

웹에서의 전거통제 : 디지털 도서관의 지식조직 시스템(KOS)과 전거 통제

Authority Control on the Web: Knowledge Organization Systems and Authority Control for Digital Libraries

윤 정 옥*

Cheong-Ok Yoon

차 례

- | | |
|--------------------|--------|
| 1. 머리말 | 4. 맺음말 |
| 2. 지식조직 시스템 | · 참고문헌 |
| 3. 지식조직 시스템과 전거 파일 | |

초 록

이 연구의 목적은 디지털 도서관에서 지식조직 시스템을 사용함으로써 전통적인 전거 파일의 역할을 확대하여 활용하는 사례를 살펴보는 것이다. 지식조직 시스템은 정보를 조직하고 지식경영을 조정하는 모든 유형의 체계로서 디지털 도서관의 핵심적 요소 중의 하나이다. 다양한 지식조직 시스템은 웹에서 목록과 디지털 자원 사이의 링크를 제공하고 이용자의 정보요구와 자원 사이의 다리 역할을 수행하면서 디지털 컬렉션으로부터 상관된 아이템의 검색을 지원하도록 콘텐츠를 조직한다. 지식조직 시스템이 코드를 전문으로, 객체를 기술 레코드로, 그리고 물리적 객체를 소재와 실물정보로 링크시키기 위하여 어떻게 사용될 수 있는가를 서술하였다.

키 워 드

전거파일, 전거통제, 디지털 도서관, 디지털 자원, 지식조직 시스템

* 한국과학기술정보연구원 정보자료센터 초청선임 연구원
(Visiting Senior Researcher, Information Resources Center, KISTI, ynco0120@kisti.re.kr)
· 논문접수일자 : 2002년 11월 9일
· 게재확정일자 : 2002년 12월 11일

ABSTRACT

The purpose of this study is to review the expanded role of authority files, which uses the Knowledge Organization System(KOS) to organize contents and manage the collections in digital libraries. The KOS plays a role as a bridge between the users' information needs and materials by providing links between catalogs and digital resources on the web and supporting retrieval of the relevant items. It is described how a KOS can be used to link a digital resource to related materials, to provide more descriptive records for entities, and access to relevant physical objects.

KEYWORDS

Authority File, Authority Control, Digital Libraries, Digital Resources, Knowledge Organization System

1. 머리말

최근 디지털 도서관이 확대되고 웹 상에서 접근할 수 있게 됨에 따라 지금까지 도서관의 목록과 색인에서 사용되던 전통적 개념의 전거통제가 다양한 형태의 지식조직 시스템을 통하여 확대될 것이 기대되고 있다. 전거통제는 도서관 소장자료를 위한 목록에서 오직 관심 대상인 레코드나 아이템만 검색해 내도록 정확성을 제공하고, 또한 상호참조를 주어 모든 상관된 자료들을 모을 수 있도록 재현성을 높여 줌으로써 목록에서의 찾기와 모으기(location and collocation)기능을 지원하는 역할을 해왔다. 전거통제를 가장 먼저 그리고 가장 널리 적용해 왔던 목록을 웹 상에서 접근할 수 있고, 색인, 초록, 이미지 및 전문(full text) 등의 디지털 자원들을 직접 링크할 수 있게 됨에 따라 이같이 보다 다양한 자원들에 대한 접근을 지원하기 위하여

전거통제를 적용할 필요성이 더욱 커지고 있다.

지난 2000년 미국 국회도서관에서 열린 'Bicentennial Conference on Bibliographic Control for the New Millenium'에서 틸렛은 웹 검색의 혼란과 불충분함에 대한 인식이 커감에 따라 전통적 도서관 이외의 영역에서도 "도서관에는 뭔가가 있을 것이다(There may be something out there)."라는 것에 대한 관심이 높아져가고 있다는 사실을 지적하였다(Tillett 2001). 전거통제는 목록에서 특정한 대상을 가리키는 여러 이름들과 개념들 가운데 어떤 형태를 선택하여 기본형으로 삼고, 그것을 그 밖의 다른 변형들과 연결해 줌으로써, 이용자들이 어떤 형태의 이름이나 개념을 갖고 검색을 시도하든지 의미 있는 결과를 얻을 수 있도록 도와준다. 따라서 도서관에서만 아니라 아카이브나 박물관, 미술관 등과 같이 보유자원을 디지털화하고, 목록과 색인을

이용하여 검색하고 원자료에 접근할 수 있도록 허용하는 다양한 기관들에서 전거통제를 적용할 필요성이 높아진다.

최근에 워너와 브라운은 미국 존스 홉킨스 대학에서 소장하고 있던 '레스터 악보 컬렉션(The Lester S. Levy Collection of Sheet Music)'을 디지털화하면서 이름을 전거통제하였던 사례를 보고하였다. 이 악보 컬렉션은 레스터 S. 레바이가 주로 1780년대부터 1980년대까지 2백여 년 동안에 간행된 미국의 대중음악 악보 29,000점 이상을 수집하여 만든 것인데, 이 대학의 아이젠하워 도서관 특별 컬렉션에 속해 있었다. 이 컬렉션을 디지털화하면서, 음악도서관협회(Music Library Association)의 지침에 따라 풍부한 온라인 색인을 만들어 주었다. 이 색인에서는 작곡가, 작사가, 편곡자와 표지 삽화가 등의 이름을 악보에 나타난 그대로 베껴 쓴(transcribed) 정보를 제공하기는 했지만, 전통적인 개념의 '전거통제'를 적용하지 않았다. 따라서 컬렉션을 탐색하는 이용자들은 동일한 인물을 나타내는 다른 인명이나 동일한 인명의 변형들로 연결되지 않았고, 자기가 이 컬렉션의 메타데이터에서 특정한 예술가나 작곡가에 관하여 제공하는 모든 것을 발견했는지 알 수 없었다는 것이다. 예를 들어 "Stephen Foster"를 찾으면 "Stephen Collins Foster" 혹은 "Stephen C. Foster"로 되어 있는 모든 사례들을 검색할 수 없었고, 이용자는 "Alice Hawthorne"라는 이름으로 탐색했을 경우에 동일 인물이 "Septimus Winner"라는 이름으로서 발표한 작품은 찾을 수가 없었다. 결국 탐색의 능력을 확대시키기 위

하여 전거통제가 필요하다는 것이 입증되었고, 자동화된 방식으로 '소급 전거통제'를 수행하였다. 이 작업을 위하여 메타데이터에 기록된 모든 작곡가, 작사가, 편곡자, 판화가, 인쇄가 및 미술가의 이름들을 추출해내고, LC의 이름전거 파일을 이용하여 이름들의 전거형태를 자동적으로 생성하도록 하였다(Warner and Brown 2001).

이 같은 사례는 전거통제를 기본적으로 도서관에서 목록과 색인을 작성할 때에만 적용하고, 특수한 컬렉션을 디지털화하고 그것의 색인을 만드는 과정에서는 필요하다고 생각하지 않았기 때문에 발생하였다고 할 수 있다. 최근 웹에서 전세계 도서관의 온라인 목록을 탐색할 수 있고, 도서관, 박물관, 미술관, 기록문서관 등이 소장하고 있는 주제 컬렉션과 각종 자원을 디지털화하여 웹에서 이용할 수 있도록 공개하고 있다. 따라서 목록으로부터 원래의 정보자원에 접근할 수 있는 가능성이 보다 확대되고, 이제는 목록만이 아니라 자원 자체로부터 특정한 이름과 관련된 정보 혹은 추가적인 정보를 제공할 필요성이 커지게 된다. 디지털 도서관에서는 목록에만 전거통제를 적용할 것이 아니라, 보다 다양한 형태의 자원들을 대상으로 전거통제를 수행해야 할 필요가 있다는 것을 의미한다.

실제로 온라인 목록의 서지 레코드 및 전거 레코드와 디지털 자원 사이에 하이퍼텍스트 링크를 만들어 줌으로써 목록은 디지털 자원에 대한 게이트웨이 역할을 하고 또 그 역순으로 디지털 자원으로부터 목록으로 접근할 수 있게 되었다. 먼저 목록의

이용자가 서지 레코드에 기술된 내용의 전문이나 거기에 인용된 발견도구, 혹은 전거 레코드에 인용된 참고도구의 서지 엔트리 등으로 직접 링크할 수 있다. 또한 참고도구나 온라인 문헌으로부터 시작하여 이용자가 선택한 다른 온라인 도서관 목록으로 링크하여, 그 개인이나 단체들이 저자이거나 주제가 되는 다른 저작들을 탐색할 수도 있게 된다.

하지(Gail Hodge)는 최근 디지털 도서관 연맹(Digital Library Federation)을 위한 보고서(2000)에서 지식조직 시스템을 전거 통제를 위하여 어떻게 사용할 수 있는가에 대하여 기술하였다. 기존에 전거통제를 위해서 가장 보편적으로 사용되었던 것은 이름전거 파일로서, 주로 개인과 단체의 다양한 이름들과 이름의 형태들을 처리하기 위해 많이 적용되었던 것이다. 그러나 최근에는 디지털 도서관에서 목록과 디지털 자원을 링크하기 위하여 여러 유형의 지식조직 시스템을 보다 확대된 개념의 전거 파일로 사용할 수 있고, 따라서 전거통제의 개념도 훨씬 넓어질 수 있다는 점을 강조하고 있다. 실제로 하지가 주목한 것은 디지털 도서관에서 특정한 지식조직 시스템이 목록과 자원 사이의 링크를 위한 중간 파일(intermediate file) 역할을 함으로써 전거통제의 기능을 한다는 것이다. 이 논문에서는 하지의 보고서를 중심으로 하여 디지털 도서관 환경에서 다양한 지식조직 시스템을 어떻게 전거 파일로 사용하고 전거통제를 적용하는가에 관하여 살펴보고자 한다.

2. 지식조직 시스템 (Knowledge Organization System : KOS)

2.1 지식조직 시스템과 디지털 도서관

지식조직 시스템(Knowledge Organization Systems : KOS)이란 용어는 미국의 펜실베이니아주 피츠버그시에서 1998년 열렸던 ACM Digital Libraries' 98 Conference에서 Networked Knowledge Organization Systems Working Group에 의하여 처음으로 사용되었다. 지식조직 시스템은 "정보를 조직하고 지식경영을 조장하는 모든 유형의 체계"라고 정의되며, 분류와 범주화 체계처럼 일반적 수준으로 자료들을 조직하는 것부터 보다 상세한 접근을 제공하는 주제표목, 지리명과 개인명과 같이 핵심적인 정보의 다양한 이형들을 제어하는 전거 파일 등을 포함한다. 또한 시소러스처럼 매우 체계적인 언어들도 포함하는가 하면, 의미망과 온톨로지처럼 훨씬 새로운 개념의 스킴들도 포함한다(Hodge 2000). 지식조직 시스템은 사실상 전혀 새로운 것을 의미하는 용어는 아니며 정보를 조직하기 위한 도구로서 대부분 지금까지 모든 도서관, 박물관과 아카이브의 중심에 있어 왔다고 할 수 있다. 여기에서는 특히 최근 디지털 도서관 환경에서는 지식조직 시스템의 목적이 "디지털 도서관 컬렉션으로부터 상관된 자료를 검색할 수 있도록 지원하기 위하여 콘텐츠를 조직"하는 것이라는 점에 주목하여 살펴보고자 한다.

지식조직 시스템은 검색의 목적으로 자료를 조직하고 컬렉션을 관리하기 위하여

사용되는데, 이용자의 정보요구와 컬렉션 내의 자료 사이의 가교 역할을 한다고 할 수 있다. 모든 디지털 도서관에서는 한 가지 이상의 지식조직 시스템을 사용하기 마련인데, 디지털 도서관에 포함되는 자원에 그것이 적용될 수 있어야 하고, 일단 컬렉션에 자료가 포함되면 그 지식조직 시스템이 이용자들에게 의미있는 것이 되어야 한다. 따라서 어떤 지식조직 시스템을 채택하여 사용할 것인가를 결정하는 것이 디지털 도서관 개발의 핵심이라고 할 수 있다. 다음에서는 지식조직 시스템의 유형을 간략하게 소개하도록 한다. 그러나 이들 가운데 이 논문에서 관심을 두고 있는 전거 파일의 역할을 하는 것은 그 일부에 해당하는 것임을 밝혀둔다.

2.2 지식조직 시스템의 유형

어떤 형태로든 지식을 조직하려는 시도는 인류의 문명과 함께 계속되어 왔다고 할 수 있다. 그러나 모든 사람이 동의하는 단일의 지식분류 시스템은 없다는 전제 하에, 하지는 특히 디지털 도서관을 조직하기 위하여 적용할 수 있는 지식조직 시스템을 분류하여 개관하였다. 하지는 이들을 크게 용어 리스트(Term lists), 분류와 범주화(Classification and categories) 및 관계 리스트(Relationship list)로 나누었는데, 그 구조, 복잡성, 기능은 상이하지만 모두 디지털 도서관에서 자료의 조직 및 자료에 대한 접근을 위해 사용될 수 있다고 하였다(Hodge 2000). 다음에서는 지식조직 시스템을 간략하게 소개한다.

2.2.1 용어 리스트

용어 리스트는 보통 용어와 정의를 함께 포함하고 있는 것을 말한다. 이들 가운데에는 전거 파일, 용어집, 사전 및 가제티어(Gazetteers) 등이 있다. 하지는 전거 파일을 “객체의 다양한 이름들이나 특정한 필드의 도메인 값들을 통제하기 위하여 사용되는 용어의 리스트”라고 정의하고 있다(Hodge 2000). 사공철 등이 편한 <정보학사전>에서는 전거 파일을 “전거통제를 위한 통제어휘집”이라고 표현하였고, 이름전거 파일과 주제전거 파일로 대별하였다(사공철 외 2001). 전거 파일은 기본적으로 선호되지 않는 용어들을 선호되는 용어들로 링크시켜 주면서, 간혹 제한된 계층구조를 보여주기도 하지만 보통 깊은 구조나 복잡한 조직을 갖지 않는다. 대표적인 것은 미국국회도서관의 ‘이름전거 파일(Library of Congress Name Authority File. LCNAF)’과 게티 보케뷰러리 프로그램(Getty Vocabulary Program)의 ‘예술가 종합인명록(Union List of Artist Names. ULAN)’을 들 수 있다.

용어집은 보통 정의를 포함한 용어 리스트로서, 특정한 주제 분야나 특별한 저작으로부터 그 용어들을 가져오기도 한다. 용어는 특정한 환경 안에서 정의되고 다양한 의미들을 포함하지는 않는다. 대표적인 것은 미국 환경청(U.S. Environmental Protection Agency. EPA)의 ‘환경용어집(Terms of the Environment)’이 있다.

사전은 단어와 그에 해당하는 정의를 수록한 알파벳순 리스트로서 보통 용어집보다 일반적인 내용을 담고 있다. 해당되는

경우에는 다양한 의미들이 제공되고, 단어의 어원, 철자의 이형, 학문별 복수의 의미 등을 포함하기도 하지만, 계층적 구조나 개념적 분류 등은 하지 않는다.

가제티어는 장소명의 리스트로서 대표적인 것으로는 미국의 인구조사국(U.S. Bureau of Census)의 'U.S. Gazetteer' 들 수 있다. 게티 보케뷰러리 프로그램의 'The Getty Thesaurus of Geographic Names(TCN)'도 시소러스라는 이름을 갖고 있으나 가제티어로 볼 수 있으며, 캘리포니아 주립대학 샌타 바바라(University of California, Santa Barbara. UCSB)에서 주관하는 'Alexandria Digital Gazetteer' 등도 주목할 만하다.

2.2.2 분류와 범주화

주제집합의 생성을 위하여 사용되는 지식조직 시스템은 크게 주제표목과 분류체계(classification schemes), 택소노미(taxonomies) 및 범주화 체계(categorization)로 나눌 수 있다. 주제표목은 컬렉션에서 아이템의 주제를 표현하기 위하여 사용하는 통제된 용어들로서 포괄적이고 넓은 주제 범주를 포함하기도 한다. 대표적인 것으로는 잘 알려져 있고 널리 사용되는 '의학 주제표목(MeSH)'과 '미국국회도서관 주제표목(LCSH)'을 들 수 있다.

분류체계, 택소노미 및 범주화체계는 주제를 숫자나 알파벳으로 표기하고 계층적으로 배열하는 특성을 갖는 것으로 세 용어들은 미세한 차이가 있기는 하지만 종종 호환되기도 한다. 이들은 보통 'ANSI NISO Thesaurau Standard(Z39.19)'에서 요

구하는 계층의 규칙을 따르지 않고, 시소러스에서 보이는 것과 같은 명백한 관계들도 보여주지 않는다. 미국국회도서관주제분류체계(LCC)나 듀이십진분류표(DDC) 등이 여기에 속한다. 택소노미는 특별한 특징에 따라 집단화된 객체들을 가리키기 위하여 객체 지향의 디자인과 지식경영 시스템에서 사용되고 있고, 범주화체계는 야후와 같은 것을 들 수 있다.

2.2.3 관계 리스트

관계 리스트는 용어와 개념 간의 관계를 중시하는 지식조직 시스템으로 시소러스, 의미망(semantic networks)와 온톨로지(ontology)를 들 수 있다. 시소러스는 개념 기반으로 용어들 사이의 관계를 제시하는데 보통 계층, 동등 및 상관관계를 BT, NT, SY, RT 등의 기호로 대표한다. 미국 의학도서관의 UMLS(United Medical Language System)가 대표적이다. 의미망은 자연어의 처리를 위하여 개념과 용어들을 계층 아닌 노드와 링크로 표현한 체계이고, 온톨로지는 가장 새로운 형태의 지식조직 시스템으로 여겨지며 주로 특정한 개념영역을 기술하고 객체간 복합관계를 표현하는데 사용된다.

3. 지식조직 시스템과 전거 파일

하지가 개관한 지식조직 시스템 가운데 전거 파일이 가장 전통적인 개념의 전거통제를 위한 도구이지만, 그 구성은 다른 지식조직 시스템들과도 밀접하게 연관되어

있다. 실제로 전자 파일을 이름전자 파일과 주제전자 파일로 구분할 때 주제전자 파일은 지식전자 시스템 가운데 '분류와 범주화' 혹은 '관계 리스트'에 속하는 것들로부터 만들어지게 된다. 즉 주제전자 파일은 주제명 표목표로부터 생성되는 주제명 색인, 분류표로부터 생성되는 분류색인 및 시소러스로부터 생성되는 키워드 색인으로 나누어질 수 있다(사공철 외 2001). 전자 파일은 배타적인 개념이 아니라 다양한 지식조직 시스템을 통하여 전자통제를 수행할 수 있도록 대표되는 개념으로도 볼 수 있는 것이다.

여기에서 디지털 자원을 상관된 자료들에 링크시키기 위하여 어떻게 지식조직 시스템을 전자 파일로서 사용하는가를 살펴 보도록 한다. 먼저 디지털 자원에 포함된 코드를 보다 설명적인 전문(full text)으로 확장하는 것, 객체에 대하여 보다 기술적 레코드(descriptive record)를 제공하는 것 및 상관된 물리적 객체에 대한 소재와 실물정보에 대한 접근을 제공한다는 측면에서 볼 수 있다.

3.1 코드를 전문으로 확장

코드를 전문으로 확장하는 것은 특정한 주제 분야 내에서 커뮤니케이션을 촉진시키기 위하여 코딩 체계를 사용하는 관행 때문에 필요하게 된다. 디지털 자원에서 발견되는 코드를 그것이 의미하는 완전명과 그것에 관한 설명을 포함하는 전문으로 연결해 줌으로써 이용자들을 지원하는 역할을 하게 된다. 대표적인 예로는 데이터뱅크

의 등록 코드와 바이오 시퀀스 데이터 사이의 링크, 산업 코드와 완전 산업명 사이의 링크를 들 수 있다.

3.1.1 시퀀스 번호와 바이오 시퀀스 데이터뱅크 링크

데이터뱅크는 분자생물학, 생명공학, 유전학 분야에서 생화학 및 유전학 시퀀스의 데이터를 저장하고 있다. 단백질 시퀀스, 뉴클리오타이드, 셀 라인 등을 포함하는 여러 데이터뱅크들이 개발되어 있는데, 인간 지놈(human genome) 매핑 정보의 데이터베이스는 최대형 데이터베이스 중의 하나로 알려져 있다.

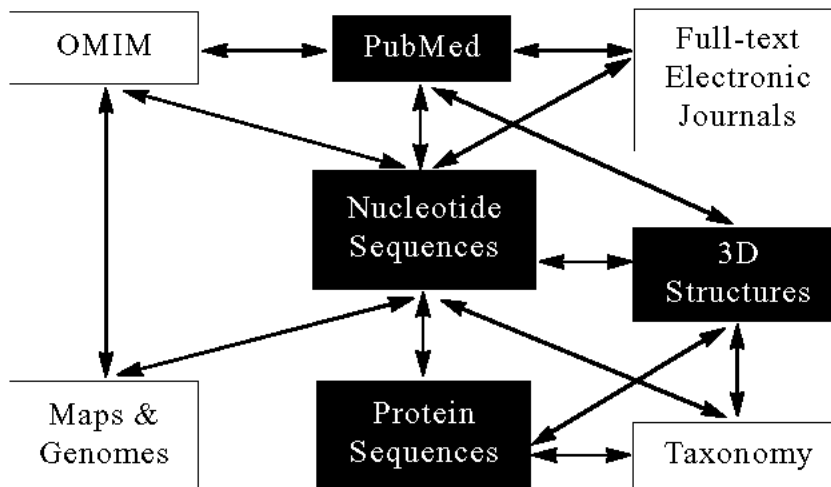
분자생물학 분야의 연구자들은 발견한 시퀀스를 과학저널에 보고하게 되는데, 시퀀스를 포함할 경우 보통 본문의 작성 뿐만 아니라 교정, 인쇄에까지 어려움을 겪게 된다. 따라서 주요한 생의학 관련 출판사들은 논문을 간행할 때 이 시퀀스들의 코드나 데이터뱅크 번호를 반드시 포함할 것을 요구하도록 잠정적으로 동의하였고, 논문이 간행될 수 있기 전에 데이터뱅크에 시퀀스 그 자체를 반드시 등록하게 하였다.

디지털 도서관에서 문헌과 데이터뱅크 사이의 링크는 탐색 프로파일, 본문분석 프로그램, 혹은 키워드 색인 등을 사용하여 본문을 분석하고, 시퀀스 데이터뱅크 번호를 식별한 후 액티브 링크를 삽입함으로써 이루어진다. 액티브 링크는 실제적 시퀀스가 저장되어 있는 데이터뱅크에서 시퀀스 번호를 찾는 탐색전략으로 구성된다. 이용자가 액티브 링크에 클릭할 때, 스크립트가 이용자의 브라우저로부터 생성되고 시작된

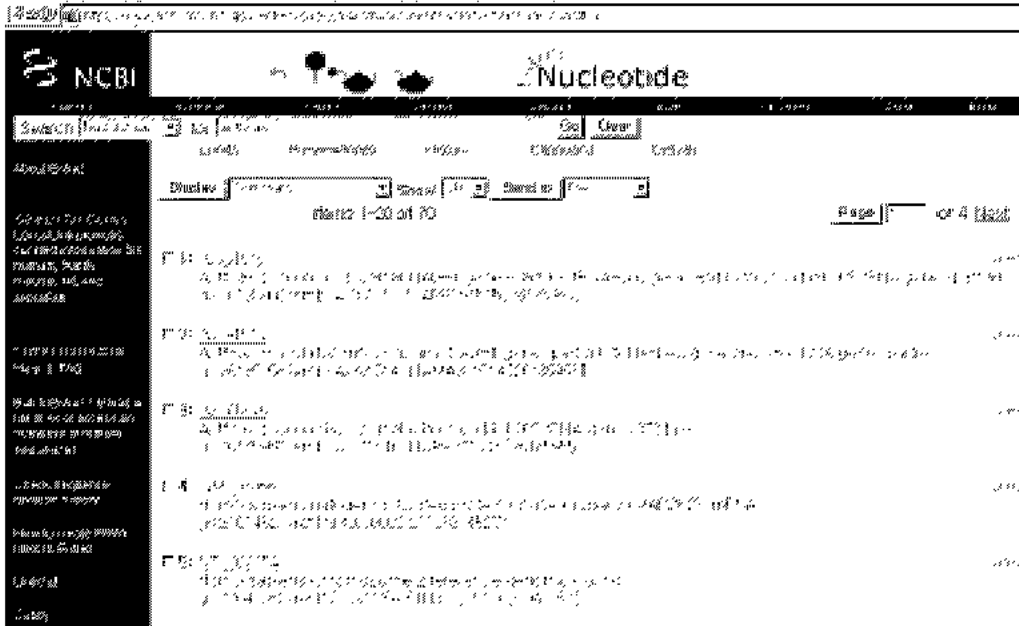
다. 웹을 통해 가동되는 데이터베이스가 탐색되고, 시퀀스 레코드가 이용자에게 보내진다. 데이터뱅크 사이트가 어떤 서비스와 도구를 제공하는가에 따라 이용자는 시퀀스를 분석할 수도 있고 조작을 위하여 시퀀스를 다운로드할 수도 있다(Hodge 2000).

이와 같은 유형의 링크는 미국의학도서관의 탐색서비스 PubMed와 미국국립생명공학정보센터(National Center for Biotechnology Information, NCBI)의 GenBank 사이에 존재한다. GenBank는 NIH가 구축하는 유전 시퀀스 데이터베이스로서 일반에게 공개된 모든 DNA 시퀀스에 대하여 주석을 단 컬렉션으로, 일본의 DNA DataBank of Japan(DDBJ)와 유럽의 European Molecular Biology Laboratory(EMBL)와 함께 국제 뉴클레오타이드 시퀀스 데이터베이스 협력(International Nucleotide Sequence Database Collabora-

tion)'을 구성하고 있다. 2002년 현재 10만 이상의 종으로부터 170억 이상의 뉴클레오타이드 베이스를 수록하고 있고, 거의 매 14개월마다 그 수록된 정보량이 배증하는 성장속도를 보이고 있다(NCBI 2002). 만약에 PubMed를 탐색하여 GenBank 번호를 갖는 레코드를 찾아내면, 이용자는 자동적으로 GenBank로부터 시퀀스 레코드들을 탐색하고 디스플레이할 수 있다. PubMed와 NCBI의 GenBank 사이의 관계는 <그림 1>에서 볼 수 있는 것처럼, GenBank의 검색 시스템인 Entrez를 통하여 뉴클레오타이드 시퀀스, 단백질 시퀀스, 3D 구조, 텍소노미 등 GenBank의 데이터베이스들을 검색하고 바로 PubMed를 거쳐 MEDLINE으로 연결될 수 있다. <그림2>는 GenBank로부터 탄저병(anthrax)의 뉴클레오타이드 시퀀스를 검색한 결과의 요약정보를 보여 주고 있는데, 각 레코드에 들어가면 시퀀스



<그림 1> GenBank와 PubMed의 관계
출처 : <http://www.ncbi.nih.gov/Database/index.html>



<그림 2> GenBank의 뉴클리오타이드 시퀀스 검색: 탄저병(anthrax)

의 정의, 접근 번호, 버전 번호, 키워드, 소스 유기체 정보, 참고문헌(저자, 제목, 저널), 코멘트, 특성 등이 기술되어 있고, 3D 구조로 볼 수도 있다.

3.1.2 개별 산업 코드와 완전 체계 링크
 산업과 경제 분야에서 종종 분류체계는 기업이나 제품에 관한 중요한 정보를 전달하는 역할을 하고, 이 분류 코드들은 일반적으로 정부, 전문기업 혹은 국제표준기구에서 통제된다. 자주 사용되는 분류체계들 중의 하나는 1987년에 최종 갱신된 '미국 표준산업분류(U.S. Standard Industrial Classification, SIC)'로서 경제의 제조업, 농업 및 서비스 부문들을 식별하기 위하여 미국의 정부, 경제학자, 재무시장, 규제 및 조달부서 등이 사용하고 있다. 이 체계는

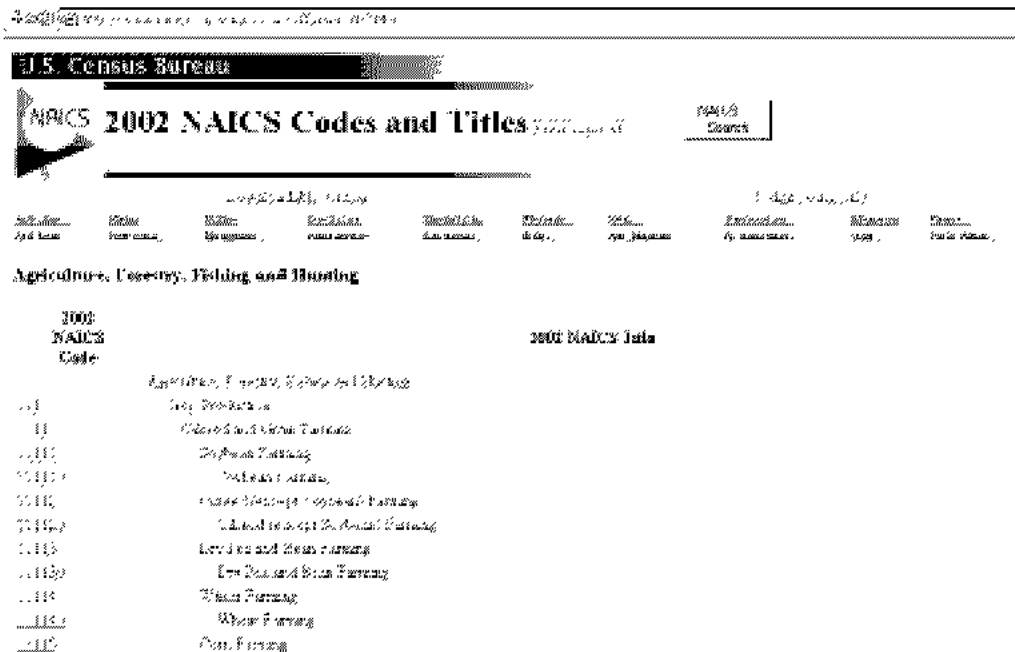
1997년 미국, 캐나다, 멕시코가 함께 개발한 새로운 시스템인 '북미 산업분류 체계(North American Industrial Classification System, NAICS)'로 대체되고 있다(Hodge 2001).

NAICS는 '11 농업, 임업, 수산업 및 수렵업, 21 광산업... 31-33 제조업, 42 도매업... 51 정보...' 등 제반 산업을 모두 스무개 부문으로 구분하고 있으며, 각 산업 아래 세부적인 산업항목을 전개하여 6자리까지 번호를 부여하고 있다. <그림 3>은 '2002 NAICS Codes and Titles'의 '11 농업, 임업, 수산업 및 수렵업'으로부터 가져온 보기이다. <그림 4>는 '51 정보' 부문에서 1997 NAICS 코드를 1987 SIC 코드와 비교하는 표의 보기이다. 각 산업에 대하여 주어진 코드로 들어가면 '514120 Libraries

and Archives 이 산업은 도서관이나 아카이브 서비스를 제공하는데 주로 종사하는 시설들로 구성된다. 이 시설들은 그 이용자들의 정보적, 연구, 교육적 혹은 오락적 요구를 충족시키기 위하여 필요한 대로 문헌(e.g., 도서, 잡지, 신문, 악보 등)의 수집을 관리하고 이 문헌들(그것의 물리적 형태와 특성에 상관없이 모든 기록된 정보)의 이용을 촉진시키는데 종사한다...' 와 같이 특정 산업의 명칭, 내용 및 정의에 대하여 상세히 기술한 내용을 볼 수 있다.

디지털 도서관에서는 이와 같은 코딩 체계를 자원에 링크하여 전자 파일로 사용함으로써 미국 인구조사국에서 유지하고 있는 공식적 정의 테이블을 탐색할 수 있다. 정의 테이블로부터 그 산업부문의 상세

한 내용을 알 수 있고, 계층구조로 제시되는 분류체계로부터는 그 산업과 다른 산업들 간의 관계를 파악할 수 있다. 디지털 본문에 1987 SIC 코드나 1997 NAICS 코드 중 하나만 갖고 있어도' 1997 NAICS and 1987 SIC Correspondence Tables'를 탐색하여 일치되는 코드를 찾아내어 필요한 상관된 정보에 접근할 수 있다. 2002 코드로는 아직 매치할 수 없다. 이 분류체계는 보편적인 이름전자 파일처럼 특정한 산업명을 다양한 형태로 접근할 수 있게 해주지는 않는다. 그러나 특정한 코드로서 대표되는 산업명의 표준화된 형태와 정의를 제공함으로써 적어도 일관성을 유지할 수 있고, 코드 혹은 명칭으로 검색할 수 있도록 지원한다.



<그림 3> NAICS 2002: 농업, 임업, 어업 및 수렵업 산업 코드

표 1. 1997년 10월 1일 기준 SIC(1987)과 NAICS(1997)의 대응

1997년 NAICS		1987년 SIC	
NAICS	1997년 NAICS의 Description	SIC	1987년 SIC의 Description
211	농업, 임업 및 어업(농업 제외)	01	농업
212	농업(농업 제외)	02	축산업
213	농업(농업 제외)	03	어업
22	에너지	13	전기
23	건설	15	건설
24	제조업	20-30	제조업
31	식품 및 음료	20	식품 및 음료
32	화학 및 allied products	28	화학
33	기계 및 장비	35	기계
34	금속	33	금속
35	전자 및 전기	36	전자 및 전기
36	항공 운송	40	항공 운송
37	자동차	37	자동차
38	가정용 기계	38	가정용 기계
39	기타 제조업	25-32	기타 제조업
40	통신	48	통신
41	에너지	49	에너지
42	운수	40-44	운수
43	서비스	50-80	서비스
44	서비스	80	서비스
45	서비스	90	서비스
46	서비스	40-44	서비스
47	서비스	53	서비스
48	서비스	54	서비스
49	서비스	55	서비스
50	서비스	56	서비스
51	서비스	57	서비스
52	서비스	58	서비스
53	서비스	59	서비스
54	서비스	60	서비스
55	서비스	61	서비스
56	서비스	62	서비스
57	서비스	63	서비스
58	서비스	64	서비스
59	서비스	65	서비스
60	서비스	66	서비스
61	서비스	67	서비스
62	서비스	68	서비스
63	서비스	69	서비스
64	서비스	70	서비스
65	서비스	71	서비스
66	서비스	72	서비스
67	서비스	73	서비스
68	서비스	74	서비스
69	서비스	75	서비스
70	서비스	76	서비스
71	서비스	77	서비스
72	서비스	78	서비스
73	서비스	79	서비스
74	서비스	80	서비스
75	서비스	81	서비스
76	서비스	82	서비스
77	서비스	83	서비스
78	서비스	84	서비스
79	서비스	85	서비스
80	서비스	86	서비스
81	서비스	87	서비스
82	서비스	88	서비스
83	서비스	89	서비스
84	서비스	90	서비스
85	서비스	91	서비스
86	서비스	92	서비스
87	서비스	93	서비스
88	서비스	94	서비스
89	서비스	95	서비스
90	서비스	96	서비스
91	서비스	97	서비스
92	서비스	98	서비스
93	서비스	99	서비스

Source: Manufacturing Industry Classification System, Bureau of Economic Analysis, Department of Commerce, 1997, 2001.

<그림 4> NAICS(1997)과 SIC(1987)의 대응 테이블

3.2 기술 레코드로 링크

개인명, 기관명이나 지리명과 같은 객체(entity)의 이름들을 그 객체에 관한 추가적인 정보로 링크하는 것은 하이퍼링크링 가장 먼저 사용한 사례들 중의 하나이다. 사전, 용어집, 분류체계 같은 지식조직 시스템들이 한 자원에 있는 객체를 그 객체에 관하여 보다 풍부한 기술정보를 포함하고 있는 다른 자원으로 링크해 줄 수 있다. 이 같은 링크는 특정 주제에 대하여 잘 알지 못하는 이용자들이나 추가적 정보를 얻음으로써 자기의 업무를 보다 효율적으로 처리할 수 있는 이용자들에게 유용한 것으로 알려져 있다.

기술 레코드로 링크하는 예로는 크게 유기체명을 그 종에 대하여 보다 완전하게 설명해주고 그 유기체의 전체적 분류체계 내에 배치하는 레코드들로 연결하는 것, 화

학을 기술 레코드와 분자구조로 연결하는 것 및 고유명사를 그 사람의 전기로 연결하는 것을 들 수 있다.

3.2.1 유기체명을 분류 레코드로 링크

종-속(Genus-Species)명은 식물, 동물, 미생물과 같은 유기체의 라틴어 학명(scientific names)들이며 보통 일반명(common names)과는 매우 다른 형태이다. 생명체를 연구하고 분류하는 분류학자들(taxonomists)은 각 유기체의 레코드를 작성하고, 이 유기체의 레코드는 일반적으로 계층구조 내에서 다른 유기체들과 상대적으로 링크된다. 유기체명과 그것이 계층구조 내에서 차지하는 위치가 전달하는 정보 외에도 분류 레코드는 그 유기체를 기술하는 분포 형태, 명명과 분류 전거, 유기체 동정(同定: identification) 일자 같은 다른 정보들을 포함한다.

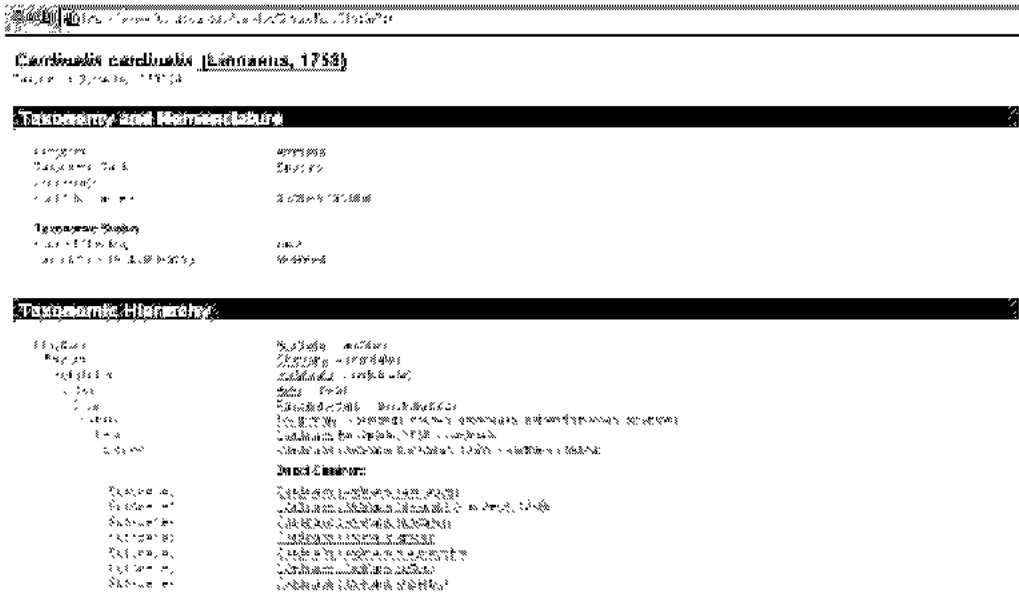
최근 국내의 사례를 보면 지난 1977년 정문기 박사의 <한국어도보(韓國魚圖譜)>에 표본관찰 없이 수록되어 국내에 서식하고 있는 것으로 알려지기는 했으나 표본이 없었기 때문에 그 존재를 입증할 수 없었던 '납작농어(학명 *lateolabrax latus*)'를 잡아, 그 측정기록과 사진을 학회에 제시함으로써 공식적으로 한국에서의 존재를 인정하게 되었다고 하였다(조선일보 2002. 11. 16 보도). 이처럼 표본이 특정한 유기체 기술의 물리적 증거 역할을 하기 때문에, 자연사 박물관, 개인 컬렉션, 개인 과학자들이 컬렉션에서 표본들을 유지하고 그것에 번호를 부여하거나 코드화함으로써 동정된 실체를 증명하는 것은 매우 중요한 과학활동의 일부이다. 때때로 표본들은 디지털화될 수 있는 사진이나 세밀화로서 지원된다.

디지털 도서관에서는 분류전자 파일을 중간 전자 파일로 이용함으로써 유기체의 이름이나 그림을 포함하는 본문이나 이미지 파일을 추가적인 상관된 정보로 링크할 수 있다. 자동적으로 본문을 처리하거나 본문에서 유기체명이나 이미지로부터 분류전자 레코드로 링크를 삽입함으로써 본문이 전달하는 지식을 확장할 수 있다. 본문은 분류 레코드에서 기술 및 역사적 정보를 포함할 수 있고 궁극적으로 사진, 그림이나 적절한 비디오, 오디오 부분으로 링크될 수 있다(Hodge 2000).

분류전자 파일의 대표적인 사례는 미국, 캐나다, 멕시코 정부, 사설단체들 및 분류 전문가들의 파트너십으로 구축한 통합분류 정보 시스템(Integrated Taxonomic Information System: ITIS)이 있다. 이 시스템은

북미주 및 세계의 동식물, 균류와 미생물에 관한 분류정보를 수록한 '신뢰할 수 있는 온라인 리스트'로서, 미국의 많은 정부기관들이 규제와 감시 목적으로 동식물을 일관성있게 명명하기 위하여 사용하고 있다. 또한 'Species 2000'과 'Global Biodiversity Information Facility (GBIF)'와도 파트너십을 맺고 있다. 디지털 도서관에서는 ITIS의 데이터베이스를 반입하여 사용할 수 있는데, FTP로 전체 데이터베이스를 반입할 수도 있고, 필요한 일부만을 반입할 수 있다. 디지털 도서관에서 본문 자료들을 ITIS 레코드로 링크하기 위해서는 모든 유기체명은 본문에서 수작업이나 자동으로 식별하고 ITIS 데이터베이스에 문의로 제출할 수 있다. ITIS는 문의로 제출된 측정된 유기체에 관한 기본적인 정보를 제공하는 레코드를 찾으려면, 이를 문의자에게 파일로서 제공한다(ITIS 2002).

ITIS의 데이터베이스에서 탐색한 분류 레코드의 보기는 <그림 5>에서 볼 수 있다. 이 레코드는 흔히 카디날(Northern Cardinal)로 알려져 있는 새의 분류와 명명 정보와 함께, 분류계층(Taxonomic hierarchy)에서 계(Kingdom) - 문(Phylum) - 아문(Subphylum) - 강(Class) - 목(Order) - 과(Family) - 속(Genus) - 종(Species) - 아종(Subspecies)에 이르는 전 계층에서 차지하는 위치 정보를 제공하고 있다. 또한 관련된 전문가, 표본정보, 서지사항 등의 정보도 포함하여, 고유한 유기체의 동정을 위하여 중요한 정보원 역할을 하고 있다.



<그림 5> ITIS: 카디날(Northern Cardinal)의 분류 레코드

3.2.2 화학명을 분자구조로 링크

화학물질의 고유성은 그 이름보다 분자구조로 식별한다. 그러나 연구문헌, 프로젝트 계획, 목록, 디렉토리 등에서는 일반적으로 화학명이 사용되는데, 각각의 화학물질에 Chemical Abstracts Service(CAS) 혹은 International Union of Pure and Applied Chemistry의 명명법에 따라 이름이 붙여지고 등록번호가 부여된다. 물론 일반적, 상업적으로 사용되는 동의어들도 있다.

화학물질명은 본문의 화학명과 분자구조 사이에 링크를 제공함으로써 식별할 수 있다. 이 링크는 특정한 명명 표준을 사용하는 화학명에 연결된 화학등록번호나 코드와 그 화학물질에 관한 추가적인 정보를 제공하는 전거 레코드 사이에 만들어질 수 있다. 이 전거 레코드는 보통 특정한 화학물질의 동의명과 그것의 화학적, 물리적 속

성에 관한 정보를 포함한다. 전거 파일로부터 화학구조 파일로 링크하는 것은 최근의 과학연구 환경에서 가장 중요한 것들 중의 하나로 여겨지고 있다. 구조 파일은 적절한 소프트웨어와 함께 사용되면서 분자구조를 시각적으로 표현하고, 삼차원 시각화, 순환 및 화학결합의 대체까지도 허용한다.

화학명을 분자구조로 링크하기 위하여 화학등록번호를 사용하는 대표적인 예로는 BIOSIS를 들 수 있다. 생물 및 생의학 데이터베이스를 구축하는 세계 최대 비영리 기관인 BIOSIS는 1993년부터 자동적으로 화학명을 식별하기 위하여 서지 인용(서명과 키워드)을 처리하기 시작하였다. BIOSIS는 이 과정에서 식별된 화학명들에 CAS 등록번호(CAS Registry Numbers :RN)를 부여한다. CAS가 미국에서 주관하는 STN International 온라인 시스템에서 BIOSIS의

이용자는 탐색결과로 나온 레코드들로부터 RN을 추출할 수 있다. 추출한 RN을 사용하여 2천 1백만 이상의 유기물질, 무기물질, 바이오시퀀스, 금속 및 합금 등 화학물질들의 레코드를 포함하고 있는 CAS Registry File을 탐색할 수 있다. 등록 파일의 레코드에는 화학명의 동의명들과 구조 파일 자체에 대한 링크도 포함되어 있다. CAS가 개발한 특별한 도구를 사용하여 구조를 볼 수도 있고 조작할 수도 있다. 또한 연구자가 구조를 조작하고 그렇게 하여 새로운 화학물질을 구상할 수 있도록 모델링 도구로 반입할 수도 있다. 역으로, 이용자는 CAS RNS를 포함하는 데이터베이스로부터 시작하여 RN을 추출하여 BIOSIS 데이터베이스에서 보완적 서지 레코드들의 탐색을 수행할 수 있다(Hodge 2000).

3.2.3 개인명을 전기정보로 링크

전거 파일의 가장 보편적인 유형은 개인명의 변형들을 통제하는 개인명 전거 파일이라고 할 수 있다. 예를 들어 미국 국회도서관의 이름전거 파일(Library of Congress Name Authority File : LCNAF)은 다양한 저자, 편자, 예술가 및 기타 개인명들을 제어하기 위하여 사용되고, 게티 보캐뷰러리 프로그램(Getty Vocabulary Program)이 개발한 '예술가 종합인명록(Union List of Artist Names : ULAN)'도 예술 분야에서 널리 사용된다.

이름전거는 편목자와 색인자가 목록에서 특정한 이름의 적절한 표준화된 형태를 채택하고 채택되지 않은 변형들과 이명들을 참조로 링크하여 그 사람에 의한 혹은

그 사람에 관한 모든 저작들이 함께 모일 수 있도록 보장하기 위하여 사용하는 도구이지만, 개인의 이름을 포함하는 서지 레코드나 문헌을 다양한 다른 관련 자료들로 링크하기 위하여 사용될 수 있다. 만약에 디지털 도서관의 자원이 이름의 표준화된 형태를 갖고 있으면, 그것의 변형들을 찾기 위하여 전거 파일을 탐색할 수 있다. 표준화된 형태와 변형 혹은 이명들은 관련된 정보를 제공할 수 있는 다양한 다른 자원을 탐색하는 과정에서 결합될 수 있다. 따라서 이름전거 파일에 특정한 인물의 다양한 이름들과 이름의 형태들이 포함되어 있으면, 서로 다른 두 개 이상의 디지털 자원에서 그 인물의 이름이 반드시 동일한 형태로 나타나지 않더라도, 자동적으로 링크해 줄 수 있다.

예를 들어 예술가의 작품 이미지나 전기, 비평본문의 디지털 도서관에서 ULAN이나 LCNAF와 같은 이름전거 파일은 추가적인 정보를 제공하기 위한 중간 파일의 역할을 한다. 디지털 도서관 컬렉션에서 나타나는 이름으로 전거 파일을 탐색하고, 해당되는 레코드가 발견되면 이용자들에게 광범한 본문 자료를 제공하면서 예술가에 관한 정보를 제시할 수 있다. 또한 예술가에 관한 중요한 전기적 혹은 비평적 저작들에 대한 인용정보를 제공하고, 그 정보원들로 직접 링크해 줄 수도 있다(Hodge 2001).

디지털 도서관에서 전기정보를 위하여 링크될 수 있는 대표적인 자원은 Gale사의 Biography Resource Center를 들 수 있다. Biography Resource Center는 Gale사에서

가장 빈번하게 사용되는 80여 종의 전기 데이터베이스들을 통합한 데이터베이스 자원으로서 250여 종의 전문(full-text) 정기간행물의 기사와 기타 주요한 전기정보원으로부터 모두 1백 25만건의 전기를 수록하고 있다. 주요한 정보원으로는 Contemporary authors, Contemporary musicians, Complete Marquis Who's Who, Dictionary of American biography 등을 포함하고 있으며, 서술적 전기, 간략한 썸네일 전기, 잡지의 전문 기사 등을 제공하는 한편 개인의 일생과 작품 웹 사이트로 링크할 수도 있다고 한다(Gale 2002). 디지털 도서관에서는 이와 같은 정보원으로부터 목록으로까지 역순으로 찾아가는 것도 허용하므로 그 중간 가교로서 전거 파일의 역할이 주목된다고 할 수 있다.

(1) ULAN

ULAN은 게티 보캐뷰러리 프로그램이 1980년대 중반부터 구축하기 시작한 것으로 22만여 예술가와 건축가의 이름과 전기 및 서지정보를 수록하고 있는 데이터베이스이다. 게티 프로그램은 ULAN과 함께 순수예술, 건축, 장식예술 및 물질문화에 관련된 정보를 기술하고 검색하기 위한 통제언어인 'Art & Architecture Thesaurus (AAT)'와 현대 세계의 모든 국가를 대표하는 90만여 지리명 레코드를 수록하고 있는 'Getty Thesaurus of Geographic Names(TGN)'을 함께 구축하였다. TGN은 장소의 현지명, 역사적 명칭, 지역 유형 및 기타 관련정보까지를 포함하고 있는 지리명의 목록이다. 게티 프로그램은 이 '통제

언어'는 '중요한 정보에 대한 접근을 촉진하기 위하여 용어와 개념의 관계를 명시하는 방식으로 조직된 용어들의 집합'이라고 정의하고 있는데, 세계의 기관들이 이 데이터를 무료로 반입하여 사용할 수 있도록 허용함으로써 예술 관련 주제를 서술하고 편목하기 위한 표준으로 널리 사용되고 있다.

ULAN에는 매우 다양한 이름의 변형들, 가명들 및 언어 변형들이 포함되어 있는데, 데이터베이스에 수록된 정보는 *Avery index to architectural periodicals, Biography of the history of art, Census of Antique Art and Architecture Known to the Renaissance, Witt checklist of painters c1200-1976*과 같은 참고도구, Canadian Centre for Architecture, Getty Conservation Institute 등의 다양한 기관과 프로젝트들로부터 수집되었다. <그림 6>은 'User's guide to the ULAN data releases'에서 보기로 제시된 영국의 화가이자 사진작가인 헨리 피치 로빈슨(Henry Peach Robinson)의 전거 레코드의 일부이다. 레코드는 기본 표목으로 정해진 그의 이름과 함께 Henry Peach Robinson, Robinson, H. P. (Henry Peach), Robinson, H.P., Robinson, Henry 등과 같이 다양한 이름의 형태들(VAR), 간략한 전기 사항 및 전기정보원들에 대하여 수록하고 있다. 또한 기술적 주기에서 그의 일생에 대한 간략한 연대기, 관련된 국가 및 활동지역, 생몰년 등의 사항을 기록해 주고 있다. ULAN 레코드의 기본적인 요소들은 이름, 역할, 생몰년 및 활동기간, 관련된 인물, 지역, 주기, 서

지 등으로 요약할 수 있으며, MARC 레코드로도 다운로드할 수 있다.

(2) LCNAF

LCNAF는 대표적인 이름전거 파일로서 미국 국내뿐만이 아니라 국제적인 도서관 커뮤니티에서도 그 중요성은 재론할 필요가 없다. 미국국회도서관은 전거 파일을 2002년 7월 1일부터 LC의 웹사이트에서 일반에게 공개하여 무료로 접근할 수 있도록 허용하는 동시에 서지목록으로 바로 링크될 수 있도록 하였다. 2002년 11월 현재 사용할 수 있는 전거 레코드의 규모를 보면 서명 파일 (Title Authority Headings File)에 35만개의 시리즈명과 통일서명이 포함되어 있고, 이름 파일(Name Authority Headings File)에는 모두 530만개의 이름표목이 포함되어 있다. 그 가운데 개인명은

380만개, 단체명이 90만개, 회의명이 12만개, 지리명이 9만개로 구성되어 있다. 주제명 파일 (Subject Authority Headings File)은 모두 26,500 개의 주제명, 이름/서명 파일 (Name/Title Authority Headings File)은 34만개의 이름/서명 표목들을 포함하고 있다(윤정옥 2002; LC 2002).

게티 프로그램의 ULAN이 예술 및 유관 분야의 주제 전문 인물명을 중심으로 구축된 것에 비하여, LCNAF는 매우 광범위한 개인명 및 기관, 단체명, 회의명의 전거 레코드를 제공한다. <그림 7>은 LCNAF로부터 ULAN과 동일한 인물의 전거 레코드(n55049097)를 탐색한 결과로서 그 채택표목의 형태는 ULAN의 것과 다르다. 그러나, ULAN의 레코드를 살펴보면 참조한 정보원으로서 LCNAF를 가리키고 있어 상호 연관되어 있음을 알 수 있다.

EXAMPLE

LEN	3282
STATUS	0
ULANIDNO	13997
PATENT	19990730
NAME	Robinson, Henry Peach--8A/p,CC/p,30/p
VAR	Henry Peach Robinson--CC/v Robinson, H. P. (Henry Peach)--CC/v Robinson, H. P.--CC/v Robinson, Henry--CC/v
BIOG	British painter, photographer, 1815-1901--8A/p British, 1815-1901--30/p English photographer and painter; born Ludlow (Shropshire, England, United Kingdom), 1815; died Tunbridge Wells (Kent, England, United Kingdom), 1901--CC/p
SOURCE	*Auer, Encyc. photographica Auer, Encyclopædie der Photographica (1885) GB *Gernsheim, Hist. Photog. Nacker, Margaret P. Henry Peach Robinson: Master of Photographic Art, 1815-1901, Oxford: Basil Blackwell, 1988 Haworth-Borch, Mark, The Golden Age of British Photography, 1839-1909, New York: Aperture, 1984., 95
	LCNAF

<그림 6> ULAN: Henry Peach Robinson의 전거 레코드

출처 : <http://www.getty.edu/research/tools/vocabulary/ulan/ULAN-Users-Manual.pdf>



LC Control Number: 503430-
READING: Peabody, C. C. (New York: 1841-191)
 000... [unreadable]
 010... [unreadable]
 020... [unreadable]
 030... [unreadable]
 040... [unreadable]
 050... [unreadable]
 060... [unreadable]
 070... [unreadable]
 080... [unreadable]
 090... [unreadable]
 100... [unreadable]
 110... [unreadable]
 120... [unreadable]
 130... [unreadable]
 140... [unreadable]
 150... [unreadable]
 160... [unreadable]
 170... [unreadable]
 180... [unreadable]
 190... [unreadable]
 200... [unreadable]
 210... [unreadable]
 220... [unreadable]
 230... [unreadable]
 240... [unreadable]
 250... [unreadable]
 260... [unreadable]
 270... [unreadable]
 280... [unreadable]
 290... [unreadable]
 300... [unreadable]
 310... [unreadable]
 320... [unreadable]
 330... [unreadable]
 340... [unreadable]
 350... [unreadable]
 360... [unreadable]
 370... [unreadable]
 380... [unreadable]
 390... [unreadable]
 400... [unreadable]
 410... [unreadable]
 420... [unreadable]
 430... [unreadable]
 440... [unreadable]
 450... [unreadable]
 460... [unreadable]
 470... [unreadable]
 480... [unreadable]
 490... [unreadable]
 500... [unreadable]
 510... [unreadable]
 520... [unreadable]
 530... [unreadable]
 540... [unreadable]
 550... [unreadable]
 560... [unreadable]
 570... [unreadable]
 580... [unreadable]
 590... [unreadable]
 600... [unreadable]
 610... [unreadable]
 620... [unreadable]
 630... [unreadable]
 640... [unreadable]
 650... [unreadable]
 660... [unreadable]
 670... [unreadable]
 680... [unreadable]
 690... [unreadable]
 700... [unreadable]
 710... [unreadable]
 720... [unreadable]
 730... [unreadable]
 740... [unreadable]
 750... [unreadable]
 760... [unreadable]
 770... [unreadable]
 780... [unreadable]
 790... [unreadable]
 800... [unreadable]
 810... [unreadable]
 820... [unreadable]
 830... [unreadable]
 840... [unreadable]
 850... [unreadable]
 860... [unreadable]
 870... [unreadable]
 880... [unreadable]
 890... [unreadable]
 900... [unreadable]
 910... [unreadable]
 920... [unreadable]
 930... [unreadable]
 940... [unreadable]
 950... [unreadable]
 960... [unreadable]
 970... [unreadable]
 980... [unreadable]
 990... [unreadable]

<그림 7> LCNAF: Henry Peach Robinson의 선거 레코드

3.3 객체명과 물리적 표본 링크

디지털 도서관 자원에서 객체명을 물리적 표본(physical specimen)으로 연결시키는 것이 가능하다. 물리적 표본이나 문화유물의 보존을 중요시하는 여러 분야에서 전시 카탈로그는 특별 전시 예술품들에 대하여 기술하고, 박물관 카탈로그는 특정한 박물관이 소장하고 있는 예술, 자연사 혹은 문화적 유물 등의 소장 현황을 알려준다. 이 카탈로그들이 디지털화되면서 박물관, 아카이브나 기타 컬렉션의 유물들에 관한 기술적 레코드를 제공하는 기존의 기능에 더하여, 실물의 소재를 지시하고, 직접 링크를 제공하는 역할까지도 하게 된다.

예를 들어 앞에서 기술한 것처럼 생물

학에서 새로운 유기체를 발견하고 기술하거나, 이미 알려진 유기체를 재분류할 경우에 그 기술과 명명의 확증을 위하여 대표 표본(type specimen)이 필요하다. 이 표본들은 보통 자연사 컬렉션에 포함되어 유지되고 다양한 분류학회들은 반드시 표본을 기탁할 것을 요구하는데, 컬렉션에서 표본들을 보존하기 위해서는 각각 동정 코드를 부여한다. 동정 코드의 일차적인 기능이 물리적인 컬렉션을 조직하는 것인 한편, 자연사 학계에서 많은 프로젝트들이 표본의 사진을 디지털화하고 그 식별자들을 포함하는 표본 레코드를 생성하고, 그것들을 쉽게 접근할 수 있게 만들기 위하여 진행되고 있다. 디지털화의 정도는 전문 분야마다 상이하지만, 식물학과 같은 분야에서는 거의

모든 중요한 연구 식물원들이 종이로 된 레코드를 유지하는 대신 자기들의 대표 표본 컬렉션을 디지털로 목록화하고, 많은 곳은 또한 웹 상에서 그 대표 표본들의 디지털 사진을 접근할 수 있게 하고 있다.

디지털 도서관 컬렉션이 동정 코드들을 포함하는 자원들을 갖고 있으면, 이 코드들은 웹 기반 목록이나 데이터베이스에 대하여 추출되고 일치된다. 이 링크는 이용자에게 소재와 접근 정보를 제공하여 디지털 도서관 자원에 언급된 물리적 대상에 접근할 수 있도록 해준다. 예술, 고고학 및 문화사적 컬렉션의 큐레이터나 등록자들도 소장품들에 대하여 인벤토리나 접근 번호를 부여한다. 자연사에서 기술된 것과 유사한 링크가 예술작품에 관련된 본문과 특정한 컬렉션에 속한 물리적 작품 사이에 만들어질 수 있다. 예술작품에 관한 논문은 본문에서 동정번호를 그 번호와 그 작품에 관한 추가적 정보를 포함하는 온라인 목록으로 연결함으로써 물리적 표본에 관한 추가적인 정보로 링크될 수 있다(Hodge 2001).

대표적인 예로서 미국의 스미소니안 국립 자연사 박물관(Smithsonian's National Museum of Natural History)이 주관하고 있는 '식물 대표 표본 등록부(Botanical Type Specimen Register)'를 보면 국립 식물표본실(U.S. National Herbarium)에서 93,000종 이상의 표본을 목록화하였다. 또한 '미국 조류 표본 목록(USNM Bird Type Catalog)'을 보면 지난 150여년 동안 조류 문헌에서 공식적인 이름을 갖도록 지명된 3,549종의 대표표본을 수록하고 있는

데 이 표본들의 상당부분은 1961년에 H.G. Deignan이 작성한 <Type specimens of birds in the United States National Museum>에 바탕을 두고 있다고 한다. 이 표본 목록은 카탈로그 번호, 조류의 과, 속, 종 및 아종으로 탐색할 수 있을 뿐 아니라 수집 연도, 수집국가, 수집 주/지역 및 수집자명으로 탐색할 수 있다. 그러나 이 표본 목록은 현재 조류의 다양한 일반명으로는 검색을 허용하지 않는다. 이 논문의 보기로 채택한 북미주에 많이 살고 있는 주홍색 새인 카디날에 대하여 탐색을 시도하였을 때, 결과의 검색에 실패하였고 라틴명인 cardinal의 과, 속, 종, 아종의 어느 것에 해당하는 이름인지 알 수 없어 각 필드를 시도해 본 끝에 속명(종명도 cardinalis임)으로 검색에 성공할 수 있었다. <그림 8>은 이 표본 목록으로부터 검색한 레코드로서 탐색의 키가 되는 필드의 해당정보 외에도 상세한 수집지역의 위도, 경도, 필드 번호, 기증자, 표본의 성별, 성체 여부, 표본 상태와 유형, 출판된 표본명, 저자와 서지 인용 등 다양한 정보가 포함되어 있다. 그러나 그 자체로는 일반명 혹은 여러 다양한 형태의 이명들을 사용한 탐색을 허용하지 않으므로, 이와 같은 표본 목록은 <그림 5>에 예시한 ITIS의 데이터베이스로부터 탐색한 카디날의 분류 및 명명 전자정보와 링크되면 이용자들이 훨씬 편리하게 검색될 수 있다. ITIS는 학명과 분류 일련 번호(Taxonomic Serial Number)뿐만 아니라 일반명으로도 탐색을 허용하며 카디날이라는 새가 분류 계층에서 어느 위치에 속하는지에 관한 정보를 매우 상세히 그리고 계층

www.kci.go.kr



Bird Types Query Result

Home | About | Search | Help

→ Index Number.....	12345
→ Family.....	CARDINALID
→ Genus.....	CARDINALIS
→ Species.....	CARDINALIS
→ Subspecies.....	CARDINALIS
→ Date Collected.....	2002-10-20
→ County.....	CHATEAUQUE
→ State/Province.....	TEXAS
→ Country/Occurrence.....	AMERICAN
→ Specific Locality.....	FAULKNER CREEK
→ Detailed Locality.....	2002-10-20, 2002-10-20, STATE PARK, 2002-10-20
→ Collector(s).....	SMITH, J.
→ Field Number.....	1234
→ Date.....	2002-10-20, 2002-10-20
→ Sex.....	M
→ Age.....	ADULT
→ Measurements.....	2002-10-20
→ Photo Status.....	None
→ Published Field Notes.....	http://www.birds.org/fieldnotes/
→ Author.....	J. Smith
→ Citation.....	Smith, J. (2002). <i>CARDINALIS</i> . In <i>Birds of Texas</i> . Vol. 1.

<그림 8> 미국조류표본목록: 카디날(Northern Cardinal)의 표본 레코드

의 순차적으로 열거하고 있다. 또한 전문가, 기타 참고정보 및 관련 출판물의 서지사항, 서식지역의 지리정보도 포함하고 있다.

4. 맺음말

이 논문에서는 디지털 도서관에서 다양한 지식조직 시스템들이 목록과 디지털 자원을 링크하는 중간 파일, 즉 전거 파일의 역할을 하는 사례들을 소개하였다. 전통적으로 이름, 서명, 주제 전거 파일 등이 도서관에서 주로 목록에서 특정한 인명에 의한, 그리고 이들에 관한 자료들을 찾고 모을 수 있도록, 또한 서명과 주제명에 관한

자료들을 모을 수 있도록 지원하는 전거통제를 위하여 사용되어 왔다. 그러나 도서관, 박물관, 미술관, 아카이브 등에서 정보자원의 디지털화가 확대되고, 이들을 웹 상에서 이용할 수 있게 됨에 따라 이용자가 관심 대상인 객체에 대한 완전한 지식 없이도 필요한 전문, 이미지, 참고도구 등에 접근할 수 있도록 이용자의 정보요구와 디지털 자원 사이의 다리로서 전거 파일의 역할을 수행하며 '찾기'와 '모으기'를 지원할 필요성이 점차 증대되었다.

지식조직 시스템은 용어 리스트, 분류와 범주화, 관계 리스트 등으로 구분될 수 있고, 디지털 도서관에서 조직과 접근을 지원하는 도구의 기능을 한다. 이 논문에서는 목록에서 디지털 자원으로, 또한 디지털 자

원에서 목록으로 링크되면서 전통적인 개념의 전거 파일보다 확대된 다양한 형태의 지식조직 시스템을 활용할 수 있다는 점에 주목하였다. 지식조직 시스템들을 코드에서 전문으로, 객체에 대한 보다 기술적 레코드로, 또한 상관된 물리적 객체에 대한 소재 및 실물정보로 접근할 수 있도록 링크를 통하여 활용하는 사례들을 중심으로 살펴 보았다.

이 논문에서 살펴 본 내용은 하지가 Digital Library Federation에 제출하였던 보고서를 중심으로 하였으므로 구미 지역의 디지털 도서관에서 디지털 자원을 조직하고 접근하기 위하여 사용하는 시스템들에 중점을 두었다. 국내에서는 가장 보편적이고 기본적인 전거통제의 도구로서 국가적인 이름 전거 파일, 주제표목표 등이 구축되지 않고 있는 것을 비롯하여, 여기에서 살펴 본 다양한 디지털 자원들의 한국 버전이라고 할 만한 정보원과 시스템이 아직 미비함을 알 수 있다.' 초고속 인터넷 통신망의 보급률 세계 최고'와 같은 가시적인 기록이 있다고 하여도, 그 통신망을 이용하여 국내의 이용자들이 관심을 갖고 있는 각종 서지 객체에 대하여, 그것이 개인, 단체, 개념, 사물, 사건, 장소이든 작품이든, 또한 이용자들이 초중등학생이든 전문 학술 연구자이든에 상관없이 적절한 수준의 정보원에 접근할 수 있도록 지원할 수 없다면 활용가치는 매우 낮다고 할 수 있다. 디지털 형태의 정보자원을 수집하고 생산하면서, 목록과 자원 사이에 만들어진 링크는 궁극적으로는 최종 이용자에게로 연결되는 것이라는 인식 아래에 다양한 지식조

직 시스템을 활용할 수 있는 기반이 마련되어야 할 것이다.

참고문헌

- 사공철, 김태수, 정영미, 최석두 편. 2001. 『정보학사전』. 서울 : 문헌정보처리 연구회.
- 윤정옥. 2002. 한국과학기술정보연구원의 회의명·단체명 전거 파일 구축 연구. 『정보관리연구』 33(3) : 69-86.
- Gale. 2000. *Biography Resource Center*. [online] [cited 2002.11.11] <<http://www.gale.com>>
- Getty Vocabulary Program. 2000a. *User's guide to the ULAN data releases*. Release version 3.0. Los Angeles, California: The J. Paul Getty Trust. [online] [cited 2002.11.11] <<http://www.getty.edu/research/tools/vocabulary>>
- Getty Vocabulary Program. 2000b. *User's guide to the TGN data releases*. Release version 2.0. Los Angeles, California: The J. Paul Getty Trust. [online] [cited 2002.11.11] <<http://www.getty.edu/research/tools/vocabulary>>
- Hodge, Gail. 2000. *System of knowledge organization for digital libraries : beyond traditional authority files*. Washington, D.C. : Digital Library Federation, Council on Library and

- Information Resources.
- Integrated Taxonomic Information System* (ITIS). [online] [cited 2002.11.11] <<http://www.itis.usda.gov>>
- Library of Congress (LC). *Library of Congress Authorities*. [online] [cited 2002.11.11] <<http://authorities.loc.gov>>
- National Center for Biotechnology Information*(NCBI). [online] [cited 2002.11.11] <<http://www.ncbi.nlm.nih.gov/>>
- Smithsonian National Museum of Natural History*. Department of Systematic Biology, Vertebrate Zoology, Division of Birds. USNM bird type catalog. [online] [cited 2002.11.11] <<http://www.nmnh.si.edu/vert/birds/brdtype.s.html>>
- Tillett, Barbara B. 2001. Authority control on the web. *Proceedings of the Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web*, Washington, D.C., November 15-17, 2000, sponsored by the Library of Congress Cataloging Directorate, edited by Ann M. Sandberg-Fox. Washington, D.C. : Cataloging Distribution Service, Library of Congress, 207-220.
- UCSB. *Alexandria Digital Library Gazetteer*[online] [cited 2002.11.11] <<http://fat-albert.alexandria.ucsb.edu:8827/gazetteer>>
- U.S. Census Bureau. *2002 NAICS Codes and Titles*. [online] [cited 2002.11.11] <<http://www.census.gov/epcd/naics02/naics02.htm>>
- U.S. Census Bureau. *U.S. Gazetteer: 2000 and 1990* [online] [cited 2002.11.11] <<http://www.census.gov/geo/www/gazetteer/gazette.html>>
- U.S. Census Bureau. *1997 NAICS and 1987 SIC Correspondence Tables*. [online] [cited 2002.11.11] <<http://www.census.gov/epcd/www/naicstab.htm>>
- Warner, James W. and Brown, Elizabeth W. 2001. *Automated name authority control*. JCDL' 01, June 24-28, 2001, Roanoke, Virginia, U.S.A.