

원저

한양방 임상연구 중례기록서의 온톨로지 구조 설계 및 웹기반 입력폼 자동 생성 알고리즘 개발

문진석¹, 박경모², 최선미¹

한국한의학연구원¹, 경희대학교 동서의료공학과²

Web Interface Generator of Ontology Based CRF for Clinical Trial in Western and Oriental Medicine

Jinseok Moon¹, Kyungmo Park², Sunmi Choi¹

Korea Institute of Oriental Medicine¹,
Dept. of Biomedical Engineering, Kyunghee Univ.²

Abstract

Objective: This study aims to develop the web-interface creator, which automatically changes the Case Report Form(CRF) web page when the protocol developer in any clinical study revises the ontology of CRF. **Methods:** This study established the conceptual model of CRF on oriental and western medicine, and developed ontologies. A program was developed to produce online-based a input form through the parser that generates automatically HTML script from OWL. **Results:** The CRF applied in this study is a draft through consensus of experts for stroke research at the Korea Institute of Oriental Medicine. The ontology of CRF consisted of Label, ControlType and Value classes and hasControl, hasValue and hasSymptoms properties. The Label is the class of question items groups, so it could have CRF questionnaire instances. The ControlType is the class that expresses controls such as checkbox, text, etc in the HTML script. The Value class represents selections for each items. Finally, the HTML script was created by XSL transformation from the OWL script. **Conclusion:** Online-based automatic interface creation, which enables immediate coping with the continuous change in the medical knowledge system, assures reduced time requirement. (*Journal of Korean Society of Medical Informatics 15-1, 59-69, 2009*)

Key words: Ontology, OWL, XSLT, Electronic Case Report Form, Clinical Research

논문투고일: 2008년 11월 18일, 심사완료일: 2009년 3월 9일

교신저자: 문진석, 대전광역시 유성구 전민동 461-24 한국한의학연구원 의료연구부 (305-811)

Tel: 042-868-94881, Fax: 042-863-9464, E-mail: moonstone2@kiom.re.kr

* 이 연구는 한국한의학연구원(K08010)의 지원을 받아 수행하였습니다.

I. 서론

임상연구를 진행하면서 개개 피험자별로 계획서에서 규정한 정보를 기록하여 의뢰자에게 전달될 수 있도록 고안한 인쇄되거나 전자 문서화된 서식인 증례기록서(CRF, Case Report Form)를 필수적으로 작성하게 된다. 최근 종이기반 CRF에서 eCRF(Electronic Case Report Form), CTMSs(Clinical Trial Management Systems) 등의 컴퓨터 기반의 임상시험 관리 시스템들의 도입으로 임상자료의 입력오류의 감소, 자료의 실시간 전송 등의 편의, 임상시험자료의 질적 가치 향상 등의 장점으로 선호되고 있다.

온톨로지(Ontology)는 단순히 개념 클래스의 의미를 정의하는 것이 아니라, 개념클래스와 관련된 관계, 개념 클래스 간의 연산함수, 지식 추론을 위한 공리규칙과 개념 인스턴스를 총체적으로 정의하여, 영역의 지식체계를 형식화하는데 목적이 있다. 즉 온톨로지는 영역 지식 표현을 위한 형식 체계인 것이다. 온톨로지는 확실적인 데이터 모델이 아니라, 더욱 대규모로 복합적인 데이터 모델 개발 시 함축적인 의미를 지닐 수 있다¹⁾.

의학 분야에서의 온톨로지는 표준용어 시스템의 필요성으로 처음 시작하게 되어, 시맨틱 웹 구현을 위한 웹 응용기술, 인공지능의 지식표현 연구 등으로 응용되고 있다²⁾.

의학용어시스템에서의 온톨로지는 GALEN, UMLS 나 SNOMED CT가 그 대표적인 예로 들 수 있다. 중국에서는 UMLS의 모델을 기반으로 UTCMLS(Unified Traditional Chinese Medical Language System)³⁾을 개발하였다. UTCMLS의 sub-ontology는 The Basic Theory of Traditional Chinese Medicine을 비롯하여, The Doctrines of Traditional Chinese Medicine and Relevant Science, Chinese Materia등 총 14개로 구성되어 있다.

국내에서 한의학과 관련된 온톨로지 연구들을 살펴보면, 대부분 domain ontology위주로 개발되어 있다. 한의학 용어의 시소러스 구축⁴⁾, 한의학 온톨로지를 이용한 논문 검색 시스템⁵⁾, 한의진단 온톨로지 개발⁶⁾, 처방지식관리시스템 개발⁷⁾등과 같은 온톨로지 연구들이 지속적으로 이루어지고 있다. 처방지식관리 시스템은 처방지식과 관련된 四診, 辨證, 論治, 方劑 4개의 영역으로 구성하고 RDFS 기반의 온톨로지 구축 방법을 제시하여 한의학 지식의 체계화 방법을 제공

하였다. 그 밖에도 온톨로지를 기반으로 3차원 EMR (Electronic Medical Record) 입력도구에 대한 연구도 있었다⁸⁾.

ISO의 TC215 Working Group에서는 clinical findings (임상소견)에 대한 용어의 conceptual framework을 표준화 하고 있다(ISO/TC 215 WG3 N231 - Draft ISO/DTS 22789(ballot version.7) Health informatics - Conceptual framework for patient findings and problems in terminologies)⁹⁾. 이 규격의 목적은 보건의료과학에서 형식적 개념 표현 시스템을 서술하는 데 필요한 기본개념을 정의하고 컴퓨터 기반 개념 표현 시스템에서 사용되는 개념과 특성을 기술하는 것을 목적으로 국제 표준에서 서술된 내용 모델을 정확하게 서술하는 것을 가능하게 한다.

온톨로지를 활용하여 임상정보들의 개념(concept)과 관계(relation)등에 대한 지식을 체계적으로 표현할 수 있을 것이다. 병원과 임상연구센터간의 CRF자료 전송 시, 온라인을 통하여 eCRF를 전송하게 되는데, 이 eCRF의 User Interface는 임상연구 프로토콜에 따라 변하게 된다. 이때 CRF의 knowledge가 변하였을 때, 웹 페이지가 자동적으로 변경될 수 있는 인터페이스가 개발된다면, 시간적 단축을 확보할 수 있을 것이다.

본 논문에서는 한양방 증례기록서의 구조를 interface ontology 관점에서 온톨로지로 설계하고, 이를 자동으로 웹페이지로 생성하는 프로그램 개발을 목적으로 한다.

II. 방법

1. 시나리오

병원과 임상연구센터간의 CRF자료 전송 시, 온라인을 통하여 eCRF를 전송하게 되는데, 이 eCRF의 User Interface는 임상연구 프로토콜에 따라 변하게 된다. 따라서 CRF의 knowledge가 변하였을 때, CRF 개발자는 온톨로지를 수정하게 되면, 웹 페이지가 자동적으로 변경된다.

CRF 개발자가 CRF내용을 ontology로 생성하거나 수정하면 WebUIMaker에서는 eCRF Web Interface인 HTML 파일로 자동적으로 변환한다. 그 결과물인 eCRF에는 임상연구자가 웹페이지에 접속하여 CRF를 입력하게 된다(Fig. 1).

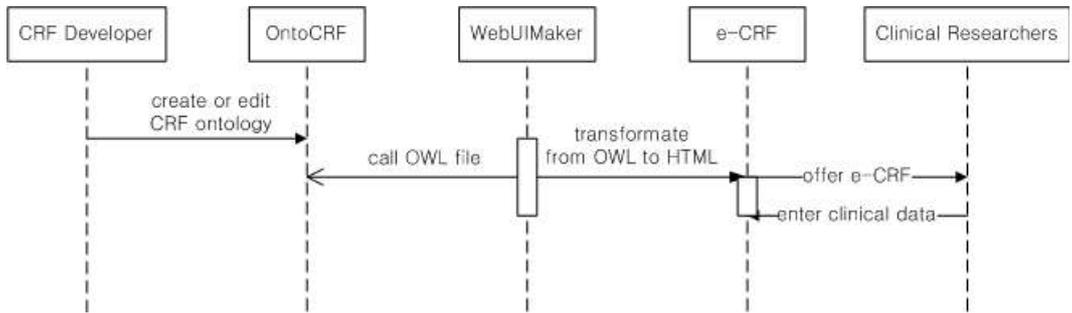


Figure 1. Sequence diagram of e-CRF

2. 개발 방법

개발 프로세스는 중풍 임상연구 CRF에 대한 conceptual model을 정립하고, ontology를 구축하여 OWL Script를 생성한 후, XSLT를 이용하여 웹기반의 HTML 코딩으로 변경하게 된다(Fig. 2).

모델링시 Conceptual Model은 ISO/TC 215 WG3 N231 - Draft ISO/DTS 22789 (ballot version.7) Health informatics - Conceptual framework for patient findings and problems in terminologies를 참고로 하여 구성하였다(Fig. 3). Characterising category간의 IS-A 관계를 명시하고, Semantic link로 연결하여 concept diagram을 구성하였다. 이의 개발 목적은 환자에게 직간접적으로 영향을 미치는 증상, 검사를 통한 결과 등의 소견들의 concept과 그들 간의 관계 정립을 통해 임상 terminology 시스템에 사용하는 것이다. 또한 이 clinical findings는 ISO17115 Health informatics-Vocabulary for terminological systems¹⁰를 기반으로 만들어졌다.

온톨로지 및 XSLT 개발을 위한 환경은 다음과 같다.

- Ontology 저작 도구: Protégé 3.1(University of Stanford, USA.)
- XSLT Validation & Transformation 도구: XMLSpy 2006 Home Edition(Altova)
- Browser: Internet Explorer 6.0

III. 결과

1. 한양방 중풍 임상연구용 CRF의 온톨로지 모델링

한국한의학연구원에서 수행중인 중풍 임상 연구에 사용되는 증례기록서를 바탕으로 Conceptual Model을 구성하였고, 온톨로지를 개발하였다. 온톨로지를 구축하기 위한 모델은 다음과 같이 표현된다(Fig. 4).

OntoCRF를 최상위 노드로 하여, Label과 ControlType, Value의 subclass가 위치하고, Label은 항목에 해당하는 그룹군으로 피험자 정보, 진단명과 같은 Label들의 상위 클래스이다. 또한 Label의 하위 클래스는 실제적인 항목을 instance로 가진다. ControlType은 항목이 가지고 입력형식에 대한 class로 그 instance들은 text, checkbox, data, time등으로 이루어져 있다. Value class는 인터페이스 상에서 Label subclass의 instance중에서 선택값이 있는 경우, hasValue관계를 가지고 있다.

예를 들어 Label class의 하위 class인 ‘진단명’ class를 살펴보면, instance로 양방진단명을 가지고 있고, 양방진단명은 Value class의 subclass인 ‘질병명’의 ‘ICH’, ‘SAH’, ‘Cerebral Infarction’ instance와 ‘hasValue’ 관계를 이루고 있는 것을 알 수 있다.

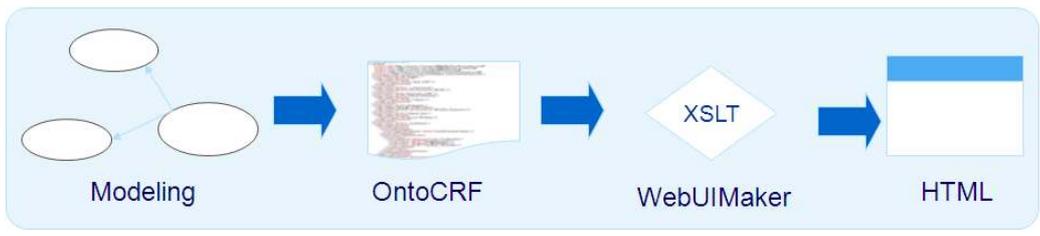


Figure 2. Process of programming

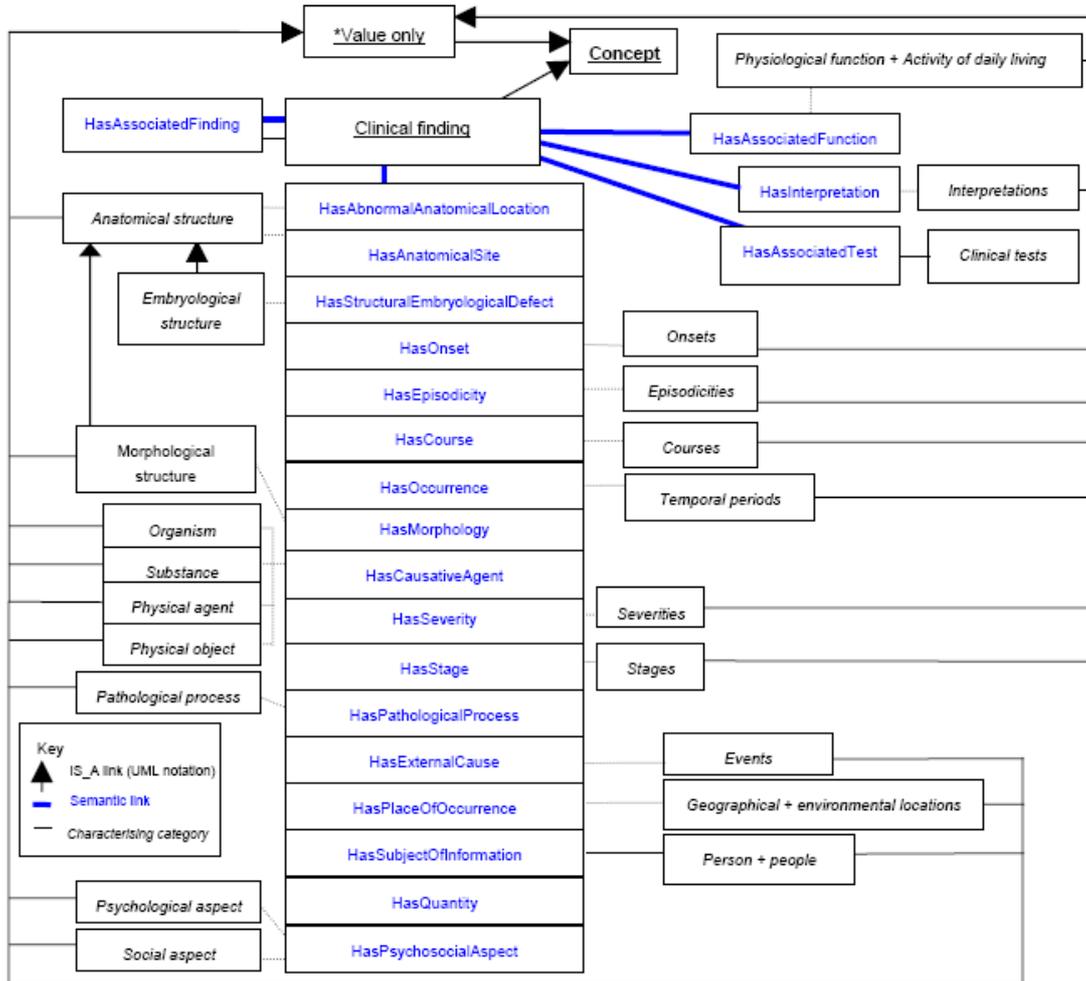


Figure 3. Conceptual framework for patient findings-ISO/TC 215 DTS 22789

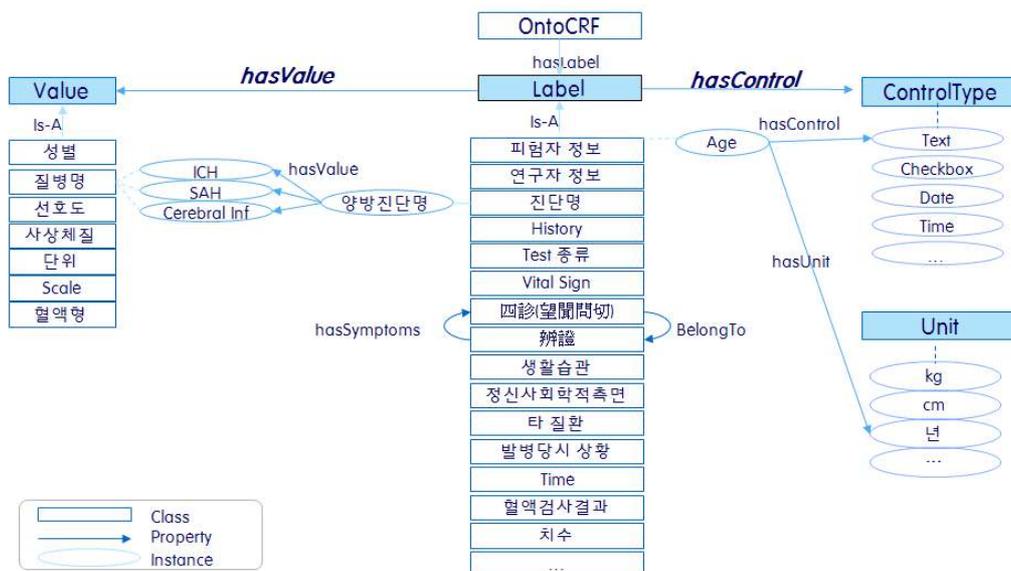


Figure 4. Conceptual model of OntoCRF in cerebrovascular disease

또한 피험자 정보에서 ‘Age’는 ‘Control Type’으로 ‘Text’를 가지는 것을 확인할 수 있다.

2. OntoCRF(Ontology of CRF)

(1) Class

Class 간의 관계는 다음과 같이 구성되어 있다. OntoCRF를 최상위 Class로 하여 ControlType, Domain, Label, Value, Unit의 하위클래스가 있고, Label의 하위 클래스로는 Person, Diagnosis, SyndromeDifferentiation (辨證, 변증; 한의학적 진단명), VitalSign, History등의 클래스로 이루어져 있다. Value는 Label에서 해당하는 선택값들에 대한 데이터를 instance로 가질 수 있다. 이 때 class name에 ‘crf:Label_History’에서 ‘Label’이 나, ‘crf:Value_Gender’에서 ‘Value’와 같이 상위 class를 표현해주어, OWL파일에서 instance를 parsing할 때, Label의 하위 class, Value의 하위 class를 구분할 수 있도록 하였다.

(2) Property

Properties는 모두 ObjectProperty로 Label과 Control과의 관계를 표현하는 hasControl, Label과 Value와의 관계를 표현하는 hasValue, 그리고 四診(FourDiagnosis; 4_Jin)과 辨證(SyndromeDifferentiation)과의 inverse property로 hasSymptoms(辨證 hasSymptom 四診), BelongTo(四診 BelongTo 辨證)가 구성되어 있다. 그 밖에도 hasUnit은 interface의 단위를 표현하는 데에 사용될 것이다 (Fig. 5).

(3) Instance

각 Label의 Symptoms, Persons등에 해당하는 “쉽게 피곤하고 힘이 없습니까?”, “나이”등 CRF 문항을 instance로 추가한다. 추후에 문항명칭이 바뀌면, instance에서 수정하게 된다.

ControlType의 Instance는 현재 CRF에서 필요한 checkbox, data, radiobutton, text, time 으로 구성되어 있다. 이는 다른 컨트롤이 필요할 경우 추가하면 된다(Fig. 6).

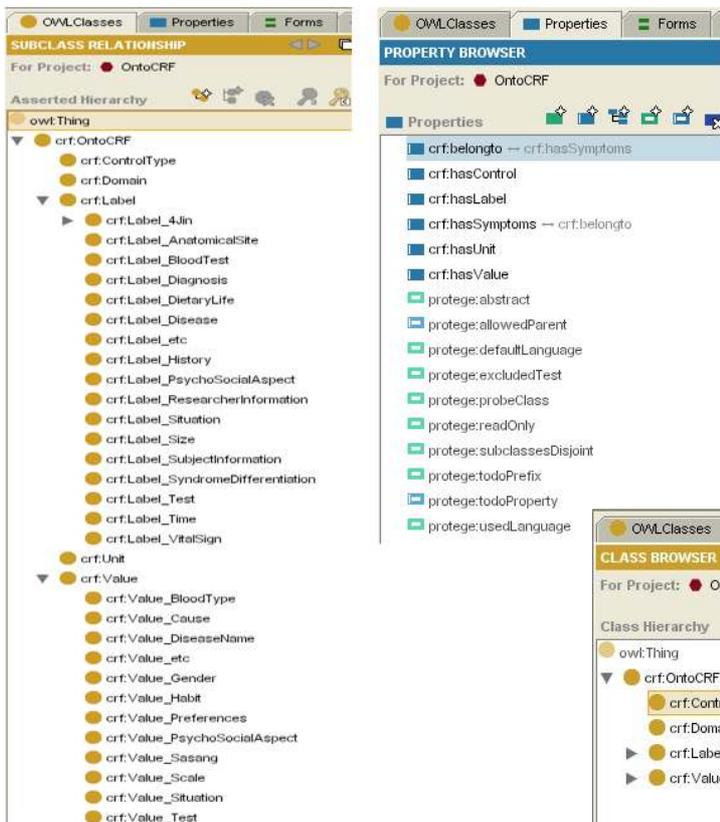


Figure 5. Classes(left) and properties(right)

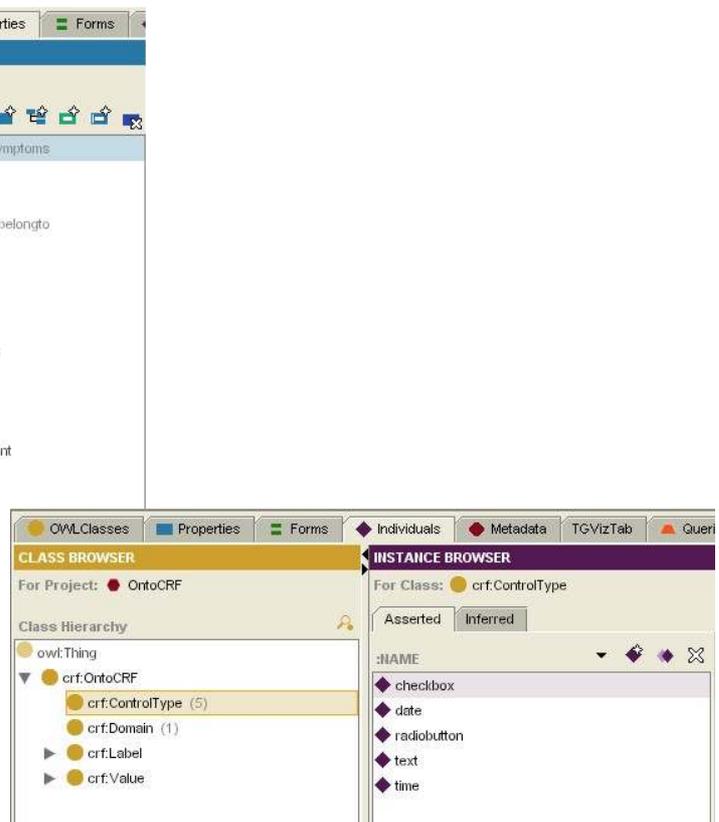


Figure 6. ControlType instance

각 instance 항목들의 Control을 지정해주기 위하여 hasControl property를 이용하여 위에서 추가하였던 ControlType의 instance를 지정해준다.

예를 들어 crf:Label▷crf:Label_4Jin▷crf:Label_Palpation (四診切診)에 해당하는 class에서 삭맥(數脈) instance는 checkbox Control을 가지고 있고(hasControl), 음허(陰虛)와 화열(火熱)의 변증의 증상으로 선택할 수 있다(Fig. 7).

만약, CRF상에서 數脈에 대해서 Yes/No의 checkbox 선택형태가 아니라, High-Medium-Low와 같은 Value

타입의 데이터 형태라면, hasValue 속성으로 관계를 가질 수 있을 것이다.

중풍에 관한 양방 진단명으로 Value 값을 Value_DiseaseName의 instance인 SAH(Sub-Arachnoid Hemorrhage; (뇌)지주막하출혈), ICH(Intra-Cerebral Hemorrhage; 뇌내출혈), Cerebral_Inf(Cerebral Infarction; 뇌경색증) 세 가지로 입력할 수 있도록 하였다(Fig. 8). 이때의 질병명은 한국표준질병사인분류(KCD)에 있는 질병명을 metadata로 사용할 수 있을 것이다.

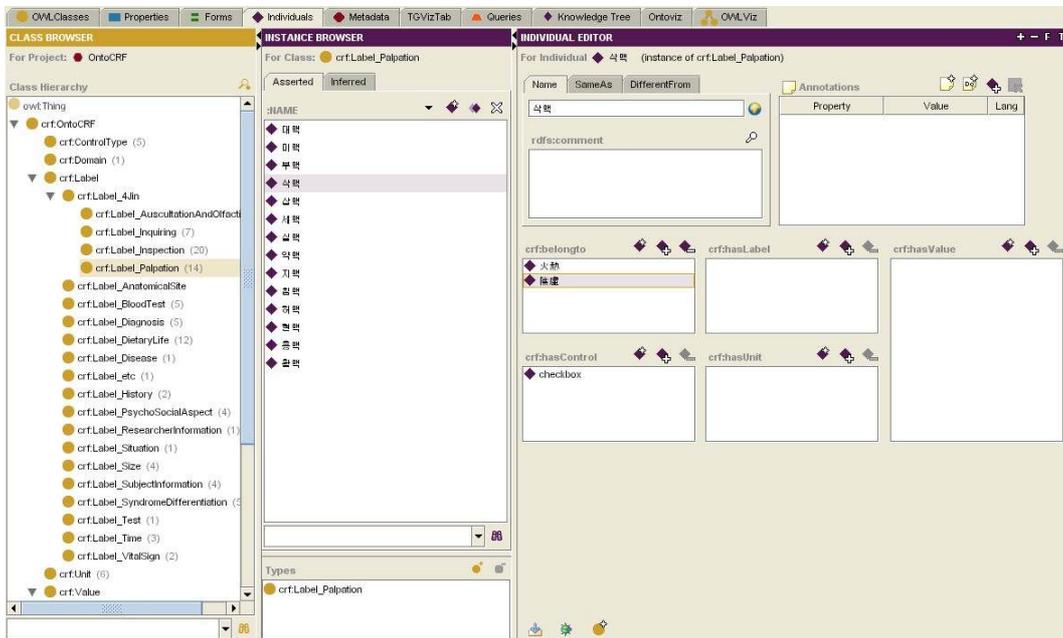


Figure 7. Instance Tab

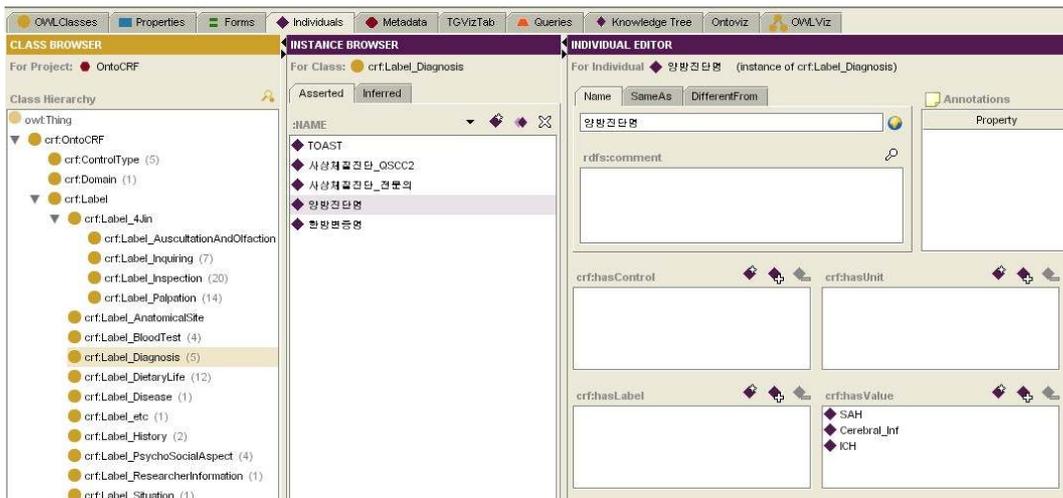


Figure 8. hasValue property

또한 단위를 웹페이지에 표시하기 위하여 단위 (Unit)도 Value instance로 추가하였다. 그리하여 둔부 들레는 hasControl 로 ‘text’를 가지고 있고, hasUnit으 로 ‘cm’을 가지고 있다(Fig. 9).

(4) Structure

Class간의 관계 및 구조를 살펴보면 다음과 같다. 대부분의 relationship은 triple 구조를 가지고 있다. OntoCRF는 Label을 가지고 있고, Label은 ControlType과

hasControl로 관계되어 있다. Label의 하위구조인 四診(4_Jin)의 望聞問切(Inspection, Inquiring, AuscultationAndOlfaction, Palpation) 4개의 Class와 SyndromeDifferentiation은 서로 belong과 hasSymptoms의 inverse관계를 가지고 있다(Fig. 10).

TGVizTab은 OntoCRF의 subclass와 각 class에 해당하는 instance, property를 보여준다. OntoCRF의 subclass인 ‘SubjectInformation’의 instance인 ‘성별’은 Value값으 로 ‘남’, ‘여’를 가지고 있는 것을 볼 수 있다(Fig. 11).

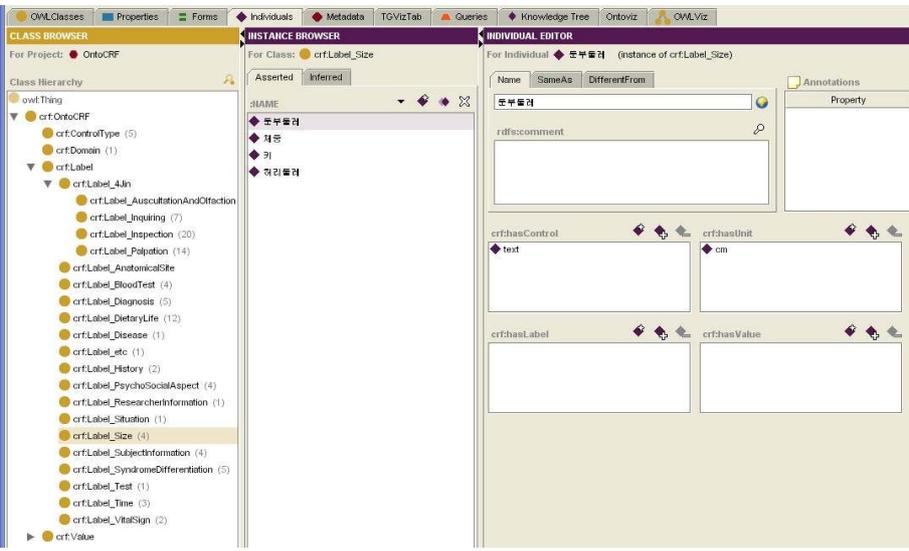


Figure 9. hasUnit property

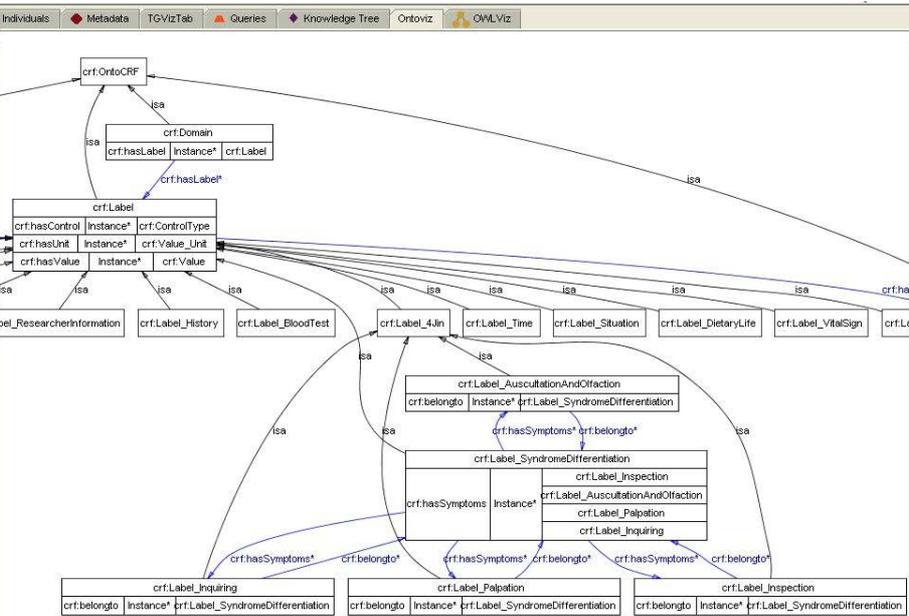


Figure 10. Structured view of OntoCRF by OntoViz

(5) Query

Protégé의 QueriesTab을 이용하여 조건에 해당하는 결과를 쉽게 검색할 수 있는 쿼리를 개발하였다(Fig. 12).

ControlType으로 checkbox를 가지고 있는 Label instance를 찾는 Query를 검색하면 ‘설사’, ‘태가 황색 이다’ 등의 결과를 볼 수 있다.

또한 Label의 하위클래스에 있는 Symptoms와 辨證 간의 관계를 이용하여 Query를 만들 수 있다. 濕痰에 해당하는 망진 증상을 찾고자 할 때, ‘Inspection belongsto 濕痰’ Query를 지정하면, “얼굴이 누렇거나 때가 낀 것 같다.”, “반대설”, “백태”, “치훈” 등의 증상결과를 얻을 수 있다.

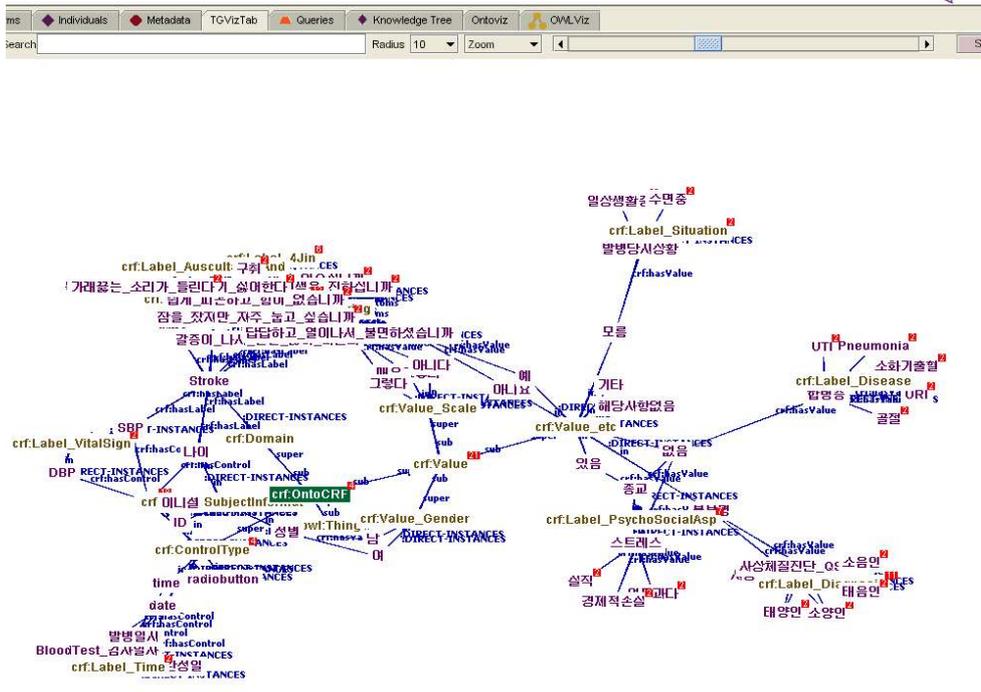


Figure 11. Relationship View using TGVizTab

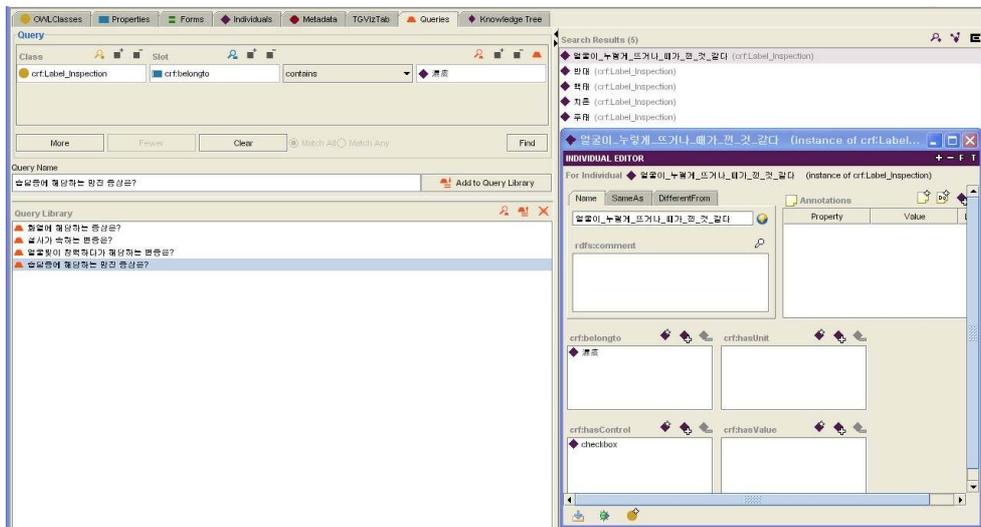


Figure 12. Queries tab to search instances

3. WebUIMaker

WebUIMaker는 OWL Script를 이용하여 질의항목들의 category(class), 질의항목(instance), 항목의 ControlType, 단위등을 자동적으로 HTML로 변환해줌으로써, CRF 변화에 따라 온톨로지를 수정하면, 웹페이지를 자동적으로 변경이 가능하게 한다.

구축된 OWL Script를 XSLT를 이용하여 HTML로 변환하는 알고리즘은 Figure 13과 같다.

XSLT를 이용하여 crf:Label의 element로 지정되어 있는 Class를 추출해내어 질의항목에 대한 상위 Class를 표현해준다. 그리고 crf:Label의 attribute인 instance를 추출하여 질의 항목을 생성한다.

다음으로 이 instance의 데이터 표현 타입에 따라 crf:ControlType에서 textbox, checkbox, date, time등으로 입력할 수 있는 란을 표시한다. 입력값이 여러 개 중 선택해야 하는 경우는 crf:hasValue의 element를 추출해내어 입력값을 나열해준다.

그리고 입력값의 단위를 표현해야 하는 경우는 crf:hasUnit의 element를 탐색하여 입력란 뒤에 표시해 주도록 하였다.

4. 자동 생성된 eCRF webpage

XSLT를 통하여 도출된 결과는 아래와 같이 웹페이지로 표현된다(Fig. 14).

어떤 도메인에 대한 CRF인지를 Domain Class의 인스턴스인 Stroke을 도출하였고, 각 Class별로 Instance를 표현해주고 있다. 또한 각 Instance의 InputType이 Text, Checkbox, Value, Date 등에 따라서 입력 Control이 위치해 있는 것을 확인 할 수 있다. 또한 단위가 필요한 경우 입력란 뒤에 표현되어 있는 것을 볼 수 있다. '구취'는 Yes/No의 binary 타입의 checkbox로 표현되어 있고, '사상체질 진단 전문의'는 '불분명', '태양인', '소음인', '태음인', '소양인'의 선택할 수 있고, '흡연량'은 textbox와 단위인 'pack-year'로 입력폼이 구성되어 있다.

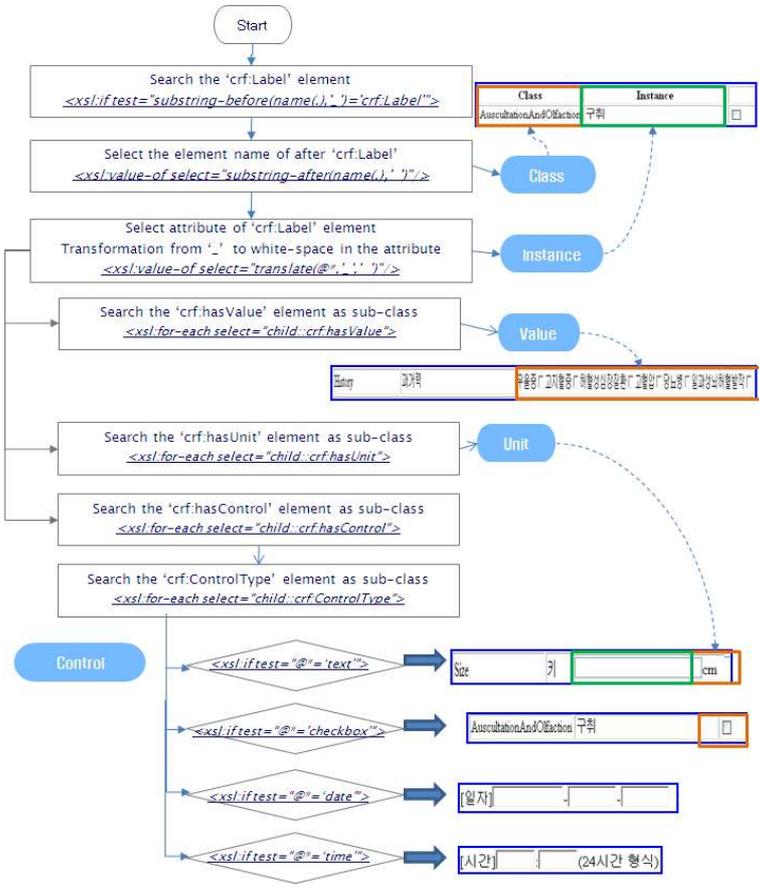


Figure 13. XSL algorithms for transformation from OWL to HTML

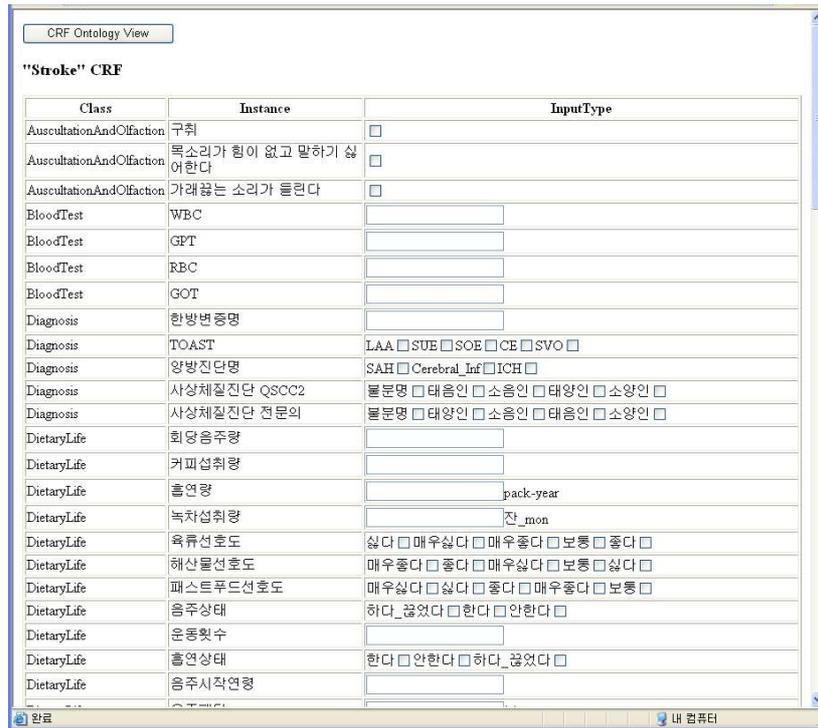


Figure 14. Output to HTML

V. 고찰

근거기반의 한의학으로 발전하기 위하여 한의학 임상시험 연구가 늘어나고 있고, WHO등에서는 한의학을 비롯하여 전통의학의 용어 및 진료의 국제 표준화를 위하여 노력하고 있다. 이에 발맞추어 한양방 융합연구를 위한 e-CRF의 개발이 필요할 것이다. 이미 다수의 한방병원에서 EMR이 개발되어 사용되고 있으며, 이를 기반으로 한의학적 개념과 양의학적 개념을 연결할 수 있는 온톨로지를 활용하여 임상연구 및 진료에 활용한다면, 보다 체계적으로 접근할 수 있리라 생각된다.

본 연구는 임상연구시 필요한 CRF를 온라인 기반 입력 폼으로 개발하기 위하여, CRF knowledge를 OWL로 표현하고, 이를 XSLT를 이용하여 HTML로 자동 변환함으로써 CRF 변화에 따라 신속하게 웹페이지를 생성, 수정할 수 있도록 하여 개발기간을 단축시키는 이점을 가지고 있다.

Ontology를 생성하기 전에 ISO의 clinical findings를 참고로 하여 한양방 임상연구시에 필요한 항목들을 보완하여 conceptual model을 구성하였다. CRF는 중풍 임상연구에 있어 한의 및 양의 융합 연구시 증례

기록서에 필요한 항목들을 domain expert들이 동의한 표준안을 바탕으로 온톨로지를 개발하였다. 온톨로지 기반으로 개념을 표현하게 되면, hierarchy 뿐만 아니라, 각 개념들간의 관계도 표현 가능하다. OntoCRF의 기본 골격은 class로 개념들을 형성하고, 개개의 항목들은 instance로 추가할 수 있도록 구성하여, OntoCRF 자체로도 CRF에 대한 개념도구로 사용할 수 있다. CRF의 ontology표현으로 OntoCRF의 subclass를 Label과 ControlType, Value로 구성하였다. Label은 질문에 대한 Instance를 추가하고, ControlType은 각 항목에 따른 HTML기반의 입력방식을 표현하였다. 그리고 Value는 각 항목의 선택값에 대한 Instance를 나타낸다.

온톨로지 저작 도구인 Protégé를 이용하여 온톨로지를 구축하였고, 이를 통해 생성된 OWL Script를 XSLT를 이용하여 HTML파일로 생성하였다. HTML파일은 CRF를 입력하기 위한 웹페이지로 사용된다.

임상정보들의 개념(concept)과 관계(relation)등에 대한 구조화된 지식인 온톨로지를 활용한다면 더욱 체계적인 자료에 대한 지식을 표현할 수 있을 것이다. 그리고 의학지식체계의 계속적 변화에 따라 웹페이지의 자동 생성으로 시간적 단축을 확보할 수 있을 것이다.

본 논문에서 개발된 중풍도메인에서의 한양방 융합 임상연구를 위한 증례기록서 ontology는 다른 질환에 적용되어 재사용 가능하다. OWL파일을 HTML기반의 입력폼으로 변환해주는 WebUIMaker는 증례기록서가 아닌 다른 도메인의 OWL파일도 Class, Instance, ControlType, Value등으로 representation하면 웹페이지로 생성이 가능하다.

본 연구를 응용한 활용방안으로는 질병명을 Value로 가지는 Label들은 UMLS나 SNOMED CT, LOINC 등과 연계하여 더욱 체계적인 의학용어를 사용할 수 있다.

온톨로지의 모델로 사용한 Conceptual framework for patient findings and problems in terminologies은 일반 진료기반에서 발생하는 도메인으로 임상연구에서 사용되는 환경과는 차이가 있다. 이에 따라 보다 임상연구에 적합한 모델 개발이 필요하다. 또한 임상연구와 관련되는 data management 프로그램으로 ICH Guideline이나 HIPAA와 같은 규정 뿐만 아니라, CDISC (Clinical Data Interchange Standards Consortium)과 같은 임상연구를 위한 국제표준등도 고려해야 할 것이다.

참고문헌

1. Lee HS. 합성 온톨로지 기반의 한의학 처방지식관리 시스템. Paju:Korean Studies Information Co.,Ltd; 2006.pp.13-16.
2. Kim HG, Kim MK. Ontology Technology in Medical Informatics. Journal of Korean Society of Medical Informatics 2003;9(3):213-219.
3. Zhou X, Wu Z, Yin A, Wu L, Fan W, Zhang R.. Ontology Development for Unified Traditional Chinese Medical Language System. Artif Intell Med 2004;32(1):15-27.
4. Knowledge of oriental medicine web service. Available at: <http://jisik.kiom.re.kr>. Accessed November 25, 2006.
5. Ko YM, Eom DM. The study on the design of korean medical article retrieval system supporting semantic navigation based on ontology. Korean Journal of Oriental Medicine 2005;11(2):35-52.
6. Park JH, Shin SW, Jung GS, Park KM, Kim SH. Development and evaluation of ontology for diagnosis in oriental medicine. Korean J. Oriental Physiology & Pathology 2006;20(1):202-208.
7. Lee HS, Lee TY. A study of the design of ontology-based prescription knowledge management system of oriental medicine. Journal of the Korean Society for Information Management 2003;20(1): 341-371.
8. Lim HS, Park KM. Interface design for structured data entry in computerized patient records system and its development. The Journal of the Korea Institute of Oriental Medical Diagnostics 2003;7(1):76-82.
9. ISO/TC 215 WG3 N231 - Draft ISO/DTS 22789 (ballot version.7) Health informatics - Conceptual framework for patient findings and problems in terminologies.
10. ISO 17115 Health informatics - Vocabulary for terminological systems.2005.