

XML-BASED BIOINFORMATIC SYSTEMS

XML 기반의 생물정보학시스템

신종환, 정무영

포항공과대학교

제품생산기술연구소/기계산업공학부

Abstract

Bioinformatics can be explained as the science of developing algorithms, application tools, and computer databases and so on, for the purpose of supporting and enhancing biological research. Bioinformatic information systems (BIS) typically handle large data sets and the amount of the data goes up exponentially. Another impediment to easy extraction and retrieval of genomic data in BIS is the need to access different sites for similar information. Recently, there has been some attempts to integrate bioinformatics data in the World Wide Web (WWW) among the bioinformatics community by the internet computing technology. However, the work to integrate bioinformatics data on a universal platform has some problems because of the lack of standard, terminologies, semantics, and ontologies about bioinformatics. In this paper, an XML-based BIS architecture is proposed as an integrated BIS framework. The XML and related technologies allow the creation of meaningful information tags to exchange data between various databases as a standard format, and to create more simple interfaces. This integrated BIS framework has bioinformatic architectural components which is used in the Corporate Information Factory (CIF) method.

Keyword : Bioinformatics, Bioinformatic Information Systems (BIS), XML, Data Integration, Corporate Information Factory (CIF)

1. 서론

최근의 분자생물학 연구의 범위는 대규모의 유

전자와 단백질 데이터의 분석을 통해 단백질의 구조와 기능을 밝혀내는데 집중되어 있다. DNA 칩과 같은 유전자를 동시에 비교할 수 있는 방법이 도입되고 전세계적으로 연구가 병렬적으로 진행되어감에 따라 분산된 대규모의 데이터베이스 시스템의 출현이 가속화 되고 있다. 기하급수적으로 증가하는 생물학적 데이터로부터 의미있는 정보를 도출하기 위해서는 보다 효율적인 생물정보학 분석 도구가 요구된다. 현재 새롭게 제작되는 생물정보학 분석도구들은 웹에서 동작 가능한 것들도 있으나, 여전히 유닉스 (Unix), 윈도우즈 (Windows), 리눅스 (Linux) 등 다양한 운영체제 기반으로 되어 있는 것들이 대부분이다. 또한 여러 데이터 소스로부터 각자 다양한 데이터 포맷을 사용하고 분석도구들간의 데이터 표준이 부재하여 생물학 연구의 공동작업, 소프트웨어 개발, 생물학적 데이터의 관리 등에 많은 시간과 비용이 든다[1].

그러므로 이질적인 데이터 포맷으로 표현된 다양한 데이터 소스로부터의 생물학적 데이터를 하나의 공통된 프레임워크로 통합하는 문제는 앞으로 생명정보학에서 새롭게 다루어야 할 문제라고 할 수 있다. BT (Bio Technology) 산업이 확장되면 생물정보시스템 (BIS : Bioinformatic Information Systems)은 데이터마이닝 기법 등을 이용한 의사결정 도구들까지 포함한 지능적인 기업형 데이터베이스 시스템으로 진화할 것으로 예상된다. 따라서 생물학 데이터의 표준과 거대한 생물학적 데이터에 대응하는 데이터베이스 시스템의 통합 프레임워크 개발은 필수적이라고 할 수 있다.

XML (eXtensible Markup Language)은 사람과 기계가 동시에 읽을 수 있고 웹을 통해 정보를 교환할 수 있는 수단이라는 장점을 가지고 있어 점차 그 적용범위가 전 산업분야로 확대되어가고 있다. 본 논문에서는 XML을 데이터 포맷, 통신 프로토콜 등 데이터 통합의 기반으로 하고, 기업정보공장 (CIF : Corporate Information Factory)[2] 개념을 적용하여 새로운 생물정보시스템 통합 프레임워크를 제안하고자 한다. 기업정보공장이란 빌 인몬 (Bill Inmon)이 제안한 웹환경에서 대규모의 기업용 정보 시스템을 관리하기 위한 패러다임으로 거대한 데이터를 수집, 보관, 가공하여 일관된 정보를 제공하는 수준의 데이터웨어하우스를 포함하고 있다.

2. XML과 생물학적 데이터의 통합

2.1 웹환경에서의 생물정보학 응용프로그램

현재 생명과학자들은 많은 양의 생물학 데이터를 찾거나 등록하는 작업을 NCBI (the National Center for Biotechnology Information)[3]나 PDB (the Protein Data Bank)[4]의 홈페이지를 통해 자유롭게 하고 있다. 웹 환경에서는 인터넷 익스플로러 (Internet Explorer)와 같은 브라우저만 있으면 다양한 플랫폼에 독립적이면서 추가 비용 발생 없이 인터넷 기반으로 쉽게 데이터 통신이 가능하기 때문에, 일반적으로 웹기반의 응용프로그램들이 널리 쓰이고 있는 실정이다. 그러나 현재 대부분의 웹페이지에서 사용중인 HTML (HyperText Markup Language)는 잘 알려진 대로 구조적 표현이 불가능하고 의미를 함축할 수 없으며, 기계가 읽을 수 없고, 확장성에 한계가 있어 복잡한 생물학적 데이터를 표현하는 데는 불완전하다.

2.2 생물정보학 공동체 내에서의 XML

XML은 SGML (Standard Generalized Markup Language)과 HTML의 한계를 극복하면서 웹을 통해 문서를 주고받을 수 있는 새로운 표준 문서 포맷으로 떠오르고 있다. 사용자에게 의해 정의된 태그에 의해 한개 혹은 그 이상의 XML 문서들이 파생될 수 있고, DTD (Document of Type Definitions)형태로 태그들에 대한 의미와 규칙들을 미리 정의해 놓을 수도 있다. XML 태그는 계층적 구조를 가지는 문서의 구성요소의 type을 명시함으로써 추론과 회귀가 가능하고, 데이터의 출판이나 생물학 데이터처럼 매우 구조적인 형태의 데이터도 표현이 가능하다. 또한 XSL (eXtensible Stylesheet Language), DOM (Document Object Model) 등과 같은 풍부한 XML 관련 기술들과 소프트웨어가 뒷받침되어 있다는 장점을 가지고 있다.

이런 XML의 장점 때문에 지놈 커뮤니티 (Geneome cummuniuty)에서도 XML을 도입하려는 움직임이 활발히 일어나고 있는 추세이다[5]. 주요 유전자 및 단백질 서열 데이터베이스인 EBML, DDJB, NCBI, PDB와 GeneOntology[6] 등의 기관에서는 부분적으로 XML로 표현된 데이터를 취급하기 시작하였다. 예를 들면, 3대 유전자 데이터베이스로 불리는 EBML, GenBank, DDBJ에서 단백질의 서열과 그래프 구조를 표현하기 위해 BSML (Bioinformatic Sequence Markup Language)[7]을 사용한다. 그 외에도 염기와 단백질의 복잡한 서열정보를 명명하기 위한 BIOML (BIOPolymer Markup Language)[8] 등도 XML을 기반으로 하고 있다. 그러나 여전히 XML 수용에 대한 찬반론이 생물정보학계에서는 여전히 제기되고 있으며 이미 수많은 데이터베이스와 웹사이트들이 서로 다른 방식으로 유전자 데이터, 질의문, 용어 등을 취급하고 있다 [9]. 그러므로 앞으로 생물정보학은 통합이라는 중요한 해결 과제를 남겨두고 있다.

2.3 생물학 데이터 통합의 두 단계

본 논문에서는 생물정보학에서의 통합을 규모를 기준으로 크게 두가지 수준, 즉 데이터 모델 수준과 기업측면의 전체 BIS 아키텍처로 나누어 다루고자 한다. 데이터 모델 수준에서의 통합은 염기와 단백질 데이터, 단백질의 2차 및 3차 구조, 그리고 명명법과 같은 관련정보 등에 대해 어떻게 데이터를 XML 과 같은 표준으로 표현하고 교환할 것인가 하는 문제를 다룬다. 그리고 기업측면의 전체 BIS 아키텍처 수준에서의 통합은 염기 혹은 단백질 서열정보와 각종 생명정보학 분석 도구 등을 포함한 대규모의 데이터베이스들을 통합하고 관리하는 문제들을 포함한다.

데이터 모델 수준에서의 통합이 필요한 가장 큰 이유는 동일한 항목에 대해 FASTA, BLAST, Genethon과 같은 여러 가지 데이터 소스들이 서로 다른 용어들을 사용하는 데서 시작된다. 동일한 데이터베이스의 'Segment ID'에 대해 다른 용어들이 사용되면 중복되는 데이터가 늘어나 결국 데이터베이스의 관리는 시간이 지날수록 복잡해지고 어렵게 된다. 예를 들어 'G Protein-coupled receptor proteins'의 화합물인 'Bovine rhodiosin (dark side)'의 구조에 대해 PDB와 GenBank에서 검색할 경우, 동일한 구조임에도 불구하고 사용자는 두 가지의 서로 다른 데이터 집합을 얻게 된다. 이는 GenBank와 PDB에서 사용하는 'Bovine rhodiosin (dark side)' 데이터 레코드가 일부는 중복되고 일부는 상이하기 때문이다. 그 밖에도 MMDB, PROMOTE, CSU, STING 등에서도 각자 상이한 데이터 포맷을 가지고 있다 [10]. 그러므로 분자생물학계와 생물정보학계가 공동으로 합의하고 표준으로 인정한 공통 용어를 가지고 XML 문서로 통합하여 데이터베이스 검색을 일원화시킬 필요가 있다. XML로 작성된 생물정보학 데이터 관련 문서는 미리 정의된 DTD에 의해 유효성 (validity)을 검증받을 수 있다. [그림 1], [그림 2], [그림 3]은 각각 임의의 가계도 데이터 집합 (pedegrees data set)에 대해 작성한 DTD, Schema, XML 문서의 예이다.

```
<ELEMENT pedigree-set (pedigree+)>
<ATTLIST pedigree-set
  title CDATA #REQUIRED
  trait CDATA #IMPLIED
>
<ELEMENT pedigree (person+)>
<ATTLIST pedigree
  title CDATA #IMPLIED
  id ID #REQUIRED
>
<ELEMENT person ( epid-data)* >
<ATTLIST person
  id ID #REQUIRED
  name CDATA
  pedigree IDREF #REQUIRED
  father IDREF #IMPLIED
  mother IDREF #IMPLIED
  sex ( male | female | unknown ) 'unknown'
  status ( affected | unaffected | unknown ) #REQUIRED
  birth-year CDATA
>
<ELEMENT epid-data ( qualitative-data | quantitative-data >
<ATTLIST epid-data
  name CDATA #REQUIRED
>
<ELEMENT qualitative-data (#PCDATA)>
<ELEMENT quantitative-data (#PCDATA)>
```

[그림 1] An example of Pedegrees DTD

```

class Pedegree-set {
set <Pedegree * > pedigrees (nonnull. card > 0);
string title (nonnull, unique);
string trait;
};

class Pedegree {
set <Person * > persons (nonnull. card > 1, inverse<Person:pedegree>);
string title (nonnull);
enum Sex {
male;
female;
unknown;
};
enum Status {
affected;
unaffected;
undetermined;
};
class Person {
string name;
Pedegree * pedegree (nonnull, inverse<Pedegree:persons>);
Person * father;
Person * mother;
Sex sex = unknown;
Status status = undetermined;
int birth-year;
Epid-data epid-data[];
};
class Epid-data {
string name (nonnull);
};
class Qualitative-data extends Epid-data {
string value;
};
class Quantitative-data extends Epid-data {
float value;
};
};
    
```

[그림 2] An example of Pedegrees Schema

```

<?xml version="1.0"?>
<!DOCTYPE document SYSTEM "pedegree.dtd"
<document title="An example of pedegree" trait="music genius">
<pedegree title="Bach family" id="pl">
<person id="1" name="Johann Sebastian" pedegree="pl"
sex="male" status="affected" birth_year="1685">
<epid-data name="pitch">
<qualitative_data> absolute pitch
</qualitative_data>
</epid_data>
</person>
<person id="2" name="Masia Barbara" pedegree="pl"
</person>
<person id="3" name="Catharina Dorotbea" pedegree="pl"
father="Johann Sebastian" mother="Maria Barbara"
sex="female" status="unknown">
</person>
<person id="4" name="Wilhelm Friedemann" pedegree="pl"
father="Johann Sebastian" mother="Marla Barbara"
sex="male" status="unknown">
</person>
<person id="5" name="Carl Philip Emmanuel" pedegree="pl"
father="Johann Sebastian" mother="Maria Barbara"
sex="male" status="affected">
</person>
<person id="6" name="Jobann Gottfried Bernhard" pedegree="pl"
father="Johann Sebastian" mother="Maria Barbara"
sex="male" status="unknown">
</person>
</pedegree>
</document>
    
```

[그림 3] An example of Pedegrees XML

한편 기업측면의 전체 BIS 아키텍처 수준에서의 통합은 실험실의 원시 데이터를 데이터웨어하우스 같은 데이터 저장소에서 정제시킨후, XML 기반의 글로벌 접속 플랫폼(global access platform)을 제공하기 위한 BIS 디자인, 최적화 작업 등을 포함된다. 본 논문에서는 기업정보공장의 기능적 구성요소들을 새롭게 정의하여 생물학 데이터의 통합에 필요한 요구사항에 적합한 통합 BIS 프레임워크를 제안하였다.

3. 통합 생물정보학 시스템 프레임워크 제안

기업정보공장은 전사적 자원관리 (ERP :

Enterprise Resource Planing)과 같은 기업의 대규모 정보시스템을 디자인하기 위한 웹 기반의 프레임워크이다. 기업정보공장은 프로세스 기반의 여러 시스템들로 구성되어 있다. 기업정보공장의 주요한 구성 요소로는 어플리케이션 환경 (application environment), 통합 및 변환 레이어 (ITL : Integration and Transformation Layer), 데이터 웨어하우스, 데이터마트 (DM : Data Mart), 작업 데이터 스토어 (ODS : Operational Data Store), 인터넷, 인터넷, 메타데이터 저장소 (metadata repository) 등이 있다.

3.1 생물학 데이터 통합에 대한 요구사항

NETTAB 2001 워크숍에서는 생물정보학과 정보기술, XML 등을 주제로 다음과 같은 문제점과의 의견을 도출한 바 있다[11].

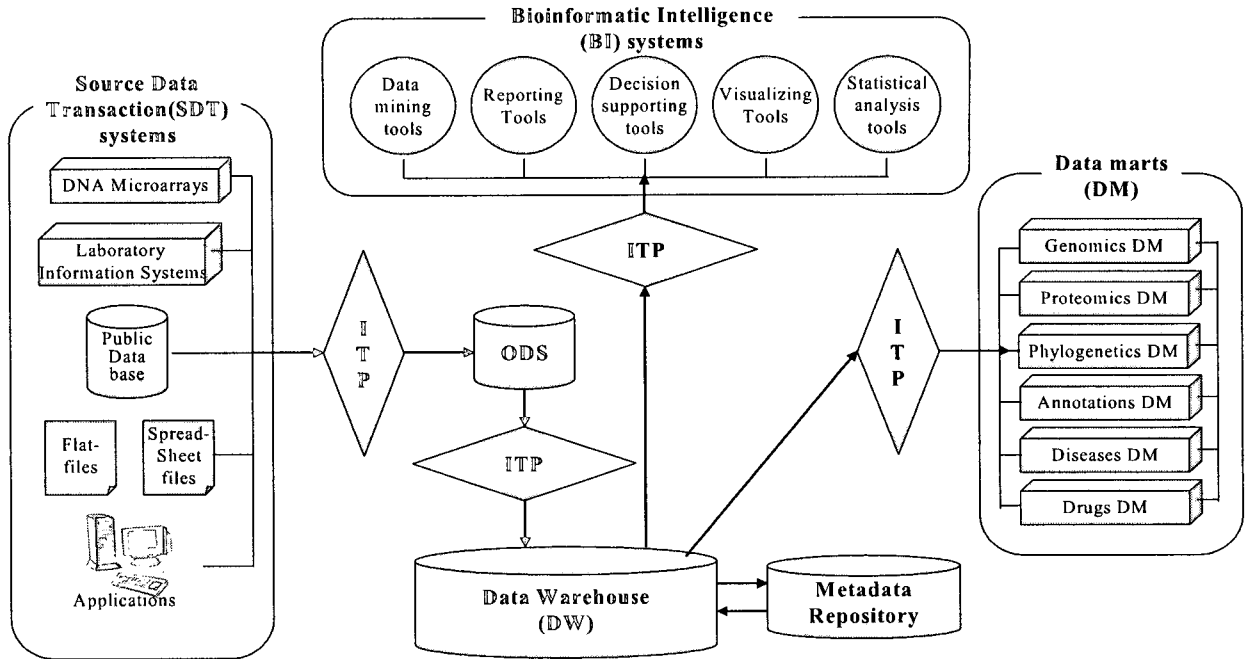
- 현재 가용한 정보 기술에 대해 생명정보학계에서 수용하려는 자세를 가져야 한다.
- 방화벽 (firewall)이 데이터 교류의 장벽이 되고 있다.
- 자신의 연구 데이터가 공개되어 기계가 읽을 수 있는 형태로 바뀌는 것에 대해 꺼리지 말아야 한다.
- 제공된 데이터에 대한 품질을 측정할 수 있는 방법이 고안되어야 한다.
- 생물정보학 관련 자원들을 통합하기 위해서는 제한된 어휘와 데이터 사전, 온톨로지 (ontology) 등이 중요한 요소들이다.

위의 결과에서 보듯이 생물정보학 데이터 통합의 핵심요소는 데이터의 내용과 관련된 의미 (Semantics)를 정의할 수 있는 언어에 대한 연구라고 할 수 있다. 이와 관련 W3C에서 RDF (Resource Description Framework)와 시맨틱웹 (Semantic Web) 작업그룹을 통해 웹의 데이터에 의미를 부여할 수 있는 방안에 대해 연구를 진행하고 있다[12]. 분자 생물학계에서도 유전자 온톨로지 컨소시엄 (Gene Ontology Consortium)[6]에서 웹에 산재한 생물학적 데이터에 대한 의미를 표현하고 기계가 읽을 수 있는 형태로 변환하기 위해 필요한 생물학적 아키텍처, 모델, 표준 등에 대해 연구를 하고 있다. 그러나 아직까지 이러한 연구는 시작 단계에 불과하며 온톨로지 교환 언어와 도구와 같은 기술적인 문제 외에도 생물학계 전체에 대한 더 많은 이해와 합의가 필요하다.

3.2 제안된 통합 생물정보학 시스템 프레임워크의 구성 요소

3.1에서 기술한 요구사항들을 충족시키기 위해 본 논문에서는 XML과 SOAP (Simple Object Access Protocol)을 기반으로 하는 통합 BIS 프레임워크를 제안하였다[그림 4]. 제안된 프레임워크는 [그림 4]와 같이 원시 데이터를 처리하는 SDT (Source Data Transactional systems), SDT로부터 파생된 데이터를 통합하고 변형하는 ODS (Operational Data Store), DW (Data Warehouses), DM (Data Marts), 메타데이터 저장소 (metadata repository), 그리고 다양한

의사 결정 도구와 데이터마이닝 및 통계 도구, 리된 온톨로지와 관련 어휘사전을 탑재한 엔진을



[그림 4] 제안된 통합 BIS 프레임워크의 구성요소

포트 출력 도구 (report generating tools), 형상화 도구 (visualization tools) 등을 포함하는 시스템인 BI (Bioinformatic Intelligence) 시스템 등으로 구성되어 있다.

3.2.1 SDT (Source Data Transaction systems)

SDT는 데이터가 만들어지는 시점에서의 정보를 저장하고 관리하는 시스템이다. SDT는 각 시점에서 대량의 세부적인 휘발성 원시 데이터를 포함하게 된다. 생물정보학의 관점에서는 실험실의 정보시스템 (LIS : Laboratory Information Systems), DNA Microarray, Microarray Scanner, 웹을 통해 자유롭게 접근가능한 공개 데이터베이스와 개인 데이터베이스, 원시 데이터를 표현한 아스키 플랫폼파일 (ASCII flatfiles), 스프레드시트파일, 다양한 응용프로그램들이 SDT에 속한다. 대표적인 SDT의 예로는 웹을 포함한 다양한 플랫폼에서 접근가능한 PDB를 들 수 있다.

3.2.2 ODS (Operational Data Store)

ODS는 기업차원의 의사결정 프로세스를 지원하기 위해 사용되는 주제 지향 (subject-oriented), 필요에 따라 삭제될 수 있는 데이터들의 집합체이다. 생물정보학 측면에서 ODS는 이질적인 SDT로부터 수집된 중심 데이터 저장소의 역할을 담당하는 데이터베이스라고 할 수 있다. ODS는 BIS 전체에서의 정보 흐름을 위해 데이터의 표준화 작업을 수행한다. 즉, BSML과 같은 XML 기반으로 제작된 데이터 포맷으로 작성되지 않은 원시 데이터들은 DTD에 의해 ODS에 입력되기 전에 발견되게 함으로써 표준화된 데이터만 수용하도록 한다. 유전자 온톨로지 컨소시엄과 같은 기관에서 제안한 표준화

ODS가 갖추어야 전체 BIS에서 중복되지 않는 데이터가 유지될 수 있다.

3.2.3 DW (Data Warehouse)

DW는 주제 지향적이면서, 데이터의 휘발성이 없으며, 시간에 따라 변하는 (time variant) 통합 데이터와 세부정보들을 담고 있는 데이터베이스 시스템이다. 단백질 데이터, 유전 데이터, 클리닉 데이터와 같은 생물정보학의 세부 주제별로 미리 정의되어 구성되므로 아미노산 서열, 단백질 3차 구조 및 기능, X-ray NMR (Nuclear Magnetic Resonance) 구조, 단백질간의 상호작용 등에 대한 주제별 카테고리화가 가능해진다.

3.2.4 DM (Data Mart)

기업에서는 특정 분야의 정보 관리와 분석을 위해 DW와 별도로 DM을 유지하기도 한다. 즉, DM은 기업내 특정사업부서가 세부적인 분석의 필요에 의해 DW로부터 불러와 한단계 요약한 맞춤형 데이터 집합이다. 생물정보학 측면에서 DM은 지노믹 (genomic) 데이터, 프로테오믹 (proteomic) 데이터, 유전표현형 (phenotypic) 데이터, 명명법 (annotation) 관련 데이터, 약품, 질병 관련 데이터 등을 유지하기 위한 데이터베이스라 할 수 있다. DM에 속한 데이터는 특정한 분석이나 보고서 출력 도구, 의사 결정 도구에서 쓰일 수 있도록 데이터 구조를 유지해야 한다. 또한 DM에서 데이터를 추출하는 과정에서의 성능을 향상시키고 단순화시키기 위해 DM 내에 쿼리 최적화 도구 (query optimizer) 등을 개발하여 사용할 수도 있다.

3.2.5 BI (Bioinformatic Intelligence) Systems

BI 시스템은 의사 결정을 지원하기 위해 필요한 데이터 마이닝 알고리즘, 통계적 분석 도구, 형상화 도구, 리포트 출력 도구 등의 BIS를 지원하는 지능적인 기능을 갖춘 시스템이다. DM이나 DW로부터 받은 데이터를 보다 정결한 고품질의 정보로 변환하기 위해 BI 시스템은 보다 스마트한 알고리즘과 분석 도구들을 갖추어야 할 필요가 있다.

3.2.6 ITP (Integration and Transformation Process) layer

ITP 레이어는 BIS 전체에서 보다 효율적인 생물학 데이터들의 교환과 흐름을 위해 디자인된 일종의 프로토콜 개념과 유사하다. SDT부터 BI까지 흐르는 데이터와 정보가 단일의 표준화된 형태로 유지되기 위해서는 BIS 구성요소간 이음새 없는 (seamless) 통과 계층이 필요하다. 데이터 내용이나 구조를 표준으로 표시할 수 있는 XML과, 방화벽 문제를 해결하면서 기존의 HTTP를 수용하는 SOAP이 해결책이 될 수 있다. 코바 (Corba)와 비교할 때 SOAP이 보다 인터넷에 접목하기 쉬우며 기존의 C/C++, Phythin 외에도 BioPERL, BioJAVA 등 생물정보학 전용 프로그래밍 언어도 활발하게 연구되고 있다는 점에서 장점을 가진다.

3.2.7 Metadata Management

메타 데이터 관리란 BIS 전체 프로세스에서 발생하는 메타 데이터들을 효율적으로 관리하기 위해 필요하다. 생물정보학 측면에서의 메타데이터는 생명정보학 메타데이터 (bioinformatic metadata)와 기술적인 메타데이터 (technical metadata) 두가지 형태로 나누어볼 수 있다. 생명정보학 메타데이터에는 각 데이터의 출처, 변환된 위치, 업데이트된 상황, 작성 및 삭제된 날짜, 작성자 등이 있을 수 있다. 또한 기술적인 메타데이터에는 BIS내의 데이터 베이스의 테이블 종류, 각종 데이터에 대한 입출력 로그 파일 등이 포함된다. 가령 데이터 트래픽 (data traffic)과 같은 데이터의 이동에 따른 부수적으로 발생하는 데이터에 대해서도 전체 BIS 디자인 시 고려해야 할 사항중의 하나이다.

4. 결론 및 추후 연구과제

본 논문에서는 생물정보학에서의 통합의 필요성을 기술하고, 통합에 관련된 요구사항에 맞도록 BIS를 두 가지 수준 즉, 데이터 모델과 기업측면의 전체 BIS 아키텍처 수준으로 나누어 XML 기반의 통합 BIS 프레임워크를 제시하고 구성 요소들을 정의하였다. 생물정보학에서의 통합에는 기술적인 요인들 외에도 분자생물학계와 생명정보학계에서 표준화된 데이터와 온톨로지에 대한 이해와 동의가 필요하다. 제안된 프레임워크에 기반하여 실제 가능한 시나리오들을 대상으로 프로토타입을 구현하는 것은 추후 연구해야 할 과제이다.

감사의 글

본 논문은 한국과학재단 목적기초연구 (2001-1-31500-005-1)의 일부 지원으로 수행되었음.

Reference

- [1] Wong, R. and Shui, W., "Utilizing multiple bioinformatics information sources : An XML database approach", Proceedings of the IEEE 2nd International Symposium on Bioinformatics and Bioengineering, pp. 73-80, 2001
- [2] Imhoff, C., Sousa, R., and Inmon, W., "Corporate information factory," 2nd ed., John Wiley & Sons, 2000
- [3] <http://www.ncbi.nlm.nih.gov/>
- [4] <http://www.rcsb.org/pdb/>
- [5] Cheung, K., Liu Y., Kumar A., Snyder M., Gerstein, M., and Miller, P., "An XML application for genomic data integration," Proceedings of the IEEE 2nd International Symposium on Bioinformatics and Bioengineering, pp. 97-103, 2001
- [6] <http://www.geneontology.org/>
- [7] <http://www.visualgenomics.com>
- [8] <http://www.bioml.com/BIOML/>
- [9] Achard, F., Vaysseix, G., and Barillot, E., "XML, bioinformatics and data integration," Bioinformatics Review, 17(2), pp. 115-125, 2001
- [10] http://www.rcsb.org/pdb/cgi/explore.cgi?job=others&pdbId=1HZX&page=20&pid=273491014508862&mia_direct=1,2,3,4,5,6,7&mia_in_expl=1/
- [11] Martin, A., "Can we integrated bioinformatics data on the internet?," Trends in biotechnology, 19(9), 2001
- [12] <http://www.w3.org/>
- [13] Mcentire, R., Karp, P., Abernethy, N., Beton, D., Helt, G., DeJongh, M., Kent, R., Kosky, A., Lewis, S., Hodnett, D., Neumann, E., Olken, F., Pathak, D., Tarczy-Hornch, P., Toldo, L., and Topaloglou, T., "An evaluation of ontology exchanges languages for bioinformatics," Proceedings of 8th Int'l Conf. on Intelligent Systems for Molecular Biology, pp. 239-250, 2000
- [14] Baxevanis, A. and Francis, B., "Bioinformatics, a practical guide to the analysis of gene and proteins," 2nd ed., John Wiley & Sons, 2001