

온톨로지를 이용한 의학용어의 개념 모델링 사례 분석 연구

A Study of the Case Analysis of Conceptual Modeling of Medical Terminologies by Ontology

이 현 실(Hyun-Sil Lee)*

초 록

최근 의학정보 분야에서는 임상 지식관리와 의학정보 검색의 효율화를 위한 수단으로 온톨로지의 개념 모델링을 이용한 의학용어 시스템에 관심이 모아지고 있다. 본 연구는 우리나라의 의학정보 분야에 이러한 시스템의 응용이나 새로운 시스템 개발에 기초적인 자료제공을 목적으로, 정보 모델링과 온톨로지의 이론에 대해 고찰하였고, 외국의 의학정보 분야에서 온톨로지를 이용한 용어 시스템이 개발된 4 가지 대표적인 사례를 분석하여 비교하였다. 연구결과 비형식적인 수준의 온톨로지인 MeSH의 의학용어 표준화와 UMLS의 용어 개념화, 형식적인 수준의 온톨로지인 ON9의 의학 온톨로지 통합의 이론화 그리고 GALEN의 의학지식의 의미 모델과 형식화로 핵심적 특징을 요약할 수 있었다. 온톨로지의 응용은 목적하는 시스템에 따른 수준적 차별화가 이루어져야 할 것이고, 본 연구의 분석 결과가 참고 될 수 있을 것이다.

ABSTRACT

Recent research in the field of medical information systems has paid much attention to an ontology based medical terminology system to support clinical study and effective information search. This study aims to conduct research for further application or construction of ontology systems in Korea. This research reviews the theory of concept modeling and ontology, and analyses 4 cases of conceptual modeling of medical terminologies by ontology. The findings of this study display these specific characteristics in medical ontologies : (1) The standardization of terminology on MeSH. (2) The conceptualization of terminology on UMLS. (1) and (2) are showed as unformal ontologies. (3) The theory of ontology integration in ON9. (4) The reference model of medical knowledge with formalization in GALEN. (3) and (4) are showed as formal ontologies. The application and construction of ontology should be differentiated according to the level of the proposed system, and then this analysis will provide useful information for the researcher and developer of the system.

키워드: 개념모델링, 온톨로지, 의학용어

MeSH, UMLS, ON9, GALEN, concept modeling, ontology, medical terminology

* 원광대학교 중앙도서관 사서 (hyunsil@wonkwang.ac.kr)

- 논문접수일자 : 2004년 8월 16일
- 게재확정일자 : 2004년 9월 18일

1. 서론

의학은 발전과 변화의 속도가 매우 빠른 학문 영역일 뿐 만 아니라, 새로운 용어가 지속적으로 출현하고 있고 생물학과 같은 인접 학문의 용어와도 관련성이 크기 때문에 복합적인 용어 체계가 방대하고 용어관리가 상당히 복잡한 영역이다. 따라서 효율적인 의학 용어의 관리를 위한 용어의 통일과 표준화에 관한 오랜 연구가 있었으며, 이 분야에서 색인과 시소러스의 발달 및 이를 탑재한 정보시스템이 먼저 발전될 수 있었다. 그러나 기존의 의학용어 시스템은 다른 분야의 시스템과 마찬가지로 문제점이 지적되고 있고 개념을 기반으로 하는 온톨로지가 의학의 새로운 용어시스템 구축의 수단으로 주목받게 되었다.

이제까지의 키워드-기반 정보 검색은 원하는 내용에 대해서 동일한 의미를 가지고 다르게 기술되는 용어로 정확한 정보에 접근하기 어렵고, 동일한 용어지만 다른 의미로 사용된 용어에 관련해서는 필요 없는 정보를 추출한다(Fensel et al. 2000). 또한 이용자가 주제의 검색에 대한 상세한 지식이 없이 목적에 맞는 질의를 잘 구성하기도 어렵다. 이에 시소러스가 의학의 주제접근에 유용하게 사용되어왔지만 시소러스 역시 용어의 관계가 보통 동등, 계층, 연관 관계로 단순하게 표현되고 관계에 대한 정의가 명확치 않으며, 이것은 주제 영역의 개념화 구조 보다는 용어의 관련구조를 형성하고 있으므로 의학의 용어적 변화를 수용하기도 어려웠다. 또한 인터넷이 발달함에 따라 다양한 시스템에 접근과 공유가 요구되고 있지만 표준적 개념을 공유하고 많은 분산된 자원

에서 정보 통합과 호환이 힘들다.

더욱이 최근에는 임상에 있어서 의학 데이터의 재사용 문제가 제기되고 있고, 환자 데이터 공유 및 의학의 지식관리에 관한 요구가 증가하고 있다. 하지만 단순한 데이터 관리 방식인 기존의 시스템에서는 이를 충족시키기가 어려운 실정이다. 이에 의미 기반의 개념적 용어관리가 이러한 문제를 해결하고 의학지식을 처리할 수 있는 선결 요건으로 대두되고 있다. 이것은 복잡한 의학용어에 대한 상세한 개념화가 의학정보시스템의 내부에서 정보의 식별과 수집에 유용한 수단이 될 뿐 만 아니라 이종의 시스템 간에도 정확한 커뮤니케이션이 이루어질 수 있는 토대가 되는 것으로, 이를 바탕으로 향상된 의학지식의 관리와 서비스를 기대할 수 있기 때문이다.

온톨로지 개념 모델링은 주제 용어를 명확하게 개념적으로 구조화하므로 이러한 표준화한 개념으로 검색 절차의 정제와 확장은 물론 시스템간의 상호운영성을 제공할 수 있으며, 지식의 처리와 추론이 가능하기 때문에 지식베이스를 구축하고 추론 메커니즘을 가져야 하는 지식관리 시스템에도 적합하다. 이와 같은 이유로 용어체계가 복잡하고 고도의 지식 관리와 처리가 요구되고 있는 의학 분야에서, 기존의 용어 시스템의 단점을 보완할 수 있는 수단으로 온톨로지의 개념 모델링을 이용한 용어시스템을 구축하여 임상적 표준적인 지식관리와 효율적인 의학 정보서비스를 추구하고 있는 것이다.

현재 우리나라 의학 분야에서도 온톨로지에 대한 관심이 확대되고 있으며 온톨로지의 개발과 활용이 절실한 시점이다. 이에 본 연구는 의학의 지식관리와 정보서비스를 향상시킬 수

있는 수단이 될 수 있는 의학 온톨로지 구축과 응용을 위한 기초적인 자료제공을 목적으로, 의학분야에서 잘 알려진 외국의 온톨로지 기반 의학용어의 개념 모델링 사례를 분석하고자 한다. 이를 위해 이론적인 배경으로서 개념 모델링의 발전 과정과 온톨로지의 개념구조 및 의학용어의 개념모델링에 대해 고찰하고, 대표적인 의학 4개의 의학 온톨로지인 MeSH, UMLS, ON9, GALEN을 조사하여 각 온톨로지의 특성을 파악하고 비교한다.

2. 이론적 배경

2.1 정보의 개념모델링

정보 모델링은 현실세계의 실체를 정적 특성(사건)과 동적 특성(행동)을 중심으로 사고하고 표현하는데 필요한 수학적으로 잘 정의된 개념들의 집합으로 정적 특성은 스키마(schema)에 정의되고 동적 특성은 기능 명세(Specification)로 정의되고 있으며, 정보 모델은 기본적으로 객체(object), 연산(operation), 제한(constraints)의 개념을 갖는다(Brodie, Mylopoulos, and Schmidt 1984). 즉 정보 모델링이란 정보의 뜻을 파악하고 유용하게 조직하는 컴퓨터 기반의 구조적 표현이라고 할 수 있다.

Mylopoulos(1998)는 정보모델링을 물리적 정보 모델링, 논리적 정보모델링 개념적 정보 모델링의 세 가지 범주로 구분하였다. 물리적 정보 모델링(physical information modeling)은 컴퓨터의 처리를 위한 기억장치에서 레

코드, 스트링 리스트 변수명 등에 응용되는 보통 데이터와 프로그래밍의 구조로 컴퓨터 시스템과 프로그램이 고려대상이다. 논리적 정보 모델링은 70년대 초반에 추상 수학적인 상징 구조를 제공하는 논리적 정보 모델링(logical information modeling)이 제안된 것에서 비롯되었으며 현실세계를 컴퓨터와는 독립적으로 자연스럽게 기술하는 것이다. 데이터베이스의 관계 모델, 계층 모델과 네트워크 모델이 논리적 정보모델의 예이다. 논리적 모델설계는 정보베이스를 구축하는데 있어서 물리적인 구체화 방법과는 무관하게 정보의 논리적 연관 관계를 구조화 할 수 있다.

개념적 정보 모델링(conceptual information modeling)은 논리적 정보 모델이 제안된 직후 인지과학의 영향을 받아 발전된 것으로, 정보 모델 응용과 정보베이스의 구축을 위한 풍부한 표현력을 가진 정보 모델로 제안되었다. 개념적 정보 모델링은 정보의 조직에 있어 일반화(generalization), 집단화(aggregation), 계층화(classification)와 같은 용어의 추상화 메커니즘 수단을 제공하며 모델링의 과정이 인간의 사고구조에 가깝다. 구현차원의 설계인 물리적 모델링과 논리적 정보 모델링의 상위 개념 모델 또는 이와는 무관하다고도 할 수 있는 개념적 모델링은 현실세계 그 자체의 인식을 바탕으로 하고 있기 때문에 설계가 편리하고 풍부하게 표현을 할 수 있는 장점이 있다. 이 때문에 개념 모델링 시스템 설계자, 특정 영역의 전문가, 최종 사용자 간의 의사소통이 매우 용이해지며, 지능형 시스템의 핵심요소로 부각되고 있다.

개념정보 모델은 1960년대 후반, Ross

Quillian이 인간기억의 구조를 모델링하기 위한 장치로서 개념간의 관계를 중심으로 그래프 형태의 의미망(semantic networks) 이론을 제시한 것에서 출발된다. 그 후 Peter Chen은 의미망 이론에서 진일보한 ER모델(entity-relationship model)이라는 의미론적 데이터 모델 설계 기법을 개발하였고 1970년대 중반에는 Marvin Minsky가 일반 상식적인 지식을 표현하는 개념 구조로 프레임(frame) 이론을 제안하였다. 1980년대 초반부터 의미망에 논리와 절차적인 요소들을 통합하려는 시도가 있었고, 1980년대 후반에는 데이터 중심의 계층 개념에 입각한 계층형 구조인 객체지향 설계 기술들이 소개되었다(Mylopoulos 1998). 그 후 웹 기술의 발전과 함께 증가되고 있는 멀티미디어 등의 다양한 정보자원의 관리를 위한 모델링 방법으로 1990년대 중반 개념을 구조화한 메타데이터 모델링의 방법들이 제안되었다. 메타데이터 모델링 방식은 종래의 다양한 개념 정보 모델링 방법을 통합한 것이라고 할 수 있으며, 이것으로 개념 정보 모델링을 위한 가장 원리적이고 실용적인 접근방식을 도출해 낼 수 있게 되었다.

메타데이터를 체계적으로 표현하고 있는 온톨로지는 명시적으로 정의되고 있으며 기계처리가 가능한 개념모델의 구조를 가지고 프레임 워크를 구성한다. Wand(1996)는 온톨로지를 메타 모델링의 기초로 보아, 온톨로지를 이용하여 모델링 방법의 평가를 위한 기초지침으로 실천적인 규칙을 만들어 제공하였고, 정보시스템의 모델링에 적용하였다. 기초적인 전제는 모델링 문법은 정보시스템의 이용자가 흥미를 갖는 실제 세계의 모든 사물을 표현할 수 있어

야 하며, 그렇지 않다면 결과로서의 모델은 불완전하다는 것이다. 이러한 관점에서 앞서의 정보 모델링 방식은 역사적 변천과 함께 다양한 방식이 제시되고 있지만, 충분한 이론적 기초를 제공하지 못하고 있다. 온톨로지는 모델링에 대한 충분한 이론적 기초를 제공하며 개념적인 표준화를 가능하게 한다.

2.2 온톨로지의 개념과 종류

정보시스템은 현실세계의 시스템을 인공적으로 표현하고 있으며, 세상의 기본적인 특징을 다루는 철학의 이론인 온톨로지가 인공지능 분야에 도입이 되어 추상적 구조를 가지고 최근 정보 모델의 기초가 되고 있다. 1990년대 초부터 지식설계, 자연어처리 및 지식표현 등을 위해 본격적으로 연구되기 시작한 온톨로지는 특정영역에 공유하는 개념의 형식 명세를 만들어 지식을 포착하고 공통적으로 합의된 이해를 제공하는 것을 그 목적으로 한다. Neches(1991)는 온톨로지를 “특정분야의 용어와 용어들 간의 관계를 정의하며, 용어의 조합규칙과 용어의 확장까지도 포함된다”고 하였고 Gruber(1993)는 “공유된 개념화에 대한 형식적이고 명시적인 명세”라고 정의하였다.

온톨로지는 특정분야의 정보 모델에 이용되어 그 분야에서 공통의 어휘를 제공하고, 그 용어의 의미와 용어간의 관계를 다양한 수준의 형식성을 가지고 정의한다. 형식적으로 정의된 온톨로지에서는 지식은 주로 클래스(class), 관계(relation), 함수(function), 공리(axiom), 인스턴스(instance)의 다섯 가지 요소를 이용한다(Corcho, Fernández-López, and Pérez

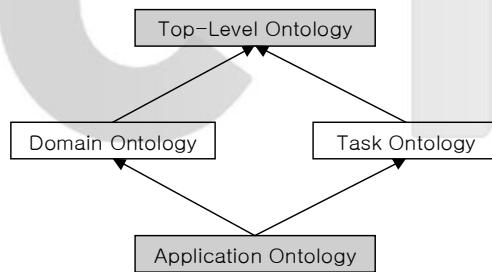
2000). 여기서 ‘클래스’는 보통 개념어에 해당되고 ‘관계’는 개념들 사이의 속성의 유형을 말한다. 함수는 관계가 특정 값을 가지게 될 때 성립되는 것이고, 공리는 논리의 전개나 추론의 기저가 되는 것으로 ‘참’으로 인정되는 문장을 말하며, 인스턴스는 요소의 실제 값이다 이러한 구성요소를 모두 갖춘 온톨로지는 완전히 형식적인 것이 되며 단지 클래스, 관계, 인스턴스로 정의된 온톨로지를 비형식적 온톨로지라고 한다. 온톨로지의 형식성은 지식을 수식적으로 설계하고 표현하는데 있어서 중요한 역할을 하므로 시스템적인 접근에는 반드시 고려되어야 할 사항이다.

온톨로지의 지식을 이용하고 추론하기 위해서는 지식을 표현하는 온톨로지 기술 언어가 역시 필요하다. 이때 지식의 표현은 보통 기술논리와 프레임을 이용하며, 온톨로지 구현언어는 이러한 기능을 지원할 수 있도록 만들어진다. 최근 온톨로지에 대한 관심이 증대됨에 따라 온톨로지의 개념구조를 체계적으로 기술하기 위한 새롭고 다양한 온톨로지 언어들이 개발되고 있다. 특히 XML을 기반으로 한 RDF/RDFS와 이를 확장한 DAML+OIL 등이 개발되었으며, W3C(2004)에서는 기술논리(description logic) 기반의 이론 체계를 갖고 있는 OWL을 온톨로지 표준 언어로 권고하였다. 또한 개념간의 관계를 RDF와는 다른 측면에서 형식화한 TopicMap도 멀티미디어 정보의 개념 구조 표현에 널리 이용되고 있다.

온톨로지는 목적에 따라서 다양하게 개발될 수 있고 일반성의 수준 또는 형식화의 정도에 따라서도 여러 유형으로 구분할 수 있는데 Guarino(1998)는 일반화의 정도에 따라서 상

위수준(top-level) 온톨로지, 도메인(domain) 온톨로지, 과업(task) 온톨로지, 응용(application) 온톨로지 등으로 구분하였다.

<그림 1>에서 상위수준 온톨로지(Top-level ontology)는 매우 일반적인 개념을 묘사하는 것으로, 다양한 커뮤니티를 위하여 통일된 상위수준에서 온톨로지에 대한 이론을 제공한다. 도메인 온톨로지(Domain ontology)는 특정분야에 한정되는 개념들을 제공하는 것으로 개념을 규정할 때 보다 정제된 정의를 요구한다. 과업 온톨로지(Task ontology)는 특정 과업 수행을 위한 개념을 기술하며, 보통 일반 또는 도메인 영역에서 사용하는 언어를 재사용하여 상위-수준 온톨로지에 도입하여 사용한다. 응용 온톨로지(Application ontology)는 특정 도메인과 과제 모두에 종속되는 개념을 묘사하는 것으로 온톨로지의 특수화라고 할 수 있다.



<그림 1> 온톨로지의 유형

이외에도 형식화의 유형에 따라서 비형식적 온톨로지 레포지터리(Informal ontological repositories), 공리화한 분류(Axiomatized taxonomy), 온톨로지 라이브러리(Ontology Library), 표현온톨로지(Representation ontologies), 일반온톨로지(Generic ontologies),

중개온톨로지(intermediate ontologies)로 구분할 수 있다. 비형식적 온톨로지 레포지터리는 공리가 없는 용어의 집합으로 용어집이나 시소러스의 예를 들 수 있다. 표현온톨로지는 개념화를 명시적으로 표현하며, 특정영역에 관계없이 지식표현 형식의 기초가 되는 온톨로지로서 프레임 온톨로지를 예로 들 수 있다. 일반온톨로지는 특정 영역과는 무관하게 일반적이고 기초적인 개념을 제공한다(Studer, Benjamins, and Fensel 1998). 중개온톨로지는 도메인의 일반적인 개념과 관계를 가지고 도메인 온톨로지와 일반 온톨로지 사이에서 인터페이스로 사용된다. 이처럼 온톨로지의 유형은 다양하지만, 지식표현에 광범위하게 사용되고 있는 수단으로서 온톨로지의 궁극적인 목적은 특정 영역의 지식을 명백하게 포착하기 위한 것이라고 할 수 있다.

2.3 의학용어 시스템과 온톨로지

의학은 복잡하고 많은 양의 정보와 지식이 유통되는 분야로 그 동안 의학용어의 효과적인 관리와 이를 활용한 의학 지식의 구조화 방안에 대한 상당한 연구가 이루어져 왔다. 1990년대 중반 국제의학정보협회인 WG6에서는 의

료 정보시스템에서 개념과 지식 그리고 기술언어에 대한 권고가 있었는데, 그 내용은 의학지식을 공유하고, 현재의 용어에 대한 구문적의 가능성을 개척하는 방법을 찾아야 한다는 것이다(McCray et al. 1995).

의학 분야에서는 복잡한 의학 용어의 표준화와 관련하여 정보시스템에서 표현의 유연성과 정확성의 문제를 해결하기 위하여 시소러스와 같은 용어체계를 결합한 용어 시스템을 크게 발전시켜왔다. 그러나 이것들은 의학 지식을 처리하기에는 미흡한 것이었고, Dessena (1999) 등은 의학 용어시스템의 미래 방향을 제시하기 위해 과거의 1세대 의학 용어시스템과 앞으로 추구되어야 할 2세대의 지식기반 용어시스템이 구현되어야 할 내용을 비교하여 각 특징을 <표 1>과 같이 요약하여 설명하였다.

<표 1>에서, 지난 제 1세대 의학 용어시스템이 컴퓨터에서 단지 기호의 코딩으로 세부적 내용을 기술하지 못하였다면 지향되어야 할 제 2세대 용어시스템은 개념이나 지식을 표현하며 지식의 세부적인 사항을 보존하여야 한다. 또한 제 1세대에서는 통계적인 수치에 의한 근사치 데이터의 제공이라면, 제 2세대에서는 치료를 준비할 수 있는 정확한 지식을 제공해야 한다. 제 1세대가 대규모의 단일주제 코딩 시스

<표 1> 용어시스템의 변천

제 1세대	제 2세대
<ul style="list-style-type: none"> • 코딩 • 세부사항 무시 • 통계적 • 오프라인 • 대규모의 '단일주제' 코딩시스템 • 분류, 시소러스, 목록 	<ul style="list-style-type: none"> • 표현 • 세부사항 보존 • 치료준비(구성, 계획, 통계)지원, 지식접근 • 온라인 • 소규모의 값 집합과 해답목록의 모음 • 다기능 시스템과 서비스 • 범세계적인 주제도구

템이었고 주로 분류나 시소러스 목록의 제공이었다면 제 2세대는 소규모의 값 집합과 이들을 논리적 추론에 의해 조합된 해답목록의 모음이고 다기능을 제공하는 시스템과 서비스로 범세계적인 주제도구가 된다고 할 수 있다. 현재 의학용어 시스템은 변화의 과정에 있고 새로운 용어 시스템은 지식접근이 가능하고 표준화 및 다양한 기능의 제공이 요구되고 있으며, 여기에 온톨로지 개념 모델링의 방법이 새로운 의학 용어시스템에 적합한 수단으로 제안된다(Burgun 1999). 온톨로지 개념 모델링의 방식은 메타데이터의 의미적체계로 의학 지식영역을 서술하는 표준적 개념과 관계를 수집하여 개념화의 집합을 구성하고, 이로써 의학적인 지식을 표현하고 처리한다.

한편 의학은 급속히 발전하고 변화하는 분야이므로 새로운 개념의 정의와 기존의 개념에 대한 수정의 요구가 생길 수 있다. 확실한 개념 분류의 지침이 제공되지 못했던 ‘病人學’의 예를 들면, 과거에 심이지장계양은 스트레스에 의한 심리적 원인이 있는 병으로만 분류되었지만 헬리코박터 감염과 심이지장계양의 상관성이 증명되어 새로운 병인 분류가 필요했다. 온톨로지 개념모델링을 이용하는 경우 개념의 집합과 관계를 만들면, 이러한 관계는 포섭(subsumption)이나 계층삽입(hierarchies inclusion)을 통한 개념간의 유사성과 차이를 다양하게 표현할 수 있고 위상 기하학적인 관계의 표현도 가능하며 이러한 개념간의 관계 모델링은 의학 용어의 영역에서 일반 이론적 기초의 결여를 보충할 수 있다(Burgun et al. 1999).

Ingenerf 등(2001)은 의학에서 이종의 컴퓨

터 응용을 위한 상호 연결시 의미적 상호운용성의 문제가 야기됨을 지적하고, 이는 전통적인 용어 표준화의 방법만으로는 해결할 수 없는 것임을 지적하였는데 그 내용은 다음과 같이 세 가지 측면에서 접근된다. 첫째, 현재 서로 다른 학문영역에서 병존하고 있는 의학용어의 다양성은 자동화된 시스템 통합에 장애가 되고 있다. 둘째 교환되는 의학 데이터는 환자의 데이터 통합과 임상적 연구는 물론 지식기반의 접근 등 다양한 목적을 위한 기계처리가 요구되고 있다. 셋째, 서비스는 단일화되고 투명하게 접근될 수 있어야 한다는 것이다. 이것은 의학정보 서비스는 더 이상 수동적인 사전식의 용어 제공만으로 충분하지 않다는 것을 의미한다. 의학 용어의 표준적 접근과 의학 지식의 처리는 물론 다양한 시스템 간을 연결할 수 있는 범세계적인 통합시스템으로 서비스될 수 있어야 한다. 기계와 인간에게 공통의 의미를 제공하는 온톨로지는 영역의 지식체계를 형식화하여 표준 모델을 제공함으로써 이종의 시스템은 물론 대규모로 통합된 소프트웨어 시스템에서 표준적인 이해의 틀을 만들 수 있다(Uschold and Gruninger). 이러한 장점으로 온톨로지 개념모델링을 이용하여 의학 정보시스템에서 프레임워크로 사용하고 있는 사례들이 있다.

3. 온톨로지 기반의 의학 용어 시스템 사례 분석

본 장에서는 의학분야에서 온톨로지 개념모델링을 이용하여 용어시스템을 구축한 대표적인 4개의 사례를 분석하고, 핵심적인 내용과

특징을 파악하여 비교하고자 한다. 분석의 기준점은 앞서 기술한 온토로지를 구성하는 중요한 사항들을 중심으로 첫째, 온토로지 개발의 목적이 무엇인가. 둘째, 용어의 개념화 여부와 개념의 규모는 어느 정도인가. 셋째, 개념 간에 설정된 특징적 관계는 무엇인가. 넷째, 온톨로지의 형식화 여부로 공리를 포함하고 있는지와 기술되는 지식 표현의 방법은 무엇인가. 다섯째, 재사용 가능한 수준의 온토로지 구현 언어를 포함하고 있는가. 마지막으로, 핵심적인 온토로지의 유형은 무엇인가 하는 관점이다.

3.1 MeSH(Medical Subject Heading)

MeSH는 미국의 국립의료원 도서관(NLM)에서 의학문헌의 색인과 검색을 목적으로 만든 통제어휘 시소러스로 비형식적인 온톨로지에 속한다고 볼 수 있다. MeSH의 구성은 광의어(Broader-than Term), 협의어(Narrower-than Term), 관련어(Related Term)의 연결로 이루어져 있으며 모두 22,000여개의 디스크립터를 가지고 있다.

MeSH에서 용어의 연결은 관련된 용어간의 관계를 보여줄 뿐만 아니라 협의에서 광의까지 상세하게 탐색할 수 있는 계층적인 구조를 제공한다. Medline의 문헌이 단순한 문헌 참조보다 상세하게 정보를 제공하는 것은 MeSH가 의학용어의 구조화를 통해 문헌의 의미적 데이터(Semantic Data)를 제공하기 때문이다(Abasolo and Gómez 2000). 이 용어시스템에서 의미적인 계층구조는 복합적인 트리도(polytree)로 표시되며 MeSH의 각 객체는 <표 2>와 같은 속성을 가지고 있다.

<표 2> MeSH 객체의 속성

Name	용어의 명칭
Definition	의학적 정의
Relation Terms	용어와 관련된 다른 용어
Subheading	용어의 의미를 한정하고 보충하는 부표목
Position on the polytree	용어의 위치에서 부모와 자식

MeSH 계층구조의 최상위 계층에는 'Anatomy', 'Organisms', 'Diseases', 'Chemicals and Drugs', 'Psychiatry and Psychology'와 같은 일반적인 의학적 의미를 가진 용어를 가장 광의의 표목어로 정하여 점차 협의어의 수준인 보다 특수한 의미를 가진 하위의 표목어로 전개된다. 이들은 최상위의 표목 값에 주어진 알파벳 기호를 동일한 식별자의 값으로 가지고 트리의 최하위까지 전개된다. <그림 2>의 예는 'Anatomy'의 [A] 기호로 전개되는 트리표현의 일부를 보인 것이다.

Anatomy[A]
Fluids and Secretions [A12]
Body Fluids [A12.207]
Blood [A12.207.152]
Fetal Blood [A12.207.152.200]
Plasma [A12.207.152.693]
Hemic and Immune Systems [A15]
Blood [A15.145]
Blood Cells [A15.145.229]
Fetal Blood [A15.145.300]
Plasma [A15.145.693]
Embryonic Structures [A16]
Fetus [A16.378]
Fetal Blood [A16.378.200]
Fetal Heart [A16.378.303]
Meconium [A16.378.529]
Vernix Caseosa [A16.378.857]

<그림 2> MeSH 계층구조 예

이와 같이 MeSH는 추상적 구조로 용어를 체계화하지만 하나의 용어가 여러 개의 하위 트리에 속하는 복합적인 트리 구조를 가지고 있다. 예로써 “Anatomy” [A]에서 Fetal Blood는 ‘Blood’, ‘Blood Cells’, ‘Fetus’의 세 종류의 상위어 아래 복합적으로 전개되고 있다. 그러나 이러한 복합적 전개는 개념분류를 위한 논리적 이론과 형식적 방법이 도입되지 않은, 개념과 관계가 의미적으로 모호한 경우이다. 여기서 ‘blood’는 ‘blood plasma’와 ‘fetal blood’ 모두의 부모 형태이나 ‘fetal blood’와는 일반화(isA)가 성립되는 관계이지만 ‘fetal blood’와의 관계는 부분사(partitive) 관계이다. 이처럼 MeSH의 상·하위어 관계는 논리성이 결여되었기 때문에 속성을 상속받을 수 있는 부모와 자식(is-A)의 관계로 모두 일반화하기가 어려우며, 부분과 전체(Part-Whole)의 관계를 구분하고 있지 않다. 이러한 점에서 MeSH의 계층구조를 형성하는 의미의 관계는 단지 ‘협의어’와 ‘광의어’의 관계를 벗어나지 못한다.

컴퓨터의 의미처리를 염두에 두고 만들어지지 않은 통제어인 MeSH는 관계와 개념이 이처럼 의미적으로 불투명하고 형식적인 개념화가 결여되어 있기 때문에, MeSH를 이용하여 자동추론을 하기는 어렵다(Hann, Romacker, and Schulz 1999). 따라서 MeSH는 의학 용어의 추상적 구조화로 표준적 용어체계를 제공하여 문헌의 검색과 색인에는 유용하지만 지식의 처리와 관리가 요구되고 있는 임상 지식관리 시스템에는 충족되지 못하고 있는 것이다.

3.2 UMLS(Unified Medical Language System)

UMLS는 미국 국립의료원 도서관(NLM)에서 의학용어의 통합을 위하여 1986년 부터 개발을 시작한 개념기반의 의학용어 시스템이다. 이 시스템의 구성은 메타시소러스(Meta-thesaurus), 의미망(Semantic Network), 전문가 사전(Specialist lexicon)의 세 종류 지식소스이다. 이러한 UMLS의 지식소스들은 의학 용어와 약물 및 생의학 관련의 다양한 어휘들로 구성된 분산된 소스를 통합하여 연결하고, 용어들을 통일된 개념을 적용하여 사용하기 위해 만들어진 대규모의 데이터베이스이다.

UMLS의 메타시소러스는 생의학적 개념 개념에 해당하는 다양한 명칭, 개념간의 관계에 대한 의미적 정보를 담고 있다. 이 시스템의 개념어 규모는 생의학적 개념 1,000,000여 개와 각종 의학 데이터베이스와 전문가 시스템에 사용되고 있는 2,800,000여 개의 개념어를 포함하고 있다(NLM 2004). 메타시소러스에서는 관련된 소스의 어휘를 개념으로 통합함으로써 어휘가 상호 연결이 되며, 통합된 개념들은 개념 식별자(CUI: Concept Unique Identifier)를 가진다. 개념 식별자는 동일한 개념이지만 용어가 달리 표현하고 있는 용어 식별자(LUIs: Language Unique Identifier)를 동일한 개념으로 연결한 것이며, 용어 식별자는 동일한 용어이지만 용어가 변형되어 나타나고 있는 문자열 식별자(SULs: String Unique Identifier)를 연결한 것이다. 여기서는 개념이나 의미에 따라 조직되는 식별자(identifier)를 중심으로 동일한 개념으로 사용되는 다른 용어나 문자열이 연결되어 같은 개념으로 인식된다.

Concept (CUI)	Terms (LIUs)	Strings (SUIs)
C0004238 Atrial Fibrillation (preferred) Atrial Fibrillations Auricular Fibrillation Auricular Fibrillations	L0004238 Atrial Fibrillation (preferred) Atrial Fibrillations	S0016668 Atrial Fibrillation (preferred)
		S0016669 Atrial Fibrillations
	L0004327 (synonym) Auricular Fibrillation Auricular Fibrillations	S0016899 Auricular Fibrillation (preferred)
		S0016900 (plural variant) Auricular Fibrillations

〈그림 3〉 문자열과 용어의 개념 연결

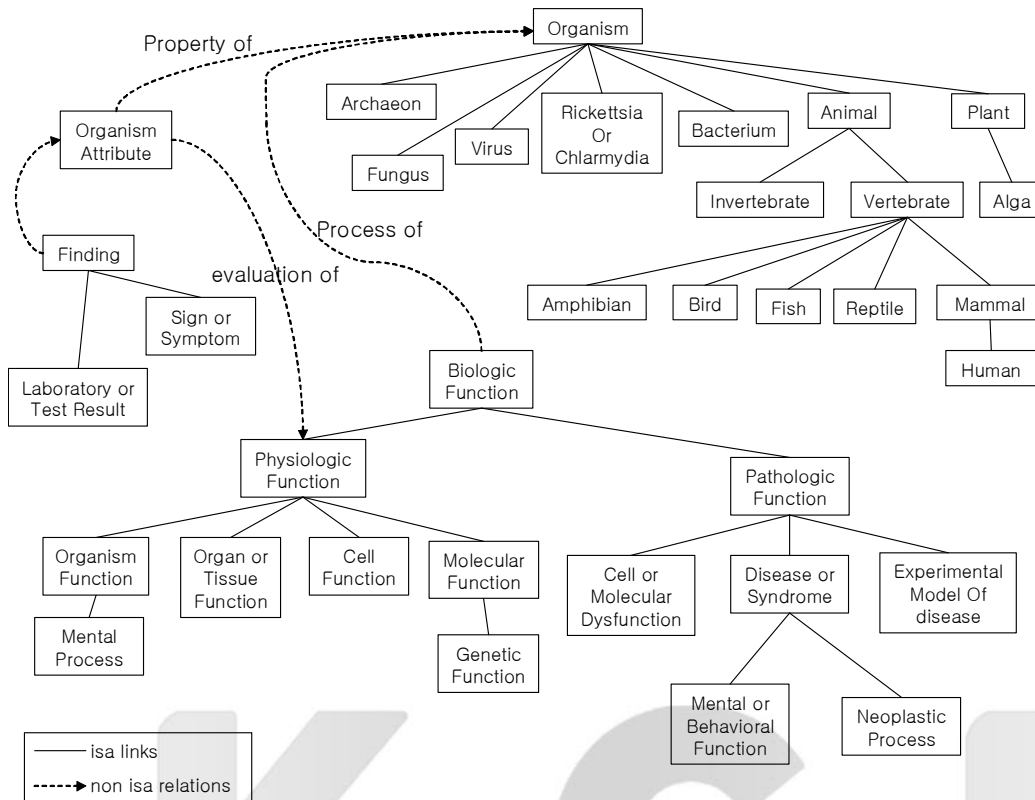
〈그림 3〉(NLM 2003)은 문자열과 용어가 개념으로 연결되는 과정을 도식화 한 것이다.

UMLS 의미망의 구성은 의미유형(Semantic Type)과 의미관계(Semantic Relation)로 이루어져 메타시소러스의 모든 개념들은 최소한 하나 이상의 의미유형 범주에 할당되고, 의미유형 사이에는 의미관계가 부여된다. 의미유형은 크게 개체(Entity)와 사건(Event)으로 분류되고, 이들은 계층적 관계의 하위 의미유형을 가지고 각각의 의미유형에 속하는 개념들을 그룹화 한다.

의미유형을 연결하는 의미관계는 계층 관계(isA)와 비계층 관계(associated with)인데 의미유형간의 연결하는 기본적인 관계인 '계층 관계'는 '개체'와 '사건'의 의미유형 연결에 근원이다. 비계층 관계는 물리적 관계(physically related to), 공간적 관계(spatially related to), 시간적 관계(temporally related to), 기능적 관계(functionally related to), 개념적 관계(conceptually related to)의 다섯 가지 관계의 집합으로 구성된다. 〈그림 4〉는 '개체'의 의미유형에 속하는 'Organism'에 관계되는

의미유형과 의미관계를 도식화한 것으로, isA 관계로 연결되는 의미유형들은 상위 의미유형 클래스에 명시된 관계를 하위 의미유형 클래스가 그대로 상속받는다.

〈그림 4〉에서 상위의 의미유형 'Organism'과 'Organism Attribute'은 비계층 개념적 관계의 일종인 property of의 관계가 있으며, 'Organism'과 'Biologic Function'은 비계층 기능적 관계의 일종인 process of의 관계가 있다. 'Biologic Function'과 'Physiological Function'은 부모자식의 관계가 성립되어 부모 클래스가 가지는 관계를 자식클래스가 상속 받을 수 있으며, 자식인 'Physiological Function'은 다시 'Organism Attribute'와 비계층 개념적 관계를 갖는다는 것을 보이고 있다. 이처럼 의미망에서는 메타시소러스의 개념들이 하나 이상의 의미유형으로 범주화되어 계층적, 비계층적인 관계 유형을 가지고 서로 연결된다. 어휘 변형 정보를 포함하고 있는 전문가 사전은 추후 전문가 시스템에 필요한 용어의 사전적 정보 제공을 목적으로 만들어진 것이다.



〈그림 4〉 UMLS 의미망

UMLS에서는 새롭게 생성되는 용어는 없고 다양한 소스에서 도입한 기존의 용어를 개념을 기반으로 추가된 값(Value)에 따라 어휘를 통합하므로 용어시스템의 기본 엔트리는 단어나 용어가 아닌 개념이다. 따라서 UMLS는 디스크립터나 통제어를 지정하지 않고, 같은 의미에 속하는 용어를 서로 연결시킴으로써, 이종의 시스템에서 용어의 차이에 관계없이 동일한 개념으로 정보의 검색이 가능하며(김혜선 2001), 다양한 관계의 유형을 제시하여 용어시스템의 유용성을 높이고 있다. 하지만 여기서도 방대한 개념어에 대한 형식적 정의와 완전한 온톨로지 체계를 갖추고 있지 않아 의학

지식을 충분히 구현하고 있지 못하며 용어의 자동분류와 자연언어의 처리가 어렵다. 최근 DAML+OIL이나 OWL 등 XML 기반의 온톨로지 언어로 UMLS의 의미 지식 표현이 시도되고 있지만 이 또한 부분적 기술로 이에 대한 평가가 아직 어려운 단계이다.

3.3 ON9

ON9은 ISTC-CNR(Institute of Cognitive Sciences and Technology of the Italian National Research Council)의 온톨로지 그룹에서 의학 용어시스템을 통합하여 대

규모의 온톨로지 도서관을 만들고, 이를 활용하여 임상정보의 지적 검색과 같은 분야에 응용하고자 ONIONS (ONtological Integra-tion Of Naive Sources) 의 방법론으로 개발한 통합 온톨로지의 용어 시스템이다.

ONIONS 방법은 도메인 온톨로지에 광범위한 공리화, 명시적 의미, 온톨로지적인 깊이를 제공하고자 의도된 것이고, 이 방법에 의하여 대규모의 온톨로지 도서관이 생성되었으며 병원의 임상 데이터를 통합하고 지적 정보검색을 실현한 응용 시스템도 개발되었다(Corcho, Fernández-López, and Pérez 2000).

ON9의 온톨로지 통합 방법론에 적용된 의학온톨로지는 UMLS, SNOMED-III, CMN, ICD10, GALEN 등으로 50 만 이상의 개념을 처리하고 있다. 온톨로지 도서관의 용어적 지식의 획득은 UMLS의 메타시소러스를 기초로 위의 시스템들의 개념적 분석을 통한 온톨로지 통합에 의해 획득되었다. 통합의 단계는 (1) 용어집을 수집하여 도메인에 타당한 언어자료 생성, (2) 용어리스트에서 핵심적 분류구조 발

견, (3) 개념의 로컬정의를 분석하여 개념 해석을 위한 분류기준 설정, (4) 멀티-로컬 정의를 분석하여 엘리먼트를 스키마화하고 개념적 원칙을 유도, (5) 온톨로지 도서관 구축, (6) 자동 분류 제공의 순이다(Steve, Gangemi, and Pisanlli 1997).

ON9 개발 방법의 핵심은 기술논리 기반의 모듈 구성과 일반 온톨로지의 이론을 확장한 모듈의 계층화이며, 온톨로지는 슬롯(slot)과 필러(filler)로 지식을 표현하는 프레임 온톨로지(frame-ontology)인 Ontolingua에서 디폴트로 제공하는 표현 온톨로지, 그리고 기술논리와 생성규칙 기반의 Loom 구문을 통해 생성된다. <그림 5>는 프레임 온톨로지 생성의 예로, UMLS로부터 통합한 의학개념 ‘viral-hepatitis type A’를 Ontolingua로 엔코딩하여 약형식적 제한이 있는 정의를 생성한 것이다.

<그림 5>에서 (1)과(2)는 ‘o-umls’로 명명된 이론에서 개념 ‘viral-hepatitis-a’가 hu-man-liver-disease, finding, morphology

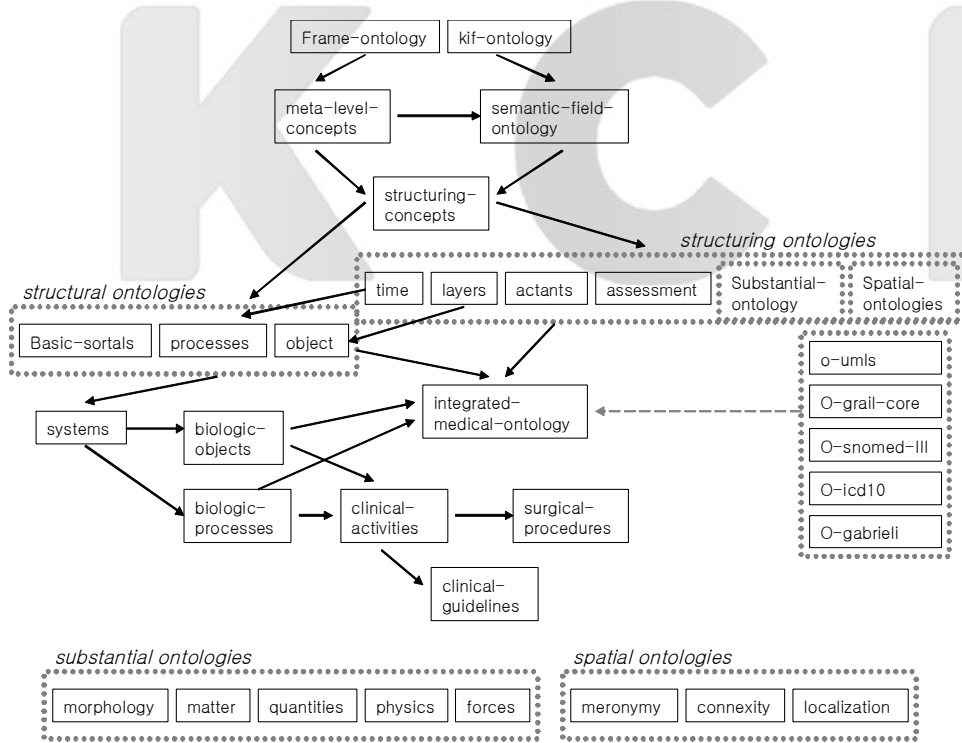
```
(in-theory 'o-umls) (1)
(define-class viral-hepatitis-a (2)
  "the disease of human liver causing a degenerative or necrotic alteration of hepatocytes; it is caused by hepatitis virus A and is transmitted by fecal contaminated food or water. It can be viewed as a morphology of liver as well as a finding"
  axiom-def (and (subclass of viral-hepatitis-a human-liver-disease)
    (subclass of viral-hepatitis-a finding)
    (subclass of viral-hepatitis-a morphology)
    (caused-by viral-hepatitis-a hepatitis-virus-a)
    (involves viral-hepatitis-a degenerative-alteration-of-hepatocytes)
    (involves viral-hepatitis-a necrotic-alteration-of-hepatocytes)
    (transmitted-by viral-hepatitis-a fecal contaminated-food)
    (transmitted-by viral-hepatitis-a fecal contaminated-water)))
```

<그림 5> ON9의 프레임 온톨로지 생성의 예

의 하위클래스로 포섭되고 있으며, 포섭된 개념들로부터 상속되는 다섯 가지 제한사항들을 기술하고 있다.

한편, ON9에서는 일반 온톨로지, 중개 온톨로지, 도메인 온톨로지가 층을 형성하여 도메인 온톨로지는 중개온톨로지 모듈에 포함되고, 중개 온톨로지는 일반 온톨로지 모듈에 포함된다. 온톨로지를 통합하는 일반 온톨로지에서 철학적 개념의 일반 온톨로지 이론들이 잘 정의되고 있고 이들 관계의 특성은 “structuring-concepts”의 이론에 정의되고 있다. 이러한 “structuring-concepts”은 “meta-level concepts”, “semantic-field-ontology”와 함께 일반 온톨로지 도서관을 가지고 있는 표현 온

톨로지에 링크되도록 정의되고 있으며 “structural ontologies”와 “structuring ontologies”의 집합이 일반 온톨로지를 내포하게 한다(Steve, Gangemi, and Pisanlli 1997). <그림 6>은 이 같은 온톨로지 삽입격차를 보여주고 있는데, Onto-lingua에서 디폴트로 제공되고 있는 표현온톨로지는 “frame-ontology”와 “kif-ontology”의 집합이다. 여기서 온톨로지는 실선박스로 표현되었고 두꺼운 점선박스는 온톨로지 집합을 나타낸 것이다. 실선 화살표의 의미는 Ontolingua의 제한사항에 따라 적용되는 포함(include in)을 나타내고, 점선 화살표는 통합(integrated-in)을 의미한다. 통합된 의학적 온톨로지는 다섯 개 용어시스템



<그림 6> ON9의 온톨로지 도서관 삽입격차

의 용어적 온톨로지 통합에 사용된 모든 일반 온톨로지를 포함한다.

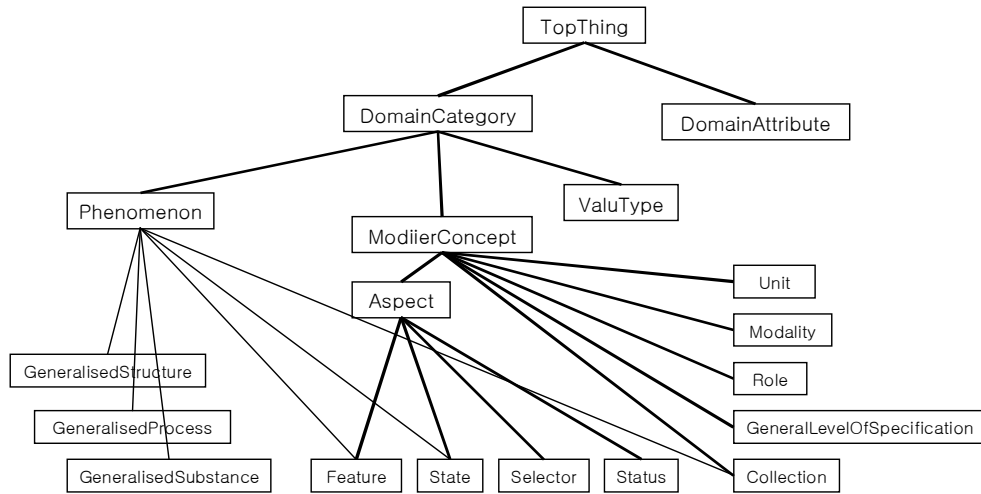
ON9의 온톨로지를 통합하는 일반 온톨로지의 이론들은 상위수준개념, 형식온톨로지(formal ontology)이론, 계층화(stratification)이론, 개체화(Individuation)이론, 행위자(Actor)이론, 인식론(Epidemiological)이론, 메타온톨로지(metaontology), 영역(domain)이론 등의 철학적 이론들이고 각 개념들은 포함(include in)의 관계를 가지고 해당 이론의 개념에 포섭된다(Gangemi, Pisanelli, and Steve 1999). 이처럼 ON9은 용어를 통합하는 상위수준에서 철학적 개념의 일반 온톨로지들을 도입하여, 의학 용어 시스템에 실세계의 존재론적인 의미를 풍부하게 담고 있다. 또한 온톨로지 소스에서 획득한 개념, 관계, 템플릿, 규칙 공리가 형식적인 방법으로 표현되고 있는 바와 같이 형식적인 표현 온톨로지인 의학의 용어 시스템을 형식적으로 통합하여 표현함으로써 다양한 시스템 간에 공유할 수 있는 표준화한 용어의 명세를 제시하고 있다.

3. 4 GALEN

GALEN(Generalized Architecture for Languages Encyclopedias and Nomenclatures in Medicine)은 의학분야에서 의학 용어의 다국어 응용을 위해 온톨로지용어 서버를 개발한 유럽지역의 프로젝트이다. 프로젝트의 목적은 온톨로지를 이용하여 서로 다른 데이터 소스에서 전자적 의학데이터를 호환할 수 있도록 의미적으로 타당한 의학용어의 모델을 만드는 것이다. 광범위하게 재사용할 수 있

는 의학용어의 의학지식베이스라고 할 수 있는 GALEN의 온톨로지 기술은 개념 모델링 언어인 GRAIL(Galen Representation And Integration Language)이라는 자체개발한 형식언어로 이루어지는데 GRAIL은 프레임 기반의 기술논리로 된 지식표현 언어이다. GRAIL을 이용하여 임상 용어의 표현모델을 제공하며 여기서 의학개념들을 재사용할 수 있도록 지원하는 기술들은 1) 전자보건기록, 2) 임상 인터페이스, 3) 분류와 코딩시스템 4) 결정 지원시스템, 5) 지식관리시스템, 6) 자연언어처리이다.

GALEN이 의학에 풍부하게 응용될 수 있는 의미 모델을 만들어 제공하고자 하는 기본원칙은 의미 모델에서 개념의 분류가 명백해야 한다는 것과 개념이 여러 축으로 분류되어 충분한 응용이 가능해야 한다는 것이다. GALEN에서 도메인 범주의 가장 일반적인 상위수준의 개념 분류는 '사물(thing)'과 '사물의 속성(properties)'이며, 사물(thing)에서는 일반화한구조(GeneralisedStructures), 일반화한물질(GeneralisedSubstances), 일반화한절차(GeneralisedProcesses)의 세 가지 유형으로 나누어진다. 이 세 가지 범주에서 객체는 일반적으로 '구조적인 속성(Constructive Attribute)'으로 알려진 'is part of', 'has function', 'is consequence of' 등의 의미링크의 가족을 이용하여 서로 연결된다. 구조의 특질 등을 나타내는 속성(properties)을 표현하는 한정자 개념(Modifier-Concepts)은 '한정자 속성(ModifierAttributes)'인 'hasShape', 'hasFrequency', 'hasLeft-RightSelector' 등의 링크를 이용하여 사물의 객체와 연결된다(Rogers



<그림 7> GALEN Core 모델의 상위수준 개념공간

and Rector 1999). <그림 7>은 GALEN의 Core 모델 상위수준의 개념을 도식화 한 것으로 구조, 물질 절차의 분류와 구조의 특성을 묘사하는 한정자 개념의 분류를 보여준다.

GALEN의 ‘용어서버’는 개념 모듈, 다국어 모듈, 코드변환 모듈의 세 가지 모듈을 통합하여 이루어진다. 개념모듈은 GRAIL(GALEN Representation and Integration Language)로 기술되며 GRAIL은 Core(Concept Reference)모델에서 개념들과 개념들의 성격 및 관계를 표현하는데 이 온톨로지에서는 5000여 개념과 1000여개 정도의 관계가 정의되고 있다. 이러한 개념과 관계를 이용하여 코딩되는 참조모델은 기본적인 의학개념(예, ‘fracture’, ‘bone’, ‘left’, ‘humerus’ 등), 개념들이 결합되어야 할 방식을 통제하는 관계(예, ‘fractures can occur in bones’), 그리고 복합적인 개념들(예, ‘fracture of the left ones’)로 구성되었다. 이처럼 GALEN에서는 해부학, 생리학, 병리학 등에서 가장 핵심이

되는 의학개념들과 의미모델을 제공하고 있으며 이를 재사용하여 의학지식을 다양하게 모델할 수 있다. GRAIL 형식론을 이용하면 이러한 개념들과 관계들을 포함하는 모델들을 만들 수 있고 기존의 모델에 구성적으로 타당한 새로운 개념들을 획득할 수 있다. GRAIL에 포함되어 있는 구성 규칙은 시스템의 술어를 만들고, 정의 규칙은 어휘를 만든다. 다양한 언어에서 동의어와 우선어 개념 매핑은 다국어 모듈에 의해 관리되고 기존의 코딩시스템에 개념 매핑은 코드변환 모듈에 의해 관리된다(Keizer, Abu-Hanna, and Zwetsloot-Schonk 2000).

<그림 8>은 Core 모델에서 ‘Fracture(골

```
Fracture which <
  hasLocation (AnatomicalNeck which
    isDivisionOf Femur) Which
  hasCause PostMenopausalChange>
```

<그림 8> GRAIL로 표현된 의미모델의 예

절)’의 개념과 그 관계의 의미모델이 GRAIL로 표현된 간단한 예이고, 여기서 ‘골절의 부위와 인과가 논리적으로 분류되고 있는 것을 볼 수 있다. 이와 같이 개념은 <그림 9>와 같이 다른 추상구조 아래 분류될 수 있다. 요구에 따라 새로운 추상화가 만들어 질 수 있고, 분류기는 관련된 추상항목을 찾아 개념을 분류한다. 사용자의 목적에 따라 다른 추상개념을 사용하여 다르게 응용할 수 있으며 다른 축을 따라 모델 할 수 있다.

```

Fracture
Fracture which hasLocation LongBone.
Fracture which hasLocation (AnatomicalNeck Which isDivisionOf Femur).
Fracture which hasLocation Thigh.
Fracture which hasLocation Hip.
Lesion Which isCausedBy Osteoporosis.
Lesion Which isCausedBy PostmenopausalChange.
etc.
    
```

<그림 9> GRAIL로 표현된 의미모델의 예

GALEN은 Part-Whole의 관계 그리고 연관된 추론의 형태를 명백히 구분하여 논리적으로 성립된 온톨로지 형식이론의 가능성을 충분히 보여준다. Part-Whole 관계는 의학의 개념표현에 있어서 매우 중요한 역할을 하며 해부학에서 가장 명백한 모델링이 되고 있지만 생리학과 질병의 절차 및 외과적 절차의 표현에도 큰 영향을 준다(Hahn, Romacker, and Schulz 1999). GALEN에서는 Winston과 Odell에 의해 발전된 Part-Whole 관계의 범주를 채용하고 있는데, 그 종류는 component of, stuff of, portion of, area of, member of, partner of, piece of의 7가지가 있으며

그중 앞부분 5가지가 주로 사용되고 있다 (Rogers and Rector 2000). <그림 10>은 GALEN에서 표현되고 있는 Part-Whole 관계의 해부학적 모델링의 예이다.

```

articular cartilage component-of
Knee joint component-of
Knee region component-of
thigh component-of lower extremity
    
```

<그림 10> Part-Whole관계 모델링 표현의 예

이처럼 GALEN은 다른 용어시스템들과는 다르게 가장 온톨로지적인 개념 구조를 잘 표현하고 있다. 이것은 시스템으로 온톨로지 표현 언어인 GRAIL은 기술논리 기반의 형식 규칙을 제공함으로써 풍부한 개념 표현의 방법을 제공하고 있기 때문이고, 온톨로지 형태의 용어 시스템에서 가장 중요한 기능인 재사용 가능한 개념으로 의학 지식을 표현하고 있는 것이다(김흥기, 김명기 2003).

종합하여, GALEN은 풍부한 의식지식을 모델하며 형식적인 체계로 온톨로지를 가장 잘 실현하고 있는 용어시스템이라고 볼 수 있다. 그러나 현실적으로 UMLS와 같은 방대한 규모의 개념을 모두 가지고 이러한 온톨로지를 구축하기는 매우 복잡하고 어려운 일이다. 지능적 처리 또는 일반적 시스템 구축 등 목적하는 바에 따라서 GALEN이나 ON9과 같은 형식체계를 갖춘 중량급 온톨로지, 혹은 UMLS와 MeSH와 같은 경량급 온톨로지를 선택하여 응용할 수 있을 것이다. <표 3>이 이상에서 고찰한 MeSH, UMLS, ON9, GALEN의 용어모델링 사례의 특징을 비교한 것이다.

〈표 3〉 의학용어의 온톨로지 개념모델링 사례 비교

비교항목	GALEN	ON9	UMLS	MeSH
목적	다국 용어의 처리 의학지식의 모델	의학 용어 시스템의 통합	의학용어의 개념화	의학용어의 표준화
기본엔트리	개념	개념	개념	용어(개념)
개념의 수	5000 이상	500,000이상	2,500,000 이상	20,000이상
특정적 관계	Part-Whole	include in	isA	BT NT
공리 (Axiom)	포함	포함	비포함	비포함
지식표현	기술논리	Frame 기술논리	의미망	비포함
온톨로지 구현언어	GRAIL	Ontolingua	비포함	비포함
핵심 온톨로지 유형	도메인온톨로지 공리화한 분류	표현온톨로지 일반온톨로지	도메인 온톨로지	비형식온톨로지 도메인온톨로지

4. 결론 및 제언

의학은 임상에 있어 데이터의 재사용과 공유의 요구가 많으며 다양하게 지적 정보 서비스가 요구되고 있는 학문 분야이다. 따라서 현재 국외의 선진 의학정보기관들은 의학 지식 서비스의 실현을 목적으로 온톨로지의 개념 모델링을 이용하여 복잡한 의학 용어를 표준화한 용어시스템을 구축하여, 이를 기반으로 의학 지식의 표현과 처리를 시도하고 있다. 온톨로지의 개념모델링을 이용한 지식베이스구축은 의학 지식을 정확하게 수집하고 제공할 수 있는 유용한 수단으로 제시되고 있기 때문에, 이를 이용한 통합된 개념체계 용어 시스템을 구성하여 검색 및 이종의 시스템과도 상호 호환이 가능한 응용의 환경을 제공하고자 하는 것이다. 본 연구에서 의학 분야에서 4개 대표적인 용어시스템의 온톨로지 개념 모델링의 사례를 분석한 결과 다음과 같은 특징을 파악할 수 있었다.

MeSH는 메타데이터로 의학용어의 의미를 표준화한 계층구조의 관계를 형성한다. 그러나

단지 ‘협의어와 광의어’의 관계로 문헌의 검색과 색인에는 유용하지만 논리적인 일관성 부족 및 온톨로지 형식성의 결여로 지식의 표현이 필요한 임상적인 요구에는 충족되지 못한다. UMLS는 의학에 관련된 전 영역의 용어를 개념화한 대 규모의 용어시스템이다. 하지만 이것 역시 응용시스템과 연결시킬 수 있는 형식적 공리가 결여되어 있으므로 데이터의 논리적인 참조를 제공하지 못한다. ON9은 의학용어를 통합하고 구조화하는데 있어, 상위수준에서 존재론적으로 풍부한 개념 통합의 이론과 형식적 표현의 표준 명세를 제공한다. GALEN은 의학지식의 의미 모델과 이것의 형식화로 온톨로지의 형식론적인 가능성을 가장 잘 보여주고 있다. 그러나 ON9과 GALEN에서 표현하고 있는 개념의 규모는 UMLS에 비하여 적는데, 이것은 대규모의 개념어 대한 형식적 표현과 추론의 기계적 절차가 그만큼 어렵다는 것을 의미한다.

이들 4개 용어시스템의 비교를 통하여 MeSH의 의학 용어의 표준화, UMLS의 용어의 개념화, ON9의 의학 온톨로지 통합의 모델

과 이론화, GALEN의 의학지식 의미의 모델과 형식화로 주요 특징을 요약할 수 있다. 이를 기초로 본 연구에서는 다음의 사항들을 제안한다.

첫째, 온톨로지는 구축과 재사용에 있어서 명확한 목적 설정과 사전지식이 중요하므로, 본 연구에서 분석한 각 의학 용어 시스템의 특징과 수준을 참조하여 목적하는 시스템에서의 지식처리 정도에 맞는 용어시스템을 공유하고 재사용할 수 있는 방법으로 응용이 이루어지고 온톨로지가 구축되어야 할 것이다.

둘째, 본 연구의 사례로 제시한 각 용어 시스템들은 오랜 기간의 연구와 개발이 이루어졌고 지속적인 개선이 추진되고 있는 것들이다. 따라서 향후에도 각 시스템에 대한 지속적인 고찰이 필요하고 시스템 변화에 따른 진화된 형태의 응용을 고려해야 한다.

셋째, ON9과 GALEN과 같이 형식적으로 개념모델링 되어진 의학용어 시스템을 재사용하여, 의학문헌 검색에 패턴 매칭이 아닌 표준

화된 개념 기반의 의학지식 제공이 가능한 정보검색 시스템 개발을 위한 연구가 이루어져야 할 것이다.

넷째, 의료정보시스템이 지능적이기 위해서는 의학 지식의 표현과 처리를 위한 추론 기술이 필수적이다. 따라서 의학지식 표현의 세부적인 방법 및 온톨로지 추론 시스템에 대한 연구가 함께 이루어져야 한다.

끝으로, 의학 온톨로지는 XML 기반의 온톨로지 언어가 개발되기 이전부터 개발되어 온 것으로 이러한 온톨로지 언어를 이용해서 기술하는 작업이 상대적으로 미흡하였다. 온톨로지 언어를 제공하고 있지 않거나 독자적인 언어를 사용해온 의학 온톨로지들은 최근 온톨로지 기술이 OWL 등으로 표준화됨에 따라 이러한 언어를 이용해서 새롭게 기술할 것으로 전망된다. 향후 의학지식을 표현함에 있어서 온톨로지 표준언어의 사용성에 대한 고찰과 이의 응용방안이 함께 연구되어야 할 것이다.

참 고 문 헌

- 김혜선. 2001. 통합의학언어 시스템(UMLS)의 구성 및 적용에 대한 고찰. 『情報管理研究』, 32(2): 26-39.
- 김흥기, 김명기. 2003. 의료정보학에서 온톨로지 기술. 『대한의료정보학회지』, 9(3): 213-219.
- Abasolo, Jose María and Mario Gómez, 2000. "MELISA. An ontology-based agent for information retrieval in medicine." [cited 2004. 7. 27] <<http://www.ics.forth.gr/isl/SemWeb/proceedings/session3-1/paper.pdf>>
- Brodie, M. L., J. Mylopoulos and J. W. Schmidt. 1984. *On Conceptual Modeling*. New York : Springer-Verlag.
- Burgun, A., G. Botti, M. Fieschi, and P. Le

- Beux. 1999. *IEEE International Conference On SMC'99*, 6: 300-305.
- Corcho, Oscar, Mariano Fernández-López, and Asunción Gómez Pérez. 2000. *OntoWeb Technical Roadmap v1.0*. IST Programme of the Commission of the European Communities as Project No. IST-2000-29243.
- Fensel, Dieter, Frank van Harmelen, Michel Klein, and Hans Akkermans. 2000. "On-To-Knowledge: Ontology-Based Tools for Knowledge Management." [cited 2003. 7. 27].
<<http://www.cs.vu.nl/~frankh/postscript/eBeW00.pdf>>
- Gangemi, Aldo, Domenico M. Pisanelli, and Geri Steve. 1999. "An Overview of the ONIONS project: Applying ontologies to the integration of medical terminologies." *Data & Knowledge Engineering*, 31: 183-220.
- Gruber, T. 1993. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5: 199-220.
- Guarino, Nicola. 1998. "Formal Ontology and Information Systems." In *Proceedings of FOIS'98, trento, Italy, 6-8 June, 1998*. IOS Press.
- Hahn, Udo, Martin Romacker, and Stefan Schulz. 1999. "How knowledge drives understanding-matching medical ontologies with the needs of medical language processing." *Artificial intelligence in medicine*, 15: 25-51.
- Ingenerf, Jose, Jörg Reiner, and Bettina Seik. 2001. "Standardized Terminological Services Enabling Semantic Interoperability Between Distributed and Heterogeneous Systems." *International Journal of Medical Informatics*, 64: 223-240.
- ISTC-CNR. "ON9." 1998. [cited 2003. 8. 10]
<<http://saussure.irmkant.rm.cnr.it/onto/>>
- Keizer, N. F., A. Abu-Hanna, and J. H. M. Zwetsloot-Schonk. 2000. "Understanding Terminological Systems I: Terminology and Typology". *Methods of Information In Medicine*, 39(1): 16-21.
- McCray, A. T., J. R. Scherrer, C. Safran, and C. G. Chute. 1995. "Concepts, Knowledge, and Language in Health-Care Information Systems." *Methods of Information in Medicine*, 34(1/2): 1-4.
- Mylopoulos, John. 1998. "Information modeling in the time of the revolution." *Information Systems*, 23(3/4): 127-155.
- Neches, Robert, R Fikes, T Finin, T Gruber, R Patil, T Senator, and W Swartout. 1991. "Enabling Tech-

- nology for Knowledge Sharing.” *AI Magazine*, 12(3): 36-56.
- NLM(National Library of Medicine). “MeSH” [cited 2004. 7. 22].
<<http://www.nlm.nih.gov/mesh/>>
- NLM(National Library of Medicine). “UMLS.” 2003. 2. [cited 2003. 7.15]
<<http://www.nlm.nih.gov/research/umls>>
- OpenGALEN. “GALEN.” [cited 2004. 7. 26].
<<http://www.opengalen.org>>
- Rogers, Jeremy, and Alan Rector 2000. “GALEN’s Model of Parts and Wholes: Experience and Comparisons.” *Proceedings American Medical Informatics Association Symposium*, 2000: 714-718.
- Rogers, J. E. and A. L. Rector. 1999. “Extended Core Model for Representation of the Common Reference Model for Procedures.” [cited 2004. 7. 26].
<<http://www.opengalen.org/open/crm/>>
- Stefano, Dessena, Moris Angelo Rossi and Galeazzi Elena. 1999. “Development of a cross-thesaurus with Internet-based refinement supported by UMLS.” *International Journal of Medical Informatics*, 54: 29-41.
- Steve, Geri, Aldo Gangemi, and Domenico M. Pisanlli 1997. “Integrating Medical Terminologies with ONIONS Methodology.” [cited 2004. 7.10].
<<http://www.loa-cnr.it/Papers/onions97.pdf>>
- Studer, Rudi, V. Richard Benjamins, and Dieter Fensel. 1998. “Knowledge Engineering : Principles and Methods.” *Data & Knowledge Engineering*, 25(1-2): 161-197.
- Wand, Yair. 1996. “Ontology as a foundation for meta-modelling and method engineering.” *Information and Software Technology*, 38: 281-187.
- W3C Web Ontology Working Group. 2004. 2. “OWL is a W3C Recommendation.” [cited 2004. 8. 10].
<<http://www.w3.org/2001/sw/WebOnt/>>