DOI:10.5392/IJoC.2010.6.3.010

Design of Subject-based Community Model by Linkage Heterogeneous Content: Focused on Field of Biological Science

Buyoung Ahn[1], Jiyoung Kim[2], Chungshick Oh[3], and Myungsun Lee[1]
Supercomputing Center[1], Knowledge Information Center[2], Information Strategy Team[3]
Korea Institute of Science and Technology Information
335 Gwahangro, Yuseong-gu, Daejeon, Korea

ABSTRACT

Researchers in Korea and elsewhere have carried out a wide variety of important research activities in their respective fields, producing valuable research results. For such diverse research results to be shared and exchanged among researchers working in the same discipline and research subject there needs to be a community environment based on free utilization of information. Against this backdrop, this study seeks to classify and reprocess the reference/factual content owned by the KISTI (Korea Institute of Science and Technology Information), a state-run distributor of information on science and technology, by the different research subjects. It also seeks to develop and provide a community model based on the concepts of open archiving and open access for the researchers specialized in the related fields of research. This community model is developed focusing on the research results from the field of bioscience, where the most extensive studies are currently being conducted. To develop the community model, this study: (a) surveys the current status of the content owned by KISTI; (b) analyzes the patterns and characteristics of biological scientific content among the KISTI-owned content; and (c) designs a web platform where researchers can freely upload/download research results.

Keywords: *Community Model, Content Linking, Factual Content, Literature Content, Patent Content, Biological Science*

1. INTRODUCTION

It is no exaggeration to say that, amid an exponential increase in the volumes of information available, we are now living in a flood of information. A slew of new disciplines are being created, as a growing number of disciplines are converged and interlinked with others, and researchers working on such disciplines are producing countless volumes of research findings. With the development of information technology and the Internet, these research findings (e.g. articles published in academic journals or presented in symposiums, research reports, patents, research notes, seminar presentations, school materials, and articles in newspapers or magazines) are being provided across a wide variety of web portals on science/technology or liberal arts/social sciences.

Researchers seeking to search research findings in a given, specialized field of research, however, find these portals rather inconvenient, as they often end up retrieving unwanted information on other fields of research as well. To address this problem, research has recently been underway on an information service combined with the semantic web, but this new information service has yet to be applied to information service platforms in Korea.

Hence this paper designs the community model based on the subjects in the field of bioscience, applying the concept of the open archive and the open access using literature, facts, and patent contents already established by KISTI focusing on the bioscience field where there are many on-going researches recently.

2. CURRENT STATUS OF KISTI CONTENT

2.1 Literature Content

Literature content which is collected, processed, and built by KISTI services about 60 million of academic papers, research reports, and the theses for degree. Theses' database provides bibliography of the domestic/oversea journals & conferences and links it to the information of original electronic text.

Table 1. Status of construction of the literature content

Field	Database	Period	Number
Journal	Korean journal	1948 ~	478,922
	English journal	1991 ~	39,529,281
	China, Japan journal	2003 ~	2,621,927
Proceedings	Korean proceedings	1972 ~	213,868
	English proceedings	1993 ~	6,881,628
INSPEC	Physics,	1969 ~	9,977,921

This is an excellent paper selected from the papers presented at ICCC 2009.
* *Corresponding author. E-mail : ahnyoung@kisti.re.kr*
Manuscript received Apr. 20, 2010 ; accepted Sep. 06, 2010

	Electronics, etc.		
FSTA	Food science	1969 ~	793,567
Bibliography	Journal		63,175
	Proceedings		203,092
Thesis	Korean thesis	1945 ~	1,072,406
Report	National R&D report	1983 ~	115,911
	NTIS report	1995 ~	146,421

INSPEC (Information Services for the Physics and Engineering Communities) is a database in physics, electric & electronics, and computer science provided by IET (The Institution of Engineering and Technology, Great Britain) and FSTA (Food Science & Technology Abstracts) is the only database in food provided by IFIS (International Food Information Service, Great Britain).

KISTI research report database is from the collection of KISTI selected from the technical reports supported by the budget of USA government and research reports supported by Korean government budget. The theses database is identical to that of National Assembly Library of Korea which collects and accumulates the master and Ph. D. theses announced by the Korean universities and colleges.

2.2 Factual Content

Fact information or factual database is the generic concept of the information to the factual data in science and technology fields such as chemical structure, physical characteristics, genetic information, and biodiversity. KISTI built and provides the factual content in the fields of physics, chemicals, bioscience, astronomy, human body, virtual science museum with the support from external experts because the fact data can generated only by the experts of those fields.

Table 2. Status of construction of the factual content

Field	Database	Period	Number
Physics /Chemistry	ChemDB	2000 ~	2,579,458
	ICSD	1913 ~	93,062
	Plasma	1995 ~	43,370
Body	Digital Korean	2003 ~	-
	Visible Korean	2000 ~	-
Bio	Gene/Protein	2002 ~	250,000,000
	Biodiversity	2000 ~	1,140,000
Astronomy	Luni-Solar calendar	918 ~	-
	Ancient astronomy	2000 ~	3,633
Cyber museum	Virtual Science Museum	1999 ~	1,948,740
	Fossil Museum	1999 ~	850
	Shell Museum	1999 ~	950
	CNIC cyber museum	2002 ~	-

2.3 Patent Content

KISTI also provides a database service on the patent information focused on USA, Europe, Japan, International, and Korea where more than 80% of the global patent is owned.

Tab. 3. Status of construction of the patent content

Country	Period	Number
Korea	1947 ~	2,970,872
USA	1976 ~	3,710,861
Japan	1976 ~	8,019,137
EU	1976 ~	1,948,740
International	1976 ~	1,486,226

Korean patent database has 2,970,872 records since 1947, and those of USA (3,710,861 records), Japan (9,019,137), Europe (1,948,740), International (1,486,226) have bibliography, abstracts, patent scope, etc. since 1976[6].

3. ANALYSIS SCHEMA OF KISTI CONTENT

3.1 Literature Content Schema

Bioscientific themes are classified in KISTI's literature content, among which the most frequently searched database items for articles published in domestic academic journals are presented in Table 4.

Table 4. Schema of literature content

Field name	Code
Control Number	CN
Title	TI
Parrel Title	TIP
Author	AU
Data Type	
Source Type	
Journal Title	SO
Publisher	PB
Publish Year	PY
Language	LA
Volume Issue	VO/IS
Pages	PG
ISSN	SN
ISBN	SB
Classification Code	DC
Keyword	KW
Copyright	
Source Access Path & Right	

3.2 Factual Content Schema

KISTI's factual content includes biological information and biodiversity information. Table 5 shows the components of Genbank, a database on genetic information that is the most widely used around the world.

Table 5. Schema of Genbank content

Field name	Code
Locus	LO

Definition	DE
Accession	AC
Version	VE
Keywords	KE
Source	SO
Organism	OR
Reference	RE
Authors	AU
Title	TI
Journal	JO
Comment	CO
Features	FE
Origin	OR

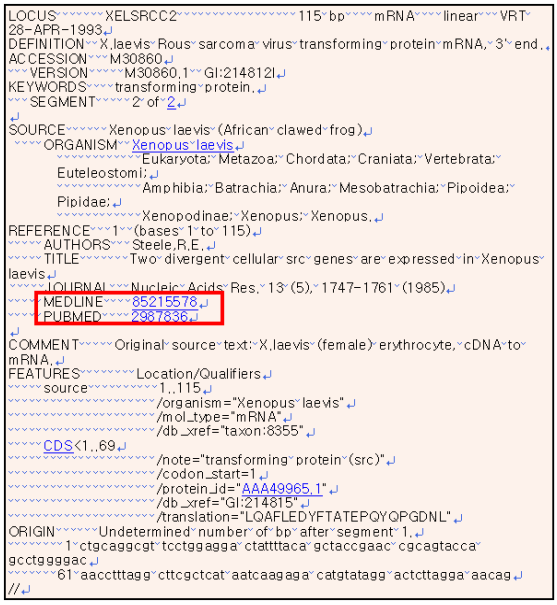
LOCUS XELSRCC2 115 bp mRNA linear VRT 28-APR-1993
DEFINITION X.laevis Rous sarcoma virus transforming protein mRNA, 3' end.
ACCESSION M30860
VERSION M30860.1 GI:214812
KEYWORDS transforming protein.
SEGMENT 2 of 2
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea;
Pipidae;
Xenopodinae; Xenopus; Xenopus.
REFERENCE 1 (bases 1 to 115)
AUTHORS Steele,R.E.
TITLE Two divergent cellular src genes are expressed in Xenopus laevis
JOURNAL Nucleic Acids Res. 13 (5), 1747-1761 (1985)
MEDLINE 85215578
PUBMED 2987836
COMMENT Original source text: X.laevis (female) erythrocyte, cDNA to mRNA.
FEATURES Location/Qualifiers
source 1..115
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
CDS <1..69
/note="transforming protein (src)"
/codon_start=1
/protein_id="AAA49965.1"
/db_xref="GI:214815"
/translation="LQAFLEDYFTATEPQYQPGDNL"
ORIGIN Undetermined number of bp after segment 1.
1 ctgcaggcgt tcctggagga ctattttaca gctaccgaac cgcagtacca
gcctggggac
61 aaccttagg cttcgctcat aatcaagaga catgtatagg actcttagga aacag
//

Fig.1. Search Result of Genbank[3]

3.3 Patent Content Schema

KISTI's patent content is not classified by bioscientific themes but by grand categories set by the International Patent Classification (IPC). Table 6 illustrates the components of the U.S. patent database among KISTI-provided patent information.

Tab. 6. Schema of patent content

Field name	Code
Control Number	CN
Title of Invention	TI
Country Code	CY
Patent Type	PT
Document Kind	REG
International Patent Classification	IC
US Patent Classification	USC
Application Number	AN
Application Date	AD
Publication Number	UN
Publication Date	UD
Registration Number	RN
Registration Date	RD
Priority Number	PRN
International Application No.	IAN
International Application Date	IAD
USC 371 (PCT) Date	
USC 102 (e) Date	
International Unexamined No.	IUN
International Unexamined Date	IUD
Drawing	
Inventor	IN
Inventor Address	INA
Inventor Country	INC
Applicant	PA
Applicant Country	
Applicant Address	PAA
Agent	AG
Abstract	AB
Claim	CM
Claim Number	BSC
Examiner	
Related Application Data	
Designated Country	DS

4. DESIGN OF COMMUNITY MODEL

4.1 Open Archiving Service

In an open archiving service, an open web—instead of a closed one—should be utilized so that anyone can access open research/academic information. This study aims to build a platform where anyone can find needed information for free and which serves as an environment for researchers themselves to evolve the community in a constructive way.

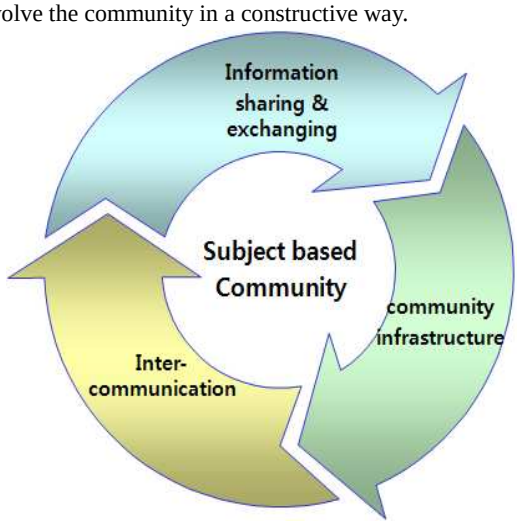

Fig.2. Concept of community model

Open archiving service should not use the closed web but the open web allowing anyone is able to utilize the open R&D and academic information. The feature that user upload information directly is essential to share data in the web environment(refer to Figure 2).

Tab. 7. Schema for metadata of shared information

Column	Description	Data type
c_num	control number	int(10)
name	title	varchar(50)
c_type	type	varchar(20)
issn_isbn	ISSN / ISBN	char(10)
summary	abstract	text
pub_date	publishing date	date
lang	language	varchar(10)
keywords	keyword	varchar(50)
g_code	classification code	char(5)
url_addr	source URL	varchar(50)
r_doi	source DOI	varchar(20)
r_gendate	registration date	date
r_moddate	Modification date	date
r_generator	registrator	char(10)

A database schema (refer to Table 7) for the metadata of shared information is designed to share the location information (URL, e-mail address, etc.) of the original files or to attach the original files by uploading any format of data (refer to Figure 3)[5].

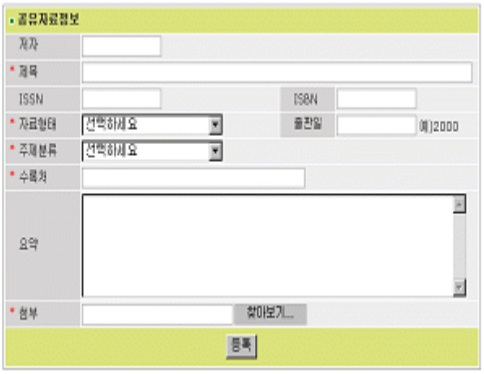

Fig.3. Screen of data uploading

4.2 Open Access Service

The design target is that researchers using this community can freely use information produced by the mutual accumulation of bioscience information which is needed to each other while the metadata of journals and articles which are open to access is collected, built, and serviced based on the free usage of information(refer to Table 8)[5].

Table 8. Schema for metadata of article

Column	Description	Data type
c_num	control number	int(10)
name	title	varchar(50)
author	first author	varchar(20)
sub_author	correspond author	varchar(50)
j_name	journal name	varchar(50)
j_spec	volume, number	varchar(30)
pub_plc	country	varchar(20)
pub_org	organization	varchar(50)
pub_date	publish date	date
lang	language	varchar(10)
summary	abstract	text
keywords	keywords	varchar(50)
g_code	classification code	char(5)
url_addr	source URL	varchar(50)
r_doi	source DOI	varchar(20)
r_gendate	registration date	date
r_moddate	Modification date	date
r_generator	registrator	char(10)

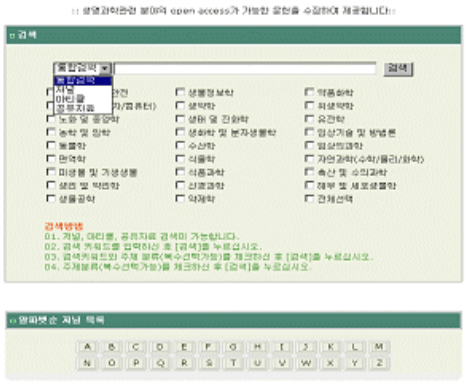

Fig.4. GUI of information retrieval

<Figure 4> is a search GUI to the content of papers (association papers, journal articles, and conference papers, etc.) in bioscience field from the legacy that was provided and built by KISTI and from that was directly uploaded by user in the bioscience. The search function is designed to enable 'search by detail subject', 'search by alphabetic order', 'search by data type', and 'integrated search'[4].

4.3 Community Management System

The community management system has features such as screening function for the data registration, modification function of database, and subscriber management function[1][8]. And it is designed to review the registered various research results, to correct if necessary, and ultimately store them in the database as the system configuration in <Figure 5>.

The system also includes the functions of membership level management (i.e. administrator, operator and ordinary members) and statistics (for monitoring how the information is being used). The administrator of the community management system has the rights to register, review, and modify data and to adjust the level of members, and thus the system was developed in a way enabling the administrator to exercise those rights.

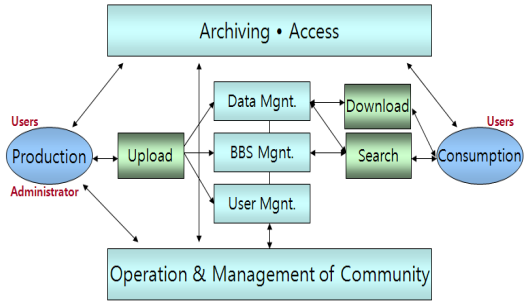

Fig.5. Configuration of community management system

5. CONCLUSION

This study has surveyed and analyzed reference, factual, and patent content that KISTI offers, and, in line with this content, developed a subject-based community model for the field of bioscience. This community model will be modified and improved, based on the feedback of bioscientific researchers and information experts, and become available for domestic researchers of bioscience via the official website of KISTI's comprehensive scientific/technology information service (http://www.ndsl.kr). Once communities for different bioscientific research themes are run in earnest, they will contribute to enhanced research efficiency for researchers by establishing a system for collection and sharing of bioscientific information among researchers, by enabling exchanges of academic information resources, and by providing a virtual laboratory of research and academic activities.

Also, this community model can be utilized further as the basis of future endeavors to interlink KISTI-owned content in other fields of research (e.g. electric/electronic engineering, mechanical engineering, and materials engineering) and build communities for the respective fields as well.

REFERENCES

[1] arXiv.org e-Print archive website, http://www.arxiv.org
[2] Bu Young Ahn and Chi Pyoung Song, "Construction of the Bibliographic Information Network Prototype for Biology & Life Science", Journal of Information Management, v.36, n.2., 2005, pp.125-151.
[3] Bu Young Ahn, Jeong Min Han, Chung Shick Oh, and Beom Jong You, "A Study on Update of Bioinformatics Content", *Proc. The 21th International CODATA Conference*, 2008, pp.61-65.
[4] DOAJ (Directory of Open Access Journal) website, http://www.doaj.org
[5] MODS (Metadata Object Description Schema) website, http://www.loc.gov/standards/mods
[6] KISTI website, http://www.ndsl.kr
[7] Factual Database website, http://fact.kisti.re.kr
[8] OAI (Open Archives Initiative) website, http://www.openarchives.org

Bu young AHN
She is senior researcher at the Supercomputing Center, Korea Institute of Science and Technology Information (KISTI), where she does research on e-Science, metadata, and factual data. She finished her Ph.D. thesis at the University of Chungnam under professors Eungbong Lee on the Open Community Framework Development based on Web 2.0.

Ji young KIM
She is senior researcher at the Knowledge Information Center, Korea Institute of Science and Technology Information (KISTI), where she does research on Information service and factual data. She finished her master thesis at the University of Chungnam under professors Kang Seong Kwon on the Reaction mechanisms on Intermetallic compounds.

Chung shick OH
He is principal researcher at the Information Strategy Team, Korea Institute of Science and Technology Information (KISTI), where his does research information security and ubiquitous. He finished his Ph.D. thesis at the University of Chungbuk under professors Younghwan Cho.

Myung sun LEE
He is principal researcher at the Supercomputing Center, Korea Institute of Science and Technology Information (KISTI), where he does research on data processing, network security, and information system security. He finished his Ph.D. thesis at the University of Hannam under professors Jeakwang Lee.