

자원공유를 위한 온톨로지 기반 컬렉션 단위 기술 모형개발 연구*

A Study on the Model of Collection-Level Description based on Ontology for Resources Sharing

이혜원(Hyewon Lee)**

초 록

본 연구는 빠르게 발전되는 네트워크 환경을 고려한 분산된 자원 활용에 필요한 기반 연구이다. 앞으로의 자원공유에 대한 노력은 기술적인 문제가 아닌 의미적인 상호운용성이 강조되어야 한다. 이를 위해 본 연구에서는 자원공유를 위한 컬렉션 단위 기술(collection level description; 이하 CLD라고 함) 모형을 제안하였다. 먼저 컬렉션의 범위, 구축목적, 관련 행위자들을 기반으로 하여 컬렉션에 대한 정의를 설명하였으며, CLD의 중요성을 강조한 연구와 그와 관련된 표준들을 비교·분석하였다. 마지막으로 기존의 CLD의 기능에서 보강되어야 할 관계 중심의 모형 구축을 시도하였으며, 이를 위해 온톨로지 개념을 사용하였다. 온톨로지 기반 CLD 모형은 새로운 관계를 창출하기 위한 다양한 표현 기술들을 정의하였다. 기존 연구와는 달리 클래스와 속성을 구별하고, 속성 중에서도 클래스 자체에 대한 특성과 클래스들 간의 관계를 연결하는 속성으로 나누어서 제시하였다.

ABSTRACT

This study is based on the practical use for distributed resources considering growing network rapidly. The focal point of this study will be argued on semantic interoperability for sharing of resources, not be emphasized the technical issues of network. The aim of this article is developing the model of Collection-Level Description(CLD) for sharing of resources. The present article consists of a definition of collection in relation to the scope, objectives, and agents of the collection and an analysis of researches about CLD strengths and standards. Lastly, it was intended to construct the model focused on relation which was needed to be strengthened the existing CLD's function, thus, this study attempted to use the concept of ontology. The model of CLD based on ontology suggested the description could represent new relations inferred between classes and properties. Distinguishing class and property, furthermore, this study suggested properties were separated the characteristic of class and the relation with classes.

키워드: 컬렉션 단위 기술, 개별자료 단위 기술, 온톨로지, 자원공유

RSLP collection description, DC collection description, natural collection description
collection level description, item level description, ontology, resources sharing

* 이 논문은 제15회 한국정보관리학회 학술대회(2008년 8월 21일, 상명대학교)에서 발표한 원고를 수정·보완한 것임.

** 서울여자대학교 사회과학대학 문헌정보학과 시간강사(hwlee@swu.ac.kr)

■ 논문접수일자: 2008년 8월 20일 ■ 최초심사일자: 2008년 8월 22일 ■ 게재확정일자: 2008년 9월 4일
■ 情報管理學會誌, 25(3): 209-230, 2008. [DOI:10.3743/KOSIM.2008.25.3.209]

1. 서론

Dempsey의 연구(2006)에서는 도서관 목록의 새로운 변화를 촉구하였으며, 특히 네트워크 환경을 고려하여 도서관간의 자원공유를 강조하였다. Dempsey(2006)는 도서관의 다양한 물리적인 서비스와 네트워크 공간에서의 서비스를 향상시키기 위해 동일한 노력이 필요하다고 제시하였지만, 빠른 발전을 이룬 네트워크 환경을 고려한다면 다양한 수준의 자원들이 온라인으로 통해 쉽게 제공될 수 있는 효율성에 대해 더 강조하였다.

빠르게 발전되는 네트워크 환경을 고려한다면 분산된 자원의 활용에 대한 고민은 당연한 일일 것이다. 이러한 고민들로 인해 도서관 및 정보를 다루는 모든 기관에서는 자관에 보관된 자원뿐만 아니라 컬렉션 수집 정책에 의하여 다른 기관의 자원을 재사용(reuse)하는 모형들을 고려하였다. 특히 Xin의 연구(2004)에서는 중국의 빠른 네트워크 환경의 발전으로 인한 국가단위 디지털 도서관들 간의 자원 활용 연구에 대한 필요성을 강조하였다.

기존에 제기된 자원공유를 위한 상호운용성은 정보를 다루는 기관을 중심으로 고민되는 사안이었다. 그렇지만 상용자원 제공자나 웹 포털사이트와 같은 다양한 정보 생산 및 수집 기관이 생겨나고, 간단한 웹 어플리케이션을 통해 웹 콘텐츠를 생산할 수 있는 이 시점에서는 특정 기관이나 개인을 중심으로 한 자원공유의 의미가 없어지고 있다.

자원공유는 기존 고려 대상이었던 기술적인 문제에서 벗어나, 의미적인 상호운용성이 강조되어야 한다. 의미적인 상호운용성은 개체에

대한 표현뿐만 아니라 표현들을 담고 있는 용기까지도 의미적으로 연결되어야 한다. 개체에 대한 표현은 개체가 나타내고자 하는 개념이며, 표현을 담고 있는 용기는 자관시스템에서 사용되는 메타데이터의 형식이라고 할 수 있다. 자관시스템에서 사용하는 메타데이터 형식은 개별자료를 위한 기술이지만 하나의 컬렉션 단위로 제시된다면 좀 더 간단하게 제시될 수 있다. 컬렉션은 한 기관에서 특정 목적으로 제작한 자원들의 집합체이며, 그 집합체를 구성하는 자원은 하나의 개별자료 단위 기술(item level description; 이하 ILD라고 함)로 설명된다. 집합체를 표현하는 기술은 해당 기관에서 특정한 목적으로 생산되거나 수집된 자원들의 '생산자가 누구인지', '관리자가 누구인지', '대주체-컬렉션의 주제-가 무엇인지' 등과 같은 가장 근본적인 요소들을 포함하고 있어야 하며, 이러한 요소들은 ILD에서는 간과되기 쉽다.

그러므로 본 연구에서는 자원공유를 위한 컬렉션 단위 기술(collection level description; 이하 CLD라고 함) 모형을 제안하였다. 먼저 컬렉션의 범위, 구축목적, 관련 행위자 등을 기반으로 하여 컬렉션에 대한 정의를 설명하였으며, CLD의 중요성을 강조한 연구와 그와 관련된 표준들을 비교·분석하였다. 마지막으로 기존의 CLD의 기능에서 보강되어야 할 관계 중심의 모형 구축을 시도하였으며, 이를 위해 온톨로지 개념을 사용하였다. 온톨로지기반 CLD 모형은 새로운 관계를 창출하기 위한 다양한 표현 기술들을 정의하였다. 기존 연구와는 달리 클래스와 속성을 구별하고, 속성 중에서도 클래스 자체에 대한 특성과 클래스들 간의 관계를 연결하는 속성으로 나누어서 제시하였다.

2. 선행연구

본 연구에서 제시한 선행연구들은 크게 두 가지의 관점으로 나누어질 수 있다. 첫 번째 관점은 CLD의 중요성을 제시하는 것이다. Lourdi & Papatheodorou(2006), Van Veen & Clayphan(2002) 등의 연구는 다양한 정보환경에서 이질의 메타데이터 스키마를 통합한 상호운용성을 강조하기 위해 CLD를 활용하였으며, Brenner, Larsen & Weston(2006)은 검색실험을 통해 CLD의 강점을 부각시켰다.

두 번째 관점과 연계되는 연구들은 온톨로지를 기반으로 한 CLD 연구이다. Kemp, Tan & Whalley(2007)과 Qin & Finneran(2002)의 연구가 대표적이며 의미적인 접근을 시도하여 상황정보(contextual information)와 주제 분석을 가능하게 하였다.

Van Veen & Clayphan(2002)은 The European Library Project(이하 TEL이라고 함)을 통해 CLD의 중요성을 부각시켰다. 특히 이들의 연구에서는 CLD 표준으로 DC를 제시하였으며, OAI 프로토콜을 사용하여 자원을 수집하고, Z39.50 탐색을 지원하는 시스템과 이용자 서비스와의 조화를 강조하였다. 또한 이들은 유럽지역의 주요 국가도서관의 자원이나 납본된 장서들의 공유를 위해 이용자 중심의 자원 수집을 강조하였다. 이를 위해 그들은 앞으로 TEL의 정책에서 CLD를 강조해야 한다고 주장하였다. CLD의 기능은 분산된 컬렉션에 대한 기술들을 수집하고, 이러한 기술들을 이용자가 입력한 질의와 매칭시킨 후, 적합하다고 판정된 컬렉션을 확인하고 연결시키는 것이다.

Lourdi & Papatheodorou(2006)은 민속학

분야 컬렉션 기술에 대한 메타데이터 정책을 제시하였으며, CLD에서 고려해야 사항으로 자원(material), 기술 정책(description policy), 구조(structure), 접근(access), 보존(preservation) 등을 언급하였다. 특히 문화적인 개체들은 다양한 컬렉션에 분산되어 존재하기 때문에, 이용자의 요구에 맞도록 관련 있는 주제나 시간, 공간 등을 중심으로 개체들을 수집하는 것이 필요하다고 제안하였다. 또한 정보를 다루는 대표적인 기관인 도서관, 박물관, 기록관과의 통합이 필요하며, 이를 위해 다양한 정보 기관에서 사용되는 이질의 메타데이터들을 분석하여 매핑할 수 있는 기반을 준비해야 한다고 제시하였다. Lourdi & Papatheodorou(2006)은 메타데이터 상호운용성을 위한 네 개의 메커니즘을 제시하였다. 첫 번째 메커니즘은 매핑테이블(crosswalk)이며, 이것은 이질의 메타데이터 요소들을 해석할 수 있는 참고자료이다. 도서관, 박물관, 기록관의 다양한 메타데이터 셋들의 상호운용성을 고려하여 요소들 간의 매핑, 계층적인 구조 연결, 개체에 대한 분석, 메타데이터 콘텐츠 변화 등이 가능하도록 설계되어야 한다. 두 번째 메커니즘은 어플리케이션 프로파일(application profile)이며, 이는 다양한 커뮤니티들의 특성을 반영한 이질의 정보환경을 고려한 것이다. 어플리케이션 프로파일은 로컬시스템에 대한 설명, 대상 이용 계층에 대한 분석 및 주된 정보 요구 분석 등을 제시하는 것이다. 세 번째 메커니즘은 메타데이터 프레임워크(metadata framework) 즉 컨테이너(container)이다. 다양한 메타데이터 스키마들이 존재하는 경우 이를 캡슐화하여 다른 시스템에 전달하는 기능을 가지고 있으며, 대표적

인 예로는 RDF나 METS 등을 들 수 있다. 네 번째 메커니즘은 프로토콜이다. 프로토콜은 분산된 컬렉션에 존재하는 개별자료의 메타데이터를 교환할 때 사용되는 표준으로 이를 통해 다양한 디지털 서비스를 가능하다. 대표적인 예로는 OAI-PHM 모형을 들 수 있다.

Brenner, Larsen & Weston의 연구(2006)에서는 사서와 기록연구사들의 공통적인 관심사인 자원 수집과 보존에 대해 언급하였다. 도서관이나 기록관에서는 자원에 대한 배열, 기술, 사용에 대한 관점에는 차이가 있으나 최종적으로 자원공유를 위해 자원을 기술하고 배열하는 그 자체는 동일하다고 지적하였다. 자원은 문서, 파일, 시리즈, 레코드, 컬렉션 등 다섯 가지 단위로 나눌 수 있으며, 그 중 컬렉션 단위에서의 기술을 강조하였다. CLD는 일반적인 문서 또는 개별자료 검색을 위한 검색도구(finding aid)로 사용될 수 있으며, Portland State University(PSU) 통합 도서관 시스템을 대상으로 이를 검증하였다.

Kemp, Tan & Whalley(2007)의 연구에서는 지구공간 분석을 위해서 의미적인 접근을 시도하였다. 분산된 환경에서 각기 다른 목적을 가진 이용자들이 통합된 정보를 활용하기 위한 시도로 온톨로지 기반 시스템을 제시하였으며, 이는 상황정보와 주제 분석을 가능하게 하였다. Kemp, Tan & Whalley(2007)는 온톨로지 기반 시스템을 지식기반 서비스라고 간주하고, 이질적인 데이터베이스를 통합하여 온톨로지 기반 시스템을 구축하였다. 이용자가 이 시스템을 기반으로 개념 검색이 가능한 서비스를 활용할 수 있도록 설계하였다.

Qin & Finneran(2002)은 컬렉션 내에 위치

한 학습자원들을 시멘틱 웹 기준인 온톨로지로 표현하여 개별 학습자원의 활용을 주장하였다. 온톨로지 기반으로 표현된 CLD는 구성요소들-개별 학습자원들-의 개념, 관계성들을 포함하고 있다. 구체적인 정보요구에 맞는 표현은 구성요소들에 대한 기술에 제시되지만 잠재적인 자원 활용을 고려하고, 구체적인 구조적인 정보를 포함하기 위해서는 온톨로지 기반의 CLD가 필요하였다.

3. 컬렉션 단위 기술의 필요성

'컬렉션(collection)'이라는 용어보다는 '장서'라는 용어가 우리에게 더 익숙하다. '장서'는 도서관이 존재하는 가장 기본적인 조건으로 도서관을 채우는 개체인 동시에 도서관 이용자들의 도구인 셈이다. 장서는 도서관과 거의 일치된 개념으로 인식되어왔다. 장서에 대한 사전적인 개념은 '책을 간직해 둠'이며, 특히 문헌정보학용어사전에서 제시된 바로는 '일정한 방침에 따라 조직한 도서관을 말함'이다. 즉 도서관이나 정보를 다루는 장소에서 장서개발 정책에 따라 수집된 자원들을 '장서'라고 한다. '장서'에 대한 개념은 네트워크의 발전이후 그 범위가 확장되었다. 기존의 인쇄자료를 대상으로 한 장서의 범위가 비도서자료, e-book, e-journal, 데이터베이스 등으로 확대되었다.

본 연구에서는 장서라는 용어대신에 컬렉션이라는 용어를 사용하였으며 그 이유는 다음과 같다. 첫째, 기존의 장서의 개념은 물리적인 공간과 함께 고려되었다. 디지털도서관이라는 개념이 출현하고 자원에 대한 소유보다는 접근에

대한 관심이 고조됨에 따라 이러한 현상은 줄어들었지만, 현재까지도 장서하면 도서관에서 이용할 수 있는 자원이라는 개념에서 쉽게 벗어날 수가 없다. 둘째, 장서는 해당 도서관이나 정보센터의 장서관리정책에 의해 수집된 결과물이었다. 윤희운(2007)의 연구에서 제시한 장서관리정책은 “도서관이 이용자의 현재적 및 잠재적 요구를 충족시킬 목적으로 수립하는 장서관리 및 관리계획을 말한다. 다시 말해서 도서관의 목적과 목표, 이용자의 요구, 자료예산 등을 반영한 장단기 장서관리계획을 수립·관리하기 위한 활동지표인 동시에 성문규정이다.”라고 정의하였다. 이와 같이 장서관리정책은 해당 도서관에 초점을 맞추고 있는 정책이며 특히 도서관의 설립목적이나 예산을 중심으로 하는 것이다.

본 연구에서 다루고자 하는 컬렉션은 물리적인 공간을 초월하는 자원을 대상으로 하는 것이며, 하나의 도서관을 중심으로 하는 자원들의 집합체보다는 여러 기관이 협력하거나 많은 이용자들이 공유할 수 있는 장서에 초점을 맞추고 있다. 그러므로 본 연구에서는 ‘장서’라는 용어보다는 ‘컬렉션’이라는 용어를 사용하게 되었다.

3.1 컬렉션의 개념

컬렉션이라는 개념을 확장시킨 연구는 디지털 도서관이라는 새로운 변화에 맞춰서 진행되었다. 송영희 등(2005)의 연구에서는 장서관리 영역의 확장 범위를 전자저널의 관리, 인터넷 정보자원의 개발과 관리, 디지털 아카이빙, 인쇄매체와 전자매체와의 통합과 관련된 정보자

원의 통합관리 등으로 확장시켰다.

Heaney(2000)의 연구에서도 아래와 같이 장서 즉 컬렉션의 범위를 확대시켰다.

- 인터넷 포털 사이트(예. Yahoo)
- 주제 게이트웨이(예. Intute: Social Sciences)
- 도서관, 미술관, 기록관의 목록
- 웹 색인(예. Alta Vista)
- 텍스트, 이미지, 음원, 데이터셋, 소프트웨어 등의 컬렉션 또는 이러한 자원들을 혼합한 형태(데이터베이스, CD-ROM, 웹 자원의 컬렉션을 포함함)
- 이벤트 중심으로 한 컬렉션(예. Follett Lecture Series)
- 도서관 및 박물관 컬렉션
- 디지털 아카이브

네트워크의 발전을 편하게 받아들이는 이용자들은 장서라는 개념으로 한정되었던 도서관 또는 정보를 다루는 기관들의 자원 집합체가 디지털 콘텐츠의 집합체로 확장됨을 기대하였다.

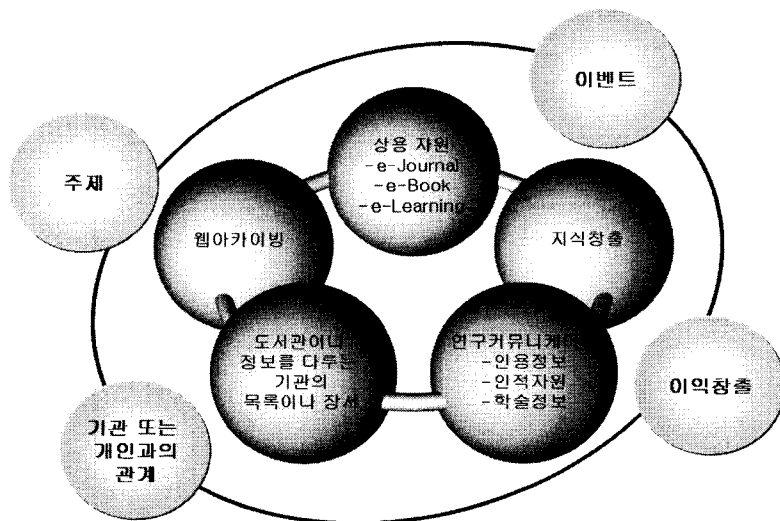
디지털 콘텐츠에 대한 개념을 본 연구에서는 크게 세 가지 영역으로 정의하였다. 첫째 영역은 디지털 콘텐츠의 구축 목적이다. 디지털 콘텐츠는 다양한 목적에 의해 수집되거나 생성된다. 둘째 영역은 디지털 콘텐츠의 범위이다. 흔히 디지털 콘텐츠라고 불릴 수 있는 자원들을 5가지 하위 영역으로 구분하고 그에 대한 실제 사례를 제시하였다. 셋째 영역은 디지털 콘텐츠와 관련된 행위 주체들이다. 디지털 콘텐츠는 다양한 개인이나 기관들에 의해 생성, 수집, 관리, 이용, 보존되며, 이를 확인하는 것은 분산된 정보 환경에서 중요하다.

〈그림 1〉의 안쪽 다섯 개의 요소는 디지털 콘텐츠의 범위를 규정하는 것이며, 바깥쪽의 네 개의 요소는 디지털 콘텐츠를 구축 목적을 제시하였다.

디지털 콘텐츠의 범위는 〈그림 1〉에서 제시한 다섯 가지의 범주로 정리된다. 첫째, 도서관이나 정보를 다루는 기관의 목록이나 장서이다. 이는 전통적인 장서의 개념을 가장 잘 반영하는 것으로 물리적인 공간에 소장되어 있는 자원을 전자적인 형태로 변형한 것이다. 둘째, 연구커뮤니케이션과 관련된 콘텐츠이다. 연구의 결과물에 포함될 여러 가지의 정보를 이용하여 그 결과물에 대한 표현을 좀 더 정확하게 제시할 수 있다. 즉 인용 정보, 연구자들 간의 교류, 결과물 내에 있는 데이터 셋, 표나 그림 등도 하나의 정보원으로 가치를 부여할 수 있다. 셋째, 이윤창출을 위해 제작되어진 전자형태의

자원인 e-Journal, e-Book, e-learning 자원이다. 이러한 자원들은 디지털 콘텐츠 또는 멀티미디어 정보원이라고 할 만큼 다양한 정보 형태를 지니고 있다. 넷째, 디지털의 특성을 고려해야 하는 웹 아카이빙이다. 가변적인 웹 콘텐츠를 유지·관리하는 것은 그리 쉬운 일은 아니다. 국가 단위의 또는 기관이나 개인의 가치 있는 정보를 지속적으로 보존하는 것은 매우 중요하다. 마지막 범위는 지식 창출이다. 도서관뿐만 아니라 정보를 다루는 모든 기관과 개인들은 수집된 자원에 새로운 가치를 창조하였다. 기관이나 개인들은 기존의 다양한 형태의 정보를 수집하고, 거기에 각자가 가지고 있는 경험과 정보를 추가시킨다. 그러한 자원들은 추가된 경험과 정보를 바탕으로 하나의 지식으로 발전하게 된다.

다음은 〈그림 1〉에서 제시한 디지털 콘텐츠

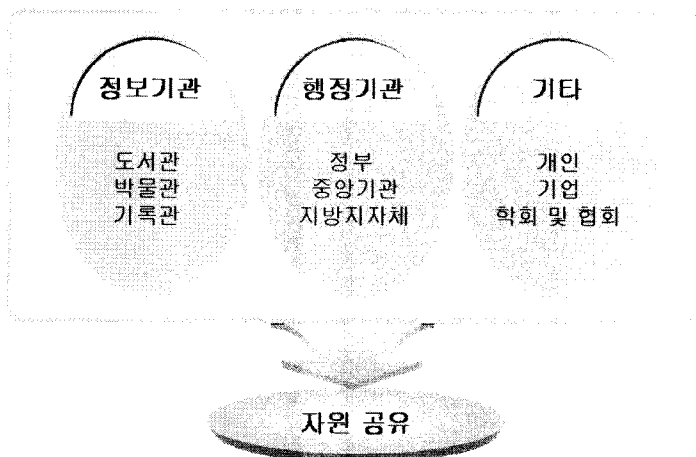


〈그림 1〉 디지털 콘텐츠의 범위와 구축 목적

구축 목적에 대한 설명이다. 첫째, 디지털 콘텐츠를 구축한 목적은 행위의 주체인 기관이나 개인과의 관계에서 찾아볼 수 있다. 해당기관은 기관의 설립 목적이 있으며 그에 따른 프로젝트나 업무가 발행하게 된다. 예를 들어 도서관은 '도서관 이용자들의 요구에 충족하는 자원 획득'을 목적으로 하고 있다. 그러므로 도서관은 다양한 자원을 수집하고 자세하게 기술을 표현한다. 둘째, 디지털 콘텐츠 구축과 관련된 이익창출은 e-Journal, e-Book, e-learning 같은 전자적인 형태의 자원을 구축하는 것과 가장 밀접한 관계를 가지고 있다. 셋째로 제시된 목적은 하나의 주제를 기반으로 하여 디지털 콘텐츠를 구축하는 것이다. 특히 도서관, 박물관, 기록관에서는 특정 주제로 표현될 수 있는 다양한 형태의 자원을 수집하고 관리한다. 주제에 따른 컬렉션 구축은 국가단위로 진행되는 웹 아카이빙, 인물중심으로 수집된 컬렉션, 예술적인 요소가 가미된 자원들을 수집하는 등 다양하게 구현될 수 있다. 네 번째는 이벤트이

다. 이벤트는 세 번째의 주제와 비슷하지만 주제와는 다른 특성을 가진다. 이벤트에 관련된 목적에는 시공간적인 특성이 부각된다. 예를 들어 특정 년도의 특정 사건에 대해서만 자원을 수집한다면, 바로 하나의 이벤트를 중심으로 한 디지털 콘텐츠 구축이 되는 것이다.

〈그림 2〉는 〈그림 1〉의 디지털 콘텐츠 구축과 관련된 주제들에 대한 내용을 좀 더 구체적으로 표현하였다. 디지털 콘텐츠를 구축하는 주체가 이것을 사용하는 주체와 동일시될 수 있다. 디지털 콘텐츠와 관련된 기관 중 가장 대표적인 기관은 도서관이며, 그와 비슷한 정보기관은 박물관과 기록관이다. 이러한 세 기관들은 정보공유를 위한 연구와 노력을 계속 진행하고 있으며 그 결과 MLA 협력 프로그램이나 자원공유를 위한 자원기술규칙 등이 제시되었다. 두 번째 기관은 행정기관이다. 정부부처나 지방 자치단체와 관련이 있으며, 활발하게 새로운 디지털 콘텐츠를 구축하고 있다. 가장 일반적인 것은 각 부처나 지자체에서 제공하는



〈그림 2〉 디지털 콘텐츠와 관련된 주체들

정보공개이며, 이를 통해 행정문서, 예산자료, 홍보 및 교육 프로그램에 대한 정보가 전달된다. 또한 이들은 각 행정관할구역의 역사나 그 지역의 특정성을 다양한 자원들을 관리한다. 마지막은 기타범주로 존재하는 모든 개인과 기관들을 포함시킬 수 있다. 기관들은 물리적인 공간을 대표할만한 웹사이트를 구축하고 그곳에 기관의 이념이나 활동들을 제시한다. 특히 웹사이트를 통해 기관의 역사와 지식을 관리하고자 노력한다. 이는 일반적인 기관뿐만 아니라 학교나 학회, 협회 등 비영리적인 단체도 마찬가지이다.

디지털 콘텐츠는 <그림 1,2>에서 살펴본 바와 같이 넓은 범위, 다양한 구축 목적, 다수의 관련 주체들과 연결된다. 이는 시공간적인 것을 뛰어넘는 일이며, 전통적인 장서를 이해하는 방식과는 다르게 접근해야한다. 이제는 디지털 콘텐츠를 생산하는 이들이 정보를 전문적으로 다루는 이들이 아닌 정보에 접근할 수 있는 모든 이들로 확대되었다. 그러므로 본 연구에서는 디지털 콘텐츠를 하나의 디지털 자원 집합체로 간주하고, 이를 컬렉션이라는 관점에서 접근하였다.

3.2 컬렉션 단위 기술

컬렉션 단위 기술(collection-level description: 이하 CLD라고 함)은 이질적인 기관들이 정보를 공유하는 것과 밀접한 관계가 있다. 즉 정보를 이용하는 이들은 본인과 관련된 기관의 틀을 벗어나 전체적인 웹상의 모든 자원을 이용하기를 원한다. 이용자들이 바라는 것은 자원에 대한 통합적인 이용이다. 통합적인 이용

이라는 것은 자원에 대한 소재를 알려주고 접근을 가능하게 하는 것만이 아니라, 단일 인터페이스를 통해서 수집된 모든 자원의 개념, 주제, 관련 인물이나 기관, 형태, 형태에 따른 매체 사용 등이 분석되고 제공되는 것이다.

그러므로 단일 인터페이스를 통한 정보제공은 분산된 자원에 대한 설명과 분석이 선행된 경우에만 가능하다. 현재 분산된 자원 자체를 표현하는 기술은 다양하게 제시된 반면에 그 자원들을 둘러싼 상황정보에 대한 기술 표준은 미비한 수준이다. 이에 본 장에서는 분산 환경에서의 CLD의 중요성을 알아보고, 다양한 CLD 표준들을 살펴보았다.

3.2.1 컬렉션 단위 기술의 중요성

디지털 컬렉션을 구성하는 개별 자원에 대한 기술과 함께 CLD에 대한 중요성을 제시한 연구로는 Foulonneau et al.(2005)와 Macgregor(2003) 등이 있다.

Foulonneau et al.(2005)의 연구에서는 개별자료 단위 메타데이터(item-level metadata) 수집을 위한 CLD의 중요성을 제시하였다. 그들의 연구에서는 개별자료 단위 메타데이터는 해당 기관의 규칙에 맞게 작성된 것이며, 그 개별자료에 대한 기술적인 세분화가 강조된 것이라고 주장하였다. 만약 개별자료 단위 메타데이터의 범위 안에 개별자료를 둘러싼 상황에 대한 정보를 포함하지 못한다면, 그 자체의 정확도가 떨어질 수 밖에 없다고 지적하였다. 특히 개별자료 단위 메타데이터만을 고려한다면 주제(subject)를 표현한 암시적이고 참고적인 정보 확보가 어렵다고 주장하였다.

Macgregor(2003)의 연구에서는 기능성 세

분화(functional granularity)에 대해 설명하였다. 이는 컬렉션에 대해 얼마나 많은 정보를 제공하는가가 따라 메타데이터의 효율성을 높여 줄 수 있다는 것이다. 컬렉션에는 다양한 수준의 하부 컬렉션을 가지고 있으며 컬렉션 간의 관계를 설명할만한 기술 요소들이 필요하다고 하였다.

자원에 대한 기술은 크게 두 가지로 나눌 수 있다. 하나는 자원 자체에 대한 표현 기술이며, 다른 하나는 자원을 둘러싼 상황을 표현하는 기술이다. 특히 앞에서 제시한 연구들이 후자 기술인 상황정보를 표현한 것을 강조한 것이다.

〈그림 3〉에서 제시한 바와 같이, 상황을 나타내는 정보인 CLD를 통해 관련된 컬렉션의 수준을 알 수 있으며, 더 나아가서는 자원의 주제와 생산자 및 소유자에 대한 정보도 얻을 수 있다. 예를 들어 '2008년 베이징올림픽 컬렉션'을 구축한다면 베이징올림픽이라는 컬렉션 이름아래 개별자료에 대한 메타데이터를 작성될 것이다. 만약 2008년 베이징올림픽이라는 컬렉

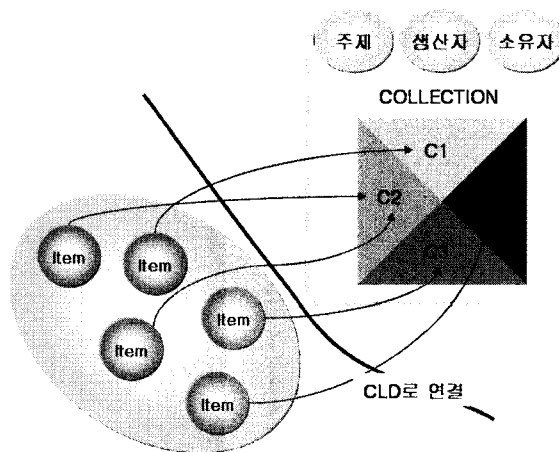
션에 대한 정보가 사라진다면 각 개별자료의 정체성을 확인하는 것은 어려울 것이다.

또한 이성숙(2005)의 연구에서는 디지털 도서관의 확산에 따라 복수 자원을 검색하고 관리하기 위한 가이드 역할을 하는 '정보지도(information map or landscape)와 같은 도구가 필요한 시점에서, 산발적으로 존재하던 장서기술을 포괄하는 CLD의 개념이 정립된 것은 그 의의가 있다고 설명하였다.

3.2.2 컬렉션 단위 기술 관련 표준

먼저 일반적인 두 개의 CLD관련 표준을 알아보고, 그 후에 이들보다 좀 더 구체적인 클래스와 속성을 정의한 Natural Collection Description에 대해 알아보았다.

UKOLN(UK Office for Library Networking)와 협력을 통해 개발된 RSLP(Research Support Libraries Program) Collection Description Schema(이하 RSLP CD라고 함)와 DC Collection Description Application Profile



〈그림 3〉 CLD를 통한 컬렉션과 개별자료와의 연결

(이하 DC CD AP라고 함)은 CLD의 대표적인 예이다. 두 스키마는 Heaney의 장서모델을 기반으로 개발되었다.

RSLP CD 스키마가 가지는 의미는 첫째, 컬렉션 데이터 모델에 기반한 스키마라는 것이다. 둘째, 특정 장서만을 대상으로 하는 것이 아니라 범용적으로 적용될 수 있는 모델로 개발된 것이다. 셋째, RSLP CD 스키마가 RDF를 사용해서 구현되었다는 것이다(이성숙 2005: Xin 2004). RSLP CD 스키마에서는 창작하다(creates), 출판하다(produces), 수집하다(collects), 소유하다(owns), 관리하다(administers), ~에 수집되다(isGatheredInto), ~에 위치하다(hasLocation), ~의 한부분이다(hasPart) 등의 속성 유형을 제시하였다.

RSLP CD 속성은 컬렉션(collection), 위치 정보(location), 에이전트(agent) 등으로 나누어져 있으며 특히 컬렉션과 위치정보 속성은 세분화되었다. 컬렉션 속성은 일반속성(general attribute), 주제(subject), 날짜(date), 관련 에이전트(associated agent), 외부와의 관계(external relationships)등으로 세분되었으며, 위치정보 속성은 일반속성(general attribute), 관련 에이전트(associated agent), 물리적인 위치(physical location), 온라인 위치(online location) 등으로 나누어져 있다(표 1 참고).

DC CD AP는 RSLP CD 스키마를 기반으로 하여 데이터 모델을 제안하였으며, 서비스(service)라는 엔티티를 추가하여 컬렉션의 개별자료들을 접근하기 위한 시스템을 고려하였다.

〈표 1〉 RSLP CD 속성

COLLECTION		Attribute	RDF property
Attribute	RDF property	Object	cld: objectName
General attributes		Name	cld: agentName
Title	dc: title	Place	dcq: spatial
Identifier	dc: identifier	Time	dcq: temporal
Description	dc: description	Dates	
Strength	cld: strength	Accumulation Date Range	cld: accumulationDateRange
Physical Characteristics	dc: format	Contents Date Range	cld: contentsDateRange
Language	dc: language	Associated agents	
Type	dc: type	Collector	dc: creator
Access Control	cld: accessControl	Owner	cld: owner
Accrual Status	cld: accrualStatus	External relationships	
Legal Status	cld: legalStatus	Sub-collection	dcq: hasPart
Custodial History	cld: custodialHistory	Super-collection	dcq: isPartOf
Note	cld: note	Catalogue or description	cld: hasDescription
Location	cld: hasLocation	Described collection	cld: isDescriptionOf
Subject		Associated collection	cld: hasAssociation
Concept	dc: subject	Associated publication	cld: hasPublication

출처: <<http://www.ukoln.ac.uk/metadata/rslp/schema/>>

DC CD AP에서는 ~에 수집되다(is-gathered-into), ~라고 기술되다(is-described-by), 수집하다(collects), 소유하다(owns), ~에 위치하다(is-located-in), ~에 의해 접근되다(is-accessed-by), 관리하다(administers), 제공하다(provides) 등의 속성 유형을 제시하였다. <표 2>는 DC CD AP의 28개 속성으로 컬렉션 단위의 요소들을 정의하였다.

<표 1, 2>를 근거로 하여 두 스키마간의 공통점과 차이점을 살펴보았다. 먼저 공통점으로는 컬렉션의 날짜정보나 관리자 및 소유권 변경에 대한 내력, 컬렉션 간의 관계를 나타내는 요소들이 같은 속성으로 표현되었다. 두 스키마 간의 차이점은 다음과 같다. 첫째, 주제에 대한 표현이 RSLP CD가 더 자세하였다. 컬렉션의 개념과 관련 개체를 바탕으로 시공간적인 배경 그리고 관계를 가지는 개인 및 단체를 표현하고자 하였다. 둘째, RSLP CD은 행위 주체들에 대한 다양한

접근을 시도하였다. 누가 컬렉션을 수집하였는지(Associated agents_Collector), 누가 컬렉션을 소유하고 있는지(Associated agents_Owner), 누가 컬렉션 위치에 대한 책임을 지고 있는지(Associated agents_Administrator)를 구별하였다.

위에서 살펴본 CLD 표준들 보다 좀 더 발전된 표현을 제시한 Natural Collection Description(이하 NCD라고 함)은 온톨로지를 기반으로 하여 클래스와 속성을 정의하였다. NCD는 the European Commission, Gordon & Betty Moore Foundation, Natural History Museum, RLG Programs, Smithsonian Institution 등의 지원을 통해 개발된 것으로 박물관 연구(natural history)에 관한 컬렉션의 기술 교환 표준이다. 컬렉션 기술인 NCD는 일반적인 자원 검색을 위해 간단한 DC 레코드를 추출할 수도 있으며 EAD(Encoded Archival Descrip-

<표 2> DC CD의 속성

Properties	Qualified Name	Properties	Qualified Name
Collection Identifier	dc: identifier	Subject	dc: subject
Title	dc: title	Spatial Coverage	dcterms: spatial
Alternative Title	dcterms: alternative	Temporal Coverage	dcterms: temporal
Description	dcterms: abstract	Accumulation Date Range	dcterms: created
Size	dcterms: extent	Contents Date Range	cld: dateContentsCreated
Language	dc: language	Collector	dc: creator
Type	dc: type	Owner	marcrel: OWN
Rights	dc: rights	Is Located At	gen: isLocatedAt
Access Rights	dcterms: accessRights	Is Accessed Via	gen: isAccessedVia
Accrual Method	dcterms: accrualMethod	Associated Publication	dcterms: isReferencedBy
Accrual Periodicity	dcterms: accrualPeriodicity	Super-Collection	dcterms: isPartOf
Accrual Policy	dcterms: accrualPolicy	Sub-Collection	dcterms: hasPart
Custodial History	dcterms: provenance	Catalogue or Description	cld: collectionDescription
Audience	dcterms: audience	Associated Collection	cld: associatedCollection

출처: <<http://www.ukoln.ac.uk/metadata/dcmi/collection-application-profile/>>

tion)처럼 풍부한 컬렉션 기술을 확보할 수도 있다(Biodiversity Information Standards 2008).

NCD의 클래스는 크게 헤더(header), 컬렉션(collection), 기관(institution), 자세한 연락처(contact details) 등으로 나누어져 있으며, 반드시 제시되어야 하는 네 개의 속성들 - 레코드의 저자(author of the record), 레코드

생성 날짜(date of record creation), 컬렉션 이름(collection name), 컬렉션 기술(collection description)- 을 지정하였다(Biodiversity Information Standards 2008).

NCD의 클래스와 속성에 대한 정의는 <표 3>과 같으며, NCD의 클래스를 온톨로지로 표현하면 <그림 4>와 같다.

<표 3> NCD 클래스와 속성에 대한 정의

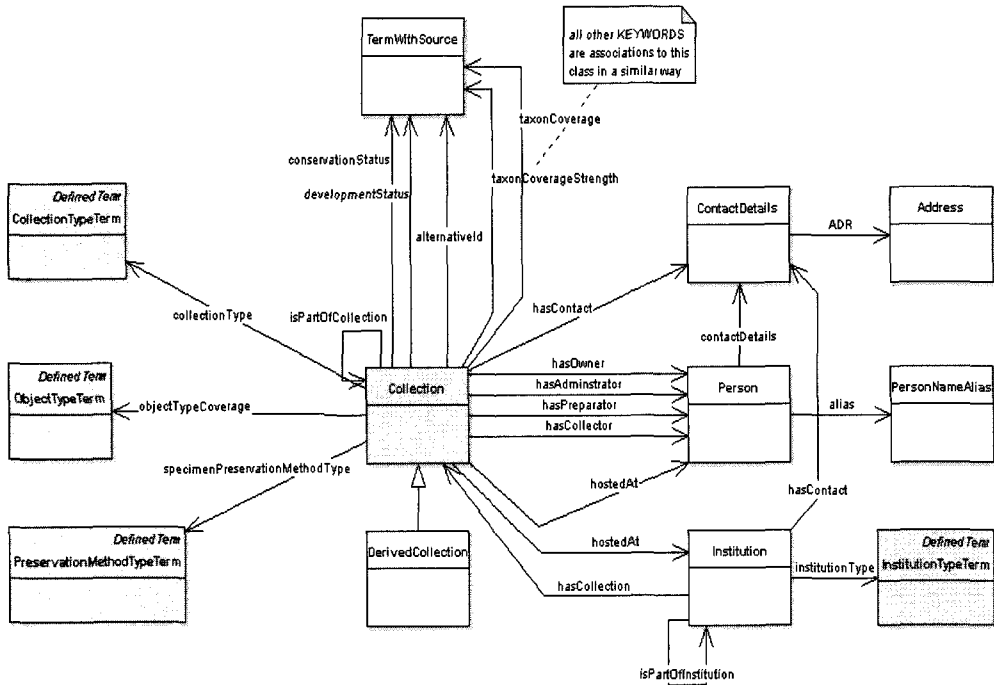
클래스	클래스 및 속성 이름	정의
Header	Record Source	(저자인 다른 사람에 의해 생성된 경우)레코드의 정보원
	Record Harvest Date	레코드 수집 날짜
	Author*	레코드 생산자
	Corporate Affiliation	저자소속기관
	Record Created Date*	레코드 생산 날짜
	Editor	편집자
	Record Edited Date	레코드 마지막 편집 날짜
	Record Rights	레코드의 지적재산권 상황
	Notes	기타정보
Collection	Derived Collection	추출된 컬렉션(개별자료 데이터베이스에서 질의에 적합한 셋)
	Collection Identifier	컬렉션 식별자(ex. URI(LSID 또는 URL))
	Alternative Identifier	대체 컬렉션 식별자
	Parent Collection Identifier	부모 컬렉션 식별자
	Acronym or Coden	두문자어 또는 약어
	Collection Name*	로컬 언어로 된 공식적인 컬렉션 이름
	Alternative Name	대체 컬렉션 이름
	Associated Person	컬렉션과 관련된 사람의 이름
	Description*	일반적인 이용을 위한 짧은 설명
	Description for Specialists	특별한 이용을 위한 기술적인 텍스트
	Extent	컬렉션의 크기나 범위를 자료롭게 서술
	Collection Type**	archival, art, audio, cell cultures 외 16개
	Common Name Coverage	로컬 언어로 된 일반적인 분류군 이름
	Conservation Status**	McGinley Level 1-10
	Conservation Status Date	Conservation Status를 평가받는 날짜
	Digital Format	디지털 포맷
	Digital Medium	디지털 매체
	Development Status**	active growth, closed, consumable, decreasing, lost, missing, passive growth, static 등
	Expedition Name	수집된 자원들을 발견한 원정의 공식적인 이름
	Formation Period	컬렉션이 형성된 기간
	Place Name Coverage	장소에 대한 명칭
	Geospatial Coordinates	십진으로 나타낸 위도와 경도
Item-Level Access	개별자료 데이터에 대한 URI	
Kingdom Coverage**	animalia, archaeobacteria, eubacteria, fungi, plantae, protista 등	

클래스	클래스 및 속성 이름	정의
Collection	Known to Contain Types	표시문자에 대한 속성
	Living Time Period	생물 자원이 생존한 기간
	Physical Location	개인소장 컬렉션일 경우 또는 컬렉션과 관련된 주된 기관과의 연결
	Primary Grouping Principle**	cultural, ecosystems, environmental, events, expeditions 외 8개
	Primary Purpose**	commercial, education, exhibition, monitoring, ornamental, personal, research, voucher 등
	Related Material	관련된 자원에 대한 짧은 설명
	Related Collection	관련된 컬렉션에 대한 짧은 설명
	Specimen Preservation Method**	controlled atmosphere, cryopreserved, dried, dried and pressed 외 17개
	Taxon Coverage	컬렉션 안에서의 가족군 수준의 분류
	Temporal Coverage	컬렉션의 자원들과 관련있는 시간대
	Access Conditions	컬렉션 접근에 관한 조건
	Usage Conditions	컬렉션 이용에 관한 조건
	Provenance	소유권의 변화와 컬렉션 관리, 감독에 관한 설명
IPR Statements	전반적인 자원들에 관한 지적재산권 상황	
Notes	기타정보	
Institution	Institution Identifier	기관을 나타내는 식별자
	Name	로컬 언어로 된 공식적인 기관 명칭
	Alternative Name	기관의 대체 명칭
	Unit Name	기관 하부 단위의 명칭
	Description	일반적인 이용을 기관 설명
	Alternative Identifier	기관의 대체 식별자
	Contact	컬렉션에 대한 정보를 얻기위한 연락처와의 연결
	Type	aquarium, archive, botanic garden, conservation 외 19개
Parent Institution or Network	모 기관의 이름, 두문자어, 약어 등	
Contact Details	Name	사람 이름
	Family Name	가족성
	Given Name	성
	Other Name	기타
	Prefix	성명 앞에 붙이는 경칭
	Suffix	성명 뒤에 붙이는 접미사
	Job Title	직업명
	Role	역할명
	Institution Name	기관 명칭
	Institution Unit	기관 하부 단위의 명칭
	Post Office Box	사서함
	Extended Address	주소
	Street Number and Name	
	Local Area name	
	Regional Name	
	Postcode or ZIP code	우편번호
	Country Name	국가명
	Telephone Number	전화번호
Fax Number	팩스번호	
Email Address	이메일주소	
URL	URL	
Logo URL	로고를 위한 URL	
Notes	기타정보	

출처: <<http://www.tdwg.org/standards/>>

* : 반드시 표현되어야 할 속성

** : 지정된 값 중에서 택일하는 속성



〈그림 4〉 온톨로지로 표현된 NCD 클래스

출처: <http://wiki.tdwg.org/twiki/pub/NCD/NCDWorkShop/NCD_ontology.pdf>

〈표 3〉과 〈그림 4〉를 근거로 살펴본 NCD의 문제는 클래스와 속성의 구별이 모호하다는 것이다. 〈표 3〉에서는 가장 상위 클래스인 헤더, 컬렉션, 기관, 자세한 연락처 등을 기준으로 하위의 클래스와 속성을 구별 없이 제안하였다. 또한 〈그림 4〉는 컬렉션 클래스를 중심으로 다른 클래스와의 관계 속성을 제시하였지만, 〈표 3〉에서 제시한 속성들을 정확하게 표현하지는 못하였다. 특히 〈그림 4〉에서 나타난 온톨로지 기반 NCD 클래스는 〈표 3〉의 헤더 클래스에 해당하는 부분을 제시하지 못했으며, 〈표 3〉과는 다르게 인적자원(person)이라는 부분을 따로 분리하여 정리하였다.

위와 같은 문제가 있음에도 불구하고, NCD

표준은 RDF 중심의 기존 데이터 모델링을 온톨로지기반으로 표현한 성공적인 연구라는 점에서 그 의미를 찾아볼 수 있을 것이다.

3.2.3 기존 컬렉션 단위 기술 분석

위에서 소개한 RSLP CD와 DC CD AP는 RDF 스키마를 기반으로 한 데이터 모델을 제시하였다. RDF는 자원들과 그들 간의 관계를 표현하는 데이터 모델이며 간단한 의미를 제공한다. 그런 반면에 NCD는 온톨로지를 기반으로 한 데이터 모델을 제공하였다. 온톨로지는 클래스와 속성에 대하여 기술할 수 있는 풍부한 어휘를 제공함으로써 지식기반 데이터 모델링을 가능하게 한다. 온톨로지 기반으로 구축

된 데이터 모델은 다차원적으로 정보를 확보하고 개체들 간의 관계를 부각시킴으로써 새로운 지식을 추론할 수 있는 기반으로 제공한다.

그러므로 4장에서는 기존 CLD에서 표현하지 못했던 정보를 다양하게 기술할 수 있는 방법으로 온톨로지 기반 데이터 모형을 개발하고자 하였다.

4. 온톨로지 기반 컬렉션 단위 기술 모형 개발

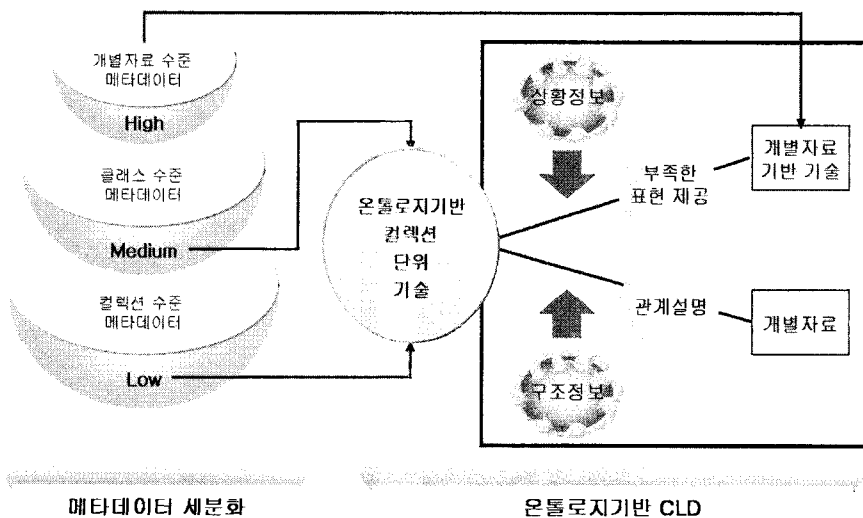
메타데이터 세분화(metadata granularity)에 대한 설명은 Calanag, Tabata & Sugimoto(2004)와 LeFurgy(2002)의 연구에서 언급되었다. <그림 5>의 왼쪽 부분을 살펴보면 개별자료로 갈수록 메타데이터 세분화가 높으며, 컬렉션 단위로 갈수록 메타데이터 세분화가 낮아진다고 할 수 있다. 중간정도의 메타데

이터 세분화라는 특성을 가지는 클래스수준의 메타데이터는 개체 속성들의 형태와 통합 시 필요한 정보를 조정하는 구조적인 정보를 말하는 것이다.

본 연구에서는 기존의 메타데이터 세분화 연구에서 제시한 바를 기초로 컬렉션 및 클래스 단위의 메타데이터를 포함하는 CLD를 제시하였다(그림 5 참고). 또한 본 연구에서는 온톨로지를 이용하여 특정 컬렉션과 개별자료들 간의 관계 그리고 그 개별자료의 다양한 특성을 제시하는 표현을 지원하였다. 특히 온톨로지 기반 CLD 모형은 상황정보를 제공함으로써 개별자료 메타데이터에서 나타낼 수 없는 표현들을 지원하며, 구조정보를 통해 개별자료들 간의 관계, 컬렉션 간의 관계를 표현할 수 있었다.

4.1 클래스 및 속성 정의

본 연구에서 제시한 클래스 및 속성은 다음



<그림 5> 수준별 메타데이터 세분화와 온톨로지 기반 컬렉션 단위 기술과의 차이

의 <표 4, 5>와 같다. 기존의 CLD 표준들에서는 클래스와 속성을 구별하지 않고 그냥 필드라는 항목을 통해 혼용하여 정의하였다. 그렇지만 본 연구에서는 클래스와 클래스의 특성을 표현하는 속성을 분리하여 정의하였다.

본 연구에서는 온톨로지 언어 OWL의 편집기인 Protégé가 지원하는 속성 표기에 맞춰서 작성하였다. Protégé에서는 속성을 오브젝트 타입(objecttype)과 데이터타입(datatype) 속성으로 나누고 있다. 오브젝트타입은 클래스간의 관계를 설명하는 것이고, 데이터타입 속성은 클래스 자체에 대한 특성을 기술하는데 사

용된다.

<표 4>에서 제시된 내용은 클래스를 중심으로 해당 클래스의 특성을 기술하는 데이터타입 속성을 제시하였으며, 특이점을 정리하면 다음과 같다. 첫째, 컬렉션 클래스는 하위 클래스를 두어서 서브 컬렉션이 계층적으로 위치하도록 설계하였다. 둘째, 개체와 주제는 각각의 하위 클래스를 가지며, 컬렉션 클래스의 하위클래스와는 다르게 클래스를 구성할 어휘집(vocabulary)을 제공하여 정적인 하위클래스를 제시한다. 셋째, 속성 값이 미리 제시되는 경우가 존재한다. NCD처럼 속성에 대한 속성값을 미리

<표 4> 본 연구에서 제시한 CLD 클래스 및 데이터타입속성

Class	Property(datatype)	Class	Property(datatype)
컬렉션 (하위클래스)	식별자	인적자원	식별자
	컬렉션명칭		이름
	대체컬렉션명칭		별명
	기술		연락처
	형태		홈페이지
	주제		이메일
	시간적 정보		인적자원특성리스트
	공간적 정보	서비스	식별자
	구축목적리스트		서비스명칭
	보존방법리스트		정보시스템
저작권상황		연결웹서비스	
기관	식별자	개체(하위클래스)	식별자
	기관명칭	개별자료	식별자
	대체기관명칭		자료명
	기관특성리스트		자료유형
	우편번호		개별자료_저작권상황
	주소	주제(하위클래스)	식별자
	전화번호	이벤트	식별자
	팩스번호		이벤트_시간적 정보
	홈페이지		이벤트_공간적 정보
	대표이메일		
	하위부서		

제시하는 방법을 채택하였으며, 보존방법리스트, 기관특성리스트, 인적자원특성리스트 등이 이에 해당한다.

〈표 5〉는 클래스간의 관계를 표현하는 오브젝트타입 속성을 제시하였다. 온톨로지 언어로 표현된 속성은 그 범위가 자유롭고 내용 확장

도 가능하며, 이러한 속성들을 기반으로 새로운 관계가 추론된다. 예를 들어 이벤트와 주체가 연결된 '이벤트와관련된주체가있다'속성은 나중에 특정사건과 주제적인 배경을 연결할 수 있는 연결고리를 제공할 것이다.

〈표 5〉 본 연구에서 제시한 CLD 오브젝트타입속성

Property(objecttype)	Domain	Range
창작하다	기관 인적자원	개체
출판하다	기관 인적자원	개별자료
수집하다	기관 인적자원	컬렉션
소유하다	기관 인적자원	컬렉션
관리하다	기관 인적자원	컬렉션
저작권이있다	기관 인적자원	컬렉션 개별자료
소속되어있다	인적자원	기관
서비스책임기관이다	서비스	기관
위치하다	컬렉션	기관 인적자원
~에 수집되다	컬렉션	개별자료
관련컬렉션이있다	컬렉션	컬렉션
관련인물이있다	컬렉션	인적자원
관련기관이있다	컬렉션	기관
관련개체가있다	컬렉션	개체
이용되다	컬렉션	서비스
주체가있다	컬렉션 개체 개별자료	주제
이벤트가있다	컬렉션 개체 개별자료	이벤트
이벤트와관련된주체가있다	이벤트	주제

4.2 분석 및 기대효과

본 연구에서 제시한 온톨로지기반 CLD를 분석하고 그에 대한 효과를 살펴보면 <그림 6>과 같다.

본 연구에서는 CLD를 통해 제시될 수 있는 항목을 크게 네 가지로 나누었다(그림 6 참고).

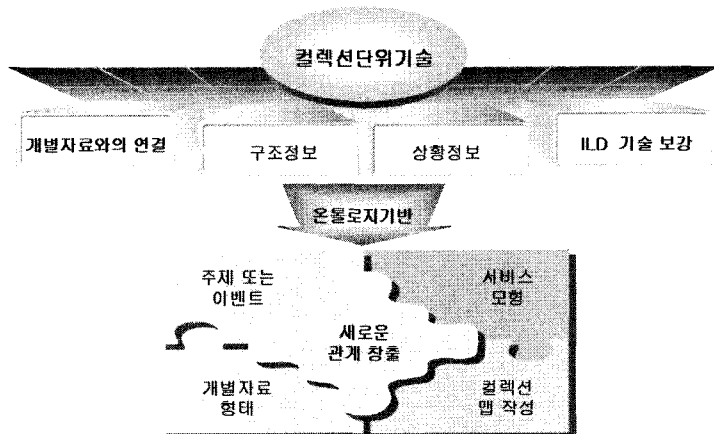
첫 번째로 제시될 수 있는 항목은 개별자료들과의 연결이다. 특정 컬렉션 안에서 각 개별자료들은 자체의 특성 표현을 제한적으로 기술된다. 하나의 개별자료는 다른 기관의 컬렉션에 존재하다가 해당 기관의 수집정책에 의해 새로운 대주제-컬렉션의 주제와 부합됨-로 해당 기관에서 재이용될 수 있으며, 이는 컬렉션과 개별자료와의 연결을 통해 제시되었다.

두 번째는 구조정보이다. 구조정보는 첫 번째의 '개별자료와의 관계' 항목에도 도움을 줄 수도 있으며, 자세한 기능은 다음과 같다. 구조정보는 첫째, 컬렉션 내의 하위 컬렉션을 연결하였고, 둘째, 주제나 이벤트 요소들을 중심으로 개별자료를 연결할 수 있다. 구조정보는 특

정 컬렉션에 존재하는 개별자료들을 군집화하여 그 군집의 속성들을 파악할 수 있는 정보를 제공하였다.

세 번째는 ILD에서 제공될 수 없는 기술을 확보하는 ILD 기술 보강이다. ILD의 구성요소들은 관련 기관이나 개인을 미리 염두하고, 구축목적이나 해당 주제를 생략하는 경우가 있다. 이를 위한 CLD 모형에서는 CLD를 통해 중점적으로 다루고 있는 주제를 설명하고, 관련 에이전트나 인적자원을 확보하였다.

네 번째는 상황정보이다. 상황정보는 세 번째의 'ILD 기술 보강' 항목에도 도움을 줄 수도 있으며, 자세한 기능은 다음과 같다. 상황정보는 특정기관에서 생성 또는 수집하고자 하는 컬렉션의 구축목적, 관련 에이전트나 인적자원을 제시할 수 있으며, 컬렉션의 주제와 이벤트 등 모든 요소들을 동시에 설명할 수도 있다. 또한 상황정보는 CLD의 가장 중요한 항목으로 ILD의 기능성 세분화를 지원할 수 있으며, 컬렉션 자체에 대한 속성을 제시하였고, 무엇보다 과거의 컬렉션 이력에 대한 공유가 가능하



<그림 6> 온톨로지기반으로 표현된 CLD 모형의 효과

게 하였다.

본 연구에서는 CLD를 통해 제시될 수 있는 항목들을 온톨로지를 기반으로 하여 재조명하였다. 온톨로지를 기반으로 한 CLD 모형의 장점은 새로운 관계를 창출할 수 있다는 것이다. 해당 기관은 하나의 주제 또는 이벤트를 중심으로 컬렉션들을 모아볼 수도 있으며, 특정 이용자들을 위한 새로운 서비스를 설계할 수도 있을 것이다. 또한 개별자료의 형태를 중심으로 한 새로운 관계는 소속된 컬렉션과는 상관없이 관련 자료들과의 연계를 통해 제시될 수 있을 것이다. 만약 한 기관이 특정 주제에 해당하는 자원들 중에서 동영상으로 표현된 것만을 제공하는 서비스를 기획한다면 <표 4>의 '개별자료' 클래스의 '자료유형' 속성을 중심으로 데이터를 수집하면 된다. 마지막의 새로운 관계 창출은 컬렉션 맵을 작성하는 것으로 표현된다. 컬렉션 맵은 컬렉션과 하위컬렉션에 대한 계층적인 구조를 설명할 수 있으며, 컬렉션과 특정 기관이나 개인과의 관계를 제시할 수도 있다.

5. 결론

웹이 일반화되어 정보의 홍수에서 살고 있는 이용자들은 많은 자원을 확보할 수 있다는 장점과 유용한 자원인가를 확인해야한다는 부담감을 동시에 가지고 있다. 또한 정보를 다루는 기관에서도 다량의 자원에 대한 확보의 즐거움과 그에 따른 수집정책 및 관리 방안에 대한 부담감을 동시에 가지게 되었다.

특히 네트워크 환경의 장점을 고려한 외부 기관과의 연계는 모든 정보를 다루는 기관의

기회이며 도전이 되고 있다. 그렇지만 각 기관에서 제공하는 자원 표현 기술의 정밀도에 따라 자원 공유에 대한 효율적인 실행의 성공은 결정된다. 자원을 표현하는 기술은 자원 자체를 표현하는 기술요소와 자원을 둘러싼 상황정보를 표현하는 기술요소로 나누어 질 수 있다. 자원 자체를 표현하는 기술은 현재 다양하게 제시되고 있는데 반해, 상황정보를 제시하는 기술요소에 대한 관련 연구는 미비한 상태이다. 이에 본 연구에서는 자원의 상황정보를 제시할 수 있는 기술을 컬렉션 단위 기술(CLD)로 규정하고 CLD의 장점과 관련 기술 표준들을 살펴보았다.

먼저 CLD의 특성을 살펴보면 다음과 같다. 첫 번째 특성은 경제성이다. CLD는 컬렉션 단위로 부여하기 때문에 한 번의 기술로 대규모의 컬렉션에 대한 설명이 가능하다. 컬렉션을 구성하는 개별자원에 대한 기술은 독립적인 것에 반해, 한 번의 작성으로 가능한 컬렉션 기술은 매우 경제적이라고 할 수 있겠다. 두 번째 특성은 대표성이다. CLD는 하나의 컬렉션에 대해 고유하고 특별한 정보만을 전달하는 것이며, 이와 동시에 분산된 기관들을 구별하기 위한 근거로 제공된다. 세 번째 특성은 상호운용성이다. CLD에 대한 연구 배경이 분산된 환경에서 자원을 공유하기 위한 노력의 시작이었듯이, CLD는 이질의 정보시스템에서 제공되는 자원들을 하나의 틀로 관리할 수 있도록 설계되었다.

위와 같은 특성을 바탕으로 한 분산 환경에서의 CLD의 기능은 개별자료 단위 기술(ILD)에서 제공할 수 없는 상황정보를 제시할 수 있다는 것이었다. ILD은 자관의 환경에 맞게 부

여되는 기술 요소로, 개별자료 자체에 대한 설명을 위주로 하고 있으며, 특히 중요한 주제 정보가 생략되기 쉬었다. CLD는 ILD들의 공통적인 특성들을 제시하는 것으로 특정 주제나 관련기관 및 개인에 대한 정보 등을 제시할 수 있었다. 이에 본 연구에서는 기존의 CLD 표준들에서 부족한 기능들을 강화하기 위해 온톨로지를 기반으로 한 CLD 모형을 제시하였다.

온톨로지를 기반으로 한 CLD 모형은 새로운 관계를 창출하기 위한 목적으로 설계되었다. RSLP CD와 NCD를 참고하여 클래스, 데이터 타입 속성, 오브젝트타입 속성을 설계하였으며, 다음과 같은 장점을 가지고 있었다.

첫 번째는 개별자료에 대한 정보를 다차원적으로 표현할 수 있다는 것이다. 이는 개별자료 자체의 기술뿐만 아니라 다른 개별자료와의 관

계를 제공하였다. 두 번째는 컬렉션과 관련된 구축목적, 범위, 해당 개체 및 개별자료, 주제나 이벤트 등 다양한 표현들을 확보할 수 있다는 것이다. 세 번째는 새로운 관계를 창출할 수 있다는 것이다. 이점은 온톨로지의 우수한 특성이 반영된 것으로 주제 또는 이벤트 중심으로 표현이 가능하며, 서비스 모형에 맞춘 컬렉션 제공이 가능하였다. 또한 개별자료의 형태를 중심으로 컬렉션을 재조정할 수 있으며 컬렉션 맵을 통해 컬렉션과 하위컬렉션에 대한 계층적인 구조, 컬렉션과 특정기관이나 개인과의 관계를 제시할 수도 있었다.

앞으로의 향후 연구에서는 지금 제시한 모형을 기반으로 온톨로지 시스템을 구축하고 그에 맞는 이용자 평가실험을 제안하고자 한다.

참 고 문 헌

- 송영희, 노진영, 권은경, 윤희영. 2005. 『디지털시대의 장서관리』. 서울: 한국도서관협회.
- 윤정욱. 2006. 미국, 영국, 호주 및 캐나다 대학교서관의 발전전략의 분석. 『한국도서관·정보학회지』, 37(1): 55-82.
- 윤희운. 2007. 『장서관리론』. 대구: 태일사.
- 이성숙. 2005. 디지털도서관에서 분산자원 검색을 위한 장서 정보에 관한 연구. 『한국문헌정보학회지』, 39(2): 185-209.
- Biodiversity Information Standards. 2008. *Natural Collections Descriptions (NCD)*:

A data standard for exchanging data describing natural history collections. [cited 2008.7.15].

<<http://www.tdwg.org/standards/>>.

Biodiversity Information Standards. *Natural Collection Description Ontology: Migrating NCD from XML schema to OWL.* [cited 2008.7.14].

<http://wiki.tdwg.org/twiki/pub/NCD/NCDWorkShop/NCD_ontology.pdf>.

Brenner, Michaela, T. Larsen, and C. Weston.

2006. "Digital Collection Management through the Library Catalog." *Information Technology and Libraries*, 25(2): 65-77.
- Calanag, M. L., K. Tabata, and S. Sugimoto. 2004. "Linking Preservation Metadata and Collection Management Policies." *Collection Building*, 23(2): 56 - 63.
- Dempsey, Lorcan. 2006. "The Library Catalogue in the New Discovery Environment: Some Thoughts." *ARIADNE*. [online], 48. [cited 2008.6.25]. <<http://www.ariadne.ac.uk/issue48/dempsey/>>.
- Dublin Core Collection Description Working Group. 2006. *Dublin Core Collection Description Application Profile*. [cited 2008.6.1]. <<http://www.ukoln.ac.uk/metadata/dcmi/collection-application-profile/>>.
- Foulonneau, Muriel, Timothy W. Cole, Thomas G. Habing, and Sarah L. Shreeves. 2005. "Using collection descriptions to enhance an aggregation of harvested item-level metadata." *Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries*, 32-41.
- Heaney, Micheal. 2000. *Analytical Model Collections and their Catalogues*. [cited 2008.7.23]. <<http://www.ukoln.ac.uk/metadata/rsrp/model/amcc-v31.pdf>>.
- Kemp, Zarine., L. Tan, and J. Whalley. 2007. "Interoperability for Geospatial Analysis: a Semantics and Ontology-based Approach." *In Proc. Eighteenth Australasian Database Conference(ADC 2007)*, edited by Bailey, J. and A. Fekete. 83-92. Ballarat, Australia: CRPIT. [cited 2008.7.1]. <<http://crpit.com/confpapers/CRPITV63Kemp.pdf>>.
- Lee, Sul H, ed. 2005. *Collection Management and Strategic Access to Digital Resources: the New Challenges for Research Libraries*. Binghamton, NY: Haworth Information Press.
- LeFurgy, William G. 2002. "Levels of Service for Digital Repositories." *D-Lib Magazine*. [online], 8(5). [cited 2008.6.19]. <<http://www.dlib.org/dlib/may02/lefurgy/05lefurgy.html>>.
- Lourdi, I. and C. Papatheodorou. 2006. "Metadata Policies for the Description of Digital Folklore Collections." *Third International Conference of Museology & Annual Conference of AVICOM*. [cited 2008.7.12]. <dlib.ionio.gr/pubs/museology2006_lourdi.pdf >.
- Macgregor, George. 2003. "Collection-level descriptions: metadata of the future?" *Library Review*, 52(6): 247-250.
- Powell, A. 2000. *RSLP Collection Description: Collection Description Schema*. [cited 2008.6.1]. <<http://www.ukoln.ac.uk/metadata/rsrp/schema/>>.
- Powell, H. and L. Dempsey 2000. RSLP Collection Description. *D-Lib Magazine*. [online], 6(9). [cited 2008.6.3].

- <<http://www.dlib.org/dlib/september00/powell/09powell.html>>
- Van Veen, T. and R. Clayphan. 2002. "Meta-data in the Context of The European Library Project." *Proc. Int. Conf. on Dublin Core and Metadata for e-Communities*, 19-26. [cited 2008.7.2.2] <<http://www.bncf.net/dc2002/program/ft/paper2.pdf>>.
- Qin, J. and C. Finneran. 2002. "Ontological representation for learning objects." In *Proceedings of the Workshop on Documents Search Interface Design and Intelligent Access in Large-Scale Collections*, JCDL'02, Portland. [cited 2008.8.2] <xtasy.slis.indiana.edu/jcdlui/papers/qin.pdf>.
- Xin, Wang. 2004. "Research and Usage of Collection Level Metadata in Chinese Digital Libraries." *The International Information & Library Review*, 36(4): 291-295.