creative commons
COMMONS DEED

저작자표시-비영리-동일조건변경허락 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.
- 이차적 저작물을 작성할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

동일조건변경허락. 귀하가 이 저작물을 개작, 변형 또는 가공했을 경우에는, 이 저작물과 동일한 이용허락조건하에서만 배포할 수 있습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

의학석사 학위논문

통제자연어 진료서술문 작성규칙 표현을 위한 온톨로지 모델링 및 인터페이스 구현

An ontology model for representing writing rules in medical template and interfaces

2012년 8월

서울대학교 대학원

의학과 의료정보학 협동과정

이 경 민

A thesis of the Master's degree

An ontology model for representing writing rules in medical template and interfaces

통제자연어 진료서술문 작성규칙 표현을 위한 온톨로지 모델링 및 인터페이스 구현

August 2012

The Department of Medical Informatics,
Seoul National University
College of Medicine
Kyung Min Lee

통제자연어 진료서술문 작성규칙 표현을 위한 온톨로지 모델링 및 인터페이스 구현

지도교수 김 홍 기

이 논문을 의학석사 학위논문으로 제출함

2012 년 6 월

서울대학교 대학원

의학과 의료정보학 협동과정

이 경 민

이경민의 의학석사 학위논문을 인준함

2012 년 6 월

위 원 장 김 명 기 (인)

부위원장 김 홍 기 (인)

위 원 이 재 일 (인)

초 록

서론: 본 연구의 목적은 통제자연어 형태의 진료서술문을 표현하기 위한 온톨로지를 시멘틱웹 온톨로지 언어 OWL을 통해 모델링하고 진료서술문 작성 시 발생할 수 있는 규칙을 시멘틱웹 규칙 표현 언어인 SWRL을 통해 표현하고 이를 기반으로 인터페이스들을 구현 하는 것이다. 더 나아가 구현 된 인터페이스에서 올바른 통제자연어 입력을 위해 제공하는 용어 추천 목록의 순서를 다양한 관점을 산출한 용어의 사용 빈도를 고려하여 정하였다.

방법: 의료 서식 온톨로지는 서식 자체에 대한 정보인 Template attribute와 서식에서 사용될 수 있는 개념의 집합 인 Clinical entity로 분류하여 모델링 하였다. Template attribute는 서식의 제목과 설명을 포함하여 규칙에서 사용할 수 있는 선택 사항 Template option을 표현한다. 예를 들어 대장암 병리보고서 작성 시 기술 하고자 하는 검체의 종류에 따라 진료서술문 기재사항의 내용이 변경되므로 검체의 종류는 Template option으로 표현 될 수 있다. Clinical entity는 의무 기록이 표현하고자 하는 대상을 묘사하기 위한 개념과 기재 사항을 표현한다. 의무 기록 작성 규칙은 서식의 선택 사항이나 특정 기재 사항의 값에 따라 다른 기재 사항 값의 범위가 변경 되는 것을 표현하였으며, 기재 사항의 값 범위 변경 시 새로운 값의 범위로 변경하는 'Replace', 혹은 기존의 값의 범위에 새로운 값들을 추가하는 'Merge'의 두 가지 방법으로 변경한다. 또한, 용어추천 시 용어자체, 의무기록자, 환자의 3가지 관점에서 사용 빈도를 산출하여 최근 많이 사용한 용어 일수록 높은 점수를 부여한다.

결과: OWL과 SWRL로 의무기록 서식 온톨로지와 규칙을 저작할 수 있는 저작 인터페이스와 저작 된 결과물을 활용하여 구조적인 진료서술문을 작성할 수 있는 통제자연어 진료서술문 작성 인터페이스로 구성 된 Owlpath를 구현하였다. 이 중, 작성 인터페이스는 규정된 문법과 의무기록 서식 온톨로지 및 규칙을 통해 사용자가 입력할 수 있는 용어를 추천하고 상이한 점이 있을시 이를 알려주어 올바른 통제자연어 진료서술문 작성을 유도한다. 작성 인터페이스로 작성 된 진료서술문은 RDF로 출력 하였다.

결론: 구현 된 Owlpath 인터페이스들을 활용하여 국내외에서 제안 된 병리보고서 기재사항 표준안 및 웹 폼 형태의 의료 서식들의 온톨로지 및 작성 규칙을 입력하고, 이를 통해 통제자연어 진료서술문을 작성한 뒤, RDF로 추출함으로써 제안 된 시멘틱 웹 기술을 이용한 의료 서식 및 규칙 표현의 의료현장 사용 가능성을 보여 주었다.

* 본 내용은 의료정보학회(Kyungmin Lee, Hong-Gee Kim, Jeong-Suk Yoo, An ontology model for representing writing rules in medical template and CNL interface, Proceedings of KOSMI, 2012; 519)에 출판 완료된 내용임.

주요어: 온톨로지, 시멘틱 웹, 규칙언어, 통제자연어, 의료서식

학 번: 2010 - 23750

목 차

초록 ·····i

목차 ·····iii

표 및 그림 목록 ·····iv

서론 ·····1

1. 연구의 필요성 ·····1
2. 시멘틱 웹 기술 기반의 진료서술문 표현 ·····5
3. 통제자연어 ·····8

연구재료 및 방법 ·····14

1. 예시로 활용한 의료서식 ·····14
2. 의료 서식 온톨로지 모델링 ·····16
3. 의무기록 작성 규칙 표현 ·····22
4. 다각도 용어 사용 빈도 ·····27

결과 ·····29

1. Owlpath 통제자연어 진료서술문 문법 ·····29
2. Owlpath 의료서식 저작 및 통제자연어 진료서술문 입력 인터페이스 ·····33
3. 의료서식 입력 및 진료서술문 작성 결과 ·····44

고찰 ·····54

참고문헌 ·····57

초록 (영문) ·····60

표 및 그림 목록

표 2.1 대장암 병리보고서 표준기재사항 절제 대장 판독용 양식15
표 2.2 개념 종류별 OWL 구문 예시(Turtle 방식) ···19
표 2.3 SWRL XML로 표현한 진료서술문 작성 규칙 26
표 3.1 형태에 따른 대장암 병리보고서 ·····31
표 3.2 다각도 용어 사용빈도 측정을 위한 SQL 구문 43
표 3.3 통제자연어 진료서술문 예시 ·····47

그림 1.1 진료서술문 표현을 위한 RDFS 및 RDF ·····7
그림 1.2 ACE 통제자연어 작성 에디터 웹 버전 ·····10
그림 1.3 BNF형식의 Ontopath 통제자연어 문법 ·····11
그림 1.4 의무기록 입력형태 별 인터페이스 ·····12
그림 2.1 Owlpath의 의료서식 온톨로지 ·····18
그림 2.2 통제자연어 진료서술문 RDF diagram ·····21
그림 2.3 진료서술문 작성규칙 예시 ·····24
그림 3.1 BNF 형태의 Owlpath 통제자연어 문법 ·····30
그림 3.2 통제자연어 진료서술문의 parsing tree ·····32
그림 3.3 Owlpath 시스템 아키텍처 ·····34
그림 3.4 의료서식 온톨로지/규칙 저작 인터페이스 ···36

그림 3.5 통제자연어 진료서술문 작성 인터페이스 ……39
그림 3.6 통제자연어 작성 인터페이스 다이얼로그 ……41
그림 3.7 용어 사용빈도 측정 ER-Diagram ……42
그림 3.8 의료서식 온톨로지 및 작성규칙 결과 ……46
그림 3.9 대장암 병리보고서 RDF diagram ……52

서 론

1. 연구의 필요성

의무기록이란 환자의 질병에 관계하는 정보와 병원이 진단과 치료를 위해 시행한 처치내역, 병력, 검사결과, 진술문, 입/퇴원 기록, 수술기록 등 모든 내용을 기록한 법적 문서이다. 종이로 작성하는 전통적인 의무기록 방식인 하드카피 방식이 가진 한계인 서식의 불일치, 자료 분실의 위험, 중복기록 및 열람의 시공간적 제약 등을 극복하고, 의무기록을 효과적으로 관리하기위해 전자의무기록(EMR)의 필요성이 강조되고 있다(1). EMR은 앞서 언급한 하드카피의 단점들을 보완함과 동시에 자료의 올바른 입력여부를 확인 할 수 있다는 자료의 정확성(1)과 자료 통계 및 의사결정 등을 위한 연구용 자료로의 활용이 용이 하다는 재사용성 그리고 환자를 비롯하여 해당 정보를 필요로 하는 모든 사람이 동시에 열람할 수 있다는 가용성에서 장점을 갖는다(2).

또한 인건비 감소, 의료비 절약과 의료업무의 효율화 등의 경제적 효과도 기대되고 있다(3). 이와 같은 장점들로 인해 미국의 EMR 도입률이 지속적으로 증가하여 미국 내 63.1%의 응급 진료 병원(Acute care hospital)를 조사한 결과 2009년 말 26.1%의 의료 기관이 EMR을 도입하였다(4). 국내에서도 2002년 의료법 개정이 개정됨에 따라 2003년 S대 병원의 종이 없는 병원을 시작으로(5), EMR 도입률이 꾸준히 증가하여 2010년 종합병원급 이상 의료기관이 37.2%, 3차 의료기관 50.2%에서 도입하였다(6).

이와 같은 EMR 사용률의 증가와 발맞춰 전산화 된 의무기록을 효과적으로 관리하기 위한 표준화가 제정되어왔다. 이를 위해 고려해야 할 EMR 문서의 특징은 구조의 복잡성, 키, 혈당과 같은 숫자, 코드화 된 텍스트, 날짜, 시간, 기간 등 형태의 다양성과 코딩시스템의 의미적 제한성, 다양한 관점의 사용자가 존재한다는 관점의 다양성 그리고 의료정보가 환자의 증세나 질병에 따라 다르게 통합될 수 있다는 통합의 다양성이 있다(1). 이러한 특징들을 해결 하고자 해외에서는 HL7의 CDA(7), OpenEHR(8) 등 의료 서식에 관련 된 표준이 개발 되어 왔으며, 국내에서도 서식모델을 지원하는 임상콘텐츠 모델(Clinical Content Model)을 개발하여 의료 문서서식 표준안 관련 연구를 수행하였다(9).

그러나 국내외 증가된 EMR 도입율과 의료 문서 표준화의 노력에도 불구하고, EMR 시스템이 포괄적으로 적용되는 사례는 드물다. 미국 내 EMR을 도입한 26.1%의 Acute care hospital 중 1.5%만이 포괄적인 EMR 시스템을 가지고 있고 기초적인 시스템을 운영하고 있는 하고 있는 병원도 7.6%에 그치고 있다(4). 국내에서도 종합 전자건강기록 시스템을 기준으로 평가할 때, 미국보다 높은 도입율을 보이지만 전체 도입률이 5%에 머물고 있는 실정이며(6), 일부 의료종사자는 기존의 하드카피 보다 업무효율이 감소하거나 입력된 자료의 질이 저하될 수 있다고 우려하고 있다(9). 이처럼 포괄적인 EMR시스템 적용률이 낮은 이유는 새로운 시스템을 구입해야 한다는 금전적인 문제를 제외하더라도, 환자 데이터를 포함하는 의무기록이 전산망에서 사용 됨에 따라 개인 정보 보호 요구가 높아진다는 점(10)과 이에 따른 보안유지문제(11), 그리고 컴퓨터 안전성 자체가 진료에 많은 영향을 미칠 수 있다는 점(3) 등 하드카피 방식의 기존 의무기록에서 대두되지

않았던 문제점들을 고려해야하기 때문이다. 그러나 더욱 근본적인 문제는 서술문 방식의 하드카피에 익숙한 의료 종사자들이 구조적 형태의 의무기록 입력방식에 익숙하지 않아 짧은 시간 내에 종이 의무기록지에 작성하듯 자유롭게 진료내용을 작성하기 어렵다는 점이다(9).

의무기록의 형태는 의사, 의무기록사 등을 포함한 의료 종사자에게 익숙한 자연어 형태인 진료 서술문(free text or medical narrative)과 지능적 관리가 용이한 구조적 형태(structured data)로 크게 나뉘며, Sharer에 따르면 전체 의무기록의 70%(2)를 진료서술문이 차지하고 있다. 그러나 자연어 형태의 정보는 수많은 언어를 저장 할 수 있는 데이터베이스가 필요하여 처리가 느리고, 입력 시 오류가 발생할 수 있다는 다양성과 오류로 인한 의미의 모호성으로 인해(1), EMR 시스템은 주로 구조적 형태의 의무기록 입력방식을 제공하는 방향으로 개발되어 왔다. 이에 따라 의료종사자에게 친숙한 입력 방식을 갖는 EMR 문서 입력 시스템의 필요성이 대두되어 자연어와 유사한 형태의 입력방식을 가지며, 입력된 자료를 구조적으로 관리할 수 있는 온톨로지 기반 통제자연어 진료서술문 입력 인터페이스 Ontopath가 개발되었다. 해당 시스템을 통해 의료종사자들에게 익숙한 진료서술문 형태로 구조적인 EMR 문서를 작성할 수 있는 가능성이 제시되었다(2).

하지만 Ontopath를 의료 현장에 적용하기에는 정보 표현을 위해 1) 시멘틱웹 온톨로지 언어 RDFS(Resource description framework scheme) (12)를 기반으로 모델링하여 의료진술문 작성 시 필요한 기재사항에 대한 제약사항 표현에 어려움이 있었으며, 2) 의료 문서 작성 시 발생할 수 있는 의미적 규칙을 표현할 수 없다는 부족함이 있었다. 또한, 통제자연어 진료서술문 작성도구에서 3) 추천되는 용어의 목록의 순서가 고정되어 있어 개인화 된 추천 기능을 제공하기

어려웠다. 이에 따라, 본 논문에서는 1) RDFS보다 표현력이 좋은 시멘틱웹 온톨로지언어 OWL(Web ontology language)을 기반으로 모델링을 수행하여 의료현장에 존재하는 개념과 개념 간의 관계 및 제약사항을 반영할 수 있도록 하였으며, 2) 의료 문서 작성 시 발생할 수 있는 의미 종속관계를 시멘틱웹 규칙 언어인 SWRL(13)로 표현하였다. 더 나아가, 3) 사용자의 용어 사용 빈도를 측정하여 사용성을 고려한 용어 추천 방법을 제안하며 4) 끝으로 3가지 사항을 반영한 온톨로지 저작 및 진료서술문 작성 인터페이스 Owlpath를 구현한다.

2. 시멘틱 웹 기술 기반의 진료서술문 표현

본 절에서는 시멘틱 웹의 온톨로지 표현방법을 설명하고 이를 이용하여 의무기록을 표현하는 방법을 설명한다. 시멘틱 웹의 대표적인 온톨로지 표현방법은 RDFS를 사용하는 방법과 이보다 표현력이 뛰어난 OWL을(14) 사용하는 방법이 있다.

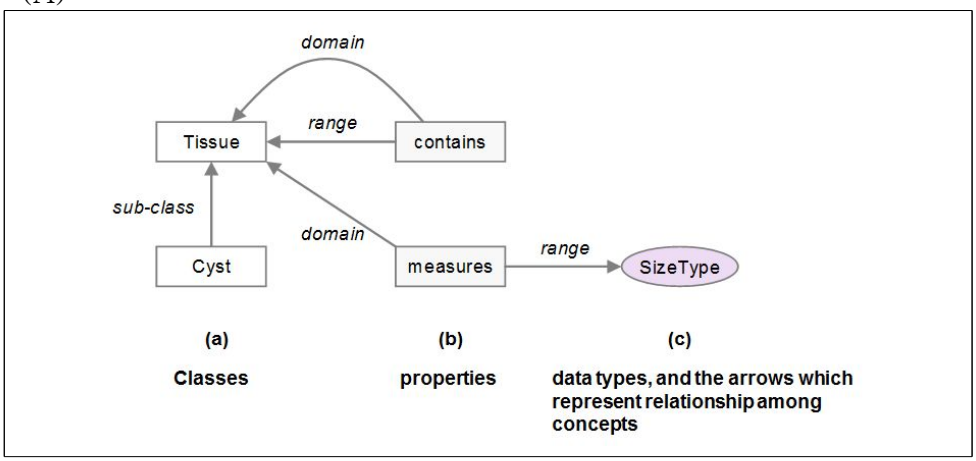
RDFS에는 클래스(class)와 속성(property)의 개념(concept)이 존재한다. class는 실세계에 존재하는 자원(resource)의 추상화 된 concept인 원형(type)을 나타내며, property를 통해 class의 속성을 나타냄으로써 class간의 관계를 표현한다. 이 때 property로 묘사하고자 하는 class인 정의역을 domain이라고 하고 property의 값을 뜻하는 공역을 range라는 개념을 사용하여 이를 제한 한다(15). 또한 추상화 되지 않는 property의 값 즉, 숫자 표현, 임의의 문자열과 같은 정보는 Datatype으로 표현한다. 이와 같이 RDFS를 통해 정의 된 온톨로지는 Instance들의 집합인 RDF(Resource description framework)를 통해 표현이 가능하다.

RDFS를 통해 “Specimen received contains a cyst which measures 2×1cm.” 라는 진료서술문 문장의 서식을 온톨로지로 표현하면 “specimen received”는 class인 Tissue로 “a cyst”는 Cyst class로 각각 추상화 된다. 또한 모든 cyst는 tissue이므로 둘의 관계를 고려하여 class의 계층구조 표현이 가능하다(그림 1.1A). 그리고 해당 문장에서 class를 연결함으로써 property 역할을 하는 “contains”와 “measures”의 domain과 range가 각각 Tissue class와 Datatype의 형태로 표시된다. 정의 된 온톨로지를 사용하면 위 문장을 RDF로 표현할 수 있다(그림 1.1B).

이와 같이 시멘틱웹 온톨로지 언어인 RDFS를 사용하여 진료서술문 형태의 의무기록 서식을 표현할 수 있으며, RDF를 통해 진료서술문을 표현함으로써 자료의 재사용성을 증대시킨다.

본 논문에서는 RDFS와 비슷하게 개념을 추상화 하고 제한사항을 두어 더 풍부한 표현력을 가지고 있으며, 규칙 표현 언어인 SWRL(Semantic web rule language)(13)과 연동이 가능한 시멘틱웹 온톨로지 표현방식 OWL을 사용하여 의무기록 서식을 표현한다.

(A)



(B)

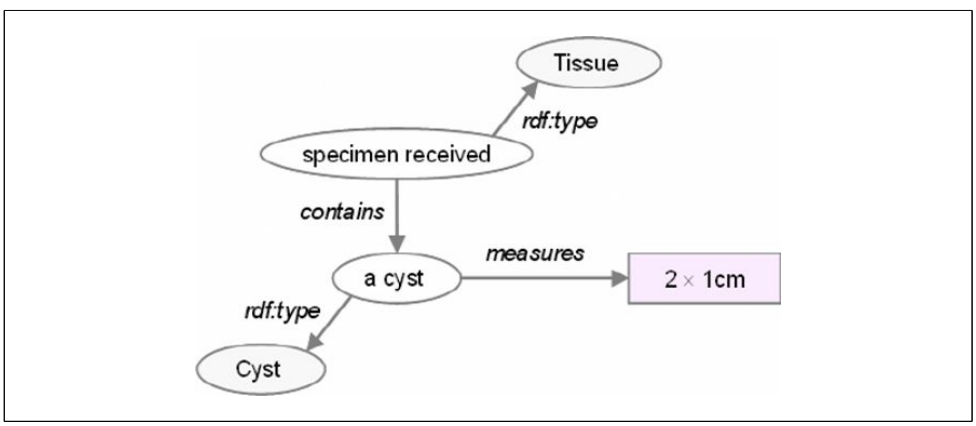


그림 1.1. 진료서술문 표현을 위한 RDFS 및 RDF

(A) RDFS를 통해 진료서술문 의료서식을 표현한 것으로 간선은 domain과 range 및 sub-class 관계를 나타내며, 타원은 data type을 직사각형은 class를 나타낸다.

(B) RDF를 통해 진료서술문을 표현한 것으로 concept은 타원으로 predicate는 간선으로 그리고 data type은 직사각형으로 표현된다. 간선의 시작점은 문장의 주어인 Subject를 나타내며 끝점은 목적어인 Object를 나타낸다.

3. 통제자연어

통제자연어(Contolled natural language)는 문법 및 어휘를 제한함으로써 자연어의 모호성 및 복잡성을 줄인 자연어의 부분집합이다. 통제자연어는 가독성을 향상시키기 위하여 짧은 문장을 사용하고 대명사 사용을 피하는 등의 제한을 두는 방식(16)과, 기계의 자동적인 언어처리를 수월하게 하기 위해 일차 술어논리와 같이 기존에 존재하는 형식 언어에 사상(Mapping) 가능하도록 형식 논리를 기초로 하는 방식이 존재한다. 이와 같은 통제자연어는 자연어의 친숙성과 형식언어의 자동 언어처리 수월성을 동시에 가진다는 장점을 지닌다.

대표적인 통제자연어는 ACE(Attempto Controlled English)(17)이며, 그 밖에 CLCE(Common Logic Controlled English)(18), Rabbit(19), PENG(Processable ENGlish)(20) 등이 존재하고 주로 기계, 전자장치의 매뉴얼 또는 안내표지판 등에서 사용자의 쉽고 명확한 이해를 돕기 위해 사용되어 왔다.

ACE는 스위스의 Zurich 대학의 연구 프로젝트인 Attemto에서 개발된 통제자연어로써 지식 표현 언어로 활용되기 위해 디자인 된 표준 영어의 부분집합이다. ACE의 문법에 맞춰 작성 된 문장은 구조적 담화 표현인 DRS(discourse representation structure) 혹은 온톨로지 언어의 출력형태인 OWL/XML 로 변환 할 수 있으며, ACE 통제자연어 작성방식에 따라 사용자가 쉽게 문장을 입력 할 수 있도록 입력 인터페이스 ACEeditor를 제공한다(그림 1.2).

CLCE는 Sowa에서 개발 된 통제자연어로 영어에 가까운 문법을 지니고 있으며, ACE보다 뛰어난 표현력을 가진다. FOL(First order logic)의 모든 연산자 및 한정사의 표현을 지원하며 자동으로 기존에 작성 된

narrative(서술문) 들을 logic, SQL, UML 등 으로 바꿀 수 있다. 또 다른 특징은 관계형 데이터베이스에 연결하여 in-line 형식의 단어정의를 허락한다는 점이다.

PENG-D는 기계처리 가능한 통제자연어로 제안 되었으며, 이는 OWL과 함께 결정 가능한 형식 논리적 문장으로 바꿀 수 있다. 그 외 온톨로지 저작을 도와주는 통제자연어 Rabbit과 RNLS(Restricted natural language statement)과 같은 통제자연어들이 존재한다.

하지만 이와 같은 통제자연어들은 영어 전반에 대한 통제자연어를 목표로 하고 있기 때문에 의료지식이 아닌 개개의 사실들을 구체적으로 기술하기 위한 목적으로 사용되며, 특정 분야에 만들어져 있는 온톨로지에 대한 고려가 없으며(2), 도메인 지식과 함께 사용하기 어렵다(17). 이에 따라 class, property와 같은 RDFS에 정의 된 개념을 활용하는 통제자연어 DESC가 개발 되었다(2)(그림 1.3). 이를 활용하면 입력방식을 구조적인 형태에서(그림 1.4A) 자연어에 가까운 형태로(그림 1.4B) 변경 할 수 있다 장점을 가지며, EMR시스템이 요구하는 형식에 맞춰 의무기록을 구조적으로 제공할 수 있다는 장점을 갖는다.

본 논문에서는 RDFS 기반의 통제자연어 문법 DESC와 유사한 OWL 기반의 통제자연어 문법을 사용하며, 이와 더불어 진료서술문 작성규칙을 SWRL로 표현하여 진료서술문 작성 시 문법과 함께 사용함으로써 좀 더 정교한 통제자연어 진료서술문 작성 가이드가 가능하도록 한다.

그림 1.2. ACE 통제자연어 작성 에디터 웹 버전

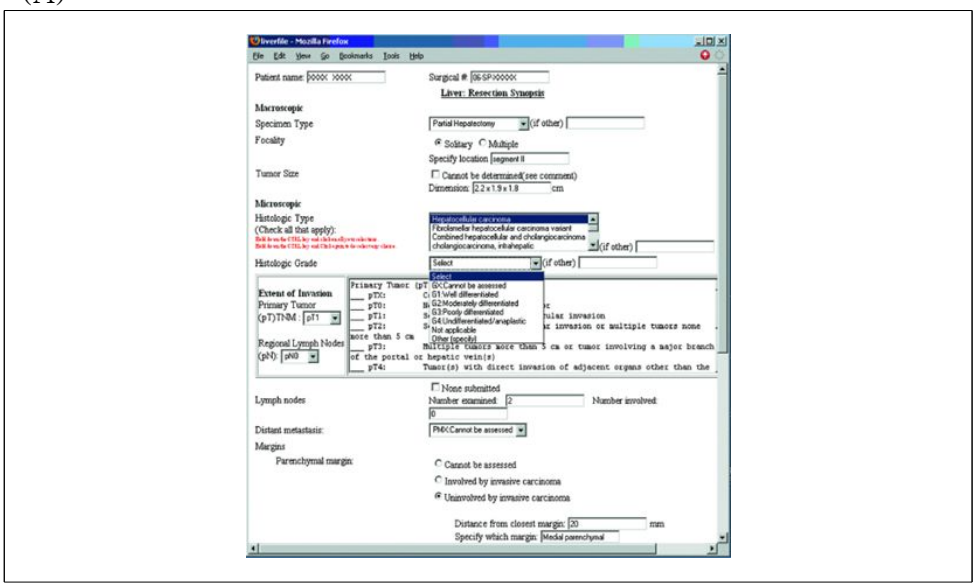
“Berlin smiles above at most 12 airlines against an apple.” 는 통제자연어로 작성된 문장이며, 각 문장은 Syntax Boxes를 통해 구분되는 모습을 볼 수 있으며 정확히 작성된 ACE 통제자연어 문장은 세 번째 박스같이 DRS로 출력된다. 하지만 문법의 정확성만을 고려하기 때문에 의미적으로 상충하거나 잘못된 내용이 작성되더라도 처리할 수 없다.

```
Sentence ::= Phrase { Conjunction Phrase } ". “
  Phrase ::= NounPhrase VerbPhrase { Conjunction VerbPhrase }
      | Instance TypePredicate class { Conjunction VerbPhrase }
  NounPhrase ::= Subject | Subject RelativePronoun VerbPhrase
      | Instance RelativePronoun TypePredicate class
  VerbPhrase ::= Predicate ObjectPhrase { Conjunction ObjectPhrase }
  ObjectPhrase ::= Object | Object RelativePronoun VerbPhrase
      | Instance RelativePronoun TypePredicate class
      | Literal
  Subject        ::= Instance | Pronoun | Definiteclass
  Predicate      ::= property
  Object          ::= class | Instance | Pronoun | Definiteclass
  Definiteclass ::= Article class
  TypePredicate ::= "is"
  Article          ::= "the"
  Pronoun          ::= "it"
  Conjunction    ::= "and"
  RelativePronoun ::= "which"
```

그림 1.3. BNF형식의 Ontopath 통제자연어 문법

Ontopath에서 사용 된 DESC는 terminal과 non-terminal의 형태로 CFG(Context Free Grammar) (2, 21)를 이용하여 통제자연어 문법을 정의하며 이는 BNF(Backus Naur Form)으로 표현가능하다.

(A)



(B)

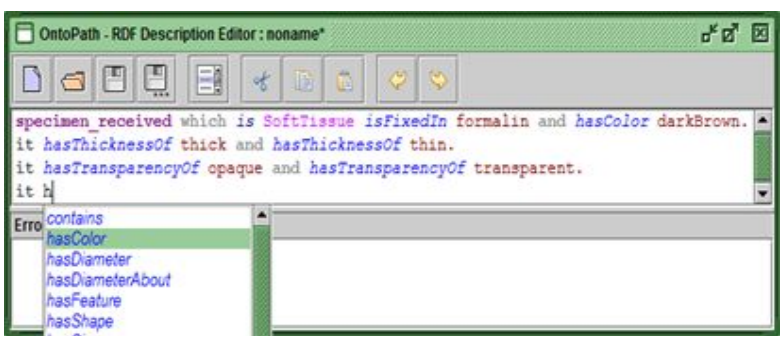


그림 1.4. 의무기록 입력형태 별 인터페이스

(A) 미국의 Rochester에 대학에서 제안 된 웹 폼 형태의 구조적인 Tumor reporting system의 예로서 절제된 폐에 대한 병리보고이다.

(B) 통제자연어를 활용한 입력방식을 사용하는 인터페이스 Ontopath의 화면이다. 각 구문이 문법과 온톨로지에 의해 분석되며 올바른 통제자연어 진료서술문 작성을 위해 필요한 용어 목록을 보여준다.

연구 목적

1. 규칙 표현에 유용한 의무기록 서식 온톨로지를 모델링하고 이를 시멘틱웹 온톨로지 언어인 OWL로 표현한다.
2. 진료서술문 작성을 위한 의미 종속성을 시멘틱 웹 규칙언어인 SWRL로 표현한다.
3. 온톨로지와 규칙에 근거한 추천 용어 목록의 순서를 사용자 용어 사용 빈도를 고려하여 갱신한다.
4. 위 3가지 사항을 반영하여 의무기록 서식 온톨로지 및 규칙을 저작하고 진료서술문을 작성할 수 있는 인터페이스 OwlPath를 구현한다.

연구재료 및 방법

본 장에서는 논문에서 활용하는 의료 서식을 설명하고, 이를 예시로 활용하여 통제자연어 진료서술문을 표현하는 온톨로지와 작성 규칙 표현에 적용한 OWL과 SWRL에 대하여 설명한다. 또한 사용성을 향상시키기 위해 고안한 용어추천 목록의 정렬 방법을 설명한다.

1. 예시로 활용 된 의료 서식

본 논문에서 예시로 활용하는 의무 기록 서식은 대한병리학회 소화기병리학 연구회에서 2006년 개발 된 “대장암 병리보고서 기재사항 표준화”(22)로써 국내 대형병원 및 일본, 미국 등에서 사용되는 대장암 병리보고서 서식을 참고하여 작성된 의무 기록 서식 표준안이다.

본 표준안의 구성은 모든 병원이 기재해야하는 암 진단에 필요한 최소 기본사항인 “표준 기재사항”과 병원 별로 선택하여 기재할 수 있는 “선택 기재사항” 그리고 “내시경 생검의 표준 기재사항”으로 되어 있으며, 병원에 따라 선택 기재사항 이 외의 항목을 추가하여 사용할 수 있다. 또한, 기재사항 별로 선택 될 수 있는 값의 범위가 지정되어 있다.

본 장에서는 표준안의 양식 중 절제 대장의 표준 기재사항 양식을 예시로 활용하여 온톨로지 모델링 및 작성규칙에 대해 설명한다(표 2.1).

해당 양식이 설명하고자 하는 검체 colon과 rectum은 SNOMED-CT의 “body structure” attribute를 기준으로 하여 “Large intestine part”로 개념화 하였다(23).

기재사항	차수	분류법
Specimen type	1	국내 대형병원
Main diagnosis	1	국내 대형병원
Histologic type	1	WHO(2000)
Location	1	ICD-O
Gross type	1	일본 육안 분류법
Tumor size	1	국내 대형병원
Depth of invasion	1~2	AJCC/UICC[1)]
Resection margin	1	국내 대형병원
Regional lymph node metastasis	0~1	미국 병리학회 프로토콜
Lymphatic invasion	1	국내 대형병원
Venous invasion	1	국내 대형병원
Perineural invasion	1	국내 대형병원
Pre existing adenoma	1	국내 대형병원
Associated findings	0~1	국내 대형병원
Separate-lesions	0~1	국내 대형병원

표 2.1. 대장암 병리보고서 표준기재사항 절제 대장 판독용 양식

표 항목 중 기재사항은 해당 표준안에서 작성되는 항목 들이며 차수는 출현 가능 횟수를 뜻한다. 0회 인 경우는 상황에 따라 작성여부가 결정된다는 것을 뜻한다. 표준안 내에서 분류법이 명시적으로 나타나지 않은 것은 국내 대형병원에서 사용하는 병리보고서 양식을 참고한 것으로 간주하였다.

1) AJCC/UICC stands for American Joint Cancer Committee/Union Internationale Contre le Cancer

2. 의료 서식 온톨로지 모델링

의료 서식 온톨로지는 개체 그리고 개체 간 관계인 속성과 속성에 대한 제한으로 표현이 가능하다. 본 연구에서는 웹상에서 사용되는 온톨로지 언어 OWL을 사용하여 의료 문서 서식을 표현할 수 있는 용어체계를 구축하고 이에 기반 하여 진료서술문을 구조적으로 저장한다.

의료 서식에 표현 된 Clinical entity(의료 기록 내용)들은 묘사하거나 묘사에 사용되는 개념을 표현하는 개체와 개체 간 관계를 설정하는 기재사항으로 구성된다. 개체는 OWL에서 entity로 표현되며 entity는 다시 개별 사실을 표현하는 individual과 individual들을 추상화한 집합인 class로 표현되며, 측정값을 묘사하는 수치나 개념화 되지 않은 문자열은 data type으로 표현된다. 예를 들어, colon, rectum 같은 대상이 개별사실로 표현 된다고 할 때 각각은 individual로 표현되며 large intestine part class로 추상화 된다. 또한, 대상이 size같은 속성을 가진다고 할 때 “[0-9]*x[0-9]*x[0-9]*cm” 과 같이 정규식 문자열로 data type을 정의한다.

기재사항은 OWL에서 개체 간의 관계를 나타내는 property로 표현될 수 있으며, 서식에서 묘사되는 대상 인 정의역과 묘사를 위해 사용하는 대상 인 공역을 각각 property의 domain과 range로 설정함으로써 표현할 수 있다. 또한 property에 cardinality를 설정하여 의무 기록에서 가능한 출현 횟수를 표현 할 수 있다. 이 때 property의 range가 class인 경우 이를 Object property라고 하며 data type을 range로 가진다면 Data property라고 한다. 이와 같이 각 항목을 표현하고 이를 통해 용어 체계 온톨로지를 구축하였다.

Clinical entity와 더불어 서식에 대한 정보인 Attribute를 제목과 간략한 설명으로 정의하고, 추론에 영향을 미치지 않고 클래스에 부가적인

정보를 추가하는 annotation인 rdfs:label[2]과 rdfs:description[3]을 통해 OWL로 표현하였다. 또한 “hasOption”이라는 기재사항을 두어 property 항목에 속하지 않지만 다른 property들의 range class member를 변경 할 수 있는 선택사항들을 반영할 수 있도록 하였다. 예를 들어 “Endoscopic mucosal resection 검체에서만 사용하는 기재사항의 값이 존재한다.”라는 규칙은 대장암 병리보고서 표준안의 기재사항 외에 존재하는 검체의 종류에 대한 선택사항에 해당하는 것이므로 “hasOption” property의 range로 정의하여 이와 연관되는 property의 range class에 속한 멤버들을 변경할 수 있다.

이와 같은 사항들을 반영하여 의료서식에 대한 정보 및 기재사항을 표현하는 온톨로지를 구축하였으며(그림 2.1), OWL로 표현하였다(표 2.2).

통제자연어로 작성 된 진료서술문 문장 “Colon which hasSize 2.5x1.5x1.5cm hasGrossType fungating.”라는 종양의 크기와 육안형태 대해 설명하는 진료서술문 문장을 RDF Subject-Predicate-Object로 기술하면 “Colon”은 Subject로 “hasSize”와 “hasGrossType”은 Predicate로 그리고 “2.5x1.5x1.5cm”와 “fungating”은 Object로 표현된다. 또한 RDF로 기술 된 사항은 RDF diagram으로 도식화 하여 표현할 수 있다(그림 2.2).

본 절에서 설명한 바와 같이 OWL을 사용하여 의료서식을 온톨로지로 표현하는 것이 가능하며, OWL은 특정 분야의 개념, 속성, 그리고 속성의 제한사한 사항 등을 나타내는 체계이므로 이를 바탕으로 수집한 정보를 다양한 방식으로 재사용할 수 있다는 가능성을 보장한다. 또한 표현된 서식의 내용에 맞게 작성 된 진료서술문은 의미적 기계처리가 가능한 형태인 RDF로 출력 될 수 있다.

---

2) rdfs:label is <http://www.w3.org/2000/01/rdf-schema#label>
3) rdfs:comment is <http://www.w3.org/2000/01/rdf-schema#comment>

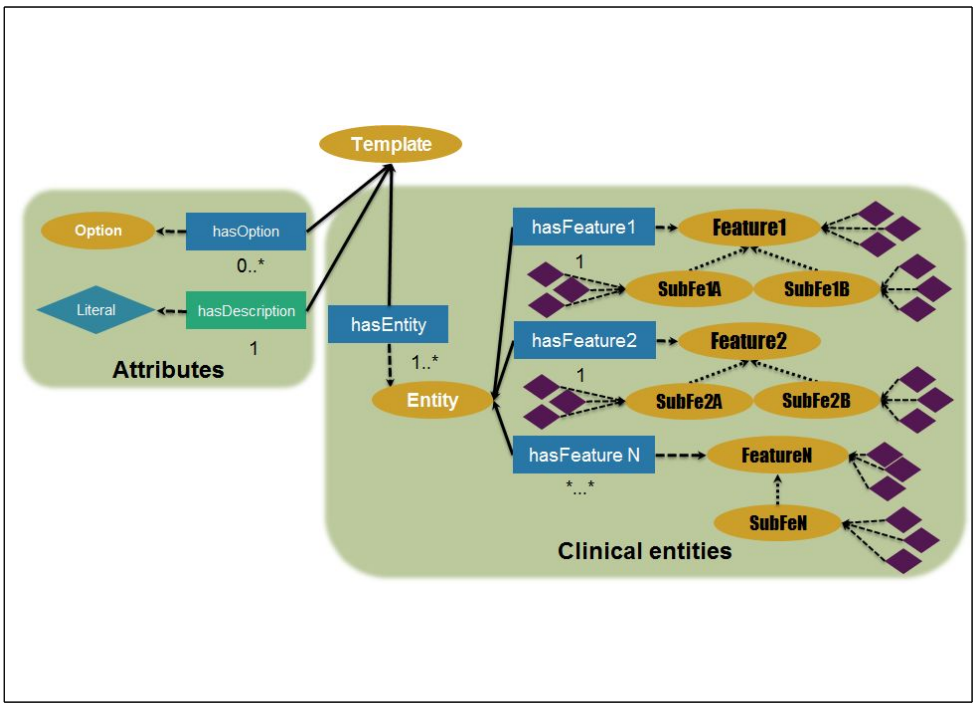


그림 2.1. Owlpath의 의료서식 온톨로지

의료서식 온톨로지는 크게 Attribute와 Clinical entity 로 구성된다. 타원은 class를 진한색의 마름모는 individual을 그리고 직사각형은 property를 나타내며 property에 연결된 화살표 중 실선은 domain, 점선은 range를 나타낸다. 또한 class에 연결 된 점선 화살표의 시작점과 도착점은 각각 하위, 상위 class를 나타낸다. property 아래에 표기 된 “x...y” 는 property의 cardinality가 “최소 x개에서 최대 y개” 표현 될 수 있다는 것을 뜻하며 이 때, 기호 “*” 는 어떠한 숫자도 들어갈 수 있다는 것을 뜻한다.

종류	OWL 구문
class	:template rdf:type OWL:class ; rdfs:label "Colorectal cancer"^^xsd:string ; rdfs:subclassOf OWL:Thing ; rdfs:comment "This is a standardized template"^^xsd:string.
individuals	:ascending_colon rdf:type :location , OWL:Namedindividual .
Object Properties	:hasAssociatedFindings rdf:type OWL:Objectproperty ; rdfs:domain :large_intestine_part ; rdfs:range[rdf:type OWL:Restriction ; OWL:onproperty :hasAssociatedFindings ; OWL:onclass :associated_findings ; OWL:minQualifiedCardinality "0"^^xsd:nonNegativeInteger] , [rdf:type OWL:Restriction ; OWL:onproperty :hasAssociatedFindings ; OWL:onclass :associated_findings ; OWL:maxQualifiedCardinality "1"^^xsd:nonNegativeInteger] .
Data properties	:hasSize rdf:type OWL:Datatypeproperty ; rdfs:domain :large_intestine_part ; rdfs:range :size .
Datatypes	:size rdf:type rdfs:Datatype ; OWL:equivalentclass [rdf:type rdfs:Datatype ; OWL:onDatatype :size ; OWL:withRestrictions ([xsd:pattern "[0-9]*X[0-9]*X[0-9]*cm"^^xsd:string])] .

표 2.2. 개념 종류별 OWL 구문 예시(Turtle 방식)

예시에서 class는 “template”을 나타내며 label과 comment을 통해 제목과 설명이 작성되었다. OWL의 top-class는 “thing”이므로 OwlPath의 최상 위 class들은 모두 “thing”의 sub-class로 구성 된다. individual은 “location” class에 속하며 “Namedindividual”로 저장된다. property는 cardinality를 restriction으로 가지며 최소를 “minQualifiedCardinalitiy”, 최대를 “maxQualifiedCardinalitiy”, cardinality가 하나 만 존재하는 경우는 “exactQualifiered Cardinality”로 표현한다. 또한 Data property의 range로 설정되는 Datatype은 OWL의 “equivalentclass”를 통해 restriction을 설정하여 값을 제한한다. 예시의 Datatype은 “size”를 “㎤”로 나타낸다.

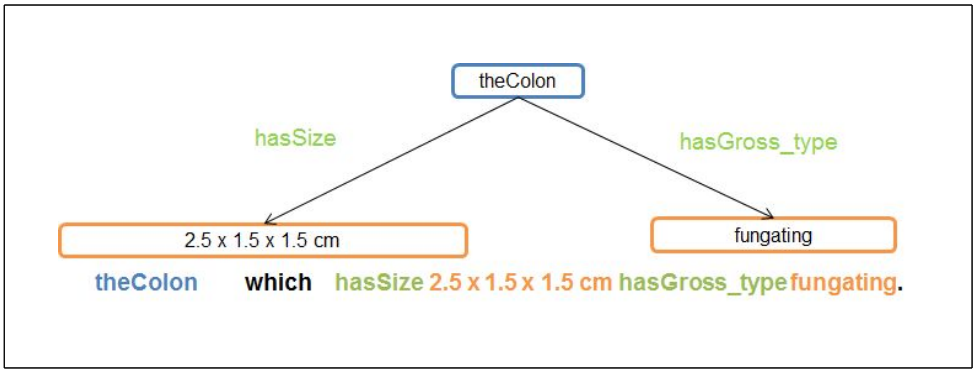


그림 2.2. 통제자연어 진료서술문 RDF diagram

화살표는 Predicate를 나타내며, 화살표의 시작점인 “theColon”은 Subject를, 화살표의 끝점인 “2.5x1.5x1.5cm”과 “fungating”는 Object를 표현함으로써 통제자연어로 작성된 문장을 RDF diagram으로 도식화 할 수 있다.

3. 의무기록 작성 규칙 표현

의무 기록 작성 시 발생할 수 있는 규칙은 시멘틱웹 규칙 언어인 SWRL의 선행조건(antecedent)과 결과(consequent)로 표현 가능하며 논리식으로 표현 된다(식 1).

$$antecedent \rightarrow consequent \quad \text{-------} \quad (1)$$

antecedent와 consequent는 공리(axiom)들의 conjunction들로 이루어지며, antecedent가 참이라면 consequent도 반드시 참이다. SWRL의 axiom은 class와 property를 표현할 수 있으며 class axiom를 명시하기 위하여 "class(x)"과 같이 표기하며 x는 individual이거나 variable 일 수 있다. 또한 property에 대한 표현으로 "x가 property의 값으로 y를 갖는다."는 axiom이 있을 경우 이를 표현하면 "property(x,y)"과 같다.

SWRL을 통해 표현 하고자 하는 통제자연어 진료서술문의 작성 규칙은 크게 1) 의료서식에 대한 선택사항 이나(그림 2.3A), 2) 한 property 값 선택에 따라 다른 property의 range class에 속하는 member가 변경 되거나(그림 2.3B), 3) 다른 property 값이 결정되는 규칙으로 정의할 수 있다(그림 2.3C).

위와 같은 규칙들을 SWRL에 적용 할 수 있도록 일반화 하면 "한 property의 값이 특정 class에 속한 individual중 에 선택되거나, 특정 individual 혹은 특정 형태의 data로 선택된다면, 다른 특정 property의 값이 class 나 individual 및 data 형태로 제한된다."라고 다시 표현 할 수 있다. 이와 더불어 consequent에 속하는 property가 Object property인 경우 property의 range가 변경 된 값들로 바뀌거나

(replace), 변경 된 값들이 추가(merge) 될 수 있다. 이 중 replace가 되는 경우는 antecedent가 만족되기 전에 range에 속해 있던 individual들은 다른 class로 속해야 하므로 해당 individual들의 class를 사용하지 않는 class로 변경 해주어야 한다. 단, data property의 경우 data type을 range로 가지므로 replace만 가능하다.

위에서 설명한 방법을 이용해 "만약 option이 Endo_mucosal_resection 이라면, 검체의 specimen type은 emrSpecimenType 로 제한한다." 라는 규칙이 존재할 때 이를 SWRL 방식으로 바꾸어 표현하면, " 'Template의 hasOption property 값은 EMR 이다.' 라는 triple이 존재할 경우, emrSpecimenType class에 속하는 individual들을 specimenType으로 Replace 한다." 와 같이 표현되며 이는 SWRL의 조건문으로 표현가능하다(식 2). 해당 식에서 "EMR"은 "Endo mucosal resection"을 뜻하며 "Trash"라는 클래스는 사용하지 않는 임의의 클래스를 뜻한다.

$$Template(z) \wedge hasOption(z, EMR) \wedge emrSpecimenType(x) \wedge specimenType(y) \rightarrow specimenType(x) \wedge Trash(y) \quad -----(2)$$

위 식을 antecedent와 consequent을 나누어 XML 형태의 SWRL로 작성할 수 있다(표 2.3).

본 절에서 설명한 바와 같이 통제자연어 진료서술문 작성 시 발생할 수 있는 작성규칙들은 SWRL의 규칙 표현 방식인 antecedent와 consequent를 통해 표현 가능하며, SWRL은 OWL과 연계가능 한 규칙 체계이므로 OWL을 통해 수집한 온톨로지 정보로 부터 다양한 추론(reasoning) 가능성을 보장한다.

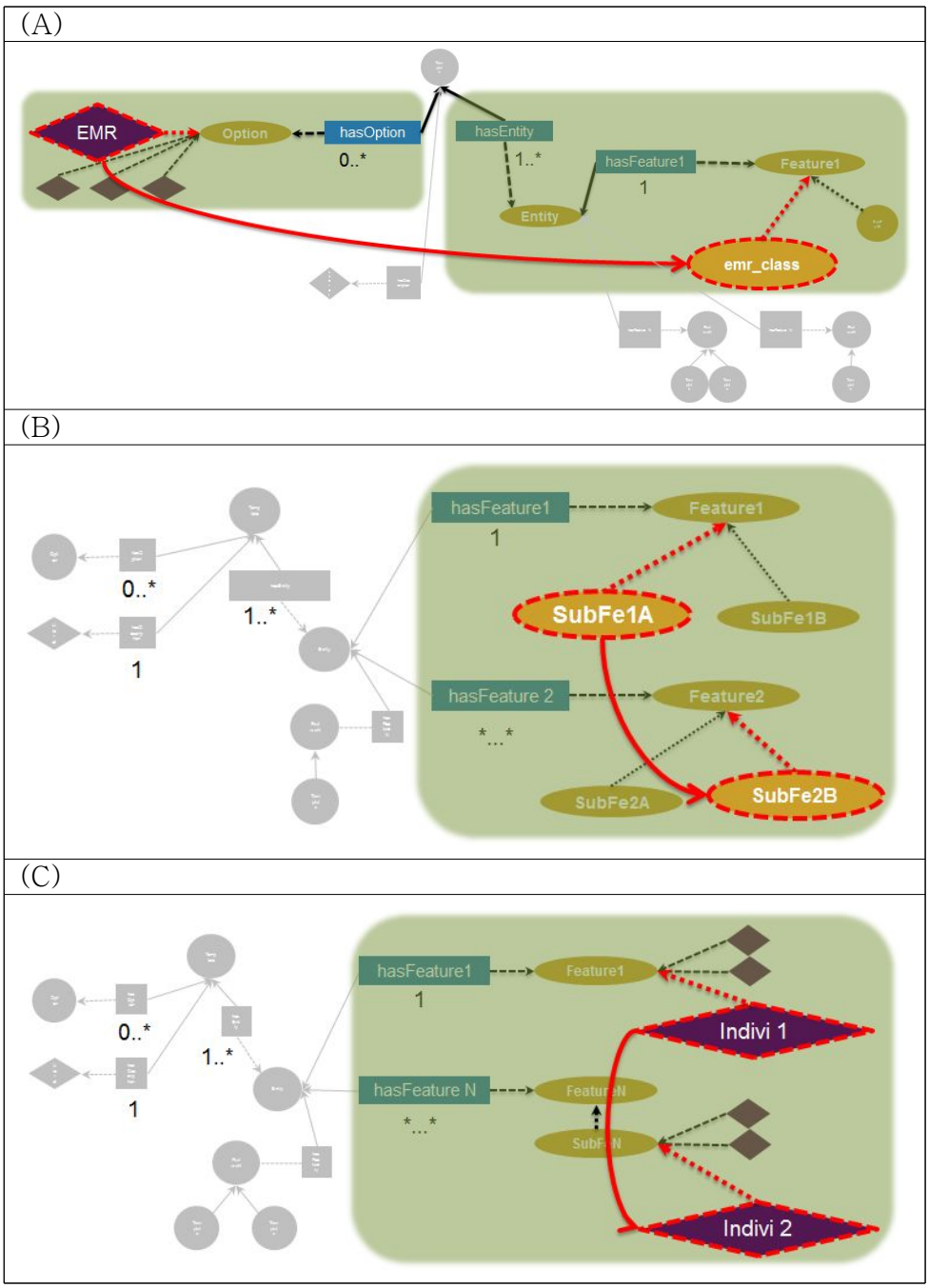


그림 2.3. 진료서술문 작성규칙 예시

굵은 화살표는 규칙을 표현하며, 화살표의 시작점은 antecedent를 끝점은 consequent를 나타낸다. 시작점에 연결된 개체가 선택되면, 끝점에 연결 된 개체 중에서 특정 property의 값이 선택되어야 한다는 것을 나타낸다.

(A) 의료서식 선택사항에 따른 규칙은 선택사항이 결정되면 property인 Feature1의 값이 SubFe1A에서 결정 되어야 하며, 이를 위해 SubFe1A class에 속하는 individual들이 Feature1에 속하게 된다.

(B) 특정 property Feature1의 값이 SubFe1A class의 member 중에서 결정 된 경우 Feature2의 값이 SubFe2B class의 member 중에서 결정 된다.

(C) 특정 property Feature1의 값이 individual인 Indivi1으로 선택된다면, Feature2의 값은 Indivi2이다.

Type	SWRL Source
Antecedent (Body)	<DLSafeRule> <Body> <classAtom> <class IRI="emr_specimen_type"/> <Variable IRI="x"/> </classAtom> <classAtom> <class IRI="specimen_type"/> <Variable IRI="y"/> </classAtom> <classAtom> <class IRI="template"/> <Variable IRI="z"/> </classAtom> <ObjectpropertyAtom> <Objectproperty IRI="hasOption"/> <Variable IRI="z"/> <Namedindividual IRI="EMR"/> </ObjectpropertyAtom> </Body>
Consequent (Head)	<Head> <classAtom> <class IRI="specimen_type"/> <Variable IRI="x"/> </classAtom> <classAtom> <class IRI="trash"/> <Variable IRI="y"/> </classAtom> </Head> </DLSafeRule>

표 2.3. SWRL XML로 표현한 진료서술문 작성 규칙

SWRL의 규칙은 “DLSafeRule” 로 표기되며 Antecedent 부분이 규칙의 body, Consequent 부분이 규칙의 head를 뜻한다. axiom에 사용 된 IRI는 특정한 자원을 지칭하기 위해 사용된 값이다.

4. 다각도 용어 사용 빈도

다각도 용어 사용 빈도(TFV: Term frequency from various angles)를 기반으로 추천 용어 목록 순서를 결정한다. 이 값을 기준으로 용어를 정렬함으로써 최근에 많이 사용 된 용어 일수록 상위 목록에 보여주어 개인적인 추천 용어 목록 순서 제공을 가능하게 한다.

다각도 용어 사용 빈도는 모든 사용자(단어 자체), 의무기록 작성자 및 환자 별로 3가지 측면으로 용어 사용 빈도를 계산한다. 이 때, 사용 시점과 현재시간을 계산을 위한 값을 시간 값이라 하며 1970년 1월 1일 기준으로 경과시간을 밀리 초로 나타낸 자연수 값이다.

이러한 시간 값을 활용하여 사용시간을 고려한 용어 사용 빈도 계산식은 다음과 같다(식 3).

$$TF = \sum_{i=1}^{n} \left(\frac{ed_i}{td} \right) \quad \text{-------} \ (3)$$

TF(Term frequency)는 용어 사용 빈도를 뜻하며 n은 사용횟수를 나타낸다. 용어가 사용 된 시점의 시간 값(ed)을 현재의 시간 값(td)으로 나누어 오래 전에 사용한 용어 일수록 0에 가깝고 최근 일수록 1에 가까운 값을 갖도록 하여 이를 합산한다. 사용시간을 포함하여 위에서 언급한 3가지 항목을 고려하여 용어별 사용 빈도를 측정한다(식 4).

$$TFV = \sum_{type} (TF_{type} \times w_{type}) \quad \text{-------} \ (4)$$

TFV 는 3가지 항목을 고려한 용어 사용 빈도를 나타내며, type은 전체, 사용자, 환자의 3가지 항목을 뜻한다. 또한 각 항목 별로 사용자로부터 설정 된 0과 1 사이의 실수 값인 가중치(w)를 곱하여 합산함으로써 요소별 중요도 설정 기능을 제공할 수 있다.

이 절에서 설명한 바와 같이 산출 된 TFV를 이용하여 사용자가 자주 그리고 최근에 사용한 용어 일수록 큰 다각도 용어 사용 빈도를 갖게 되며 이를 이용하여 의무기록 작성 시 입력 인터페이스의 용어추천 목록 순서를 변경함으로써 사용성을 반영한다.

결 과

본 장에서는 온톨로지 및 규칙과 더불어 진료서술문 형태의 의무기록 작성을 위해 사용된 통제자연어를 설명하고 온톨로지 및 규칙을 입력 받도록 구현 된 인터페이스 Owlpath의 각 부분인 Authoring interface와 통제자연어 진료서술문을 입력받기 위한 인터페이스 CNL editor interface를 설명한다.

1. Owlpath 통제자연어 진료서술문 문법

OwlPath의 통제자연어 문법은 시멘틱 언어로부터 subject, predicate, object를 참조하는 DESC(2)와 유사하게 정의한다. OwlPath에서 사용된 문법은 OWL의 개념(concept)과 문법에서 사용되는 용어인 예약어(keyword) 그리고 매끄러운 자연어 형태의 진료서술문 작성을 위해 사용하는 지시자(identifier) 및 대명사(pronoun)을 Token으로 사용한 통제자연어로써 BNF 형태로 표현가능하다(그림 3.1).

이와 같은 통제자연어 문법을 활용하여 기존의 개조식 형태의 의무기록을 진료서술문 형태로 작성할 수 있다(표 3.1). 또한 규정된 문법에 의해 작성 된 통제자연어 진료서술문 문장의 구조는 별도의 작업 없이 명확한 파악이 가능하다(그림 3.2).

```
Sentence ::= Phrase { “,” Conjunction Phrase } "."

Phrase ::= NounPhrase VerbPhrase { Conjunction VerbPhrase }
          | Subject TypePredicate class { Conjunction VerbPhrase }

NounPhrase ::= Subject
          | Subject RelativePronoun VerbPhrase
          | Subject RelativePronoun TypePredicate class

VerbPhrase ::= Predicate ObjectPhrase { Conjunction ObjectPhrase }

ObjectPhrase ::= Object
          | Object RelativePronoun VerbPhrase
          | Object RelativePronoun TypePredicate class
          | Literal

Subject  ::= individual | Pronoun | Definiteindividual | Identifier
Predicate::= property
Object   ::= individual | Pronoun | Definiteindividual | Literal

Definiteindividual ::= Article individual
TypePredicate ::= "is"
Article ::= "the"
Pronoun ::= "it"
Conjunction ::= "and"
RelativePronoun ::= "which" | "that"
Identifier ::= "specimen_received" | "specimen"
```

그림 3.1. BNF 형태의 Owlpath 통제자연어 문법

TypePredicate, Article, Pronoun, Conjunction, RelativePronoun, Identifier 및 문장 마침표 “.” 는 환경에 따라 지정 가능하며, class, individual, property의 온톨로지 정보는 서식 정보로부터 지정된다. Literal은 공백문자로 분리되지 않은 모든 문자열을 수용한다.

개조식 대장암 병리보고서	통제자연어 대장암 병리보고서
Colon, anterior_resection: · Carcinoma_with_adenomatous_component · Adenocarcinoma, moderately_differentiated 1. Location : signomid colon 2. Gross_type : fungating 3. Size : 2.5×1.5×1.5 cm 4. Depth_of_invasion: invades_proper_muscle(pT2) 5. Resection_margins : free_from_carcinoma – safety_margins : proximal 3.2cm, distal 6.6 cm 6. Regional_lymph_node_metastasis : no_metastasis_in_all_20_regional_lymph_nodes(pN0) (pericolic 0/10, “IMA root” 0/10) 7. Lymphatic_invasion : present 8. Venous_invasion : not identified 9. Perineural_invasion : not_identified 10. Pre-existing_adenoma : tubular_adenoma_with_high_grade_dysplasia 1) size : 1.5×1×0.5 cm 2) resection_margin : free_from_adenoma	theColon hasSpecimen_Type anterior_resection and hasMaindiagnosis carcinoma_with_adenomatous_component. theColon which hasSize 2.5 x 1.5 x 1.5 cm hasGross_type fungating. the_received_specimen hasLocation sigmoid colon. the_received which hasSafetymargins proximal 3.2cm, distal 6.6cm hasDepth_of_invasion "invades_proper_muscle(pT2)" and hasResectionmargins free_from_carcinoma. the_received hasRegional_lymph_node_metastasis no_metastasis_in_all 20 regional_lymph_nodes(pN0) and pericolic 0/10, "IMA root" 0/10. It hasLymphatic_invasion present and hasVenous_invasion not_identified and hasPerineural_invasion not_identified. the received hasPre-existing_adenoma tubular_adenoma_with_high_grade_dysplasia that hasSize 1.5 X 1 X 0.5 cm and hasResection_margin free_from_adenoma.

표 3.1. 형태에 따른 대장암 병리보고서

개조식 대장암 병리보고서: colon에 대해 anterior_resection을 수행한 절제 대장에 대한 병리보고서로써, 검체 종류 및 주 진단명과 상태묘사를 포함하여 장소, 육안 검사형태, 크기, 침윤의 깊이, 절제연, 지역적 임파선 전이, 임파선, 혈관, 신경적 전이 상태와 기존의 선종에 대한 총 13가지 정보를 표현하고 있다.

통제자연어 대장암 병리보고서: 13가지 정보를 7개의 통제자연어 문장으로 표현하였다.

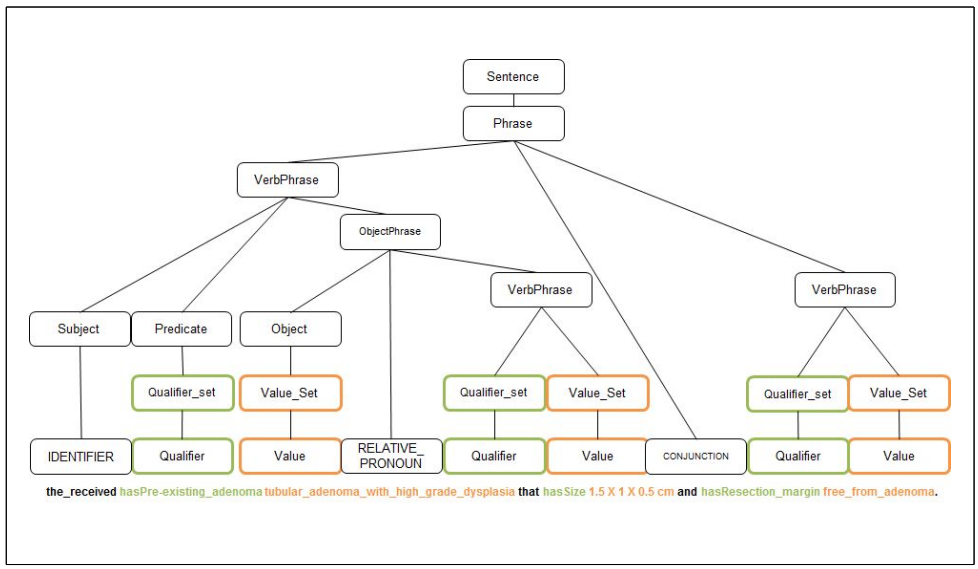


그림 3.2. 통제자연어 진료서술문의 parsing tree

통제자연어로 표현 된 기존 선종(Pre-existing adenoma)에 대한 문장 “the received hasPre-existing_adenoma tubular_adenoma_with_high_grade_dysplasia that hasSize 1.5 X 1 X 0.5 cm and hasResection_margin free_from_adenoma.” 의 부분별 문법 요소를 분석하면 각각 Identifier, Qualifier, Value, RELATIVE_PRONOUN, Qualifier, Value, Conjunction, Qualifier, Value로 구분되며, 구분 된 문법 요소들이 계층적 구조로 연결되어 하나의 문장(Sentence)를 구성하게 된다.

2. Owlpath 의료서식 저작 및 통제자연어 진료서술문 입력 인터페이스

본 절에서는 통제자연어로 입력되는 진료서술문 형태의 의료서식 온톨로지 및 작성규칙을 저작할 수 있고, 이에 따라 의무기록을 작성할 수 있도록 구현 된 Owlpath의 각 인터페이스들을 설명한다.

2.1. 시스템 아키텍처

OwlPath는 의무기록의 양식에 대한 정보를 작성할 수 있는 저작 인터페이스(Authoring interface)와 통제자연어의 의무기록을 작성할 수 있는 통제자연어 진료서술문 작성 인터페이스(Reporting interface), 2개의 인터페이스로 구성된다(그림 3.3).

의료종사자는 저작 인터페이스를 통해 의무기록 서식 온톨로지와 작성규칙을 입력할 수 있으며, 입력 된 OWL과 SWRL을 기반으로 통제자연어 진료서술문 작성 인터페이스를 통해 의무기록을 작성할 수 있다. 이와 같이 작성 된 의무기록은 RDF 형태로 출력 가능하다. 또한, 각 요소에 따라 용어 추천 목록 순서를 정렬하기 위한 정보인 의료종사자, 환자 및 용어 사용기록을 데이터베이스로 관리한다.

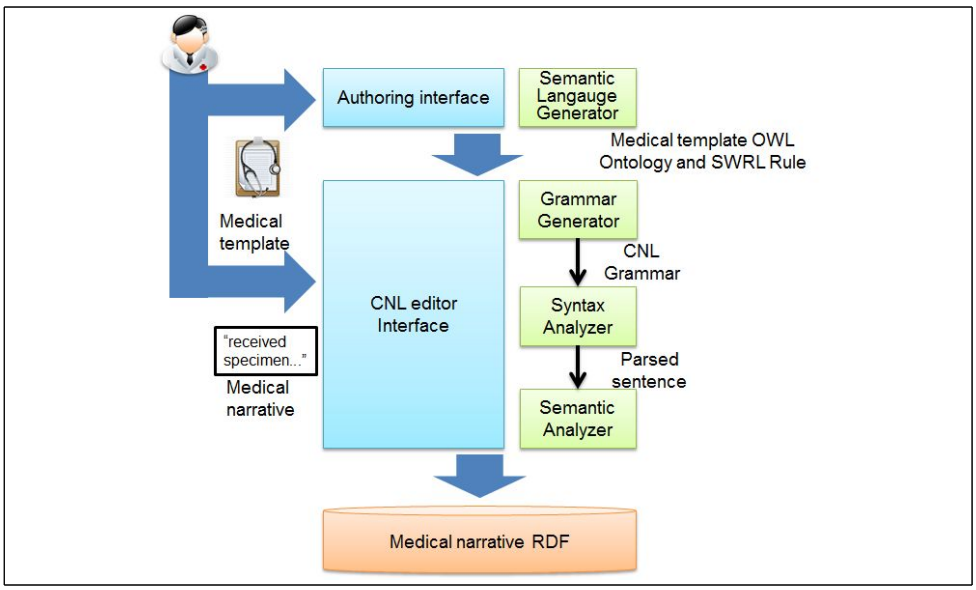


그림 3.3. Owlpath 시스템 아키텍처

저작 인터페이스를 통해 입력 된 의료서식 정보가 OWL 온톨로지 및 SWRL 규칙으로 변환된다. 작성 인터페이스는 저작 인터페이스를 통해 작성 된 서식파일과 CNL문법을 결합하여 통합된 문법을 생성한 뒤, 구문 분석기, 의미 분석기를 통해 사용자로부터 입력 받는 진료서술문이 올바른 통제자연어 문법으로 작성될 수 있도록 가이드하게 된다. 작성된 결과물은 RDF로 출력될 수 있다.

Owlpath에서 통제자연어 생성을 위해 사용한 JavaCC[4] 버전은 5.0이며, 구문 분석을 위해 사용한 JavaCC JTB는 1.4.4 이다. 온톨로지 및 규칙 작성과 분석을 위해 사용한 OWL API[5]는 3.2.4 버전을 사용하였다. 또한 각 요소별 TFV를 측정하고 추천 용어를 정렬하기 위해 사용한 Database는 sqlite[6] 0.56버전이며, 작성 된 통제자연어 진료서술문을 RDF로 출력하기 위해 Jena 2.7.1[7]을 사용하였다. 끝으로 화면 구성을 위해 eclipse RCP(Rich client platform)[8]를 사용하였다.

---

4) Available at: http://javacc.java.net
5) Available at: http://owlapi.sourceforge.net
6) Available at: http://www.sqlite.org/
7) Available at: http://jena.apache.org/download/index.html
8) Available at: http://www.eclipse.org/home/categories/rcp.php

2.2. 의료서식 온톨로지 및 작성규칙 저작 인터페이스

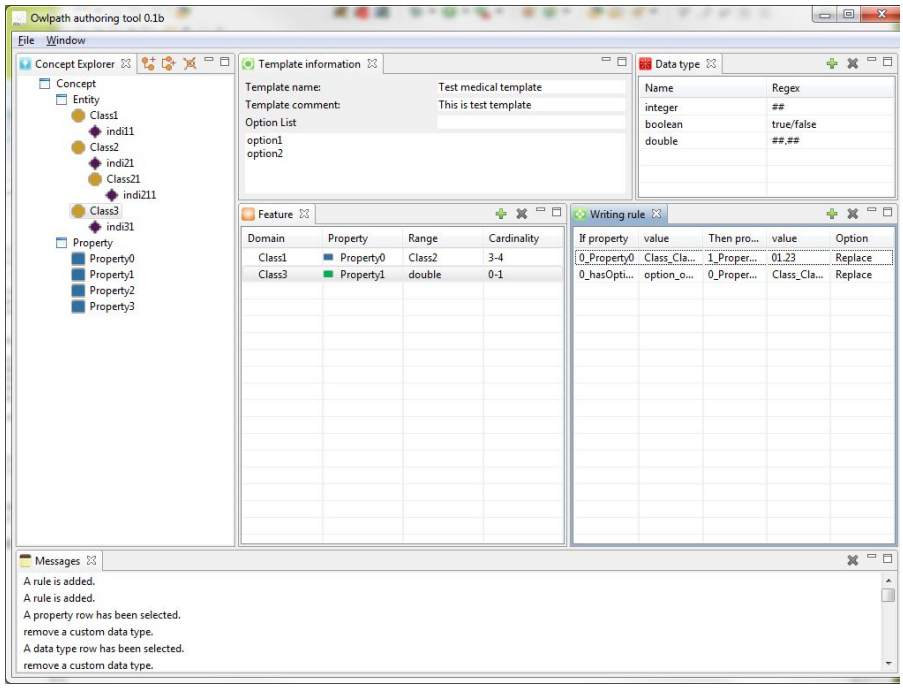
Owlpath의 의료서식 온톨로지 및 작성규칙을 저작 하는 인터페이스는 상단의 메뉴바와 6개의 창(Viewpart)으로 이루어진다(그림 3.4A).

상단의 메뉴 바의 “File” 메뉴를 통해 새로운 template ontology를 작성하거나 기존의 OWL파일을 불러오거나 현재 작성중인 내용을 저장할 수 있다. “Window” 메뉴를 통해 꺼져있는 창을 켤 수 있다.

concept창 상단의 아이콘 중 왼쪽 첫 번째 “Add child” 아이콘을 클릭하면 현재 선택한 concept의 하위에 개념이 생성되고 두 번째 “Add sibling” 아이콘을 클릭하면 현재 선택한 concept의 형제(sibling)으로 개념이 생성된다. template 창을 통해 template의 이름(title) 및 설명(comment)을 작성할 수 있으며 선택사항을 리스트(option list) 형식으로 입력 가능하다. Feature 창에서는 한 줄을 추가, 선택 후 property를 끌어 놓음(drag and drop)으로써 property를 추가하고 해당 property의 domain, range 또한 drag and drop으로 선택가능하며 cardinality의 범위를 하이픈 “-” 이나 물결기호 “~” 를 통해 지정하며, 최소를 “minQualifiedCardinalitiy”, 최대를 “maxQualifiedCardinalitiy” 로 저장한다. 이 때 범위를 나타내는 수를 한 개만 입력하면 “exactQualifieredCardinality” 로 cardinality를 저장하게 된다. 또한 data type 창을 통해 새로운 자료형을 정규식 텍스트로 표현하여 입력 가능하다. 작성 규칙 입력 시는 writing rule 창의 추가 버튼을 클릭하여 규칙입력 윈도우를 통해 새로운 규칙을 규정할 수 있다(그림 3.4B).

본 절에서 설명한 “OwlPath: 저작인터페이스” 를 이용하여 작성된 의무기록 양식의 온톨로지와 작성규칙은 OWL의 저장형태인 OWL/XML, Turtle의 형태로 저장, 불러오기를 수행할 수 있다.

(A)



(B)

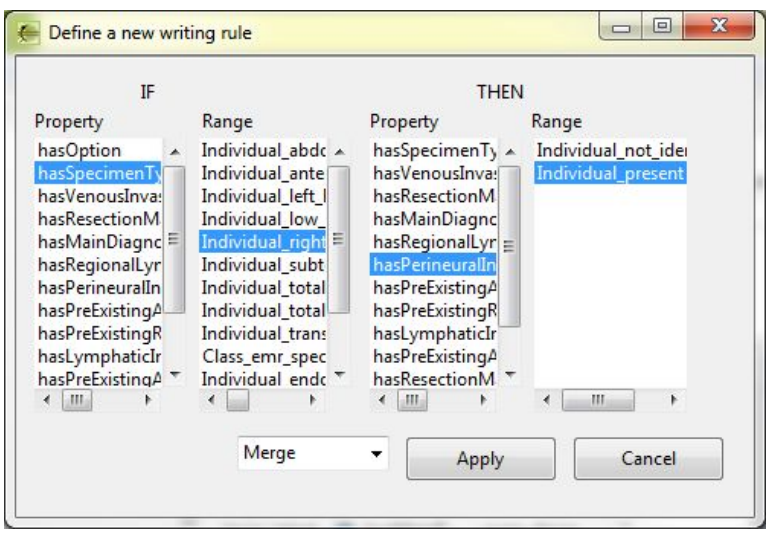


그림 3.4. 의료서식 온톨로지/규칙 저작 인터페이스

(A) 저작 인터페이스의 화면구성은 시계방향으로 좌 상단에는 OWL의 entity인 class와 individual 및 property의 구성 보여주고 설정할 수 있는 concept 창이 존재하며, 상단의 양식의 정보와 option을 설정할 수 있는 template창, 그리고 중앙에는 property의 domain, range와 restriction인 cardinality를 선택할 수 있는 feature 창이 존재한다. 우 상단에는 data type을 설정할 수 있는 data type 창이 존재하며 의료서식의 작성규칙을 입력할 수 있는 writing rule 창이 존재한다. 끝으로 하단에 위치한 메시지 창을 통해 메시지를 볼 수 있으며 각 창은 앞서 설명한 위치에 고정되어 있지 않고 이동, 크기변경, 켜기, 끄기 등을 수행할 수 있다.

(B) 새로운 규칙을 규정하는 창인 “Define a new writing rule” 창의 if 에는 SWRL의 antecedent에 해당하는 의료서식 선택사항의 목록 및 선택 된 property range로 설정 된 class와 individual 혹은 data type이 선택 될 수 있으며, then 부분도 마찬가지로 SWRL의 consequent에 해당하며 임의의 class, individual 및 data type이 선택될 수 있다. 또한 class나 individual이 consequent로 지정된 경우 “Replace” 와 “Merge” 의 선택사항을 지정할 수 있다. 단, data type으로 지정된 경우 “Replace” 로 고정된다.

2.3. 통제자연어 진료서술문 작성 인터페이스

Owlpath 진료서술문 작성 인터페이스는 사용자가 올바른 통제자연어로 진료서술문을 작성하도록 유도하는 인터페이스이다(그림 3.5).

진료 서술문을 작성하기 위해 로그인(그림 3.6A)을 해야 하며, 이 후 “File” 메뉴의 “Load template”을 통해 저작 인터페이스로 작성한 서식 온톨로지 파일을 로드해야 한다. 만약 새로운 사용자의 경우 “New Account” 다이얼로그를 통해 등록 할 수 있다(그림 3.6B). 서식 온톨로지 로딩이 완료 되면, 작성하고자 하는 진료서술문에 대한 정보를 입력하는 “Information” 다이얼로그를 통해 진료서술문의 제목, 환자 이름 및 지시자(identifier)가 지칭하는 개념을 선택할 수 있다(그림 3.6C). 이 때 입력한 환자가 새로운 환자인 경우 “New patient” 다이얼로그를 통해 환자 정보를 입력할 수 있다(그림 3.6D).

서식 로딩 및 진료서술문 정보 입력 후 진료서술문 작성 시 올바른 통제자연어 진료서술문 작성을 위한 용어 추천을 제공하고, 추천되는 용어와 달리 문법 및 의미적으로 잘못 작성 된 사항은 이탤릭체로 변경 후 밑줄을 긋고 각각 빨강, 주황색으로 해당 용어 및 문장을 표시한다. 또한, 개인화 된 추천용어 목록을 지원하기 위해, 구축 된 데이터베이스를(그림 3.7) 통해 용어별 TFV를 산출한다(표 3.2).

작성 된 통제자연어 진료서술문은 진료서술문의 식별번호와 identifier 지칭 용어 정보를 포함한 텍스트 파일형태로 저장가능하며, RDF 출력형태인 n-triple 로 추출할 수 있다. 또한 RDF로 통제자연어 진료서술문 저장 시, 서식 정보에 해당하는 서식 제목(rdfs:label)과 설명(rdfs:comment)이 저장되고, 진료서술문 정보인 제목(hasName)과 작성자(hasClinicianSeq) 및 환자(hasPatientSeq) 정보가 저장되며, 적용된 서식의 선택사항(hasTemplateOption)이 저장된다. 이와 더불어 진료서술문내의 모든 entity는 “hasClinicalEntity” property로 보고서 class와 연결된다.

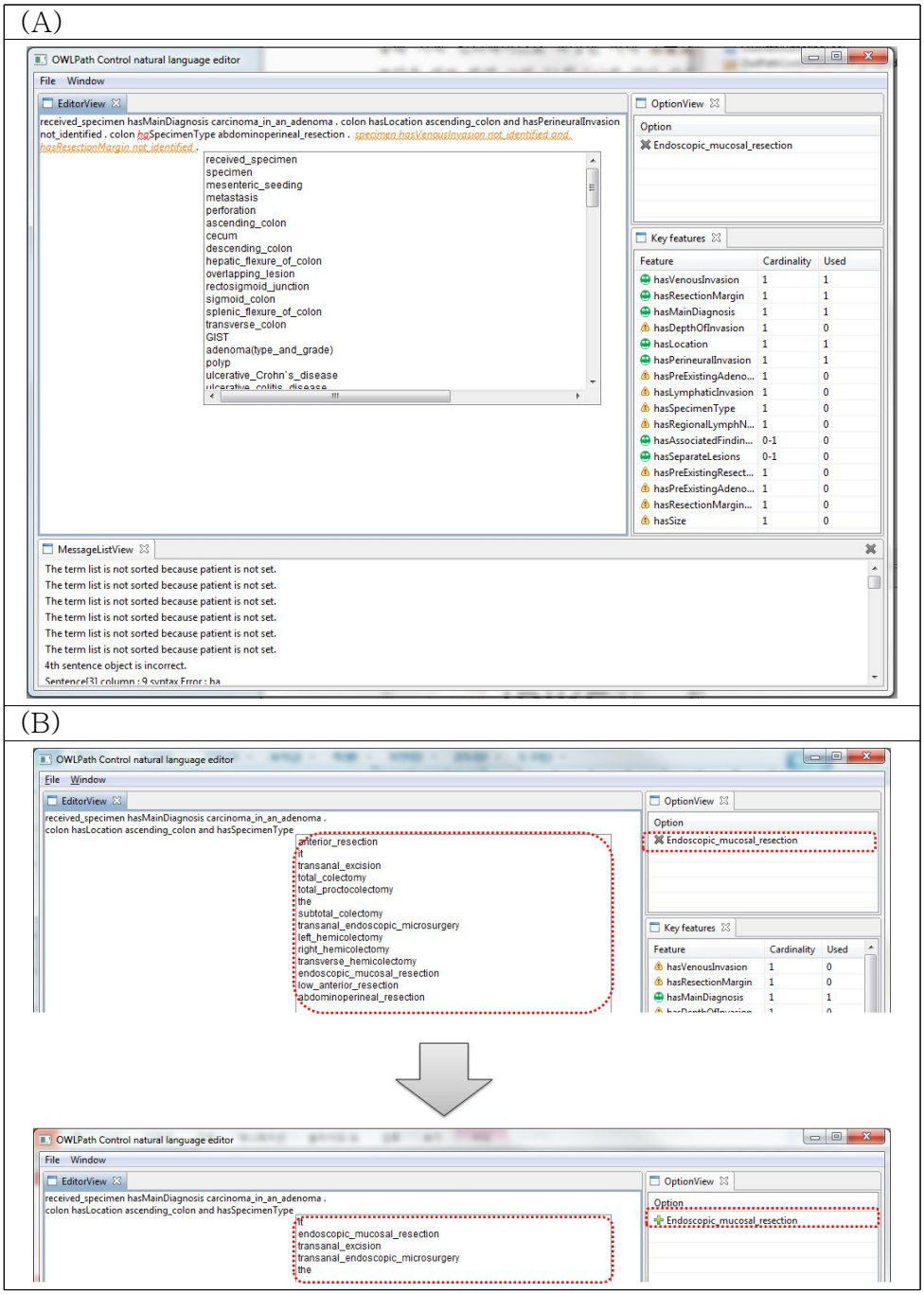


그림 3.5. 통제자연어 진료서술문 작성 인터페이스

(A) 주 화면은 좌 상단부터 시계 방향으로 진료 서술문을 입력하는 “Editor view”와 서식의 선택사항을 변경할 수 있는 “OptionView” 그리고 property가 cardinality에 맞게 작성 되었는지 알려주는 “Key features”view가 위치하며, 끝으로 하단은 저작 인터페이스과 마찬가지로 메시지 화면을 보여주는 “MessageListView”로 구성 되어 있다. “Editor view”를 제외한 다른 view들은 이동, 확대, 축소, 열기, 닫기를 수행할 수 있으며, 최상단에 위치한 메뉴 바의“Window”메뉴를 통해 view를 켤 수 있다.

(B) 의료서식 선택사항 “Endoscopic_mucosal_resection” 이 선택되면, “hasSpecimenType” property의 range class member가 변경된다는 규칙이 존재 할 경우 이를 고려하여, 추천되는 용어 목록이 상단의 14개 목록에서 하단의 5개의 목록으로 변경된다.

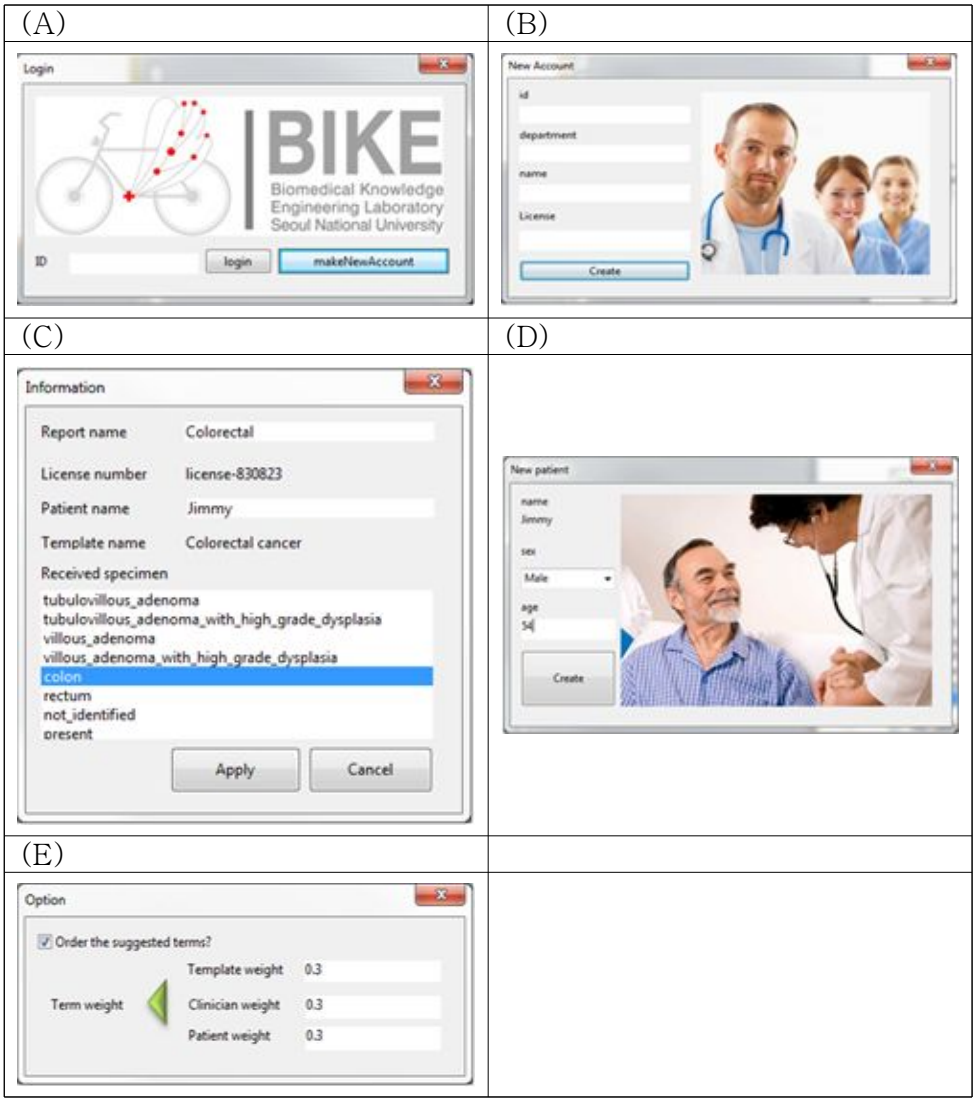


그림 3.6. 통제자연어 작성 인터페이스 다이얼로그

(A) Login 다이얼로그: 로그인 기능을 제공한다.

(B) New Account 다이얼로그: 새로운 계정 생성 기능을 제공한다.

(C) Information 다이얼로그: 진료서술문 정보 입력 기능을 제공한다.

(D) New patient 다이얼로그: 새로운 환자 생성 기능을 제공한다.

(E) Option 다이얼로그: TFV의 가중치 변경 옵션을 제공한다.

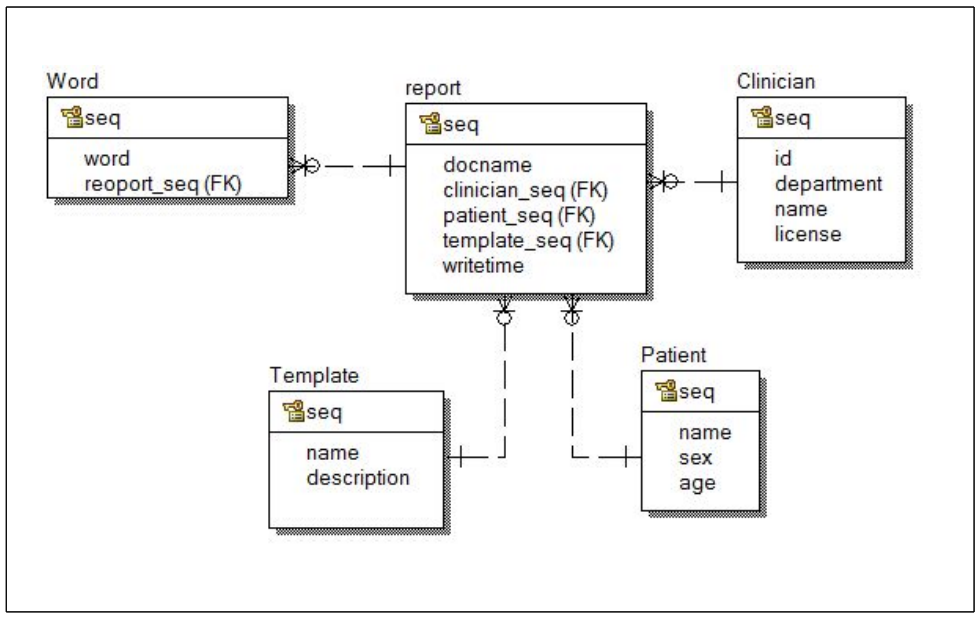


그림 3.7. 용어 사용빈도 측정 ER-Diagram

데이터베이스의 중심적인 테이블인 report테이블은 진료서술문에 대한 정보인 제목(docname), 작성시간(writetime)을 비롯하여 작성자(clinician), 환자(patient), 서식(template)의 식별번호(seq)를 가지게 된다. 또한 사용용어를 나타내는 word 테이블은 용어가 사용된 진료서술문 식별번호와 용어정보를 지닌다.

SQL	Comment
select	Start of SQL
(SELECT sum(writetime/[current time]) FROM word,report where word.report_seq = report.seq and report.template_seq = [template seq] and word.word='[term]') * templateWeight +	Usage frequency of term itself.
(SELECT sum(writetime/[current time]) FROM word,report where word.report_seq = report.seq and report.template_seq = [template seq] and word.word='[term]' and report.patient_seq = [patient seq])* patientWeight +	Usage frequency of term for a patient.
(SELECT sum(writetime/[current time]) FROM word,report where word.report_seq = report.seq and report.template_seq = [template seq] and word.word='[term]' and report.clinician_seq = [clinician seq]) * clinicianWeight;	Usage frequency of term for a clinician.

표 3.2. 다각도 용어 사용빈도 측정을 위한 SQL 구문

단어자체, 환자, 작성자 별 TF를 산출하여 통합하는 SQL 구문을 통해 한 용어에 대한 TFV를 산출하게 된다. “current time”은 현재의 시간 값을 나타낸다.

3. 의료서식 입력 및 진료서술문 작성 결과

본 절에서는 앞서 설명 된 인터페이스들을 이용하여 의료현장에서 사용 가능한 의료 서식들을 입력하고, 이를 통해 생성된 OWL과 SWRL을 바탕으로 통제자연어 진료서술문을 작성함으로써 Owlpath의 적용 가능성을 보인다.

의료서식은 2장에서 설명한“대장암 병리보고서 기재사항 표준화”와 2005년에 제안 된“위암 병리보고서 기재사항 표준화”(24)이며, 해외 사례는 2007년 ASCP(American Society for Clinical Pathology)에 제안된 웹 폼 형태의 병리보고서“Tumor reporting system”(25)의 신장 종양, 폐종양 병리보고서 서식정보를 입력 하였다.

국내 서식의 온톨로지 입력 기준은 해당 논문들에 기재된 사항을 준수하였으며(그림 3.8A), 규칙은“(양식의)이탤릭체는 EMR 검체에서만 사용하는 항목임.” 이라는 것을 SWRL로 표현하기 위해 이탤릭체로 표기된 것을 “emr_”이라는 접두사를 붙인 class로 설정 후 규칙에 적용하였다. 또한 “절제연이 involved_by_carcinoma인 경우에는 safety margin을 0cm 로 적는다.”라는 규칙을 적용하기 위해 해당 속성 값이 선택되면 “hasResectionMarginSafetyMargin”의 값을 모두 0cm가 되도록 규칙을 규정하였다(그림 3.8B).

웹 폼 형태의 해외 의료서식의 온톨로지 입력 기준은 웹 폼의 각 컴포넌트들의 특성을 고려하여 설정하였다. object property인 선택사항 컴포넌트를 표현하기 위해, 체크박스는 최소 0개에서 최대 체크박스 개수, 라디오 버튼은 최소 0개에서 최대 1개 그리고 콤보박스는 최소, 최대 모두 1로 설정하였다. 또한 text field로 구성된 입력사항은 data property로 입력하였다. 해외 의료서식의 규칙은“RegionalLymphNode가 pNX인 경우 examine 되거나 involve된 lymphnode의 수는 0이다.”라는 구문을

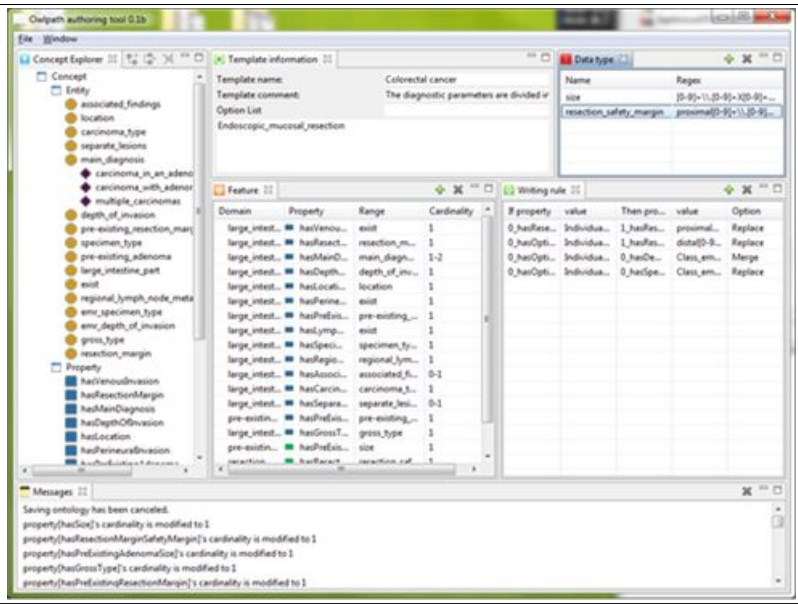
data type의 규칙으로 규정하였으며, 절제연에 대한 규칙은 국내 서식에서와 유사하게 “involved_by_invasive_tumor”가 선택 된 경우 “hasResectionSafetyMargin”가 모두 0이 되도록 규정하였다.

입력결과 “대장암 병리보고서 표준 기재사항”의 class는 15개, property 16개(object 13개, data 3개), replace rule 3개, merge rule 1개였으며, “위암 병리보고서 표준 기재사항”의 경우 class 16개, property 18개(object 15개, data 3개), replace rule 3개, merge rule 1개였다. 해외 의료서식 경우 신장암 병리보고서 양식의 경우 class 14개, property 18개(object 14개, data 4개), replace rule 3개가 존재하였으며, 폐암의 경우 class 15개, property 18개(object 14개, data 4개), replace rule 2개가 존재하였다.

이와 같이 입력된 각 양식의 온톨로지와 규칙을 준수하여 작성된 4가지 병리보고서는 통제자연어(표 3.3) 및 RDF 형태(그림 3.9)로 저장 가능하였다. 대장암, 위암, 신장암, 폐암의 진료서술문 순으로 각각 8개, 9개, 8개, 10개의 문장으로 작성되었다. 또한, 추출 된 RDF는 웹상의 RDF validator[9]를 활용하여 문법적 오류를 확인 후, W3C에서 제공하는 RDF visualizer[10]를 활용하여 다이어그램으로 표현하였다.

9) Available at: http://www.rdfabout.com/demo/validator/index.xpd
10) Available at: http://www.w3.org/RDF/Validator/

(A)



(B)

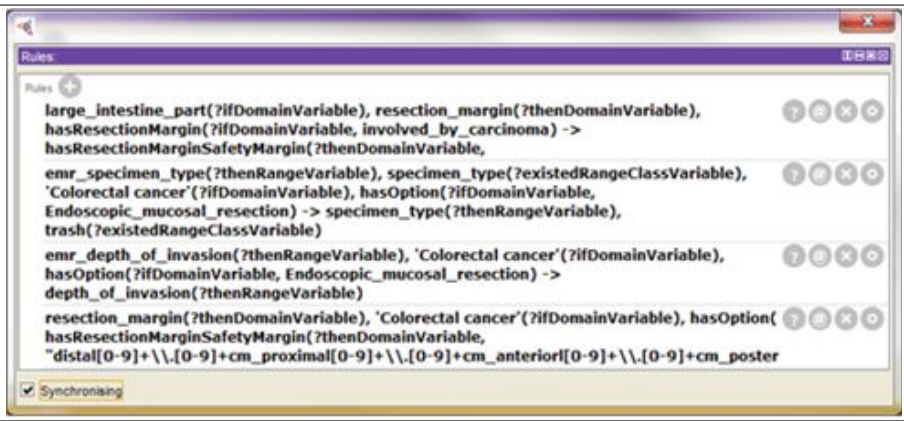


그림 3.8. 의료서식 온톨로지 및 작성규칙 결과

(A) 대장암 병리보고서 기재사항 표준화를 입력한 결과이다.

(B) 작성된 OWL파일에 표현된 규칙 논리식 예시이다.

Title	ColorectalOwlpathTest
Identifier word	colon
Appliyed rules	none

colon hasSpecimenType anterior_resection.

specimen hasMainDiagnosis carcinoma_with_adenomatous_component and hasCarcinomaType adenocarcinoma_moderately_differentiated.

colon hasLocation sigmoid_colon and hasGrossType fungating.

specimen hasSize 2.5X1.5X1.5_cm.

colon hasDepthOfInvasion invades_proper_muscle(pT2) and hasResectionMargin free_from_carcinoma which hasResectionMarginSafetyMargin proximal3.2cm_distal6.3cm.

received_specimen hasRegionalLymphNodeMetastasis no_metastasis_in_all_20_regional_lymph_nodes(pN0)(pericolic_0/10_`IMA_root`_0/10).

colon hasLymphaticInvasion present and hasVenousInvasion not_identified and hasPerineuralInvasion not_identified.

colon hasPreExistingAdenoma tubular_adenoma_with_high_grade_dysplasia which hasPreExistingAdenomaSize 1.5X1.0X0.5_cm and hasPreExisting ResectionMargin free_from_adenoma.

표 3.3-A. 통제자연어 진료서술문 예시) 대장암 병리보고서

Title	GastricOwlpathTest
Identifier word	stomach
Appliyed rules	TemplateOption - Endoscopic_mucosal_resection

stomach hasSpecimenType endoscopic_mucosal_resection.

specimen hasMainDiagnosis early_gastric_carcinoma.

stomach hasLocation antrum and hasLocation lesser_curvature.

stomach hasGrossType EGC_type_IIb and hasGrossType EGC_type_IIc.

received_specimen hasHistologicType tubular_adenocarcinoma,_moderately _differentiated.

stomach hasHistologicTypeByLauren intestinal, and it hasSize 2.2x1.5x0.15cm.

stomach hasDepthOfInvasion invades_submucosa(depth_of_sm_ invasion: 0.03_cm)(pT1b).

stomach hasResectionMargin free_from_carcinoma which hasResectionSafetyMargin distal1.0cm_proximal0.6cm_anterior0.4cm_ posterior0.7cm_deep0.05cm.

stomach hasLymphaticInvasion present and hasVenousInvasion not_ identified and hasPerineuralInvasion not_identified .

표 3.3-B. 통제자연어 진료서술문 예시) 위암 병리보고서

Title	KidneyOwlpathTest
Identifier word	left_kidney
Appliyed rules	regional_lymph_node is pNX
	resection_margin is involved_by_invasive_tumor

left_kidney hasTumorSite upper_pole and hasTumorFocality unifocal and hasSurgicalProcedure Total_nephrectomy.

left_kidney hasTumorSize 7.2X1.5X1.5_cm, and it hasTumorExtent Tumor_extends_into_major_veins.

left_kidney hasHistologicType Clear_cell_(conventional)_renal_carcinoma and hasHistologicGrade Fuhrman_Grade_2.

left_kidney hasPrimaryTumor tumor_extends_into_major_veins_or_perinephric_tissues_but_not_into_the_ipsilateral_adrenal_gland_and_not_beyond__fascia_of_Gerota(pT3).

received_specimen hasRegionalLymphNodes pNX_Regional_lymph_nodes_cannot_be_assessed and hasExaminedRegionalLymphNodes 0 and hasInvolvedRegionalLymphNodes 0.

left_kidney hasMetastasis no_distant_mestatasis(pM0).

left_kidney hasAngiolymphaticInvasion present.

received_specimen hasResectionMargin Involved_by_invasive_tumor which hasResectionSafetyMargion distal0cm_radial0cm_parenchymal0cm.

표 3.3-C. 통제자연어 진료서술문 예시) 신장암 병리보고서

Title	LiverOwlpathTest
Identifier word	liver_structure
Applied rules	none

liver_structure hasSpecimenType partial_hepatectomy, and specimen hasFocality solitary which hasFocalityLocation segmentII .

liver_structure hasTumorSize 2.2x1.9x1.8_cm.

specimen hasHistologicType hepatocelluar_carcinoma.

liver_structure hasExtentOfInvasion solitary_tumor_without_vascular_invasion(pT1) and hasRegionalLymphnode no_regional_lymph_node_metastasis(pN0).

specimen hasSubmittedLymphNode present and hasExamined LymphNodes 2 and hasInvolvedLymphNodes 0.

specimen hasDistantMetastasis cannot_be_assessed(PMX).

liver_structure hasParenchymalMargin uninvolved_by_invasion_carcinoma which hasParenchymalMarginType medial and has DistantFromClosest Margin 20mm.

received_specimen hasBileDuctMargin cannot_be_assessed and hasOtherMargin vascular_designated_by_surgeon_uninvolved_by_invasive_carcinoma .

specimen hasVenousInvasion absent.

liver_structure hasAdditionalPathologicFindings cirrhosis/fibrosis .

표 3.3-D. 통제자연어 진료서술문 예시) 폐암 병리보고서

표의 “Identifier word”는 해당 진료서술문에서 identifier가 지칭하는 단어이며, “Applied rule”은 antecedent가 만족된 rule을 뜻한다. 표 7의 내용인 위암 병리보고서 경우 서식의 선택사항으로 “Endoscopic_mucosal-resection”이 설정되었으므로 “specimen_type”이 “endoscopic_mucosal_resection”으로 설정되었으며, “depth_of_sm_invasion”이 “depth _of_invasion”에 추가되었다. 또한 표 8의 내용인 신장 병리보고서의 경우 “regional_lymph_node”로 “pNX”가 선택되어 “examined_regional_lymphnodes”와 “involved_regional_lymphnode” 모두 0으로 작성되었다. 또한 “resection_margin”이 “involved_by_invasive_tumor”로 선택되어“resection_safety_margin”의 모든 값이 0으로 작성되었다.

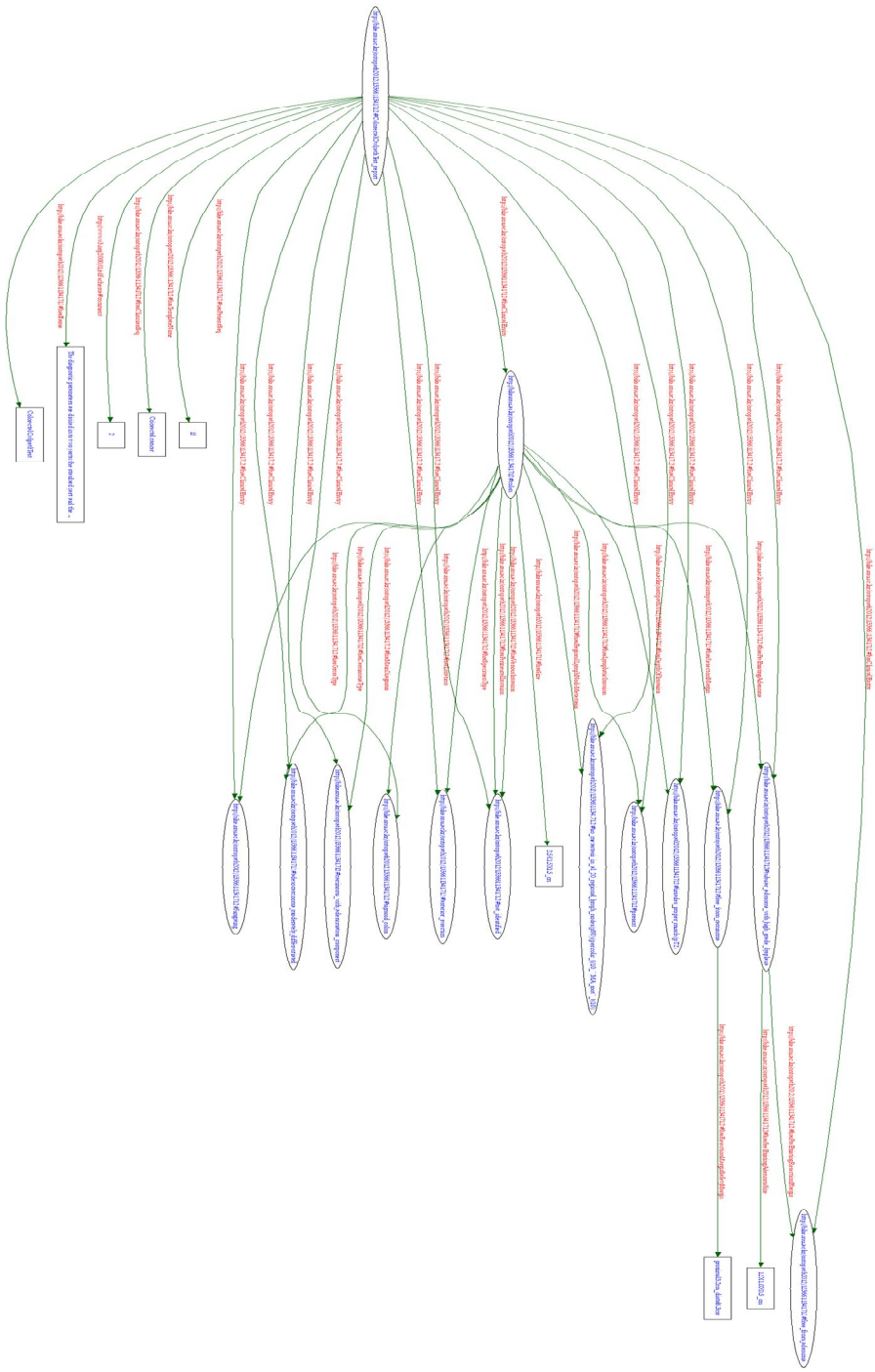


그림 3.9. 대장암 병리보고서 RDF diagram

작성 된 “대장암 병리보고서 기재사항 표준화”의 통제자연어 진료서술문을 rdf diagram으로 표현한 결과이며, 최상단에 진료서술문 individual이 존재하며 해당 individual은 “hasTemplateName”으로 서식이름과 연결되며, “rdfs:comment”로 서식설명 인 “The diagnostic parameters are divided into two parts: the standard part and the ...”과 연결된다. 또한 진료서술문에 포함된 Clinical entity 들은 "hasClinicalEntity" 라는 property로 연결되어 있다.

고 찰

본 논문에서는 기존 EMR의 의료종사자에게 익숙하지 않은 입력방식으로 인해 적절히 활용되지 못하던 단점을 보완하고자 개발되었던 Ontopath를 개선하여 새로운 인터페이스 Owlpath를 구현하였다.

Ontopath에 요구되던 개선사항은 1)모델링을 위해 사용했던 RDFS에서는 사용자가 입력 하는 용어를 제한하기 위해 property의 domain과 range만 활용되어 다른 property의 값들 간 상관관계를 표현할 수 없어 의미적으로 상충하는 부분에 대해 제약할 수 없다는 사항이 존재하였다. 또한, 2) 의무기록이 작성되는 환경 등의 선택사항에 따라 property의 range class member를 변경할 수 없던 사항, 그리고 3) property의 사용 횟수 제한을 표현할 수 없던 사항들을 표현 할 수 없다는 사항 이었다.

위의 요구사항들을 반영하기 위해 property의 cardinality와 같은 제한사항의 표현이 수월한 1) OWL를 활용하여 온톨로지를 표현하고, 2) 진료서술문 작성 시 발생할 수 있는 작성규칙을 SWRL로 표현하였다. 또한, 3) 용어 사용 빈도를 전체, 사용자, 환자별로 측정하여 이를 용어 추천 순서에 반영한 뒤, 위와 같은 3가지 개선점을 반영하여 4) 의료서식 온톨로지와 작성규칙을 각각 OWL과 SWRL을 통해 저작하고, 통제자연어 진료서술문 작성 기능을 지원하는 인터페이스 Owlpath를 구현하였다.

구현 된 Owlpath를 이용하여 국내외 의무기록 서식 및 이를 활용한 통제자연어 진료서술문 입력함으로써 의료 종사자가 다양한 환경에 맞게 기재사항 그리고 기재사항의 값들을 입력할 수 있는 적용가능성을 보여주었다.

설명한 바와 같이 구축 된 온톨로지와 작성규칙을 통해 통제자연어 입력을 수행하게 되면, 지정된 문법과 의미뿐 아니라 특정 기재사항을 선택함으로써 발생하는 의미적 상관관계를 고려한 입력을 유도함으로써 입

력된 통제자연어 진료서술문의 정확도가 향상될 것으로 기대된다. 더 나아가 SWRL 규칙의 consequent의 property값이 한 개로 지정되는 규칙에 대하여 antecedent가 만족되었을 경우 consequent에 해당하는 구문이 생략되어도 해당 부분의 의미분석이 가능할 것이다. 또한, 최근 사용이 급증하고 있는 스마트폰이나 휴대용 단말기를 이용한 음성인식 활용 시에도 문법과 의미의 상관관계를 통한 규칙으로 용어의 자유도를 제한하여 보다 정확한 음성 의료 기록을 저장할 수 있을 것으로 생각된다.

하지만 실제로 의료현장에 적용하기 위해서는 EMR의 온톨로지 지원 기능이 추가되어야 하며, 온톨로지를 다루는 관점으로 데이터가 다뤄져야 할 것이다. 또한, 다양한 분야, 기관 간 용어의 상호관련 정보를 올바르게 표현하기 위해 참조의료용어체계를 적절히 이용해야 할 것이다.

향후 연구계획은 통제자연어 자유도, 의료현장 적용력, 사용성 증대 및 평가방법을 개선하는 것이다. 먼저, 1) 작성 된 통제자연어 진료서술문이 S-P-O의 트리플을 문장형태로 변환 하는 현재의 방식보다 자연어에 가까운 통제자연어 진료서술문 작성을 위해 다양한 문법이 지원되거나, 사용자가 본인에게 맞는 문법을 쉽게 지정할 수 있는 방법을 모색함으로써 통제자연어 문법에 관련 된 연구가 필요로 된다. 또한 통제자연어 사용에 대한 학습량을 줄이기 위해 통제자연어의 문법을 사용자의 의무기록 작성 패턴과 유사하게 바꿔주는 방법에 대한 연구가 필요하다. 2) 의료현장의 적용력을 높이기 위해 본 논문에서 사용한 규칙 표현방식 SWRL 외에 의료지식을 표현하기 위해 사용하는 Arden syntax(26)와 연계하는 방향도 고려해야할 것이다. 이와 더불어 3)사용성을 증대시키기 위해 사용자가 사용한 용어의 빈도를 측정 것 외에 Hidden markov chain(27)를 활용하여 사용 된 용어의 패턴을 추천 용어 순서 변경에 적용할 수 있을 것이다. 또한, 본 논문에서는 의료서식

표준안에서 제시하는 샘플 데이터를 입력하는 제한적 평가 방법이 사용되었으나, 실제 의무기록을 통한 표현력, 정확도 및 사용성 측면의 평가 수행이 가능할 것이다. 먼저, 특정 의료분야 혹은 부서의 의무기록 중 왜곡되거나 빠짐없는 표현 가능 비율을 측정하는 것과 더불어 온톨로지 내 중복 입력 된 용어의 비율을 산출함으로써 온톨로지와 규칙 모델링의 표현력을 평가할 수 있을 것이다. 또한, 의료종사자를 통해 기존의 의무기록을 통제자연어로 표현하고, 이를 다시 원래의 의무기록으로 바꾸어 자연어와 통제자연어 진료서술문 변환 전 후의 의무기록을 비교함으로써 정확도를 평가할 수 있을 것이다. 끝으로 추천용어 목록 자동 순서 변경, 통제자연어 자유도 및 서식 입력의 용이성을 설문을 통해 평가함으로써 사용성을 평가할 수 있을 것이다.

참 고 문 헌

1. George, J., An ontology model for clinical documentation templates, MIT Thesis, 2005, 70
2. 하병현, 김홍기, 이재일, 김명기, 진료서술문의 구조적인 입력을 위한 온톨로지 기반의 통제자연어와 인터페이스 개발, 대한의료정보학회지, 2006, 12(3):11.
3. 장정화, 종이의무기록과 전자의무기록의 마취기록 기재 충실도 비교, 연세대 학위논문, 2010, 56
4. Ashish K. Jha, M. D., M.P.H., etal., Use of Electronic Health Records in U.S. Hospitals, N Engl J Med, 2009, 360:38:00
5. EMR활용 사례 및 발전방향, 분당 서울대병원, 2007, 93
6. Dukyung Yoon, Byung-chul Chang, Adoption of electronic health records in Korean tertiary teaching and general hospitals, International Journal of Medical Informatics, 2012, 196-203:8
7. Boone, K. W., The HL7 Clinical Document Architecture, The CDA TM BOOK, 2011, 5
8. Kohl CD, Garde S, Knaup P., Facilitating secondary use of medical data by using openEHR archetypes., Stud Health Technol Inform, 2010, 160(Pt 2):1117-21
9. 김윤, 안선주, 조경희, 류상희 고영택 등, 임상콘텐츠모형 개발 및 구조화된 자료입력 도구, EHR핵심 공통 기술연구개발사업단 연구보고서, 2009, 130
10. 전영주, 의료법상 의료정보 보호방안-의무기록보호 중심으로, 법학연구 한국법학회 제28집, 2007, 465-483

11. 이인영, 전자의무기록에 관한 법적 문제, 한양대학교 법학 논총 제28집 제1호, 2011, 24
12. http://www.w3.org/TR/rdf-schema/, RDF Vocabulary Description Language 1.0, W3C, 2004
13. Seon Gyu Park, H.-G. K., Dental Decision Making on Missing Tooth Represented in an Ontology and Rules, ASWC, 2006, LNCS 4185:07
14. http://www.w3.org/TR/owl-features/, OWL overviews, W3C, 2004
15. Kim, H. G., B. H. Ha, et al., A multi-layered application for the gross description using Semantic Web technology, Int J Med Inform, 2005, 74(5): 399-407
16. Sharon O'Brien, Controlling Controlled English - An Analysis of Several Controlled Language Rule Sets, Proceedings of EAMT-CLAW, 2003
17. Fuchs NE, Schwertel U, Schwitter R, Attempto Controlled English: Not just another logic specification language, Lecture Notes in Computer Science, 1999, 1559: 1-20
18. Sowa JF, Graphics and languages for the Flexible Modular Framwork, Lecture Note in Artificial Intelligence, 2004, 3124:31-51
19. Glen Hart, M. J., and Catherine Dolbear, Rabbit: Developing a Control Natural Language for Authoring Ontologies, ESWC, 2008, LNCS 5021: 13
20. Schwitter R, Tillbrook M, Controlled natural languages meets the Semantic Web, Proceedings of the Australasian Language Technology Workshop, 2004, pp. 55-62

21. Hyun Namgoong, H.-G. K., Ontology-based Controlled Natural Language Editor Using CFG with Lexical Dependency, ISWC'07/ASWC'07, 2007, 14
22. 장희진, 박. 외, 대장암 병리보고서 기재사항 표준화, The Korean Journal of Pathology, 2006, 40:11
23. Lowe Hj, Huang Y, Regula DP., "Using a statistical natural language Parser augmented with the UMLS specialist lexicon to assign SNOMED CT codes to anatomic sites and pathologic diagnoses in full text pathology reports.", AMIA, 2009, 2009:386-90
24. 김우호, 박. 외., 위암 병리보고서 기재사항 표준화, The Korean Journal of Pathology, 2005, 39:13
25. Zhenhong Qu, Shibu Ninan, etc., Synoptic Reporting in Tumor Pathology, Am J Clin Pathol, 2007, Jun 127(6):845-7
26. Samwald M, Fehre K, de Bruin J, Adlassnig KP., The Arden Syntax standard for clinical decision support: Experiences and directions., J Biomed Inform, 2012, DOI j.jbi.2012.02.001
27. Lawrence R. Rabiner, A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, IEEE, 1989

Abstract

Introduction: The purpose of this study was to model an OWL based ontology for representing medical template, define SWRL rules of writing medical narratives, and implement OwlPath the controlled natural language editor and an authoring tool. Moreover we devised the way to sort suggested terms according to term frequency of usage from various angles.

Methods: The medical template ontology represented in OWL consists of information about template, and the data of medical entities. The information about template refers to options which can be used as precondition in writing rules such as whether EMR specimen or not. Writing rules mean that precondition such as selecting a individual can decide the ranges of certain properties. For example, if the EMR option is selected as a template option, then the histologic type should be selected in a certain range like a emr_histologic_type class. Moreover, in order to improve usability, term frequency was calculated from various anlgles such as term itself, clinicians, and patients.

Results: Considering all functions explained above, we implemented Owlpath supporting both an authoring interface and a CNL editor interface. By using those interfaces, users can define medical template ontology and writing rule for CNL

medical narratives on the authoring interface, and also they could be guided to write correct CNL medical narrative on the CNL editor interface. While writing medical narratives, users can see the suggested terms to write correct CNL sentences. By doing so, the consistency could be assured between the semantic to which a user intends and the semantic which is understood by machines.

Conclusions: The Owlpath interfaces showed sufficient usability in medical field by inserting template ontologies and writing rules represented in pathology reports which are pathology reporting standards or web pathology reporting system.

* This work is published in KOSMI(Kyungmin Lee, Hong-Gee Kim, Jeong-Suk Yoo, An ontology model for representing writing rules in medical template and CNL interface, Proceedings of KOSMI, 2012; 519).

Keywords: Ontology, Semantic web, Rule language, Controlled natural langauge, Medical template

Student Number: 2010-23750