



국가 차원의 장기적 계획 필요

인터넷 자원은 수시로 변경·추가·삭제되는 동적 문서이므로 장기적인 보존과 접근·이용에 어려움이 따른다. 따라서 인터넷 자원을 체계적으로 보존하여 활용할 수 있는 국가차원의 종합적인 인터넷 자원 보존소를 구축할 필요성이 제기되고 있다.

■ 안계성/한국데이터베이스진흥센터 연구조사부 정책연구담당 부장

21

세기 뉴밀레니엄 시대가 도래하였다. 21세기 지식 기반사회에서 한 나라의 국부는 사이버 공간에 축적된 디지털화된 자원의 질과 양에 의해 결정될 것이다. 디지털 자원은 지식사회를 이루는 핵심으로 그 가치가 날로 증가되고 있다. 디지털화된 지식은 국가적 지식과 경험에 대한 인식 증진과 접근 향상에 기여할 뿐만 아니라 각 산업의 발전과 기회확대를 가져올 수 있다.

이처럼 인간의 지식과 경험을 기록하기 위한 매체로서, 급증하고 있는 인터넷 자원은 지식기반사회의 새로운 정보매체로서 사회·문화적 가치가 높기때문에 국가 유산으로 간주될 수 있다. 그러나 인터넷 자원은 수시로 변경·추가·삭제되는 동적 문서이므로 장기적인 보존과 접근·이용에 어려움이 따르고 있기 때문에, 인터넷 자원을 국가적인 차원에서 체계적으로 보존하여 활용할 수 있는 국가차원의 종합적인 인터넷 자원 보존소를 구축할 필요가 있다.

인터넷 자원 보존소는 디지털 형태의 사회·경제·문화적 정보 자원을 수집·저장하여 장기적으로 접근·활용할 수 있도록 하는 개념으로, 실제로 미국, 영국, 호주 등은 자국의 인터넷 자원을 보존하는 국가 프로젝트를 추진하고 있다. 이제 국내에서도 유통되고 있는 인터넷 자원을 체계적으로 조사하고 망라적으로 수집하여 장기적인 활용을 위한 국가 보존 시스템을 구축해야 한다.

이를 위해 선진 각국의 인터넷 자원 보존 국가프로젝트의 현황을 조사하고, 각 프로젝트의 비교·분석을 통하여 국내에 적합한

보존모델을 도출해야 하며, 대상 선정 기준을 마련을 위해 장기 보존할 인터넷 자원의 대상과 세부 선정기준을 마련해야 한다.

또한 인터넷 자원 조사·수집, 인터넷 자원 보존 시스템 구축, 인터넷 자원의 저장을 위한 시스템 설계, 보존된 디지털자원의 접근방법 설계, 실제 데이터베이스 구축, 시스템 평가 및 유지보수 등의 계획을 더 이상 미뤄서는 안된다.

인터넷자원 장기보존의 배경

인터넷의 장점은 새로움과 역동성이라고 할 수 있다. 인터넷에는 매일 1,500만 쪽 이상의 자료가 증가하고 있으며, 현재는 약 10억 쪽에 달하는 자료가 디지털화된 형태로 존재하는 것으로 예측되고 있다.

그러나 전체 인터넷 자원의 1%가 1주일 후에는 사라지며 인터넷 자원의 평균 예상 수명은 대략 70일 정도로 예측된다. 어떤 인터넷 사이트를 검색하여 자료를 확인했다고 그 자료가 언제나 여전히 그곳에 존재하고 있을지 혹은 그 내용이 여전히 동일할지 보장할 수 없다. 이 중에는 가치를 신뢰할 수 없는 인터넷 자원도 많은 반면, 중요한 자원도 많다. 그러나 안정성의 수준은 보장되는 것이 아니다.

인터넷이 빠르게 확산됨에 따라 인터넷 자원을 제공하고 평가하는 것이 어려워졌으며 이에 따라 이제는 저장과 보존방법을 개발할 때가 되었다. 매체에 상관없이 국가 문화유산은 보존되어야 하기 때문에 이를 디지털화된 인터넷 자원은 저장 및 접근 체계를 가지고 있는 지식 자원으로 수집되어야 하고 수집된 후에는

디지털 자원으로서는 가지는 장기 보존상의 여러 문제를 해결해야 할 것이다.

보존 분야에서 직면하고 있는 가장 분명하고 두드러진 변화는 현재 정보가 저장되어 있는 매체의 비영속성과 취약성이다. 더욱 커다란 문제는 디지털 기술이 빠른 속도로 노화된다는 것이다. 디지털 형태의 정보를 관독하는 데에는 소프트웨어와 하드웨어가 필요하다. 이 설비는 자주 변화하고 수년 사이에 시장에서 완전히 사라질지도 모른다.

인터넷 자원 장기 보존의 개요

디지털 형태의 자원을 보존한다는 것은 그 내용 뿐만 아니라 상황정보와 구조정보를 보존하는 것이다. 구조정보가 없으면 자원의 본래 형태대로 재구성하는 것이 불가능하다. 또한 이름이나 제목, 관련문서와의 연결, 날짜소인, 네트워크 시스템에 저장되어 있는 관련 데이터와 같은 상황정보를 보존하지 못한다면 이후 디지털 문서의 신빙성을 확인할 수 없다.

특히, 네트워크 환경에서는 자원의 수명주기의 초기단계에 어떤 자원을 보존할지에 대한 결정이 내려져야 한다. 오래된 자료가 가치있는 자료가 되었던 관행과는 달리 디지털 자원의 경우는 어떤 레코드를 계속 보유할지 판단하기 위해 몇십년을 기다린다는 것은 불가능하다. 그것은 결국 노후한 포맷의 레코드를 보유하는 것이 되며 연구에 이용할 수 있는 충분한 상황정보를 얻지 못할 수도 있기 때문이다.

인터넷 자원의 장기 보존은 기술적인 문제에서 시작되었다. 그러나 오히려 해결하기 어려운 문제는 조직적, 경제적, 관리적 측면에서 지적 재산권의 보호와 보존주체, 보존대상에 관한 것이다.

자원을 보존할 일차적인 책임은 그 자원의 생산자에게 있으며, 이밖에 자원의 제공자, 소유자가 자신의 정보를 보존할 책임을 지게 된다. 그러나 자원을 생산하는 기관이 보존책임까지를 모두 맡을 것인가는 의문이다.

국가 디지털 보존기구는 자료가 분산되는 특성에 따라 분산된 자원을 운영하며 다양한 매체 및 형식을 가지고 있는 자료들을 적절한 기관에 분산하여 보존하는 방식으로서, 일부 국가적 관심의 대상이 되는 자료는 국가도서관이 보존하도록 할 수 있다. 국가 디지털보존기구의 구축은 기존 디지털 자원의 검증에서부터

시작되어야 할 것이며 보존될 디지털 자원의 국가적 등록체제를 필요로 할 것이다.

디지털 보존의 본래 성격과 규모 자체는 기관간의 협력 활동을 필요로 한다. 단일기관이 모든 디지털 자료를 보존하는 책임을 맡지는 못할 것이다. 관리기관간 협정을 체결해야하고 협의체를 구성해야 한다. 다양한 주제분야나 중복 정도 등에 의해 책임을 분산하는 것도 필요하다.

장기적인 해결을 위해서는 기존의 경계를 구분하지 않는 협력을 필요로 한다. 여기에는 전통적으로 보존역할을 담당해 왔던 도서관과 기록보존소 이외의 대학 전산기관, 전자레코드 관리자, 데이터 제공자, 공공부문 등이 포함될 것이다.

인터넷 자원의 장기 보존을 위한 지침

인터넷 자원의 장기 보존을 위해서는 우선적으로 선정정책이 필요하며 선정정책의 필요성에 대해서는 여러 의견이 있지만 점차 인터넷 자원의 가치를 따져서 양질의 자료를 선별해야 한다는 방향으로 모아지고 있다.

단지 인터넷 자원에 접근할 권리를 획득한다고 해서 인터넷 자원을 보존할 수는 없다. 실제로 출판물의 사본을 수집하고 저장해야 한다. 영구적인 파일 포맷을 사용하며 장기적으로 데이터의 가치와 출처의 신빙성을 확보해야 한다. 디지털 자원은 이용이 적거나 많거나 간에 점차 그 가치가 상실된다. 따라서 자료를 살아있게 보존하지 않으면 이후에 디지털자원에 접근하는 것이 쉽지 않게 될 것이다.

또한 적절한 접근방법이 없다면 디지털 객체를 보존한다는 것은 시간과 자원의 낭비일뿐이다. 이를 위해 메타데이터를 작성해야 하지만, 검색을 위한 메타데이터와 장기보존을 위한 메타데이터는 다르다. 보존을 위한 메타데이터는 디지털 객체의 보존을 지원하는 기술정보를 저장하는 수단으로 사용될 수 있는 관리용 메타데이터의 형태가 된다. 이러한 메타데이터라면 디지털 자원과 메타데이터의 관계를 재설정하고 메타데이터 요소를 상세히 해야하며 그러한 필요성은 이미 여러곳에서 제기되어 왔다.

다양한 디지털 자료의 비용·효과적인 장기 보존을 달성하기 위해서 필요한 가장 중요한 원칙은 모든 것이 단순하게 유지되어야 한다는 것과 디지털 보존용 자료는 표준 포맷과 표준매체와 표준 운영시스템으로 운영해야 한다는 것이다.



해외 사례

최근들어 각국에서는 자국의 인터넷 자원을 보존하기 위한 움직임이 보이고 있다. 이러한 사례에는 광범위한 이용자에게 개방되는 자료를 대상으로 하는 사례와 국가도서관이 주축이 되어 추진하는 사례로 나눌 수 있는데 대표적인 사례는 <표>와 같다.

국가 인터넷자원 장기 보존체제 모형

인터넷자원 장기보존 국가프로젝트의 선행사례는 대부분 아직 실용화 단계에 있지는 않고 프로젝트가 진행중이거나 시범서비스 단계에 있다. 이는 기본적인 자원의 식별과 기술을 위한 체제와 다양한 형태의 자원을 수집하기 위한 기제가 모두 마련되어 있다는 전제에서 장기 보존을 위한 체제가 설립될 수 있으며 다양한 형태의 디지털 자원을 장기 보존하기 위한 효과적인 기술과 전략이 아직 궁극적인 해결책에 도달하지 못하였기 때문이다.

따라서 인터넷자원 보존 국가 프로젝트를 구축하기 위해서는 다음과 같은 사항을 결정해야 한다.

첫째, 국가차원을 전제로한 책임기관을 설정하고, 설정시 중앙집중식과 분산방식중 어떤 것이 효과적인가 타당성을 검토해야 한다.

둘째, 수집자원의 선정정책을 설정해야 한다. 자원의 선정정책은 책임기관의 설정과도 관련되어 있다. 이와같은 기본적인 사항을 결정한 후에 보다 구체적인 국가 인터넷자원 보존을 구축하기

위한 논리적, 기능적, 기술적 필요사항을 도출하고 상세화해 나가야 할 것이다.

모형의 기본적인 기능은 수집·저장·접근을 기본으로 한다. 이를 위해 책임기구를 설정해야하는데, 새로운 기관을 설립할 수도 있으며 기존의 국립중앙도서관이 모든 기능을 맡아 기존의 체제를 확장하는 방법과 국가전자도서관 프로젝트를 보완한 형태로 운영할 수도 있다.

또한 인터넷 자원의 선별정책과 기준이 설정되어야 하는데, 일반적인 기본 지침은 장기적으로 혹은 영구적으로 보존할만한 중요한 학술적 가치나 역사적 가치가 있는 자료로 한정시켜야 한다는 것이고, 중복된 자원에 중복된 투자가 이루어지지 않도록 해야 한다는 것이다. 따라서 선정기준은 자국자료로 한정하고, 인쇄형태로도 존재하는 자료는 제외해야하며, 상업적인 온라인 서비스, 광고나 홍보자료, 다른 자원을 안내하는 서비스는 제외하며, 개인 홈페이지는 제외해야 한다.

보존체제에서 보존기술정보는 매우 중요한데, 특히 보존기술정보로서 필요한 메타데이터는 모두 기술해야 한다. 보존을 위한 필수 메타데이터 요소로는 식별기호, 수집, 진위여부, 자료의 물리적인 특징, 기술적 관리 접근제어와 저작권 관리 등이다.

국가 인터넷자원 보존을 위한 시스템은 수집기능, 저장기능, 접근기능, 관리기능 등으로 구성되며 이를 설명하면, 수집은 웹 로봇을 이용한 자동수집을 전제로 하며 수동선별이 병행되는 방식

<표> 주요국의 인터넷 자원 보존 프로젝트

프로젝트	국 가	시 기	수집방법	수집범위
Kulturarw3	스웨덴	1996.9~	로봇을 이용한 자동 수집	스웨덴의 모든 웹사이트
	전자출판물 장서의 기초, 능동적 자료 수집			
INDOREG	덴마크	1996~	수집. 이후에 출판사의 직접 등록으로 바뀔 계획	몇 가지 사항을 제외한 모든 웹사이트
	국가서지등록 시스템 구축, 전자문헌등록과 전자문헌 제어 중심			
EVA	핀란드	1997.1~1998.6	로봇에 의한 자동 수집	수집 모든 웹사이트
	문헌관리를 위한 기술적 문제 고려(파일 포맷 등)			
PANDORA	호주	1997~	SCOAP 지침에 따른 수집	강력한 선정기준에 따른 전자출판물
	전자 보존소 구축, 선정과 접근 중심의 프로젝트			
NEDLIB	EU	1998.1~2000.12	검토 중	검토 중
	전자출판물 보존 시스템을 위한 국가간 협력, 표준 정의			
CEDARS	영국	1998.4~2000.4	DESIRE 선정원칙에 따른 수집	선정기준에 부합한 전자출판물
	eLib의 3단계 프로젝트, eLib, JISC의 다른 프로젝트와 연결			

미국의 경우는 1996년 WAIS를 개발한 Brewster Kahle이 동료들과 함께 Internet Archiv를 개발하여 미국의 인터넷 공공자료를 수집하고 있음(<http://www.archiv.org>)
Sun Microsystems, Apple Computer, UC Berkeley, 스미소니안 박물관 등의 후원으로 수행 및 유지·관리되고 있음

으로 추진하고, 저장기능은 시간이 지나서 새로운 기술로 이전하는데 관련된 문제를 최소화하기 위해 표준방식으로 저장하는 것을 기본으로 해야하며 디지털 객체와 메타데이터를 나누어 저장해야 한다. 접근기능은 분산되어 있는 개별 자원의 메타데이터 저장소에의 접근을 통합하는 서비스를 지원해야 하며, 관리기능은 저장소의 모든 디지털 서비스의 통합관리를 지원해야 할 것이다.


결론

인터넷 자원의 폭증과 가치 증대로 인해 이들 자원의 보존이 시급히 요청되고 있으며, 이는 국가적 유산으로서 고려되어야 하며, 국가는 기존의 기록물 보존과 관련된 사업과 마찬가지로 인터넷 자원에 대한 인식을 제고하고 더욱 방대한 자원이 생성되기 전에 이를 단계적으로 해결해야 한다. 정보의 바다라 불리는 인터넷에서, 유통되는 각종 자원 가운데 가치가 높은 자원의 발굴과 보존이 선별적으로 추진되어야 하고, 이러한 사업은 적극적인 하향식 정책 추진 뿐만 아니라 모든 국민이 참여할 수 있는 상향식 정책 전개가 보다 효율적이다.

그리고 범 국가적인 인터넷 자원의 장기 보존체제를 수립하고 이를 계속적으로 유지해야 한다. 이는 인터넷 자원이 지식기반사회의 새로운 정보매체로서 국가 유산으로 간주될 수 있기 때문에

이를 효율적으로 보존·활용할 수 있는 국가의 장래를 위한 인터넷 정보자원 보존 정책이 수립되어야 한다. 또한 인터넷 자원을 국가적인 차원에서 체계적으로 보존하여 활용할 수 있도록 종합적인 인터넷 자원 보존소를 구축하여 국가정보자원관리를 효율화해야 할 것이다.

미국을 비롯한 선진 각국에서는 인터넷 자원을 보존하는 국가 프로젝트를 추진하고 있으며, 자국내에서 유통되고 있는 인터넷 자원을 체계적으로 조사하고 망라적으로 수집하여 장기적 활용을 위한 국가 보존 시스템을 구축함으로써 영구적인 저장 방안을 모색하고, 재난 방지 및 복구방안을 마련하고 있다. 이를 위해서는 인터넷 자원의 범위 설정, 효율적인 보존 시스템 개발, 인터넷 주소체계의 효과적인 파악, 저작권 관리체계의 효율화 등이 선행되어야 할 것이다.

이와같이 국가적 차원에서 인터넷 자원을 보존하는 것은 자국의 자원을 발굴하여 널리 활용함으로써 지식기반사회의 인프라를 마련하고, 새로운 매체인 인터넷자원은 사회·문화적 가치가 높아 국가 유산으로 간주될 수 있기 때문에 인터넷 자원을 보존·활용함으로써 사회적 기록유산의 수호 역할을 할 수 있으며, 이를 통해 새로운 밀레니엄의 지식기반사회를 선도해 나갈 수 있도록 해야 할 것이다. 

정기구독안내

1. 구독신청방법

1. 일단, 02-318-5050(ext 119)번으로 전화하여 안내를 받으실 수 있습니다.
2. 아래의 은행구좌로 구독료를 입금하신 다음 데이터베이스월드 담당자와 통화하시면 됩니다.
3. 구독자 또는 구독기관명, 구독기간, 책을 받아보실 주소, 신청인 주소와 전화번호 등을 적어서 02-318-5040번 팩스로 넣어주셔도 정기구독자로 등록됩니다.

2. 구독료 입금계좌

조흥은행 수송동지점 390-03-003978
 국민은행 세종로지점 023-25-0008-729
 * 예금주 : 한국DB진흥센터

3. 정기구독료

- 6개월 : 24,000원
- 1년 : 44,000원
- 2년 : 88,000원
- * 권당 가격은 4,000원입니다.
- * 정기구독을 신청하시면 편안히 책을 받아보실 수 있습니다.

재단법인 한국데이터베이스진흥센터
 100-180 서울시 중구 다동 10번지 한국관광공사빌딩 9층

데이터베이스월드
 The Database World