SMS : SBML 문서관리기

임정곤* 김태경** 정태성** 조완섭*
*충북대학교 경영정보학과, **충북대학교 정보산업공학과
{tinnom, misoh049}@hanmail.net, mispro97@naver.com, wscho@cbnu.ac.kr

SMS : An SBML Document Manager

Jioungkon Lim*, Taekyong Kim**, Taesung Jeong**, Wansup Cho*
*Dept. of Management Information System,
**Dept. of Information Industrial Engineering
Chungbuk National University

요 약

최근 이슈가 되고 있는 시스템 생물학(Systems Biology)은 생물학적인 이론과 컴퓨터의 계산적인 모델링 그리고 실험의 상호 의존적인 통합으로써 특징 지워진다. 그 중 컴퓨터의 계산적인 모델링에 대한 연구가 무엇보다 중요한 비중을 차지하고 있다. 하지만 계산적인 모델링에서 여러 자원을 통합하기 위한 공통의 기반 구조나 표준에 대한 연구는 미흡한 실정이다. 이러한 문제점을 해결하기 위해 XML 기반의 형식을 갖춘 SBML(Systems Biology Markup Language)이 시스템 생물학의 표준으로 개발되어 연구 중에 있다. 현재 개발 중인 시뮬레이션과 데이터 분석을 위한 다양한 응용 어플리케이션이 이미 SBML 문서를 지원하고 있다. 본 연구에서는 시스템 생물학 분야에서 SBML 표준에 대한 중요성을 인식하여, 객체지향 바이오 데이터베이스로부터 질의의 결과를 SBML 문서로 변환하고, 반대로 SBML 문서를 객체지향 데이터베이스에 저장하는 변환기를 제안하고자 한다.

1. 서 론

최근 생물학 실험에 대한 데이터의 양산과 함께 정보 기술을 적용하여 시뮬레이션이나 데이터의 분석을 수행하는 다양한 어플리케이션이 개발되고 있다. 하지만 이러한 어플리케이션에 맞는 표준적인 데이터 형식이 부재하여 데이터의 공유 및 교환에 있어 어려움을 겪고 있었다. 이러한 문제점을 해결하기 위해 시스템 생물학 데이터의 표준 형식인 SBML(Systems Biology Markup Language)[1]이 데이터 표준 형식으로 개발 되었고, 현재 계속 연구 중에 있다. SBML은 응용 어플리케이션 사이에서 데이터의 교환 및 시각화를 위한 표준으로 사용 된다. 본 연구의 목적은 객체 DBMS를 이용하여 SBML 문서를 저장하고, 데이터베이스에 대한 질의 결과를 SBML 문서 형태로 사용자에게 제공하여 객체 DBMS와 SBML의 통합된 환경을 제공하는데 있다. 기존의 관계형 데이터베이스보다 객체지향 데이터베이스를 이용하는 것은 XML 스키마[2] 기반인 SBML 스키마가 객체 모델을 지향 한다는 점에서 양자간에 유사점이기 때문이다. 때문에 객체지향 데이터베이스를 이용함으로써 보다 쉽고 빠른 매핑 방법으로 SBML 문서를 객체지향 데이터베이스로 변환할 수 있는 장점이 있다. 또한 반대로 객체 질의의 결과를 SBML 문서 형태로 변환하여 데이터 교환과 시뮬레이션 관련 도구의 활용에서 용이하도록 한다.

본 논문의 구성을 보면 2장에서는 관련연구로 SBML과 생물학 데이터베이스에 대해 살펴본다. 3장에서는 객체지향 데이터베이스와 SBML의 스키마의 매핑을 다룬다. 4장에서는 시스템의 전체적인 구성에 대한 내용을 다루고, 5장으로 결론을 내린다.

2. 관련연구

2.1 SBML(Systems Biology Markup Language)

SBML은 생물학 분야에 대한 수많은 연구를 기술하고 있으며 생화학적 반응에 대한 시스템을 네트워크로 묘사

본 연구는 한국과학재단 특정기초 연구사업(R01-2003-000-11723-0)으로부터 지원을 받았음

하고 있는 언어이다. 또한 SBML은 cell signaling pathways, metabolic pathways, biochemical reactions, gene regulation등 기타 여러 분야를 포함하여 기술하고 있다. SBML의 기본적인 목적은 분산되어 있는 수많은 데이터에 대해 표준을 정하고 있으며, 데이터의 교환과 상호 운용적인 사용을 위해 개발된 언어이다. 현재 SBML을 활용하기 위한 어플리케이션이 많이 개발되고 있다. 그중 SBW(Systems Biology Workbench)[4]라는 프로젝트에서 시스템 생물학에 대한 대사경로와 반응모델을 구축하는 시뮬레이션이나 분석을 위한 시각화 어플리케이션이 개발 되고 있다. 이처럼 SBML이 활용되고 있는 분야가 확산되고 있다.

2.2 생물학 데이터베이스

시스템 생물학(Systems Biology)[4] 분야에서 생화학 실험에 대한 데이터를 저장하고 관리하는 대표적인 데이터베이스는 KEGG[5], EcoCyc, Enzyme등을 들 수 있다. 이러한 데이터베이스는 실험 반응에 대한 대사경로에 관련된 데이터를 관리하기 위해 사용된다. 이처럼 쓰이는 용도는 비슷하나 데이터의 형식이나 표준은 제각기 다르므로 각 데이터베이스마다 그에 대응하는 어플리케이션을 새롭게 만들어야 하는 노력과 데이터의 교환에 문제점을 가지고 있다. 최근 JST ERATO Kitano symbiotic Systems Project에서는 KEGG(Kyoto Encyclopedia of Genes and Genomes) 데이터베이스에 저장된 정보를 SBML 문서로 변환[6]하고, 변환된 SBML 문서를 활용하는 연구가 진행되고 있다. KEGG는 관계형 데이터베이스를 이용하고 있다. 관계형 데이터베이스는 테이블의 참조 관계를 키로 연결하므로 테이블 수가 많아지고 복잡하다. 또한 복잡한 질의시 조인 비용이 높아진다. SBML을 활용하는 측면에서 관계형 데이터베이스의 문제점을 보완하기 위해 객체지향 데이터베이스에 관한 연구가 선행되어야 한다.

3. 객체 데이터베이스와 SBML 스키마의 매핑

본 연구에서 제안하는 전체적인 시스템 구성 중 가장 중요한 비중을 차지하는 부분은 데이터베이스 스키마와 SBML 스키마를 매핑하는 방법이다. 기존에 연구에서는 관계형 데이터베이스를 이용하여 SBML을 변환하고 있다. 하지만 관련연구에서 보았듯 몇 가지 문제점이 발생한다. 본 연구에서 제안하는 SBML 변환기는 객체지향 데이터베스와 XML 스키마 기반인 SBML 스키마를 매핑한다. 객체지향 데이터베이스의 특징은 계층 구조와 상속관계 그리고 집합값을 쉽게 표현해 줄 수 있다. 게다가 객체를 참조하는 구조로 되어 있어 질의문이 간단해 질수 있는 장점을 가진다. 객체지향 데이터베이스의 스키마처럼 SBML 스키마도 XML의 객체 지향적인 데이터 모델의 특성을 그대로 유지하게 되므로 SBML 스키마와 객체지향 데이터베이스 스키마와의 매핑이 쉬워진다. 매핑이 쉬워지므로 변환 과정이 보다 쉽고 간결해지는 장점을 얻을 수 있다. 표 1에서는 SBML 표준 스키마의 UML 표기법의 예를 보여주고 있다. 반면 표 2에서는 실제 구현에 쓰인 객체지향 데이터베이스 스키마의 예를 보여주고 있다. 표와 같이 두 스키마가 전달하는 의미가 매우 유사함으로써 서로 매핑이 쉬운 장점을 가질 수 있다.

[표 1] SBML 스키마의 UML 표기

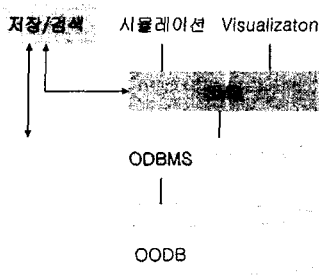
Model
id : Oid {use="optional"} name : string {use="optional"} functionDefinition : FunctionDefinition [0...*] unitDefinition : UnitDefinition [0...*] compartment : Compartment [0...*] species : Species [0...*] parameter : Parameter [0...*] rule : Rule [0...*] reaction : Reaction [0...*] event : Event [0...*]

[표 2] OODB 스키마

Model
id : Oid name : string functionDefinition : FunctionDefinition {SET} unitDefinition : UnitDefinition {SET} compartment : Compartment {SET} species : Species {SET} parameter : Parameter {SET} rule : Rule {SET} reaction : Reaction {SET} event : Event {SET}

4. 시스템 구성

본 연구에서 제안하는 SBML 변환기의 전체적인 시스템 구성을 살펴보면 그림 1과 같다.



[그림 1] SMS 구성도

외부의 SBML 문서를 객체지향 데이터베이스에 저장할 수 있으며, 객체지향 데이터베이스에서 사용자가 검색한 결과를 SBML 문서로 변환하여 시뮬레이션 하거나 시각화에 응용하는 시스템 구조이다.

4. 1 객체지향 데이터베이스 와 SBML 연동

본 시스템은 데이터베이스에서 SBML 문서로 변환 과정과 반대로 SBML 문서를 데이터베이스에 입력하는 두 과정으로 나눌 수 있다. 첫째 외부의 수많은 어플리케이션에서 작성된 SBML 문서나 기타 방법으로 새로 작성된 SBML 문서를 검증하여 SBML 변환기를 거쳐 객체지향 데이터베이스에 저장하는 시스템 구조를 가지고 있다. 둘째 객체지향 데이터베이스에서 사용자의 질의에 대한 결과가 SBML 변환기를 거치게 되면 SBML 문서가 자동적으로 생성된다. 3장에서와 같이 객체지향 데이터베이스의 스키마와 SBML 스키마가 쉽고 간결하게 매핑 되므로 변환 알고리즘이 간단하며, 시스템 성능을 향상 시킬 수 있다. 이와 같이 전체적인 시스템 구성은 객체지향 데이터베이스와 SBML이 양방향으로 처리 되는 구조이다. 시스템 구축을 위한 데이터베이스는 UniSQL을 사용 하였으며 데이터베이스의 연동을 위해서 UniSQL에서 제공하는 JDBC 드라이버를 이용하였다.

4. 2 SBML 문서 검증 및 평가

4. 1에서와 같이 일차적으로 작성된 SBML 문서를 어떻게 신뢰 할 것인가에 대한 문제를 제기할 수 있을 것이다. 신뢰성을 평가하기 위해 SBML 스키마에서 정의하는 표준과 변환기에 의해 작성된 SBML 문서를 비교하고 검증하여 최종적으로 SBML 문서를 완성하게 된다. 이러한 과정을 거치는 이유는 다른 응용 어플리케이션에서 작동될 때 발생할 수 있는 에러를 사전에 방지하기 위함이다.

4.3 응용

최종 완성된 SBML 문서는 여러 응용 분야에서 쓰일 수 있게 된다. 예를 들어 SBML 문서 내용에 대한 시뮬레이션이나 분석을 위한 어플리케이션의 시각화에 적용시킬 수 있다. SBML 문서를 이용하여 시뮬레이션이나 시각화를 수행하는 다양한 도구들이 제안되고 있으며 여기서는 이들을 생략한다. 본 연구에서는 완성된 SBML 문서를 시각화 어플리케이션인 JDesinger[7]에 테스트 하고 있다.

5. 결론

최근 IT 분야에서 등장한 XML은 데이터의 교환 및 이질적인 환경을 통합하여 데이터를 상호 운용적으로 사용하도록 하는 공헌을 하고 있다. 현재 부각되고 있는 시스템 생물학(Systems Biology) 분야에서도 XML 기반인 SBML의 표준이 개발되어 분산된 데이터를 상호 운용적으로 사용할 수 있게 되었다. 결론적으로 데이터베이스의 정보들을 SBML로 변환하여 응용 분야에 활용하기 위해서, 기존의 관계형 데이터베이스를 이용하여 SBML로 변환 하는 것보다 객체지향 데이터베이스를 사용하여 변환하는 것이 쉽게 처리 될 수 있다. 향후 연구 과제로 전체적인 시스템 설계와 구현을 할 것이다.

6. 참고문헌

[1] Andrew Finney, Micael Hucka Systems Biology Markup Language(SBML) Level 2 : Structrues and Facilities for Model Definitions 2003. 6

[2] David C. Fallside(IBM) XML Schema Part 0 : Primer 2001. 5.2

[3] Michael Hucka, Andrew Finney, Herbert Sauro, Hamid Bolouri, John Doyle, Hiroaki Kitano The ERATO Systems Biology Workbench : Architectural Evolution 2001. 11 The Proceedings of the Second International Conference on Systems Biology(ICSB)

[4] Stefan Hohmann, Jens Nielsen, Hiroaki Kitano Yeast Systems Biology - Concepts 2004. 1

[5] Minoru Kanehisa, Susumu Goto, Shuichi Kawashima and Akihro Nakay The Kegg database at GenomeNet 2001. 9.26

[6] Akira Funahashi, Hiroaki Kitano Converting KEGG DB to SBML

[7] Hervert M sauro A introduction to JDesigner