

온톨로지 품질평가를 위한 평가항목 추출에 관한 연구

Extracting Evaluation Criteria for Evaluating Ontology Quality

김성훈 (Seonghun Kim)*

오삼균 (Sam Gyun Oh)**

초 록

온톨로지의 평가는 잘 구축된 기존 온톨로지와 비교하는 방법, 활용될 애플리케이션에 적용해보는 방법, 원천데이터와의 적합성·관련성을 판단해보는 방법을 통해 이뤄지고 있다. 이와 같은 방법론은 온톨로지를 통해 얻게 된 결과에 치중되어 온톨로지의 체계, 의미표현, 상호운용성과 같은 내재적인 영역의 평가에 어려움이 있다. 본 연구는 온톨로지 전문가를 통해 온톨로지 품질평가를 위한 항목을 도출하였다. 문헌조사를 통해 온톨로지의 내재적 평가를 위한 범주를 추출하였고, 각 범주에 대한 평가항목을 델파이조사를 통해 전문가들에게 수집한 뒤, 수집된 평가항목을 재검증하였다. 그 결과, 처음 수집된 70개의 평가 항목에서 최종적으로 53개의 평가항목을 선정하였다. 또한 수집된 평가항목을 온톨로지 평가에 활용하여 봄으로써 평가항목의 신뢰도를 측정하였다.

ABSTRACT

The focus of traditional evaluations of ontologies is largely performance-based. A comparison of a new ontology with well-established ones, testing of ontologies in different applications, as well as any judgment of an ontology's appropriateness and relatedness to source data heavily rely on what results that ontology seems to manifest. This study, on the other hand, is an attempt to evaluate the quality of a particular ontology as manifested by its structure, representation, and interoperability. To that end, major categories of quality evaluations were first identified through an extensive survey of literature. Evaluation questions were formulated from these categories using the Delphi method and were validated by ontology experts. The entire process produced a set of 53 evaluation questions, which was then employed to test the quality of a newly-developed smartphone ontology.

키워드: 온톨로지, 온톨로지 평가방법, 온톨로지 평가항목, 델파이 조사
ontology, ontology evaluation method, ontology evaluation items, delphi

* 성균관대학교 문헌정보학과 박사과정수료(godwmaw@skku.edu, godwmaw0278@gmail.com)

** 성균관대학교 문헌정보학과 교수(samoh@skku.edu, samoh21@gmail.com) (교신저자)

■ 논문접수일자: 2015년 6월 11일 ■ 최초심사일자: 2015년 6월 11일 ■ 게재확정일자: 2015년 6월 21일

■ 정보관리학회지, 32(2), 193-219, 2015. [<http://dx.doi.org/10.3743/KOSIM.2015.32.2.193>]

1. 서론

1.1 연구의 필요성 및 목적

CERN(Conseil Européenne pour la Recherche Nucléaire)에서 연구 중이던 Tim Berners-Lee가 국적에 관계없이 연구자들 간에 문서의 상호호환성 문제를 해결하고 정보를 공유하기 위해 1989년 World Wide Web(Web)을 처음 고안한 이래, 웹은 그 개방성과 데이터 통신 기술의 뒷받침에 힘입어 아이부터 노인들까지 모두에게 큰 영향을 주며 괄목할만한 성장을 이뤄왔다. 그러나 점점 더 방대해지는 정보의 양은 키워드 기반의 검색뿐 아니라 의미론적 검색의 필요를 대두시켰으며, Tim Berners-Lee는 웹 개발이 10년도 채 되지 않은 시점인 1998년 시맨틱 웹(Semantic Web)을 제안하였고, 시맨틱 웹 표준화 작업이 계속적으로 진행 중에 있다.

시맨틱 웹은 웹상에 분산되어 있는 정보자원에 온톨로지의 개념을 도입하여 의미의 혼동이 없이 명확하게 정의된 웹이라고 할 수 있다. 온톨로지는 라틴어로는 ontologia, 그리스어로는 'on(존재자)'과 'logos(논)'을 어원으로 하며, 전통적으로 철학 분야에서 존재론으로 명명하여 연구해 왔다. 그리고 정보관리 분야에서는 '사람의 마음 속에 존재하는 내재적 생각이나 외재적 세계의 현상에 대하여 공유하는 개념을 명확하고 명시적으로 정의하고 규정하는 것'이라 정의하고 있다(고영만, 서태설, 2005). 철학에서 존재자들(물리적, 현상적, 개념적, 추상적, 감성적)의 본성과 존재자들 간의 관계를 정의하는 온톨로지의 개념을 활용하여 웹자원을 정의하고 개념간의 관계를 명세화한 것이 시맨틱 웹의

개념이다.

시맨틱 웹을 구현하기 위해 웹상에 존재하는 정보자원에 URI를 할당하여 식별하고, 자원이 가지는 속성과 자원 간의 관계를 표현해 주기 위해 RDF(Resource Description Framework) 구조를 활용한다. RDF 자체적으로는 클래스, 클래스 간의 위계, 제약, 추론규칙을 표현할 어휘를 보유하지 않기 때문에 RDFS, OWL과 같은 웹온톨로지 언어를 사용하며, Dublin Core, SKOS, FOAF와 같은 표준 어휘를 사용하여 자원의 표현하고 상호운용성을 확충하고 있다.

온톨로지는 자원에 관한 정의와 관계를 표현하고, 웹이 더욱 의미를 가진 형태로 발전하기 위한 밑바탕이 되며, 다양한 어플리케이션에 지식베이스로 사용하기 위한 연구가 이뤄지고 있다(함영균, 서지우, 황도삼, 최기선, 2014). 이렇게 온톨로지의 활용이 많아지는 상황에서 필연적으로 온톨로지의 품질이 중요한 문제가 되었으며, 온톨로지의 품질을 평가하거나 적합한 온톨로지를 선별하는 연구가 진행되어 왔다.

온톨로지의 품질평가는 잘 구축되었다고 인식되는 기존 온톨로지와 비교하는 방법, 활용될 어플리케이션에 적용해본 결과로 평가하는 방법, 원천데이터와의 적합성·관련성을 판단해서 평가하는 방법, 미리 구축된 온톨로지 평가 준거를 적용하여 평가하는 방법을 통해 이뤄지고 있다. 그러나 기존의 평가 방법론들은 온톨로지를 활용해 응용프로그램의 성능을 향상시키는 데 치중되어 있으며, 온톨로지의 체계, 의미표현, 상호운용성 등과 같은 내재적인 영역의 평가에 대해서는 구체적인 도움을 주기 어렵다(박진수, 조원진, 노상규, 2008). 또한 다양한 온톨로지 구축 방법론이 오래전부터 연구되어 왔

으나 주로 온톨로지 구축 절차를 체계화하여 개발자들이 단계적으로 온톨로지를 개발할 수 있도록 지침을 줄 뿐, 온톨로지의 품질까지 보증한다고 보기는 어렵다.

이에 본 연구에서는 온톨로지 전문가들을 통해 온톨로지의 품질을 평가할 수 있는 항목을 제안받고, 전문가들의 검증을 통해 평가항목을 엄선하였다. 또한 평가항목을 특정 도메인의 온톨로지 평가에 직접 활용하였고 신뢰성을 검증하였다. 본 연구의 결과는 다양한 도메인에서 구축되는 온톨로지의 품질을 검증하는 평가항목으로 활용될 수 있고, 온톨로지 구축시 체크리스트로도 활용가능할 것으로 기대한다.

1.2 연구 방법 및 제한점

본 연구의 방법은 첫째, 문헌조사를 통해 온톨로지의 내재적 평가를 위한 범주를 추출하였다. 둘째, 각 범주에 대한 평가항목을 조사하기 위해 온톨로지 전문가를 선별하여 델파이조사를 수행하였다. 온톨로지 전문가들에게 개방형 질문을 통해 온톨로지의 구문적 부분, 의미적 부분, 실제적 부분에 관한 8개 하위범주의 평가항목을 수집하기 위해 1차 델파이 조사를 수행하였다. 그리고 수집된 평가항목을 검증하기 위해 2차 델파이 조사를 수행하여 삭제할 항목을 조사하였다. 또한 삭제대상 항목에 대해 전문가를 개별 접촉하여 삭제의사를 재확인한 뒤, 철회의사가 없으면 평가항목에서 삭제하였고, 최종 평가항목을 도출하였다. 셋째, 도출된 평가항목으로 기구축된 스마트폰 온톨로지 평가에 적용해 봄으로써 신뢰성을 검증하는 방식으로 최종 평가항목을 점검하였다.

본 연구의 제한점은 다음과 같다. 첫째, 연구에 가담한 온톨로지 전문가는 국내 온톨로지 전문가였으며, 그 범위를 온톨로지 관련 컨퍼런스에서 발표한 전문가, 온톨로지 관련 강의, 프로젝트를 수행하고 있는 교수나 연구자 등으로 제한한 점이다. 국외 온톨로지 전문가의 의견을 반영하지 못하였으며, 전문가의 기준은 연구자의 시각에 따라 달라질 수 있다. 둘째, 델파이 조사의 특성상 2~3회 반복적인 연구를 수행해야 하는 관계로 처음부터 명확하게 연구참여 의사를 밝힌 10명 내외의 온톨로지 전문가를 대상으로 연구한 점이다. 보다 많은 수의 전문가가 연구에 가담하면 보다 신뢰성 및 객관성이 높아질 수 있겠지만 국내 온톨로지 전문가가 많지 않은 점, 반복되는 질적연구에 가담하는 것이 까다로운 점, 질적연구의 특성상 표본의 수보다는 표본의 질이 중요한 점을 고려하여 연구결과를 이해하는 것이 필요하다.

2. 선행연구

2.1 온톨로지 구축 방법론

국·내외에서 연구된 온톨로지 구축방법론을 살펴보면, 해당 도메인의 특성에 따른 온톨로지 구축 방법을 제시하는 연구와 몇 개의 구축방법을 혼합하여 보다 보편적인 구축 방법을 제시하는 연구로 크게 구분해 볼 수 있다.

김정민, 최병일, 김형주(2006)는 철학 지식 온톨로지를 구축하면서 하향식 기법을 적용해 텍스트 기반 온톨로지 구축 방법론을 제시하였고, 이해윤(2006)은 외국어가 아닌 한국어 기반

으로 온톨로지를 구축하는 4단계의 방법을 제시하는 연구를 수행하였으며, 정태영 등(2011)은 '현대상한론'에 기재된 조문을 중심으로 상한론 온톨로지를 구축하면서 그에 적절한 4단계 구축방법론을 제시하였다. 조나혜와 강영옥(2013)은 한국의 해안지형에 관한 정보를 온톨로지 구축하면서 6개의 단계로 온톨로지를 구축하였는데, 특이한 점은 마지막 단계에서 온톨로지 평가를 텍사노미 검증과 인스턴스 검증으로 나누어 실시한 점이다. 텍사노미 평가를 위해 비일관성, 불완전성, 중복성을 검증했으며, 인스턴스 검증은 일관성, 완성도, 간결성, 확대가능성, 민감성을 평가하였다. 또 심재후와 최명길(2013)은 창업학 지식구조 분석결과를 이용해서 창업 온톨로지를 구축하면서 Noy와 McGuinness(1996), Cimiano(2006)의 온톨로지 구축 절차를 참고하여 3단계로 구축하였다.

김은경(2004)은 소프트웨어 개발 방법 표준인 IEEE Standard 1074-1997과 OTK, CommonKADS, ONIONS, Ontology Development 101 온톨로지 구축방법론을 비교하였다. 각 구축방법의 장단점을 수용, 보완함으로써 1) 사전 개발단계로서 타당성 조사단계, 2) 초기 구축단계로서 온톨로지 구축 목표 성립단계, 3) 지식 재사용을 위한 온톨로지 재사용 단계, 4) 주제 영역의 개념과 관계를 규정하는 온톨로지 구성단계, 5) 평가 검증단계, 6) 유지 관리를 포함한 사후개발 단계로 진행되는 구축방법론을 제시하였다.

Grüniger와 Fox(1994, 1995a, 1995b)는 기업에서 사용할 수 있는 보편적인 온톨로지 구축 방법인 TOVE(TORonto Virtual Enterprise)를 제안하였다. 구축 과정은 첫째, Motivation scenarios - 기업의 당면한 문제 및 온톨로지

구축을 위한 출발점 확인단계, 둘째, Informal competency questions - 온톨로지가 대답해 줄 수 있는 질문을 일상적으로 사용하는 용어로 작성, 셋째, Terminology specification - 객체(Object)와 속성(Attribute), 관계(Relation)들을 공식 명세화, 넷째, formal competency questions - 공식적으로 정의된 전문용어를 사용하여 온톨로지 요구사항 공식화, 다섯째, axiom specification - 용어의 정의를 구체화 및 공리 설정, 여섯째, Completeness theorems - 최초 단계에서 제시한 문제 해결을 판단하는 단계로 구성된다. Uschold와 King(1995)은 기업의 비즈니스 양상을 파악하고 필요사항에 대한 해결 방안 수립과 비교를 컴퓨터 기반으로 가능하게 하는 도구를 개발하는 Enterprise Project의 한 부분으로 Enterprise Ontology 구축방법을 제안하였는데, 4단계로 진행되며 TOVE의 방법과 유사하다. 또 Prieto-Diaz(2002)는 온톨로지 구축단계 중 도메인 분석(Domain Analysis)단계에서 패킷분류 이론을 접목한 접근이 가미된 온톨로지 구축방법을 제안하였고, Ra(2012)는 하향식(Top-down) 및 상향식(Bottom-up) 구축 방법의 특징을 결합한 방법론인 혼합형 온톨로지 구축방법론(Mixed Ontology Building Methodology)을 제안하였다.

Uschold(1996a, 1996b)는 많은 온톨로지 구축방법론이 제시되었지만, 특정한 도메인에 한정되어 구축된 온톨로지 구축방법론이라는 것을 지적하였다. 그리고 어떤 분야에도 적용할 수 있는 범용적인 가이드라인을 제시하기 원했다. 그래서 Uschold는 TOVE와 Enterprise 온톨로지 구축방법론을 일반화하고 병합하는 것을 시도했고, 다양한 온톨로지 구축방법론에 대한 문

현조사를 통해 자신의 방법론에 세부적인 단계마다 참조하여 개선하는 연구를 수행하였다.

Gomez-Perez, Fernandez, De Vicente(1996)와 Fernandez, Gomez-Perez, Juristo(1997)는 Methontology라는 온톨로지 구축 방법을 제안하였다. Methontology는 명세화(specification), 지식획득(knowledge acquisition), 개념화(conceptualization), 통합(integration), 구현(implementation), 평가(evaluation), 문서화(documentation)의 일곱 단계로 구성된다.

여기까지 살펴본 온톨로지 구축방법에 관한 연구들은 온톨로지 구축절차를 소개하고 도메인에 따른 특성을 중심으로 기술된 것이 많으며 온톨로지 구축에 효율성을 줄 수는 있지만 품질에 대한 보증을 하지는 않는 것으로 판단된다. 구축방법 내에 평가에 대한 부분이 존재하기도 하나, 조나혜와 강영옥(2013)이 비일관성, 불완전성, 중복성, 일관성, 완성도, 간결성, 확대가능성, 민감성을 평가한 것 외에는 주로 온톨로지를 통해 구축목적이 얼마만큼 성취되는가를 평가하는 방식이 주를 이루었고, 온톨로지의 내적 품질과 완성도를 평가하는 연구는 찾아보기 힘들었다.

2.2 온톨로지 평가 방법론

온톨로지의 품질을 평가하는 방법론을 4가지 카테고리로 분류할 수 있다(Brank, 2006). 첫째, 잘 구축되었다고 판단되는 온톨로지(gold standard)와 새롭게 구축된 온톨로지와 비교하는 방법이다. 둘째, 구축된 온톨로지를 실제 적용될 응용프로그램에 적용해 봄으로써 온톨로지의 기능이 제대로 되는지를 판단해 보는 방

법이다. 셋째, 온톨로지와 그것이 소스(source)로 삼고 있는 데이터 집합사이의 적합성 및 관련도를 평가함으로써 온톨로지의 질을 평가하는 방법이다. 넷째, 미리 정의된 지표(criteria)나 표준(standard), 필요사항(requirement)을 가지고 구축된 온톨로지를 전문가들이 평가하는 방법이다.

기존의 온톨로지와 새롭게 구축된 온톨로지와 비교하는 방법으로서 Maedche와 Staab(2002)는 온톨로지를 확장하거나 수정, 혹은 더 큰 범위의 온톨로지와 비교할 경우, 어떤 기준으로 수행할 것인지 프레임워크를 제시하였는데, 어휘적 비교(Lexical comparison)와 위계적 비교 및 관계적 비교(Taxonomies and relations comparison)를 중심으로 실증적 연구를 수행하였다.

구축된 온톨로지를 실제 적용될 응용프로그램에 적용해 보는 방법으로서 Porzel과 Malaka(2004)는 효과적인 온톨로지란 주어진 과업환경에서 수행능력 평가를 통해 가려질 수 있다는 비교적 단순한 논리를 가지고 평가하고자 하는 온톨로지를 과업환경에 적용하여 개념어휘(concept vocabulary), 위계 및 인스턴스(hierarchy/granularity), 의미관계(semantic relation)를 측정하여 과업기반의 온톨로지 평가(task-based ontology evaluation)를 시도하였다.

온톨로지와 온톨로지가 원천 데이터로 삼고 있는 데이터집합 사이의 적합성 및 관련도를 평가하여 온톨로지의 질을 평가하는 방법에 관한 연구는 Brewster, Alani, Dasmahapatra, Wilks(2004)의 데이터 기반 온톨로지 평가(Data-driven Ontology Evaluation)에 관한

연구가 있었다. 이 연구는 크게 4단계로 이루어 지는데, 첫째, 시나리오 단계이다. 실험자가 온톨로지가 사용될 특정 애플리케이션과 주제분야에서 필요로 하는 요구사항에 대해 고려하되, 도메인 내 용어들을 추출한다. 그리고 현존하는 온톨로지 내에서 애플리케이션과 함께 사용할 수 있는 가장 적합한 것들을 선별하여 놓는다. 두 번째 단계에서는, 첫 번째 단계에서 추출한 용어들과 연구대상 온톨로지 및 비교대상 온톨로지와의 자동매핑을 테스트한다. 추출한 용어에는 있지만 온톨로지에는 없거나, 온톨로지에는 있지만 추출한 용어에는 없는 것을 기준으로 단순비교를 수행하거나 벡터비교를 수행한다. 세 번째로는 보다 현학적인 방법으로 온톨로지 평가의 적합성을 얻는 단계이다. 이에 대해 세 가지 단계가 수행된다. 상세화하면 ① 키워드 식별, ② 질의문 확장, ③ 온톨로지 매핑이다. 네 번째 단계로는 개연성에 의한 접근법으로 온톨로지를 평가한다. 이병길과 김희섭(2013)은 온톨로지의 소스 데이터가 되는 국가기록원 나라기록포털 '새마을운동'에 관한 기록물 376건을 대상으로 토픽(topic, 혹은 질의)에 적합한 기록물을 미리 선정해 놓은 뒤 기존 검색시스템과 온톨로지 기반 검색시스템을 비교하여 평가하였다.

미리 정의된 지표, 표준, 요구사항을 가지고 전문가들이 온톨로지 평가를 수행하는 연구로서 조나혜와 강영옥(2013)은 지리개념의 이해를 돕는 지오 온톨로지를 구축하면서 제작한 온톨로지를 텍사노미와 인스턴스 측면에서 8개의 하위범주로 평가영역을 구성하여 4명의 온톨로지 전문가들과의 인터뷰를 통해 검증하였다. Lozano-Tello와 Gómez-Pérez(2004)는 AHP

기법(Analytic Hierachy Process method)을 이용하여 적절한 온톨로지를 선별하기 위한 방법론인 ONTOMETRIC 평가도구를 개발하였다. ONTOMETRIC은 도구, 언어, 내용, 방법론, 비용적 측면에서 17개의 하위항목과 매우 세분화된 차하위 항목들로 평가항목을 수립하였다. 박진수, 조원진, 노상규(2008)는 기호학 이론에 기초하여 구문적(syntactic), 의미적(semantic), 실질적(pragmatic), 그리고 사회적(social) 측면에서 온톨로지의 질을 평가한 Burton-Jones 등(2005)의 프레임워크 연구를 참고하여 온톨로지 특성을 평가하는 프레임워크를 제시하였다.

2.3 온톨로지 성능을 통한 평가

대부분의 온톨로지 구축과정에서 온톨로지를 검증하기 위해 평가를 실시하는데, 검색결과나 추론에 적용해서 성능을 평가하는 방법이 주로 이뤄지고 있다.

Noh(2011)는 제나(Jena) 검색엔진을 사용해 구축한 온톨로지를 정확률(Precision)과 재현률(Recall)을 이용해 성능을 평가하였다. 김수경과 안기홍(2007)은 온톨로지의 검증을 위해 온톨로지 추론 실험을 수행하여 실험 시스템과 비교 대상 시스템 간 재현율과 정확율을 비교하였다. 이정희와 김희섭(2007)은 기존의 전자기록관리시스템과 온톨로지 기반의 검색시스템을 장문 10개와 단문 10개로 이뤄진 질의문에 대해 재현율과 정확률로 그 성능을 비교하였다.

또 김수경과 안기홍(2008)은 표준화되지 못한 온톨로지 평가방법의 문제점을 지적하며, 상하위

개념 간 관계설정에 대해 포섭(Subsumption) 검증, 특정개념과 사례 간의 정확한 추론에 대해 인스턴스 검증, 용어적 공리와 선언적 공리의 추론결과에 대해 일관성(Consistency) 검증을 수행하였다. 이정연, 이재윤, 정한민, 강인수, 신숙경(2007)은 심사자 자동추천을 위하여 온톨로지를 구축하고 전문가들을 통해 온톨로지가 심사자를 추천하는 방식과 배제 규칙에 대해 검증하는 방식으로 연구를 수행하였다.

이상의 온톨로지 구축방법 내의 평가, 평가방법론, 온톨로지 성능을 통한 평가방식은 구체적인 내용을 제시하는 데 미비하였고, 온톨로지 평가범주는 제시하였으나 구체적인 온톨로지 품질평가항목이 미비하였으며, 또 재현률과 정확률을 통한 성능평가가 주를 이루어, 체계적으로 온톨로지의 품질을 평가할 수 있는 방법론이 필요한 것으로 여겨진다.

3. 온톨로지 평가항목 추출을 위한 기본범주의 선택

이상의 선행연구 조사를 통해서 기본적인 범주를 가지고 온톨로지를 평가한 연구와 범주를 파악할 수 있었다. 조나혜와 강영옥(2013)이 온톨로지의 텍사노미 검증과 인스턴스 확인을 위해 비일관성, 불완전성, 중복성, 일관성, 완성도, 간결성, 확대가능성, 민감성을 평가한 연구와 Lozano-Tello와 Gómez-Pérez(2004)의 ONTOMETRIC, Burton-Jones 등(2005)이 수행한 연구를 참고하여 만든 박진수 등(2008)의 평가방법이 그 예이다. 이들을 보다 심층적으로 비교분석하여 온톨로지 평가를 위한 범주

를 선택할 필요가 있다. 이는 기 수행된 연구와 연계를 갖도록 하여 객관성을 확보하고, 연구에 소모되는 시간과 비용 절감에 도움을 줄 것으로 기대된다.

조나혜와 강영옥(2013)은 온톨로지 평가를 위해 클래스 및 클래스의 계층구조가 일관성을 띄는지, 범위가 적절하고 중복되는 것은 없는지 확인하였고, 또 인스턴스의 명명체제와 URI체제가 일관되는지, 수량은 충분하며 중복되는 것은 없는지, 차후 확장을 고려하여 설계되었는지 등을 검증하였다(〈표 1〉 참조).

〈표 1〉 조나혜와 강영옥(2013)이 온톨로지 평가를 위해 사용한 평가범주

대분류	소분류
텍사노미 평가	비일관성(inconsistency)
	불완전성(incompleteness)
	중복성(redundancy)
인스턴스 확인	일관성(consistency)
	완성도(completeness)
	간결성(conciseness)
	확대가능성(expandability)
	민감성(sensitiveness)

이 평가범주는 온톨로지 평가를 텍사노미와 인스턴스 평가로 집중하여 평가대상의 모호성을 배제하고, 평가범주 간 중복이 발생하지 않도록 세밀하게 설계되었다고 평가된다. 그러나 평가범주를 좁은 범위로 제한하였기 때문에 온톨로지 내 개념과 속성의 정의, 개념들 간의 관계, 제약사항의 정의 등 더 많은 영역에서 평가항목이 추출되는 것을 제한할 가능성이 있다. 또 텍사노미 평가의 하위범주는 비일관성, 불완전성, 중복성과 같이 부정적 의미의 범주가 사용되었고, 인스턴스 확인의 하위범주는 일관성,

완성도, 간결성 등 긍정적 의미의 범주가 사용되어 평가항목 추출시 일관성이 저해될 가능성도 있다고 평가된다.

Lozano-Tello와 Gómez-Pérez(2004)는 온톨로지 기반의 어플리케이션이 많아지는 상황에서 적절한 온톨로지를 선별하기 위한 방법론인 ONTOMETRIC을 개발하였다. ONTOMETRIC은 도구(Tool), 언어(Language), 내용(Content), 방법론(Methodology), 비용(Cost)적 측면에서 17개의 하위범주와 언어(Language) 부분에서 9개의 차하위 범주로 구성되었고, 평가 범주에 대한 129개의 세분화된 평가항목들을 제시하여, 총 160개 텍사노미 구조로 이뤄진 평가항목을 제시하고 있다. 범주의 내용은 <표 2>와 같고, 비용(Cost)에 대한 하위범주가 없는 관계로 본 표에서는 생략하였으며, 평가항목도 생략하였다.

Lozano-Tello와 Gómez-Pérez(2004)가 제안한 평가범주는 먼저 도구(Tool)부분에서 온톨로지 생성, 시각화, 유지관리, 협업 등에 관한

제 영역을 다루고 있으며, 언어(Language) 부분은 온톨로지의 문법, 위계체계, 공리, 추론의 내용을 포함하고 있다. 내용(Content)부분은 온톨로지가 다루고 있는 개념 및 관계의 개수, 계층구조의 단계, 하위 클래스의 수를 다루며, 방법론(Methodology) 부분은 명세서가 정확히 구성되어 있는지, 매뉴얼 구성이 예제와 함께 되어있는지, 생성된 온톨로지의 수와 품질이 우수한지를 평가하는 내용을 다루고, 비용(Cost) 부분은 온톨로지를 사용하는 경우 어떤 라이선스 비용을 평가하고 있다. 이 평가방법은 제작한 온톨로지의 품질을 측정하는 것보다는 적절한 온톨로지를 선별하여 사용하기 위해 시각화, 데이터 반입, 데이터 반출기능, 사용성, 매뉴얼 구성, 비용과 같은 측면을 고려하고 있다.

박진수 등(2008)은 온톨로지 추출도구의 성능을 평가하기 위한 프레임워크를 제안하였는데, 추출 후의 온톨로지 특성 평가부분에서 구문적 부분, 의미적 부분, 실제적 부분을 온톨로지 품질평가를 위한 범주로 제시하였다. 온톨로

<표 2> ONTOMETRIC이 사용한 온톨로지 평가범주

대분류	중분류	대분류	중분류	소분류
도구 (Tool)	성능	언어 (Language)	도메인 지식	개념/인스턴스
	시각화			속성
	에디션			패킷
	상호작용			관계
	방법론적 측면			텍사노미
	협업적 측면			공리
	전환기능			제작원칙
	통합기능		추론원리	추론가능성
내용 (Content)	개념	방법론 (Methodology)	정확도	추론엔진
	관계			사용성
	텍사노미			완성도
	공리			

〈표 3〉 박진수 등(2008)의 평가 프레임워크 품질평가 관련 범주

프레임워크 내 단계	특성	세부 평가 범주	
추출후	온톨로지 특성	구문적 부분	합문법성
			풍부성
		의미적 부분	해석 가능성
			일관성
			명료성
		실제적 부분	정확성
			포괄성
			관련성

지 추출도구 평가 프레임워크 중 온톨로지 품질 평가에 관련된 부분은 〈표 3〉과 같다.

박진수 등(2008)이 제안한 프레임워크 중 온톨로지의 품질에 관한 범주들은 문법적으로 정확하고 풍부한 표현력을 활용했는지 평가하는 구문적 부분과 문법적인 문제를 넘어 의미 전달에 관해 해석이 명확하고 전체적으로 일관되며 명료하게 표현하였는지를 평가하는 의미적 부분, 그리고 온톨로지가 다루고 있는 내용이 정확하고 범위가 적절하며 타 온톨로지와의 연계를 다루는 실제적 부분으로 구성되어 있다. 크게 3부분으로 온톨로지 평가 범주를 구분한 내용이 온톨로지의 특성을 잘 반영하도록 구성되었고 하위 범주도 각각 2~3가지로 크게 구분하여 필요에 따라 다양한 평가항목들을 담을 수 있게 설계되었다고 평가된다.

박진수 등(2008)이 제안한 평가범주가 범위 해석 부분에 있어 혼동이 있을 수 있으나, 조나혜와 강영옥(2013)이 제안한 평가범주가 텍사노미와 인스턴스에 치중한 점, Lozano-Tello와 Gómez-Pérez(2004)가 제안한 범주가 지나치게 항목이 많고 복잡한 점을 고려할 때, 박진수

등(2008)이 제안한 평가범주가 가장 활용가능성이 높은 범주로 판단되었다. 온톨로지 평가항목 추출을 위한 온톨로지 범주를 결정하기 위해, 3장에서 조사한 세가지 평가범주를 적절히 혼합하여 사용할 것도 고려하였으나, 범주 간 중복될 가능성이 높아서 본 연구에서는 박진수 등(2008)이 제안한 평가범주를 그대로 사용하기로 하였다.

4. 온톨로지 평가항목추출을 위한 델파이 조사

본 장에서는 앞서 선별한 온톨로지 평가 범주를 토대로 범주별 온톨로지 평가항목을 추출하였다. 온톨로지 평가항목을 추출하기 위해 온톨로지 전문가 패널을 구성하였고, 총 2회의 델파이 조사를 실시하였다. 제1차 델파이 조사는 2015년 2월 11일부터 2015년 2월 27일까지 총 17일에 걸쳐 패널 11명을 대상으로 실시하였고, 제2차 델파이 조사는 2015년 3월 2일부터 2015년 3월 11일까지 총 10일 동안 실시하였다.

〈표 4〉 차수별 델파이 조사 회수율

확정된 패널 표본수 (명)	제1차 델파이 조사		제2차 델파이 조사	
	최종 응답수(명)	응답률(%)	최종 응답수(명)	응답률(%)
11	11	100.0	9	81.8

제 2차 델파이 조사이후 항목별 삭제 및 수정사항에 대해 전문가들에게 개별 이메일, 전화, 인터뷰를 통해 항목 오해 여부를 확인하고 삭제 의사 및 수정내용을 재확인 하였다. 델파이 조사를 위한 설문지의 작성, 배포와 수집은 구글 드라이브의 서베이를 이용한 설문지 작성, 이메일을 통한 배포, 그리고 구글드라이브를 통한 실시간 수집방식으로 결과를 종합하였다. 제1차 델파이 조사 회수율과 제2차 델파이 조사 회수율은 〈표 4〉와 같으며, 1차 이후 전문가들의 일정으로 인해 2명이 제외되었다.

4.1 온톨로지 전문가 패널

델파이 조사방법은 이론적 배경이 확실하지 않은 내용에 대해 전문가들의 권위를 신뢰하며

그들의 합치된 의견을 도출하는 방법이므로, 적합한 전문가들의 선택이 필수적이다. 이종승(2009)은 델파이 조사를 위한 전문가의 기준으로 주어진 질문에 요구되는 지식이 평균이상이어야 하며, 편향되지 않은 사고를 할 수 있는 지적능력 및 객관성에 대한 기준과 함께 지리적으로 한 지역으로 편중되지 않을 것과 2~3회 반복되는 조사기간 중 시간을 낼 수 있어야 한다는 실제적인 조건을 제시하였다. 이에 따라 본 연구에서는 〈표 5〉와 같이 전문가 기준을 가지고 전문가를 선별하였다.

〈표 5〉와 같은 전문가 선정기준에 따라 본 연구에서 선별한 전문가 패널 목록은 〈표 6〉에서와 같이 다양한 현장에서의 직급, 전문성과 함께 확인할 수 있다.

〈표 5〉 온톨로지 전문가 선정 기준

항목	조건
온톨로지관련 경력	<ul style="list-style-type: none"> • 주제분야에 관계없이 공개된 온톨로지 구축 경험이 있는 사람 • RDF/OWL 및 각종 open vocabulary에 대한 지식이 있는 사람
전공분야	<ul style="list-style-type: none"> • 온톨로지의 개념이 철학에서 파생하였고, 온톨로지 모델링 및 실구축은 문헌정보학, 산업공학, 컴퓨터 공학을 비롯한 다양한 분야에서 이뤄지고 있음 • 본 연구에서 전문가의 전공에는 제한을 두지 않음 • 특정 전공에서만 전문가를 추출하여 연구에 편향을 줄 수 있기 때문
실험참여 가능성	<ul style="list-style-type: none"> • 선행연구에 의하면 델파이 3회 조사의 경우 최소 45일 정도의 시간이 소요됨 • 따라서 45일간의 조사기간 동안 접촉 할 수 있고 3회까지 응답이 가능한지 이메일 및 유선을 통해 확실히 파악된 사람만 전문가 패널로 인정

〈표 6〉 온톨로지 평가항목 추출을 위한 델파이 조사 패널 내역

번호	직급	전문성	비고
1	부장	국립중앙도서관 생물정보, 의학도서관 등 온톨로지 구축 및 서비스 구축 외 다수 / 대학에서 시맨틱 웹 강의 / KSWC 및 LODAC 등 관련 행사에서 연 2~3차례 발표 / 국내·외 관련 논문 20편 이상 투고	
2	부장	국립중앙도서관 디지털도서관 전문연구(2007-2013) / 한중일 국립디지털도서관 프로젝트 진행 중 국립중앙도서관 LOD 프로젝트 수행 및 서지, 주제, 저자명 모델링 / 온톨로지 주제로 석사학위논문 / KISTI 도서관 링크드 데이터 사례 지식리포트 작성(2013.9) / 온톨로지 관련 기업 기술연구소 국가 링크드 데이터 포털 PM	
3	부장	국립중앙도서관 온톨로지 모델링 및 구축 / 국립수목원 온톨로지 구축 / 한국과학기술정보연구원 NDSL 온톨로지 구축 / RDF 관리 시스템 온톨로지 구축 / TV 온톨로지 모델링 및 구축 / NTIS 온톨로지(초기버전) 모델링 및 구축 / KDATA 온톨로지 모델링 및 구축 / 사내 직원 관리 온톨로지 모델링 및 구축 / 트리플 변환 온톨로지 모델링 및 구축 외 다수	
4	교수	다수의 온톨로지 구축 / 추론엔진 설계 / 논문 다수	
5	대표이사	다수의 온톨로지 모델링 및 실구축 / Triple Store 제작 기술 보유 / 온톨로지 강의 경험	
6	차장	시맨틱 웹을 디자인하고 구축하는 업무 담당 중 / 링크드 데이터, AI, 데이터마이닝을 위해 온톨로지를 설계하고 활용 업무 중 / 석사논문 및 관련논문 다수	
7	차장	소프트웨어 엔지니어로서 온톨로지 관련 프로젝트를 수 차례 수행 / 온톨로지 모델링, 데이터 구축 / 온톨로지 기반 기술 및 모델링 강의 및 발표	
8	과장	인공지능 석사 / 관련 실무 경력 8년 / 온톨로지 연관 연구 6건 / 온톨로지 연관 프로젝트 2건 / 온톨로지 연관 논문 3건 / 온톨로지 연관 저술 1건	1차 설문 이후탈락
9	차장	온톨로지 프로젝트 수행하며 다양한 도메인 모델링 및서비스 구축 / 스마트폰 내부 데이터를 활용한 모델 설계, 공공서 민원 데이터를 활용한 모델 설계, 연구시설장비를 활용한 모델 설계 / 정보공개 관점에서 행정, 시설, 문화재, 기술정보, 지리정보, 물 관련 정보 등에 대해 LOD 구축을 위한 모델 설계 및 구축	
10	차장	시맨틱웹 기반 프로젝트 다수 수행 / 온톨로지 구축방법론 SBM5 구축 / 미래부 미래학습플랫폼 과제 - 학습정보 스키마 설계 및 LOD 구축 / 미래부 엑소브레인 과제 한국어 지식베이스 모델링 및 구축 중	1차 설문 이후탈락
11	교수	bibleontology 모델링 및 구축(lod cloud 등재) / 국립중앙도서관 근대문학 온톨로지 구축 / 차세대융합기술연구원 Linked Data / Linked Data 관련 논문 저술 / 온톨로지 강의 다수	

4.2 제1차 델파이 조사

제1차 델파이 조사에서는 〈표 7〉에서 보는 것처럼 박진수 등(2008)이 제안한 프레임워크 중 온톨로지의 품질에 관한 범주들을 활용하여 개방형으로 전문가들의 의견을 수집하였다. 제1차 델파이 조사에서 수집된 답변을 종합하여 최초의 평가항목을 [부록 1]에 정리하였다. 제

1차 델파이 조사를 통해 추출한 온톨로지 평가항목은 합문법성 13개 항목, 풍부성 8개 항목, 해석가능성 14개 항목, 일관성 9개 항목, 명료성 9개 항목, 정확성 7개 항목, 포괄성 7개 항목, 관련성 5개 항목이었다. 대분류를 중심으로 요약하면 〈표 8〉과 같다. 추출된 항목들은 제2차 델파이 조사 질문지에 반영되었다.

〈표 7〉 박진수 등(2008)의 연구를 활용한 델파이 1차 개방형 질문

박진수 등 (2008)의 프레임워크		1차 델파이조사 질문	
구문적 부분	합문법성	구축된 온톨로지의 구문이 작성된 언어 규칙에 어느 정도까지 맞게 작성되었는가?	구문적으로 합문법적인 온톨로지라면 어떤 요소들을 가진다고 생각하십니까?
	풍부성	구축된 온톨로지가 개념들과 공리들을 모두 포함하는가? 아니면 개념들만 가지고 있는가?	구문적으로 풍부한 온톨로지라면 어떤 요소들을 가진다고 생각하십니까?
의미적 부분	해석 가능성	온톨로지에서 표현된 언어가 현실에서도 실제 사용되는 올바른 의미를 가지고 있는가?	해석에 어려움이 없고, 의미가 잘 통하는 온톨로지는 어떤 요소들을 가진다고 생각하십니까?
	일관성	그 용어가 온톨로지 속에서 일관된 의미를 가지고 있는가?	일관성 있게 의미를 전달하는 온톨로지는 어떤 요소들을 가진다고 생각하십니까?
	명료성	용어의 맥락이 분명한가?	용어가 나타내는 것이 명확한 온톨로지는 어떤 요소들을 가진다고 생각하십니까?
실제적 부분	정확성	구축된 온톨로지에 표현되고 있는 개념이나 관계가 실제 사실인가?	정확한 온톨로지가 되기 위해 구체적으로 가져야 할 요소는 무엇이라고 생각하십니까?
	포괄성	구축된 온톨로지가 충분한 범위를 포괄하고 있는가?	온톨로지가 필요로 하는 충분한 범위를 포괄하기 위해 가져야 할 요소는 무엇이라고 생각하십니까?
	관련성	온톨로지가 사용자의 특정 필요사항을 만족시키는가?	온톨로지가 관련성이 높기 위해 가져야 할 요소는 무엇이라고 생각하십니까?

〈표 8〉 대분류 별 온톨로지 평가항목 수

대분류	소분류 및 개수	합계
구문적 부분	합문법성 13개	21개
	풍부성 8개	
의미적 부분	해석가능성 14개	30개
	일관성 9개	
	명료성 9개	
실제적 부분	정확성 7개	19개
	포괄성 7개	
	관련성 5개	
총합계		70개

4.3 제2차 델파이 조사

제2차 델파이 조사는 제1차 델파이 조사를 통해 추출된 70개의 온톨로지 평가 항목을 검증 받는 형태로 진행하였다. 조사에 참여한 전문가 패널은 온톨로지 평가항목에 대해 삭제, 표현 수정, 항목유지에 대한 의사를 표현할 수 있었고,

제2차 델파이 조사 중 패널들이 삭제를 요구한 항목에 관해서는 원칙적으로 삭제하는 것으로 하였다. 델파이 조사 방법이 실험에 참여하는 전문가들의 전문성을 신뢰하고 의견이 합치하는 것을 찾고자 하기 때문이다. 단, 전체 패널 수의 10% 이하인 소수의 전문가들이 삭제 요구를 한 경우, 삭제요구를 한 소수의 전문가에게 상황을

알리고 삭제 철회의사가 변함이 없는지 확인하는 검증과정을 거쳤다. 삭제된 항목과 주요 이유에 대해서는 평가 범주 별로 <표 9>, <표 10>, <표 11>에 정리하였다. <표 9>에서 보는 것처럼 구문적 부분에서는 9개의 삭제요청이 제안되었는데, 합문법성에서는 문법성의 궁극적 목적에서 벗어난 항목과 문법성의 범위를 기준으로 삭제 요구가 있었다. 풍부성에서는 선택적으로 적용할 수 있는 추론을 필요조건으로 오해하게 하는 항목과 함께 풍부성에 대한 객관적인 기준으로 보편적이지 않다고 판단하는 항목에 대해 삭제 요구가 있었다. <표 10>에서 제시된 의미적 부분

의 삭제 요구는 dereferencing과 같이 즉시 확인하기 어렵거나, 웹상으로 관련 문서 발행과 같이 반드시 필요하다고 보기 어려운 항목이거나, 의미전달이 불분명한 항목에 대한 삭제 요구가 있었다. <표 11>은 실제적 부분에서의 삭제 요구와 이유에 대한 것인데 추론 결과의 참, 거짓 판별성과 상위 온톨로지 확보가 좋은 온톨로지의 조건이 아닌 것을 이유로 삭제요구가 있었다. 특히 포괄성 영역에서 차후 확장가능성에 대한 항목은 다수의 전문가가 찬성하였으나 확장가능성 때문에 결과값이 모호해지며 명료성이 저해될 수 있다는 소수의 의견이 반영되어 삭제되었다.

<표 9> 제 2차 델파이 조사를 통해 구문적 부분에서 삭제된 항목 및 이유

분류	삭제 항목명	삭제 이유
합문법성	개발되는 어휘는 실제계의 개념을 충실히 반영하면서도 문법적으로 이상이 없이 작성되어 있는가?	이 항목은 합목적성 부분에서 궁극적으로 평가해야 하는 내용인데 하나의 항목으로 물어보기에는 문제가 있다고 평가됨
	기존의 용어집을 수정하여 활용하는 경우, 필요에 맞게 수정하되 문법적으로 정확하게 수정하였는가?	기존의 용어집을 참고하려면 정확하게 따라서 사용하는 것이 좋고, 아니라면 새롭게 설계하는 것이 좋으므로 이 항목은 옳지 않은 개념을 담고 있음
	Class의 restriction은 실제계를 잘 반영하면서도, 합문법적으로 작성되어 있는가?	실세계를 반영하고 있는가 여부를 확인하는 것은 문법성 확인에 부적절한 항목임
	objectProperty를 사용할 때 더 구체적으로 표현력을 올리기 위한 조치를 하고 있는가? (예를들면, owl:SymmetricProperty 등)	표현력과 합문법성은 별개의 관점이라 평가되었고, 풍부성 영역으로 이동하도록 제안하여 풍부성 항목으로 이동 조치
	Class와 property 작성시 네이밍 규칙을 잘 지키고 있는가?	네이밍 규칙은 문법적인 것과 별개로 특정 영역에서 정한 규칙이므로 합문법성의 조건이 될 수 없음
풍부성	다양한 추론(inference configuration)을 적절하게 활용하는가? (symmetricity 또는 intransitivity 등)	추론이 적절하게 활용하였는지를 풍부성으로 평가한다면 추론이라는 것 자체가 반드시 해야 된다는 의미로 받아들일 수 있음. 추론은 필요에 따라서 하는 것이지 어떤 평가에 만족하기 위해 추론을 하는 것은 옳지 않다고 판단함
	타 온톨로지 혹은 타 객체로 이동할 수 있도록 설계되었는가?	온톨로지는 활용 목적에 따라 특정 도메인 로컬 시스템에서만 활용 가능할 경우도 있으므로 이 항목과 적정하지 않고 속성을 얼마나 풍부하게 정의하고 있는가가 타당한 요건임
	subject와 object에 해당하는 객체들은 literal이 아니라 URI를 갖는 객체로 표현되었는가?	URI로 표현될 때 연결성이 좋아져서 내용이 풍부해지는 것이 문법적으로 풍부해 지는 것은 아니기에 삭제하는 것이 적절함
	시맨틱 웹 표준으로 제안된 구문들을 충분히 활용하였는가?	시맨틱 웹 표준(W3C 기준)은 권고사항으로 풍부성 보다는 합문법성과 관련된 항목으로 풍부성 측정을 위해서는 타당한 항목이라 보기 힘들

〈표 10〉 2차 델파이 조사를 통해 의미적 부분에서 삭제된 항목 및 이유

분류	삭제 항목명	삭제 이유
해석 가능성	기계가 읽고 처리할 수 있는 수준의 엄격한 스키마 정의가 설정되었는가?	의미적 해석이 스키마 정의로만 이뤄지는 것이 아님
	Class 명에서 상하위 클래스를 이해할 수 있도록 지어졌는가?	상하위 클래스 명을 이해할 수 있는 것이 반드시 필요하지 않음
	property 명에서 domain과 range가 이해될 수 있도록 지어졌는가?	반드시 필요하지는 않은 내용이므로 삭제
	Class와 property를 정의한 문서가 웹에 발행되어 있는가?	웹에 발행될 수 없는 경우도 많이 존재하기 때문에 이 항목으로 해석가능성을 논하기 어려움
	인스턴스 명에서 Class 타입이 이해될 수 있도록 지어졌는가?	반드시 필요한 항목이 아니라고 파악됨
	Class 또는 property 간에 상하위 관계를 잘 구축했는가?	상하위 클래스명을 이해하는 것과 domain, range의 범위 등은 반드시 필요하지 않음(모든 온톨로지 설계에 공통적으로 필요한 요소는 아님)
일관성	상위 클래스의 property로 잘 정의되었는가?	질문의 의미가 모호하며 중요하지 않은 내용이라고 판단됨
	하나의 Class에 너무 많은 인스턴스가 집중되지 않도록 모델 설계하였는가?	경우에 따라 인스턴스가 집중되는 경우가 있으므로 적절하지 않음
명료성	URI로 표현되는 객체가 상·하·좌·우 관계가 명확하게 표현되는가?	URI에 표현된 개념은 모델링하고자하는 도메인 전문가 혹은 일반인들이 이미 인식하고 있는 세계관에 따라 판단이 되는 항목으로 URI 명명화로 명료성 판단을 하기는 부적절하며 원천소스가 존재할 경우 원천소스의 통제어휘 등이 반영된 것인지를 확인하는 것이 필요함
	URI가 dereferencing이 가능하여 의미를 명확히 알리게 되도록 작성되었는가?	구축된 데이터의 현 상태로 판단하기 어려우며 발행된 후에 확인 가능한 항목으로 판단됨

〈표 11〉 2차 델파이 조사를 통해 실제적 부분에서 삭제된 항목 및 이유

분류	삭제 항목명	삭제 이유
정확성	온톨로지를 기반으로 추론을 수행했을 때, 참·거짓으로 명확히 판별가능한 추론결과가 도출되는가?	추론의 결과가 참거짓으로 표현되는 것이 아님
포괄성	도메인을 표현하는데 충분히 활용할 수 있는 upper온톨로지를 확보하였는가?	upper온톨로지가 있다고 해서 좋은 온톨로지는 아님
	온톨로지 구축 이후의 데이터의 관리 및 운영측면을 고려하여 Class 및 property를 결정하였는가?	온톨로지 구축은 현 시점에서의 문제해결을 위한 모델링화로 향후 확장성을 지나치게 고려하여 class, property를 결정할 경우 포괄성은 증가할 수 있으나 모델링에 따른 질의결과의 모호성이 증가할 수 있어 부적절함
관련성	owl:inverseFunctionalProperty를 활용하여 특정 개념에 대해 흠어진 지식을 모을수 있도록 하고 있는가?	owl:inverseFunctionalProperty의 목적이 흠어진 개념을 모으는 것만이 아니기 때문에 객관적인 평가항목으로 제시하기는 어려움

이상과 같이 온톨로지 전문가 패널을 통한 검증과정을 거쳐 최종 완성된 온톨로지 평가 항목은 〈표 12〉, 〈표 13〉, 〈표 14〉와 같다. 먼저 〈표

12〉에서 구문적 부분은 합문법성 8개 항목, 품부성 4개 항목으로 총 12개 항목으로 구성되었다. 합문법성에서는 W3C 권고안의 준행, 텍사

〈표 12〉 구문적 부분(합문법성, 풍부성) 최종 평가항목

대분류	중분류	평가항목명
구문적 부분	합문법성	W3C 권고안(RDF, OWL)을 충실히 이행하였는가?
		Class의 계층관계가 재귀적인 문제없이 잘 구성되었는가?(재귀: 원래의 자리로 되돌아가거나 되돌아 오)
		property를 만들 때 objectProperty와 datatypeProperty를 잘 구별하였는가?
		objectProperty를 만들 때 domain과 range가 적절히 설정되었고, 그에 맞게 사용되었는가?
		rdfs:range에서 boolean, interger 등 치역이 합법적으로 취해지도록 조치가 되어 있는가?
		기존의 용어집(foaf, skos, skos-xl) 등을 활용한 경우, 용어의 명세에 맞게 활용하고 있는가?
		is-a, has-a와 같이 방향성을 가지는 property의 표현이 명확히 이뤄졌는가?
		인스턴스들이 모델링 규칙에 잘 맞게 합문법적으로 입력되었는가?
	풍부성	Class 간의 관계, 인스턴스의 property를 풍부하게 표현하고 있는가?
		잘 알려진 어휘(DC, SKOS 등)를 활용하여 property를 표현하였는가?
		온톨로지의 목적에 맞는 수준으로 어휘를 적절하게 활용하였는가?
		해당 도메인의 현실세계 모습을 충분히 반영하기 위해 공리(axiom)나 제약조건(restriction)을 잘 식별하여 반영하였는가?

〈표 13〉 의미적부분(해석가능성, 일관성, 명료성) 최종 평가항목

대분류	중분류	평가항목명
의미적 부분	해석 가능성	URI가 직관적으로 이해하기 좋게 명명되어 있는가?
		URI 충돌을 피하기 위해서 숫자로 된 URI를 만들었다면, rdfs:label과 같은 레이블을 활용해서 이용자들이 이해할 수 있게 구성하였는가?
		Class 명이 직관적이면서 명확히 이해가 되도록 명명되어 있는가?
		property 명이 직관적이면서 명확히 이해가 되도록 명명되어 있는가?
		Class명이 real world의 concept name으로 잘 표현되어 있는가?
		인스턴스 명이 real world의 concept name으로 잘 표현되어 있는가?
		property명이 수여, 소유, 가용 등의 방향성을 잘 표현한 이름으로 명명되었는가?
		domain과 range를 명시하여 명확한 의미를 전달하는가?
		인위적으로 anonymous class를 사용하지 않고 명확한 어휘로 표현되었는가?(anonymous class는 URI를 갖지 않고 Class를 모아주는데 사용하는 blank node와 같은 의미)
		이미 존재하는 용어사전이나 상위 온톨로지를 선별적으로 활용하여 해석 가능성을 높이고 있는가?
	일관성	스키마에 입각해서 인스턴스들이 입력되었는가?
		Class가 중복되지 않도록 계층구조가 일관성있게 이루어졌는가?
		property가 중복되지 않으며 계층구조가 일관성있게 이루어졌는가?
		property를 사용할 때 온톨로지 설계 시 정해진 domain과 range를 정확하게 준수하는가?
		적절한 restriction을 통해 일관성을 유지시켜 주고 있는가?
		Class의 restriction 정의시 논리적인 충돌이 없도록 적절히 구성했는가?
		Class 레벨의 적정성, 확장 가능성을 고려한 모델링으로 장기적으로 일관성 유지를 고려했는가?
		용어에 대한 URI정책이 명확한가? (URI할당의 적절성)
	명료성	URI의 가독성이 좋은가?
		URI로 표현되는 객체가 유일하게 식별되는가?
		domain과 range를 명확하게 표현하였는가?
		해당 온톨로지의 관점을 벗어난 개념, property를 포함하지는 않도록 명료하게 설계하였는가?
		Class와 인스턴스가 명료하게 구분되도록 구성되었는가?
		온톨로지 내의 용어에 대해 name, label, annotation 등을 충분히 활용하여 명확하게 의미를 표현해 주었는가?

노미 구조의 정확성, 프로퍼티의 명확한 사용, 스키마에 맞는 인스턴스 입력과 같은 항목들이 도출되었으며, 풍부성에서는 클래스 간 관계와 인스턴스의 풍부성, 다양한 표준어휘를 잘 활용하는 것, 도메인의 특성을 표현하기 위한 노력에 관한 항목이 도출되었다.

〈표 13〉에서 보는 것처럼, 의미적 부분은 해석가능성이 10개 항목, 일관성이 7개 항목, 그리고 명료성이 7개 항목으로 평가항목이 추출되어서 총 24개의 평가항목으로 구성되었다. 해석가능성에서는 직관적인 URI, 클래스명, 프로퍼티명, 인스턴스명에 대한 평가항목이 있었고, rdfs:label을 활용하여 사람들이 보기에 이해할 수 있도록 하는 항목도 존재했다. 일관성에서는 명확한 스키마, 중복되지 않는 계층구조, 제약 조건, 확장성을 고려한 모델 등이 주요 평가항

목으로 도출되었다. 하지만 일관성 부분에서 클래스, 프로퍼티 구조의 일관성과 합문법성에서 제시된 계층구조의 재귀적 문제에 관한 항목이 다소 중복되는 성향이 있음이 발견되었다. 이는 합문법성과 일관성의 유사한 속성에서 기인한 것으로 파악되며, 항목이 완전히 일치하는 것이 아니며 다소 평가의 방향이 차이가 있으므로 무방할 것으로 판단하였다.

끝으로 〈표 14〉에서 실제적 부분은 정확성이 6개 항목, 포괄성이 7개 항목, 관련성이 4개 항목으로 구성되어 총 17개 항목으로 이뤄져 있는 것을 볼 수 있다. 정확성은 온톨로지가 참고하고 있는 정보원을 평가하고, 목표한 정확한 결과가 도출되는지와 차후 정보수정 계획에 대한 항목들로 구성되어 있다. 포괄성에는 온톨로지 범위에 관한 항목과 온톨로지 범위에 대해 현재

〈표 14〉 실제적부분(정확성, 포괄성, 관련성) 최종 평가항목

대분류	중분류	평가항목명
실제적 부분	정확성	정확하고 권위있는 참고원 또는 도메인 전문가를 통해 획득한 정보로 구성되어 있는가?
		온톨로지의 목적에 맞는 결과가 정확히 나오는가?
		목적한 개념적 레벨까지 정확하게 표현해 주는가?
		중의적 개념을 갖지 않고 명확한 개념으로 구축되었는가?
		온톨로지의 소스가 되는 정보가 존재한다면, cleansing과 refinement 작업을 잘 거친 후 온톨로지가 구축되었는가?
		정보원에서 차후 추가되거나 수정, 삭제되는 부분이 온톨로지에 잘 반영되도록 계획되어 있는가?
	포괄성	구축하고자 하는 온톨로지 도메인에 대한 명확한 정의가 있는가?
		해당 도메인에 대한 충분한 지식을 가지고 설계되었는가?
		온톨로지가 다뤄야할 범위에 대한 구체적인 정의가 되어있는가?
		표현하고자 하는 대상을 추상화하여 개념을 충분히 추출했는가?
		온톨로지서 활용되는 용어에 대한 사전이 구축되어 있으며, 해당 용어 및 용어들의 관계가 온톨로지에 반영되어 활용되고 있는가?
		개념 간의 관계를 충분히 표현해 주고 있는가?
	관련성	온톨로지 구축 이후의 데이터의 관리 및 운영측면을 고려하여 Class 및 property를 결정하였는가?
		우수한 naming sense를 바탕으로 URI의 체계 및 naming이 간결하고 의미 전달이 명확한가?
		공유를 고려한 URI 정책을 가지고 있는가?
		외부 데이터와의 인터링킹 정보가 포함 되어 있는가?(owl:sameAs, skos:closedMatch 등의 활용)
		누구나 쉽게 알 수 있는 공통어휘를 사용하고 있는가?

와 미래에 모두 충족시킬 수 있는지에 대한 항목이 있다. 관련성은 온톨로지를 쉽게 이해할 수 있으며, 외부 자원에 인터링킹이 잘 되는지에 관한 항목으로 구성되어 있다.

본 연구에서 추출한 온톨로지 평가항목은 구문적 부분에서 12개 항목, 의미적 부분에서 24개 항목, 실제적 부분에서 17개 항목으로 도합 53개 항목이다.

5. 스마트폰 온톨로지 평가에 적용을 통한 신뢰성 측정 및 활용방향 제시

제1차 델파이 조사와 제2차 델파이 조사 및 개별 확인과정을 거쳐 추출한 온톨로지 평가항목을 기초로 스마트폰 온톨로지를 평가하였다. 스마트폰 온톨로지는 2013년과 2014년 사이에 출시된 스마트폰 31종에 대해 9개의 상위 클래스

(Agent, Function, Operation, Person, Place, Product, Spec, Time, Type), 26개의 하위 클래스, 123개의 프로퍼티, 총 21,137개의 트리플로 구성되어 있으며, 스마트폰에 관한 상세한 정보를 제공하는 것을 목적으로 하는 온톨로지이다. 스마트폰 온톨로지를 평가하는 목적은 추출해낸 온톨로지 평가항목의 신뢰성을 검증해보고, 실제적인 활용의 예를 보이는 것이다. 평가기간은 2015년 3월 21일부터 2015년 4월 7일까지 18일 동안 진행하였다. 평가는 정량적 평가로 진행되었으며 리커드 7점 척도로 진행하였다. 각 항목은 1) 매우 나쁨 2) 대체로 나쁨 3) 약간 나쁨 4) 보통 5) 약간 좋음 6) 대체로 좋음 7) 아주 좋음으로 구성하였다. 평가대상은 온톨로지 전문가 9명이었고, 설문지는 구글 온라인 서비스를 이용하여 배포하였다. 평가자들에게 스마트폰 온톨로지 파일과 온톨로지 명세서, 그리고 온톨로지 데이터 사전을 제공하였다. 온톨로지 평가항목의 분석을 위한 부호화는 <표 15>와 같다.

<표 15> 온톨로지 평가항목 부호화

대분류	소분류	항목	부호화
구문적 부분	합문법성	전체	a
		합문법성 1) 항목 ~ 합문법성 8) 항목	a1~a8
	풍부성	전체	b
		풍부성 1) 항목 ~ 풍부성 4) 항목	b1~b4
의미적 부분	해석 가능성	전체	c
		해석가능성 1) 항목 ~ 해석가능성 10) 항목	c1~c10
	일관성	전체	d
		일관성 1) 항목 ~ 일관성 7) 항목	d1~d7
	명료성	전체	e
		명료성 1) 항목 ~ 명료성 7) 항목	e1~e7
실제적 부분	정확성	전체	f
		정확성 1) 항목 ~ 정확성 6) 항목	f1~f6
	포괄성	전체	g
		포괄성 1) 항목 ~ 포괄성 7) 항목	g1~g7
	관련성	전체	h
		관련성 1) 항목 ~ 관련성 4) 항목	h1~h4

5.1 신뢰도 분석

여러 개의 항목으로 하나의 개념을 얼마나 잘 표현하는지를 나타내는 것을 내적 일치도(internal consistency)라 하며 이를 수치화한 것을 Cronbach's α 라 한다. 이를 문항의 신뢰도라 하며, Cronbach's α 계수는 0~1의 값을 가지며, 높을수록 바람직하나 반드시 몇 점 이상이어야 한다는 기준은 없다. 흔히 0.8~0.9 이상이면 바람직하고 0.6~0.7이면 수용할 만한 것으로 여겨진다. 그러나 0.6보다 작으면 내적 일관성을 결여한 것으로 받아들여진다. 일반적으로 내부일관성에 대한 지수를 구하거나, 개념을 측정하는 항목 중 어느 항목을 제외시켜야 하는지 결정하는 근거로 사용한다. 본 연구에서 사용하는 평가항목들은 온톨로지 전문가들이 제안하고 서로 간의 검증을 거친 항목들이기에 신뢰도 분석을 통해 항목을 제외시키는 시도는 하지 않았다. 다만, 신뢰도 분석을 통해 내적 일관도를 저해하는 항목의 내용을 분석하여 온톨로지를 수정하는데 사용하였다. 온톨로지 평가항목의 신뢰도 분석을 위해 SPSS Ver. 21을 이용하였으며 분석 결과는 <표 16>과 같다.

각 평가항목의 Cronbach's α 값을 살펴보면 관련성($\alpha = 0.587$)을 제외하면 최소값이 0.608로 평가 결과를 수용할 만한 것으로 나타났다. 합문법성의 경우 “Class의 계층관계가 재귀적인 문제없이 잘 구성되었는가?” 항목과 “is-a has-a와 같이 방향성을 가지는 property의 표현이 명확히 이뤄졌는가?” 항목을 제거할 경우 신뢰도가 0.814로 향상되는 것으로 나타났다. 해석가능성의 경우 “URI 충돌을 피하기 위해서 숫자로 된 URI를 만들었다면, rdfs:label과 같은 레이블을 활용해서 이용자들이 이해할 수 있게 구성하였는가?”, “domain과 range를 명시하여 명확한 의미를 전달하는가?”, “인위적으로 anonymous class를 사용하지 않고 명확한 어휘로 표현되었는가?” 항목을 제거하는 경우 0.785로 향상 가능하며 명료성에서는 “domain과 range를 명확하게 표현하였는가?” 항목을 제거하는 경우 신뢰도가 0.689로 향상되는 것으로 나타났다. 유일하게 Cronbach's α 값이 0.6 이하로 나온 관련성에서는 “우수한 naming sense를 바탕으로 URI의 체계 및 naming이 간결하고 의미 전달이 명확한가?” 항목과 “공유를 고려한 URI 정책을 가지고 있는가?” 항목

<표 16> 온톨로지 평가항목 신뢰도 분석결과

대분류	평가항목	Cronbach's α	비고
구문적 부분	합문법성	0.651	0.814 (a2, a7 제거)
	풍부성	0.750	
의미적 부분	해석가능성	0.608	0.785 (c2, c8, c9 제거)
	일관성	0.828	
	명료성	0.631	
실제적 부분	정확성	0.769	0.706 (h1, h2 제거)
	포괄성	0.855	
	관련성	0.587	

을 제거하는 경우 신뢰도가 0.706으로 향상되는 것으로 나타났다. 이는 두 문항이 전체적으로 일관성을 저해하고 있는 항목인 것으로 이해할 수 있다. 실제적으로 h1과 h2는 명명규칙과 URI정책에 관한 내용이며, 관련성 내의 타 항목은 외부데이터와의 인터링킹과 상호운용성을 위한 공통어휘 사용에 관한 내용일 것을 볼 수 있다. 따라서 h1과 h2를 제거하거나 별도의 범주로 구분하면 좀 더 신뢰성 있고 일관성 있는 평가항목이 될 것으로 기대된다.

5.2 온톨로지 평가항목의 평균차이 검정

본 연구에서 추출한 온톨로지 평가항목은 온톨로지 구축시 좋은 품질을 위한 체크리스트와 같이 사용할 수 있고, 동일한 도메인의 온톨로지를 비교하거나 평가할 때 평가도구로 사용할 수 있을 것이다. 본 절에서는 앞 절에서 리커드 척도로 평가한 스마트폰 온톨로지의 평가결과를 가지고 통계적 방법을 활용해 품질이 좋은 온톨로지인지 검증을 시도해보았다. 현재 온톨로지의 품질을 평가하는데 있어 절대적인 기준

이 되는 점수는 없지만, 스마트폰 온톨로지 평가를 위해 활용한 리커드 척도에서 “약간 좋음”을 나타내는 5점을 객관적으로 좋은 온톨로지의 기준으로 설정하고, 5점과 얼마나 유의한 차이가 있는지 비교해서 통계적으로 유의한 차이를 나타내는 지 확인하고자 하였다. 정규성 검정을 시도하여 t-test를 사용할 수 있는지 확인하였고, 정규성이 만족되지 않는 경우 Wilcoxon 부호검정순위를 사용하여 검증해보았다. 결과의 예로서, 합문법성의 경우 전체 평균이 5.986으로 대체로 좋은 방향으로 평가된 것으로 나타났다. 각 항목에서도 최소 평균이 5.444로 “약간 좋음”보다 높게 평가되는 것으로 나타났다. 이러한 차이가 통계적으로도 의미가 있는지 살펴보기 위해 먼저 정규성 검정을 한 결과 모든 항목에서 정규성을 만족하는 것으로 나타났다. 따라서 t-test를 통해 평균차이 검정을 수행해도 문제가 없는 것으로 판단되었고, 본 연구에서 설정한 좋은 온톨로지의 기준인 5점과 비교하였다(〈표 17〉 참조).

유의수준 5%하에서 평균차이 검정을 수행한 결과 합문법성 전체에 대한 결과에서 t-test의

〈표 17〉 합문법성의 평균차이 검정

	항목	평균	표준편차	정규성검정 p-value	t-test		Wilcoxon
					t	p-value	p-value
합문법성	전체	5.986	0.494	0.843	5.990	0.000	0.011
	a1	6.667	0.500	0.091	10.000	0.000	0.006
	a2	6.444	0.726	0.270	5.965	0.000	0.009
	a3	6.000	1.000	0.491	3.000	0.017	0.030
	a4	5.556	0.726	0.118	2.294	0.051	0.059
	a5	5.444	0.882	0.109	1.512	0.169	0.157
	a6	5.556	1.424	0.112	1.170	0.276	0.143
	a7	5.889	1.167	0.075	2.286	0.052	0.075
	a8	6.333	0.500	0.091	8.000	0.000	0.006

p-value가 < 0.001로 통계적으로 유의한 차이가 있는 것으로 나타났으며 이때 평균이 5.986으로 “약간 좋음”보다 높은 만족도를 보이는 것으로 나타났다. 표본크기가 9이기 때문에 결과의 신뢰성을 위해 평균차이 검정의 비모수적 방법인 Wilcoxon 부호 순위 검정으로 검정해본 결과 t-test의 결과와 마찬가지로 5% 유의수준에서 결과가 통계적으로 유의한 차이가 있는 것으로 나타났다. 각 항목을 살펴보면 “W3C 권고안(RDF, OWL)을 충실히 이행하였는가?”, “Class의 계층관계가 재귀적인 문제없이 잘 구성되었는가?”, “property를 만들 때 objectProperty와 datatypeProperty를 잘 구별하였는가?”, “인스턴스들이 모델링 규칙에 잘 맞게 합문법적으로 입력되었는가?” 항목에서 “약간 좋음”보다 높은 만족도를 보이는 것으로 나타났으며 “objectProperty를 만들 때 domain과 range가 적절히 설정되었고, 그에 맞게 사용되었는가?”, “rdfs:range에서 boolean, interger 등 치역이 합법적으로 취해지도록 조치가 되어 있는가?”, “기존의 용어집(foaf, skos, skos-xl) 등을 활용한 경우, 용어의 명세에 맞게 활용하고 있는가?”, “is-a, has-a와 같이 방향성을 가지는 property의 표현이 명확히 이뤄졌는가?” 항목은 유의수준 5%하에서 통계적으로 유의하지 않은 것으로 나타났다. 이는 “약간 좋음”보다 만족도가 높다는 근거를 찾을 수 없다는 것을 의미하며, 즉, “약간 좋음” 정도의 만족도를 보인다는 것을 의미한다. 동일한 방법으로 다른 7개의 하위 범주도 평가해본 결과, 풍부성 총 4개 중 1개, 해석가능성 총 10개 중 4개, 일관성 총 7개 중 2개, 명료성 총 7개 중 1개, 정확성 총 6개 중 1개, 포괄성 총 8개 중 0개, 관련성 총 4개 중

1개가 5점 척도와 비교하여 통계적으로 유의한 차이가 나지 않아서 5점, 즉 약간 좋은 정도의 수준이며, 이 외의 항목들은 모두 7점 만점에서 6점에 해당하는 “좋음” 수준인 것으로 확인되었다.

이와 같은 방식으로 스마트폰 온톨로지 외에도 다양한 도메인의 온톨로지를 품질측정하는 도구로서 사용할 수 있으며, 대부분의 온톨로지 구축방법론에서 제시하는 것처럼 평가 이후 수정 및 재구축을 통해 온톨로지 품질향상에 기여하도록 할 수 있다.

6. 결론

다국어 서비스 보다 나은 검색결과를 제공하기 위한 지식베이스로 온톨로지를 활용하고자 하는 연구가 증가하고 있고, 최근 각광을 받고 있는 빅데이터 분석에도 온톨로지를 활용하여 성능을 높이고자 하는 시도들이 이뤄지고 있다 (함영균, 서지우, 황도삼, 최기선, 2014; 유은지, 김정철, 이춘열, 김남규, 2012). 그러나 지금까지의 온톨로지 평가에 관한 연구는 평가범주를 제시하였으나 구체적인 온톨로지 품질평가항목이 미비하였고, 재현률과 정확률을 통한 성능평가가 주를 이루어, 체계적으로 온톨로지의 품질을 평가할 수 있는 방법론이 시급한 실정이었다.

본 연구는 온톨로지 품질측정을 위한 구체적인 평가항목을 추출하는 시도를 하였다. 먼저 문헌 조사를 통해 다양한 온톨로지 구축방법과 평가방법 내에서 온톨로지 평가를 위한 기본 범주를 추출하는 시도를 하였으며, 방만하지 않으면서도 비교적 넓은 범위를 다루고 있다고 평가되

는 박진수 등(2008)의 연구를 참고하여 온톨로지 평가를 위한 기본범주를 추출하였다. 온톨로지 전문가에 대한 기준을 수립하고 개별적인 검증과정을 거쳐 온톨로지 전문가 패널을 구성한 뒤, 1, 2차에 걸쳐 델파이 조사를 수행하였고 추가적인 검증을 시도하였다. 이를 통해 총 3개의 상위범주와 8개의 하위범주에서 53개의 온톨로지 평가항목을 추출할 수 있었다. 또한 본 연구에서는 추출한 53개의 온톨로지 평가항목을 가지고 스마트폰 온톨로지를 평가하는 시도를 통해, 첫째, 추출한 평가항목의 신뢰도를 검증하였고, 둘째, 좋은 온톨로지를 검증하는 사례를 선보였다. 평가 결과, 추출한 온톨로지 평가항목 중 8개의 하위범주에서 1개의 범주를 제외하고는 모두 신뢰도를 받아들일 수 있는 수준이라는 것을 알 수 있었고, 좋은 온톨로지를 검증하는 방법론에서도 스마트폰 온톨로지는 총 53개의 평가항목 중 39개의 항목에서 “ 좋음 ” 수준을 나타내는 온톨로지임을 확인할 수 있었다.

본 연구에서 추출한 온톨로지 평가항목은 다음과 같은 의미를 지닌다. 첫째, 온톨로지 구축의 경험이 많은 전문가들이 본인들의 실제 구축 경험에서 나오는 노하우를 바탕으로 제시한 항목들이라는 점이다. 둘째, 델파이 조사방법론의 특성상 전문가들이 개별적으로 개방질문을 통해 응답을 받았기 때문에 포커스 인터뷰 방법과는 달리 실험실 내의 분위기나 영향력 있는 몇 사람에게 의해 결과가 영향받지 않고 전문가들이

자신의 의견을 소신있고 충분히 개진한 결과라는 점이다. 마지막으로 소수의 전문가라도 삭제요구를 한 항목에 대해서는 검증을 거친 뒤 삭제하였고, 현 53개의 항목들은 전문가 패널 전체가 합의한 항목들이기에 객관성을 띠다고 할 수 있다는 점이다. 즉, 어떤 도메인에서든지 보편적으로 받아들일 수 있는 내용만 평가항목으로 추출되었다는 것이다.

이러한 온톨로지 평가항목들은 다음과 같이 활용될 수 있을 것으로 예상된다. 첫째, 온톨로지 구축을 위한 체크리스트이자 간단한 가이드라인으로서 활용할 수 있다. 둘째, 온톨로지 구축시 도메인에 구애받지 않고 평가를 위한 도구로 사용할 수 있다. 셋째, 도메인의 특성에 맞는 평가방법론을 개발할 때, 기초자료로 사용되어 도메인의 특성을 반영한 평가방법론 개발에 도움을 줄 수 있다.

웹 기반의 지식정보화 사회가 심화되면서 정보는 많지만 이 정보들을 구조화하고 분석해낼 수 있는 기초자료에 대한 요구가 많아지는 상황에서 온톨로지의 중요성은 더욱 커질 것으로 예상된다. 고품질의 온톨로지에 대한 요구와 품질 평가방법에 대한 요구도 같이 상승할 것을 쉽게 예측할 수 있다. 이런 상황을 감안할 때, 온톨로지와 그 품질 평가에 대한 후속연구가 계속되어 다양한 도메인에서 생산되는 온톨로지의 품질을 객관적으로 검증할 수 있도록 더 합리적이고 확실한 평가방안이 강구되어야 할 것이다.

참 고 문 헌

- 고영만, 서태설 (2005). 온톨로지 기반 메타데이터 명명 규칙에 관한 연구. 정보관리학회지, 22(4), 97-109.
- 김수경, 안기홍 (2007). 지능형 이미지 검색 시스템을 위한 추론 기반의 웹 온톨로지 구축. 정보관리학회지, 24(3), 119-147. <http://dx.doi.org/10.3743/KOSIM.2007.24.3.119>
- 김수경, 안기홍 (2008). 기술논리와 SWRL 기반의 웹 온톨로지 모델링. 정보관리학회지, 25(1), 149-171. <http://dx.doi.org/10.3743/KOSIM.2008.25.1.149>
- 김은경 (2004). 시맨틱웹을 위한 온톨로지 구축방법에 관한 비교 연구. 석사학위논문, 중앙대학교 대학원, 문헌정보학과.
- 김정민, 최병일, 김형주 (2006). 텍스트 내용 지식 기반의 철학 온톨로지 구축. 정보과학회논문지: 컴퓨팅의 실제, 11(3), 275-283.
- 노영희 (2011). 문헌정보학 학술지를 대상으로 한 온톨로지 구축에 관한 연구. 정보관리학회지, 28(2), 177-193. <http://dx.doi.org/10.3743/KOSIM.2012.28.2.177>
- 박진수, 조원진, 노상규 (2008). 온톨로지 자동추출도구의 기능적 성능 평가를 위한 평가지표의 개발 및 적용. 지능정보연구, 14(4), 69-87.
- 심재후, 최명길 (2013). 창업학 지식구조 분석결과를 활용한 창업 온톨로지 구축. 한국정보기술응용학회, 20(2), 161-176.
- 유은지, 김정철, 이춘열, 김남규 (2012). 시맨틱 텍스트 마이닝을 위한 온톨로지 활용 방안. 정보시스템연구, 21(3), 137-161.
- 이병길, 김희섭 (2013). 새마을운동 기록물의 개체기반 온톨로지 검색시스템 설계 및 평가. 한국기록관리학회지, 13(3), 67-97.
- 이정연, 이재윤, 정한민, 강인수, 신숙경 (2007). 확률적 온톨로지와 연구자 네트워크를 이용한 심사자 자동 추천에 관한 연구. 정보관리학회지, 24(3), 43-65. <http://dx.doi.org/10.3743/KOSIM.2007.24.3.043>
- 이정희, 김희섭 (2007). 대학 전자기록물을 위한 온톨로지 기반 검색시스템 설계 및 구현. 정보관리학회지, 24(3), 343-362. <http://dx.doi.org/10.3743/KOSIM.2007.24.3.343>
- 이해운 (2006). 동사 온톨로지 구축 방법론 - 한국어 이동동사. 언어와 언어학, 38, 89-108.
- 정태영, 김희열, 박종현 (2011). 傷寒論 온톨로지 구축 방법론 연구. 동의생리병리학회지, 25(5), 765-772.
- 조나혜, 강영옥 (2013). 지리개념 이해를 위한 지오 온톨로지 구축: 해안지형을 사례로. 한국지도학회지, 13(3), 75-87.
- 함영균, 서지우, 황도삼, 최기선 (2014). 프레임넷을 통한디비피디아 온톨로지 인스턴스 생성의 커버리지 개선, 제 26회 한글 및 한국어 정보처리 학술대회 논문집.

- Brank, J., Mladenic, D., & Grobelnik, M. (2006). Gold standard based ontology evaluation using instance assignment. The 4th International Workshop on Evaluation of Ontologies for the Web(EON) at the 15th International World Wide Web Conference.
- Brewster, C., Alani, H., Dasmahapatra, S., & Wilks, Y. (2004). Data driven ontology evaluation. International Conference on Language Resources and Evaluation. (LREC 2004), 24-30 May 2004, Lisbon, Portugal.
- Burton-Jones, A., Storey, V., Sugumaran, V., & Ahluwalia, P. (2005). A semiotic metrics suite for assessing the quality of ontologies. *Data & Knowledge Engineering*, 55: 84-102.
- Cimiano, P. (2006). *Ontology learning and population from text: Algorithms, evaluation and application*. New York: Springer.
- Fernandez, M., Gomez-Perez, A., & Juristo, N. (1997). METHONTOLOGY: From ontological art towards ontological engineering. AAAI-97 Spring Symposium on Ontological Engineering, Stanford University, March 24-26th.
- Gomez-Perez, A., Fernandez, M., & De Vicente, A. J. (1996). Towards a method to conceptualize domain ontologies. ECAI-96 Workshop on Ontological Engineering, Budapest.
- Grüniger, M., & Fox, M. S. (1994). The design and evaluation of ontologies for enterprise engineering. Workshop on Implemented Ontologies, European Conference on Artificial Intelligence 1994. Amsterdam, NL.
- Grüniger, M., & Fox, M. S. (1995). Methodology for the design and evaluation of ontologies. IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing, Montreal, August 19-20th.
- Grüniger, M., & Fox, M. S. (1995). The role of competency questions in enterprise engineering. *Benchmarking – Theory and Practice*. 22-31. Springer, US.
- Lozano-Tello, A., & Gómez-Pérez, A. (2004). Ontometric: A method to choose the appropriate ontology. *Journal of Database Management*, 2(15), 1-18.
- Maedche, A., & Staab, S. (2002). Measuring similarity between ontologies. *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web*. 251-263.
- Noy, N. F., & McGuinness, D. L. *Ontology development 101: A guide to creating your first ontology*. Retrieved from [http:// protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html](http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html)
- Porzel, R., & Malaka, R. (2004). A task-based approach for ontology evaluation. Proc. ECAI 2004 Workshop on Ontology Learning and Population, 9-16.

- Prieto-Díaz, R. (2003). A faceted approach to building ontologies. In Information Reuse and Integration, 2003. IRI 2003. IEEE International Conference on (pp. 458-465). IEEE.
- Ra, M., Yoo, D., No, S., Shin, J., & Han, C. (2012). The mixed ontology building methodology using database information. Proceedings of the International MultiConference of Engineers and Computer Scientists, Vol. 1.
- Uschold, M. (1996a). Converting an informal ontology into ontolingua: Some experiences, ECAI-96 Workshop on Ontological Engineering, Budapest, August 13th.
- Uschold, M. (1996b). Building ontologies: Towards a unified methodology, Proc. Expert Systems 96, Cambridge, December 16-18th.
- Uschold, M., & King, M. (1995). Towards a methodology for building ontologies. IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing, Montreal, Canada.

<p>• 국문 참고문헌에 대한 영문 표기 (English translation of references written in Korean)</p>
--

- Hahm Younggyun, Seo, Jiwoo, Hwang, Dosam, & Choi, Key-sun (2014). DBpedia ontology population coverage enhancement with frameNet, Annual Conference on Human & Cognitive Language Technology.
- Jung, Taeyoung, Kim, Heeyeol, & Park, Jonghyun (2011). Study on a methodology for developing shanghanlun ontology. Journal of Physiology & Pathology in Korean Medicine, 25(5), 765-772.
- Kim, Eunkyong (2004). A comparative study on methodologies of building ontologies towards semantic web. M.A. thesis, The Graduate School Chung-Ang University, Department of Library and Information Science.
- Kim, Jung-Min, Choi, Byoung-Il, & Kim, Hyoung-Joo (2006). Implementation a philosophy ontology based on knowledge of text contents. Journal of KIISE: Computing Practices and Letters, 11(3), 275-283.
- Kim, Su-Kyoung, & Ahn, Kee-Hong (2007). An implementation of inference-based web ontology for intelligent image retrieval system. Journal of the Korean Society for Information Management, 24(3), 119-147. <http://dx.doi.org/10.3743/KOSIM.2007.24.3.119>
- Kim, Su-Kyoung, & Ahn, Kee-Hong (2008). Web ontology modeling based on description logic and SWRL. Journal of the Korean Society for Information Management, 25(1), 149-171. <http://dx.doi.org/10.3743/KOSIM.2008.25.1.149>

- Ko, Young Man, & Seo, Tae-Sul (2005). A study on the naming rules of metadata based on ontology. *Journal of the Korean Society for Information Management*, 22(4), 97-109.
- Lee, Byunggil, & Kim, Heesop (2013). Design and evaluation of an individual instance-based ontology retrieval system for archival records of the "Saemaul Movement". *Journal of Korean Society of Archives and Records Management*, 13(3), 67-97.
- Lee, Hae-yun (2006). A case study on building verb-ontology: Korean locomotion verbs. *Language and linguistics*, 38, 89-108.
- Lee, Junghee, & Kim, Heesop (2007). A design and implementation of ontology-based retrieval system for the electronic records of universities. *Journal of the Korean Society for Information Management*, 24(3), 343-362. <http://dx.doi.org/10.3743/KOSIM.2007.24.3.343>
- Lee, Jungyeoun, Lee, Jae-Yun, Jung, Hanmin, Kang, In-Su, & Shin SukKyung (2007). Automatic recommendation of panel pool using a probabilistic ontology and researcher networks. *Journal of the Korean Society for Information Management*, 24(3), 43-65. <http://dx.doi.org/10.3743/KOSIM.2007.24.3.043>
- Noh, Younghee (2011). A study on constructing the ontology of LIS journal. *Journal of the Korean Society for Information Management*, 28(2), 177-193. <http://dx.doi.org/10.3743/KOSIM.2012.28.2.177>
- Park, Jinsoo, Cho, Wonchin, & Rho, Sangkyu (2008). Measurement criteria for ontology extraction tools. *Journal of Intelligent Information Systems*, 14(4), 69-87.
- Shim, Jaehu, & Choi, Myeonggil (2013). An establishment of entrepreneurship ontology through analysis of intellectual structure in entrepreneurship research. *Journal of Information Technology Applications & Management*, 20(2), 161-176.
- Yu, Eunji, Kim, Jungchul, Lee, Chunyol, & Kim, Namgyu (2012). Using ontologies for semantic text mining. *The Journal of information systems*, 21(3), 137-161.

[부록 1] 제 1차 델파이 조사 결과 종합

대분류	소분류	추출된 평가 항목
구문적 부분	합문법성	<ul style="list-style-type: none"> • 개발되는 어휘는 실세계의 개념을 충실히 반영하면서도 문법적으로 이상이 없이 작성되어 있는가? • 기존의 용어집(foaf, skos) 등을 활용한 경우, 용어의 명세에 맞게 활용하고 있는가? • 기존의 용어집을 수정하여 활용하는 경우, 필요에 맞게 수정하되 문법적으로 정확하게 수정하였는가? • Class의 계층관계에 있어 *재귀적인 문제가 발생하지는 않는가?(*재귀: 원래의 자리로 되돌아가거나 되돌아 옴) • Class의 restriction은 실세계를 잘 반영하면서도, 합문법적으로 작성되어 있는가? • property를 만들 때 domain과 range가 적절히 설정되었고, 그에 맞게 사용되었는가? • property를 만들 때 objectProperty와 datatypeProperty를 잘 구별하였는가? • objectProperty를 사용할 때 더 구체적으로 표현력을 올리기 위한 조치를 하고 있는가? (예를들면, owl:SymmetricProperty 등) • rdfs:range에서 boolean, interger 등 치역이 합법적으로 취해지도록 조치가 되어 있는가? • W3C 권고안(RDF, OWL)을 충실히 이행하였는가? • is-a has-a 와 같은 directive property에 대한 명확한 표현이 준수되었는가? • Class와 property 작성시 네이밍 규칙을 잘 지키고 있는가? • 생성된 인스턴스들이 모델링 규칙에 잘 맞게 구성되었는가?
	풍부성	<ul style="list-style-type: none"> • 해당 도메인의 현실세계 모습을 충분히 반영하기 위해 공리(axiom)나 제약조건(restriction)을 잘 식별하여 반영하였는가? • 다양한 추론(inference configuration)을 적절하게 활용하는가? (symmetricity 또는 intransitivity 등) • Class 간의 관계, 인스턴스의 property를 풍부하게 표현하고 있는가? • 타 온톨로지 혹은 타 객체로 이동할 수 있도록 설계되었는가? • subject와 object에 해당하는 객체들은 literal이 아니라 URI를 갖는 객체로 표현되었는가? • 시맨틱 웹 표준으로 제안된 구문들을 충분히 활용하였는가? • 잘 알려진 어휘(DC, SKOS 등)를 활용하여 property를 표현하였는가? • 온톨로지의 목적에 맞는 수준으로 어휘를 적절하게 활용하였는가?
의미적 부분	해석 가능성	<ul style="list-style-type: none"> • 기계가 읽고 처리할 수 있는 수준의 엄격한 스키마 정의가 설정되었는가? • URI가 직관적으로 이해하기 좋게 명명되어 있는가? • URI 충돌을 피하기 위해서 숫자로 된 URI를 만들었다면, rdfs:label과 같은 레이블을 활용해서 이용자들이 이해할 수 있게 구성하였는가? • Class와 property 명이 직관적으로 이해가 되도록 명명되어 있는가? • Class 명에서 상하위 클래스를 이해할 수 있도록 지어졌는가? • property 명에서 domain과 range가 이해될 수 있도록 지어졌는가? • 인스턴스 명에서 Class 타입이 이해될 수 있도록 지어졌는가? • Class명과 인스턴스 명이 real world의 concept name으로 잘 표현되어 있는가? • Class와 property를 정의한 문서가 웹에 발행되어 있는가? • domain과 range를 명시하여 명확한 의미를 전달하는가? • anonymous class*의 사용은 최대한 자제하고 명확한 어휘로 표현되었는가?(*anonymous class는 URI를 갖지 않고 Class를 모아주는데 사용하는 blank node와 같은 의미) • 이미 존재하는 용어사전이나 상위 온톨로지를 활용하여 해석 가능성을 증하고 있는가? • Class 또는 property 간에 상하위 관계를 잘 구축했는가? • property를 설계할 때 수여, 소유, 가용 등의 방향성을 잘 표현하여 설계하였는가?

대분류	소분류	추출된 평가 항목
의미적 부분	일관성	<ul style="list-style-type: none"> 스키마에 입각해서 인스턴스들이 입력되었는가? Class가 중복되지 않도록 계층구조가 일관성 있게 이루어졌는가? 유사한 의미의 property가 많지 않은가? 상위 클래스의 property로 잘 정의되었는가? property를 사용할 때 온톨로지 설계시 정해진 domain과 range를 정확하게 준수하는가? restriction을 통해 표현력을 제공하면서도 일관성을 유지시켜주고 있는가? Class의 restriction 정의시 논리적인 충돌이 있지는 않은가? 하나의 Class에 너무 많은 인스턴스가 집중되지 않도록 모델 설계하였는가? Class 레벨의 적정성, 확장 가능성을 고려한 모델링으로 장기적으로 일관성 유지를 고려했는가?
	명료성	<ul style="list-style-type: none"> 용어에 대한 URI정책이 명확한가? (URI할당의 적절성) URI의 가독성이 좋은가? URI로 표현되는 객체가 유일하게 식별되는가? URI로 표현되는 객체가 상·하·좌·우 관계가 명확하게 표현되는가? URI가 dereferencing이 가능하여 의미를 명확히 알게 되도록 작성되었는가? domain과 range를 명확하게 표현하였는가? 해당 온톨로지의 관점을 벗어나거나 과도하게 세부적인 개념, property를 포함하지는 않는가? Class와 인스턴스가 잘 구분되는가? 온톨로지 내의 용어에 대해 name, label, annotation 등을 충분히 활용하여 언어적 표현의 한계를 극복하고 의미를 명확하게 전달하였는가?
실제적 부분	정확성	<ul style="list-style-type: none"> 정확하고 권위있는 참고원 또는 도메인 전문가를 통해 획득한 정보로 구성되어있는가? 온톨로지의 목적에 맞는 정확한 결과값이 도출되는가? 목적한 개념적 레벨까지 정확하게 표현해 주는가? 중의적 개념을 갖지 않고 명확한 개념으로 구축되었는가? 온톨로지를 기반으로 추론을 수행했을 때, 참·거짓으로 명확히 판별 가능한 추론결과가 도출되는가? 온톨로지의 소스가 되는 정보가 존재한다면, cleansing과 refinement 작업을 잘 거친 후 온톨로지가 구축되었는가? 정보원에서 차후 추가되거나 수정, 삭제되는 부분이 온톨로지에 잘 반영되도록 계획되어 있는가?
	포괄성	<ul style="list-style-type: none"> 구축하고자 하는 온톨로지 도메인에 대한 명확한 정의가 있는가? 해당 도메인에 대한 충분한 지식을 가지고 설계되었는가? 온톨로지가 다뤄야할 범위에 대한 구체적인 정의가 되어있는가? 표현하고자 하는 대상을 추상화하여 개념을 충분히 추출했는가? 개념 간의 관계를 충분히 표현해 주고 있는가? 온톨로지에서 활용되는 용어에 대한 사전이 구축되어 있으며, 해당 용어 및 용어들의 관계가 온톨로지에 반영되어 활용되고 있는가? 도메인을 표현하는데 충분히 활용할 수 있는 upper온톨로지를 확보하였는가? 온톨로지 구축 이후의 데이터의 관리 및 운영측면을 고려하여 Class 및 property를 결정하였는가?
	관련성	<ul style="list-style-type: none"> 우수한 naming sense를 바탕으로 URI의 체계 및 naming이 간결하고 의미 전달이 명확한가? 공유를 고려한 URI 정책을 가지고 있는가? 외부 데이터와의 인터링킹이 정보가 포함 되어 있는가? (owl:sameAs, skos:closedMatch 등의 활용) owl:inverseFunctionalProperty를 활용하여 특정 개념에 대해 흩어진 지식을 모을 수 있도록 하고 있는가? 누구나 쉽게 알 수 있는 공통어휘를 사용하고 있는가?