

DNA microarray chip 을 위한 LIMS

이유진*, 차재혁**, 임상택*, 노정호*, 심진욱*

*한양대학교 정보통신대학원 석사과정

**한양대학교 정보통신대학원 조교수

e-mail : eu98@ihanyang.ac.kr

LIMS for DNA microarray chip

Yu-jin Lee*, Jae-Hyuk Cha**, SangTeak Lim*, Jeongho Rho*, Jinwook Shim*,

* The Graduate School of Informations & Communications, Hanyang University

** The Graduate School of Informations & Communications, Hanyang University, Assistant Professor

요 약

본 논문은 DNA microarray chip 을 사용한 실험 결과로 생산되는 대량의 데이터를 효율적으로 관리하기 위한 LIMS 개발에 대해 기술한다. 기존의 상용 LIMS 는 보편적 패턴과 방식을 정규화하여 제공하기 때문에 실험실의 고유한 방식을 포함하긴 어렵다. 본 논문에서는 유연성 있는 LIMS 를 개발하기 위해 특정 실험 중심으로 설계하면서 MAGE-OM 의 표준을 따르도록 디자인 하였고, HYLIMS manager 라는 Local Application 과 검색을 주로 이용하는 사용자를 위하여 Web 검색 시스템을 구현하였다. 데이터베이스의 부하를 줄이기 위해 데이터 저장용 DB 와 검색용 DB 를 구분하였고, 데이터를 타입과 처리 형태에 따라 분류하여 관리하였으며 데이터 보안을 위해 실험 관리자가 사용자의 접근 제한을 설정 할 수 있도록 하였다.

1. 서 론

DNA microarray 는 기존의 유전 공학 방법보다 획기적으로 많은 수의 유전자를 빠른 시간 내에 검사 할 수 있는 방법이다. 수 cm^2 이하 크기의 유리 또는 실리콘으로 된 chip 위에 수만에서 수십만 가지의 서로 다른 DNA 분자를 2 차원 매트릭스 형태로 집약해 놓았기 때문에 한꺼번에 많은 DNA 를 검사 할 수 있다. DNA microarray 의 활용으로 짧은 시간 내 대량의 데이터가 만들어지며, 이러한 작업은 하루에도 여러 번 이루어져 수많은 데이터를 관리하기 위해 전산화가 필요하다. 실험실의 실험들은 많은 수작업을 포함하고 있다. 정보 및 데이터의 수집, 저장, 분석 등의 과정에 많은 시간이 소비되고 있기 때문에 이를 개선하기 위하여 체계적인 실험실 정보 관리 시스템이 필요하다. 이러한 실험실 정보관리를 위한 전산화 시스템을 LIMS(Laboratory Information Management System)라 한다. 국내 실험실에서는 미국 FDA, EPA 등의 기준에 부합되고 시스템 패키지의

완성도도 높은 시스템을 도입하여 사용한다. 그러나 시스템 도입 비용이 상대적으로 높고, 실험실마다 상이한 실험 디자인과 노하우를 가지고 있기 때문에 일반적인 개념을 포함하고 있는 기존 LIMS 는 특정 실험에 대해 적합하지 않아 활용이 제한적이다.

본 연구에서 제안하는 LIMS 는 특정 연구실에서 실험하는 cDNA microarray chip 의 특징을 포함하고 다른 연구실의 데이터를 통합 관리가 가능하도록 설계하고자 한다. 이를 위해 본 연구의 LIMS 가 Object management Group, Inc(OMG)라는 국제기관에서 제안한 MicroArray and GeneExpression Object Model(MAGE-OM)이라는 표준안을 따르도록 하고 실험 진행과정 관리를 위해 Local Application 을 구현하며 검색을 위한 Web 검색 시스템을 구현하여 효율적인 데이터 관리와 검색이 이루어 지도록 한다. 또한 데이터 부하를 줄이기 위해 목적에 따라 데이터베이스를 구분하여 설계하였고 데이터 보안을 위해 접근 권한을 사용자 별로 설정 하고자 한다.

2. DNA microarray chip 실험에서 기존 LIMS 문제점

DNA microarray chip은 생물체로부터 얻은 DNA 또는 RNA와 같은 시료에 형광물질을 넣은 것과 반응하도록 한다. 이 반응은 Hybridization이라 부르는 것으로 DNA 또는 RNA가 서로 염기서열이 상보적인 것끼리 결합을 하려는 성질을 이용한 것이다. 즉, 어떤 생물체 시료에 약물의 투여와 같은 조작을 가하거나 또는 압조직과 같은 것에서 DNA 나 RNA를 추출하여 이에 형광물질 표지를 붙이면 그 결합 정도가 바로 얻고자 하는 데이터가 된다. 이것은 결합에 참여하지 못한 것을 씻어낸 다음 남아 있는 형광물질의 양을 측정하여 분석결과를 얻어낸다. 그러나 위의 실험과정에서 많은 에러가 발행 하기 때문에 이를 줄이기 위해 실험실 마다 다른 노하우를 가지고 있다. 예를 들면, 형광물질의 양을 측정하기 위해 scanner를 통해 이미지를 만들고 그 이미지를 추출하고자 할 때, 정확한 측정을 위해 이미지 spot 추출 방법과 noise제거 및 normalization을 하기도 하지만 초기에 microarray 디자인을 할 때 같은 gene을 여러 개 존재하도록 하기도 한다. 이렇듯 다른 실험 설계 및 실험실 특징 때문에 일반적인 LIMS는 사용자 용도에 정확하게 부합하지 않는다.

LIMS와 다른 회사나 자체 개발한 분석 도구 및 mining tool을 사용한다면 LIMS를 이용하여 실험 과정을 자동화하기는 어렵다.

기존 LIMS는 데이터베이스를 이용하여 데이터 관리 및 검색도 가능 하게 하였으나 이는 같은 LIMS를 사용할 때만 가능하다.

3. HYLIMS 제안

한양대 의대의 DNA microarray chip 실험을 관리하기 위한 LIMS로 한양 LIMS 라는 의미의 HYLIMS를 개발하였다. HYLIMS는 한양대 의대 실험실에서 산출되는 데이터를 관리하기 위해 설계되었으며 다른 실험실의 데이터도 관리 할 수 있도록 일반적인 방법을 포함한 형태로 구현하였다.

3.1 보편적 방법의 수용

Object management Group, Inc(OMG)는 MicroArray and GeneExpression Object Model(MAGE-OM)을 제시하였다. 이것의 주요한 목적은 재사용성, 이식가능성과 분산된 이종의 환경에서 객체지향의 소프트웨어가 상호작용 할 수 있도록 하는 것이다.

MAGE-OM은 [그림 1]의 Experiment, BioAssay DesignElement, ArrayDesign, Array, BioSequence, BioAssayData, BioMaterial, BioEvent, Measurement 등 17개의 Object package로 구성되어 있다.

우리가 제시하는 DNA microarray 실험을 위한 LIMS가 MAGE-OM을 포함하도록 하여 다른 기관의 실험 데이터에 대해서도 재사용 가능하고 확장 가능하

도록 디자인 하였다.

3.2 특정 목적의 수용

Affymetrix®LIMS Manager가 Affymetrix®Microarray Suite와 NetAffx®WebSite라는 data mining tool을 포함하는 것처럼 LIMS는 다른 어플리케이션을 포함할 수 있다.

HYLIMS는 한양대 의대의 실험 데이터 형식이 표준과 차이가 나기 때문에 이를 저장하고 검색 가능하도록 디자인 하였다. 또한 microarray 이미지 분석 도구로 자체 개발한 MAAS(MicroArray Analysis System)를 사용한다. MAAS는 microarray 이미지의 농도를 분석한 엑셀파일, 분석된 이미지의 재실행을 편리하게 하는 cod 파일, 분석 결과 보고서인 HTML 파일을 생성한다. HYLIMS는 MAAS의 데이터를 처리하기 위한 방법과 저장 구조를 구현하였다.

이는 표준을 따르면서 특정 목적에 맞는 어플리케이션을 제작하여 사용자 인터페이스의 편의를 제공하도록 한 것이다.

3.3 데이터형과 관련된 조작들과 그 표현 형태의 결합

표준 모델을 따르면서 MAAS라는 프로그램을 호출하고 그 실험결과를 관리하기 위해 데이터를 구분하여 관리한다. 실험 과정 데이터는 테이블 구조로 저장하고 MAAS 실험 결과 데이터는 BLOB 형태로 저장한다

4. HYLIMS 데이터베이스 스키마 디자인

MAGE-OM의 표준을 따라 데이터베이스 스키마를 디자인 하여 다른 실험데이터와 통합 관리 할 수 있도록 하면서 MAAS 데이터 및 한양대 의대 실험 데이터를 관리할 수 있도록 디자인 하였다.

Experiment 테이블은 Project와 Sample이라는 상위 테이블을 가지고 있다. 실험 그룹을 Sample이라고 하고, 실험에 대한 다른 분류를 그룹 지어 놓은 것이 Project 테이블이다.

각 실험 과정을 저장 하기 위한 테이블로는 GridAlignment, Giomaterial, Hybridization, BioAssay, Treatment, Reporter, Composite, S/W, H/W, Scan이 있다.

파일 관리를 하는 DataManagement 테이블을 만들어서 그 테이블 하위에 Jpg File, Xls File, Cod File, Cye3 Tif File, Cy5 Tif File, Csv File을 각각 관리 할 수 있도록 하였다.

실험에 따른 실험자의 데이터 접근 권한을 주기 위한 SecurityGroup이란 테이블을 만들었다. 이 테이블은 Experiment와 Person과의 관계에서 생긴 테이블이다.

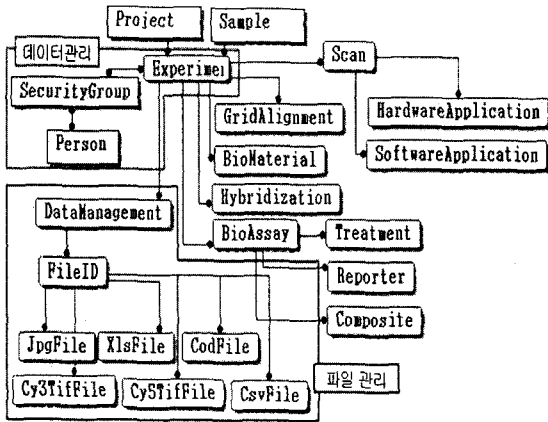


그림 1 HYLIMS 데이터베이스 스키마 디자인

5. HYLIMS 아키텍처

client/server 구조를 참조한 모델로 client 는 web 기반의 query interface 를 이용하여 검색할 수 있는 검색 전용 시스템이고 server 는 LIMS manager 라는 application 을 이용하여 실험을 관리할 수 있는 부분과 ProcessDB 와 PublishDB 를 포함한 소켓 통신 부분으로 구성하였다.

Web query Interface 를 이용함으로써 실험 참여자 외의 사용자가 완료된 실험에 대해 어디서, 어떠한 환경이든 검색 가능 하도록 구현하였다. HYLIMS manager 는 Local application 으로 수정, 배포의 용이함과 플랫폼의 제한을 받지 않도록 Java 가상 머신에서 실행하도록 구현하였다.

5.1 HYLIMS manager

Local Application 인 LIMS manager 는 Process 와 Publish, 그리고 Role 모듈로 구분한다. Process 는 실험과정을 입력하는 모듈이고, Publish 는 local hard drive 에 임시로 저장된 실험 내용을 원격 데이터베이스에 저장한다. Role 은 실험에 참여한 사람들의 권한을 설정해 줄 수 있고 포함된 사람들의 권한과 개인 정보를 확인할 수 있다.

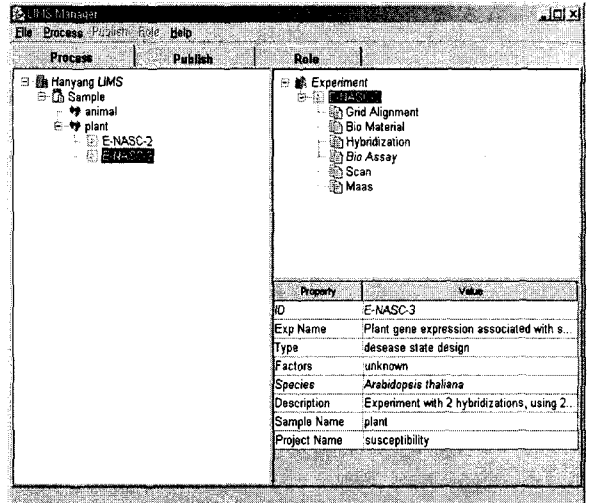


그림 3 HYLIMS Manager

5.2 HYLIMS Web 검색 시스템

실험자와 일반 사용자가 어디서나 환경의 제약 없이 Web query interface 를 이용하여 실험을 검색할 수 있도록 시스템을 구현하였으며 실험 과정의 정보관리와 검색을 분리하여 데이터베이스의 부하를 줄였다. 검색의 편의를 위해 검색어를 사용자가 입력 하는 것이 아니라 이미 데이터베이스에 저장되어 있는 값들만을 선택하여 검색할 수 있도록 하였다. 검색결과와 항목에 링크를 두어 관련항목들 사이에 유기적으로 연관되도록 하였다.

[표 1]은 검색어 리스트로 Experiment, BioMaterial & BioAssay, Protocol 과 Member 를 검색어로 이용할 수 있다. 각 검색어 리스트가 상하로 나뉘어 있는 것은 각 검색어에 여러 개가 결과를 가지는 것과 검색어에 하나의 결과를 주는 것으로 구분한 것이다.

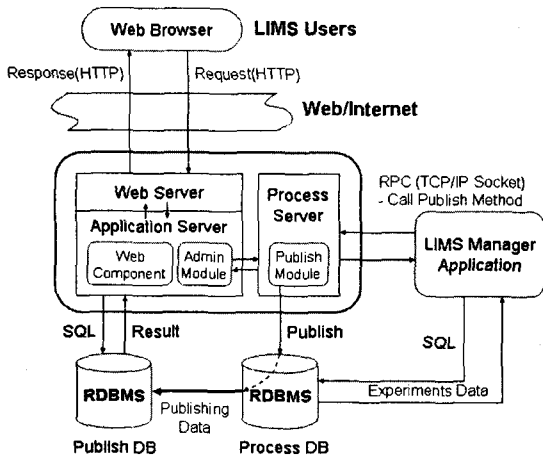


그림 2 HYLIMS 아키텍처



그림 4 HYLIMS Web 검색 시스템의 메뉴

그림 5 HYLIMS Web 검색 시스템 Experiment 검색

표 1 HYLIMS 검색시스템의 검색어 리스트

Experiment	BioMaterial & BioAssay
Sample name	BioMaterial Type
Project name	BioMaterial Character category
Experiment type	BioAssay Type
Experiment factor	
Experiment species	
Experiment name	BioMaterial Character Value BioMaterial Accession Number BioAssay Unigene BioAssay Cloneid
Protocol	Member
Protocol Type	Group
S/W name	Person name
H/W name	

5.3 Process DB & Publish DB

Process DB는 진행중인 실험의 과정의 정보를 저장하는 데이터베이스이고 Publish DB는 완료된 실험을 저장하는 데이터베이스이다. Publish DB는 HYLIMS manager의 Publish 모듈에서 실험이 완료되었을 때 완료된 실험 정보를 Publish DB에 저장하도록 한다. 실험 과정 정보뿐 아니라 실험 결과 파일을 다운로드 할 수 있도록 파일구조를 가지고 있고,

검색이 가능 하도록 인덱스 구조도 가지고 있다.

6. 결론

HYLIMS manager와 Web 검색 시스템을 구현하여 MAAS 데이터를 포함한 특정 실험 데이터를 체계적으로 관리하도록 하였다. 이 시스템은 특정 실험뿐 아니라 다른 실험도 포함할 수 있는 유연성을 가지도록 하기 위해 MAGE-OM의 표준을 따르도록 하였고 데이터 관리를 효율적으로 하기 위해 데이터 타입과 처리방법에 따라 구분하여 저장하였다. 실험 과정에서 산출된 데이터를 저장하는 데이터베이스와 완성된 결과를 검색하기 위한 데이터베이스를 분리하여 구성함으로써 시스템의 부하도 줄이고 데이터 관리도 체계적으로 할 수 있게 하였다.

HYLIMS가 특정 실험 목적을 위해 만들었지만 보편적인 특징을 수용하기 때문에 다른 실험 데이터를 수용할 수 있었다. 그러나 실험의 특수성을 포함하여 개발하기 위해서는 관련 실험자의 협력이 필요하다.

참고문헌

- [1] OMG, Inc. "Gene Expression Specification" Draft Adopted Specification, February 2002
- [2] EMBL-EBI ArrayExpress Query
"http://www.ebi.ac.uk/arrayexpress/query/entry"
- [3] 황승용 "DNA Chip 기술" 정보과학회지 제 18 권 8 호 pp.23-28
- [4] 원세연 "생물정보 분야의 개괄과 전망" 정보과학회지 제 17 권 5 호 pp. 49-56
- [5] AffyMetrix, Inc AffyMetrix@Laboratory
"Information Management System User's Guide version 3.0"
- [6] Perter E. Schilling "Technical Strategy for laboratory information management systems" Laboratory Information Management 26 (1994) pp.143-151
- [7] J.F. Tfigg, J.A.P.Smith, Y.Lerat "Do end-users meet the LIMS requirements?" Laboratory Information Management 26 (1994) pp.181-187