

OAI 프로토콜을 이용한 P2P 파일 검색 시스템

임수정⁰ 장태무
동국대학교 컴퓨터공학과
(cryss⁰, jtm)⁰@dongguk.edu

A P2P File Search System Using OAI Protocol

Sujeong Im⁰ Taemu Chang
Department of Computer Engineering, Dongguk University

요 약

최근까지 개발되어 사용되고 있는 대부분의 P2P 파일 검색 및 공유 시스템들은 파일 이름과 확장자를 이용한 단순 키워드 검색 방식을 취하고 있다. 점점 기술이 발전되고 많은 정보가 쏟아져 나오면서 많은 자료 중에서 필요한 정보를 빠르고 쉽게 찾을 수 있는 능력이 필요하게 되었다. 또한 단순히 파일 검색을 떠나 콘텐츠를 공유하고 배포하는 경우 파일의 이름만을 가지고는 파일의 역할과 내용을 알 수 없어 파일에 대한 정보를 담은 메타데이터에 관한 연구와 그 중요성이 증가하고 있는 상황이다. 본 논문에서는 P2P 파일 공유 및 검색 시스템에서 검색의 정확성과 효율성을 높일 수 있도록 메타데이터를 이용한 파일 검색 시스템을 제안한다. 이를 위해 메타데이터로 자료의 검색이 가능한 OAI 프로토콜을 이용하였고 제안한 시스템을 구현해 실험해보았다.

1. 서 론

최근 인터넷 사용량이 증가함에 따라 유통되는 자료의 절대량이 증가하고 서비스 방식도 점점 복잡해지고 있다. 최근까지도 네트워크 환경에서 자료의 저장과 관리를 위해 가장 널리 사용되고 있는 클라이언트/서버 모델은 사용자와 통신량이 증가하면서 데이터 관리와 분산 처리에 어려움을 갖는다.

반면 P2P(Peer to Peer) 환경에서는 서버와 클라이언트 중심의 구조를 벗어나 파일 전송 프로토콜, 공유 드라이브, 그룹웨어 기술 등을 통합하여 사용자들 간의 직접적인 정보 교환을 가능하게 해주었고 현재 여러 응용에 쓰이고 있으며 미래에도 많은 분야에서 사용될 거라 예상된다.

다수의 P2P 시스템들이 파일과 데이터를 공유하기 위해 개발되었고 Napster, Gnutella, Freenet, Edongkey, Jxta 등의 프로그램들이 그 연장선 위에 있다. P2P 기술이 파일 검색 및 공유를 위해서만 사용되는 것은 아니지만 바로 이 파일 공유의 특징 때문에 많은 사람들에게 인기를 끌고 있다.

최근까지도 개발되어 사용되고 있는 대부분의 P2P 파일 검색 및 공유 시스템들은 파일 이름과 확장자를 이용한 단순 키워드 검색 방식을 취하고 있다. 초기에는 검색 능력이 떨어지더라도 파일 자체를 공유한다는 것 자체만으로도 충분했지만 점점 기술이 발전되고 많은 정보가 쏟아져 나오면서 많은 내용 중에서 필요한 정보를 빠르고 쉽게 찾을 수 있는 능력이 필요하게 되었다. [1]

또한 단순히 파일 검색을 떠나 콘텐츠를 공유하고 배포하는 경우 파일의 이름만을 가지고는 이 파일의 역할과 내용을 알 수 없어 파일에 대한 정보를 담은 메타데이터에 관한 연구와 그 중요성이 증가하고 있는 상황이다.

본 논문에서는 P2P 파일 공유 및 검색 시스템에서 검색의 정확성과 효율성을 높일 수 있도록 메타데이터를 이용한 파일 검색 시스템을 제안한다. 이를 위해 메타데이터로 자료의 검색이 가능한 OAI 프로토콜을 이용하였다.

2. 관련 연구

본 장에서는 P2P의 개념과 논문의 기본 기술이라 할 수 있는 OAI 프로토콜에 대해 소개한다.

2.1 P2P 개념

P2P는 각 컴퓨터가 동등한 능력을 가지고 있어, 어떤 컴퓨터에서라도 통신 세션을 시작할 수 있는 통신 모델을 지칭한다. 이는 네트워크에 연결되어 있는 모든 컴퓨터들이 서로 대등한 동료의 입장에서 데이터나 주변장치 등을 공유할 수 있다는 의미를 담고 있다. [2]

P2P 컴퓨팅은 파일 서버에 전적으로 의존하지 않고 직접적인 자원 교환을 지원하는 프로그램이나 네트워크 솔루션이라 정의할 수 있다. 그러므로 모든 클라이언트와 서버 양쪽으로 활동할 수 있는 피어가 되어 기존 클라이언트 서버 환경의 로드를 덜어냄으로써 고가의 성능에 방해가 되는 업그레이드의 필요성을 줄이는 장점을 지닌다.

또한 P2P 기술을 사용하면 PC 사용자가 별도의 서버나 고정 IP 없이도 인터넷으로 서로의 컴퓨터를 자유롭게 왕래하며 필요한 자료를 쉽게 주고받을 수 있다. P2P가 가지는 가장 중요한 장점 중의 하나가 자원의 분산에 있다고 볼 수 있다.

2.2 OAI 프로토콜

OAI(Open Archives Initiative)에서는 개방형 정보 제공 프레임워크와 프로토콜을 개발하고 있는 단체이다. 이 단체는 디지털 콘텐츠의 효율적인 유통 및 활용을 목적으로 상호운영성을 갖는 표준을 개발하고 보급하는 역할을 담당한다.[3]

OAI 프로토콜은 메타데이터의 수확(Harvesting)을 위한 프로토콜로서 공식 명칭은 OAI-PMH(Protocol for Metadata Harvesting)라 한다.

메타데이터 수집을 위한 OAI-PMH는 하나의 서비스 시스템에서 타 서비스 시스템으로 메타데이터 레코드의 정규화된 수집 동작을 지원하는 단순 프로토콜이며 하나의 시스템

에서 다른 시스템으로 메타데이터를 전송할 수 있는 기능을 갖는다. OAI-PMH의 특징은 첫 번째로 HTTP(HyperText Tranfer Protocol), XML(eXtensible Markup Language), XML 스키마와 같은 웹 표준을 기반으로 동작하기 때문에 구현하기가 쉬우며, 둘째로 프로토콜에 의해 지원되는 동작 체계가 단순하다.[4]

OAI 프로토콜의 운용은 서비스 제공자(Service Provider)와 데이터 제공자(Data Provider)의 단순한 모델로 설명이 가능하다. 서비스 제공자는 데이터 제공자로부터 수확한 메타데이터를 기반으로 검색, 브라우징, 원문 제공 등과 같은 부가가치 서비스를 제공하는 기능을 수행한다.

데이터 제공자는 자체적으로 각종 디지털 장서를 수집, 보유하고 있는 시스템이며, 서비스 제공자의 수확요청에 대응하여 적합한 메타데이터를 제공하는 역할을 수행한다.

OAI-PMH를 구성하는 요소는 다음과 같다.[5]

가. 하비스터(Harvester)

하비스터(harvester)는 OAI-PMH 요청을 생성하는 클라이언트 애플리케이션이다. 하비스터는 서비스 제공자에 의해 동작되며 리퍼지토리로부터 메타데이터를 수집하는 기능을 수행한다.

나. 리퍼지토리(Repository)

리퍼지토리는 OAI-PMH 요청을 처리할 수 있는 웹서버 및 데이터베이스로서 데이터 제공자에서 운영된다. 리퍼지토리는 전문과 메타데이터 모두 저장할 수 있다. 또한 메타데이터만을 저장할 수도 있다.

다. 아이템(item)

아이템(컨텐츠)은 유통시킬 자원에 대한 메타데이터이며 리퍼지토리의 구성요소이다. 하나의 아이템에 대하여 여러 형태의 메타데이터를 생성할 수 있으며 각 아이템은 OAI-PMH를 통하여 레코드로서 수집될 수 있다. 또한 각 아이템은 리퍼지토리 내에 유일한 identifier를 갖는다.

라. 레코드(record)

레코드는 하나의 형태로 표현되는 메타데이터로서 XML로 인코딩된 바이트 스트림 형태로 제공된다.

마. Set

set은 레코드들에 대해서 선택적으로 수집하기 위해 아이템을 그룹화할 수 있는 선택적 구조체이다. set은 여러 개의 최상위 계층의 노드를 갖을 수 있으며 set의 계층적 구조는 setSpec 파라미터의 구문으로 표현될 수 있다.

3. 시스템 설계

본 논문은 메타데이터를 이용하여 파일 검색의 정확성을 높이는데 그 목표가 있다. 특히 OAI 프로토콜을 이용한 기존의 클라이언트/서버 환경의 시스템을 P2P 환경에 맞게 변형하여 P2P 환경에서도 메타데이터를 이용하여 파일 검색이 가능하도록 한다.

이를 위해 OAI 프로토콜 기반의 서비스 제공자(Service Provider) 역할의 피어와 데이터 제공자(Data Provider) 역할의 피어를 설계하였다.

다음 그림 1은 설계한 시스템의 모식도로 Sn은 서비스 제공자 피어이고 Dn은 데이터 제공자 피어이다.

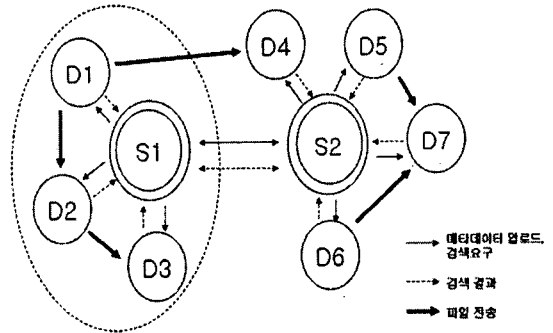


그림 1 시스템의 전체 모식도

데이터 제공자 역할의 피어가 하는 일은 기존 OAI 프로토콜 시스템의 데이터 제공자의 기능과 거의 비슷하다. 데이터 제공자 피어는 소유한 파일을 같은 역할의 피어들과 공유하며 공유된 파일들의 메타데이터를 생성하여 서비스 제공자 피어에 제공한다. 또한 메타데이터를 이용하여 파일을 검색할 수 있는 기능을 가진다.

서비스 제공자 피어는 일종의 슈퍼 피어(Super Peer)로써 데이터 제공자 피어의 기능을 가지며 데이터 제공자 피어와 연결을 관리한다. 그리고 데이터 제공자 피어가 제공하는 메타데이터를 주기적으로 수집하여 데이터 제공자 피어가 이 메타데이터를 이용하여 파일을 검색할 수 있도록 한다.

4. 시스템 구현

구현한 P2P 파일 검색 시스템은 제안한 시스템 구조를 테스트해 보기 위해 기능을 간소화한 프로그램이다. 자바 언어의 소켓 함수를 이용해 구현하였으며 자바 언어의 특성상 리눅스, 윈도우즈, 유닉스 등 대부분의 범용 플랫폼에서도 동작이 가능하다.

구현 시스템은 두 개의 프로그램으로 구성되어 있다. 첫째 서비스 제공자 피어 기능을 갖는 프로그램으로 데이터 제공자 피어들의 연결 상태를 관리하고 주기적으로 데이터 제공자 피어들이 가지고 있는 메타데이터를 수집하는 기능을 가지고 있다.

데이터 제공자 피어 기능을 갖는 프로그램으로 자신이 가진 파일들을 다른 데이터 제공자 피어들과 공유하며 보유한 파일의 메타데이터를 생성하고 이 메타데이터를 서비스 제공자 피어에 제공하는 기능을 갖는다. 생성된 메타데이터는 xml파일 구조를 지니고 있으며 서비스 제공자 피어 프로그램은 xml검색을 통해 데이터 제공자 피어 프로그램이 요청하는 결과 값을 전달하게 된다.

그림 2는 데이터 제공자 피어 프로그램에서 각 파일의 메타데이터를 입력하는 화면이다. 이 과정을 거친 후 파일에 대한 메타데이터가 담긴 별도의 파일이 생성되며 서비스 제공자 피어 프로그램에 의해 주기적으로 서비스 제공자 피어에 모이게 된다.

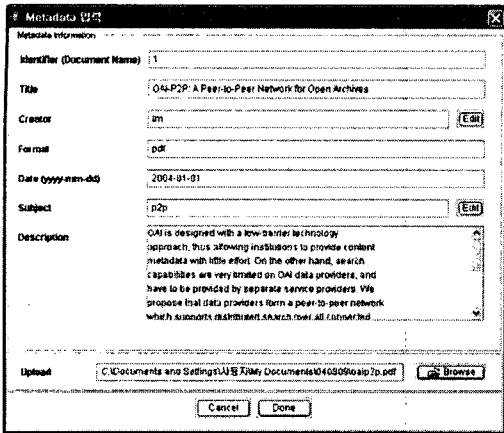


그림 2 메타데이터 추가 화면

그림 3은 p2p라는 주제로 메타데이터를 이용하여 검색한 후의 결과 화면이다. 파일명에 p2p라는 단어가 없는 파일도 검색됐음을 볼 수 있다.

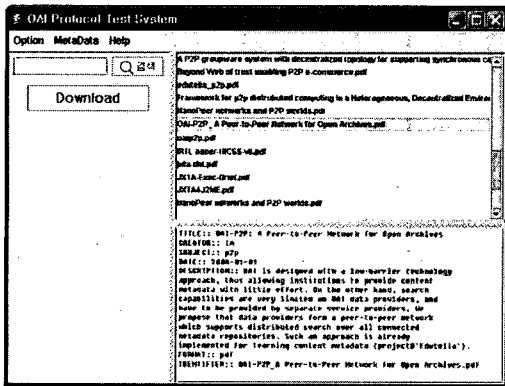


그림 3 메타데이터 검색 결과

그림 4는 기존의 파일명을 이용한 검색 화면으로 그림 3과 비교해보았을 때 훨씬 빈약한 결과가 나왔음을 확인할 수 있다.

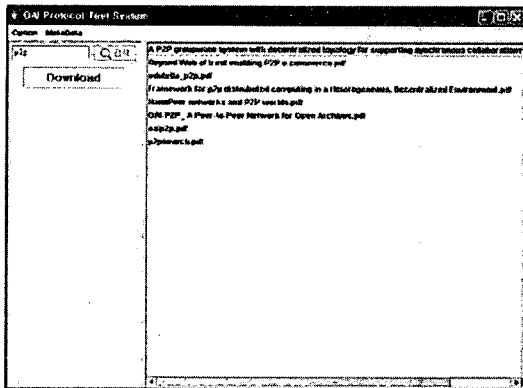


그림 4 파일 이름 검색 결과

5. 결론 및 향후 연구 계획

본 논문에서 OAI 프로토콜을 기반으로 P2P 환경에서도 메타데이터를 이용하여 파일 검색이 가능한 시스템을 제안하였다. 이를 통해 구현된 시스템은 기존의 파일 이름을 위주로 한 키워드 검색에 비해 더욱 신뢰적이고 정확한 결과 값을 얻을 수 있었다.

또한 기존의 OAI 프로토콜을 이용한 대부분의 시스템이 클라이언트/서버 구조로 이루어져 있어 시스템 확장 및 설치에 비용 및 노력이 많이 소요되었으나 제안된 시스템에서는 P2P 환경에서도 OAI 프로토콜을 이용할 수 있는 시스템이기 때문에 시스템의 확장과 분산 처리에 더욱 효율적이다.

그러나 본 논문에서 제안한 시스템 구조는 서비스 제공자가 데이터 제공자 피어의 메타데이터를 수집 보관해야 하기 때문에 서비스 제공자 피어에 부담을 줄 수 있다. 그리고 서비스 제공자 피어의 경우 다른 데이터 제공자 역할의 피어가 다수 몰릴 경우 부하가 발생하여 원활한 시스템이 구성될 수 없고 메타데이터를 수집하는 주기에 따라서는 시스템이 큰 영향을 받을 수 있다.

또한 구현 후 실험 결과를 통해 검색의 정확성을 측정하여 성능을 분석하는데 어려움이 있다. 왜냐하면 메타데이터 검색 특성 상 간단한 파일명 검색보다 시간이 많이 걸리고 메타데이터 입력의 충실성에 따라 검색의 정확도가 결정되기 때문에 단순히 기존의 파일명 검색과 성능을 비교할 수 없다.

이러한 단점을 해결하고 더욱 안정적이고 효율적인 시스템 개발을 위해서 서비스 제공자 피어의 메타데이터 수집 주기와 접속 피어의 관리 등에 대한 연구와 검색의 정확성을 높이고 이를 측정할 수 있는 방식의 연구가 더욱 요구된다.

6. 참고문헌

- [1] Xin Xiang, Yuanchun Shi, Ling Guo "Rich Metadata Searches Using the JXTA Content Manager Service" AINA'04, 2004
- [2] Stoica, R. Morris, D. Karger, F. Kaashoek "Chord: A scalable peer-to-peer lookup service for Internet applications," in Proc. ACM SIGCOMM, 2001
- [3] Open Archives Initiative, <http://www.openarchives.org>
- [4] 이수상, "학술정보 유통에 있어 OAI 프로토콜의 적용에 관한 연구", 한국 도서관 정보학회지(제 35권 제 2호), 2003
- [5] 송중철, 문병주, "OAI의 메타데이터 Harvesting 관련 기술 동향", 선진이러닝 기술표준 연구그룹, 2002
- [6] Joseph S. "NeuroGrid: Semantically Routing Queries in Peer-to-Peer Networks", In Proceedings of the International Workshop on Peer-to-Peer Computing, 2002
- [7] Sumeet Thadani, LimeWire LLC "Meta Information Searches on the Gnutella Network" http://www.limewire.com/developer/metainfo_searches.html
- [8] Hao Ding, Ingeborg Solvberg, "Metadata Harvesting Framework in P2P-Based Digital Libraries" DCCConf'04, 2004
- [10] Open Archives Forum, <http://www.oaforum.org/>