

2003

2 0 0 4

論文 碩士學位 論文 提出

2004 年 1 月

梨花女子大學校 科學技術大學院

李 煥 全

李 焘 全 碩 士 學 位 論 文 認 准

指導教授

審查委員

梨花女子大學校 科學技術大學院

I .	-----	1
1.1	-----	1
1.2	-----	2
II .	-----	4
2.1	-----	4
2.1.1	-----	4
2.1.2 Axis	-----	6
2.1.3 IBM WSTK	-----	8
2.2 BioDAS	-----	10
III .	-----	13
3.1	-----	13
3.2	-----	14
3.2.1	-----	14
3.2.2	-----	17
3.3	-----	18
3.3.1	-----	19
3.3.2	-----	20

3.3.3	-----	21
3.3.4 DASGFF	-----	23
3.3.5	-----	25
IV .	-----	29
4.1	-----	29
4.2	-----	30
4.2.1	-----	30
4.2.2	-----	39
V .	-----	42
	-----	44
	-----	46

[2.1]	-----	5
[2.2] BioDAS	-----	10
[3.1]	-----	13
[3.2] BioDAS	WSDL -----	15
[3.3]	-----	17
[3.4]	-----	19
[3.5]	-----	22
[3.6] DASGFF	-----	23
[3.7]	-----	25
[3.8]	-----	28
[4.1]	-----	30
[4.2] DASDSN	-----	31
[4.3] DASDNA	-----	33
[4.4] rice	-----	34
[4.5] P0047B08	-----	36
[4.6] OSJNBa0048101	1bp~3000bp -----	37
[4.7] P0497A05	20000bp~30000bp -----	38
[4.8]	-----	39
[4.9]	-----	40
[4.10]	-----	41

[2.1] GFF	-----	11
[3.1]	-----	12
[4.1]	-----	29
[4.2]	-----	32

論文概要

가

가

BioDAS

GFF

BioPerl

GFF

I.

1.1

(Bioinformatics)

[1,2].

가

[3].

가

FTP

가

[4].

가

(Web Services)

[5].

1.2

가

BioDAS(Distributed Sequence Annotation System)

DAS

OmniGene

. BioDAS

HTTP

가

Sanger Institute

GFF

DASGFF

XML

[4].

BioDAS

가

DASGFF

[6].

가

가

가

HTTP

가

가

가,

가

, 가

가

. 2

가

BioDAS

. 3

. 4

3

5

II.

가

가

SOAP(Simple Object Access Protocol) API

3 Axis(Apache eXtensible Interaction System) ,

IBM WSTK(Web Services ToolKit) .

가

BioDAS .

2.1

Axis IBM WSTK .

2.1.1

, RPC(Remote Procedure Calls)

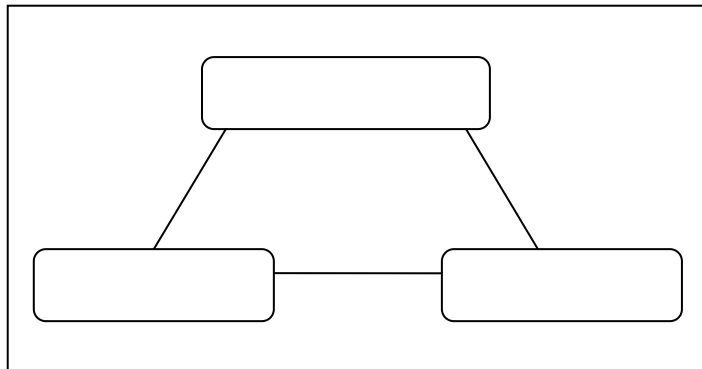
HTTP XML

가 . , URL

, 가

(role) (operation) 가

가 [7]. [2.1]



[2.1]

, ,
, ,

()가 .

□ (description) :

Web Service Description Language (WSDL)

WSDL XML

URL

- (publish) (find) :

Universal Description, Discovery, and Integration (UDDI)

UDDI

. (UDDI (description))

UDDI

(UDDI (discovery))

- :

Simple Object Access Protocol (SOAP)

SOAP XML

. XML

[8].

2.1.2 Axis

Axis(Apache eXtensible Interaction System)

SOAP

(chain) (handler)

Axis 1.1

SOAP

Axis SOAP SAX
가 Axis 가
Axis 가
SMTP, FTP, MOM
SOAP (sender) (listener)
Axis
[8].

- SOAP 1.1
-
- SOAP JWS , 가
- /
- /
- 'List' SOAP
- SOAP RPC
- WSDL
- WSDL Wsd12Java
- WSDL Java2Wsd1
- 2.2 /
- EJB EJB

HTTP

HTTP (standalone)

, Axis

. Axis

WSDD(Web

Service Deployment Descriptor)

Axis

. WSDD Axis

(chain)

가

SOAP

가

2.1.3 IBM WSTK

IBM (WSTK, Web Services ToolKit)

. WSTK

[9].

, IBM

. IBM WSTK

가

. WSTK

SOAP, UDDI, WSDL

WSTK

- UDDI save, delete, find, get
- UDDI4J API
- UDDI publish, unpublish, find
- API
- WSDL 1.1, WSFL(Web Services Flow Language), WSIL(Web Services Inspection Language), HTTPR(Reliable HTTP)
-
- WSDL Java2WsdI
- WSDL WsdI2Java
- Axis
- WSTK customization WSTK
- (Utility Web Services)
- (Utility Web Services Portal Tool)
-
-
- WAS(Web Application Server) WSAD(WebSphere Application Server)
- UDDI stack
- JavaDoc WSDL
- WSDLdoc
- WSDL WSDL4J preview
- UDDI v2 UDDI UDDI4J v2 preview

2.2 BioDAS

가

(robust)

BioDAS(Distributed Sequence

Annotation System)

BioDAS

가

HTTP

가

XML

DAS

DAS XML

DAS

XML (specification)

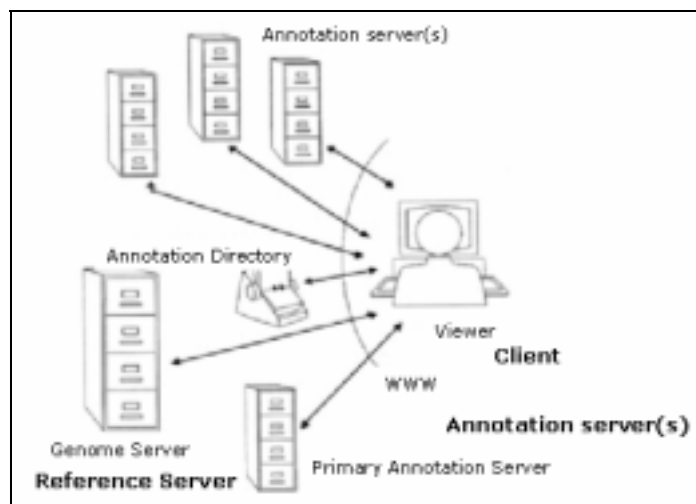
BioDAS

[

2.2] . DAS

(

)



[2.2] BioDAS

(source)

DASDSN, DASDNA, DASGFF XML . DASDSN XML
 DASDNA XML
 DNA . DASGFF GFF Sanger Institute
 General Feature Format Gene Finding Format . [2.1] GFF
 [10].

[2.1] GFF

GFF	
seqname	Sequence EMBL/GenBank/DDBJ public DB accession number .
source	Feature source , public DB annotation .
feature	Feature type Sanger feature set 가 .
start, end	start end . Sequence numbering 1 .
score	Floating point EMBL ‘.’ .
strand	‘+’, ‘-’, ‘.’ . ‘.’ strand RNA ‘.’
frame	0, 1, 2, ‘.’ 가 가 .
attribute	tag . (GFF ver.2)(A- Za-z][A-Za-z0-9]*) free text value 가 .

GFF

. , 가

DAS

TIGR (<http://www.tigr.org/>), UCSC

(<http://genome.uscs.edu/>), Wormbase (<http://www.wormbase.org/>), Ensembl (<http://www.ensembl.org/>),

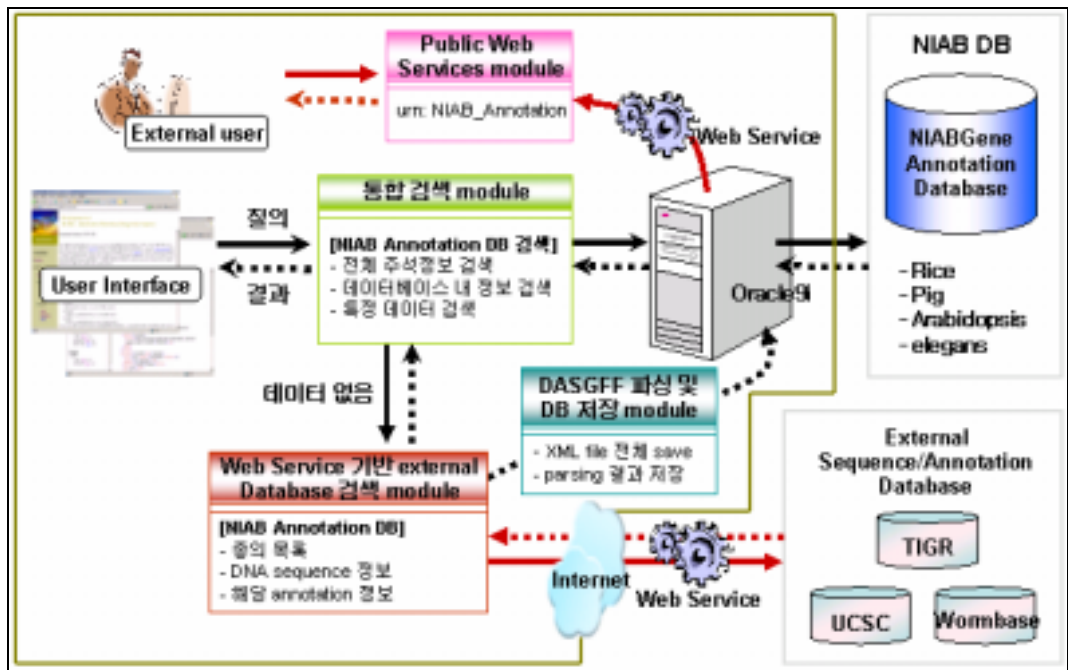
Flybase (<http://www.flybase.org>)

. DAS

DAS

3.1

[3.1]



[3.1]

(NIAB DB)

TIGR, UCSC, Wormbase

DAS

rice, pig, Arabidopsis, elegans

locus id

가

가

HTTP

3.2

3.2.1

DAS

OmniGene

[11]. OmniGene

DAS

가

MIT

J2EE

OmniGene

OmniGene

Axis, IBM WSTK

[3.2] BioDAS

WSDL(Web Service Description Language)

GenomicWebService.wsdl

[12].

```
<?xml version="1.0" encoding="UTF-8" ?>
- <definitions name="GenomicWebService" targetNamespace="http://www.biodas.org/names/wsdl"
  xmlns:tns="http://www.biodas.org/names/wsdl" xmlns="http://schemas.xmlsoap.org/wsdl/"
  xmlns:soap="http://schemas.xmlsoap.org/wsdl/soap/" xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xmlns:ls="http://www.ics.org/names/core" xmlns:das="http://www.biodas.org/names/annotation"
  <import namespace="http://www.biodas.org/names/annotation" location="../../das/annotation.xsd" />
- <types>
  - <schema targetNamespace="http://www.biodas.org/names/wsdl" xmlns: xsi="http://www.w3.org/2001/XMLSchema-instance"
    xmlns:soap-enc="http://schemas.xmlsoap.org/soap/encoding/" xmlns:wsdl="http://schemas.xmlsoap.org/wsdl/"
    xmlns="http://www.w3.org/2001/XMLSchema" >
    - <complexType name="ArrayOfLSID">
      - <complexContent>
        - <restriction base="soap-enc:Array">
          <attribute ref="soap-enc:arrayType" wsdl:arrayType="ls:lsid[]" />
        </restriction>
      </complexContent>
    </complexType>
    - <complexType name="ArrayOfType">
      - <complexContent>
        - <restriction base="soap-enc:Array">
          <attribute ref="soap-enc:arrayType" wsdl:arrayType="das:Type[]" />
        </restriction>
      </complexContent>
    </complexType>
    - <complexType name="ArrayOfDataSource">
      - <complexContent>
        - <restriction base="soap-enc:Array">
          <attribute ref="soap-enc:arrayType" wsdl:arrayType="das:DataSource[]" />
        </restriction>
      </complexContent>
    </complexType>
  </schema>
</types>
<message name="getDataSources" />
- <message name="getDataSourcesResponse">
  <part name="result" type="tns:ArrayOfDataSource" />
</message>
- <message name="getAnnotatedResources">
  <part name="DataSource" type="das:DataSource" />
</message>
```

[3.2] BioDAS

WSDL

WSDL IBM WSTK Wsd12java , BioDAS 가

[13].

□ *getDataSources()*

- 가 source .

□ *getAnnotatedResources(org.biodas.www.DataSource datasource)*

- 가 (locus) .

□ *getAnnotationTypes(org.i3c.www.Lsid_Type lsid)*

- (identified) 가 가

□ *getAnnotations(org.biodas.www.Type[] types, org.i3c.www.Lsid_Type lsid)*

- locus 가

□ *getAnnataionsByLocataion(org.biodas.www.Type[] types, org.i3c.www.Lsid_Type lsid, java.math.BigInteger start, java.math.BigInteger end)*

- .

TIGR, Wormbase, UCSC

Rice, Arabidopsis, Elegans, Pig

DASGFF XML

SAX

Rice,

Arabidopsis, Elegans, Pig

300

가

3.2.2

DASGFF XML

DASGFF XML DTD

DASGFF

DTD 가

GFF

가

GFF

가

가

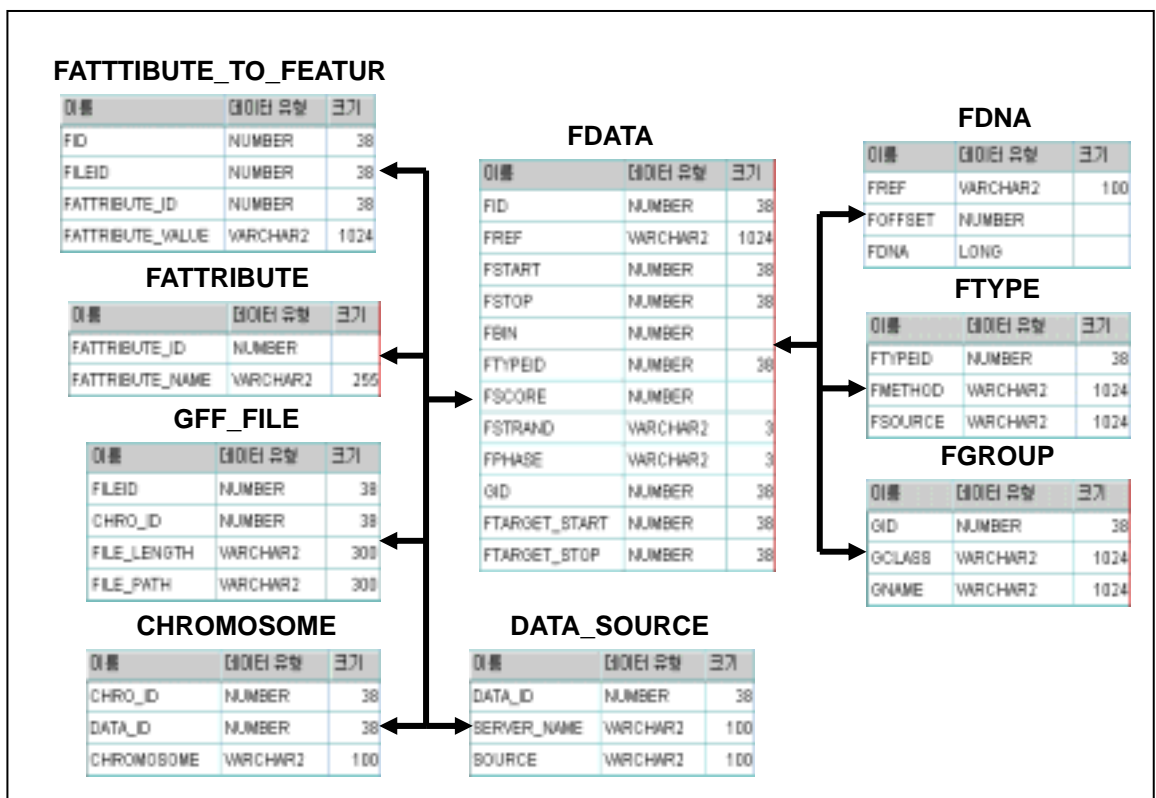
가

가 BioPerl

GFF

[14]. [

3.3]



[3.3]

BioPerl (FGROUP, FTYPE, FDNA, FDATA, FATTRIBUTE, FATTRIBUTE_TO_FEATURE) 6

가 (GFF_FILE, DATA_SOURCE, CHROMOSOME) 3 , 9

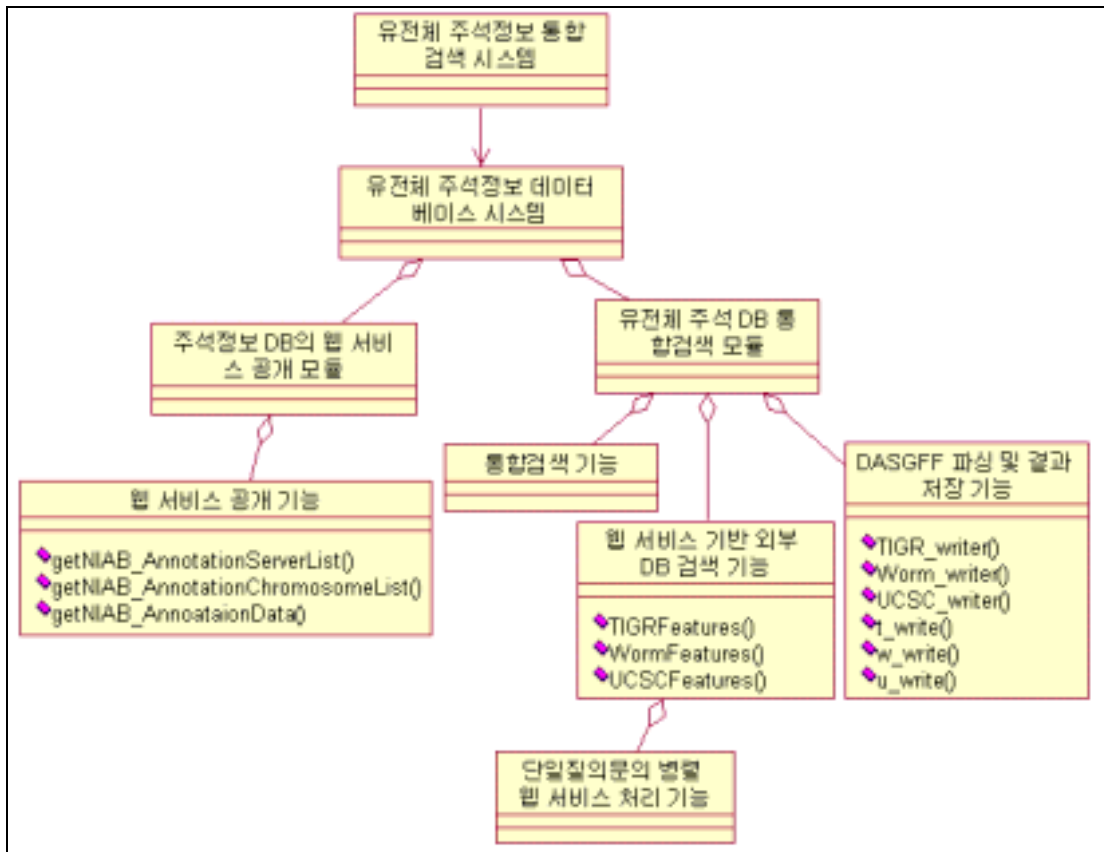
. [3.1]

[3.1]

GFF_FILE	DASGFF
DATA_SOURCE	
CHROMOSOME	locus
FGROUP	(Accession number,)
FTYPE	Feature (, source)
FDNA	DNA
FDATA	feature
FATTRIBUTE	FDATA DASGFF FEATURE TYPE 6
FATTRIBUTE_TO_FEATURE	FATTRIBUTE (value)

3.3

. [3.4]



[3.4]

3.3.1

. 가 가
, DBMS .

-
- locus
- locus

가 가 ,
[3.1] (TIGR, Wormbase,
UCSC) 가 .

UI
가 .

3.3.2

TIGR, Wormbase, UCSC

XML

XML .

□ **DASDSN XML**

- 가 URL XML

□ **DASDNA XML**

- locus DNA
XML

□ **DASGFF XML**

- locus XML

3.3.3

가

(multithreading)

(multitasking)

가 가

[15].

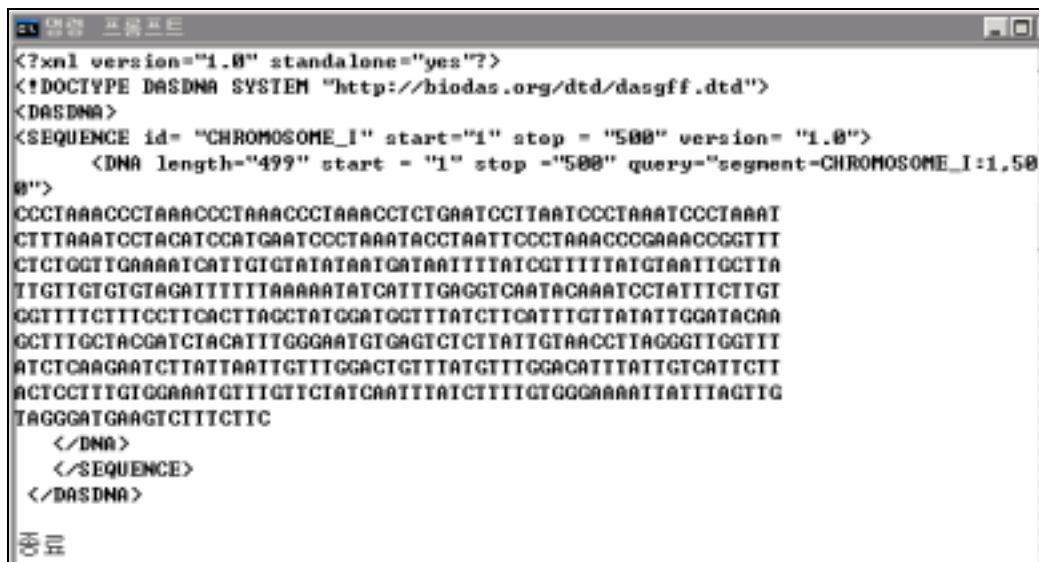
가

arabidopsis CHROMOSOME_I 1bp ~ 500bp DNA

```
MyThread first = new MyThread("arabidopsis", "CHROMOSOME_I", 1, 500);  
MyThread second = new MyThread("arabidopsis", "CHROMOSOME_I", 1, 500);
```

. [3.5]

```
while ((c = in.read()) != -1) {  
    synchronized(this) {  
        if(stopped) break;  
    }  
    System.out.print((char)c);  
};
```



```
<?xml version="1.0" standalone="yes"?>  
<!DOCTYPE DASDNA SYSTEM "http://biodes.org/dtd/dasgff.dtd">  
<DASDNA>  
<SEQUENCE id= "CHROMOSOME_I" start="1" stop = "500" version= "1.0">  
  <DNA length="499" start = "1" stop = "500" query="segment-CHROMOSOME_I:1,500">  
CCCTAAACCCCTAAACCCCTAAACCCCTAAACCCCTGAAATCCCTAAATCCCTAAATCCCTAAAT  
CTTTAAATCCTACATCCATGAATCCCTAAATACCTAATTCCTAAACCCGAAACCGGTTT  
CTCTGGTTGAAATCATTGIGTATATAATGATAATTTTATCGTTTTATGTAATTGCTTA  
TTGTTGTGIGTAGATTTTTAAAAATATCATTGAGGTCAATACAAATCCATTTCITGI  
GGTTTTCTTTCCTTCACTTACCTATCGATCGTTTATCTTCATTTGTTATATTGGATACAA  
GCTTTGCTACGATCIACATTTGGGAATGIGAGTCICTTATTGTAACCTTAGGGTTGGTTT  
ATCTCAGAAATCTTATAATTGTTGGACTGTTTATGTTTGGACATTTATTGTCATTCTT  
ACTCCTTTGIGGAATGTTTGTTCTATCAATTTATCTTTTGGGAAATTTATTAGTTG  
TAGGGATGAAGTCITTC  
  </DNA>  
</SEQUENCE>  
</DASDNA>
```

[3.5]

3.3.4 DASGFF

DASGFF XML

XML

SAX

가

```
<?xml version="1.0" standalone="yes" ?>
<!DOCTYPE DASGFF (View Source for full doctype...)>
- <DASGFF>
- <GFF version="1.0" href="http://www.tigr.org/docs/tigr-scripts/nhgi_scripts/das/dasview_rice_exonsTC/features">
- <SEGMENT id="Sequence:OSJNBa0015N08" start="1" stop="10000" version="1.0">
- <FEATURE id="TC121731" label="TC121731">
  <TYPE id="exon" category="transcription" reference="no" subparts="no">transcript</TYPE>
  <METHOD id="curated">curated</METHOD>
  <START>9623</START>
  <END>10138</END>
  <SCORE>-</SCORE>
  <ORIENTATION>+</ORIENTATION>
  <PHASE>0</PHASE>
  <NOTE>Cytochrome P450 71E1 EC 1141368 4hydroxyphenylacetaldehyde oxime monooxygenase Start:
  Stop:</NOTE>
- <GROUP id="TC121731">
  <LINK href="http://www.tigr.org/docs/tigr-scripts/tgi/tc_report.pl?
  species=rice;tc=TC121731">TC121731</LINK>
</GROUP>
</FEATURE>
- <FEATURE id="BM422175" label="BM422175">
  <TYPE id="exon" category="transcription" reference="no" subparts="no">transcript</TYPE>
  <METHOD id="curated">curated</METHOD>
  <START>2002</START>
  <END>2022</END>
  <SCORE>-</SCORE>
  <ORIENTATION>+</ORIENTATION>
  <PHASE>0</PHASE>
  <NOTE>Not Assigned Start: Stop:</NOTE>
- <GROUP id="BM422175">
  <LINK href="http://www.tigr.org/docs/tigr-scripts/tgi/est_report.pl?
  species=rice;GB=BM422175">BM422175</LINK>
</GROUP>
</FEATURE>
- <FEATURE id="BQ908149" label="BQ908149">
```

[3.6] DASGFF

[3.6] TIGR 가 Rice

'OSJNBa0015N08' 1bp ~ 10000bp(Base Pair)

DASGFF XML [3.6] FEATURE

SAX

(dasview_rice_exonsTC, Sequence:OSJNBa0015N08, 1, 10000, curated, 9623,

10138, -, +, 0, Cytochrome P450....., TC121731) FGROUP, FTYPE, FDATA

FATTRIBUTE FATTRIBUTE_TO_FEATURE

[3.7] [3.6] Rice 'OSJNBa0015N08' 1bp ~

10000bp DASGFF

FGROUP									
GID	GCLASS							GNAME	
373	Cytochrome P450 71E1 EC 1141368 4hydroxyphenylacetaldehyde oxime monooxygenase							TC121731	
FTYPE									
FTYPEID	FMETHOD	FSOURCE							
14	curated	dasview_rice_exonsTC							
FDATA									
FID	FREF	FSTART	FSTOP	FBIN	FTYPEID	FSCORE	FSTRAND	FPHASE	
619	Sequence:OSJNBa0015N08	9623	10138	100000	14		+	0	
						GID	FTARGET_START	FTARGET_STOP	
						373			

(a)

FATTRIBUTE	
FATTRIBUTE_ID	FATTRIBUTE_NAME
1	FEATURE: id
2	FEATURE: label
3	TYPE: id
4	TYPE: category
5	TYPE: reference
6	TYPE: subparts

FATTRIBUTE_TO_FEATURE				
FID	FILEID	FATTRIBUTE_ID	FATTRIBUTE_VALUE	
@19	27	1	TC121731	
@19	27	2	TC121731	
@19	27	3	exon	
@19	27	4	transcription	
@19	27	5	no	
@19	27	6	no	

(b)

[3.7]

3.3.5

3.3.1

가

getNIAB_AnnotationServerList()

-

getNIAB_AnnotationChromosomeList()

-

locus

□ *getNIAB_AnnotationXML(java.lang.String source, java.lang.String sp, java.lang.Long s_length, java.lang.Long e_length)*

- 가

□ *getNIAB_AnnotationTABLE(java.lang.String sp)*

- 가 locus

Axis

, chainable

Axis

WSDD(Web Service Deployment Descriptor)

SOAP 2.0

. WSDD

[13].

```
<deployment xmlns="http://xml.apache.org/axis/wsdd/"
  xmlns:java="http://xml.apache.org/axis/wsdd/providers/java">
  <service name="urn:NIAB_Annotation" provider="java:RPC">
    <parameter name="className" value="project.NIAB_Annotation" />
    <parameter name="allowedMethods" value="*"/>
  </service>
</deployment>
```

WSDD

WSDD

, WSDD

<deployment>

Axis

<undeployment>
 . <deployment> WSDD , 'java'
 . <service> .
 (chain) . .
 (,),
 . RPC 'java:RPC' .
 urn:NIAB_Annotation ,
 가 4 . Axis
 org.apache.axis.client.AdminClient .
 WSDD Axis
 urn:NIAB_Annotation 가 .
 URL
 '?WSDL' WSDL . , Axis
 WSDL , XML
 .
 IBM UDDI . UDDI(Universal
 Description Discovery and Integration) (publishing) (finding)
 ,
 . UDDI
 . IBM UDDI
 가
 [9]. [3.8] IBM UDDI

UDDI Business Test Registry

Universal Description, Discovery, and Integration

Service Details

The details of the selected service are shown below.

Service Information			
Key			
7216AFE0-07C9-11D8-95A0-000629DC0A53			
Owning Business		Owner Key	
Gene Annotation Database		3AC21080-07C8-11D8-95A0-000629DC0A53	
Service Name(s)			
Name			Language
urn:NIAB_Annotation			en
Service Description(s)			
Description			Language
Provide the result of the search for gene annotation			en
Access Point(s)			
Protocol	Address	Description	Actions
http	http://203.255.177.216:8000/axis/body.html	None	Details

[3.8]

4.1

[4.1]

[4.1]

	Windows XP Professional
DBMS	Oracle 9i release 2
	Tomcat 4.0.6
	JAVA, JSP, JAVA Bean
	IBM WSTK 3.3.2, Axis 1.1, IBM UDDI registry

DASGFF XML

BioPerl

1.2.3 GFF

XML

SAX

Xerces

1.4.4

BioDAS

DAS

OmniGene

1.1.2

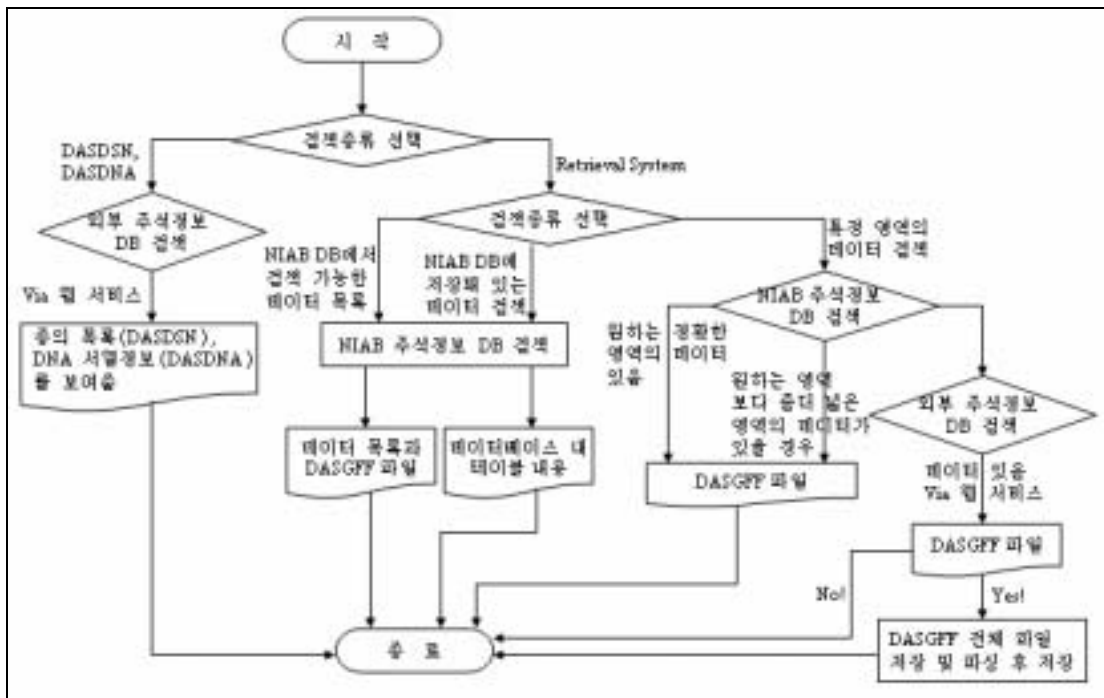
J2SE

1.4.1 SDK

4.2

4.2.1

[4.1]



[4.1]

DASDSN, DASDNA

UI

가

DASDSN XML, 가

locus

DNA

DASDNA XML

[4.2]

가 TIGR

TIGR 가

DASDSN XML

Select DAS server - DASDSN

TIGR
 WormBase
 UCSC

(a) TIGR

NIAB Gene Annotation Database
Development Based on Web Service

Waiting - (데이터 원본 부 가용을 누르십시오.)

+ Source ID:

```

<?xml version="1.0" standalone="yes" ?>
<!DOCTYPE listlist [view source for fulldoctype ...]
- <DSN>
- <DSN>
  <SOURCE id="arabidopsis">Arabidopsis Reference Server</SOURCE>
  <MAPMASTER>http://www.tigr.org/docs/tigr-
  scripts/nhgj_scripts/das/arabidopsis/</MAPMASTER>
  <DESCRIPTION>This is the Arabidopsis Reference Server database - (tested with Genefix)
  </DESCRIPTION>
</DSN>
- <DSN>
  <SOURCE id="dasview_human_TCG">Alignment of TCGs</SOURCE>
  <MAPMASTER>http://servlet.sanger.ac.uk:8080/das/ensembl729/</MAPMASTER>
  <DESCRIPTION>This is the Alignments of TCGs against the Human Genome- (Alignments can be seen
  on Ensembl.org)</DESCRIPTION>
</DSN>
- <DSN>
  <SOURCE id="dasview_human_mouseTCG">Alignment of TCGs</SOURCE>
  <MAPMASTER>http://servlet.sanger.ac.uk:8080/das/ensembl729/</MAPMASTER>
  <DESCRIPTION>This is the Alignments of TCGs against the Human Genome- (Alignments can be seen
  on Ensembl.org)</DESCRIPTION>
</DSN>
- <DSN>
  <SOURCE id="dasview_mouse_mixedTCG">Alignment of TCGs</SOURCE>
  
```

(b)

[4.2] DASDSN

[4.2] DASDSN XML <SOURCE>

URL <MAPMASTER>

<DESCRIPTION>

DASDSN

SOURCE id

[4.2]

[4.2]

TIGR	arabidopsis, dasview_humanTC, dasview_rice_exonsTC, dasview_rice_exons dasview_human_mouseTC, dasview__BGI_Indica, dasview_drosophila_TC, dasviwe_mouse_mixedTC, ...
Wormbase	elegans
UCSC	mm2, mm3, mm4, sc1, ci1, ce1, zooCat3, zooRat3, zooPig3, zooDog, zooTetra, zooZebrafish3, zooFugu3, zooBaboon3, hg13, hg15, hg16, ft1, falciparum, sacCer1, ...

[4.3] 가 Wormbase , elegans

ZK154 1bp ~ 1000bp DNA

[Select DAS server - DASDNA]

TIGR WormBase UCSC

(a) WormBase



(b)

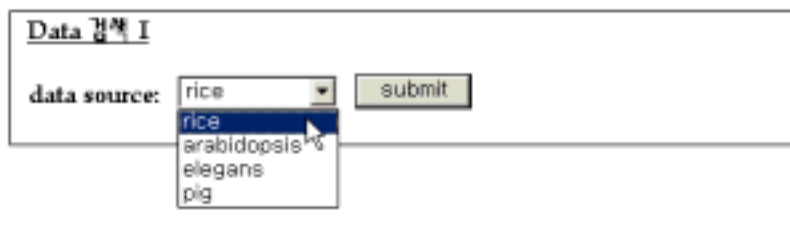
[4.3] DASNMA

(Retrieval System)

가 UI ([4.3]) ,
가 가

([4.3] Data I)
가 ,
locus id

[4.4] rice, Arabidopsis, elegans, pig rice



(a)



(b)

rice

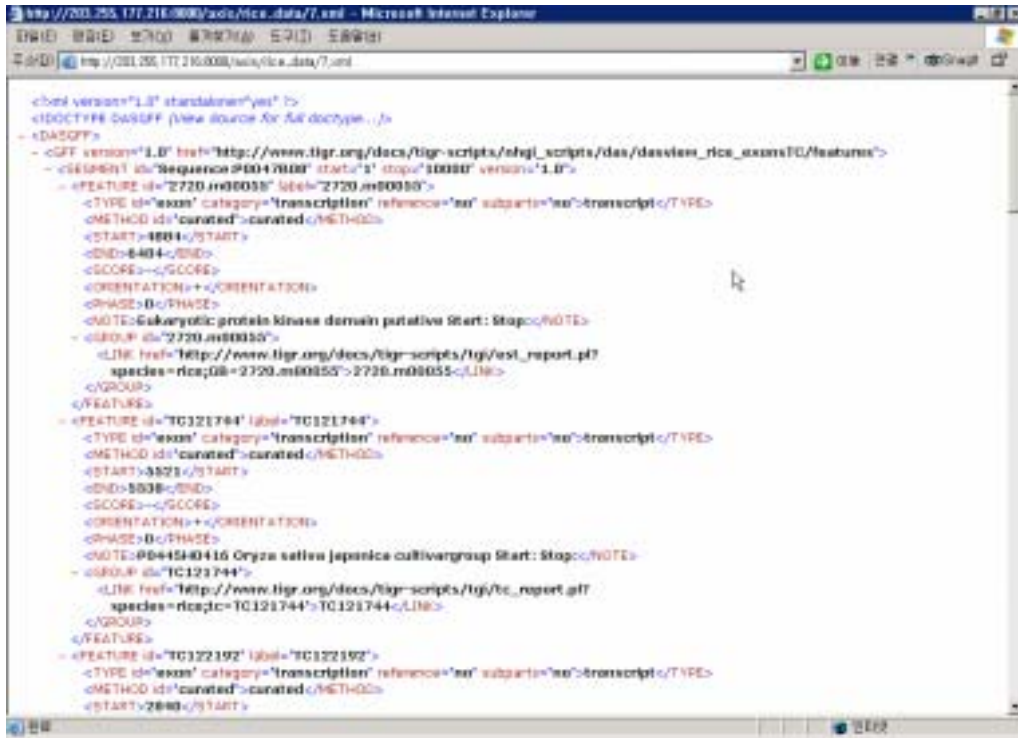
[4.4] rice

[4.4] rice

id (locus id)

DASGFF XML

XML



(b) P0047B08 DASGFF

[4.5] P0047B08

가 locus id

DASGFF XML

가 가

가 가 ,

가

[4.6] rice OSJNBa0048I01 1bp ~ 3000bp

Data 검색 II

source chromosome start end

Between 1 - 10000 Data [Click!!](#)

(a)

```
<?xml version="1.0" standalone="yes" ?>
<!DOCTYPE DASGFF (View Source for full doctype...)>
- <DASGFF>
- <GFF version="1.0" href="http://www.tigr.org/docs/tigr-
scripts/nhgi_scripts/das/dasview_rice_exonsTC/features">
- <SEGMENT id="Sequence:OSJNBa0048I01" start="1" stop="10000" version="1.0"
- <FEATURE id="2722.m00185" label="2722.m00185">
  <TYPE id="exon" category="transcription" reference="no" subparts="no">tra
  <METHOD id="curated">curated</METHOD>
  <START>35</START>
  <END>361</END>
  <SCORE>-</SCORE>
  <ORIENTATION>-</ORIENTATION>
  <PHASE>0</PHASE>
  <NOTE>conserved hypothetical protein Start: Stop:</NOTE>
- <GROUP id="2722.m00185">
  <LINK href="http://www.tigr.org/docs/tigr-scripts/tgi/est_report.pl?
species=rice;GB=2722.m00185">2722.m00185</LINK>
</GROUP>
</FEATURE>
- <FEATURE id="2722.m00185" label="2722.m00185">
  <TYPE id="exon" category="transcription" reference="no" subparts="no">tra
  <METHOD id="curated">curated</METHOD>
```

(b)

[4.6] OSJNBa0048I01 1bp ~ 3000bp

[4.7] rice P0497A05 20000bp ~ 30000bp

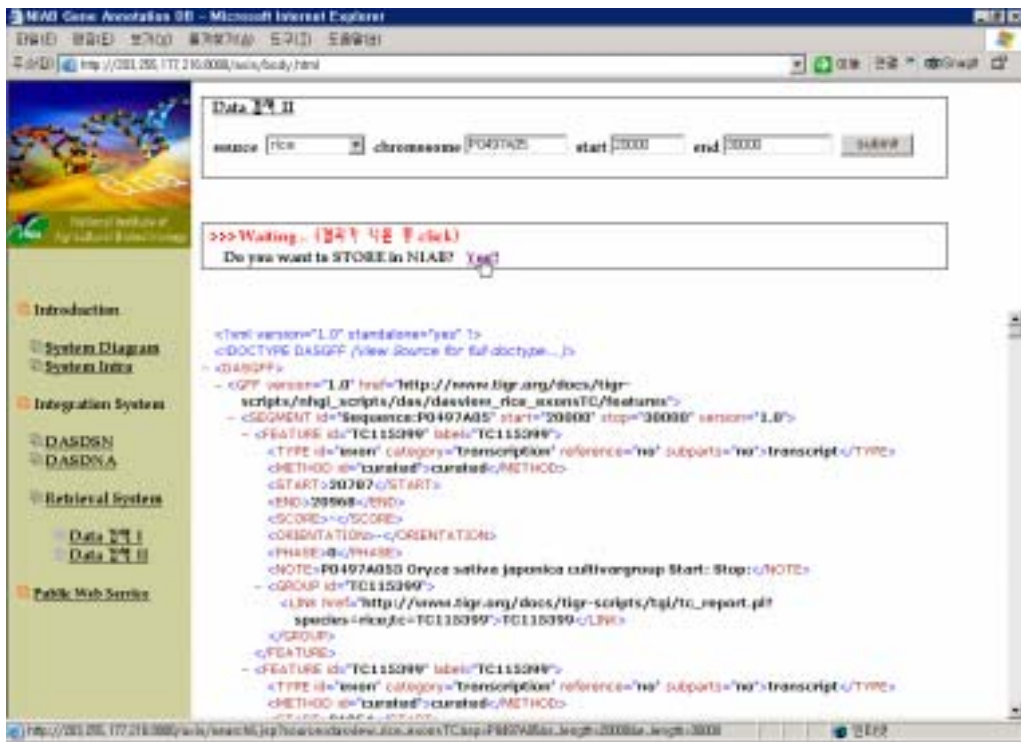
가

Data 검색 II

source chromosome start end

Not Exist Data!!
 Search External Annotation Database (Based on Web Services) [Click!!](#)

(a)



(b)

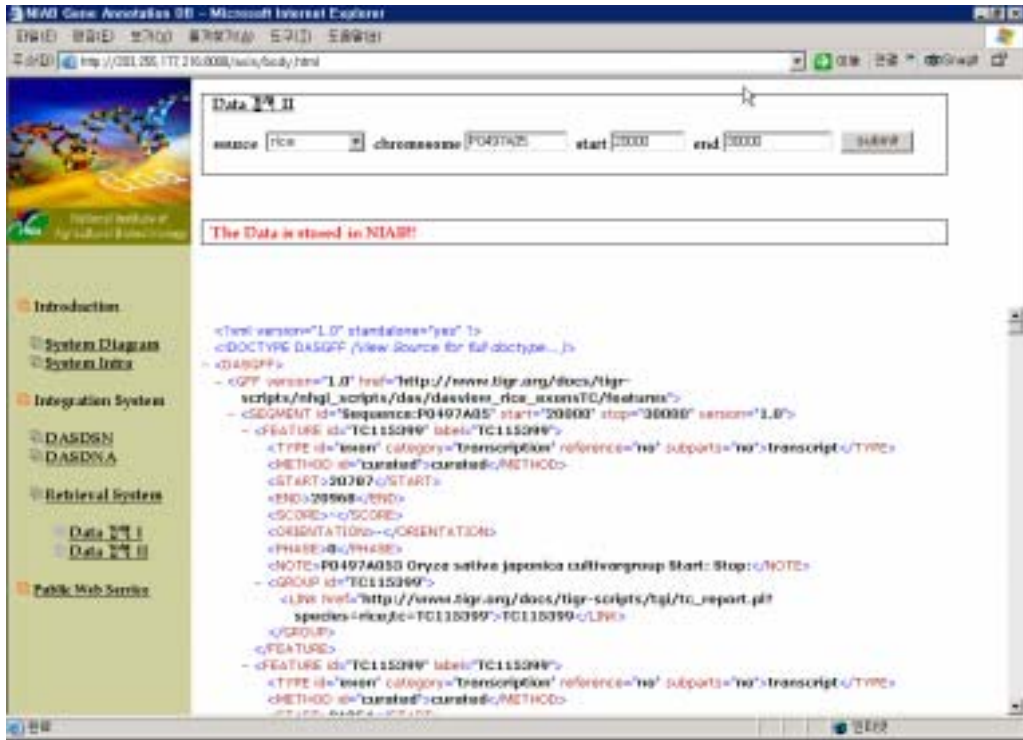
[4.7] P0497A05 20000bp ~ 30000bp

[4.7]

DASGFF

DASGFF XML

. [4.8]



[4.8]

4.2.2

가

3.3.5

WSDD

API

WSDL

[4.9]

API

WSDL Package Class

project
Class NIAB_Annotation

java.lang.String

See:

Method Summary	
java.lang.String	getNIAB_AnnotationServerList() - Return list of server name and source id from NIAB database
java.lang.String	getNIAB_AnnotationChromosomeList() - Return list of Chromosome from NIAB database
java.lang.String	getNIAB_AnnotationXML(java.lang.String source, java.lang.String sp, java.lang.Long s_length, java.lang.Long e_length) - Return XML of annotation information from NIAB database
java.lang.String	getNIAB_AnnotationTABLE(java.lang.String sp)

Method Detail

(a)

API



(b) WSDL

[4.9]

[4.9]

가

가

[4.10]

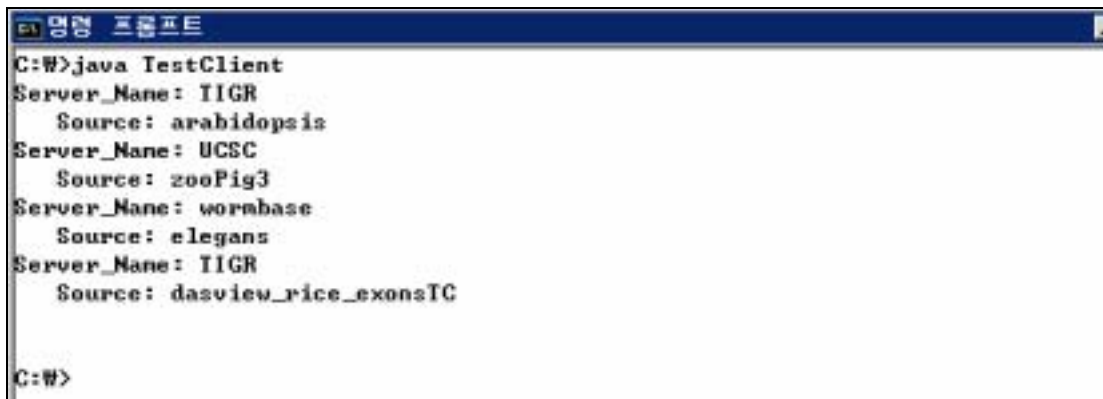
가

getNIAB_AnnotationServerList

가

```
String endpoint =  
"http://203.255.177.216:8000/axis/services/urn:NIAB_Annotation";  
String methodName = "getNIAB_AnnotationServerList";  
  
Service service = new Service();  
Call call = (Call)service.createCall();  
  
call.setTargetEndpointAddress(new java.net.URL(endpoint));  
call.setOperationName(methodName);  
  
String ret = (String) call.invoke(new Object[]{});
```

(a)



```
C:\W>java TestClient  
Server_Name: IIGR  
  Source: arabidopsis  
Server_Name: UCSC  
  Source: zooPig3  
Server_Name: wormbase  
  Source: elegans  
Server_Name: IIGR  
  Source: dasview_rice_exonsTC  
  
C:\W>
```

(b)

[4.10]

▪

BioDAS

BioPerl

GFF

가

가

- [1] Cynthia Gibas, *Bioinformatics Computer Skills*, O'REILLY, 2002.
- [2] Tisdall, James, *Beginning Perl for Bioinformatics*, O'REILLY, 2001.
- [3] Jurgen Kaljuvee, *Bioinformatics Data Integration Approaches via Soap Web Services and XML*, O'REILLY Bioinformatics Technology Conference, February, 2003.
- [4] Robin D. Dowell and Lincoln Stein, *The Distributed Annotation System*, Research article, BMC Bioinformatics, 2001.
- [5] , , , 2002.
- [6] Brian Gilman, *Using The Web Services Model to Solve Data Integration Problems in Bioinformatics*, Whitehead Institute.
- [7] Heather Kreger, *Web Services Conceptual Architecture*, IBM Software Group, White Paper, 2001.
- [8] S.Jeelani Basha, *Professional Java Web Services*, wrox, 2002.
- [9] www-3.ibm.com/software/solutions/webservices/uddi/, Web Services.

[10] <http://www.sanger.ac.uk/>, GFF: an Exchange Format for Feature Description.

[11] <http://omnigene.sourceforge.net/>, Overview of OmniGene.

[12] <http://www.biodas.org/>, Brian King, A Web Services Description of DAS, 2002.

[13] <http://ws.apache.org/axis/index.html/>, Axis User's Guide.

[14] <http://www.bioperl.org/>, A tutorial for BioPerl.

[15] Herbert Schildt, Java2: A Beginner's Guide, McGraw-Hill, 2001.

ABSTRACT

Development of Integrated Retrieval System Based on Web Service for Genome Annotation Database

Department of Computer Science & Engineering

Ewha Institute of Science and Technology

Lee Hee Jeon

With the rise of active research about various genomes and as a result of its production, a huge genome annotation data has been generated in recent bioinformatics area. Accordingly, the questions have been raised about how to share its meaningful annotation data among multiple researchers.

The study on the existing annotation data management has been performed by calculative and experimental ways but these approaches lack with which to become integration system of genome annotation information due to technical problem.

In this thesis, to efficiently solve how to share annotation data, we designed and developed the integrated retrieval system among decentralized annotation database servers using Web Service to take over object technology and BioDAS which is a great practical example on how to begin the

integration process. In order to construct database in retrieval system, we received genome annotation data via Web Service from decentralized servers. The format of received data is GFF(General Feature Format) and we used the GFF schema determined by BioPerl to load it in database. Integrated retrieval system provided users with expedience by the ability of the meta search and is equipped with easiness of system expansion by providing the ability of the stored result, as well as, reduced the processing time by multi-thread and published Web Service to external users, based on the constructed database.