

# 콘텐츠 연계를 통한 주제기반 커뮤니티 모델 개발 연구 -생명과학 분야를 중심으로-

안부영, 최선희, 신용주, 김순영  
한국과학기술정보연구원 콘텐츠융합팀  
e-mail:{ahnyoung, sunny.choi, yjshin, maya}@kisti.re.kr

## A Study on Development of Subject-based Community Model by Link of Content -Focused on Life Science-

Bu-Young Ahn, Seon-Heui Choi, Yong-Ju Shin, Soon-Young Kim  
Content Convergence Team, Korea Institute of Science & Technology Information

### 요 약

국내의 연구자들은 각자의 분야에서 다양하고 중요한 연구를 수행하면서 그 연구결과물을 생산하고 있다. 연구결과물의 형태는 학회지 및 학술대회 논문, 연구보고서, 특허, 연구노트, 세미나 발표자료, 학교교재, 신문 및 잡지의 기사 등 매우 다양하다. 이런 다양한 연구결과물을 같은 학문 분야, 같은 주제의 연구자들끼리 서로 공유하고 교환하기 위해서는 정보의 자유로운 이용에 근거한 커뮤니티 환경이 필요하다. 이에, 국가 과학기술정보 유통기관인 한국과학기술정보연구원(KISTI)에서 보유하고 있는 문헌 콘텐츠와 사실 콘텐츠를 주제별로 분류하고 재가공하여 특정 주제분야 전문 연구자들을 위한 오픈 아카이빙, 오픈 액세스 개념을 적용한 커뮤니티 모델을 개발하여 제공하고자 한다. 본 커뮤니티 모델은 요즘들어 가장 많은 연구가 진행되고 있는 생명과학 분야의 연구결과물을 중심으로 개발하였다. 커뮤니티 모델을 개발하기 위하여 1) KISTI가 보유하고 있는 콘텐츠 현황을 조사하고, 2) 그 중에서 생명과학분야 콘텐츠의 형태와 특성을 분석하고, 3) 연구자들이 연구결과물을 자유롭게 업로드/다운로드할 수 있는 웹 환경의 플랫폼을 설계하였다.

### 1. 서론

현대는 정보의 폭발적인 증가로 인하여 정보 홍수속에서 살고 있다고 해도 과언이 아니다. 학문분야 또한 학문과 학문의 결합 등으로 새로운 복합학문분야가 등장하고 있으며 이런 많은 학문분야에서 연구하는 연구자들의 연구결과물도 무수히 생산되고 있는 실정이다. 이렇게 생산된 연구결과물(학회지 및 학술대회 논문, 연구보고서, 특허, 연구노트, 세미나 발표자료, 학교교재, 신문 및 잡지의 기사 등)은 정보기술과 인터넷의 발달로 인해 과학기술 또는 인문사회분야 등 다양한 포털사이트에서 산재되어 서비스되고 있다.

그러나 특정 전문분야의 연구결과물만을 검색하고자 하는 연구자들에게는 이런 포털사이트는 원하지 않는 분야 정보까지 찾아주기에 도리어 불편함을 주고 있다. 이런 불편을 해소하고자 요즘들어 시맨틱웹과 정보서비스를 접목한 정보서비스 연구도 진행되고 있지만 국내 정보서비스 플랫폼에 적용된 사례는 아직은 없다[1]. 이에 본 논문에서는 요즘들어 전세계적으로 가장 많은 연구가 진행되고 있는 생명과학 분야를 중심으로 기 구축된 KISTI 콘텐츠를 활용하여 오픈 아카이빙 및 오픈 액세스 개념을

적용한 주제기반 커뮤니티 모델을 제안하고자 한다

### 2. KISTI 콘텐츠 현황

#### 2.1 문헌 콘텐츠 구축 현황

KISTI에서 수집, 가공, 구축한 문헌 콘텐츠는 (표 1)과 같이 학술논문, 연구보고서, 학위논문 등이 있다. 학술논문 데이터베이스는 국내외 학술지와 학술대회 논문정보의 서지사항 및 전자원문 링크정보를 제공하고 있다. INSPEC (Information Services for the Physics and Engineering Communities) 데이터베이스는 영국 공학기술학회에서 제작한 물리, 전기전자, 컴퓨터분야 데이터베이스이며, FSTA(Food Science & Technology Abstracts)는 영국 IFIS(International Food Information Service)가 제공하는 식품관련한 유일한 데이터베이스이다. 연구보고서는 국가에서 예산을 지원받아 수행한 연구보고서 및 미국정부 예산지원을 받아 수행한 기술보고서중에서 KISTI가 소장한 자료로 구분된다. 학위논문은 국회도서관에서 수집, 축적한 국내 각 대학에서 발표된 석박사 학위논문 데이터베이스로 국회도서관 서비스 내용과 동일하다.

(표 1) KISTI 문헌 콘텐츠 구축 현황(2008년 8월)

구분	데이터베이스 내용	수록기간	구축건수
학술지	국내 학술지	1948 ~	478,922
	해외 학술지(영미권)	1991 ~	39,529,281
	해외 학술지(중국, 일본)	2003 ~	2,621,927
회의자료	국내 학술대회 발표논문	1972 ~	213,868
	해외 학술대회 발표논문	1993 ~	6,881,628
INSPEC	물리, 전기, 전자, 컴퓨터 분야	1969 ~	9,977,921
	식품과학 분야	1969 ~	793,567
서지	저널 서지정보		63,175
	프로시딩 서지정보		203,092
학위논문	국내 석박사 학위논문	1945 ~	1,072,406
연구보고서	국가 연구개발 보고서	1983 ~	115,911
	NTIS 보고서	1995 ~	146,421

2.2 사실 콘텐츠 구축 현황

사실정보 또는 사실정보 데이터베이스(Factual DB)란 화합물 구조, 물질 특성, 유전정보, 생물종 정보 등과 같이 과학기술 분야의 실물에 관한 정보를 총칭하는 개념이다 [2]. 사실정보는 그 분야의 전문가가 아니면 생산할 수 없는 정보이기에 KISTI에서는 외부전문가의 도움을 받아 (표 2)와 같이 사실 콘텐츠를 구축하여 제공하고 있다.

(표 2) KISTI 사실 콘텐츠 구축 현황(2008년 8월)

분야	데이터베이스 내용	수록기간	구축건수
물리·화학	화합물정보(ChemDB)	2000 ~	2,579,458
	무기결정구조(ICSD)	1913 ~	93,062
	플라즈마 물질	1995 ~	43,370
인체	디지털 코리언(Digital Korean)	2003 ~	-
	한국인의 인체영상(Visible Korean)	2000 ~	-
생명정보	유전자/단백질 정보	2002 ~	2억5천만건
	생물다양성정보	2000 ~	1,140,000
천문	한국의 표준연력(918-2050)	2000	-
	고천문(역사, 사상, 의기 등)	2000 ~	3,633
가상과학관	가상과학학관	1999 ~	1,948,740
	화석(Fossil) 박물관	1999 ~	850점
	패류(Shell) 박물관	1999 ~	950점
	중국 CNIC 가상과학관	2002 ~	-

2.3 특허 콘텐츠 구축 현황

KISTI에서는 (표 3)과 같이 전세계 특허의 80% 이상을 차지하고 있는 미국, 유럽, 일본, 국제, 한국 특허를 중심으로 데이터베이스를 구축하여 제공하고 있다.

(표 3) KISTI 특허 콘텐츠 구축 현황(2008년 8월)

구분	수록 내용	수록기간	구축건수
한국특허	공개/등록특허, 공개실용신안, 등록실용신안, 디자인의장에 대한 서지사항, 초록, 청구범위 전문	1947 ~	2,970,872
미국특허	1976년부터 공개/등록된 특허를 중심으로 서지사항, 초록, 청구범위	1976 ~	3,710,861
일본특허	1976년부터 공개된 특허를 중심으로 서지사항, 초록, 청구범위	1976 ~	8,019,137
유럽특허	1976년부터 공개된 특허를 중심으로 서지사항, 초록, 청구범위	1976 ~	1,948,740
국제특허	국제특허청(WIPO)를 통해 1976년부터 공개된 특허를 중심으로 서지사항, 초록, 청구범위	1976 ~	1,486,226

3. KISTI 콘텐츠 형태 분석

3.1 문헌 콘텐츠 형태 분석

KISTI 문헌 콘텐츠에는 생명과학 주제가 분류되어 있다. 이 중에서 연구자들이 가장 많이 검색하는 국내 학술지 논문의 데이터베이스 항목은 (표 4)와 같다.

(표 4) 국내 학술지 논문 데이터베이스 항목

한글명칭	영문명칭	코드
관리번호	Control Number	CN
논문명	Title	TI
대등논문명	Parrel Title	TIP
저자	Author	AU
자료유형		
원문형태		
학술지명	Journal Title	SO
발행기관	Publisher	PB
발행년	Publish Year	PY
언어	Language	LA
권호정보	Volume Issue	VO/IS
페이지	Pages	PG
ISSN	ISSN	SN
ISBN	ISBN	SB
주제분류	Classification Code	DC
키워드	Keyword	KW
저작권(출처)	Copyright	
원문접근경로	Source Access Path & Right	
초록	Abstract	AB

3.2 사실 콘텐츠 형태 분석

KISTI 사실 콘텐츠에는 생명정보 및 생물다양성정보가 포함되어 있다. 생명정보중에서 전세계적으로 가장 많이 활용되는 유전자정보 데이터베이스인 GenBank의 구성 항목은 (표 5)와 같다.

(표 5) 유전자정보(Genbank) 데이터베이스 항목

한글명칭	영문명칭	코드
유전자 위치	LOCUS	
생물학적 특성	DEFINITION	
고유 식별자	ACCESSION	
등록번호	VERSION	
서열 핵심어	KEYWORDS	
일반명/학명	SOURCE	
종정보	ORANISM	
인용논문정보	REFERENCE	
저자	AUTHORS	
논문제목	TITLE	
저널명	JOURNAL	
주석 및 해설	COMMENT	
생물학적 특징	FEATURES	
서열의 원소스	ORIGIN	

3.3 특허 콘텐츠 형태 분석

KISTI 특허 콘텐츠는 생명과학 주제분류는 되어 있지 않고 국제특허분류(IPC)의 대분류로만 분류되어 있다. KISTI에서 제공하는 특허정보 중에서 미국특허 데이터베이스 구성항목은 (표 6)과 같다.

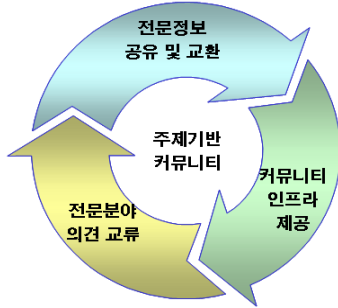
(표 6) 미국특허 데이터베이스 항목

한글명칭	영문명칭	코드
관리번호	Control Number	CN
발명명칭	Title of Invention	TI
특허청구분	Country Code	CY
특허구분	Patent Type	PT
상태	Document Kind	REG
국제특허분류	International Patent Classification	IC
미국특허분류	US Patent Classification	USC
출원번호	Application Number	AN
출원일자	Application Date	AD
공개번호	Publication Number	UN
공개일자	Publication Date	UD
등록번호	Registration Number	RN
등록일자	Registration Date	RD
우선권번호	Priority Number	PRN
국제출원번호	International Application Number	IAN
국제출원일자	International Application Date	IAD
P371	USC 371 (PCT) Date	
P101	USC 102 (e) Date	
국제공개번호	International Unexamined Number	IUN
국제공개일자	International Unexamined Date	IUD
대표도면	Drawing	
발명자	Inventor	IN
발명자주소	Inventor Address	INA
발명자국적	Inventor Country	INC
출원인	Applicant	PA
출원인국적	Applicant Country	
출원인주소	Applicant Address	PAA
대리인	Agent	AG
초록	Abstract	AB
대표청구항	Claim	CM
청구항수	Claim Number	BSC
심사관	Examiner	
관련자료	Related Application Data	
지정국	Designated Country	DS
원문		

#### 4. 커뮤니티 모델 설계

##### 4.1 오픈 아카이빙 서비스

오픈 아카이빙 서비스에서는 폐쇄적인 웹이 아니라 개방적인 웹을 이용하여 공개된 연구학술정보를 누구나 활용 가능하여야 한다. 누구나 본인에게 필요한 정보를 찾을 수 있고, 대가없이 제공할 수 있는 플랫폼을 구축하여, 이 플랫폼을 이용하는 연구자 스스로 커뮤니티를 발전적인 방향으로 진화시킬 수 있는 환경을 제공하고자 한다.



(그림 1) 주제기반 커뮤니티 개념도

웹 환경에서 자료를 공유하려면 이용자들이 스스로 자료를 업로드하는 기능이 필요하다. 어떤 형태의 자료이든 업로드하여 원본 파일을 첨부하거나 또는 원본 파일이 있는 소재정보(URL, 이메일 등)를 공유할 수 있도록 (표 7)과 같이 공유자료 메타 데이터베이스 스키마를 설계하였고, (그림 2)와 같이 자료 업로드 화면을 설계하였다.

(표 7) 공유 메타 DB 스키마

컬럼명	내용	데이터타입	필수
c_num	제어번호	int(10)	Y
name	자료 제목	varchar(50)	Y
c_type	자료 유형	varchar(20)	N
issn_isbn	ISSN 또는 ISBN	char(10)	Y
summary	요약	text	Y
pub_date	발간일	date	Y
lang	언어정보	varchar(10)	N
keywords	키워드	varchar(50)	N
g_code	분류코드	char(5)	Y
url_addr	원문 URL	varchar(50)	Y
r_doi	원문 DOI	varchar(20)	N
r_gendate	등록일자	date	Y
r_moddate	수정일자	date	N
r_generator	등록자	char(10)	Y

(그림 2) 자료 업로드 화면

##### 4.2 오픈 액세스 서비스

정보의 자유로운 이용을 기반으로 오픈 액세스가 가능한 생명과학 관련 저널과 아티클 메타데이터를 수집, 구축하여 서비스함과 동시에, 본 커뮤니티를 이용하는 연구자들이 서로 필요한 생명과학 정보를 축적하면서, 관련 주제별 전공자들의 개방적인 질문과 답변을 주고받을 수 있도록 플랫폼을 설계하였다.

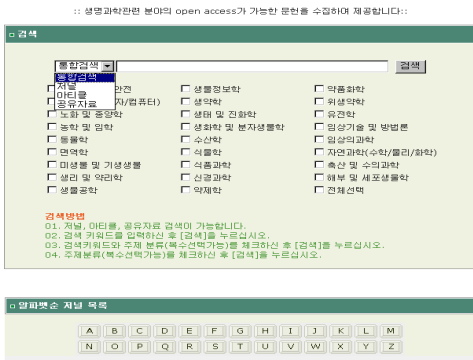
생명과학 이용자들이 생명과학 분야의 정보를 자유롭게 이용할 수 있도록 KISTI가 기 구축한 콘텐츠중에서 생명과학분야 논문만을 추출하여 메타 데이터베이스를 구축하고자 (표 8)과 같이 논문 스키마를 설계하였다. 이외에도 이용자들이 스스로 업로드하는 자료를 저장·관리할 수 있는 공유자료(단행본, 세미나자료 등) 테이블, 이용자

들간 교류한 의견을 저장·관리하기 위한 BBS 테이블, 저널정보를 관리할 수 있는 저널 테이블 등을 설계하였다.

(표 8) 논문 메타 DB 스키마

컬럼명	설명	데이터타입	필수
c_num	제어번호	int(10)	Y
name	논문 제목	varchar(50)	Y
author	주저자	varchar(20)	Y
sub_author	공동저자	varchar(50)	N
j_name	수록처(저널명)	varchar(50)	Y
j_spec	수록처 권호정보	varchar(30)	N
pub_plc	발행국가(지역)	varchar(20)	N
pub_org	발행기관	varchar(50)	Y
pub_date	발행일	date	Y
lang	언어정보	varchar(10)	N
summary	요약	text	N
keywords	키워드	varchar(50)	N
g_code	분류코드	char(5)	Y
url_addr	원문 URL	varchar(50)	Y
r_doi	원문 DOI	varchar(20)	N
r_gendate	등록일자	date	Y
r_moddate	수정일자	date	N
r_generator	등록자	char(10)	Y

(그림 3)은 이용자들 스스로 업로드한 자료와 기존 KISTI에서 구축하여 제공중인 콘텐츠(국내학회논문, 해외 학술지 논문, 학술대회 논문 등) 중에서 생명과학 분야의 콘텐츠 검색화면이다. 생명과학분야의 세부주제별 검색과 알파벳순 검색 및 자료 유형별 통합검색이 가능하도록 화면을 설계하였다.



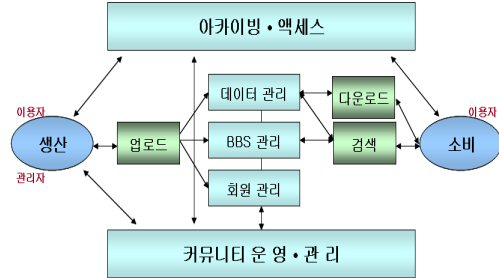
(그림 3) 통합검색 화면

### 4.3 커뮤니티 관리 시스템

커뮤니티 관리시스템은 자료등록을 위한 심사와 데이터베이스 수정 기능 및 회원관리 기능이 있으며, 각종 연구결과물이 등록되면 내용을 검토하고, 수정사항이 있으면 수정하여 데이터베이스에 최종적으로 저장되도록 (그림 4)의 시스템 구성도와 같이 설계하였다.

또한 회원 레벨(관리자, 운영자, 일반 이용자)을 조정할 수 있는 기능과 정보 이용 현황을 모니터링하는 통계기능

을 포함하였다. 관리시스템을 운영·관리하는 관리자는 자료의 등록, 심사, 수정 등을 권한과 회원의 등급을 조정할 수 있는 권한도 가지고 있다.



(그림 4) 관리시스템 구성도

### 5. 결론 및 향후 연구방향

지금까지 KISTI에서 제공하고 있는 문헌, 사실, 특허 콘텐츠를 조사 분석하고 이를 연계하여 생명과학 특정 주제 기반 커뮤니티 모델을 설계하였다. 본 커뮤니티 모델은 생명과학 연구자 및 정보전문가의 피드백을 거쳐 기능을 수정 보완 및 구현하여 KISTI 과학기술정보통합서비스 웹사이트(<http://www.ndsl.kr>)를 통해 국내 생명과학 연구자들에게 제공될 계획이다.

본 커뮤니티가 본격적으로 서비스되면 연구자들간의 생명과학 관련 정보수집 및 공유체제 확립, 학술정보 자원의 교환, 가상의 연구·학술공간의 제공 등으로 연구자의 연구효율을 높일 수 있을 것으로 사료된다. 또한 본 커뮤니티 모델을 기반으로 생명과학 주제분야뿐만 아니라 KISTI가 보유하고 있는 전기전자, 기계, 재료 등 타 분야의 콘텐츠 연계를 통한 과학기술 각 주제별 커뮤니티 구축도 가능할 것으로 기대된다.

### 참고문헌

- [1] 정한민, 이미경, 성원경, 박동인, "OntoFrame-K: 협업 연구지원을 위한 시맨틱 웹 기반 지식정보 공유·유통 플랫폼", 한국정보과학회 2006년 한국컴퓨터종합학술대회 논문집B: 100-102, 2006.7.
- [2] 안부영, 이상호, "사실정보 데이터베이스에 관한 고찰", 한국콘텐츠학회지 5(2): 37-46, 2007.12.
- [3] 과학기술정보 통합서비스 데이터베이스, 한국과학기술정보연구원, 2008.8.
- [4] KISTI 과학기술정보 통합서비스 웹사이트, <http://www.ndsl.kr>