

시맨틱 웹과 바이오 온톨로지

경기대학교 김인철*

1. 서론

웹에 관한 비전 중에 하나는 현재 기계 가독형(machine readable), 그리고 인간 이해형(human understandable)의 형태에서 기계 가독과 이해형(machine readable and understandable)의 형태로 변환되는 것이다[9]. 이것은 사용자의 중간 개입 없이도 응용프로그램들과 에이전트들이 웹 상의 정보를 이용할 수 있다는 것을 의미한다. 웹 자원의 내용을 기술하기 위해 온톨로지(ontology)들을 이용하는 것이 시맨틱 웹(semantic web)의 핵심이다. 일반적으로 하나의 온톨로지는 하나의 영역에 대한 한 집단의 지식을 구체적인 형태로 개념화한 것이다. 현재 두 가지 언어인 OIL[4]과 DAML(DARPA Agent Markup Language)은 DAML+OIL[7](<http://daml.org>)이라는 하나의 언어로 결합되었고, 이것은 웹 상의 온톨로지들을 인코딩하고, 관리하고, 전달하는 표준 메커니즘으로 제안되어 있다. W3C가 시맨틱 웹 생성을 위해 지식이 핵심이라는 사실을 인식하는 것과 마찬가지로 시맨틱 생물정보 자원의 핵심 또한 지식이다.

대부분의 생물학 연구는 미지의 개체에 대해 이미 알고 있는 사전 지식을 적용함으로써 진행될 수 있다. 더욱이 생물정보학 데이터베이스에 저장된 복잡한 생물학 데이터들은 종종 그 데이터베이스에 허용 가능한 데이터 값들에 대한 제약을 명시적으로 표현하는 지식을 요구한다. 생물정보학 프로그램들과 데이터베이스들에 그러한 지식을 제공하는 한 가지 방법은 바로 바이오 온톨로지(bio-ontology)를 사용하는 것이다[13]. 이러한 바이오 온톨로지의 사용은 다양한 생물정보 데이터에 대한 자동화된 기계 처리 가능성을 높이게 될 것이다. 따라서 향후 지식이 풍부한

생물정보 자원들이 다양한 응용프로그램들 안에서 상호 운용이 가능하게 됨으로써 생물정보 자원들의 유용성을 높이게 될 것으로 예상된다.

본 논문에서는 이러한 관점에서 대표적인 바이오 온톨로지와 관련 시스템들을 소개하고 이들의 특징을 서로 비교하려고 한다. 현재 생물정보학(bioinformatics) 분야를 중심으로 한 바이오 온톨로지에 관한 연구는 바이오 온톨로지의 용도에 따라 크게 두 가지 카테고리로 나누어 볼 수 있다. 첫째는 온톨로지를 하나의 데이터 저장소 혹은 서로 다른 데이터 저장소들 간에 공유할 수 있는 통제어휘(controlled vocabulary)로 이용하는 것이다. 일반적으로 온톨로지는 데이터 항목을 나타내는 용어들의 의미를 정의함으로써 각 용어들이 하나의 공통된 의미를 가지고 일관되게 쓰일 수 있도록 한다. 또 온톨로지 구조는 질의 조정(query refinement)에도 도움이 된다. 즉 검색시스템 이용자는 온톨로지 구조 안에서 위 아래로 이동함으로써 검색 질의에 쓰인 질의 용어(query term)들을 더욱 일반화 하거나 세분화 할 수 있다. 뿐만 아니라 그러한 통제어휘들은 검색 재현율(recall)과 정확도(precision)를 높일 수 있다. GO(Gene Ontology)와 MBO(Molecular Biology Ontology)는 이러한 연구의 대표적인 예들이다. 둘째는 생물정보학 응용시스템들에 지식 서비스를 제공하기 위해 온톨로지들을 사용하는 것이다. TAMBIS, RiboWeb, EcoCyc과 같은 시스템들에서는 데이터베이스 스키마를 제공하기 위해 온톨로지를 이용하고 있다. 그들은 표현력에서 기존의 스키마 언어들보다 뛰어나며, 따라서 데이터 인스턴스들을 매우 충실하게 묘사할 수 있다. 특히 RiboWeb시스템에서 온톨로지는 리보솜(ribosome)의 구조를 분석하기 위한 지식을 제공하며, 이용자에게 어떤 분석방법을 적용해야 할지를 알려주고, 현재 보유하고 있는 지식과의 모순점을 찾아주는 역할을 한다.

* 중신회원

표 1 바이오 온톨로지들의 특징 요약

Ontology	Application Scenario	Modularised?	Domain-oriented Component	Task-oriented Component	Generic Component	Instances	Detail level	KR
RiboWeb	database schema	✓	Ribosome Components, covalently bonded molecules, biological macromolecules, regions of molecules	Experimental detail techniques for analysing data, publication	✓	✓	high	frames
EcoCyc	database schema	✓	E. coli genes, metabolism, regulation, signal transduction and metabolic pathways	Visualisation of biochemical reactions and layout of genes with chromosome	✓	✓	high	frames
MBO	community reference	✓	Shallow	Shallow	✓	X	low	X
GO	controlled vocabulary for database annotation	partially	Drosophila, mouse and yeast gene function, gene product function, process and cellular location and structure.	X	X	✓	high	X
TaO	common access ontology-based search	partially	Proteins, enzymes, motifs, secondary and tertiary structure, functions and processes, subcellular structure and chemicals, including cofactors. The larger model includes nucleic acid and genes	Bioinformatics search and analysis tasks	✓	X	high	DLs

표 1은 현재 이미 서비스하고 있거나 개발 중인 대표적인 바이오 온톨로지들인 RiboWeb, EcoCyc, MBO, GO, TaO 등의 특징을 표로 요약하여 보여주고 있다. 본 논문에서는 이들 바이오 온톨로지 시스템 각각의 특징에 대해 간략히 살펴보고자 한다.

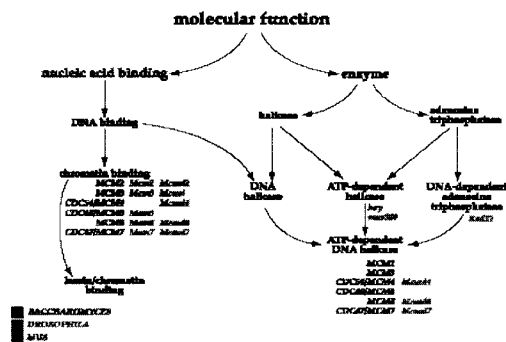


그림 1 Gene Ontology의 예

2. GO

GO(Gene Ontology)는 뒤에서 소개되는 MBO와 마찬가지로 생물정보 데이터베이스의 주석(annotation)이 주된 목적이다[14]. 하지만 GO는 외부 연구 기관이나 표준화 기관에서 제안된 것이 아니라 데이터베이스 그룹 안에서 시작하여 형성되었다. 또 GO는 데이터베이스 그룹들에 포함될 수 있는 모든 분자 생물 지식을 표현하려고 하지 않고, 하나의 생물체를 중심으로 그 안에서 유전자들의 역할에 대한 정보만

을 표현하려고 하기 때문에, 범위가 광범위하지 않고 비교적 좁다. 원래 GO 컨소시엄은 FlyBase, MGI(Mouse Genome Informatics), SGD(Saccharomyces Genome Database) 등 대표적인 세 가지 생물체 데이터베이스에 대한 연합 프로젝트이다. 이 컨소시엄의 목표는 유전자들이 각 생물체에서 수행하는 역할을 기술하기 위한 구조화된 공통의 통제어(controlled vocabulary)를 생성하는 것이다.

다른 바이오 온톨로지들에 비해 GO는 상위레벨 온톨로지가 부족하다. GO는 생물학적 프로세스 온톨로지(biological process ontology), 분자 기능 온톨로지(molecular function ontology), 세포 구성원 온톨로지(cell component ontology) 등 서로 다른 3가지 온톨로지들로 구성된다. 생물학적 프로세스 온톨로지는 DNA 물질대사(metabolism)를, 분자 기능 온톨로지

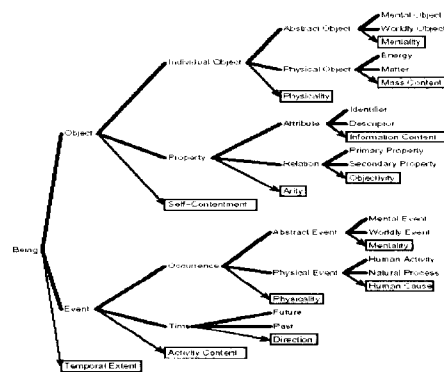


그림 2 Molecular Biology Ontology의 예

는 유전자 산물의 기능(function)을, 세포 구성원 온톨로지는 세포의 위치와 구조를 각각 나타낸다. 그림 1은 GO의 한 예로서, 분자 기능 온톨로지 부분을 보여주고 있다. GO는 “is a subclass of” 관계와 “is part of” 관계를 이용하여 유전자 산물의 역할을 표현하고 있다. 현재 GO는 약 5000여 개가 넘는 개념들을 온톨로지에 포함하고 있다.

3. MBO

MBO(Molecular Biology Ontology)는 분자생물 데이터베이스들 간의 상호 통신과 운용을 위한 시도이다[12]. 이질적인 데이터베이스들이 공통적으로 MBO를 사용함으로써 Gene과 같은 개념을 사용할 때 발생할 수 있는 의미론적 혼란을 피하고자 하였다. Schulze-Kremer는 하나의 공통 온톨로지를 서로 공유함으로써 서로 다른 데이터베이스들로부터 추출된 데이터 항목들 간의 관계에 대한 의미적 불확실성과 오해를 제거할 수 있다고 보았다. MBO는 생물학적 객체, 실험적 절차, 분자생물학의 계산적 측면을 표현하기 위한 개념(concept)과 관계(relation)들을 포함하고 있다. MBO 온톨로지는 매우 광범위하며, 개념과 인스턴스들을 표현하는 약 1200 개가 넘는 노드들을 가지고 있다.

MBO의 개념 부분에 사용되는 주된 관계는 “is a subclass of” 관계이다. 그림 2는 MBO의 상위 레벨의 한 예를 보여주고 있다. 그림에서 각 링크들은 “is a subclass of” 관계를 나타내며, 인스턴스들(instances)은 표시되지 않는다. 클래스 구분의 기준은 화살표와 박스로 표시 되었으며, 굵은 선들은 하위 서브클래스들로의 분화를 나타낸다. 뿌리노드인 Being 클래스는 시간적 범위(temporal extent)에 따라 Object 클래스와 Event 클래스로 나뉜다. 그리고 다시 Object 클래스는 자기-만족(self-contentment) 여부에 따라 Individual Object와 Property 서브클래스들로 나뉜다. 즉, Individual Object는 독자적으로 존재할 수 있

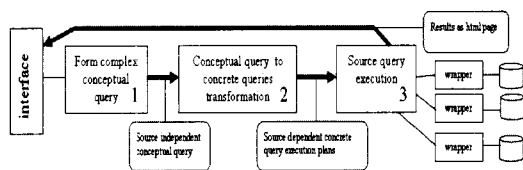


그림 3 TAMBIS 시스템의 질의 처리과정

으나 Property는 언제나 참조할 다른 Object나 Event를 필요로 한다는 점에서 서로 구분된다. 현재 MBO는 분자생물학 데이터베이스 항목들을 중심으로 약 1300여 개의 개념들을 포함하고 있으며, 계속 확장 보완 중이다.

4. TAMBIS

TAMBIS(Transparent Access to Multiple Bioinformatics Information Sources) 시스템에서는 생물학자들로 하여금 하나의 공통 질의 인터페이스를 사용하여 다수의 외부 데이터베이스들에 질의를 할 수 있도록 하고 있으며, 이를 위해 온톨로지를 이용하고 있다[5]. TAMBIS 시스템의 온톨로지인 TaO는 광범위한 생물정보학 작업과 자원들을 표현하고 있으며, TAMBIS 시스템 안에서 가장 중추적인 역할을 수행한다.

TaO는 다른 바이오 온톨로지들과는 달리 인스턴스(instance)들을 포함하지 않는다. TaO는 단지 생물정보학에 관한 지식과 분자생물학 개념들, 그리고 그들의 관계들만을 포함하고 있으며, 그들이 표현하는 인스턴스들은 외부 데이터베이스들에 놓여 있다. 개념들이 인스턴스들을 나타내기 때문에, 하나의 개념은 하나의 질의로 쓰일 수 있다. 예컨대 Receptor Protein이라는 개념은 수용체(receptor) 기능을 가진 단백질 인스턴스들을 나타내며, 이러한 인스턴스들을 수집하는 일은 곧 그러한 질의에 대한 답을 찾는 일이다.

TaO는 개발자에 의한 새로운 지식의 개념화나 인코딩 작업 없이도 스스로 성장할 수 있다는 의미에서 하나의 동적 온톨로지이다. 이와는 반대로 대부분의 다른 바이오 온톨로지들은 새로운 개념들을 생성하기 위해서는 개발자가 직접 간여하여 인코딩 작업을 수행해야 한다는 점에서 정적인 온톨로지들이다. TaO에서는 새로운 개념을 생성하기 위해 규칙(rule)들을 이용하는데, 이것들은 어떤 개념들이 관계를 통해 다른 개념들과 결합될 수 있는지를 나타낸다. 따라서 TaO에서는 관계를 매우 중요시 한다. TAMBIS 시스템 사용자는 관계를 이용하여 다음과 같은 방식으로 여러 자원들에 관련된 복잡한 질의를 생성할 수 있다. 예컨대, 사용자는 Protein 개념에서 시작해서, Protein 개념을 다른 개념들과 결합하기 위해 어떤 관계를 사용할 수 있을지 TaO를 참조할 수 있다. 많은

관계들 가운데 “is homologous to Protein”과 “has AccessionNumber AccessionNumber”가 제시되면, 원래 Protein 개념은 새로운 개념인 Protein isHomologous to Protein으로 확장되고, 두 번째 Protein은 hasAccessionNumber AccessionNumber로 확장된다. 그리고 최종적인 결과 개념인 Protein homologue of Protein with Accession Number는 특정 접근번호를 가진 단백질 상동성을 갖는 단백질들을 나타낸다. 이 개념은 그러한 질의에 어떻게 답을 해야 할지에 대한 정보를 가지고 있지 않은 자원-독립적인(source-independent) 질의로 쓰일 수 있다.

TAMBIS 시스템의 나머지 부분들은 이러한 개념적 질의를 받아들인 뒤, 외부 자원들을 대상으로 하나의 실행 프로그램을 수행하여 이 질의를 처리한다.

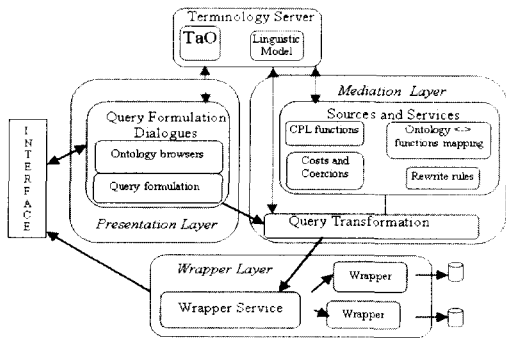


그림 4 TAMBIS 시스템의 구조

그림 3은 TAMBIS 시스템의 일반적인 3단계 질의 처리과정을 나타내고 있다. 개념적 질의(conceptual query) 단계에서는 사용자가 온톨로지를 브라우징하면서 생물학적으로 일관성 있으면서도 자원-독립적인 복잡한 하나의 개념적 질의를 구성하는 단계이다. 구체화된 질의(concrete query) 단계에서는 TAMBIS가 적절한 자원들을 결정하고, 실행 계획을 세워 선언적인 개념적 질의를 하나의 순서화된 자원 위주의 호출들로 변환한다. 마지막으로 자원 질의(source query) 단계에서는 실행 계획들이 적절하게 포장된 정보 서비스들에 주어져 실행된다. 그리고 그 실행 결과들은 html형태로 사용자의 웹 브라우저에 반환된다. 그림 4는 3계층 구조(three layer architecture)를 형성하고 있는 TAMBIS 시스템의 구성요소들과 시스템의 전체구조를 나타내고 있다. 최상위 계층인 제시 계층(presentation layer)은 사용자가 지식베

스로부터 얻은 용어들을 결합하여 자원-독립적이고 선언적인 질의를 생성할 수 있도록 지능적인 사용자 인터페이스를 제공한다. 중재 계층(mediation layer)은 질의를 만족시킬 수 있는 적절한 자원들을 알아낸 다음, 질의를 자원-독립적인 절차들의 순서 집합으로 변환한다. 최하위 계층인 포장 계층(wrapper layer)은 외부 자원들을 관리하는 역할을 하며, 이질적인 각 외부 자원들을 포장하여 통신과 형식, 그리고 네트워크 면에서 투명성을 줄 수 있는 공통의 인터페이스를 제공한다.

현재 TaO는 단백질들에만 초점을 맞춘 작은 모델과 핵산들을 포함하는 대규모 모델, 이렇게 두 가지 형태로 제공되고 있다. 250여 개의 개념들과 60여 개의 관계들을 포함하고 있는 작은 모델은 단백질들과 효소들뿐만 아니라 그들의 모티프, 이차 및 삼차 구조, 기능과 프로세스들을 담고 있다. 약 1500여 개의 개념들을 포함하고 있는 대규모 모델은 이러한 영역을 더욱 넓혀 핵산과 유전자들에 관한 개념들까지 담고 있다.

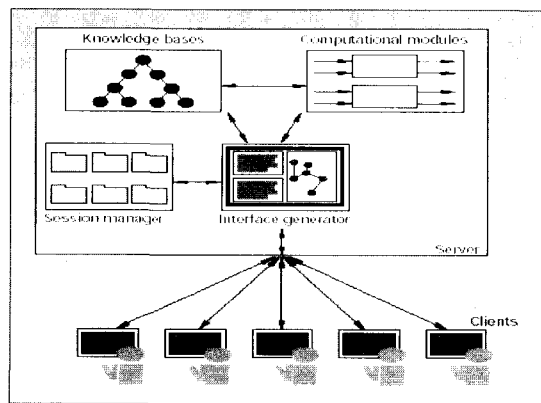


그림 5 RiboWeb시스템의 구조

5. RiboWeb

RiboWeb[1]은 원핵생물의 리보솜(prokaryotic ribosome) 구조 데이터를 해석하기 위한 하나의 통합 데이터 자원이다. RiboWeb은 이미 발표된 모델들과 그 모델들의 기초가 되는 데이터 자원들을 긴밀히 연결함으로써, 네트워크 기반의 시스템에서 사용자가 특정 데이터에 대한 해석을 변경하거나 이 해석들로부터 새로운 모델들을 생성할 수 있도록 지원한다. 이 시스템은 데이터에 관한 표준화된 개념(concept) 집

합, 속성(attribute) 집합, 관계(relation) 집합, 그리고 기본적인 추론(default reasoning) 등을 포함하고 있다. 그리고 이러한 의미론을 명시하기 위한 하나의 언어로서 온톨로지들을 제공한다.

RiboWeb은 전세계적으로 분포하고 있는 리보솜 연구자들에게 인터넷을 통해 데이터 자원에 접근할 수 있도록 하고 있다. 그림 5는 RiboWeb시스템의 전체 구조를 보여주고 있다. 그림에서 보듯이 개별 클라이언트들은 서버의 세션 관리자(session manager)와 인터페이스 생성자(interface generator)와 상호작용을 가진다. 그리고 이들은 다음과 같은 순환루프를 수행한다 :

- 사용자에게 관심 있을 계산모듈을 결정한다
- 지식베이스로부터 계산에 필요한 관련 데이터를 검색한다
- 지식베이스로부터 계산모듈에 관한 정보를 찾아낸다
- 모듈을 실행한다
- 결과를 저장한다

계산모듈들은 지식베이스를 브라우징하거나, 발표된 데이터를 선택하거나, 선택된 데이터를 이용하여 새로운 3차원 모델을 계산하거나, 3차원 모델과 데이터를 비교하거나, 발표된 데이터들 사이의 불일치성을 찾는 등의 매우 다양한 작업들을 수행한다. 다수의 클라이언트들은 계산을 수행하거나 지식베이스를 브라우징하기 위해 인터페이스 생성자와 통신할 수 있다. 현재 RiboWeb시스템의 인터페이스 생성자는 Unix 워크스테이션 상에 Perl로 구현되어 있고, 계산모듈들은 C, C++, Perl, 또는 Lisp 등으로 구현되어 있으며, 지식베이스는 Sophia시스템으로 구현되어 있다.

RiboWeb시스템의 온톨로지들은 다음과 같은 특징을 가진다.

- 일반적인 것부터 세부적인 것까지 계층적인 개념(concept)들의 계통도
- 기본값과 제약을 허용하는, 각 개념과 연관된 속성(attribute)들의 리스트
- 온톨로지의 연결을 의미하는 개념들간의 관계(relation)

개념들과 그들간의 관계들은 RiboWeb시스템의 온톨로지를 형성한다. 또 각 개념에 대해 속성값을 갖는 인스턴스(instance)들을 생성할 수 있다. 따라서 이와 같은 온톨로지와 인스턴스들을 합쳐 다시

RiboWeb시스템의 지식베이스를 형성한다. RiboWeb 시스템에서는 Ontolingua[3] 지식모델을 채용하고 있으며, 속성계승은 지원하나 상위클래스-하위클래스 관계(superclass-subclass relationship)를 제외하고는 자동화된 추론이나 제한을 적용하지 않는다.

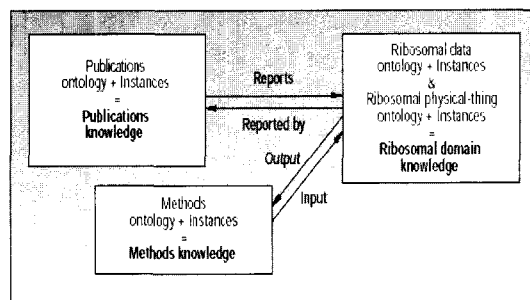


그림 6 RiboWeb시스템에서 온톨로지들간의 관계

RiboWeb시스템은 지식베이스의 스키마로서 온톨로지들을 사용한다. 이 시스템은 사물 온톨로지(physical-thing ontology), 데이터 온톨로지(data ontology), 참고 온톨로지(reference ontology), 수단 온톨로지(methods ontology) 등 4가지 유형의 온톨로지들을 이용한다. 사물 온톨로지는 리보솜의 분자구조와 공통인자들에 대한 명세로서, 리보솜의 구조에 관한 데이터를 나타내는데 결정적인 사물과 관계들을 명시한다. 데이터 온톨로지는 생물학적 실험들로부터 수집된 데이터의 유형들을 명시하며, 참고 온톨로지는 심사 레벨을 포함하여 출간물들의 유형을 명시한다. 끝으로 수단 온톨로지는 RiboWeb이 취할 수 있는 행동들을 명시하고 이러한 행동들을 실행하기 위해 필요한 핵심 속성들을 선언한다. 이러한 온톨로지들 외에 RiboWeb시스템은 사용자에게 관한 기본정보, 패스워드, 권한, 세션, 생성된 결과들과 그러한 결과들을 생성한 계산모듈들 간의 연관성 등을 저장하기 위해 표준 데이터베이스 기술을 이용하고 있다. 그림 6은 RiboWeb시스템에서 온톨로지들 간의 상호작용을 보여주고 있다. 지식베이스의 출간물 부분은 기본적인 서지정보를 보유하고 있으며, 이 지식은 Reports와 Reported-by 링크를 통해 리보솜 영역 지식에 연결된다. 또한 RiboWeb의 분석능력에 관한 정보를 담고 있는 지식베이스의 수단 부분은 영역 지식베이스로부터 받은 정보를 입력으로 삼고 출력으로서 그 결과를 지식베이스에 저장한다.

6. EcoCyc

EcoCyc 데이터베이스는 물질대사 데이터(metabolic data)를 제공하는 온라인 데이터 자원이다[10]. 특히 EcoCyc는 *E. coli*에 대한 완전한 생화학적 네트워크를 표현하고자 하는 목표를 가지고 있다. EcoCyc 역시 RiboWeb과 같이 데이터베이스 스키마를 명세하기 위한 용도로 온톨로지를 이용한다. EcoCyc는 온톨로지를 바탕으로 생물학자들이 마치 하나의 백과사전을 이용하듯 물질대사 데이터가 이용하도록 하고 있다. EcoCyc 온톨로지는 *E. coli* 유전자들과 물질대사(metabolism) 등을 포함하고 있으며, 현재 지식베이스는 4391 개의 *E. coli* 유전자들과 695 개의 효소들, 그리고 904 개의 물질대사 및 상호작용들을 기술하고 있다. 이와 같은 EcoCyc를 이용해 생물학자들은 *E. coli* 염색체 안에서의 유전자들의 배열과 개별 생화학 반응을 시각화해 볼 수 있다.

데이터베이스 스키마를 정의하기 위해 온톨로지를 이용하는 EcoCyc의 장점은 높은 표현력과 생물학적 정보에 필요한 신속한 스키마 변경이다. 하지만 사용자는 온톨로지의 사용을 알지 못한다. 예컨대 EcoCyc에서 Gene의 개념은 Polypeptide product, gene name, synonyms, identifiers와 같은 다양한 속성들을 가진 하나의 클래스로 표현되며, 이들 속성은 Gene 개념을 다른 개념들과 연결시켜 주는 역할을 한다. 현재 EcoCyc는 다음의 두 가지 형태로 제공되고 있다 :

- URL <http://ecocyc.PangeaSystems.com/ecocyc/>을 통해 인터넷 웹으로 접근 가능
- X-windows 버전(네비게이션 소프트웨어와 EcoCyc DB)

7. 결론

본 논문에서는 대표적인 바이오 온톨로지들을 소개하고 그의 특징들을 간략히 살펴보았다. 현존하는 바이오 온톨로지들은 많은 영역지식을 담고 있다. 이 온톨로지들에 담겨 있는 내용과 개념화는 그 온톨로지의 구체적인 용도가 무엇이나에 따라 편중되어 있음이 분명하다. 하지만 그럼에도 불구하고 새로운 온톨로지를 구축하는데 드는 막대한 노력 때문에 온톨로지들을 재사용 하고자 하는 욕구가 더욱 증가하고 있다. 온톨로지 인코딩 작업은 개념화가 서로 다를 수 있다는 어려움 외에 지식표현 언어의 이질성으로 인한 또 다른 어려움을 내포하고 있다. 현재 다양한 유

형의 지식표현 언어들이 있으며, 이들은 표현력, 정형성, 엄격성 면에서 서로 다르다. 관용구(phrase) 기반의 온톨로지들은 수작업으로 만든 자연어 용어들의 계통도(taxonomy)들이다. GO(Gene Ontology)는 관용구 기반의 온톨로지의 한 예이다. 이러한 관용구 기반의 온톨로지들은 자연어 형태를 사용하기 때문에 표현력은 좋으나, 정형성과 엄격성이 부족하여 간혹 재사용에 어려움이 있을 수 있다. 프레임(frame) 기반의 온톨로지 역시 수작업으로 만들지만 개념들은 클래스들로, 각 개념의 성질들은 다른 클래스들에 대한 관계들로 표현되는 객체지향적 관점을 가진다. 따라서 이들은 좀더 정형적이나, 표현력은 관용구 기반의 온톨로지에 비해 부족할 수도 있다. 하지만 프레임 기반의 온톨로지들이 갖는 객체지향적 관점은 많은 이용자들에게 매우 직관적이어서 관심이 높다. 표현 논리(description logics) 기반의 온톨로지들은 표현력이 뛰어나고 정형적으로 정의된 의미론(semantics)을 갖는다. 그리고 표현 논리를 이용한 추론은 다른 지식표현 언어들에서는 찾아 볼 수 없는 엄격함을 제공한다. 하지만 논리형태는 많은 이용자들이 사용하기에는 어려워 보인다는 단점이 있다.

OIL은 프레임 기반의 패러다임과 표현 논리 기반의 패러다임의 장점을 결합하고 있다.

OIL에서는 개념들이 클래스들로 표현되는 프레임 기반의 구조를 이용한다. 하지만 클래스와 슬롯들은 매우 표현력이 높은 개념식(concept expression)들로 표현될 수 있다[8]. 논리식 추론을 통해 정의들의 논리적 일관성을 검사할 수도 있고, 개념 묘사들로부터 하나의 포함 계층 구조(subsumption hierarchy)를 유도해낼 수도 있다. 따라서 OIL은 프레임 기반의 시스템이 갖는 모델의 직관성과 표현 논리 기반의 시스템이 갖는 표현력과 엄격성을 함께 가진다. 따라서 OIL로 인코딩된 온톨로지들은 점차 더 엄격하고 서술적인 형태로 진화할 수 있다. OIL은 다양한 모델링 스타일과 정형식에 적합할 뿐 아니라 현존하는 몇 가지 다른 인코딩 방법들로부터 변환하거나 이들로 역변환이 가능하여 하나의 교환언어(exchange language)로도 적합하다. 따라서 향후 이러한 장점을 지닌 OIL에 기초한 다양한 바이오 온톨로지 개발연구가 활발히 진행될 것으로 예상된다.

참고문헌

- [1] R. Altman et al., "RiboWeb: An Ontology-

- Based System for Collaborative Molecular Biology," IEEE Intelligent Systems, Vol.14, No.5, pp.68-76, 1999.
- [2] P.G. Baker et al., "An Ontology for Bioinformatics Applications," Bioinformatics, Vol.15, No.6, pp.510-520, 1999.
- [3] A. Farquhar, R. Fikes, and J. Rice, "The Ontolingua Server: A Tool for Collaborative Ontology Construction," Technical Report KSL-96-26, Stanford University, 1996.
- [4] D. Fensel et al., "OIL in a Nutshell," Proceedings of EKAW-2000, LNAI, 2000.
- [5] C. Goble et al., "Transparent Access to Multiple Bioinformatics Information Sources," IBM Systems Journal, Vol.40, No.2, 2001.
- [6] C.D. Hafner and N. Fridman, "Ontological Foundations for Biology Knowledge Models," Proceedings of ISMB-1996, pp.78-87, AAAI Press, 1996.
- [7] J. Hendler and D.L. McGuinness, "The DARPA Agent Markup Language," IEEE Intelligent Systems, Jan. 2001.
- [8] Ian Horrocks et al., "OilEd: A Reasonable Ontology Editor for the Semantic Web," Proceedings of IJCAI-2001, 2001.
- [9] T. Berners Lee, Weaving the Web, Orion Business Books, 1999.
- [10] P.D. Karp et al., "The EcoCyc and MetaCyc Databases," Nucleic Acids Research, Vol.28, pp.56-59, 2000.
- [11] T.E. Rothenfluh et al., "Reusable Ontologies, Knowledge Acquisition Tools, and Performance Systems: PROTEGE II Solutions to Sisyphus-2," Int'l Journal of Human Computer Studies, Vol.44, No.3, p.303-332, 1996.
- [12] S. Schulze Kremer, "Ontologies for Molecular Biology," Proceedings of the 3rd Pacific Symposium on Biocomputing, pp.693-704, AAAI Press, 1998.
- [13] R. Stevens, C.A. Goble, and S. Bechhofer, "Ontology-based Knowledge Representation for Bioinformatics," Briefings in Bioinformatics, Vol.1, No.4, pp.398-416, 2000.
- [14] The Gene Ontology Consortium, "Gene Ontology: Tool for the Unification of Biology," Nature Genetics, Vol.25, pp.25-29, 2000.

김 인 철



1985 서울대학교 수학과 학사
1987 서울대학교 계산통계학과 석사
1995 서울대학교 계산통계학과 박사
1989~1995. 12 경남대학교 전산통계학과 전임강사, 조교수
1996~현재 경기대학교 정보과학부 부교수
관심분야 : 지능형 에이전트, 데이터마이닝, 바이오인포매틱스
Email : kic@kyonggi.ac.kr
