.

커뮤니케이션 분야의 웹 정보원 분석에서 저널

리즘, 수사학, 컴퓨터 매개 커뮤니케이션, 매스 커뮤니케이션, 커뮤니케이션 일반 등의 하위주제 영역이 뚜렷이 형성되었으며, 커뮤니케이션 분야의 전자잡지를 대상으로 한 분석은 문화연구, 전자적 커뮤니케이션, 매스 커뮤니케이션 영역이 하위주제분야로 나타났다. 전자잡지를 대상으로 한 분석에서는 분석대상의 수가 많았음에도 불구하고 하위주제영역이 세 분야로만 구분된 이유는, 전자잡지들이 방송, 오락, 문화 등 사회전반에 걸친 주제를 다루는 경우가 많기 때문에 상호 밀접하게 연관되어 있는 것으로 분석되었다.

인쇄문헌에 대한 인용분석이 저자와 출판되는 과정에서 일정 기간이 소요된 문헌들을 대상으로 한다면, 웹을 대상으로 한 인용분석은 현재 연구자들의 관심분야가 되는 주제들에 대한 분석이 가능하여 보다 동적인 학문 경향을 파악하고 이를 통해 미래의 학문경향을 예측할 수 있다는 것이다.

이러한 결과는 분석의 대상이 개별적인 문헌이나 저자가 아닌 색인이나 기관, 개인의 홈페이지와 같은 정보원인 경우에도 동시인용분석이 가능하며, 아울러 그 방법론이 인쇄문헌을 대상으로 실시한 경우와 거의 동일하므로 계량정보학적 연구의 대상으로 웹 정보원이 이용될 수 있다는 근거를 보여주고 있다.

웹 정보원을 대상으로 한 계량정보학적 연구가 활성화되기 위한 몇 가지 사항을 제언할 수 있다.

첫째, 웹 정보원의 구조를 규명할 수 있는 보다 정교하고 다양한 검색엔진이 필요하며, 웹 정보원을 동시인용분석의 대상으로 선정시 적절한 기준설정이 요구된다. AltaVista와 같은 검색엔진으로 URL, 타이틀, 도메인, 링크 등의 검색이 가능하지만 앞으로 색인전문 검색엔진의 활용으로 더욱 다양하고 폭 넓은 분석이 필요하다. 대상 선

정시에도 인쇄문헌과 달리 웹 상의 무수한 정보원들은 여과과정을 거치지 않기 때문에 대상 정보원을 선정하는데 어려움이 따르게 된다.

둘째, 웹 상에서 자신의 저작물을 게재하는 저자들이 적절한 링크를 생성하는 것이 요구된다. 인쇄문헌의 저자들이 인용을 자신의 저작물에 대한 근거로 제시하는 것처럼 웹 상의 저자들도 자신과 관련된 웹 정보원에 대한 적절한 링크를 생성할 필요가 있으며, 링크는 동시인용분석의 도구가 될 뿐만 아니라 이용자들이 다른 정보원을 검색할 수 있는 검색도구로 활용할 수 있다.

셋째, 웹 정보원은 계속적으로 변화하고 있으므로 시간의 경과에 따른 지속적인 연구로 학자

들의 관심사와 학문 구조의 변화를 파악함으로써 현재와 미래의 지식 구조를 이해하는데 필수적인 도구가 될 수 있다. 예를 들어 웹을 통한 계량정보학적 연구는 도메인별로 분석이 가능하므로 국가별, 도시별, 영역별, 지역별로 학문의 발전상이나 그 구조를 용이하게 파악할 수 있다.

지금까지 인쇄문헌을 중심으로 적용되었던 다양한 계량정보학적 법칙과 방법이 웹 정보원에도 적용될 수 있음이 밝혀지고 있다. 웹 정보원에 대한 연구와 분석은 미래의 계량정보학을 위한 도구로서 많은 가능성을 지니고 있기 때문에 효과적인 방법론의 개발 등 지속적인 연구가 요구된다.

참 고 문 헌

- Almind, T. C. & P. Ingwersen. 1997. "Informetrics Analysis on the World Wide Web: Methodological Approaches to 'webometrics'." *Journal of Documentation* 53(4): 404-26.
- Berghel, H. 1995. "Digital Village: Maiden Voyage." *Communications of ACM* 38(11): 25-7.
- Bossy, M. J. 1995. "The Last of the Litter: Netometrics." [cited 1999.3.20]. <<http://www.info.unicaen.fr/bnum/jelec/Solaris/d02/2bossy.html>>.
- Cameron, R. D. 1995. "A Universal Citation Database as a Catalyst for Reform in Scholarly Communication." [cited 1999. 3. 20]. <<http://elib.cs.sfu.ca/projects/papers/citebase/citebase.html>>.
- Ingwersen, P. 1998. "The Calculation of Web Impact Factors." *Journal of Documentation* 54(2): 236-43.
- Kitao, K. 1988. *Internet Resources: ELT, Linguistics, and Communication*. Tokyo: Eichosha.
- Rousseau, R. 1997. "Citation: an Exploratory Study." In *Cybermetrics 1*: [cited 1999.3.21]. <<http://www.cindoc.csic.es/cybermetrics/articles/vlilp1.html>>
- Zakon, R. H. 1988. "Hobbes' Internet Timeline." [cited 1999.3.15]. <<http://198.6.250.9/guest/zakon/Internet/History/HIT.html>>.

〈부록〉 커뮤니케이션 분야 웹 정보원에 대한 선정후보 목록

정 보 원	URL	링크수
American communication Association	http://www.americancomm.org/ http://cavern.uark.edu/comminfo/www/ACA.html	551
Association for Educational Communications and Technology	http://www.aect.org/	1034
Communication Institute for Online Scholarship	http://www.cios.org/	742
ICA International Communication Association	http://www.icahdq.org http://www.io.com/~icahdq/ica/ica.html	316
The Public Relations Society of America	http://www.prsa.org/	1014
Southern States Communication Association	http://ssca.net/	163
Analysys Telecoms Virtual Library	http://www.analysys.com/vlib	475
Computer mediated communication Magazines	http://www.december.com/cmc/mag	2183
Journal of computer-Mediated Communication	http://shum.huji.ac.il/jcmc/jcmc.html http://jcmc.huji.ac.il http://www.ascusc.org/jcmc	997
Advertizing	http://advweb.cocomm.utexas.edu/world	294
Vanderbilt Television News Archive	http://tbnews.vanderbilt.edu/	1682
Debate Central	http://beluga.uvm.edu/	294
Debate.net	http://debate.net/	311
Center for Media Education	http://www.cme.org/cme	147
Communication & Education Resources on the web	http://alnilam.ucs.indiana.edu:1027/	390
CompuServe's journalism Forum	http://www.jforum.org/	199
FACSNET	http://www.facsnet.org/	815
Media and Communication Studies site	http://www.aber.ac.uk/~dgc/media.html	794
Media Studies Center	http://www.mediastudies.org/	1135
Megasources	http://www.acs.ruerspm.ca/~journal/megasources/html	340
Reporter.org	http://www.reporter.org/	942
WWW Virtual Library-Journalism	http://cais.com/makulow/vlj.html	757
Information Sources: the Internet and Computer-Mediated Communication	http://www.december.com/cmc/info/index.html	334
Department of speech Communication: Univ. of Illinois	http://www.spcomm.uiuc.edu/	738
Department of Advertising & Public Relations	http://www.apr.ua.edu/	123
European speech Communication Association	http://ophale.icp.grenet.fr/esca/esca.html	529
NCA Undergraduate student Organization	http://cotton.uamont.edu/~roiger/scaclub/start.htmlx http://cotton.uamont.edu/~scauso/	171
Sarah Zupko's Cultural Studies Center: Calls for Papers/ conferences	http://www.mcs.net/~zupko/popcult.htm	368
Electronic Journ@list, The	http://www.spj.org/	468
Online journalism Review	http://www.ojr.org/	163
Ryerson Review of Journalism Online	http://www.ryerson.ca/rj/	115
JournalismNet	http://www.journalismnet.com/	159

〈부록〉 커뮤니케이션 분야 웹 정보원에 대한 선정후보 목록(계속)

정 보 원	URL	링크수
Links to friendly communication homepages	http://www.csufresno.edu/speechcomm/wsca link.htm	13
Speech Communication Association	http://cotton.uamont.edu/~scauso/	60
Undergraduate Student Organizations		
Tennessee Speech Communication Association	http://www.utm.edu/research/tsca/tsca2.htm	22
WCA World Communication Association	http://ilc2.doshisha.ac.Jp/users/kkitao/organi/	15
WCA The Western states Communication Association	http://www.csufresno.edu/speechcomm/wsca.htm	73
Communication Links	http://www.hamline.edu/depts/commdept/interlinks.html	20
Communication Bookmarks	http://www.bc.edu/bc_org/avp/cas/comm/links/default.html	6
Electronic Resources for Research in Communication	http://www.vanderbilt.edu/~parker/comm/comm.html	3
Internet Resources	http://cavern.uark.edu/libinfo/guide/comm.htm	4
WWW Communication Resources	http://www2.soc.hawaii.edu/css/dept/com/resources/com-resources.html	13
Canadian Journal of Communication CCA's Journal	http://hoshi.cic.sfu.ca/calj/cjc/index.html	6
Communication Quarterly ECA	http://www.as.wvu.edu/~sbb/eca/cq.htm	4
Communication Research Reports ECA	http://www.as.wvu.edu/~sbb/eca/crr.htm	4
Current issue of CIOS/Comserve Intercom Newsletter	http://www.cios.org/www/intercom/inttoc.htm	3
EJC/REC: The Electronic Journal of Communication /La Revue Electronique de Communication	http://www.cios.org/www/ejcmain.htm	48
Human Communication Research ICA	http://www.ica hdq.org:80/publications.html	2
Journal of Communication Studies Arkansas State Communication Association	http://cotton.uamont.edu/~arkcomm/jcs/jcs.html	3
The Southern States Communication Association Newsletter	http://www.vanderbilt.edu/~parker/COMM/v15n2.html	3
Southern Communication Jomal SSSA	http://ssca.net/about/scj.htm	13
Tennessee Communicator	http://www.utm.edu/research/tsca/tsca2.htm	14
Center for dewey Studies	http://www.siu.edu/library/showcase/dewey.html	4
RhetoricNet	http://english-www.hss.cmu.edu/rhetoric.html	40
Project McLuhan	http://www.vyne.com/Mcluhan/	34
Journalism Resources on the Web	http://www.missouri.edu/~jschool/hotlinks/jourlinks.html	20
Mediapolis	http://www.partal.com/mediapolis/ANG/index.html	64
Resources for Rhetoric and Technical Communicators	http://rhetoric.agof.umn.edu/Rhetoric/misc/i links.html	25

〈부록〉 커뮤니케이션 분야 웹 정보원에 대한 선정후보 목록(계속)

정 보 원	URL	링크수
Communication & Education Resourcers	http://alnilam.ucs.indiana.edu:1027/textonly.html	11
Galaxy Communication List	http://www.einet.net/galaxy/Arts-and-Humanities/Language-and-Literature/Commcations.html	6
Southern States Communication Association	http://cotton.uamont.edu/~adams/ssca.htmlx	20
The Electronic Journal for Computer Writing, Rhetoric and Literature	http://auden.fac.utezas.edu:80/~cwl/	16
Vanderbilt University Department of Communication Studies and Theatre	http://www.vanderbilt.edu/AnS/Comm/	15
Communication Programs	http://alnilam.ucs.indiana.edu:1027/sources/programs.html	61
N.C. State U. Department of Communication	http://www2.ncsu.edu/ncsu/chass/communication/index.html	22
Southwest Texas State University	http://www.finearts.swt.edu/	45
The Debate Home Page	http://www.urich.edu/~msl8c/debate.htm	10
Uw - Computer Assisted Journalism	http://weber.u.washington.edu/~scnuweb/classes/316/	17
The Forensics Capital of The Web	http://rhf.bradley.edu/~mcdonald/speech/speech.html	26
Canadian University Society for Intercollegiate Debate	http://www.cam.org/~buchang/cusid/	24
Yahoo List of Public Relations Firms	http://www.yahoo.com/Business/Corporations/Corporate_Services/public_Relations/	2
PRSSA	http://weber.u.washington.edu/~short/	13
Central States Communication Association	http://hydra.bgsu.edu/~jburns/cscsa.html	5
New Jersey Communication Association	http://www.scils.rutgers.edu/njca/njca.htm	16
Communication Studies Resources at Indiana University.	http://alnilam.ucs.indiana.edu:1027/sources/comm.html	90
Content Analysis Resources	http://www.gsu.edu/~wwwcom/content.html	74
Communication Resources	http://www.fau.edu/divdept/ccommcatn/resource.htm	36
Media and Communication Studies Page	http://www.aber.ac.uk/~dgc/media.html	24
California Journalism Online	http://www.csne.org	40
NewsPort	http://newsport.sfsu.edu/	81
Uomini e Media	http://www.vol.it/olgiata/uomini_e_media	76
Wordbiz.Net	http://www.wordbiz.com/	76
World Database of Journalists	http://www.shef.ac.uk/~shep/data/jolist/	17