

바이오인포매틱스 Web Services

조희형, 안부영, 안성수, 박재홍, 박형선
한국과학기술정보연구원 바이오인포매틱스센터
e-mail:choh2@kisti.re.kr

Web Services of Bioinformatics

Hee-Hyung Cho, Bu-Young Ahn, Sung-Soo Ahn,
Jae-Hong Park, Hyung-Seong Park
Dept of CCBB, KISTI

요 약

근래 들어 생명과학 및 생명정보에 대한 관심이 높아지면서, 활발한 연구 활동의 결과물로 많은 정보를 얻을 수 있게 되었다. 방대한 생명정보를 분석할 수 있는 많은 알고리즘이 개발되었으며, 웹 기술이 발달함에 따라 생물정보의 활용도가 높아져서 다양한 방법으로 정보 서비스가 가능하게 되었다. Web Services는 XML, SOAP, WSDL, UDDI를 사용하는 분산 컴포넌트 기반의 컴퓨팅 기술로 서비스를 제공하는 Service Provider와 서비스 제공자의 정보를 저장하는 Service Registry 및 서비스를 사용하는 Service Requestor로 구성된다. 본 논문에서는 바이오인포매틱스 분야에서 Web Services를 구성하는 모든 요소들을 개발한다. 특히 One-Stop Web Services는 GUI 환경의 인터페이스로 사용자에게는 용이한 실행 환경을 제공하고 있으며, Web Services 개발자에게는 사용자 인터페이스 개발 부담을 줄일 수 있다.

1. 서론.

세계적으로 생명과학 및 생명정보에 대한 관심이 높아짐에 따라 활발한 연구 활동으로 방대한 생명 정보가 생산되었고, 앞으로도 생산될 것이다. 특히 Human Genome Project 수행 후, 많은 유전정보가 데이터베이스화 되었으며, 유전정보를 분석할 수 있는 도구가 개발되었다. 대표적인 데이터베이스로는 GenBank, PDB, dbEST, SWISS-PROT 등이 있으며, 분석도구로는 BLAST, FASTA, ClustalW, InterProScan 등이 있다.

또한 최근 웹 분야 기술이 발달함에 따라 생물정보의 활용도가 높아지고 다양한 방법으로 정보를 서비스할 수 있게 되었다. 인터넷상에서 지리적으로 분산된 데이터베이스를 검색, 활용하는 기술로 Web Services 기술이 있다. 이 기술은 데이터 표현방식으로 XML를 사용하고, XML 문서를 교환하는 프로토콜은 SOAP을 사용한다. 클라이언트에서 서버의 서비스를 호출할 수 있도록 서비스의 내용을 기술하는

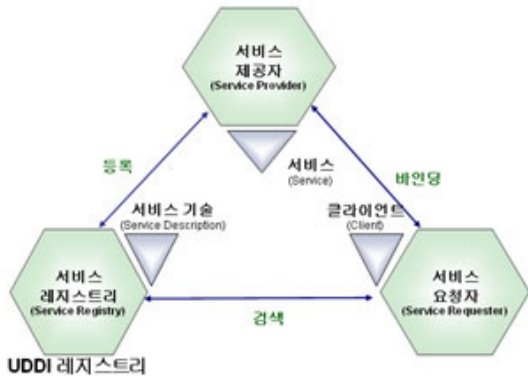
언어는 WSDL이며 서비스를 UDDI 저장소에 저장하고 검색하는 구조를 가지고 있다.

본 논문에서는 바이오인포매틱스 분야의 FASTA, ClustalW, DIP, BIND 등을 Service Provider로 제공하고, Service Registry인 UDDI와 사용자가 Web Services를 활용하는데 손쉬운 방법을 제공하는 GUI 환경의 응용프로그램인 One-Stop Web Services 개발에 관하여 기술하고자 한다.

2. Web Services

Web Services는 인터넷 어딘가에 자리잡고 요청자에게 데이터와 서비스를 제공하는 응용프로그램 로직의 단위로 응용프로그램들은 각각의 Web Services가 어떻게 구현되었는지 인지할 필요 없이 HTTP, SOAP과 같은 프로토콜로 Web Services에 접근한다. Web Services 이용은 웹사이트에 접속하는 것처럼 간단할 수도 있고, 다층 조직적인 구조를 가지는 복잡한 일일 수도 있다. 현재 많이 사용하는

기술인 CORBA, CGI 스크립트, J2EE 등을 Web Services 기술로 분류할 수 있지만, 표준화된 XML을 사용한다는 점에서 다르다고 말할 수 있다. <그림 1>은 Web Services의 구조를 나타낸다.



<그림 1> Web Services 구조

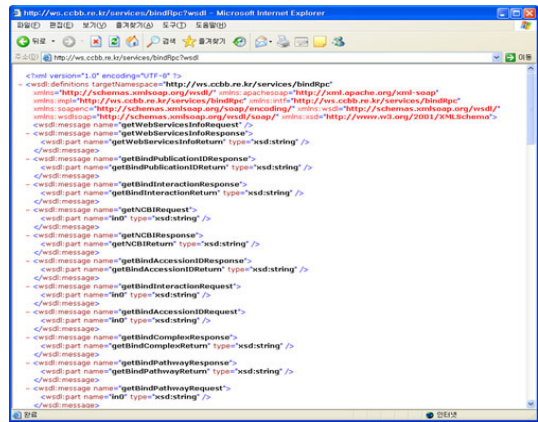
2.1 SOAP(Simple Object Access Protocol)

SOAP는 XML 프로토콜 표준으로 Web Services의 요청 및 응답에서 사용되어지는 XML 메시지 형식과 통신 프로토콜과의 바인딩 방법을 정의하고 있다. SOAP은 분산된 컴퓨팅 환경에서 서비스 요청자와 서비스 제공자 사이에 구조적이고 유형화된 정보를 교환하는데 있어 필요한 간단하면서도 경량의 XML 기반 프로토콜이라는 특징을 가지며, HTTP 프로토콜을 그대로 이용할 수 있어 기존 웹 환경을 확장한 형태로 분산컴퓨팅 환경을 구현할 수 있다. 또한 CORBA, RPC 등과 같은 기존 분산컴퓨팅 솔루션 이용시 발생할 수 있는 방화벽에 의한 통신 장애에 대한 문제를 고려할 필요가 없으며, 임의의 포트 번호를 위한 소켓 설정이 필요치 않다는 장점도 가지고 있다.

2.2 WSDL(Web Services Description Language)

WSDL은 Web Services의 바인딩 및 실행을 위해 필요한 인터페이스와 입출력 메시지의 형식, 포트 정보 등과 같은 기술정보들을 정의하는데 사용되며, CORBA나 COM의 IDL(Interface Definition Language)과 유사한 역할을 수행한다. WSDL은 네트워크 서비스에서 메시지를 교환할 수 있는 통신 종점(endpoint)들의 집합으로 정의할 수 있도록 XML 문법을 사용함으로써 Web Services의 기술에 필요한 정보를 정의할 수 있게 한다. WSDL 서비스 정의는 분산시스템에 대한 설명서를 제공하고 응용 프로그램간의 통신과 관련된 자세한 정보의 정의와 이용을 자동화할 수 있도록 한다.

Web Services는 일반 사용자나 어플리케이션 등의 서비스 요청자가 요구되는 서비스에 대한 메시지 작성을 한 후, SOAP이나 다른 특정 프로토콜을 사용하여 전달하고, 이 메시지를 받은 서비스 응답자는 응답 메시지를 작성하여 다시 서비스 요청자에게 보내어지는 일련의 과정으로 설명할 수 있다. SOAP을 통한 Web Services 이용시 WSDL을 이용하여 Web Services 정보를 기술하고, 전송에 사용되는 SOAP 메시지는 여러 파트들로 구성되는데 각 메시지 파트들은 XML Schema를 이용해 정의된 데이터 형들을 이용하게 된다. <그림 2>는 CCBB(Center for Computational Biology & Bioinformatics)에서 제공하는 BIND의 WSDL 문서이다.

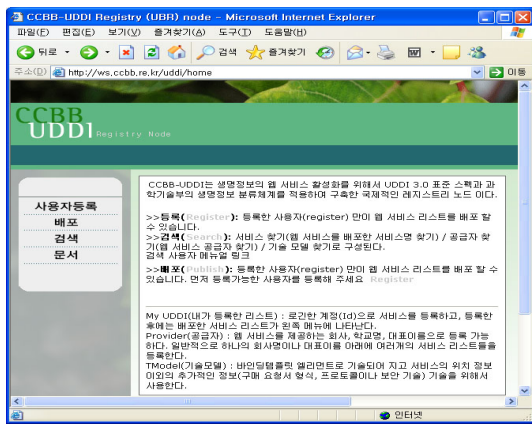


<그림 2> CCBB에서 제공하는 BIND의 WSDL

2.3 UDDI 레지스트리

UDDI(Universal Description, Discovery and Integration) 표준안은 Web Services 관련 정보를 등록하고 검색할 수 있는 방법을 정의한다. “Web Services”라는 용어는 인터넷 연결을 통해 기업 또는 소프트웨어 프로그램이 서비스를 이용할 수 있는 방법을 제공할 목적으로 한 기업의 특정 업무 기능으로 정의된다. 현재 Web Services는 전자상거래를 위한 중요한 프로그래밍 기술로 자리를 잡아가고 있다. 한 예로, 기업들은 인터넷 연결을 통해 구매신청서를 전송하거나 배송할 물품의 크기, 무게, 거리 등의 조건을 통해 운송비용을 계산하는데 있어 다른 기업의 서비스를 이용할 수 있다. 이러한 예에서 이미 알고 있는 비즈니스 파트너와 전자상거래 게이트웨이를 이용한다면 Web Services의 발견 과정이 필요하지 않을 수도 있다. 그러나 현재 알거나 이용해보지 않았던 특정 서비스나 이를 공급하는 제공자를 검색하는 경우에는 Web Services의 발견 기능에 대한 지원이 절실히 요구되어진다. 서비스 발견이라는 문제를 풀기 위한 방법은 각 기업의 웹 사이트에 Web Services

관련 정보를 기술한 파일을 저장시키고, 이를 이용하여 특정 업체에서 제공하는 서비스 목록을 검색하는 것이다. 이러한 과정은 웹 로봇(또는 크롤러)과 같은 프로그램이 등록된 URL을 통해 각 기업의 사이트에 접속하여 웹 페이지의 텍스트들을 검색하고 인덱싱함으로써 가능해진다. 분산적인 접근 방법은 잠재적으로는 안정적인 방법이 될 수 있으나, 서비스 기술(description) 형식과 서비스 변경 이력의 처리 등의 문제에 있어 일관성을 보장하기 어려운 구조를 가진다. UDDI에서는 비즈니스에 대한 분산 레지스트리 방식을 지원하고, 비즈니스의 서비스 기술 정보를 일반적인 XML 형식으로 정의하여 이용할 수 있는 방법을 제공한다. <그림 3>은 CCBB의 UDDI 레지스트리이다.



<그림 3>은 CCBB의 UDDI 레지스트리

2.4 One-Stop Web Services

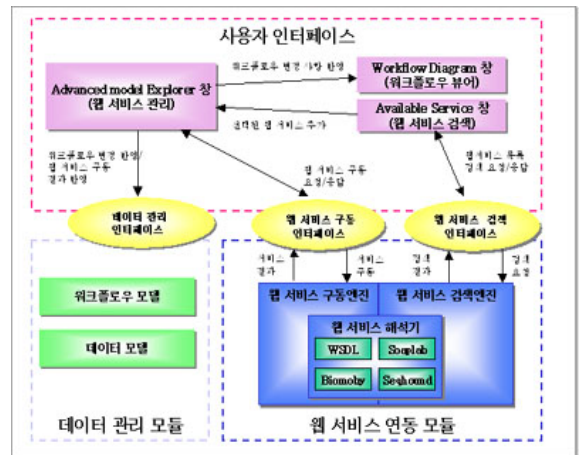
바이오인포매틱스에서는 전세계적으로 배포되어 있는 여러가지 Web Services의 효율적인 관리를 통하여 사용자에 대한 서비스 강화 및 분산 구축되어 있는 생명정보 자원의 활용성을 높이고자 하였다.

생명정보자원에 대한 활용성을 높이려면 CCBB, NCBI, EBI, DDBJ 등의 사이트에서 배포한 Web Services에 접근하여 입력 값과 출력된 결과를 확인 및 유지하고, 객체들의 흐름을 볼 수 있는 프로그램의 개발이 필요하다. 이에 One-Stop Web Services 관리 프로그램을 개발하였다.

2.4.1 시스템 개요

One-Stop Web Services 관리 프로그램은 Web Services 검색 및 구동을 담당하는 Web Services 연동 모듈, Web Services 워크플로우 및 Web Services 구동 결과 값을 관리하는 데이터 관리 모듈, 그리고 사용자가 Web Services 검색, 워크플로우 생성 작업을 할 수 있도록 지원하는 사용자 인터

페이스로 구성되어 있다. <그림 4>는 One-Stop Web Services의 시스템구성도이다.

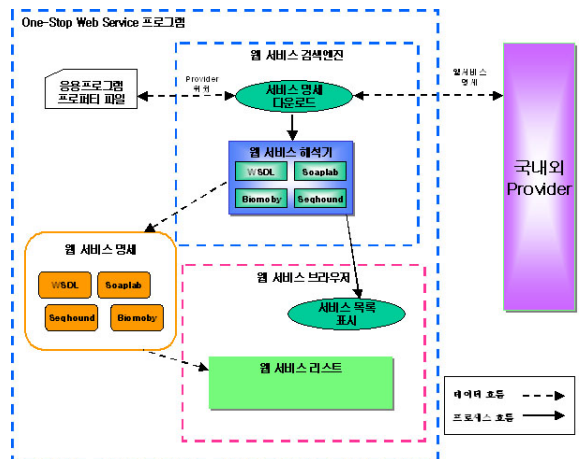


<그림 4> One-Stop Web Services의 시스템구성도

2.4.2 One-Stop Web Services 기능

- Web Services 검색

CCBB, NCBI, EBI, DDBJ, SEQHOUND 등 국내외 생명정보 Web Services Provider로부터 Web Services 명세를 다운로드 받고, 이를 각각의 Web Services 종류 (WSDL, Soaplab, Biomoby, Seqhound)에 따라 해석하여 사용자가 확인할 수 있게 한다. <그림 5>는 Web Services 검색기능 흐름도를 나타낸다.



<그림 5> Web Services 검색기능 흐름도

- 워크플로우 생성/수정

서비스 브라우저에 표시된 Web Services 목록에서 사용자가 원하는 서비스를 자유롭게 선택하여 등록/삭제함으로써 생명정보를 가공/처리할 수 있는 객체흐름을 생성/수정할 수 있는 기능을 제공한다. 워크플로우 데이터모델은 Web Services 함수 및 각각의 함수를 구동하기 위한 입/출력 값 그리고 여러

Web services 함수들을 연결하여 데이터 객체들의 흐름을 나타내기 위한 링크정보 등으로 구성하였다.

- 워크플로우 뷰어

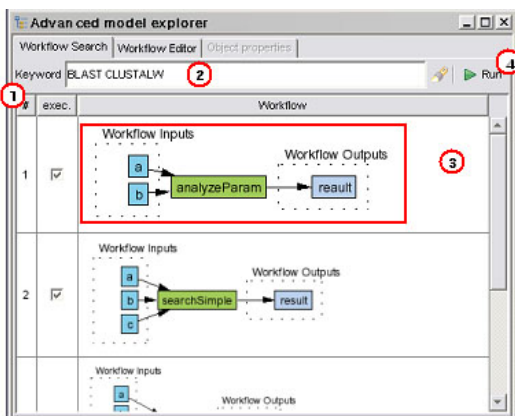
Web Services 검색 창에서 사용자가 생성/수정 중인 워크플로우를 다이어그램으로 표시하며, 작업 중인 워크플로우에 대한 변경사항을 반영하는 기능을 가지는 뷰어를 구현하였다.

- 워크플로우 실행

객체흐름을 구성하는 각각의 Web Services 함수들을 구동할 수 있는 프로세스들을 생성하여 Web Services로부터 결과를 얻어오는 기능으로 프로세스 생성 시에는 Web Services 해석기를 통하여 해당 Web Services의 종류(WSDL, Soaplab, etc)를 파악하고 적합한 프로토콜을 사용하는 프로세스를 하고, 각 프로세스들을 객체흐름 순서에 맞도록 수행시키는 Web Services 스케줄러에 의하여 프로세스를 실행시킨다.

- 워크플로우 라이브러리 검색 및 활용

컴퓨터에 친숙하지 않은 사용자의 경우 Web Services 검색에서 직접 워크플로우를 생성/수정하고 이를 실행시켜 결과를 얻어내는데 어려움을 겪을 것으로 예상된다. 그래서 워크플로우 라이브러리에서는 자주 사용될 것으로 예상되는 Web Services를 선정하여 워크플로우를 생성한 후, 라이브러리로 제공하므로 사용자가 쉽게 찾아 실행할 수 있도록 지원하도록 하였다. <그림 6>은 워크플로우 라이브러리를 검색한 결과이고 결과에 대한 설명은 다음과 같다.



<그림 6> 워크플로우 라이브러리 검색

① Web Services 검색의 Workflow Search에서 그

림과 같이 라이브러리의 워크플로우 목록을 검색할 수 있다.

② Blast/Fasta/Clustalw 등 Web Services의 종류를 나타내는 단어를 입력한다. 각 단어들은 공백으로 구분되어 여러 개를 입력할 수 있으며, 샘플 워크플로우 중 해당되는 종류의 Web Services가 포함된 워크플로우가 검색된다.

③ 목록에 표시된 각각의 워크플로우를 선택하면, Workflow Editor와 Workflow Diagram 창에 해당 워크플로우가 반영되어 자세한 내용을 확인할 수 있다.

④ Run 버튼을 누르면 목록의 워크플로우 중 exec 체크박스가 선택된 워크플로우를 실행하게 된다.

3. 결론

Web Services는 분산 컴포넌트 기반의 컴퓨팅 기술로 산업 전반에 이용할 수 있는 기술이다. 본 논문은 바이오인포매틱스 분야에서 분산 구축되어 있는 생명정보자원을 활용하는 Web Services를 제공하며, 이를 사용자가 쉽게 실행할 수 있도록 워크플로우 라이브러리를 개발하여 검색, 활용할 수 있도록 하였다. CCBB, NCBI, EBI, DDBJ 등의 사이트에서 배포한 Web Services에 접근하여 입력 값과 출력된 결과를 확인, 유지하고, 객체들의 흐름을 볼 수 있는 One-Stop Web Services는 GUI 환경의 인터페이스로 사용자에게는 용이한 실행환경 제공을, Web Services 개발자에게는 사용자 인터페이스 개발부담을 줄일 수 있을 것으로 사료된다.

참 고 문 헌

[1] 안성수, 생물다양성데이터 검색포탈 구축, 한국콘텐츠학회, 2005.
 [2] 한국과학기술정보연구원, UDDI 기반 생명정보 웹 서비스 시스템 개발, 최종연구보고서, 2004.10
 [3] Paul Lipton, Composition and Management of Web Services, SYS-CON, 2004.
 [4] Wang.L.Riethoven.J.J and Robinson.A, XEMBL: distributing EMBL data in XML format, Bioinformatics, 2002.
 [5] Lee.j.M and Sonnhammer.E.L, Genomic gene clustering analysis of pathways in eukaryotes, Genome Res, 2003.
 [6] Tom Oinn, Taverna:a tool for the composition and enactment of bioinformatics workflows, Bioinformatics, 2004.