

# 학제간 연구를 통한 IPR 데이터베이스 구축의 쟁점

김동용\*, 박영철  
경북대학교 컴퓨터학부

## Issues of IPR Database Construction through Interdisciplinary Research

Dong Yong Kim\*, Young Chul Park

School of Computer Science and Engineering, Kyungpook National University

**요약** 인문·사회과학 연구원들과 데이터베이스 전문가들이 태평양문제연구원(IPR)가 작성한 IPR 자료를 데이터베이스로 구축하기 위해 함께 하였다. 본 논문은 그 데이터베이스의 구축에서 IPR 자료의 품질을 보장하기 위한 쟁점과 해결책을 제시한다. 우리는 데이터베이스의 접근성을 위해 웹에 데이터베이스를 유지하여 연구원들이 웹 브라우저로 접근할 수 있도록 하며, 데이터베이스 구축의 편리성을 위해 통합 인터페이스를 제공하여 연구원들이 그 속에서 모든 작업을 수행하게 하며, 구축한 IPR 자료의 완전성을 위해 입력된 각 IPR 자료의 책임 소재를 밝히는 책임 입력과 책임 승인을 지원하며, 승인의 즉시성을 위해 연구원들의 입력을 원활히 하는 대화형 승인 과정을 지원한다. 또한 구축한 IPR 자료의 정확성을 보장하기 위해 우리는 데이터베이스 설계, 질의 처리, 트랜잭션 관리, 검색과 정렬 기법들을 사용한다. 특히, 레코드들 간의 존재 종속관계를 이용한 동시성 제어를 통하여 운영체제 파일과 그 파일의 경로 간의 정확성을 보장한다. 향후 연구로 내용 검색, 데이터베이스의 다운로드와 업로드, IPR 자료의 저작권 관련 작업을 포함한다.

• 주제어 : 태평양문제연구원, 웹, 데이터베이스, 트랜잭션, 공동연구, 학제간 융합

**Abstract** Humanities and social sciences researchers and database experts have teamed up to build a database of IPR materials prepared by the Institute of Pacific Relations (IPR). This paper presents the issues and solutions inherent in the database construction for ensuring the quality of IPR materials. For the accessibility of the database, we maintain the database on the Web so that researchers can access it via web browsers; for the convenience of the database construction, we provide an integrated interface that allows researchers to perform all tasks in it; for the completeness of IPR materials constructed, we support the responsible input and the responsible approval that identify responsibilities of each IPR material entered; and for the immediacy of the approval, we support an interactive approval process facilitating the input of researchers. We also use database design, query processing, transaction management, and search and sorting techniques to ensure the correctness of IPR materials entered. In particular, through concurrency control using existence dependency relationships between records, we ensure the correctness between the operating system files and their paths. Our future studies include content search, database download and upload, and copyright related work on IPR materials.

• Key Words : IPR, Web, Database, Transaction, Collaborative Research, Interdisciplinary Convergence

\*Corresponding Author : 김동용(qwe7087@naver.com)

Received May 26, 2017

Accepted August 20, 2017

Revised July 24, 2017

Published August 28, 2017

## 1. 서론

태평양문제연구원(Institute of Pacific Relations, IPR)은 태평양 연안 국가들 사이의 문제와 관계에 대한 논의를 위해 1925년에 설립되고 1960년에 해산된 국제민간 학술단체이다. IPR의 국제사무국은 1925년에서 1952년 사이에 IPR에 의해 출판된 1,408건의 자료들의 목록을 1953년에 발간하였다[1]. 우리는 그 목록을 IPR 자료 목록, 그 목록에 수록된 자료를 IPR 자료, 각 IPR 자료에 대한 그 목록의 정보를 그 자료의 기본 정보라 한다.

경북대학교 아시아연구소는 한국연구재단의 지원으로 2016년 9월부터 3년 기간으로 IPR 자료를 습득하고 분석하여 연구용 정보를 생성하고 그를 데이터베이스로 구축하여 재단에 제출하는 사업을 수행하고 있다. 우리는 그 정보를 IPR 자료의 상세 정보, 그 데이터베이스를 IPR 데이터베이스(이하 IPR DB)라 한다. 본 사업의 참여 인력은 인문·사회과학 연구원 9명[2,3], 데이터베이스 전문가 3명, 그리고 법률자문위원 1명[4]으로 구성된다. 인문·사회과학 연구원들은 본 사업을 공동 연구로 수행하지만 사업 추진을 위해 관리자 1명, 공동 연구원 4명, 전임 연구원 3명, 연구 보조원 1명으로 업무가 구분된다.

관리자는 본 사업의 총 책임자로서 각 연구원에게 업무를 배정하고 그들의 업무를 점검하며 IPR DB에 접속하는 사용자의 생성과 삭제 등의 책임을 가진다. 전임 연구원들은 IPR 자료를 습득하고 분석하여 상세 정보를 IPR DB에 입력한다. 본 사업은 전임 연구원들이 입력한 IPR 자료에 대한 정확성과 완전성을 검증하기 위해 공동 연구원들의 승인 과정을 두었다. 공동 연구원들은 전임 연구원들이 승인을 요청한 자료를 검토하고 승인하며 자료의 수정을 요구할 수 있다. 보조 연구원은 관리자로부터 권한을 위임 받은 업무를 수행한다. 본 사업의 참여 인력과 모든 일반 사용자들은 IPR 자료를 검색하고 열람할 수 있다.

본 사업의 참여 인력들이 함께 결정한 IPR DB의 품질을 보장하기 위한 IPR DB 구축의 중요 쟁점과 해결책은 다음과 같다.

첫 째, IPR DB는 접근성을 보장해야 한다. 이를 위해 우리는 특정 웹 서버(web server)에 IPR DB 구축 프로그램을 두고, 그 웹 서버만이 IPR DB를 가지며, 모든 연구원들은 웹 브라우저(web browser)를 이용하여 IPR DB의 웹 페이지에 접근하여 실시간으로 업무를 수행하도록 한다.

둘 째, IPR DB는 편리성을 보장해야 한다. 이를 위해 우리는 하나의 통합 인터페이스를 제공한다. 그 통합 인터페이스는 상단에 대분류의 틀을, 좌측에 각 대분류에 대한 중분류의 틀을 고정해 두고, 선택된 분류에 대한 화면을 가운데 나타낸다.

셋 째, IPR DB는 구축한 IPR 자료의 완전성을 보장해야 한다. 이를 위해 우리는 각 IPR 자료의 상세 정보는 오직 한 명의 전임 연구원에 의해 생성, 수정, 삭제되며, 오직 한 명의 공동 연구원에 의해 승인되도록 하는 책임 입력과 책임 승인을 지원한다.

넷 째, IPR DB는 IPR 자료의 승인의 즉시성을 보장해야 한다. 이를 위해 우리는 승인의 전 과정을 통합 인터페이스 내에서 대화형으로 지원한다.

다섯 째, IPR DB는 IPR 자료의 정확성을 보장해야 한다. 이를 위해 데이터베이스 전문가들은 데이터베이스 설계, 질의 처리, 트랜잭션 관리, 검색과 정렬 기법들을 사용한다. 특히, 레코드들 간의 존재 종속관계를 이용한 동시성 제어를 통하여 운영체제 파일과 그 파일의 경로간의 정확성을 보장한다.

본 논문의 관련 연구는 다음과 같다. IPR DB는 IPR 자료들을 <지역, 분야, 문서형태>로 분류한다. 분류와 관련하여 문화자원을 <지역, 시대, 분야>로 분류하는 분류체계를 제시하는 연구도 있다[5]. IPR DB는 IPR 자료들이 35년의 단기간에 구축되었으므로 <시대> 분류를 가지지 않는다. 우리는 우리의 분류체계를 IPR DB에 정확히 반영하는 방법 또한 제시한다.

IPR DB는 IPR 자료에 대한 기본적인 검색과 정렬 기능을 제공한다. 그 검색의 인터페이스를 OPAC(Online Public Access Catalog)의 표준 기능성 체크리스트를 참고하여 설정할 수도 있다[6]. 그러나 IPR DB의 기본적인 검색 기능은 IPR DB 구축의 지원에 있으므로 원래의 목적에 충분하다고 본다. 그러나 이는 향후 IPR DB의 검색 인터페이스의 확장에 참고가 될 것이다.

IPR DB에서 검색과 정렬의 결과는 IPR 자료의 목록과 링크(link)이며 그 링크를 클릭함으로써 IPR 자료의 상세 정보에 연결된다. 검색 기능을 사용해 자료의 정보에 연결된다는 점에서 국내 포털의 검색기능을 사용해 국내 대학도서관 종합목록 서비스 OPAC에 접근하고 그를 통하여 그 자료를 소장한 대학도서관 OPAC의 서지 화면에 연결되는 모델을 제시하는 연구도 있다[7]. 그러나 IPR DB는 IPR 자료가 소장된 대학도서관 OPAC에

접근하는 것이 아니라 본 사업에서 제작한 IPR 자료의 상세 정보에 접근함이 근본적인 차이이다.

현재까지 우리가 조사한 바로는 본 논문과 같이 다양한 분야의 전문가들이 특정 자료를 장기간의 공동 연구를 통해 데이터베이스로 구축하는데 발생하는 쟁점들을 식별하고 그 해결책을 제시하는 문헌들을 찾지 못하였다. 특히, 웹 환경에서 자료의 삽입, 삭제, 변경과 관련하여 그를 수행하는 데이터베이스 트랜잭션의 처리[8], 데이터베이스 설계에서 존재 종속관계와 로크 모드(lock mode)의 설정[9]의 관계 등이 구축된 데이터베이스의 정확성에 미치는 영향에 대한 논의는 그 중요성에도 불구하고 전무한 실정이다. 본 논문은 IPR DB 구축의 중요 쟁점들을 분석하고 그 해결책의 상세 내용을 제시한다.

## 2. IPR 자료의 기본 정보와 상세 정보

IPR 자료 목록은 대상지역을 35개로 구분한다. 그 분류를 바탕으로 각 IPR 자료에 대한 기본 정보는 <일련번호, 제목, 저자, 대상지역, 발행처, 발행연도, 페이지 수>의 속성들을 가진다. '일련번호'는 대상지역을 특정 순으로 나열한 뒤 각 대상지역에 속하는 IPR 자료들을 제목의 알파벳순으로 정렬하여 1부터 시작하는 정수를 부여한 것이다. '발행처'는 모두 85개이다.

IPR 자료의 상세 정보는 기본 정보에 <대상분야1, 대상분야2, 문서형태>의 분류를 추가하고 <제공처, 제목\_한글, 목차\_원문, 목차\_한글, 초록\_원문, 초록\_한글, 해제, 키워드\_원문, 키워드\_한글, PDF 파일, 텍스트 파일, 표지 이미지 파일>의 속성들을 추가로 가진다. 대상분야는 여덟 가지로 구분되며 '대상분야1'과 '대상분야2'는 그들 중 두 가지를 나타낸다. '문서형태'는 그 자료의 형태로써 여섯 가지로 구분된다. '제공처'는 본 사업에 자료를 제공한 기관을 나타낸다. '제목\_한글'은 '제목\_원문'을 한글로 번역한 것이다. '목차\_원문'과 '목차\_한글'은 그 자료의 목차를, '초록\_원문'과 '초록\_한글'은 그 자료의 초록을 각각 원문과 한글로 보인 것이다. '해제'는 그 자료를 소개하는 내용으로써 한글로 작성된다. '키워드\_원문'과 '키워드\_한글'은 그 자료가 가진 키워드(keyword)들을 각각 원문과 한글로 나타낸 것이다. 'PDF 파일'은 그 자료의 전체 내용을 담고 있는 PDF(Portable Document Format) 파일이며, '텍스트 파일'은 그 PDF 파일을 텍스트(text) 형태로 전환한 파일이며, '표지 이미지 파일'은 그 자료의

표지를 이미지(image)로 가지는 파일이다. 이들 각각은 여러 파일들로 구성될 수 있다.

본 논문은 IPR 자료의 상세 정보에서 '키워드\_원문'과 '키워드\_한글'을 키워드 정보라 하며, 'PDF 파일', '텍스트 파일', '표지 이미지 파일'을 대용량 파일 정보라 하고, 키워드 정보와 대용량 파일 정보를 제외한 나머지를 일반 정보라 한다.

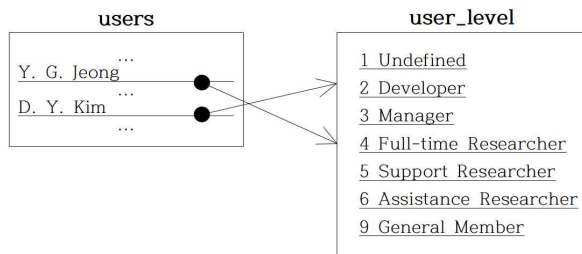
## 3. IPR DB의 구축 환경

IPR DB를 구축하기 위하여 개발자가 개인용 컴퓨터에서 구동하는 데이터베이스 구축 프로그램을 개발하고, 각 연구원은 그 프로그램을 이용하여 자신의 컴퓨터에 자료를 입력한 후, 주기적으로 자료를 취합하는 방법이 있을 수 있다. 이 방법에서 각 연구원의 데이터베이스는 연구원들 간에 공유되지 않으므로 IPR DB가 가져야 할 쟁점들을 어느 것도 만족하지 못한다.

우리는 특정 웹 서버에 IPR DB 구축 프로그램을 두고, 그 웹 서버만이 IPR DB를 가지며, 모든 연구원들은 웹 브라우저를 이용하여 IPR DB의 웹 페이지에 접근하여 실시간으로 업무를 수행하도록 하였다. 우리는 웹 서버의 안정적 관리를 위해 특정 웹 호스팅(hosting) 업체를 선택하였다. 그 업체가 제공하는 서버의 운영체제는 리눅스(Linux)이며, HTTP 웹 서버는 아파치(Apache)[10]이고, 데이터베이스 관리시스템은 MySQL[11]이다. 우리는 스크립트 언어(script language)로 Javascript[12]와 PHP[13]를 사용하였다.

IPR DB는 사용자의 등급 구분을 보관하는 테이블 user\_level과 사용자의 정보를 보관하는 테이블 users를 둔다. 테이블 user\_level은 <사용자등급, 사용자등급정의>의 칼럼(column)들을 가진다. 테이블 users는 <사용자식별자, 비밀번호, 사용자명, 사용자등급, ...>의 칼럼들을 가진다. '비밀번호'는 사용자가 설정한 비밀번호를 암호화하여 보관한다. 사용자가 IPR DB에 연결할 때 제시하는 <사용자 식별자, 비밀번호>를 그 사용자 식별자와 제시한 비밀번호의 암호화된 값을 테이블 users에서 검색한다. [Fig. 1]은 특정 두 사용자들이 각각 전임 연구원과 개발자의 역할임을 나타낸다.

IPR DB에서 데이터베이스 관리시스템에 연결하는 connect 문의 사용자 식별자와 비밀번호는 항상 특정 계정의 사용자 식별자와 비밀번호를 사용하여 데이터베이

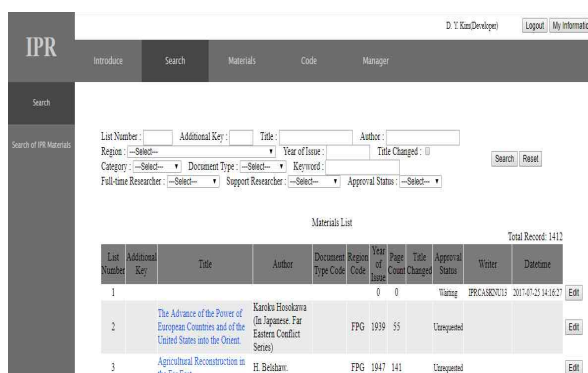


[Fig. 1] Table users and user\_level

스 세션(session)을 형성하며, 매번 사용자가 요구한 작업을 종료할 때는 그 연결을 해제하는 disconnect 문을 사용하여 데이터베이스 세션을 종료한다. 이 작업은 사용자가 IPR DB에 연결할 때의 <사용자 식별자, 비밀번호>의 확인 절차는 물론이고 그 이후 각 사용자 요구의 수행에도 동일하게 적용된다.

우리는 사용자와 IPR DB 간의 로그인(login)에서 로그아웃(logout)까지의 세션을 데이터베이스 세션과 구분하기 위해 웹 세션이라 한다. 웹 세션은 IPR DB에 로그인 시에 <사용자식별자, 사용자이름, 사용자등급>의 정보를 그 웹 세션의 세션 변수들에 보관하며 로그아웃 시에 그들을 초기화한다. 웹 세션의 세션 변수들은 사용자의 웹 브라우저가 아닌 서버에 존재하며, 사용자가 웹 브라우저에 접속하여 작업을 하고 그 웹 브라우저를 종료할 때까지 유지되고 종료하면 사라진다.

IPR DB의 통합 인터페이스는 화면 상단에 대분류를 나타내며, 각 대분류에 대한 중분류들을 화면 좌단에 두어 선택할 수 있게 하며, 각 대분류를 선택하면 그 대분류의 대표 중분류가 묵시적으로 선택되며, 선택된 대분류의 중분류 화면이 화면 중앙에 나타난다. 로그인 후의 최초 대분류를 '자료검색'으로 중분류를 '검색'으로 자동 설정하였다. [Fig. 2]는 그 통합 인터페이스를 나타낸다.



[Fig. 2] The Integrated User Interface of the IPR Database

#### 4. IPR 자료의 상세 정보 구축

IPR DB는 IPR 자료의 기본 정보를 테이블 publications\_abstract에 보관하며 상세 정보를 여러 테이블들에 분할하여 저장한다. 일반 정보는 테이블 publications, 키워드들은 테이블 keyword\_list에 보관한다. 대용량 파일들은 운영체제의 파일들로 보관하며 그 파일들의 경로(path)를 유형별로 테이블 pdf\_file, txt\_file, cover\_image\_file에 보관한다. IPR DB는 상세 정보의 입력이력 정보와 자료 승인의 승인이력 정보를 테이블 makeup\_history와 approval\_history에 각각 보관한다.

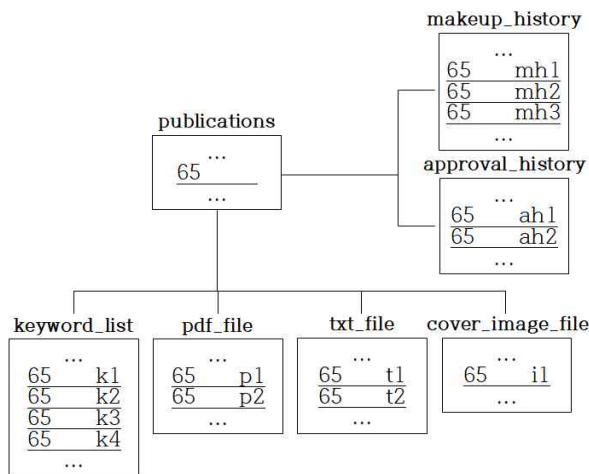
IPR DB 구축의 초기에 테이블 publications는 <일련번호>를 주키(primary key)로 가졌다. 그러나 일련번호 1271의 IPR 자료가 서로 다른 제목과 저자의 3 개의 세부 자료들로 되어 있음을 알게 됨으로써 IPR 자료 목록의 자료인 원 자료와 그 세부 자료들 모두를 수용하도록 칼럼 '부가번호'를 추가하여 <일련번호, 부가번호> 쌍을 주키로 설정하였다. 세부 자료가 있는 경우, 원 자료는 기본 정보 이외의 상세 정보를 가지지 않는다. 테이블 publications는 테이블 publications\_abstract의 내용으로 초기화된 후 전임 연구원들에 의해 변경된다.

하나의 IPR 자료에 대한 테이블 publications의 '제목\_원문'이 테이블 publications\_abstract의 '제목'과 다른 경우가 발생할 수 있다. 이는 IPR 자료 목록의 원 제목에 오류가 있거나 원 제목과 달리 제목을 설정해야 할 필요가 있는 경우이다. 그 경우, 공동 연구원이 그를 쉽게 파악하도록 테이블 publications에 칼럼 '제목\_변경'을 추가하였다. 그 칼럼은 IPR 자료를 저장할 때 저자의 필드가 변경되면 테이블 publications\_abstract와 비교하여 다를 경우 참(TRUE)으로 설정된다.

[Fig. 3]은 일련번호 65의 IPR 자료가 하나의 일반 정보 레코드를 정점으로 키워드, 대용량 파일, 그리고 입력이력과 승인이력의 레코드들이 하나의 계층을 형성함을 보인다. 그들은 일반 정보의 주키를 외래키(foreign key)로 참조하며 존재 종속관계에 있다.

IPR 자료의 상세 정보의 분류 칼럼들은 동일 구분에 대하여 동일 값을 가져야 한다. 우리는 이를 다음의 방법으로 지원한다.

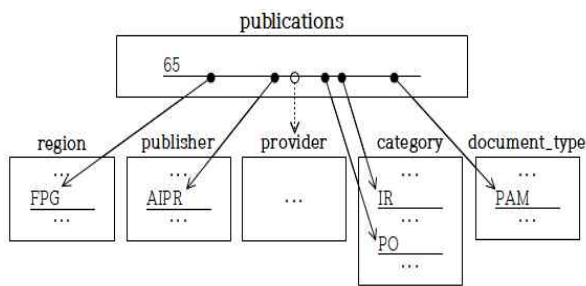
첫 째, 각 분류에 대해 별도의 테이블을 두어 그 분류의 구분들을 보관하며, 각 구분은 <코드값, 실제값, 한글 설명>의 속성을 가지며, 각 테이블에서 코드값을 주키로



[Fig. 3] Existence Dependence Relationship among IPR Database Tables

그리고 실제값을 유일키(unique key)로 설정하였다. 분류 ‘대상지역’은 테이블 region, ‘발행처’는 publisher, ‘제공처’는 provider, ‘대상분야1’과 ‘대상분야2’는 category, ‘문서형태’는 document\_type에 보관한다.

둘 째, 테이블 publications의 분류 칼럼들 각각을 해당 분류 테이블의 주키를 참조하는 외래키로 설정하였으며, IPR 자료의 변경 시에 각 분류 칼럼의 값은 해당 분류의 코드 값들 중의 하나 또는 널(NULL, 값없음)을 선택하도록 통합 인터페이스를 작성하였다. 예를 들어, [Fig. 4]는 일련번호 65의 IPR 자료의 그 칼럼들의 코드 값들을 나타낸다. ‘제공처’의 경우 그 값은 널이다.



[Fig. 4] References using Code Values

셋 째, IPR 자료 목록은 ‘발행처’를 위한 실제값 ‘Hansung Publishing Company’와 ‘Harper & Brothers’에 대한 코드값을 각각 ‘HP’와 ‘Hp’로 가진다. 우리는 그 코드값들을 그대로 유지하면서 질의에서 이들을 구분하기 위해 ‘발행처’의 유형을 VARBINARY로, 그 외 분류의 유형을 VARCHAR로 하였다.

넷 째, 통합 인터페이스에 각 분류의 구분들을 추가, 수정, 삭제할 수 있는 화면을 제공한다. [Fig. 5]는 분야 분류의 관리 화면으로써 새로운 분야를 추가할 수 있으며 테이블 category의 분야들을 나열하여 각 항목 우측의 버튼(button)을 이용해 수정과 삭제를 할 수 있다.

Category Code      Category(English)      Category(Korean)      Add

---

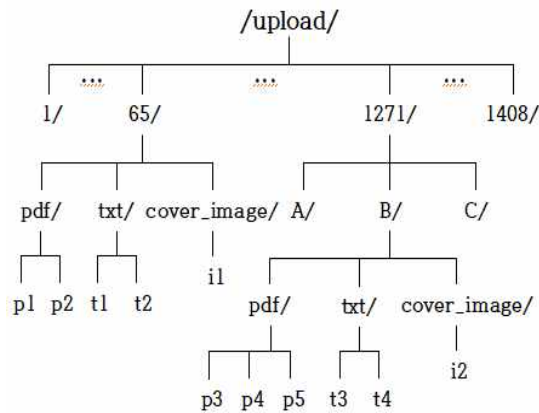
Category List Total: 8

Category Code	Category(English)	Category(Korean)	Edit	Delete
SC	Social and Cultural Affairs	사회문화	Edit	Delete
PO	Politics	정치	Edit	Delete
LA	Law	법	Edit	Delete
IR	International Relations	국제관계	Edit	Delete
HG	Human Geography	인문지리	Edit	Delete
HI	History	역사	Edit	Delete
EC	Economy	경제	Edit	Delete
ETC	.etc	기타	Edit	Delete

[Fig. 5] User Interface for the Code Management of Target Field Classification

테이블 keyword\_list는 <일련번호, 부가번호, 영어\_한글, 키워드순서, 키워드>로 구성된다. 각 키워드는 ‘영어\_한글’로 영어 또는 한글인지를 나타내며, ‘키워드순서’로 그 키워드가 그 자료의 영어 또는 한글 키워드 리스트에서 몇 번째 키워드인지를 나타내며, ‘키워드’로 값을 나타낸다. <일련번호, 부가번호>의 쌍은 테이블 publications를 참조하는 외래키이다.

IPR DB는 별도의 디렉토리(directory) upload를 두고 그 아래에 각 IPR 자료의 일련번호의 디렉토리를 두며 그 아래에 각 파일 유형별 디렉토리를 두어 그 자료의 파일들을 저장한다. 부가번호를 가진 IPR 자료의 경우 일련번호 아래에 그 부가번호의 디렉토리를 둔다. 대용량 파일 경로를 보관하는 테이블들은 <일련번호, 부가번호, 파일순서, 파일경로, 업로드시각>으로 동일하다. 각 파일의 경로는 해당 유형의 테이블에 하나의 레코드로 저장된다. ‘파일순서’는 그 파일이 그 IPR 자료에서 해당 유형의 몇 번째 파일인지를 나타내며, ‘파일경로’와 ‘업로드시각’은 그 파일이 업로드된 경로와 시각을 각각 나타낸다. 이들 각 테이블에서 <일련번호, 부가번호>의 쌍은 테이블 publications를 참조하는 외래키이다. [Fig. 6]은 일련번호 65와 일련번호 1271의 부가번호 ‘B’의 IPR 자료들의 파일들을 나타낸다. 일련번호 65의 IPR 자료의 PDF 파일 “p1”과 “p2”가 그 순으로 업로드 되었다면 그들의 <파일순서, 파일경로>는 각각 <1, “/upload/65/pdf/p1”>, <2, “/upload/65/pdf/p2”>이다.

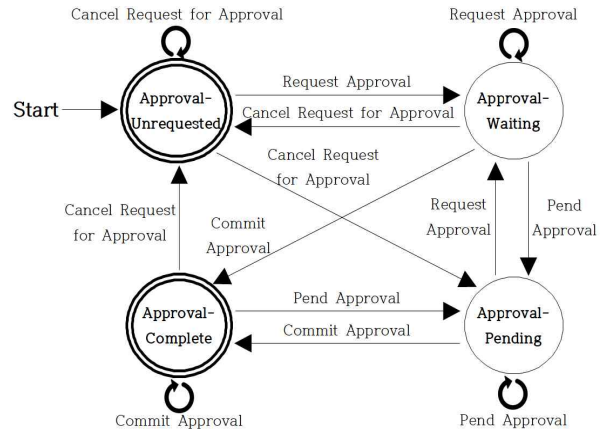


[Fig. 6] File Paths of the Uploaded Files

테이블 publications는 IPR 자료의 책임 입력을 위해 칼럼 ‘전임연구원\_식별자’와 ‘마지막 입력시각’을 가진다. ‘전임연구원\_식별자’는 IPR 자료가 테이블 publications\_abstract에서 복사되어 초기화될 때 널 값을 가지며, 그 자료를 최초로 수정하는 경우 그를 수행한 전임 연구원의 식별자로 설정되며, 테이블 publications\_abstract에 존재하지 않는 새로운 IPR 자료를 테이블 publications에 생성할 경우 그를 수행하는 전임 연구원의 식별자로 설정된다. 그 값은 설정되면 그 레코드가 삭제될 때까지 변경되지 않는다. IPR 자료의 생성, 수정, 삭제의 요구에 대해 항상 ‘전임연구원\_식별자’를 확인하고 그 값이 널이거나 널이 아닐 경우에는 그 작업을 요구한 전임 연구원의 식별자와 동일한 경우에만 그 작업을 허용함으로써 IPR 자료의 책임 입력을 보장한다. ‘마지막 입력시각’은 전임 연구원이 마지막으로 입력한 시각을 나타낸다. IPR DB는 IPR 자료의 입력이력을 테이블 makeup\_history에 기록한다. 각 입력이력은 <일련번호, 부가번호, 이력순서, 입력시각, 입력IP주소, 수정사항>으로 구성된다. 그 테이블은 IPR 자료의 일반 정보와 키워드 정보의 이력을 가지며 대용량 파일과 승인의 이력을 가지지 않는다.

테이블 publications는 IPR 자료의 책임 승인을 위해 칼럼 ‘공동연구원\_식별자’를 가진다. 전임 연구원은 자신이 입력한 IPR 자료에 대해 “승인신청” 또는 “승인신청취소”를 요구할 수 있으며, 공동 연구원은 전임 연구원이 승인을 신청한 자료에 한하여 “승인보류” 또는 “승인완료”를 수행할 수 있다. [Fig 7]은 승인의 상태전이도이다. 타원은 승인상태, 화살표는 상태 전이, 화살표 위의 레이블(label)은 행위를 나타낸다. IPR DB는 IPR 자료의 승인 이력을 테이블 approval\_history에 기록한다. 승인이력은

<일련번호, 부가번호, 승인이력순서, 사용자식별자, 승인 설명, 직후승인상태, 입력시각>으로 구성된다.



[Fig. 7] State Transition Diagram of Approval

## 5. IPR 자료의 제작 인터페이스

본 장은 통합 인터페이스가 제공하는 기능들 중에서 IPR 자료의 입력, 승인, 검색, 삭제를 중심으로 설계의 중요 사항들을 제시한다.

### 5.1 IPR 자료의 자료입력 화면

자료입력 화면은 “블러오기 폼”, “자료승인 폼”, “자료작성 폼”, “파일처리 폼”의 네 폼(form)으로 구성된다.

“블러오기 폼”은 ‘일련번호’와 ‘부가번호’로 기존 자료를 블러오거나 새로운 자료를 생성할 수 있다. 기존 자료의 일반 정보, 키워드 정보, 입력이력 정보는 “자료작성 폼”, 대용량 파일 경로 정보는 “파일처리 폼”, 그리고 승인이력 정보는 “자료승인 폼”에 배치된다.

“자료작성 폼”은 IPR 자료의 ‘일련번호’와 ‘부가번호’를 제외한 일반 정보와 키워드 정보를 입력 또는 수정할 수 있는 폼이며 변경사항을 간단히 기입할 수 있는 필드와 모든 입력이력을 나열하는 필드를 제공한다. “자료작성 폼”에 입력된 정보는 그 폼의 <저장> 버튼을 클릭하면 <일련번호, 부가번호, 변경된 정보>가 서버로 전송되어 하나의 트랜잭션으로 다음과 같이 수행된다. (1) 트랜잭션의 격리도를 SERIALIZABLE로 설정한다. (2) 변경된 일반 정보가 있다면 그를 테이블 publications에 반영한다. (3) 변경된 키워드 정보가 있다면 테이블 keyword\_list에서 그 자료의 해당 영어 또는 한글의 키워

드들을 모두 삭제한 후 전송된 키워드들을 테이블 keyword\_list에 차례로 하나씩 삽입한다. (4) 전송된 작업 정보를 테이블 makeup\_history에 삽입한다. (5) COMMIT을 수행하여 트랜잭션을 종료한다.

“파일처리 폼”은 대용량 파일의 업로드와 삭제관련 폼이다. [Fig. 8]은 일련번호 65의 IPR 자료가 [Fig. 6]의 상태에서 세 번째 PDF 파일로 자신의 컴퓨터에 있는 특정 파일을 선택한 상태를 보인다.

IPR Materials(PDF File)	Choose File The Future of... Pacific.pdf	Upload	Delete All
	1 p1.pdf		
	2 p2.pdf		
IPR Materials(TXT File)	Choose File No file chosen	Upload	Delete All
	1 t1.txt		
	2 t2.txt		
IPR Materials Cover(Image File)	Choose File No file chosen	Upload	Delete All
	1 i1.png		

[Fig. 8] Form for the File Processing

“파일처리 폼”에서 파일의 업로드는 한 번에 하나의 파일만을 대상으로 하며 그 수행 과정은 다음과 같다. 첫 째, <파일 선택> 버튼을 클릭하여 “업로드할 파일 선택창”을 띄워 업로드할 파일을 선택함으로써 그 파일의 이름이 <파일 선택> 버튼 옆에 표시되고 그 파일의 사용자 컴퓨터에서의 경로를 보관한다. 둘째, <업로드> 버튼을 클릭하여 <그 자료의 일련번호와 부가번호, 파일의 포맷 정보, 선택된 파일과 그 파일의 정보(포맷 정보, 파일명 등)>을 서버로 전송한다. 전송된 파일은 서버의 임시 디렉토리에 저장된다. 셋 째, 서버는 그 파일을 동일 파일명으로 해당 업로드 디렉토리에 저장하고 그 경로를 해당 유형의 테이블에 삽입한다. 넷 째, 사용자 화면에 파일의 업로드가 성공하였음을 보인다. [Fig. 8]에서 파일의 업로드가 완료되면 ‘IPR 자료(PDF 파일)’에 <3, 파일명>이 링크로 추가된다.

“파일처리 폼”에서 파일의 삭제는 한 번에 그 IPR 자료의 특정 유형의 모든 파일들을 대상으로 한다. 그 수행 과정은 다음과 같다. 첫 째, 사용자는 삭제할 파일의 유형에 보인 <모두 삭제> 버튼을 클릭함으로써 <IPR 자료의 일련번호와 부가번호, 파일의 유형 정보>가 서버로 전송된다. 둘째, 서버는 해당 유형의 테이블에서 그 IPR 자료에 관련된 모든 레코드들을 삭제한다. 셋 째, 그 IPR 자료의 업로드 디렉토리의 해당 유형의 디렉토리 아래의 모

든 파일들을 삭제한다. 넷 째, 사용자 화면에 파일의 삭제가 성공하였음을 보인다.

“파일처리 폼”에서 파일의 다운로드 는 나열된 파일 링크를 클릭함으로써 수행된다. 그러면 <일련번호, 부가번호, 파일 유형, 파일의 업로드된 순서> 정보가 서버로 전송되고 그 파일 유형의 테이블에서 그 파일의 경로를 식별하고 그 파일을 조회하여 “내려받기 창”을 나타나게 한다. 그 창에서 “내려받을 위치”를 지정함으로써 해당 파일을 지정한 위치에 내려 받는다.

“자료승인 폼”은 ‘승인상태’, ‘신청/신청취소’, ‘코멘트’의 필드들과 “승인이력 창”으로 구성된다. “승인이력 창”은 그 자료의 승인이력 레코드들을 승인이력순서의 정렬된 순서로 ‘사용자식별자’가 현 사용자의 식별자와 동일할 경우에는 그 창의 우측에 다를 경우에는 좌측에 나타낸다. [Fig. 9]는 전임 연구원이 <승인신청> 버튼을 누른 직후의 모습을 나타내는 것으로 메시지는 <사용자식별자, 시각, “승인신청”>으로 구성됨을 볼 수 있다.

Approval Status	Approval-Urequested	Approval-Waiting	Approval-Pending	Approval-Complete
Request Approval	Request Approval			
Cancel Request for Approval	Cancel Request for Approval			
Comment	Input			
Support Researcher	IPRCASNU13 2017-07-25 05:16:36 Request Approval			Full-time Researcher

[Fig. 9] Data Approval Form of Full-Time Researchers

### 5.2 공동 연구원의 자료승인 화면

공동 연구원들의 자료승인 화면은 “자료승인 폼”과 “자료읽기 폼”의 두 폼으로 구성된다. “자료읽기 폼”은 자료입력 화면의 “자료작성 폼”과 “파일처리 폼”을 합한 것에서 모두 읽기 전용으로 설정되어 있음을 제외하면 동일하다. “자료승인 폼”은 ‘승인상태’, ‘승인보류/승인완료’, ‘코멘트’의 필드들과 “승인이력 창”으로 구성된다.

[Fig. 10]은 [Fig. 9]의 IPR 자료의 승인대기 상태에 대해 공동 연구원이 <승인보류> 버튼을 누르고 ‘코멘트’ 필드에 “제공처가 입력되지 않았고 표지 이미지 파일도 업로드되지 않았습니다.”라는 메시지를 입력한 후 <입력> 버튼을 누른 것에 대해 해당 전임 연구원이 자신의 “자료승인 폼”에서 ‘코멘트’로 “수정 뒤에 다시 승인 신청

하겠습니다.”라는 메시지를 쓴 후 <입력> 버튼을 누른 직후의 그 공동 연구원의 “자료승인 폼”의 모습이다.

[Fig. 10] Data Approval Form of Support Researchers

### 5.3 자료검색 화면

자료검색 화면은 검색 조건을 명시하는 폼과 검색 결과를 나열하는 테이블로 구성된다. 검색 조건은 여러 항목들 간의 논리곱인 AND로, 각 항목의 내부는 띄어쓰기로 논리합인 OR로 수행된다. 제목, 저자, 키워드는 부분 일치(partial match)로, 나머지 항목들은 정확한 일치(exact match)로 수행된다. 검색 조건에 키워드가 포함된 경우, 주어진 키워드가 부분 일치되는 키워드의 순서에 따라 가중치를 두며 그 가중치의 합이 그 자료가 가지는 자료의 값이고 그 값의 감소 순으로 나열한다. 키워드 순서 1은 가중치 10, 2는 9로 차례로 1씩 감소하다가 10 이상은 모두 1의 가중치 값을 부여한다. 그 외의 경우에는 일련번호와 부가번호의 증가 순으로 나열한다. 명시된 검색 조건은 <검색> 버튼을 클릭하면 서버로 전송되어 질의(query)를 구성한 후 그를 수행한다. 우리는 자료검색을 위한 트랜잭션 격리도를 READ UNCOMMITTED로 하였다.

[Fig. 11]에서 제목은 입력한 “china”, 지역분류는 테이블 region에 등록된 값들을 드롭다운 리스트로 나열한 것들 중의 하나를 선택한 “CH: China”, 키워드는 입력한 “자본 market”를 검색 조건으로 한다. 그 검색 조건을 만족하는 IPR 자료는 현재 네 건이며 그들의 키워드 검색의 가중치의 합을 감소 순으로 정렬한 결과이다.

검색결과 테이블의 자료 목록은 현재 질의 결과를 대상으로 각 열의 제목을 클릭할 때마다 그 열의 오름차순 또는 내림차순으로 정렬되도록 하였다. 검색결과 테이블에서 제목에 부여된 링크를 클릭하면 현재 접속한 계정의 등급에 따라 그 자료의 자료입력 화면 또는 자료열람

화면으로 연결된다. IPR 자료의 자료열람 화면은 자료입력 화면과 유사하지만 모두 읽기 전용이다.

[Fig. 11] Retrieval Interface for IPR Materials

### 5.4 IPR 자료의 자료삭제 화면

IPR 자료의 자료삭제 화면은 [Fig. 12]와 같이 현재의 전임 연구원이 작성한 모든 IPR 자료들 중에서 특정 자료 또는 전체를 구하는 폼과 그 결과를 나열하는 검색결과 테이블로 구성된다.

검색결과 테이블에서 각 항목에 대해 수정 또는 삭제를 수행할 수 있다. 수정의 경우 그 IPR 자료의 자료입력 화면으로 이동하며 삭제의 경우 그 IPR 자료의 식별자 서버로 전송되어 그 자료가 삭제된다.

[Fig. 12] Removal Interface for IPR Materials

## 6. 대용량 파일과 트랜잭션

운영체제 파일의 삭제를 포함하는 데이터베이스 트랜잭션은 그 삭제를 미처리 연산(pending actions)으로 처리해야 한다[14,15]. 미처리 연산이란 트랜잭션의 완료가 결정된 후에 그리고 그 트랜잭션의 종료 이전에 수행하



는 연산을 의미한다. 그러나 현재 우리가 사용하는 데이터베이스 관리시스템은 그를 지원하지 않는다. 파일 업로드의 경우 IPR DB와 같이 그 파일이 하나의 운영체제 파일로 생성되면 그 정확성을 데이터베이스 트랜잭션으로 보장할 수 없다.

우리는 데이터베이스에는 경로가 존재하지만 그 경로에 해당하는 운영체제 파일이 존재하지 않는다면 그 경로를 허상 경로(dangling path)라 하며, 데이터베이스와 관련된 운영체제 파일들 중에서 그 경로가 데이터베이스에 존재하지 않으면 그 파일을 무참조 파일(unreferenced file)이라 한다. 무참조 파일의 존재는 정확성에 영향을 미치지 않지만 존재하지 않아야 한다. 그러나 허상 경로는 정확성에 치명적 결함이 된다. 우리는 IPR DB에 허상 경로가 존재하지 않도록 파일 업로드의 경우 파일을 업로드한 후 데이터베이스에 그 파일의 경로를 삽입하며, 파일 삭제의 경우 데이터베이스에서 그 경로를 삭제한 후 그 파일을 삭제한다.

IPR DB에서 대용량 파일 정보를 변경하는 연산은 파일 업로드, 파일 삭제, IPR 자료 삭제, 무참조 파일 삭제의 네 가지이다. 본 장은 그 연산들을 정확히 수행하는 방법을 제시한다.

파일 업로드는 다음과 같이 하나의 트랜잭션으로 수행된다. (1) 트랜잭션 격리도를 SERIALIZABLE로 설정한다. (2) 테이블 publications에서 그 IPR 자료에 대해 UPDATE 로크(U lock)를 잡는다. (3) 그 파일을 저장할 디렉토리를 결정하며, 그 디렉토리가 생성되지 않았다면 그를 생성하며, 임시 디렉토리에 보관된 그 파일을 그 디렉토리로 이동한다. (4) 해당 파일 경로 테이블에서 그 IPR 자료의 레코드들을 읽어 마지막으로 업로드된 파일의 파일 순서를 구한다. 기존 레코드가 없다면 파일 순서를 1로 그렇지 않으면 구한 그 순서의 다음 순서로 파일 순서를 정한다. (5) 해당 파일 경로 테이블에 <일련번호, 부가번호, 파일 순서, 파일 경로, 업로드한 시각>의 정보를 삽입한다. (6) COMMIT을 수행한다.

이 작업은 운영체제 파일의 업로드를 하나의 트랜잭션 범위 내에서 그 IPR 자료의 일반 정보의 레코드에 U 로크를 잡은 상태에서 수행함을 특징으로 한다. 위 작업에서 (3)을 (5)의 다음 순서로 수행하여도 동일 트랜잭션 내이므로 정확성에는 문제가 없다.

파일 삭제는 다음의 두 단계로 수행된다. 단계 1. 아래의 트랜잭션을 수행한다. (1) 트랜잭션 격리도를

SERIALIZABLE로 설정한다. (2) 테이블 publications에서 그 IPR 자료에 대해 U 로크를 잡는다. (3) 해당 파일 경로 테이블에서 그 IPR 자료에 관련된 모든 레코드들을 삭제한다. (4) COMMIT을 수행한다. 단계 2. 그 IPR 자료의 업로드 디렉토리의 해당 유형의 디렉토리 아래의 모든 파일들을 삭제한다.

파일의 삭제에서 단계 1을 성공적으로 수행하였으나 단계 2를 완료하지 못하면 그 파일들은 무참조 파일이 된다. 단계 2의 운영체제 파일의 삭제를 단계 1의 (2)와 (4) 사이의 어디에 두어도 상황은 동일하다. 그러나 단계 1과 단계 2의 순서를 바꾸거나 운영체제 파일의 삭제를 단계 1의 (1)과 (2) 사이에 둔다면 허상 경로의 발생 가능성이 존재한다.

IPR 자료 삭제는 두 단계로 수행된다. 단계 1. 아래의 트랜잭션을 수행한다. (1) 트랜잭션 격리도를 SERIALIZABLE로 설정한다. (2) 테이블 publications에서 그 IPR 자료에 대해 U 로크를 잡는다. (3) 테이블 keyword\_list, pdf\_file, text\_file, cover\_image\_file, makeup\_history, approval\_history에서 그 자료의 레코드들을 삭제한다. (4) 테이블 publications에서 그 자료의 레코드들을 삭제한다. (5) 그 자료의 '일련번호'의 자료가 테이블 publications\_abstract에 있다면 그를 복사하여 테이블 publications에 삽입한다. (6) COMMIT을 수행한다. 단계 2. 그 IPR 자료의 업로드 디렉토리의 모든 파일 디렉토리 아래의 파일들을 삭제한다.

IPR 자료의 삭제에서 단계 1을 성공적으로 수행하였으나 단계 2를 완료하지 못하면 그 파일들은 무참조 파일이 된다. 운영체제 파일의 삭제를 단계 1의 (2)와 (5) 사이의 어디에 두어도 상황은 동일하다. 그러나 단계 1과 단계 2의 순서를 바꾸거나 운영체제 파일의 삭제를 단계 1의 (1)과 (2) 사이에 둔다면 허상 경로의 발생 가능성이 있다. 단계 1에서 테이블 publications의 종속 테이블들에서 그 IPR 자료의 관련 레코드들을 삭제한 후에 테이블 publications에서 그 IPR 자료의 레코드들을 삭제한다. 그 순서를 바꾸면 참조 무결성 위배(referential integrity violation)의 오류가 발생한다.

무참조 파일 삭제는 관리자 화면의 <무참조 파일 삭제> 요구로 다음과 같이 하나의 트랜잭션으로 수행된다. (1) 트랜잭션 격리도를 SERIALIZABLE로 설정한다. (2) 테이블 publications의 모든 레코드들에 U 로크를 설정한다. (3) 테이블 publications에 저장된 일련번호의 최댓값

을 구한다. (4) 일련번호 1부터 최댓값까지 각 일련번호에 대하여 아래 작업을 차례로 수행한다. 4.1) upload 디렉토리에 그 일련번호의 디렉토리가 있는지 검사한다. 없다면 그 일련번호의 작업을 종료하며 있다면 나머지를 수행한다. 4.2) 그 일련번호 디렉토리의 하위 디렉토리의 상황을 기억한다. 4.3) 파일 경로 테이블들에서 그 일련번호의 모든 경로들을 구하여 기억한다. 4.4) 4.2에서 기억한 경로들 중에서 4.3에서 구한 경로들에 존재하지 않는 경로의 파일들을 삭제한다. (5) COMMIT한다.

무참조 파일들을 삭제하는 트랜잭션은 IPR 데이터베이스의 테이블들의 어떠한 레코드도 변경하지 않는다. 다만 대용량 파일과 관련하여 다른 수정하는 트랜잭션들의 작업을 잠시 블록(block)할 뿐이다.

## 7. 논의 및 결론

학제간 연구를 통한 IPR DB 구축의 쟁점들에 대한 우리의 기여는 다음과 같다. 첫 째, 웹 브라우저를 이용하여 IPR DB에 접근함으로써 접근성을 높였다. 둘째, 통합 인터페이스를 제공함으로써 IPR DB 구축의 편리성을 높였다. 셋 째, IPR DB의 완전성을 높이기 위해 IPR 자료의 상세 정보에 대한 책임 입력과 책임 승인을 지원하였다. 넷 째, 통합 인터페이스의 틀 내에서 원활한 대화형 승인 과정을 지원하였다. 다섯 째, 다양한 데이터베이스 기법들과 정렬 기능을 동원하여 IPR DB의 정확성을 보장하였다. 특히, 트랜잭션 격리도와 IPR DB의 존재 종속관계의 레코드들의 특징을 이용한 동시성 제어를 통하여 운영체제 파일과 그 파일의 경로 간의 정확성을 보장하였다.

추후 연구로 내용 검색과 IPR DB의 다운로드와 업로드를 지원하며, IPR 자료의 저작권 관련 작업을 수행할 예정이다.

## REFERENCES

- [1] Institute of Pacific Relations, IPR Publications on the Pacific, 1925-1952; a catalog of the publications of the IPR national councils and the International Secretariat, 1 East 54th St., New York 22, N.Y., 1953.
- [2] K. J. Lee, "Political Meaning and Utilization Plan of IPR Materials," in Proceedings of the 19<sup>th</sup> International Conference of Asia Research Center, *IPR Materials - Their Bibliographical Significance*, pp. 5-20, May 17, 2017.
- [3] J. W. Kim, "The Bibliographical Significance of IPR Materials for the Study of 'Modern History' in Western Scholarship," in Proceedings of the 19<sup>th</sup> International Conference of Asia Research Center, *IPR Materials - Their Bibliographical Significance*, pp. 53-75, May 17, 2017.
- [4] S. Y. Cha, "Database and Copyright Issues - Focused on IPR Cases," in Proceedings of the 19<sup>th</sup> International Conference of Asia Research Center, *IPR Materials - Their Bibliographical Significance*, pp. 98-119, May 17, 2017.
- [5] M. G. Lee, "A Study on the Classification Scheme of Cultural Resource in ACIA", Journal of the Korean Library and Information Science Society, Vol. 49 No. 1, pp. 319-340, 2015.
- [6] C. O. Yoon, "A Study on the Search Interface of Archival Databases in Korea: With a Focus on Search Functions", Journal of Korean Library and Information Science Society, Vol. 42, No. 2, pp. 257-276, 2011.
- [7] J. Cho, J. W. Lee, "The Study on the Method for Linking Portal Sites and OPAC of University Libraries", Journal of Information Management, Vol. 37 No. 2, pp. 75-92, 2006.
- [8] J. Gray and A. Reuter, Transaction Processing: Concepts and Techniques, Morgan Kaufmann Publishers, Inc., 1993.
- [9] R. Elmasri and S. B. Navathe, Database Systems: Models, Languages, Design, and Application Programming, Sixth Edition, Pearson, 2011.
- [10] E. L. Thompson, H. E. Goodman, S. D. Nowicki, and A. Cove, Professional PHP5, Wrox, 2004.
- [11] L. Welling and L. Thomson, PHP and MySQL Web Development (5th Edition), Sams Publishing, 2016.
- [12] N. C. Zakas, Professional JavaScript for Web Developers, Wrox, 2011.

- [13] E. Naramore, J. Gerner, Y. L. Scouarnec, J. Stolz, and M. K. Glass, *Beginning PHP5, Apache, and MySQL Web Development*, Wrox, 2005.
- [14] C. Mohan, D. Haderle, B. Lindsay, H. Pirahesh, and P. Schwarz, "ARIES: a transaction recovery method supporting fine-granularity locking and partial rollbacks using write-ahead logging," *ACM Transactions on Database Systems*, Vol. 17, No. 1, pp. 94-162, 1992.
- [15] J. H. Park, Y. C. Park, "An Efficient Scheme of Performing Pending Actions for the Removal of Database Files," *Journal of KISS : Database*, Vol. 28, No. 3, pp. 494-511, 2001.

저자소개

김 동 용(Dong Yong Kim)

[정회원]



- 2016년 2월 : 경북대학교 경제통상학부 (이학사)
- 2016년 3월 ~ 현재 : 경북대학교 IT대학 컴퓨터학부 석사과정

<관심분야> : 데이터베이스, 데이터마이닝, 미시경제학

박 영 철(Young Chul Park)

[정회원]



- 1977년 2월 : 서울대학교 전기공학부 졸업 (학사)
- 1981년 8월 : 서울대학교 전기공학부 졸업 (석사)
- 1986년 6월 : Northwestern University (전산학, 석사)

- 1989년 12월 : Northwestern University (전산학, 박사)
- 1981년 9월 ~ 1993년 2월 : 울산대학교 전자계산학과 부교수
- 1990년 3월 ~ 1992년 2월 : 울산대학교 중앙전자계산소 소장
- 1991년 10월 ~ 1992년 2월 한국전자통신연구소 초빙연구원
- 1992년 3월 ~ 1994년 2월 한국정보과학회 데이터베이스연구회 운영위원
- 1993년 ~ 현재 : 경북대학교 IT대학 컴퓨터학부 교수

<관심분야> : 데이터베이스 시스템, 멀티미디어 데이터베이스, XML 데이터베이스, 트랜잭션 처리 등