An Experimental Study on Web Link Analysis of the Korean National Archives

국가기록원 웹사이트 링크정보 분석에 관한 실험적 연구

Zi-Young Park*

ABSTRACT

In this study, web link analysis of the Korean National Archives has been performed using LexiURL to show the dynamic web flow of the National Archives and to identify which closely related institutions the archives should seek to cooperate with for developing an institutional strategic plan. Inlink and co-inlink analysis were performed and, to identify related institutions of the Korean National Archives, a comparison was made with other countries such as Australia, England and the United States. Through web link analysis, several features can be detected and future directions are suggested. Most inlinks and co-inlink of the Korean National Archives were from public institutions, indicating that relationships with educational or research institutions are weak. Moreover, Korean National Archives involve fewer web links related to international activities than other countries. Proceeding from these results, educational function, research function and international activities should be fortified. Korean National Archives need to provide more materials for supporting educational and research activities and more cultural interchange among related institutions. Further research with different heritage institutions such as libraries or museums is needed.

초 록

본 연구에서는 우리나라 국가기록원의 발전 전략 개발을 위해, 국가기록원 웹사이트의 웹링크를 분석하고, 협력이 필요한 관련 기관을 도출하였다. 웹링크의 수집과 분석에는 LexiURL을 이용하였으며, inlink와 co-inlink를 분석하였다. 그리고 우리의 경우와 비교하기 위해 해외 국가기록원(미국, 영국, 호주)의 웹링크도 함께 분석하였다. 웹링크 분석을 통해서 우리나라 국가기록원의 현황 파악과 발전전략 수립을 위한 몇 가지 제안사항을 도출하였다. 현재 우리나라 국가기록원은 대부분의 inlink와 co-inlink가 공공기관에서 온 것으로 교육기관이나 연구기관과의 연계성이 약하다. 또한 국제 활동 현황을 볼 수 있는 링크도 다른 국가들에 비해 부족하다. 이러한 결과로 볼 때, 앞으로는 교육기능과 연구기능, 국제교류 측면을 강화해야 한다. 국가기록원은 학습과 연구활동을 지원할 기록정보를 제공하고, 관련 기관과의 문화교류를 확대할 필요가 있다. 향후에는 도서관이나 박물관과 같은 문화유산 관리기관과의 비교 분석도 이루어질 필요가 있다.

Keywords: link analysis, LexiURL, inlink, co-inlink, informetrics, National Archives, Korean National Archives, 링크분석, 동시링크분석, 계량정보분석, 국가기록원, 영구기록물관리기관

---

* Lecturer, Hansung University(hobbit@yonsei.ac.kr)

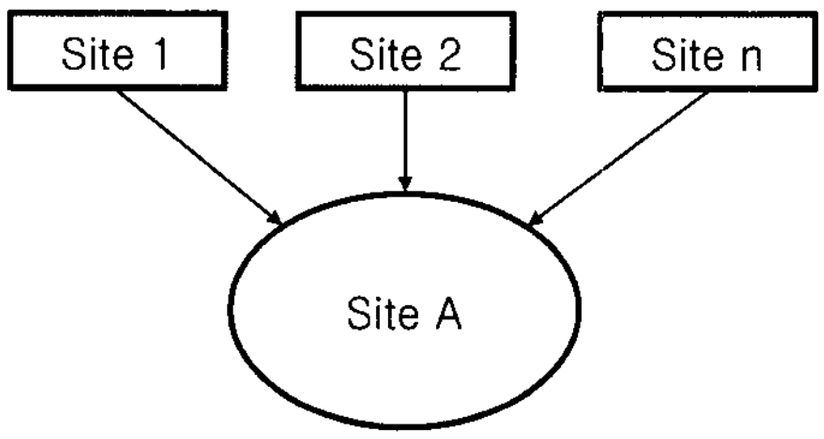
- Received : 22 May 2008 ▪ Revised : 30 May 2008 ▪ Accepted : 4 June 2008
- Journal of the Korean Society for Information Management, 25(2): 167-181, 2008. [DOI:10.3743/KOSIM. 2008.25.2.167]

1. Introduction

Due to the development of information and technology, geographic and physical borderlines in the knowledge information service areas have become vague. World wide access to information is not a dream. In this context, web links play an important role in delivering information. We can make our own web logs and collect any information we need using web links. Institutional web sites also have links of their favorite subjects.

Moreover, we can track the evidence of this information flow through web links analysis. In this study, two categories in link analysis were considered: inlinks and co-inlinks.

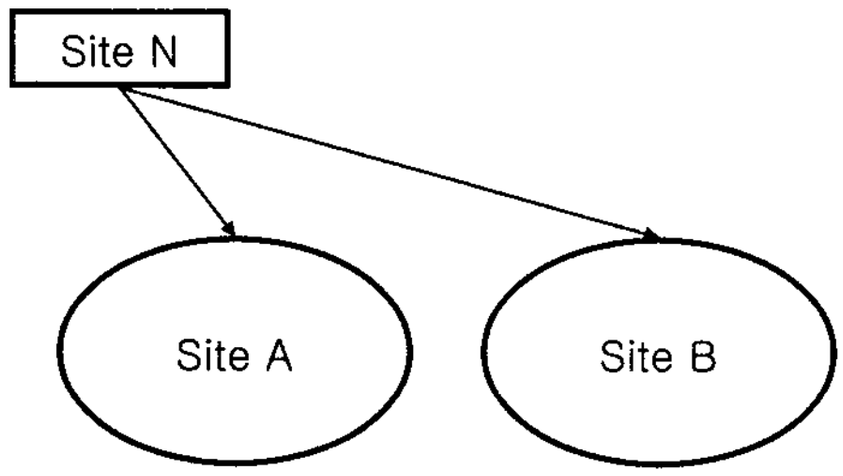
An inlink is a link to a web page, meaning that the link should come from outside of the specified web page(Thelwall, 2004)(see Figure 1).



(Site 1, 2, ..., n are inlinks of site A)

〈Figure 1〉 Ininks between web sites

When two pages both have links from a third page, we can say the two pages are co-linked(or are co-inlink sites)(Thel-wall, 2004)(see Figure 2).



(Site A and B are co-linked)

〈Figure 2〉 Co-linked web sites

In this study, web link analysis of the Korean National Archives has been performed to show the dynamic web flow of the National Archives and to identify which closely related institutions the archives should seek to cooperate with for developing an institutional strategic plan. Web link analysis of National Archives abroad was also performed to compare with the Korean one.

At present, there are three types of research about web link analysis for categorizing or comparing related web sites. All of this research attempts to define the relationship among web sites using link data or classification systems.

It is possible to use web link information for the analysis of inter-relationships among institutions or research areas. Thelwall(2004) mainly detected research trends in universities and relationships among research institutions using university web site link information. Björneborn(2001) analyzed researcher homepages, published bookmark lists and found 'weak ties'(called 'so called co-linkage chains' by him). Adding to this, Jung and Chung (2007) performed comparative analysis of interdisciplinarity in eight scientific subject areas using link information. Both of these papers used link

information for setting up relationships among web sites.

Cho & Kim(1997) and Oh & Myaeng(2002) used link information for measuring document similarities and categorized web pages. This research also calculated link information for web document analysis.

There are also researchers focused on domain analysis. Koehler(1999) and Yang et al.(2002) used information shown in domains. In particular, Koehler described the structural characteristics of URLs below(1999).

▸ transport medium://
server.level.domain/directory/...

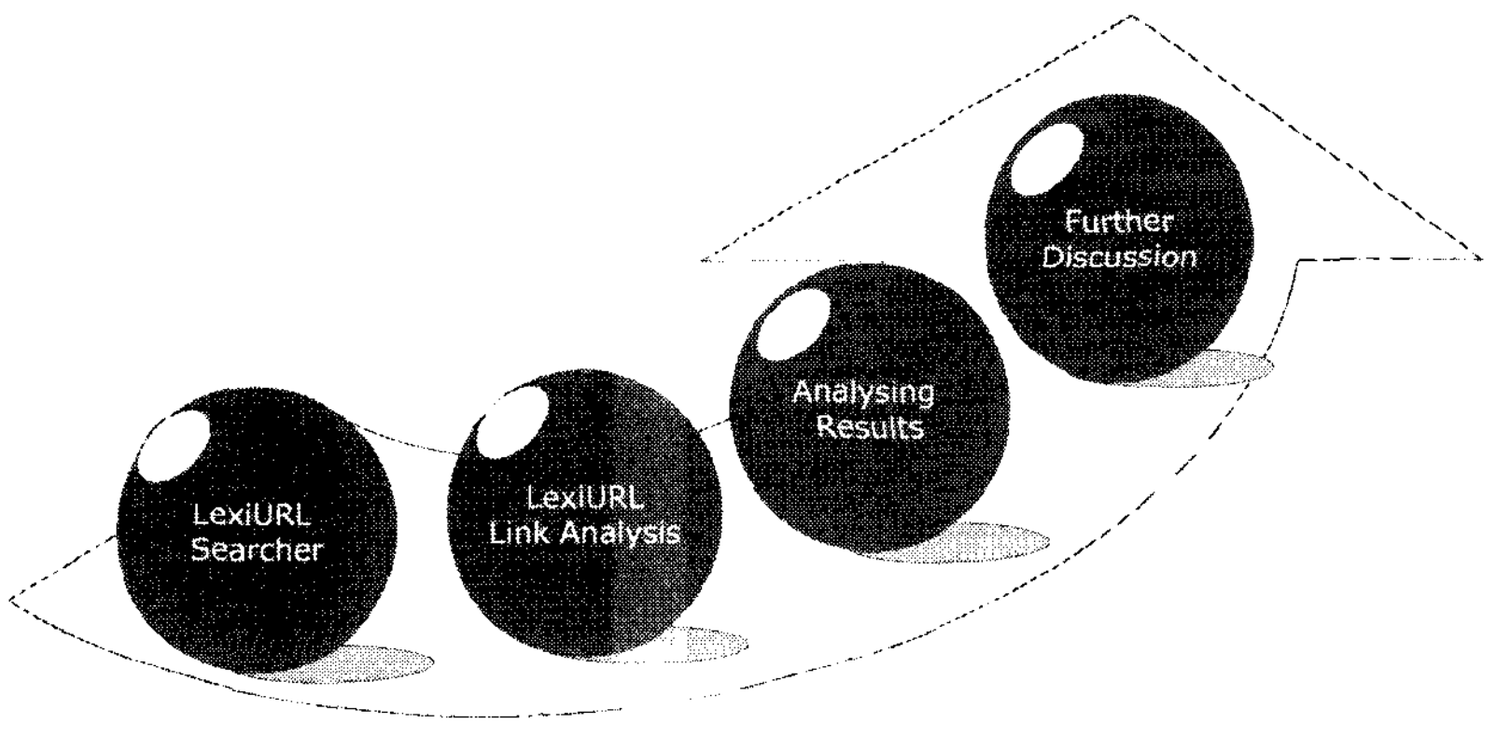
- Transfer protocol: "transport medium" can be used for cataloging purposes(MARC 856 field ($2)). ex) gopher, http, ftp etc.
- Server-level domain(SLD): A form of address that is resolved to the underlying IP number.
- Top-level domain(TLD): Publisher type and rightmost-portion of the server-level domain.
  - TLD Type1) Functional TLD: .com, .edu, .gov, .net etc.
  - TLD Type2) Geographic TLD: kr(Korea), fr(France) etc.

Using this URL structure, we can classify web sites by country, publisher etc.

2. Research Methods

In this study, LexiURL Searcher, launched in March 2007, has been used to collect link data from Yahoo. After that, through LexiURL link analysis, inlink and co-inlink information has been obtained.

The resulting link information has been categorized into areas and countries according to domain. SLD and TLD were used for domains, and some link information has been acquired through a direct check of sites for detailed analysis. Each step of this procedure is detailed in <Figure 3>.



〈Figure 3〉 Overview of Research

2.1 Data Collection

Link data has been collected through Yahoo sites using Yahoo Site Explorer and LexiURL. SocSciBot also can be used as a link crawler, but with large sites it is more time-consuming to collect inlink information than LexiURL, and inlink information of target sites cannot be checked in detail. For this reason, LexiURL has been used in this study. LexiURL has limits in crawling options and amount of collected data, but is better in analysing collected inlinks and is faster than SocSciBot in crawling time.

Before analysing collected data, crawling inlink data can be performed with both Yahoo Site Explorer and LexiURL Searcher. When using Yahoo Site Explorer, it is possible to save crawling data in .tsv(Tab Separated Values) format and analyse the results with LexiURL. However, this method cannot be applied with Korean letters, Hangeul. For this reason, Yahoo API for LexiURL Search is used for data crawling.

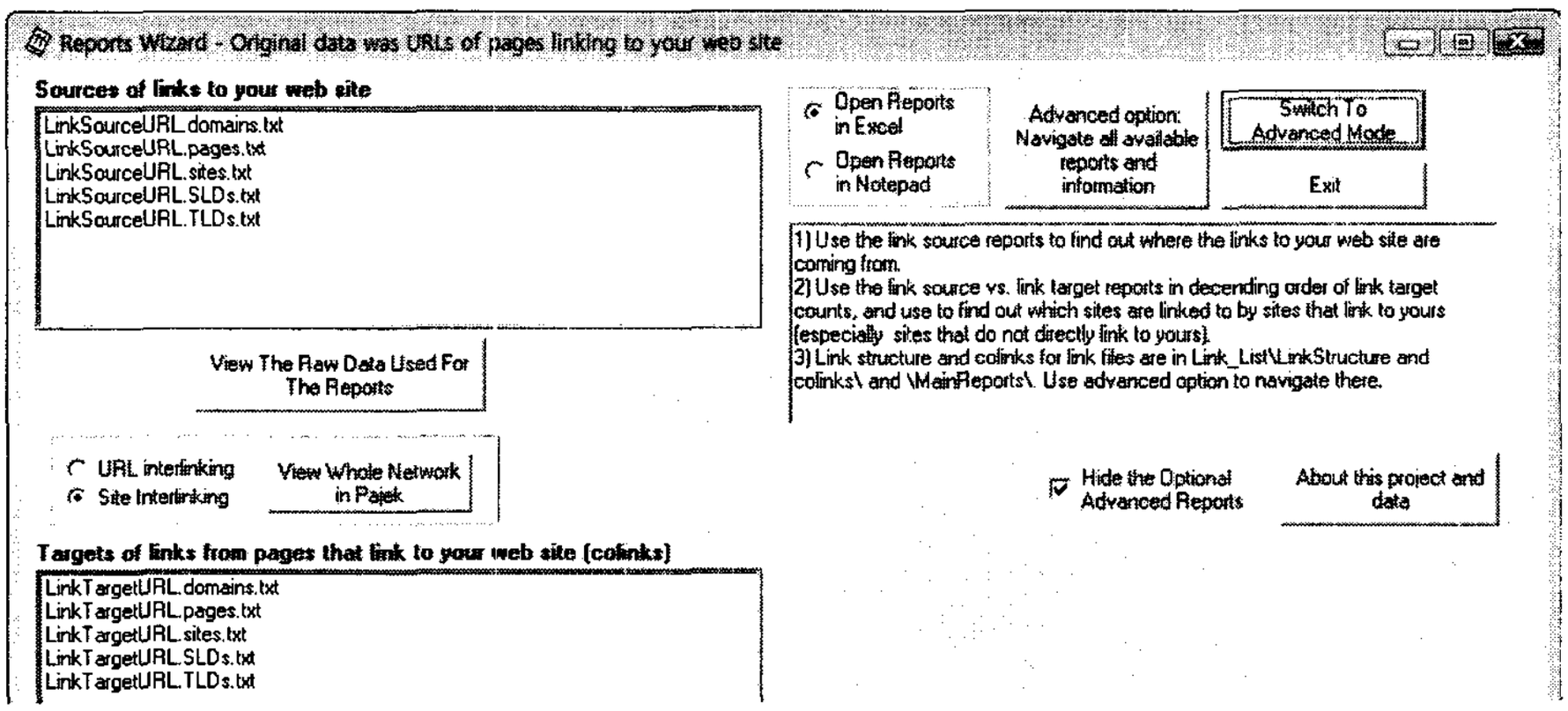
For data analysing, LexiURL has a simple interface(see Figure 4) and the results are provided in two types below:

- sources of links(inlink)
- targets of links(co-inlink)

Furthermore these two types of reports are classified by domain, web page and web site.

With the Korean National Archives, it took 20 minutes to analyze 666 files(self-links were excluded). LexiURL Searcher crawled a total of 1000 links but 334 files were missing URLs or non-ASCII webpage.

In this study, among reports provided by LexiURL, inlink information by web site, SLD, and TLD reports were used. Inlink information by site(LinkSource URL_sites.txt, LinkTargetURL_sites.txt) indicates domain names which include TLDs and SLDs.



〈Figure 4〉 Link Analysis Interface - Report Wizard

2.2 Link Data Classification

Domain names were used for link data classification. For the TLD, which means top domain, nation-based categories were extracted, and for the SLD, second top domain, subject area categories were extracted. Added to this, manually classified categories are also used for detailed link analysis.

Classification Category Mapping by Domains

- TLD classification category(nations): Categories were produced using org., edu., gov., which are limited for registration. National top-level domain codes, a 2-character nation code which is defined by ISO-3166, were used for the national categories.
- SLD classification category(subjects): Categories were produced using or., ac., go., which are limited for registration, and the regional KR domain(in the case of Korea).

Manually Classified Category

For detailed analysis of the Korean National Archives, top 30 inlinks and co-inlinks were manually categorized by information professionals(see Table 1).

These four categories can be mapped by the types of users of the National Archives such as archivists, public officers, students, and information professionals. There are more categories - personal homepages and portal sites, and these are dealt with as a minor part.

3. Link Analysis Results

In this part of the study, three types of data were analyzed for the four countries.

- Target countries: Rep. of Korea, Australia, United Kingdom, United States.
- Types of analysis
  - ① inlink by domain, SLD, TLD
  - ② co-inlink by domain, SLD, TLD
  - ③ comparison of sources vs. target link(inlink & co-inlink)
  - ④ comparison of results for Korea and other countries

Each type of analysis was conducted in two aspects - the top 30 web site analysis(only for Rep. of Korea) and domain type analysis. The former focused on the top 30 inlinks of the Korean

〈Table 1〉 Assigned Categories

Category
Archival institutions: Archival institutions except the Korean National Archives
Administrative institutions: Governmental department
Educational institutions: Universities etc.
Cultural institutions: Libraries, museums etc. (except archival institutions)

National Archives and the latter categorized the web sites using domain information.

3.1 Korean National Archives

3.1.1 Inlink analysis

With an inlink of the Korean National Archives, an inlink indicates a link from another site to the Korean National Archives. In this study, inlink analysis was conducted by site and domain.

Inlink analysis by site

In the 255 inlink web sites of the Korean National Archives, 136 sites with more than 30 link pages were analysed in detail, and types of sites and nationalities were extracted. Among the 136 sites, 107 sites(excluding personal blogs/homepages, news sites, portal sites) were analysed and the results were shown in the order of Administrative institutions > Educational institutions > Archival institutions > Cultural institutions(see Table 2).

Administrative institutions are top in the order. This is because most government web sites have link pages for the Korean government organization chart and the Korean National Archives is one of the public institutions. Therefore this type of link is generated automatically in the context of a member of the Korean government, not its functionaries or specific archival services.

Regarding links from the educational institutions, department web sites of archival science or history made links for the National Archives, some educational institutions have National Archives links on job information.

Archival institutions were third in inlink number and the links mostly came from abroad(most from China; others include Iran, Turkey, and Mexico).

The other sites(38%) are mostly commercial sites such as private institutions for examination. It is possible to think that it shows some aspects of training archival professionals in Korea but not considered in this study. Commercial sites also include architecture web sites as these sites have 'favorite links' to Korean government organizations(it is the same reason that the Korean National Archives has inlinks from other government web sites).

From the above analysis, the context of the inlink can be summarized in the order of 'as a government organization' > 'as a related organization' > 'as an archival institution'.

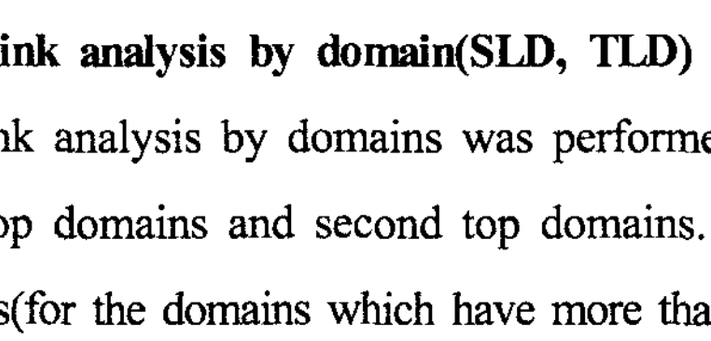
〈Table 2〉 Inlink Categories

Site Type	Rate(%)
Administrative institutions	21%
Educational institutions	20%
Archival institutions	12%
Cultural institutions	9%
The others	38%

Most links point the main page(www.archives.go.kr) and links to the archival resources or specific services provided by the Korean National Archives are hard to find.

Inlink analysis by domain(SLD, TLD)

Link analysis by domains was performed for the top domains and second top domains. TLD results(for the domains which have more than two links) show in the order of 'kr' > 'com' > 'jp' = 'cn' > 'org'(see Figure 5).

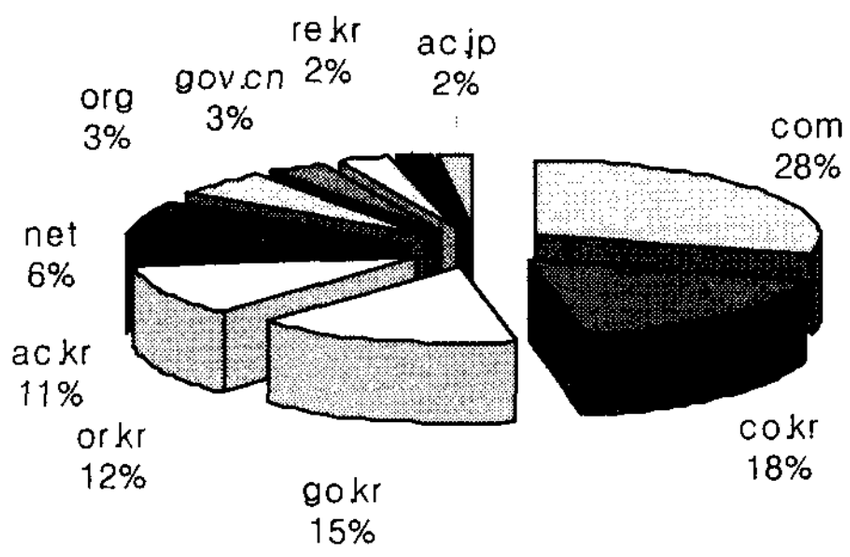


〈Figure 5〉 TLD inlink(Kor)

Among domains, 'kr' and 'com' are at the top. This is because many public institutions, universities, and commercial sites often use these two domains.

As for Chinese domains, most links were connected with the Chinese National Archives. On the other hand, in the case of Japan, most were personal blogs.

For SLD, the results are similar with the TLD(domains linked more than four times were analyzed) and show in the order of 'com' > 'co.kr' > 'go.kr' > 'or.kr' > 'ac.kr'(see Figure 6).



〈Figure 6〉 SLD inlink(Kor)

'com' and 'co.kr' are highly linked because these are very common domains in commercial sites. The domain 'or.kr' representing research institutions and 'ac.kr' representing university sites are 23% of the total and 'go.kr' representing government institutions is 15%. The domain analysis also indicates that the relationship with the governmental and educational institutions is strong.

One interesting point is that most Chinese domains were represented as 'gov' and most Japanese domains were represented as 'ac'. This indicates that these are links centered on the Chinese National Archives and links related to Japanese universities.

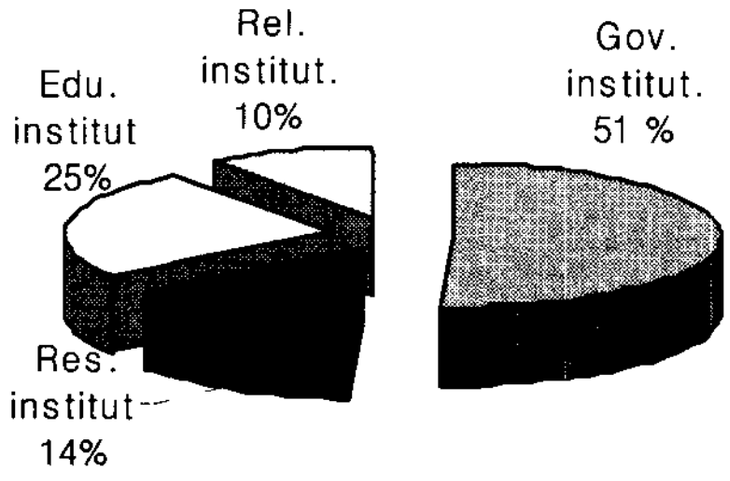
3.1.2 Co-inlink analysis

In this chapter, one of the two sites is the Korean National Archives and the other is analysed by site and domain.

Co-inlink analysis by site

118 co-linked sites with more than 7 link pages were analysed by site type and country. In 118 co-linked sites, 65 sites except portals and personal blogs were included in the results(see Figure 7).

More than half of the co-linked sites of the Korean National Archives have links to the other governmental institutions(51%) and the most domestic sites. We therefore can see that the links of the Korean National Archives are provided with other administrative institutions(not cultural or information management institutions).

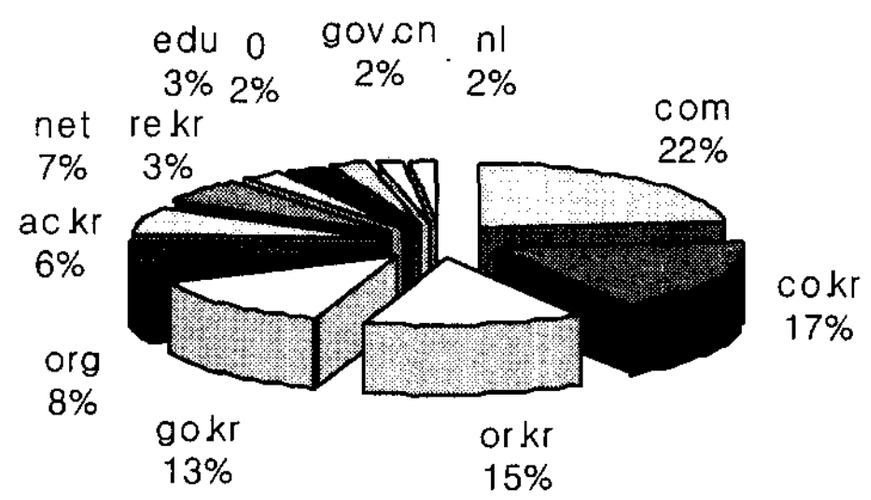


〈Figure 7〉 Co-inlink by Site(Kor)

The proportion of co-inlinks with educational or research institutions was next. It is worth considering that inlinks from university sites were mainly domestic one but co-linked university sites were located in the United States or China. The difference between inlinks and co-inlinks was caused by the web page of 'directory of Archives and Records Management Resources' containing National Archives link and related educational institutions (domestic, international), and research institutions on the same web page. Related institutions - libraries, museums etc. were contained with the links to the Korean National Archives as well.

Co-inlink analysis by domain(SLD, TLD)

As for co-inlink by TLD, among 160 domains, 10 domains having more than 35 links were analyzed. The most frequent was kr, with others following in the order of com>org>net>cn>jp(see Figure 8).

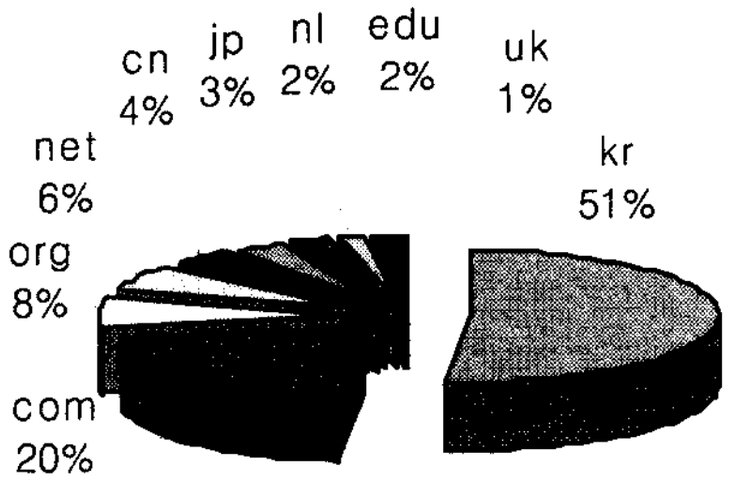


〈Figure 8〉 SLD Co-inlink(Kor)

A key difference here from inlink is the appearance of the 'edu' domain. French national domains have been removed and domains from England and the Netherlands are shown. Most 'edu' domains were universities in the United States and China. This means educational institutions other than universities seldom link to the National Archives in Korea, but there are cases of the third institutions simultaneously linking to the National Archives and to universities abroad.

SLD analysis by domain was performed for web site which have more than 100 links based on domain. The result shows a pattern similar to the inlink SLD analysis results. It also show a similar shape of details of domains which have many links in TLD(see Figure 9).

There were decreases in 'ac.kr' and increases in 'org'. domain compared with Inlink analysis by domain(see Figure 6 and 8). Moreover, the portion of Chinese domains was decreased and Japanese domains were not ranked in the top level.



〈Figure 9〉 TLD Co-inlink(Kor)

This shows that Korean National Archives web site is closer to institutional web site than university web site. Link for 'edu.cn' domains numbered 97 and for 'ac.jp' domains 94, which shows a similar inlink proportion with China and Japan.

3.1.3 Comparison of inlink and co-inlink

A comparative analysis for second top domains was performed by SLD using inlink and co-inlink.

There is a big difference in link numbers between inlink and co-inlink. Therefore the number of links was divided by the total number of link numbers in each column, and then multiplied by 100. After that, 10 from the top were selected among inlink and co-inlink. 'difference' refers to the sequence of the numbers according to the order. These numbers are inlink minus co-inlink(see Table 3).

As shown in Table 3, top ranked domains are similar, and they show a similar pattern with individual inlink and co-inlink analysis results acquired in advance. The biggest difference is found in the 'ac.kr' domain, which indicates Korean university web site. This means that Korean university web site seldom link to the National Archives, and the third web site seldom link simultaneously to universities and the National Archives, either.

3.2 Comparison with National Archives Abroad

The link of the Korean National Archives was

〈Table 3〉 Normalized sources/ target link

SLD	inlink	co-inlink	Differ.
ac.kr	8.89	4.81	+4.08
com	22.96	19.17	+3.79
go.kr	11.85	10.76	+1.09
ac.jp	1.48	0.99	+0.76
ne.kr	1.85	0.56	+0.49
net	5.93	5.88	+0.05
gov.cn	2.22	1.46	-0.83
co.kr	14.07	14.95	-0.88
re.kr	1.48	2.38	-0.9
edu	0.37	2.16	-1.79
or.kr	9.26	12.15	-2.89
org	2.96	6.52	-3.56

compared with three foreign National Archives, those in Australia(NAA-National Archives of Australia), England(TNA-The National Archives), and the United States(NA-National Archives). Abbreviations were used in the table below are 'kr' for Korea, 'au' for Australia, 'uk' for England, and 'us' for the United States.

The commercial '.com' domain has the highest proportion regardless of county, but is not mentioned here because the domain gives little implications for this study.

3.2.1 Inlink comparison

TLD Comparison of inlink

Top five TLDs of the four countries were selected according to link numbers(see Table 4).

With foreign countries it is more difficult to analyse TLD than Korea because most TLD belong to the agency located in United States. In these circumstances, the results of the comparison of TLD shows as follows.

- national domains: With Korea and England, the proportion of national domains is high. With Korea, the proportion of foreign national domains such as 'cn' and 'jp' is high.
- educational institutions: There is no edu domain for Korea, because in Korea the domain 'ac.kr' is used for universities.
- groups/institutions domains: For this type of domain there is no limitation on registration. Therefore, only with domains, web site characteristics can not be detected. 'Org' domains were frequent in the order of 'us' > 'uk' > 'au'. There being no Korean domain is because that inlink of the Korean National Archives appear mainly on the web directories of public institutions.

SLD Comparison of inlink

A top ten SLDs comparison of inlink data for the four target nations was performed. In this comparison, domains were analyzed instead of sites(Table 5).

As shown in <Table 5>, the 'com' domain is top for all nations. The characteristics of the remaining nine domains are as follows:

- national domains: the Korean domain, 'kr', showed in the order of 'co.kr' > 'go.kr' > 'or.kr' > 'ac.kr'. With Australia, the 'au' domain showed in the order of 'co.au' > 'gov.au' >

〈Table 4〉 TLD Comparison of inlink data

TLD_KR	Sites	TLD_AU	Site	TLD_UK	Sites	TLD_US	Sites
kr	132	com	154	uk	245	com	266
com	59	au	259	com	205	org	191
net	12	net	30	org	92	edu	79
jp	13	org	62	edu	57	net	60
cn	11	edu	33	net	43	us	38

〈Table 5〉 SLD Comparison of inlink data

SLD_KR	Domains	SLD_AU	Domains	SLD_UK	Domains	SLD_US	Domains
com	62	com	168	com	215	com	281
co.kr	38	com.au	97	co.uk	104	org	191
go.kr	32	gov.au	80	org	92	edu	106
or.kr	25	org	64	org.uk	69	net	63
ac.kr	24	org.au	58	edu	64	gov	53
net	16	edu.au	46	gov.uk	45	mil	18
org	8	edu	36	net	43	nl	5
gov.cn	6	net	32	ac.uk	35	co.uk	4
ne.kr	5	net.au	20	ca	12	tx.us	3
ac.jp	4	co.uk	19	ie	6	be	3

'edu.au' > 'net.au'. With England, 'uk' showed in the order of 'co.uk' > 'org.uk' > 'gov.uk' > 'ac.uk'. As for the United States, 'us' showed in 'tx.us' only.

- governmental/public institution domains: In the case of Korea, the number of links from Korean and Chinese public institutions was high. Australia also shows many links. England and the US show comparatively fewer.
- educational institution domains: the US shows the largest number of educational institutions domains. After the US, the order is 'uk' > 'au.kr'. In Korea, there are 'ac.kr' and 'ac.jp' domains; in Australia, 'edu.au' and 'edu'; in England, 'edu' and 'ac.uk'; and in the United States, there is the 'edu' domain.
- foreign domains: in TLD, only Korea hosted domains from foreign nations. In SLD, England, Irish and Canadian domains are shown. With the United States, English, Dutch, and Belgian domains appear. In Korea, Chinese public institution domains and Japanese university domains are shown.
- etc: In the US, unlike in other countries, a 'mil' domain is used to indicate military service organizations.

3.2.2 Co-inlink comparison

TLD Comparison of co-inlink are as follow(see Table 6):

- national domains: The ratio of links with other web site within the country appears highest in Korea. With Korea, there were many co-inlink cases with China and Japan, which geographically are located closely. With Australia, the co-inlink ratio with England was highest, followed by the Netherlands, Canada, and New Zealand. England co-linked with Germany and Australia the most, followed by the Netherlands, Canada, and Ireland. In the US, co-linked with England or Germany was most frequent, followed by the Netherlands and Canada. English and Dutch domains commonly appeared in every nation.

〈Table 6〉 TLD Comparison of co-inlink data

TLD_KR	Sites	TLD_AU	Sites	TLD_UK	Sites	TLD_US	Sites
kr	3727	com	8740	com	9348	com	19759
com	1499	au	6038	uk	5930	org	10812
org	562	org	3755	org	4876	net	2121
net	438	uk	1694	net	1316	uk	1100
cn	297	net	1100	de	1023	de	999
jp	225	edu	697	au	567	edu	952
0	184	de	436	edu	559	us	790
nl	129	nl	390	nl	354	gov	655
edu	116	ca	309	ca	283	nl	329
uk	39	nz	297	ie	248	ca	277

• educational institution domains: In Korea, the co-inlink rate with edu domains is relatively low compared with other countries.

In the case of SLD, a more detailed comparison can be performed than the case of TLD(see Table 7). The result of SLD comparison is same in general as with TLD, but here the analysis is focused on the differences.

• educational institution domains: in Korea, public web site are very often co-linked. However, the ratio of co-inlink with educational institutions is relatively low compared with public institutions. Korea shows a low co-inlink ratio of educational institutions compared with Australia, England, and the United States.
• public institution domains: In Korea, the ratio of cocitation with public institutions is high compared with other countries. Australia shows a similar pattern to Korea, which shows a high portion of public institutions. Only Korea shows a high ratio of co-inlink with foreign(Chinese) public institutions.

4. Conclusion

Three results were derived from link analysis in order to develop the policies of the Korean National Archives.

Educational function

The Korean National Archives shows a lower number of links from educational institutions than is the case in other developed countries. Current links from educational institutions are focused on universities only. Moreover, most links from educational institutions show the URL of the homepage, not the URL of a specific service.

Therefore, the National Archives should develop educational content to reinforce links from educational institutions(such as links from 'es.kr'(elementary

〈Table 7〉 SLD Comparison of co-inlink data

SLD_KR	Domains	SLD_AU	Domains	SLD_UK	Domains	SLD_US	Domains
com	1821	com	9842	com	11357	com	23344
co.kr	1420	com.au	4198	org	5490	org	11956
or.kr	1154	org	4021	co.uk	3152	edu	3873
go.kr	1022	edu	2099	org.uk	2119	gov	2497
org	619	gov.au	1455	edu	1979	net	2392
net	559	net	1281	net	1485	de	1241
ac.kr	457	co.uk	1020	de	1293	co.uk	638
re.kr	226	org.au	981	gov.uk	917	mil	598
edu	205	ca	756	ac.uk	855	ca	591
gov.cn	139	edu.au	688	ca	666	nl	489

schools), 'ms.kr'(middle schools), and 'hs.kr'(high schools)). Moreover, several services using educational content should also continue for sustained impact in citing materials developed by Korean National Archives.

This should be considered to provide valuable materials of the Korean National Archives to connect more efficiently to extra educational curricula.

Research function

Most inlinks and co-inlink of the Korean National Archives were from public institutions, indicating that relationships with research institutions are weak. The National Archives of Australia also shows similar pattern compared with United Kingdom or United States. This signifies a relatively weak relationship with research institutions. Conversely, the United States, England, and Australia show active web links from the third institutions, not only limited to public institutions.

With the Korean National Archives, links from research institutions were those of heritage institutions or institutes of historical research. It is natural to have a relationship with these institutions, development of multiple user groups using materials of Korean National Archives is needed.

Public records of government departments have various subject backgrounds and manuscripts from many donators can be a rich source of research as well. More links from research institutions will increase the reputation of archives for supporting research activities.

International activities

Inlinks and co-inlinks of Korean National Archives are from China or Japan. This is natural given that Korean, China and Japan are adjacent countries and form a cultural sphere using Chinese Characters. English speaking cultures(Australia, United Kindom and United States) also have links to and from each other.

However, co-inlink analysis(SLD) indicates something important. This is that the countries of English speaking culture are co-linked from web sites of third countries, but the Korean National Archives receive few links from third countries.

This means that Korean National Archives involve fewer web links related to international activities than the countries of English speaking culture. There is a need for more cultural interchange among related institutions.

Summary and suggestion

A summary, through web analysis, web citation and trends in other countries was produced, and compared with Korea to identify our country's current position. Future suggestions to develop the National Archives are those to fortify education, research functions and international activity. This study showed the possibility of applying web link analysis to suggest the future direction and policies of a current representative national organization, the Korean National Archives.

Further research is needed in different heritage institutions link libraries or museums. Currently, the web linkage among these heritage institutions is weak in Korea, but in the future we must consider the service linkage among these related institutions more seriously.

References

Korean National Archives. [cited. 2007.12.24]. <http://www.archives.go.kr>.

Björneborn, Lennart. 2001. "Small-world linkage and co-linkage." Conference on Hypertext and Hypermedia(Session: 3b-Short Papers), 133-134.

Cho Kwangje, Kim Juntae. 1997. "Automatic Categorization of HTML Documents by using Hyperlink Information." *Proceedings of the Korea Information Science Society Conference*, 24(2): 277-280.

Jung Ho-Yeun, Chung Young-Mee. 2007. "A Comparative Study on Interdisciplinarity in the Fields of Science and Technology Based on Journal Citation and Web Link Analyses." *Journal of Korea Society for Information Management*, 24(3): 179-200.

Koehler, Wallace C. Jr. 1999. Classifying Web Sites and Web pages: the use of metrics and URL characteristics as markers.

LexiURL - Links to a Web site or web page case study. [cited. 2008.5.23]. <http://lexiurl.wlv.ac.uk/help/linksToWebSiteCa-seStudy.htm>.

NA - National Archives. [cited. 2007.12.24]. <http://www.archives.gov/>.

NAA - National Archives of Austalia. [cited. 2007.12.24]. <http://www.naa.gov.au/>.

Oh Hyo-Jung, Myaeng Sung-Hyoun. 2002. "A Hypertext Categorization Method using Incrementally Computable Class Link Information." *Journal of the Korea Information Science Society*, 29(7, 8): 498-509.

Thelwall, Mike. 2004. Link Analysis: An Information Science Approach. Elsevier. chapters: 7-9.

TNA-The National Archives. [cited. 2007.12.24]. <http://www.nationalarchives.gov.uk/>.

Yiming Yang, Sean Slattery, & Rayid Ghani. 2002. "A study of approaches to hypertext categorization." *Journal of Intelligent Information Systems*, 18(2, 3): 219-241.