웹 서비스 기반 유전체 주석정보 통합검색 시스템 구축
(Development of Integrated Retrieval System Based on Web Service for Annotation Database)

이 희 전*, 용 환 승*

(Hee-Jeon Lee, Hwan-Seung Yong)

초 록

최근 바이오인포매틱스 분야에서는 수많은 유전체에 대한 연구들이 활발하게 진행됨에 따라 그에 따른 산출물로 방대한 양의 유전체 주석정보 데이터들이 생겨나고 있다. 본 논문에서는 주석정보 데이터들의 공유문제를 효과적으로 해결하기 위하여 분산 객체 기술의 계보를 잇는 웹 서비스 기술과 오늘날 바이오인포매틱스 분야에서 데이터 통합에 관한 실질적인 방법으로 대두되고 있는 BioDAS 개념을 응용하여 분산된 주석정보 데이터베이스 서버들간의 통합검색 시스템을 설계, 구현하였다. 통합검색 시스템은 메타검색 기능을 통해 사용자에게 편의를 제공해 주며 결과 저장기능 제공을 통해 시스템 확장의 용이성을 갖추고 있다.

Abstract

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. 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. 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 function of result storing.

1. 서 론

최근 빠르게 발전하고 있는 학문인 바이오인포매틱스(bioinformatics)는 생물학 데

* 이화여자대학교 컴퓨터학과 데이터베이스 연구실

* 본 논문은 2003년도 정통부 IMT-2000 출연금 기술개발지원사업의 결과임.

이터의 관리와 분석에 컴퓨터학 분야의 첨단 기술을 이용하여 이를 자동화, 전산화하는 응용분야이다. 바이오인포매틱스 분야의 발전과 더불어 유전체 데이터에 대한 다양한 결과와 예측 보고를 다루는 방대한 양의 주석정보 데이터들이 쏟아져 나오고 있다[1,2]. 그러나 대부분의 유전체 주석정보 데이터들은 주석정보 데이터 관련 연구자에 의해 제한된 자원들을 가지고서 중앙집중적으로 관리되고 있을 뿐이다. 즉, 다수의 주석정보 데이터 관련 연구자들 사이에 데이터들을 공유하는 방법이 명쾌하게 존재하지 않고 있다[3].

주석정보 데이터의 관리에 있어 이러한 문제를 해결하기 위하여 관련 연구자들은 계산적인 방법과 실험적인 방법 등을 통해 유전체 서열 데이터의 주석정보들을 통합해 나가고 있다. 그러나 이들은 유전체 주석 데이터들을 위해 편리한 원스톱 소스를 제공하는 것이 아니라, 어떤 특정 영역에 대한 정보를 얻기 위해 다양한 웹사이트들을 체크하여 주거나 여러 다른 형태의 데이터들을 FTP를 통해 다운로드받거나 혹은 전체 그림을 얻기 위해 메뉴얼하게 통합을 수행하는 수준이다. 이러한 접근방법들은 중앙 데이터베이스의 사용자 제한과 같은 정책상의 문제나 실시간 통합이 용이하지 않은 등의 기술적인 문제로 인한 한계점을 가지고 있기 때문에 유전체 주석정보의 통합시스템으로 부족한 면을 지니고 있다[4].

그러므로 주석정보 데이터들의 공유문제를 효과적으로 해결하기 위해서는 분산된 데이터베이스들간의 통합방안에 대한 적절한 논의가 필요하다. 이를 위해 본 논문에서는 분산객체 기술의 계보를 잇는 웹 서비스(Web Services) 기술을 이용한 통합방안에 대해 살펴보겠다[5]. 인터넷과 같은 네트워크를 통하여 기술하고 배포하며 실행시킬 수 있는 모듈화된 어플리케이션인 웹 서비스를 기반으로 분산된 데이터베이스 중, 특히 유전체 주석정보 데이터와 관련된 기존 데이터베이스들로부터 통합검색 시스템을 설계하고 구축하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 분산 시스템간의 통합을 가능하게 하는 신기술인 웹 서비스 기술에 대해 소개하고 웹 서비스 기술과 관련하여 최근 연구 동향에 관해 살펴본다. 또한 바이오인포매틱스 분야에서 데이터 통합에 관한 실질적인 방법으로 대두되고 있는 BioDAS에 대해서도 살펴본다. 3장에서는 분산된 주석정보 데이터들을 통합하기 위하여 웹 서비스 기술을 사용하여 설계한 통합검색 시스템에 대하여 전체 시스템을 소개하고 주석정보 데이터베이스의 구성에 대해 살펴보며 설계한 통합검색 시스템의 기능에 관하여 자세히 기술한다. 4장에서는 3장의 설계를 바탕으로 실제 구현된 웹 서비스 기반 유전체 주석정보 통합검색 시스템의 결과에 대해 기술한다. 5장에서는 결론 및 향후 연구과제를 제시한다.

2. 관련 기술 및 연구 동향

본 장에서는 통합 기법으로 웹 서비스 기술에 대해 살펴보기로 한다. 특히 웹 서비스의 최근 동향으로 아파치 SOAP(Simple Object Access Protocol) API에서 제 3세대에 해당하는 Axis(Apache eXensible Interaction System)에 대해 살펴보기로 한다. 또한 바이오인포매틱스 분야에서 데이터 통합에 대한 가장 실질적인 방법으로 최근 대두되고 있는 클라이언트-서버 시스템인 BioDAS에 대해 살펴본다.

2.1 웹 서비스 기술

웹 서비스란, 인터넷이나 네트워크로 다른 객체에 RPC(Remote Procedure Calls)를 수행하는 기술로써 플랫폼 중립적 표준인 HTTP나 XML을 사용함으로써 사용자에게

전체 시스템 구현을 숨길 수 있다는 장점을 가지고 있다. 즉, 사용자는 서비스의 URL과 메소드 호출에 사용될 데이터형은 알아야 하나, 서비스가 어떤 플랫폼에서 구현되었는지는 알 수 없다. 이는 웹 서비스 모델이 어느 시스템에서나 존재하는 요소인 역할(role)과 오퍼레이션(operation)이라는 두 가지 요소에 초점을 맞추고 있기 때문에 가능하다[6].

2.2 Axis 소개

Axis는 기존의 SOAP 구현을 재설계한 것으로서 메시지 체인과 핸들러 객체를 기반으로 하고 있다. Axis는 메시지를 처리하는 엔진이라고 할 수 있다. Axis를 이용하여 웹 서비스를 처리하는 과정은 다음과 같다. 웹 서비스는 WSDD(Web Service Deployment Descriptor)라는 설치 디스크립터를 이용하여 Axis 메시지 처리 노드에 설치된다. WSDD는 Axis 노드로 설치된 여러 컴포넌트들이 어떻게 서로 맞물려서 입출력 메시지를 처리해야 하는지에 대해 기술되어 있다. 이러한 체인 정의는 레지스트리를 통해 해석되어 실행시간에 사용 가능하게 된다. 실행 시에 SOAP 요청은 일련의 핸들러를 거치게 되고, 이런 과정에서 메시지는 헤더 부분이 추가 및 삭제되거나 바디 부분이 변경될 수도 있다.

2.3 BioDAS

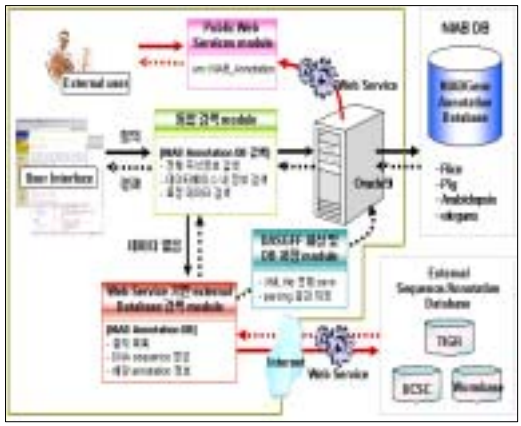
주석정보 시스템이 이상적으로 추구하는 것은 주석데이터 관련 전문가들에게 축적된 주석정보들을 빠르고 견고하며 큰 어려움 없이 새로운 정보를 덧붙일 수 있게 제공해 주어야 한다는 점이다. 이러한 요구사항들을 만족시키기 위하여 등장한 것이 BioDAS(Distributed Sequence Annotation System) 오픈 프로젝트이다. 즉 BioDAS는 유전체 서열 데이터 주석정보의 통합을 위한 방안으로, 하나의 클라이언트가 다양한 서버들로부터 정보들을 통합하는 클라이언트-서버 시스템이다. 이것은 HTTP를 통해 클라이언트가 요청한 데이터에 대한 주석처리된 정보를 XML 문서로 보여준다. 이것은 다시 말해 DAS 안에서 클라이언트와 서버 사이의 통신은 DAS XML을 통해 이루어지므로 어떠한 클라이언트나 서버라도 DAS XML 명세에 맞추면 시스템에 참여할 수 있다.

3. 유전체 주석정보 통합검색 시스템 설계

본 장에서는 주석정보 데이터에 대한 분산된 데이터베이스 시스템들을 통합하기 위하여 본 논문에서 구현한 웹 서비스 기반 통합검색 시스템의 소개와 데이터베이스 구성, 기능에 대해 살펴본다.

3.1 시스템 구성

다음 [그림 1]은 유전체 주석정보 통합검색 시스템의 전체 구성도이다.

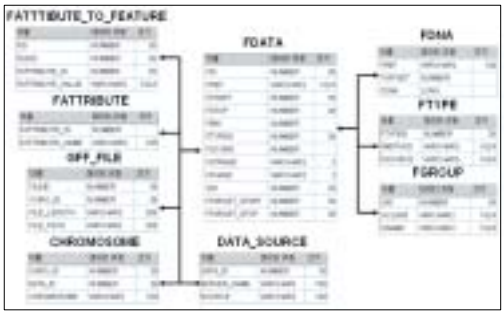


[그림 1] 통합검색 시스템의 전체 구성도

로컬 데이터베이스(NIAB)에는 기본적으로 TIGR, UCSC, Wormbase 등과 같은 DAS 서버들로부터 웹 서비스를 통해 서비스 받은 rice, pig, arabidopsis, elegans 등의 유전체에 관한 주석 데이터들이 로드되어 있다. 사용자는 웹으로 구현된 사용자 인터페이스를 통하여 원하는 주석정보를 검색 요청할 수 있다. 통합검색 시스템을 통하여 사용자는 로컬 데이터베이스 내에 저장되어 있는 전체 데이터의 목록과 특정 데이터에 대한 데이터베이스 내 정보, 특정 유전체의 염색체 혹은 locus id에 대한 특정 영역의 주석정보를 검색할 수 있다. 시스템은 사용자의 질의요청에 대해 로컬 데이터베이스 내에 데이터가 존재할 경우는 결과를 리턴해 주며, 만약 요청한 주석정보가 로컬 데이터베이스 내에 존재하지 않을 경우에는 외부 주석정보 데이터베이스들로부터 HTTP를 통해 웹 서비스 기반으로 주석정보를 서비스 받는다. 서비스 받은 데이터는 사용자의 요청 혹은 로컬 데이터베이스 관리자의 필요에 의해 데이터 파싱과 같은 전처리 과정을 거쳐 데이터베이스에 로드할 수 있다. 또한 시스템은 구현한 로컬 데이터베이스의 데이터들을 외부 사용자에게 웹 서비스를 통해 제공한다.

3.2 주석정보 데이터베이스의 구성

DAS에 참여하고 있는 서버들로부터 웹 서비스 기반으로 유전체 데이터를 서비스 받기 위하여 OmniGene 프레임웍을 사용한다[7]. OmniGene은 DAS 클라이언트가 되기 위한 모듈의 하나로 MIT에서 개발 중에 있는 오픈 프로젝트이다. 또한 웹 서비스와 J2EE 기술을 통해 생물학 관련 데이터들을 교환하는 표준화 작업 중의 하나이다. 본 논문에서는 OmniGene에서 제공하는 라이브러리와 Axis, IBM WSTK를 사용하여 분산된 데이터베이스들로부터 주석정보를 서비스 받는다. BioDAS가 외부 사용자에게 제공하는 자바 함수를 이용하여 주석정보 데이터를 제공하는 각각의 서버들로부터 유전체에 관한 주석정보 데이터인 DASGFF XML을 서비스 받아 SAX 파서로 데이터를 파싱하는 전처리 과정을 거쳐 데이터베이스에 로드한다. DASGFF XML 데이터를 로컬 데이터베이스에 로드하기 위하여 DASGFF XML DTD를 분석한다. 분석 결과 각각의 서버마다 서비스하는 DASGFF 데이터의 DTD가 조금씩 다르다는 것을 알 수 있다. 그러므로 주석정보 데이터베이스를 구축하기 위해서는 GFF 형식을 위한 표준 데이터베이스 스키마가 필요한 것으로 판단된다. 이에 따라 본 논문에서는 주석정보 데이터베이스 구축을 위하여 데이터베이스 스키마는 GFF 데이터 전체가 표현가능하며 데이터 변화에 대하여 유연하게 대처가 가능한 BioPerl에서 정한 GFF 스키마를 기반으로 한다[8]. [그림 2]는 데이터베이스 스키마이다.



[그림 2] 데이터베이스 스키마

스키마는 BioPerl에서 제안한 기본 테이블(FGROUP, FTYPE, FDNA, FATTRIBUTE, FDATA, FATTRIBUTE_TO_FEATURE) 6개와 본 논문을 통해 통합검색 시스템 구축을 위하여 추가한 테이블(DATA_SOURCE, GFF_FILE, CHROMOSOME) 3개, 총 9개의 테이블로 구성된다. 다음 [표 1]은 각 테이

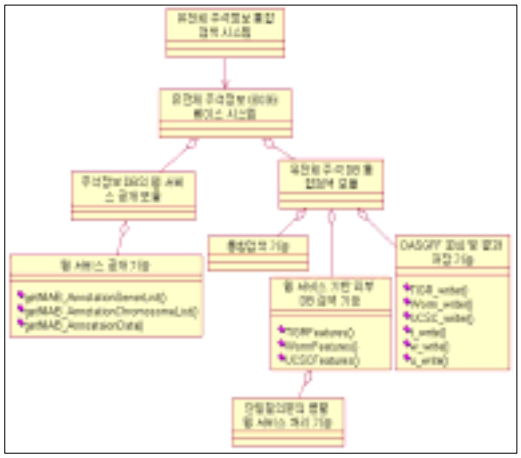
블에 대한 상세 설명이다.

[표 1] 테이블의 상세 설명

테이블명	설 명
GFF_FILE	로컬 데이터베이스에 저장되어 있는 DASGFF의 기본적인 데이터 정보
DATA_SOURCE	데이터베이스에 로드되어 있는 데이터들이 어느 서버로부터 받은 것인지에 관한 정보
CHROMOSOME	로컬 데이터베스에 로드되어 있는 염색체 혹은 locus 정보
FGROUP	주석정보 데이터의 그룹 정보(Accession number, 기타 정보)
FTYPE	Feature 종류에 관한 일반적인 정보(메소드, source)
FDNA	해당 영역에 대한 간단한 정보와 그에 대한 DNA 서열 정보
FDATA	전체 feature 정보
FATTRIBUTE	FDATA 테이블에서 다루지 못한 DASGFF 파일의 FEATURE와 TYPE 애트리뷰트의 6개 항목명
FATTRIBUTE_TO_FEATURE	FATTRIBUTE 테이블의 항목명에 해당하는 값(value)

3.3 유전체 주석정보 통합검색 시스템의 기능

본 절에서는 본 논문을 통해 구현한 웹 서비스 기반 유전체 주석정보 통합검색 시스템의 각 기능에 대하여 살펴본다. 다음 [그림 3]은 시스템의 전체 클래스 다이어그램이다.



[그림 3] 전체 클래스 다이어그램

① 통합검색 기능

사용자는 필요한 주석정보 데이터를 본 논문에서 구축한 주석정보 데이터베이스 시스템으로부터 검색할 수 있다. 사용자는 검색 요청한 주석정보 데이터가 로컬 데이터베이스에 존재할 경우, 시스템은 DBMS를 거쳐 결과 데이터를 리턴해 준다. 그러나 사용자가 요청한 데이터가 로컬 데이터베이스에 존재하지 않을 경우, 시스템은 웹 서비스를 통하여 [그림 1]의 외부 주서데이터 서버들로부터 주석정보 데이터를 서비스 받아 사용자가 요청한 데이터를 리턴해 준다. 이를 통해 사용자는 필요한 주석정보를 찾기 위하여 여러 사이트에 접속할 필요 없이 본 논문에서 구현한 통합검색 UI를 통해 원하는 정보를 찾을 수 있는 메타 검색 기능이 가능하다. 본 논문에서 구현한 주석정보 데이터베이스로부터 다음과 같은 내용을 검색할 수 있다.

- □ 로컬 데이터베이스 내에 저장되어 있는 전체 주석정보 데이터의 목록 검색
- □ 특정 염색체 혹은 locus 데이터에 대한 데이터베이스 내 정보 검색
- □ 특정 염색체 혹은 locus 데이터에 대한 특정 영역의 주석정보 검색

② 웹 서비스 기반 외부 주석정보 데이터베이스 검색 기능

웹 서비스 기반의 외부 주석데이터 서버를 통한 검색은 TIGR, Wormbase, UCSC 등 웹 서비스를 통해 데이터를 제공하는 서버들로부터 생물학 관련 정보 데이터들의 XML 형식 데이터들을 받을 수 있다. 본 논문에서 구현한 통합검색 시스템을 통해 서비스 받을 수 있는 생물학 관련 XML 데이터의 종류는 다음과 같다.

- □ DASDSN XML
  - 서버가 제공하는 유전체 목록과 URL 정보를 보여주는 XML 문서
- □ DASDNA XML

- 유전체의 염색체 혹은 locus에 대한 특정 영역의 DNA 서열 정보를 보여주는 XML 문서

□ DASGFF XML
- 유전체의 염색체 혹은 locus에 대한 특정 영역의 주석정보를 보여주는 XML 문서

③ 단일 질의문의 병렬 웹 서비스 처리 기능

자바의 멀티스레드 기능을 이용, 차후 웹 서비스를 제공하는 주석데이터 서버들이 늘어날 경우, 사용자가 요청한 질의에 대해 같은 유전체 정보를 제공하는 서버들 사이에 먼저 처리된 결과를 리턴 함으로써 처리시간 단축을 꾀할 수 있다. 이를 위해 멀티스레드를 써서 먼저 수행된 스레드의 수행이 끝나면 나머지 스레드를 강제 종료시킨다.

④ DASGFF 파싱 및 데이터베이스 저장 기능

사용자의 요청 혹은 로컬 데이터베이스 관리자의 필요에 의해 웹 서비스를 통해 외부 주석 데이터 서버로부터 서비스 받은 DASGFF XML 파일을 데이터베이스에 로드하는 기능이다. 이를 통해 서비스 받은 XML 파일 전체를 저장할 수 있으며 SAX 파서로 데이터 파싱의 전처리 과정을 거쳐 각 항목을 데이터베이스 내 테이블에 저장 가능하다.

⑤ 유전체 주석 데이터베이스의 웹 서비스 공개 기능

본 논문에서는 구축한 통합검색 시스템이 웹 서비스 모델에서 서비스 제공자의 역할을 하기 위하여 ①에서 소개한 검색 기능을 웹 서비스로 이용 가능하도록 한다. 다음은 주석정보 통합검색 시스템에서 서비스 하고자 하는 자바 함수명과 그에 대한 설명이다.

□ *getNIAB_AnnotationServerList()*
- 주석 데이터베이스 내에 존재하는 유전체 목록을 리턴

□ *getNIAB_AnnotationChromosomeList()*
- 주석 데이터베이스 내에 존재하는 염색체 혹은 locus 목록을 리턴

□ getNIAB_AnnotationXML(java.lang.String source, java.lang.String sp, java.lang.Long s_length, java.lang.Long e_length)
- 사용자가 요청한 유전체의 염색체 중 특정영역 해당 주석정보를 리턴

□ getNIAB_AnnotationTABLE(java.lang.String sp)
- 사용자가 요청한 유전체의 염색체 혹은 locus에 대한 데이터베이스 내 테이블 내용을 리턴

4. 유전체 주석정보 통합검색 시스템 구현

본 장에서는 분산된 주석정보 데이터 서버들을 통합하기 위한 방안을 3장에서 설계한 웹 서비스 기반 주석정보 통합검색 시스템에 대한 구현환경과 구현결과를 살펴본다.

4.1 구현환경

본 논문에서 구현한 주석정보 통합검색 시스템의 구현환경은 [표 2]와 같다.

[표 2]

운영체제	Windows XP Professional
DBMS	Oracle 9i release2
서블릿 엔진	Tomcat 4.06
개발언어	JAVA, JSP, JAVA Bean
웹 서비스 개발도구	IBM WSTK 3.3.2, Axis 1.1, IBM UDDI registry

유전체 주석정보 데이터인 DASGFF XML을 위한 표준 스키마는 BioPerl 버전

1.2.3의 GFF 스키마를 사용하였으며 XML 파서로는 SAX 파서인 xerces 버전 1.4.4를 사용하였다. 또한 BioDAS에서 DAS 서버들로부터 주석정보를 웹 서비스 기반으로 서비스 받기 위하여 OmniGene 프레임웍 버전 1.1.2를 사용하였으며 자바 툴킷은 J2SE 1.4.1 SDK를 사용하였다.

4.2 구현결과

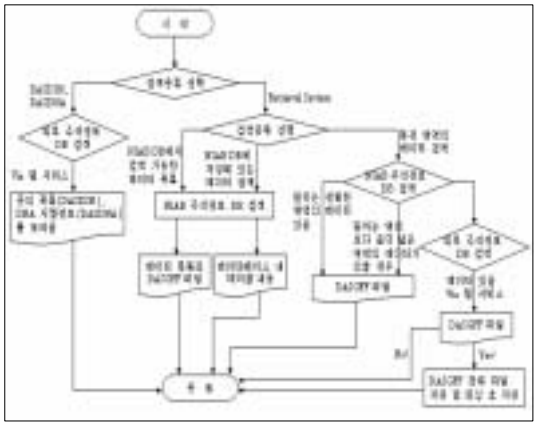
본 절에서는 통합검색 시스템에서 제공하는 통합검색과 관련한 기능들과 웹 서비스 제공 기능에 대한 구현결과를 살펴본다.

4.2.1 통합검색 기능 구현

[그림 4]는 통합검색 시스템에서 검색과 관련한 전체 프로세스 흐름도이다.

① DASDSN, DASDNA 검색

사용자는 웹으로 구현된 UI를 통해 검색 종류를 선택할 수 있다. 서버가 제공하는 유전체의 종류를 보여주는 문서인 DASDSN XML, 사용자가 요청한 유전체의 염색체 혹은 locus의 특정 영역에 대한



[그림 4] 시스템 프로세스 흐름도

DNA 서열정보를 보여주는 문서인 DASDNA XML에 대한 검색은 웹 서비스를 통해 외부 주석정보 데이터베이스를 검색하여 사용자에게 결과를 리턴해 준다. [그림 5]는 사용자가 TIGR 서버를 선택하였을 때 TIGR 서버가 제공하는 유전체 종류들을 보여주는 DASDSN XML을 웹 서비스를 통해 서비스 받아 사용자에게 결과를 보여주는 화면이다.

[그림 5] DASDSN 검색

② 통합검색(Retrieval System)

사용자가 UI를 통해 통합검색 시스템을 선택하였을 경우, 사용자는 본 논문에서 구현한 시스템에서 제공하는 두 가지 종류의 검색 중 한가지를 고를 수 있다.

첫 번째 검색은 로컬 주석 데이터베이스 내에 저장되어 있는 유전체 중 하나를 사용자가 선택하면 선택한 유전체에 대해 로컬 데이터베이스 내에 저장되어 있는 전체 염색체 목록 혹은 locus id 목록을 볼 수 있다. [그림 6]은 rice, arabidopsis, elegans, pig 유전체 중 rice 유전체를 선택한 결과화면이다.

(a) 유전체 선택 화면

(b) 주석 데이터베이스 내에 저장되어 있는 rice 염색체 목록

[그림 6] rice 유전체 선택 화면

[그림 6]의 rice 유전체 선택 결과화면에서 염색체 id(혹은 locus id)를 클릭할 경우 사용자는 선택한 염색체에 대해 로컬 주석 데이터베이스 내에 저장되어 있는 테이블의 전체 내용을 볼 수 있다. 또한 특정 영역별로 DASGFF XML 파일이 링크되어 있으므로 사용자는 전체 XML 데이터를 확인할 수 있다.

두 번째 검색은 사용자가 주석정보를 얻기 원하는 유전체의 염색체 혹은 locus id에 대하여 특정 영역의 데이터를 검색할 수 있다. 이때의 검색결과는 DASGFF XML이다. 시스템은 사용자가 요청한 주석정보가 로컬 데이터베이스 내에 존재하면 결과를 리턴해준다. 만약 로컬 데이터베이스가 정확한 영역의 값을 가지고 있지 않을 경우, 시스템은 검색을 요청한 영역을 포함하는 더 큰 영역의 데이터를 찾는다. 만약 더 큰 영역의 데이터도 존재하지 않는다면 통합검색 시스템은 웹 서비스를 통해 사용자가 요청한 영역의 데이터를 외부 데이터베이스 서버로부터 웹 서비스를 통해 서비스 받아 사용자에게 보여준다. [그림 7]은 rice 유전체의 P0497A05 염색체 중 20000bp~30000bp 영역의 검색 요청 시 사용자에게 로컬 주석 데이터베이스 내에 원하는 정보가 없음을 알리고 외부 주석 데이터베이스 서버를 검색해 웹 서비스를 통하여 데이터를 서비스 받는 화면이다.

(a) 검색결과 일치하는 값이나 더 큰 영역의 값도 없음을 알림

(b) 웹 서비스를 통해 외부 주석데이터베이스 서버로부터 값을 받음

[그림 7] P00497A05의 20000bp~30000bp 영역의 검색 결과 화면

[그림 7]과 같이 외부 주석정보 데이터베이스로부터 서비스 받은 DASGFF 파일은 사용자의 요청 혹은 데이터베이스 관리자의 필요에 의해 로컬 데이터베이스 내에 저장할 수 있다. 저장을 요청하면 시스템은 DASGFF XML 파일 전체를 저장하고 파싱 과정을 거쳐 각 항목을 데이터베이스 내 테이블에 로드한다.

4.2.2 웹 서비스 공개 기능 구현

본 논문을 통해 구현한 유전체 주석정보 통합검색 시스템이 웹 서비스 모델의 역할 중 웹 서비스 제공자가 되기 위하여 WSDD를 이용하여 웹 서비스를 배포한다. 시스템은 웹 서비스를 이용하기 원하는 외

부 사용자들을 위하여 시스템에서 웹 서비스를 통해 제공하는 자바 함수에 대한 API 형태의 문서와 WSDL 문서를 제공한다.

5. 결론 및 향후 과제

본 논문에서는 유전체 주석정보에 대한 분산된 데이터베이스 서버들간의 데이터 통합문제를 해결하기 위하여 웹 서비스 기술과 BioDAS 개념에 대하여 살펴보았다. 살펴본 기술들을 이용하여 BioPerl이 정한 GFF 스키마를 기반으로 유전체 주석정보 통합검색 시스템을 설계, 구현하였다. 본 시스템은 유전체 주석정보에 대한 메타검색 기능을 제공하므로 사용자는 여러 사이트에 접속하여 자신이 원하는 정보를 개별적으로 찾는 번거로운 검색작업을 하지 않아도 된다. 뿐만 아니라 저장 기능을 통한 시스템 확장의 용이성도 갖추고 있다. 또한 차후 웹 서비스를 제공하는 유전체 주석정보 관련 데이터베이스 서버들이 늘어날 경우, 본 논문을 통해 구현한 통합검색 시스템의 수행 처리 성능을 향상시키기 위하여 본 시스템에 병렬 웹 서비스 처리 기능을 덧붙였다. 이러한 유전체 주석정보 통합검색 시스템의 설계와 구현을 통해 앞으로 방대한 양의 유저체 정보를 보다 체계적이며 조직적으로 관리할 수 있을 것으로 기대한다. 또한 구축한 주석정보 데이터베이스를 기반으로 본 시스템에서는 웹 서비스를 공개함으로써 외부 사용자는 웹 서비스 기술을 통해 유전체 주석정보 데이터를 서비스 받을 수 있을 뿐만 안이라 본 논문을 통해 구현한 시스템과 유사한 시스템을 로컬 시스템에서 쉽게 구현할 수 있다. 향후 과제로써 사용자에게 검색 결과를 좀더 사용자 친화적으로 보일 수 있게 하기 위하여 사용자 인터페이스에 관한 연구와 함께 웹 서비스 기반 통합검색 시스템을 통해 얻은 유전체 주석정보 데이터를 효율적으로 이용하는 방안에 대한 연구가 필요할 것으로 보인다.

참고문헌

[1] Cynthia Gibas, Bioinformatics Computer Skill, O'RELLY, 2002.

[2] Tisdall, James, Beginning Perl for Bioinformatics, O'REILLY, 2001.

[3] Jurgen Kaljuvee, Bioinformatics Data Integration Approach 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] 정지훈, 웹 서비스, 한빛미디어, 2002.

[6] Healther Kreger, Web Services Conceptual Architecture, IBM Software Group, White Paper, 2001.

[7] http://omnigene.sourceforge.net/, Overview of OmniGene

[8] http://www.bioperl.org/, A tutorial for BioPerl.