

웹 아카이빙 인터페이스 유형 및 특성 분석*

- 국가주도 웹 아카이빙 사업을 중심으로 -

The Analysis of Interface Types and Characteristics of Web Archiving: With a Focus on the Web Archiving Project Initiated by the Nation

김 희 정(Heejung Kim)**

목 차

1. 서 론	3.3 캐나다 GCWA
2. 선행연구	3.4 호주 PANDORA
2.1 웹 아카이빙 선행연구	3.5 싱가포르 WAS
2.2 웹 인터페이스 선행연구	4. 웹 아카이빙 인터페이스 유형 및 특성분석
3. 국가주도 웹 아카이빙 사업 인터페이스	5. 적용방안 및 제언
3.1 미국 LCWA	6. 결 론
3.2 영국 UK Web Archive	

<초 록>

본 연구에서는 해외 국가 웹 아카이빙 사례들을 인터페이스 유형 및 특성들을 항목별로 분석하고 살펴보았다. 이를 위하여 다섯 개의 인터페이스 영역과 열네 개의 평가 항목을 선정하였으며, 해외사례 및 국내 OASIS를 대조하여 분석한 결과, 국내의 경우에는 검색 인터페이스 및 지원기능 인터페이스 영역에의 보강이 필요함을 확인하였다. 또한 해외 사례의 참조를 통하여 국내의 경우 수집 대상 웹 콘텐츠 주제 영역의 다양화, 웹 자원 수집의 분담화, 특별 컬렉션의 구성, 아카이빙 시점 및 흐름의 시각화, 체계적인 아카이빙 전략 수행, 그리고 국제기구와의 적극적 협력 등이 강화되어야 함을 확인하였다.

주제어: 웹 아카이빙, 장기보존, 디지털 보존, 디지털 아카이빙, OASIS, 인터페이스, 이용자 인터페이스, 사용자 인터페이스, 웹 인터페이스

<ABSTRACT>

In this study, national web archiving cases were reviewed and interface types and characteristics were examined. Five interface categories and fourteen evaluation criteria were selected. As for the OASIS, Korean national web archiving project, functions of retrieval interface and user support interface were insufficient and need to be reinforced. Other recommended improvement plans were: diversifications of the subject area of web contents, sharing the web acquisition, make up the special collection, visualization of archiving time and trends, systematic performance of the archiving strategy, and active collaboration with international organization.

Keywords: web archiving, long term preservation, digital preservation, digital archiving, OASIS, interface, user interface, web interface

* 본 논문은 한국기록관리학회 추계학술발표(2010. 10. 29) 내용 중 웹 아카이빙 프로젝트 부분을 재구성하여 정리한 것임.

** 국제백신연구소 정보자료실장(heejung@yonsei.ac.kr)

■ 접수일자 2010년 11월 30일 ■ 수정일자 2010년 12월 21일 ■ 게재확정일자 2010년 12월 22일

1. 서론

최초의 웹 아카이빙 프로젝트인 캐나다 국립 도서관의 Electronic Publications Pilot Project (EPPP)가 1994년에 수행된 이후 16년에 이르는 긴 시간동안 해외 각국에서는 다각적인 웹 아카이빙 사업이 추진되어 왔다.

웹 정보자원은 변경되거나 소멸되기 쉬운 휘발적 성격을 갖고 있으나, 현대 사회에서 지식정보 공유와 커뮤니케이션을 위한 웹에 대한 의존도는 거의 절대적으로 증가하고 있다. 따라서 가치 있는 웹 자원에 대해서는 별도의 아카이빙 작업을 수행하는 것이 매우 필요하다.

이러한 배경 속에서 장기보존의 필요성이 있는 웹 정보자원들은 생산자나 생산기관, 또는 국가들이 주축이 되어 아카이빙을 수행하고 있다.

해외 주요 국가들이 웹 자원의 아카이빙에 대한 중요성을 인지하고 국제적인 협력 움직임을 나타낸 것은 2003년 7월부터인데, 당시 미국, 영국, 캐나다, 프랑스, 호주, 덴마크, 핀란드, 아이슬란드, 이탈리아, 노르웨이 그리고 스웨덴의 국가도서관과 미국의 Internet Archive가 함께 하였다. 이 국가들은 웹 기반 자원을 보존하고 후속 세대에게 전승하기 위해서는 국제 협력이 필요함을 인지하고 웹 자원 보존을 위한 국제컨소시엄인 International Internet Preservation Consortium (IIPC)을 구성하였다.

웹 아카이빙의 궁극적인 목적은 이용자에게 유용한 웹 자원을 제공하는 데에 있다. 즉, 장기간 이후의 미래 이용자 뿐만 아니라 단기간 이후의 이용자에게도 아카이빙된 웹 자원을 적절

하고도 효율적인 방법으로 제공함으로써 웹 자원이 효과적으로 활용되게끔 하는 데에 있는 것이다. 따라서 아카이빙 대상이 되는 웹 자원의 수집 정책 뿐만 아니라 최종이용자를 고려한 적절한 인터페이스 환경 역시 지원되어야 할 필요가 있다. 즉, 장기보존이 필요한 웹 콘텐츠들을 아카이빙한 웹 사이트에서의 지속적인 접근성 및 효과적인 인터페이스의 제공이 필요하다.

이에 본 논문에서는 웹 아카이빙 인터페이스, 즉, 해외 주요 웹 아카이빙 사례 사이트의 인터페이스 유형 및 특성을 분석함으로써 국내의 경우에 적용할 수 있는 개선 방안 및 발전 방향을 제안해보고자 한다. 구체적인 분석을 위하여 웹 인터페이스 선행연구를 중심으로 평가요소들을 선정하였으며, 이를 기준으로 해외 웹 아카이빙 사이트의 인터페이스를 분석하였다.

대상이 되는 해외 웹 아카이빙 사업으로는 국가가 주도가 되는 대표 사례들을 선정하였다. 국가 선정의 기준은 IIPC에 소속된 국가들 중 국가 웹 아카이빙 사이트에의 접근을 공개적으로 허용하고 있는 국가들을 대상으로 하였으며, 그 중 지역별(북미지역, 유럽, 아시아 등) 대표 사례들을 선정하였다.

2. 선행연구

2.1 웹 아카이빙 선행연구

국내에서의 웹 아카이빙 관련 연구들은 2004년부터 간헐적으로 수행되어 왔다. 국가기관을 중심으로 한 공문서 기록 관리에 대한 비중과 관심이 상대적으로 큰 국내의 기록관리 환경

속에서는 공기록 관리에 초점을 맞춘 전자기록 관리 및 디지털 아카이빙에 대한 연구들이 다수 수행되어 왔다. 한편, 웹 기반 정보자원의 장기보존을 지향하는 웹 아카이빙에 초점을 맞춘 연구들은 소수 연구자들에 의해서 수행되어 왔는데, 이에 대한 내용을 정리하면 다음 <표 1>과 같다.

<표 1>에서 살펴볼 수 있듯이, 그동안 국내에서 수행된 웹 아카이빙 연구는 법 제도 또는 정책 관련 연구 및 메타데이터를 중심으로 한 연구들이 주로 수행되어 왔다.

서혜란(2004)은 Kulturarw3, PANDORA 그리고 Internet Archive의 사례 소개와 함께 웹 아카이빙의 기술적, 정책적 과제에 대하여 제시하였다. 김희정 외(2007)는 선택적 웹 아카이빙에 있어서 필요한 관리적 메타데이터 요소들을 제안하였다. 김유승(2007)은 저작권, 납본과 수집, 그리고 진본성을 중심으로 한 법 제도적 문제에 대한 고찰에 이어, 선택적 웹 아카이빙과 포괄적 웹 아카이빙을 병행하는 복합적 웹 아카이빙 정책을 제안하였다(김유승 2008). 오상훈 외(2008)는 웹 아카이빙을 위한 디지털 자원 납본 주체와 대상을 정의하고, 납본 프로세스의 단계별 정의와 기능을 명시하였다. 차승준

외(2008)는 웹 기록물 아카이빙 워크플로우 및 아키텍처 제안에 이어서 공공기관 심층 웹 기록물의 장기보존을 위하여 장기보존포맷과 더블린 코어를 매핑하여 KoDeWeb 메타데이터 요소들을 제안하였다.

2.2 웹 인터페이스 선행연구

웹 인터페이스(web interface)란 웹 서버 기반에서 이용자 대 웹 브라우저와 웹 페이지 간에 이루어지는 상호작용을 의미한다(Computer Desk Encyclopedia).

미국의 심리학자인 Norman(1991)이 이용자 인터페이스에 대한 개념을 인지과학적인 측면에서 제시한 이래 인터페이스 관련 연구는 심리학, 인지과학 및 컴퓨터과학 영역에서 학제적으로 접근하며 활발히 연구가 수행되어 오고 있다. 이에 따라서 이용자 인터페이스에 관한 연구관심의 초점은 하드웨어적인 물리적 제어 요소나 구성 요소에 대한 관점과 콘텐츠의 표현과 해석을 돕는 인지과학적인 정보 전달 관점 등에 주안점을 두고 진행되어 왔다(Myers et al. 1998; Shneiderman 1998; Mitta et al. 1995).

<표 1> 웹 아카이빙 관련 국내 수행 연구주제

저자	연도	연구주제
차승준 외	2009	공공기관 심층 웹 아카이빙을 위한 메타데이터 설계
차승준 외	2008	웹 아카이빙 워크플로우 및 메타데이터 요소 개발
김유승	2008	프랑스 국립도서관을 중심으로 한 복합적 웹아카이빙 정책 고찰
오상훈 외	2008	웹 아카이빙을 위한 납본 프로세스 개발 및 기능 설계
김유승	2007	웹 아카이빙의 법 제도적 문제 고찰
김희정 외	2007	선택적 웹 아카이빙을 위한 메타데이터 요소 개발
서혜란	2004	웹 아카이빙의 해외 사례 소개 및 과제 제시

한편 웹 인터페이스의 평가 및 분석과 관련된 주요 핵심 요소들에 대하여 Vertelney et al은 세가지 요소를 제안하고 있는데, 이는 유용성(사용성)(usability), 시각화(visualization), 기능성(functionality), 그리고 접근성(acceptability)이다. 유용성은 얼마나 쉽고 편리하고, 그리고 정확하게 부여된 업무(task)를 처리하고 제공하는가에 관한 내용이고, 시각화는 웹 사이트를 구성하는 아이콘, 그림, 색상 등에 있어서 불필요한 정보를 최소화하고, 핵심정보를 얼마나 효과적으로 전달하는가와 관련된 내용이다. 기능성은 이용자와의 상호작용성, 피드백 등을 포함한 커뮤니케이션 기능 요소들을 구성하고 있는가와 관련된다. 마지막으로 접근성은 웹 사이트 가독성에 장애를 갖고 있는 사용자층이나 환경에서 대안적으로 제공할 수 있는 접근 루트를 구성하고 있는가와 관련된 웹 접근성 내용이다.

Zhang et al(1999)는 웹 인터페이스 설계와 평가를 위한 이론적 프레임워크를 제시하였는데, 프레임워크를 구성하는 사용자 중심의 웹 인터페이스 세부 영역을 10여개의 항목으로 정의하고 있는데, 이는 직관적 외양(outlook), 기본 기능, 피드백 또는 응답, 접근 제한성, 프라이버시 및 데이터 보안성, 네비게이션, 정보검색 업무, 정보 콘텐츠의 프리젠테이션과 구성 등이다.

Ivory(2001)는 웹 인터페이스 평가와 관련하여 웹 인터페이스 디자인(web interface design) 영역을 언급하고 있는데, 여기에는 정보 디자인(information design), 네비게이션 디자인(navigation design), 그리고 그래픽 디자인(graphic design) 등이 포함된다.

최정태 등(2001)은 웹 인터페이스의 품질 평가를 위한 영역으로서 다섯 개의 영역 및 열 다섯 개의 하위 요소들을 설정하였다. 다섯 개의 상위 영역들은 정보의 구성, 네비게이션의 용이성, 사이트의 구성요소, 레이블의 명확성, 그리고 페이지의 디자인 측면이다. 정보의 구성 영역은 정보의 단위, 정보구조의 폭과 깊이 등을 포함한다. 네비게이션의 용이성은 위치 파악의 용이성, 이동의 효율성 및 일관성, 유연성 등을 포함하고, 사이트의 구성요소 측면은 메뉴 바, 팝업 창, 풀다운 메뉴 등을 의미한다. 레이블의 명확성 영역은 네비게이션 및 표제 레이블의 정확성을 의미한다. 페이지의 디자인 측면은 페이지 크기, 길이의 적절성 일관성 등을 반영한다.

3. 국가주도 웹 아카이빙 사업 인터페이스

앞 단원에서 살펴보았듯이 그동안의 국내 웹 아카이빙 관련 연구들은 정책 관련 연구 및 메타데이터 관련 연구들이 주로 수행되어 왔다. 한편 기록관리 및 아카이빙 영역의 패러다임이 그동안의 정책 및 보존 중심의 관점으로부터 이용 및 접근성 확대의 관점으로 확산됨에 따라(이재윤 외 2007) 이용자 접근성의 제고를 위한 고려가 필요하다.

이를 위하여 앞 단원에서 살펴본 웹 인터페이스 평가 관련 선행 연구에서 제시한 요소들을 정리하여, 해외의 국가 주도 웹 아카이빙 사례에 적용하여 분석해보고자 한다.

선행 연구에서 공통적으로 제시한 웹 인터페이스

이스 평가 요소들을 정리하여 제시하면 <표 2>와 같다. 각 연구자들이 제시한 요소들을 유사한 내용끼리 재배치하여 정리하였다.

<표 2>에서 정리된 공통적인 요소들을 실제 웹 아카이빙 사이트의 인터페이스 분석을 위하

여 평가 가능한 항목으로 구체화하여 <표 3>에서 제시하였다. 구체화된 항목 내용들은 선행 연구에서 세부적으로 기술하고 있는 해당 항목 내용 및 제시한 평가표 항목, 또한 각 영역에 대한 설명 등을 기반으로 하였다.

<표 2> 웹 인터페이스 평가 요소 정리

	Vertelney et al.	Zhang et al.	Ivory	최정태 등	인터페이스 영역
1	시각화	직관적 외양	정보 디자인 그래픽 디자인	페이지의 디자인	시각 디자인 인터페이스
		프리젠테이션과 구성	텍스트·링크·그래픽 구성요소		
2	웹 접근성	접근 제한성			웹 접근성 인터페이스
3		프라이버시 및 데이터 보안성			보안 인터페이스
4	유용성	정보검색 업무		사이트 구성요소	검색 인터페이스
5	기능성	기본 기능	네비게이션 디자인	레이블의 명확성	지원기능 인터페이스
		피드백 / 응답		네비게이션의 용이성	
		네비게이션			

<표 3> 웹 아카이빙 사이트 인터페이스 평가 항목

	인터페이스 영역	상세 설명	평가 항목
1	시각 디자인 인터페이스	웹 페이지를 구성하는 시각적 요소가 과다하지 않고, 이용자 친화적임.	1.1 웹 사이트를 구성하는 웹 페이지들이 일관성 있는 시각적 디자인을 유지함. 1.2 한 주제와 관련된 아이콘과 그림, 색상등 이 과다하지 않음. 1.3 프레임의 길이와 크기가 크지 않아 한 화면 내에서 스크롤바 사용을 최소화함.
2	웹 접근성 인터페이스	가독성에 장애를 갖는 이용자층이나 환경에도 보편적인 접근성을 보장해야 함.	2.1 색깔로 정보가 구별되지 않도록 하고 흑백으로도 이해될 수 있도록 함. 2.2 시청각콘텐츠의 대체 텍스트를 제공함.
3	보안 인터페이스	프라이버시 관련 내용은 공개가 제한되고, 보안을 대비한 인터페이스 요소를 구성하고 있음.	3.1 프라이버시 관련 내용을 보호하는 인터페이스 요소가 제공됨. 3.2 공개가 제한된 정보에 대한 보안 조치 인터페이스 요소를 구성하고 있음.
4	검색 인터페이스	이용자가 찾고자 하는 정보를 빠르고 정확하고 효율적으로 검색할 수 있는 기능들을 구성함.	4.1 검색 콘텐츠 유형에 적절한 상세검색 기능(단순키워드 외 다양한 검색)이 제공됨. 4.2 키워드 검색 외에도 메뉴바, 팝업창, 풀다운 메뉴 등을 통하여 다각적인 검색 접근점을 제공함. 4.3 검색된 컬렉션 내에서의 상세 검색이 가능하며, 부가적인 검색 안내에 대한 인터페이스를 제공함.

	인터페이스 영역	상세 설명	평가 항목
5	지원기능 인터페이스	이용자의 해당 웹 사이트의 브라우저를 기능적으로 지원하며, 커뮤니케이션 할 수 있는 상호작용적 기능요소들 구성하고 있음.	5.1 이용자를 지원하는 피드백 및 응답 기능이 제공됨. 5.2 위치과약을 용이하게 지원하는 네비게이션 기능이 제공됨. 5.3 웹 사이트의 소속 페이지들의 표제 레이블이 정확하게 제공됨. 5.4 불필요한 링크 정보를 최소화하고, sub 메뉴간의 깊이(web depth)가 상호 균형을 이루며 적절히 제공됨.

위 주제 영역 중 웹 접근성 영역은 단독 주제로 많은 연구가 이루어지고 있는 영역이다. 주로 자동분석 도구를 이용하여 특정 주제영역의 웹 사이트들을 평가하는 방식인데, 문헌정보학과 웹사이트와 대학도서관 웹 사이트의 접근성을 분석한 사례(Schmetzke 2001), 보건정보 관련 웹 사이트의 접근성을 분석한 사례(McCord et al. 2002), 국제적 순위 100위권 내 대학 웹사이트를 분석한 사례(Kane et al. 2007) 등이 있다. 자동분석 도구로는 주로 A-Prompt, Bobby, 그리고 WebXACT 등이 많이 활용되는데, 이러한 자동분석 도구는 모두 웹 접근성 평가와 관련된 국제표준인 WCAG(Web Content Accessibility Guideline)을 준수하고 있다. 이에 따라 위 주제 영역 중 웹 접근성 영역에 해당하는 세부 항목은 WCAG에서 제시하고 있는 상위 카테고리 평가항목들을 반영하여 기술하였다.

3.1 미국 LCWA

3.1.1 개관

미국 국회도서관은 2000년부터 국가의 웹기반 디지털 문화유산을 후대에 전승하기 위한 노력을 시작하였다. 초기의 웹 아카이빙 프로젝트는 MINERVA로 불리었고, 주요 정치경제적 사건 컬렉션 등을 강화해가면서 현재는

The Library of Congress Web Archives(이하 LCWA)라고 불리고 있다. LCWA에서는 특히 최근의 이슈가 되는 사건들과 관련된 웹 콘텐츠 등을 구축해오고 있는데, 이는 미국의 선거, 9·11 테러, 이라크 전쟁 등이 포함된다.

LCWA는 복합적인 수집방식을 취하고 있다. 즉, 전문가를 중심으로 한 선택적 아카이빙 및 웹 크롤러를 이용한 포괄적 아카이빙을 함께 병행하고 있다. 또한 LCWA에서는 자체 개발 툴인 Digiboard를 이용하여 주제전문사서들 외에도 일반 국민들을 대상으로 아카이빙 대상이 될만한 웹사이트들의 추천도 받고 있다.

LCWA는 다수의 관련 기관들과 파트너십을 이루고 있지만, 특히 Internet Archive와의 지속적 협력으로 약 400억개 정도의 웹 페이지들을 아카이빙하고 있으며, 2010년 11월 현재 LCWA는 200 테라바이트 이상의 데이터를 수집하였다.

3.1.2 인터페이스

LCWA의 영역별 인터페이스를 살펴보기 위하여 해당 웹 사이트를 확인한 결과, 먼저 첫 화면에 미국 선거, 911 테러, 이라크 전쟁 등 미국 관련 주요 주제 컬렉션 리스트를 제공하고 있고, 이 외에 탐색(search)과 브라우징(browse) 메뉴를 함께 제공하고 있다.

탐색 화면에서의 탐색 필드는 오른쪽의 박스

에서 나타나는 것과 같이 7개의 필드(이름, 표제, 주제, 초록, 언어, 연도, 레코드 ID)를 제공하고 있다. 이 중 주제(subject)는 아카이빙된 각각의 웹사이트에 부여된 LC 주제명표목(Library of Congress Subject Headings, LCSH)을 부여하고 있고, 아카이빙된 웹 사이트에서 추출된 비통계 주제 키워드 스트링도 포함되어 있다. 연도(year captured)는 해당 웹사이트가 아카이빙된 연도를 나타낸다.

특정 주제 검색에 대한 결과 화면에서는, 해당 웹 사이트에 대하여 그동안 수행된 아카이빙 히스토리 정보를 제공하고 있다. 다음 그림은 링컨(Lincoln)에 대한 LCWA 검색 결과 화면이다(그림 1 참조).

LCWA에서 사용하는 웹 크롤러(web crawler)인 Heritrix 크롤러는 스트리밍 미디어(streaming

media), 심층웹(deep web) 또는 이용자 인증을 요구하는 데이터베이스 콘텐츠는 아카이빙할 수 없다. 또한 웹 사이트 중 현재 사라진 옛 기술을 적용했거나, 아니면 보편적이지 않은 특수한 기술을 활용한 웹사이트도 아카이빙할 수 없다. Heritrix 웹 크롤러는 Internet Archive에서 활용하고 있는 확장가능형 오픈소스이기도 하다.

아카이빙된 웹 콘텐츠에의 접근과 관련해서는 WERA와 Wayback access tools, 그리고 the NutchWAX 검색 엔진을 기술적으로 활용하고 있다. 이러한 기술적 환경(technical environment)은 리눅스와 같은 오픈 플랫폼에 기반하고 있다.

LCWA의 웹 아카이빙 사이트의 항목별 인터페이스를 점검해 본 결과는 <표 4>와 같다.



<그림 1> LCWA 결과 제공 화면 [cited 2010. 11. 5].

<http://lcweb2.loc.gov/diglib/lcwa/html/lcwa-home.html>

〈표 4〉 미국 LCWA 웹 아카이빙 인터페이스

	인터페이스 영역	평가 항목	적용결과
1	시각 디자인 인터페이스	1.1	웹 사이트의 하위 웹 페이지들이 해당 사업에서 사용하는 색상과 레이아웃을 일관성 있게 사용하고 있음.
		1.2	단일 주제에 대한 아이콘과 색상의 사용이 적음.
		1.3	화면 내 프레임 사용이 적고 길이가 길지 않아 별도의 스크롤바의 사용이 적음.
2	웹 접근성 인터페이스	2.1	텍스트 기반으로 정보가 제공되고 있으므로 보편적인 서비스 제공이 가능함.
		2.2	웹 사이트 자체적으로 제공되고 있는 멀티미디어 형태의 콘텐츠는 적음.
3	보안 인터페이스	3.1	프라이버시와 관련된 정보의 제공이 제한된다는 설명을 제시함.
		3.2	공개에 제한을 갖는 정보 보안 및 저작권 관련 설명을 제시함.
4	검색 인터페이스	4.1	기본적인 검색 접근점 외에 브라우징 등의 정보 검색 접근점을 제공함.
		4.2	메뉴바 팝업창 풀다운메뉴 등의 제공은 활발하지 않음.
		4.3	검색된 콘텐츠와 유사한 유형별 주제별 검색확장을 안내하고 있음.
5	지원기능 인터페이스	5.1	이용자에 대한 전체적인 FAQ와 기본적인 피드백 정보가 제공됨.
		5.2	네비게이션 기능은 위치확인에 대한 간략한 라인 커멘트(line comment)로 제공됨.
		5.3	모든 소속 웹 페이지의 제목은 LCWA 로 제시되고 있음.
		5.4	불필요한 링크정보가 없고, web depth는 최종 콘텐츠에 이르기까지 대략 3~5 path를 이루고 있음.

3.2 영국 UK Web Archive

3.2.1 개관

UK Web Archive 사업은 영국 국립도서관을 주축으로 2004년부터 진행해 오고 있으며, 대상이 되는 웹 사이트들은 영국과 관련된 다양한 영역의 활동, 사회상, 역사, 가치 등을 포함하는 사이트들로서 정치적, 경제적, 사회문화적, 종교적, 과학적 가치를 반영하는 사이트들이다.

UK Web Archive는 주제전문사서들과 아키비스트들이 중심이 되어 사업을 수행하고 있으나, 보다 포괄적이고도 적극적인 수집을 위하여 일반 국민들로부터 수집할 필요가 있는 웹사이트들을 항시적으로 추천받고 있다. 추천할 때에는 추천 사유 및 추천자의 정보를 제공할 것을 요구하고 있으며, 이와 함께 추천하는

사이트의 특징, 즉, 소멸예정 이라든지 데이터 베이스와 연관된 심층 웹 인 경우 등 아카이빙 작업에 참고해야 할 내용들을 명기해 줄 것을 요청하고 있다.

웹 사이트의 아카이빙은 6개월간의 간격을 두고 정기적으로 수행해 오고 있으나 상황에 따라서 그 빈도를 조정하기도 한다. UK Web Archive에서는 매달 아카이빙 된 규모를 제시하고 있는데 2010년 11월 현재의 규모는 〈표 5〉와 같다.

UK Web Archive 사업 추진과 관련하여 파트너십을 이루고 있는 기관들에는 영국 국가기록원, 스코틀랜드 국립도서관, 웨일즈 국립도서관, JISC, 웰컴 도서관 등이 있으며 그 외에도 컬렉션 구성과 관련하여 협력해온 다수의 전문 기관들이 포함된다.

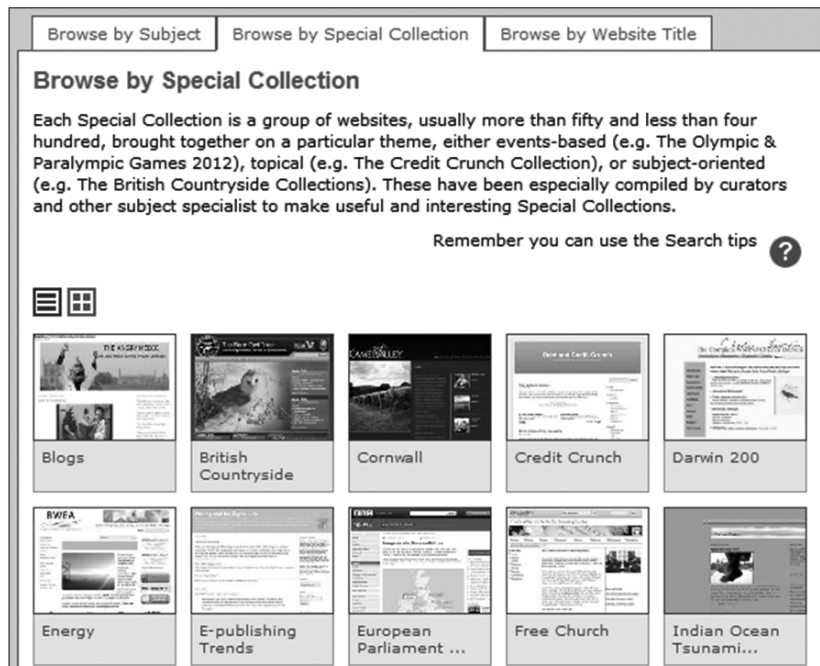
〈표 5〉 UK Web Archive 규모 현황

UK Web Archive	2010년 11월	2010년 10월	이번달 증가분
아카이빙된 웹사이트 수	8,153	7,325	828
아카이빙된 인스턴스(instance) ¹⁾ 수	31,508	27,740	3,768
데이터 크기	7.57TB	6.64TB	931.72GB

3.2.2 인터페이스

UK Web Archive에서도 탐색(search) 메뉴 및 브라우징(browse) 메뉴를 중심으로 한 인터페이스를 제공하고 있다. 탐색 메뉴를 통해서는 웹사이트의 타이틀 또는 URL을 이용하여 검색할 수 있다. 브라우징(browse) 메뉴

를 통해서 검색할 수도 있는데, 브라우징 메뉴는 세 가지 하위 메뉴를 포함하고 있다. 세 가지 하위 메뉴는 주제 메뉴, 특별 컬렉션(Special Collection) 메뉴, 그리고 웹사이트 제목 메뉴인데, 이는 〈그림 2〉와 같다.



〈그림 2〉 UK Web Archive 브라우징(browse) 화면 [cited 2010. 11. 5].

(<http://www.webarchive.org.uk/ukwa/>)

1) 인스턴스(instance)란 아카이브에 보내어진 단일 스냅샷(snapshot) 또는 웹사이트의 복사본을 의미한다. 동일한 웹사이트도 시간이 경과함에 따라서 여러번 아카이빙되므로, 캡처된 각각의 단일 웹사이트는 구별하여 인스턴스라고 불린다.

〈그림 2〉에서 나타나는 것처럼 브라우징 화면은 세 개의 하위 메뉴로 제시하고 있는데, 먼저 주제 메뉴에서는 인문학, 경영경제, 교육연구, 정부·법·정책, 의학·건강, 과학기술, 그리고 사회문화 라는 일곱 개의 하위 주제영역을 제시하고 있다.

특별 컬렉션 하위 메뉴에서는 UK Web Archive에서 특히 부각하여 제공할 수 있는 컬렉션들을 화면 이미지와 함께 제시한다. 또한 웹 사이트 제목 메뉴에서는 UK Web Archive에 아카이빙된 사이트들의 제목들을 알파벳순으로 제시한다.

UK Web Archive에서 제공하는 특정 주제에 검색에 의한 검색 결과 화면은 〈그림 3〉과 같다.

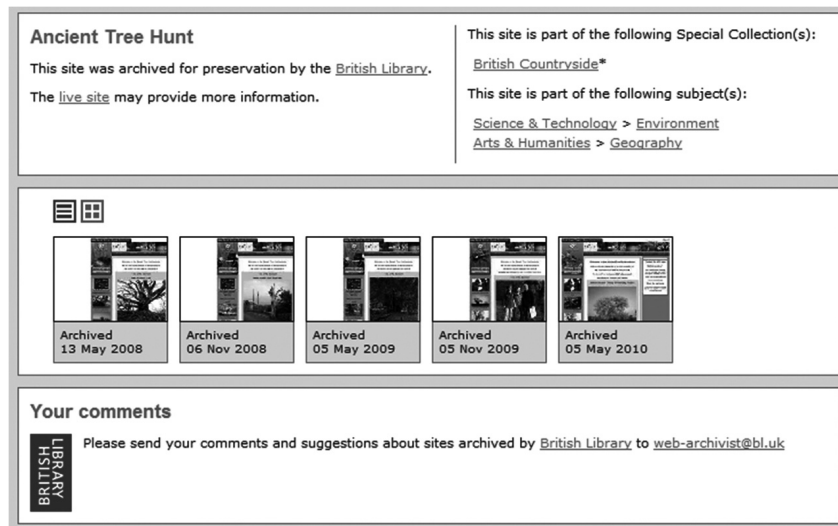
〈그림 3〉은 ‘Ancient Tree Hunt’라는 키워드에 대하여 웹 아카이빙된 콘텐츠를 제공하는 화면이다. 화면에서 확인할 수 있듯이 해당 사

이트는 2008년 5월부터 5개월 단위로 아카이빙되어 왔으며, 해당 시점에서의 아카이빙된 화면이 썸네일로 제시되고 있다.

또한, 해당 웹 사이트의 현재 사이트(live site)와 전체 아카이빙 컬렉션 내에서의 위치정보, 그리고 이용자로부터의 커멘트나 제안 등의 피드백을 위한 공간 등이 구성되어 있는 것을 확인할 수 있다.

UK Web Archive도 Heritrix 크롤러를 사용하여 웹사이트들을 수집하는데, 현재 캡처하지 못하는 유형의 웹 페이지들은 다음과 같다.

- 동적으로 생성된 콘텐츠(Dynamically generated content)
- 검색 엔진을 통해서만 웹사이트에 접근되는 콘텐츠
- 자바스크립트 기반 메뉴 중 일부 유형
- 유튜브 동영상, 플래시 동영상 및 유사한 스트리밍 오디오 또는 비디오



〈그림 3〉 UK Web Archive 결과 제공 화면 [cited 2010. 11. 5].
 (http://www.webarchive.org.uk/ukwa/)

- 이용자 인증을 거치는 웹 사이트

UK Web Archive의 웹 아카이빙 사이트의 항목별 인터페이스를 점검해 본 결과는 <표 6>과 같다.

3.3 캐나다 GCWA

3.3.1 개관

캐나다에서는 국가차원에서의 웹 아카이빙 사업이 2005년 12월에 시작되었다. 사업명은 Government of Canada Web Archive(GCWA)이고, 캐나다 국립 도서관 기록관(Library and Archives Canada, 이하 LAC)에서 주체가 되어 수행하였다. 아카이빙 대상이 되는 웹 사이트

는 캐나다 정부관련 웹사이트이다. 아카이빙은 1년에 두 번씩 수행되어 오고 있으므로, 6개월 미만의 짧은 생명을 갖고 있는 웹사이트들은 포함되지 않을 수도 있다는 단점이 있다.

2007년 가을에는 그동안 아카이빙 된 약 1억 개의 디지털 객체(4 테라바이트 이상)를 LAC 웹사이트를 통하여 공개하였다. 2010년 현재 GCWA에는 17억개의 디지털 객체 및 7 테라바이트 이상의 데이터가 포함되어 있다.

GCWA의 콘텐츠들은 키워드 탐색, 부서명 탐색, 또는 URL 탐색 접근을 통하여 제공되고 있으며, 특정 포맷이나 유형별 탐색(예: PDF 유형 탐색 등)도 가능하다.

<표 6> 영국 UK Web Archive 웹 아카이빙 인터페이스

	인터페이스 영역	평가 항목	적용결과
1	시각 디자인 인터페이스	1.1	웹 사이트의 하위 웹 페이지들이 해당 사업에서 사용하는 색상과 레이아웃을 일관성있게 사용하고 있음.
		1.2	비주얼 아이콘을 비롯한 단일 주제에 대한 아이콘과 색상이 많이 사용되고 있음.
		1.3	화면 내 프레임 사용이 적고 길이가 길지 않아 별도의 스크롤바의 사용이 적음.
2	웹 접근성 인터페이스	2.1	색상의 대조 속에서 표현된 텍스트의 정보 전달력이 이용능력에 제한이 있는 이용자에게 제한될 수 있음.
		2.2	시청각 콘텐츠가 포함되어 있지 않고, 텍스트 위주로 표현됨.
3	보안 인터페이스	3.1	프라이버시와 관련된 정보의 제공이 제한된다는 설명을 제시함.
		3.2	공개에 제한을 갖는 정보 보안 및 저작권 관련 설명을 제시함.
4	검색 인터페이스	4.1	기본적인 검색 접근점 외에 특별 컬렉션의 이미지 썸네일을 통하여 이용자 친화적 검색 접근점을 제공하고 있음.
		4.2	메뉴바 팝업창 풀다운메뉴 등의 제공은 활발하지 않음.
		4.3	검색된 콘텐츠와 유사한 유형별 주제별 검색확장을 안내하고 있음.
5	지원기능 인터페이스	5.1	이용자에 대한 전체적인 FAQ와 기본적인 피드백 정보가 제공됨.
		5.2	네비게이션 기능은 위치확인에 대한 간략한 라인 커멘트(line comment)로 제공됨.
		5.3	모든 소속 웹 페이지의 제목은 UK Web Archive로만 제시되고 있음.
		5.4	불필요한 링크정보가 없고, web depth는 최종 콘텐츠에 이르기까지 대략 3~5 path를 이루고 있음.

3.3.2 인터페이스

GCWA의 웹 콘텐츠 접근은 기본 키워드 검색 및 상세 검색 화면을 제공하고 있다. 상세 검색에서는 키워드, 아카이빙된 시기, 대상 웹 문서의 포맷 유형(이미지, 비디오, PDF, DOC, HTML), 그리고 포함되는 URL을 중심으로 검색할 수 있다.

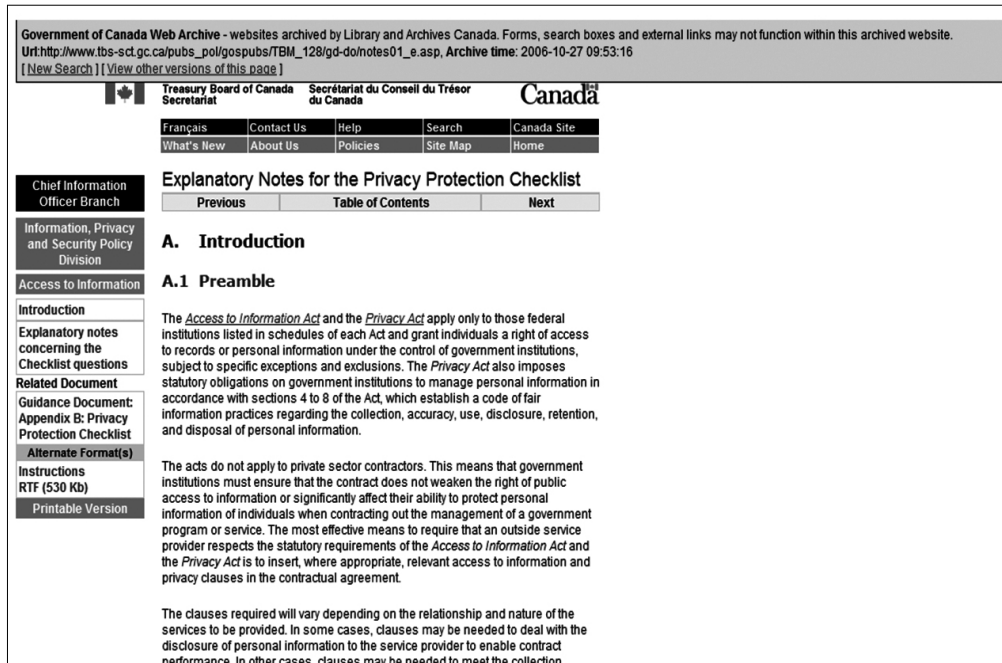
해외 다른 사례에 비하여 비교적 단순한 인터페이스를 구성하고 있었으며, 검색 결과에 대한 간략 설명 내용도 텍스트 기반으로만 결과를 제시하고 있다. 특정 주제어에 대한 검색을 수행한 결과 화면은 <그림 4>와 같다.

GCWA는 아카이빙의 횟수도 아직은 단회성에 그치는 경우가 많았다. 검색 화면에는 그림에서 나타나는 바와 같이 윗부분에 배너로

LAC에서 아카이빙한 정보를 수록하고 있다.

LAC은 오픈 소스 툴들을 이용하여 캐나다 웹 아카이브를 구축하였는데, LCWA와 마찬가지로 다음과 같은 동일한 세 가지 유형의 소프트웨어가 활용되었다.

- 웹 크롤링 소프트웨어: Heritrix v1.12.1
- 데이터베이스 / 리포지터리 소프트웨어: NutchWAX v0.10.0 NutchWAX란 is “Nutch(<http://www.nutch.org>)”와 “Web Archive eXtensions”의 합성어로서, 웹 아카이브 컬렉션의 탐색을 돕는다.
- 접근: Wayback v0.8.0 GCWA의 웹 아카이빙 사이트의 항목별 인터페이스를 점 검해 본 결과는 <표 7>과 같다.



<그림 4> GCWA 검색 결과 화면 [cited 2010. 11. 5].

<<http://www.collectionscanada.gc.ca/webarchives/search/index-e.html>>

〈표 7〉 캐나다 GCWA 웹 아카이빙 인터페이스

인터페이스 영역	평가 항목	적용결과	
1	시각 디자인 인터페이스	1.1	웹 사이트의 하위 웹 페이지들의 일부만 대표 로고 및 색상, 레이아웃 등을 사용하고 있음.
		1.2	아이콘, 색상 등에 대한 사용이 적음.
		1.3	웹 페이지의 길이는 적절하여 별도의 스크롤바의 사용이 적음.
2	웹 접근성 인터페이스	2.1	기본적인 텍스트 중심으로 표현하고 있어 이용 환경에 제한이 있는 이용자에게도 보편적인 서비스 제공이 가능함.
		2.2	시청각 콘텐츠가 포함되어 있지 않고, 텍스트 위주로 표현됨.
3	보안 인터페이스	3.1	프라이버시와 관련된 설명이 제시되지 않음.
		3.2	공개에 제한을 갖는 보안 관련 설명은 없으나 해당 캐나다 법령에 대한 설명을 제공함.
4	검색 인터페이스	4.1	기본적인 검색 접근점 외에 상세 검색화면, URL 기반 검색 등의 접근점을 제공함.
		4.2	메뉴바 팝업창 풀다운메뉴 등의 제공은 활발하지 않음.
		4.3	검색된 콘텐츠와 관련된 검색확장 기능이 부족함.
5	지원기능 인터페이스	5.1	이용자에 대한 전체적인 FAQ와 기본적인 피드백 정보가 제공됨.
		5.2	네비게이션 기능은 위치확인에 대한 간략한 라인 커멘트(line comment)로 제공됨.
		5.3	각 웹 페이지의 해당 주제에 대한 레이블이 적절히 지원되고 있음.
		5.4	불필요한 링크정보가 없고, web depth는 최종 콘텐츠에 이르기까지 대략 3~5 path를 이루고 있음.

3.4 호주 PANDORA

3.4.1 개관

PANDORA(Preserving and Accessing Networked Documentary Resources of Australia)는 1996년 호주 국가도서관에 의하여 시작된 국가 웹 아카이빙 사업으로서 현재 10여 개의 호주 도서관 및 문화 유산 기관들과 협력 컨소시엄을 이루고 있다. 각 협력 기관들은 전문 영역을 중심으로 역할을 분담하여 웹 수집 작업을 진행하고 있다. 즉, The National Film and Sound Archive 에서는 호주 필름 및 음악과 관련된 웹 사이트들을 수집하고 있으며, the Australian War Memorial archives는 군 역사와 관련된 웹 자원들을, 그리고 AIATSIS (Australian Institute of Aboriginal and Torres

strait Islander Studies) archives는 원주민 관련 웹 정보자원을 수집한다.

PANDORA는 국제적으로 대표적인 국가 웹 아카이브인 만큼 아카이브를 구성하는 각 모듈 및 시스템들이 체계적으로 운영되고 있다. 즉, PANDORA 디지털 아카이빙 시스템인 PANDAS(PANDORA Digital Archiving System, 이하 PANDAS)에서는 아카이브의 각 아이템들의 타이틀 레벨에서 컴포넌트 파일들에 이르기까지 식별자(persistent identifiers)를 부여함으로써 수집된 웹 자원의 진본성(authenticity)과 무결성(integrity)이 유지되도록 지원하고 있다. PANDAS는 선택적 웹 아카이빙을 수행하고, 타이틀 정보를 기록하는 워크플로우를 지원하는 컬렉션관리 시스템(collection management system)이다.

또한, 장기보존을 위한 시스템으로는 DOSS (Digital Object Storage System)를 활용하고 있다. PANDORS에서 수집한 디지털 객체는 두 가지 방식으로 보존되고 있는데, 보존 마스터 (preservation master)와 접근 마스터(access master) 방식이다. 접근 마스터는 PANDORA 디스플레이 머신(display machine)의 유닉스 파일 시스템에 보존함으로써 신속한 접근을 돕는다. 보존 마스터는 DOSS 시스템에 디지털 객체를 전송함으로써 안전한 장기 보존을 지원한다.

PANDORA 웹 아카이브의 규모 및 현황은 <표 8>과 같다.

3.4.2 인터페이스

PANDORA에서의 탐색은 별도로 구현한 Trove 온라인 탐색 엔진 웹 사이트를 통하여 진행된다. Trove는 특히 심층 웹의 아카이빙에 초점을 맞추고 있으며, 아카이빙된 콘텐츠 자체 외에도 해당 주제와 연관된 다양한 콘텐츠들을 제공하며, 더 나아가 트위터 운영, 이용자들이 태깅, 커멘트 등을 부여할 수 있는 기능 등을 마련함으로써 소셜 네트워크 커뮤니케이

션을 돕고 있다.

탐색방식은 첫 화면에서 이미지를 중심으로 한 콘텐츠의 유형별 브라우징을 지원하고, 이 외에도 기본 탐색, 키워드 불리언 탐색, 구절 (phrase)탐색 등을 지원한다. Trove에서의 연관 콘텐츠들의 탐색 확장은 Trove 사이트에서 계속 진행이 되지만, 최종 검색된 아카이브 콘텐츠의 경우에는 클릭과 함께 다시 PANDORA 화면으로 되돌아와서 검색 결과를 제시하게 된다. 검색 결과에는 아카이빙한 날짜들과 기관, 그리고 해당 사이트에 대한 설명이 제시된다(그림 5, 그림 6 참조).

PANDORA의 웹 아카이빙 사이트의 항목별 인터페이스를 점검해 본 결과는 <표 9>와 같다.

3.5 싱가포르 WAS

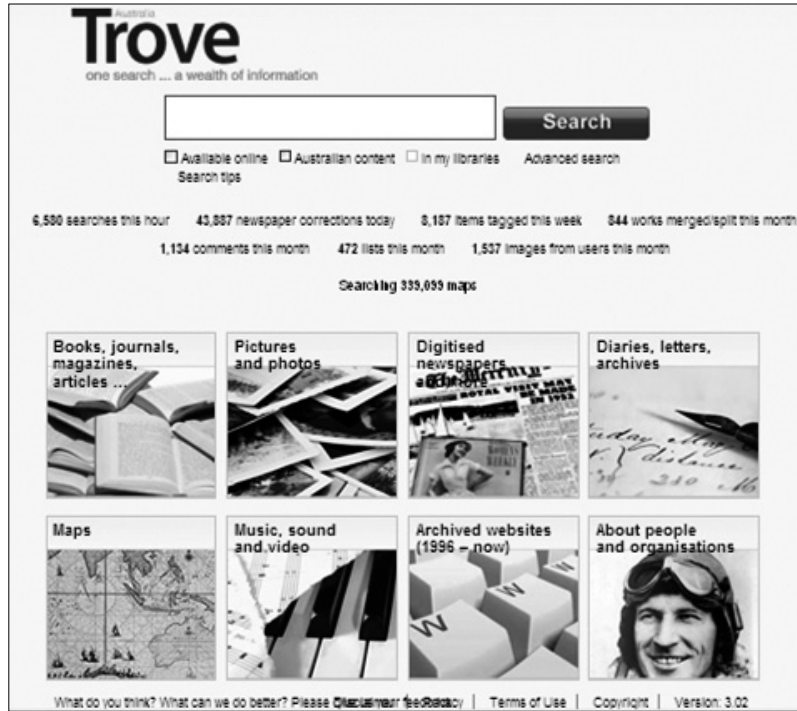
3.5.1 개관

싱가포르의 국가 웹 아카이브는 국가도서관인 The National Library Board(이하 NLB)를 중심으로 수행되고 있으며, 프로젝트 명은 Web Archive Singapore(이하 WAS)이다. 싱가포르 관련 온라인 콘텐츠들은 싱가포르의

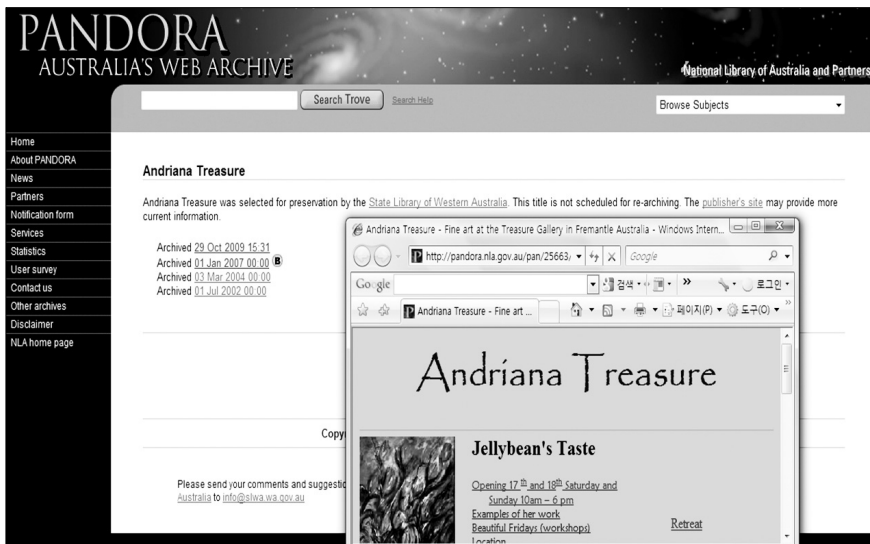
<표 8> PANDORA Web Archive 규모 현황

PANDORA	2010년 10월	2010년 9월	이변달 증가분
아카이빙된 타이틀 ²⁾ 수	26,311	26,183	128
아카이빙된 인스턴스(instance) 수	59,018	58,399	619
아카이빙된 파일 수	99,838,229	97,269,674	2,568,555
데이터 크기	4.47 TB	4.37 TB	102 GB
정부간행물 보유량	대략 전체 컬렉션의 50%		
2009-2010 1년간의 활용도	4,985,676 페이지 열람		

2) 타이틀(title)은 아카이빙을 위해 선정된 엔티티로서 목록 레코드가 만들어져 있다. 웹 사이트 전체 또는 부분일 수 있다.



〈그림 5〉 Trove 탐색 화면 [cited 2010. 12. 13].
〈<http://trove.nla.gov.au/>〉



〈그림 6〉 PANDORA 검색 결과 화면 [cited 2010. 11. 5].
〈<http://pandora.nla.gov.au/about.html>〉

〈표 9〉 호주 PANDORA 웹 아카이빙 인터페이스

	인터페이스 영역	평가 항목	적용결과
1	시각 디자인 인터페이스	1.1	PANDORA 메인 페이지와 Trove 사이트의 연결된 이미지 및 레이아웃은 약함. 단 Trove 내에서의 자체적인 시각적 연결 이미지와 레이아웃은 적절히 제공되고 있음.
		1.2	단일 주제에 대한 아이콘과 색상이 적절하게 사용되고 있음.
		1.3	화면 내 프레임 사용이 적고 길이가 길지 않아 별도의 스크롤바의 사용이 적음.
2	웹 접근성 인터페이스	2.1	Trove 내에서는 이미지 썸네일을 많이 사용하고 있어서 대체 텍스트의 제공이 필요함.
		2.2	시청각 콘텐츠의 포함은 적으나 이미지 썸네일은 대체 텍스트가 필요함.
3	보안 인터페이스	3.1	프라이버시와 관련된 정보의 제공이 제한된다는 설명을 제시함.
		3.2	공개에 제한을 갖는 정보 보안 및 저작권 관련 설명을 제시함.
4	검색 인터페이스	4.1	Trove 첫 화면에서는 유형별 카테고리 및 이미지를 썸네일을 통하여 제시함으로써 이용자 친화적 검색 접근점을 제공하고 있음.
		4.2	메뉴바 및 풀다운메뉴 등의 제공이 활발함.
		4.3	검색된 콘텐츠와 유사한 유형별 주제별 검색확장을 안내하고 있음.
5	지원기능 인터페이스	5.1	이용자에 대한 전체적인 FAQ와 기본적인 피드백 정보외에도 태그, 트위터, 커멘트 등 적극적인 커뮤니케이션 통로가 제공됨.
		5.2	현재 위치를 쉽고 빠르게 이해할 수 있는 네비게이션 기능은 전반적으로 미약함.
		5.3	모든 소속 웹 페이지의 제목은 Trove로만 제시되고 있음.
		5.4	관련된 링크정보가 많이 제공되고 있고, web depth는 최종 콘텐츠에 이르기까지 대략 3~5 path를 이루고 있음.

사회문화상, 생활상 등을 다각적으로 보여준다. 주제영역은 정부 행정부처, 교육영역으로부터 예술, 레크레이션 관련 콘텐츠에 이르기까지 국가의 유산을 이루는 웹 콘텐츠들은 모두 포함하고 있다.

NLB에서 WAS를 구축하면서 기준으로 삼고 있는 아카이빙 대상 웹 사이트 선정 기준은 다음과 같다.

- 대상 웹사이트는 국가적 중요성과 의미가 있는가? 국가적으로 중요한 이벤트(선거 등) 등을 포함하는 사이트인가?
- 대상 웹사이트는 역사적 가치가 있는가? 또는 사회문화적, 정치경제적인 사회상을 반영하고 있는가?
- 대상 웹사이트는 연구자들이 관심을 가질

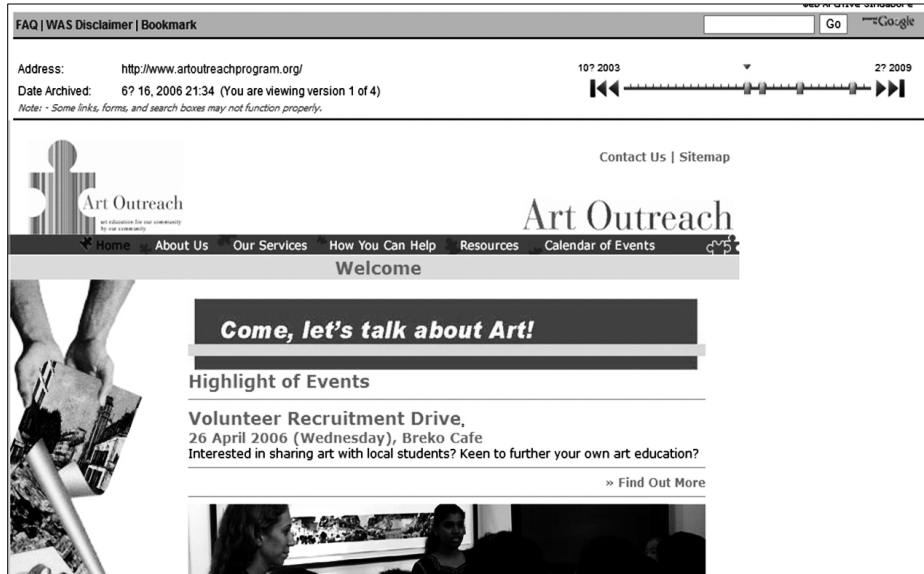
만한 내용을 수록하고 있는가?

- 대상 웹사이트는 싱가포르에 대한 정보를 제공하고 홍보를 할 수 있는 내용을 수록하고 있는가?
- 대상 웹사이트는 주제 분야 전문가나 권위 있는 기관에서 생산했는가?

3.5.2 인터페이스

WAS에서는 키워드 탐색이나 주제 브라우징을 통하여 접근점을 제공하고 있으며, 결과는 웹페이지 타이틀을 중심으로 한 알파벳순 리스트로 제시된다.

WAS 웹 아카이브의 검색 결과에 대한 화면은 〈그림 7〉과 같다.

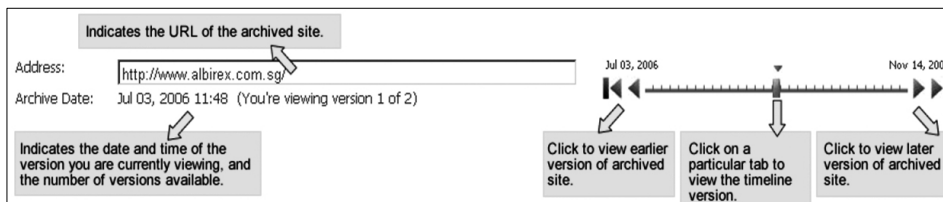


<그림 7> WAS 웹 아카이브 검색 결과 화면 [cited 2010. 11. 5].
 <<http://was.nl.sg/>>

<그림 7>에서 보면 상단 프레임에 아카이빙과 관련된 Timebar가 제시되고 있는데, 이는 이용자들이 서로 다른 시간의 차원을 갖는 동일한 웹 사이트들을 효율적으로 네비게이트할 수 있도록 지원한다. Timebar에서는 각각의 시간에 따른 버전이 시간의 흐름에 따른 순서대로 눈금으로 표시되고 있다. 이용자들은 해당 눈금을 클릭하거나 양쪽 끝단에 있는 화살표를 클릭함으로써 시간대를 선택할 수

있다. Timebar에 대한 그림 및 설명은 <그림 8>과 같다.

그림에서 나타난 것과 같이 화면의 상단에 기본적인 아카이빙 관련 정보 제시와 함께 우측에 Timebar를 제시함으로써, 시간의 흐름에 따라 여러번 아카이빙 된 현황을 제시하고, 이용자가 원하는 시점으로 바로 이동할 수 있도록 이용자 중심의 검색 인터페이스를 제공하고 있다.



<그림 8> WAS에서 사용하고 있는 Timebar [cited 2010. 11. 5].
 <<http://was.nl.sg/>>

또한, WAS에서는 국제 정보교환 및 효율을 위하여 IIPC에서 권장하는 틀인 Heritrix 크롤러를 이용하여 아카이빙 할 웹 사이트들을 수집하고 있으며, 웹 아카이브의 색인과 탐색을 위하여서는 NutchWAX를 사용하고 있다. 또한, 웹 아카이브의 탐색과 네비게이션을 위하여서는 WERA 어플리케이션을 활용하고 있다.

WAS의 웹 아카이빙 사이트의 항목별 인터페이스를 점검해 본 결과는 <표 10>과 같다.

4. 웹 아카이빙 인터페이스 유형 및 특성분석

이상에서 살펴본 바와 같이 해외 각국들은

수집 정책 및 활용하고 있는 틀, 이용자 인터페이스 등에 있어서 대체적으로 유사한 양상을 보이고 있었다. 먼저 국가별 웹 아카이빙 사례를 중심으로 한 기본적인 사항을 비교하여 정리하면 <표 11>과 같다.

웹 인터페이스 유형별 분석을 국가별로 상대적으로 평가하여 정리하면 <표 12>와 같다.

표에서 살펴볼 수 있듯이, 각 국가별로 인터페이스 유형에 있어서 우수한 영역 등은 조금씩 차이가 있다. 호주와 영국의 경우에는 이미지 컬렉션을 활용한 검색 인터페이스가 우수했고, 호주는 이용자와의 커뮤니케이션 채널을 적극적으로 개발함으로써 지원 기능 인터페이스도 우수했다. 미국의 경우에는 대체적으로 각 영역별 인터페이스가 우수한데 비하여, 캐

<표 10> 싱가포르 WAS 웹 아카이빙 인터페이스

	인터페이스 영역	평가 항목	적용결과
1	시각 디자인 인터페이스	1.1	웹 사이트의 하위 웹 페이지들이 해당 사업에서 사용하는 색상과 레이아웃을 일관성있게 사용하고 있음.
		1.2	단일 주제에 대한 아이콘과 색상 등의 표현이 적음.
		1.3	화면 내 프레임 사용이 적고 길이가 길지 않아 별도의 스크롤바의 사용이 적음.
2	웹 접근성 인터페이스	2.1	텍스트 위주로 표현되어서 이용능력에 제한이 있는 이용자 층에게도 보편적인 서비스 제공이 가능함.
		2.2	시청각 콘텐츠가 포함되어 있지 않고, 텍스트 위주로 표현됨.
3	보안 인터페이스	3.1	프라이버시와 관련된 내용을 웹 사이트에서 제시하고 있음.
		3.2	정보 보안 관련 규정 및 법령 등을 설명하고 있음.
4	검색 인터페이스	4.1	기본적인 검색 접근점 외에 브라우징 검색법을 제공하지만 이미지 기반의 검색 접근점이 없고 다소 단순함.
		4.2	메뉴바 팝업창 폴다운메뉴 등의 제공은 활발하지 않음.
		4.3	검색된 콘텐츠와 유사한 유형별 주제별 검색확장이 없음.
5	지원기능 인터페이스	5.1	이용자에 대한 전체적인 FAQ가 제공됨.
		5.2	네비게이션 기능은 위치확인에 대한 간략한 라인 커멘트(line comment)로 제공됨.
		5.3	각 웹 페이지의 제목은 적절하게 레이블되고 있음.
		5.4	불필요한 링크정보가 없고, web depth는 최종 콘텐츠에 이르기까지 대략 3~5 path를 이루고 있음.

〈표 11〉 국가별 웹 아카이빙 사례 기본 사항 정리

	미국 LCWQ	영국 UK Web Archive	캐나다 GCWA	호주 PANDORA	싱가포르 WAS
수집방식	주제전문가 웹크롤러	주제전문가 웹크롤러	웹크롤러	주제전문가 웹크롤러	웹크롤러
활용 툴	Heritrix	Heritrix	Heritrix	PANDAS	Heritrix
	Wayback	Wayback	Wayback	DOSS	NutchWAX
	NutchWAX	NutchWAX	NutchWAX		WERA
	WERA				
규모	200 Tb	7.57 Tb	7 Tb	4.47 Tb	

〈표 12〉 국가별 웹 아카이빙 인터페이스 비교 분석(△부족 ○우수 ◎매우 우수)

인터페이스 영역	평가 항목	관련 내용	미국 LCWA	영국 UKWA	캐나다 GCWA	호주 PANDORA	싱가포르 WAS
1 시각 디자인 인터페이스	1.1	웹 사이트 하위 웹 페이지들의 일관성	○	○	△	△	○
	1.2	적절한 아이콘, 그림, 색상 등	○	△	○	○	○
	1.3	프레임의 길이	○	○	○	○	○
2 웹 접근성 인터페이스	2.1	색상으로 정보가 제한되지 않음	○	△	○	△	○
	2.2	시청각콘텐츠의 대체 텍스트 제공	○	○	○	△	○
3 보안 인터페이스	3.1	프라이버시 관련 내용의 보호	○	○	△	○	○
	3.2	보안정보 공개제한	○	○	△	○	○
4 검색 인터페이스	4.1	콘텐츠유형에 적절한 상세검색 기능	○	◎	○	◎	△
	4.2	메뉴바, 팝업창, 플다운메뉴 제공	△	△	△	○	△
	4.3	검색된 컬렉션 내에서의 확장검색	○	○	△	◎	△
5 지원기능 인터페이스	5.1	이용자 피드백 및 응답 기능	○	○	○	◎	○
	5.2	위치과악 지원하는 네비게이션 기능	○	○	○	△	○
	5.3	소속 페이지들의 표제 레이블	△	△	○	△	○
	5.4	적절한 링크 정보 및 web depth	○	○	○	○	○

나다는 보안 인터페이스가 상대적으로 약한 것으로 나타났고, 싱가포르의 검색 인터페이스 영역이 부족한 것으로 나타났다. 호주의 경우

에는 보편적인 서비스를 제공하는 웹 접근성 인터페이스 영역이 부족하였다.

5. 적용방안 및 제언

우리나라의 국가 웹 아카이빙 사업은 국립중앙도서관이 주축이 되어서 2001년부터 시범시스템을 개발한 이후 2005년에 웹 기반 OASIS (Online Archiving & Searching Internet Source) 시스템 확장 개선작업이 완료되었다. 국내에서 국가적으로 수행하고 있는 웹 아카이빙 사업과 관련해서는 OASIS 외에도 국가기록원에서 수행하는 대통령기록 웹 아카이빙 사업이 있다. 그러나 본 논문에서 분석한 해외 사례들이 대통령기록에만 국한한 기록이 아닌 국가적 차원에서의 웹 아카이빙 사례이므로, OASIS 만 그 대상으로 선정하여, 앞 장에서 살펴본 동일한 인터페이스 영역별 요소들을 적용하여 분석하였다(표 13 참조).

표에서 살펴볼 수 있듯이, 국내의 OASIS의 경우에는 시각적인 인터페이스, 웹 접근성, 보안 인터페이스에 있어서는 기본적인 요구사항이 충족되어 우수한 것으로 나타났으나, 검색 인터페이스와 지원기능 인터페이스에는 미흡한 것으로 나타났다.

현재 제공되고 있는 평면적인 검색 인터페이스의 보강이 필요하고, 이용자와의 커뮤니케이션 강화를 지원하는 서비스 제공이 필요하다.

또한, 해외사례의 경우, 각 사례마다 특징이 되는 포커스를 갖고 집중하여 사업을 진행하고 있다. 즉, 미국의 경우에는 최신 사회 이슈 및 사건들을 중심으로 한 아카이빙을 강화한다든지, 호주의 경우 모듈별로 자체 시스템을 구축하고 전문적 분업을 진행하고, 소셜 네트워크를 강화한 이용자 지원 특화 서비스를 한다는

〈표 13〉 한국 OASIS 웹 아카이빙 인터페이스

	인터페이스 영역	평가 항목	적용결과	
1	시각 디자인 인터페이스	1.1	웹 사이트의 하위 웹 페이지들이 해당 사업에서 사용하는 색상과 레이아웃을 일관성있게 사용하고 있음.	○
		1.2	단일 주제에 대한 아이콘과 색상이 적게 사용됨.	○
		1.3	화면 내 프레임 사용이 적고 길이가 길지 않아 별도의 스크롤바의 사용이 적음.	○
2	웹 접근성 인터페이스	2.1	텍스트 기반의 정보제공이므로 보편적인 웹 접근성 제공이 가능함.	○
		2.2	시청각 콘텐츠가 포함되어 있지 않고, 텍스트 위주로 표현됨.	○
3	보안 인터페이스	3.1	프라이버시와 관련된 정보의 제공이 제한된다는 설명을 제시함.	○
		3.2	공개에 제한을 갖는 정보 보안 및 저작권 관련 설명을 제시함.	○
4	검색 인터페이스	4.1	키워드 및 브라우징 검색 접근점이 제공되고 있으나, 브라우징의 범위와 깊이가 제한적임.	△
		4.2	메뉴바 팝업창 풀다운메뉴 등의 제공은 활발하지 않음.	△
		4.3	검색된 콘텐츠와 관련된 검색확장을 제공되지 않음.	△
5	지원기능 인터페이스	5.1	이용자와 커뮤니케이션할 수 있는 피드백 정보가 부족함.	△
		5.2	위치정보를 지원하는 구체적인 네비게이션 기능이 없음.	△
		5.3	모든 소속 웹 페이지의 제목은 해당 주제에 따라서 제공되고 있음.	○
		5.4	불필요한 링크정보가 없고, web depth는 최종 콘텐츠에 이르기까지 대략 3~5 path를 이루고 있음.	○

지 하는 사례가 그에 해당한다.

국내의 OASIS의 경우에도 국내 디지털 콘텐츠 중 특화할 수 있는 컬렉션에 중점을 두고 아카이빙 하는 전략이 필요하다.

이 외에도 선진국의 국가 웹 아카이빙 사례를 통하여 국내에 적용하여 개선할 수 있는 방안들은 다음과 같다.

1) 수집 대상 웹 콘텐츠 주제 영역의 다양화

OASIS에서는 총류, 철학, 종교, 사회과학, 순수과학, 기술과학, 예술, 언어, 문학, 역사에 이르는 10개 주제 영역의 브라우징과 함께 직접 사이트 검색을 할 수 있는 인터페이스를 제공하고 있다. 학문 영역을 중심으로 한 웹 콘텐츠의 수집에 특히 강점을 보이고 있는데, 대상 영역을 확장하여서 LCWA의 사례와 같이 최신 사회 문화적 이슈 및 사건 등의 콘텐츠의 수집도 고려해 볼 수 있다.

2) 웹 자원 수집의 분담화

PANDORA의 경우에는 호주 관련 웹 콘텐츠들을 해당 주제 영역에 따라서 관련 기관들이 웹 수집 역할을 효율적으로 분담해오고 있다. 즉, 호주 전쟁기념관(Australian War Memorial)에서는 호주 전쟁 관련 웹 기록물들을, 호주 필름 사운드 아카이브(National Film and Sound Archive)에서는 필름 관련 주제 기록물들을 수집하고 있다. 국내에서도 아카이빙 대상 웹 사이트 수집 작업에 관련 기관들이 참여하는 방안을 고려해 볼 수 있다.

3) 특별 컬렉션의 구성

UK Web Archive에서는 브라우징 메뉴 밑

에 특별 컬렉션을 구성하여 제공하고 있다. 가장 경쟁력을 갖고 있는 특별 아카이브를 제시함으로써 차별화된 서비스를 제공하고 있는 것이다. OASIS에서도 구축된 기존 아카이브 컬렉션 중 경쟁력을 갖는 특정 컬렉션들을 별도로 구성하여 접근점을 제공하는 방안을 고려해 볼 수 있다.

4) 아카이빙 시점 및 흐름의 시각화

UK Web Archive, GCWA, 그리고 WAS에서 살펴볼 수 있듯이, 여러 차례 수행되는 아카이빙 버전을 용이하게 식별하고 적절히 제공하기 위하여 아카이빙 시점 및 흐름과 변화를 시각화하여 제시하는 경향이 늘어나고 있다. 현재 OASIS에서는 아카이빙 회차에 대한 정보를 간략히 텍스트 기반으로 제시하고 있지만, 직관적인 이해를 돕는 시각적 방법으로 표현하는 방법을 고려해볼 수 있다.

5) 체계적인 아카이빙 전략 수행

대부분의 국가에서 수립한 원칙에 따라서 아카이빙 작업이 정기적으로 수행되고 있다. 또한, 주제 영역의 규모적 편차도 어느 정도 균형을 이루고 있다. 현재 OASIS는 웹사이트의 보유량이 언어(7)에서 사회과학(111)에 이르기까지 카테고리에 따른 편차가 심하고 규모도 작다. 체계적인 아카이빙 전략 수행과 함께 아카이빙 웹 콘텐츠의 양적 확충이 필요하다.

6) 국제기구와의 적극적 협력 및 범용 툴 활용

IIPC를 구성하는 많은 국가들이 웹 아카이빙의 주요 이슈에 대하여 정기적으로 논의하고

합의를 이루어 오고 있다. 또한 많은 국가들이 IIPC에서 제안하고 있는 범용 툴을 활용하고 있다. 웹 아카이빙 관련 툴들은 아직까지 모두 완료된 것이 아니고 지금도 현재 진행형으로 계속되고 있다. 우리나라도 웹 아카이빙 관련 국제 협회에 적극적으로 참여해나감으로써 국제 트렌드에 동참할 필요가 있다.

6. 결 론

본 연구에서는 미국의 LCWA, 영국의 UK Web Archive, 캐나다 GCWA, 호주 PANDORA, 그리고 싱가포르의 WAS를 대상으로 선정하여, 국가 웹 아카이빙 사례의 웹 인터페이스를 살펴보았다. 인터페이스 영역은 총 다섯 개로서 시각 디자인 인터페이스, 웹 접근성 인터페이스, 보안 인터페이스, 검색 인터페이스 그리고 지원기능 인터페이스이다.

분석 결과, 국가마다 강점을 보이는 영역 인터페이스에는 차이가 있었는데, 영국과 호주의

경우에는 이미지 컬렉션을 활용한 검색 인터페이스가 우수했고, 호주는 이용자와의 커뮤니케이션 채널을 적극적으로 개발함으로써 지원 기능 인터페이스가 우수했다.

국내 OASIS의 경우에는 시각적인 인터페이스, 웹 접근성, 보안 인터페이스 영역은 우수하였으나, 검색 인터페이스와 지원기능 인터페이스는 미흡하여 관련 기능의 보강이 필요함을 확인할 수 있었다.

아울러 선진국 사례를 참조하여 우리나라 국가 웹 아카이빙 사업인 OASIS에 적용할 수 있는 방안으로는, 수집 대상 웹 콘텐츠 주제 영역의 다양화, 웹 자원 수집의 분담화, 특별 컬렉션의 구성, 아카이빙 시점 및 흐름의 시각화, 체계적인 아카이빙 전략 수행, 그리고 국제기구와의 적극적 협력 및 범용 툴 활용 등을 제안하였다.

디지털 유산에 해당하는 가치 있는 웹 자원의 아카이빙 노력 및 적극적 활용을 위한 인터페이스 강화 노력은 사회 환경의 변화 및 정보 기술의 트렌드와 병행해 나가면서 앞으로도 지속적으로 수행해 나가야 할 것이다.

참 고 문 헌

김유승. 2007. 웹 아카이빙의 법·제도적 문제에 대한 고찰. 『한국문헌정보학회지』, 41(3): 5-24.

김유승. 2008. 복합적 웹 아카이빙 정책에 관한 고찰. 『한국문헌정보학회지』, 42(4): 159-179.

서혜란. 2004. 웹 아카이빙의 성과와 과제. 『한국

비블리아학회지』, 15(1): 5-22.

오상훈 외. 2008. 디지털 자원의 웹 아카이빙을 위한 납본 프로세스 개발 및 기능 설계. 『정보관리학회지』, 25(4): 5-23.

이재운 외. 2007. 텍스트마이닝을 이용한 국내 기록관리학 분야 지적구조 분석. 『한국문헌정보학회지』, 41(1): 345-372.

- 차승준 외. 2008. 웹 기록물 아카이빙 기반기술 연구 개발. 『지식정보산업연합회 창립기념 학술대회』, 12: 359-368.
- 차승준 외. 2009. 공공기관 심층 웹 기록물 아카이빙을 위한 메타데이터 설계. 『한국전자거래학회지』, 14(4): 181-193.
- 최정태 외. 2001. 웹사이트 사용자 인터페이스의 품질 평가. 『한국도서관·정보학회지』, 32(4): 1-32.
- Brown, Adrian. 2006. *Archiving Websites*. Facet Publishing.
- Gladney, H. M. 2007. *Preserving Digital Information*. Berlin: Springer.
- Kane, S. K. et al. 2007. A Web Accessibility Report Card for Top International University Websites. *W4A 2007 Communications Paper*, May, 07-08: 148-156.
- Kim, Heejung and Hyewon Lee. 2007. Development of Metadata Elements for Intensive Web Archiving. *Journal of the Korean Society for Information Management*, 24(2): 143-160.
- Masanès, Julien. 2002. Towards Continuous Web Archiving. *D-Lib Magazine*, 8(12).
- Masanès, Julien. 2005. Web Archiving Methods and Approaches: A Comparative Study. *Library Trends*, 54(1): 72-90.
- Masanès, Julien. 2006. *Web Archiving*. Springer.
- McCord, S. K. et al. 2002. An Accessibility Assessment of Selected Web-based Health Information Resources. *Library Hi Tech*, 20(2): 188-198.
- Mitta, D., C. Delk and W. Lively. 1995. Selecting System Functionalities for Interface Evaluation. *Human Factor*, 37(4): 817-834.
- Myers, B., J. Hollan, and I. Cruz. 1998. Strategic Directions in Human Computer Interaction. *ACM Computing Survey*, 28(4): 794-809.
- Norman, D. A. 1991. Cognitive Artifacts. In J. M. Carroll (Ed.), *Designing Interaction: Psychology at the Human-Computer Interface*. New York: Cambridge University Press.
- Schmetzke, A. 2001. Web Accessibility at University Libraries and Library Schools. *Library Hi Tech*, 19(1): 35-49.
- Shneiderman, B. 1998. *Designing the User Interface, Reading*. Boston: Addison Wesley.
- Zhang, P. et al. 1999. Websites that Satisfy Users: A Theoretical Framework for Web User Interface Design and Evaluation. *Proceedings of the 32nd Hawaii International Conference on System Sciences*, 1-8.
- Australia's Web Archive: PANDORA. [cited 2010. 11. 5]. <<http://pandora.nla.gov.au/about.html>>.
- Computer Desk Encyclopedia. [cited 2010. 12. 13]. <<http://computer.yourdictionary.com>>.
- Government of Canada Web Archive. [cited

2010. 11. 5].
〈<http://www.collectionscanada.gc.ca/webarchives/search/index-e.html>〉.
- Library of Congress Web Archives. [cited 2010. 11. 5].
〈<http://lcweb2.loc.gov/diglib/lcwa/html/lcwa-home.html>〉.
- OASIS. [cited 2010. 11. 5].
〈<http://www.oasis.go.kr>〉.
- Trove. [cited 2010. 12. 13].
〈<http://trove.nla.gov.au>〉.
- UK Web Archive. [cited 2010. 11. 5].
〈<http://www.webarchive.org.uk/uk/wa/>〉.
- Vertelney, L., M. Arent, and H. Lieberman. Two Disciplines in Search of an Interface. [cited 2010. 12. 13].
〈<http://cim.mcgill.ca/~jer/courses/hci/ref/vertelney.pdf>〉.
- Web Archive Singapore. [cited 2010. 11. 5].
〈<http://was.nl.sg>〉.
- W3C. Web Accessibility Guidelines. [cited 2010. 12. 15].
〈<http://www.w3c.org/>〉.