

## 디지털도서관 상호운용성 관련 표준 및 기술

충남대학교 이만호\*

### 1. 서론

분산컴퓨팅 환경은 개별적으로 운용되는 여러 컴퓨터 시스템들을 이용하여 이들이 상호 협력을 유지하며 일관성 있고 편리한 서비스를 사용자에게 제공할 수 있는 환경을 말한다. 이런 환경은 상호 협력 관계에 참여하고 있는 다양한 컴퓨터 시스템들이 몇 가지의 기술적 표준을 공동으로 적용하고 있어야 한다. 상호 정보 교환을 필요로 하는 디지털도서관 시스템들은 분산컴퓨팅 환경의 좋은 예이다.

디지털도서관은 교육기관, 연구기관, 정부기관, 기업체 등과 같은 다양한 형태의 기관에서 운영하고 있다. 이들은 소장하고 있는 컬렉션의 내용, 제공하고 있는 서비스의 종류 등에서 다른 디지털도서관과 구별되고 있고, 사용자가 필요로 하는 모든 정보를 다 갖추고 있는 디지털도서관은 거의 없다고 볼 수 있다. 따라서, 사용자들은 자신들이 필요로 하는 정보를 얻기 위해서 여러 디지털도서관에서 제공하고 있는 컬렉션과 서비스를 이용하여 필요한 정보를 얻어야 할 필요가 있다. 그렇다면, 사용자들은 자신이 필요로 하는 정보를 가지고 있을 수 있는 여러 디지털도서관을 어떻게 알아내서 어떤 방법으로 자신이 필요로 하는 정보를 얻을 수 있을 지가 문제이다. 그렇게 하기 위해서는 많은 디지털도서관을 알고 있어야 하며, 이들이 필요로 하는 접근 방식을 익숙하게 알고 있어서 그 방법에 따라 자신이 필요로 하는 정보를 얻을 수 있을 것이다. 하지만, 이런 방식의 작업은 그다지 효율적이라고 말할 수 없을 것이다. 디지털도서관 초기부터 이런 문제를 고려하여, 여러 디지털도서관을 연계하여 하나의 통합된 환경 하에서 사용자가 하나의 접근 방식을 이용하여 여러 디지털도서관

을 동시에 검색할 수 있는 환경 구축에 많은 노력을 해 왔다.

여러 디지털도서관이 참여하여 통합된 환경을 제공하는 형태로는, 비슷한 내용의 컬렉션을 제공하는 디지털도서관을 연계하는 형태와, 이질적 내용의 컬렉션을 제공하는 디지털도서관을 연계하는 형태가 있다. 전자의 대표적인 것으로는 NDLTD(Networked Digital Library of Theses and Dissertations)[13]와 NCSTRL(Networked Computer Science Technical Reference Library)[12]이 있고, 후자의 대표적인 것으로는 NSDL(National SMETE Digital Library, SMETE는 Science, Mathematics, Engineering and Technology Education을 의미한다.)[8,14]이 있다. NDLTD는 학위 논문을 컬렉션으로 가지고 서비스하는 전 세계의 많은 대학의 디지털도서관이 참여하고 있고, NCSTRL은 컴퓨터 분야의 테크니컬 레포트를 컬렉션으로 가지고 서비스하는 기관들이 참여하고 있다. NSDL은 기존에 교육 목적으로 개발되어 있는 과학 및 공학 분야의 여러 디지털도서관을 하나의 통합된 환경으로 하여 다양한 서비스를 제공하는 디지털도서관 시스템을 구축하는 프로젝트로서 미국의 NSF가 주도하고 있다.

어떤 형태이든지, 여러 디지털도서관이 참여하여 통합된 환경을 제공하려면 상호간에 정보 교환이 원만하게 이루어져야 한다. 이런 운용 방식의 특성을 상호운용성(interoperability)이라고 한다. 본 고에서는 디지털도서관 사이의 상호운용성을 보장하기 위해 필요한 기술과 표준 관점에서 고찰하고자 한다.

### 2. 상호운용성

다양한 형태의 기관에서 운영하고 있는 디지털도서관 시스템은 각각 다음과 같은 면에서 다른 시스템

\* 종신회원

과 구별된다; 가지고 있는 컬렉션의 종류, 제공하는 서비스, 운영 기관의 조직, 관리 방식, 사용된 표준, 적용 기술, 등. 이들 특성들은 시간이 지남에 따라, 정보의 양이 늘어나면서 컬렉션이 증가하고, 사용자의 요구가 다양해짐에 따라 새로운 서비스가 개발되기도 하며, 새로운 표준의 등장, 신 기술의 개발 등과 같은 요인으로 꾸준히 변화하고 있기도 하다.

이와 같이 다양한 특성을 가지는 여러 디지털도서관을 연계해서 사용자에게 통일된 서비스를 제공하는 환경을 구축하고자 하는 것이 상호운용성이 추구하는 목표이다[2]. 이러한 목표를 달성하기 위해서는 하드웨어, 데이터 포맷, 프로토콜과 같은 기술적 문제, 언어, 메타데이터, 시각적 표현, 시멘틱 등과 같은 정보 표현 문제, 그리고, 기관간의 협력, 법적 권리와 책임, 문화적 차이 등과 같은 사회적 문제가 해결되어야 한다[3]. 이들 문제점을 기반으로 하여 통합 환경에 참여하는 디지털도서관 사이에 다음과 같은 합의가 이루어져야 한다[2].

- 기술의 합의 : 정보 교환을 위한 포맷, 프로토콜, 보안 시스템 등
- 내용의 합의 : 정보의 표현을 위한 데이터, 메타데이터, 시멘틱 등
- 조직의 합의 : 접근 규칙, 컬렉션의 저장, 과금, 인증 등

그러나, 여러 디지털도서관을 운영하는 기관들 사이에 이와 같은 합의가 이루어 진다면 상호운용성은 훌륭하게 보장될 수가 있다. 하지만, 일반적으로 이들 합의가 포괄적으로 이루어지기는 매우 어려워, 각 기관에서 이들과 같은 합의 사항의 일부라도 받아들일 수 있도록 동기를 부여하는 것이 중요하다고 한다.

상호운용성 보장을 위해 비교적 용이하게 합의할 수 있는 것이 표준화이다. 각 기관이 상호운용성을 보장하기 위한 표준을 받아 들이도록 하는데 영향을 미치는 중요한 요인을 비용과 기능면에서 설명할 수 있다. 표준을 채택함으로써 많은 비용이 소요된다면, 얻을 수 있는 기능이 꼭 필요한 소수의 기관만이 표준을 채택할 것이다. 반대로 적은 비용이 소요된다면, 비록 기능이 떨어진다고 해도 많은 기관이 채택할 것이다. 이와 같이 비용과 기능 면에서 장단점이 발생하게 되는데, 바람직한 경우는 적은 비용으로 강력한 기능을 가지는 디지털도서관을 제공하는 경우가 될 것이다. 이와 같은 경향을 그래프로 표현하면 그림 1과 같다[2]. 그림 1에서 세로축은 표준을 채택

함으로써 기관이 지불해야 하는 비용이고, 가로축은 표준을 채택함으로써 기관이 얻을 수 있는 기능이다.

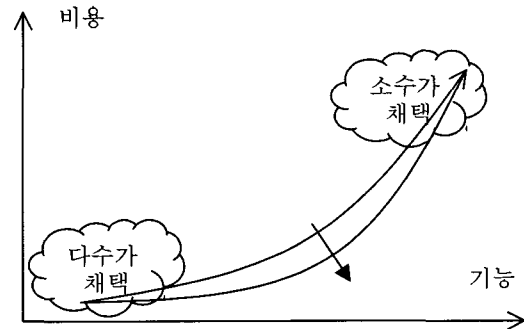


그림 1 표준 채택의 비용과 기능

디지털도서관 관련 기술은 꾸준히 개발되고 있고 표준도 계속해서 새로운 표준이 제정되고 있기 때문에, 그림 1에 나타나는 곡선의 모양과 위치는 다양하게 나타날 수 있다. 높은 비용과 강력한 기능 방향으로 이행하는 여러 표준에 대해서 곡선의 형태가 아래에 위치하도록 하는 표준이 있다면, 그것을 채택함으로써 적은 비용으로 강력한 기능을 얻을 수 있게 되어, 많은 기관이 그 표준을 채택하게 될 것이다.

이후에는 상호운용성의 보장을 위해서 널리 사용되는 여러 가지 표준에 대하여 논의하게 되는데, 각 표준의 평가는 그림 1의 비용과 기능 관점에서 다루어진다.

### 3. 데이터의 표현

인간의 의사를 기계를 이용해 전달하기 시작하면서 정보를 표현하기 위해 여러 가지 형태로 코드를 사용해 왔다. 컴퓨터를 이용하기 시작하면서부터는 문자를 표현하기 위해 ASCII를 사용하여 문서를 작성하였다. 오늘날 ASCII를 사용한 정보의 표현은 가장 단순하고 가장 널리 이용되고 있다. ASCII로 작성된 문서가 제공할 수 있는 기능은 매우 제한적이어서 거의 없다고 볼 수 있다. 하지만 ASCII를 사용하기 위한 비용은 거의 필요하지 않기 때문에 모든 시스템이 지원하고 있다.

ASCII는 영어 문자를 표현하기에 적합한 코드이므로, 영어 이외의 외국어 문자를 표현하기에는 부적절하다. 영어 이외의 외국어를 표현하기 위해서, 각 언어별로 ASCII를 변형한 코드를 이용해 그들의 문

자를 표현해 왔다. 이런 방식으로는 언어가 다른 문서 사이의 정보 교환에는 해결해야 할 문제들이 많이 존재하기 때문에 상호운용성이 현저하게 떨어진다. 이에 대한 해결방법으로 전세계에서 사용되는 모든 문자를 표현할 수 있는 문자 표준으로 Unicode가 이용된다[15]. Unicode를 사용하면 전 세계의 모든 언어가 충돌 없이 표현될 수 있다. 또한 Unicode의 인코딩 방식으로 매우 효과가 좋은 UTF-8이 제안되어 문자로 인한 상호운용성은 현저하게 향상되었다.

ASCII를 이용한 정보 표현과는 대조적으로, 매우 많은 정보를 포함하고 있으며 정보의 표현 방법도 매우 융통성 있게 활용할 수 있는 SGML 언어가 있다. SGML은 기능이 매우 강력하여 오랜 동안 문서 교환의 표준으로 자리잡아 왔다. 그러나, 배우기가 어려울 뿐만 아니라 이용하기도 쉽지 않은 단점이 있어, 널리 사용되지는 못하고, 극히 일부 전문가들 사이에서만 사용되어 왔다. 이런 점은 사용하기 어려운 단점 보다는 기능상의 장점이 훨씬 큰 경우에만 해당된다.

SGML로부터 발생된 HTML 언어는 배우기 쉽고 사용하는데 적은 비용이 들어 많은 사람들이 널리 사용할 수 있는 기본적인 조건은 갖추었다. 때마침 도래한 인터넷 환경에서 제공할 수 있는 기능도 많아, 모든 브라우저가 지원하고 있으며, 전 세계적으로 널리 사용되는 표준 언어가 되었다. 하지만 SGML에 비하면 기능 면에서 현저하게 떨어지고, 인터넷 이외의 문서 교환 방식의 표준으로서는 부적절하다.

XML 언어는 SGML의 기능 일부를 제한하여 만들어진 표준이지만, SGML과 거의 비슷한 기능을 가지고 있어 문서 교환을 위한 표준으로 널리 자리잡아가고 있다. XML이 점차 널리 사용됨에 따라, XSL, XPath, XLink, XQuery 등과 같은 XML 관련 표준들이 속속 제안되고 있어 XML은 적은 비용에 더욱 강력한 기능을 갖추게 되었다. 이런 추세대로 나간다면 문서 교환을 위한 표준으로 XML이 차지하는 비중이 절대적이 될 것이고, 그렇게 되다면 문서 교환으로 인한 상호운용성 문제는 완전히 해결될 것이다.

위에서 열거한 데이터의 표현 방식과는 약간 다른 개념이지만, PDF(Portable Document Format)로 문서를 표현할 수 있다. PDF는 문서를 페이지 단위로 이미지로 만들어 저장하고 표현할 수 있도록 개발된 페이지 기술 언어이다. 디지털도서관은 문서를 디지털화하여 저장한다. 문서를 디지털화한 형태에는 워드프로세서로 작성된 문서, 이미지로 스캔한 문서,

XML 문서 등과 같은 여러 가지 형태가 있을 수 있다. 이런 다양한 형태의 문서는 해당 문서 형태를 위한 특별한 처리기가 있어야 사용자가 볼 수 있다. 그러나, PDF 문서는 Adobe사가 무료로 제공하는 Acrobat reader를 다운 받아 설치하면 어떤 환경에서도 읽을 수 있다. 그러므로 PDF 문서 작성기를 사용하여 모든 문서를 PDF 문서로 변환해 두면 누구나 환경에 구애 받지 않고 읽을 수 있게 된다. 이런 면으로 인해서 사용자에게 제공하는 문서는 PDF 문서로 제공하는 디지털도서관이 매우 많다.

#### 4. 메타데이터

메타데이터는 문서의 특성을 기술하기 위한 데이터로서, 각 디지털도서관마다 선호하는 메타데이터가 사용되고 있다. 메타데이터를 서로 교환하는 경우에 상호운용성을 보장하기 위해서는 메타데이터의 형태에 대한 약속이 있어야 한다. 이와 같이 원래의 메타데이터 형태를 원시 메타데이터라고 하고, 메타데이터 교환을 위한 형태를 정규화 메타데이터라고 하자.

메타데이터 교환이 필요할 경우, 원시 메타데이터가 정규화 메타데이터와 다르다면, 원시 메타데이터를 정규화 메타데이터로 변환하여 전송해야 하고, 전송 받은 메타데이터는 역으로 원시 메타데이터로 변환하여 처리해야 한다.

정규화 메타데이터로는 주로 더블린코어(Dublin Core) 메타데이터가 사용된다[9]. 더블린코어 메타데이터는 표현이 간단하고 확장성이 뛰어나 상호운용성을 보장하기 위한 메타데이터로 많이 사용된다. 실제로 OAI 프로토콜에서 메타데이터를 수확할 때 더블린코어 메타데이터 형태로 전송이 일어난다. 또한, NDLTD에서는 학위 논문을 ETD(Electronic Theses and Dissertations) 표준에 따라 작성하고, 메타데이터는 ETD-MS(ETD-Metadatas Standard)를 따른다[10]. ETD-MS는 더블린코어를 기반으로 하여 학위 논문의 특성을 반영할 수 있도록 일부 확장한 것이다.

XML이 문서 교환의 표준으로 널리 사용됨에 따라, 메타데이터도 XML을 사용하여 표현하고 있는 것이 보통이다.

#### 5. 정보 교환 방식

디지털도서관 하나로서는 정보를 요구하는 사용자들의 요청을 충족시킬 수 있을 정도로 충분한 양의 컬렉션을 가지고있을 수가 없으므로 디지털도서관 사이의 정보 교환이 필수적이다. 정보 교환을 위해서는 정보 교환을 위한 약속을 기술한 프로토콜이 있어야 하고 이런 프로토콜을 지원하는 디지털도서관 사이에 정보 교환이 가능한 것이다.

정보 교환은 단순수집(gathering) 방식, 연합(federation) 방식, 수확(harvesting) 방식의 세 가지 수준에서 이루어진다[2].

단순수집 방식은 정보 교환의 가장 기초적인 형태로서, 정보 교환을 위한 어떤 공식적인 협정도 필요 없고, 단지 인터넷 상에 공개된 정보만을 대상으로 정보를 수집하여 서비스하는 방식이다. 이 방식은 정보 수집에 비용이 들지 않으나, 대단히 많은 사이트를 대상으로 하고 있으므로 수집된 정보는 질적으로 떨어질 수 밖에 없으나, 공식적인 협력 관계 없이 많은 디지털도서관의 정보를 수집할 수 있다는 장점이 있다. 따라서, 이런 방식을 사용하는 예가 매우 많고, 웹을 이용한 검색엔진이 이런 방식의 전형적인 예이다.

디지털도서관의 상호운용성을 말할 때 습관적으로 떠올리는 방식이 연합 방식이다. 이 방식은 참여하는 기관들이 협약에 의해 상호 동의하는 서비스를 정의하고 특정 프로토콜을 사용하여 구현한다. 대표적인 예로 Z39.50을 사용하는 기관은 이 프로토콜에 의해 모든 정보 교환을 수행하게 되는데 이런 연합 방식을 구성하고 구현하기 위해서는 상당한 비용의 초기 투자가 필요하며 지속적인 변화를 수용해야 하는 추가 부담이 따른다. 반면에 연합 방식을 통해 상호운용성을 추구하는 경우 매우 고 품질의 서비스를 제공할 수 있다. 따라서 이런 방식은 소수의 회원 기관들 중심으로 운영되고, 많은 기관들이 회원으로 참여하기는 어려운 방식이다. 기존의 많은 디지털도서관 시스템들이 이런 방식에 속하고, NDLTD, NCSTRL등을 비롯하여, NSDL 구축에 참여하는 많은 시스템들도 이 방식에 속한다.

수확 방식은 OAI 프로토콜에서 사용하는 방식으로서 단순수집 방식과 연합 방식의 중간 수준에 있다고 볼 수 있다. 수확 방식의 기본 개념은 데이터제공자와 서비스제공자를 분리하여 독립적으로 운용하도록 하는 것이다. 이런 개념은 기존의 디지털도서관들이 데이터 제공자이면서 동시에 서비스도 제공하는 구조와는 다른 형태이다. 기존의 디지털도서관 시스

템들은 자신이 데이터 컬렉션을 가지고 있으므로 사용자가 요청하는 정보를 즉시 서비스할 수 있다. OAI 프로토콜에서 서비스제공자는 여러 데이터제공자로부터 정보를 가져와서 사용자의 요청을 처리하는 서비스를 제공해야 한다. 서비스제공자는 여러 데이터제공자로부터 정보를 가져와야 하므로 각 데이터제공자가 가지고 있는 모든 컬렉션을 가져와 저장한다는 것은 문제가 있다. 그래서, 데이터제공자는 자신이 가지고 있는 컬렉션에 대한 메타데이터를 구축하고, 이를 서비스제공자가 가져갈 수 있도록 서버를 운영한다. 서비스제공자는 관심 있는 데이터제공자로부터 정기적으로 또는 필요에 따라 메타데이터를 가져와서 메타데이터만으로 구성된 데이터베이스를 구축하고 사용자 질의를 받아 처리한다. 이와 같이 서비스제공자가 데이터제공자로부터 메타데이터를 가져오는 것을 수확(harvest)이라고 한다. 메타데이터를 수확할 때, 서비스제공자의 특성에 따라 선택적으로 메타데이터를 가져올 수 있다. 사용자는 서비스제공자에 접속하여 검색을 수행한 후 관심 있는 정보에 대해서는 원문을 직접 데이터제공자에게 요청할 수도 있어, 단순히 메타데이터 검색에 그치지 않고 다른 시스템과 비교해도 서비스의 질적 수준이 저하되지 않는다.

단순수집 방식은 어떤 협정도 필요하지 않고, 연합 방식은 공식적인 협정이 필요할 뿐만 아니라 복잡하고 비용이 많이 요구되는 Z39.50과 같은 프로토콜을 준수해야 한다. 반면에 수확 방식은 매우 단순하지만 강력한 기능을 제공할 수 있는 OAI 프로토콜만 구현한다면 어떤 협정이나 계약도 필요 없다. 컴퓨터 시스템에 대한 제약, 운영 환경에 대한 제약도 없으므로, 이기종, 이질적 환경 사이의 상호운용성 면에서는 어떤 방식 보다도 우수하다고 볼 수 있다.

## 6. 통신 프로토콜

디지털도서관에서 상호운용성은 클라이언트가 디지털도서관 시스템의 서버에 연결하여 메시지를 보내고, 사용자가 요구하는 정보를 서버에서 클라이언트로 보낼 수 있는 통신 프로토콜이 필요하다.

인터넷에서 정보 교환 프로토콜로 가장 많이 사용되는 것이 HTTP이다[11]. HTTP는 웹 브라우저와 웹 서버 사이에서 메시지를 주고 받는 프로토콜로서, 먼저 웹 브라우저와 웹 서버 사이를 연결하고, 웹 서

버로부터 웹 브라우저로 파일을 복사해 오거나, 웹 서버의 내용을 변경하는 작업을 한 후에, 연결을 종료하는 방식으로 동작하는 프로토콜이다. 이런 프로토콜은 그 기능이 복사, 저장, 삭제 등과 같은 단순한 작업을 일회성으로 끝내고, 계속적인 세션 작업이 이루어지지 않는 등 매우 단순하다. 하지만, HTTP는 모든 웹 브라우저가 지원하고 있는 기능이고, 웹 브라우저는 무료로 사용할 수 있기 때문에, 이것을 이용하는 데는 비용이 전혀 들지 않는다. 따라서, 비록 기능은 단순해도, 비용이 들지 않기 때문에 가장 많이 사용되고 있는 프로토콜이다.

디지털도서관 시스템의 상호운용성을 위해서 가장 많이 사용되는 프로토콜은 Z39.50으로 알려져 있다[6]. Z39.50은 도서관 연합체에서 개발한 것으로서 디지털도서관의 목적에 잘 부합된다. Z39.50에서 클라이언트는 서버와 연결하고, 대화식으로 세션이라고 부르는 일련의 작업을 진행할 수 있다. 클라이언트가 필요로 하는 모든 작업이 끝났을 때, 클라이언트와 서버 사이의 연결을 종료한다. 여기서, Z39.50의 중요한 면은 세션을 유지할 수 있다는 것이다. 세션이 유지되는 동안에 진행되는 모든 작업은 그 상태를 기억할 수 있어 대단히 많은 기능을 구현할 수 있다. 또한, Z39.50은 서버가 지원하는 데이터베이스의 종류에 구애 받지 않으며, 사용자 인터페이스나 클라이언트와의 연결 방법에 대한 아무런 제약을 두지 않고 있다. 이런 면들은 Z39.50이 상호운용성 면에서 매우 뛰어나다고 볼 수 있다.

Z39.50이 기능 면에서 매우 뛰어나지만 이를 구현하는 데는 많은 비용이 요구된다. 적은 비용에 저렴한 기능을 제공하는 HTTP와는 정 반대로 많은 비용에 강력한 기능을 제공하는 프로토콜이 Z39.50이다. 디지털도서관에서는 기능이 강력하기 때문에 비록 많은 비용이 요구된다고 해도, 대부분의 디지털도서관 시스템이 Z39.50을 지원하고 있는 상태이다. 디지털도서관 시스템에 많은 예산을 투자할 수 없는 작은 기관에서는 이런 고비용의 프로토콜을 채택하기는 쉽지 않다.

그림 1의 곡선에서 보듯이 비교적 적은 비용으로 비교적 강력한 기능을 제공할 수 있는 프로토콜에 대한 연구 결과로서 출현한 프로토콜이 있는데, 그것이 바로 OAI(Open Archives Initiative)이다[4, 5].

OAI의 구조는 그림 2에서 보는 바와 같이, 데이터를 가지고 있는 데이터저장소와 이들로부터 데이터를

를 수확해 오는 데이터수확기라는 두 개의 구성 요소를 필요로 한다. OAI에서는 데이터 저장소 기능을 하는 것을 데이터제공자(data provider)라고 하고, 데이터 수확 기능을 하는 것을 서비스제공자(service provider)라고 한다. 데이터제공자는 데이터 컬렉션을 가지고 있으면서 이를 제공할 수 있는 디지털도서관으로서, 자신이 가지고 있는 문서에 대한 메타데이터를 서비스제공자가 이용할 수 있도록 한다. 서비스제공자는 여러 데이터제공자로부터 메타데이터를 수확(harvesting)해 와서 이들을 사용자가 검색할 수 있도록 하는 서비스를 제공한다.

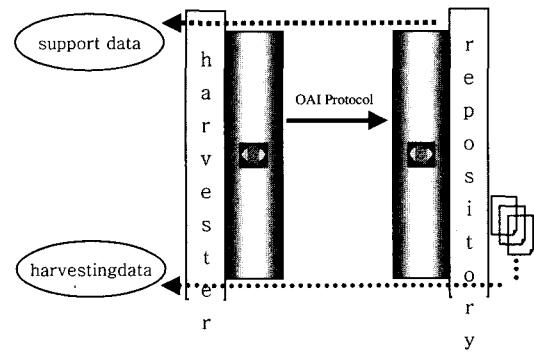


그림 2 OAI 구조

이들 두 구성 요소간의 프로토콜은 지원프로토콜(supporting protocol)과 수확프로토콜(harvesting protocol)로 크게 구 가지로 구분된다. 지원프로토콜은 서비스제공자가 데이터를 수확해 올 데이터제공자의 특성을 파악하기 위해 데이터제공자에게 요청하는 것으로, 데이터제공자는 요청 종류에 따라 해당 정보를 서비스제공자에게 보낸다. 수확프로토콜은 서비스제공자가 데이터제공자가 가지고 있는 메타데이터 정보를 요청하는 것으로, 데이터제공자는 요청된 메타데이터 정보를 서비스제공자에게 보낸다.

지원프로토콜에는 다음과 같은 세 가지의 요청이 있다.

- Identity : 데이터제공자를 인식할 수 있는 정보를 요청한다.
- ListMetadataFormats : 데이터제공자가 지원하고 있는 메타데이터에 대한 정보를 요청한다.
- ListSets : 데이터제공자가 가지고 있는 데이터들을 분류한 상태인 분류정보를 요청한다.

수확프로토콜에는 다음과 같은 세 가지의 요청이

있다.

-GetRecord : 데이터제공자가 가지고 있는 메타데이터 레코드 중에서 하나를 지정하여 요청한다.

-ListIdentifiers : 데이터제공자가 가지고 있는 메타데이터 레코드들의 identifier 리스트를 요청한다. 이때 조건을 지정할 수 있다.

-ListRecords : 데이터제공자가 가지고 있는 메타데이터 레코드들을 요청한다. 이때 조건을 지정할 수 있다.

위에 열거한 프로토콜이 OAI에서 사용되는 프로토콜 전체이다. 매우 간단하지만 어떤 정보라 하더라도 이들 프로토콜을 이용하여 데이터제공자로부터 가져올 수 있을 정도로 강력하다. 프로토콜 자체가 간단하므로, 매우 적은 비용으로 짧은 시간 안에 이를 구현할 수 있다. 그러므로, OAI 그룹에 가입한 어떤 디지털도서관 시스템을 가지고 데이터제공자가 될 수가 있고, 이미 가입된 수 많은 데이터제공자로부터 메타데이터를 수확하여 서비스제공자가 될 수도 있다.

OAI 개념은 1999년 10월에 Santa Fe에서 첫 모임을 가진 후, 급속한 속도로 많은 디지털도서관 시스템에 의해서 지원되고 있어, 상호운용성을 위한 표준으로 강력하게 부상하고 있다. 예를 들면, NDLTD, NCSTRL, NSDL 등과 같은 기존의 디지털도서관 시스템들이 모두 OAI를 지원하는 데이터제공자이다. 앞으로 더욱 빠른 기세로 많은 디지털도서관들이 OAI를 지원하게 될 것으로 생각한다.

국내에서 개발되어 활용되고 있는 프로토콜에는 K-프로토콜이 있다. K-프로토콜은 각 디지털도서관 혹은 사이트가 보유하고 있는 컬렉션의 특성을 메타데이터 형태로 서로 교환하여 분산검색을 가능하게 한다[7]. 즉, 각 사이트에서는 자신이 보유하고 있는 문서 컬렉션의 특성 및 시스템의 특성을 메타데이터의 형태로 구성하여 제공하고, 그 이외에 개별 문서에 대한 어떤 작업도 요구하지 않는다. 통합된 검색 서비스를 하고자 하는 경우, 각 사이트에 표준 형식의 질의를 보내면 각 사이트에 설치된 소프트웨어는 이 표준 질의를 각 사이트 별로 요구되는 질의 형식으로 자동 변환하여 검색을 수행한 후 검색 결과를 다시 표준 형식으로 변환하여 통합 서비스 사이트로 전송한다. 각 사이트에서 전송된 결과들은 통합되어 사용자에게 제시된다. 이 방법은 단순한 수준의 상호

운용성 제공 방식이 갖는 정보검색 결과의 품질 저하 문제를 사이트 특성에 대한 메타데이터를 이용하여 극복하면서 각 사이트의 자율권을 보장할 수 있다.

## 7. 상호운용성의 발전 방향

이상으로 상호운용성을 보장하기 위한 여러 가지 표준에 대해 살펴 보았다. 상호운용성을 보장하기 위해서는 적은 비용으로 많은 기능을 얻을 수 있는 표준을 채택하는 것이 이상적이다.

각각의 항목에서 상호운용성을 향상시키기 위해 계속 발전하리라고 생각하는 것들은 다음과 같다.

- 문서는 UTF-8로 문자를 표현하여 XML로 작성한 문서.
- 메타데이터는 더블린코어
- 정보 교환 방식은 수확 방식
- 통신 프로토콜은 OAI 프로토콜.

표준은 기술 발전과 사용자의 요구에 의해서 꾸준히 변화하고 발전해 나간다. 결국 이들 표준은 상호운용성이 더 적은 비용으로 더 많은 기능을 얻을 수 있는 방향으로 발전되어 나갈 것이다.

## 참고문헌

- [1] [Arms-book] W. Y. Arms, Digital Libraries, MIT Press, 2000.
- [2] W. Y. Arms, et al., A Spectrum of Interoperability, D-Lib Magazine, 8 (1), Jan. 2002.
- [3] E. A. Fox, G. Machionini, G., Toward a worldwide digital library, Communications of the ACM, 41(4), Apr. 1998, pp28-32.
- [4] C. Lagoze, H. Van de Sompel (editor), The Open Archives Initiative Protocol for Metadata Harvesting, Protocol Version 2.0 of 2002-06-14.
- [5] C. Lagoze, H. Van de Sompel, The Open Archives Initiative: Building a low-barrier interoperability framework, JCDL '01, June 17-23, 2001, Roanoke, VA.
- [6] C. A. Lynch, et al., The Z39.50 Information Retrieval Standard Part I: A Strategic View of Its Past, Present and Future, D-Lib Magazine, 3(4), Apr. 1977.
- [7] S. H. Myaeng, et al., A Protocol-Based Architecture for Federated Searching in Digital

Libraries, Proc. Of the 4th International Conference of Asian Digital Libraries, Dec. 2001, Bangalore, India.

[8] L. L. Zia, The NSF National, Science, Technology, Engineering, and Mathematics Education Digital Library (NSDL) Program: New Projects and a Project Report, D-Lib Magazine, 7(11), Nov. 2001.

[9] Dublin Core 홈페이지, <<http://dublincore.org/>>

[10] ETD-ms: an Interoperability Metadata Standard for Electronic Theses and Dissertations-version 1.00, <<http://www.ndltd.org/standards/metadata/current.html>>

[11] HTTP 홈페이지, <<http://www.w3.org/Protocols/>>

[12] NCSTRL 홈페이지, <<http://www.ncstrl.org/>>

[13] NDLTD 홈페이지, <<http://www.ndltd.org/>>

[14] NSDL 홈페이지, <<http://www.nsd.org/>>

[15] Unicode 홈페이지, <<http://www.unicode.org/>>

이 만 호



1975 서울대학교 공과대학 응용수학과 (학사)  
 1977 KAIST 전산학과(석사)  
 1991 Indiana University 전산학과(PhD)  
 1977~1980 ADD연구원  
 1980~1984 충남대학교 계산통계학과 조교수  
 1991~현재 충남대학교 정보통신공학부 교수  
 2002~2001 Virginia Tech. Digital Library Research Laboratory 방문

관심분야: 디지털도서관, 디지털컨텐츠  
 E-mail: mhlee@cs.cnu.ac.kr

● 소프트웨어공학 기술 합동 워크샵 2002 ●

- 일 자 : 2002년 8월 22~24일
- 장 소 : 경주 하일라 콘도
- 주 최 : 한국정보과학회 · 한국정보처리학회
- 문 의 처 : 대전대학교 황선명 교수

E-mail: sunhwang@dragon.taejon.ac.kr