

효율적 정보자원 공유 및 유통을 위한 P2P 기반 플랫폼 설계

Designing a P2P-based Platform for Effective Information Share and Distribution

구남앙·이응봉, 충남대학교 대학원 문헌정보학과
nak0519@kisti.re.kr, eblee@cnu.ac.kr

Kuh, Nam-Ang, Lee, Eung-Bong

Graduate School of Library & Information Science, Chungnam National University

연구 개발자들은 자신의 정보활동에 있어 전체 활동의 20~40%를 관련 정보의 탐색 및 분석에 사용하고 정보활동 결과는 개인 컴퓨터에 저장한다. 그러나 개인 컴퓨터상의 유용한 정보는 정보 관리 및 공유 도구의 부재로 인하여 가치 및 유용성을 상실하고 있다. 본 연구에서는 데스크탑 서치와 P2P 기반의 정보공유 기능을 갖는 시스템을 제안한다. 데스크 탑 서치를 통하여 연구 개발자는 본인의 컴퓨터에 저장된 정보를 효율적으로 활용할 수 있다. 연구 개발자들이 많은 시간을 들여 축적한 유용한 정보자원을 쉽게 공유할 수 있게 하는 P2P 기반의 정보공유는 연구개발 활동의 전체적인 생산성을 높일 수 있다.

1. 서론

연구 개발자들은 전체 활동의 20~40%를 관련 정보의 탐색 및 분석에 소비하고 있다 (Larry Bielawski 1997). 과학기술의 급격한 발전으로 인하여 연구 개발자들이 탐색 및 분석해야 할 정보의 양이 폭증함에 따라 관련 정보의 수집 및 분석 시간은 점점 증가하게 될 것이다. 따라서 관련 정보의 효율적 입수 방법은 연구활동의 성공에 결정적인 역할을 하게 된다 (KISTI 2004)

인터넷 및 개인 컴퓨터가 일반화되지 않은 시기에는 대부분의 정보가 종이 매체에 기록되었고 도서관을 통해서 입수할 수 있었다. 웹이 일반화된 현대에는 많은 정보가 종이 매체뿐만 아니라 전자파일 형태로 유통되고 있다. 연구 개발자들은 인터넷 상에 존재하는 수많은 정보

자원들 중에서 자신에게 꼭 필요한 정보를 찾기 위해서 인터넷 포탈 또는 출판사 홈페이지에서 상당한 시간동안 검색을 한다. 이렇게 탐색된 전자파일은 연구 개발자의 유용성 판단에 따라 개인 컴퓨터에 저장된다.

그런데 연구 개발자들은 정보 수집 활동을 통해서 자신의 개인 컴퓨터에 유용한 정보자원을 매우 많이 저장하고 있음에도 불구하고 사장시키는 경우가 많이 존재한다. 이러한 현상은 개인 컴퓨터의 저장 용량의 증가로 인하여 컴퓨터에 존재하는 전자파일의 수는 매우 많이 증가되었으나 효율적으로 관리 및 검색할 수 있는 시스템의 부재 때문에 발생한다.

정보 수집 활동의 대부분 시간은 무수히 많은 정보자원들 중에서 관련된 정보들을 선별하기 위하여 사용된다. 과학기술의 발전에 따라

서 단독 연구보다는 동일 분야의 연구 개발자들간 협력 연구가 중요시되고 있는 오늘날 정보 탐색 과정 및 결과의 공유를 통한 협력연구가 더욱 중요해지고 있다(Hendry & Harper 1997). 연구 개발자 개인 컴퓨터상에 저장된 정보자원들은 정보 수집 활동의 결정체로서 타 연구 개발자들과의 효율적 공유 체계는 연구개발의 생산성을 높일 수 있다.

본 논문에서는 컴퓨터상에 존재하는 전자파일을 효율적으로 관리하고 공유할 수 있는 시스템을 제안한다. 데스크탑 서치 기능을 통하여 전자파일의 이름이 아닌 내용을 기반으로 컴퓨터상에 있는 파일을 효율적으로 탐색할 수 있다. P2P(Peer-to-Peer)기반의 전자파일 전송 기능을 통하여 연구 개발자들은 정보자원들을 상호 공유한다. 전자파일이 전송될 때는 해당하는 메타데이터가 항상 같이 전송되어 저작권 보호 및 부가정보 공유가 이루어질 수 있다.

2. 관련연구

현재 개인 컴퓨터의 정보자원에 대한 효율적인 관리시스템으로 수많은 데스크탑 서치 엔진들이 제기되고 있다. 또한 정보공유 시스템으로 전자게시판의 Q&A 방식이나 P2P 기반의 개인의 파일 전송 시스템 방식을 이용하여 정보공유가 이루어지고 있는 실정이다.

우선 효율적인 정보관리와 관련하여 다양한 벤더에 의한 데스크탑 서치가 출시되었고, 지속적으로 개선되고 있다. 이를 통해서 이용자는 자신의 컴퓨터에 저장된 정보를 내용기반의 정보검색으로 쉽게 찾을 수 있다. 그러나 정보자원에 대한 다양한 형태의 파일검색 지원 불가능이나 정보자원에 대한 색인정보의 누락·부족으로 인한 정보관리 측면의 문제점이 해결과제로 남아있다(이현주 2005).

연구 개발자들의 생산성 향상을 위해서는 공유를 통한 협력체계의 구축이 필요하다. 현재 공유체계로는 협력연구 그룹간 웹 기반의 커뮤니티

형태 형성을 통한 전자게시판의 운영이나 P2P 기반의 정보공유를 통한 협력연구 체제가 있다.

전자게시판은 다수의 이용자들간 각종 정보를 교환하거나 프로그램을 공유할 수 있는 시스템이다. 이는 이용자들간 의사소통이 순조롭게 이루어질 수 있는 장점이 있다. 그러나 전자게시판의 경우 단방향 정보공유 시스템으로 Q&A 방식이 대부분이고, 익명성이라는 점을 감안하면 역기능이 다수 발생할 수 있다.

P2P 기반 정보공유 시스템은 개인 대 개인이 파일을 주고받을 수 있는 시스템이다. 이는 다양한 콘텐츠 및 서비스 지원을 할 수 있다는 장점이 있다. 그러나 현재 기존 P2P 시스템의 문제점 중 하나는 검색의 효율성 문제이다. 단순히 파일명이나 확장자 등에 대해 일치하는 검색기능을 제공하는 것으로는 이용자의 검색의도를 만족시킬 수 없다.

이에 Mei Li(2004)는 P2P의 공유 정보자원에 대한 검색 효율화를 위해 해당 정보자원의 소유자 ID와 데이터나 파일의 위치, 서비스 등에 대한 색인을 통하여 질의를 처리하는 시스템의 개발을 시도하였다. 이를 통하여 다양한 형태의 질의나 범주, 속성 등에 대한 검색의 효율성을 증가시켰다. 본 플랫폼은 이러한 공유자원의 색인정보를 공유자원 속에 포함시켜 유통시킴으로써 검색의 효율성을 향상시키고자 한다.

현재 이용자의 정보자원을 관리하는 효율적인 문서관리 시스템과 정보자원의 공유에 대하여 P2P 기반의 정보공유 시스템이 각각 독립적으로 존재하고 있다. 그러나 독립적인 시스템은 문서관리를 효율적으로 지원할 수 있지만 공유에 있어 정보자원의 누락이라는 문제가 발생할 수 있다. 따라서 개별 시스템을 결합하여 효율적인 정보유통을 위한 새로운 시스템의 개발이 필요하다. 이 시스템은 효율적인 문서관리를 위한 데스크탑 서치 기능을 바탕으로 색

인정보를 해당 정보자원과 결합하여 유지함으로써 효율적인 문서검색 및 관리 기능을 제공할 수 있다. 또한 검색 및 관리대상이 되는 정보자원은 색인정보를 유지하는 형태로 정보자원을 공유함으로써 P2P 기반의 정보유통 체제에서 효율적인 정보유통이 이루어질 수 있다.

3. 지식파일

데스크탑 서치는 파일의 내용을 색인하여 단순히 내용기반 검색을 지원할 뿐 정보자원의 효율적인 관리 및 유통에 대해서는 고려하지 않는다. 일반적으로 이용자는 컴퓨터의 디렉토리 혹은 데이터 셋에서 원문정보를 전부 열람하여 파일을 식별하기는 어렵다. 또한 정보공유에 있어서 일반파일이 유통되면 이용자의 정보탐색의 시간이 오래 걸린다는 효율성 측면의 문제가 있다. 따라서 관리가 용이하도록 파일에 일정한 식별 정보를 부여함으로써 이용자로 하여금 필요한 정보를 추출해 주어 사장될 가능성이 있는 정보자원의 이용을 돕고, 공유파일의 검색 효율성에 관하여 수많은 정보이용자에게 해당 파일을 공유할 때 정보의 탐색을 도와줄 수 있는 체계적이고 객관적인 판단 정보가 필요하다.

본 연구에서는 효율적인 문서관리와 유통을 위하여 지식파일이라는 개념을 도입하였다. 본 플랫폼에서 지식파일은 파일의 실제 내용을 담고 있는 원문정보와 원문파일의 다양한 메타데이터 정보를 논리적으로 구조화한 파일이다.

지식파일 메타데이터의 구성요소는 형태적 메타데이터, 내용적 메타데이터, 관리용 메타데이터로 구분할 수 있다. 형태적 메타데이터는 단순히 관리를 위한 파일의 형태적인 측면에서 파일명이나 생성자, 파일타입 등의 정보를 기재한다. 내용적 메타데이터에서는 정보자원의 관리를 위한 분류정보로 활용하기 위한 분야정보나 검색의 효율성을 증가시킬 수 있는 키워드를 기재한다. 또한 공유된 정보자원의 정보

탐색에 있어 요약이나 평가정보를 활용하여 탐색시간을 단축시킬 수 있다. 관리용 메타데이터는 버전관리 및 유통정보를 유지할 수 있는 식별자를 두고 공유된 정보자원의 위조 및 변조를 방지하기 위한 접근제어 정보를 기재한다.

4. 문서관리와 유통을 위한 플랫폼 설계

4.1 시스템의 전체 구성도

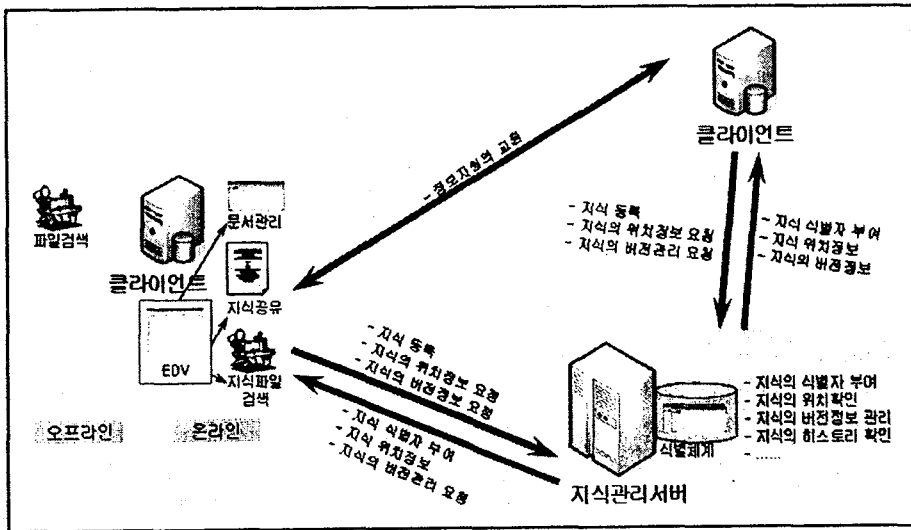
효율적인 문서관리와 유통을 위한 플랫폼은 크게 세 가지 요소로 구성된다.

① 지식파일을 생성하고 이를 조회 및 편집 가능한 도구로써 Electronic Document Viewer, ② 컴퓨터상의 정보자원에 대한 검색 및 분류, 공유설정 등의 기능을 하는 지식관리 클라이언트, ③ 컴퓨터상의 정보자원에 대한 버전관리 및 정보의 유통상황을 확인할 수 있는 식별자의 관리를 위한 지식관리 서버가 그 구성요소이다(그림 1 참조).

4.2 Electronic Document Viewer

일반파일은 원문파일과 원문파일의 메타데이터 정보가 결합되어 물리적으로 하나의 지식파일로 생성된다. 앞서 이후 지식파일의 열람, 편집을 위해서는 원문파일에 대한 메타데이터 정보를 입력·수정할 수 있는 도구가 필요하다. 이를 위하여 본 플랫폼은 일종의 컨테이너로써 “문서관리와 유통을 위한 Electronic Document Viewer”를 개발한다(이하 EDV라 칭함). 컨테이너란 “인터넷 익스플로러”와 같은 웹 브라우저에서 파일을 실행하면 실행되는 형태는 인터넷 익스플로러와 유사하게 동작한다. 그러나 실제 구현되는 기능은 해당 파일의 실제 소프트웨어 기능이 실행된다.

EDV도 마찬가지로 앞서 언급한 지식파일의 논리적 구조화에서 해당 파일의 포맷과 연결되는 “연결 프로그램” 정보를 메타데이터 부분에 포함시킴으로써 EDV를 통해 메타데이터의 보



<그림 1> 효율적인 문서관리 및 유통 시스템

기와 편집 그리고, 지식파일의 보기 및 편집이 가능하다. EDV는 컨테이너 역할을 하고, 실제로는 해당 연결 프로그램이 실행되게 된다. 즉, 지식파일은 해당 파일의 연결 프로그램에 관계 없이 EDV를 통해 실행된다.

EDV는 논리적으로 구조화된 지식파일을 원문정보와 메타데이터를 분리하여 지식의 편집과 수정기능을 제공하고 더불어 지식파일의 변경에 관한 정보의 식별자를 유지하는 기능을 한다. 해당 지식파일이 EDV를 통해서 수정되면 변경 정보를 서버에 보내 지식파일의 버전을 관리·유지한다.

4.3 지식관리 클라이언트

문서관리와 유통을 위한 플랫폼의 클라이언트는 컴퓨터상에 저장된 이용자의 모든 파일을 색인한다. 색인결과 중에서 지식파일로 관리하기 위해 선택된 파일은 원문정보와 결합시켜 문서관리와 유통을 위한 플랫폼의 고유파일로 변환된다. 또한 지식파일로 변환되지 않는 일반파일에 대해서도 파일의 색인 결과는 이용자의 컴퓨터상의 파일검색을 위해 유지된다. 이

이용자의 관리대상 파일에 대하여 지식파일이 생성되면 이후 지식맵에 등록된다.

지식맵은 이용자 임의대로 생성 가능한 일종의 분류체계로 윈도우 운영체제의 하위 폴더이다. 윈도우 폴더구조와 마찬가지로 이용자 임의대로 폴더의 생성·수정·삭제가 가능하다. 지식맵은 이러한 이용자의 분류체계에 의해 컴퓨터에 저장된 모든 파일의 색인결과와 해당 파일의 분야정보를 바탕으로 지식파일을 생성·등록한다. 이는 윈도우에서 기본적으로 제공되는 탐색기를 통해서도 탐색이 가능하다. 이후 정보자원의 공유는 지식맵의 공유를 의미하는데 지식맵에 등록되어 있는 파일 중에서 지식파일을 대상으로 공유가 이루어진다.

지식파일의 등록대상은 ① 내 컴퓨터 파일의 검색결과와 ② 지식맵, 실제 윈도우에 설치된 “문서관리와 유통을 위한 플랫폼 폴더의 지식맵”에 저장된 일반파일로 지식파일로 등록되지 않은 파일, ③ P2P를 통해 다운로드한 공유 지식파일이 그 대상이 된다. 즉, ① 내 컴퓨터의 파일 검색결과는 이용자의 검색결과로 실제 윈도우 폴더의 파일이다. 이들은 지식맵에

등록되기 위해 앞서 말한 지식파일로 변환된다. 이 과정에서 메타데이터와 원문정보가 논리적으로 구조화되고 등록된다. ② 지식맵에 저장된 파일이지만 논리적으로 구조화되지 않은 파일은 프로그램의 별도 메시지의 송출을 통해 이용자에 의해 지식파일로 변화되어 등록된다. ③ P2P를 통하여 다운로드한 공유 지식파일은 최초 컴퓨터상의 특정 폴더로 전송된 후 이용자가 임의로 재분류한다

문서관리를 위하여 색인된 모든 파일(지식파일과 일반파일 모두 포함)은 3개의 검색으로 구분하여 검색기능을 제공한다. ① 컴퓨터의 모든 파일을 대상으로 하는 “내 컴퓨터 파일 검색”기능과 ② 지식맵에 등록된 지식파일의 검색을 위한 “내 지식파일 검색”기능, ③ P2P 상에서 공유되고 있는 지식파일을 검색하기 위한 “P2P 지식파일 검색”기능으로 각각 구별하여 제공한다.

위의 검색을 구분한 이유는 검색의 종류별로 로그인과 로그오프 상태를 구분하여 파일의 버전에 관련된 정보를 일원화시키는데 있다. 내 컴퓨터 파일검색은 단순한 데스크탑 서치 기능으로 내 컴퓨터에 저장된 문서를 검색하여 그 결과를 제공하는데 그 목적이 있다. 따라서 로그인 상태나 로그오프 상태에서도 검색결과를 제공한다. 그러나 내 지식파일 검색과 P2P 지식파일 검색은 파일의 검색결과를 통해 검색결과를 브라우징 하는데는 문제가 없지만 이용자에 의한 지식파일의 변경과 관련하여 파일의 버전과 유통정보 관리를 위해 해당 파일의 다운로드 및 편집 시에 변경된 식별자 정보가 서버로 보내져 관리되기 위함이다.

4.4 지식관리 서버

현재 일반적인 P2P 시스템의 의미에서는 서버와 클라이언트가 존재하지만, 최근 개발되고 있는 P2P 시스템은 관리자만 존재할 뿐 서버와 클라이언트의 기본 개념은 사라지고 피어가

서버와 클라이언트의 역할을 수행하는 형태이다.

그러나 본 플랫폼은 컴퓨터상의 파일에 대한 버전관리와 유통되는 지식파일의 유통상황에 관한 식별자 유지를 위하여 서버를 두어야 한다. 지식관리 서버는 개별 이용자에게 유일한 ID를 부여하고 해당 파일에 대한 지식파일 등록 시 해당 지식파일에 대한 원본 생산자임을 증명할 수 있는 식별자와 각 클라이언트로부터 받은 지식파일의 생성에 관한 메타데이터 정보를 유지한다. 이를 위하여 각 지식파일에 식별자를 조합하여 부여한다. 또한 지식파일 식별자의 변경에 대하여 변경사항을 식별자를 통하여 각각 재배분하여 지식파일을 일관성 있게 관리하여 파일에 대한 버전유지 및 유통상황을 추적할 수 있도록 한다. 또한 서버는 P2P의 부적합 정보의 유통방지를 위하여 이용권한을 제어하기 위한 사용자 정보를 유지해야 한다.

그러나 지식 수정 시 이전 버전에 관한 식별자를 모두 유지하는 것은 원문 부분보다 메타데이터 부분이 커질 수 있기 때문에 문제가 될 수 있다. 따라서 식별자의 유지는 해당 지식파일이 변경되면 지식파일의 식별자는 최초 원본에 관한 정보, 현재 지식파일 이전 버전에 관한 식별자 정보와 현재 지식파일 버전에 대한 식별자 정보로 구성된다. 이러한 식별자를 이용한 버전유지는 지식파일 공유에 있어서도 동일하게 해당 지식별 메타데이터에 식별자 유통 정보를 유지하여 각 식별자 정보를 서버를 통해 연계하여 원본파일의 식별이 가능하다.

5. 결론

현재 연구 개발자들은 개인 컴퓨터를 이용하여 수많은 정보자원을 생산하고 정보활동을 통해 수집된 수많은 정보자원을 축적한다. 그러나 컴퓨터상에 저장되어 있지만, 이용자에 의해 검색과 재이용이 어려운 정보자원들이 많이 있다. 본 연구에서는 이러한 정보자원의 관리

를 위하여 데스크탑 서치를 활용하여 내용기반 검색기능을 제공하고, P2P 기반의 정보공유 시스템을 활용하여 관련 연구분야의 정보활동 시간을 축소시켜 줄 수 있다. 이를 통하여 보다 나은 성과를 산출하기 위하여 본 플랫폼을 제안하였다.

향후 연구로는 정보탐색 및 분석, 협력연구 지원을 통하여 연구 생산성을 높이기 위하여 유사분야 연구 개발자를 쉽게 찾을 수 있는 시스템 개발이 필요하다. 유사분야 연구 개발자 검색 시스템은 본 플랫폼의 이용자 개인정보나 플랫폼을 활용한 정보탐색 행태 등을 분석한 정보를 통하여 구현할 수 있을 것이다.

참고문헌

----. 2004. 2004년 디지털콘텐츠백서 제 3부. 디지털콘텐츠 유통, p. 381.

양광민. 2003. JXTA를 이용한 P2P 기반 자료 공유시스템(JDSS)의 구현. 『정보기술과 데이터베이스 저널』, 10(3).

오동근. 2004. 메타데이터의 이해. p. 21~23.

이정기. 2002. 네트워크상에서 안전한 정보 공유에 관한 연구. 『전자정보통신연구소논문집(조선대)』, 2002(2). p.

이현주. 2005. Desktop 검색도구를 이용한 개인 정보의 효율적 관리방안에 관한 연구 : 디지털 정보자료를 중심으로. 『정보관리연구』, 36(2). p. 41~42.

한국과학기술정보연구원, 2004. 과학기술동향 모니터링 사업(Technology Trends Monitoring). p. 3.

Narvenkar Harshad. 2001. An Overview of Technology and Opportunities, p. 1~2.

Biström Johnny. 2005. Peer-to-Peer Networks as Collaborative Learning Environments. Helsinki Univ. of Technology, p. 2~3.

Bielawski Larry. 1997. Electronic Document

Management System. Prentice Hall PTR, p. 20.

Mei Li. 2004. Efficient Peer-to-Peer Information Sharing over Mobile Ad Hoc Networks. Pennsylvania State Univ.

Tang Li. 2002. Identifying Resource Authenticity in P2P Networks. Tsinghua Univ.(China).

Tony White. ----. Sharing@the Edge: Secure Information Sharing(Carleton Univ.), p. 2~5.

<http://www.informit.com/article> [cited 2004. 08. 10]

<http://www.uwebc.org/reports/> [cited 2004. 08. 10]

<http://www.jxtakorea.net/> [cited 2004. 08. 10]