

# MIRAGE의 OAI지원 기능을 위한 Data Provider 설계 와 구현

지용인<sup>0</sup>, 이종필, 이현숙, 이만호  
충남대학교 컴퓨터학과  
(jiyongin, jplee, hylee, mhlee)@cs.cnu.ac.kr

## A Design and Implementation of Data Provider for Developing Features Supporting OAI in MIRAGE

Yong-In Ji<sup>0</sup>, Jong-Phil Lee, Hyun-Sook Lee, Mann-Ho Lee  
Dept. of Computer Science, Chungnam National University

### 요 약

MIRAGE는 XML을 기반으로 하는 가상문서라는 개념을 도입하여 인터넷 상의 정보 공유를 가능하게 하여 새로운 지식을 생성할 수 있도록 한 디지털도서관 시스템이다. OAI는 컨텐츠에 대한 메타데이터를 이용하여 디지털도서관 안에 있는 정보를 상호이용하기 위한 프로토콜이다. 본 논문에서는 MIRAGE에 OAI 지원 기능을 첨가하여 MIRAGE 시스템에 의해 만들어진 가상문서에 대한 메타데이터를 제공하는 data provider를 구현하여 다른 디지털도서관과 더불어 정보를 상호이용할 수 있도록 하였다. 또한 원문에 대한 접근을 원하는 사용자를 위해서 메타데이터의 원문에 직접 접근할 수 있도록 하였다.

### 1. 서론

기존의 디지털도서관 시스템의 문제점은 각 디지털도서관마다 접근하기 위한 방법이 다르며, 제공하는 정보의 형태도 다양하다는 것이다. 이러한 문제들로 인해 디지털도서관 사이의 상호이용에 어려움이 있었다. OAI(Open Archive Initiative)는 이러한 디지털도서관 사이의 상호이용의 문제점을 해결하기 위한 규약으로서, 이 규약을 따르는 디지털도서관 사이에는 상호간의 정보 교환이 가능하게 되었다. OAI는 OAI 프로토콜에 따라 각 디지털도서관이 가지고 있는 컨텐츠에 대한 메타데이터를 제공하기 위한 data provider와 여러 data provider들로부터 메타데이터를 수집하고 수집한 정보를 이용해 유용한 서비스를 제공하기 위한 service provider라는 두 개의 프레임워크를 정의하였다. 1999년 10월에 열린 첫 회의에서 규약이 정해진 이후로, 여러 연구 기관에서 이를 구현하여 운용해 보고 그 효과가 인정되어 빠른 속도로 퍼지고 있는 기술이다.

본 논문에서는 인터넷 상의 문서를 일부 또는 전부를 링크로 연결함으로써 새로운 문서를 생성할 수 있는 가상문서라는 개념을 도입한 MIRAGE(Multimedia Information Retrieval And Gathering Environment) 디지털도서관 시스템[Myaeng 99]에 OAI지원 기능을 첨가하여 MIRAGE 시스템에 의해 만들어진 가상문서에 대한 메타데이터를 제공하는 data provider를 설계 및 구현하였다. 이로써 다른 디지털도서관과 더불어 정보를 상호이용할 수 있도록 하였으며, 메타데이터의 원문에 직접 접근할 수 있는 방법도 제공하고 있다.

본 논문의 구성은 다음과 같다. 2절에서는 OAI에 대한 개념과 framework 그리고 protocol에 대해 다루고, 3절에서는 OAI protocol을 지원하는 MIRAGE 시스템 구조와 data provider의 supporting protocol과 harvesting protocol에 대해 설명한다. 4절에서는 결론 및 향후 연구 방향에 대하여 살펴본다.

### 2. OAI(Open Archive Initiative)

#### 2.1 OAI의 정의 및 개념

Repository는 HTTP protocol을 통해 OAI protocol request를 받을 수 있는 네트워크를 통한 접근이 가능한 서버이다. OAI protocol을 따르는 repository들의 메타데이터에 접근할 수 있으며, repository는 하나의 item마다 메타데이터를 제공한다. 이 메타데이터는 record형태로 제공된다. Record는 XML-encoded byte stream 형태로 제공되며, 크게 다음 3가지로 구성되어 있다.

- Header : 모든 Record들에게 공통되고 harvesting process 위해 필요한 정보로서 두개의 부분으로 구성되어 있다.
  - unique identifier : repository 안에 있는 한 item에 대한 metadata를 추출하기 위한 key이며, metadata prefix와 결합하여 한 item에 대해서 지정된 metadata format을 가진 record를 요청하기 위해 사용된다.
  - timestamp : item이 생성, 삭제된 날짜 또는 item의 내용이 변경된 가장 최근의 날짜를 기록
- Metadata : 한 item의 메타데이터에 대한 단 하나의 표현으로 한 record는 하나의 메타데이터 format으로만 표현된다.
- About : record안에 있는 metadata part에 관한 data를 유지하기 위한 optional container이다

Set은 record들에 대해서 선택적으로 수집(harvesting)을 하기 위해 repository안에 있는 item들을 그룹화 하기 위한 optional container이다. 각 repository는 각 repository 나름대로 item들의 계층적 구조를 정의할 수 있으며 계층에서 각 node는 하나의 set이다. repository안에 있는 각 item들은 한 개 이상의 set에 포함될 수 있고, 또는 어떤 set에도 포함되지 않을 수도 있다. set의 의미, 구성에 대한 것은 OAI protocol에 정의 되어있지 않고 각 community에 달려 있다. set은 3

가지 요소로 구성된다.(setTag, setSpec, setName)

2.2 OAI framework

Data provider는 디지털도서관 시스템안에 있는 정보에 대한 메타데이터를 제공하기 위한 방법으로 OAI protocol을 지원하는 시스템이다. 즉, HTTP protocol을 통해서 요청되는 client 질의어가 OAI protocol에 맞으면 질의어를 만족하는 메타데이터를 제공한다. Service provider는 data provider들에게 OAI protocol request를 보내고 그 결과로써 받은 메타데이터를 이용하여 사용자에게 적절한 서비스를 제공하는 시스템이다. 즉, data provider들로부터 수집(harvesting)한 메타데이터를 저장하고 이를 통해 검색과 같은 유용한 서비스를 사용자에게 제공한다.

2.3 OAI protocol request

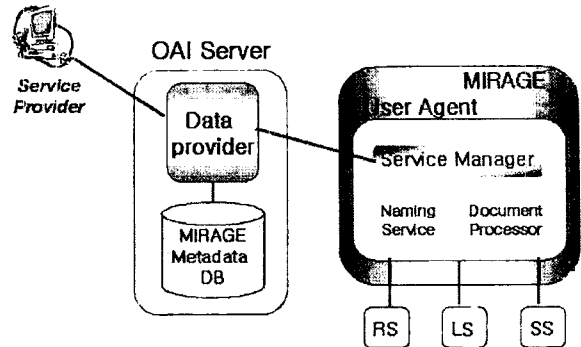
OAI protocol request는 크게 data provider의 자체 정보를 요구하는 supporting protocol request와 data provider가 가지고 있는 메타데이터를 요구하는 harvesting protocol request가 있다. Request시 사용되는 각 verb의 용도 및 기능은 아래의 [표 1]과 같다.

	Verb	용도
Supporting protocol request	Identify	Repository의 일반적인 정보 검색한다.
	ListMetadataFormats	Repository로부터 사용 가능한 metadata formats을 검색한다.
	ListSets	Repository 안에 있는 set의 구조를 검색한다.
Harvesting protocol request	GetRecord	Repository 안에 있는 item으로부터 각각의 record(metadata)을 검색한다.
	ListIdentifiers	Repository로부터 수집할 수 있는 record들의 identifier를 검색한다.
	ListRecords	Repository로부터 record들을 수집한다.

[표 1] OAI protocol features

3. OAI를 지원하는 MIRAGE 시스템

기존의 MIRAGE 시스템이 OAI 기능을 지원하기 위해서는 MIRAGE 시스템이 가지고 있는 컨텐츠에 대한 메타데이터를 제공하는 data provider를 구현하여야 한다. 이를 위해 본 논문에서 설계한 data provider는 MIRAGE 시스템에 있는 정보에 대한 메타데이터를 제공하기 위해 MIRAGE 시스템과 연계하여 동작하여야 한다. MIRAGE 시스템을 통해서 가상문서를 제작하여 저장할 경우에 그 가상문서에 대한 메타데이터 정보가 data provider로 넘겨지며, 이 정보를 데이터베이스에 저장한다. [그림 1]은 OAI 기능을 지원하는 MIRAGE 디지털도서관 시스템의 구조를 나타내고 있다.

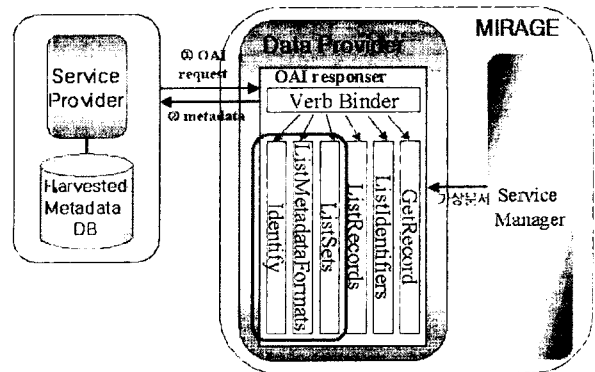


[그림 1] OAI를 지원하는 MIRAGE 시스템 구조

3.1 Data provider 설계 및 구현

3.1.1 Supporting protocol

Service provider는 data provider의 특성을 파악하기 위해 OAI protocol request를 보내게 되는데 data provider가 service provider로부터 받은 OAI protocol에 맞는 질의어(request)를 이해하기 위해 verb binder를 통해 파싱을 하게 된다.



[그림 2] Supporting protocol 구조

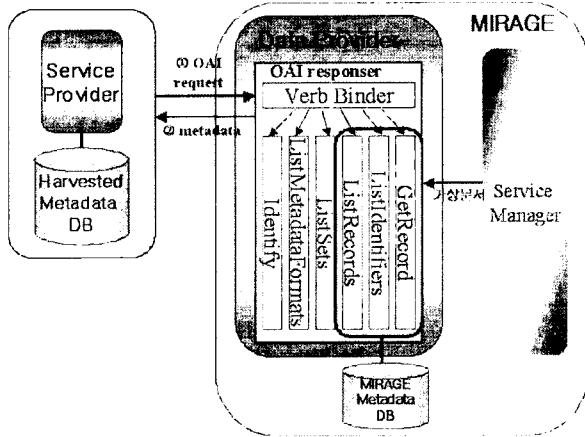
만약 질의어가 Identify, ListMetadataFormats, ListSets등과 같은 supporting protocol일 경우에는 [그림 2]의 구조와 같이 각 질의어를 만족하는 정보를 XML형태로 service provider에게 제공한다. supporting protocol에서 제공되는 것은 repository의 일반적인 정보(admin e-mail, repository name), repository로부터 사용 가능한 메타데이터의 형태에 관한 정보(metadataPrefix, schema) 그리고 set의 구조 등 OAI server내의 일반적인 정보와 정책에 관한 것이다. 다음은 service provider가 data provider의 특성을 파악하기 위해 data provider에게 보내는 request의 한 예이다.

```
http://an.oa.org/OAI-script?verb=Identify
http://an.oa.org/OAI-script?verb=ListMetadataFormats
&identifier=oai:arXiv:hep-th/9901001
http://an.oa.org/OAI-script?verb=ListSets
```

3.1.2 Harvesting protocol

Service provider는 data provider가 가지고 있는 메타데이터를 수집(harvesting)하기 위해 OAI protocol request를 보내고 data provider가 verb binder를 통해 이해한 질의어(request)가 ListIdentifiers, ListIdentifiers, GetRecord등과 같은 harvesting protocol일 경우에는 아래의 [그림 3]구조에서 보는 바와 같이 데이터베이스에 저장되어 있는 MIRAGE 시스템의 가상문서에 관한 정보를 각 질의에 맞게 검색을 수행하여 서비스를 하게 된다.

즉, MIRAGE 시스템의 저작도구(AT)를 이용해서 가상문서를 제작하고 저장 하였을 경우 data provider에도 같이 저장되며, 가상문서에서 메타데이터 부분만 파싱하여 데이터베이스에 저장을 하게 된다. 그리고 각 질의에 맞는 정보를 데이터베이스로부터 검색한 후 XML형태로 변환하여 service provider에게 보내게 된다.



[그림 3] Harvesting protocol 구조

다음은 service provider가 data provider의 메타데이터를 수집하기 위해 data provider에게 보내는 request의 한 예이다.

`http://an.oa.org/OAI--script? verb=GetRecord& identifier=oai:arXiv:quant--ph/9901001& metadataPrefix=oai_dc`

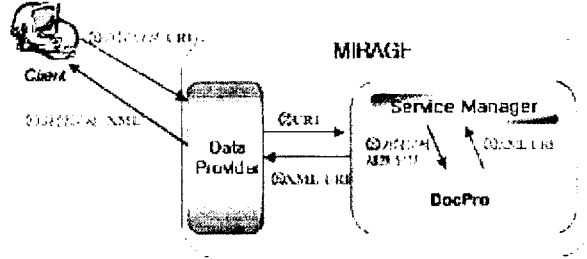
`http://an.oa.org/OAI--script?verb=ListIdentifiers &from=1998-01-15&set=physics:hep`

`http://an.oa.org/OAI--script? verb=ListRecords& from=1998-01-15&set=physics:hep& metadataPrefix=oai_rfc1807`

### 3.1.3 원문요청 서비스

사용자가 service provider를 통해 검색한 콘텐츠에 대한 메타데이터 중에서 이와 관련된 원문에 직접 접근하기 원할 경우에는 MIRAGE 시스템이 가지고 있는 문서에 직접 접근해야만 한다. 하지만 MIRAGE 시스템이 가지고 있는 문서는 링크요소로만 되어있는 가상문서이기 때문에 일반적인 URI로는 접근할 수가 없다. 따라서 MIRAGE 시스템에서 가상문서의 처리를 담당하는 DocPro(document process)를 통해 서비스가 이루어져야만 하는데, DocPro는 링크요소로만 되어있는 가상문서를 사용자가 가상문서를 작성한 의도대로 볼 수 있도록 하기 위해 가상문서를 실제 XML로 변환하고 적절한 스타일시트를 적용하는 문서변환기의 기능을 가지고 있다. 가상문서를 이 기능을 이용하여 문서변환 과정을 거친 후 저장을 하게 되며, 저장된 문서의 URI를 이용하여 data provider가 직접 MIRAGE 시스템

에 접근하여 원문을 서비스 하게 된다.



[그림 4] 원문요청 서비스 구조

## 4. 결론

OAI는 각 디지털도서관 시스템들이 제공하는 서로 다른 접근 방법들과 제공되는 정보의 다형성으로 인한 정보 교환의 문제를 해결하여 준다. 본 논문에서는 MIRAGE에 OAI기능 지원을 첨가하여 MIRAGE 시스템에 의해 만들어진 가상문서에 대한 메타데이터를 제공하는 data provider를 구현하여 service provider를 통하여 정보를 상호이용할 수 있도록 하는 방법을 제시하였다.

## 5. 참고 문헌

[OAI] Open Archives Initiative. (<http://www.openarchives.org/>).

[강지훈 99] 강지훈, 맹성현, 이만호, "분산 디지털 도서관 시스템에서 XML을 이용한 가상문서의 표현처리," 제 2회 디지털 도서관 컨퍼런스, 서울, 1999년 11월

[Mya 99] S. H. Myaeng, M.-H. Lee, and J.-H. Kang, "Virtual Documents: a New Architecture for Knowledge Management in Digital Libraries," Proc. Asian Digital Libraries Conf., Taipei, Taiwan, Nov. 1999.

[Mya 2000] S. H. Myaeng, M.-H. Lee, J.-H. Kang, E.-I. Cho, Y.-B. Lee, D.-S. Lim, J.-M. Lim, H.-J. Oh, and J.-S. Yang, "A Digital Library System for Easy Creation/Manipulation of New Documents over Existing Resources", RIAO 2000, Paris, France, April 2000.

[XML 98] W3C, Extensible Markup Language (XML) 1.0, Recommendation, Feb. 1998. (<http://www.w3.org/TR/REC-xml>).

[DOM 98] W3C, Document Object Model (DOM) Level 1, Recommendation, Oct. 1998. (<http://www.w3.org/TR/REC-DOM-Level-1>).