

# 웹페이지의 지속성에 관한 연구

## A Study on the Persistence of Web Pages

장우권, 전남대학교 문헌정보학과, wk1961@jnu.ac.kr  
Woo-Kwon Chang, Dept. of LIS, Chonnam National University

이 연구는 인용된 웹자원에서 웹사이트의 지속성을 알아보기 위한 것이다. 연구에서 사용된 데이터는 한국정보관리학회지에 수록된 각 연구논문의 참고문헌이다. 연구논문 대상 참고문헌은 2001년부터 2002년까지이다. 이를 통해 웹페이지의 콘텐츠 효용성을 알 수 있을 것이다.

### 1. 서론

연구자들이 글을 쓸 때 수 많은 문헌정보가 필요하다. 이러한 정보자원은 어디에서 찾고 어떻게 활용할 것인가에 대한 탐색정보도 중요하다. 인터넷과 월드와이드웹(WWW)의 출현은 정보자원의 탄생을 기하급수적으로 가져왔고 이를 활용한 학술도구와 매체들 또한 많은 변화와 발전을 가져왔다.

연구자들은 학술적 가치가 높은 정보들을 토대로 궁리와 궁리 속에 이론의 틀을 완성하고 실무에 적용할 수 있는 실용적 지식을 생산한다. 이러한 지식의 원천이 문헌정보와 참고문헌이다.

참고문헌은 연구자가 참고한 문헌자료로서 단행본, 저널, 학위논문과 글, 기사 등으로 다양한 연구적 요소와 그 내용들을 함유하고 있다. 연구에 필요한 지식의 원천으로서 중요한 참고문헌은 1990년대 말 이후 많은 웹자원으로서 수록되고 있으며 연구의 가용성과 효용성을 나타내고 있다.

따라서 이 연구에서는 각 연구논문에서 수록된 참고문헌에 나타난 웹자원의 페이지(웹 콘텐츠)의 지속성을 알아보기 위한 것이다. 즉 연구결과를 발표당시에 수록된 웹페이지가 현

재까지 존재하는지 여부를 조사하여 분석하였다.

### 2. 이론적 배경

#### 2.1 참고문헌

참고문헌(參考文獻, references)은 논문, 도서 또는 도서의 각장의 끝에 저자가 참고한 자료목록을 말하며 통상 본문에 표시한 번호와 동일한 순서로 배열한다(문헌정보학용어사전 2010:330). 또한 학회지에 따라서 주석을 참고문헌으로 대체하는 경우도 있다 이 경우도 참고문헌에 포함한다.

#### 2.2 인터넷정보와 웹페이지

인터넷 정보는 웹페이지의 형식안에서 조직되고 제공된다. 웹페이지는 HTML, XML 등의 마크업언어(컴퓨터 프로그래밍 언어)로 작성되며 그림, 소리, 동화상, 하이퍼링크(문서이동) 등이 가능하고 서로 접근점을 가지고 연결되어지며 정보를 담은 하나의 구조(하이퍼텍스트)를 이룬다(조찬식, 한혜영 2005:91) 즉 웹페이지(web page)는 자원통일소재기호(URL)

로 접근 가능한 웹상의 하이퍼미디어 문서이다(문헌정보학용어사전 2010:244). 또한 이를 웹문서라고도 칭한다.

일반적으로 URL과 웹페이지의 구조는 다음과 같이 나타난다.

①	②	③	④
<a href="http://www.loc.gov/marc/dcqualif.html">http://www.loc.gov/marc/dcqualif.html</a>			
①프로토콜 ②정보자원을 가진 컴퓨터의 위치			
③파일 디렉토리 ④자원이름			

### 3. 연구 문제와 연구 방법

#### 3.1 연구 문제 및 연구 모형

이 연구에서는 참고문헌에 나타난 웹페이지의 지속성(durability) 여부를 다루었다. 여기에서는 웹페이지에 대한 정보획득차원에서 URL(Uniform Resource Locator, 자원 위치 지정자)와 웹페이지 존재 여부를 알아보기 위해 다음과 같은 연구문제를 설정하였다.

1. URL에 대한 지속성 여부는 어떠한가?
2. 웹페이지의 내용에 대한 지속성 여부는 어떠한가?

#### 3.2 조사 항목 설계

참고문헌에서 일반문헌과 웹페이지 여부를 알기 위해 1차와 2차로 나누어 한국정보관리학회지에 수록된 연구논문을 조사하였다. 1차에서는 수록된 연구논문의 참고문헌에서 일반문헌과 웹페이지 등을 파악하였다. 1차 조사 후 2차로 URL과 웹페이지 지속성 여부를 조사하였다.

### 3.3 연구방법

이 연구에서 사용된 연구의 표본은 정보관리학회지 제17권(1, 2, 3, 4호; 2000년 3월, 6월, 9월, 12월)과 제18권(1, 2, 3, 4호; 2001년 3월, 6월, 9월, 12월)에 수록된 각 연구논문의 참고문헌이다. 조사기간은 1차와 2차로 나누어 실시하였다. 1차는 2013년 6월 15일부터 30일까지, 2차는 동년 7월 1일부터 25일까지이다. 조사방법은 위의 학회지에 수록된 각 연구논문의 참고문헌을 하나하나 면밀히 대조하며 일반문헌과 웹페이지들을 계량화 하였다. 이후 URL과 웹페이지들을 일일이 타이핑하며 존재여부를 탐색하였다.

조사내용은 1차와 2차로 구분하였으며, 1차는 참고문헌에서 일반문헌과 URL 여부, 2차는 URL과 웹페이지 내용 존재 여부로 구성되었다. 자료 분석은 SPSS 18.0 프로그램을 이용하였다. 연구문제에 대한 빈도(N), 퍼센트(%) 등의 기술 통계량과 그 값을 산출하였다.

## 4. 조사분석 결과

### 4.1 1차 조사

#### 1) 참고문헌

정보관리학회지 제17권(2000년)에서는 44개의 논문에 979개의 참고문헌이 수록되었다. 여기에서 일반문헌이 801개, 웹페이지가 178개로 나타났다. 제18권(2001년)에서는 45개의 논문에 983개의 참고문헌이 수록되었으며, 이중 794개가 일반문헌, 189개가 웹페이지이다. 일반문헌에는 단행본, 학회지 논문, 기사, 사전 등이 포함된 것이다. 전체 49개의 연구논문에 1,962개의 참고문헌이 수록되었다. 평균적으로 논문 당 참고문헌 수는 40개로 나타났다. 이중 일반문헌은 81.4%(1,595개), 웹페

이지는 18.6%(367개)이다.

(표 1) 참고문헌 분포도

연도	논 문 수	참고문헌						계	
		일반문헌			URL				
		국문	일문	영문	국문	일문	영문		
2000	44	280	5	516	10	9	159	979	
2001	45	307	8	479	43	0	146	983	
계	수	89	587	13	995	53	9	305	1,962
	%		29.9	0.7	50.7	2.7	0.5	15.5	100

2) URL

367개의 URL중 14.4%(53개)가 한글 정보 자원이며, 일문은 2.5%(9개), 영문이83.3%(305개)로 나타났다.

4.2 2차 조사

1) URL

1차 조사에서 수록된 367개의 URL 중 208개(56.7%)는 지속적으로 존재하고 있고, 60개(16.3%)는 변경되어 나타났다. URL 변경의 원인은 웹정보자원을 가진 컴퓨터의 위치 즉 기관명의 변경으로 기관의 성격은 변화되지 않은 것이다.

예) ① <http://www.nso.go.kr/> (☞) 통계청 CI 개편 및 홈페이지 주소가 2009.9.1.부터 아래 주소로 변경되었습니다.

바로가기 : <http://kostat.go.kr>

② <http://www.arttoday.com/>(☞) 주소변경  
<http://www.clipart.com/en/>

또한 기관의 성격이 변화되어 나타난 것도 있다. 이것은 도메인 주소는 같으나 기관의 역할이 전혀 다르게 나타난 경우이다.

예) <http://user.chollian.net/>(☞) 웹자원 DB는 없고 홈페이지 관리가 주.

반면에 URL이 존재하지 않은 것은 99개(27.0%)였다.

(표 2) URL 분포도

연도	논 문 수	URL			계	
		존재		미 존재		
		존재	변경			
2000	44	107	29	42	178	
2001	45	101	31	57	189	
계	수	89	208	60	99	367
	%		56.7	16.3	27.0	100

2) 웹페이지

(표 3) 웹페이지 분포도

연도	논 문 수	웹페이지			계	
		존재		미 존재		
		원소스 (URL)	다른 웹페이지에 존재			
2000	44	48	1	129	178	
2001	45	53	5	131	189	
계	수	89	101	6	260	367
	%		27.5	1.6	70.9	100

웹페이지 내용여부에서 27.5%(101개)는 12~13년이 지난 현재도 존재하고 있으며, 1.6%(6개)는 변경되어 존재하고 있다.

Instone, Keith. 2000. Information Architecture and Personalization. <[http://argus-acia.com/white\\_papers/personalization.html](http://argus-acia.com/white_papers/personalization.html)> ☞ Information Architecture and Personalization. An Information Architecture-Based Framework for Personalization Systems. By Keith Instone, Argus Associates December 2000

This white paper demonstrates the use of information architecture components as a foundation for thinking about personalization. After defining the information architecture components, it describes a model that combines the components into a complete

personalization system. This model could be used to guide your personalization system development methodology, evaluate a set of personalization systems, or merely to give you the terminology to help you communicate about personalization.

Please direct your comments and suggestions to Keith Instone (keith@instone.org).

[View or Download White Paper \(51K PDF\)](#)

변경되어 존재한 경우는 URL이 존재하지 않으나 다른 웹페이지에 존재하는 경우이다.

http://educorner.com/courses/ia/ "Information architecture is the process of organizing, labeling, designing navigation and ...

mysite.pratt.edu/~giannini/ia.html.

이것은 관련 유사 기관이나 개인이 학술적 또는 비즈니스적 목적으로 수집하여 웹 콘텐츠를 구축하여 제공하고 있는 것이다. 수록된 웹페이지의 70.9%(260개)가 현재 존재하지 않은 것으로 나타났다.

다음은 발견되지 않은 이유를 설명한 다양한 메시지이다.

① Not Found  
The requested URL /newsletter/korean.shtml/ was not found on this server.  
Additionally, a 404 Not Found error was encountered while trying to use an ErrorDocument to handle the request  
② Resource not found  
Sorry, specified resource could not be found. It is possible it has been deleted. If you believe that this is due to a system error, please let us know and we will look into it.  
Home Back Please wait...  
③'www.cencus.gov/srd/www/abstract/rr94-4.htm'에 대한 검색결과가 없습니다.  
· 검색어의 단어 수를 줄이거나, 보다 일반적인 단어로 검색해 보세요.  
· 두 단어 이상의 키워드로 검색 하신 경우, 정확하게 띄어쓰기를 한 후 검색해 보세요.  
· 키워드에 있는 특수문자를 뺀 후에 검색해 보세요.

## 5. 결론

이 연구는 연구논문 말미에 수록된 참고문헌에서 웹페이지 지속성 여부를 알아보기 위한 것이다. 1차 조사와 2차 조사 결과를 정리하면 다음과 같다.

1) 전체 49개의 연구논문에 1,962개의 참고문헌이 수록되었으며, 평균적으로 논문 당 참고문헌 수는 40개로 나타났다. 이중 일반문헌은 81.4%, 웹페이지는 18.6%이다.

2) 367개의 URL중 14.4%가 한글 정보자원이며, 일문은 2.5%, 영문이 83.3%로 나타났다.

3) URL 중 56.7%는 지속적으로 존재하고 있고, 16.3%는 변경되어 나타났다. 반면에 URL이 존재하지 않은 것은 27%였다.

4) 웹페이지 지속성에서 27.5%는 12~13년이 지난 현재도 존재하고 있으며, 1.6%는 변경되어 존재하고 있다. 또한 수록된 웹페이지의 70.9%가 현재 존재하지 않은 것으로 나타났다.

따라서 중요한 웹페이지는 수시로 웹DB로 구축하여 축적하는 것이 바람직하다. 이것은 연구자의 웹페이지의 가용성과 효용성 측면에서 웹정보자원의 가치를 높이는 것이다.

## 참고문헌

김규환. 2012. 국내학술지 논문에 인용된 인터넷 참고문헌의 특성 및 활동성 연구. 『한국도서관·정보학회지』, 43(3):385-405.  
조찬식, 한혜영. 2005. 웹페이지를 통한 의학도서관 정보서비스 실태조사. 『정보관리학회지』, 22(2):87-101.  
한국도서관협회. 문헌정보학용어사전. 2010.  
Casserly, M. F. & Bird, J. E. Web Citation Availability: A Follow-up Study. LRTS, 52(1):42-53.