

연구자 간 협업 지원 서비스를 위한 지식 베이스 설계

정한민 성원경 박동인
한국과학기술정보연구원 정보시스템부
{jhm, wksung, dipark}@kisti.re.kr

A Design of Knowledge Base for the Service of Collaboration Support between Researchers

Hanmin Jung Won-Kyung Sung Dong-In Park
Division of Information System, KISTI

요 약

지식 베이스는 대량의 인적·물적 자원을 투입하고 철저한 설계에 기반하여 구축하여야 하는 부담이 큰 자원이다. 그렇지만, 지식 베이스 없이 다양한 응용 분야와 서비스를 만족시키는 시스템을 구축할 수 없기 때문에 그 중요성을 인정하지 않을 수 없다. 기존의 지식 베이스 구축에 있어서 간과했던 것들의 하나는 응용 분야와 서비스를 명확히 하고, 해당 분야에 맞는 지식 베이스를 선택해야 한다는 것이다. 어휘의미망을 포함하여 시소러스를 범용 또는 여러 분야에서 구축하고 있으나 이들이 어떤 서비스에 도움이 되는지를 확실히 분석하지 못함으로 인해 응용 시스템에서 제 기능을 다하지 못하고 있으며, 온톨로지 또한 근본적으로 분야와 활용 목적에 의존적이어서 이를 고려하여 설계를 하지만 시소러스나 용어 사전과 같은 기본적인 지식 베이스와의 연계성을 통한 효율성 재고라는 측면에서 약점을 보여왔다. 본 연구에서는 사용자 질의와 실 데이터 간의 어휘 불일치 해소를 위해 시소러스를 설계·구축하고 온톨로지의 사례화 결과인 인스턴스와 연동시킴으로써 연구자 간 협업 지원 서비스를 제공할 수 있는 기반 자원으로서의 지식 베이스 구축에 초점을 둔다.

1. 서론

인터넷의 등장과 함께 정보가 폭발적으로 늘어나면서 이들을 체계적으로 관리하고 공유할 필요성이 증대하였다. 또한, 시멘틱 웹이 생겨나면서 서비스가 아닌 서비스 대상에 대해 온톨로지를 구축하고자 하는 노력이 필요하게 되었다. 특히, 과학기술 분야는 그 특성상 발전 속도가 다른 분야에 비해 빠르며 해당하는 정보의 양 또한 상대적으로 크기 때문에 이러한 지식 베이스 (Knowledge Base) 구축 필요성이 더욱 절실히 느껴진다.

지식 베이스는 응용 분야에 맞추어 세심하게 설계되고 구축되어야 한다. 응용 분야를 고려하지 않고 구축된 지식 베이스는 범용적인 용도로 밖에 사용할 수 없으며, 특정 응용에 재적용하고자 할 때의 추가 고려 사항으로 인해 그 구조가 깨질 수 있는 위험성도 내포하고 있다.

본 논문에서는 응용 분야를 연구자 간 협업을 지원하

는 서비스에 초점을 맞추고, 이를 지원할 수 있는 지식 베이스로서의 시소러스와 온톨로지 설계를 소개한다. 어휘 불일치 (Lexical Gap) 해소를 위한 시소러스와 실 데이터 표현을 위한 온톨로지 설계를 통해 지식 베이스의 연계와 활용 가능성을 보여준다.

본 논문의 구성은 다음과 같다. 2장에서는 본 논문이 지향하는 응용 분야인 연구자간 협업 지원을 위한 서비스를 소개하고, 3장에서는 서비스와 연계된 지식 베이스 설계 및 고려 사항을 기술한다. 4장에서는 지식 베이스 구성 요소인 시소러스, 온톨로지, 인스턴스의 구체적인 구조를 설명하고, 지식 베이스 구축 현황을 5장에서 소개한다.

2. 연구자 간 협업 지원 서비스

과학 기술 분야에서의 연구자 간 협업 지원 서비스는

정보 유통 플랫폼 상에서 이루어진다¹. 연구자 간의 원활한 협업을 위해 정보 유통 플랫폼은 검증된 정보 교환, 정보 수정, 정보 등록 및 정보 제공 서비스를 지원한다. 제공되는 정보 서비스는 다음의 세 가지로 구분된다.

(1) 연구자 네트워크 (Communities of Practice): 사업/학술 관점에서 특정 주제(들)에 대해 연구자 네트워크를 획득하고, 해당 분야의 전문성에 따른 전문가 추천 순위 정보를 제공한다.

(2) 연구자 추적 (Researcher Tracking): 정해진 기간 내에서 특정 연구자를 선택하여 전체 또는 특정 주제에 대해 사업/학술 정보의 연도별 규모를 보여준다.

(3) 연구 맵 (Research Map): 정해진 기간 내에서 지역/부서를 선택하여 전체 또는 특정 주제에 대해 사업/학술 정보의 연도별 규모를 보여준다.

그림 1은 상기 시나리오를 포함하는 정보 유통 플랫폼 상에서의 연구자 간 협업 지원 서비스를 보여준다. 각 연구자는 정보를 중앙 서버에 등록하고, 중앙 서버는 이들을 검증하고 URI를 할당한다. 연구자가 별도로 입력한 과제/학술 등의 과학 기술 기반 정보는 온톨로지에 맞추어 사례화 (Instantiation)되고 RDF Repository에 저장된다. 중앙 서버에 등록된 문서는 다른 연구자가 다운로드 할 수 있으며, 권한이 맞는 경우에 수정도 가능하다. 연구자가 상기 세 가지 서비스를 이용하고자 하는 경우에 중앙 서버에 자연어 또는 RDQL 형태의 질의로 요청할 수 있으며, 중앙 서버는 시소러스, 온톨로지, 인스턴스를 이용하여 추론하고 그 결과를 연구자에게 시각화 (Visualization)하여 제시한다.

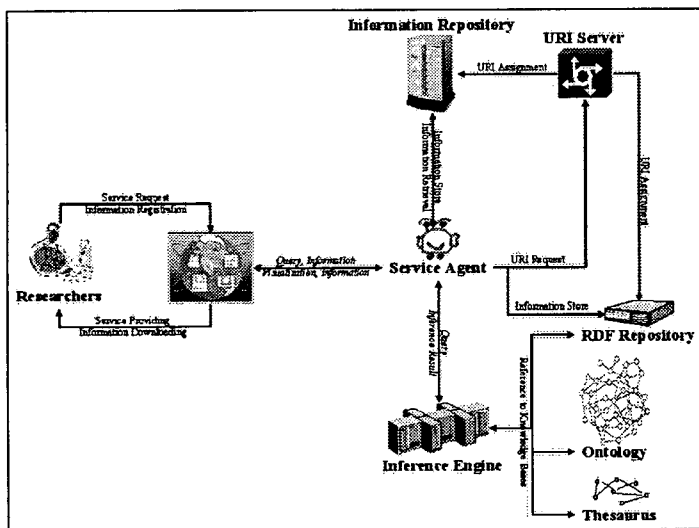


그림 1. 연구자 간 협업 지원 서비스

¹ 특허 출원 번호: 10-2005-0078461

3. 지식 베이스 설계

지식 베이스를 구축함에 있어서 다음 사항들을 고려해야 한다.

(1) 관점 (Viewpoint): 사물이나 분야를 바라보는 관점은 구축자마다 모두 다를 수 밖에 없다. 이는 구축자의 경험, 상식, 레벨을 포함하는 복합적 요소에 의해 좌우된다. 그렇지만, 지식 베이스의 일관성 있는 구축과 설득력 있는 구조를 얻어야 하기 때문에 다양한 관점을 반영할 수 있는 지식 베이스 설계가 이루어져야 한다.

(2) 선택적 제약 (Selectional Restriction): 응용 분야가 단순한 정보 검색이 아닌 질의 응답이나 추론을 포함하는 복잡한 서비스를 요구할수록 지식 베이스 상에서의 확장에 있어 기하급수적인 계산량과 잘못된 확장 가능성을 내포하게 된다. 온톨로지에서 관계를 설계할 때와 시소러스에서 상·하위 관계를 맺을 때 선택적 제약을 반영해야 효율적인 실시간 확장을 보장할 수 있다.

(3) 어휘적 불일치 (Lexical Gap): 사용자는 시스템 내부의 실 데이터 (Instance)가 어떠한 형식과 내용으로 구성되어 있는지 알지 못하기 때문에 사용자 질의에 대한 어휘적 불일치를 해소할 수 있는 방안을 마련해야 한다. 시소러스는 도메인 의존적이기도 하지만, 언어 의존적인 성격이 강하므로 이러한 지식 베이스를 이용하는 것이 효율적인 해법이 될 수 있다.

또한, [4]에서 언급하고 있듯이 지식 베이스 구축을 위한 기반 도구도 동시에 고려하여야 한다.

(a) 구축 인터페이스: 효율적인 지식 베이스 구축을 위해서는 다른 지식 베이스와의 호환, 사용자 편의성을 고려하여야 하며, 기존의 구축 인터페이스를 활용하는 경우에는 설계한 지식 베이스의 특징을 이것이 처리할 수 있는 지를 살펴보아야 한다. 온톨로지를 위한 OWL, 실 데이터/프레임워크를 위한 RDF, 어휘의미망이나 시소러스를 위한 SKOS, 상호 인터페이스를 위한 Web Service 등 다양한 표준이 제공되고 있으므로, 이들에 대한 분석을 통해 재활용이 가능한 지를 먼저 확인해 보아야 한다.

(b) 통계 기반 말뭉치 분석 도구: 타 분야와 달리 과학 기술 분야의 발전 속도는 눈부시며, 불과 몇 년 전에 활발히 사용된 용어들이 이미 사라져버리는 현실이 다반사로 일어난다. 또한, 전문가 집단에 의해 시소러스가 구축되는 경우 특정 분야에 편중하는 경향을 방지하여야 한다. 응용 분야를 고려하여 서비스 대상을 선정하고 서비스 범위를 예측하는 것이 무엇보다 먼저 선행되어야 하고, 이를 지원하는 도구로서 현재 상황을 정확히

분석할 수 있는 통계 기반 말뭉치 분석 도구가 필요하다. 지식 베이스를 구축함에 있어 가장 핵심이 되는 용어 선정은 이러한 맥락에서 이루어져야 한다.

(c) 교육 기반: 대규모 지식 베이스를 구축함에 있어 구축자 간의 협업, 그리고 이를 지원할 수 있는 지침 및 교육이 필요하다. 일관성 있는 지식 베이스는 서비스의 질을 향상시킬 뿐만 아니라 확장성도 보장한다. 지침은 지식 베이스의 구축과 검증 모두에 해당하며, 전체 프로세스를 정의하고 각 작업에 대한 상세한 절차와 주의를 포함하여야 한다. 이러한 지침은 결국 작업자의 자의성을 제어할 수 있는 유일한 방안이 된다.

(d) 데이터베이스 관리 시스템: 지식 베이스 관리 환경으로서의 데이터베이스는 서비스 속도 및 성능에 직접적인 영향을 미친다. 또한, 지식 베이스의 확장에 따른 부하를 충분히 견딜 수 있도록 설계되어야 하며, 지식 베이스 설계 단계에서 미리 예상되는 부하와 사용자의 기대치를 반영하여야 한다.

4. 지식 베이스 구성 요소

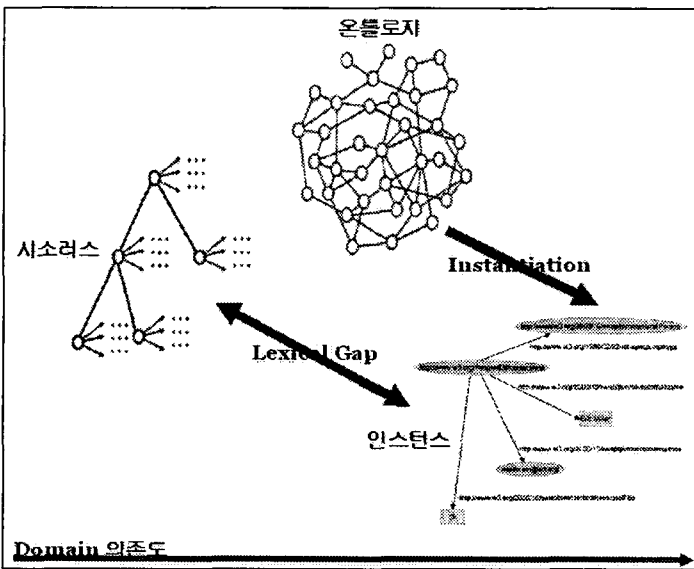


그림 2. 지식 베이스 구성 요소 간의 관계

지식 베이스는 언어, 응용 분야, 서비스 목표에 따라 종류, 규모, 구조가 결정된다. 본 논문에서는 응용 분야로서 연구자 간의 협업 지원 서비스를 목표로 하고 있으므로 이에 맞는 지식 베이스를 선택하고 설계한다. 어휘 불일치를 해소하기 위해 어휘의미망 (Lexical Semantic Network)이나 시소러스 (Thesaurus)를 사용할 수 있는데, 이들의 구성 요소인 용어 집합을 고려한다면 그 해답을 얻을 수 있다. 과학 기술 분야에서의 용어는 전문 용어로서 이들은 명사 위주이며, 복잡한 구조나 선택적 제약의 필요성이 약하다. 즉, 반의어, 항의 관계 등

의 복잡한 정보를 배제할 수 있다. 또한, 하향식 확장 검색이 주된 연산이므로, 트리 형태의 구조를 가지는 것이 바람직하다. 본 논문에서는 이러한 요구 사항에 부합하는 시소러스를 지식 베이스의 한 구성 요소로서 추가한다.

그림 2는 지식 베이스의 구성 요소인 시소러스, 온톨로지, 인스턴스의 관계를 보여준다. 실 데이터는 온톨로지의 사례화 (Instantiation)를 통해 지식 베이스로 변환된다. 시소러스는 사용자 질의와 인스턴스 간의 어휘 불일치를 해소하기 위한 목적으로, 전체 분야 또는 해당 응용 분야에 대해 구축한다.

(1) 시소러스: 질의어 확장을 위한 시소러스의 기본 구조는 상위 개념 (BT; Broader Term)와 하위 개념 (NT; Narrower Term) 간의 연결 관계를 가진다. 본 논문에서는 이러한 상·하위 관계에 대해서 관점을 부여한 관계 패시 (Relational Facet)를 추가하여 확장 시에 선택적 제약으로서 활용한다. 시소러스의 최상위 개념 (TT; Top Term)의 개수는 시소러스의 균형 (Balancing)을 맞추는 측면을 제외하면 중요성이 크지 않다. 인스턴스에 대한 높은 Coverage를 유지하기 위해서는 지식화하는 실 데이터 (Instance)를 충분히 고려해야 한다. 우선어와 비우선어의 구축은 어휘 불일치를 해소하기 위한 정보로서 활용될 수 있다. 실 데이터에서 사용하는 용어와 사용자 질의에서 발견되는 용어가 우선어와 비우선어 관계인 경우에 시소러스를 통해 정규화된다.

(2) 온톨로지: 응용 분야에서의 목표 서비스에 따라 설계가 이루어져야 한다. 온톨로지의 클래스 (Class)들, 그리고 이들이 가지는 관계 (Relation)는 서비스에서 해당 내용이 얼마나 중요하냐에 의존한다 [5]. 특히, 관계에 있어서의 Range를 클래스로 해야 하는지, 단순한 값 (Value)으로 해야 하는지의 문제 역시 응용 분야와 서비스에 의존할 수 밖에 없다. 온톨로지의 구성 요소로는 클래스, 관계, 함수, 공리 (Axiom), 인스턴스가 있다.

(3) 인스턴스는 RDF 기반의 Triple로 표현함으로써 온톨로지의 Relation (Domain, Range) 관계와 Class를 사례화한다.

5. 지식 베이스 구축

[9]는 지식 베이스 구축 (특히, 시소러스)에 있어서 좋은 가이드라인을 제시하고 있다. Nilsson은 Semantic Web 메타데이터 구조가 가져야 하는 조건으로 (a) subjective and non-authoritarian, (b) evolving, (c) extensible, (d) distributed, (e) flexible, (f) conceptual을 언급했다. (a)의 경우, 다양한 관점을 만족시키기 위해 우리는 개념 패시 (Conceptual Facet)과 관계 패시 (Relational Facet)를 도입하며, (b)를 만족시키기 위해 용

어를 생명 주기 측면에서 관리하며, (d)를 만족시키기 위해 철저한 지침에 의한 구축자 간 협업을 가능하도록 한다. (c), (e), (f)는 앞에서 언급한 방안들을 도입함으로써 자연스럽게 해결되는 조건들이다.

(1) 시소러스

본 논문에서는 시소러스를 구성하는 성분인 용어를 추출하기 위하여 용어 생명 주기 기반의 용어 추출 시스템을 사용한다² [2] [3]. 우리는 용어 생명주기 관리를 위해 용어지배값 (TDV; Term Dominance Value)을 정의하여 사용한다. 용어지배값은 성장하고 있거나, 쇠퇴하고 있는 정도를 일정기간 내에서의 관찰을 통해 수치화한 값으로 이 값의 추이에 따라 용어의 현재 단계를 추정할 수 있다. 시작 (start)과 종료 (end) 시점 사이의 기간은 TDV 추이를 일관성 있게 관찰하기 위해 고정한다. 전문용어 관점에서 본다면 그 생명주기가 지식보다 상대적으로 짧을 수 밖에 없으므로, 우리는 한정된 자원을 효율적으로 사용하기 위해 쓰임새의 변화가 심한 용어 특성을 반영하여 용어 생명주기를 재정의하고 이를 이용한 관리방안을 제안한 것이다.

기존에 구축된 시소러스들에서 개념에 대한 정보를 백과사전적인 접근에 의존하여 본말이 전도되는 현상과 가용한 용어 자원이 제약되는 현상이 나타났으나 [1] [7], 우리는 시소러스 상의 각 개념에 대한 정의를 구축자 간 의사소통이 원활히 이루어질 수 있는 수준으로 최소화한다. 즉, 개념을 대표하는 우선어/비우선어에 대한 예제들을 필수적으로 선택하고 범위 주기 (Scope Note)에 대해서는 동형어의어나 다의어적 성격을 가진 개념어들에 대해서만 부여한다.

관련 개념 (RT; Related Term)는 개념들 간 (Domain 과 Range)의 관계를 표현하는 부분이지만 기존의 시소러스들은 Range의 명시에만 치중하고 어떤 관계 (Relation)인지에 대한 정보는 소홀히 하였다 [7]. 우리는 이러한 부분이 일관성을 결여 시킬 수 있는 점이라고 판단하여 관련 개념 구축을 본 시소러스에서는 제외한다.

기존 시소러스들이 많은 비판을 받는 이유 중의 하나는 다양한 관점을 반영하지 못한다는 것이다. 전문가들 뿐만 아니라 해당 시소러스에 관심이 있는 사용자 입장에서 시소러스를 보는 관점은 모두 다를 수 밖에 없으며, 이들의 입장을 반영하고 일관성 있는 구축을 보장하기 위해서 우리는 개념 패싯 (CF; Conceptual Facet)과 관계 패싯 (RF; Relational Facet)을 도입한다. 관계 패싯은 다시 속성 관계 패싯 (ARF; Attribute Relational Facet), 의미역 관계 패싯 (TRF; Thematic Role Relational Facet)과 범주 관계 패싯 (Category Relational

Facet)으로 나누어진다. 개념 패싯은 개념의 대표 의미 속성 분류 체계로서 의미 본질 보다는 활용적 측면을 고려한다. 예를 들어, ‘백신’은 컴퓨터 분야에서 사용되는 프로그램으로서의 의미와 의학에서 약의 통칭으로서의 의미가 있다. 전자의 경우에는 ‘내용’이라는 개념 패싯을, 후자의 경우에는 ‘물질·재료’라는 개념 패싯을 부여한다. 현재 ‘감각·감정,’ ‘위치·공간,’ ‘시간,’ ‘생물’ 등을 포함하여 15개를 정의하여 사용한다. 관계 패싯은 상위 개념 (BT)이 하위 개념 (NT)을 바라보는 관점을 의미한다. 서술형 명사를 중심어로 가지는 “인터넷 접속,” “원격 탐사”와 같은 용어들에는 의미역 관계 패싯이 추가로 부여된다. 상위 개념과 하위 개념 간의 자질 상속을 고려하여 하위 개념에 붙는 변별적 자질을 키워드화하여 이를 속성 키워드 (AK; Attribute Keyword)로 명명한다. 속성 관계 패싯과 의미역 관계 패싯은 속성 키워드를 판단하여 부여된다. 속성 관계 패싯은 개념 관계 패싯 15개에 ‘사례’를 추가하여 사용한다. 의미역 관계 패싯은 Fillmore로부터 발전한 동사의 의미적 수행에 필요한 의미역을 개선하여 ‘근원,’ ‘대상,’ ‘도구,’ ‘목표’ 등 9개를 정의하여 사용한다. 범주 관계 패싯은 일반화, 전체부분, 사례 관계로 나누어 상위 개념에 대한 하위 개념의 유형을 세분화한다. “메모리 → 플래시 메모리”는 일반화 관계, “심장 → 심근”은 전체부분 관계, “운영 체제 → MS Windows”는 사례 관계의 예이다. 그림 3은 동위원소 (isotope)를 상위 개념으로 가지는 부분 트리에서의 개념 패싯과 관계 패싯의 사용 예이다. 기존 시소러스에서는 하위 개념 그룹을 분할하거나 구별할 방법을 가지지 못하여 서로 상이한 개념들이 혼재되었으며, 하위로 내려갈수록 상위 개념들과의 상하 관계가 점점 모호해지는 문제가 발생할 가능성이 커진다.

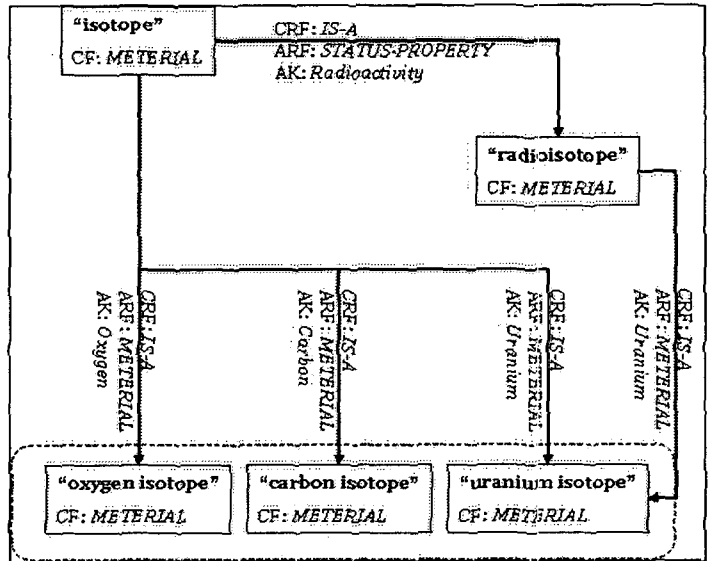


그림 3. 개념 패싯과 관계 패싯의 사용 예

개념 패싯과 관계 패싯은 다음의 측면에서 그 유용함을 확인할 수 있다.

² 특허 출원 번호: 10-2005-0040251

(a) 확장 검색에서의 제약: 모든 확장 검색에서 상·하위 관계만을 이용하여 확장하는 것보다는 사용자 질의에 따라 관계 패시를 제약하여 확장 검색하는 것이 정확을 향상에 도움이 된다.

(b) 동형이의어에 대한 용이한 구분: 동형이의어의 구분을 범위 주기나 정의문 등의 비정형적인 표현에만 의존하는 경우에는 기계적 처리가 힘든 경우가 생기므로, 이를 보완하는 목적으로 개념 패시를 활용할 수 있다.

(c) 정보 검색에서의 효율적 질의 생성: 동일 개념 패시에 대해서는 'OR'를, 상이한 개념 패시에 대해서는 'AND'를 적용하는 방식으로 자연어 질의어의 불리언 질의로의 변환을 가능하게 한다.

(d) 의미적 유사도 반영: 동일한 하위 개념 그룹일지라도 관계 패시에 따라 세부적으로 그룹화될 수 있으며, 이 경우에 그룹 내의 의미적 거리가 그룹 간의 의미적 거리보다 가까워 진다.

(e) 시소러스 확장에서의 일관성 보장: 관계 거리가 멀어지더라도 동일한 관계 패시로 유지되는 Path는 일관성을 유지 측면에서 유리하다.

현재까지 12,000여 개의 용어³들에 대해 시소러스를 구축하였으며, 다음은 그 예를 보여준다.

표 1. 패시를 적용한 상·하위 관계의 예

상·하위 개념	우선어	개념 패시	관계 패시		
			의미역 관계 패시	속성 관계 패시	범주 관계 패시
BT	검사	행위			
NT	LCD 검사	행위	대상	기기·장치	일반화
NT	내시경 검사	행위	도구	기기·장치	일반화
NT	유전자 검사	행위	대상	조직	일반화

(2) 온톨로지

온톨로지 구축을 위해서는 먼저 응용 분야와 서비스 목표 (온톨로지의 범위)를 정해야 한다. 본 연구에서는 국가 과학 기술 R&D 기반 정보를 온톨로지로 표현하여 연구자 간 협업 지원 시스템을 구축한다. 다음으로는 온톨로지에 있어서 중요한 키워드들을 찾아내고 실 데이

³ 본 시소러스에서 사용된 용어들은 범용 과학 기술 분야에 맞도록 최대 11년치 분량의 IT/경제/일반신문들에서 추출하여 TDV에 의해 순위화되고 수작업으로 정제된 것들이다.

터를 분석하는 작업이 필요하다. 예를 들어, 다음과 같은 스키마들을 놓고 이들간의 상관 관계와 온톨로지 범위 내에서 포함/배제시켜야 하는 요소들을 찾아낸다.

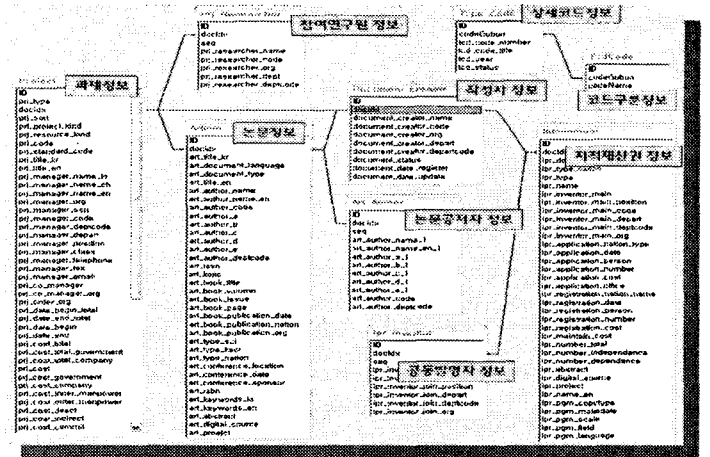


그림 4. 스키마 분석의 예

클래스를 정의하고 클래스들 간의 계층 구조를 설정하는 단계가 필요하다. 본 연구에서는 KA2 [6]과 Hyphen.Info [8] 온톨로지를 참조하여 구축하고 있다. 마지막으로 클래스들의 관계 (Relation or Property)를 정의하고 Range의 값들에 대한 타입, 범위를 정한다. 이때, 고려해야 하는 사항으로는 Range를 클래스화할지 단순한 값으로 처리할 지의 여부이며, 서비스 목표에 따라 달라질 수 있다.

현재 온톨로지에 대한 설계와 구축 지침을 마련한 단계이며, 서비스 목표에 맞추어 연구 보고서, 과학 기술 인력, 과학 기술 학술지, 과학 기술 전문 정보, 학위 논문을 포함하는 국가 과학 기술 R&D 기반 정보에 대해 온톨로지를 구축한다.

(3) 인스턴스

연구자 간 협업 지원을 위한 실 데이터는 크게 사업 정보와 연구 정보로 나누어진다. 사업 정보는 프로젝트에 대한 상세 정보이며, 연구 정보는 논문, 저서, 지적재산권을 포함하는 성과물에 대한 상세 정보이다. 본 연구에서 한국과학기술정보연구원의 연구자들에 대한 프로젝트와 성과물을 기반으로 인스턴스를 추출하여 구축한다.

6. 결론

기존의 지식 베이스 구축에 있어서 간과했던 것들의 하나는 응용 분야와 서비스를 명확히 하고, 해당 분야 맞는 지식 베이스를 선택해야 한다는 것이다. 어휘의 의미를 포함하여 시소러스를 범용 또는 여러 분야에서 구축하고 있으나 이들이 어떤 서비스에 도움이 되는지를 확실히 분석하지 못함으로 인해 응용 시스템에서 제 기

능을 다하지 못하였다. 온톨로지는 근본적으로 분야와 활용 목적에 의존적이어서 이를 고려하여 설계를 하지 만 시소러스나 용어 사전과 같은 기본적인 지식 베이스와의 연계를 통한 효율성 재고라는 측면에서 약점을 보여왔다. 본 연구에서는 사용자 질의와 실 데이터 간의 어휘 불일치 해소를 위해 시소러스를 설계·구축하고 온톨로지의 사례화 결과인 인스턴스와 연동시킴으로써 질의 응답이나 추론과 같은 고급 서비스에 맞는 지식 베이스를 구축할 수 있는 기반을 마련하였다.

앞으로의 과제 중 하나는 현재 정의된 시소러스 관계를 좀더 확장하여 사용자 질의에 대한 Coverage를 높이는 것이다. 예를 들어, 사용자 질의 중 “~이 아닌”과 같은 부분에 대한 처리를 반의어 관계를 추가하는 데 있어서 반의어 관계를 추가하거나, OWL의 Disjoint 제약을 시소러스에 도입할 수 있을 것이다.

또한, 의미역 관계 패킷에 대한 명확한 부여를 위한 방안을 보완할 필요가 있다. 예를 들어, “~게임,” “~메일”과 같은 복합어들에서 중심어에 대해 서술형 명사인지 판단하기가 쉽지 않다. 이러한 단어들은 단순 명사에서 서술형 명사로 쓰임새가 확장되고 있는 과정에 놓인 것으로 파악되기 때문이다. 현재는 ‘행위’ 외의 다른 개념 패킷을 부여 받을 수 있는 개념에 대해서는 의미역 관계 패킷을 할당하지 않는다. 서술형 명사를 중심으로 포함하는 복합어의 경우에도 전문용어 임에도 불구하고 애매성이 상당히 많이 발생한다. 예를 들어, “네트워크 접속”이란 용어는 “네트워크(에) 접속(하다),” “네트워크(에서) 접속(하다),” “네트워크(로) 접속(하다)” 모두 가능할 수 있기 때문에 의미역 관계 패킷들에 대한 문형을 만들어 다중으로 할당하고 있으나 좀더 많은 연구가 필요하다.

참고 문헌

- [1] 정영미, 김명옥, 이재윤, 한승희, 유재복, 과학기술 분야 통합 개념체계의 구축 방안 연구, 정보관리학회지 19 (1), 2002.
- [2] 정한민, 구희관, 이병희, 성원경, 효율적인 자원 운영을 위한 전문용어 생명주기 관리 연구, *Korea Computer Congress*, 2005.
- [3] H. Jung, H. Koo, B. Lee, and W. Sung, Toward Managing the Life Cycle of Terms Using Term Dominance Trend, *Proceedings of Pacific Association of Computational Linguistics*, 2005.
- [4] S. Nirenberg and V. Raskin, *Ontological Semantics*, MIT Press, 2004.
- [5] N. Noy and D. L. McGuinness, *Ontology Development 101: A Guide to Creating Your First Ontology*, SMI Technical Report, SMI-2001-0880, 2001.
- [6] S. Staab, et al., *Semantic Community Web Portals*, *Proceedings of the 9th International World Wide Web Conference*, 2000.

[7] <http://jisik.kiom.re.kr/th/>

[8] <http://www.hyphen.info/ontology/index.html>

[9]

<http://wwwconf.ecs.soton.ac.uk/archive/00000221/01/index.html>