

Application of UML (Unified Modeling Language) in Object-oriented Analysis of Microarray Information System

UML을 활용한 마이크로어레이 정보시스템의 객체지향 분석

Ji Yeon Park, Hee Joon Chung, Ju Han Kim*

Seoul National University Biomedical Informatics, Seoul National University College of Medicine, Seoul 110-799, Korea

*To whom correspondence should be addressed. E-mail: juhan@snu.ac.kr

Abstract

Microarray information system is a complex system to manage, analyze and interpretate microarray gene expression data. Establishment of well-defined development process is very essential for understanding the complexity and organization of the system. We performed object-oriented analysis using Unified Modeling Language (UML) in specifying, visualizing and documenting microarray information system. The object-oriented analysis consists of three major steps: (i) use case modeling to describe various functionalities from the user's perspective (ii) dynamic modeling to illustrate behavioral aspects of the system (iii) object modeling to represent structural aspects of the system. As a result of our modeling activities we provide the UML diagrams showing various views of the microarray information system. We believe that the object-oriented analysis ensures effective documentations and communication of information system requirements. Another useful feature of object-oriented technique is structural continuity to standard microarray data model MAGE-OM (Microarray Gene Expression Object Model). The proposed modeling efforts can be applicable for integration of biomedical information system.

1. 서론

유전자 마이크로어레이 기술의 도입으로 유전자 발현 데이터가 급증하게 되었고 이를 효과적으로 처리할 수 있는 정보시스템의

개발이 필요하게 되었다. 정보시스템의 목적은 방대하고 복잡한 데이터를 체계적으로 관리하고, 다양한 외부자원과 연결하여 유용한 지식발견을 촉진하는 것이다 [1]. 정보시스템의 개발공정은 분석, 설계, 구현, 테스트 과정으로 구성된다. 분석과정에서는

성공적인 시스템 개발에 필요한 사항들을 식별하고 시스템의 기능과 수행 범위를 명시한다. 분석과정은 소프트웨어 개발 초기의 과정이며, 이 과정에서 원하는 요구사항을 분명히 파악하지 못하면 시스템 개발 전반에 많은 영향을 끼치게 됨으로 매우 중요한 과정이다 [2]. 특히 마이크로어레이 정보 시스템과 같이 다양한 요구조건을 반영해야 하고 요구조건이 변화하기 쉬운 영역에서는 문제를 명확히 정의하고 분석하는 체계적인 방법론의 지원이 중요하다.

소프트웨어 개발 방법론이란 소프트웨어 시스템을 구축하는데 필요한 여러 가지 일들의 수행방법과 이러한 일들을 효율적으로 수행하려는 과정에서 필요한 각종 기법과 도구를 체계있게 정리하여 표준화한 것이다. 그 중 객체지향 개발방법은 다른 기법에 비해 풍부한 모델링을 지원하기 때문에 급속히 보급되고 사용되고 있다.

본 논문에서는 신뢰성 있는 소프트웨어 개발을 위해 보편화되고 있는 객체지향 분석 과정을 통하여 마이크로어레이 정보시스템을 명시하고자 하였다. 마이크로어레이 정보시스템의 구조나 행위에 대한 여러 관점의 모델을 구축하여 시스템을 시각화 하므로 개발하고자 하는 시스템을 정확히 이해할 수 있게 하고 시스템의 확장과 변경에 유연하게 대응할 수 있게 된다. 또한 표준 모델링 표기체계인 UML (Unified Modeling Language)을 사용하여 분석결과를 문서화하므로 사용자와 개발자간의 의사소통을 원활히 하고 시스템 개발경험을 공유할 수 있게 하고자 하였다.

본 논문은 다음과 같이 구성된다. 먼저 마이크로어레이 정보시스템, 객체지향 분석, UML에 관한 연구 배경을 소개하고 3장에 UML 기반의 객체지향 분석과정을 단계별로 설명한다. 4장에서는 마이크로어레이 정보시스템의 객체지향 분석결과를 종류별로 UML 모델을 통해 제시한다. 마지막으로 마이크로어레이 정보시스템에서 객체지향 분

석방법을 활용하였을 때의 이점들을 정리하고 활용가능성을 제시한다.

2. 연구 배경

1) 마이크로어레이 정보시스템

유전자 마이크로어레이는 한 생물체의 유전자 전체 또는 일부에 대해 유전자 발현양을 측정할 수 있는 도구이다. 유전자 발현 데이터는 개개 유전자 및 조직표본에 대한 정보와 실험조건, 실험방법들에 대한 상세한 주석이 있어야 해석이 가능하기 때문에 다차원적인 자료구조를 필요로 한다 [3]. 이렇게 형태가 다양하고 복잡한 구조의 마이크로어레이 데이터를 연구자들 간에 공유하기 위해 MGED (Microarray Gene Expression Data) 단체에서 표준화 작업을 진행하고 있다. MAGE-OM (Microarray Gene Expression Object Model)은 마이크로어레이 데이터와 모든 실험정보를 세밀하게 표현한 객체지향의 표준 데이터 모델이고, MAGE-ML (Microarray Gene Expression Markup Language)은 MAGE-OM을 XML 형식으로 구현한 표준 데이터교환 포맷이다. MGED 표준화 제안이 공식적으로 인정받게 되면서 표준을 지원하는 정보시스템의 개발이 요구된다 [4].

2) 객체지향 분석

객체지향 분석은 시스템의 문제영역을 현실의 실체를 담은 '객체'를 중심으로 생각하고 문제를 해결해 나가는 방식이다. 객체는 데이터와 기능을 하나의 단위로 하여 현실 세계를 그대로 반영하는 구조를 가지고 있다. 객체와 객체들의 상호작용이라는 관점으로 시스템을 이해하는 객체지향 기법은 기능 중심의 구조적 기법이나 자료구조 중심의 정보공학 기법보다 실세계를 자연스럽게 표현할 수 있는 체계이다. 객체지향 기

술은 최근 확장성과 재사용성을 높이는 핵심기술로 부상하고 있다 [5].

3) UML

UML은 객체지향 기반의 표준 모델링 표기 체계이다. 여러 개의 그래픽 기호로 구성되며 각 기호들은 정확한 의미를 가지고 있다. UML은 하나의 시스템을 용도와 목적에 따라 다양하게 표현할 수 있도록 여러 종류의 다이어그램을 제공한다. 정보시스템의 분석, 설계, 구현 과정에서의 모든 중요한 결정에 대해 명세서를 다룰 수 있다 [6].

3. 연구 방법

정보시스템 개발에서 분석과정은 사용자의 관점에서 요구사항을 파악해서 정의하고 시스템의 개념적 논리적 모습을 대략적으로 표현하는 단계이다. UML 기반의 객체지향 분석과정에서 요구분석은 유스케이스 모델링을 통하여 표현되고 시스템의 논리적 개념적 모델은 객체 동적 모델링과 객체 모델링을 통하여 구축된다 [그림 1].

1) 유스케이스 모델 작성과 객체 식별

유스케이스 모델은 시스템 개발 착수 후 최

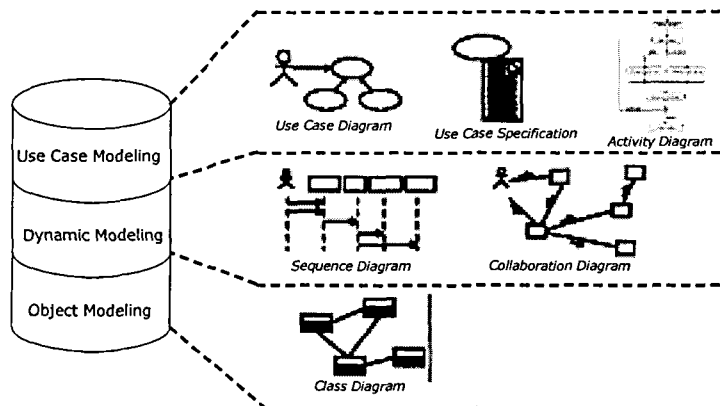
초로 작성되는 모델로 시스템의 기능적 측면을 사용자 관점에서 설명하는 명시적인 모델이다. 객체지향 분석, 설계 과정 전반에 걸쳐서 기준이 되고 뼈대가 되는 중요한 모델이다. 유스케이스 모델링의 결과 유스케이스 다이어그램 (use case diagram)과 유스케이스 명세서 (use case specification)가 작성된다. 유스케이스 다이어그램은 시스템이 제공하는 서비스 (유스케이스)와 외부 환경 (액터) 간의 관계를 도식화 한 것이고, 유스케이스 명세서는 각각의 유스케이스의 작업 흐름을 기술한 문서이다. 활동 다이어그램 (activity diagram)은 유스케이스 명세서에 기술된 처리 흐름을 간결하게 표현하는 용도로 사용될 수 있다. 유스케이스 모델을 작성한 후 유스케이스 명세서의 처리흐름에서 시스템에 필요한 객체를 도출하고 추상화하여 클래스를 정의하므로 객체지향 모델의 기본 체계를 구축한다.

2) 객체간 동적 모델 작성

객체간 동적 모델링이란 시간개념 하에서 시스템 내부의 객체간 처리순서를 정의한 것이다. 동적 모델링은 유스케이스 명세서에서 역할을 수행하는 객체들을 표현하고 객체간에 주고받는 메시지를 통하여 가시화된다. 2가지 관점에서 모델링이 가능하다. 시간적 관점에서 객체간 메시지 순서를 강조하는 순차 다이어그램 (sequence diagram)

[그림 1]

객체지향 분석과정에서 작성되는 모델
 객체지향 분석과정은 시스템의 기능분석을 위한 유스케이스 모델링, 시스템의 동적인 측면의 분석하는 동적 모델링, 시스템의 정적인 측면을 분석하는 객체 모델링으로 구성된다.



과 구조적 관점에서 메시지 순서를 강조하는 협력 다이어그램 (collaboration diagram) 을 작성할 수 있다. 동적 모델링을 수행하면 불필요한 실행모듈을 줄이고 시스템을 오류없이 정확하게 동작하게 하고 프로그램의 코딩 근거를 제시하게 된다.

3) 객체간 정적 모델 작성

객체 모델링은 시스템 범위에 속하는 객체 혹은 클래스들 간의 협력관계를 정적인 관계성으로 표현하여 도식화 하는 과정이다. 객체모델링의 결과 클래스 다이어그램 (class diagram)이 작성되는데 UML의 모델 가운데 가장 공통적으로 많이 쓰인다. 클래스 다이어그램은 클래스 (class)와 관계 (relationship)로 구성된다. 클래스는 공통 성질을 가지고 있는 객체를 추상화 한 것이며 속성 (attribute)과 오퍼레이션 (operation)으로 구성된다. 속성은 객체들이 유지할 정보를 정의한 것이고 오퍼레이션은 객체의 행위를 정의한 것이다. 클래스간 관계는 크게 연관 (association)과 일반화 (generalization) 관계로 구분된다. 두 클래스가 연관관계에 있다면 한쪽 객체에서 상대편 객체를 참조할 수 있음을 의미한다. 일반화는 객체의 특성 중 상속 (inheritance)을 표현하는 관계이다. 상속이란 상위 클래스의 모든 특성을 하위 클래스가 이어받음으로 중복을 없애는 방법을 제공한다.

4. 연구 결과

1) 마이크로어레이 정보시스템의 기능적 분석

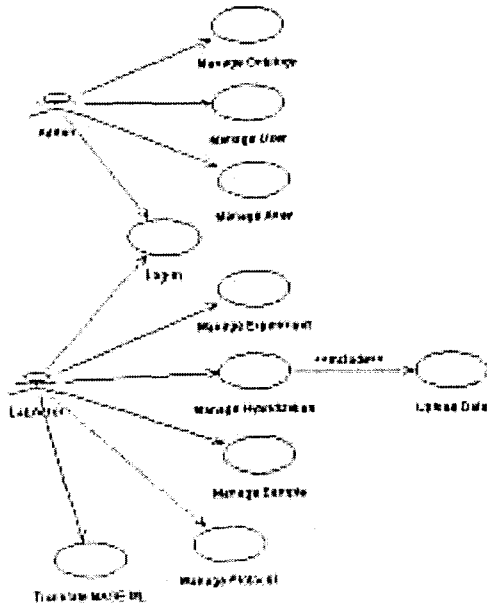
마이크로어레이 정보시스템의 예상되는 사용자들로부터 개발하고자 하는 시스템에 요구되는 기능들을 수집하여 정리하였다. [표 1]은 요구사항 수집을 통해 분리한 요구사항의 예이다.

[표 1] 마이크로어레이 정보시스템의 기능적 요구사항

-
- 실험에서 생성되는 다양한 데이터를 저장한다.
 - 마이크로어레이 유전자 정보를 저장한다.
 - 실험에 사용된 생물학적 시료와 실험 조건에 대한 정보를 저장한다.
 - 실험에 사용된 실험방법 프로토콜을 저장한다.
 - 프로젝트 별로 데이터를 관리한다.
 - 같은 연구 그룹에 소속된 연구자만 데이터를 열람한다.
 - MGED 온톨로지 용어를 사용한다.
 - MAGE-ML로의 데이터 전환이 가능해야한다.
-

시스템이 무엇을 해야 하는지를 정확히 파악하기 위해 추출된 요구사항을 가지고 유스케이스 모델을 작성하였다. 유스케이스 다이어그램은 사용자의 관점에서 시스템을 모델링 한 것으로 사용자가 시스템의 동작 방식을 쉽게 이해할 수 있도록 시각화한 것이다. 유스케이스 다이어그램을 작성하기 위해서는 먼저 액터 (actor)를 선정하여야 한다. 액터는 시스템 외부에서 시스템과 상호작용 하는 것으로 시스템에 서비스를 요청하는 사람이나 다른 시스템이 된다. 마이크로어레이 정보시스템의 경우 시스템을 작동시키는 대상으로 시스템 운영자 (Admin)와 실험자 (LabActor)를 선별하였다. 시스템을 사용하는 모든 사용자는 로그인을 통해 사용자 인증을 받아야 한다. 유스케이스는 특정 액터가 시스템을 통해 수행하는 일련의 동작을 모아놓은 것이다. 사용자의 요구사항의 항목 하나가 하나의 유스케이스가 될 수 있고, 관련있는 여러 개의 항목들을 합쳐서 하나의 유스케이스로 표현할 수 있다 [그림 2].

유스케이스 다이어그램이 작성되면 각각의 유스케이스가 수행할 작업흐름을 유스케이스 명세서로 작성한다. 유스케이스 명세서는 사용자의 관점에서 논리적인 일처리 흐름을 간결하고 명확하게 기술한 것이다 [표 2]. 유스케이스 명세서에 기술된 내용이 복



[그림 2] 마이크로어레이 정보시스템의 유스케이스 다이어그램

시스템을 작동시키는 액터는 막대모양의 사람 형태로, 시스템이 제공하는 기능을 정의한 유스케이스는 타원으로 표기된다. 액터와 유스케이스, 유스케이스와 유스케이스 사이를 연결하는 실선은 둘 사이에 교류가 이루어지고 있음을 나타낸다.

잡할 경우 활동 다이어그램으로 도식화하여 이해를 용이하게 한다. [그림 3]은 'Manage Hybridization' 유스케이스 명세서에서 hybridization의 등록부분에 초점을 두고 작성한 활동 다이어그램이다.

2) 마이크로어레이 정보시스템의 동적 분석

동적 모델이란 외부의 요청이 있는 시점부터 그 요청에 대한 처리를 다할 때까지 객체간의 내부적인 상호작용을 모델링 한 것이다. 유스케이스 명세서에서 역할을 수행하는 객체를 표현하고 시간의 흐름에 따라 객체와 객체간의 메시지 전달 흐름을 순차 다이어그램으로 표현하였다. 유스케이스별로 하나씩 작성할 수도 있고 한 개의 유스케이스 내부에 여러 가지 흐름이 존재하게 되면 각 흐름마다 모델링을 하게 된다. [그

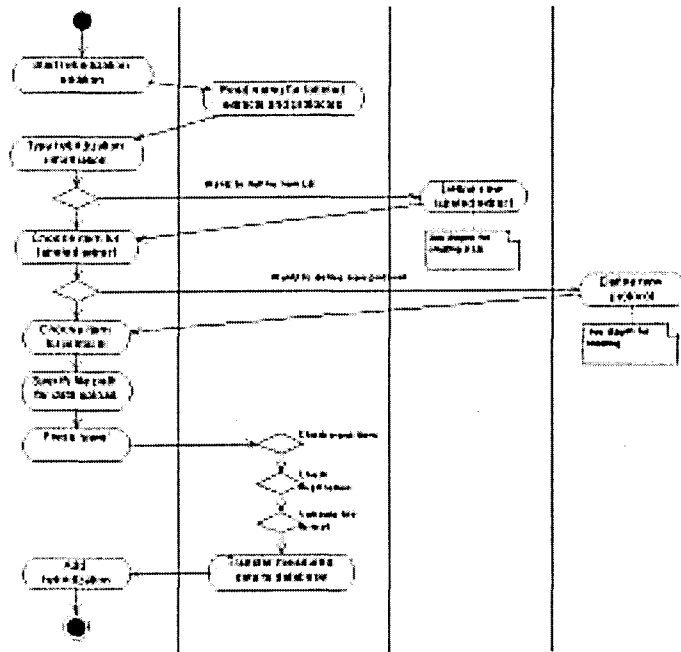
[표 2] 'Manage Hybridization' 유스케이스 명세서 이벤트 흐름 (Flow of events) 부분은 액터와 유스케이스 사이의 처리흐름을 상세히 정의하는 부분이며 두 종류의 흐름이 있다. 기본 흐름 (Basic flow)은 유스케이스가 실행되면 반드시 발생하는 기본적 처리흐름을 기술하고 선택 흐름 (Alternative flow)은 기본 흐름과는 달리 상황에 따라 달리 실행되는 처리흐름에 대해 기술한다.

유스케이스명	Manage Hybridization
액터명	LabActor (실험자)
개요	이 유스케이스는 Hybridization 실험정보와 실험결과 파일들을 등록하고 관리하는데 필요한 수행과정에 대한 문서
이벤트 흐름	
기본 흐름	<ol style="list-style-type: none"> 1. 실험자가 등록, 수정, 삭제 중 하나의 기능을 선택한다. 2. 선택된 기능에 따라 등록, 수정, 삭제 중 하나의 흐름이 수행된다. <ol style="list-style-type: none"> 2-1. 등록 <ol style="list-style-type: none"> ① Hybridization 실험정보를 입력한다. ② Labeled Extract와 Protocol은 선택항목 중 하나를 선택한다. ③ Hybridization 데이터 파일이 위치한 경로를 지정한다. ④ 저장 버튼을 누르면 Hybridization 이름, Array 일련번호, 파일이름의 입력과 중복여부를 검사하고 파일형식을 확인한다. Quantitation 파일의 수치데이터는 데이터베이스에 이전되어 저장된다. 2-2. 수정 (생략) 2-3. 삭제 (생략)
선택 흐름	<ol style="list-style-type: none"> 2-1 <ol style="list-style-type: none"> ② 원하는 Labeled Extract와 Protocol이 선택항목에 없으면 새로 추가해서 지정한다. ④ Hybridization 이름, Array 일련번호, 파일경로가 입력되지 않았거나 중복되었다면 에러 메시지를 내고 입력화면으로 돌아간다. 파일의 형식이 맞지 않으면 해당 메시지를 내고 입력화면으로 돌아간다.
사전 조건	Hybridization을 등록할 Experiment를 먼저 등록하여야 한다.
사후 조건	

림 4]는 'Manage Hybridization' 유스케이스의 hybridization 등록과정에서 시스템 내부의 동적인 처리 흐름을 보여주는 순차 다이어그램이다.

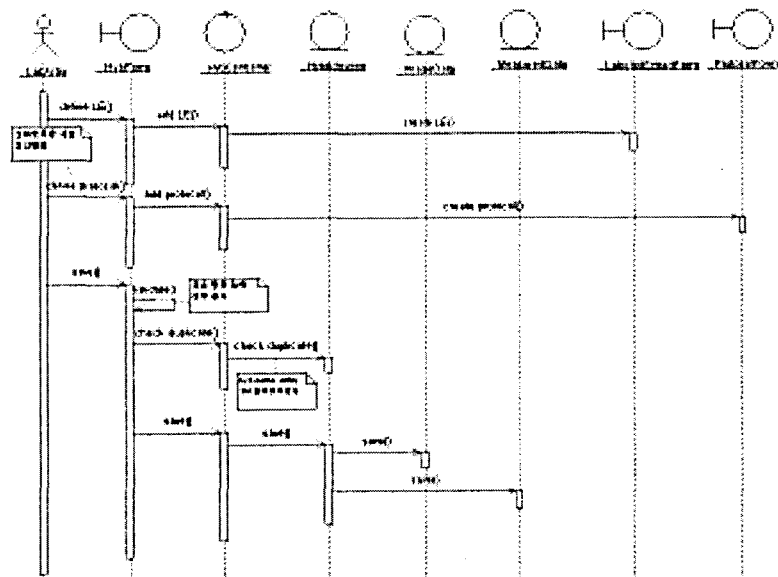
3) 마이크로어레이 정보시스템의 정적 분석

유스케이스 명세서를 참고하여 식별된 객체를 바탕으로 클래스를 정의하고 클래스간의 관계를 표현하는 과정이다. 객체 모델링 결



[그림 3] 'Manage Hybridization' 유스케이스에서 등록과정을 그린 활동 다이어그램

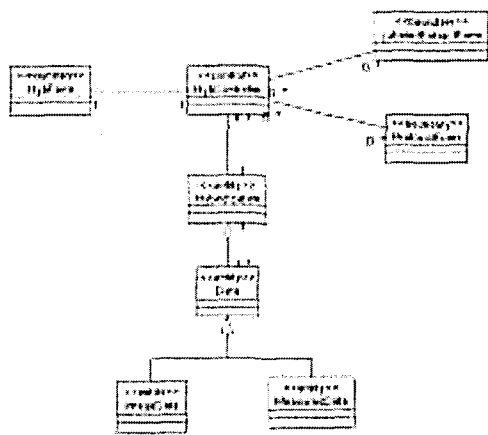
Hybridization 등록하는 과정을 처리단계, 처리과정, 결정위치, 분기처리 등을 사용하여 시각적으로 보여준다.



[그림 4] 'Manage Hybridization' 유스케이스에서 등록과정에 대한 순차 다이어그램

Hybridization 등록요청을 해결하기 위해 필요한 객체들과 그 객체들이 참여한 시간적, 순서적 처리흐름을 표현하는 다이어그램이다.

과인 클래스 다이어그램은 시스템의 정적인 구조를 나타내게 된다. 마이크로어레이 정보시스템 같이 규모가 큰 시스템에서 전체에 나타나는 모든 클래스를 가지고 각각의 관계성을 나타내면 시스템을 이해하는데 더욱 복잡함을 더하게 된다. 따라서 각각의 유스케이스에만 해당되는 클래스들만을 가지고 클래스 다이어그램을 작성하였다. 세부적으로 클래스를 식별하는 과정이 각각의 유스케이스마다 이루어지므로 클래스 다이어그램들을 통합하여 전체적으로 모순이 없는지 검증하였다. 또한 작성된 클래스 다이어그램을 MAGE-OM의 클래스 다이어그램과 비교하여 일관성을 가지는지 확인하였다. [그림 5]는 'Manage Hybridization' 유스케이스에서 식별된 클래스들 간의 관계를 나타내는 클래스 다이어그램이다.



[그림 5] 'Manage Hybridization' 부분에 대한 클래스 다이어그램
클래스는 클래스 이름, 속성, 오퍼레이션으로 구성되거나 간결하게 표현하기 위해 클래스의 속성과 오퍼레이션들은 생략하였다.

5. 결론 및 고찰

본 연구는 마이크로어레이 정보시스템의 개발과정을 풍부한 모델링을 지원하는 객체지

향 기법으로 분석하고 그에 따른 구체적인 단계와 활동을 다양한 UML 모델로 정의하였다. UML 기반의 객체지향 분석방법을 마이크로어레이 정보시스템 개발과정에 적용하여 다음과 같은 유용한 효과를 얻을 수 있었다. 마이크로어레이 정보시스템은 데이터의 구조가 복잡하고 연구자들의 다양한 요구사항을 반영하여야 하므로 시스템의 기능과 범위를 명확히 하는 것이 중요하다. 객체지향 방식으로 시스템을 분석하고 그래픽 요소가 강한 표준 언어인 UML로 그 결과를 표현하면 사용자와 개발자간의 의사소통을 원활히 하고 요구사항을 정확히 파악하는데 도움을 준다. 또한 요구사항을 문서화해서 관리할 수 있으므로 요구에 맞게 시스템이 개발되고 있는지에 대해 반복적으로 검증할 수 있고 생략되거나 불일치되는 모델링 구조에 대한 지적도 용이하게 한다. 또한 객체지향 모델인 MAGE-OM과도 일관된 체계를 형성하기 때문에 표준 데이터 모델을 기반으로 구축할 수 있다.

유전자 마이크로어레이 기술은 생물 제분분야의 연구에 사용될 뿐 아니라 의학 분야까지 널리 활용될 수 있으므로 차후 마이크로어레이 정보시스템은 약학의료 분야 시스템과의 통합되어야 한다 [7]. 이질적인 시스템을 통합하기 위해서는 각 시스템의 역할을 명확히 규정할 수 있는 객체지향 방식의 방법론이 필수적이다. 그러나 객체지향은 다른 방식에 비해 아직 활성화 되지 오래되지 않았기 때문에 생물의료 시스템 개발에 활용된 사례가 많지 않고 전문적인 객체지향 사고와 기법을 갖춘 전문가도 부족한 상황이다. 따라서 본 연구에서는 마이크로어레이 정보시스템 개발에 UML 기반의 객체지향 방식을 적용하여 정형화된 절차와 결과들을 제공하였으며 이는 마이크로어레이 정보시스템을 다른 생물의료 시스템과 통합하는데 활용할 수 있을 것이다.

6. 참고 문헌

Pharmacogenomics Journal, 1(3), 2001,
167-170

- [1] O. Ermolaeva, M. Rastogi, K. D. Pruitt, G. D. Schuler, M. L. Bittner, Y. Chen, R. Simon, P. Meltzer, J. M. Trent, M. S. Boguski, .Data management and analysis for gene expression arrays, *Nature Genetics*. 20(1), 1998, 19-23.
- [2] V. Aggarwal, The application of the unified modeling language in object-oriented analysis of healthcare information systems, *Journal of Medical Systems*, 26(5), 2002, 383-397
- [3] 김주한, 유전체학과 정보의학, 대한의료정보학회지, 9(2), 2003, 79-89
- [4] P. T. Spellman, M. Miller, J. Stewart, C. Troup, U. Sarkans, S. Chervitz, D. Bernhart, G. Sherlock, C. Ball, M. Lepage, M. Swiatek, W. L. Marks, J. Goncalves, S. Markel, D. Iordan, M. Shojatalab, A. Pizarro, J. White, R. Hubley, E. Deutsch, M. Senger, B. J. Aronow, A. Robinson, D. Bassett, C. J. Stoeckert Jr, A. Brazma, Design and implementation of microarray gene expression markup language (MAGE-ML), *Genome Biology*, 3(9), 2002, research0046.1-0046.9
- [5] 삼성 SDS 기술사회 핵심 정보기술총서 2, 데이터베이스/소프트웨어 공학, 한울아카데미
- [6] K. Saleh, Documenting electronic commerce systems and software using the unified modeling language, *Information and Software Technology*, 44, 2002, 303-311
- [7] T. E. Klein, J. T. Chang, M. K. Cho, K. L. Easton, R. Ferguson, M. Hewett, Z. Lin, Y. Liu, S. Liu, D. E. Oliver, D. L. Rubin, F. Shafa, J. M. Stuart, R. B. Altman, Integrating genotype and phenotype information: an overview of the PharmGKB project. *Pharmacogenetics Research Network and Knowledge Base*,